
Dissertation

zur Erlangung des Doktorgrades (Dr. rer. nat.)

angefertigt an der Fakultät für Mathematik, Informatik und Naturwissenschaften im Fachbereich Chemie der Universität Hamburg,
Institut für Pharmazie

Identifizierung von neuen, selektiven Inhibitoren der Dual Leucine Zipper Kinase (DLK)

vorgelegt von

Christina Maria Painer (geb. Brinker)

aus Haselünne

Hamburg 2022

Erstgutachter: Dr. Thomas Lemcke

Zweitgutachter: Prof. Dr. Andrew Torda

Tag der Disputation: 16.12.2022

Die vorliegende Arbeit wurde in der Zeit von April 2016 bis August 2022 am Institut für Pharmazie im Fachbereich Chemie der Universität Hamburg angefertigt. Teile der experimentellen Arbeiten wurden am Institut für Pharmakologie und Toxikologie, Universitätsklinikum Eppendorf, Hamburg, durchgeführt.

Die Arbeit entstand unter der Leitung von

Herrn Dr. Thomas Lemcke,

dem ich an dieser Stelle für die freundliche Aufnahme in seinen Arbeitskreis, für die Überlassung des interessanten Themas, für seine engagierte Betreuung und Unterstützung sowie für zahlreiche wertvolle Anregungen und Diskussionen während der Durchführung dieser Arbeit danken möchte.

Mein besonderer Dank gilt ebenso

Frau Prof. Dr. Elke Oetjen

für die Co-Betreuung, Unterstützung und die wertvollen Gespräche im Rahmen dieser Arbeit.

Danken möchte ich zudem

Herrn Prof. Dr. Andrew Torda

für die Übernahme des Zweitgutachtens.

I. Publikationen und Tagungsbeiträge

- 2021 Duque Escobar, J., Kutschenko, A., Schröder, S., Blume, R., Köster, K.-A., Painer, C., Lemcke, T., Maison, W., Oetjen, E. Regulation of dual leucine zipper kinase activity through its interaction with calcineurin. *Cellular Signalling* 2021, 82, 109953.
- 2019 Painer, C., Mazurek, A., Martens, M., Lemcke, T., Is a selective virtual screening possible? Posterbeitrag DPhG-Jahrestagung 2019, *Pharmaceutical research: From basic research to medical applications*, Heidelberg.
- 2018 Painer, C., Lemcke, T., Structure-Based Virtual Screening for Novel Potential Inhibitors of the Dual Leucine Zipper Kinase (DLK), Posterbeitrag DPhG-Jahrestagung 2018, *Pharmaceutical Science: Structure, Function and Application*, Hamburg.
- 2016 Brinker, C., Lemcke, T., 3D-QSAR-Analysis of Dual Leucine Zipper Kinase Inhibitors, Posterbeitrag DPhG-Jahrestagung 2016, *We live interdisciplinarity*, München.

II. Inhaltsverzeichnis

1	Zusammenfassung	1
2	Abstract	4
3	Einleitung	7
3.1	Diabetes mellitus.....	7
3.2	Proteinkinasen als potenzielle Targets in der Diabetestherapie.....	16
3.2.1	DLK	24
3.3	ATP-kompetitive Inhibitoren der DLK	30
3.4	Wirkstoffsuche.....	36
3.4.1	Computerbasierte Methoden zur Wirkstoffsuche und -entwicklung	37
3.5	Ziel der Arbeit.....	40
4	Material und Methoden	44
4.1	Kraftfelder	44
4.1.1	Das OPLS-Kraftfeld.....	45
4.1.2	Das TIP3P-Wassermodell.....	47
4.2	Vorbereitung von Proteinstrukturen.....	47
4.2.1	Loop Prediction und Refinement.....	50
4.3	Vorbereitung von Liganden	52
4.3.1	VS-Testdatensatz und Decoygenerierung (DUD-E)	57
4.4	Docking	59
4.4.1	Glide.....	63
4.4.2	HYDE	70
4.4.3	Prime MM/GBSA	73
4.4.4	PLP-Score	74
4.4.5	FlexX-Score.....	74
4.4.6	Consensuscore	75
4.4.7	Datenfusionierung und -reduktion.....	76
4.4.8	Validierungsmethoden und -metriken	78
4.5	Molekulardynamik	81
4.5.1	NPT-Ensemble und Langevin-Dynamik	84
4.6	RMSD und RMSF	87
4.7	Clustering	89
4.7.1	Variablen und Proximitätsmaße.....	90

4.7.2	Fusionierungsalgorithmen.....	93
4.8	QSAR	95
4.8.1	3D-QSAR: CoMFA und CoMSIA.....	97
4.8.2	Statistische Auswertung von 3D-QSAR-Modellen	100
4.9	Zusätzlich verwendete Analysesoftware	104
4.9.1	POVME	104
4.9.2	KNIME.....	105
4.9.3	Gnuplot.....	106
4.9.4	R.....	106
4.10	Biologische Prüfungen	107
4.10.1	Semiquantitative Bestimmung der DLK-Menge mit SDS-PAGE.....	109
4.10.2	ADP-Glo™ Kinase Assay.....	110
4.10.3	Microscale Thermophoresis (MST).....	114
4.10.4	KINOMEscan®	118
5	Ergebnisse.....	121
5.1	Zielstrukturen.....	121
5.1.1	Analyse von DLK-Röntgenkristallstrukturen	121
5.1.2	Loop Modelling des Activation Loops mit Prime	131
5.1.3	MD-Simulationen und Auswahl von Konformationen.....	133
5.2	Analyse und Vorbereitung der Liganden	148
5.2.1	Analyse bekannter DLK-Inhibitoren.....	148
5.2.2	Ligandendatensätze	155
5.2.3	Filterprozess der ZINC-Clean-Drug-like-Datenbank	157
5.3	3D-QSAR	163
5.3.1	Alignment durch Docking	163
5.3.2	Test- und Trainingdatensatz und Berechnung der QSAR-Modelle.....	169
5.3.3	Ergebnisse und Auswahl des besten QSAR-Modells	170
5.4	Strukturbasiertes virtuelles Screening.....	179
5.4.1	Vorbereitende Untersuchungen.....	179
5.4.2	ZINC-Screening.....	198
5.5	Postprocessing	204
5.5.1	Einfluss verschiedener Scoringfunktionen auf die Anreicherung und die Korrelation mit Aktivitätswerten	204
5.5.2	Normalisierung der Scorewerte und Kombination ausgewählter Scoringmethoden.....	207

5.5.3	Postprocessing des ZINC-Screenings.....	214
5.5.4	Leader-Clustering mit RDKit-Fingerprints.....	217
5.5.5	Substanzauswahl	219
5.6	Biologische Prüfung.....	227
5.6.1	ADP-Glo™ Kinase Assay.....	227
5.6.2	Microscale Thermophoresis.....	238
5.6.3	KINOMEscan®.....	245
6	Diskussion.....	249
7	Experimenteller Teil	262
7.1	Alignment von Proteinstrukturen und Berechnung von RMSD-Werten.....	262
7.2	Berechnung von Volumina und Oberflächen mit POVME	262
7.3	Proteinvorbereitungen der Röntgenkristallstrukturen mit ProtPrep.....	263
7.4	MD-Simulationen.....	265
7.4.1	Relaxationsprotokoll	266
7.4.2	Clustering der MD-Frames	268
7.5	Glide-Docking.....	268
7.5.1	Grid-Generierung	268
7.5.2	Docking.....	269
7.5.3	Berechnung der RMSD-Werte von Dockingposen	271
7.6	Berechnung der Validierungsmetriken und Diagrammerstellung	271
7.7	Erstellung von Pharmakophormodellen.....	272
7.8	Erstellung von QSAR-Modellen.....	274
7.9	Postprocessing	275
7.9.1	ChemScore, PLP-Score, FLeX-Score	275
7.9.2	HYDE-Score	276
7.9.3	Prime MM/GBSA	276
7.9.4	Rescoring mit dem QSAR-Modell	277
7.9.5	IFP-Tanimotokoeffizient.....	278
7.9.6	Normalisierung der Scorewerte und Kombination von Scorewerten	279
7.9.7	Berechnung der RDKit-Fingerprints.....	280
7.9.8	Leader-Clustering	280
7.10	Ableiten von Bindungsmodi nichtselektiver DLK-Inhibitoren.....	281
7.11	Erfassung und Bearbeitung von Liganden	281
7.11.1	LigPrep.....	282

7.11.2	Bearbeitung der Ligandenlisten vor bzw. nach LigPrep.....	282
7.11.3	Konvertierung von Moleküldateiformaten.....	285
7.12	Berechnung von physikochemischen bzw. molekularen Eigenschaften und Erstellung von Histogrammen	286
7.13	Darstellung von dreidimensionalen Strukturen und Erstellen von Abbildungen und Diagrammen.....	286
7.14	Kauf von Testsubstanzen	286
7.15	Charakterisierung der verwendeten DLK-Lösungen.....	286
7.15.1	Durchführung der SDS-PAGE.....	287
7.16	Durchführung des ADP-Glo™ Kinase Assays	289
7.17	Vorbereitung und Durchführung der MST (Microscale Thermophoresis)	293
7.18	Hochleistungs-Flüssigchromatografie (HPLC)	296
8	Literaturverzeichnis.....	298
	Anhang.....	341
A.	Übersicht über die verwendeten Knime-Nodes und deren Funktion	341
B.	Zusammensetzung der Ligandenliste c	342
C.	QSAR-Selektionen.....	342
D.	Statistische Parameter weiterer QSAR-Modelle.....	356
E.	Proteinkonformationen für Docking- und VSW-Experimente	357
F.	Ergebnisse des OPLS2005-A-1-, OPLS3-B-1- und OPLS3A-2-Workflows.....	358
G.	Ergebnisse der Bewertung des ZINC-Actives-Screenings mit den Einzelscorewerten.....	366
H.	Referenz- und Testsubstanzlösungen.....	367
I.	Chromatogramme der HPLC-Analytik	368
J.	Gefahrstoffverzeichnis.....	376
	Danksagung	377
	Eidesstattliche Versicherung.....	378

III. Abkürzungsverzeichnis

2D bzw. 3D	- zwei- bzw. dreidimensional
4-PL	- 4-Parameter-Logarithmische Funktion
5-PL	- 5-Parameter-Logarithmische Funktion
AA	- <i>all atom</i> (dt. alle Atome)
Abb.	- Abbildung
ADP	- Adenosindiphosphat
Ala	- Alanin
ALS	- amyotrophische Lateralsklerose
AMP	- Adenosinmonophosphat
AMPK	- Adenosinmonophosphat-aktivierte Proteinkinase
ANP	- Adenylylimidodiphosphat
ANOVA	- <i>Analysis of variance</i> (dt. Varianzanalyse)
APS	- Ammoniumperoxodisulfat
ASL	- <i>Atom Specification Language</i>
Asn	- Asparagin
Asp	- Asparaginsäure
ATP	- Adenosintriphosphat
BSA	- bovines Serumalbumin
bzw.	- beziehungsweise
CaCl ₂	- Calciumchlorid
cAMP	- cyclisches Adenosinmonophosphat
C _α	- Alpha-Kohlenstoffatom
ca.	- circa
CoMFA	- <i>Comparative molecular field analysis</i>
CoMSIA	- <i>Comparative molecular similarity indices analysis</i>
(c)log P	- (<i>calculated</i> = dt. berechneter) logarithmierter Octanol-Wasser-Verteilungskoeffizient
CPU	- <i>Central processing unit</i> (dt. Zentraleinheit)
CREB	- <i>cAMP response element binding protein</i>
CRTC	- <i>cAMP regulated transcriptional coactivators</i>
Cys	- Cystein
DFG	- Asparaginsäure-Phenylalanin-Glycin
d.h.	- das heißt
dH ₂ O	- destilliertes Wasser
DLK	- Dual Leucine Zipper Kinase
DMSO	- Dimethylsulfoxid
DNA	- <i>Deoxyribonucleic acid</i> (dt. Desoxyribonukleinsäure)
dt.	- deutsch
DTT	- 1,4-Dithiothreitol
DUD(-E)	- <i>Directory of Useful Decoys(, Enhanced)</i>
EDTA	- Ethylendiamintetraacetat
EF	- <i>Enrichment Factor</i> (dt. Anreicherungsfaktor)
et al.	- <i>et alii</i> (dt. und andere)
f.	- folgende
FDA	- <i>Food and Drug Administration</i>
ff.	- fortfolgende
ggf.	- gegebenenfalls

GIP	- glucoseabhängiges insulinotropes Polypeptid
Glide	- <i>Grid-based Ligand Docking with Energetics</i>
Gln	- Glutamin
GLP-1	- Glucagon-like Peptid 1
Glu	- Glutaminsäure
Gly	- Glycin
GSK3	- Glycogensynthase-Kinase 3
GUI	- <i>Graphical User Interface</i> (dt. graphische Benutzeroberfläche)
HBA	- Wasserstoffbrückenakzeptoren
HbA _{1c}	- Hämoglobin A _{1c} (= Glykohämoglobin)
HBD	- Wasserstoffbrückendonoren
HTS	- <i>High Throughput Screening</i> (dt. Hochdurchsatzscreening)
HTvS	- <i>High Throughput virtual Screening</i> (dt. virtuelles Hochdurchsatzscreening)
HPLC	- <i>High Performance Liquid Chromatography</i> (dt. Hochleistungsflüssigkeitschromatographie)
HYDE	- <i>Hydrogen Bond and Dehydration</i>
IC ₅₀	- mittlere inhibitorische Konzentration bei halbmaximaler Hemmung
ID	- Identifikationsnummer
i.d.R.	- in der Regel
IFP	- <i>Structural interaction fingerprints</i>
IL-1(β)	- Interleukin-1(β)
Ile	- Isoleucin
Inh.	- Inhibitoren
IR-Laser	- Infrarot-Laser
IRS-1	- Insulinrezeptorsubstrat 1
JIP1	- <i>JNK-interacting protein 1</i>
JNK	- c-Jun-N-terminale Kinase
Kap.	- Kapitel
K _d	- Dissoziationskonstante
K _i	- inhibitorische Konstante
Konz.	- Konzentration
LBVS	- ligandbasiertes virtuelles Screening
LD	- <i>Loading dye</i>
LE	- <i>Glide Ligand Efficiency</i>
LED	- <i>Light-emitting diode</i> (dt. lichtemittierende Diode)
Leu	- Leucin
Lig.	- Ligandenliste
Lys	- Lysin
MafA	- <i>V-maf musculoaponeurotic fibrosarcoma oncogene homologue A</i>
MAPK	- Mitogenaktivierte Protein-Kinase
MAP2K	- Mitogenaktivierte Protein-Kinase-Kinase
MAP3K	- Mitogenaktivierte Protein-Kinase-Kinase-Kinase
MAP4K	- Mitogenaktivierte Protein-Kinase-Kinase-Kinase-Kinase
Max. bzw. max.	- Maximum bzw. maximal
MBP	- <i>Maltose binding protein</i> (dt. Maltose-Bindeprotein)
MD	- <i>Molekular Dynamics</i> (dt. Molekulardynamik)
Met	- Methionin
Min. bzw. min.	- Minimum bzw. minimal

mind.	- mindestens
MgCl ₂	- Magnesiumchlorid
MnCl ₂	- Manganchlorid
MLK	- Mixed-Lineage Kinasen
MM	- Molekülmechanik
MM/GBSA	- <i>Molecular Mechanics – Generalized Born Surface Area</i>
MST	- <i>Microscale Thermophoresis</i>
MW	- <i>Molecular Weight</i> (dt. Molekulargewicht)
NaCl	- Natriumchlorid
NHS	- N-Hydroxysuccinimid
NPT	- Teilchenzahl (N), Druck (P), Temperatur (T) konstant
Nr.	- Nummer
NVT	- Teilchenzahl (N), Volumen (V), Temperatur (T) konstant
o.g.	- oben genannte(n/r)
OPLS	- <i>Optimized Potential for Liquid Simulations</i>
PAINS	- <i>Pan Assay Interference Compounds</i>
PBS	- <i>Phosphate Buffered Saline</i> (dt. Phosphat-gepufferte Salzlösung)
PDB	- <i>Protein Data Bank</i> bzw. Datei(format) aus dieser Datenbank
Phe	- Phenylalanin
PKB bzw. AKT	- Proteinkinase B
pK _d	- negativer dekadischer Logarithmus der Dissoziationskonstante K _d
pK _i	- negativer dekadischer Logarithmus der inhibitorischen Konstante K _i
PK-Modell	- Pharmakophormodell
P-Loop	- phosphatbindender Loop
PLP	- <i>Piecewise Linear Potential</i>
PLS	- <i>Partial Least Squares</i> (dt. partielle kleinste Quadrate)
PMSF	- Phenylmethylsulfonylfluorid
PPAR-α bzw. -γ	- Peroxisomen-Proliferator-aktivierter Rezeptor α bzw. γ
Pro	- Prolin
PrSA	- <i>Protein Structure Alignment</i>
PSA	- <i>Polar Surface Area</i> (dt. Polare Oberfläche)
QM	- Quantenmechanik
QSAR	- <i>Quantitative Structure-Activity Relationship</i> (dt. Quantitative Struktur-Wirkungs-Beziehung)
RB	- rotierbare Bindungen
RLU	- <i>Relative Light Units</i>
RMSD	- <i>Root mean square deviation</i> (dt. mittlere quadratische Abweichung)
RMSE	- <i>Root mean square error</i> (dt. Wurzel des mittleren quadratischen Fehlers)
RMSF	- <i>Root mean square fluctuation</i> (dt. mittlere quadratische Fluktuation)
s.	- siehe
S.	- Seite
SBVS	- strukturbasiertes virtuelles Screening
SD	- <i>Standard Deviation</i> (dt. Standardabweichung)
SDS	- <i>sodium dodecyl sulfate</i> (dt. Natriumdodecylsulfat)
SDS-PAGE	- <i>sodium dodecylsulfate polyacrylamide gel electrophoresis</i> (dt. Natriumdodecylsulfat-Polyacrylamidgelelektrophorese)
Ser	- Serin
s.o.	- siehe oben

sog.	- sogenannte(n/r)
SP	- <i>Standard Precision</i>
Tab.	- Tabelle
TEMED	- N,N,N',N'-Tetramethylethyldiamin
TEV	- Tobacco-Etch-Virus
Thr	- Threonin
TIP3P	- <i>Transferable intermolecular potential 3 point</i>
TNF- α	- Tumornekrosefaktor α
(TR-)FRET	- <i>(Time-resolved) Fluorescence resonance energy transfer</i> (dt. zeitaufgelöster Fluoreszenzenergietransfer)
Tris-HCl	- Tris-(hydroxymethyl)-aminomethanhydrochlorid
Trp	- Tryptophan
Tyr	- Tyrosin
u.	- und
Val	- Valin
VEGFR-2	- <i>Vascular endothelial growth factor receptor 2</i>
vgl.	- vergleiche
\bar{x}	- Mittelwert
XP	- <i>Extra Precision</i>
u.a.	- unter anderem
vdW	- van-der-Waals
VK	- Variationskoeffizient
vs.	- versus
VS	- Virtuelles Screening
VSGB	- <i>Variable dielectric surface generalized Born model</i>
VSW	- Virtual-Screening-Workflow
WBB	- Wasserstoffbrückenbindung(en)
z.B.	- zum Beispiel
ZNS	- Zentralnervensystem

1 Zusammenfassung¹

In dieser Arbeit wurde ein strukturbasiertes virtuelles Screening zur Identifizierung von neuen ATP-kompetitiven Inhibitoren des potenziell diabetesrelevanten Targets DLK unter Anwendung des molekularen Dockings durchgeführt (s. Übersicht in Abb. 1-1).

Vier Röntgenkristallstrukturen der DLK-Kinasedomäne – eine Apo-Struktur und drei Strukturen mit gebundenem Liganden in der ATP-Bindetasche – wurden als Ausgangspunkt für das strukturbasierte Design genutzt. Der Bindungsmodus gebundener Liganden wurde analysiert. In den Bindetaschen wurden vier konservierte Wassermoleküle identifiziert, die entweder Kontakte zum Protein ausbilden oder an der Ligandeninteraktion beteiligt sind. MD-Simulationen zur Generierung von variablen Proteinkonformationen für das nachfolgende Docking wurden durchgeführt.

Als Ausgangspunkt für die Substanzbibliothek wurde der *Clean-Druglike*-Datensatz der ZINC-Datenbank mit 5.142.887 Substanzen verwendet. Eigene Filterkriterien (z.B. #HBA, #HBD, #Heteroatome, #RB, #Ringe) und die Generierung von möglichen Protonierungszuständen und Tautomeren führte zu 3.801.953 *Single States*, entsprechend 5.941.050 Strukturen (*Multiple States*).

Verschiedene Settings des GLIDE-Dockings (Kraftfelder, Constraints) der SCHRÖDINGER-Software wurden anhand des Re- und Crossdockings bekannter DLK-Inhibitoren sowie des Screenings von Testdatensätzen aus Aktiven und Decoys unter Verwendung verschiedener DLK-Proteinkonformationen, die sich in ihrer Konformation bzw. der Anwesenheit konservierter Wassermoleküle unterscheiden, vor dem eigentlichen ZINC-Screening untersucht. Ein Ensemble von sechs Proteinkonformationen mit besonders guten Ergebnissen hinsichtlich Dockingqualität (Richtigkeit

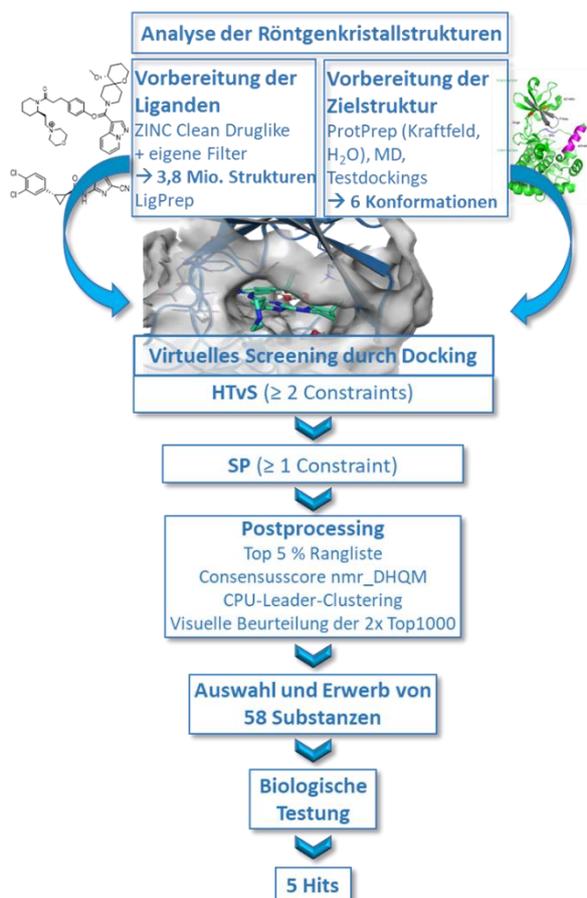


Abb. 1-1 – Virtual-Screening-Workflow

¹ Die Bezifferung weicht vom Rest der Arbeit ab.

der Pose), Anreicherung von Aktiven sowie Korrelation zwischen Scorewert und experimentellem Aktivitätswert wurden für das ZINC-Screening ausgewählt.

Die vorbereitete ZINC-Datenbank sowie ein Testdatensatz aus Aktiven und Decoys wurden in zwei aufeinander aufbauenden Schritten in die sechs DLK-Proteinkonformationen gedockt: zuerst ein HTvS unter der Bedingung, mindestens zwei der drei Wasserstoffbrückenbindungen zur Hinge-Region der DLK auszubilden. In diesem Schritt wurden insgesamt 811.489 ZINC-Strukturen (21 %), bzw. 10 % der Decoys und 80 % der Aktiven (Testdatensatz) gedockt. Nach erneuter Ligandenpräparation wurde im zweiten Schritt ein SP-Docking (GLIDE) in das Sechspeiteinkonformations-Ensemble durchgeführt (mind. 1 Constraint). Die Zahl der insgesamt gedockten Liganden blieb nahezu gleich, allerdings verbesserte sich die Anreicherung der Aktiven und die Korrelation zwischen Scorewert und experimenteller Aktivität im Vergleich zum HTvS deutlich. Die besten 5 % der nach Dockingscore sortierten Gesamtangliste (123.630 *Single States*) dieses Dockings wurden weiterverwendet, da hier der Anteil an Aktiven besonders hoch war.

Um die Anreicherung der aktiven Substanzen zu verbessern wurden verschiedene Scoringfunktionen an den Dockingposen eines Testdatensatzes aus Aktiven und Decoys hinsichtlich Anreicherung und Score-Aktivität-Korrelation getestet. Zusätzlich wurde ein QSAR-Modell auf Basis bekannter DLK-Inhibitoren entwickelt, welches zur Vorhersage von Aktivitätsdaten als zusätzliche Bewertungsmethode eingesetzt wurde. Es wurde ein Consensuscore **nmr_DHQM** entwickelt, welcher den Dockingscore (D), den HYDE-Score (H), die vorhergesagte Aktivität mit dem QSAR-Modell (Q) sowie die freie Bindungsenergie mit der PRIME MM/GBSA-Methode (M) kombiniert, der eine besonders hohe frühe Anreicherung von Aktiven ermöglicht. Nach Sortierung der Rangliste nach dem Consensuscore wurde ein Leader-Clustering der RDKit-Fingerprints dieser Strukturen durchgeführt, um ähnliche Moleküle zu gruppieren und nur die Clusterrepräsentanten mit dem besten Consensuscore weiter zu berücksichtigen. Auf diese Weise wurden zwei unabhängige Clusterings, mit und ohne Begrenzung des Molekulargewichts, durchgeführt. Die Interaktionen der jeweils 1.000 besten ZINC-Strukturen in der Bindetasche wurden visuell beurteilt. Es wurden 58 Substanzen ausgewählt und erworben.

Diese 58 Substanzen wurden in einem ADP-GloTM Kinase Assay auf ihre DLK-hemmende bzw. mit der *Microscale Thermophoresis* (MST) auf ihre DLK-bindende Fähigkeit unter Verwendung des murinen DLK-Fusionsproteins MBP-TEV-DLK₁₄₀₋₅₁₇ untersucht. 36 der 58 Verbindungen wurden zusätzlich mit der KINOMEScan[®]-Methode bei der Firma Eurofins DiscoverX Corporation in San Diego an der humanen DLK getestet. Insgesamt gingen fünf Verbindungen als potenzielle Hits

mit K_d -Werten zwischen 2,5 und 24 μM (**CP01**, **CP02**, **CP03** und **CP05** im KINOMEscan®) bzw. 52,8 μM (**CP04** im MST-Assay) aus den Untersuchungen hervor² (s. Abb. 1-2). Ob die Substanzen die DLK auf ATP-kompetitive Weise hemmen kann abschließend erst nach weiteren Untersuchungen, idealerweise auch durch kristallographische Experimente an Cokristallen, beurteilt werden.

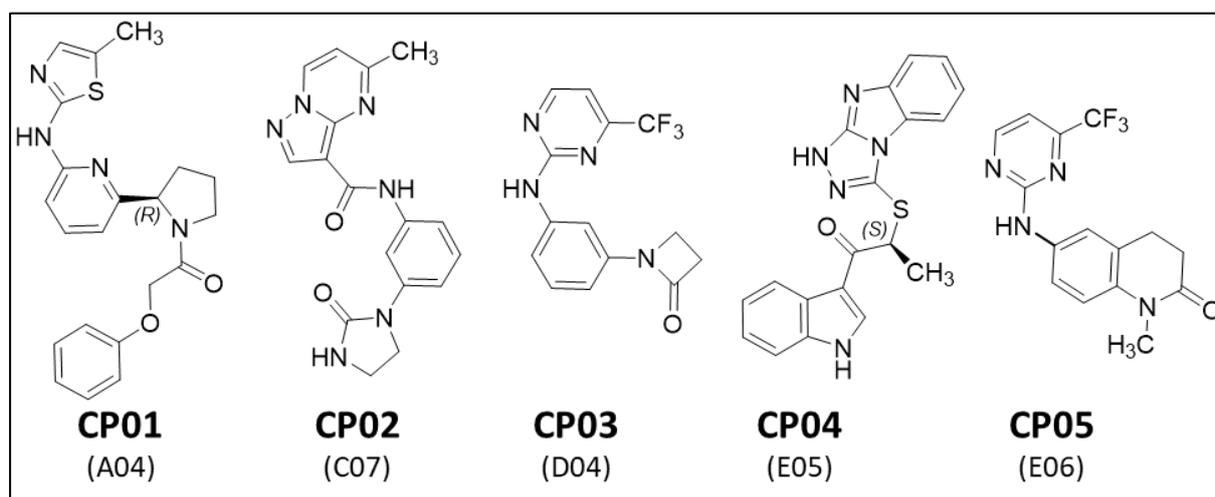


Abb. 1-2 – Hit-Strukturen des strukturbasierten virtuellen Screenings

² **CP01** und **CP04** gingen aus dem virtuellen Screening als Enantiomer hervor, die biologische Aktivitätsmessung erfolgte am Racemat.

2 Abstract³

In this work, a structure-based virtual screening was conducted to identify novel ATP-competitive inhibitors of the potentially diabetes-relevant target DLK using molecular docking (Fig. 2-1).

Four X-ray crystal structures of the DLK kinase domain – one apo structure and three structures with a bound ligand in ATP-binding pocket – were used as starting points for the structure-based design. The binding mode of the bound ligands were analysed. Four conserved water molecules were identified in the binding pockets which either form contacts to protein or to ligand. MD simulations were carried out to generate variable protein conformations for the following docking experiments.

The *clean druglike* dataset of the ZINC database with 5,142,887 substances was used as the starting point for the substance library. Own filter criteria (e.g. #HBA, #HBD, #hetero atoms, #RB, #rings) and the generation of possible protonation states and tautomers

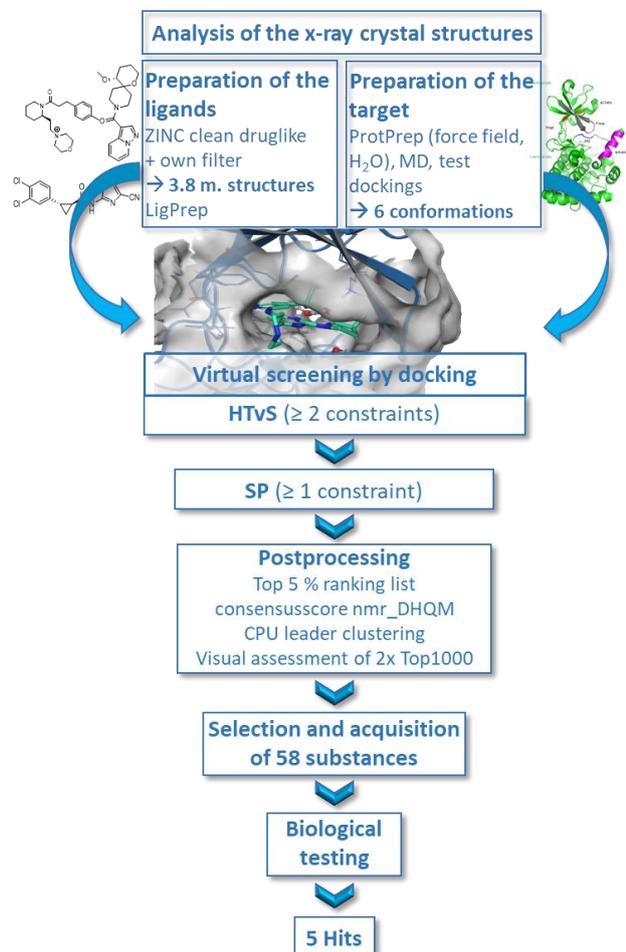


Fig. 2-1 – Virtual screening workflow

led to 3,801,953 single states, corresponding to 5,941,050 multiple states.

Before executing the actual ZINC screening, different settings of the GLIDE docking (force fields, constraints) of the SCHRÖDINGER *software* were investigated with the help of re-docking and cross-docking experiments of known DLK inhibitors as well as with screening experiments of actives-and-decoys-testsets using different DLK protein conformations, which differ in conformation or the presence of conserved water molecules. An ensemble of six protein conformations with particularly good results in terms of docking quality (accuracy of pose), enrichment of actives as well as correlation between score and experimental activity value were chosen for the ZINC screening.

³ The numbering differs from the rest of the work.

The prepared ZINC database and a test data set of actives and decoys were docked into the six DLK protein conformations in two successive steps: first an HTvS under the condition that at least two of the three hydrogen bonds to the hinge region of the DLK were formed. A total of 811,489 ZINC structures (21 %), or 10 % of the decoys and 80 % of the actives (test data set) were docked in this step. In the second step, after renewed ligand preparation, SP docking (GLIDE) was performed in the six-protein conformation ensemble (at least 1 constraint). The total number of ligands docked remained almost the same, but the enrichment of the actives and the correlation between score and experimental activity value improved significantly compared to HTvS. The best 5 % of the total ranking list sorted by docking score (123,630 single states) of this docking was further used, as the proportion of actives was particularly high here.

To improve the enrichment of actives, different scoring functions were tested on the docking poses of a test data set of actives and decoys regarding enrichment and score-activity correlation. Additionally, a QSAR model based on known DLK inhibitors was developed, which was used as an additional assessment method to predict activity data. A consensus score **nmr_DHQM** was developed, which combines the dockingscore (D), HYDE score (H), the predicted activity value with QSAR model (Q) as well as the free binding energy with PRIME MM/GBSA method (M). The use of the consensus score enables a particularly high early enrichment of actives. After sorting the ranking list by consensus score, a leader clustering of the RDKit fingerprints of these structures was carried out in order to group similar molecules and to further consider only the cluster representatives with the best consensus score. Two independent clusterings were performed in this way, with and without limiting the molecular weight. The interactions of the 1,000 best ZINC structures in the binding pocket were assessed manually. 58 substances were selected and acquired.

These 58 substances were tested in an ADP-Glo™ kinase assay for their DLK inhibiting capability and with *Microscale Thermophoresis* (MST) to their DLK binding capability using the murine DLK fusion protein MBP-TEV-DLK₁₄₀₋₅₁₇. Additionally, 36 of these 58 compounds were tested with the KINOMEscan® method by the company Eurofins DiscoverX Corporation in San Diego on human DLK. In total, five compounds emerged as potential hits with K_d values between 2.5 and 24 μM (**CP01**, **CP02**, **CP03**, **CP05** in the KINOMEscan®) and 52.8 μM (**CP04** in the MST assay), respectively⁴ (Fig. 2-2). If the substances inhibit the DLK in an ATP-competitive way can only be

⁴ **CP01** and **CP04** emerged from virtual screening as enantiomers, the activity measurements were carried out with the racemates.

finally assessed after further investigations, ideally also by crystallographic experiments on cocrystals.

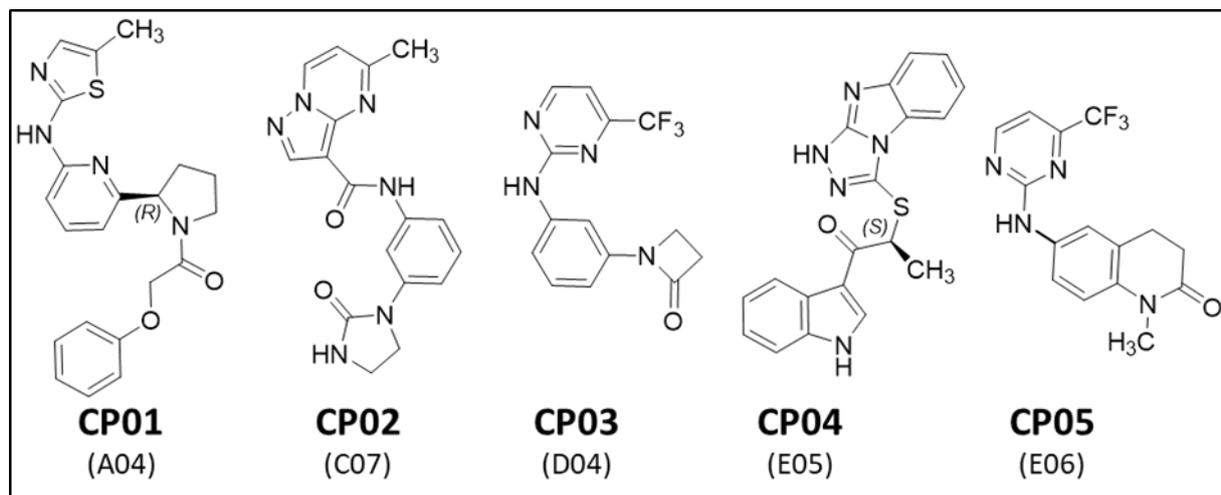


Fig. 2-2 – Hit structures of the structure-based virtual screening.

3 Einleitung

3.1 Diabetes mellitus

Die weltweite Ausbreitung von Diabetes mellitus wird mittlerweile als „eskalierende Diabetes-Pandemie“ [1] bezeichnet, da nicht mehr nur die Bevölkerung in den Industriestaaten und nicht mehr vorwiegend nur die Älteren betroffen sind. Die Gesundheitsausgaben für Diabetes weltweit beliefen sich im Jahr 2013 auf 548 Milliarden US-Dollar [1]. Die Zahl der an Diabetes erkrankten Menschen weltweit im Jahr 2019 wird auf 463 Millionen (Prävalenz von 9,3 %) geschätzt und für die Jahre 2030 und 2045 wird ein weiterer Anstieg der Erkrankten um 25 % (2030) bzw. 51 % (2045) prognostiziert [2]. Diabetes-Erkrankte weisen eine zwei- bis vierfach erhöhte Sterblichkeit auf [3]. Weltweit starben im Jahr 2021 schätzungsweise 6,7 Millionen Menschen an Diabetes und dessen Folgeerkrankungen [4]. Myokardinfarkt, Schlaganfall und die koronare Herzkrankheit (Makroangiopathien) sowie Nerven-, Nieren- und Netzhautschäden (Mikroangiopathien) können als Folge der erhöhten Blutglucosespiegel auftreten. Eine schwerwiegende Spätkomplikation als Folge der Gefäß- und Nervenschäden ist das diabetische Fußsyndrom, welches eine Amputation zur Folge haben kann [5]. Diabetiker haben darüber hinaus eine erhöhte Infektanfälligkeit [6] und ein erhöhtes Risiko, im späteren Verlauf an Krebs zu erkranken [7].

Unter Diabetes mellitus werden Stoffwechselstörungen zusammengefasst, die von einer erhöhten Blutzuckerkonzentration (Hyperglykämie) geprägt sind. Die Ursache liegt immer in einem Funktionsverlust der β -Zellen in den Langerhans'schen Inseln der Bauchspeicheldrüse. [8]

Die zwei Hauptformen des Diabetes sind der insulinabhängige Diabetes mellitus Typ 1, welcher vorrangig auf einer autoimmunvermittelten Zerstörung der β -Zellen in der frühen Kindheit und einem damit einhergehenden absoluten Insulinmangel beruht und der Diabetes mellitus Typ 2, welcher von einem relativen Insulinmangel aufgrund einer reduzierten Insulinsensitivität der Muskel- und Fettzellen (Insulinresistenz) geprägt ist. Mit 90 – 95 % der Diabetesfälle stellt dieser Typ die häufigste Form dar. Weitere Diabetestypen sind spezifische Formen (z.B. durch monogenetische Defekte, die als *Maturity Onset Diabetes of the Young* (MODY) zusammengefasst werden, sowie durch Infektionen oder chemisch induzierte Formen), erstmals in der Schwangerschaft diagnostizierter Diabetes (Gestationsdiabetes) sowie unklassifizierter Diabetes. [8]

Zur Diagnosestellung gehört die Feststellung erhöhter Blutglucosespiegel (Nüchtern-Plasmaglukose $\geq 7,0$ mmol/l) bzw. eines erhöhten HbA_{1c}-Werts ($\geq 6,5$ %). Symptomatisch zeigt sich die Hyperglykämie durch vermehrtes Wasserlassen (Polyurie) und dadurch bedingt vermehrtem Durst

(Polydipsie). Insbesondere bei Diabetes mellitus Typ 1 sind diese Symptome besonders akut und es kommt zu einem starken Gewichtsverlust. Diabetes mellitus Typ 2 hingegen bleibt oft über mehrere Monate oder Jahre unerkannt, sodass sich in dieser Zeit bereits Schäden an den Gefäßen und Nerven manifestieren können. Häufig weisen Typ-2-Diabetiker weitere Krankheitsbilder wie Hypertonie, Hyperlipidämie und Adipositas auf, die bei gleichzeitigem Auftreten mit Diabetes als metabolisches Syndrom zusammengefasst werden. Dieses stellt ein eminentes Risiko für die Entstehung von Herz-Kreislauf-Erkrankungen dar. [9]

Pathogenese des Diabetes mellitus Typ 2

Die Entstehung von Diabetes mellitus Typ 2 ist ein multifaktorielles Geschehen. Sie wird begünstigt durch den westlichen Lebensstil, welcher von Überernährung und wenig körperlicher Bewegung geprägt ist. Übergewicht bzw. Fettleibigkeit, insbesondere ein erhöhter Körperfettanteil in der Bauchregion, sind typische Merkmale eines Typ-2-Diabetikers. Oftmals besteht eine genetische Prädisposition. [8]

Eine Übersicht über die Entwicklung von Diabetes mellitus Typ 2 gibt Abb. 3-1. Bei Nahrungsüberschuss und geringem Energieverbrauch gelangen die Fett-, Muskel- und Leberzellen an die Grenze ihrer Nährstoffspeicherkapazität. Dies führt zu erhöhtem zellulären Stress und einer Aktivierung der c-Jun-N-terminalen Kinasen (JNKs) und der I κ B-Kinase β (IKK- β). Diese Kinasen führen durch eine hemmende Serin-Threonin-Phosphorylierung des Insulinrezeptorsubstrats 1 (IRS-1) zu einer Abschwächung des Insulinsignalwegs in den Fett- und Muskelzellen. Die Zellen werden unempfindlicher gegenüber Insulin. Die verminderte Insulinsensitivität führt zu einer verstärkten Lipolyse in den Fettzellen, d.h. zu einer verstärkten Freisetzung freier Fettsäuren. Darüber hinaus kommt es durch den oxidativen zellulären Stress und die JNK-Aktivierung zur vermehrten Produktion von proinflammatorischen Zytokinen (z.B. TNF- α = Tumornekrosefaktor α) im Fettgewebe. Es kommt somit zu einer subakuten Entzündung des Gewebes. TNF- α wiederum steigert die Lipolyse. Die freien Fettsäuren und die Zytokine steigern die Insulinresistenz durch Hemmung der Insulin-Signaltransduktion. [10, 11]

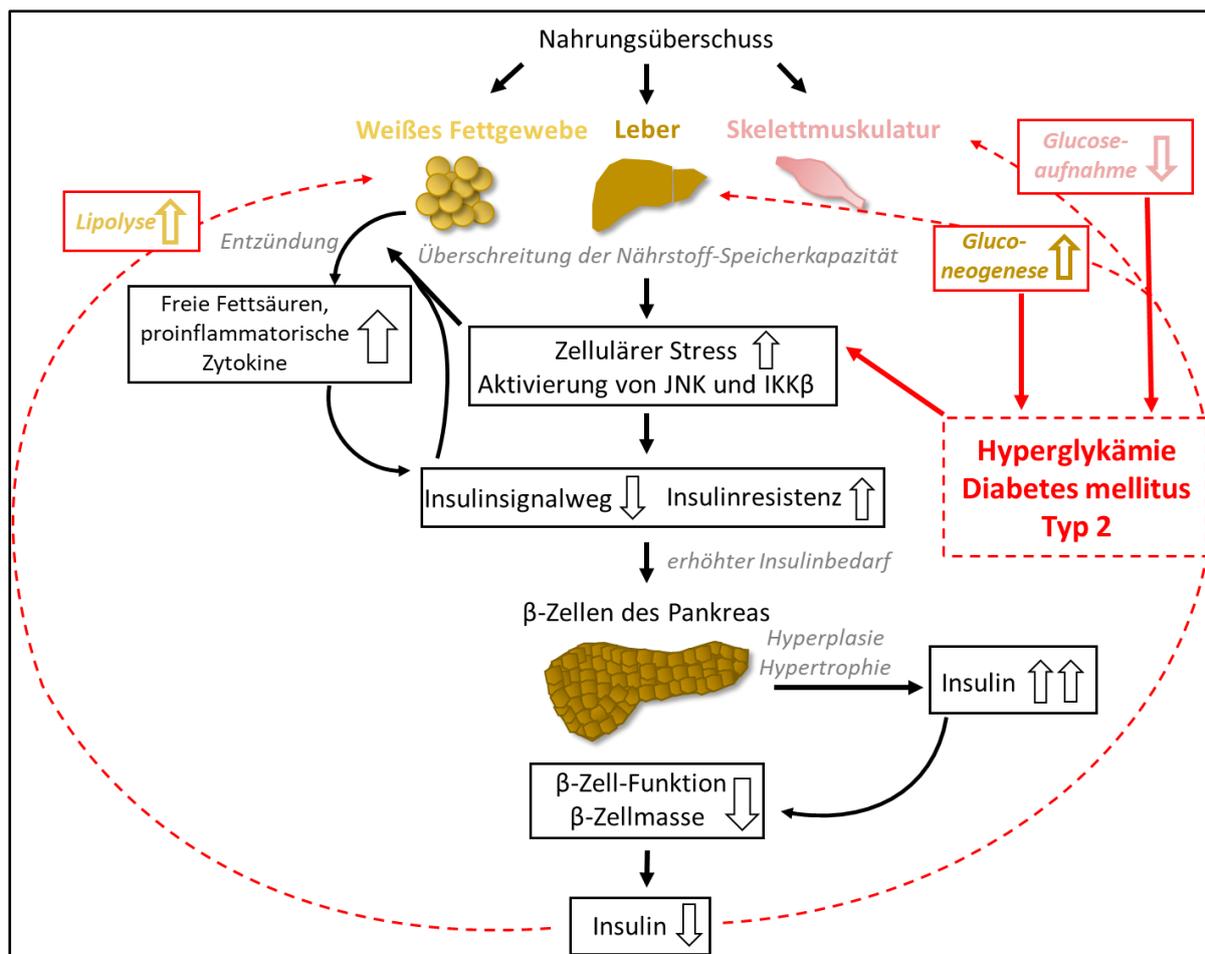


Abb. 3-1 – Entstehung von Diabetes mellitus Typ 2. Die erhöhte Nährstoffzufuhr führt zu einer Überschreitung der Speicherkapazität in den Fett-, Leber und Muskelzellen, sodass zellulärer Stress entsteht und Proteinkinasen aktiviert werden, welche wiederum den Insulinsignalweg in den Fett-, Leber- und Muskelzellen hemmen. Zellulärer Stress, JNK-Aktivierung und die Insulinresistenz führen zu einer Entzündung im Gewebe. Aus Fettzellen werden vermehrt freie Fettsäuren und Zytokine freigesetzt, welche zum Voranschreiten der Insulinresistenz beitragen. Dadurch entsteht ein erhöhter Insulinbedarf, der zunächst noch durch eine gesteigerte Insulinproduktion in den β -Zellen der Langerhans'schen Inseln des Pankreas ausgeglichen werden kann. Im späteren Verlauf kann diese starke Insulinproduktion durch Funktionsverlust der β -Zellen und Abnahme der β -Zellmasse nicht mehr aufrechterhalten werden. Die Hyperglykämie selbst verstärkt den zellulären Stress. Abb. modifiziert nach KAISER UND OETJEN (2015) [10] und STUMVOLL ET AL. (2005) [11]

Als Kompensationsmechanismus zu der verminderten Insulinsensitivität findet zunächst eine erhöhte Insulinsekretion statt (Prädiabetes). Diese kann allerdings im späteren Verlauf durch die weiter zunehmende Insulinresistenz aufgrund einer zunehmenden β -Zelldysfunktion nicht mehr weiter gesteigert werden kann [12]. Die abnehmende Insulinsekretion aus den β -Zellen der Bauchspeicheldrüse sowie die verminderte Insulinwirkung an den Zielzellen führen zu einer verringerten Glucoseverwertung in den Muskelzellen und zu einer gesteigerten glucagoninduzierten Gluconeogenese in der Leber, sodass sich eine Hyperglykämie manifestieren kann. Die Hyperglykämie verstärkt den zellulären Stress und die Entzündung. Die Lipolyse in den Adipozyten wird durch den relativen Insulinmangel weiter gesteigert [10, 11]. Eine optimale antidiabetische Therapie sollte diesen sich selbst verstärkenden Kreislauf durchbrechen, indem zum einen die

Entzündung im stoffwechselaktiven Gewebe gehemmt und zum anderen die β -Zellfunktion und -masse aufrechterhalten wird [10].

Therapie des Diabetes mellitus Typ 2

Eine gesunde Ernährung und Gewichtsreduktion durch gezielte Diät sowie die Steigerung der körperlichen Betätigung sind die wichtigste Grundlage für die Diabetes-Typ-2-Behandlung. Wenn diese Lebensstiländerungen allein nicht greifen wird der Einsatz von Antidiabetika erforderlich, um Akutkomplikationen und Langzeitschäden an den Gefäßen und Organen zu verhindern. [13] In der Regel wird ein HbA_{1c} -Wert von $< 6,5\%$ angestrebt [14].

Eine Übersicht über die verschiedenen Arzneistoffklassen (Insulin ausgenommen) zur Behandlung des Diabetes mellitus Typ 2 sowie deren pharmakologische Angriffspunkte liefert Tab. 3-1. Die Arzneistoffe zielen über unterschiedliche physiologische Ansätze auf eine Verminderung der Blutglucosekonzentration ab:

Die älteste Wirkstoffgruppe der Sulfonylharnstoffe und die Gruppe der Glinide steigern die Insulinsekretion aus den β -Zellen durch Schließung spannungsabhängiger Kaliumkanäle [13]. Der HbA_{1c} -Wert wird effektiv gesenkt, allerdings tritt häufig eine Gewichtszunahme auf und es besteht insbesondere bei den Sulfonylharnstoffen ein erhöhtes Hypoglykämierisiko, welches in seltenen, schweren Fällen zu Koma, Krampfanfällen oder kardiovaskulären Ereignissen führen kann [15]. Da β -Zellfunktion und -masse im Laufe der Erkrankung immer weiter abnehmen, kann die Therapie mit diesen Wirkstoffen mit der Zeit versagen [13].

Das Biguanid Metformin hemmt die Gluconeogenese in der Leber. Der genaue Mechanismus ist weitgehend unbekannt. Eine Stimulation der Adenosinmonophosphat-aktivierten Protein-kinase (AMPK) scheint hierbei eine wichtige Rolle zu spielen. Die Effektivität hinsichtlich der HbA_{1c} -Senkung ist mit der der Sulfonylharnstoffe zu vergleichen. Häufig treten gastrointestinale Beschwerden unter der Therapie mit Metformin auf. In sehr seltenen Fällen kann eine lebensbedrohliche Laktatazidose entstehen, was den Einsatz u.a. bei niereninsuffizienten Patienten einschränkt. [13]

α -Glucosidasehemmer hemmen die Kohlenhydratspaltung im Dünndarm, sodass es zu einer geringeren postprandialen Glucoseresorption kommt. Der Einfluss auf den HbA_{1c} -Wert ist mäßig. Die häufig auftretenden gastrointestinalen Beschwerden schränken den Einsatz dieser Wirkstoffgruppe erheblich ein. [13]

Tab. 3-1 – Übersicht über die eingesetzten Antidiabetika in der Diabetes Typ 2-Therapie (Insulin nicht mit aufgeführt). SUR-1: Sulfonylhurea-Rezeptor 1. AMPK: AMP-aktivierte Proteinkinase. PPAR- γ : Peroxisomen-Proliferator-aktivierter Rezeptor γ . GLP-1: Glucagon-like-Peptide 1. DPP4: Dipeptidylpeptidase 4. GIP: glucoseabhängiges insulinotropes Polypeptid. SGLT-2: Sodium glucose linked transporter 2. Die Informationen stammen aus Veröffentlichungen von AHREN [16], FONSECA ET AL. [17], KAHN ET AL. [18], PHILIPPE UND RACCAH [15], SCHMITZ ET AL. [19], STUMVOLL ET AL. [11], VIRALLY ET AL. [13] und TSOUTSOUKI ET AL. [20].

Arzneistoffgruppe	Pharmakologisches Target	Effekt	Verwendung seit	Vertreter (Bsp.)	Nebenwirkungen
Sulfonylharnstoffe (oral)	SUR-1 \rightarrow Schließung spannungsabhängiger Kaliumkanäle	Insulinsekretion \uparrow	1946	Glibenclamid, Glimepirid	(schwere) Hypoglykämien, Gewichtszunahme
Biguanide (oral)	Nicht bekannt; u.a. Stimulierung der AMPK	Gluconeogenese in der Leber \downarrow , Glucoseverwertung im peripheren Gewebe \uparrow	1957	Metformin als einziger Vertreter	gastrointestinale Beschwerden, Laktatazidose (sehr selten)
α-Glucosidasehemmer (oral)	Hemmung der α -Glucosidase	Kohlenhydratspaltung im Dünndarm \downarrow \rightarrow postprandiale Glucoseresorption \downarrow	1995	Acarbose, Miglitol	gastrointestinale Beschwerden
Thiazolidine (oral)	Agonist am PPAR- γ	Veränderung des Fettstoffwechsels: Konzentration freier Fettsäuren u. Zytokine \downarrow \rightarrow Insulinsensitivität \uparrow	1997	Pioglitazon, Rosiglitazon	Ödeme, Frakturen, Gewichtszunahme, kardiovaskuläre Ereignisse (Myokardinfarkt, Herzinsuffizienz)
Glinide (oral)	SUR-1 und Kir 6.2 \rightarrow Schließung spannungsabhängiger Kaliumkanäle	Insulinsekretion \uparrow	1997	Nateglinid, Repaglinid	Gewichtszunahme, (geringes) Hypoglykämierisiko
GLP-1-Rezeptor-Agonisten (parenteral)	Aktivierung der GLP-1-Rezeptoren	glucoseabhängige Insulinsekretion \uparrow , β -Zellmasse \uparrow , Glucagonsekretion \downarrow , Magenentleerung \downarrow , Sättigungsgefühl \uparrow , Körpergewicht \downarrow	2005	Exenatid, Liraglutid	Übelkeit, Erbrechen, Assoziation mit Pankreatitis
Amylin-Analoga (parenteral)	Amylinrezeptor	Glucagonsekretion \downarrow , Magenentleerung \downarrow	2005 (nicht in Dtl.)	Pramlintid	gastrointestinale Beschwerden (z.B. Übelkeit)
DPP-4-Inhibitoren (oral)	Hemmung der DPP-4 und somit Verhinderung des Abbaus der Inkretine GLP-1 und GIP	glucoseabhängige Insulinsekretion \uparrow , Glucagonsekretion \downarrow	2006	Sitagliptin, Vildagliptin	Atemwegsinfektionen, Pankreatitis
Gallensäurebindende Resine (oral)	Bindung v. Gallensäuren	Blutglucose- und -lipidkonzentration \downarrow ; blutzuckersenkender Mechanismus ist unklar	2008	Colesevelam	Gastrointestinale Beschwerden (Obstipation, Dyspepsie)
SGLT-2-Inhibitoren (oral)	SGLT-2-Hemmung (natriumabhängiger Glucosetransporter in der Niere)	Glucoseausscheidung über die Niere \uparrow \rightarrow Blutglucosekonzentration \downarrow , Körpergewichts \downarrow	2012	Canagliflozin, Dapagliflozin	Harnwegsinfektionen, genitale Pilzinfektionen, euglykämische diab. Ketoazidose, Fournier-Gangrän (selten), Amputations- und Frakturrisiko

Thiazolidine (Glitazone) agonisieren den Peroxisomen-Proliferator-aktivierten Rezeptor γ (PPAR- γ) in Fett-, Muskel- und Leberzellen. Die Freisetzung freier Fettsäuren wird reduziert und die Insulinsensitivität der Fett-, Leber- und Muskelzellen erhöht [13]. In Deutschland ist nur noch das Pioglitazon zugelassen, dessen Einsatz jedoch vom Bundesinstitut für Arzneimittel und Medizinprodukte (BfArM) aufgrund eines erhöhten Blasenkarzinomrisikos nicht empfohlen wird [21]. Glitazone erhöhen das Körpergewicht durch Wassereinlagerung, das Risiko einer Herzinsuffizienz sowie das postmenopausale Frakturrisiko [13].

Die subkutan zu verabreichenden Glucagon-like Peptid 1 (GLP-1)-Analoge stimulieren die Insulinsekretion glucoseabhängig, hemmen die Glucagonsekretion und erhöhen die β -Zellmasse durch Agonismus am GLP-1-Rezeptor. Sie hemmen die Magenentleerung, verstärken das Sättigungsgefühl und sind gewichtsreduzierend. Zu den Nebenwirkungen gehören Übelkeit und Erbrechen. Es besteht eine Assoziation mit dem Auftreten einer Pankreatitis. Dipeptidylpeptidase-4 (DPP-4)-Inhibitoren (Gliptine) hemmen den Abbau des körpereigenen Inkretins GLP-1 und stimulieren daher ebenso die glucoseabhängige Insulinsekretion und reduzieren die Glucagonsekretion. Anders als GLP-1-Analoge wirken sie sich nicht auf das Körpergewicht aus [15]. Als Nebenwirkung werden Atemwegsinfektionen [15] und Pankreatiden [20] beschrieben.

Die neueste Wirkstoffklasse bilden die Natrium-Glucose-Transporter-2 (SGLT-2)-Inhibitoren, welche die Glucoseresorption in der Niere hemmen und so die Glucoseausscheidung über die Niere fördern. Dies führt zu einer Blutglucosesenkung und zu einer Gewichtsabnahme. Zu den Nebenwirkungen gehören Harnwegsinfektionen, Pilzinfektionen im Genitalbereich und eine euglykämische diabetische Ketoazidose. Darüber hinaus gibt es Assoziationen mit einem erhöhten Risiko für Amputationen kleiner Gliedmaßen und einem erhöhten Frakturrisiko. [22]

Weniger häufig eingesetzt werden die Amylin-Analoge, welche die Glucagonsekretion und die Magenentleerung hemmen. Der einzige Vertreter Pramlintid wird aktuell in den USA ergänzend zur Insulintherapie eingesetzt [15]. Darüber hinaus gibt es die gallensäurebindenden Resine, deren blutzuckersenkender Wirkmechanismus unbekannt ist [18]. Bei beiden Wirkstoffgruppen stehen gastrointestinale Beschwerden im Vordergrund [15, 17].

Bei den aktuell in der Diabetes-Typ-2-Therapie eingesetzten Wirkstoffen bestehen demnach vorrangig Nebenwirkungen wie Hypoglykämien, Gewichtszunahme und gastrointestinale Beschwerden. Das bereits in den 1950er Jahren entwickelte Metformin wird aufgrund seiner effektiven HbA_{1c}-senkenden Wirkung und seines positiven Sicherheitsprofils noch immer vorrangig eingesetzt. Der Einsatz ist kosteneffizient [20]. Metformin hat allerdings keinen Einfluss auf die

β -Zelldysfunktion. Die Ursache von Diabetes wird nicht behoben [13]. Abgesehen von den GLP-1-Rezeptor-adressierenden Wirkstoffen zielt keiner der vorhandenen Wirkstoffe auf den Schutz der β -Zellen ab. Zudem ist die langfristige Blutzuckerkontrolle mit den vorhandenen Wirkstoffen oftmals unzureichend. Zwar führt eine Kombination dieser Wirkstoffe zu einer besseren Blutzuckerkontrolle als die Erhöhung der Dosis eines einzelnen Wirkstoffs. Dennoch schaffen es viele Diabetespatienten (etwa 30-60 % der betroffenen Diabetiker in der westlichen Bevölkerung) nicht langfristig, ihren Blutzuckerwert im Sollbereich zu halten, sodass sie ein erhöhtes Risiko für Spätschäden haben [23]. All diese Aspekte inklusive der stetig zunehmenden Diabetesprävalenz in der Bevölkerung machen deutlich, dass eine Erforschung neuer Antidiabetika dringend erforderlich ist, welche optimalerweise die Insulinsensitivität der Zielzellen im Fett-, Muskel- und Lebergewebe steigern und die β -Zellfunktion in der Bauchspeicheldrüse verbessern.

Aktuelle Forschung

Aktuell gibt es weltweit mehr als 7.000 Studien, die sich mit Diabetes mellitus Typ 2 befassen [24]. PERREAULT ET AL. untersuchten diese Studien im Hinblick auf Wirkstoffe mit neuen Wirkmechanismen und fanden dabei 43 Wirkstoffe, die sich in klinischer Entwicklung befinden. Die Wirkstoffgruppen werden in Tab. 3-2 nach ihren pharmakologischen Targets und Effekten aufgeführt. Sie lassen sich grob in vier Klassen hinsichtlich ihrer Effekte einteilen: [24]

1. *Stimulation der Insulinsekretion aus den β -Zellen*: Glucokinase-Aktivatoren (GKAs) und G-Protein-gekoppelte Rezeptor-40 (GPR-40)-Agonisten (s. S. 15, Tab. 3-2, grün) bewirken eine gesteigerte Insulinsekretion unabhängig von der Glucosekonzentration im Plasma. [24]

2. *Nachahmung der Inkretineffekte*: Anders als die Wirkstoffgruppen in 1.) löst diese Klasse nur bei erhöhter Plasmaglukosekonzentration eine Insulinsekretion aus (s. S. 15, Tab. 3-2, orange). Sie führen zu einem erhöhten cAMP-Spiegel in den β -Zellen, was schlussendlich zu einer erhöhten Pro-Insulin-Biosynthese und einer gesteigerten Insulinfreisetzung aus den β -Zellen führt. Aus dieser Gruppe sind die GLP-1-Rezeptor-Agonisten bereits seit 15 Jahren als Arzneimittel in der Diabetestherapie etabliert (s. Tab. 3-1, S. 11). Neu ist die Entwicklung von Substanzen, die gleichzeitig auch als GIP (glucoseabhängiges insulinotropes Polypeptid)-Rezeptor-Agonisten bzw. als Glukagon-Rezeptor-Agonisten wirken. Es befinden sich darüber hinaus dreifache GLP1-GIP-Glukagon-Rezeptor-Agonisten in der Entwicklung. [24]

3. Reduktion der hepatischen Glucosesynthese bzw. Steigerung der hepatischen Glucoseaufnahme: Die Wirkstoffe aus dieser Gruppe wirken alle über die Reduktion der Glucagon-Wirkung (s. S. 15, Tab. 3-2, blau). Da Glucagon die Gluconeogenese in der Leber fördert, bedeutet wenig zirkulierendes Glucagon eine geringere hepatische Gluconeogenese. [24]

4. Steigerung der Insulinsensitivität: In dieser Klasse befindet sich eine Vielzahl von unterschiedlichen Targets, durch deren Beeinflussung die Insulinsensitivität an Muskel-, Fett- bzw. Leberzellen verbessert wird (s. Tab. 3-2, gelb). Dies geschieht beispielsweise durch den Eingriff in den Fettstoffwechsel (z.B. durch PPAR- α/γ -Agonisten) oder durch Hemmung der geringgradigen Entzündung (z.B. durch Nuklearfaktor- κ B (NF κ B)-Inhibitoren), welche im Fettgewebe des Diabetikers vorherrscht (vgl. Abb. 3-1, erhöhte Freisetzung von Zytokinen im Diabetesgeschehen). [24]

Zusätzlich sind in Tab. 3-2 mögliche Probleme aufgeführt, die sich aus der Wirkstoffanwendung ergeben können, wobei zu beachten ist, dass die Wirkstoffe sich noch in der Entwicklung befinden (z.B. Dosisfindungsstudien) und die therapeutische Dosis zum Teil noch unbekannt ist.

Tab. 3-2 – Wirkstoffgruppen für die Diabetes-Typ-2-Therapie in der aktuellen Forschung. Grün: Stimulation der Insulinsekretion; orange: Nachahmung der Inkretineffekte; blau: Reduktion der hepatischen Glucosesynthese bzw. Steigerung der hepatischen Glucoseaufnahme; gelb: Steigerung der Insulinsensitivität. Apo-B = Apolipoprotein B, Appl. = Applikation, cAMP = cyclisches Adenosinmonophosphat, Glc = Glucose, GLUT4 = Glucosetransporter Typ 4, IKK β = I κ B-Kinase β , IP3 = Inositoltriphosphat. IRS-1 = Insulinrezeptorsubstrat 1, (V)LDL = (Very low density lipoprotein. Informationen aus PERREAULT ET AL. (2021) [24] und den in der Tabelle angegebenen Quellen.

Wirkstoffgruppe	Pharmakologischer Effekt	Mögliche Probleme
Glukokinase-Aktivatoren (GKAs)	Umwandlung \uparrow von Glc in Glc-6-Phosphat \rightarrow glc-stimulierte Insulinsekretion \uparrow	Hypoglykämien [25] Wirkungsverlust wg. β -Zellverlust (?)
G-Protein-gekoppelte Rezeptor 40 (GPR-40)-Agonisten	Phospholipase C (PLC) $\uparrow \rightarrow$ IP3 $\uparrow \rightarrow$ Ca ²⁺ _(iZR) $\uparrow \rightarrow$ Insulinsekretion \uparrow	Obstipation, Übelkeit [26, 27], hämatologische Veränderungen [26], Schmerzen, Diarrhoe [27]
Glucagon-like Peptid 1 (GLP-1)-glucoseabhängige insulinotrope Polypeptid (GIP)-Rezeptor-Glucagonrezeptor-Triple-Rezeptoragonisten duale GLP1-GIP-Rezeptor-Agonisten duale GLP1-Glucagon-Rezeptor-Agonisten	Stimulation einer der Rezeptoren \rightarrow cAMP \uparrow \rightarrow a) Proteinkinase A (PKA) $\uparrow \rightarrow$ Pro-Insulin-Biosynthese \uparrow \rightarrow b) EPAC2 \rightarrow RAP1A (= <i>Ras-related protein Rap-1A</i>) \rightarrow Insulin-Exozytose \uparrow	Übelkeit, Erbrechen, Diarrhoe [28, 29, 30], Assoziation mit Pankreatitis (GLP-1-Rezeptoragonisten, s. Tab. 3-1)
G-Protein-gekoppelte Rezeptor 119 (GPCR-119)-Agonisten		
Oxyntomodulin-Analoga	Stimulation des GLP-1- u. Glucagonrezeptors (s. dort)	
Glucagonrezeptor-Antagonisten (Leberzelle)	Glucagon-Wirkung $\downarrow \rightarrow$ cAMP $\downarrow \rightarrow$ Proteinkinase A (PKA) $\downarrow \rightarrow$ hepatische Glc-Produktion \downarrow (Glycogenolyse, Gluconeogenese \downarrow)	gastrointestinale Beschwerden, Kopfschmerzen [31]
Antisense-Oligonukleotid-Inhibitoren für Glucagonrezeptor-mRNA		parenterale Appl., Anstieg hepat. Transaminasen [32]
Duale Amylin-Calcitonin-Rezeptor-Agonisten (DACRAs)	Amylin-/Insulin-Sekretion \uparrow aus β -Zellen \rightarrow Glucagonsekretion aus α -Zellen \downarrow	gastrointestinale Beschwerden, Übelkeit, Erbrechen [33]
Hepatosensitive Glukokinaseaktivatoren (GKAs)	hepatische Glucoseaufnahme \uparrow (Umwandlung von Glc in Glc-6-Phosphat)	Triglyceride \uparrow [25]
Antisense-Oligonukleotid-Inhibitor für Protein-Tyrosin-Phosphatase 1B (PTP1B) mRNA	PTP1B $\downarrow \rightarrow$ hemmende Dephosphorylierung des Insulinrezeptors und IRS-1 $\downarrow \rightarrow$ Insulinsignalweg \uparrow	parenterale Appl.; nur für Patienten, die auf Insulintherapie nicht mehr ansprechen (Add-on) [34]
(Partiell) Duale Peroxisom-Proliferator-aktivierter Rezeptor α/γ (PPAR- α/γ)-Agonisten Zulassung Chitagalazar 10/2021 in China [36]	PPAR- α : Lipoprotein-C-III-Expression $\downarrow \rightarrow$ VLDL-Produktion $\downarrow \rightarrow$ Triglyceride \downarrow PPAR- γ : \rightarrow Differenzierung von Prä-Adipozyten zu reifen Adipozyten \uparrow , subkutane Fettdepots $\uparrow \rightarrow$ Insulinsensitivität \uparrow	Hyperlipidämie, Hypoglykämie, Gewichtszunahme, Ödeme, Frakturen, Anämien, kardiovaskuläre Ereignisse [35,36]
selektiver allosterischer Aktivator der LYN-Kinase	a) IRS-1-Tyrosinphosphorylierung $\uparrow \rightarrow$ Insulinsignalweg \uparrow b) GLUT-4-Translokation $\uparrow \rightarrow$ Glc-Aufnahme \uparrow	Kopfschmerzen, Somnolenz [37]
Stimulator des GLUT-4 (Glucose-transporter Typ 4)	GLUT-4-Translokation $\uparrow \rightarrow$ Glc-Aufnahme \uparrow	unveröffentlicht
Analoga des Fibroblasten-Wachstumsfaktors 21 (FGF21)	Hepatische Lipogenese \downarrow , braunes Fettgewebe \uparrow , Insulinsensitivität und -sekretion \uparrow	parenterale Appl., Überempfindlichkeit, Leberenzyme \uparrow [38]
Inhibitoren der Diacylglycerin-Acetyl-Transferase 1 (DGAT1)	Acetylierung von DAG (Di-) zu TAG (Triacylglycerin) \downarrow (TAG-Akkumulation \sim Insulinresistenz)	gastrointestinale Beschwerden, Kopfschmerzen [39]
Inhibitoren des enterozytären mikrosomalen Triglycerid-Transfer-Proteins (MTP)	Triglyceridübertragung auf Apo-B-haltige Lipoproteine \downarrow (LDL \downarrow)	gastrointestinale Beschwerden, Transaminase \uparrow , Fettleber [40]
NF- κ B (Nuklearfaktor- κ B)-Inhibitoren	IKK β /NF- κ B-Weg $\downarrow \rightarrow$ IRS-1-Serinphosphorylierung $\downarrow \rightarrow$ Insulinsignalweg \uparrow	gastrointestinale Beschwerden [41]
nikotinischer α 7-Rezeptor-Ligand	proinflammatorische Zytokinproduktion \downarrow	gastrointestinale Beschwerden [42]
Ghrelin-Analoga		parenterale Appl., Diarrhoe, Hämatome [43]
Ghrelin/Growth-Hormone-(GH)-Rezeptor-Coagonisten	Stimulation der IRS-1-Produktion $\uparrow \rightarrow$ Insulinsignalweg \uparrow	parenterale Appl., Arthralgie, Kopfschmerzen, Myalgie, periphere Ödeme, Übelkeit, Diarrhoe, Müdigkeit, Atemwegsinfektionen [44, 45]
Aktivin II-B-Rezeptormodulator	Wirkmechanismus unbekannt; Muskelwachstum \uparrow , Fettabbau	Lipase \uparrow , Pankreatitis, Lungenentz., Durchfall, Muskelkrämpfe, Leberenzyme \uparrow [46]
Imeglimin	Oxidative Phosphorylierung \downarrow , Stimulation der Mitochondrienfunktion	gastrointestinale Beschwerden, Kopfschmerzen [47]

3.2 Proteinkinasen als potenzielle Targets in der Diabetestherapie

Eine Targetklasse, welche noch nicht von den aktuell in der Praxis eingesetzten Antidiabetika direkt angesprochen wird und schon seit längerer Zeit erforscht wird, ist die Gruppe der Proteinkinasen. Die Proteinphosphorylierung wurde erstmals im Jahr 1955 von FISCHER UND KREBS [48] sowie von SUTHERLAND UND WOSILAIT [49] am Beispiel der Phosphorylasekinase im Glycogenstoffwechsel entdeckt. Mitte der 1980er Jahre wurde die Bedeutung der Tyrosinkinaseaktivität des Insulinrezeptors für die Insulinwirkung nachgewiesen [50]. Ein weiteres Beispiel ist die Aufdeckung der Beteiligung der JNK an der durch oxidativen Stress reduzierten Insulingenexpression [51] sowie der Nachweis der hemmenden Serinphosphorylierung des IRS-1 durch JNK und die damit einhergehende Induktion einer Insulinresistenz [52, 53].

Funktion von Proteinkinasen

Proteinkinasen spielen eine zentrale Rolle in der intrazellulären Signalweiterleitung. Prozesse wie die Zellproliferation und -differenzierung, das Wachstum der Zelle sowie die Apoptose werden durch das Zusammenspiel von Proteinkinasen wesentlich gesteuert [54–56]. Ihre Funktionsweise besteht in der Übertragung der γ -Phosphatgruppe von ATP auf eine hydroxylhaltige Aminosäure eines Proteins oder Peptids [57, 58]. Man unterscheidet hauptsächlich zwischen Serin/Threonin-Kinasen und Tyrosin-Kinasen, je nach Aminosäure, die phosphoryliert wird [59]. Die Phosphorylierung führt durch die Ausbildung eines Wasserstoffbrückennetzwerks zwischen dem Phosphat und in der Nähe befindlichen Aminosäuren des Proteins zu einer Konformationsänderung und so zu einer Veränderung der Aktivität des Proteins, da aktive Bindestellen freigelegt oder verschlossen werden [60]. Dieser Vorgang ist reversibel: Phosphatasen katalysieren die entgegengesetzte Reaktion, nämlich die Phosphatabspaltung von dem Protein [61]. Proteinkinasen wie auch Phosphatasen sind häufig integriert in Signaltransduktionskaskaden – so werden beispielsweise Signale, die mittels Rezeptoren an der Zelloberfläche durch Bindung extrazellulärer Botenstoffe empfangen werden, mithilfe hintereinandergeschalteter Proteinkinasen in den Nucleus weitergeleitet [62]; auch eine Quervernetzung zwischen den Signalwegen sowie regulatorische Feedbackmechanismen zwischen einer untergeordneten und einer übergeordneten Kinase finden statt [63]. Eine Störung dieser Signalketten, z.B. durch eine genetisch bedingte Überaktivität einer Proteinkinase, kann zur Manifestation verschiedener Erkrankungen wie Tumoren, neurodegenerativen Erkrankungen oder Diabetes führen [64–67].

Klassifizierung von Proteinkinasen

MANNING ET AL. stellten nach Analyse von cDNA (complementary DNA)- und EST (Expressed sequence tag)-Sequenzen fest, dass 518 Gene des menschlichen Genoms für Proteinkinasen codieren [68]. Nach ihrer Sequenzähnlichkeit in der katalytischen Domäne lassen sich die eukaryotische Proteinkinasen in verschiedene Gruppen einteilen: **TK** (= Tyrosinkinase, z.B. ABL, Januskinasen, Insulinrezeptor), **AGC** (z.B. Proteinkinase A und B), **CAMK** (= Ca²⁺/Calmoculin-abhängige Kinasen, z.B. Myosin-leichte-Ketten-Kinase (MLCK), MAP-Kinase aktivierte Proteinkinasen (MAPKAPKs)), **CMGC** (z.B. cyclinabhängige Kinasen (CDKs), mitogenaktivierte Proteinkinasen (MAPK), Glycogensynthase-Kinasen (GSK)), **STE** (Serin-Threonin-Proteinkinasen, z.B. diverse MAP2- und MAP3-Kinasen), **TKL** (= Tyrosinkinase-like, z.B. Mixed-Lineage-Kinasen (MLK), Raf-Kinasen), **CK1** (= Casein-Kinase 1) und **atypische Proteinkinasen** [68, 69].

Aufbau von Proteinkinasen

Im Aufbau der Tertiärstruktur der katalytischen Einheit ähneln sich die Proteinkinasen sehr stark (s. Abb. 3-2): Sie besteht aus einer großen C-terminalen Domäne aus α -Helices und aus einer kleineren N-terminalen Domäne, bestehend aus β -Faltblättern und einer α C-Helix. Diese beiden Domänen sind durch die *Hinge*-Region („Scharnier“) verbunden. In diesem Bereich findet die ATP-Bindung statt. Der *P-Loop* („Phosphatbindender Loop“, auch als *Glycine rich loop* bezeichnet) in der C-

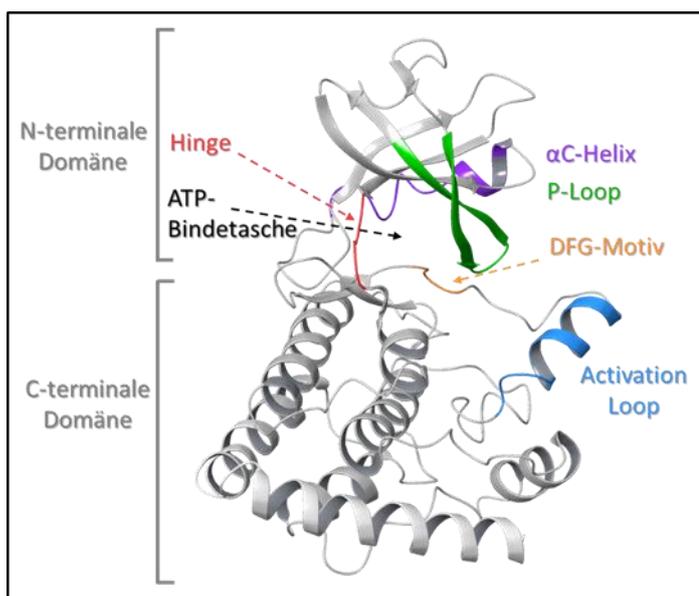


Abb. 3-2 – Aufbau der katalytischen Einheit von Proteinkinasen am Beispiel der DLK (PDB: 5CEN, DLK-Apoenzym aus PATEL ET AL. (2015) [70])

terminalen Domäne stellt die Begrenzung der ATP-Bindetasche gegenüberliegend von der Hinge-Region dar. Er besteht aus der Aminosäureabfolge G-X-G-X-y-G (y = Tyr oder Phe). Durch den hohen Glycingehalt kann sich der P-Loop gut an die Phosphatgruppen des ATP anpassen und ist ebenso anpassungsfähig gegenüber potenziellen Inhibitoren. Die Seitenkette eines Lysins in der ATP-Bindetasche (N-terminale Domäne) interagiert mit der α - und β -Phosphatgruppe des ATP-Moleküls. Diese Lysinseitenkette geht im aktiven Zustand der Kinase auch eine ionische

Wechselwirkung mit einem stark konservierten Glutamat der α C-Helix ein. Die α C-Helix steht ebenso mit dem *Activation Loop* (dt. Aktivierungsschleife) in der Nähe des DFG-Motivs in Kontakt. Dies ist ein konserviertes Motiv aus den drei Aminosäuren Asp, Phe und Gly (= DFG) nahe der ATP-Bindetasche, auf das der Activation Loop folgt, welcher die Phosphorylierungsstellen (z.B. Tyr) enthält. Wird der Loop an diesen Stellen phosphoryliert und die Kinase somit aktiviert, gibt der Activation Loop in einer mit der α C-Helix synchronisierten Bewegung die ATP-Bindetasche frei, sodass ATP und ein Substrat – beispielsweise eine untergeordnete Proteinkinase – binden können (aktiver Zustand). [71]

Der aktive Zustand ist demnach zum einen gekennzeichnet durch die ionische Wechselwirkung zwischen Lysin und dem Glutamat der α C-Helix, zum anderen ist die Seitenkette des DFG-Aspartats zur ATP-Bindetasche hin orientiert (sog. „DFG-in“-Konformation). Im inaktiven Zustand („DFG-out“-Konformation) ist die DFG-Aspartatseitenkette von der ATP-Bindetasche weg gerichtet und die ionische Wechselwirkung zwischen Glutamat und Lysin existiert nicht. [72]

ATP-Bindetasche von Proteinkinasen und die Phosphatübertragung

Die ATP-Bindetasche lässt sich hinsichtlich der Lage der verschiedenen Gruppen des ATP in der Bindetasche und aufgrund der Besonderheiten der verschiedenen Wechselwirkungen von Seiten des Proteins in verschiedene Bereiche einteilen:

Adeninregion: Der Adeninring des ATP bildet zwei Wasserstoffbrückenbindungen zur Hinge-Region aus (s. Abb. 3-3). Die gleichen Interaktionen werden häufig von ATP-kompetitiven Inhibitoren genutzt, welche in diesem Bereich in

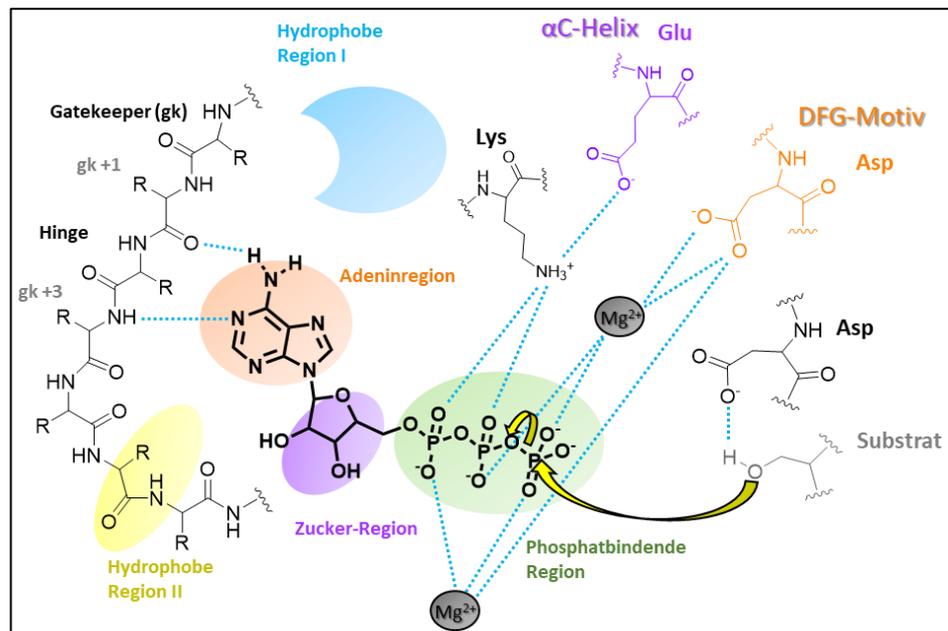


Abb. 3-3 – Bindungsmodus von ATP in der Bindetasche von Proteinkinasen sowie Reaktionsmechanismus der Phosphatübertragung auf ein Substrat (z.B. Tyrosin, Threonin, Serin). Abb. modifiziert nach LIU u. GRAY (2006) [73] und ADAMS (2001) [74]

der Regel auch einen Aromaten aufweisen [73]. Das aromatische System wird sandwichartig durch hydrophobe Aminosäuren in Position gehalten. [75]

Zucker-Region: Der Riboseteil ragt in eine nichtkonservierte, polare Tasche der C-terminalen Domäne (Zucker-Region, s. Abb. 3-3) und bildet über die beiden Hydroxylgruppen Wasserstoffbrückenbindungen zum Protein aus [76].

Phosphatbindende Region: Die drei Phosphatgruppen werden von zwei Magnesium- (auch möglich: Mangan-)Ionen koordiniert, welche wiederum ionische Wechselwirkungen zum Protein ausbilden, z.B. zum Asp des DFG-Motivs (s. Abb. 3-3). Darüber hinaus werden zahlreiche direkte Wasserstoffbrückenbindungen zwischen Protein und den Phosphatgruppen ausgebildet (s. Abb. 3-3⁵) [74]. Der P-Loop steht in engem hydrophilen Kontakt mit dem Triphosphatrest [71, 76].

Die Phosphorylierung eines Substrats (z.B. einer nachgeschalteten Proteinkinase) erfolgt, indem ein konserviertes Aspartat der Proteinkinase als Base die Deprotonierung des Substrats erleichtert und der Sauerstoff der Substrathydroxylgruppe anschließend nukleophil das γ -Phosphat des ATP angreift, wodurch die Übertragung der Phosphatgruppe stattfindet (s. Abb. 3-3) [74, 77]. Die Substratspezifität der Proteinkinase hängt stark von den Aminosäuren neben den Phosphorylierungsstellen des Substrats ab [74].

Hydrophobe Region I: Im hinteren Teil der ATP-Bindetasche befindet sich eine hydrophobe Tasche („Hydrophobe Region I“, s. Abb. 3-3), in die das ATP nicht hineinragt. Der *Gatekeeper* – eine nicht konservierte Aminosäure der Hinge-Region, welche je nach ihrer Größe den Zugang eines Liganden zur hydrophoben Tasche versperrt oder erleichtert – bestimmt die Größe dieser Tasche und hat einen wesentlichen Einfluss auf die Selektivität von Inhibitoren: Inhibitoren, die diese Tasche mit einem Molekülteil besetzen, können durch Veränderung dieses Teils in ihrer Selektivität verändert werden. [76]

Hydrophobe Region II: Diese Region stellt einen zum Lösungsmittel ausgerichteten hydrophoben Kanal dar, der ebenso nicht von ATP besetzt wird und von Inhibitoren genutzt werden kann [76].

⁵ Hier ist nur das konservierte Lys dargestellt, da die weiteren mit dem Triphosphat interagierenden Aminosäuren sich je nach Kinase unterscheiden können.

Proteinkinase-Inhibitoren

Es gibt verschiedene Ansätze, Proteinkinase-Inhibitoren in Klassen einzuteilen. Eine bekannte Klassifizierung von ROSKOSKI [78] berücksichtigt die adressierte Bindestelle (ATP-kompetitiv, Besetzung der tiefen Tasche, allosterische Bindestelle), die Reversibilität der Bindung, die Konformationen des DFG-Aspartats, der α C-Helix und des sogenannten *R-Spine*. Letzteres ist ein regulatorisches Rückgrat aus vier hydrophoben Aminosäuren, welches linear (= Proteinkinase aktiv) oder verzerrt (= Proteinkinase inaktiv) vorliegen kann [78].

Typ-I-Inhibitoren werden von der aktiven Form der Proteinkinase („DFG-in“, „ α C-Helix-in“⁶, linearer R-Spine) gebunden. Im Gegensatz dazu binden die Typ-I $\frac{1}{2}$ -Inhibitoren in der ATP-Bindetasche im Zustand einer inaktiven DFG-in- und (zumeist) α C-Helix-out- und R-Spine-verzerrten Konformation der Proteinkinase. Die ebenso ATP-kompetitiven Typ-II-Inhibitoren binden an die inaktive Kinase („DFG-out“) [78]. Durch die Rotation des DFG-Motivs in den „DFG-out“-Zustand wird eine zusätzliche hydrophobe Tasche freigelegt [79], in die die Typ-II-A-Inhibitoren hineinragen, die Typ-II-B-Inhibitoren jedoch nicht [78]. Typ-III-Inhibitoren besetzen eine allosterische Bindetasche der Proteinkinasen außerhalb, aber in der Nähe der ATP-Bindetasche innerhalb des Spalts zwischen der C- und N-terminalen Domäne, während Typ-IV-Inhibitoren in größerer Entfernung zur ATP-Bindestelle außerhalb des Spalts allosterisch binden. Typ-V-Inhibitoren binden an zwei verschiedenen Stellen der Proteinkinase. Darüber hinaus gibt es die kovalenten Inhibitoren (Typ VI), welche die Proteinkinase im Gegensatz zu den anderen Inhibitortypen irreversibel hemmen [78].

Es ist bekannt, dass man mit der Entwicklung von Typ-II-Inhibitoren generell höhere Chancen auf eine hohe Selektivität hat als mit Typ-I-Inhibitoren, weil die Aminosäuren in der zusätzlichen tiefen Tasche der DFG-out-Konformation weniger konserviert sind im Vergleich zur vorderen ATP-Bindetasche [73, 80, 81]. Es gibt aber auch durchaus Fälle, in denen Typ-II-Inhibitoren weniger selektiv sind bzw. eine hohe Selektivität mit Typ-I-Inhibitoren erreicht werden kann [82, 83]. Bei der Entwicklung selektiver Typ-I-Inhibitoren versucht man Interaktionen mit Aminosäuren zu nutzen, die nicht direkt an der ATP-Bindung beteiligt sind (s.o., Abb. 3-3, Hydrophobe Regionen), sondern in der Nähe der ATP-Bindung eine einzigartige Bindungsregion darstellen, die so nur in dieser speziellen Proteinkinase zu finden ist; dies ist ebenso durch das Berücksichtigen des Gatekeepers möglich, da dieser je nach Proteinkinase variieren kann [73].

⁶ α C-Helix-in = Die Salzbrücke zwischen Lysin und Glutamat (s. S. 18) wird ausgebildet.

Zum einen kann eine Off-Target-Aktivität von wenig selektiven Proteinkinaseinhibitoren zu Nebenwirkungen führen, andererseits können „Multikinase-Inhibitoren“ auch den Vorteil haben, dass sie durch die Hemmung weiterer Kinasen, die ebenso an der Pathogenese einer Erkrankung beteiligt sind, ihre Wirksamkeit erhöhen [84]. Dies wurde beispielsweise für Sunitinib beschrieben, welches unter anderem die Rezeptortyrosinkinasen PDGFR (*Platelet-derived growth factor receptors*), VEGFR-2 (*Vascular endothelial growth factor receptor 2*) und AXL hemmt, die allesamt die Krebsentstehung bzw. -ausbreitung begünstigen [85].

2001 wurde Imatinib als erster Proteinkinaseinhibitor zugelassen – dieser BCR-ABL-Tyrosinkinasehemmer wird für die Therapie der chronisch-myeloischen Leukämie eingesetzt [86]. In der von CARLES ET AL. entwickelten, öffentlich zugänglichen, monatlich aktualisierten Datenbank „PKIDB“ (*Protein Kinase Inhibitor Database*) [87, 88] werden alle Proteinkinaseinhibitoren aufgeführt, die sich in klinischen Prüfungen befinden bzw. bereits zugelassen sind. Nach dem letzten Update der Datenbank am 28.07.2021 handelte es sich um 283 Substanzen, von denen 71 Wirkstoffe insgesamt in unterschiedlichen Ländern zugelassen wurden – 69 davon in den USA (FDA-Zulassung). Der größte Anteil dieser Wirkstoffe (60/69) wird in der Krebstherapie eingesetzt. Die restlichen Wirkstoffe werden in der Therapie von Autoimmunerkrankungen – insbesondere der Rheumatoiden Arthritis –, der Glaukomtherapie, der Myelofibrose, der idiopathischen pulmonalen Fibrose sowie in der Behandlung der Graft-versus-Host-Reaktion eingesetzt. Im 2021 wurden von der FDA bisher bereits sechs neue Proteinkinaseinhibitoren zugelassen, zuletzt der ROCK (*Rho-associated coiled-coil-containing protein kinase*)-Inhibitor Belumosudil (REZUROCK™, Kadmon Pharmaceuticals) im Juli 2021 zur Behandlung der Graft-versus-Host-Reaktion [89].

Proteinkinasen mit Einfluss auf die Insulinsensitivität

Wie zuvor schon aufgeführt (s. S. 15, Tab. 3-2), befinden sich bereits Aktivatoren von Proteinkinasen, nämlich der Glucokinase und der LYN-Kinase, zur Diabetesbehandlung in der klinischen Entwicklung. Desweiteren haben viele der in klinischer Entwicklung befindlichen Wirkstoffe einen Einfluss auf nachgeschaltete Proteinkinasen, wie z.B. auf die Proteinkinase A, welche von den GLP-1-, GIP- und Glucagonrezeptor-Agonisten durch gesteigerte cAMP-Bildung aktiviert wird und die Insulinsekretion steigert. Zahlreiche weitere Kinasen sind im Glucosestoffwechsel involviert und können potenziell durch Aktivatoren bzw. Inhibitoren hin zu einer verbesserten Insulinsensitivität bzw. einer verbesserten Glucoseverwertung beeinflusst werden, um die Blutglucosekonzentration effizient zu senken (s. Abb. 3-4).

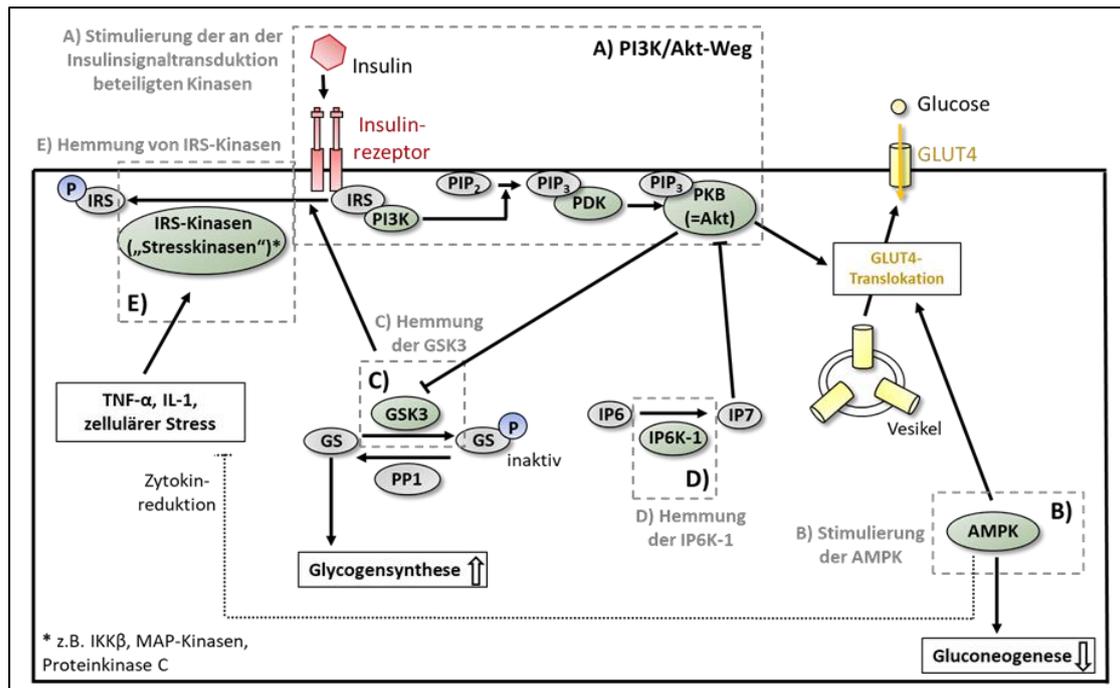


Abb. 3-4 – Möglichkeiten (A-E) für die Verbesserung der Insulinsensitivität durch Beeinflussung von Proteinkinasen. Grüne Ovale: Proteinkinasen. AMPK = AMP-aktivierte Proteinkinase, GLUT4 = Glucosetransporter Typ 4, GS = Glycogensynthese, GSK3 = Glycogen-Synthase-Kinase 3, IKK β = I κ B-Kinase β , IL-1 = Interleukin-1, IP6 = Inositolhexakisphosphat, IP7 = Diphosphoinositolpentakisphosphat, IP6K1 = Inositolhexakisphosphat-Kinase-1, IRS = Insulinrezeptorsubstrat, MAP-Kinasen = Mitogenaktivierte Proteinkinasen, P = Phosphat, PDK = 3-Phosphoinositid-abhängige Proteinkinase, PIP₂ = Phosphatidylinositol-4,5-bisphosphat, PIP₃ = Phosphatidylinositol-3,4,5-triphosphat, PI3K = Phosphoinositid-3-Kinase, PKB = Proteinkinase B, PP1 = Proteinphosphatase 1, TNF- α = Tumornekrosefaktor α . Abb. modifiziert nach SCHULTZE ET AL. (2012) [90]

A) Stimulierung der Kinasen der Insulinsignalübertragung (PI3K/Akt-Weg): Die Bindung von Insulin an den Insulinrezeptor, einer Rezeptor-Tyrosinkinase, setzt die Insulinsignaltransduktion in Gang: es kommt nachfolgend zu einer Tyrosin-Phosphorylierung des Insulinrezeptorsubstrats (IRS-1). Folglich wird die Phosphoinositid-3-Kinase (PI3K) aktiviert, wodurch die intrazelluläre Phosphatidylinositol-3,4,5-triphosphat (PIP₃)-Konzentration steigt. Dies führt zu einer Aktivierung der 3-Phosphoinositid-abhängigen Proteinkinase (PDK), welche die Proteinkinase B (PKB, AKT) stimuliert. Die erhöhte PKB-Aktivität führt zu einer gesteigerten Integration des Glucosetransporters Typ 4 (GLUT-4) in die Zellmembran, wodurch Glucose aufgenommen und verwertet werden kann, sodass die Plasmaglukosekonzentration sinkt. [91–93]

Die Steigerung der Aktivität bzw. Expression der Kinasen in diesem Signalweg würde demnach zu einer verminderten Insulinresistenz führen und ist daher als Ansatz für eine Diabetestherapie interessant. Hier wurden beispielsweise schon Aktivatoren des Insulinrezeptors untersucht [94]. Der Einsatz von Aktivatoren des PI3K/Akt-Signalwegs könnte allerdings die Krebsentstehung begünstigen, da ebenfalls die Zellproliferation gefördert und die Apoptose gehemmt wird [95–97].

B) Stimulierung der AMPK: Die AMP-aktivierte Proteinkinase (AMPK) wird physiologisch durch ein hohes AMP:ATP-Verhältnis in der Zelle stimuliert, welches zustandekommt, wenn oxidativer

Stress oder eine Hypoxie auftritt, wenn niedrige Glucoselevel vorliegen oder es zu einer vermehrten Muskelkontraktion z.B. durch körperliche Aktivität kommt [90].

Die AMPK-Aktivierung führt unabhängig vom Insulinsignalweg ebenso zu einer erhöhten GLUT-4-Translokation an die Zelloberfläche (s. S. 22, Abb. 3-4); dieser AMPK-Signalweg wird von Metformin aktiviert. Die AMPK hemmt die Gluconeogenese, indem sie die Genexpression von an dieser Reaktion beteiligten Enzymen reduziert, und reduziert die Zytokinwirkung.

C) Hemmung der GSK3: Eine weitere Kinase im Netzwerk des Glucosestoffwechsels ist die GSK3 (Glycogensynthase-Kinase 3), deren Konzentration und Aktivität bei Diabetikern erhöht ist [98]. Sie unterdrückt die Glycogensynthase-Aktivität durch Phosphorylierung und hemmt dadurch die Glycogensynthese [91], sodass mehr freie Glucose zirkuliert (s. S. 22, Abb. 3-4). Die PKB des Insulinsignalwegs hemmt die GSK3, sodass die Glycogensynthese gesteigert wird [99]. Die GSK3 führt zu einer Abschwächung des Insulinsignals durch Phosphorylierung des IRS-1 (s. S. 22, Abb. 3-4) [100]. GSK3-Inhibitoren können somit die Insulinresistenz vermindern und die Glycogensynthese steigern. Für Diabetes befindet sich noch kein GSK3-Inhibitor in klinischer Entwicklung. Allerdings befindet sich Tideglusib, ein Nicht-ATP-kompetitiver GSK3 β -Inhibitor in Phase II/III zur Behandlung der kongenitalen myotonen Dystrophie [101]. Eine Phase-II-Studie dieses Wirkstoffs von 2011 bis 2012 zur Behandlung von Alzheimer-Erkrankungen konnte den primären Endpunkt nicht erreichen [102].

D) Hemmung der IP6K1: Die Inositolhexakisphosphatkinase-1 (IP6K1) überträgt eine Phosphatgruppe auf Inositolhexakisphosphat (IP6), sodass ein Pyrophosphatderivat (Diphosphoinositolpentakisphosphat = IP7) entsteht, welches durch Bindung an PKB die Interaktion mit PIP3 verhindert (s. S. 22, Abb. 3-4). Dadurch unterbleibt die Phosphorylierung der PKB durch PDK, sodass der Insulinsignalweg an dieser Stelle gestoppt wird und die Insulinresistenz zunimmt. Durch IP6K1-Inhibitoren könnte die Insulinsensitivität gesteigert werden. [103]

E) Hemmung von IRS-Kinasen: Bei Typ 2-Diabetes wird der Insulinsignalweg gestört, indem das IRS-1 an den Serinresten durch Serin-Threonin-Kinasen phosphoryliert wird und so nicht mehr mit dem Insulinrezeptor assoziieren kann [91]. Die durch erhöhte DAG (Diacylglycerin)-Konzentration stimulierte Proteinkinase C (PKC) bewirkt auf diesem Wege eine Abnahme der Insulinsignaltransduktion; darüber hinaus agiert PKC als vorgeschaltete Kinase als Stimulator entzündlicher Proteinkinase-Kaskaden [104]. Diese bestehen aus Serin-Threonin-Kinasen, welche ebenso durch proinflammatorische Zytokine (TNF- α , IL-1 = Interleukin-1), freie Fettsäuren und zellulären Stress aktiviert [105] und daher auch als „Stresskinasen“ [92] bezeichnet werden. Auch sie führen

eine hemmende Serinphosphorylierung des IRS-1 durch und erhöhen so die Insulinresistenz [91]. Beispiele dafür sind die κ B-Kinase β (IKK- β) [106] und Vertreter der MAP-Kinasen (extrazelluläre signalregulierte Kinasen (ERK) [107], p38 mitogenaktivierte Proteinkinase [108, 109], c-Jun N-terminale Kinase (JNK) [52, 53]) [92].

Eine der entzündlichen Proteinkinase-Kaskaden ist der stressaktivierte MAP-Kinasesignalweg, in dem die *Dual Leucine Zipper Kinase* (DLK) als eine MAP-Kinase-Kinase-Kinase an der Aktivierung von JNK und p38 beteiligt ist (s. Abb. 3-5, S. 25). Die Rolle der DLK für die Diabetesentstehung sowie ihr Potenzial für die Prävention und Behandlung von Diabetes mellitus Typ 2 sollen im Folgenden näher dargestellt werden.

3.2.1 DLK

Die Familie der Mixed-Lineage-Kinasen (MLKs)

Die Dual Leucine Zipper Kinase (DLK, MAP3K12, auch: ZPK, MUK) gehört zur Gruppe der Mixed-Lineage-Kinasen (MLKs), welche als Serin/Threonin-Kinasen fungieren [110, 111]. Die MLK-Familie wird wegen ihrer Aminosäuresequenzähnlichkeit und ihrer biologischen Funktion zur Klasse der TKL (*Tyrosinkinase like*) zugeordnet [68]. Es wurden bisher sieben MLK-Vertreter in Säugetieren identifiziert, die auf Basis ihrer Sequenzähnlichkeit in drei Unterfamilien eingeteilt werden: Die MLKs (MLK1-4), die DLKs (*Dual leucine zipper bearing kinases*: DLK und LZK⁷) und die ZAK (*Zipper sterile- α -motif kinase*) [111].

Die Gemeinsamkeit aller sieben MLKs ist das Vorhandensein einer Kinasedomäne und eines Leucin-Zippers. Die aus 11 Subdomänen bestehende Kinasedomäne der MLKs ähnelt zum Teil den Serin/Threonin-Kinasen, andererseits weist sie aber auch Ähnlichkeit mit der Tyrosinkinase-Familie auf [110, 111]. Diese ursprüngliche Unklarheit, ob es sich um Serin/Threonin- oder Tyrosinkinasen handelt, führte zu der Bezeichnung *mixed lineage* (dt. gemischte Abstammung) [111]. Der Leucin-Zipper dient der Ausbildung von sogenannten „Coiled Coils“ zur Proteindimerisierung [111] – zwei Helices, die sich umeinander winden und somit eine Superhelix ausbilden [112]. Hydrophobe, nicht aromatische Aminosäuren wie das Leucin oder Isoleucin stabilisieren diese Superhelix [111, 112].

⁷ LZK = *Leucin Zipper Bearing Kinase*

Der Signalweg der DLK

Bereits in der Fruchtfliege konnte die Rolle der MLKs für den JNK-Signalweg, welcher den Dorsschluss im Embryonalstadium unterstützt, nachgewiesen werden [113].

In Säugetierzellen sind alle Vertreter der MLK-Familie am MAP-Kinase-Signalweg beteiligt [114]. Die Aktivierung von MAP3-Kinasen kann durch eine Vielzahl von Stimuli erfolgen. Solche Stimuli, wie z.B. Wachstumsfaktoren, Zytokine oder Hormone, binden extrazellulär an G-Protein-gekoppelte Rezeptoren (GPCRs) oder Rezeptor-Tyrosinkinasen (RTKs), die wiederum MAP3-Kinasen aktivieren [115]. Auch intrazellulär, durch MAP4Ks oder oxidativen Stress, können MAP-Kinase-Signalwege stimuliert werden [116]. Diverse Stimuli wurden bisher für die DLK-Aktivierung konstatiert: eine Erhöhung des cAMP-Spiegels und die darauffolgende Aktivierung der Proteinkinase A (PKA) führen zu einer erhöhten DLK-Aktivität in Nervenzellen [117]. Ebenso erhöhen JNK [118], oxidativer Stress sowie der Entzug des Nervenwachstumsfaktors (NGF) die DLK-Aktivität [119, 120]. Für die proinflammatorischen Zytokine TNF- α und IL-1 β wurde in einer β -Zelllinie eine DLK-Aktivierung festgestellt [121]. In murinen, embryonalen Stammzellen führt eine Hemmung der Proteinkinase B (PKB, AKT) zu einer erhöhten DLK-Aktivität [122].

Die DLK aktiviert zum einen die MAP-Kinase JNK über den Zwischenschritt der MKK4/7-Phosphorylierung, zum anderen die MAP-Kinase p38 über den Zwischenschritt der MKK3/6-Phosphorylierung (s. Abb. 3-5) [111]. Die MAP-Kinasen aktivieren nachgeschaltete Transkriptionsfaktoren im Nucleus, deren Transkriptionsaktivitäten unterschiedliche Auswirkungen auf den Zellstoffwechsel haben können; so werden u.a. das Wachstum der Zelle, die Differenzierung, die Apoptose oder Entzündungsprozesse beeinflusst [123].

JNK fördert nach Aktivierung über diesen

Signalweg durch Aktivierung von Transkriptionsfaktoren den Abbau von Axonen [124]. In einem

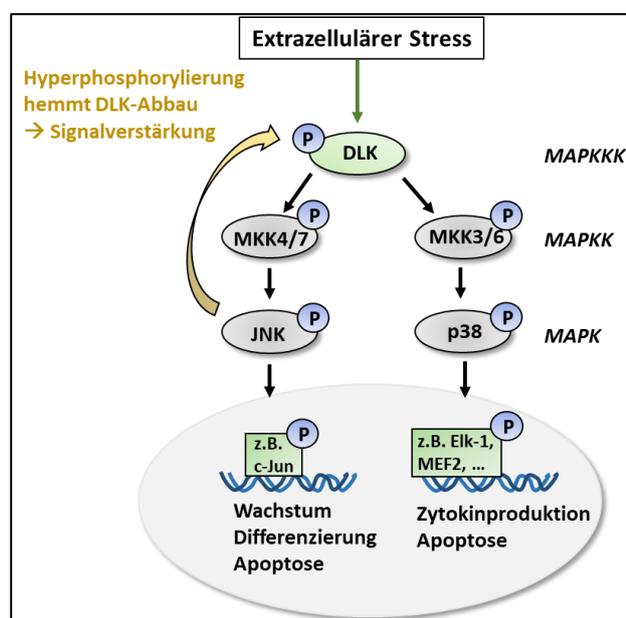


Abb. 3-5 – Stimulierung von MAPK-Signalwegen durch die DLK; Elk-1: ets-like-1; MEF2: Myozyten-spezifischer Enhancer-Faktor 2; Abb. modifiziert nach GALLO U. JOHNSON (2002) [111], MORRISON (2012) [123] und HUNTWORK-RODRIGUEZ ET AL. (2013) [118].

Rückkopplungsmechanismus phosphoryliert JNK die DLK, wodurch diese weniger stark ubiquitinyliert und somit weniger stark abgebaut wird, sodass der Signalweg verstärkt wird („Feed-Forward-Loop“) [118].

Aufbau und Lokalisation der DLK

Die murine DLK wurde erstmals 1994 von HOLZMAN ET AL. [110] bei der Suche nach entwicklungsabhängigen Proteinkinase aus einer embryonalen Mausniere isoliert und sequenziert. Sie besteht aus 888 Aminosäuren mit einer Gesamtmasse von 96 kDa. [110]

Die humane DLK besteht aus 859 Aminosäuren (s. Abb. 3-6). N-terminal gibt es eine Glycin-Prolin-rei-

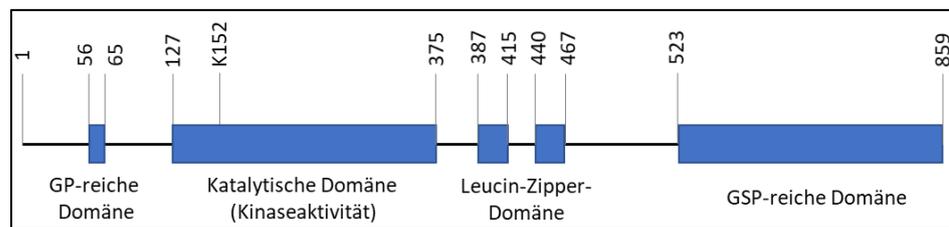


Abb. 3-6 – Darstellung der humanen DLK-Domänen. Die Zahlen entsprechen der jeweiligen Nummer der Aminosäure im Protein. GP: Glycin-Prolin; GSP: Glycin-Serin-Prolin. Abb. modifiziert nach FERRARIS ET AL. (2013) [125].

che Domäne. Die katalytische Domäne befindet sich zwischen den Aminosäuren 127 und 375. Das konservierte Lysin 152 in diesem Abschnitt ist maßgeblich an der ATP-Bindung beteiligt und für die Kinaseaktivität unentbehrlich. 2015 wurde die erste Röntgenkristallstruktur der katalytischen Untereinheit der DLK von PATEL ET AL. [70] veröffentlicht. Nach einem kurzen Abstand von 11 Aminosäuren zur katalytischen Domäne folgen zwei Leucin/Isoleucin-Zipper-Motive und eine Glycin-Serin-Prolin-reiche C-terminale Domäne. [125]

Der Vergleich der Aminosäuresequenzen muriner und humaner DLK ergibt eine Übereinstimmung von 99 % in der katalytischen Domäne und von 95 % in der gesamten Primärstruktur. Der größte Unterschied besteht in der N-terminalen Region außerhalb der katalytischen Domäne, in der 33 zusätzliche Aminosäuren in der Mausequenz zu finden sind. [126]

Die DLK ist ein hochkonserviertes Enzym in verschiedenen Organismen in der Evolution; ein Homolog der DLK ist ebenso in der Fruchtfliege *Drosophila (wallenda)* [127] und im Fadenwurm *C. elegans (DLK-1)* [128] zu finden. Die DLK wird in vielen Gewebetypen von Säugetieren exprimiert, insbesondere im Nervengewebe [129], zentral sowie peripher [130]. Auch im Fettgewebe [131], in der Lunge und in der Niere [110] sowie in den β -Zellen der Langerhans'schen Inseln der Bauchspeicheldrüse [132] wurde die DLK bislang identifiziert. In der Epidermis wurde der Einfluss der DLK auf die Keratozytendifferenzierung festgestellt [133].

Dimerisierung und Autophosphorylierung der DLK

Im inaktiven, unphosphorylierten Zustand der DLK liegt ein Heterodimer mit dem Gerüstprotein JIP1 (*JNK-interacting protein 1*) vor, welcher in dieser Form nicht mit einem weiteren DLK-Monomer homodimerisieren kann [134, 135]. Während dieser Komplex durch JIP1-Tyrosinphosphorylierung durch Kinasen der SRC-Familie stabilisiert wird [136], wird dessen Dissoziation durch Thr103-Phosphorylierung des JIP1 durch JNK gefördert [135]. Die Homodimerisierung von zwei DLK-Proteinen mithilfe der Leucin-Zipper geht der Phosphorylierungsaktivität der DLK voraus; erst dann kann eine Autophosphorylierung und eine Signalweiterleitung (z.B. JNK-Aktivierung) stattfinden [137]. Eine Heterodimerisierung über Leucine-Zipper mit anderen Proteinkinasen der MLK-Familie findet nicht statt, während eine Heterodimerisierung mit LZK mithilfe des N-Terminus möglich ist, welche mutmaßlich durch JIP1 unterstützt wird [111].

Rolle der DLK bei der Pathogenese neurodegenerativer Erkrankungen

Schon früh stellte man fest, dass die DLK in Nervenzellen und hier in besonders hoher Konzentration in den präsynaptischen Endigungen vorliegt [129]. Die DLK nimmt in den Neuronen eine zentrale Regulierungsfunktion bei der Erkennung und Reaktion auf axonale Verletzungen ein. Während DLK-Knockoutmäuse ein reduziertes Axonwachstum sowie eine langsamere Neuronenzellmigration aufweisen und perinatal sterben [138], kann bei erwachsenen Mäusen durch eine spätere DLK-Deletion keine Abnormalität und keine erhöhte Letalität festgestellt werden [139]. Sowohl ein gewebespezifischer DLK-Knockout in Mäusen als auch eine pharmakologische DLK-Hemmung schützen die Neuronen vor der verletzungsinduzierten Degeneration [140, 141]. Nach SIU ET AL. fördert die DLK je nach Umgebungsbedingung die Axonregeneration oder die -degeneration [142]. Die Stimulierung der Axondegeneration wurde im zentralen Nervensystem von Säugetieren beobachtet [141, 143]. Hier stellt die DLK daher ein interessantes Target für die Behandlung neurodegenerativer Erkrankungen wie Parkinson, Alzheimer und amyotrophische Lateralsklerose (ALS) dar [141]. Ebenso könnte das Glaukom mit DLK-Inhibitoren behandelt werden, da es sich auch hierbei um eine neurodegenerative Erkrankung handelt; bei dieser Erkrankung sterben retinale Ganglienzellen ab, wodurch es zum Sehverlust kommt [143].

Rolle der DLK bei der Pathogenese von Diabetes mellitus Typ 2

In den insulinproduzierenden β -Zellen des Pankreas kann die DLK, stimuliert von den prädiabetischen Zytokinen TNF- α und IL-1 β , über zwei Wege zu einer reduzierten Insulinproduktion führen und somit zur Diabetesentwicklung beitragen [121, 144]:

1. Reduktion der CREB-Transkriptionsaktivität: Die DLK inhibiert die Transkriptionsaktivität von CREB (*cAMP response element binding protein*), welcher für das Überleben und für die insulinproduzierende Funktion der β -Zellen von Bedeutung ist [132]. CREB benötigt die beiden Koaktivatoren CRTC (*cAMP regulated transcriptional coactivators*) und CBP (*CREB binding protein*) für seine Transkriptionsaktivität [145]. Die DLK interagiert mit dem CREB-Koaktivator CRTC und phosphoryliert ihn, wodurch die nukleäre Lokalisation des CRTC und dessen Assoziation mit dem Transkriptionsfaktor CREB verhindert wird. Dies hat die Hemmung der CREB-abhängigen Gentranskription zur Folge [146]. Die von CREB regulierten Gene codieren unter anderem für das Insulinrezeptorsubstrat 2 (IRS-2), welches Bestandteil des Insulinsignalwegs und bedeutsam für die β -Zellfunktion ist [147], sowie für das anti-apoptotische Protein Bcl-2 (*B-cell lymphoma 2*), welcher die β -Zellen vor Apoptose schützt [148, 149]. Daher führt eine CREB-Hemmung durch DLK-Aktivität zu einem β -Zellverlust.

Für die Hemmung der CREB-abhängigen Transkription ist eine nukleare Translokation der DLK erforderlich. Für die Translokation der DLK in den Zellkern ist keine Homodimerisierung der DLK notwendig, allerdings muss die Fähigkeit zur ATP-Bindung erhalten sein. Eine erhöhte JNK-Aktivität fördert den Translokationsprozess, vermutlich über den o.g. Mechanismus der JIP1-Phosphorylierung und der darauffolgenden Dissoziation der DLK von dem JIP1-Komplex. Eine DLK-Anreicherung im Zellkern konnte in den Langerhans'schen Inseln des Pankreas bei Mäusen gefunden werden, die fett- und zuckerreich ernährt und so in einen prädiabetischen Zustand versetzt wurden. [121]

2. Reduktion der MafA-Transkriptionsaktivität: Die DLK führt durch eine Verminderung der transkriptionellen Aktivität des Transkriptionsfaktors MafA (*V-maf musculoaponeurotic fibrosarcoma oncogene homologue A*) zu einer verminderten Insulingentranskription. Dies erfolgt über die DLK-induzierte Stimulation des JNK-Signalwegs, woraufhin JNK in den Zellkern transloziert und den Transkriptionsfaktor MafA am Ser65 phosphoryliert. Dadurch kommt es zum Abbau und Funktionsverlust von MafA. [150]

MafA reguliert die Transkription verschiedener Gene, die für die β -Zellfunktion wichtig sind wie z.B. Gene für Insulin, für einen Glucosetransporter, für den GLP-1-Rezeptor und die Glukokinase (Funktion s. S. 15, Tab. 3-2) und für PC-1/3 (Prohormonkonvertase 1 bzw. 3) für die Weiterverarbeitung von Proinsulin [151]. Eine Hemmung der DLK und/oder der JNK würde diesen positiven Effekt des MafA auf die β -Zellfunktion und Insulinproduktion stärken.

Abgesehen von der Beeinflussung von Transkriptionsprozessen durch die DLK mit negativen Auswirkungen auf die β -Zelle spielt die DLK eine übergeordnete Rolle bei der Phosphorylierung des IRS-1 durch JNK (s. S. 23, Unterpunkt E). Durch Verstärkung der JNK-Aktivität durch die DLK könnte der Insulinsignalweg verstärkt gehemmt werden.

Da die DLK-Hemmung sowohl der Verminderung der β -Zellmasse als auch der β -Zellfunktion entgegenwirkt, stellt sie eine interessante Option dar, in die Entwicklung des Diabetes mellitus Typ 2 einzugreifen. Ebenso könnten, wie oben beschrieben, auch Glaukom-, Alzheimer-, ALS- und Parkinsonpatienten zukünftig von der Erforschung und Weiterentwicklung selektiver DLK-Inhibitoren profitieren.

Hemmung der DLK

Es gibt verschiedene Möglichkeiten, die DLK zu hemmen (s. Abb. 3-7). An dieser Stelle möchte ich mich auf die Möglichkeiten der direkten Beeinflussung der DLK-Aktivität beschränken:

Zum einen wäre die Anwendung von Proteinkinase-Inhibitoren möglich, welche die DLK ATP-kompetitiv, allosterisch oder irreversibel durch kovalente Bindung hemmen könnten. Vorteilhaft wäre die generell mögliche orale Applikation in Form von Tabletten oder Kapseln sowie der einfache strukturelle

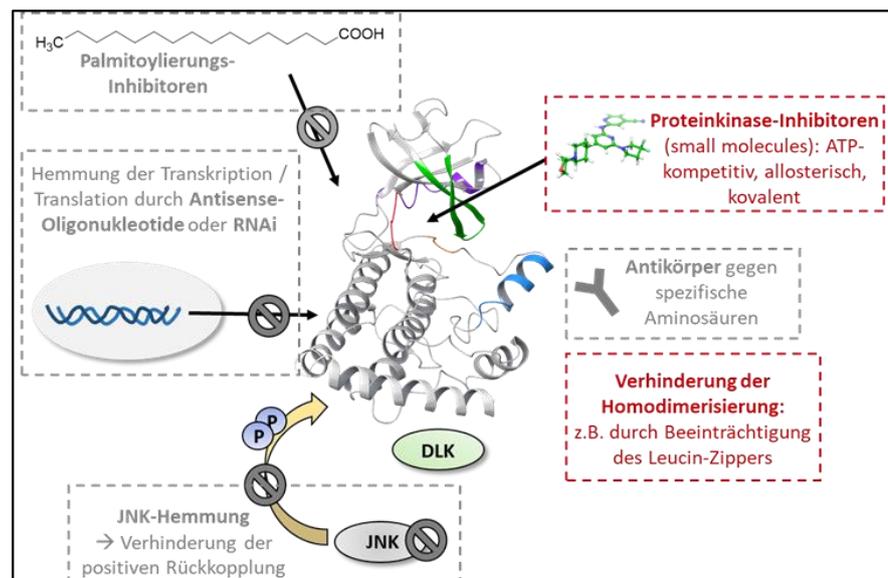


Abb. 3-7 – Mögliche Ansätze zur DLK-Hemmung. Die Möglichkeiten der direkten Beeinflussung der DLK-Aktivität sind rot hervorgehoben. Informationen für die Abbildung stammen aus [78, 152, 153, 154].

Aufbau (*small molecules*) und somit zumeist einfache Synthese dieser Substanzen.

Die Verhinderung der DLK-Homodimerisierung wäre ein weiterer Therapieansatz; durch Zerstörung der α -helikalen Strukturen des Leucin-Zippers wird die Homodimerisierung unterbrochen, und somit auch die Autophosphorylierung und Aktivität der DLK beeinträchtigt [137]. Nach OETJEN UND LEMCKE [152] könnte die Homodimerisierung durch eine Überexpression kinasetoter Mutanten der DLK verringert werden. Beispiele für solche Mutationen sind die K152A-Mutation sowie Mutationen an bestimmten Serinen des Activation Loops, welche für die Autophosphorylierung notwendig sind. Desweiteren wäre es denkbar, die Homodimerisierung der DLK durch *small molecules* zu stören [152]. Solche Moleküle wurden noch nicht für die DLK beschrieben.

Im Folgenden werden die aktuell vorhandenen Möglichkeiten der ATP-kompetitiven Hemmung der DLK näher erörtert. Aufgrund der Tatsache der ersten veröffentlichten Röntgenkristallstrukturen der Kinasedomäne der DLK im Jahr 2015 scheint dies eine Wirkstoffklasse zu sein, die man mithilfe eines strukturbasierten virtuellen Screenings hinsichtlich der Strukturdiversität weiter expandieren könnte.

3.3 ATP-kompetitive Inhibitoren der DLK

Die ersten Ansätze einer DLK-Hemmung basieren auf den semisynthetischen Pan-MLK-Inhibitor-Derivaten von **CEP-1347 (1)** (s. Abb. 3-8), welcher in den ersten Untersuchungen hemmende Eigenschaften auf die Motoneuronen-Apoptose aufwies [155] und aus dem Naturprodukt **K-252a** hervorgeht [156, 157]. **K-252a** ist ein Indolcarbazol-Alkaloid, welches aus der Aktinomyceten-Gattung *Nocardopsis* sp. stammt und ursprünglich als PKC-Inhibitor entdeckt wurde [158]. **K-252a** ist strukturverwandt mit dem Pan-Kinase-Inhibitor **Staurosporin** [159]. Es hemmt ATP-kompetitiv, aber unselektiv eine Vielzahl von Serin-Threonin-Kinasen [160]. Aufgrund der modulierenden Eigenschaften von **K-252a** auf den Neurotrophinstoffwechsel und die Erkennung des Potenzials, neurodegenerative Erkrankungen behandeln zu können, wurde **K-252a** zur Leitstruktur für die Entwicklung neurotropher Wirkstoffe [161]. Aus **K-252a** sollten neue Verbindungen mit höherer Selektivität für bestimmte Kinasen abgeleitet werden, da der Multikinasehemmer **K-252a** wegen seiner Unselektivität unerwünschte Effekte hervorrief [156].

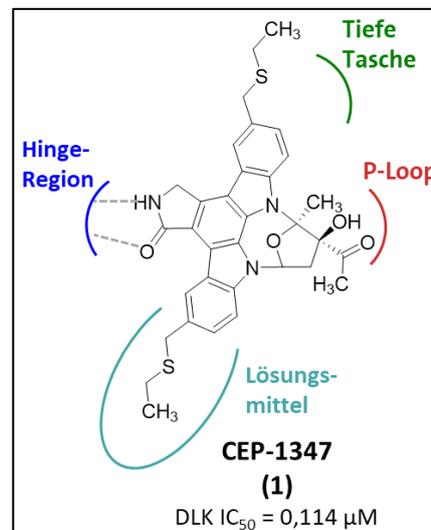


Abb. 3-8 – **CEP-1347 (1)** und dessen mutmaßlicher Bindungsmodus in der ATP-Bindetasche der DLK. Die Informationen über die Bindungsaffinität stammt aus HUDKINS ET AL. (2008) [163].

Indolcarbazol-Derivate wie **CEP-5214** und **CEP-701** weisen ebenso eine DLK-Hemmaktivität auf [162]. Die CEP-Strukturen sind selektiver für die MLK-Unterfamilie als für die DLK. Mit **CEP-6331** wurde erstmals 2008 eine Röntgenkristallstruktur von MLK1 veröffentlicht und somit der Bindungsmodus der Indolcarbazole in MLKs deutlich [163], sodass der Bindungsmodus für die DLK abgeleitet werden kann. Das aromatische System legt sich wie ein Sandwich in den Spalt (Adenin-Position) zwischen der C- und N-terminalen Domäne und bildet mit dem Amid der Pyrrolidinon-Teilstruktur zwei Wasserstoffbrücken zur Hinge-Region aus (s. Abb. 3-8 am Beispiel von **CEP-1347 (1)**). Ein weiteres Indolcarbazolderivat mit nachgewiesener DLK-Bindungsaffinität ist **Staurosporin** (K_d DLK = 1,500 μ M [164]). Diverse Derivate aus einer Veröffentlichung von TAO ET AL. [165], in denen der Indolring durch einen Tetrahydroindazolring ersetzt wurde, weisen DLK- IC_{50} -Werte zwischen 0,001 und 0,118 μ M auf (dargestellt in Kap. 5.2.1, S. 152, Abb. 5-41).

Aufgrund der Sequenzähnlichkeit – insbesondere der ATP-Bindetasche – der DLK zu anderen Serin/Threonin- und auch Tyrosinkinase weisen manche Verbindungen, die ursprünglich zur Hemmung anderer Proteinkinasen entwickelt wurden, eine (unselektive) DLK-Hemmung auf (s. Abb. 3-9). Die Strukturen werden in dieser Arbeit in ihrer Ausrichtung so dargestellt, wie sie in den ATP-Bindestellen anderer Proteinkinasen in veröffentlichten Röntgenkristallstrukturen (s. Kap. 5.2.1, S. 153 f., Abb. 5-42 bis Abb. 5-44 und Kap. 7.10, S. 281, Tab. 7-6) vorliegen. Der mit der Hinge-Region interagierende Teil ist immer ein stickstoffhaltiger, aromatischer Ring (z.B. Oxindol, Indazol, Chinolin, Pyrimidin). Oftmals trägt der Inhibitor einen hydrophilen Substituenten, der aus der Bindetasche hinausragt. Beispielhaft seien an dieser Stelle die DLK-Bindungsaffinitäten für die Tyrosinkinase-Inhibitoren **Sunitinib (2)** (K_d DLK = 100 nM [164]), **Crizotinib (3)** (K_d DLK = 170 nM [125]) und **Bosutinib (4)** (K_d DLK = 670 nM [125]) genannt. **Tozasertib (5) (VX-680)**, DLK K_d = 190 nM [164], ein Pan-Aurora-Kinase-Inhibitor, wurde in einer Studie zur Hemmung der

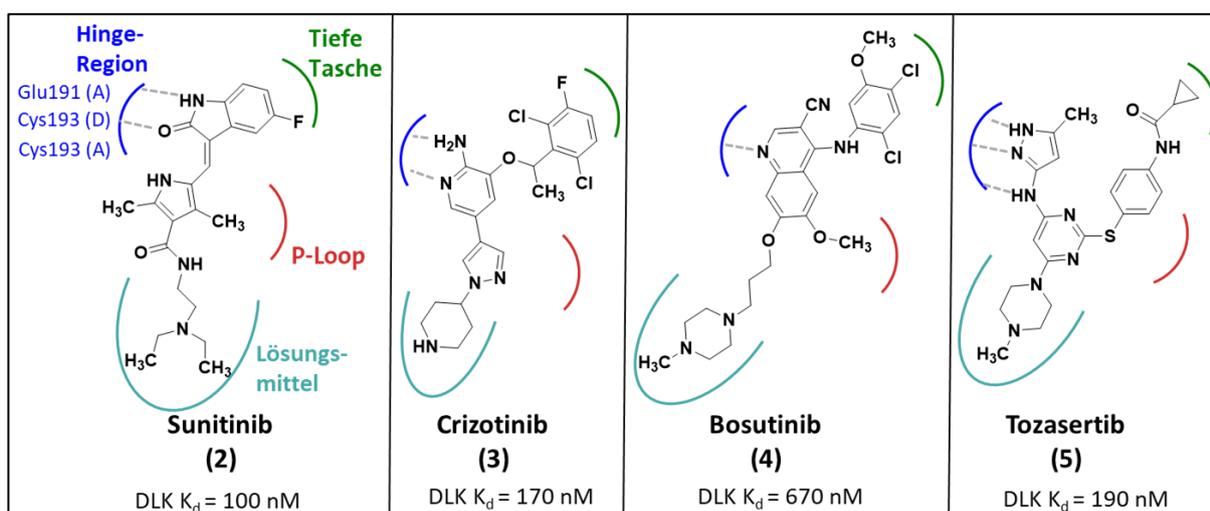


Abb. 3-9 – Diverse Serin/Threonin- bzw. Tyrosinkinaseinhibitoren mit unselektiver DLK-Hemmaktivität und mutmaßlichem Bindungsmodus in der ATP-Bindetasche der DLK

DLK in Nagetiermodellen für Glaukom und traumatische Optikusneuropathie eingesetzt und konnte das Überleben von retinalen Ganglienzellen dosisabhängig steigern [143].

PATEL ET AL. [166] entwickelten ausgehend von einem High-Throughput-Screening-Hit **(6)** (s. Abb. 3-10) eine Reihe von hoch aktiven und selektiven DLK-Inhibitoren. Der erste Hit **(6)** enthält ein Grundgerüst mit zwei Heteroaromaten, die über eine Aminofunktion miteinander verknüpft sind. Der Thiazolring bildet mutmaßlich drei Wasserstoffbrücken zur Hinge-Region der DLK aus. Sie optimierten den Hit zu der Leitstruktur **GNE-3511 (7)** (s. Abb. 3-10) mit verbesserter inhibitorischer Aktivität (DLK $K_i < 0,5$ nM) und verbesserter Fähigkeit, bei oraler Einnahme die Blut-Hirn-Schranke zu überwinden. Die Selektivität gegenüber anderen Kinasen ist sehr hoch: die Hemmung der MLKs 1/2/3 erfolgt mit IC_{50} -Werten von 67,8 nM/767 nM/602 nM und die Hemmung der untergeordneten MAP2-Kinasen MKK4/7 erfolgt mit IC_{50} -Werten von jeweils > 5000 nM. **GNE-3511 (7)** zeigt drei für Kinaseinhibitoren typischen Wasserstoffbrückenbindungen zur Hinge-Region (Glu191 und Cys193) in der Abfolge *Donor-Akzeptor-Donor* (in Abb. 3-10 dargestellt). Der difluorierte Pyrrolidinring bildet hydrophobe Interaktionen mit dem P-Loop aus, während der Piperidinrest aus der Bindetasche und zum Lösungsmittel hinausragt. Dieser Substituent ist – insbesondere nach einer möglichen Protonierung – besonders hydrophil. Da die Bindung in die ATP-Bindetasche der aktiven Form der Kinase („DFG-in“) erfolgt, kann man bei **GNE-3511 (7)** von einem Typ-I-Inhibitor sprechen.

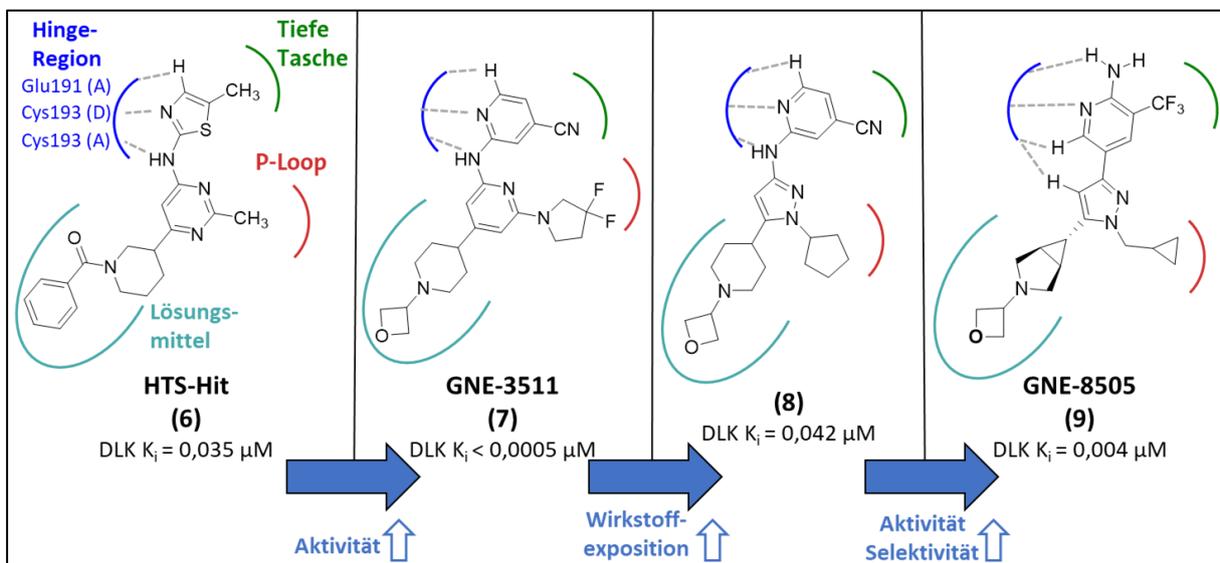


Abb. 3-10 – Entwicklung selektiver, ATP-kompetitiver DLK-Inhibitoren. Substanz **(6)** und **(7)** von PATEL, COHEN ET AL. (2014) [166], Substanz **(8)** von PATEL, HARRIS ET AL. (2015) [70], Substanz **(9)** von PATEL, MEILANDT ET AL. (2017) [167]

Zur Verbesserung der physikochemischen Eigenschaften, insbesondere der zu hohen Clearance im Rattenmodell, ersetzten PATEL ET AL. [70] im Sinne eines Scaffold-Hoppings den Pyrimidinkern durch einen Pyrazolkern (s. Abb. 3-10). Diese Verbindung **(8)** ist ebenso hochselektiv und überwindet die Blut-Hirn-Schranke; die DLK-Bindungsaffinität ist mit 42 nM immer noch sehr hoch,

aber geringer als von **GNE-3511 (7)** ($< 0,5$ nM). Der Bindungsmodus dieser Substanzen wurde mithilfe von Röntgenkristallstrukturen der DLK-Kinasedomäne von PATEL ET AL. bestimmt (PDB [168, 169]: **5CEO**, **5CEP**, **5CEQ** und **5CEN**). Sie wurden jeweils mit **GNE-3511 (7)**, mit Verbindung **(8)** und einem weiteren Inhibitor aus dieser Entwicklungsreihe sowie als Apoenzym erhalten. Die Proteinkonformationen befinden sich in einem Zustand, der dem aktiven Zustand der Kinase wegen der hier vorliegenden DFG-in-Konformation nahekommt, allerdings ist die ionische Wechselwirkung zwischen der α C-Helix und Lysin 152 nicht ausgebildet. Diese Besonderheit des großen Abstands zwischen den Partnern Lys152 und Asp161 (α C-Helix) kommt dadurch zustande, dass die α C-Helix durch das Vorhandensein des Asp161 statt des sonst konservierten Glu an dieser Stelle einen Knick aufweist. Die physiologische Rolle dieser besonderen Konformation ist noch unklar. [70]

Da das Vorkommen des Gatekeepers Met190 und der Hinge-Aminosäure Phe192 sowie die Aminosäure Gln197 in der DLK in dieser Konstellation eine Besonderheit im Vergleich zu anderen Proteinkinasen darstellt [142], wurden die DLK-Inhibitoren im weiteren Verlauf durch die Forschungsgruppe an ihren Hinge-interagierenden Stellen hin zu einer verbesserten Selektivität verändert (s. S. 32, Abb. 3-10, **GNE-8505 (9)**) [167]. Die strukturellen Veränderungen führten darüber hinaus zu einer höheren Bindungsaffinität im Vergleich zu **(8)** bei gleichbleibender ZNS-Gängigkeit [167]. Der Bindungsmodus dieses neuen Inhibitorotyps wurde später mithilfe von zwei weiteren DLK-Röntgenkristallstrukturen (PDB: **5VO1** und **5VO2**) bestätigt.

Acht Patente beinhalten einen großen Datensatz von insgesamt rund 2000 DLK-Inhibitoren, die strukturell den von PATELET AL. entwickelten selektiven Verbindungen ähneln: Sechs Patente von F. HOFFMANN-LA ROCHE und GENENTECH, INC.: WO 2013/174780 [170] („Substituted dipyridylamines and uses thereof“), WO 2014/111496 [171] („3-Substituted pyrazoles and use as DLK inhibitors“), WO 2014/177060 [172] („Biheteroaryl compounds and uses thereof“), WO 2014/177524 [173] („C-Linked heterocycloalkyl substituted pyrimidines and their uses“), WO 2015/091889 [174] („Pyrazole derivatives and uses thereof as inhibitors of DLK“) und WO 2016/142310 [175] („Tricyclic DLK inhibitors and uses thereof“); sowie zwei Patente von dem BOARD OF REGENTS DER UNIVERSITY OF TEXAS SYSTEM: WO 2018/044808 [176] („Inhibitors of dual leucine zipper (DLK) kinase for the treatment of disease“) und WO 2018/107072 [177] („Bicyclo[1.1.1]pentane inhibitors of dual leucine zipper (DLK) kinase for the treatment of disease“). In jeder Patentschrift herrscht jeweils eine Grundstruktur vor (s. Abb. 3-11, fett hervorgehoben). Ausgehend von dem bekannten Bindungsmodus von **GNE-3511 (7)** (s.S. 32, Abb. 3-10) kann aufgrund der hohen strukturellen Ähnlichkeit zu den Strukturen aus den Patentschriften von einem analogen Bindungsmodus ausgegangen werden; es handelt sich also ebenfalls um ATP-kompetitive Typ-I-Inhibitoren.

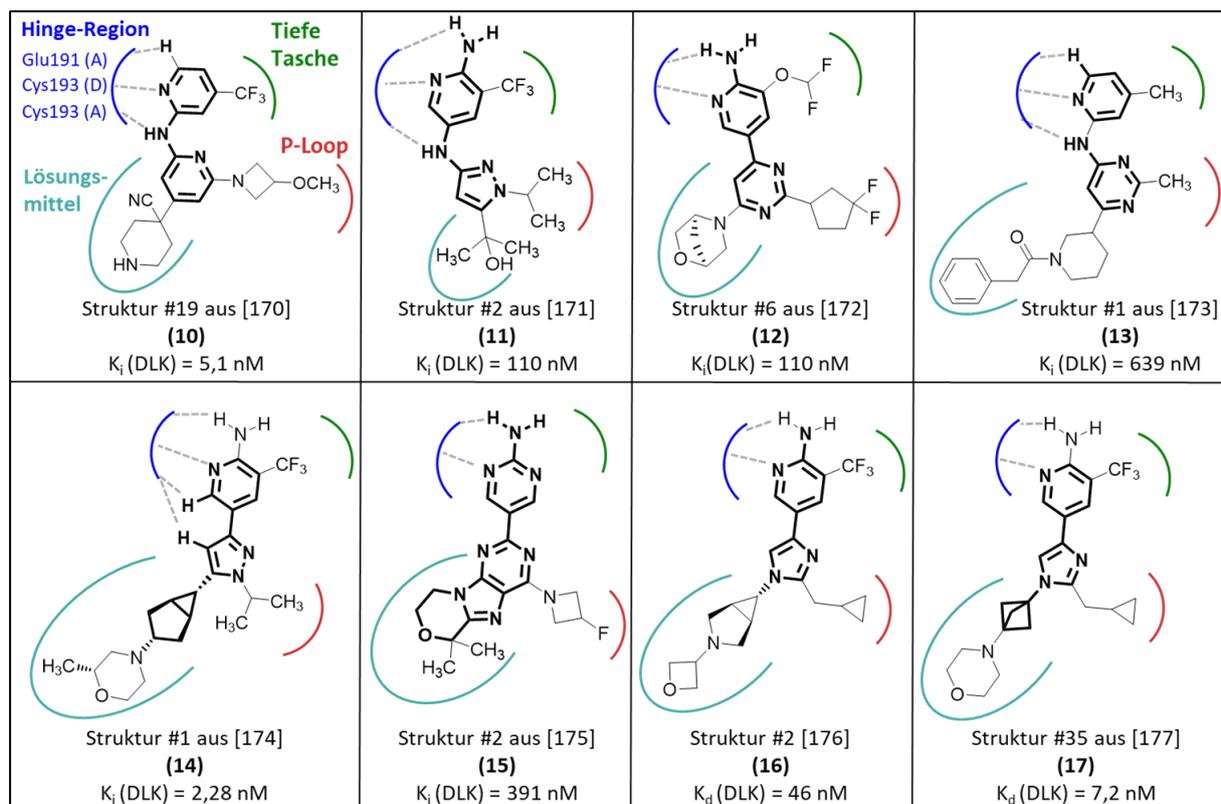


Abb. 3-11 – Typische Vertreter der ATP-kompetitive DLK-Inhibitoren aus den Patentschriften von F. Hoffmann-La Roche und Genentech Inc. [171-175] sowie von dem Board of Regents der University of Texas System [176, 177]. Grundstruktur ist fett hervorgehoben.

Die K_i -Werte der Verbindungen **(10)** bis **(15)** wurden mittels TR-FRET (*Time-resolved fluorescence resonance energy transfer*) bestimmt, während die K_d -Werte von **(16)** und **(17)** mit der KINOMEscan[®]-Methode (KdELECT[®]-Assay) von DiscoverX bestimmt wurden.

LE PICHON ET AL. testeten die DLK-Inhibitoren **GNE-3511 (7)** und **GNE-8505 (9)** (s. S. 32, Abb. 3-10) im transgenen SOD1^{G93A}-Mausmodell der neurodegenerativen Erkrankung ALS, die vom Verlust von Motoneuronen und neuromuskulärer Verbindungen geprägt ist. Beide DLK-Inhibitoren konnten die Zahl der Zellen reduzieren, welche positiv für das phosphorylierte c-Jun waren. **GNE-3511 (7)** konnte das Überleben der Motoneuronen verlängern und die Denervierung der neuromuskulären Verbindungen und somit das Fortschreiten der ALS-Erkrankung verlangsamen. [141]

Mit dem ebenfalls von GENENTECH, INC. entwickelten DLK-Inhibitor **GDC-0134** (Struktur nicht veröffentlicht) wurde im Jahr 2016 eine Phase-I-Studie (NCT02655614 bzw. GN29823) an ALS-Patienten zur Untersuchung der Sicherheit, Verträglichkeit sowie der Pharmakokinetik begonnen und im März 2020 abgeschlossen [178]. Darüber hinaus lief eine Bioverfügbarkeits-Studie (NCT03237741 bzw. GP39778) im Jahr 2017 mit gesunden weiblichen Probanden, in der der Einfluss von Nahrung und Protonenpumpenhemmern auf die Pharmakokinetik des **GDC-0134** untersucht wurde [179]. Ergebnisse der Studie wurden bisher nicht veröffentlicht.

Die hier dargestellten selektiven DLK-Inhibitoren gehen strukturell alle aus einem Hit hervor und zeigen keine große strukturelle Diversität. Darüber hinaus lag ein Fokus in den bisherigen Entwicklungen auf eine Verbesserung der ZNS-Gängigkeit der Verbindungen, da sie zukünftig bei neurodegenerativen Erkrankungen eingesetzt werden sollten.

3.4 Wirkstoffsuche

Von der ersten Forschung bis zur Marktreife eines Arzneimittels vergehen in der Regel mindestens 10 Jahre und es entstehen Kosten von 450 Mio. bis 2,3 Mrd. EUR [180]. Sehr viele Wirkstoffe scheitern in Phase II der klinischen Prüfung, am häufigsten (zu 52 %) aufgrund mangelnder Wirksamkeit. Die Zahl der von der FDA zugelassenen Arzneimittel stagniert seit etwa 2010 und liegt bei etwa 30 Zulassungen pro Jahr, während die Ausgaben für Forschung und Entwicklung weiter steigen [181].

Nach Identifizierung und Validierung einer biologischen Zielstruktur (*Target*), die für eine Erkrankung eine Rolle spielt, erfolgt die Entdeckung bzw. Entwicklung einer Leitstruktur (engl. *lead*) [182] – diese erzeugt an der Zielstruktur eine biologische Wirkung, ist aber hinsichtlich seiner biologischen Aktivität und Selektivität an der Zielstruktur sowie der Bioverfügbarkeit noch entwicklungsfähig [180]. Zu diesem Zweck werden im Anschluss Leitstrukturoptimierungen durchgeführt, indem Derivate der Leitstruktur erzeugt und getestet werden. Vielversprechende Arzneistoffkandidaten gehen in die präklinischen und – falls diese erfolgreich waren – in die klinischen Prüfungen [182].

Das High-Throughput-Screening (HTS) ist eine der wichtigsten Methoden zur Auffindung aktiver Verbindungen (*Hits*), aus denen Leitstrukturen entwickelt werden können. Mit dieser Methode wird die Interaktion zwischen Testsubstanz und Target in automatisierbaren, miniaturisierten *in-vitro*-Assays gemessen; hunderttausende Substanzen können innerhalb weniger Tage auf diese Weise gescreent werden [183]. Die Optimierung der Assay-Technologie hinsichtlich des geringstmöglichen Substanzverbrauch und schneller, automatisierbarer Durchführbarkeit sowie die Kombination mit der kombinatorischen Chemie, welche einen großen Pool von Verbindungen in das Screening einspeist, ließ auf große Erfolge hoffen. Die Trefferquoten sind allerdings ernüchternd gering [183]. Die synthetisierten Substanzdatenbanken haben in der Regel keine große strukturelle Vielfalt und bilden nur einen kleinen Teil des großen chemischen Raums ab [184], welcher schätzungsweise aus 10^{23} [185] bis 10^{60} Strukturen [186] besteht. Nicht zu vernachlässigen ist der finanzielle und zeitliche Aufwand für die Synthese bzw. den Kauf von hunderttausenden Verbindungen sowie für die Assay-Verbrauchsmaterialien und das Entwickeln bzw. Instandhalten der automatisierten Testsysteme für die biologische Prüfung dieser Vielzahl von Substanzen [187, 188].

3.4.1 Computerbasierte Methoden zur Wirkstoffsuche und -entwicklung

Eine Möglichkeit zur Reduktion von Kosten und Arbeitsaufwand ist die Anwendung von computerbasierten (*in-silico*)-Methoden, welche sowohl für die Vorfilterung von Datenbanken, als auch zur Leitstrukturidentifizierung und -optimierung eingesetzt werden können. Eine rationale Vorauswahl von Strukturen mithilfe von computergestützten Methoden (CADD = *Computer-aided drug design/discovery*) reduziert die Anzahl der zu testenden Verbindungen *in vitro*, ermöglicht eine gezieltere Suche und erhöht die Wahrscheinlichkeit des Auftretens eines Hits in der biologischen Testung. Somit stellen die CADDs eine zeit- und kosteneffektive Alternative und Ergänzung zum HTS dar. Sie lassen sich in strukturbasierte und ligandbasierte Methoden einteilen. [189]

Strukturbasierte Methoden: Die strukturelle Information über ein Targetprotein, experimentell bestimmt mit der Röntgenkristallstrukturanalyse, der NMR-Spektroskopie oder der Cryoelektronenmikroskopie, ist Voraussetzung für die Anwendung strukturbasierter Methoden [189, 190]. Eine bekannte Quelle für solche 3D-Strukturen ist die im Jahr 1971 eingerichtete Proteindatenbank PDB (*Protein Data Bank*) [168, 169], in der aktuell 181.969 Proteinstrukturen abrufbar sind (Stand: 14.09.2021). Gibt es keine 3D-Struktur, lässt sich diese, bei Vorhandensein einer primären Aminosäuresequenz des Proteins, häufig durch Homologie-Modelling auf Basis eines 3D-Templates eines strukturell ähnlichen Proteins am Rechner generieren [183]. Folgende Techniken finden im strukturbasierten Design Anwendung:

a) *Docking:* Bei dieser Methode werden Liganden virtuell in die Bindetasche des Proteins eingepasst und mit einer Scoringfunktion bewertet. Auf Basis dieser berechneten Interaktionsenergien, welche die Bindungsaffinität der Liganden widerspiegeln, können gezielt geeignete Moleküle für die Testung ausgewählt werden. [183]

b) *De-novo-Design:* Ein Ligand wird sukzessive in der Proteinbindetasche entworfen [183].

c) *MD (Molecular Dynamics)-Simulationen:* Simulationen von Molekülbewegungen werden eingesetzt, um die Stabilität eines Protein-Liganden-Komplexes in einer definierten Umgebung (pH-Wert, Wassermoleküle, Ionen) zu überprüfen oder um Proteinkonformationen zu generieren, die wiederum für ein Ligandendocking eingesetzt werden können. Desweiteren können Bindungsaffinitäten aus der MD-Simulation abgeleitet werden. [191]

d) *Molekulare Interaktionsfelder:* Nach Analyse der Bindetasche werden dort Regionen für geeignete interagierende Gruppen (z.B. Wasserstoffbrückenakzeptoren (HBA) / -donoren (HBD), aromatische Ringe, saure oder basische Gruppen) eines potenziellen Liganden identifiziert. Dies

geschieht beispielsweise durch das Abtasten der Bindetasche mit einer Sonde in Form eines definierten Interaktionstyps (z.B. HBA). Aus der räumlichen Anordnung der Funktionalitäten ergibt sich ein bestimmtes Muster (= Pharmakophor), auf dessen Basis ein De-Novo-Design oder ein Screening durchgeführt werden kann. [180] Pharmakophore können statt vonseiten der Bindetasche auch von bekannten, bindenden Liganden ausgehend generiert werden; diese Variante zählt dann zu den ligandbasierten Methoden.

Ligandbasierte Methoden: Die ligandbasierten Methoden benötigen die Kenntnis über die 2D- bzw. 3D-Informationen aktiver bzw. inaktiver Liganden und finden oft Anwendung, wenn die Targetstruktur nicht verfügbar ist. [189]

a) *QSAR (Quantitative structure-activity relationships):* QSAR-Modelle werden aus einer Reihe aktiver Liganden erstellt, um aus den strukturellen Informationen dieser Liganden die biologische Aktivität von Strukturen anderer Liganden unbekannter Aktivität vorherzusagen. Die Struktur wird somit mit der biologischen Aktivität in eine mathematische Beziehung gesetzt [183].

b) *Pharmokophormodelle:* In diesem Fall nutzt man die strukturellen Informationen vorhandener aktiver Liganden zur Erstellung eines Pharmakophors. [189]

c) *Ähnlichkeitssuche:* Eine aktive Verbindung dient als Ausgangssubstanz, um eine Moleküldatenbank nach Molekülen mit Ähnlichkeit zu dieser Referenzverbindung zu durchmustern. Hier liegt die Annahme im Vordergrund, dass strukturell ähnliche Moleküle auch ähnliche Eigenschaften z.B. an der Zielstruktur aufweisen. Die Suche erfolgt häufig auf Basis von sogenannten molekularen Fingerprints. Diese speichern strukturelle Informationen eines Moleküls in Bitstrings, mithilfe derer ein Abgleich effizienter möglich ist. [192]

Das **virtuelle Screening (VS)**, als *in-silico*-Pendant zum *in-vitro*-Screening, ist die Anwendung von computerbasierten Methoden für das Durchmustern, Bewerten, Filtern und Sortieren von Moleküldatenbanken, um potenziell aktive Substanzen für die Synthese, den Erwerb und/oder für die biologische Testung auszuwählen [193]. Hierbei können Millionen von virtuellen Verbindungen innerhalb einer vertretbaren Zeit und ohne die Aufwendung von Kosten für „reale“ Substanzen mit dem Computer gescreent werden [194]. Mehrere *in-silico*-Methoden können im Sinne von Filtern hintereinandergeschaltet werden, um so die Zahl der Moleküle schrittweise immer weiter zu reduzieren. Dadurch ist es möglich, eine große Molekülanzahl durch eine Methode mit geringem Rechenaufwand zu screenen und für die weitere Verarbeitung der erhaltenen reduzierten Anzahl von Kandidaten eine Methode höherer Komplexität einzusetzen [183]. Je nachdem, welche

der *in-silico*-Methoden verwendet werden, spricht man von einem strukturbasierten (SBVS) oder ligandbasierten virtuellen Screening (LBVS) [189].

Ausgangspunkt und Voraussetzung für das SBVS ist die strukturelle Information über das Target. Der Screeningworkflow besteht in der Regel aus den folgenden hintereinander ablaufenden Schritten: [195, 196]

1.) *Vorbereitung der Proteinstruktur*: Die 3D-Struktur eines Targets aus der Röntgenkristallographie, der NMR-Spektroskopie oder eines Homology-Modells wird für die Kompatibilität mit der entsprechenden Docking-Software analysiert und vorbereitet. Fehlende Informationen wie z.B. über formale Ladungen, fehlende Atome oder Bindungsordnungen werden ergänzt. Durch Modellierung solcher Fehlstellen ist es möglich, vollständige Proteinstrukturen zu erhalten, die weiterverarbeitet werden können. Die Proteinstruktur wird anschließend in eine energetisch günstige Konformation gebracht. [195, 196]

2.) *Vorbereitung der Screeningdatenbank*: Die Substanzbibliothek kann kommerziell verfügbare Substanzen, leicht synthetisierbare Substanzen oder virtuell *in-silico* konstruierte Substanzen enthalten. Der wichtigste Schritt für die Anwendbarkeit computerbasierter Methoden ist die Überführung der Moleküle von 2D- in chemisch korrekte 3D-Konformationen. Unterschiedliche Tautomere bzw. Protonierungszustände werden generiert. Filter können eingesetzt werden, um Moleküle mit unerwünschten Eigenschaften für die Weiterentwicklung zu einem Arzneistoff zu entfernen. [195]

3.) *Screening (z.B. durch Docking)*: Die Liganden aus der Substanzbibliothek werden in die Targetbindetasche gedockt. Die Liste von Dockingposen wird nach einem Scorewert sortiert, der aus den Interaktionen zwischen Ligand und Protein berechnet wird und die Bindungsaffinität des Liganden widerspiegelt [195]. Das Dockingprogramm kann in einem Fall zu besonders guten Ergebnissen bzw. neuen Hits führen, während die Anwendung bei einer anderen Proteinstruktur keine korrekten Dockingposen dieser Hits (*Dockingproblem*) oder keine korrekte Bewertung von Posen (*Scoringproblem*) ergibt [197]. So können im ersten Fall Strukturen vernachlässigt werden, die biologisch aktiv sind, aber am Computer schlecht bewertet wurden (falsch Negative), oder es kommt zu einer Auswahl von fälschlicherweise gut bewerteten Molekülen ohne tatsächliche biologische Aktivität (falsch Positive) [198].

4.) *Postprocessing*: In der Regel gibt es viele Dockinglösungen, die in ihrer Menge weiter reduziert und von falsch positiven Hits weitestgehend befreit werden sollten. Zusätzliche Scoringfunktionen können eingesetzt werden, um die Anreicherung potenziell aktiver Liganden zu

verbessern [196]. Mit Clustermethoden werden ähnliche Moleküle zusammengefasst, um die Datenmenge zu reduzieren. Die besten Dockingposen werden meistens visuell durch einen Experten für medizinische Chemie beurteilt und eine Auswahl für die biologische Testung getroffen [195].

Oftmals wird das strukturbasierte virtuelle Screening mit ligandbasierten Methoden kombiniert, wenn sowohl strukturelle Informationen über das Target als auch über aktive Liganden vorliegen [199–202]. Vorteilhaft bei der Anwendung ligandbasierter virtueller Screeningmethoden im Vergleich zum strukturbasierten Screening ist die häufig schnellere Durchführbarkeit (z.B. Ähnlichkeitssuche über Fingerprints) [203] und das Hervorbringen aktiverer Hits (i.d.R. besser als 1 μM beim LBVS vs. 1-100 μM beim SBVS). Ein Nachteil des ligandbasierten virtuellen Screenings liegt darin, dass die Hits den eingesetzten bekannten, aktiven Liganden strukturell stark ähneln können und die Chance auf die Entdeckung neuer Grundgerüste gering ist [204]. Im Gegensatz dazu zielt das strukturbasierte virtuelle Screening auf die Wechselwirkung zwischen Ligand und Proteinbindetasche ab, sodass neue chemische Grundgerüste gefunden werden können [203, 204], die ebenso wie bereits bekannte Liganden in die Bindetasche passen und ähnliche oder sogar neue Interaktionsmuster ausbilden.

3.5 Ziel der Arbeit

Wie zuvor beschrieben spielt die DLK eine zentrale Rolle bei der Pathogenese neurodegenerativer Erkrankungen sowie von Diabetes mellitus Typ 2. Die Forschung konzentriert sich dabei bisher vorrangig auf die Entwicklung und Optimierung von Blut-Hirn-Schranken-gängigen DLK-Inhibitoren für die Behandlung neurodegenerativer Erkrankungen. Dagegen stellt die Entwicklung nicht ZNS-gängiger DLK-Inhibitoren zur potenziellen Behandlung des Diabetes mellitus Typ 2 einen neuen Forschungsansatz dar. Einen Grundstein für die Entwicklung solcher Wirkstoffe kann durch die Identifizierung von DLK-Inhibitoren mit neuartigen Grundgerüsten und Substitutionsmustern gelegt werden. In dieser Arbeit soll ein strukturbasiertes virtuelles Screening zur Identifizierung neuer kompetitiver Inhibitoren der ATP-Bindungsstelle der DLK geplant und durchgeführt werden. Als Screeningtool soll dabei das Docking von kommerziell erhältlichen Molekülkandidaten dienen. Daraus resultierende Hits können in späteren Projekten zu Leitstrukturen weiterentwickelt werden.

Der Workflow zur Lösung dieser Aufgabe soll folgende zentrale Punkte umfassen (s. Abb. 3-12):

Analyse der Zielstruktur

In der Literatur veröffentlichte Röntgenkristallstrukturen der DLK und insbesondere deren ATP-Bindetasche sollen hinsichtlich ihrer Größe, Flexibilität und charakteristischen Interaktionsmöglichkeiten analysiert und mit den entsprechenden Tools vorbereitet werden. Ziel ist das Verständnis der Bindungsverhältnisse und die Auswahl repräsentativer, diverser DLK-Proteinkonformationen als Ensemble für das virtuelle Screening. Konservierte Wassermoleküle in der Bindetasche und die Ergebnisse von *Molecular-Dynamics*-Simulationen können bei der Auswahl eine Rolle spielen. Die Kenntnis des Bindungsmodus bekannter DLK-Inhibitoren soll in dem späteren Dockingprozess der Liganden einbezogen werden.

Vorbereitung des Screening-Datensatzes

Die ZINC-Datenbank (*Zinc is not commercial*) [205] mit über 100 Millionen erwerbbarer Substanzen soll Ausgangspunkt der Screeningbibliothek für das Docking sein. Zwecks geplanter Testung im Anschluss an das Screening sollen nur die kommerziell erhältlichen Strukturen gescreent werden.

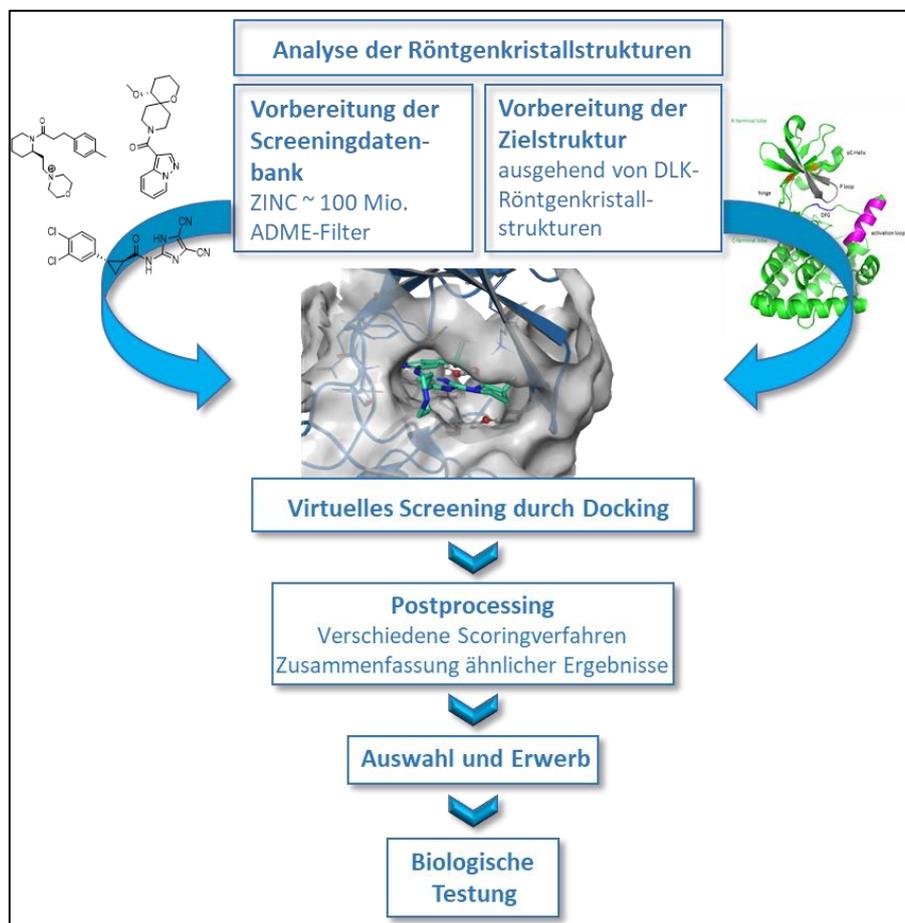


Abb. 3-12 – Geplanter Workflow für die Durchführung des strukturbasierten virtuellen Screenings

Da aufgrund eines zu hohen Rechenaufwands nicht alle ZINC-Substanzen gedockt werden können, soll die Bibliothek vor dem Docking nach unterschiedlichen Kriterien (z.B. *Lipinski's Rule of Five*) gefiltert werden, um beim Screening Substanzen mit potenziell ausreichender Bioverfügbarkeit zu finden (*drug-like structures*). Hierbei soll auf einen *Hit-like*-Charakter der Strukturen geachtet werden (z.B. geringes Molekulargewicht). Der Vergleich bestimmter struktureller und physikochemischer Eigenschaften mit bekannten DLK-Inhibitoren ist hierbei von großer Bedeutung, um einerseits vorrangig Strukturen mit bisher nicht verwendeten Strukturelementen und -eigenschaften in der Ausgangsdatenbank zu erhalten und eine große strukturelle Diversität zwischen den Liganden zu generieren, und um andererseits die wesentlichen Informationen dieser bekannten Strukturen zur Stärkung der Bindung, wie typische Interaktionsmuster und Kinaseinhibitor-Eigenschaften, in die Entscheidungskriterien für die finale Auswahl neuartiger Strukturen mit einfließen lassen zu können.

Virtuelles Screening durch Docking

In einer schrittweise hierarchischen Vorgehensweise sollen die Substanzen in die repräsentativen Proteinensemble-Strukturen mit steigender Genauigkeit gedockt werden. Der Screeningprozess soll mithilfe eines kleineren Testdatensatzes aus bekannten aktiven DLK-Inhibitoren und Decoys im Vorwege optimiert werden. Hierbei soll eine möglichst gute frühe Anreicherung bekannter DLK-Inhibitoren erreicht werden. Ebenso soll die Durchführung des eigentlichen ZINC-Screenings durch Zusatz bekannter DLK-Inhibitoren zur Screeningdatenbank stetig kontrolliert und gesteuert werden.

Postprocessing

Die voraussichtlich große Anzahl von Dockingergebnissen soll so weiterverarbeitet und reduziert werden, dass Strukturen erhalten und in der Rangfolge weit vorne stehen werden, welche biologisch aktiv sind. Durch den Einsatz zusätzlicher Scoringmethoden zum Dockingscore könnte ein kombinierter Score berechnet werden, der unterschiedliche Bewertungsmethoden vereinigt und so das Scoringproblem vermindert. Hierbei soll möglichst auch ein QSAR-Modell basierend auf den literaturbekannten DLK-Inhibitoren zur Vorhersage der Aktivität der gedockten Liganden als eine zusätzliche Bewertungsmethode eingesetzt werden. Clustermethoden könnten eingesetzt werden, um die Menge der Dockinglösungen zu reduzieren.

Auswahl und Erwerb

Beurteilungskriterien bei der Durchsicht einer festgelegten Menge bestbewerteter Posen sollen aus den Erkenntnissen der Workflow-Erstellung und des Postprocessings abgeleitet werden. Etwa 50-100 Substanzen sollen ausgewählt und käuflich erworben werden.

Biologische Testung

Die erworbenen Hit-Strukturen sollen in verschiedenen, voneinander unabhängigen Testsystemen auf ihre DLK-inhibitorische (IC_{50}) bzw. -bindende Aktivität (K_d) untersucht werden. Hierzu ist die Überprüfung auf Eignung und die Anwendung verschiedener selbst durchgeführter Assay-Methoden geplant, aber auch externe Testungen sollen in Betracht gezogen werden.

4 Material und Methoden

Die Theorie angewandter Methoden und Modelle des computergestützten Designs sowie die Funktionsweise verschiedener Programme, die in dieser Arbeit verwendet wurden, sollen in diesem Kapitel erläutert werden.

4.1 Kraftfelder

Grundlage zur Beschreibung von Bindungen und Kräften zwischen einzelnen Atomen und letztendlich zur Beschreibung und Optimierung von Molekülstrukturen in dieser Arbeit war die Anwendung sog. Kraftfelder. Um die Energie und Geometrie von Molekülen zu berechnen, können Methoden der Quantenmechanik (QM) oder der Molekülmechanik (MM) angewendet werden. Die Verwendung von *ab-initio*-Verfahren (QM) mit Berücksichtigung aller Elektronen wäre für große Moleküle wie Proteine zu komplex und rechenintensiv. *Semiempirische* Verfahren (QM) können die Berechnungszeit verkürzen, indem z.B. nur Valenz-Elektronen mit einbezogen und viele Parameter basierend auf empirischen Daten abgeschätzt werden. [206]

Die Kraftfeldmethoden (*Force Field, FF*) hingegen erlauben eine einfache und schnelle Berechnung der Energie einer dreidimensionalen Molekülstruktur bzw. von Komplexen. In diesem Modell, basierend auf der klassischen Molekülmechanik (MM), werden die Atome als Massekugeln behandelt, deren Größe und Eigenschaften sich je nach Atomtyp unterscheiden. Zwischen den Kugeln wirken mechanische Federkräfte entlang der kovalenten Bindungen. Die Energieberechnung beruht auf dem Hook'schen Gesetz, welches die Federkraft abhängig von der Federkonstante und Auslenkung berechnet. Die aus dem Kugel-Feder-Modell berechneten kovalenten Energien aus Bindungslängen, -winkeln und -rotationen sowie die nicht-kovalenten van-der-Waals (vdW)- und elektrostatischen Wechselwirkungsenergien werden in einer Funktion zur Beschreibung des molekülmechanischen Kraftfelds zusammengeführt. [206]

Die Parameter für jeden Atomtyp und jede Bindungsart werden aus experimentellen Daten abgeleitet oder aus quantenmechanischen Berechnungen gewonnen. [183]

Die gesamte mechanische Energie E_{MM} eines Moleküls lässt sich mit folgender allgemeiner Gleichung beschreiben (s. auch Abb. 4-1): [207]

$$E_{MM} = E_{\text{bond}} + E_{\text{angle}} + E_{\text{dihedral}} + E_{\text{electr}} + E_{\text{vdW}} \quad (1)$$

- Kovalente Beiträge:

E_{bond} – Energie zur Streckung bzw. Stauchung einer Bindung

E_{angle} – Winkeldeformationsenergie

E_{dihedral} – Torsionswinkelenergie

- Nicht-kovalente Beiträge:

E_{electr} – elektrostatische Wechselwirkungsenergie

E_{vdW} – van-der-Waals-Wechselwirkungsenergie

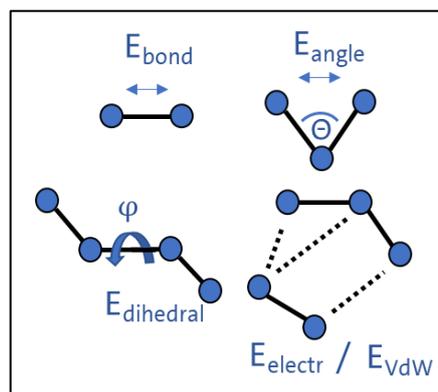


Abb. 4-1 – Darstellung der einwirkenden Kräfte (Pfeile = kovalent, gestrichelt = nicht-kovalent) mit den zugehörigen Energiebeiträgen in der Molekülmechanik. Abb. modifiziert nach KLEBE (2009) [183], S. 228.

Häufig verwendete molekularmechanische Kraftfelder sind

das AMBER [208, 209]-, OPLS [210]- und CHARMM [211]-Kraftfeld. All diese Kraftfelder berücksichtigen die Polarisierbarkeit nicht explizit, sondern setzen feste Ladungen ein. Sie verwenden alle Lennard-Jones-Potenziale zur Beschreibung des vdW-Terms und es werden harmonische Streckungs- und Deformationsterme sowie Torsionsterme basierend auf dem Flächenwinkel eingesetzt. [212]

In dieser Arbeit wurden für Liganden- und Proteinminimierungen sowie für die Postdocking-Minimierungen der Posen das OPLS3 [212]- und OPLS2005 [213]-Kraftfeld verwendet. Für die Moleküldynamiksimulationen wurden OPLS3 und, seit der Neueinführung im Jahr 2019, OPLS3e eingesetzt.

4.1.1 Das OPLS-Kraftfeld

Das OPLS (*Optimized Potentials for Liquid Simulations*)-Kraftfeld, welches sich seit der Veröffentlichung [214] durch William L. Jorgensen an der Purdue Universität im Jahr 1984 in einem stetigen Entwicklungsprozess befindet, wurde insbesondere für die Simulation von organischen Molekülen in Lösungen konzipiert. Es gibt grundsätzlich zwei Varianten dieses Kraftfelds: zum einen das OPLS-UA (*united atom*)-Kraftfeld, welches jedes Kohlenstoffatom mit den daran gebundenen Wasserstoffatomen als ein vereinigt Atom zusammengefasst betrachtet [214], und zum anderen das OPLS-AA (*all-atom*)-Kraftfeld, welches jedes einzelne Atom berücksichtigt. In der frühen Entwicklung für kleine Moleküle wie einfache Kohlenwasserstoffe [215] erprobt, wurde im weiteren Verlauf unter Hinzunahme eines großen Teils der Streckungs- und Winkeldeformations-

parameter aus dem AMBER-AA-Kraftfeld [216] und unter Verwendung von *ab-initio*-Berechnungen für den Torsionsterm ein OPLS-AA-Kraftfeld für viele gängige organische funktionelle Gruppen implementiert. Somit werden auch die strukturellen Elemente von Proteinen mit diesem Kraftfeld abgedeckt [210].

Die potenzielle Energie eines Moleküls wird mit dem OPLS-Kraftfeld in folgender Weise durch bindende und nichtbindende Wechselwirkungen beschrieben: [212]

$$E = \sum_{i < j} [q_i q_j e^2 / r_{ij} + 4 \epsilon_{ij} (\sigma_{ij}^{12} / r_{ij}^{12} - \sigma_{ij}^6 / r_{ij}^6)] f_{ij} \quad (2)$$

$$+ \sum_{\text{bonds}} K_r (r - r_{eq})^2$$

$$+ \sum_{\text{angles}} K_\theta (\theta - \theta_{eq})^2$$

$$+ \sum_{\text{dihedrals}} \left[\frac{V_1}{2} (1 + \cos \varphi) + \frac{V_2}{2} (1 - \cos 2\varphi) + \frac{V_3}{2} (1 + \cos 3\varphi) + \frac{V_4}{2} (1 + \cos 4\varphi) \right]$$

Nichtkovalente Beiträge: Der erste Term beinhaltet die Summe aus den elektrostatischen und den van-der-Waals-Wechselwirkungsenergien zur Berechnung des nichtbindenden Potentials. Zur Berechnung der elektrostatischen Energie wird das Coulomb-Potential auf Basis der Atomladungen q_i und q_j , der Elektronenladung e und des Abstands zwischen diesen Atomen r_{ij} verwendet. Zum anderen geht das Lennard-Jones-Potential zur Beschreibung der van-der-Waals-Interaktionen zwischen ungeladenen Atomen i und j in Abhängigkeit vom Abstand r_{ij} dieser Atome in den nichtbindenden Term ein. Das Lennard-Jones-Potential setzt sich zusammen aus einem stark abstoßenden Term, der Pauli-Repulsion $\sigma_{ij}^{12}/r_{ij}^{12}$, welcher mit kleiner werdendem Abstand r rasch ansteigt, und aus einem anziehenden Term σ_{ij}^6/r_{ij}^6 (Dispersion). Für den Abstand $r_{ij} = \sigma_{ij}$ nimmt die Energie den Wert Null an. Der Parameter ϵ_{ij} entspricht der Tiefe des Potentials und gibt die Stärke der Anziehung wieder. [212]

Kovalente Beiträge: Die Bindungsstreckungs- und die Winkeldeformationsenergie (Term 2 und Term 3) folgen dem Hook'schen Gesetz mit einer Kraftkonstanten K_r für die Bindungslänge bzw. K_θ für den Bindungswinkel und der Differenz aus der aktuellen Bindungslänge r bzw. des aktuellen Bindungswinkels θ und der Referenz im Gleichgewicht (eq) zum Quadrat. Der vierte Term behandelt die Torsionswinkelenergie als periodische Funktion. In diesen Beitrag fließen der aktuelle Torsionswinkel φ und die Koeffizienten (V) der Fourier-Reihe mit ein. Die Fourierkoeffizienten sind ein Maß für die Größe der Barriere für die Rotation. [212]

Die in dieser Arbeit verwendeten OPLS-Modelle bauen auf dem OPLS-AA-Modell auf. Im OPLS2005-Kraftfeld wurden die nichtkovalenten Terme unverändert von dem OPLS-AA-Kraftfeld übernommen, während für eine bessere Erfassung der funktionellen Gruppen von Liganden zusätzliche Streckungs-, Winkeldeformations- und Torsionsparameter eingeführt wurden [213]. Für

die Parametrisierung des Torsionsterms wurden Daten von nur etwa 600 Verbindungen eingesetzt, während hingegen das neuere OPLS2.0-Kraftfeld ab-initio-Daten von etwa 11.000 Molekülen verwendet [217].

Das im Jahr 2016 veröffentlichte OPLS3-Kraftfeld stellt eine 30%ige Verbesserung gegenüber OPLS2005 bezüglich der Genauigkeit bei der Posenvorhersage dar [212, 218]. Im Vergleich zu den Vorgängerversionen wurde eine umfangreiche Parametrisierung der Valenz- und Torsionsterme auf Basis quantenchemischer Berechnungen vorgenommen und ein verbessertes Ladungsmodell eingesetzt, welches Ladungsstellen außerhalb der Atome (*sog. off-atom charge sites*) für eine höhere Genauigkeit der Berechnungen mit einbezieht. Bei diesen Ladungsstellen handelt es sich um freie Elektronenpaare an aromatischen Stickstoffatomen und um das Sigma-Loch bei Halogenatomen [212].

2019 wurde OPLS3 in der SCHRÖDINGER SUITE durch OPLS3e (*extended*) ersetzt, welches ein neues Torsionsmodell auf Grundlage eines erweiterten Trainingsdatensatzes beinhaltet und eine neue Methode der Ladungszuweisung nutzt. Diese Methode basiert auf on-the-fly semiempirischen QM-Berechnungen partieller Ladungen für Ligandbestandteile, insbesondere für neuartige Heterozyklen, welche außerhalb des Kraftfeldtrainingsatzes liegen [218]. OPLS2005 steht optional immer noch zur Verfügung in der SCHRÖDINGER SUITE.

Kraftfelder finden auch Anwendung in Moleküldynamiksimulationen (s. Kap. 4.5).

4.1.2 Das TIP3P-Wassermodell

In den MD-Simulationen dieser Arbeit wurde das von Jorgensen et al. [219] entwickelte Wassermodell TIP3P (*Transferable intermolecular potential 3 point*) verwendet. Es handelt sich um ein kraftfeldbasiertes Modell, welches Wasser als starres Molekül mit drei elektrostatischen Wechselwirkungspunkten darstellt. Das Sauerstoffatom erhält eine negative Partialladung von -0,834 und die beiden Wasserstoffatome jeweils eine positive Partialladung von +0,417. Bei der Berechnung der vdW-Wechselwirkungen mithilfe des Lennard-Jones-Potenzials wird nur das Sauerstoffatom berücksichtigt. Die Geometrie des Wassermoleküls entspricht der experimentellen Geometrie in der Gasphase. [219]

4.2 Vorbereitung von Proteinstrukturen

Die Durchführung eines strukturbasierten virtuellen Screenings erfordert eine umfassende Vorbereitung der Zielstruktur zur Präzisierung von Atompositionen, -bindungen und -ladungen, welche nicht aus den experimentellen Proteinstrukturdaten hervorgehen. [220]

Die Koordinaten der verwendeten Röntgenkristallstrukturen wurden aus der „Protein Data Bank“ (PDB) [168, 169] abgerufen, einer der weltweit führenden Ressourcen für dreidimensionale Biomolekülstrukturen.

Experimentelle Röntgenkristallstrukturdaten sind fehlerbehaftet und die dargestellten Atome und Bindungen sind nicht als „fixiert“ anzusehen. Bei der Röntgenkristallstruktur handelt es sich vielmehr um ein Modell in optimaler Übereinstimmung mit der aufgenommenen Elektronendichtekarte gemittelt über die Aufnahmezeit [221]. Bei einer Auflösung von 2,5 Å schätzt man die Standardabweichung der Atomkoordinaten auf etwa 0,4 Å [222].

Eine hohe Auflösung⁸ von Proteinstrukturen ermöglicht im Allgemeinen ein genaueres Docking. Daher sollte schon bei der Strukturauswahl ein Augenmerk auf die Auflösung gelegt werden, um Dockingergebnisse schlechter Qualität zu vermeiden. Die Auflösung von Proteinstrukturen liegt i.d.R. zwischen 1,5 und 3 Å [183]. Eine Auflösung von 2,5 Å oder besser wird in der Literatur häufig als akzeptabel für Dockingexperimente bzw. für den Einsatz im virtuellen Screening angesehen [223–225].

Im Allgemeinen ist die Auflösung von Röntgenkristallstrukturen zu gering, sodass die schlecht beugenden Wasserstoffatome nicht beobachtet werden können. Diese müssen daher vor der Weiterverarbeitung ergänzt werden. Die Abwesenheit von Wasserstoffatomen in den Röntgenkristallstrukturen führt auch dazu, dass Protonierungszustände saurer oder basischer Aminosäuren und von Liganden nicht festgelegt sind. [222]

Stark bewegliche Sequenzabschnitte werden bei der Röntgenkristallographie wegen der geringen Elektronendichte nicht immer erfasst. Wenn solche Sequenzlücken in der Nähe der Bindetasche auftreten, sollten sie vor dem Docking aufgefüllt werden. Dies muss aber individuell entschieden werden. Der Einfluss von Wassermolekülen, Ionen und Cofaktoren auf die Protein-Ligand-Interaktionen und auf die Stabilität der Proteinstruktur muss geprüft werden. Auch hier gilt einzuschätzen, ob solche Moleküle und Ionen beibehalten oder entfernt werden sollen. [226]

⁸ d.h. kleinere Werte für die Auflösung

Aufgrund ähnlicher Ordnungszahlen des Sauerstoff- und Stickstoffatoms sind deren Positionen im Amidrest des Asparagins bzw. Glutamins nicht eindeutig definiert [222]. Zudem lassen sich rotamere und tautomere Zustände von Imidazolringen wie im Histidin nicht unterscheiden (s. Abb.

4-2). Die Analyse von Wasserstoffbrückenbindungen (WBB) zu benachbarten Aminosäureketten oder von *Clashes*⁹ ermöglicht häufig eine Entscheidung über die richtige Konformation [222, 228]. Die Orientierung der Wasserstoffatome in Hydroxyl- und Thiolgruppen geht ebenfalls nicht aus der Röntgenkristallstruktur hervor [228].

Die oben genannte Rotation von Gruppen und die Veränderung der Ionisierungs- bzw. Tautomerisierungszustände werden von Algorithmen (wie *protassign* in PROTPREP) so vorgenommen, dass Wasserstoffbrückenbindungen im gesamten System optimiert werden. Auch Wassermoleküle werden zu diesem Zweck umorientiert. Der Algorithmus *protassign* bewertet die Anzahl und Qualität ausgebildeter Wasserstoffbrückenbindungen auf Grundlage der Geometrie und enthält Strafterme für zu nahe Kontakte zwischen zwei Wasserstoffatomen. Die Wahrscheinlichkeit von (De-)Protonierungen saurer bzw. basischer Gruppen wird mithilfe von PROPKA berechnet. [228] Es ist also erforderlich, Bindungen, Bindungsordnungen (Hybridisierungszustände) und formale Ladungen in einer PDB-Struktur zu analysieren und abhängig vom verwendeten Kraftfeld klar zu definieren. [228]

In dieser Arbeit wurden alle Protein Vorbereitungen mit dem PROTPREP-Tool (*Protein Preparation Wizard* [229]) von SCHRÖDINGER durchgeführt. Alle hier beschriebenen Probleme können mit diesem Tool in einem hierarchischen Prozess in den folgenden Schritten behandelt werden: [230]

1. *Preprocess*: Atom- und Bindungstypen werden analysiert und entsprechend des Kraftfelds zugeordnet. Wasserstoffatome werden unter Verwendung von Standardgeometrien hinzugefügt. Disulfidbrücken können gebildet und Selenomethionin in Methionin umgewandelt werden. Vorhandene Liganden erhalten – unter Verwendung von EPIK (s. Kap. 4.3, S. 54) und eines definierten pH-Bereichs – wahrscheinliche Ionisierungs- bzw. Tautomerisierungszustände. Endständige Aminosäuren können mit N-Acetyl- (am C-terminalen Ende) bzw. N-Methylamidgruppen (am N-terminalen Ende) gekappt werden. [231, 232] Fehlende Seitenketten

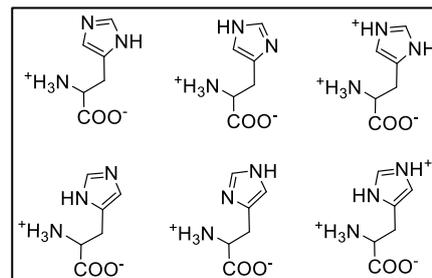


Abb. 4-2 – Die sechs möglichen rotameren und tautomeren Zustände des Histidins. Abb. modifiziert nach KIM ET AL. (2013) [227]

⁹ Clashes = Überlappungen der vdW-Radii zweier oder mehrerer Atome

und *Loops*¹⁰ können mit PRIME ergänzt werden (s. Kap. 4.2.1, S. 50 f.). Wassermoleküle können in diesem Schritt entfernt werden.

2. *Review and Modify*: Die vorverarbeitete Proteinstruktur wird auf Fehler analysiert. Unerwünschte Strukturen – z.B. Peptidketten, Wassermoleküle, Liganden – können gezielt entfernt und weitere Ionisierungs- bzw. Tautomerisierungszustände von Liganden erzeugt werden.

3. *Refine – H-bond assignment*: Die Optimierung des Wasserstoffbrückenbindungsnetzwerks wird durchgeführt. Dabei werden die Positionen der Wasserstoffatome von Wassermolekülen, Hydroxy-/Thiolgruppen, der Amidgruppen von Asparagin und Glutamin sowie des Histidin-Imidazolrings angepasst. (De-)Protonierungen an sauren bzw. basischen Proteinseitenketten werden vorgenommen. Nach der Optimierung können optional Wassermoleküle gelöscht werden.

4. *Refine – Restrained minimization*: Am Ende der Proteinvorbereitung steht die Minimierung der Struktur mit dem ausgewählten Kraftfeld (SCHRÖDINGER MAESTRO: OPLS3(e) oder OPLS2005, vgl. Kap. 4.1.1). Alle Nicht-Wasserstoffatome werden in ihrer Position innerhalb eines definierten Bereichs fixiert (SCHRÖDINGER MAESTRO: 0,3 Å). Wasserstoffatome werden nicht in ihrer Beweglichkeit eingeschränkt. Eventuell vorhandene Clashes können durch die Minimierung entfernt werden. [228]

In dieser Arbeit wurden verschiedene für die Protein-Ligand-Wechselwirkung bedeutsame Wassermoleküle in der Bindetasche (sog. *structural waters*, s. Kap. 5.1.1, S. 128 f. u. Kap. 5.4.1.2, S. 184 f.) in der Proteinvorbereitung berücksichtigt, um die Auswirkungen der unterschiedlichen Wasserpositionierungen auf die Dockingexperimente zu untersuchen.

4.2.1 Loop Prediction and Refinement

Fehlende Koordinaten von Loops und Seitenketten können mit dem Programm PRIME [233] (SCHRÖDINGER, LLC) rekonstruiert werden. Dazu wird die entsprechende unvollständige Proteinstruktur auf eine Sequenz als Vorlage übereinandergelagert, welche die Aminosäuren des fehlenden Bereichs enthält. Der in PRIME implementierte Ab-initio-Algorithmus PLOP (*Protein Local Optimisation Program* [234]) baut den Loop von beiden Seiten schrittweise auf. [235]

Für die Suche nach passenden Konformationen der – von der Sequenzvorlage kopierten – Peptidbausteine werden Rotamer-Bibliotheken verwendet. Diese enthalten bevorzugte *Backbone*¹¹-

¹⁰ Loop = Schleifenbereich als lokales Strukturelement einer Proteinstruktur / Polypeptidkette

¹¹ Backbone = Rückgrat eines Makromoleküls (z.B. Protein-Backbone)

Torsionswinkel, die durch Analyse aus Proteinstrukturdaten mit hoher Auflösung erhalten wurden. Konformationen angefügter Peptidbausteine, deren Atome mit anderen Atomen kollidieren¹², werden verworfen [235]. Abhängig von der Länge des Loops kann zwischen drei Genauigkeitsstufen für das Sampling gewählt werden: Default (bis zu 5 Aminosäuren), Extended (6-11 Aminosäuren) und Ultra Extended (ab 12 Aminosäuren) [236]. Auf Basis einer Seitenkettenrotamer-Bibliothek werden Seitenkettenkonformationen gewählt, die frei von Clashes mit dem Loop-Backbone und den übrigen Proteinatomen (außerhalb des Loops) sind [234]. Die so generierten Loopkandidaten werden vollständig auf Clashes überprüft. Zur Verminderung des Rechenaufwands für den Optimierungsprozess wird die Zahl der Loopkandidaten durch ein RMSD-basiertes k-means-Clustering reduziert. Der Loopkandidat jedes Clusters, der dem Clusterzentrum am nächsten liegt, wird beibehalten [235].

Für die anschließende Seitenkettenoptimierung der Clusterrepräsentanten werden Seitenketten-Rotamer-Bibliotheken mit erhöhter Auflösung verwendet. Zunächst werden dem Loop-Backbone alle Seitenketten in einem zufälligen Rotamerzustand neu hinzugefügt. Alle Rotamere einer Seitenkette werden gescreent und das Rotamer mit der niedrigsten Energie ausgewählt. Dadurch dass dieser Prozess für jede Seitenkette nacheinander durchgeführt wird, können sich die energetischen Zustände von zuvor optimierten Torsionswinkeln wieder verändern. Der Prozess wird iterativ solange durchgeführt, bis für mehr als 95 % der Seitenketten Rotamere mit der niedrigsten Energie eingesetzt werden konnten (Konvergenz). Für die Energieberechnungen wird das OPLS3(e)- bzw. OPLS2005-Kraftfeld (s. Kap. 4.1.1) in Kombination mit dem Surface Generalized Born Lösungsmittelmodell [237, 238] verwendet. Zum Schluss wird eine vollständige Minimierung des Loop-Backbones und der Seitenketten zur Anpassung der Bindungslängen und -winkel sowie der Positionen von unpolaren Wasserstoffatomen durchgeführt. [235]

¹² Zur Überprüfung von Kollisionen (Clashes) wird mit einem definierten Grenzwert für den Abstand zwischen zwei Atomen gearbeitet.

4.3 Vorbereitung von Liganden

Wie Proteinstrukturen (s. Kap. 4.2) müssen auch Liganden vor einem Dockingexperiment bzw. einem virtuellen Screening vorbereitet werden. In einem mehrstufigen Prozess werden die Strukturen gefiltert und cheminformatisch bearbeitet, um arzneistoffähnliche (*drug-like*) Verbindungen in 3D-Konformation für das Docking bereitzustellen [239]. Die Ligandenvorbereitung hat einen großen Einfluss auf die Anreicherung im virtuellen Screening [220].

Filter

Ein vorausgehendes Filtern der Screeningdatenbank wirkt sich günstig auf den Rechenaufwand für das Docking aus. Aus den physikochemischen Eigenschaften bekannter, bioverfügbarer Arzneistoffe können Grenzwerte abgeleitet werden, um Strukturen aus dem Datensatz zu entfernen, die unvorteilhafte Eigenschaften bezüglich der peroralen Bioverfügbarkeit besitzen und voraussichtlich nicht zu einem Arzneistoff weiterentwickelt werden können [239]. Unzureichende pharmakokinetische Eigenschaften sind einer der häufigsten Gründe für das Scheitern der Wirkstoffentwicklung [240]. Von den Autoren COUPEZ UND LEWIS [197] werden folgende Filterkriterien vorgeschlagen:

- Lipinski's Rule of Five [241] (s. Tab. 4-1)
- Anzahl rotierbarer Bindungen (#RB), Polare Oberfläche (PSA)
- Abwesenheit von Toxophoren

Die Lipinski's Rule of Five (s. Tab. 4-1), veröffentlicht von LIPINSKI ET AL. im Jahr 1996, sind vier Regeln, die von mehr als 90 % der oral bioverfügbaren Verbindungen erfüllt wurden. Liegen diese molekularen Eigenschaften unterhalb der vorgegebenen Grenzwerte, so ist dies mit einer optimalen Löslichkeit und Permeabilität und somit mit einer hohen oralen Bioverfügbarkeit assoziiert [241]. Dazu gehört ein Molekulargewicht von deutlich unter 500 Da, eine Begrenzung der Anzahl von Wasserstoffbrückendonoren (#HBD) auf maximal 5, von Wasserstoffbrückenakzeptoren (#HBA) auf 10 und ein log P-Wert von maximal 5. Der log P-Wert nimmt mit der Lipophilie einer Substanz zu. Einen Ersatz für den experimentell ermittelten log P-Wert stellt der berechnete clog P-Wert (*calculated log P*) dar. Mit dem Programm ClogP [242], basierend auf der fragmentbasierten Leo-Hansch-Methode [243], können clog P-Werte von Molekülstrukturen abgeleitet werden.

Tab. 4-1 – Lipinski's Rule of Five. MW = Molekulargewicht, #HBD bzw. #HBA = Anzahl an Wasserstoffbrückendonoren bzw. -akzeptoren, log P = logarithmierter Octanol-Wasser-Verteilungskoeffizient

Lipinski's Rule of Five	
MW	≤ 500 Da
#HBD	≤ 5
#HBA	≤ 10
logP	≤ 5

Rotierbare Bindungen (RB) sind Nicht-Wasserstoff-Bindungen, die sich nicht in einem Ring befinden, und deren Drehung eine Änderung der relativen Atompositionen hervorruft [244]. Eine erhöhte #RB geht mit einer Zunahme der Molekülgröße und einer geringeren Bioverfügbarkeit einher [245], sowie mit einer Erweiterung des Torsionsraums und somit einem erhöhten Rechenaufwand beim Docking [246]. Der Entropieverlust bei einer Protein-Ligand-Komplexbildung nimmt tendenziell mit steigender #RB zu, sodass einer Bindung an die Zielstruktur verstärkt entgegenwirkt wird. In Scoringfunktionen wird ein solcher Rotations-entropieverlust häufig als Strafterm berücksichtigt [247]. Als geeigneter Grenzwert – basierend auf einem Rattenmodell – für eine noch akzeptable Bioverfügbarkeit gilt $\#RB \leq 10$ [245].

Desweiteren spricht eine geringe polare Oberfläche ($PSA \leq 140 \text{ \AA}^2$) eines Moleküls für eine gute Bioverfügbarkeit. Die PSA korreliert mit der Anzahl an Wasserstoffbrückendonoren und -akzeptoren. In der Regel nimmt sie mit steigendem Molekulargewicht zu. [245]

Molekülstrukturen mit reaktiven bzw. toxischen funktionellen Gruppen sollten vorab entfernt werden. Desweiteren kann eine Filterung zugunsten definierter Pharmakophore, funktioneller Gruppen oder Atomtypen sinnvoll sein, für die bekannt ist, dass sie auf jeden Fall für die Wirkung am Zielort erforderlich sind. [239]

Pan Assay Interference Compounds (sog. PAINS), sind Verbindungen, die aufgrund ihrer Reaktivität mit mehreren Targets unselektiv reagieren und in experimentellen Tests daher häufig ein falsch-positives Ergebnis liefern. Diese können mit sog. PAINS-Filtern auf Basis definierter Substrukturen aus Substanzbibliotheken eliminiert werden. [248]

In *Drug-like* (Arzneistoff-ähnlichen)-, *Lead-like* (Leitstruktur-ähnlichen)- und *Fragment-like* (Fragment-ähnlichen)-Filtern gelten – in der Reihenfolge – immer strengere Grenzwerte für die Filterkriterien. Als *Drug-like*-Filter kann die Lipinski's Rule of Five dienen. Leitstrukturen hingegen haben i.d.R. ein geringeres Molekulargewicht, eine geringere Anzahl an RB und Ringen und sind hydrophiler [249]. Als *Lead-like*-Filter wird u.a. die „Rule of 3.5“ ($150 \leq MW \leq 350$, $\log P \leq 4$, $\#HBD \leq 3$, $\#HBA \leq 6$) eingesetzt [250]. Leitstrukturen können aufgrund ihrer geringeren Größe gut durch Anhängen von funktionellen Gruppen an das Grundgerüst in ihrer Bindungsaffinität optimiert und in einen Wirkstoff überführt werden [251]. Für *Fragment-like*-Filter gelten noch strengere Grenzwerte (z.B. „Rule of 2.5“: $MW \leq 250$, $-2 \leq \log P \leq 3$, $\#HBD < 3$, $\#HBA < 6$, $\#RB < 3$) [250]. Sie werden vorbereitend für fragmentbasierte virtuelle Screenings mit sehr kleinen Molekülen

angewendet, welche nur in Teilbereiche einer Target-Bindetasche gedockt werden und für gewöhnlich eine sehr schwache Bindungsaffinität aufweisen. Hits dienen als Ausgangspunkt für eine anschließende strukturbasierte Ligandenoptimierung [252].

Generierung von 3D-Strukturen, Ladungszuständen und Tautomeren

Oftmals sind nur 2D-Informationen von Ligandenstrukturen in einer Datenbank (z.B. im SMILES-Format [253]) verfügbar. Auch die Erstellung von Liganden durch Zeichnen mithilfe von Cheminformatik-Tools wie CHEMDRAW [254] oder dem 2D-SKETCHER (SCHRÖDINGER MAESTRO [255]) erfolgt zunächst im 2D-Format. Als Pendant zum PROTPREP-Tool (s. Kap. 4.2) wurde in dieser Arbeit das LIGPREP-Tool von SCHRÖDINGER zur Vorbereitung der Liganden für das GLIDE-Docking verwendet. Es ermöglicht die Konvertierung zweidimensionaler Strukturen in energetisch günstige 3D-Konformationen auf Basis des OPLS2005- bzw. OPLS3(e)-Kraftfelds (s. Kap. 4.1) zur Optimierung von Bindungslängen und -winkeln [256]. Atomtypen, Bindungsordnungen, Wasserstoffatome und Formalladungen werden mithilfe von LIGPREP zugeordnet [256].

Die Berücksichtigung von Stereoisomeren führt zu vielfach größeren Ligandendatenbanken. Mit dem LIGPREP-Tool kann vorgegeben werden, dass definierte Stereozentren beibehalten werden sollen. Es können aber auch ausgehend von einem Stereoisomer alle möglichen Stereoisomere generiert werden [257].

Verschiedene Protonierungs- und Tautomerisierungszustände (im Folgenden: *States*) eines Moleküls sind für einen Dockingalgorithmus unterschiedliche Substanzen, sodass unterschiedliche Dockingposen resultieren können. Daher ist es sinnvoll, alle vorstellbaren States in einem definierten pH-Bereich zu generieren. Im Allgemeinen wird ein pH-Wert von 7,4 als physiologischer pH-Wert festgelegt. Die Berechnung verschiedener States geschieht im LIGPREP-Tool mit der Anwendung von EPIK [258, 259]. EPIK ist ein Programm zur Berechnung von pK_a -Werten von Molekülen unter Verwendung der Hammett- und Taft-Methodologie. Für jeden Protonierungszustand wird eine *Epik State Penalty* mit der Einheit [kcal/mol] berechnet. Diese Zustandsstrafe beziffert den Energieaufwand, der für die Zustandsgenerierung erforderlich ist. Diese kann zur Bewertung von Dockingposen mit einbezogen werden und ergibt zusammen mit dem GlideScore den Dockingscore (vgl. Kap. 4.4.1, S. 68) [220].

Das in dieser Arbeit verwendete Dockingprogramm GLIDE (s. Kap. 4.4.1) beinhaltet ein Konformationssampling für Liganden, sodass in der Ligandenvorbereitung keine unterschiedlichen Ligandenkonformere berechnet werden müssen.

ZINC-Datenbank

Für diese Arbeit sollten die ausgewählten Substanzen nach dem virtuellen Screening kommerziell erworben werden. Daher wurde die online frei verfügbare ZINC-Datenbank [205, 260–262] (*Zinc is not commercial*) als Ausgangspunkt für die Screeningdatenbank genutzt. Der Vorteil dieser Datenbank ist, dass die meisten Verbindungen aufgrund der laufenden Aktualisierung der dort hinterlegten Herstellerkataloge¹³ jederzeit erhältlich sind. Vor dem Herunterladen der Strukturdaten ist es möglich, Eigenschaftsfilter zu verwenden, um fokussierte Substanzbibliotheken zu erhalten (z.B. *Drug-like*, *Lead-like*, *Clean*). Die Dateien können im SMILES-, mol2-, SDF- und DOCK Flexibase-Format heruntergeladen werden. [260]

Zur Aufnahme von Substanzen aus den Herstellerkatalogen in die ZINC-Datenbank gibt es festgelegte Standardregeln¹⁴ bezüglich der erlaubten Atomelemente, des Molekulargewichts oder des Vorhandenseins reaktiver funktioneller Gruppen. Biologisch aktive Verbindungen werden ohne Prüfung auf reaktive funktionelle Gruppen direkt in die ZINC-Datenbank aufgenommen. Hier werden die Grenzen für definierte Kriterien (z.B. Molekulargewicht, #Stereozentren) weiter ausgedehnt [261].

Innerhalb von ZINC können dann unter Verwendung verschiedener Filter Teilmengen definiert werden. Für diese Arbeit wurde festgelegt, dass nur erwerb- bare Substanzen (*In-Stock*, d.h. sofort verfügbar ohne Wartezeit) mit der Subset-Option *Clean Drug-like* infrage kommen sollten. Der *Clean*-Datensatz enthält noch PAINS, ist aber frei von reaktiven Gruppen, wie z.B. Michaelakzeptoren, Aldehyden oder Thiolen. Mit dem *Drug-like*-Filter werden Deskriptor-Grenzen festgelegt, die der Lipinski's Rule of Five (s. S. 52, Tab. 4-1) entsprechen und zusätzlich die Anzahl an rotierbaren Bindungen (RB) und die polare Oberfläche limitiert (s. Tab. 4-2), sodass man von einer hohen Bioverfügbarkeit der Verbindungen in dieser Teilmenge ausgehen kann. [205, 261]

Tab. 4-2 – Filterregeln für die Erstellung eines „Drug-like“-Datensatzes in der ZINC-Datenbank (Quelle: <http://zinc12.docking.org/browse/subsets/special>, zuletzt aufgerufen am 05.10.21).

Molekulare Eigenschaft	Grenze
Molekulargewicht (MW)	150 - 500 Da
Berechneter log P (xlogP)	≤ 5
Anzahl Wasserstoffbrückendonoren (#HBD)	≤ 5
Anzahl Wasserstoffbrückenakzeptoren (#HBA)	≤ 10
Polare Oberfläche (PSA)	≤ 150 Å ²
Anzahl rotierbarer Bindungen (#RB)	≤ 7

¹³ Über 400 Herstellerkataloge waren im Januar 2021 hinterlegt (zu finden unter <http://zinc15.docking.org/catalogs/>).

¹⁴ Diese sind zu finden unter: http://blaster.docking.org/filtering/rules_default.txt (letzte Abfrage 12.01.2021)

Es wurden die 2D-Daten verwendet, da sie im Rahmen der Ligandenvorbereitung mit LIGPREP ohnehin in 3D überführt werden sollten. Zum Zeitpunkt des Downloads für diese Arbeit im September 2017 handelte es sich nach Anwendung des *Clean-Drug-like*-Filters um 5.142.887 Substanzen¹⁵. Diese schon stark fokussierte Datenbank kann noch nicht ohne Weiteres für das Screening eingesetzt werden. Sie enthält noch verschiedene Protonierungszustände und Tautomere¹⁶ und unter Umständen Duplikate.

Mona

Für die Berechnung molekularer Eigenschaften und für die darauf basierende weitere Filterung der über 5 Millionen Ligandenstrukturen des *Clean-Druglike*-ZINC-Datensatzes vor der Ligandenaufbereitung mit LIGPREP [256] wurde das Cheminformatik-Programm MONA [263, 264] (*Molecular Filtering and Visualization Tool based on Naomi*) verwendet. Die Software basiert auf NAOMI [265], welche das Einlesen und die Konvertierung gängiger Moleküldateiformate (SDF, SMILES, MOL2) ermöglicht. Die Daten in MONA werden in relationalen Datenbanken verarbeitet und können über eine einfache graphische Benutzeroberfläche interaktiv visualisiert, analysiert und verarbeitet werden. [264]

Zu den Funktionen, die in der aktuellen MONA-Version 2.1 eingesetzt werden können, gehören u.a. die Berechnung von physikochemischen Eigenschaften und der Einsatz darauf basierender Filter, die Konvertierung von Molekül-Dateiformaten, die Bereinigung der Datenbank von Duplikaten und die Vereinigung verschiedener Protonierungs- und Tautomerisierungszustände sowie verschiedener Stereoisomere. [263, 264]

Jedes Molekül wird beim Importieren mit einer eindeutigen topologischen Bezeichnung versehen. Verschiedene Topomere eines Moleküls werden zusammengefasst. In dem Moleküleintrag taucht dafür eine Zahl für die Instanzen auf, an der man das Vorkommen von Duplikaten erkennen kann. Es ist nicht möglich, Konformationen zu unterscheiden bzw. die Moleküle in 3D darzustellen. Standardmäßig werden verschiedene Protonierungs- und Tautomerisierungszustände sowie Stereoisomere separat behandelt. Optional können diese auch zusammengefasst werden, welche sich dann auch in der Instanzzahl wiederfinden. Beinhaltet ein Eintrag mehrere nicht-verbundene Komponenten, wird nur die größte Komponente behalten. [264]

¹⁵ Gezählt mithilfe von MONA, s. nächster Abschnitt.

¹⁶ Diese States sollten in dieser Arbeit erst im LIGPREP-Schritt generiert werden.

Die ZINC-Strukturen der *Clean-Drug-like*-Teilmenge wurden in MONA mit der Einstellung *Topology, Charges, Tautomers equal* geladen, sodass Topomere, Tautomere und unterschiedliche Protonierungszustände zusammengefasst wurden. Mithilfe von MONA wurde die Ligandenbibliothek auf Basis eigener Filterkriterien weiter reduziert (s. Kap. 5.2.3, S. 157).

4.3.1 VS-Testdatensatz und Decoygenerierung (DUD-E)

Das virtuelle Screening (VS) soll idealerweise dazu in der Lage sein, aktive von inaktiven Verbindungen zu trennen. Um diese Fähigkeit zu überprüfen, setzt man Testdatensätze (auch: Benchmarking-Datensätze) bestehend aus aktiven und inaktiven Verbindungen ein. Diese können jeweils separat oder gemeinsam dem VS-Prozess unterworfen werden. [266]

Ähnlichkeitsclustering von Aktiven

Die Auswahl von Aktiven sollte so erfolgen, dass eine möglichst große Diversität zwischen den Strukturen besteht, da eine starke Ähnlichkeit der Aktiven untereinander gegenüber einer breiteren strukturellen Vielfalt von inaktiven Verbindungen zu einer Überschätzung der Anreicherung führt (Analog-Bias) [267]. In dieser Arbeit sollte das Problem durch ein Ähnlichkeitsclustering der Aktiven und die Verwendung der Clusterrepräsentanten minimiert werden (s. Kap. 5.2.2, S. 156, Ligandenliste c).

Decoygenerierung

Für den Datensatz inaktiver Substanzen wurden in dieser Arbeit sogenannte *Decoys* (dt. Köder) eingesetzt. Dabei handelt es sich um Moleküle, die mutmaßlich nicht an die Zielstruktur binden [266]. Decoys sollten den Aktiven in ihren physikochemischen Eigenschaften ähneln. Bei einem größeren Molekulargewicht von Aktiven im Vergleich zu den Decoys kann es dazu kommen, dass die größeren Aktiven fälschlicherweise wegen der durch ihre Größe höheren Anzahl an Interaktionsmöglichkeiten mit der Bindetasche von der Scoringfunktion besser bewertet werden als die kleineren Decoys. Im umgekehrten Fall würden die Decoys eine zu gute Bewertung erhalten. In beiden Fällen handelt es sich um eine Verzerrung, die zu einer falschen Interpretation der Anreicherung führen würde [266].

Für die Decoygenerierung wurde die Online-Plattform¹⁷ der DUD-E (*Directory of Useful Decoys, Enhanced*) -Datenbank [268] verwendet, mit der es möglich ist, ausgehend von eigenen aktiven Liganden nach einem festgelegten Protokoll Decoys zu erstellen.

¹⁷ <http://dude.docking.org/generate> (zuletzt aufgerufen am 13.01.2021)

Die DUD-E-Datenbank hält vorbereitete Benchmark-Datensätze aus Aktiven und Decoys für die Evaluierung von Docking- und Screeningexperimenten bereit. Mit über 1,4 Millionen Decoys generiert auf der Basis von 22.886 Liganden mit Affinitäten gegenüber 102 Targets ist DUD-E eine der größten Datenbanken für Decoys [269]. Die Aktiven der in DUD-E hinterlegten Datensätze wurden aus der ChEMBL09 [270]-Datenbank ab einem Aktivitäts- bzw. Affinitätswert von $\leq 1 \mu\text{M}$ extrahiert und zur Erhöhung der Diversität geclustert, während die Decoys aus der ZINC-Datenbank [260] unter der Voraussetzung aufgenommen wurden, dass sie bis zu einer Konzentration von $30 \mu\text{M}$ keine Affinität besaßen.

Der Anwender kann online auf der DUD-E-Webseite mit einem/mehreren eigenen bereitgestellten aktiven Liganden Decoys erzeugen. Die Decoys werden in DUD-E so generiert, dass sie den Aktiven bezüglich ihrer physikochemischen Eigenschaften möglichst ähnlich sind, aber topologisch (strukturell) so verschieden, dass sie nicht am Target binden sollten. Dafür werden die vier Parameter der Lipinski's Rule of Five zuzüglich der Anzahl der rotierbaren Bindungen und der Nettoladung bei der Decoy-Aufnahme berücksichtigt. Für die Garantie der topologischen Unähnlichkeit zu Aktiven wird die Tanimoto-Distanz basierend auf 2D-Fingerprints (ECFP4¹⁸) im Vergleich zu den Aktiven berechnet und die 25 % unähnlichsten Decoys werden behalten. Daraus werden für jeden Protonierungszustand¹⁹ eines aktiven Liganden schließlich 50 Decoys zufällig selektioniert. [268]

¹⁸ ECFP4 = *Extended-Connectivity Fingerprints*: zirkuläre topologische Fingerprints bis zu vier Bindungen [271].

¹⁹ Die Erzeugung wahrscheinlicher Protonierungszustände erfolgt im DUD-E-Protokoll mit SCHRÖDINGER'S EPIK [258] in einem pH-Bereich von 6-8.

4.4 Docking

Unter *Docking* (engl. für Andocken bzw. Ankoppeln) versteht man den Versuch der Platzierung virtueller dreidimensionaler Strukturen (meist Liganden) mithilfe eines Computerprogramms (*in silico*) in eine Zielstruktur (*Target*, Protein) [183].

Das Docking lässt sich in zwei Schritte unterteilen: Zum einen untersucht ein Algorithmus den konformativen Raum des Liganden (Konformationsgenerierung) und platziert ihn in die Binde-tasche des Targets. Zum anderen wird die Protein-Ligand-Komplementarität anhand aller Inter-aktionen zwischen den Bindungspartnern mit einer Energiefunktion bewertet (*Scoring*). Die energetische Bewertung einer erfolgten Platzierung (*Pose*) wird mit einem Score-Wert zum Aus-druck gebracht, mit dem die Bindungsaffinität zum Protein abgeschätzt werden kann. Mithilfe dieser Score-Werte lassen sich Dockingposen in eine Rangordnung (*Ranking*) überführen. Mitun-ter werden mehrere Posen desselben Liganden durch das Dockingprogramm angeboten. Ener-getisch oder sterisch ungünstige Dockinglösungen, z.B. bedingt durch zu geringe Abstände zwi-schen Atomen (*Clashes*) oder elektrostatische Abstoßungen, können durch das Programm wie-der verworfen werden. Die erschöpfende Untersuchung der Beweglichkeit der Bindungspartner (*Sampling*) erfolgt meist nur am Liganden, da ein Sampling des Proteins wegen der großen An-zahl an drehbaren Bindungen zu rechenintensiv wäre. [272]

Herausforderungen

Da die Simulation der Proteinbeweglichkeit aufgrund der großen Anzahl an Bindungen beson-ders anspruchsvoll und rechenintensiv ist, wird sie bei Dockingberechnungen, insbesondere beim virtuellen Screening, in der Regel vernachlässigt [204]. Es gibt aber verschiedene Lösungs-ansätze. Beim Induced-Fit-Docking werden z.B. die vdW-Radien von Rezeptor- und Ligandenatomen abgeschwächt, definierte Proteinseitenketten vorübergehend durch Alaninsei-tenketten ersetzt oder die Proteinseitenkettenkonformationen variiert [273–277]. Eine weitere Möglichkeit ist das Docking in ein Ensemble aus mehreren starren Proteinkonformationen (z.B. in FLEXE [278]). Solche multiplen Konformationen lassen sich z.B. mithilfe von *Molecular-Dyna-mics*-Simulationen (s. Kap. 4.5) generieren [279–281].

Um ein strukturbasiertes virtuelles Screening effektiv durchführen zu können, müssen zwei ele-mentare Herausforderungen des Dockings geklärt werden. Die erste Schwierigkeit besteht in der korrekten Positionierung eines aktiven Liganden innerhalb der Bindetasche (*Docking-Problem*).

Wenn die wahre Pose eines Liganden nicht erzeugt werden kann, obwohl die Bindung geometrisch und chemisch möglich ist, könnte es zum Verlust dieser Substanz im Rahmen eines virtuellen Screenings kommen. Nach dem Positionieren muss die korrekte Pose als solche unter zahlreichen möglichen Dockinglösungen erkannt werden (*Scoring-Problem*). Hierin liegt die größere Herausforderung, da noch keine Scoringfunktion alle Anforderungen an Genauigkeit und Geschwindigkeit der Bindungsaffinitätsvorhersage erfüllt. So kommt es häufig zu falsch positiven Treffern in den Ergebnissen eines virtuellen Screenings. Aktive Verbindungen könnten wiederum durch die Bewertungsfunktion einen zu niedrigen Score-Wert erhalten. Gründe für dieses Scoring-Problem liegen hauptsächlich in der fehlenden energetischen Berücksichtigung der Proteinbeweglichkeit und der damit zusammenhängenden Anpassung des Liganden in der Bindetasche. Darüber hinaus legen die meisten Scoringfunktionen aufgrund der leichteren Abschätzung ein stärkeres Gewicht auf enthalpische als auf entropische Effekte. Solvatisierungseffekte und Änderungen in der Entropie sind schwierig vorherzusagen. [282, 283]

Die Stärken und Schwächen der verschiedenen Arten von Scoringfunktionen werden ab S. 61 beschrieben.

Suchalgorithmen

Die Dockingprogramme verwenden verschiedene Dockingalgorithmen für die Einpassung der Liganden in die Bindetasche und somit zur Lösung des Docking-Problems. Dockingalgorithmen behandeln entweder beide Bindungspartner als starr (rigides Docking), nur den Liganden flexibel (semi-flexibles Docking) oder beide Partner flexibel (flexibles Docking). Es gibt drei grundlegende Arten von Suchalgorithmen, die auch in einer hierarchischen Abfolge (z.B. Kombination von systematischen und stochastischen Algorithmen in GLIDE [284]) miteinander verknüpft werden können: [272, 285]

1. Deterministische Algorithmen: Die gleiche Ausgangskonformation eines Liganden führt immer zum gleichen Ergebnis (generierte Konformation), d.h. es gibt keine zufälligen Komponenten. Zu den deterministischen Methoden (auch: Simulationsmethoden) gehören z.B. die Molekulardynamik und die Energieminimierung.

2. Systematische Algorithmen: Die veränderbaren Eigenschaften der Konformation des Liganden (auch bezeichnet als *Freiheitsgrade*, z.B. Torsionswinkel, Drehung von Bindungen) werden in vordefinierten Abständen variiert. Auf diese Weise wird der Konformationsraum

sehr umfassend abgetastet. Für wenig komplexe Moleküle sind diese Algorithmen leistungsstark. Problematisch kann eine mögliche kombinatorische „Explosion“ bei einer zu feinen Diskretisierung der Freiheitsgrade und bei großen Molekülen, z.B. Peptiden oder Proteinen, sein. Zu den systematischen Algorithmen gehört z.B. die inkrementelle Konstruktion (z.B. FLEXX [286]). Hierbei wird der Ligand in Fragmente zerlegt, welche gesondert in die Bindetasche gedockt werden. Daraufhin verbleibt das Fragment mit der besten Bewertung als Basisfragment in der Bindetasche und die restlichen Fragmente werden sukzessive wieder angebaut. Parallel dazu ermittelt eine Bewertungsfunktion die besten Fragmentplatzierungen. Eine weitere systematische Methode ist die erschöpfende Suche (*exhaustive search*): Um eine kombinatorische Explosion auszuschließen, wird zuerst der abzusuchende Raum eingeschränkt. Die Konformationen werden durch systematisches Drehen aller rotierbaren Bindungen des Moleküls in einem definierten Intervall erzeugt. Das Dockingprogramm GLIDE [284] nutzt eine solche erschöpfende Suche als Teil seines hierarchischen Filters.

3. Stochastische Algorithmen: Ausgehend von einer Startkonformation werden die Freiheitsgrade des Liganden zufällig variiert. Jeder Änderung wird mit einer im Algorithmus festgelegten Wahrscheinlichkeitsfunktion bewertet. Zu den wichtigsten stochastischen Methoden gehören sowohl das Monte-Carlo-Verfahren unter Verwendung der Boltzmann-Wahrscheinlichkeitsfunktion (z.B. als Teil des hierarchischen Filters als letzter Minimierungsschritt in GLIDE) als auch genetische Algorithmen (z.B. GOLD [287], AUTODOCK [288]).

Scoringfunktionen

Wenn nun durch den Suchalgorithmus ein korrekter Bindungsmodus eines aktiven Liganden vorhergesagt wurde, ist dies erst als Erfolg sichtbar, wenn die Bindungsenergie dieser Pose durch eine differenzierende Scoringfunktion realitätsnah abgeschätzt wird. Hinsichtlich der Posenbewertung gibt es nicht DIE eine Scoring-Funktion, welche für alle Zwecke geeignet ist, da keine Funktion als vollkommen betrachtet werden kann. Hier ist es wichtig, eine geeignete Bewertungsfunktion für ein vorgesehene Dockingprojekt herauszuarbeiten. Ziel ist eine möglichst realistische Einschätzung der Bindungsaffinität des Liganden zur Zielstruktur und bestenfalls eine Unterscheidung von bindenden und nicht-bindenden Molekülen. Die Genauigkeit der Energiebewertung nimmt mit der Anzahl an berücksichtigten Parametern in der Scoringfunktion zu, jedoch steigt auch die Rechenintensität mit steigender Variablenanzahl. Dadurch werden Geschwindigkeit und Effizienz bei der Bewertung einer großen Anzahl an Posen, wie sie beim vir-

tuellen Screening großer Substanzbibliotheken mit überlicherweise Millionen Strukturen entstehen, in Mitleidenschaft gezogen. Eine gute Energiefunktion muss daher immer einen Kompromiss zwischen Schnelligkeit und Genauigkeit bieten. Weit verbreitet ist die Einteilung von Scoring-Funktionen in drei Klassen: [289–292]

1. Kraftfeldbasiert: Die Wechselwirkungsenergien zwischen den Bindungspartnern als auch die innere Energie des Liganden werden unter Verwendung klassischer Potenzialterme aus der Molekülmechanik (z.B. Coulomb-, Lennard-Jones-Potenzial) berechnet und aufsummiert. Kraftfeldbasierte Scoringfunktionen sind z.B. in GOLD und AutoDock zu finden. Defizite gibt es bezüglich der Einbeziehung entropischer Effekte und bei der Berücksichtigung von Lösungsmittelleffekten. Allerdings gibt es erweiterte Scoringfunktionen mit einem abstandsabhängigen Dielektrikum zur Simulation des Lösungsmittelleffekts.

2. Empirisch: Ausgangspunkt für die Erstellung empirischer Scoringfunktionen ist ein Trainingsdatensatz aus Protein-Ligand-Komplexen (z.B. in Form von hochauflösenden Kristallstrukturen) mit bekannten Bindungsenergien zur Durchführung einer Regressionsanalyse. Daraus werden die Koeffizienten zur Gewichtung einzelner Wechselwirkungsterme erhalten. Jeder dieser Terme beschreibt eine nichtkovalente Wechselwirkung, die zur Bildung des Protein-Liganden-Komplexes beiträgt (sog. *Deskriptor*), z.B. Wasserstoffbrückenbindungen, polare oder der vdW-Wechselwirkungen. Darüber hinaus sind Terme für die Desolvatisierung und für entropische Effekte enthalten. Das Ergebnis ist sehr stark von den verwendeten Trainingsdaten abhängig. Diese Energieterme sind stark vereinfacht im Vergleich zu kraftfeldbasierten Funktionen und liefern daher eine schnellere Bewertung der Posen. Beispiele für empirische Methoden sind die LUDI-Scoringfunktion [293], FLEXX [294] und der GlideScore [284].

3. Wissensbasiert: Für die Vorhersage der Bindungsenergie wird eine Datenbank von Protein-Ligand-Komplexen bezüglich der vorhandenen Kontakte verschiedener Atomtypen statistisch ausgewertet. Hierbei steht die Annahme im Vordergrund, dass statistisch häufig auftretende Kontakte zwischen Atomtypen energetisch vorteilhaft sind und so verstärkt zur Bindungsaffinität beitragen. Mittels einer Wahrscheinlichkeitsverteilung (z.B. der Boltzmann-Statistik) werden aus der Häufigkeitsverteilung der verschiedenen Kontakte Wechselwirkungspotenziale zwischen Atompaaren abgeleitet. Im Vergleich zur kraftfeldbasierten und empirischen Methode werden Genauigkeit und Geschwindigkeit mit der wissensbasierten Methode (z.B. DrugScore [295, 296], PMF [297, 298]) am besten in Einklang gebracht.

Eine Weiterentwicklung dieser Ansätze stellt die Kombination verschiedener Scores zur Schätzung der Bindungsaffinität dar (*Consensus Scoring*), wodurch Unzulänglichkeiten einzelner Scoringfunktionen ausgeglichen und die Anreicherung tatsächlich interagierender Liganden verbessert werden sollen. [247, 299, 300]

4.4.1 Glide

GLIDE (*Grid-based Ligand Docking with Energetics*) [284, 301, 302], integriert als Modul in der SCHRÖDINGER-Software, wurde in dieser Arbeit für alle Dockingexperimente angewendet. Es handelt sich hierbei um ein Dockingtool, welches in hoher Geschwindigkeit in hierarchischer Weise die räumliche Anpassung des Liganden an die Bindetasche vornimmt und bewertet. [284]

Vor dem Docking mit GLIDE müssen Protein(e) und Ligand(en) in eine ordnungsgemäße dreidimensionale Struktur überführt werden (s. Kap. 4.2 und 4.3). Der Softwarehersteller SCHRÖDINGER, LLC empfiehlt für diesen Zweck die Anwendungen PROTPREP für die Proteinvorereitung und LIGPREP für die Ligandenvorbereitung vor Durchführung des GLIDE-Dockings.

Grid-Generierung

Mit der unter Anwendung eines Kraftfelds (s. Kap. 4.1) vorbereiteten Proteinstruktur kann die Erstellung des Grids durchgeführt werden. Mit dem „RECEPTOR GRID GENERATION“-Panel von GLIDE muss zunächst ein dreidimensionales *Grid* (dt. Gitter) in der Bindetasche der Zielstruktur erzeugt werden. An jedem Gitterpunkt werden die Wechselwirkungsenergien zu Proteinatomen in der Umgebung mithilfe von Sonden definierten

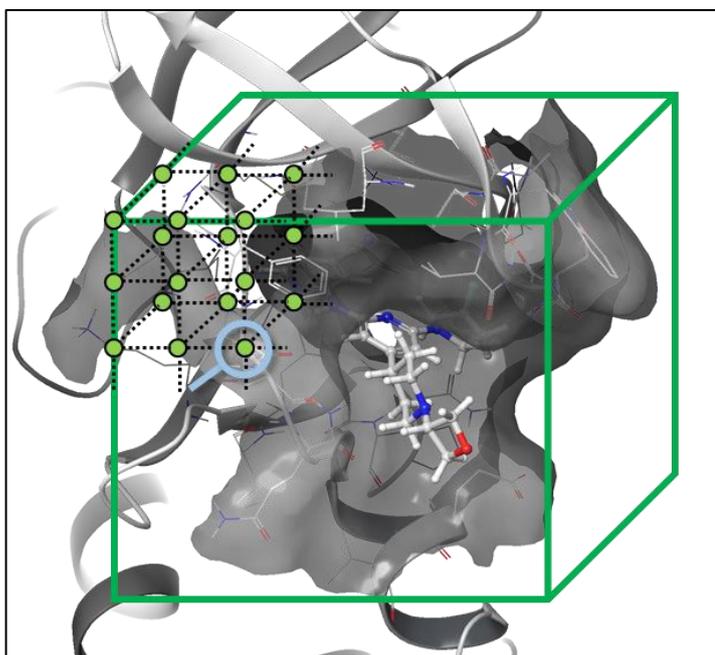


Abb. 4-3 – Vereinfachte Darstellung der Erstellung eines Grids in einer Bindetasche. Es wird eine Box definierter Größe um einen Bezugsliganden bzw. einen definierten Punkt in der Bindetasche gelegt und in äquidistante Gitterpunkte (hellgrün) aufgeteilt. An jedem Gitterpunkt werden die Wechselwirkungsenergien berechnet (dargestellt mit der Lupe) und in dem Grid gespeichert.

diesen (s. Abb. 4-3). Diese Daten werden in der Grid-Datei gespeichert und während des Dockingprozesses automatisch für jedes Atom eines platzierten Liganden abgerufen basierend auf dessen definierter Position im Grid. Bindungsenergien können so schneller erhalten werden, indem

die relevanten gespeicherten Daten bezüglich der Wechselwirkungsmöglichkeiten des Rezeptors berücksichtigt werden. [303] Zudem ist ein genaueres Scoring der Ligandenposen möglich [304].

Bei Anwesenheit eines Liganden in der Bindetasche lässt sich der Schwerpunkt des Liganden als Mittelpunkt des Grids definieren. Ohne einen komplexierten Liganden lässt sich ein Bezugspunkt auf Basis auszuwählender Aminosäuren der Bindetasche oder durch die Eingabe definierter Koordinaten festlegen. Die Größe des Grids kann an die Größe des komplexierten Liganden angepasst werden.

Durch eine Skalierung der vdW-Radien von unpolaren Proteinatomen lässt sich das Potenzial solcher Regionen abschwächen und somit die Wahrscheinlichkeit des Auftretens enger Kontakte erhöhen. Mit dieser Option lässt sich ein gewisses „Nachgeben“ der Bindetasche und somit ein Induced-Fit-ähnlicher Effekt erzeugen.

Definierte Anforderungen an eine Ligand-Protein-Interaktion (sog. *Constraints*) werden bei der Grid-Erstellung bereits gespeichert, um sie im Falle eines Dockings gezielt auszuwählen. So kann man mit Positions-Constraints Regionen in der Bindetasche festlegen, die beim Docking besetzt werden müssen. Proteinatome, welche an Wasserstoffbrückenbindungen beteiligt sein sollen (*Hbond-Constraints*), können ausgewählt werden. Mit *Metal-Constraints* werden erwünschte Ligand-Metall-Interaktionen vorgegeben. Darüber hinaus kann man Regionen festlegen, die während des Dockings nicht besetzt werden dürfen (sog. *excluded volumes*).

Rotationen von Hydroxylgruppen in Serin, Threonin und Tyrosin, sowie der Thiolgruppe in Cystein innerhalb der Bindetasche können zugelassen werden, um möglichst angepasste Wechselwirkungen mit dem Liganden zu realisieren.

Die gewisse Flexibilität solcher Gruppen lässt sich, einmal im Grid gespeichert, im Gegensatz zu den Constraints nicht wieder deaktivieren und verursacht einen etwas größeren Rechenaufwand beim Docking. Dieser ist aber deutlich geringer, als wenn man für jede Möglichkeit dieser Hydroxyl- und Thiolgruppenorientierungen ein eigenes Grid erstellen und in jedes dieser Grids docken würde.

Dockingmodi

GLIDE bietet drei Präzisionsstufen des Dockings an: den HTvS- (*High Throughput virtual Screening*), SP- (*Standard Precision*) und XP (*Extra Precision*)-Modus. [304]

Der HTvS-Modus ist die schnellste Methode und kann eine größere Datenbank virtueller Strukturen mit einer Geschwindigkeit von etwa 2 Sekunden pro Ligand [305] docken. Das SP-Docking verwendet die gleiche Bewertungsfunktion wie das HTvS-Docking, führt aber eine umfassendere Suche im Konformationsraum zur Ligandenkonformationsgenerierung durch und ist daher genauer, aber auch langsamer als der HTvS-Modus (etwa 10 Sekunden pro Ligand [306]). Mit dem SP-Modus können zehn- bis hunderttausende Liganden gedockt werden [307]. Wegen des ausgewogenen Verhältnisses zwischen Genauigkeit und Schnelligkeit [284] eignet sich das SP-Docking z.B. für eine Leitstruktur-Identifizierung aus großen Datensätzen. Um die Anzahl falsch positiver Treffer zu reduzieren und die Korrelation zwischen der experimentellen und berechneten Bindungsaffinität zu erhöhen bietet GLIDE zudem den XP (*Extra Precision*)-Dockingmodus an. Dieser tastet zum einen unter Verwendung eines Anchor-and-Grow-Algorithmus die Bewegungsmöglichkeiten des Liganden gründlicher ab (verbessertes Sampling). Zum anderen verwendet der XP-Modus eine verbesserte, strengere Scoringfunktion als der HTvS- und SP-Modus mit höheren Ansprüchen an die Komplementarität zwischen Liganden und Bindetasche [302]. Wegen der deutlich höheren erforderlichen Rechenzeit (etwa 2 Minuten pro Ligand [305]) wird empfohlen, nur Liganden mit der besten Bewertung im XP-Modus zu docken, die aus einem virtuellen Screening im HTvS- oder SP-Modus hervorgegangen sind [304].

Während der HTvS- und SP-Modus häufig in Virtual-Screening-Experimenten für die Suche nach Hits eingesetzt werden, stellt der XP-Modus eine geeignete Methode für die Leitstrukturoptimierung dar. [284, 307]

Hierarchischer Dockingprozess

Zur Lösung des *Dockingproblems* nutzt GLIDE eine Reihe von hierarchischen Filterprozessen mit zunehmender Präzision.

Vor dem eigentlichen Docking müssen Grids, wie zuvor beschrieben, auf Basis der Bindetasche generiert werden, welche die Geometrien und Wechselwirkungsmöglichkeiten der Proteinbindetaschen widerspiegeln. Die Gridberechnung ist für jeden Rezeptor nur ein Mal nötig. Für die Generierung der Ligandenkonformationen wird eine nahezu erschöpfende Suche im Torsionsraum durchgeführt. Darauf folgt der Prozess des Dockings, den man aufgrund der schrittweisen Zunahme der Genauigkeit und des Aussortierens von Ligandenkonformeren als Trichter (s. Abb. 4-4) darstellen kann. [284]

Die Beschreibung des Dockingprozesses beruht auf der Veröffentlichung von FRIESNER ET AL. [284]:

1 – Site-point search: Zuerst erfolgt ein erstes grobes Einpassen der Liganden (flexibel) in ein 2 Å-Raster der Bindungsstelle auf Basis von Bindestellenpunkten. Dies geschieht durch einen Abgleich der Abstände zwischen Bindestellenpunkt und Rezeptoroberfläche mit den Abständen eines definierten Ligandenzentrums und der Ligandenoberfläche.

2 a – Diameter test: Im zweiten Schritt erfolgt die Prüfung auf sterische Kollisionen von Atomen nahe der Ligandenachse²⁰ (*diameter*) je nach deren Ausrichtung.

2 b – Subset test: Daraufhin erfolgt die Bewertung von Wasserstoffbrückenbindungen und Ligand-Metall-Interaktionen durch Drehen des Liganden um die Ligandenachse.

2 c – Greedy score: Bei ausreichend guter Beurteilung dieser Kontakte werden alle übrigen Wechselwirkungen ebenfalls bewertet. Die Bewertung dieser einzelnen Einpassungsschritte (2 a-c) werden mit dem *Greedy Score* vorgenommen, welcher eine Version des ChemScores, einer empirischen Bewertungsfunktion, darstellt, und ebenso die hydrophoben Wechselwirkungen, Wasserstoffbrückenbindungen und Metall-Ligand-Wechselwirkungen erkennt und sterische Kollisionen sanktioniert. Der *greedy-scoring*-Algorithmus berechnet das Interaktionspotenzial jedes Ligandenatoms an seiner Position und den bestmöglichen Scorewert durch Bewegung des entsprechenden Atoms um +/- 1 Å in x-, y-, und z-Richtung.

2 d – Refinement: Der Ligand wird dann im Ganzen starr um +/- 1 Å in alle drei Richtungen bewegt und nochmals bewertet.

3 – Grid minimization + Monte Carlo: Eine kleine Anzahl (ca. 100-400) der besten Posen im Komplex mit dem Protein werden einer Minimierung im OPLS-AA-Kraftfeld [210] in Kombination mit einem entfernungsabhängigen Ladungsmodell unterzogen. Im HTvS-Modus erfolgt diese Minimierung im OPLS2005-Kraftfeld, während im SP- und XP-Modus standardmäßig das OPLS3-Kraftfeld verwendet wird (s. Kap. 4.1). Eine kleine Anzahl (ca. 3-6) der energetisch günstigsten

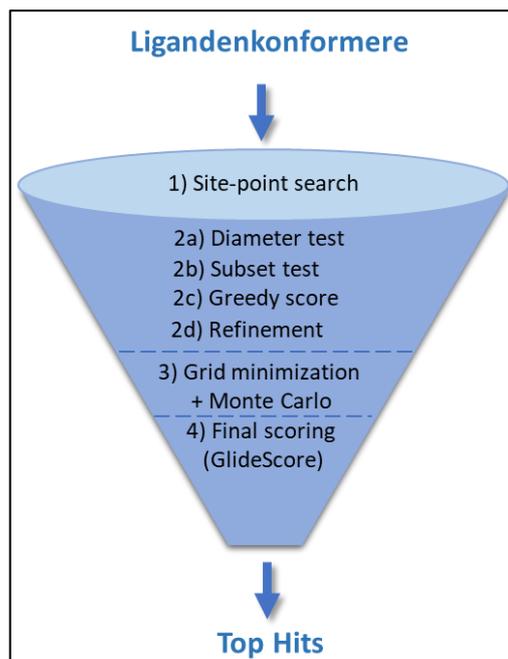


Abb. 4-4 – Darstellung der Hierarchie des GLIDE-Dockingalgorithmus. Abb. modifiziert nach FRIESNER ET AL. (2004) [284]

²⁰ definiert als die Linie durch die am weitesten voneinander entfernten Atome

Posen auf Basis des Greedy Scores werden mithilfe von *Monte-Carlo-Simulationen* hinsichtlich ihrer Torsionswinkel, und somit in ihrem Energiewert, verbessert.

4 – Final scoring: Zum Schluss werden die minimierten Posen mit dem GlideScore zur Abschätzung der freien Bindungsenergie bewertet und sortiert. Mit dem *Emodel*-Wert lassen sich die verschiedenen gedockten Konformere eines Liganden bewerten.

Das umfangreichere XP-Docking startet zunächst in gleicher Weise, nur dass der hier beschriebene Trichter (s. S. 66, Abb. 4-4) breiter angelegt ist, d.h. es werden anfänglich mehr gedockte Strukturen generiert. In dem anschließend angewandten *Anchor-and-Growth*-Verfahren werden Ankerfragmente (meist Ringe) des gedockten Liganden aus den verschiedenen Posen ausgewählt und die verschiedenen Posen des gleichen Ankerfragments geclustert. Ausgehend von jedem Clusterrepräsentant wird der Ligand nach und nach wieder aufgebaut und die resultierenden Posen werden vollständig minimiert und dem XP-Scoring unterzogen. Dieser Prozess des Ankerwachstums erfolgt in mehreren Zyklen mit dem Ziel, möglichst gut bewertete Posen zu erhalten. Die Bewertung erfolgt mit dem XP-GlideScore. [302, 304]

Scoringfunktionen

Der zur Bewertung und Sortierung von HTvS- und SP-Dockinglösungen in GLIDE herangezogene GlideScore²¹ basiert auf der empirischen Scoringfunktion **ChemScore** (3), der sich aus vier verschiedenen Wechselwirkungstermen zur Abschätzung der Bindungsenergie eines Protein-Ligand-Komplexes zusammensetzt: [308]

$$\Delta G_{\text{binding}} = \Delta G_0 + \Delta G_{\text{hbond}} \sum_{i,l} g_1(\Delta r) g_2(\Delta \alpha) + \Delta G_{\text{metal}} \sum_{a,m} f(r_{aM}) + \Delta G_{\text{lipo}} \sum_{i,l} f(r_{iL}) + \Delta G_{\text{rot}} H_{\text{rot}} \quad (3)$$

Die jeweiligen Koeffizienten (ΔG) der einzelnen Kontaktterme wurden von ELDRIDGE ET AL. [308] mit einem Trainingsdatensatz aus 82 Protein-Ligand-Komplexen mit bekannten Bindungsaffinitäten durch Regressionsanalysen aus experimentellen Daten (empirisch) ermittelt. Die Formel berücksichtigt:

- $\Delta G_{\text{hbond}} \sum_{i,l} g_1(\Delta r) g_2(\Delta \alpha)$: die Energie der Wasserstoffbrückenbindungen zwischen Ligand(*i*)- und Proteinatomen (*l*) auf Grundlage der Abweichung des Winkels α vom idealen Winkel 180° und des Abstands *r* vom idealen Abstand von $1,85 \text{ \AA}$ für Wasserstoffbrückenbindungen; auch wasservermittelte Kontakte werden berücksichtigt, allerdings ohne Differenzierung zwischen geladenen und ungeladenen Interaktionen

²¹ im Folgenden bezeichnet als SP-GlideScore zur Unterscheidung vom XP-GlideScore

- $\Delta G_{\text{metal}} \sum_{aM} f(r_{aM})$: die Energie der Ligand(a)-Metall(M)-Interaktionen abhängig vom Abstand r
- $\Delta G_{\text{lipo}} \sum_{iL} f(r_{iL})$: die Energie der hydrophoben Wechselwirkungen zwischen lipophilen Ligandenatomen (i) und lipophilen Proteinatomen (L) abhängig vom Abstand (r) zwischen diesen Atomen
- $\Delta G_{\text{rot}} H_{\text{rot}}$: die Rotationsenergie resultierend aus der eingeschränkten Flexibilität des gebundenen Liganden (Flexibilitätsstrafe)

Der **SP-GlideScore** (4) stellt eine Erweiterung des ChemScores dar und berechnet die Protein-Ligand-Wechselwirkungen mit folgender Gleichung: [284]

$$\begin{aligned} \Delta G_{\text{bind}} = & C_{\text{lipo-lipo}} \sum f(r_{lr}) + C_{\text{hbond-neut-neut}} \sum g(\Delta r)h(\Delta \alpha) + \\ & C_{\text{hbond-neut-charged}} \sum g(\Delta r)h(\Delta \alpha) + \\ & C_{\text{hbond-charged-charged}} \sum g(\Delta r)h(\Delta \alpha) + C_{\text{max-metal-ion}} \sum f(r_{lm}) + \\ & C_{\text{rotb}} H_{\text{rotb}} + C_{\text{polar-phob}} V_{\text{polar-phob}} + C_{\text{coul}} E_{\text{coul}} + C_{\text{vdW}} E_{\text{vdW}} + \\ & \text{solvation terms} \end{aligned} \quad (4)$$

Die empirisch zu bestimmenden Koeffizienten werden hier mit C bezeichnet. Im ersten Term des SP-GlideScores werden lipophile Kontakte wie auch im ChemScore bewertet. Der Term für die Wasserstoffbrückenbindungen wird hier in drei Terme aufgeteilt (Terme 2-4), je nach vorhandener Ladung am Akzeptor oder Donor. Bei der Berechnung von Ligand-Metall-Interaktionen (Term 5) werden nur die Interaktionen mit Akzeptor-Anionen toleriert. Bei Vorhandensein mehrerer Metall-Ligationsmöglichkeiten wird die bestbewertete Metall-Anion-Ligation bevorzugt. Die Funktionen f , g und h in Gleichung (4) ergeben eine volle Punktzahl (1,00), wenn die Abstände oder Winkel innerhalb vorgegebener Grenzen liegen. Außerhalb dieser Grenzen, aber innerhalb größerer Schwellenwerte, ergeben die Funktionen eine Teilpunktzahl (0,00-1,00). Die Rotationsenergie (Term 6) wird wie im ChemScore auch berücksichtigt. Term 7 belohnt Wechselwirkungen von polaren, nicht wasserstoffbrückenbildenden Atomen in einer hydrophoben Umgebung. Darüber hinaus spielen Coulombsche (Term 8) und vdW-Kräfte (Term 9) zwischen Liganden und Protein eine wichtige Rolle bei der Berechnung des SP-GlideScores. Auch Terme für Solvatisierungseffekte auf empirischer Basis werden eingesetzt. [284]

Der **Dockingscore** beinhaltet zusätzlich zum SP-GlideScore die *Epik State Penalty*. EPIK [258, 259] als Anwendung der SCHRÖDINGER SUITE berechnet als Bestandteil der Ligandenvorbereitung die wahrscheinlichen Ionisierungs- und Tautomerzustände von Molekülen innerhalb eines definierten pH-Bereichs (s. Kap. 4.3). Für jeden dieser Zustandsübergänge wird der Energieaufwand (*Epik State Penalty*) berechnet. Durch Einbeziehung dieser energetischen Strafe kann der Dockingscore für manche Dockingposen weniger negativ sein als der SP-GlideScore. Die Anreicherung aktiver

Substanzen soll durch Berücksichtigung der energetischen Strafe in der Scoringfunktion deutlich verbessert werden. [259]

Der **Emodel-Wert** wird zur Bewertung verschiedener Posen desselben Liganden herangezogen und sollte nicht für den Vergleich verschiedener Liganden verwendet werden. Er besteht aus dem SP-GlideScore, der inneren Energie²² (E_{internal}) des Ligandenkonformers und einem besonderen Term für Coulomb- und vdW-Interaktionen (*CvdW energy*). [304]

Die **Glide Ligand Efficiency** (LE) ist der Dockingscore dividiert durch die Anzahl der schweren Atome. [304] Somit ist der Wert abhängig von der Molekülgröße. Ein kleiner Ligand mit einer für seine Größe hohe Bindungsaffinität (bzw. mit einem guten Dockingscore) ist für eine nachträgliche Optimierung besser geeignet als ein großer Ligand mit vergleichbarem Dockingscore. Der Wert sollte daher möglichst niedrig (negativ) sein.

Der **XP-GlideScore** (5) zur Bewertung der XP-Dockinglösungen bestraft im Vergleich zum SP-GlideScore Desolvatationseffekte und erkennt strukturelle Motive, die einen großen Beitrag zur Affinität beisteuern. Er setzt sich neben den Coulomb'schen (E_{coul}) und vdW-Wechselwirkungsenergien (E_{vdW}) aus einem bindungsaffinitätssteigernden (E_{bind}) und einem bindungsaffinitätsreduzierenden (E_{penalty}) Term zusammen: [302]

$$\text{XP GlideScore} = E_{\text{coul}} + E_{\text{vdW}} + E_{\text{bind}} + E_{\text{penalty}} \quad (5)$$

$$E_{\text{bind}} = E_{\text{hyd_enclosure}} + E_{\text{hb_nn_motif}} + E_{\text{hb_cc_motif}} + E_{\text{PI}} + E_{\text{hb_pair}} + E_{\text{phobic_pair}} \quad (6)$$

$$E_{\text{penalty}} = E_{\text{desolv}} + E_{\text{ligand_strain}} \quad (7)$$

Der XP-GlideScore belohnt in dem Term E_{bind} (6)

- den Einschluss hydrophober Ligandenatome von zwei Seiten durch hydrophobe Proteinatome (*hydrophobic enclosure*, $E_{\text{hyd_enclosure}}$)
- spezielle Wasserstoffbrückenbindungen zwischen neutralen ($E_{\text{hb_nn_motif}}$) bzw. geladenen Atomen ($E_{\text{hb_cc_motif}}$)
- π -Stacking und π -Kation-Interaktionen (E_{PI})
- Wasserstoffbrückenbindungen ($E_{\text{hb_pair}}$) und lipophile Wechselwirkungen ($E_{\text{phobic_pair}}$) zwischen Ligand und Rezeptor (typische ChemScore-Terme)

E_{penalty} (7) erniedrigt die Bindungsaffinitätsvorhersage aufgrund der Desolvatisierung polarer oder geladener Gruppen (E_{desolv}) und aufgrund hoher energetischer Spannungen innerhalb des gebundenen Liganden ($E_{\text{ligand_strain}}$). [302]

²² gemeint ist die überschüssige innere Energie des Konformers im Vergleich zur Pose mit der niedrigsten Energie, berechnet mit einer Version des OPLS2005-Kraftfelds

Induced-Fit-Docking

Die Verwendung starrer Proteinstrukturen beim Docking führt nicht immer zu den erwarteten Dockingposen. Die Berücksichtigung der Beweglichkeit von Seitenketten und Protein-Backbone kann zu verbesserten Dockingergebnissen führen, weil die hervorgerufenen Veränderungen eine Formanpassung an den Liganden zur Folge haben können (*Induced Fit*).

Die SCHRÖDINGER-Software bietet ein Induced-Fit-Docking an unter der Verwendung von GLIDE als Dockingtool und PRIME als Proteinmodellierungstool. [309] Die Protein- und Ligandenvorbereitung entspricht der Vorgehensweise wie für das GLIDE-Docking. Zuerst wird ein GLIDE-Docking mit niedrigskalierten vdW-Radien der Protein- und Ligandatome durchgeführt. Hier kann optional ein *Trimming* (dt. Abschneiden, Kürzen) von Seitenketten erwirkt werden, um mehr Platz für Liganden in der Bindetasche zu schaffen. Die Auswahl der relevanten Seitenketten erfolgt manuell oder automatisch auf Basis des *B-Faktors*²³. Für den Dockingprozess werden die ausgewählten Seitenketten zur Alaninseitenkette mutiert. Nach dem Docking erfolgt für die jeweils besten 20 Posen eine Vorhersage der Seitenkettenpositionen mit PRIME in einem Abstand von standardmäßig 5 Å zu jeder Ligandenpose und eine PRIME-Minimierung der Posen unter Anwendung von Kraftfeldparametern (OPLS) und dem impliziten *Generalized-Born*-Lösungsmittelmodell. Proteinstrukturen mit einer *Prime-Energy*²⁴-Abweichung von maximal 30 kcal/mol von der Struktur mit der niedrigsten Energie werden beibehalten. Mit diesen nun „angepassten“ Proteinkonformationen wird ein erneutes GLIDE-Docking mit Standardparametern, d.h. ohne niedrigskalierte vdW-Radien bzw. Trimming, vollzogen. Für jede ausgegebene Pose wird ein *IFDScore* (8) berechnet zur Abschätzung der Bindungsaffinität: [273]

$$\text{IFDScore} = 1,0 * \text{GlideScore} + 0,05 * \text{PrimeEnergy} \quad (8)$$

4.4.2 HYDE

Die empirische Scoringfunktion HYDE [311–313] (*Hydrogen Bond and DEhydration*) zur Bewertung der Bindungsaffinität berücksichtigt die Desolvatisierung von Protein und Ligand und den Aufbau neuer Wasserstoffbrückenbindungen, um die freie Energie des Protein-Liganden-Komplexes

²³ Der B-Faktor (auch: Temperaturfaktor) quantifiziert die Fluktuation der Atomposition in experimentellen Röntgenstrukturen um den Mittelwert bedingt durch eine Abschwächung der Beugungsintensität infolge thermischer Atombewegungen. Hohe B-Faktoren lassen auf eine hohe Flexibilität schließen. [310].

²⁴ Die *Prime Energy* umfasst die innere Energie des Proteins nach Minimierung und Solvatationsenergien.

zu berechnen. Die Besonderheit an dieser Scoringfunktion ist die Berücksichtigung destabilisierender Effekte auf die Bindung. So kann z.B. durch die Desolvatation bzw. Dehydratation von wasserstoffbrückenbildenden Ligandenatomen, welche innerhalb der Proteinbindetasche entweder neue Wasserstoffbrücken suboptimaler Geometrie oder gar keine Wasserstoffbrücken mehr ausbilden, die Bindungsenergie negativ beeinflusst werden. [313]

Zur Berechnung des **HYDE-Scores** (9) werden die Beiträge aller Rezeptor- und Ligandenatome i zur Dehydratisierung und zur Ausbildung neuer Wasserstoffbrückenbindungen aufsummiert: [313]

$$\Delta G_{\text{HYDE}} = \sum_{\text{atoms } i} \Delta G_{\text{dehydration}}^i + \Delta G_{\text{H-bonds}}^i \quad (9)$$

Bei dem Dehydratations-Term ($\Delta G_{\text{dehydration}}^i$) wird zwischen hydrophoben und hydrophilen Atomen unterschieden, da eine Dehydratisierung von hydrophoben Gruppen die Bindung begünstigt, während das Abstreifen von Wassermolekülen von hydrophilen Gruppen als ungünstig zu bewerten ist: [312]

$$\Delta G_{\text{dehydration}}^{i,\text{hydrophobic}} = -2,3 \cdot RT \cdot p \log P^i \cdot (acc_{\text{unbound}}^i - acc_{\text{bound}}^i) \quad (10)$$

$$\Delta G_{\text{dehydration}}^{i,\text{hydrophilic}} = -2,3 \cdot RT \cdot p \log P^i \cdot f_{\text{bur}}^i \cdot f_{\text{water}}^i \cdot \sum_{\text{H-bond functions } j} w^j \cdot p_{\text{dehyd}}^j \quad (11)$$

- $p \log P^i$ – partieller logP-Wert eines Atoms i
- R – universelle Gaskonstante
- T – absolute Temperatur
- $acc_{(\text{un})\text{bound}}^i$ – Oberfläche des Atoms i , die dem Lösungsmittel im (un)gebundenen Zustand zur Verfügung steht
- p_{dehyd}^j – Dehydratisierungswahrscheinlichkeit eines Wasserstoffbrückendonors bzw. -akzeptors (j); beträgt 1 für tief in der Bindetasche liegende Donoren/Akzeptoren und für solche, die Wasserstoffbrücken mit der Zielstruktur ausbilden; p_{dehyd}^j nimmt ab, je mehr Platz für ein Wassermolekül in der bevorzugten Richtung für eine Wasserstoffbrückenbindung vorhanden ist;
- w^j – Gewichtungsfaktor für Donoren/Akzeptoren, die mehr als eine Wasserstoffbrückenbindung ausbilden können
- f_{bur}^i – Skalierungsfaktor für hydrophile Gruppen im ungebundenen Zustand nach dem Grad ihrer Exposition zu Wasser; beträgt 1 für hydrophile Ligandenatome und nimmt für hydrophile Proteinatome ab, je tiefer sie in der Bindetasche „vergraben“ sind;
- f_{water}^i – Korrekturfaktor, der die Überlappung und Wechselwirkung von ideal platzierten Wassermolekülen beschreibt und somit die Qualität der Hydratisierung charakterisiert

Die partiellen logP-Werte für die einzelnen Atome wurden durch die Kalibrierung experimenteller logP-Werte kleiner Moleküle aus einer Datenbank gewonnen. In die Berechnung der Dehydrationsenergie hydrophober Atome ($\Delta G_{\text{dehydration}}^{i, \text{hydrophobic}}$) geht die durch Ligandenbindung bedingte Änderung der Oberfläche der Atome i , die dem Lösungsmittel zur Verfügung stehen, ein. Der energetische Term für die Dehydratisierung hydrophiler Atome ($\Delta G_{\text{dehydration}}^{i, \text{hydrophilic}}$) berücksichtigt die Wahrscheinlichkeit der hydrophilen Atome, dehydratisiert zu werden und bezieht Faktoren mit ein (s.o.), um die Qualität und Quantität der Wasserstoffbrücken ausgehend von diesen Atomen zu bewerten. Beispielsweise können hydrophile Atome in der Proteinbindetasche für Wassermoleküle nur begrenzt erreichbar sein oder mehrere hydrophile Atome mit dem gleichen Wassermolekül (oder vice versa) interagieren, sodass die Dehydratisierungsenergie für diese Atome niedriger ausfällt. [312]

Der Term zur Berechnung der Wasserstoffbrückenbindungsenergie ($\Delta G_{\text{H-bonds}}^i$) hat eine ähnliche Form wie der Dehydratisierungsterm für hydrophile Atome, da es sich um einen inversen Prozess handelt (Aufhebung vs. Wiederaufbau von Wasserstoffbrückenbindungen): [312]

$$\Delta G_{\text{H-bond}}^i = \frac{-2,3 \cdot RT}{F_{\text{sat}}(T)} \cdot p \log P^i \cdot f_{\text{bur}}^i \cdot \sum_{\text{H-bond } j} w^j \cdot f_{\text{dev}}^j \quad (12)$$

- $F_{\text{sat}}(T)$ – temperaturabhängiger Sättigungsfaktor: beschreibt den Sättigungsgrad des Wasserstoffbrücken-Netzwerks; $F_{\text{sat}} = 1$, wenn alle vier möglichen Positionen jedes Wassermoleküls jeweils eine ideale Wasserstoffbrückenbindung ausbilden
- f_{dev}^j – Faktor zur Bewertung der Qualität einer Wasserstoffbrückenbindung (abhängig von der Abweichung vom idealen Abstand und Winkel)
- $w^j, p \log P^i, f_{\text{bur}}^i$ – siehe oben (Dehydratationsterm)

Der HYDE-Score kann für alle Arten von Zielstrukturen verwendet werden, weil für die Erstellung der Scoringfunktion nicht auf experimentelle Bindungsaffinitäten oder Protein-Ligand-Komplex-Daten zurückgegriffen wurde. Da der HYDE-Score atombasiert ist, kann er auf die Liganden- und Proteinatome als Farbschema projiziert werden und so die Analyse von Dockingergebnissen erleichtern. [312]

SCHNEIDER ET AL. konnten unter Verwendung eines DUD-Datensatzes [314] zeigen, dass bindende gegenüber nicht-bindenden Liganden im Vergleich zu anderen weit verbreiteten Scoring-Methoden (z.B. GLIDE SP [284], GOLD CHEMPLP [315], ICM [283], etc.) genauso gut angereichert werden konnten. [312]

Das HYDE-Scoring kann über die Kommandozeile ausgeführt werden (s. Kap. 7.9.2). Das Softwarepaket LEADIT [316] kann zur Erzeugung der Bindetaschen-Dateien (.fxx-Format) unter Verwendung eines Referenzliganden und einem benutzerdefinierten Radius um den Liganden herum definiert werden. Mit der LeadIT-GUI kann der Protein-Ligand-Komplex insbesondere unter Verwendung des oben genannten Farbschemas visualisiert werden.

4.4.3 Prime MM/GBSA

Mit dem in der SCHRÖDINGER SUITE enthaltenen Modul PRIME MM/GBSA (*Molecular Mechanics – Generalized Born Surface Area*) wird die freie Bindungsenergie ΔG_{bind} (13) von Protein-Ligand-Komplexen nach folgender Gleichung berechnet: [317]

$$\Delta G_{\text{bind}} = E_{\text{complex}(\text{minimized})} - E_{\text{ligand}(\text{minimized})} - E_{\text{receptor}(\text{minimized})} \quad (13)$$

Nach einer vorausgehenden Minimierung des Protein-Ligand-Komplexes, des freien Proteins und des freien Liganden werden die Kraftfeldenergien (E) (vgl. Kap. 4.1) für jede dieser drei Strukturen in einem impliziten Lösungsmittelmodell berechnet. Die Energieberechnung beinhaltet Terme für vdW- und elektrostatische Interaktionen, für die Desolvatisierungsenergie für den Liganden und für die Deformationsenergien von Liganden und Proteinen. In einer reinen Kraftfeldberechnung können entropische Effekte nicht berücksichtigt werden. Im VSGB²⁵-Modell [318] hingegen wird die Entropie des Lösungsmittels berücksichtigt [319]. Als Kraftfeld steht das OPLS3(e) bzw. OPLS2005 zur Auswahl. Desweiteren kann zwischen verschiedenen Lösungsmittelmodellen ausgewählt werden (VSGB, Vakuum, Chloroform) [317].

Neben der berechneten freien Bindungsenergie ΔG_{bind} erhält man auch die *Ligand Strain Energy* (Ligand-Deformationsenergie) als Ergebnis für jede Pose, welche als ein zusätzliches Kriterium bei der Auswahl zu testender Hits herangezogen werden kann. Dieser Wert ergibt sich aus der Differenz der Energie des freien Liganden und des gebundenen Liganden (mit veränderter Geometrie, unter „Spannung“ bzw. Deformation) [257]. Man kann eine Pose als vorteilhaft ansehen, wenn der Ligand nicht unter Aufwendung einer hohen *Ligand Strain Energy* in den deformierten Zustand im Komplex gebracht werden muss. Daher sollte der Wert für die *Ligand Strain Energy* möglichst klein sein.

Anders als die herkömmliche MM/GBSA-Methode [320–322], welche ein Ensemble von Protein-Ligand-Komplex-Konformationen aus MD-Simulationen basierend auf einem Kontinuum-Lö-

²⁵ VSGB = *Variable dielectric surface generalized Born model*

sungsmittelmodell (*Generalized Born Model*) bewertet, benötigt PRIME MM/GBSA für die Abschätzung der Freien Bindungsenergie nur einen einzigen minimierten Komplex. Daher ist die PRIME MM/GBSA-Methode in Kombination mit molekularem Docking eine geeignete Methode für das Rescoring einer großen Anzahl von Posen mit relativ geringem Rechenaufwand [323]. Die Methode eignet sich besonders gut, um eine Reihe von gleichartigen Strukturen nach ihrer relativen Freien Bindungsenergie zu sortieren. Dies wird z.B. für eine strukturbasierte Leitstruktur-Optimierung einer chemischen Reihe von Inhibitoren für Proteinkinasen eingesetzt (vgl. [324, 325]).

4.4.4 PLP-Score

PLP (*Piecewise Linear Potential*) [326, 327] ist eine empirische Scoringfunktion, welche die sterischen Wechselwirkungen und Wasserstoffbrückenbindungen zwischen schweren Protein- und Ligandenatomen paarweise aufsummiert. Diesen Atomen wird je nach Atomtyp ein Wechselwirkungstyp zugeordnet (WBB-Akzeptor, -Donor, Akzeptor und Donor, unpolar). [327]

Die entsprechenden Terme werden nicht wie sonst größtenteils üblich mit der „glatten“ Lennard-Jones-Funktion (s. Kap. 4.1.1, S. 46), sondern mit einer abschnittsweise linearen Funktion berechnet. Bei einem Abstand zwischen zwei Atomen gegen Null ergibt diese Funktionsform im Gegensatz zum Lennard-Jones-Potenzial einen endlichen Wert. (vgl. [327] und [328])

In dieser Arbeit wurde die Implementierung des PLP-Scores in LeadIT für das Rescoring von Dockingposen verwendet (s. Kap. 5.5.1, S. 204 ff. u. Kap. 7.9.1, S. 275 f.).

4.4.5 FlexX-Score

Die empirische **FlexX-Scoringfunktion** (14) zur Berechnung der freien Bindungsenergie ΔG ähnelt der von BÖHM ET AL. [293] entwickelten Scoringfunktion und lautet: [329]

$$\Delta G = \Delta G_0 + \Delta G_{\text{rot}} \cdot N_{\text{rot}} + \Delta G_{\text{hb}} \sum_{\text{neutral H-bonds}} f(\Delta R, \Delta \alpha) + \Delta G_{\text{io}} \sum_{\text{ionic int.}} f(\Delta R, \Delta \alpha) + \Delta G_{\text{aro}} \sum_{\text{aro int.}} f(\Delta R, \Delta \alpha) + \Delta G_{\text{lipo}} \sum_{\text{lipo.cont.}} f^*(\Delta R) \quad (14)$$

- ΔG_0 – konstanter Term, der den Wegfall der Translations- und Rotationsentropie beschreibt
- ΔG_{rot} – Energieverlust durch die Immobilisierung frei rotierbarer Bindungen
- N_{rot} – Anzahl frei rotierbarer Bindungen, deren Beweglichkeit durch die Bindung verloren geht
- $\Delta G_{\text{hb}}, \Delta G_{\text{io}}, \Delta G_{\text{aro}}, \Delta G_{\text{lipo}}$ – Wasserstoffbrückenbindungs-(hb), ionische (io), aromatische (aro) bzw. lipophile (lipo) Wechselwirkungsbeiträge zur freien Bindungsenergie bei idealer Geometrie
- $f(\Delta R, \Delta \alpha)$ bzw. $f^*(\Delta R)$ – Straffunktion zur Berücksichtigung von Abweichungen von der idealen Interaktionsgeometrie (Abstand R , Winkel α)

Die Implementierung des FlexX-Scores in LeadIT (s. Kap. 7.9.1, S. 275 f.) wurde in dieser Arbeit verwendet (s. Kap. 5.5.1, S. 204 ff.).

4.4.6 Consensuscore

Unter Consensus-Scoring versteht man allgemein die kombinierte Anwendung verschiedener Scoringfunktionen für die Bewertung von (Docking-)Posen. Dies soll die Wahrscheinlichkeit für falsch-positive Treffer reduzieren und die resultierende neue Rangfolge durch Kombination mehrerer „Stichproben“ (Scoringfunktionen) näher an die „wahre“ Rangfolge bringen. [299, 300]

Für die Bildung eines kombinierten Scores gibt es unterschiedliche Herangehensweisen: [300]

a) *rank-by-number*: Für jede Pose wird der Mittelwert aus den Scorewerten unterschiedlicher Scoringfunktionen gebildet. Es wird allerdings vorausgesetzt, dass die Scorewerte die gleiche Einheit besitzen (z.B. kcal/mol für berechnete Bindungsaffinitäten). Freie Bindungsenergien können z.B. nicht mit kraftfeldbasierten Energien kombiniert werden.

b) *rank-by-rank*: Hier spielt die Einheit des Scorewerts keine Rolle, da jeder Ligand, sortiert nach dem entsprechenden Score, eine Rangpositionsnummer erhält. Der **Consensuscore** (15) wird nach folgender Formel berechnet: [330]

$$\text{Consensuscore} = \frac{\sum_{i=1}^n r_i}{n} \quad (15)$$

n – Anzahl verschiedener Scoringfunktionen, die in die Bewertung einfließen sollen

r_i – Rangpositionsnummer, die sich aus der Sortierung nach den Score-Werten aus Scoringfunktion i ergibt

Der Consensuscore einer Pose wird demnach als Mittelwert aus den Rangpositionsnummern berechnet, welche sich aus der Sortierung nach den Score-Werten der jeweiligen Scoringfunktion ergeben.

c) *rank-by-vote*: Eine Pose erhält eine „Stimme“ (z.B. +1), wenn sie nach Sortierung nach dem entsprechenden Score zu einem definierten Prozentsatz der bestbewerteten Posen gehört. Alle „Stimmen“ eines Liganden werden addiert zu einem Consensuscore.

Die *rank-by-vote*-Strategie hat den Nachteil, dass quantitative Informationen verloren gehen und die Kategorisierung der Liganden grob erfolgt. In einer Untersuchung von WANG ET AL. schneidet sie bezüglich der Trefferquoten am schlechtesten ab. [300]

In dieser Arbeit wurden Scoringfunktionen miteinander verglichen, die Score-Werte mit unterschiedlichen Einheiten hervorbringen. Da die *rank-by-rank*-Strategie im Gegensatz zur *rank-by-*

number-Strategie unabhängig von den Score-Einheiten ist und einen repräsentativen und überlegenen Consensus-Ansatz darstellt [300, 330], wurde sie für diese Arbeit ausgewählt, um die Dockingergebnisse neu zu sortieren. Auf Basis des Consensuscores können auch Ligandenlisten aus verschiedenen Dockingexperimenten zusammengefasst werden (s. Datenfusionierung, nächster Absatz). Die Auswahl von drei bis vier unterschiedlichen Scoringfunktionen für den Consensuscore ist für eine Verbesserung der Trefferquote bereits ausreichend. [300]

4.4.7 Datenfusionierung und -reduktion

Nach dem Docking in mehrere Proteinkonformationen (Grids) erhält man für jedes Grid eine Rangliste der Dockingposen. Eine Verbindung kann durch unterschiedliche Protonierung bzw. Tautomerisierung (s. Kap. 4.3) mehrmals in der Rangliste vorkommen (im Folgenden als *States* bezeichnet).

Um eine bestimmte Anzahl an Posen in einem nächsten Experiment mit einem Dockingalgorithmus höherer Genauigkeit zu docken oder um eine Gesamtangliste aus den verschiedenen Ranglisten zu erhalten, ist es sinnvoll, nur jeweils die Verbindung mit der besten Bewertung zu verwenden.

Das Vereinen von Ranglisten unter Beibehaltung der Pose mit der besten Bewertung ist über den Kommandozeilenbefehl *Glide Ensemble Merge* in der SCHRÖDINGER SUITE möglich (s. Kap. 7.11.2, S. 285). Hier wird eine Inputdatei im Moleküldateiformat (z.B. .MAE) benötigt. Es wird nur diejenige Pose behalten, welche die beste Bewertung (*Default*: nach Dockingscore) erzielt hat. Gleiche Verbindungen (dazu zählen auch die verschiedenen States) werden anhand der gleichen Bezeichnung (*Title*) erkannt. Nach Anwendung des *Glide Ensemble Merge* erhält man einen Datensatz mit einmalig vorkommenden Verbindungen, d.h. Duplikate und andere States der gleichen Verbindung werden zugunsten der Pose mit der besten Bewertung (*Default*: nach Dockingscore) entfernt. [257]

Die Scorewerte aus den Rescorings stammen aus unterschiedlichen Programmen und sind nicht in den MAE-Dateien hinterlegt. Für das Zusammenfügen aller Scorewerte²⁶ zu den Liganden in einer gemeinsamen Liste, für die Sortierung nach dem jeweiligen Scorewert und für die Vereinigung von Ranglisten aus verschiedenen Grids im Rahmen des Postprocessings wurde KNIME [331] verwendet (s. Kap. 4.9.2 u. 7.9.6). KNIME kann große Textdateien bzw. Tabellen, welche die Ligadentitel und Scorewerte enthalten, verarbeiten. Die Ergebnisse (z.B. die vereinigte

²⁶ HYDE, ΔG_{bind} mit PRIME MM/GBSA, vorhergesagter pK_i -Wert auf Basis des QSAR-Modells, Dockingscore, Consensuscore und die jeweils normierten Scorewerte (siehe *norm_score* (16) in diesem Abschnitt).

Rangliste nach definiertem Score sortiert) wurden zur weiteren Verarbeitung, z.B. zur Erstellung von Anreicherungskurven bzw. zur Berechnung von Enrichmentfaktoren (s. S. 79 f.) verwendet.

Eine Besonderheit bei der Fusionierung der Ranglisten (Top 5%-SP-Dockingergebnisse) aus verschiedenen Grids stellte die Normalisierung der Scorewerte dar. Sie wird nötig, wenn sich die Scorewerte in einem *Grid a* beispielsweise von -12 bis 0 erstrecken kann, in einem anderen *Grid b* wiederum von -9 bis 2. Eine Zusammenfassung ohne Normierung würde zu einer Liste führen, in der die besten Posen nur aus dem *Grid a* stammen, sodass man die Posen aus *Grid b* vernachlässigen würde. Genauso muss auch der Consensuscore normiert werden, da sich die Anzahl an gedockten Liganden in den verschiedenen Grids unterscheidet. Eine Rangposition von „99“ in einem Grid ist unterschiedlich zu bewerten, je nachdem ob in das Grid 100 oder 10.000 Liganden gedockt werden konnten.

Um die Scorewerte aus den verschiedenen Grids miteinander vergleichbar zu machen, wurden die Scorewerte und auch der Consensuscore einer Liste daher nach folgender Formel normalisiert: [332]

$$\text{norm_score} = \frac{\text{raw_score} - \bar{x}_{\text{all_rawscores}}}{s_{\text{all_rawscores}}} \quad (16)$$

Die Differenz aus dem Scorewert eines Liganden (*raw_score*) und dem Mittelwert \bar{x} aller Scorewerte einer Dockingliste aus einem Grid wird durch die Standardabweichung *s* aller Scorewerte geteilt. Der auf diese Weise normalisierte Wert wird auch als Z-Score bezeichnet. Es handelt sich um eine dimensionslose Größe [332, 333]. Je negativer der Z-Score, desto stärker die geschätzte Bindung.

4.4.8 Validierungsmethoden und -metriken

Zur Validierung von Dockingexperimenten wurden die folgenden Methoden und Metriken herangezogen.

Validierung des Dockings

Ob ein Dockingprogramm oder ein Virtual-Screening-Workflow für den vorgesehenen Zweck geeignet ist (*Validierung*), lässt sich insbesondere anhand der folgenden Kriterien überprüfen: [334, 335]

1. Fähigkeit, einen kokristallisierten Liganden, entnommen aus seinem Komplex und mit willkürlich veränderter Konformation, in seiner ursprünglichen Konformation und Orientierung in die Bindetasche zurückzudocken (*Re-Docking*); hiermit lässt sich die Genauigkeit des Dockings überprüfen. Der RMSD-Wert der generierten Pose in Bezug auf seine Ausgangsposition, idealerweise kleiner als 2 Å, gibt den Hinweis auf eine ausreichende Genauigkeit.
2. Fähigkeit, (bekannte) aktive Liganden in den vorderen Positionen einer nach einem Score sortierten Liste zu finden und so von nicht-bindenden Molekülen (z.B. Decoys) abzugrenzen (*Enrichment*, dt. Anreicherung). Dies lässt sich mit einem Testdatensatz, welches man dem Dockingprozess unterwirft, überprüfen. Hierbei wird die Anreicherung berechnet und / oder in Diagrammen (Anreicherungskurven oder *Receiver Operating Characteristic* (ROC)-Kurven) dargestellt.
3. Fähigkeit, eine Liste aktiver Liganden mit dem Dockingprogramm in die richtige Rangfolge mit absteigender biologischer Aktivität zu überführen. Dies lässt sich mit einem Testdatensatz aus aktiven Liganden und der Berechnung der Korrelation zwischen experimentellen Aktivitäts- und Score-Werten überprüfen.

Das Docking stellt ein zentrales Werkzeug dieser Arbeit dar. So wurde zunächst mit Re- und Crossdockingexperimenten getestet, ob und mit welcher Genauigkeit sich die Liganden aus den DLK-Röntgenkristallstrukturen in die vorbereiteten Proteinstrukturen zurückdocken ließen (**Kriterium 1** der Validierung, s. Kap. 5.4.1.1). Die Bewertung der Posen erfolgte durch visuelle Inspektion und anhand der generierten Score- und RMSD-Werte.

Des Weiteren wurden Dockingexperimente an einem Ensemble an Proteinkonformationen vorgenommen (s. Kap. 5.4.1.2, 5.4.1.3 und 5.4.2), um die Flexibilität der Proteinzielstruktur (DLK) zu berücksichtigen.

Durch Docking eines Testdatensatzes aus literaturbekannten DLK-Inhibitoren und potenziell nicht-bindenden Molekülen (Decoys) wurde ein Virtual-Screening-Workflow erarbeitet und optimiert (s. Kap. 5.4.1.3). Der Workflow sollte in der Lage sein, die bekannten DLK-Inhibitoren in die verschiedenen Proteinkonformationen einzupassen (*Docking-Problem*) und besser zu bewerten als die Decoys (*Scoring-Problem*, **Kriterium 2**).

Im zentralen Schritt des Workflows würden Millionen Substanzen aus der ZINC-Bibliothek mithilfe des Dockings in die Bindetasche ausgewählter DLK-Proteinkonformationen eingepasst und bewertet werden (s. Kap. 5.4.2). Dieser Schritt sollte zum einen nicht passende Strukturen aussortieren und zum anderen die gedockten Substanzen nach ihrem Score-Wert sortieren. Das Docking könnte hier neue Liganden mit neuen Interaktionsstellen z.B. zum P-Loop der DLK hervorbringen.

Nach dem Docking erfolgte eine Neubewertung (*Re-Scoring*) der Dockingposen (s. Kap. 5.5.1 bis 5.5.3). Hier wurden verschiedene Scoringfunktionen und Vorhersagemodelle nach **Kriterium 3** überprüft und im Sinne des Consensus Scoring zugunsten einer verbesserten Anreicherung Aktiver miteinander kombiniert.

Des Weiteren wurde das *Alignment* (dt. Ausrichtung, Übereinanderlagerung) literaturbekannter DLK-Inhibitoren zur Erstellung von QSAR-Modellen mithilfe eines Dockings in die ATP-Bindetasche der DLK generiert (s. Kap. 5.3.1).

Enrichment Factor

Der *Enrichment Factor* (*EF*, dt. Anreicherungsfaktor) gibt an, um welchen Faktor die Aktiven gegenüber potenziell Inaktiven (z.B. Decoys) im Vergleich zu einer rein zufälligen Verteilung aller Verbindungen im Datensatz angereichert bzw. abgereichert werden. Der *Enrichment Factor* für die besten $x\%$ ($EF_{\text{top } x\%}$) eines Datensatzes berechnet sich nach: [336]

$$EF_{\text{top } x\%} = \frac{\%Aktive_{\text{top } x\%}}{\%Datensatz_{\text{top } x\%}} = \frac{\frac{Hits_{\text{top } x\%}}{N_{\text{top } x\%}}}{\frac{Hits_{\text{total}}}{N_{\text{total}}}} \quad (17)$$

$\%Aktive_{\text{top } x\%}$	– prozentualer Anteil der Aktiven in den besten $x\%$ des Datensatzes an allen Aktiven
$\%Datensatz_{\text{top } x\%}$	– prozentualer Anteil des gescreenteten Datensatzes (entspricht x)
N_{total}	– Anzahl Verbindungen in der gesamten Datenbank
$N_{\text{top } x\%}$	– Anzahl Verbindungen in Top $x\%$ der Datenbank
$Hits_{\text{total}}$	– Anzahl Aktiver in der gesamten Datenbank
$Hits_{\text{top } x\%}$	– Anzahl Aktiver in den Top $x\%$ der Datenbank

Ein $EF_{\text{top } 1\%}$ von 10 bedeutet, dass in der ausgewählten Fraktion der besten 1 % der Datenbank 10 Mal mehr Aktive vorhanden sind als in den besten 1 % der zufällig sortierten Datenbank. Ein $EF < 1$ bedeutet, dass die Scoringmethode eine geringere Anreicherung zur Folge hat als es bei einer zufälligen Verteilung der Aktiven der Fäll wäre. Dies entspricht einer Abreicherung von Aktiven. [336]

Die Anreicherung wird in dieser Arbeit mit *Enrichment Plots* (dt. Anreicherungsdiagrammen, vgl. Abb. 4-5) dargestellt. Der Anteil der gescreenten Datenbank ($\% \text{Datensatz}_{\text{top } x\%}$) wird auf der x-Achse

in logarithmischer Darstellung aufgetragen, während der Anteil der Aktiven an der Gesamtzahl der Aktiven ($\% \text{Aktive}_{\text{top } x\%}$) auf der y-Achse aufgetragen wird. [334]

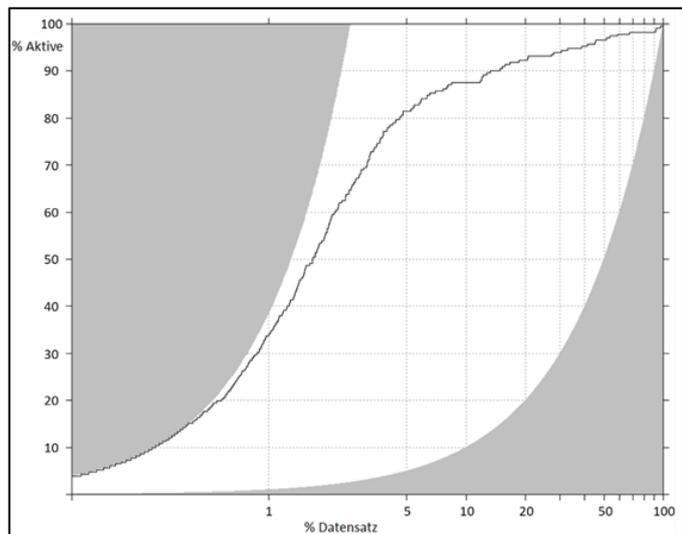


Abb. 4-5 – Exemplarischer Enrichmentplot: obere graue Kurve = perfekte Anreicherung (alle ersten Plätze werden von den Aktiven besetzt); untere graue Kurve = Anreicherung für zufällige Verteilung der Strukturen im Datensatz ($x = y$); schwarze Kurve = berechnete Anreicherung aus einem Beispieldatensatz: in diesem Beispiel sind in den Top 5 % des Datensatzes mehr als 80% der Aktiven enthalten. Dies entspricht einem Anreicherungsfaktor von über 16. Das Diagramm wurde mit GNU PLOT [337] auf Basis eigener Daten erstellt.

Korrelationskoeffizienten

Es gibt verschiedene Arten von Korrelationskoeffizienten, um zu überprüfen, inwieweit der Scorewert aktiver Liganden mit der biologischen Aktivität korreliert. Vorversuche in dieser Arbeit zeigten, dass aus der Verwendung der Pearson-Korrelation, der Kendall's tau-Rangkorrelation (tau-a, tau-b) und der Spearman-Rangkorrelation die gleichen Schlussfolgerungen gezogen werden konnten. Daher wird im Folgenden lediglich der Pearson-Korrelationskoeffizient r aufgeführt. Dieser wurde eingesetzt zur Beschreibung des linearen Zusammenhangs zwischen dem experimentellen pK_i -Wert (x) und einem Score (y). Ohne linearen Zusammenhang ist $r \approx 0$. Bei $r > 0$ spricht man von einer positiven Korrelation, d.h. größere Werte von x gehen mit größeren Werten von y einher. Bei $r < 0$ liegt eine negative Korrelation vor, d.h. höhere Werte für x hängen mit niedrigeren Werten für y zusammen.

Der Pearson-Korrelationskoeffizient (auch: Produkt-Moment-Korrelation) berechnet sich nach: [338]

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \cdot \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (18)$$

4.5 Molekulardynamik

Wie bereits in Kapitel 4.4 (S. 59) beschrieben ist es heute weithin anerkannt, dass die Target- und Ligandenflexibilität für die korrekte Posenvorhersage berücksichtigt werden sollte. Dies kann mit dem Computer durch die Anwendung von Molekulardynamik (MD)-Simulationen verwirklicht werden. Die MD-Simulation, in den späten 1950er Jahren von Alder und Wainwright [339] entwickelt, ist eine deterministische Methode zur Erzeugung einer Serie von Konformationen eines chemischen oder biologischen Systems (z.B. eines Protein-Liganden-Komplexes in Lösung) durch die Integration der Newton'schen Bewegungsgesetze über die Zeit. Mit dieser Anwendung lassen sich Molekülbewegungen über einen bestimmten Zeitraum untersuchen. Das Ergebnis einer MD-Simulation ist eine Trajektorie, welche die Veränderung der Atompositionen und -geschwindigkeiten in Abhängigkeit von der Zeit darstellt. [206]

Eine zu simulierende Ausgangsstruktur wird virtuell in eine Lösungsmittelbox getaucht. Das System wird mit dem Zusatz von Ionen neutralisiert. Jedem Atom wird eine zufällige Anfangsgeschwindigkeit in eine beliebige Richtung zugeordnet. Dann werden alle auf die Atome einwirkenden Kräfte auf Grundlage klassischer Molekülmechanik (s. Kap. 4.1) berechnet [183]. Die nächste Position wird nach einem definierten Zeitintervall basierend auf dem zweiten Newton'schen Gesetz durch schrittweise Integration für jedes Atom gleichzeitig erzeugt [340]:

$$\frac{d^2 r_i(t)}{dt^2} = \frac{F_i}{m_i} \quad (19a)$$

$$F_i = \frac{-\partial U(r_1, \dots, r_N)}{\partial r_i} \quad (19b)$$

Gleichung (19a) beschreibt die Bewegung eines Atoms i mit der Masse m_i entlang einer Koordinate r_i unter Einwirkung der Kraft F_i auf das Atom. Der Term $\frac{d^2 r_i(t)}{dt^2}$ entspricht der Beschleunigung als zweite Ableitung der Atomposition r_i in Bezug auf die Zeit t . Die dazu proportionale Kraft F_i erhält man aus der negativen ersten Ableitung der potenziellen Energie U des Systems in Bezug auf die Atompositionen (Gleichung (19b)). Die potenzielle Energie wird mithilfe eines Kraftfelds berechnet.

Die Integration erfolgt über sehr kleine Zeitintervalle Δt . Als Zeitintervall wird aufgrund der hohen Anzahl an schnell ablaufenden Vorgängen in der Regel eine Femtosekunde (10^{-15} Sekunden) eingesetzt [183]. Zur numerischen Integration der Differentialgleichung (19a) wird z.B. der Verlet-Algorithmus [341], dessen Modifikation Velocity-Verlet-Algorithmus [342] oder der leap-frog Algorithmus [343] verwendet.

Aufgrund des hohen Rechenaufwands für die Integration über jedes einzelne Atom gleichzeitig beschränkt man sich in der Regel auf Simulationszeiträume im Piko- bis Nanosekundenbereich [344]. Ausgewählte Momentaufnahmen²⁷ (sog. *Frames*) können aus den Trajektorien extrahiert werden, um sie in einem Dockingprojekt zu verwenden [281] oder diverse Eigenschaften, wie die Freie Energie oder Distanzen zwischen Atomen und RMSD-Werte (s. Kap. 4.6) der beteiligten Moleküle, zu berechnen [345, 346].

Im Bereich der strukturbasierten Wirkstoffentwicklung gibt es diverse Anwendungsgebiete für MD-Simulationen:

- Untersuchung der Dynamik eines Systems, z.B. Überprüfung der Flexibilität bestimmter Molekülregionen zur Aufklärung von Mechanismen [347], Vergleich der Bewegungsabläufe von Proteinen nach Veränderungen im System z.B. durch Mutation von Aminosäuren [348], Phosphorylierungen an Aminosäuren wie Tyrosin [349], Hinzufügen oder Entfernen eines Liganden in der Bindetasche [350] u.v.m.
- Stabilitätsuntersuchungen oder Refinement von Protein-Ligand-Komplexen nach dem Docking [351–353] bzw. von Homologiemodellen zur Validierung der Modellqualität [354, 355]
- Generierung von unterschiedlichen Proteinkonformationen für ein Ensemble-basiertes Docking zur Berücksichtigung der Proteinflexibilität [281, 356]
- Analyse der Bindungskinetik von Wirkstoffmolekülen [357, 358]
- Berechnung von Bindungsaffinitäten (Freie Energie) [359, 360]

Alle MD-Simulationen in dieser Arbeit wurden mit DESMOND [361, 362], integriert in der SCHRÖDINGER SUITE, ausgeführt.

DESMOND nutzt ein explizites Lösungsmittelmodell. Die eingesetzten Strukturen müssen zuvor einer Protein- bzw. Ligandenvorbereitung unterworfen werden. Das System muss dann mithilfe des DESMOND SYSTEM BUILDER vor der Simulation unter Auswahl des Lösungsmittelmodells für Wasser in einer Box mit periodischen Randbedingungen eingerichtet werden. Form und Größe dieser Wasserbox können variiert werden. Desweiteren kann eine Auswahl von Ionen zur Neutralisierung des Systems getroffen werden und das anzuwendende Kraftfeld wird ausgewählt.

²⁷ Momentaufnahmen = dreidimensionale Darstellung aller Atompositionen zu einem bestimmten Zeitpunkt

Im „DESMOND MOLECULAR DYNAMICS“-Panel legt man die Zeitabstände zwischen den zu speichernden Momentaufnahmen und die Gesamtsimulationszeit fest. Desweiteren wird ein Variablen-satz von experimentellen Kontrollparametern (thermodynamisches Ensemble²⁸) ausgewählt, um zu definieren, unter welchen Bedingungen das Experiment durchgeführt werden soll.

Vor der eigentlichen Simulation muss eine Überführung des Systems in ein lokales Energieminimum vorgenommen werden (*Relaxation*), da auf die einzelnen Atome sonst zu Beginn der Simulation zu große Kräfte einwirken könnten, die zu einer „Explosion“ der Simulation führen würden [363]. Danach erfolgt die Erwärmung auf eine konstante (benutzerdefinierte) Temperatur²⁹. Jedes Atom erhält eine Startgeschwindigkeit [368]. Die Kraftberechnung unter Verwendung eines Kraftfeldmodells, welche den größten Teil der Rechenleistung in Anspruch nimmt, sowie die Integration der Newton'schen (bzw. Langevin'schen) Bewegungsgesetze zur Berechnung der Atompositionen und -geschwindigkeiten wird für die entsprechenden Zeitintervallen durchgeführt.

Wenn die vordefinierte Temperatur erreicht ist, hat das System aber noch nicht den erwünschten thermodynamischen Zustandspunkt erreicht. Daher folgt eine Äquilibration zur Erreichung der ausgewählten Ensemble-Bedingungen. Im äquilibrierten Zustand sind nur noch geringe Schwankungen der thermodynamischen Variablen³⁰ um einen Mittelwert herum zu beobachten. Dann starten die eigentliche MD-Simulation und die Aufnahme der Trajektorie [363]. DESMOND ist kompatibel mit den Kraftfeldern CHARMM [369], AMBER [369, 370], OPLS-AA [210], GROMOS [371] und MMFF [372]. In der SCHRÖDINGER-Implementierung kann man zwischen OPLS3(e) und OPLS2005 wählen. Der implementierte RESPA (*Reference System Propagator Algorithm*)-Algorithmus [373] reduziert den Rechenaufwand dadurch, dass zwar nach jedem Zeitschritt alle nah wirkenden Kräfte, aber die Fernkräfte nur alle n Zeitschritte berechnet werden.

Für eine schnelle Berechnung elektrostatischer Wechselwirkungen, welche mit zunehmendem Abstand gemäß $E_{\text{coul}} \sim \frac{1}{r}$ (vgl. Kap. 4.1) abnehmen, verwendet DESMOND ab einer Distanz r von 9 Å standardmäßig die Particle-Mesh-Ewald (PME)-Methode [374]. Basierend auf der Ewald-Summation [375] werden auch bei dieser Methode die Coulomb-Wechselwirkungen zwischen zwei Punktladungen in einen Anteil mit kurzer Reichweite, der im realen Raum unter Verwendung

²⁸ Thermodynamische Ensembles sind statistische Ensembles. Sie vereinen alle Mikrozustände eines Systems, welche den gleichen Makrozustand (z.B. NVT) repräsentieren.

²⁹ Häufig wird eine Temperatur von 300 K als Zielwert eingestellt (vgl. [364] [365–367]).

³⁰ Es können auch andere Parameter eingesetzt werden, um zu beurteilen, ob sich das System im Gleichgewicht befindet, z.B. der RMSD.

eines Cutoffs berechnet wird, und einen Anteil mit langer Reichweite, der im reziproken Raum nur näherungsweise mit Fourier-Methoden berechnet wird, aufgeteilt [344], [376]. Eine vorherige Neutralisierung des Systems als formale Bedingung für die Berechnung und zur Vermeidung von Konvergenzproblemen ist erforderlich [377]. Aufgrund der großen Reichweite elektrostatischer Wechselwirkungen müssen periodische Randbedingungen eingehalten werden [206]. Diese Bedingung wird erfüllt, indem das in Lösungsmittel befindliche Protein als abgeschlossenes System – wie eine Einheitszelle (Box) – betrachtet wird und diese Einheitszelle in unendlicher Zahl in alle Raumrichtungen mit dem gleichen Inhalt fortgesetzt wird. Nähert sich ein Molekül einer Wand dieser Box, tritt es von der gegenüberliegenden Seite wieder in die Box ein [376]. Für große Systeme ist die gitterbasierte³¹ PME-Methode besser geeignet als die Ewald-Summation durch den Einsatz der schnellen Fast-Fourier-Transformation (FFT) [206].

Desweiteren verwendet DESMOND eine Variante des SHAKE-Algorithmus [378], um die Vibrationen der Bindungen zwischen schweren Atomen und Wasserstoffatomen durch „Einfrieren“ der Bindungslängen auf einen Idealwert einzuschränken. Diese Vibrationen gehören mit einer Dauer von 10 fs zu den schnellsten periodischen Bewegungen und können in der Simulation vernachlässigt werden, um die Zeitschritte zu erhöhen und somit den Rechenaufwand zu reduzieren.

Zur Auswertung der Daten aus den MD-Simulationen wurden die Tools DESMOND SIMULATION EVENT ANALYSIS und DESMOND SIMULATION INTERACTIONS DIAGRAM verwendet. Die extrahierten Eigenschaften aus den Trajektorien, wie RMSD-, RSMF-Daten und Abstandsberechnungen zwischen definierten Aminosäuren, können so exportiert und graphisch dargestellt werden.

4.5.1 NPT-Ensemble und Langevin-Dynamik

Die MD-Simulationen wurden mit dem isotherm-isobaren (NPT) Ensemble, d.h. mit konstanter Teilchenanzahl, unter konstantem Druck und konstanter Temperatur, durchgeführt. Die Teilchenzahl wird durch die Bedingung eines geschlossenen Systems mit fiktiv angrenzenden identischen Wasserboxen an allen sechs Flächen der Wasserbox konstant gehalten. [379]

Zur Temperaturkontrolle werden Thermostat-Algorithmen angewendet, in denen die Newton'sche Bewegungsgleichung durch eine modifizierte Bewegungsgleichung ersetzt sind, um die Atomzustände bei konstanter Temperatur korrekt wiederzugeben [380]. Die Temperatur wurde mit dem Langevin-Thermostat konstant gehalten. Unter Anwendung dieses Thermostats

³¹ Der FFT-Anwendung geht eine diskrete Verteilung der Partikelladungen auf Gitterpunkte (mesh) voraus.

bewegen sich die Teilchen wie in einem Wärmebad aus vielen kleineren Teilchen (Lösungsmittel) [381].

Simulationen, welche die Langevin- statt der Newton'schen Bewegungsgleichung verwenden, werden als Langevin-Dynamiken bezeichnet [382]. Sie gehören zu den stochastischen MD-Methoden und beruhen auf folgender stochastischer Differentialgleichung (nach [383]):

$$m \frac{d^2 r_i}{dt^2} = -\frac{\partial U}{\partial r_i} - m\gamma \frac{dr_i}{dt} + F_i(t) \quad (20a)$$

Basierend auf der von Paul Langevin im Jahr 1908 [384] entwickelten Theorie werden der Newton'schen Bewegungsgleichung zur Simulation des Wärmebads zwei Kraftterme hinzugefügt: Zum einen handelt es sich um eine Reibungskraft $-m\gamma \frac{dr_i}{dt}$ mit dem Reibungskoeffizienten γ . Diese der Bewegung eines Teilchens i mit der Masse m entgegengesetzte Kraft ist proportional zu dessen Geschwindigkeit. Zum anderen wird eine fluktuierende Zufallskraft $F_i(t)$ addiert, welche die ungeordneten Teilchenkollisionen mit dem Lösungsmittel beschreibt [385]. Die Zufallskraft folgt der Gauß-Verteilung und ergibt im Mittel Null. Sie ist δ -korreliert und wird auch als weißes Rauschen bezeichnet [386]. Sie wird mit folgender Formel beschrieben: [383]

$$\langle F_i(t) \cdot F_j(t') \rangle = \delta_{ij} \delta(t - t') 6k_B m T \gamma \quad (20b)$$

mit der Boltzmannkonstante k_B , der Teilchenmasse m , der Temperatur T , dem Reibungskoeffizienten γ , dem Kronecker delta (δ_{ij}) und der δ -Funktion ($\delta(t-t')$). Die stochastische Kraft ist zu jedem Zeitpunkt t völlig unkorreliert zu allen anderen Zeitpunkten $t' \neq t$. Dies wird mit der δ -Funktion ausgedrückt [387]. Der Rausch-Term, verbunden mit der Temperatur, fügt dem System Energie zu, während der Reibungsterm diesem die Energie entzieht. Über die Zeit gleichen sich der Reibungs- und Rausch-Term aus [388].

Die Langevin-Dynamik ist ergodisch, d.h. nach sehr langen Simulationszeiten werden alle Zustände durchlaufen [380, 389]. Ein weiterer Vorteil des Langevin-Thermostats besteht in einer Reduktion des benötigten Rechenaufwands, weil viele Atome durch stochastische Terme zusammengefasst behandelt werden. Die zu simulierenden Zeitschritte können wegen des Reibungsterms größer gewählt werden [390].

Ein Nachteil bei der Anwendung der Langevin-Dynamik ist, dass eine Wiederholung der Simulation wegen der eingeführten Zufallskraft nicht zu genau der gleichen Trajektorie führt und es

daher sinnvoll ist, mehrere Simulationen ausgehend von der gleichen Startposition durchzuführen [391]. Darüber hinaus reproduziert die Langevin-Dynamik nicht die korrekte Hydrodynamik eines Systems, da die Impulserhaltung nicht berücksichtigt wird [392].

Bei dem in DESMOND implementierten NPT-Ensemble nach Langevin (L_NPT) [393] handelt es sich um eine Kombination aus dem NVT-Ensemble nach Langevin (L_NVT) und dem Piston_NPH (konstanter Druck und Enthalpie; piston engl.: Kolben). Zur Druckregulierung wird eine zusätzliche stochastische Differentialgleichung nach Langevin-Piston eingesetzt, die ursprünglich auf die Methode von Andersen [394] zurückgeht (nach [395] und [393]):

$$\frac{dr_i}{dt} = \frac{p_i}{m_i} + \frac{1}{3} \frac{d \ln V}{dt} r_i \quad (21a)$$

$$\frac{dp_i}{dt} = f_i - \frac{1}{3} \frac{d \ln V}{dt} p_i \quad (21b)$$

$$\frac{d^2 V}{dt^2} = \frac{1}{W} [P(t) - P_{\text{ext}}] - \gamma \frac{dV}{dt} + R(t) \quad (21c)$$

- r_i – Position des Teilchens i
- t – Zeit
- p_i – Impuls des Teilchens i
- m_i – Masse des Teilchens i
- V – Volumen der kubischen Simulationszelle (Box)
- f_i – auf Atom i einwirkende Kraft
- W – Masse des Kolbens
- $P(t)$ – Druck zur Zeit t
- P_{ext} – der durch den Kolben aufgebrachte, externe Druck
- γ – Kollisionsfrequenz
- $R(t)$ – Zufallskraft, basierend auf einer Gauß'schen Verteilung einem Mittelwert und einer

$$\text{Varianz von Null (analog zu Gleichung (20b)): } \langle R(0) \cdot R(t) \rangle = \frac{2\gamma k_B T \gamma \delta(t)}{W}$$

Der Term $\frac{1}{3} \frac{d \ln V}{dt} r_i$ beschreibt die Anpassung des Wasserbox-Volumens zum Ausgleich des internen und angewandten Drucks und entspricht der Funktion eines Kolbens. Die Besonderheit nach Langevin-Piston ist die Berücksichtigung einer partiellen Dämpfung durch die Langevin-Gleichung in (21c). [393]

Das Protokoll zur Equilibrierung des Systems vor der eigentlichen Simulation für das NPT-Ensemble der DESMOND MD-Simulation bestand, bei Auswahl des Langevin-Thermostats und -Barostats, in dieser Arbeit aus den folgenden Schritten: [396]

1. Simulation mit Brown'scher Dynamik im NVT-Ensemble bei 10 K und kleinen Zeitschritten
2. Simulation im NVT-Ensemble mit Langevin-Thermostat für 12 ps bei 10 K
3. Simulation im NPT-Ensemble mit Langevin-Thermostat und Langevin-Barostat für 12 ps bei 10 K und 1 atm
4. Simulation im NPT-Ensemble mit Langevin-Thermostat und Langevin-Barostat für 12 ps bei 300 K und 1 atm
5. Simulation im NPT-Ensemble mit Langevin-Thermostat und Langevin-Barostat für 24 ps bei 300 K und 1 atm

In den Schritten 1-4 wirken Rückhaltekräfte (sog. *Restraints*) auf die schweren Atome (= Nicht-Wasserstoffatome) des Proteins, damit sich die Gesamtstruktur nicht zu stark verändert. Diese Kraft wird in Schritt 5 aufgehoben. Nach Schritt 5 beginnt die eigentliche Simulation und die Aufnahme der Trajektorie.

4.6 RMSD und RMSF

RMSD

Zur Analyse von MD-Simulationen und für die Beurteilung der Qualität von Dockingexperimenten wurde unter anderem der RMSD (*Root mean square deviation*)-Wert als Ähnlichkeitsmaß eingesetzt. Der RMSD-Wert ist die Wurzel der mittleren quadratischen Abweichung über alle (definierten) Atome im Vergleich zu einer Referenzposition [206]:

$$\text{RMSD} = \sqrt{\frac{\sum_{i=1}^{N_{\text{atoms}}} d_i^2}{N_{\text{atoms}}}} \quad (22)$$

N_{atoms} – Anzahl der zu berücksichtigenden Atome in der Struktur

d_i – Abstand zwischen den Koordinaten des Atoms i in den zwei übereinandergelagerten Molekülen

Er ist ein Maß für den durchschnittlichen Abstand zwischen den Atomen übereinandergelagerter Moleküle, z.B. zwischen einer experimentell ermittelten Ligandenposition in einer Bindetasche und einer gedockten Pose desselben Liganden. Er beschreibt, wie sehr sich die Lage und Konformation einer Struktur von einer Vergleichsposition und -konformation unterscheidet. Je höher der RMSD-Wert, desto unähnlicher sind sich zwei Posen. Dockingposen mit RMSD-Werten unter 2 Å gelten allgemein als korrekt reproduziert [397–399]. Wichtig zu betonen ist, dass der RMSD-Wert nur zwischen Strukturen gleicher Atomanzahl berechnet werden kann [400]. Desweiteren

ist zu beachten, dass er von der Molekülgröße abhängig ist. Der RMSD-Wert vergrößert sich im Allgemeinen mit der Zunahme der Anzahl der berücksichtigten Atome [401].

Der Verlauf der Veränderung der RMSD-Werte zwischen Startstruktur und simulierten Konformationen über die Zeit in einer MD-Simulation gibt Aufschluss über die Stabilität der Proteinkonformationen. RMSD-Werte in einem Größenbereich von bis zu 3 Å werden im Allgemeinen als akzeptabel angesehen [402–405]. Höhere RMSD-Werte können darauf hinweisen, dass während der Simulation größere Konformationsänderungen im Vergleich zur Startkonformation aufgetreten sind. In einer stabilen MD-Simulation sollten sich die RMSD-Werte der Proteinkonformationen um einen konstanten Wert einpendeln. Wenn sich die Fluktuationen gegen Ende der Simulation einem konstanten RMSD-Wert annähern, ist dies ein Hinweis darauf, dass sich die Simulation im Gleichgewicht befindet. Wenn er gegen Ende der Simulation weiter steigt oder fällt, ist die Simulation möglicherweise nicht lang genug durchgeführt worden [406].

Es ist zu berücksichtigen, dass nicht nur der absolute RMSD-Wert einzelner Proteinkonformationen für die Bewertung entscheidend ist, sondern auch das Auftreten von starken Fluktuationen oder Trends. Für eine bessere Bewertung von MD-Simulationen bietet es sich immer an, mehrere Trajektorien zu generieren und den Verlauf der konformativen Veränderungen miteinander zu vergleichen. [407]

RMSF

Um die Flexibilität von unterschiedlichen Regionen in einem Protein zu analysieren, wurden RMSF (*Root mean square fluctuation*) -Werte des Protein-Backbones aus den MD-Trajektorien berechnet. Dieser Parameter ist definiert als die durchschnittlicher Positionsänderung im Laufe der Zeit für jede Aminosäure im Vergleich zu einer Referenzstruktur: [408]

$$\text{RMSF}_i = \sqrt{\frac{1}{T} \sum_{t=1}^T \langle [r'_i(t) - r_i(t_{\text{ref}})]^2 \rangle} \quad (23)$$

T – Anzahl der Frames in der Trajektorie

r'_i – Atompositionen in der zu betrachtenden Aminosäure eines Frames nach Überlagerung auf die Referenzstruktur

$r_i(t_{\text{ref}})$ – Position der Aminosäure i zum Referenzzeitpunkt t (Referenz-Frame)

Die RMSF-Werte werden i.d.R. gegen die Aminosäuren in einem RMSF-Plot aufgetragen. Peaks im RMSF-Plot zeigen Proteinregionen an, deren Positionen während der Simulation am stärksten variieren. [406]

RMSF vs. RMSD

Im Gegensatz zum RMSD-Wert kann mit RMSF-Werten, dargestellt als Funktion der Position der Atome oder Aminosäuren, die Flexibilität bestimmter Proteinbereiche bzw. -atome sichtbar gemacht werden, allerdings kann keine zeitliche Veränderung des Gesamtproteins dargestellt werden, da der RMSF-Wert über die Zeit mittelt. Die Darstellung der RMSD-Werte als Funktion der Zeit wiederum ermöglicht die Nachverfolgung der Strukturveränderung, wohingegen sie keine direkten Informationen darüber liefert, welche Bereiche des Proteins von der Veränderung betroffen sind. Somit ist es ratsam, RMSF- und RMSD-Werte ergänzend zur Analyse von MD-Trajektorien einzusetzen.

4.7 Clustering

In dieser Arbeit wurden Clustermethoden angewendet, um ähnliche Liganden bzw. Proteinstrukturen zu gruppieren.

Clustering (dt. Gruppierung) ist ein Prozess zur Zusammenfassung ähnlicher Objekte aus einer Datenmenge in Gruppen (sog. *Cluster*). Ein Cluster besteht somit aus Gruppenmitgliedern, die sich alle untereinander in einer definierten Eigenschaft ähneln, während sie sich von Mitgliedern aus anderen Clustern deutlich unterscheiden. [409]

Clusteranalysen können eingesetzt werden, um z.B. einzelne Proteinkonformationen aus einer MD-Simulation zusammenzufassen und die Clusterrepräsentanten genauer zu analysieren und ggf. für ein Docking einzusetzen [281, 410]. Desweiteren kann eine Ligandbibliothek geclustert werden, um nur mit den Clusterrepräsentanten ein (virtuelles) Screening durchzuführen und damit den Rechen- bzw. Arbeitsaufwand zu reduzieren [411]. Eine weitere Möglichkeit ist das Clustering von Dockinglösungen, um die Anzahl der zu analysierenden Strukturen zu verringern und für die Auswahl von vielversprechenden Substanzen einen Fokus auf die jeweiligen Clusterrepräsentanten zu legen [412].

Zur Durchführung einer Clusteranalyse muss zum einen ein Maß für die Ähnlichkeit bzw. Distanz zwischen den Objekten (sog. Proximitätsmaße) auf der Basis von Variablen definiert werden. Zum anderen muss mit der Auswahl eines Fusionierungsalgorithmus festgelegt werden, auf welche Weise die Objekte miteinander verknüpft werden sollen. [413]

4.7.1 Variablen und Proximitätsmaße

Zu Beginn einer Clusteranalyse wird zuerst eine Rohdatenmatrix erstellt, welche die Variablenwerte (wie z.B. Fingerprints) den jeweiligen Objekten (z.B. Molekülen) zuordnet. Daher muss zu Beginn eine geeignete Variable (oder auch: Attribut, Beschreibungsmerkmale) oder mehrere Variablen der Objekte ausgewählt werden, deren Ausprägungen miteinander verglichen werden sollen. Basierend auf den Ausprägungen der Variablen werden Distanz- bzw. Ähnlichkeitswerte zwischen den Objekten berechnet und eine Distanz- bzw. Ähnlichkeitsmatrix generiert. Diese Matrix stellt alle Objekte gegenüber und enthält die berechneten Maßzahlen. [413]

Es werden zwei Proximitätsmaße unterschieden: [413]

- Ähnlichkeitsmaße: Maß für die Ähnlichkeit zwischen zwei Objekten; für zwei Objekte, die sich nicht ähnlich sind, beträgt das Ähnlichkeitsmaß Null. Je größer der Wert, desto ähnlicher sind sich die zwei Objekte. Bei einem Wert von 1 sind die Objekte vollständig identisch. [409]
- Distanzmaße: Umgekehrt zum Ähnlichkeitsmaß bemisst das Distanzmaß die Unähnlichkeit zwischen zwei Objekten. Der Wert Null wird für identische Objekte erreicht, während der Wert steigt, je unähnlicher sich die Objekte sind. [409]

Für eine detaillierte Übersicht über verschiedene Metriken, Ähnlichkeits- und Distanzmaße sei an dieser Stelle auf Kapitel 8.2.1 in „Multivariate Analysemethoden“ von BACKHAUS ET AL. [413] verwiesen.

Als Distanzmaß für das in dieser Arbeit verwendete RMSD-basierte Clustering von MD-Frames wird der RMSD-Wert (s. Kap. 4.6) verwendet. Für die Ligandenclusteranalysen wurde der Tanimotokoeffizient als Ähnlichkeitsmaß auf Basis von linearen binären Fingerprints und von RDKit-Fingerprints eingesetzt.

Tanimotokoeffizient

Ein häufig eingesetztes Ähnlichkeitsmaß für binäre Merkmale zum Strukturvergleich von Molekülen ist der Tanimotokoeffizient [414]. Er berechnet den relativen Anteil gemeinsamer Attribute (c) der Objekte (Moleküle) an der Gesamtzahl dieser Attribute in Objekt a und b: [415]

$$T = \frac{c}{a+b-c} \quad (24)$$

T – Tanimotokoeffizient

a, b – Anzahl der Merkmale in Objekt a bzw. b

c – Anzahl der gemeinsamen Merkmale (Überschneidung)

Der Tanimotokoeffizient kann Werte zwischen $0 \leq T \leq 1$ annehmen. Wenn die Objekte identisch sind, beträgt der Tanimotokoeffizient $T=1$. Wenn es keine gemeinsamen Merkmalsausprägungen gibt, ist $T=0$. In der Regel werden für den Strukturvergleich binäre Fingerprints eingesetzt, welche Hinweise auf An- oder Abwesenheit bestimmter struktureller Eigenschaften geben [416]. In dieser Arbeit wurden lineare binäre Fingerprints und RDKit-Fingerprints zur Berechnung von Tanimotokoeffizienten verwendet (s. Abschnitt „Molekulare Fingerprints“).

Der Tanimoto-Koeffizient gilt als Standard für die Ähnlichkeitssuche in der Cheminformatik [416, 417]. Allerdings gibt es bei der Verwendung der Fingerprint-Tanimoto-Kombination eine Größenverzerrung: Größere bzw. komplexere Moleküle haben eine höhere Dichte von „1“-Bits als kleinere Moleküle, sodass dies zu falsch-hohen Tanimotokoeffizienten für große Moleküle führt. Je kleiner der Bit-Datensatz bzw. je kleiner das Molekül, desto höher ist die scheinbare Unähnlichkeit. Größere Moleküle sind sich scheinbar ähnlicher als kleine Moleküle [418, 419].

Molekulare Fingerprints

Molekulare *Fingerprints* (dt. Fingerabdrücke) sind Darstellungen struktureller Eigenschaften von Molekülen in Form von binären *Bitstrings* (dt. Zeichenketten). Diese Art der kompakten Speicherung molekularer Informationen beschleunigt die Ähnlichkeitssuche und Analyse in Moleküldatenbanken. Zur Erzeugung des Bitstrings wird das Molekül fragmentiert. Bei Anwesenheit eines definierten Fragments wie z.B. einer bestimmten funktionellen Gruppe, wird ein entsprechendes Bit (z.B. „1“) in die Zeichenkette platziert, bei Abwesenheit wird dieses Bit auf „0“ gesetzt (vgl. Abb. 4-6). [419]

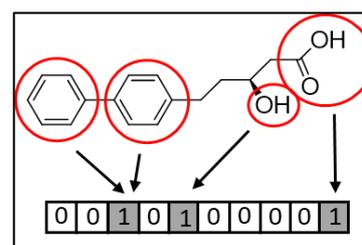


Abb. 4-6 – Erstellung eines Fingerprints aus einer Molekülstruktur. Abb. modifiziert nach GIMENO ET AL. (2019) [420]

Gruppe, wird ein entsprechendes Bit (z.B. „1“) in die Zeichenkette platziert, bei Abwesenheit wird dieses Bit auf „0“ gesetzt (vgl. Abb. 4-6). [419]

Es wird zwischen ghashten (*hashed*, dt. zerhackt) und nicht-ghashten Fingerprints unterschieden: in nicht-ghashten Fingerprints entspricht ein Bit einem definierten Fragment [421]. Die Zuordnungen zwischen funktionellen Gruppen und Bitpositionen werden aus einem vordefinierten „Wörterbuch“ entnommen [422].

Zur Erzeugung des ghashten Fingerprints wird kein Wörterbuch benötigt, stattdessen wird eine definierte Atomumgebung in numerische Werte (auch: *Pattern*) konvertiert, welche wiederum festlegen, an welcher Bitposition in dem Bitstring dieses Fragment abgebildet wird. Diese Zuordnung des Patterns zu einer Position in einem Bitstring festgelegter Länge erfolgt mit einer Hash-Funktion [423, 424]. Im Gegensatz zu den nicht-ghashten Fingerprints kann ein einziges Bit für

mehrere Fragmente verwendet werden [421]. Daher kann eine konstante Größe des Bitstrings festgelegt und Speicherplatz eingespart werden [425], nachteilig kann in diesem Zusammenhang ein möglicher struktureller Informationsverlust durch das Zusammenfallen unterschiedlicher struktureller Informationen auf die gleiche Bitposition (sog. Bitkollisionen) sein [421]. Desweiteren ist der Kodierungsprozess mit einer Hashfunktion irreversibel [426], d.h. von dem Bitstring kann nicht die chemische Struktur abgeleitet werden.

Hash-Fingerprints können, je nach Aufzählungsart der Fragmente, in pfadbasierte (= topologische) und zirkuläre Fingerprints weiter eingeteilt werden: Bei der pfadbasierten Methode werden die Fragmente des Moleküls entlang eines Pfads definierter Länge analysiert und anschließend jeder Pfad über das Hashing in den Fingerprint überführt, während die Analyse des Moleküls zur Erstellung von zirkulären Fingerprints in definierten Radien ausgehend von den Atomen erfolgt. [427]

Für die Berechnung von linearen binären Hash-Fingerprints mit dem Cheminformatik-Softwarepaket CANVAS [428], verwendet in dieser Arbeit, wird eine Struktur auf allen linearen Bindungen mit einer Länge von bis zu sieben rotierbaren Bindungen „zerlegt“ und mithilfe eines Hashing-Algorithmus in Bitstrings überführt. Für Ringschlüsse werden maximal 14 Bindungen berücksichtigt. [429]

Bei den in dieser Arbeit verwendeten RDKit-Fingerprints [432] handelt es sich um Daylight[433]-ähnliche, topologische Hash-Fingerprints. Für die Erzeugung von Daylight-Fingerprints werden für jedes Atom, für jedes

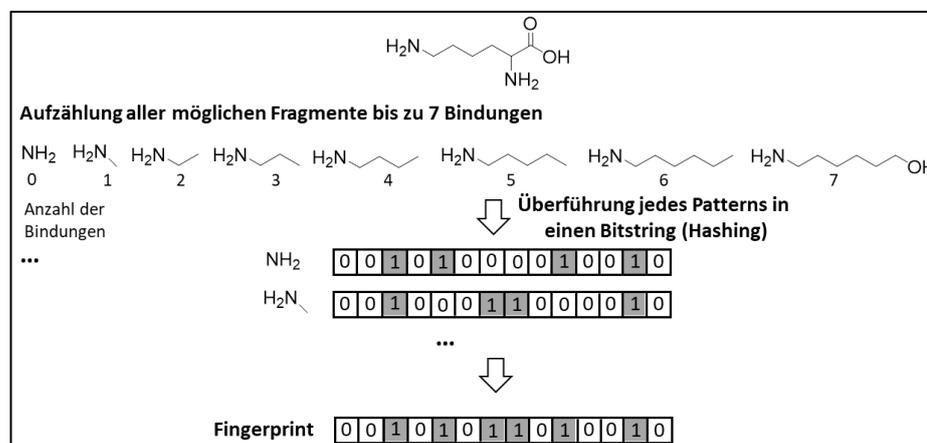


Abb. 4-7 – Schematische Darstellung der Generierung von Daylight-Fingerprints. Die Aufzählung aller möglichen Fragmente bis zu 7 Bindungen geht von jedem Atom aus (hier nur exemplarisch ausgehend von einem Stickstoffatom dargestellt). Die generierten Patterns werden durch einen Hash-Algorithmus in Bitstrings überführt (hier nur exemplarisch an einem sehr kurzen Bitstring dargestellt). Abb. modifiziert nach CERETO-MASSAGUÉ ET AL. (2015) [430] und HUANG ET AL. (2018) [431].

Atom mit seinen nächsten Nachbaratomen und -bindungen und für Atome und Bindungen in Abständen von 3-7 Bindungen jeweils ein *Pattern* generiert (s. Abb. 4-7). Diese Patterns werden gehashed, sodass für jedes Pattern vier bis fünf Bits in einem Bitstring belegt werden. [433]

Interaction Fingerprints

Ein binärer Bitstring lässt sich auch aus den dreidimensionalen Informationen der Protein-Ligand-Wechselwirkungen generieren. Eine Methode zur Berechnung solcher Interaction Fingerprints (IFP bzw. SIFt = *Structural interaction fingerprints*) wurde erstmalig von DENG ET AL. [434] beschrieben. Bis zu einem Abstands-Cutoff ausgehend von einem Ligandenatom werden die interagierenden Aminosäuren und die verschiedenen Wechselwirkungsarten berücksichtigt [434]. Der Tanimotokoeffizient ($t = 0-1$) solcher IFPs (IFP-Tanimotokoeffizient) vergleicht die Interaktionen des Rezeptor-Liganden-Komplexes mit einem Referenzliganden gebunden (oder gedockt) im gleichen Rezeptor. Je näher an 1, desto ähnlicher die Interaktionen. In dieser Arbeit wurde die Implementierung von SCHRÖDINGER (MAESTRO-GUI) zur Berechnung von Interaction Fingerprints verwendet. Hier ist eine Auswahl von bis zu neun Wechselwirkungstypen möglich.

4.7.2 Fusionierungsalgorithmen

Im weiteren Verlauf der Clusteranalyse dienen die Proximitätsmaße bzw. die aufgebaute Distanz- oder Ähnlichkeitsmatrix (s. Kap. 4.7.1) als Grundlage für den angewandten Fusionierungsalgorithmus, mit dem die Objekte gruppiert werden [413, 435]. Im Folgenden werden die Besonderheiten von hierarchischen und inkrementellen Fusionierungsalgorithmen näher beschrieben.

Hierarchische Methoden

Bei den hierarchischen Methoden werden unter Verwendung von zuvor festgelegten Startclustern die Folgecluster aufgebaut. Entweder werden größere Anfangscluster zu kleineren Clustern (divisiv) aufgegliedert oder kleinere Cluster zu größeren Clustern (agglomerativ) vereinigt. Es findet kein Austausch der Objekte zwischen den Clustern statt [413, 435, 436]. Zu den agglomerativen hierarchischen Methoden zählen die Single-Linkage-, Complete-Linkage- und Average-Linkage-Verfahren sowie die Centroid-, Median- und Ward-Verfahren [413]. Mit der Auswahl der *Linkage* (dt. Verknüpfungs)-Methode wird die Art der Abstandsberechnung zwischen den Clustern festgelegt, auf deren Basis die Cluster miteinander verknüpft werden sollen. Die Single-Linkage-Methode verwendet den kleinsten Abstand zwischen Objekten aus zwei Clustern. Mit dieser Methode können zwar Ausreißer identifiziert werden, allerdings werden i.d.R. viele kleine Cluster gebildet und Objekte der (wenigen) großen Cluster nicht weiter aufgetrennt (Kettenbildung). Die Complete-Linkage-Methode nutzt hingegen den größten Abstand zwischen den Objekten zweier Cluster. Diese Methode neigt zur Bildung kleiner Gruppen und ist nicht in der Lage, Ausreißer zu erkennen. Ausreißer sollten vor Anwendung dieses Verfahrens entfernt werden,

weil sie zur Verzerrung des Clusterings führen würden [413]. Die Average-Linkage-Methode umgeht das Problem der Empfindlichkeit gegenüber Ausreißern durch Verwendung des durchschnittlichen Abstands zwischen Objekten aus zwei Clustern und stellt einen Kompromiss zwischen der Single- und Complete-Linkage-Methode dar [437]. Bei der Anwendung agglomerativer hierarchischer Clusteranalysen in dieser Arbeit für das RMSD-basierte Clustering von Proteinkonformationen aus den MD-Trajektorien (s. Kap. 5.1.3) und für das Ligandenclustering auf Basis linearer binärer Hash-Fingerprints (s. Kap. 5.2.2, Ligandenliste c) wurde daher die Average-Linkage-Methode gewählt.

Inkrementelle Methoden

Bei den inkrementellen Clustermethoden werden einzelne Objekte zu bestehenden Clustern schrittweise hinzugefügt oder bilden einen neuen Cluster. Durch diese Vorgehensweise wird der Berechnungsaufwand bei einem wachsenden Datenbestand minimiert, da ein vollständiges neues Clustern des gesamten Datensatzes vermieden wird [438]. Zu dieser Methode gehört der Leader-Algorithmus [439, 440]. Es handelt sich um einen Single-Pass-Algorithmus [439], der das erste Objekt (bzw. Molekül) in einer sortierten Liste zum Clusterrepräsentanten (= Clusterleader) macht und zu allen folgenden Strukturen (bzw. ihren Fingerprints) das Ähnlichkeitsmaß berechnet. Alle Strukturen, die bis zu einem definierten Grenzwert t mit dem Clusterleader übereinstimmen, werden diesem zugeordnet und gehören somit zum selben Cluster. Danach wird in der Liste die erste Substanz, die nicht Teil des zuvor gebildeten Clusters ist, zum Clusterleader des nächsten Clusters und mit der Ähnlichkeitssuche auf gleiche Weise fortgefahren. Dies wird so oft wiederholt, bis alle Liganden geclustert sind [441]. Die Ergebnisse der Clusteranalyse sind von der Reihenfolge des zu clusternden Datensatzes abhängig. Die Clusteranzahl kann nicht festgelegt werden [442]. Stattdessen wird der Grenzwert t vom Anwender vor der Durchführung definiert [439]. Ein Vorteil ist die hohe Geschwindigkeit dieser Methode [443], daher eignet sie sich gut für das Clustering eines sehr großen Datensatzes an Molekülstrukturen wie in dieser Arbeit. Hier wurde das Leader-Clustering an den nach Consensuscore sortierten Dockingposen vorgenommen, sodass der Clusterrepräsentant automatisch zum Liganden mit der besten Posenbewertung in dem Cluster wurde (s. Kap. 5.5.4). Das Clustering wurde auf Basis von RDKit-Fingerprints mit einer Länge von je 2048 Bits durchgeführt.

4.8 QSAR

QSAR-Modelle auf Basis aktiver Liganden werden für die Bindungsaffinitätsvorhersage von Molekülen unbekannter Aktivität eingesetzt und können daher im virtuellen Screening-Prozess – wie in dieser Arbeit – zur zusätzlichen Bewertung von Posen im Rahmen des Postprocessings verwendet werden (vgl. [189, 444]).

Die Gemeinsamkeit aller QSAR-Methoden (QSAR – *Quantitative structure-activity relationship*, dt. Quantitative Struktur-Wirkungs-Beziehung) ist die Grundannahme, dass die biologische Aktivität eines Moleküls eine Funktion seiner chemischen Struktur ist: [183]

$$\text{Aktivität} = f_{\text{Struktur}} \quad (25)$$

Dabei wird davon ausgegangen, dass Unterschiede in den physikochemischen Eigenschaften (Hansch-Analyse [445]), die An- oder Abwesenheit von Strukturelementen (Free-Wilson-Analyse [446]) oder Unterschiede in den molekularen Wechselwirkungsfeldern (3D-QSAR der Verbindungen QSAR [447, 448]) bei den Wechselwirkungen mit der Zielstruktur eine Rolle spielen. QSAR-Methoden werden bei der Optimierung von Leitstrukturen in der Arzneistoffentwicklung häufig verwendet, wenn eine 3D-Struktur des Targets fehlt. [449]

In der Wirkstoffentwicklung werden hauptsächlich die klassischen 2D-QSAR-Methoden (z.B. Hansch-Analyse [445], Free-Wilson-Analyse [446]) und die moderneren 3D-QSAR-Methoden (z.B. CoMFA [447, 448], CoMSIA [447, 448], s. Kap. 4.8.1, S. 97 ff.) eingesetzt.

Voraussetzungen für die Erstellung validierter QSAR-Modelle

Um QSAR-Modelle zu erstellen, besteht der erste Schritt in der Sammlung von Molekülstrukturen mit bekannten biologischen Aktivitätsdaten. Die allgemeinen Voraussetzungen für die Strukturen sind eine gewisse Ähnlichkeit untereinander, derselbe Wirkmechanismus auf dieselbe Zielstruktur, ein biologisches Aktivitätsintervall von mehr als 10^2 - 10^3 und ein einfaches biologisches Testsystem, welches die Aktivitätsdaten liefert [183, 450, 451].

Die zweite Bedingung ist die Auswahl von Deskriptoren, um sie mit den Aktivitätsdaten zu korrelieren. Deskriptoren kodieren die molekularen Eigenschaften in Form von numerischen Werten. Dies können eindimensionale Deskriptoren sein, die sich aus der Summenformel ergeben (z.B. das Molekulargewicht). Häufiger werden 2D-Deskriptoren auf Basis der Strukturformel und 3D-Deskriptoren auf Basis der Konformation des Moleküls verwendet: [452, 453]

Zu den 2D-Deskriptoren für die Erstellung „klassischer“ QSAR-Modelle gehören topologische Deskriptoren, welche die Verknüpfungen zwischen den einzelnen Atomen im Molekül in numerischen Werten angeben [454], sowie physikochemische Deskriptoren, welche die sterischen [455], hydrophoben oder elektronischen [456] Eigenschaften eines Moleküls beschreiben.

3D-Deskriptoren (auch: geometrische Deskriptoren) werden aus den Atomkoordinaten des Moleküls abgeleitet [454]. Im besonderen Fall der gitterbasierten 3D-QSAR-Analysen sind es die molekularen Wechselwirkungsfelder (CoMFA-Analysen, s. Kap. 4.8.1, S. 97 ff.) bzw. Ähnlichkeitsindices (CoMSIA-Analysen), berechnet an den einzelnen Gitterpunkten. Sie sind in der Regel abhängig von der Ausrichtung der Moleküle im Raum [457].

Darüber hinaus gibt es 3D-Deskriptoren, die unabhängig von der räumlichen Ausrichtung der Moleküle sind. Dazu gehören z.B. die *GRid-INdependent* (GRIND)-Deskriptoren [458], die WHIM (*Weighted holistic invariant molecular*) [459]-Deskriptoren sowie 3D-Deskriptoren, die das Volumen und die Oberfläche von Molekülen [460] sowie Ergebnisse quantenchemischer Berechnungen (Atomladungen, molekulare Orbitalenergien, Polarisierbarkeit) [461] beschreiben. Diese wurden allerdings in den hier erstellten QSAR-Modellen nicht verwendet.

Für die Validierung von QSAR-Modellen wird der Moleküldatensatz in einen Trainings- und einen Testdatensatz aufgeteilt. Der Trainingsdatensatz dient als Grundlage für den Aufbau des Modells, während der Testdatensatz zur Überprüfung der Qualität des Modells (korrekte Vorhersage) eingesetzt wird [457]. Für ein zuverlässiges, validiertes QSAR-Modell sollte der Testsatz aus mindestens fünf Strukturen bestehen [462]. Im Hinblick auf die Diversität des Trainingsdatensatzes gibt es „lokale“ Modelle, welche homologe Reihen von Strukturen enthalten, und „globale“ Modelle, welche einen großen strukturellen Raum durch Einsatz eines größeren heterogenen Datensatzes abdecken [453]. Letzteres Modell kann eher dazu in der Lage sein, die Aktivität strukturell unterschiedlicher Moleküle vorherzusagen [453].

Die Einteilung in Test- und Trainingsdatensatz kann auf Basis der strukturellen Ähnlichkeit erfolgen, damit in beiden Datensätzen repräsentative Strukturen vorhanden sind. Zusätzlich sollte der Datensatz so geteilt sein, dass der Aktivitätsbereich in beiden Datensätzen möglichst homogen abgedeckt wird. Darüber hinaus ist auch eine rein zufällige Verteilung möglich. Üblicherweise werden 70-80 % des gesamten Datensatzes dem Trainingsdatensatz und 20-30 % dem Testdatensatz zugeordnet. [463]

2D-QSAR: Hansch- und Free-Wilson-Analyse

Der Grundstein für die klassischen 2D-QSAR-Methoden wurde 1964 mit der Hansch-Analyse [445] und unabhängig davon im gleichen Jahr mit der Free-Wilson-Methode [446] gelegt.

Die Hansch-Analyse korreliert die Aktivität mit den physikochemischen Eigenschaften einer Molekülstruktur unter Verwendung der multiplen linearen Regression: [453]

$$\log \frac{1}{C} = a(P_{\text{hydrophob}}) + b(P_{\text{elektronisch}}) + c(P_{\text{sterisch}}) + k \quad (26)$$

C – molare Konzentration des Moleküls, die einen biologischen Effekt hervorruft

a, b, c – Regressionskoeffizienten

k – konstanter Term

P – Parameter

Die chemische Struktur wird durch verschiedene physikochemische Deskriptoren, z.B. zur Beschreibung sterischer, hydrophober und elektronischer Eigenschaften, repräsentiert. [453]

Bei der Free-Wilson-Methode wird die biologische Aktivität $\log \frac{1}{C}$ (C = Stoffkonzentration) als Summe der Aktivitätsbeiträge a_i einzelner Substituenten zu einem Grundbetrag μ einer Grundstruktur berechnet. [457]

$$\log \frac{1}{C} = \sum a_i + \mu \quad (27)$$

Die An- bzw. Abwesenheit von Substituenten wird mit 1 bzw. 0 beschrieben. Die daraus erhaltene Strukturmatrix wird mit der biologischen Aktivität ins Verhältnis gesetzt. [457]

4.8.1 3D-QSAR: CoMFA und CoMSIA

Mithilfe von 3D-QSAR-Methoden werden Eigenschaften, abgeleitet aus den 3D-Koordinaten der Molekülkonformationen, mit den entsprechenden biologischen Aktivitäten korreliert. Als Deskriptoren in den gridbasierten CoMFA- und CoMSIA-Methoden dienen molekulare Felder, die aus den Wechselwirkungsenergien zwischen Sondenatom und Molekül, eingebettet in einem dreidimensionalen Gitter, berechnet werden [454].

Die Methode zur Erstellung von CoMFA-Modellen (*Comparative molecular field analysis*, dt. Vergleichende Molekularfeldanalyse) wurde 1988 von CRAMER ET AL. [447] veröffentlicht und wird als „erste echte 3D-QSAR-Methode“ [457] bezeichnet. Vorteil der CoMFA-Methode im Vergleich zu den älteren Methoden basierend auf der Hansch- bzw. Free-Wilson-Analyse ist die Berechnung ausgehend von putativ bioaktiven Ligandenkonformationen und somit die Beschreibung von

Wechselwirkungen mit der Bindetasche. Ein zusätzlich zur Aktivitätsvorhersage erhaltenes Ergebnis aus diesen Methoden ist die Möglichkeit der Visualisierung der Interaktionsfelder über Konturdiagramme, die die Analyse der einzelnen Wechselwirkungen ermöglichen sowie positive und negative Fragmenteigenschaften erkennbar machen [457].

Um die Wechselwirkungsfelder in dem CoMFA-Modell vergleichbar zu machen, ist die wichtigste Voraussetzung die räumliche Übereinanderlagerung (*Alignment*) der Strukturen. Ihre Lage im Raum, idealerweise so, wie sie in der Bindetasche vorliegen, bildet die Grundlage für die Berechnung der Wechselwirkungsenergien der molekularen Felder [183]. Das Alignment ist der schwierigste Schritt in dieser Methode, gerade wenn die bioaktive Konformation der Moleküle nicht bekannt ist [464], die eingesetzten Strukturen sehr viele rotierbare Bindungen enthalten oder nicht alle aus einem ähnlichen Grundgerüst bestehen [457]. Das Alignment kann beispielsweise über die Auswahl gemeinsamer Pharmakophore oder durch Docking (s. Kap. 4.4) in die Bindetasche der Zielstruktur erfolgen [453].

Zur Erstellung des CoMFA-Modells werden die überlagerten Moleküle in ein 3D-Gitter eingebettet. Sonden mit Eigenschaften unterschiedlich definierter Atomtypen werden auf den Gitterpunkten in definierten Abständen (1-2 Å) positioniert und die sterischen und elektrostatischen Wechselwirkungsenergien dieser Sonden zum Molekül werden berechnet. Die Summe aller Wechselwirkungsbeiträge ergibt das jeweilige molekulare Interaktionsfeld, das mit der biologischen Aktivität korrelieren soll. [183]

Für die Berechnung der elektrostatischen Wechselwirkungen wird das Coulomb-Potenzial verwendet, während die Berechnung der sterischen Wechselwirkungen auf dem Lennard-Jones-Potenzial beruht. In der CoMFA-Methode werden definierte Grenzwerte für das Lennard-Jones- und das Coulomb-Potenzial gesetzt, um unendlich

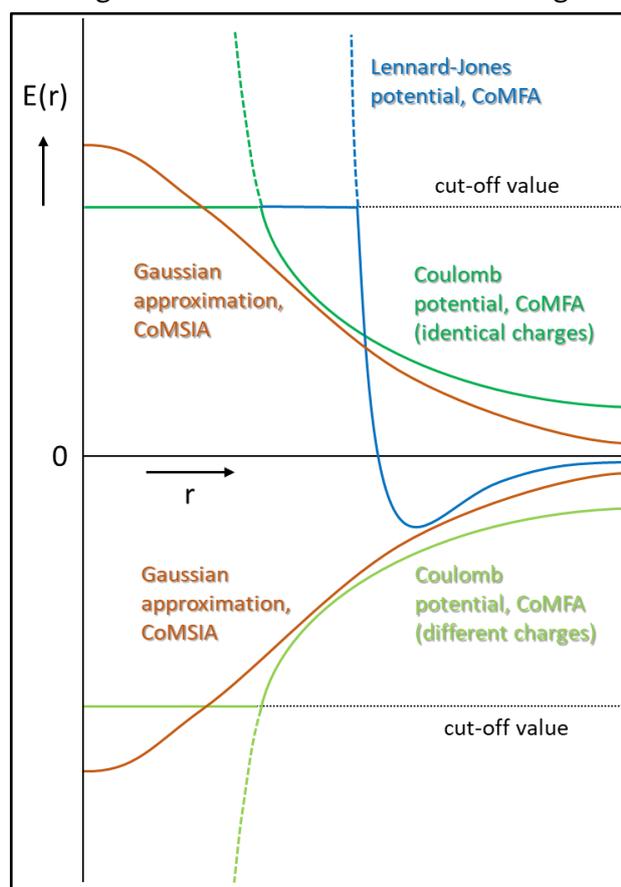


Abb. 4-8 – Approximation der Lennard-Jones- und Coulomb-Potenziale mit einer Gaußfunktion in der CoMSIA-Methode. In der CoMFA-Methode werden stattdessen Grenzwerte gesetzt, um hohe Energiewerte bei geringem Abstand r zwischen Sonde und Atom zu vermeiden. Abb. modifiziert nach KUBINYI (1997) [457].

große Energiewerte bei geringen Abständen zwischen Sonde und Molekülatom zu vermeiden (vgl. Abb. 4-8) [447]. Dies macht die Interpretation von Konturdiagrammen schwierig, da die dargestellten Regionen häufig nicht zusammenhängend sind. Ein weiterer Nachteil dieser Methode liegt in der Vernachlässigung entropischer Effekte [448].

In diesem Zusammenhang stellt die CoMSIA-Methode (*Comparative molecular similarity indices analysis*), entwickelt im Jahr 1994 von KLEBE ET AL., eine Modifikation der CoMFA-Methode dar. Die Abstandsabhängigkeit der Potenziale wird mithilfe von Gaußfunktionen approximiert (s. Abb. 4-8). Ein Grenzwert ist daher nicht erforderlich. Die Konturdiagramme sind weniger stark fragmentiert und leichter zu interpretieren, da die Bereiche am Molekül hervorgehoben werden, welche die Aktivität beeinflussen, während die CoMFA-Konturdiagramme nur Regionen außerhalb der Moleküle darstellen, in denen eine Wechselwirkung erwartet wird. [448]

Statt auf Wechselwirkungsfeldern beruht die CoMSIA-Methode auf der Berechnung von Ähnlichkeitsindizes zwischen Sonde und Molekül j an jedem Gitterpunkt q mit dem Programm SEAL [465]: [448]

$$A_{F,k}^q(j) = - \sum_{i=1}^n w_{\text{probe},k} w_{ik} e^{-\alpha r_{iq}^2} \quad (28)$$

Der Ähnlichkeitsindex A am Gitterpunkt q als Summe über alle n Atome i des untersuchten Moleküls j hängt von den physikochemischen Eigenschaften k des Sondenatoms ($w_{\text{probe},k}$) und der Atome i (w_{ik}) des zu untersuchenden Moleküls ab. r_{iq} beschreibt den Abstand zwischen Sondenatom und den Molekül-Atomen. Der benutzerdefinierte Attenuierungsfaktor α liegt in der Regel um den Wert 0,3. Je größer α , desto schneller fällt die Abstandsfunktion ab. Kleine Werte für α verbreitern die Gaußfunktion und es werden globale Informationen über die Eigenschaft erfasst. [448, 466]

Neben den sterischen und elektrostatischen Eigenschaften können in der CoMSIA-Methode auch Wasserstoffbrückendonor- und -akzeptor-Eigenschaften und hydrophobe Wechselwirkungen berücksichtigt werden [183]. Die berechneten Ähnlichkeitsindizes werden wie die molekularen Felder bei der CoMFA-Methode mit den biologischen Daten mittels multivariater Datenanalyse (z.B. PLS-Analyse, s. S. 100) korreliert [467].

Der Trainingsdatensatz in 3D-QSAR-Modellen liegt in einer Größenordnung von 28-237 Molekülen und der Testdatensatz bei etwa 6-124 Molekülen [462, 468, 469]. Entscheidend ist die Zusammensetzung der beiden Datensätze. Mit dem 3D-QSAR-Modell können wesentliche Vorhersagen nur dann getroffen werden, wenn die Substituenten der zu testenden Moleküle in Bereichen auftreten, in denen die Moleküle des Trainingssets eine strukturelle Variation aufweisen.

Die Feldbeiträge der Strukturen werden interpoliert und es kann keine Extrapolation auf Bereiche erfolgen, die von den Molekülen des Trainingssets nicht abgedeckt werden [183]. Dies verdeutlicht die Abhängigkeit der Vorhersagekraft des Modells von der Auswahl der Strukturen im Trainingsdatensatz.

Die in dieser Arbeit mit PHASE [470–472] erstellten FIELD-BASED QSAR-Modelle sind ähnlich zu CoMSIA- und CoMFA-Modellen, verwenden aber einen bestimmten Satz an Parametern: Die Atomladungen für die elektrostatischen Felder sowie die sterischen Lennard-Jones-Potenziale stammen aus dem OPLS2005-Kraftfeld. Die Parameter von GHOSE ET AL. [473] werden für die Berechnung der hydrophoben Felder verwendet, während die Donor- und Akzeptorfelder für Wasserstoffbrückenbindungen aus dem Pharmakophorenmodul von PHASE abgeleitet werden [474]. Die Bezeichnung der CoMSIA-ähnlichen Modelle in der FIELD-BASED QSAR-Anwendung von PHASE ist „GAUSSIAN“, hier können sterische (S), elektrostatische (E), hydrophobe (H), und Wasserstoffbrückendonor (D)- und -akzeptorfelder (A) für die Modellerstellung verwendet werden. Die CoMFA-ähnlichen Modelle werden als „FORCE FIELD“ bezeichnet und verwenden nur sterische und/oder elektrostatische Felder [474].

4.8.2 Statistische Auswertung von 3D-QSAR-Modellen

PLS-Analyse

Zur Korrelation der berechneten Wechselwirkungsfelder mit den Aktivitätsdaten wird eine Datenmatrix aus den Aktivitätsdaten des Trainingsdatensatzes und den entsprechenden Feldwerten an den Gitterpunkten in tausenden von Spalten angelegt [457]. Die PLS-Analyse (*Partial least squares*, dt. partielle kleinste Quadrate) als Ansatz der multivariaten Analyse ist für die Datenauswertung besonders gut geeignet, wenn die Zahl der Deskriptoren (unabhängige x-Variablen, hier: Feldwerte) die Zahl der abhängigen y-Variablen (hier: Aktivitätsdaten) deutlich übersteigt, wie es bei der Erstellung von 3D-QSAR-Modellen der Fall ist. Nachteilig ist die schwierigere Interpretation der PLS-Gleichung, da sie im Vergleich zu MLR³²-Modellen sehr abstrakt ist [467].

Mit der PLS-Analyse werden PLS-Vektoren (= latente Variablen, Komponenten) aus den Aktivitätsdaten (Y-Matrix) und den Feldwerten (X-Matrix) extrahiert [183, 457]. Mit diesem Vorgehen wird die Dimensionalität der Daten zugunsten relevanter Daten reduziert [467].

³² Multiple lineare Regression

Diese Vektoren sollen

1. die maximale Varianz der X-Matrix darstellen und somit effektiv die x-Variablen abdecken und
2. möglichst gut mit den Aktivitätsdaten (Y-Matrix) korrelieren und somit eine gute Vorhersagekraft besitzen. [475]

Der größte Varianzanteil wird von der ersten Komponente beschrieben, während jede zusätzliche Komponente einen immer kleineren zusätzlichen Beitrag zur Varianz liefert [476]. Jede zusätzliche Komponente ergibt ein neues QSAR-Modell [257], in der das Ausmaß der Beiträge der Gitterpunkte für die biologische Aktivität angegeben wird [183].

Die Ergebnisse der PLS-Analyse können mit statistischen Kenngrößen bewertet werden. Das Bestimmtheitsmaß R^2 ist ein Maß dafür, wie gut ein bestimmter Datensatz modelliert wird [477]:

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2} = 1 - \frac{\text{RSS}}{\text{TSS}} \quad (29)$$

y_i – experimentell bestimmte Werte

\bar{y} – experimenteller Mittelwert

\hat{y}_i – berechnete (= vorhergesagte) Werte für die Strukturen im Trainingsdatensatz

RSS – *Residual sum of squares* (Residuale Quadratsumme): bezeichnet die Summe der Quadrate der Differenz zwischen den experimentell ermittelten Werten (y_i) und den durch das Modell vorhergesagten Werten (\hat{y}_i); diese Abweichung wird als „Residuum“ bezeichnet und beschreibt die Ungenauigkeit des Modells;

TSS – *Total sum of squares* (Gesamtsumme der Quadrate): bezeichnet die Summe der Quadrate der Differenz zwischen experimentell bestimmten Werten (y_i) und ihrem Mittelwert (\bar{y})

Ein gutes Modell hat eine kleine Residuenquadratsumme (RSS) und ein hohes R^2 (perfektes Modell: $R^2=1$). R^2 steigt mit der Anzahl der Komponenten ($\#k$) des Modells, d.h. der Fit zwischen den X- und Y-Blöcken wird verbessert. Wenn jedoch zu viele Komponenten ausgewählt werden, kann es vorkommen, dass das Modell überangepasst ist (*Overfitting*) [477]. Es ist empfehlenswert, die Zahl der Komponenten nicht weiter zu erhöhen, wenn die Standardabweichung (SD) der Regression etwa dem experimentellen Fehler entspricht [474].

SD gibt die Standardabweichung der Regression an: [406]

$$\text{SD} = \sqrt{\frac{\text{RSS}}{df_2}} \quad (30)$$

mit $df_2 = n - m - 2$

(df_2 – Anzahl der Freiheitsgrade, n – Anzahl der Moleküle im Trainingsset, m – Anzahl der PLS-Faktoren im Modell)

Mit zunehmender Anzahl von Molekülen im Trainingsset und Komponenten im Modell nimmt die Standardabweichung ab und R^2 zu. F macht eine Aussage über die allgemeine Signifikanz des Modells: [406]

$$F = \frac{\text{TSS}/df_1}{\text{RSS}/df_2} \quad (31)$$

mit $df_1 = m + 1$

Die Varianz des Modells (Zähler) hängt mit der Varianz der experimentell ermittelten Aktivität (Nenner) zusammen. Hohe F -Werte weisen auf eine hohe statistische Signifikanz hin. [474]

Interne Validierung

Die optimale Anzahl der Komponenten $\#k$ und die statistische Signifikanz der Ergebnisse aus der PLS-Analyse werden durch eine Kreuzvalidierung als interne Validierungsmethode überprüft [478]. Dazu werden ein (*leave-one-out, loo*) oder mehrere (*leave-more-out, lmo*) Liganden iterativ aus dem Trainingsdatensatz entfernt, das Modell mit den verbleibenden Strukturen erstellt und dann die Aktivität der jeweils herausgenommenen Substanz vorhergesagt. Der R^2_{cv} -Wert ($cv = \text{cross-validation}$) kann verwendet werden, um eine Aussage über die Robustheit und Qualität des Modells zu treffen [479]:

$$R^2_{cv} = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2} \quad (32)$$

y_i – experimentell bestimmte Werte

\bar{y} – experimenteller Mittelwert

\hat{y}_i – berechnete (= vorhergesagte) Werte für jeden Liganden, der nicht im Trainingsatz enthalten ist

R^2_{cv} kann Werte zwischen $-\infty$ und $+1$ (perfektes Modell) annehmen. Für eine ausreichende Robustheit sollte das Modell einen R^2_{cv} -Wert von mindestens 0,5 besitzen [480]. Das Modell mit dem höchsten R^2_{cv} enthält die optimale Anzahl an Komponenten [481]. Allerdings reicht der R^2_{cv} -Wert als alleiniges Kriterium für den Beweis der Vorhersagekraft eines QSAR-Modells nicht aus. Dafür muss das Modell zusätzlich mit einem externen Testdatensatz validiert werden [482].

Externe Validierung

Die Vorhersage der Aktivitäten eines Testdatensatzes ist notwendig, um die Vorhersagekraft eines trainierten Modells zu testen. Der Testdatensatz besteht aus Liganden, die denen aus dem Trainingsdatensatz ähnlich sind, aber nicht zur Erstellung des Modells verwendet wurden. Das Modell hat eine akzeptable Vorhersagekraft, wenn der Q^2 -Wert größer als 0,6 ist. [482]

Im Vergleich zum Regressionskoeffizienten R^2 , welcher sich auf den Trainingsdatensatz bezieht, berücksichtigt Q^2 den Testdatensatz. Ansonsten ist die Formel identisch: [477]

$$Q^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2} \quad (33)$$

Im Gegensatz zu R^2 bezieht sich \hat{y}_i hier auf die berechneten Werte im Testdatensatz.

Der RMSE (*Root mean square error*) ist die Wurzel aus dem mittleren Fehlerquadrat und gibt an, wie gut die Funktion an die Daten des Testdatensatzes angepasst ist: [406, 477]

$$\text{RMSE} = \sqrt{\frac{1}{n_T} \sum_{j \in T} (\hat{y}_j - y_j)^2} \quad (34)$$

T – Testsatz der Moleküle

n_T – Anzahl der Moleküle im Testdatensatz

Wie klein der RMSE sein sollte, hängt von der beabsichtigten Verwendung des Modells ab. Der RMSE-Wert sollte klein sein, allerdings besteht ein Unterschied, ob Aktivitätswerte beispielsweise im nano- oder mikromolaren Bereich vorhergesagt werden sollen. [477]

Als Schwellenwerte für die Akzeptanz eines QSAR-Modells wurden in dieser Arbeit für $R^2_{cv} > 0,5$ und $Q^2 > 0,6$ festgelegt. Gleichzeitig wurde ein hoher R^2 -Wert und kleine Werte für SD und RMSE angestrebt. Erfüllten mehrere Modelle (unterschiedliche $\#k$) einer PLS-Analyse diese Kriterien, wurde ein Kompromiss gewählt zwischen möglichst hohen R^2_{cv} - und Q^2 -Werten und einer möglichst niedrigen Komponentenzahl, um ein *Overfitting* zu vermeiden. Führt eine höhere Komponentenzahl $\#k$ nicht zu einer signifikanten Verbesserung dieser Werte, wurde das Modell mit der niedrigeren Komponente bevorzugt.

Das Ergebnis der PLS-Analyse ist mit Konturdiagrammen visuell interpretierbar. In diesen farbigen Darstellungen werden Volumenbereiche dargestellt, in denen die Änderung der Struktur und damit der Feldbeiträge zu einer höheren oder niedrigeren Bindungsaffinität führt [483]. Für eine Leitstrukturoptimierung stellt dies eine Möglichkeit dar, die Struktur an solchen Stellen gezielt zu verändern, um die Affinität zu erhöhen [183]. Die Analyse von Konturdiagrammen wurde für einige Strukturen beispielhaft durchgeführt.

4.9 Zusätzlich verwendete Analysesoftware

4.9.1 POVME

POVME (*POcket Volume MEasurer*) [484–486] ist ein in Python geschriebenes Programm zur Bindetaschenanalyse durch Beschreibung der Form, des Volumens und der Flexibilität von Hohlräumen. Dafür wird eine Voxel-/Gitter-basierte Darstellung der Bindetaschen genutzt. [485]

Voraussetzungen und Ablauf für die Berechnung des Bindetaschenvolumens

POVME benötigt als Input die strukturellen Informationen des Zielproteins im PDB-Format. Für den Vergleich verschiedener Bindetaschenkonformationen (z.B. aus einer MD-Trajektorie) ist die Überlagerung der Snapshots der Trajektorie erforderlich, da der POVME-Algorithmus für eine Trajektorie immer die gleichen festgelegten Koordinaten verwendet und die Bindetasche daher nicht räumlich versetzt oder gedreht werden darf [486]. Für dieses Alignment wurde das Programm VMD [487] (*Visual Molecular Dynamics*) verwendet. Mit diesem Programm ist auch die Konvertierung verschiedener Moleküldateiformate in das PDB-Format möglich.

Nach dem Alignment legt der Benutzer die Einschlussregion (*Inclusion sphere*) und optional Ausschlussregion(en) (*Exclusion sphere(s)*) fest. Die Einschlussregion sollte alle Bindetaschenkonformationen der Trajektorie abdecken. Mit der Ausschlussregion definiert der Benutzer Bereiche der Einschlussregion, die für die Berechnung des Bindetaschenvolumens nicht berücksichtigt werden sollen, da sie z.B. nicht Teil der Bindetasche sind (s. Abb. 4-9). Die Einschlussregion wird von POVME mit Kugeln gleichen Abstands (*Grid spacing*, standardmäßig: 1,0 Å) ausgefüllt und anschließend werden die Kugeln, welche in der Ausschlussregion liegen, wieder entfernt.

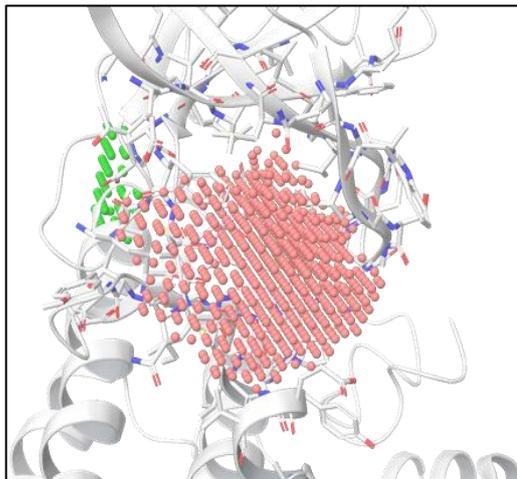


Abb. 4-9 – Exemplarische Darstellung der übrig gebliebenen Punkte in der Bindetasche (rot) nach Entfernen der an das Protein angrenzenden Atome. Die grünen Punkte liegen in einer vom Benutzer definierten Ausschlussregion, da sie nicht mit der Bindetasche verbunden sind, und werden von der Volumenberechnung ausgeschlossen.

Daraufhin werden alle Kugeln in der Nähe von Proteinatomen³³ entfernt, sodass idealerweise nur die Kugeln in der Bindetasche zurückbleiben. Optional können Kugeln entfernt werden, die aus der Bindetasche in den Lösungsmittelraum herausragen. Jede Kugel beinhaltet ein identisches Volumen. Bei einem Abstand der Punkte von 1,0 Å hat jede Kugel ein Volumen von 1,0 Å³.

³³ Die „Nähe“ wird definiert als der vdW-Radius des Proteinatoms zuzüglich einer benutzerdefinierten Toleranz (standardmäßig: 1,0 Å; 1,09 Å entspricht dem Radius eines Wasserstoffatoms). [486]

Zur Berechnung des Bindetaschenvolumens werden die Volumina aller übrig gebliebenen Kugeln, die in der Bindetasche liegen, aufaddiert. [486]

In dieser Arbeit wurde die POVME-Version 3.0 zur Berechnung des Volumens und der Oberfläche der vier verschiedenen Bindetaschen der DLK-Röntgenkristallstrukturen und der MD-Konformationen verwendet (s. Kap. 5.1.1 u. 5.1.3).

4.9.2 KNIME

KNIME (*Konstanz Information Miner*) [488] ist eine Software für die Verarbeitung und Analyse großer Datenmengen. Die graphische Benutzeroberfläche bietet die Möglichkeit, Module für die Datenverarbeitung, -analyse und -visualisierung via Drag-and-Drop miteinander zu verbinden und auf diese Weise einen visuellen und interaktiven Workflow zu erstellen. Mit Plug-Ins (z.B. auch für die SCHRÖDINGER-Software) sind unterschiedliche Module integrierbar, sodass KNIME je nach den Ansprüchen des Anwenders erweitert werden kann. Die Module, bezeichnet als *Nodes* (dt. Knoten), stellen Algorithmen dar, die bestimmte Aufgaben erfüllen (s. Abb. 4-10). Je nach Funktion besitzen sie Ein- und/oder Ausgänge, um sie mit anderen Nodes zu verbinden. So gibt es Nodes für das Einlesen von Daten (s. Abb. 4-10, 1.). Die eingelesenen Daten können durch das Hinzufügen von Datenverarbeitungs-Nodes z.B. gefiltert oder zusammengefügt werden (s. Abb. 4-10, 2.). Auch mathematische Operationen z.B. für statistische Analysen oder Clustering sind

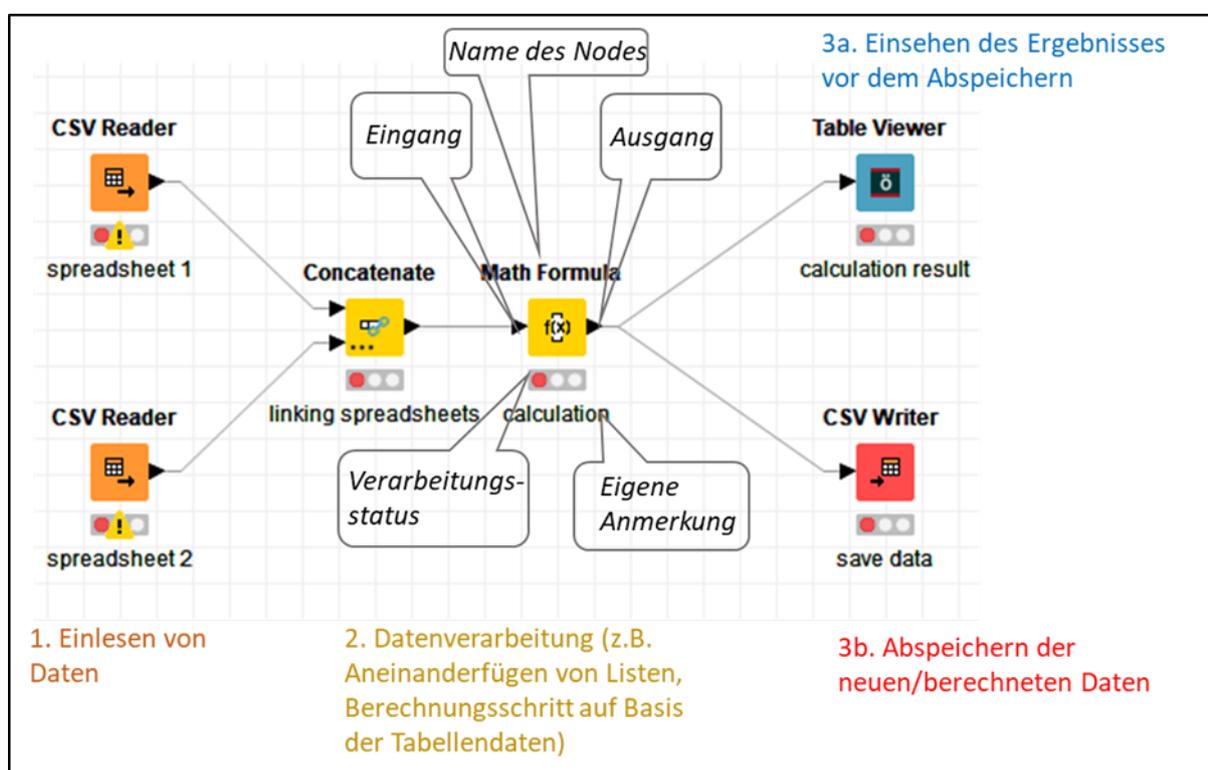


Abb. 4-10 – Exemplarische Darstellung eines kurzen KNIME-Workflows bestehend aus verschiedenen Nodes zum Einlesen (orange), zur Datenverarbeitung (gelb) und zum Abspeichern der generierten Daten (rot)

möglich. Als Abschluss eines Workflows wird häufig ein Node zur Visualisierung oder zum Abspeichern der Daten genutzt (s. Abb. 4-10, 3a. bzw. 3b). Zwischenergebnisse lassen sich einfach veranschaulichen, indem die Informationen jedes Nodes abgegriffen werden können. So können weitere Teilschritte einfach und schnell angepasst werden. [488]

Der größte Vorteil von KNIME ist die Möglichkeit zur Verarbeitung großer Datenmengen. Die Millionen Liganden konnten in Form von Tabelleneinträgen verarbeitet werden. Mit KNIME wurde der Consensuscore berechnet (s. Kap. 5.5.2 u. 7.9.6) und die Ergebnisse der verschiedenen Scorings (s. Kap. 7.9.6) zusammengefügt. Ranglisten aus verschiedenen Grids konnten miteinander vereint und Tabellen für die Erstellung der Anreicherungskurven generiert werden. Die Korrelationskoeffizienten (s. Kap. 4.4.8 u. Kap. 7.6) zwischen experimentellen Aktivitätsdaten und den Scorewerten wurden mit KNIME berechnet. Darüber hinaus wurden die RDKit-Fingerprints (s. Kap. 4.7.1 u. 7.9.7) mit einem KNIME-Workflow generiert. Die Funktionen der in dieser Arbeit verwendeten Nodes werden im Anhang (A, S. 341 f.) erläutert.

4.9.3 Gnuplot

GNU PLOT [337] ist ein Grafikprogramm zur Erstellung von 2D- und 3D-Plots, welches über die Kommandozeile oder über Skripte gesteuert wird. Für folgende grafische Darstellungen wurde GNU PLOT v.4.6 aufgrund der effektiven Verarbeitung großer Datenmengen verwendet:

- Erstellung der Anreicherungskurven
- Erstellung von Histogrammen zur Darstellung der Häufigkeitsverteilung der verschiedenen Scorewerte bzw. Moleküleigenschaften

4.9.4 R

R [489] ist eine freie Software für Statistik-Berechnungen und grafische Darstellungen. Sie wurde unter Verwendung des Erweiterungspakets „drc“ (*Analysis of dose-response curves*) [490] für die Erstellung der Dosis-Wirkungs-Kurven und deren statistische Evaluation basierend auf den Messergebnissen der Verdünnungsreihen im ADP-GloTM Kinase Assay genutzt (s. Kap. 4.10.2 u. 5.6.1).

4.10 Biologische Prüfungen

Für die biologische Prüfung der ausgewählten Substanzen aus dem virtuellen Screening (s. Kap. 5.6) auf ihre DLK-inhibitorische Aktivität sollten Methoden ausgearbeitet werden.

Die DLK überträgt als Phosphoryltransferase das γ -Phosphat von ATP auf Serin- und Threoninreste eines Substrats. In biochemischen Assays zur Messung der Kinaseaktivität wird daher häufig die Menge des phosphorylierten Substrats bestimmt. Dies kann zum einen mithilfe von radioaktiv markiertem ATP in γ -Position (^{32}P oder ^{33}P) erfolgen. Nach Ablauf der Kinasereaktion wird die Menge des phosphorylierten Substrats aufgrund des radioaktiven Anteils mittels Szintillationszählung bestimmt [491]. Eine weitere Möglichkeit zur Quantifizierung des phosphorylierten Substrats ist der Einsatz substratspezifischer Antikörper z.B. mittels ELISA (*Enzyme-linked immunosorbent assays*). Mithilfe von Waschschritten vor der Detektion werden eventuell vorhandene Störverbindungen entfernt. Die Detektion basiert i.d.R. darauf, dass ein enzymgekoppelter Sekundärantikörper hinzugefügt wird, welcher ein zugesetztes Substrat in einen Farbstoff umwandelt. Dieser wird mit einem Photometer quantifiziert [492]. Desweiteren ist der Einsatz fluoreszenzbasierter Methoden möglich. Dazu zählen FRET (*Fluorescence resonance energy transfer*)- bzw. TR (*time-resolved*)-FRET-Assays sowie die Fluoreszenzpolarisation (FP). Im FRET-Assay regt ein Donor-Fluorophor (z.B. gebunden am Kinase-Substrat) über Energietransfer ein in räumlicher Nähe befindliches Akzeptor-Fluorophor (z.B. spezifischer Antikörper gegen das phosphorylierte Substrat) zur Fluoreszenzemission an, welche gemessen werden kann. Im Gegensatz zu FRET werden in TR-FRET-Assays Fluorophore mit langer Fluoreszenzlebensdauer (z.B. Lanthanide) verwendet und die Fluoreszenz nach der Anregung als Funktion der Zeit aufgenommen [493]. Bei der Fluoreszenzpolarisation (FP) wird linear polarisiertes Licht in die Probe eingestrahlt. Durch Bindung eines fluoreszenzbasierten Liganden an ein Protein verändert sich seine Rotationsbewegung, wodurch der zu messende Polarisationswert (= Anisotropie) zunimmt. Auch hier werden häufig spezifische Antikörper gegen das phosphorylierte, in diesem Fall fluoreszenzmarkierte, Substrat eingesetzt [491].

Die Testsysteme in dieser Arbeit sollten möglichst schnell und einfach durchführbar sein, möglichst ohne lange Inkubationszeiten und sich wiederholende Waschschriffe. Sie sollten geringe Falsch-positiv- bzw. Falsch-negativ-Raten sowie eine geringe Störanfälligkeit durch andere Substanzen aufweisen. Für die Arbeit mit radioaktivem Material stand kein entsprechend ausgestattetes Labor zur Verfügung.

In dieser Arbeit wurden drei unterschiedliche Verfahren zur Überprüfung der hemmenden bzw. bindenden Eigenschaften der ausgewählten Substanzen eingesetzt: Im ADP-Glo™ Kinase Assay erfolgt die Messung der Kinaseaktivität basierend auf der Quantifizierung des Verbrauchs von ATP bzw. der Bildung von ADP [494]. Mit der *Microscale Thermophoresis* (MST) wird die Bewegung eines fluoreszenzmarkierten Moleküls (hier: DLK) entlang eines Temperaturgradienten gemessen, welche durch Bindungsprozesse beeinflusst wird [495]. Desweiteren fand eine externe Hit-Testung mit dem KINOMEscan®-Assay von Eurofins DiscoverX (San Diego, CA, USA) zur Überprüfung der laboreigenen Methoden statt. Dieser setzt DNA-markierte DLK in einem Kompetitionsassay ein und quantifiziert die an der freien DLK gebundene DNA mittels RT-PCR (*Real-time polymerase chain reaction*) [496, 497]. Die Messprinzipien dieser drei Testsysteme werden ab S. 110 detailliert vorgestellt.

Eingesetzte DLK in dieser Arbeit

Bei dem exprimierten Protein zur Anwendung der in dieser Arbeit selbst durchgeführten Assays handelt es sich um die MBP-TEV-DLK₁₄₀₋₅₁₇, welche am Universitätsklinikum Eppendorf im Arbeitskreis von Prof. Dr. Oetjen von Dr. Jorge Enrique Duque Escobar hergestellt und aufgereinigt wurde. Zur Herstellung wurde die kodierende DNA-Sequenz der murinen DLK₁₄₀₋₅₁₇ zuvor in den pMAL-c2X-Vektor inkloniert, welcher einen MBP-Tag (*Maltose binding protein*, benötigt für die Aufreinigung mittels Affinitätschromatographie) besitzt. Zusätzlich wurde von ihm eine TEV (Tobacco-Etch-Virus)-Spaltungserkennungssequenz hinzugefügt. Die Aminosäuresequenz 140-517 beinhaltet die katalytische Domäne der DLK sowie beide Leucin-Zipper-Motive. Bei dem im letzten Schritt der Proteinaufreinigung verwendete Elutionspuffer handelt es sich um einen Tris-Puffer (pH = 8,0), dem zusätzlich 10 % Glycerol zur Vermeidung der Degradierung des Proteins bei -80 °C zugesetzt wurde (s. Kap. 7.15, S. 286, Tab. 7-8). [498]

Bei der im KINOMEscan® eingesetzten DLK handelt es sich um die humane Wildtyp-DLK in voller Länge [499].

Referenzsubstanzen

Als Referenzsubstanzen bei den DLK-Assays wurden Substanzen mit bekannter DLK-Bindungsaffinität eingesetzt: Sunitinib (**2**), Crizotinib (**3**), Bosutinib (**4**) und GNE-3511 (**7**) (s. Abb. 4-11).

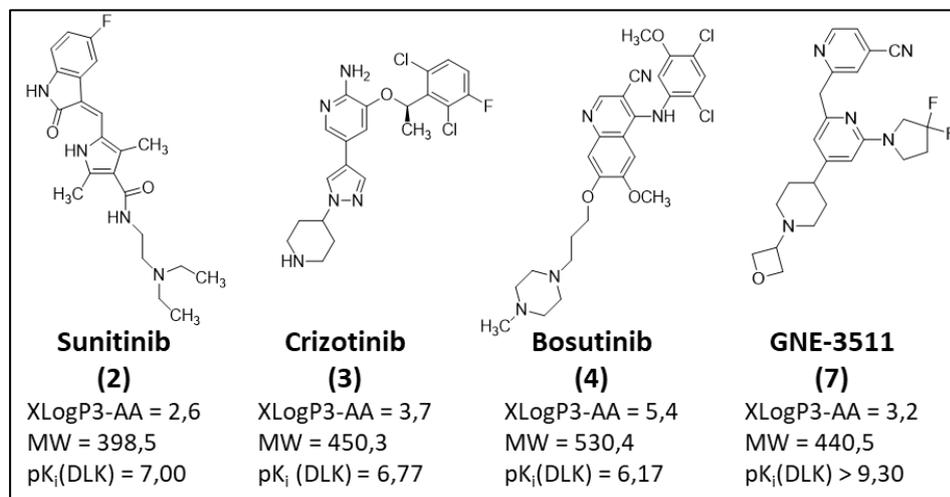


Abb. 4-11 – Referenzsubstanzen für die Bioassays; Daten für XLogP3-AA und MW stammen aus der PubChem-Datenbank [500]; pK_i-Werte für Crizotinib (**3**) und Bosutinib (**4**) (= SKI-606) aus FERRARIS ET AL. (2013) [125], für GNE-3511 (**7**) aus PATEL ET AL. (2014) [166], für Sunitinib (**2**) aus KARAMAN ET AL. (2008) [164].

4.10.1 Semiquantitative Bestimmung der DLK-Menge mit SDS-PAGE

Da bei der Aufreinigung des MBP-TEV-DLK₁₄₀₋₅₁₇-Proteins durch Dr. Duque Escobar mehrere Fraktionen der Proteinlösung unterschiedlicher Konzentrationen aufgefangen wurden, mussten diese jeweils mittels semiquantitativer SDS-PAGE bestimmt werden.

Mit der von U. LAEMMLI entwickelten SDS-PAGE (Sodium dodecylsulfate polyacrylamide gel electrophoresis) werden Proteine anhand ihrer Molekülmasse (kDa) in einem elektrischen Feld auf einem SDS-Polyacrylamidgel als Träger aufgetrennt. [501]

Zunächst wurden die einzelnen Fraktionen in jeweils der gleichen Menge mit der SDS-PAGE analysiert und nur solche Fraktionen mit einer Bande starker Intensität bei 94 kDa (für MBP-TEV-DLK₁₄₀₋₅₁₇) weiterverwendet. Die semiquantitative Bestimmung erfolgte dann durch Erstellung einer Verdünnungsreihe der ausgewählten Fraktionen im Vergleich zu einer MBP-TEV-DLK₁₄₀₋₅₁₇-Positivkontrolle (1 µg in 7,5 µl Lösung).

Das Gel für die SDS-PAGE (s. Kap. 7.15.1, S. 287, Tab. 7-9) setzt sich aus dem Sammelgel (pH = 6,8) zum Anreichern der Proteine vor der eigentlichen Trennung und dem Trenngel (pH = 8,8) zur Auftrennung der Proteine nach ihrer Größe zusammen. Diese werden im Vorfeld selber hergestellt. Die verwendeten Proben werden mit einem Loading Dye versetzt. Dieser enthält u.a. das SDS, welches die Proteine mit negativen Ladungen „befrachtet“, sodass die Proteine vom Plus-

Pol angezogen werden. Desweiteren enthält es Mercaptoethanol, welches Disulfidbrücken spaltet und somit zur Denaturierung beiträgt. In eine der Probentaschen für die SDS-PAGE wird „Ladder X“ als Größenmarker zur Abschätzung der Molmassen der Proben gegeben.

Die Herstellung des Gels und die Durchführung der Elektrophorese erfolgten analog zu der Arbeit von Dr. Duque Escobar [498] (s. Kap. 7.15.1).

4.10.2 ADP-Glo™ Kinase Assay

Eine Möglichkeit zur Quantifizierung des ATP-Verbrauchs einer Kinase ist der Einsatz der Luciferase: dieses Enzym aus dem Glühwürmchen katalysiert die Reaktion von ATP mit Luciferin zu Oxiluciferin und AMP unter Freisetzung von Lichtquanten. Mit dem Kinase-Glo™ Assay von PROMEGA kann das in der Kinasereaktion nicht umgesetzte ATP gemessen werden. Vorteilhaft ist die Anwendbarkeit des Assays auf alle Kinasen und Substrate [493]. Desweiteren werden keine spezifischen Antikörper benötigt. Es können Kinase-Assay-Kits von diversen Herstellern käuflich erworben werden.

Jedoch ist es möglich, dass die Luciferase durch kleine Moleküle gehemmt wird. In einer Untersuchung von KASHEM ET AL. [502] konnte aber herausgestellt werden, dass die Wahrscheinlichkeit sehr gering ist. Nur drei von 10.208 Strukturen konnten die Luciferase um mehr als 30 % hemmen. Für ein robustes Signal mit ausreichendem Signal-Rausch-Verhältnis im Kinase-Glo™ Assay ist es erforderlich, dass ein ausreichend hoher ATP-Verbrauch von mindestens 20 % stattfindet. Dies kann dazu führen, dass höhere Enzymkonzentrationen eingesetzt werden müssen. Eine Verbesserung hinsichtlich der Robustheit und Sensitivität bei geringer Kinasekonzentration ist der nachfolgend entwickelte ADP-Glo™ Kinase Assay. Hier wird nicht das restliche ATP, sondern das gebildete ADP durch den ATP-Verbrauch gemessen [494]. Ein Kinase-Inhibitor würde zu einem Signalanstieg im Kinase-Glo™ Assay, aber zu einem verringerten Signal im ADP-Glo™ Kinase-Assay führen [491]. In dieser Arbeit sollte der ADP-Glo™ Kinase-Assay aufgrund der Verfügbarkeit des Fertig-Kits, der postulierten leichten Anwendbarkeit und der hohen Sensitivität und Robustheit durchgeführt werden.

Für die Durchführung werden Multiwell-Platten verwendet und ein Reaktionsvolumen von 12,5 µl oder 25 µl eingesetzt. Der Assay wird in drei Schritten durchgeführt (s. Abb. 4-12):

Nach der Kinasereaktion, in der die DLK in einer 12,5 µl bzw. 25 µl Lösung das zu phosphorylierende Substrat (in dieser Arbeit: β -Casein) unter Verbrauch von ATP phosphoryliert und dabei ADP produziert („0. Schritt“, s. Abb. 4-12), wird im ersten Schritt das gleiche Volumen an ADP-

Glo™ Reagenz (12,5 µl bzw. 25 µl) hinzugefügt, um die Kinasereaktion zu beenden und das verbliebene ATP zu verbrauchen (40 min. Inkubation). Im zweiten Schritt wird das Kinase-Detektions-Reagenz (25 µl bzw. 50 µl) hinzugefügt, um das entstandene ADP in ATP zu überführen und dieses ATP wiederum mit

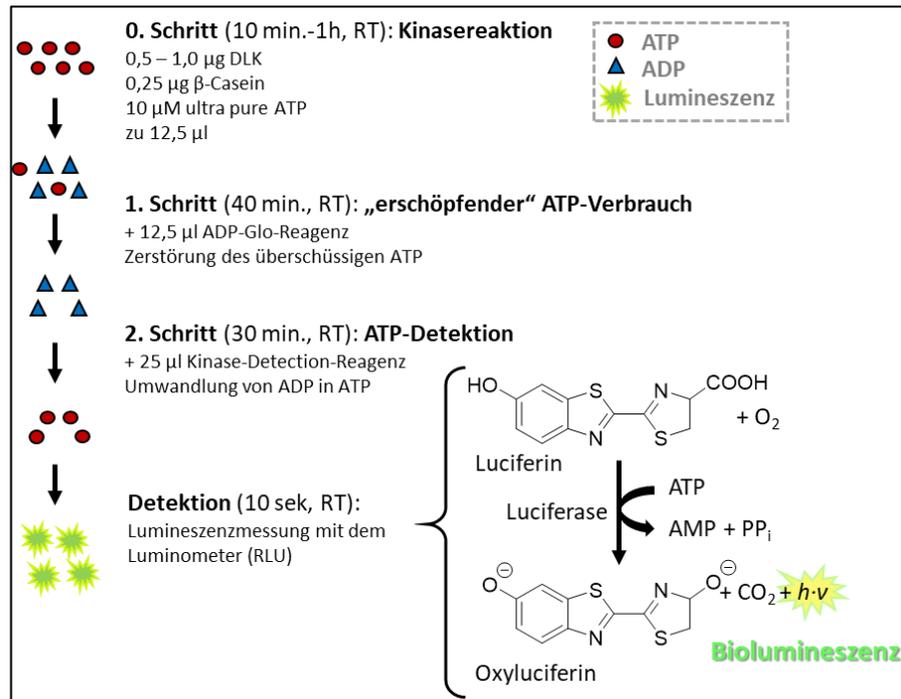


Abb. 4-12 – Links: Ablauf des ADP-Glo™ Kinase Assays. Abb. modifiziert nach PROMEGA-Handbuch [503]. Rechts: Ablauf der Luciferase/Luciferin-Reaktion (rechts), Abb. modifiziert nach MÜLLER-ESTERL (2018) [492], S. 540.

einer Luciferase/Luciferin-Reaktion zu messen (s. Abb. 4-12, rechts). [503]

Mit diesem Assay kann eine ADP-Konzentration ab 20 nM und generiertes ADP linear bis zu einer äquivalenten ATP-Konzentration von 1 mM detektiert werden. Das Lumineszenzsignal (Einheit: RLU = *Relative Light Units*) verhält sich proportional zur ADP-Konzentration und korreliert mit der Kinaseaktivität. [503]

Manche Chemikalien können mit der Luciferase-Reaktion interferieren. Generell gibt es eine Empfehlung aus dem Handbuch des Assay-Kit-Anbieters PROMEGA, welche Konzentrationen dieser Stoffe maximal toleriert werden (s. Tab. 4-3). Dies spielt eine Rolle bei der Auswahl des Lösungsmittels für die zu testenden virtuellen Hits und die Referenzsubstanzen. In unserem Assay sollten die potenziellen Inhibitoren in 100 % DMSO gelöst werden. Für die Kinasereaktion darf eine DMSO-Konzentration von 5 % nicht überschritten werden. Dies hat Auswirkungen auf die maximal einsetzbare Inhibitor-Konzentration im Assay.

Tab. 4-3 – Mit dem ADP-Glo™ Kinase Assay kompatible Lösungsmittel und Chemikalien [503]

Chemikalie	Max. tolerierte Konzentration
NaCl	200 mM
CaCl ₂	2,5 µM
Calmodulin	5 µM
DTT	10 mM
MgCl ₂	50 mM
MnCl ₂	5 mM
DMSO	5 %
Natriumorthovanadat	500 µM
Tergitol-NP-9	0,1 %
Tween-20	0,2 %
Triton X-100	0,2 %

Ablauf der Assay-Vorbereitung und der Hit-Testung

1. *Bestimmung der Aktivität der DLK-Frak-tionen:* Die Kinaseaktivitäten der einzelnen DLK-Frak-tionen wurden mit dem ADP-Glo™ Kinase Assay im Vergleich einer Positivkontrolle bekann-ter Konzentration (1 µg MBP-TEV-DLK₁₄₀₋₅₁₇ in 7,5 µl Lösung) bestimmt. Dabei sollte das Volu-men an DLK eingesetzt werden, welche laut der semiquantitativen SDS-PAGE-Bestimmung 1 µg DLK enthält.
2. *Optimierung der Substratkonzentration:* Mit einer ausgewählten Fraktion, welche vergleich-bare Ergebnisse wie die Positivkontrolle erzielt, wurde die optimale Substratkonzentration (β-Casein) bestimmt. Dazu wird eine serielle Verdünnung von β-Casein eingesetzt und die Lumineszenzmessung nach Durchführung des ADP-Glo™ Kinase-Protokolls mit und ohne Kinase durchgeführt. Die optimale Substratkonzentration erhält man bei der größten Diffe-renz in der Lumineszenz zwischen den Wells mit stattfindender Kinasereaktion und den Wells ohne Kinase [503].
3. *Optimierung der DLK-Konzentration:* Laut Herstellerangaben ist ein Vorteil des ADP-Glo™ Kinase Assays, dass selbst sehr kleine ADP-Mengen mit einem hohen Signal-Rausch-Verhält-nis detektiert werden [503]. Daher reicht es aus, eine kleine DLK-Menge einzusetzen, sodass nur eine geringe ATP-zu-ADP-Konversion stattfindet. Für die Bestimmung der optimalen DLK-Konzentration wurde eine DLK-Verdünnungsreihe hergestellt. Die optimale DLK-Kon-zentration erhält man aus der Kinase-Titrationskurve an der Stelle, wo das Lichtsignal im li-nearen Bereich liegt. HENNEK ET AL. [504] setzten für die Standardisierung von Kinaseaktivitä-ten eine Kinasemenge ein, die einer 10-30 %igen Konversion von ATP zu ADP entsprach bei einer ATP-Konzentration von 10 µM. Dies sollte in dieser Arbeit mit einer entsprechenden ATP-ADP-Konversionskurve analog erreicht werden.
4. *Aktivitätsbestimmung mit Referenzsubstanzen und den virtuellen Hits:* Mit der optimierten DLK- und Substratmenge und einer ATP-Konzentration von 10 µM wurde die Auswirkung der Referenzsubstanzen (s. S. 109, Abb. 4-11) auf das Messsignal mit einer Konzentration von 500 µM überprüft. Unter gleichen Bedingungen wurden die virtuellen Hits getestet. Sub-stanzen mit einem deutlichen Hemmsignal sollten in einer Verdünnungsreihe erneut getes-tet werden, um einen IC₅₀-Wert (da ATP-abhängig) zu erhalten.

Modellierung der Messdaten zur IC_{50} -Bestimmung

Für die Erstellung der IC_{50} -Kurven und deren statistische Evaluation wurde die freie Software R [489] unter Verwendung des Erweiterungspakets „drc“ (*Analysis of dose-response curves*) genutzt [490].

Die hierfür genutzten Gleichungen für die Anpassung der Daten an das 4-PL-Modell (4-Parameter-Logarithmische Funktion, Hill-Funktion [505], s. Formel (35a)) und an das 5-PL-Modell (5-Parameter-Logarithmische Funktion [506], s. Formel (35b)) lauten:

$$f(x) = c + \frac{d-c}{1+(\frac{x}{e})^b} \quad (35a)$$

$$f(x) = c + \frac{d-c}{(1+(\frac{x}{e})^b)^f} \quad (35b)$$

Die Variablenbezeichnungen wurden aus R übernommen. Für die Parameter des 4-PL-Modells bzw. des 5-PL-Modells gelten für die praktische Anwendung der Kurvenanpassung auf die Messergebnisse des ADP-Glo™ Kinase Assays: [490, 506]

- b - 4-PL: „Hillslope“ = Steigung an der steilsten Stelle der Kurve
- 5-PL: beeinflusst zusammen mit f die Steigung $(\frac{-bf}{2^{1+f}})$ der Kurve
- c - Messsignal der Asymptote, an der sich die Anpassungskurve bei höchster Inhibitorkonzentration (niedrigste Kinaseaktivität) annähert (= Minimum)
- d - Messsignal der Asymptote, an der sich die Anpassungskurve bei niedrigster Inhibitorkonzentration (höchste Kinaseaktivität) annähert (= Maximum)
- e - 4-PL: Inhibitorkonzentration am Wendepunkt der Kurve (= IC_{50})
- 5-PL: e beeinflusst die Position des Übergangsbereich und somit auch den IC_{50} -Wert
- x - Inhibitorkonzentration
- f - nur im 5-PL-Modell: beeinflusst zusammen mit b die Annäherungsgeschwindigkeit an die Asymptote c (vgl. Steigung: $(\frac{-bf}{2^{1+f}})$)

Der IC_{50} -Wert entspricht demnach im 4-PL-Modell dem Parameter e . Aus dem 5-PL-Modell berechnet er sich nach: $IC_{50} = e \cdot (2^{\frac{1}{f}} - 1)^{\frac{1}{b}}$ [506].

Die 5-PL-Gleichung (35b) unterscheidet sich von der 4-PL-Gleichung (35a) durch einen zusätzlichen Parameter (f), welcher einer möglichen Asymmetrie der Kurve Rechnung trägt. Das 5-PL-Modell ist daher besser geeignet, wenn die Dosis-Wirkungs-Kurve nicht symmetrisch verläuft [507]. Für den statistischen Vergleich der beiden Modelle wurde eine ANOVA (*Analysis of variance*)-Analyse mit R durchgeführt (F-Test, s. Kap. 5.6.1) [490]. Dadurch wird überprüft, ob das 4-PL-Modell für die Kurvenanpassung ausreicht. Ziel ist eine gute Datenanpassung mit einem

Modell möglichst geringer Parameteranzahl, da eine Überparametrisierung zu einer Verzerrung der Konzentrationsschätzungen führen kann [506].

4.10.3 Microscale Thermophoresis (MST)

Zur Untersuchung der Bindung der Testsubstanzen an die MBP-TEV-DLK₁₄₀₋₅₁₇ wurde die *Microscale Thermophoresis* (MST) eingesetzt. Mit der MST wird die Bewegung von Molekülen entlang eines mikroskopischen Temperaturgradienten gemessen, welche von verschiedenen bindungsinduzierten Veränderungen der molekularen Eigenschaften (z.B. Molekülgröße, Ladung, Solvatationshülle) abhängt. Ein IR-Laser wird zur Erzeugung des Temperaturgradienten eingesetzt. Die Aufzeichnung des Messsignals erfolgt über eine Fluoreszenzmessung [495]. Zur Untersuchung von Protein-Ligand-Interaktionen muss entweder der Ligand oder das Protein fluoreszenzmarkiert sein. Dieses Labeling ist mit dem Fertig-Reagenzsatz und den Aufreinigungssäulen von NanoTemper relativ einfach durchzuführen. Das Messgerät Monolith NT.115 von NanoTemper Technologies GmbH wurde verwendet. Vorteile ergeben sich aus den schnell durchführbaren Messungen an dem Gerät sowie aus dem geringen Probenverbrauch. Darüber hinaus eignet sich die MST als ergänzende Methode zum ADP-Glo™ Kinase-Assay, da es sich um ein anderes Messprinzip handelt.

Erzeugung des Temperaturgradienten und Fluoreszenzmessung

„*Thermophoresis describes the directed diffusion of molecules through temperature gradients, typically from a hot to a cold region.*“ (Gerätehandbuch Monolith®NT.115 [508]). Dieser Temperaturgradient wird durch einen IR-Laser lokal begrenzt in der Probenkapillare erzeugt. Die Diffusion des Proteins wird durch die Bindung eines Liganden beeinflusst. Die Bindung kann die Größe des Gesamtkomplexes und der Solvatationshülle beeinflussen. Ebenso kann die Schnittstelle zwischen Molekül und Lösungsmittel sowie ggf. auch die Gesamtladung des Moleküls verändert werden. Diese bindungsinduzierten Änderungen resultieren in eine veränderte Beweglichkeit, welche ein verändertes MST-Signal zur Folge hat. Dieser messbare Unterschied ermöglicht die Berechnung von Bindungsaffinitäten. Für die Messbarkeit muss einer der beiden Bindungspartner fluoreszenzmarkiert sein. Eine LED im Messgerät regt die Fluoreszenz an. Die gemessene Fluoreszenzintensität verändert sich zum einen durch den Temperaturanstieg selbst (TRIC = *Temperature related intensity change*, beschrieben durch $c \frac{\partial F}{\partial T}$) und zum anderen durch

die Konzentrationsänderung durch die verstärkte Diffusion aus der erwärmten Region heraus ($F \frac{\partial c}{\partial T}$). Die Gesamtveränderung der Fluoreszenz nach der IR-Laser-Aktivierung ergibt sich daher aus: [508]

$$\frac{\partial}{\partial T}(cF) = F \frac{\partial c}{\partial T} + c \frac{\partial F}{\partial T} \quad (36)$$

Es gibt zwei wichtige veränderbare Parameter am MST-Messgerät: Zum einen kann die MST-Power eingestellt werden. Dieser Wert gibt die Laserleistung an, welche den Temperaturgradienten erzeugt. Je größer die MST-Power, desto schneller kommt der Temperaturgradient zustande und desto größer ist dieser [508]. Desweiteren wird mit der LED-Power festgelegt, wie stark das Fluorophor angeregt wird. Mit Erhöhung der LED-Power erreicht man ein verstärktes Fluoreszenzsignal [509]. Die LED-Power kann zwischen 5 und 95 % eingestellt werden, während 40 % MST-Power bestenfalls nicht überschritten werden sollten [510, 511].

Labeling der DLK

Für die MST-Messungen musste die DLK mit einem Fluoreszenzmarker kovalent verbunden werden. Hierfür wurde das Monolith NT™ Protein Labeling Kit RED-NHS verwendet. Dieses Kit ist für die Markierung von Proteinen mit einem Molekulargewicht über 5 kDa optimiert und enthält alle Reagenzien und Materialien, die zur Durchführung der Markierungsreaktion und zur Trennung der resultierenden Konjugate vom überschüssigen Farbstoff erforderlich sind. Spin- und Schwerkraftsäulen werden verwendet, um einen Pufferaustausch durchzuführen und das markierte Protein zu reinigen. Der reaktive NT-647-NHS-Farbstoff³⁴ (NHS = N-Hydroxysuccinimid) reagiert effizient mit primären Aminofunktionen des Proteins zu Amiden. Primäre Amine befinden sich z.B. in Lysinen, welche häufig an der Proteinoberfläche zum Lösungsmittel ausgerichtet und daher gut für die Labeling-Reaktion geeignet sind. Die so markierten Proteine besitzen ein Fluoreszenzexcitationsmaximum von 650 nm und ein Fluoreszenzemissionsmaximum von 670 nm. [512]

Das Labeling der DLK wurde gemäß Herstellerprotokoll [512] durchgeführt (s. Kap. 7.17). Vor dem eigentlichen Labeling musste ein Pufferaustausch durchgeführt werden, weil der Puffer, in dem das MBP-TEV-DLK₁₄₀₋₅₁₇-Protein vorliegt, primäre Amine (Tris-HCl) enthält, welche den Farbstoff binden würden, und das reduzierend agierende DTT (Dithiothreitol), welches laut Handbuch von NanoTemper ebenso vermieden werden soll. Die Zusammensetzung des Labeling-Puffers wird vom Hersteller nicht preisgegeben. Es könnte sich um einen PBS-Puffer oder eine vergleichbare

³⁴ Die Strukturformel des Farbstoffs ist urheberrechtlich geschützt und ist daher nicht bekannt.

Puffersubstanz ohne primäre Amine und ohne freie Thiolgruppen handeln. Nach dem Labelingprozess wird mit dem MST-optimiertem Puffer (Tris-Puffer, pH 7,4, s. Kap. 7.17, S. 294, Tab. 7-19) gespült, welcher laut Hersteller für die Herstellung von Verdünnungen für die MST-Messungen geeignet ist [511].

Interpretation von MST-Diagrammen und K_d -Bestimmung

Das Ergebnis für ein typisches Bindungsexperiment mit der MST ist das Diagramm der normalisierten Fluoreszenzzeitspur (s. Abb. 4-13), welche die Veränderung der Fluoreszenz über die Zeit wiedergibt.

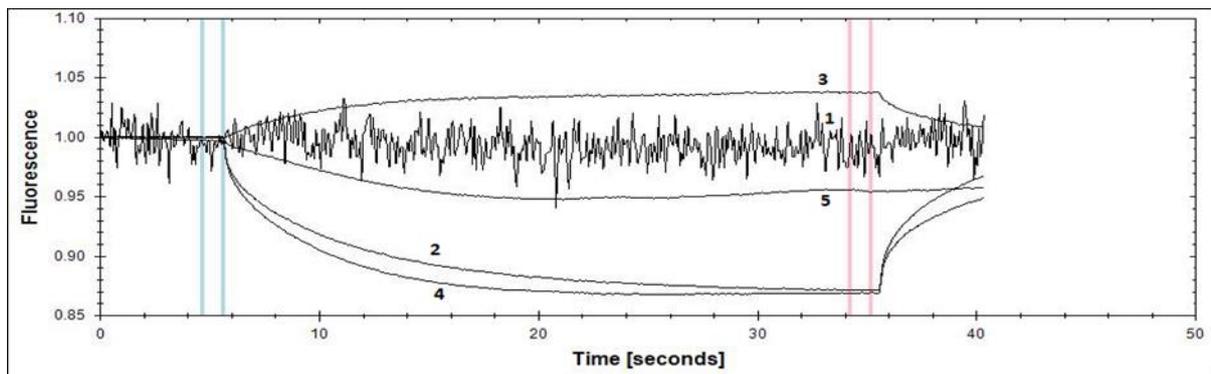


Abb. 4-13 – Normalisierte Fluoreszenzzeitspur, aufgenommen mit dem Monolith NT.115; **1** – Blank ohne Protein, **2** – unkomplexiertes Protein; **3,4, 5** – Protein-Ligand-Komplexe mit verschiedenen Liganden

Vor dem Anschalten des IR-Lasers (bis zur zweiten blauen Markierung) sind die fluoreszierenden Proteine bzw. Protein-Ligand-Komplexe homogen verteilt und es wird eine initiale Fluoreszenz detektiert. Nach dem Anschalten des IR-Lasers zur Erhitzung der Kapillare (ab der zweiten blauen Markierung) verändert sich das Fluoreszenzsignal wegen der thermophoretischen Bewegung der fluoreszierenden Moleküle aus dem erhitzten Bereich heraus. Nach dem steilen Abfall der Fluoreszenzintensität direkt nach dem Anschalten des IR-Lasers (*T-jump*, Temperatursprung) wird nach 10-30 Sekunden ein Steady-State erreicht, weil sich die Wärmediffusion und die gewöhnliche Diffusion ausgleichen. Nach Ausschalten des IR-Lasers (hinter der zweiten roten Markierung) steigt das Fluoreszenzsignal wieder an, weil die Moleküle aufgrund der Temperaturabnahme zurückdiffundieren. [512, 513]

Die gemessenen Fluoreszenzwerte werden als normalisierte Fluoreszenz (F_{norm}) angegeben, welche sich aus der Fluoreszenz vor (F_{cold}) und nach (F_{hot}) der Laseraktivierung ergibt:

$$F_{\text{norm}} = \frac{F_{\text{hot}}}{F_{\text{cold}}} \quad (37)$$

Die Werte für F_{hot} und F_{cold} entsprechen dem berechneten Durchschnitt der Fluoreszenzintensitäten zwischen den definierten Bereichen, die durch die beiden roten bzw. blauen Markierungen festgelegt sind [513]. Die Position der roten Markierungen kann hier z.B. auch im Bereich des Temperatursprungs liegen oder manuell festgelegt werden [508, 509]. Je höher die zugesetzte Ligandenkonzentration, desto stärker führt dies zu einer Änderung der Thermophorese und somit zu einer Änderung von F_{norm} . Eine Bindungskurve wird erhalten, indem man die normalisierte Fluoreszenz F_{norm} gegen den dekadischen Logarithmus der verschiedenen Ligandenkonzentrationen (serielle Verdünnung) aufträgt. Von dieser sigmoid verlaufenden Bindungskurve kann die Bindungskonstante K_d abgeleitet werden [513].

Da die Bindung eines ATP-kompetitiven DLK-Inhibitors 1:1 in der ATP-Bindetasche der DLK erfolgt, wurde das K_d -Modell der Software MO.AFFINITY ANALYSIS verwendet. Die Berechnung des K_d erfolgt mit dieser Software mithilfe des Massenwirkungsgesetzes: [508]

$$K_d = \frac{[A] \cdot [L]}{[AL]} \quad (38a)$$

K_d – Dissoziationskonstante

[A] – Konzentration des freien, fluoreszierenden Proteins (DLK)

[L] – Konzentration des freien Liganden (z.B. Crizotinib)

[AL] – Konzentration des Protein-Liganden-Komplexes

Die freie Liganden- und Proteinkonzentration ergibt sich aus: [508]

$$[A] = [A_0] - [AL] \quad (38b)$$

$$[L] = [L_0] - [AL] \quad (38c)$$

$[A_0]$ und $[L_0]$ als Ausgangskonzentrationen des Proteins bzw. des zugesetzten Liganden sind bekannt. $[AL]$ wird aus der quadratischen Anpassungsfunktion berechnet: [508]

$$[AL] = \frac{1}{2} * (([A_0] + [L_0] + K_d) - (([A_0] + [L_0] + K_d)^2 - 4 * [A_0] * [L_0])^{\frac{1}{2}}) \quad (38d)$$

$[A_0]$ bleibt während der Messungen konstant, während $[L_0]$ innerhalb der Verdünnungsreihe variiert wird. F_{norm} aus der MST-Messung gibt den Anteil $[AL]$ an der Ausgangsproteinkonzentration $[A_0]$ wider und kann daher verwendet werden, um mit den o.g. Gleichungen K_d zu ermitteln. F_{norm} wird in %o gegen die Ligandkonzentration auf einer \log_{10} -x-Achse aufgetragen. [508]

Ablauf der Assay-Vorbereitung und der Hit-Testung

1. *Optimierung der Enzymkonzentration:* Eine serielle Verdünnung des proteinhaltigen Eluats aus dem Labeling- und Aufreinigungsprozess wurde mit den Standardeinstellungen für LED (50 %)- und MST-Power (40 %) einer MST-Messung unterzogen, um die Verdünnung zu ermitteln, welche eine optimale Fluoreszenzintensität erzeugt (optimal: zwischen 200 und 1500 Fluorocounts laut Handbuch). Typischerweise wird für die MST eine Proteinkonzentration von 5-100 nM eingesetzt. Die eingesetzte MBP-TEV-DLK₁₄₀₋₅₁₇-Konzentration in dieser Arbeit für die Testung der Substanzen betrug unter Berücksichtigung der Ergebnisse der semi-quantitativen SDS-PAGE-Bestimmung (s. Kap. 7.15.1) und nach den Verdünnungsschritten rechnerisch 7,1 nM. Eine quantitative Bestimmung war nicht möglich. Desweiteren ist die Bewertung der Peakqualität erforderlich. Die Peaks müssen gleichmäßig und symmetrisch sein.
2. *Präzision des Messsignals:* Die Abweichung der Messsignale zwischen mehreren Kapillaren der gleichen proteinhaltigen Lösung wurde bestimmt. Folgt man den Angaben des Geräteherstellers, so sollten die Fluoreszenzintensitäten nicht stärker als +/- 10 % vom Mittelwert abweichen. [511]
3. *Messung einer Verdünnungsreihe der Referenzsubstanzen:* Mit dem zuvor bestimmten optimalen DLK-Eluatanteil wurden Verdünnungsreihen der Referenzsubstanzen, ausgehend von 500 µM als höchste Konzentration in der zu vermessenden Lösung, aufgenommen und der K_d -Wert bestimmt.
4. *Testung der virtuellen Hits:* Es wurden Einpunktmessungen bei 500 µM der ausgewählten virtuellen Hits im Vergleich zu einer Referenzsubstanz im Rahmen einer Doppelbestimmung durchgeführt. Für Substanzen mit einem abweichenden Messsignal von der Proteinlösung ohne Testsubstanz wurden Verdünnungsreihen zur Bestimmung des K_d -Werts erstellt.

4.10.4 KINOMEScan®

Eurofins DiscoverX (San Diego, CA, USA) bietet die Durchführung von Liganden-Bindungsassays unter Verwendung der KINOMEScan®-Technologie für mehr als 480 Kinasen [497] an.

Diese Assays ermöglichen das Erzielen von robusten und reproduzierbaren Messergebnissen bezüglich der Bindung der Testsubstanz in die ATP-Bindetasche einer ausgewählten Kinase (hier: DLK) [514]. Falls ein bindender Ligand mithilfe der KINOMEScan®-Technologie identifiziert werden kann, ist im Gegensatz zur MST der Bindungsort – nämlich die ATP-Bindetasche der Kinase –

bekannt. Daher sollte eine – wegen der Kosten reduzierte – Auswahl der Hits mit KINOMEScan[®] getestet werden. Für Liganden mit einer signifikanten Bindung im scanELECT[®]-Assay (Einpunktmessung unter Verwendung der KINOMEScan[®]-Technologie) soll im KdELECT[®]-Assay (11-Punktmessung zur Erstellung einer Dose-Response-Kurve) eine K_d -Bestimmung erfolgen.

Messprinzip

Ein Ansatz aus humanen, DNA-markierten Kinasen (DLK in dieser Arbeit) und aus immobilisierten Liganden³⁵ mit bekannter Kinasebindungsaffinität ($< 1 \mu\text{M}$) wird eingesetzt, zu dem die Testsubstanz gegeben wird (s. Abb. 4-14) [496]. Durch Bindung der Testsubstanz an die ATP-Bindungsstelle der Kinase kann der immobilisierte Ligand kompetitiv verdrängt werden. Dies führt zu freier Kinase in der Lösung. Die Menge der an den immobilisierten Liganden gebundenen Kinasen wird reduziert. Falls die Testsubstanz nicht an

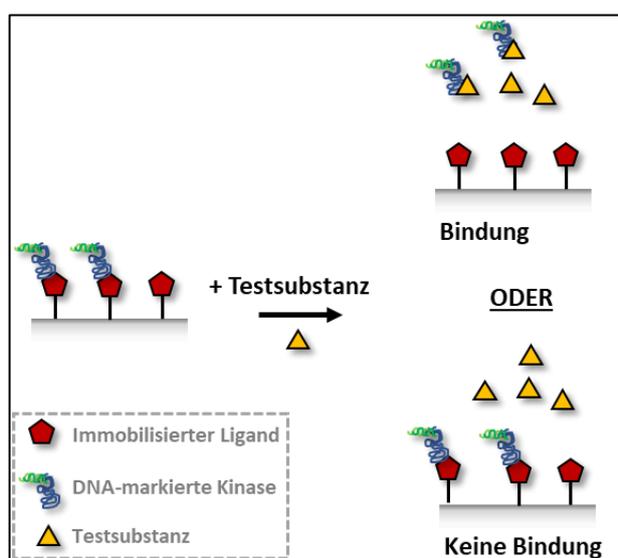


Abb. 4-14 – Prinzip des Bindungsassays KINOMEScan[®]. Abb. modifiziert nach einer Darstellung von DiscoverX [514, 515].

die Kinase bindet, bleibt die Menge der Kinasen gebunden an den Liganden auf dem festen Träger unverändert. Die noch am immobilisierten Liganden gebundene Kinase wird mithilfe der qPCR-Methode (= RT-PCR, quantitative Echtzeit-Polymerase-Kettenreaktion) durch Nachweis der DNA-Markierung an der Kinase gemessen [514].

In den KINOMEScan[®]-Assays wird kein ATP eingesetzt, da die Ergebnisse aufgrund der Messung der Bindung der Testsubstanz statt einer Messung der ablaufenden Kinasereaktion (wie im ADP-Glo[™] Kinase Assay) unabhängig von ATP sind. Daher werden im Gegensatz zum ADP-Glo[™] Kinase Assay keine IC_{50} -, sondern K_d -Werte bestimmt. [514]

Das Ergebnis der Einpunktmessung im scanELECT[®]-Assay wird mit folgender Formel erhalten: [514]

$$\%Ctrl = \left(\frac{\text{Signal mit Testsubstanz} - \text{Positivkontrollsignal}}{\text{Negativkontrollsignal} - \text{Positivkontrollsignal}} \right) * 100 \quad (39)$$

³⁵ Die immobilisierten Liganden sind über Biotin an Streptavidin-überzogenen magnetischen Kügelchen („Beads“) gebunden.

Das Ergebnis wird als Prozentsatz der Kontrolle angegeben. Die Negativkontrolle ist reines DMSO (100 % Ctrl) statt der Testsubstanz. Die Positivkontrolle ist die Kontrollsubstanz (0% Ctrl). Ein Wert von 100 % Ctrl für die Testsubstanz lässt darauf schließen, dass die Substanz nicht bindet, während ein Wert von 0 % Ctrl auf vollständige Verdrängung des immobilisierten Liganden durch die Testsubstanz hinweist. Je geringer also der % Ctrl-Wert, desto stärker die Bindung.

Im KdELECT®-Assay wird aus den Ergebnissen der Durchführung des Assays mit einer Verdünnungsreihe der Testsubstanz eine 11-Punkt-Dosis-Wirkungs-Kurve erstellt. Die Bindungskonstante (K_d) wird anhand einer Standard-Dosis-Wirkungs-Kurve unter Verwendung der Hill-Gleichung berechnet: [515]

$$\text{Response} = \text{Background} + \frac{\text{Signal} - \text{Background}}{1 + \left(\frac{K_d}{\text{Dose}}\right)^{\text{Hill Slope}}} \quad (40)$$

Response	– Effekt durch die eingesetzte Konzentration (Dose) der Testsubstanz
Background	– Effekt ohne Testsubstanz (Wert von Y für die minimale Asymptote)
Signal	– Effekt durch unendlich hohe Konzentration der Testsubstanz (Wert von Y für die maximale Asymptote)
Hill Slope	– Steigung an der „steilsten Stelle“ der Kurve; hier wird für den Hill Slope -1 eingesetzt
K_d	– Bindungskonstante; Wendepunkt der Dosis-Wirkungs-Kurve
Dose	– Konzentration der Testsubstanz

Die Kurven wurden von DiscoverX mithilfe des Levenberg-Marquardt-Algorithmus zur Lösung nichtlinearer Regressionsprobleme nach dem Prinzip der kleinsten Quadrate angepasst [515].

Die virtuellen Hits und die Referenzsubstanzen wurden von DiscoverX zunächst mit einer Konzentration von $10 \mu\text{M}^{36}$ gegen die DLK im scanELECT®-Assay getestet. Hits mit einem Ergebnis < 75 % Ctrl wurden für den nächsten Schritt der K_d -Bestimmung mit KdELECT® eingesetzt. Darüber hinaus wurden zur zusätzlichen Überprüfung auch die Substanzen, welche ein Hemm- bzw. Bindungssignal im ADP-Glo™ Kinase Assay bzw. in der MST ergeben haben, mit KdELECT® bestimmt.

³⁶ Nur GNE-3511 wurde wegen seiner sehr hohen Hemmpotenz als $1 \mu\text{M}$ -Lösung gemessen.

5 Ergebnisse

5.1 Zielstrukturen

5.1.1 Analyse von DLK-Röntgenkristallstrukturen

Zu Beginn dieses Projekts waren vier Röntgenkristallstrukturen der humanen DLK in der Proteindatenbank (PDB [168, 169] verfügbar.³⁷ Die PDB-Codes und einige Zusatzinformationen sind in Tab. 5-1 dargestellt.

Tab. 5-1 – 2015 in der Proteindatenbank veröffentlichte DLK-Röntgenkristallstrukturen von PATEL ET AL. (2015) [70].

PDB-Code	R [Å]	Organismus	Ligand	Sequenz	Fehlende Aminosäuren
5CEN	1,70	Homo sapiens	-	Leu117 - Ile397	-
5CEO	2,28	Homo sapiens	GNE-3511 = Compound 2 in [70] bzw. Compound 26 in [166]	Leu117 - Lys398	Thr257 - Ser269
5CEP	1,99	Homo sapiens	Compound 3 in [70]	Leu117 - Lys398	Thr257 - Ser269
5CEQ	1,91	Homo sapiens	Compound 11 in [70]	Leu117 - Lys398	Thr257 - Ser269

Die Werte für die Auflösung der DLK-Röntgenkristallstrukturen von 1,70-2,28 Å liegen im üblichen Bereich für Proteinstrukturen und werden für den Einsatz im virtuellen Screening als ausreichend hoch angesehen (vgl. dazu Kap. 4.2). Für die Zielstrukturanalyse wurden die Rohdaten der PDBs ohne Bearbeitung mit dem PROTPREP-Tool verwendet. Nur die DLK-Röntgenkristallstruktur mit dem PDB-Code **5CEN** ohne einen gebundenen Liganden (Apoenzym) besteht aus der vollständigen katalytischen Einheit der DLK ohne Unterbrechungen, während in den experimentellen Daten der drei übrigen Strukturen die Daten der Aminosäuren von Thr257 bis Ser269 fehlen. Diese fehlende Sequenz folgt auf das DFG-Motiv und ist Teil des Activation Loops. Wegen der hohen Flexibilität dieses Activation Loops konnte der Loop in der Röntgenstrukturanalyse nicht beobachtet werden (vgl. dazu Kap. 4.2). Zusätzlich fehlt dem Apoenzym im Vergleich zu den anderen Strukturen am C-terminalen Ende eine Aminosäure (Lys398).

Vergleich der Bindetaschenkonformationen

Die Bindetasche des **5CEN** ist durch die Positionierung des P-Loops aufgrund des fehlenden Inhibitors am kleinsten (s. Abb. 5-1). Im Vergleich zu den anderen Röntgenkristallstrukturen gibt es weniger Platz für die Bindung von potenziellen Inhibitoren. Die Bindetaschen von **5CEO**, **5CEP** und **5CEQ** sind an die jeweiligen gebundenen Liganden angepasst. Die Übereinanderlagerung

³⁷ Im Jahr 2017 folgten zwei weitere Röntgenkristallstrukturen der DLK (**5VO1** und **5VO2**), die aber aufgrund des bereits fortgeschrittenen Projekts und der Ähnlichkeit in den Bindetaschen nicht berücksichtigt wurden.

der PDBs zeigt zunächst keine große Diversität der Bindetaschen und keine stark unterschiedliche Ausdehnung der Liganden.

Sowohl die Konformationen des Backbones als auch der Seitenketten der Hinge-Regionen der vier verschiedenen Röntgenkristallstrukturen sind praktisch identisch (s. auch Abb. 5-2 A). Die berechneten RMSD-Werte (s. Kap. 4.6, Durchführung der RMSD-Messungen: s. Kap. 7.1) der Atome der Hinge-Aminosäuren Glu191, Phe192

und Cys193 zwischen **5CEO** und den übrigen Röntgenkristallstrukturen liegen alle in einem sehr niedrigen Bereich von 0,31-0,39 Å. Dadurch wird die Rigidität dieses Bereichs deutlich. Dagegen scheinen der P-Loop und das DFG-Motiv flexibler zu sein (vgl. Abb. 5-2 B): Auffällig sind die unterschiedlichen Konformationen von Gln136 als Teil des P-Loops und von Asp254 als Teil des DFG-Motivs. Während diese Aminosäuren in **5CEP** und **5CEQ** recht ähnlich ausgerichtet sind, gibt es in **5CEO** und **5CEN** abweichende Ausrichtungen. Hier wird eine grundsätzliche Beweglichkeit des P-Loops und des DFG-Motivs im Vergleich zur starren Hinge-Region deutlich.

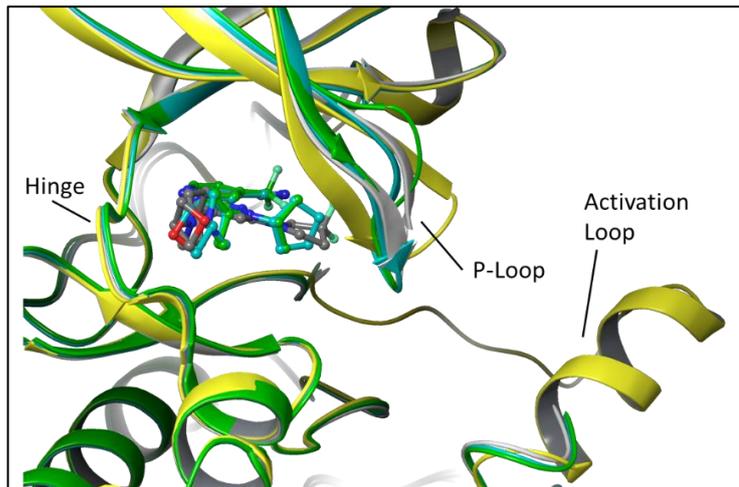


Abb. 5-1 – Übereinanderlagerung der vier PDBs **5CEN** (gelb), **5CEO** (grau), **5CEP** (grün) und **5CEQ** (cyan); **5CEN** ist die einzige PDB mit einem vollständigen Activation Loop (Helix rechts, gelb), außerdem ist der P-Loop im Vergleich zu den anderen PDBs stärker zur Bindetasche hin gerichtet.

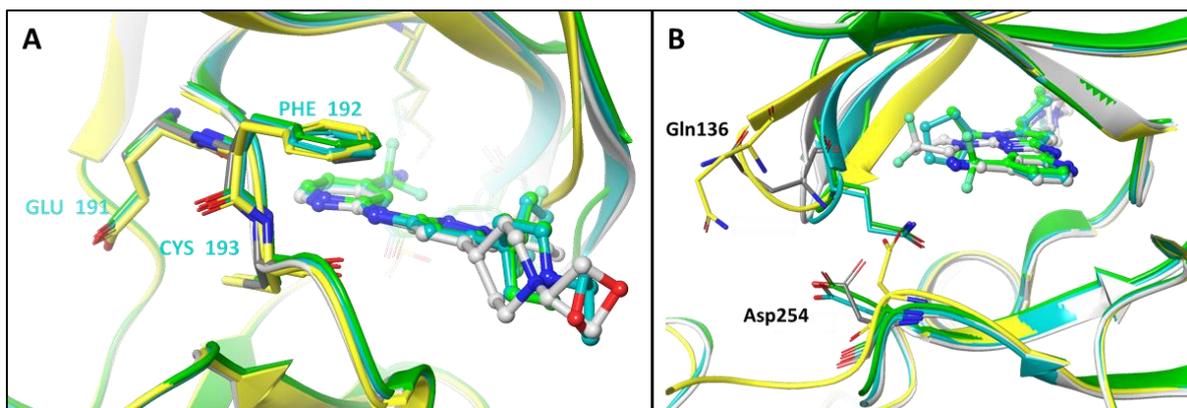


Abb. 5-2 – A – Überlappung der Aminosäuren der Hinge-Region beim Ausrichten der PDBs auf **5CEO**; die Aminosäuren überlappen fast vollständig. **B** – Perspektive von der Rückseite auf die Bindetasche der übereinandergelagerten PDBs; **5CEQ** (cyan) und **5CEP** (grün) überlappen recht gut, während **5CEO** (grau) mit den Aminosäuren Gln136 und Asp254 stark abweicht.

Die Aminosäuren Gln136 bzw. Asp254 können durch ihre Ausrichtungen in einigen der Röntgenkristallstrukturen eine Wasserstoffbrückenbindung zum Lys152 ausbilden (s. Abb. 5-3, A-D): Eine Wasserstoffbrückenbindung ausgehend von Gln136 (P-Loop) zum Lys152 ist in **5CEP** (B) und **5CEQ** (C) möglich, während die Abstände in **5CEN** (D: 5,64 Å) und in **5CEO** (A: 6,14 Å) dafür zu groß sind. Eine Wasserstoffbrückenbindung zwischen Asp254 und Lys152 ist nur in **5CEN** zu beobachten. Das Asp254 in **5CEN** nimmt etwa die Position des Gln136 in **5CEP** und **5CEQ** ein.

Alle Strukturen zeigen die für den aktiven Zustand typische DFG-in-Konformation, d.h. die Seitenkette des Asp254 zeigt zur Bindetasche hin. Dies wird auch durch die Nähe zu Lys152 deutlich. Hier beträgt der Abstand bei **5CEN** lediglich 2,85 Å, während dieser Abstand in den drei ligandhaltigen PDBs über 4 Å beträgt. Dennoch liegt auch in diesen Strukturen die DFG-in-Konformation vor, weil das Asp254 nicht vollständig von der Bindetasche weg orientiert ist. Bei den gebundenen Liganden

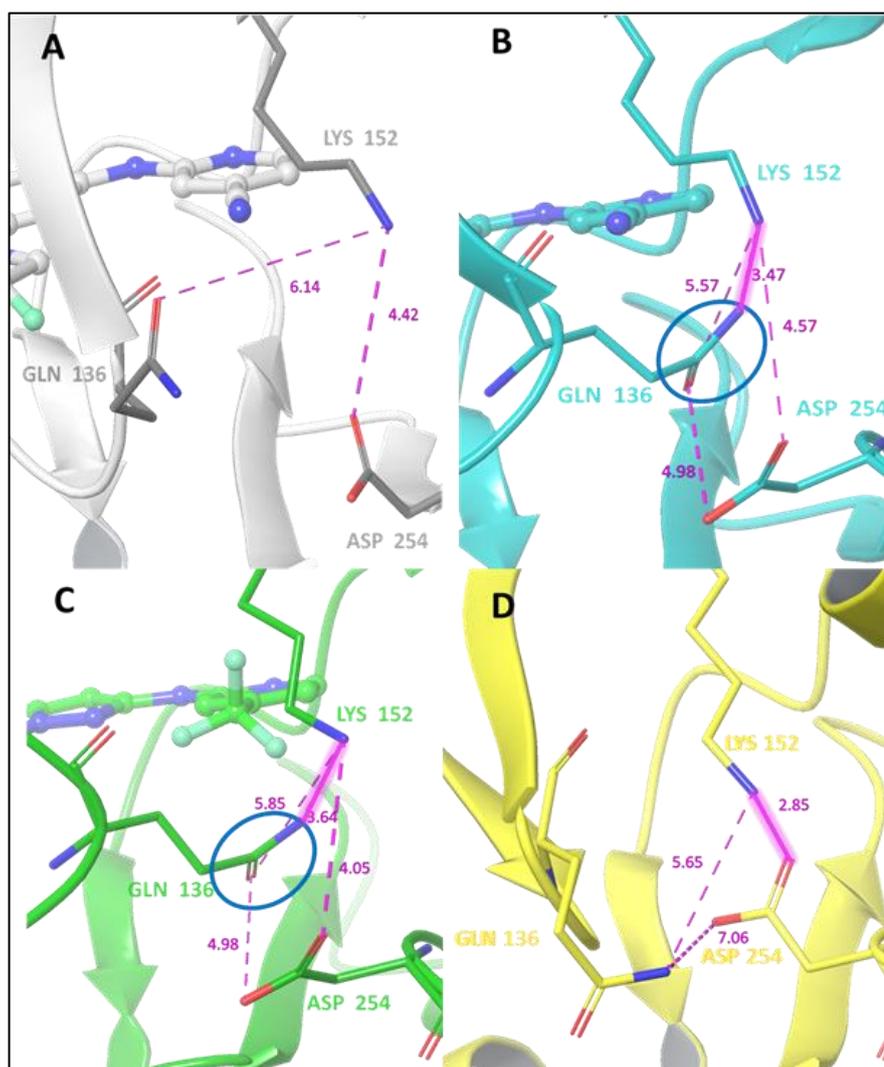


Abb. 5-3 – Räumliche Orientierung der Aminosäuren Gln136 (P-Loop), Lys152 und Asp254 (DFG-Motiv) zueinander; **A** – **5CEO** (grau), **B** – **5CEP** (cyan), **C** – **5CEQ** (grün), **D** – Apo-Kristallstruktur **5CEN** (gelb). In allen Strukturen liegt die DFG-in-Konformation vor (aktiver Zustand). Mögliche Wasserstoffbrückenbindungen werden pink leuchtend mit durchgezogener Linie dargestellt. In **5CEP** und **5CEQ** ist die Atomzuordnung in Gln136 (blau eingekreist) falsch – die korrekte Zuordnung erfolgte mithilfe von PROTREP.

handelt es sich um Typ-I-Inhibitoren. Diese binden an die DFG-in-Konformation einer Kinase.

Um eine Aussage über das zur Verfügung stehende Platzangebot für potenzielle Liganden treffen zu können, wurde die Größe der Bindetaschen analysiert und verglichen. Dafür wurde zum einen der Abstand zwischen den Alpha-Kohlenstoff (C_{α})-Atomen von Met190 (Gatekeeper) und Ala135 (P-Loop) gemessen, zum anderen der Abstand zwischen den C_{α} -Atomen von Phe192 (Hinge-Region) und Gly132 (P-Loop) (vgl. Abb. 5-4). Darüber hinaus wurden die Volumina der vier verschiedenen Bindetaschen mit dem Bindetaschenanalyse-Tool POVME (vgl. Kap. 4.9.1 u.

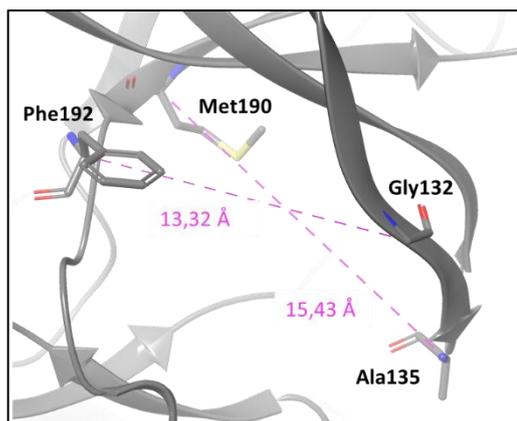


Abb. 5-4 – Messung der Abstände zwischen zwei Paaren von Aminosäuren in der Bindetasche der Röntgenkristallstrukturen, dargestellt am Beispiel **5CEO**. Je größer die Abstände, desto größer die Bindetasche und desto mehr Platz bietet sie für das Docking.

Kap. 7.2) berechnet. Die Ergebnisse dieser Berechnungen (s. Tab. 5-2) zeigen, dass die Größe der Bindetasche in der Reihenfolge **5CEN** < **5CEP** < **5CEQ** < **5CEO** zunimmt. Dies entspricht auch der Reihenfolge für die Größe der darin enthaltenen Liganden (s. Abb. 5-5). Der größere Abstand zwischen den gewählten Aminosäuren geht mit einer Vergrößerung des Volumens der Bindetasche einher.

Tab. 5-2 – Vergleich der vier Bindetaschen in Bezug auf Größe und Ausdehnung. Dist = Abstand zwischen den Alpha-C-Atomen; Vol = Bindetaschenvolumen (berechnet mit POVME)

PDB	Vol [\AA^3]	Dist [\AA] M190-A135	Dist [\AA] F192-G132
5CEN	613	13,89	11,90
5CEO	806	15,43	13,32
5CEP	717	15,35	12,67
5CEQ	725	15,26	12,80

Gebundene Liganden und ihre Position in der Bindetasche

Die in den DLK-Röntgenkristallstrukturen gebundenen Liganden (s. S. 121, Tab. 5-1) stammen aus der Arbeitsgruppe von Patel [70, 166] und gehören strukturell zum gleichen Chemotyp. **GNE-3511 (7)**, der Ligand in **5CEO**, ist der potenteste DLK-Inhibitor unter den drei Liganden (DLK $K_i < 0,0005 \mu\text{M}$). Der Aufbau der drei Liganden folgt dem typischen Aufbau der bekannten, sehr selektiven DLK-Inhibitoren und besteht aus den folgenden strukturellen Einheiten³⁸ (s. Abb. 5-5):

- **Core 1**: Ein tiefsitzender Aromat geht – gemeinsam mit der Aminofunktion, welcher Core 1 mit Core 2 verknüpft – drei Wasserstoffbrückenbindungen mit der Hinge-Region der DLK-Bindetasche ein. Im Falle der **5CEO-(7)**, **5CEP-(18)** und **5CEQ-Liganden (8)** handelt es sich um einen Pyridinring, welcher in *ortho*-Position über eine Aminogruppe mit Core 2 verbunden ist. Desweiteren ist in *para*-Position ein kleiner Substituent (R1) vorhanden.

³⁸ Die hier verwendeten Bezeichnungen der strukturellen Einheiten und deren Funktionen werden im Detail in Kapitel 5.2.1 erläutert.

- *R1*: Der Substituent ragt in eine tief liegende, vorrangig hydrophob ausgekleidete Tasche der DLK-Bindetasche hinein. Der Substituent – in GNE-3511 (**7**) und im **5CEQ**-Liganden (**8**) eine Cyano-Gruppe, im **5CEP**-Liganden (**18**) eine Trifluormethyl-Gruppe – bildet als Akzeptor eine Wasserstoffbrücken-

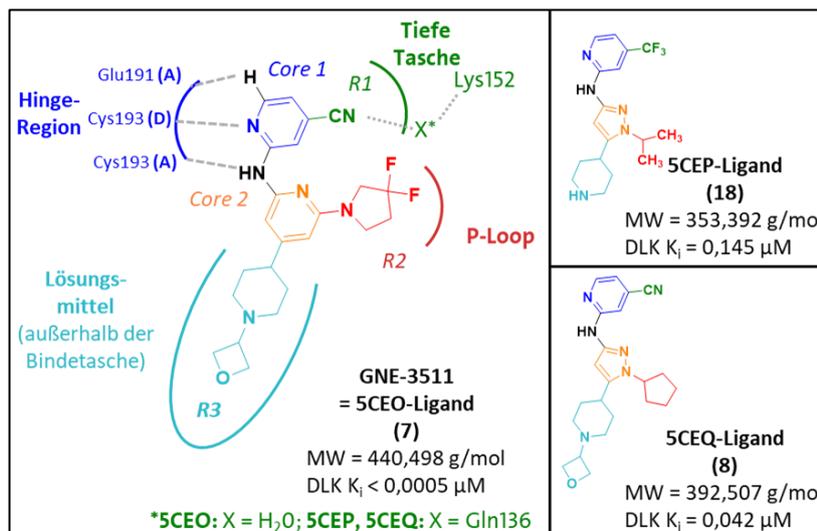


Abb. 5-5 – Aufbau und Interaktionen des GNE-3511 (**7**) (**5CEO**-Ligand) in der DLK-Bindetasche (links). Zum Vergleich: Aufbau des **5CEP**- (**18**) und **5CEQ**-Liganden (**8**) (rechts).

bindung zu einem Wassermolekül (**5CEO**) bzw. zu Gln136 (**5CEP**, **5CEQ**) aus, welches wiederum mit Lys152 über eine Wasserstoffbrückenbindung wechselwirkt.

- *Core 2*: Als zentraler Baustein des Liganden übernimmt Core 2 vorrangig eine rein geometrische Funktion zur räumlichen Anordnung der verschiedenen Substituenten oder Molekülteile. Beim GNE-3511 (**7**) handelt es sich um einen Pyridinring. Ein gemeinsames Merkmal, das zwischen dem **5CEQ**- (**8**) und **5CEP**-Liganden (**18**) hervorzuheben ist, ist der Pyrazolring anstelle des Pyridinrings in dieser Position. Core 2 trägt an der 3. und 5. Position zur 1. Substitution jeweils einen Substituenten, im Folgenden als R2 und R3 bezeichnet.

- *R2*: Der eine Substituent ist primär hydrophob und interagiert über vdW-Kräfte mit dem P-Loop. Bei GNE-3511 (**7**) handelt es sich um einen Difluorpyrrolidin-, beim **5CEP**-Liganden (**18**) um einen Isopropyl- und beim **5CEQ**-Liganden (**8**) um einen Cyclopentylrest.

- *R3*: Bei dem weiteren Substituenten an Core 2 handelt es sich um einen nach außen in das Lösungsmittel hineinragenden Piperidinring, welcher im Falle der **5CEO**- (**7**) und **5CEQ**-Liganden (**18**) seinerseits am Stickstoff mit einem Oxetan substituiert ist.

Um die Position des Liganden in der Bindetasche zu beschreiben, wurden Abstände von bestimmten Ligandenatomen zu definierten Aminosäuren, die alle Richtungen in der Bindetasche abdecken, ermittelt. Man kann hier vier wesentliche Kontaktregionen herausstellen:

1) Wasserstoffbrückenbindungen zur Hinge-Region:

Die charakteristischen Interaktionspartner der Hinge-Region sind die Carbonylsauerstoffatome des Glu191- und Cys193-Backbones als Akzeptoren und das Amidwasserstoffatom des Cys193-Backbones als Donor. In ihrer räumlichen Ausrichtung bilden sie die für die Hinge-Region von Kinasen typische Abfolge *Akzeptor-Donor-Akzeptor*. Die gemessenen Abstände zwischen 2,79 und 3,25 Å von den Ligandenatomen zur Hinge-Region (s. Tab. 5-3) entsprechen alle den Anforderungen für eine Wasserstoffbrückenbindung. Diese drei Wasserstoffbrückenbindungen werden später für das Docking von besonderer Bedeutung sein, da sie eine optimale Positionierung des Liganden in der Tasche ermöglichen. In der Mehrzahl der Inhibitor-Kinase-Interaktionen stellen sie ein konserviertes Merkmal dar. Die Abstände zwischen den Ligandenatomen und der Hinge-Region unterliegen nur geringen Abweichungen, wenn man die einzelnen PDBs miteinander vergleicht (s. Tab. 5-3).

2) Hydrophobe Wechselwirkungen und eine WBB (**5CEQ/5CEP**) zum P-Loop:

Zur rechten Seite wird die Bindetasche durch den P-Loop begrenzt, der in den Abstandsberechnungen durch Gly132 definiert wird. Im Gegensatz zur Abstandsberechnung zur Hinge-Region wurden die Abstände zum C_α-Atom von Gly132 zu verschiedenen Ligandenatomen berechnet, da sich der Substituent R2, der mit dem P-Loop interagiert, zwischen den drei Liganden unterscheidet. Die Abstände von rund 4 Å zum P-Loop und die hydrophoben Interaktionspartner auf beiden Seiten sprechen für hydrophobe Wechselwirkungen in diesem Bereich. Durch die gleichartige Ausrichtung der P-Loop-Aminosäure Gln136 in Richtung des Liganden in **5CEP** und **5CEQ** (vgl. Abb. 5-3 und Abb. 5-6) ergibt sich im Gegensatz zu **5CEO** eine mögliche Wasserstoffbrückenbindung zum Substituenten R1. In **5CEO** ist genau an dieser Position ein Wassermolekül zu finden, welches dem GNE-3511 (**7**) als Wasserstoffbrückenbindungspartner zur Verfügung steht und gleichzeitig eine Wasserstoffbrückenbindung zum Lys152 und Asp254 ausbildet (s. Abb. 5-6).

Tab. 5-3 – Berechnete Abstände [Å] zwischen den Atomen definierter Aminosäuren zum nächsten Atom des entsprechenden Liganden (angegeben in Klammern). SW = Sandwich. In der letzten Zeile wird der Interaktionstyp (Int.) angegeben: A = Akzeptor, D = Donor, H = hydrophob, E = elektrostatisch.

PDB	E191O <i>Hinge</i>	C193N <i>Hinge</i>	C193 O <i>Hinge</i>	G132 C _α <i>P-Loop</i>	Q136 N/O <i>P-Loop</i>	M190 S <i>Rückwand</i>	K152 N <i>Rückwand</i>	A150C _β <i>SW oben</i>	L243C _{δ1} <i>SW unten</i>
5CEO	3,25	3,05	2,83	3,88 (F)	8,85 (N)	4,02	5,93 (N)	3,47	3,46
5CEP	3,15	2,92	2,79	4,05 (C)	3,07 (F)	4,01	5,79 (F)	3,52	3,47
5CEQ	3,11	2,88	2,80	3,77 (C)	3,67 (N)	3,90	5,68 (N)	3,46	3,44
\bar{x}	3,17	2,95	2,88	3,90	5,20	3,98	5,80	3,48	3,46
<i>s</i>	0,07	0,09	0,02	0,14	3,18	0,07	0,13	0,03	0,02
Int.	A	D	A	H	D/E	H	E	H	H

3) Kontakte zur „Rückwand“ der Bindetasche:

Der Gatekeeper Met190 begrenzt wie auch Lys152 die Bindetasche nach hinten. Die kürzesten Abstände der verschiedenen Liganden zum Schwefelatom des Gatekeepers Met190 als rückwärtiger Teil der Bindetasche sind nahezu gleich (s. S. 126, Tab. 5-3). Es handelt sich hierbei um hydrophobe Interaktionen. Die Abstände zum Lys152 wurden zum Stickstoffatom des endständigen primären Amins ermittelt. Hier ist, wie in Abb. 5-6 dargestellt, ein Wassermolekül (**5CEO**) oder Gln136 (**5CEP/5CEQ**) zwischengeschaltet und es kommt dadurch zur Ausbildung eines Wasserstoffbrückennetzwerks.

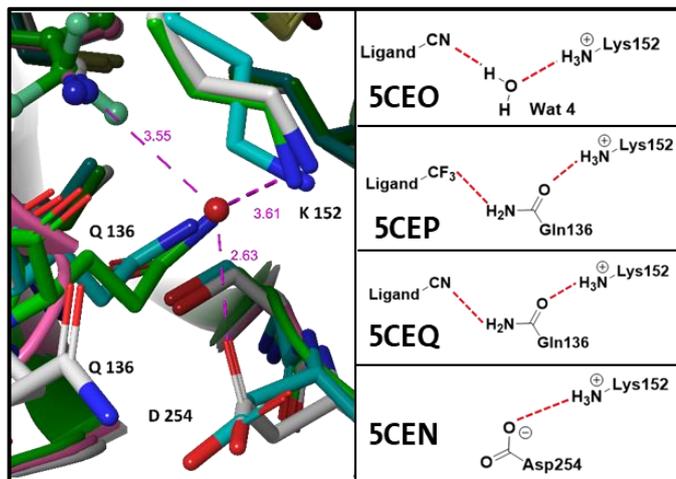


Abb. 5-6 – Wasserstoffbrückenbindung von Gln136 in **5CEP** (cyan) und **5CEQ** (grün) zum Liganden (hier noch falsche Zuordnung von N und O in Gln136, da noch nicht mit PROTPREP bearbeitet). In **5CEO** (grau) sitzt anstelle von Gln136 ein Wassermolekül (rot). Nach PROTPREP lassen sich die Interaktionen wie rechts schematisch dargestellt beobachten.

4) Sandwichstruktur durch die Aminosäuren ober- und unterhalb des Liganden:

Der tief in die Bindetasche hineinragende Pyridinring des Liganden bildet mit Ala150 (oben) und Leu243 (unten) in vdW-Abständen eine Sandwich-Struktur aus.

Die Abstände zu Ala150 oberhalb des Liganden (**5CEO**: 3,47 Å, **5CEP**: 3,52 Å, **5CEQ**: 3,46 Å) und Leu243 unterhalb des Liganden (**5CEO**: 3,46 Å, **5CEP**: 3,47 Å, **5CEQ**: 3,44 Å) sind im Vergleich der drei Liganden nahezu gleich (s. Tab. 5-3).

Insgesamt lässt sich festhalten, dass die Ausrichtung der Hinge-Region sowie des Gatekeepers Met190 in den verschiedenen DLK-Röntgenkristallstrukturen nahezu identisch ist. Ebenso wird die Bindetasche nach oben und unten in gleicher Weise von Ala150 und Leu243 begrenzt. Im Gegensatz dazu stehen der P-Loop (z.B. Gly132) und die tiefe Tasche nach hinten (z.B. Lys152). Hier lassen sich die größten Unterschiede zwischen den verschiedenen Röntgenkristallstrukturen feststellen. Eine Besonderheit stellt die unterschiedliche Ausrichtung des Gln136 dar, sodass es nur in **5CEP** und **5CEQ** zur Ausbildung einer Wasserstoffbrückenbindung zum Liganden kommt.

Konservierte Wassermoleküle

Um im weiteren Verlauf die Auswirkungen von unterschiedlichen Wasserpositionierungen auf die Dockingexperimente zu untersuchen, wurden die Wassermoleküle in den Bindetaschen der DLK-Röntgenkristallstrukturen analysiert. Alle DLK-Röntgenkristallstrukturen wurden im Raum gleich ausgerichtet (s. Kap. 7.1) und nach ähnlich positionierten Wassermolekülen untersucht. Wenn diese Wassermoleküle Bindungspartner von Liganden oder Protein sind, kann dies Aufschluss über konservierte Wassermoleküle in der Bindetasche geben. Unter der Bedingung, dass in den vier übereinanderliegenden Röntgenkristallstrukturen mindestens drei Wassermolekül-Sauerstoffatome (im Folgenden als „**Wat**“ bezeichnet) mit einem maximalen Abstand von 1,0 Å zueinander clustern und diese einen maximalen Abstand von 5 Å vom Liganden haben, wurden drei Cluster von **Wats** (s. Abb. 5-7) gefunden.

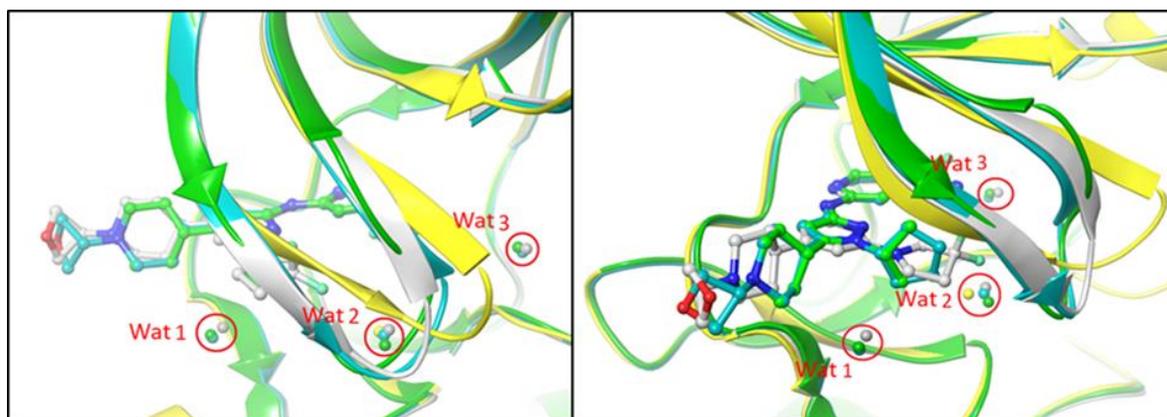


Abb. 5-7 – Cluster von konservierten Wassermolekülen in der Bindetasche übereinandergelagerter DLK-Röntgenkristallstrukturen aus verschiedenen Perspektiven. Max. Abstand von 1 Å zwischen den Wassermolekülen innerhalb eines Clusters. Max. Abstand von 5 Å zum Liganden. **5CEO** (grau), **5CEP** (cyan), **5CEQ** (grün), **5CEN** (gelb).

Die Abstandswerte wurden anhand der Rohdaten der Röntgenkristallstrukturen bestimmt. Jedes der drei **Wats** hat mehrere Interaktionspartner in der Bindetasche. In den folgenden Beschreibungen und in Tab. 5-4 werden nur Abstände von < 3,4 Å zu den entsprechenden **Wats** berücksichtigt. Wenn der Abstand zu einem Heteroatom einer Aminosäure kleiner als 3,4 Å war, wurde ein Kontakt zu Kohlenstoffen derselben Aminosäure nicht zusätzlich aufgeführt. Im Vergleich zu **Wat 1** und **Wat 3** interagiert **Wat 2** mit der höchsten Anzahl von Aminosäuren (s. Tab. 5-4). Es kann daher davon ausgegangen werden, dass **Wat 2** am stabilsten gebunden ist.

Tab. 5-4 – Interaktionspartner (< 3,4 Å) der clusternden Wassermoleküle. Zahl in Klammern gibt die Anzahl der Wasserstoffbrückenbindungen an.

	5CEN	5CEO	5CEP	5CEQ
Wat 1	-	Gln197 (2) Ligand	2 Wats (je 1) Gln197 (1) Ligand	2 Wats (je 1) Gln197 (1) Glu200
Wat 2	Wat (1) Asp254 (1) Pro240 (1)	Ala135 (1) Pro240 (1) Gln197 (1)	Pro240 (1) Gln197 (1) Gln136 (1) Asn241 (1)	Gln197 (1) Pro240 (1) Gln136 (1)
Wat 3	-	Wat 4 (1) Asp254 (1)	Gln136 (1) Asp254 (1) Wat (1)	Gln136(1) Asp254(1) Wat (1)

Wat 2 ist als einziges der clusternden Wassermoleküle auch im **5CEN** vorhanden. **5CEN** hat eine Sonderstellung als Apoenzym im Vergleich zu den anderen Röntgenkristallstrukturen. Durch die Abwesenheit eines Liganden ist die Bindetasche mit vielen beweglichen Wassermolekülen gefüllt. Die Tatsache, dass in **5CEN** nur **Wat 2** konserviert ist, kann auf die große Anzahl von Wasserstoffbrückenbindungen zu den Proteinresten und die daraus resultierende starke Bindung in der Tasche zurückzuführen sein. Die Abstände zu **Wat 2** sind in allen PDBs immer die kürzesten aller Wasserstoffbrückenbindungen, die von den drei beschriebenen **Wats** ausgehen. Andererseits kann die Anwesenheit von **Wat 1** und **Wat 3** in den Protein-Liganden-Komplexen und deren Abwesenheit im Apoenzym auch so interpretiert werden, dass sie besonders wichtig für die Bindung eines Liganden sind. Der Kontakt von **Wat 2** mit Asp254 besteht nur in **5CEN**, da Asp254 etwas näher zur Bindetasche ausgerichtet ist als in **5CEO**, **5CEP** und **5CEQ** (ebenfalls in Abb. 5-3, S. 123 zu sehen).

Wat 3 hat in **5CEO** Kontakt zu einem weiteren Wassermolekül-Sauerstoffatom **Wat 4** (2,51 Å), das durch seine Nähe zum Liganden eine wichtige Rolle spielen könnte. Nach PROTPREP (s. Kap. 7.3, S. 264, Tab. 7-1, **5CEO_ops3_minwat**), interagiert **Wat 4** mit einem Wasserstoff über eine Wasserstoffbrückenbindung (Abstand: 2,25 Å) mit dem Stickstoff der Cyanogruppe des GNE-3511. **Wat 4** stellt daher ein interessantes Verbindungsmolekül zwischen Ligand und Protein dar. Die Beobachtung dieses zusätzlichen Wassermoleküls wurde auch schon gemacht, als es um die unterschiedlichen Ausrichtungen des Gln136 in den verschiedenen Röntgenkristallstrukturen ging (s. S. 127, Abb. 5-6). Hier wurde bereits festgestellt, dass ein Wassermolekül (nämlich **Wat 4**) im **5CEO** an der Stelle des Gln136 (**5CEP** und **5CEQ**) sitzt. Es bildet darüber hinaus eine Wasserstoffbrückenbindung zu Asp254 sowie zu Lys152 aus. In **5CEN** sitzt anstelle des **Wat 4** die Carboxylgruppe des Asp254 und geht seinerseits die gleiche Wechselwirkung zum Lys152 und mit dem jeweils anderen Sauerstoffatom der Carboxylfunktion eine Wasserstoffbrückenbindung zum Gln136 ein.

Die vier hervorgehobenen Wassermoleküle **Wat 1-4** sollten für die Proteinvorbereitungen für das Docking berücksichtigt werden.

Zusammenfassung der Targetanalyse

Die verschiedenen Liganden von **5CEO**, **5CEP** und **5CEQ** weisen keine große strukturelle Vielfalt und kaum unterschiedliche Ausdehnungen in der Bindetasche auf, insbesondere die Abstände

zur Hinge-Region und zu den Aminosäuren oberhalb, unterhalb und hinter (Met190) dem Liganden sind nahezu gleich. Strukturell sind die drei Liganden insbesondere im „Kern“ sowie im Substituenten R1 (s. Abb. 5-8) nahezu deckungsgleich.

Alle drei Inhibitoren zeigen die gleichen, kinasetypischen Interaktionsmuster in Form von Wasserstoffbrückenbindungen in der Abfolge *Donor-Akzeptor-Donor* zur Hinge-Region (s. Abb. 5-8, rot gestrichelte Linien). Darüber hinaus wird der tief hineinragende Aromat durch hydrophobe Wechselwirkungen zu den Seitenketten der Aminosäuren Ala150 und Leu243 (s. Abb. 5-8, grün gestrichelte Linien) als Sandwichkomplex zusätzlich in seiner Position fixiert.

Der Substituent R2, welcher hydrophobe Wechselwirkungen zum P-Loop eingeht, ist der Ligandenteil mit der größten Variabilität. Die Bindetasche kann sich mithilfe des P-Loops insbesondere an verschiedene Liganden anpassen. Der in die tiefe Tasche zum Lys152 hineinragende Substituent R1 agiert als Wasserstoffbrückenakzeptor und interagiert mit dem Wassermolekül **Wat 4** (**5CEO**) oder mit Gln136 (**5CEP**, **5CEQ**), welches wiederum mit Lys152 eine Wasserstoffbrückenbindung eingeht (s. Abb. 5-8 B). Durch die stark unterschiedliche Position des Gln136 zeigt hier wieder der P-Loop eine starke Flexibilität. In **5CEN** sitzt in dieser Position anstelle des **Wat 4** bzw. Gln136 die Carboxylgruppe des Asp254. Der dem Lösungsmittel zugewandte Ligandenteil R3 zeigt durch den Mangel an Protein-Interaktionspartnern eine gewisse Flexibilität.

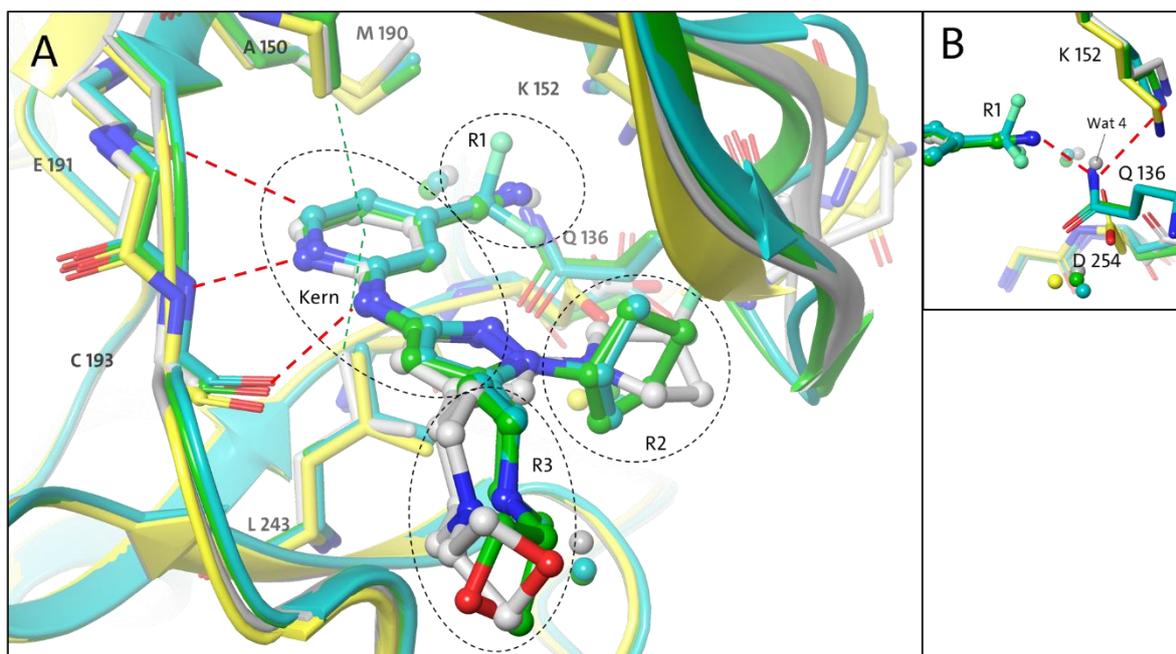


Abb. 5-8 – Übereinandergelagerte Darstellung von **5CEN**, **5CEO**, **5CEP** und **5CEQ**. Der mit der Hinge-Region über WBB (rot gestrichelt) interagierende Kern der Liganden nimmt eine nahezu identische Konformation ein. Der Aromat im hinteren Teil wird zusätzlich durch die hydrophoben Interaktionen mit den Seitenketten des Ala150 und L243 in Position gehalten (grün gestrichelte Linie). Der Substituent R1 interagiert über ein Wasserstoffbrückenbindungsnetzwerk über **Wat 4** bzw. Gln136 mit Lys152 (in **B** vergrößert dargestellt). Diversität zwischen den Liganden herrscht vor allem im Ligandenteil, der mit dem P-Loop interagiert (R2) und zum Lösungsmittel gerichtet ist (R3).

Die in diesem Kapitel analysierte Größe und räumliche Ausdehnung der verschiedenen Bindetaschen aus den unterschiedlichen PDBs kann im weiteren Verlauf eine wichtige Erklärung dafür liefern, ob Liganden in alle Bindetaschen gedockt werden können. Wie in diesem Kapitel dargestellt war es möglich, Wassermoleküle in allen Röntgenkristallstrukturen zu identifizieren, welche die gleiche räumliche Position aufweisen und die Teil eines Netzwerks von Wasserstoffbrückenbindungen zu anderen Wassermolekülen und Aminosäuren der DLK sind. Sie können als ein wichtiger Teil der Bindetasche interpretiert werden und spielen eine wichtige Rolle in der Umgebung der gebundenen Liganden.

5.1.2 Loop Modelling des Activation Loops mit Prime

Der in **5CEO**, **5CEP** und **5CEQ** fehlende Activation Loop wurde für die Durchführung der MD-Simulationen mithilfe des in der SCHRÖDINGER MAESTRO SUITE enthaltenen Loop-Modelling-Tools im Rahmen des PROTPREPS ergänzt (s. Kap. 7.3, **PL**-Strukturen). Pro Inputstruktur wurden fünf Loopkonformationen nach dem *Loop Refinement* generiert und nach ihrer *Prime Energy* (s. Tab. 5-5) sortiert. Jeweils eine der fünf Konformationen wurde für die Durchführung einer MD-Simulation (s. Kap. 5.1.3) ausgewählt. Die Auswahl erfolgte primär nach dem visuellen Vergleich mit dem vollständig aufgelösten Activation Loop einer PROTPREP-Version der **5CEN**-Kristallstruktur³⁹, da sich die Prime-Energy-Werte nur geringfügig voneinander unterschieden (s. Tab. 5-5).

Tab. 5-5 – *Prime Energies* der generierten Activation Loops für **5CEO**, **5CEP** und **5CEQ**

Loop Conformation #	Prime Energy 5CEO	Prime Energy 5CEP	Prime Energy 5CEQ
1	-11396,15	-11398,41	-11437,63
2	-11395,55	-11387,59	-11435,00
3	-11395,31	-11387,39	-11431,11
4	-11390,61	-11384,58	-11430,43
5	-11388,41	-11380,95	-11427,93

Keiner der generierten Loops bildete eine Helixstruktur der Länge wie in **5CEN**.

Für die Durchführung der MD-Simulationen wurde Loop Nr. 1 (*Prime Energy* -11396,15) ausgewählt, da dieser Loop (s. Abb. 5-9, roter Loop) dem **5CEN**-Loop (s. Abb. 5-9, gelber Loop) in der Orientierung am ähnlichsten ist. Von den **5CEP**-Activation-Loops wurde Loop Nr. 2 (s. Abb. 5-10, rot) ausgewählt. Von den **5CEQ**-Activation-Loops wurde Loop Nr. 3 ausgewählt (s. Abb. 5-11, rot).

Im Folgenden werden die hier ausgewählten **5CEO**-, **5CEP**- und **5CEQ**-Strukturen mit vollständigem Activation Loop als **PL**-Strukturen⁴⁰ (z.B. **5CEO-PL**-Struktur) bezeichnet.

³⁹ Dabei handelt es sich um die **5CEN_opl3_ow** Struktur (s. Kap. 7.3).

⁴⁰ PL = Prime Loop

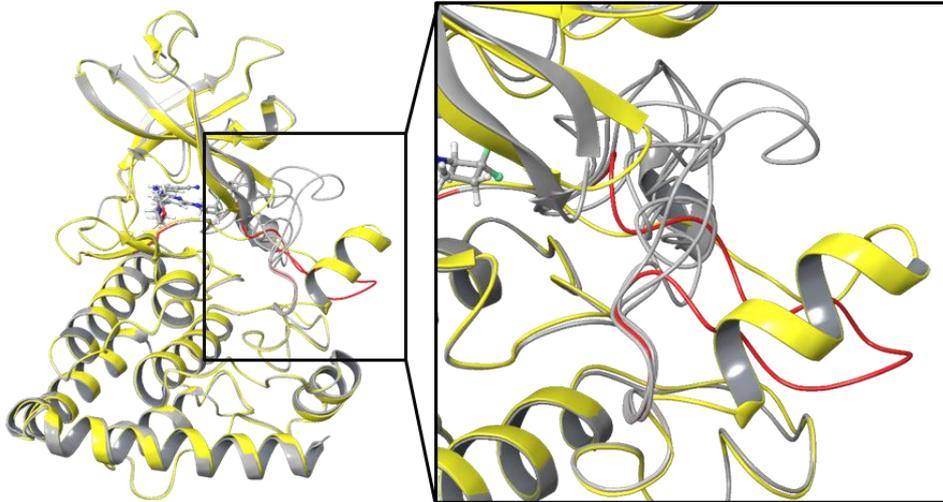


Abb. 5-9 – Darstellung der fünf generierten Loopkonformationen des **5CEO** (grau und rot) mit **5CEN_opls3_ow** (gelb) im Vergleich. Ausgewählt wurde der rot dargestellte Loop (#1) für die Durchführung der MD-Simulationen.

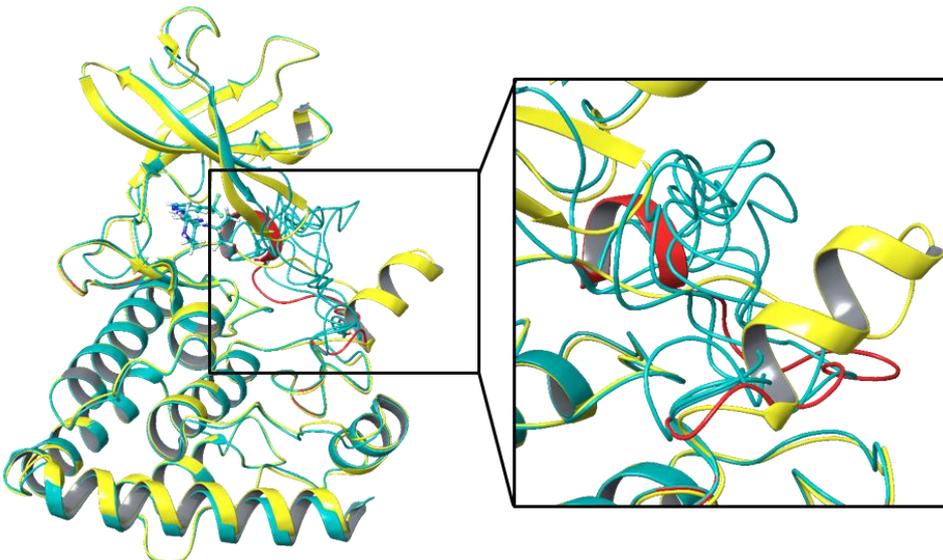


Abb. 5-10 – Darstellung der fünf generierten Loopkonformationen des **5CEP** (cyan und rot) mit **5CEN_opls3_ow** (gelb) im Vergleich. Ausgewählt wurde der rot dargestellte Loop (#2) für die Durchführung der MD-Simulation.

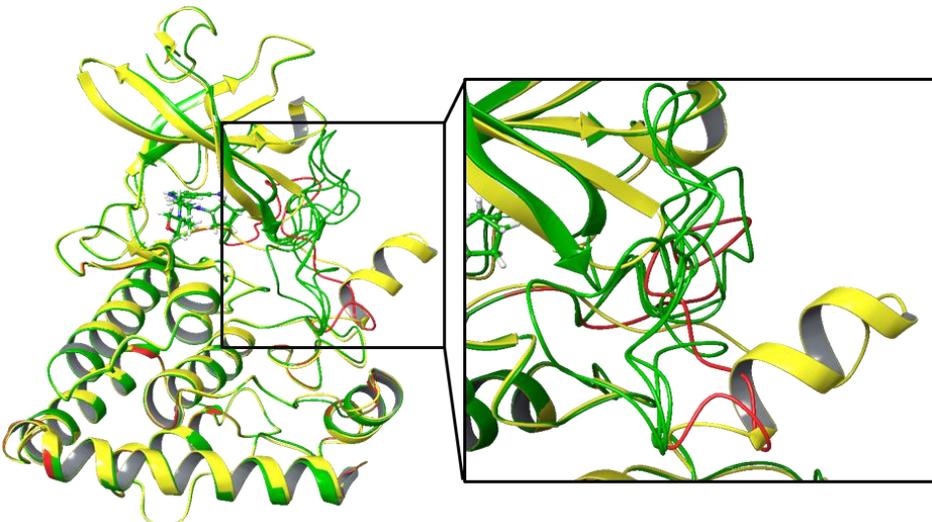


Abb. 5-11 – Darstellung der fünf generierten Loopkonformationen des **5CEQ** (grün und rot) mit **5CEN_opls3_ow** (gelb) im Vergleich. Ausgewählt wurde der rot dargestellte Loop (#3) für die Durchführung der MD-Simulation.

5.1.3 MD-Simulationen und Auswahl von Konformationen

Flexibilität und Anpassungsfähigkeit der Bindetasche sind wichtig für die Bildung eines Protein-Ligand-Komplexes (s. Kap. 4.4, S. 59). Für das virtuelle Screening ist ein variabler konformativer Raum der Bindetasche notwendig, um eine Vielzahl unterschiedlicher Chemotypen docken zu können. [516]

Die Bindetaschen der DLK-Strukturen **5CEO**, **5CEP** und **5CEQ** sind an die gebundenen Liganden angepasst und könnten in dieser Position zu starr oder zu klein für andere Inhibitoren sein. In der Zielstrukturanalyse (s. S. 124) zeigte sich für das Apoenzym **5CEN** bereits, dass es sich hierbei um eine relativ kleine Bindetasche handelt.

Für die MD-Simulation wurden die vorbereiteten Röntgenkristallstrukturen **5CEO**, **5CEP** und **5CEQ** mit vervollständigtem Activation Loop (**PL**-Strukturen, s. Kap. 7.3) sowie die vorbereitete **5CEN**-Struktur⁴¹ verwendet. Es wurden MD-Simulationen mit drei Durchläufen pro Struktur und einer Dauer von 30 ns (3001 Frames⁴²) durchgeführt (s. Kap. 7.4). Die generierten MD-Trajektorien wurden im ersten Schritt hinsichtlich der Stabilität der RMSD-Verläufe analysiert und die Flexibilität einzelner Proteinbereiche mithilfe von RMSF-Daten herausgearbeitet. Im zweiten Schritt wurden Frames aus den MD-Trajektorien ausgewählt, um sie für die nachfolgenden Dockinguntersuchungen (Alignment für QSAR, Vorbereitungen für das virtuelle Screening) zu verwenden.

Analyse der strukturellen Stabilität und Flexibilität von Protein und Protein-Liganden-Komplex

Die Trajektorien wurden visuell und mithilfe von RMSD-Daten hinsichtlich ihrer Stabilität überprüft. Die Flexibilität einzelner Proteinbereiche wurde mithilfe berechneter RMSF-Werte der Backbone-Aminosäuren untersucht. Insbesondere der Vergleich der MD-Trajektorien des **5CEN** (Apoenzym ohne gebundenen Liganden) zu den MD-Trajektorien der übrigen drei ligandhaltigen Röntgenkristallstrukturen **5CEO**, **5CEQ** und **5CEP** kann Aufschluss über den Einfluss von Liganden auf die Dynamik des Proteins bzw. der Bindetasche geben. Ein gebundener Ligand kann zu Konformationsänderungen des Proteins führen, aber auch zur Stabilisierung des Proteins beitragen. Er kann seinen Bindungsmodus über die Simulationszeit verändern oder sich aus der Bindetasche herausbewegen.

⁴¹ **5CEN_opls3_ow** (s. Kap. 7.3)

⁴² Nummerierung der Frames von 0 bis 3000

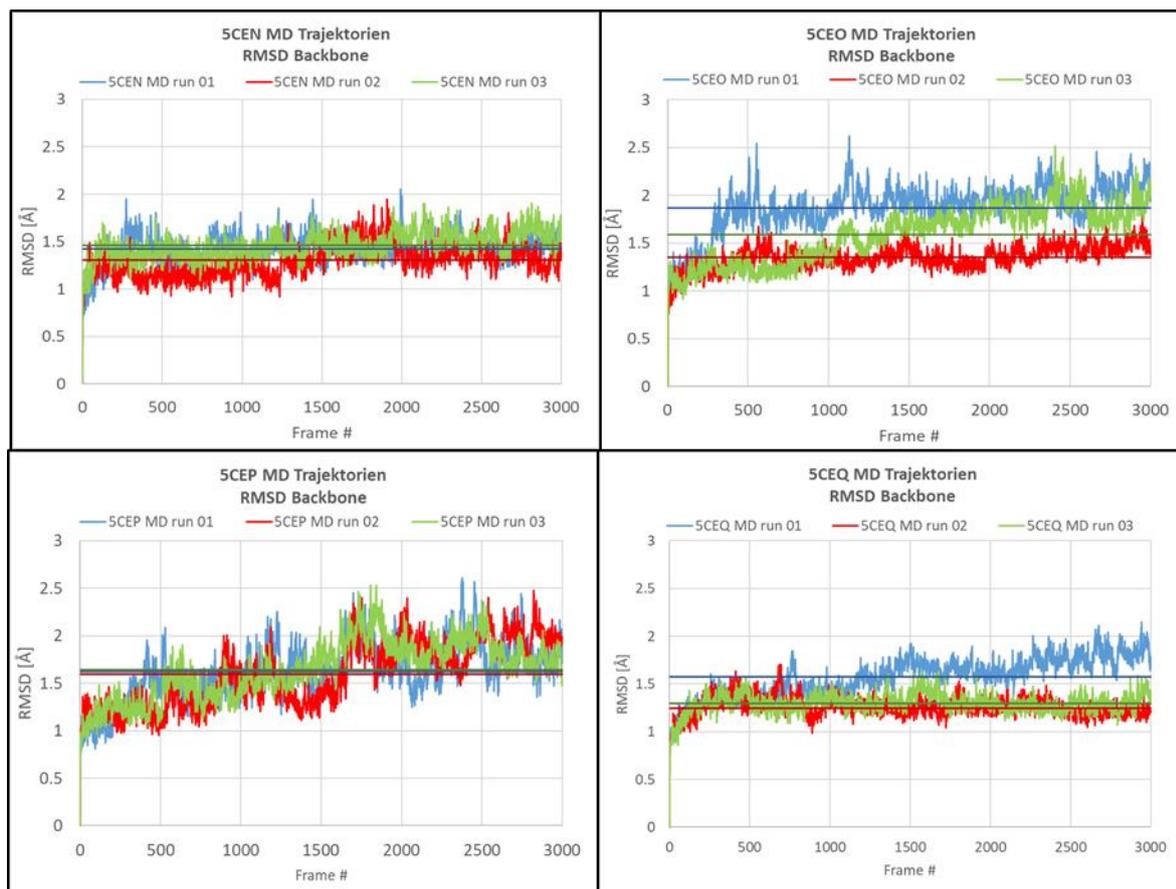


Abb. 5-12 – RMSD-Plots des Protein-Backbones basierend auf den vier MD-Ausgangsstrukturen für **5CEN**, **5CEO**, **5CEP** und **5CEQ**. 1000 Frames entsprechen 10 ns Simulationszeit. Die mittleren RMSD-Werte werden als Linien gleicher Farbe dargestellt.

Die RMSD-Werte der generierten Proteinkonformationen (s. Abb. 5-12) im Vergleich zur jeweiligen Ausgangsstruktur stiegen in allen MD-Trajektorien zu Beginn aufgrund von thermischen Fluktuationen an, gefolgt von einem Plateau bzw. von Fluktuationen um einen Mittelwert herum. Alle Trajektorien zeigen, dass die Proteinkonformationen als stabil bewertet werden können, da die RMSD-Werte unter 3 Å liegen und zum Ende der Simulation kein starker Trend zu höheren RMSD-Werten zu beobachten ist. Der mittlere RMSD-Wert über die jeweilige gesamte Trajektorie bewegt sich zwischen 1,297 Å (**5CEQ Run 03**) und 1,868 Å (**5CEO Run 01**).

Tab. 5-6 – RMSD-Mittelwerte des Protein-Backbones (PB) und des Liganden (L, falls vorhanden) der MD-Trajektorien sowie Variationskoeffizienten (VK) der RMSD-Werte als Maß für die Schwankungen während der Simulationszeit. Gesamt = über alle Frames gemittelt. Kursiv dargestellt: hohe Fluktuationen.

	5CEN 01	5CEN 02	5CEN 03	5CEN gesamt
\bar{x} PB [Å]	1,423	1,306	1,462	1,397
VK PB [%]	11,57	12,93	9,82	12,36
	5CEO 01	5CEO 02	5CEO 03	5CEO gesamt
\bar{x} PB [Å]	1,868	1,352	1,590	1,603
VK PB [%]	13,69	9,33	<i>18,73</i>	<i>19,84</i>
\bar{x} L [Å]	0,468	0,693	0,536	0,565
VK L [%]	35,86	29,19	28,52	35,22
	5CEP 01	5CEP 02	5CEP 03	5CEP gesamt
\bar{x} PB [Å]	1,624	1,595	1,643	1,621
VK PB [%]	<i>18,72</i>	<i>21,43</i>	<i>18,66</i>	<i>19,65</i>
\bar{x} L [Å]	0,668	0,644	0,637	0,650
VK L [%]	31,83	31,47	30,50	31,36
	5CEQ 01	5CEQ 02	5CEQ 03	5CEQ gesamt
\bar{x} PB [Å]	1,576	1,261	1,297	1,378
VK PB [%]	13,18	8,23	8,04	14,77
\bar{x} L [Å]	0,621	0,678	0,766	0,688
VK L [%]	30,96	32,45	24,01	30,21

Die **5CEN**-MD-Trajektorien (Apoenzym ohne Liganden) lassen hierbei hinsichtlich der geringen Fluktuationen über alle drei Trajektorien hinweg und ihrer geringen mittleren RMSD-Werte von unter 1,5 Å auf eine besonders hohe Stabilität der Proteinstruktur schließen.

Die **5CEQ**-MD-Trajektorien kommen den **5CEN**-MD-Trajektorien hinsichtlich ihrer Stabilität am nächsten: zwei (*Run 02* und *Run 03*) von den drei **5CEQ**-MD-Trajektorien liegen im RMSD-Mittelwert und im Variationskoeffizienten sogar noch unter den entsprechenden Werten der **5CEN**-Trajektorien (s. Abb. 5-12 u. Tab. 5-6). Nur der RMSD-Verlauf der ersten **5CEQ**-MD-Trajektorie *Run 01* weicht etwas stärker von den RMSD-Verläufen der beiden anderen **5CEQ**-MD-Trajektorien ab.

Eine der drei **5CEO**-MD-Trajektorien (*Run 02*) hat einen ähnlich stabilen RMSD-Verlauf mit geringem RMSD-Mittelwert (1,352 Å) und geringen Fluktuationen (VK = 9,33 %) wie die drei **5CEN**-MD-Trajektorien und zwei der drei **5CEQ**-MD-Trajektorien. In der ersten **5CEO**-MD-Trajektorie wich die Struktur schon sehr früh von der Ausgangsstruktur ab und erreichte mit Frame 1128 mit 2,622 Å den höchsten RMSD-Wert, was unter allen 12 MD-Simulationen die größte Abweichung vom Ursprung darstellte. Dennoch konnte nach dem raschen RMSD-Anstieg eine stabile Trajektorie bis zum Ende der Simulation auf einem relativ hohen RMSD-Level (RMSD-Mittelwert = 1,868 Å) mit vergleichsweise geringen Fluktuationen (13,69 %) beobachtet werden. Die dritte **5CEO**-MD-Trajektorie hingegen hat einen ähnlichen Verlauf zu den **5CEP**-MD-Trajektorien: Bei einem mittleren RMSD-Wert von 1,590 Å (**5CEP** im Vergleich ebenso um 1,6 Å) sind die anfänglichen RMSD-Werte bis etwa zur Hälfte der Simulationszeit geringer als in der zweiten Hälfte, d.h. das Plateau wird erst später erreicht und die Fluktuationen sind höher als in den übrigen **5CEO**-Runs (VK = 18,73 %).

Die **5CEP**-MD-Trajektorien weisen alle gleichermaßen einen vergleichsweise hohen mittleren RMSD-Wert von etwa 1,6 Å auf. Die RMSD-Mittelwerte sind – im Gegensatz zu den MD-Simulationen der anderen PDBs – alle nahezu gleich. Bis zu ~18 ns nahm der RMSD zu. Von da an kann von einer gewissen Stabilität ausgegangen werden, da es keine Auf- und Abwärtstendenz der RMSD-Werte mehr gibt. Alle **5CEP**-MD-Trajektorien weisen vergleichsweise die höchsten Fluktuationen auf (VK = 18,66-21,43 %). Auffällig ist, dass diese starke Mobilität des Protein-Backbones in allen **5CEP**-Simulationen gleichermaßen beobachtet werden konnte, während sich in den anderen MD-Simulationen von **5CEN**, **5CEO** und **5CEQ** höchstens eine Trajektorie als andersartig bzw. stärker fluktuierend erwies.

In den **5CEO**-, **5CEP**- und **5CEQ**-MD-Simulationen konnte beobachtet werden, dass die Protein-Ligand-Bindung über die gesamte Simulationszeit bestehen blieb und der Ligand immer in der

ursprünglichen Pose in der Bindetasche verblieb (vgl. Kap. 5.1.1, S. 130, Abb. 5-8). Die RMSD-Plots der **5CEQ**-Trajektorien (insb. *Run 02* und *Run 03*) zeigen im Vergleich zu den anderen Protein-Ligand-Komplexen **5CEP** und **5CEO**, dass der gebundene **5CEQ**-Ligand zu den geringsten Strukturabweichungen im Protein-Backbone führt. Gleiches trifft für den **5CEO**-Liganden in der zweiten **5CEO**-MD-Trajektorie zu. Die Liganden selbst haben in diesen Trajektorien einen vergleichsweise hohen RMSD-Wert (**5CEQ**-Liganden im Mittel 0,688 Å, **5CEO**-Ligand in *Run 02*: 0,693 Å), was zunächst in einem Gegensatz zu der relativ geringen Protein-Backbone-Beweglichkeit dieser Runs steht. Es könnte sein, dass der Ligand einen stabilisierenden Effekt auf das Protein hat, indem der Ligand Wechselwirkungen zum Protein eingeht und durch diese selbst stärker bewegt wird. In allen anderen MD-Trajektorien (**5CEQ**: *Run 01*, **5CEO**: *Run 01* und *Run 02*, **5CEP**: alle drei Runs) ist eine im Vergleich zum **5CEN**-Apoenzym erhöhte Beweglichkeit der Proteinstruktur zu beobachten. Entsprechend der Bindungsaffinität der Liganden hätte man in der Reihenfolge einen stabilisierenden Effekt auf die Bindetasche erwartet: **5CEO** > **5CEQ** > **5CEP**. In den Ergebnissen der hier durchgeführten MD-Simulationen weist **5CEP** die stärksten Fluktuationen auf, während die **5CEQ**-Trajektorien vergleichsweise am stabilsten sind. **5CEO** nimmt eine Zwischenposition ein.

Analyse der RMSF-Plots

Mithilfe von RMSF-Plots (s. Abb. 5-13) kann man feststellen, welche Proteinbereiche für die Fluktuationen in den RMSD-Werten verantwortlich sind. Je höher der RMSF-Wert, desto stärker hat sich der Backbone einer definierten Aminosäure während der Simulation bewegt. Die variablen

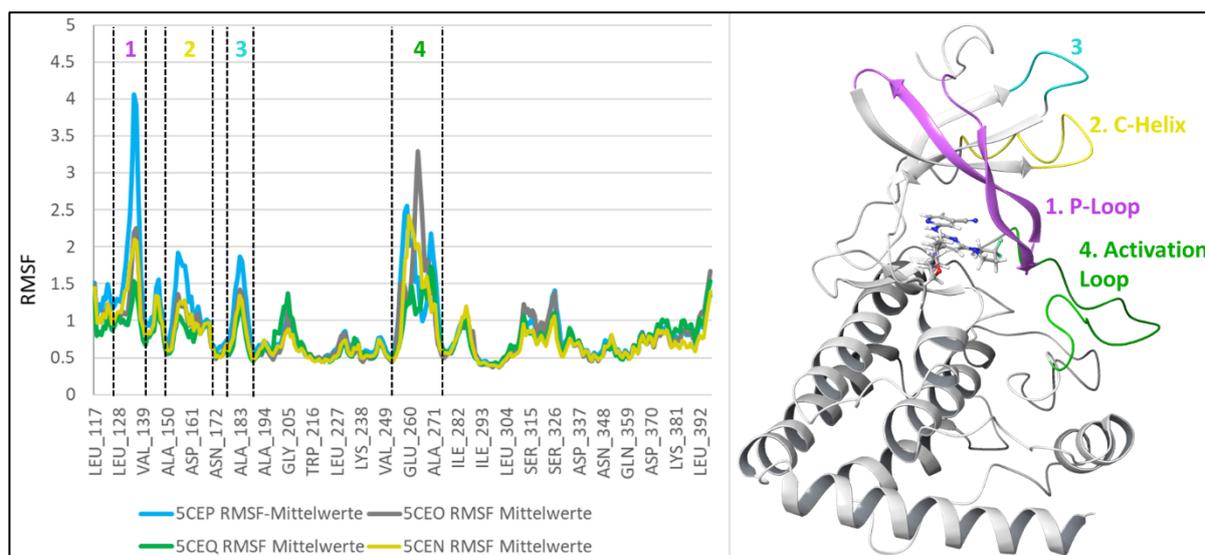


Abb. 5-13 – RMSF-Mittelwerte der jeweils drei MD-Simulationen einer PDB. Die flexiblen Bereiche werden rechts an der **5CEO-PL**-Struktur dargestellt.

Regionen des Proteins werden als Peaks angezeigt (vgl. Kap. 4.6): In allen MD-Strukturen gehören der P-Loop (1), die C-Helix (2), ein kurzer Bereich im N-Lobe (3) sowie der Activation Loop (4) zu den beweglichsten Regionen. Die Mobilität des P-Loops, der C-Helix und des Bereichs (3) war in den **5CEP**-Simulationen am stärksten ausgeprägt, während die Beweglichkeit des Activation Loops vergleichbar mit den anderen PDB-Trajektorien ist. Daraus lässt sich schließen, dass die starken Abweichungen und Fluktuationen in den zuvor beschriebenen RMSD-Plots in den **5CEP**-Trajektorien hauptsächlich durch Positionsänderungen des P-Loops und damit auch durch Änderungen in der Bindetasche verursacht werden. Die Flexibilität des P-Loops ist in den **5CEO**- und **5CEN**-MD-Trajektorien ähnlich hoch, während sich der Activation Loop insbesondere in den **5CEO**-MD-Trajektorien im Mittel am stärksten bewegt. Die Flexibilität der vier hier beschriebenen Bereiche ist in den **5CEQ**-MD-Trajektorien im Mittel am geringsten.

Bezüglich der Bindetaschen-Flexibilität, die maßgeblich vom P-Loop beeinflusst wird, lässt sich eine Abstufung von **5CEP** > **5CEO** > (=) **5CEN** > **5CEQ** vornehmen. Diese Feststellung wird in Abb. 5-14 bestätigt, in der die fünf Aminosäuren mit den höchsten und niedrigsten RMSF-Werten für jede Röntgenkristallstruktur dargestellt werden.

Auch hier befinden sich die Regionen mit den höchsten mittleren RMSF-Werten im Activation Loop und im P-Loop. Hier zeigt sich deutlich der stark bewegliche P-Loop in der **5CEP**-Simulation, während **5CEQ** im Vergleich dazu kaum stark bewegliche Bereiche aufweist.

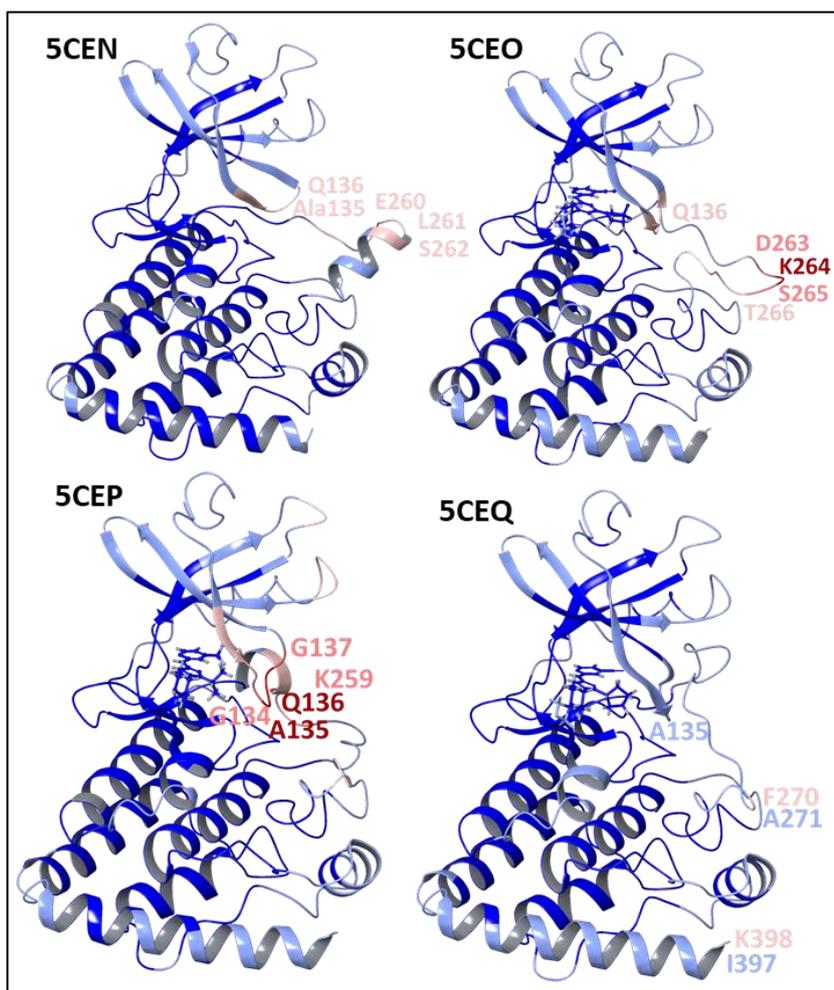


Abb. 5-14 – Darstellung der vier mit PROTREP vorbereiteten Strukturen **5CEN**, **5CEO**, **5CEP** und **5CEQ** mit nach RMSF-Mittelwerten eingefärbten Proteinbereichen von dunkelblau (geringe RMSF-Werte) über weiß (mittlere RMSF-Werte von etwa 2 Å) bis rot (hohe RMSF-Werte). Die fünf Aminosäuren mit den höchsten bzw. niedrigsten RMSF-Werten im Mittel über die drei Trajektorien sind in der Darstellung gekennzeichnet.

Betrachtet man die mittleren RMSF-Werte ausgewählter Aminosäuren der Bindetasche in allen MD-Simulationen, so zeigt sich, dass **5CEP** die höchsten durchschnittlichen RMSF-Werte der Bindetaschenamino-säuren und damit die höchste Flexibilität im Vergleich zu den anderen eingesetzten PDB-Strukturen aufweist (s. Tab. 5-7).

Tab. 5-7 – RMSF-Mittelwerte des Backbones ausgewählter Aminosäuren der Bindetasche, die aus den MD-Simulationen erhalten wurden. Damit kann die Flexibilität der Bindetasche in Abhängigkeit von der Trajektorie beurteilt werden. SW = Sandwich.

MD Trajektorie	E191 <i>Hinge</i>	C193 <i>Hinge</i>	G132 <i>P-Loop</i>	M190 <i>Rückwand</i>	K152 <i>Rückwand</i>	A150 <i>SW oben</i>	L243 <i>SW unten</i>
5CEN	0,542 Å	0,633 Å	1,429 Å	0,485 Å	0,615 Å	0,593 Å	0,522 Å
5CEO	0,535 Å	0,541 Å	1,100 Å	0,491 Å	0,690 Å	0,543 Å	0,485 Å
5CEP	0,639 Å	0,635 Å	2,072 Å	0,564 Å	0,931 Å	0,652 Å	0,567 Å
5CEQ	0,526 Å	0,562 Å	0,939 Å	0,493 Å	0,610 Å	0,581 Å	0,518 Å

In allen MD-Simulationen waren die RMSF-Werte der Hinge-Region, des Gatekeepers Met190 und von Leu243 vergleichsweise gering, während Gly132 als Teil des P-Loops den bei weitem höchsten mittleren RMSF-Wert hatte. Dieses Resultat passt auch zu den Ergebnissen der Zielstrukturanalyse (s. Kap. 5.1.1, S. 126), in der festgestellt wurde, dass die Konformationen der Hinge-Region, der Rückseite sowie der Begrenzung nach oben und unten zwischen den verschiedenen PDBs nahezu identisch sind, während das Lys152 und der P-Loop die größte Variabilität in ihrer Position aufweisen.

Im Durchschnitt war die Mobilität der Hinge-Region (hier dargestellt durch die mittleren RMSF-Werte von Glu191 und Cys193) bei **5CEO** und **5CEQ** im Vergleich zu den beiden anderen MD-Ausgangsstrukturen am geringsten. Dies traf auch auf Gly132 des P-Loops und die Aminosäuren oberhalb und unterhalb des Liganden (Ala150 und Leu243) zu.

Fasst man alle Ergebnisse zusammen, so zeigte **5CEQ** die geringste Flexibilität der Bindetasche, während **5CEP** die höchste Mobilität in allen ausgewählten Bindetaschen-Aminosäuren ohne Ausnahme aufwies. **5CEN** und **5CEO** nehmen eine Zwischenstellung ein. Die RMSD- und RMSF-Plot zeigen aber eher, dass **5CEO** insgesamt flexibler ist als **5CEN**. So lässt sich die Flexibilität der MD-Ausgangsstrukturen in dieser Reihenfolge festhalten: **5CEQ < 5CEN < 5CEO < 5CEP**.

Auswahl von MD-Frames

Für die Erstellung von QSAR-Modellen sollten die bekannten DLK-Inhibitoren in der richtigen Position in die ATP-Bindetasche von DLK-Proteinkonformationen gedockt werden, um so ein Alignment der verschiedenen Inhibitoren zu erhalten. Mit einer hohen Anzahl verschiedener Proteinkonformationen ist die Wahrscheinlichkeit höher, möglichst viele korrekte Posen für das

Alignment zu erhalten. Die Anzahl sollte jedoch begrenzt sein, da eine hohe Zahl an verschiedenen Proteinkonformationen gleichzeitig auch mehr Aufwand und Rechenarbeit bei den folgenden Dockingexperimenten bedeutet.

Als Konsequenz aus den ersten Testdockings (s. Kap. 5.4.1.1), in denen **5CEO** und **5CEQ** gute Dockingergebnisse erzielten, wurde die **5CEO**- und die **5CEQ-PL**-Struktur ausgewählt, um möglichst viele von den bekannten DLK-Inhibitoren durch Docking in die Bindetasche in der korrekten Pose zur Deckung zu bringen. Zusätzlich wurden weitere Proteinkonformationen aus den MD-Trajektorien – zunächst ohne Clustering der MD-Frames – ausgewählt. Die Auswahl sollte die konformative Vielfalt des Proteins abbilden, um möglichst viele der bekannten DLK-Inhibitoren docken zu können.

Da die Apoproteinstruktur **5CEN** selbst keine ausreichend guten Dockingergebnisse in den Testdockings erzielte (s. Kap. 5.4.1.1) und dies an der zu kleinen Bindetasche liegen könnte, wurde ein **5CEN**-Frame mit einer vergrößerten Bindetasche ausgewählt.

Die zweite **5CEN**-MD-Trajektorie (s. Abb. 5-15) zeigte einen Anstieg des Bindetaschenvolumens ab etwa 15 ns. Frame 2661 der zweiten **5CEN**-Trajektorie zeigte eine besonders große Bindetasche mit einem großen Abstand zwischen den C_α-Atomen von Phe192 und Gly132 (15,21 Å gegenüber 11,90 Å in der Ausgangsstruktur) und einem erhöhten Bindetaschenvolumen (949 Å³ gegenüber 663 Å³ in der Ausgangsstruktur).

Aus der zweiten **5CEO**-MD-Trajektorie (vgl. Abb.

5-12, S. 134), wurde Frame 2000 mit einem leicht erhöhten Bindetaschenvolumen und leicht vergrößertem Abstand zwischen den C_α-Atomen von Phe192 und Gly132 im Vergleich zur **5CEO-PL**-Struktur ausgewählt (s. Tab. 5-8).

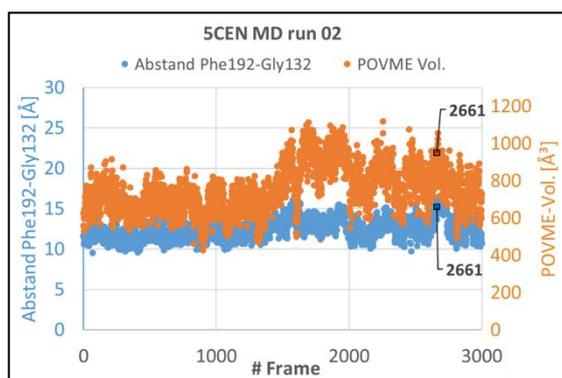


Abb. 5-15 – Bindetaschenvolumen (orange) und Abstand zwischen zwei Bindetaschenamino­säuren im Verlauf der MD-Simulation (**5CEN Run 02**). Auswahl des Frames 2661 wegen des erhöhten Volumens. 1000 Frames = 10 ns.

Tab. 5-8 – Eigenschaften der ausgewählten DLK-Proteinkonformationen zur Erstellung des Docking-Alignments für QSAR. PB = Protein-Backbone; Vol.: Bindetaschenvolumen (POVME), Dist. = Abstand, L = Ligand.

	RMSD PB	Vol.	Dist. F192 - G132	Dist. M190 - A135	RMSD L
5CEN Run 02 Frame 2661	1,515 Å	949 Å ³	15,21 Å	15,78 Å	Kein Ligand
5CEO Run 02 Frame 2000	1,387 Å	889 Å ³	13,27 Å	15,66 Å	0,779 Å
5CEQ Run 03 Frame 3000	1,433 Å	676 Å ³	12,96 Å	15,99 Å	0,727 Å
5CEO-PL-Struktur	0 Å	808 Å ³	13,20 Å	15,45 Å	0 Å
5CEQ-PL-Struktur	0 Å	746 Å ³	12,79 Å	15,46 Å	0 Å

Die MD-Simulationen von **5CEQ** lieferten im Vergleich zu den anderen MD-Simulationen die stabilsten Ergebnisse. Insbesondere der dritte Lauf zeigte eine hohe Stabilität bezüglich der niedrigen RMSD-Werte gegenüber der Ausgangsstruktur (vgl. Abb. 5-12, S. 134). Hier wurde der letzte Frame (Frame 3000) als Repräsentant der stabilen **5CEQ**-MD-Trajektorie für das Docking für die QSAR-Vorbereitungen verwendet.

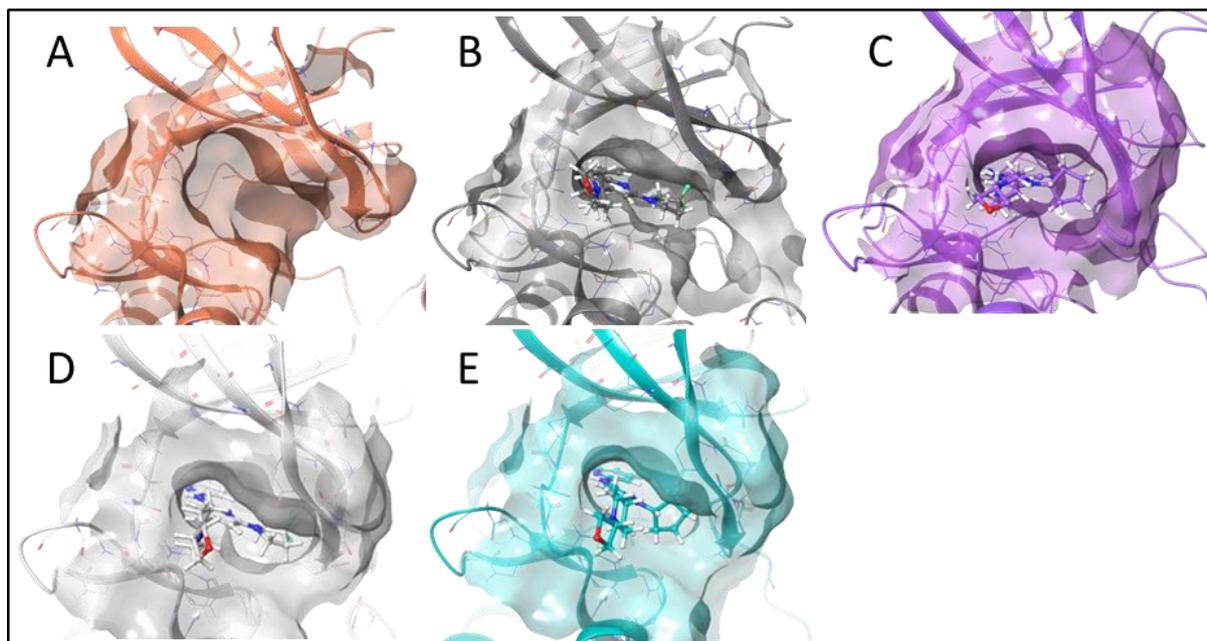


Abb. 5-16 – Darstellung der Bindetaschen der fünf ausgewählten DLK-Proteinkonformationen für das Docking-Alignment für QSAR. **A** – **5CEN** Run 02 Frame 2661, **B** – **5CEO** Run 02 Frame 2000, **C** – **5CEQ** Run 03 Frame 3000, **D** – **5CEO-PL**-Struktur, **E** – **5CEQ-PL**-Struktur.

Alle Zielstrukturen (s. Abb. 5-16) wurden vorbereitend für das Docking (s. Kap. 5.4.1.2 f.) im OPLS3-Kraftfeld minimiert (s. Kap. 7.3, S. 264, Tab. 7-1, „*MD-Frames für Docking (QSAR-Alignment)*“).

Um aus den 36.000 Frames (4 PDBs à 3 Simulationen mit je 3000 Frames) Kandidaten für das virtuelle Screening auszuwählen, die die strukturelle Variabilität der DLK noch besser repräsentieren würden, wurde zunächst ein Backbone-RMSD-basiertes Clustering der jeweils 3000 Proteinkonformationen einer MD-Trajektorie durchgeführt (s. Kap. 7.4.2) und anschließend geeignete Clusterrepräsentanten nach folgenden Kriterien ausgewählt:

- a) Clustergröße bzw. Stabilität: Auswahl von Proteinkonformationen, die repräsentativ für eine Reihe ähnlicher Proteinkonformationen einer MD-Trajektorie sind. Vor allem Clusterrepräsentanten von großen, (nach RMSD) stabil verlaufenden Clustern waren für die Auswahl interessant, da man von ihnen eine große Auftrittswahrscheinlichkeit erwartet.
- b) Diversität (Bindetaschengröße und RMSD-Werte als Beurteilungskriterien): Auswahl von Proteinkonformationen, die möglichst unterschiedliche Bindetaschen haben, sodass eine hohe

konformative Diversität sichergestellt wird. Die hochbeweglichen Aminosäuren im P-Loop erlauben eine starke Veränderung der Bindetasche im Verlauf der MD-Simulation und damit eine Auswahl verschiedener Bindetaschengrößen in Abhängigkeit vom Grad der Auslenkung des P-Loops. Das POVME-Volumen und der Abstand zwischen den C_{α} -Atomen der Bindetaschenamino-säuren Met190-Ala135 und Phe192-Gly132 diene als Hinweis für das Maß der Auslenkung des beweglichen P-Loops. Darüber hinaus wurden RMSD-Werte für die Auswahl hinzugezogen, um möglichst unterschiedliche Proteinkonformationen auszuwählen und um eine ausreichend große Unähnlichkeit zur Ausgangsstruktur zu erzielen.

Die Clusterergebnisse der **5CEO**-MD-Trajektorie sind in Abb. 5-17 dargestellt. In der ersten Trajektorie gibt es ein Cluster (1.3) mit einer besonders hohen Anzahl an Proteinkonformationen. Es ist der größte Cluster der Trajektorie und deckt 65 % der Simulationszeit ab.

Der Clusterrepräsentant dieses Clusters (Frame 2106, s. Abb. 5-18) wurde ausgewählt. Das Bindetaschenvolumen von 1112 \AA^3 ist um einiges größer als das Volumen der Ausgangsstruktur (808 \AA^3) und bietet Platz für sperrige Liganden. In der Darstellung der Bindetasche (s. Abb. 5-20) ist zu erkennen, dass dieser Volumengewinn auf die Verschiebung des P-Loops nach außen zurückzuführen ist.

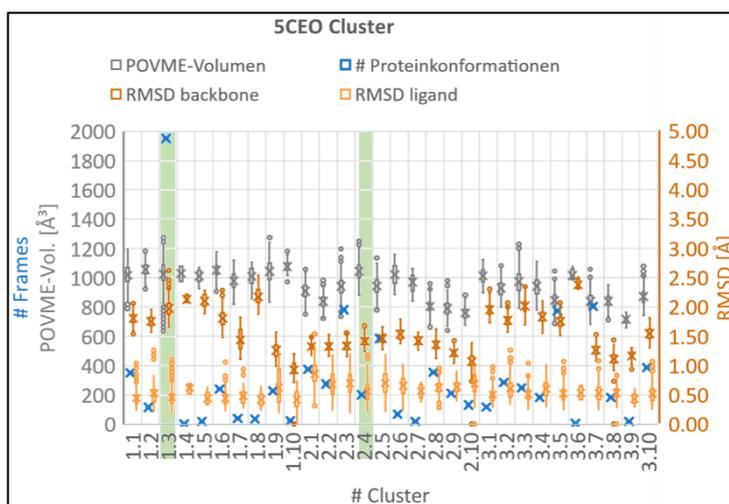


Abb. 5-17 – Ergebnisse des Clustering der Frames der jeweiligen **5CEO**-MD-Trajektorien. Cluster 1.1 bis 1.10 gehören zu *Run 01* etc.; Cluster, aus denen je ein Frame ausgewählt wurde, sind grün hinterlegt.

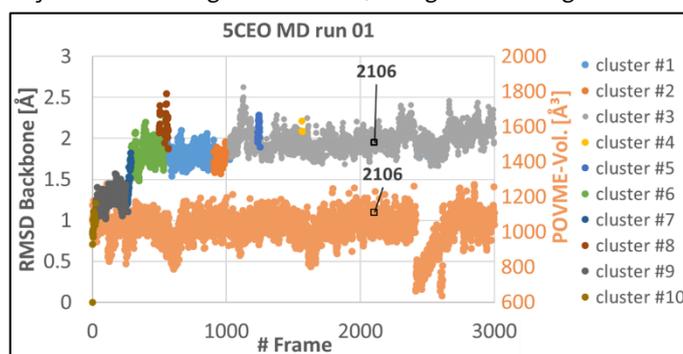
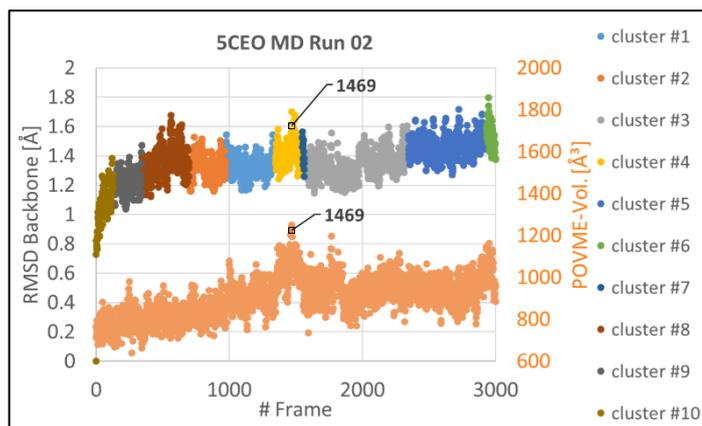


Abb. 5-18 – RMSD-Verlauf der ersten **5CEO**-MD-Trajektorie. Frame 2106 als Clusterrepräsentant des größten Clusters (#3) wurde für weitere Dockingversuche, vorbereitend für das virtuelle Screening, ausgewählt.

Cluster #4 aus der zweiten **5CEO**-MD-Trajektorie beinhaltet Proteinkonformationen mit tendenziell großem Bindetaschenvolumen. Der Frame mit dem größten Bindetaschenvolumen sollte ausgewählt werden. Bei der Auswahl auf Basis des Bindetaschenvolumens wurde versehentlich immer ein Frame



verschoben zum eigentlich gewünschten Frame ausgewählt, da die Ergebnisse der POVME-Berechnungen eine Nummerierung der Frames von 1 bis 3001 enthielten, während MAESTRO die Trajektorie von 0 bis 3000 nummeriert. Der ausgewählte Frame 1469 aus der

Abb. 5-19 – Verlauf des RMSD Backbone der zweiten **5CEO**-MD-Trajektorie. Der Verlauf des Bindetaschenvolumens (hellorange) zeigt ein Maximum im Cluster #4. Auswahl des Frames 1469 aufgrund eines hohen Bindetaschenvolumens.

zweiten **5CEO**-Trajektorie mit 1125 Å³ (s. Abb. 5-19) hat daher „nur“ das zweitgrößte Bindetaschenvolumen aller Proteinkonformationen dieser Trajektorie.

Die Gegenüberstellung der beiden ausgewählten Frames im Vergleich zur **5CEO-PL**-Struktur (s. Tab. 5-9 u. Abb. 5-20) veranschaulicht die Variabilität zwischen den drei ausgewählten Strukturen aus der **5CEO**-Simulation für die Dockingversuche.

Tab. 5-9 – Eigenschaften der **5CEO-PL**-Struktur und der zwei ausgewählten Frames aus den **5CEO**-MD-Simulationen für Dockingversuche zur Vorbereitung des virtuellen Screenings. PB = Protein-Backbone; Vol.: Bindetaschenvolumen (POVME), Dist. = Abstand, L = Ligand.

	RMSD PB	Vol.	Dist. F192 - G132	Dist. M190 - A135	RMSD L
5CEO-PL-Struktur	0 Å	808 Å ³	13,20 Å	15,45 Å	0 Å
5CEO Run 01 Frame 2106	1,948 Å	1112 Å ³	12,80 Å	13,14 Å	0,371 Å
5CEO Run 02 Frame 1469	1,604 Å	1225 Å ³	13,50 Å	14,05 Å	0,633 Å

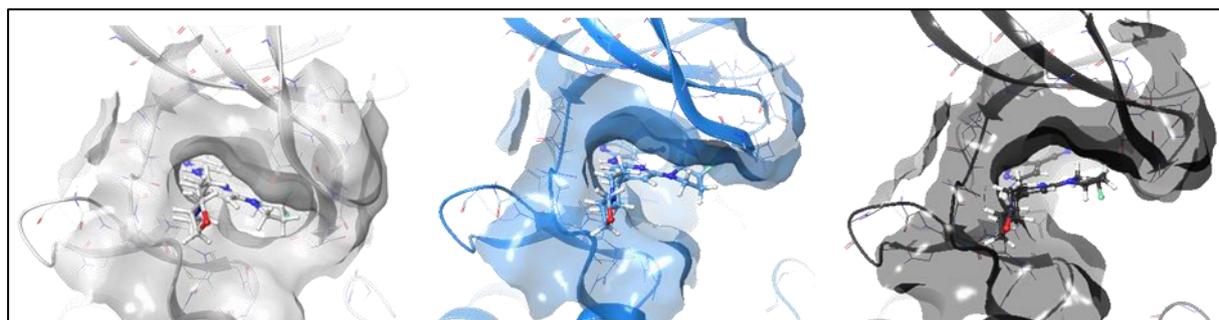


Abb. 5-20 – Bindetaschen der drei ausgewählten DLK-Proteinkonformationen auf Basis von **5CEO**. Von links nach rechts: **5CEO-PL**-Struktur (weiß), Frame 2106 *Run 01* (blau), Frame 1469 *Run 02* (schwarz).

Nach Clustering der drei **5CEN**-MD-Trajektorien (s. Abb. 5-21) wurden Proteinkonformationen mit besonders hohem Bindetaschenvolumen in Cluster 1.10 identifiziert, die relativ stark von der **5CEN**-Ausgangsstruktur abweichen (durchschnittlicher RMSD des Clusters #10 = 1,788 Å).

Frame 2006 aus diesem Cluster (s.

Abb. 5-22) enthält wegen der um 1 verschobenen Nummerierung bei der Auswahl (s. S. 142) ein Bindetaschenvolumen von 1128 Å³ (statt 1235 Å³) und befindet sich auf dem 19. Platz bezüglich des Bindetaschenvolumens. Zum Vergleich: Die MD-Ausgangsstruktur **5CEN_opls3_ow** hat lediglich ein Bindetaschenvolumen von 663 Å³.

Als zweiter interessanter Cluster stellte sich Cluster 3.6 (s. Abb. 5-21) heraus, da hier die meisten Proteinkonformationen enthalten sind und dies auf eine besondere Stabilität dieser Proteinkonformationen hinweist. Cluster #6 enthält 1194 Frames und macht damit mehr als ein Drittel der Trajektorie aus. Der ausgewählte Frame 1702 (s. Abb. 5-23) wurde ausgewählt, weil er die größte Bindetasche in diesem Cluster besitzen sollte (923 Å³). Das Volumen (wegen fehlerhafter Nummerierung, s.o.) beträgt allerdings 828 Å³, welches immer noch erhöht ist im Vergleich zur Ausgangsstruktur **5CEN_opls3_ow** (663 Å³) und dem anderen ausgewählten Frame (1128 Å³, s. Tab. 5-10 und Abb. 5-24).

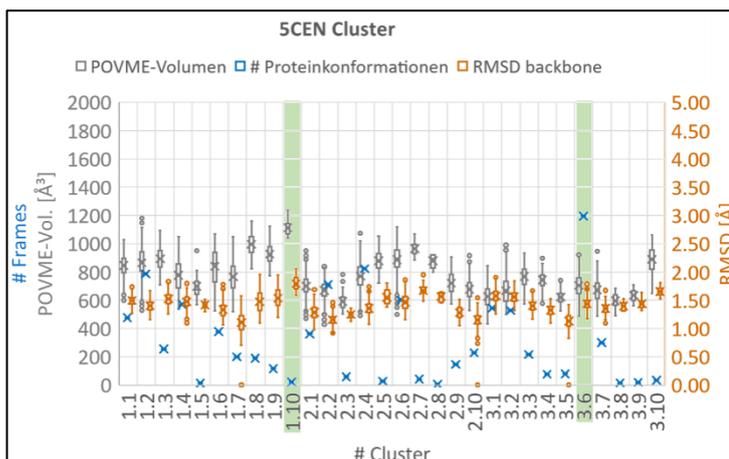


Abb. 5-21 – Ergebnisse des Clusterings der Frames der jeweiligen **5CEN**-MD-Trajektorien. Cluster 1.1 bis 1.10 gehören zu *Run 01* etc.; Cluster, aus denen je ein Frame ausgewählt wurde, sind grün hinterlegt.

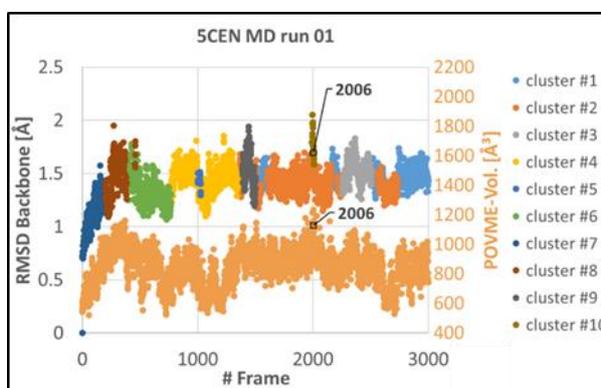


Abb. 5-22 – Verlauf des RMSD Backbone der ersten **5CEN**-MD-Trajektorie. Der Verlauf des Bindetaschenvolumens (hellorange) zeigt ein Maximum im Cluster #10. Auswahl des Frames 2006 aus diesem Cluster.

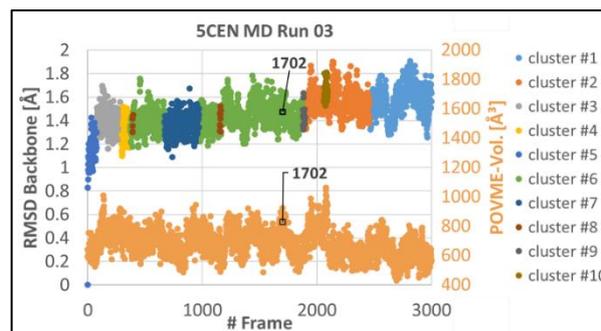


Abb. 5-23 – Verlauf des RMSD Backbone der dritten **5CEN**-MD-Trajektorie. Auswahl des Frames 1702 aus dem größten Cluster #6.

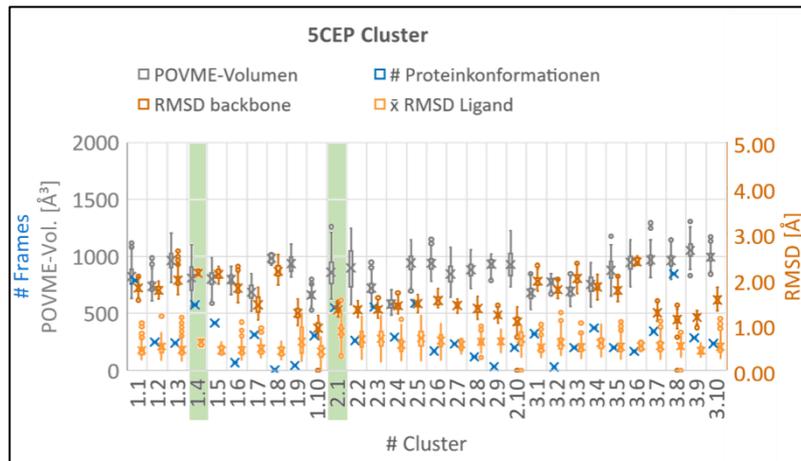
Tab. 5-10 – Eigenschaften der vorbereiteten **5CEN**-Struktur **5CEN_opls3_ow** und der zwei ausgewählten Strukturen aus den **5CEN**-MD-Simulationen für die Dockinguntersuchungen. PB = Protein-Backbone; Vol.: Bindetaschenvolumen (POVME), Dist. = Abstand, L = Ligand.

	RMSD PB	Vol.	Dist. F192 - G132	Dist. M190 - A135
5CEN_opls3_ow	0 Å	663 Å ³	11,90 Å	14,03 Å
5CEN Run 01 Frame 2006	1,697 Å	1128 Å ³	12,10 Å	16,99 Å
5CEN Run 03 Frame 1702	1,472 Å	828 Å ³	14,67 Å	14,01 Å



Abb. 5-24 – Bindetaschen der drei ausgewählten DLK-Proteinkonformationen auf Basis von **5CEN**. Von links nach rechts: **5CEN_opls3_ow** (gelb), Frame 2006 Run 01 (rot), Frame 1702 Run 03 (orange).

Das Clustering der **5CEP**-MD-Trajektorien ergab kleinere Cluster (s. Abb. 5-25) im Vergleich zu den anderen Trajektorien, welches wiederum auf die höhere Variabilität der **5CEP**-MD-Trajektorien hindeutet.



Der größte Cluster in der ersten Trajektorie (1.4, s. Abb. 5-25 und Abb. 5-26) enthält 26 % der Proteinkonformationen.

Abb. 5-25 – Ergebnisse des Clusterings der Frames der jeweiligen **5CEP**-MD-Trajektorien. Cluster 1.1 bis 1.10 gehören zu Run 01 etc.; Cluster, aus denen je ein Frame ausgewählt wurde, sind grün hinterlegt.

Aus diesem Cluster wurde der Clusterrepräsentant (Frame 2492) für die Vorbereitungen des virtuellen Screenings ausgewählt. Das Bindetaschenvolumen von Frame 2492 mit 818 Å³ entspricht etwa dem Volumen der **5CEO-PL**-Struktur (808 Å³), stellt aber im Vergleich zur **5CEP-PL**-Struktur (724 Å³) eine Vergrößerung dar.

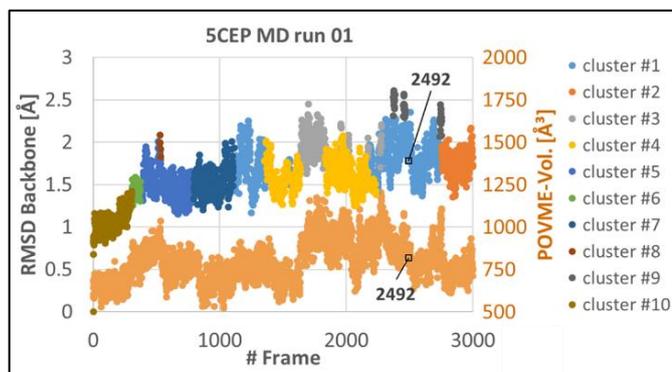


Abb. 5-26 – RMSD-Verlauf der ersten **5CEP**-MD-Trajektorie mit farbiger Darstellung der unterschiedlichen Cluster. Frame 2492 als Clusterrepräsentant des größten Clusters (#4) wurde für weitere Dockingversuche, vorbereitend für das virtuelle Screening, ausgewählt.

Cluster 2.1 (s. Abb. 5-25) zählt mit 550 Frames zu den großen **5CEP**-Clustern und enthält Proteinkonformationen mit hohem Bindetaschenvolumen.

Aus diesem Cluster wurde der Frame 1089, ausgewählt (s. Abb. 5-27). Er enthält aufgrund der starken Auslenkung des P-Loops nach außen ein stark vergrößertes Bindetaschenvolumen (1159 Å³, s. Abb.

5-28). Diese Ausdehnung ist auch aus dem vergrößerten Abstand zwischen Phe 192 und Gly132 (14,06 Å) und zwischen Met190 und Ala135 (18,2 Å) ersichtlich (s. Tab. 5-11).

Tab. 5-11 – Eigenschaften der **5CEP-PL**-Struktur und der zwei ausgewählten Frames aus den **5CEP**-MD-Trajektorien für Dockingstudien. PB = Protein-Backbone; Vol.: Bindetaschenvolumen (POVME), Dist. = Abstand, L = Ligand.

	RMSD PB	Vol.	Dist. F192 - G132	Dist. M190 - A135	RMSD L
5CEP-PL-Struktur	0 Å	724 Å ³	12,79 Å	15,47 Å	0 Å
5CEP Run 01 Frame 2492	1,872 Å	818 Å ³	13,01 Å	15,40 Å	0,709 Å
5CEP Run 02 Frame 1089	1,618 Å	1159 Å ³	14,06 Å	18,28 Å	0,653 Å

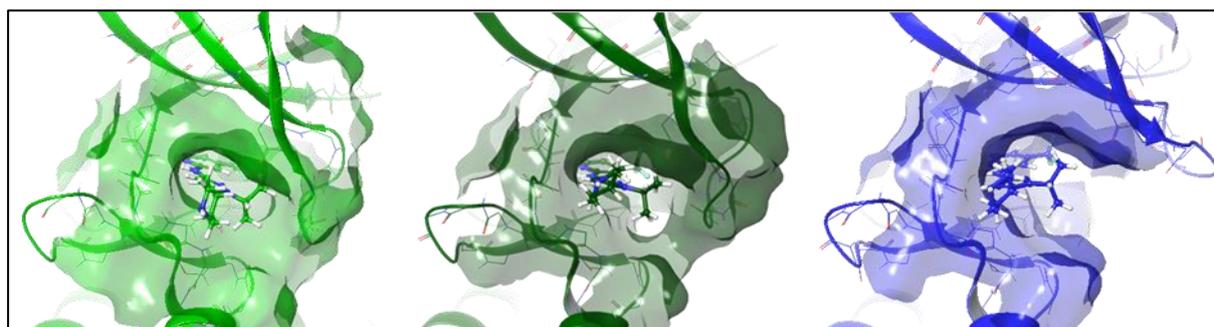


Abb. 5-28 – Bindetaschen der drei ausgewählten DLK-Proteinkonformationen auf Basis von **5CEP**. Von links nach rechts: **5CEP-PL**-Struktur (grün), Frame 2492 *Run 01* (dunkelgrün), Frame 1089 *Run 02* (blau).

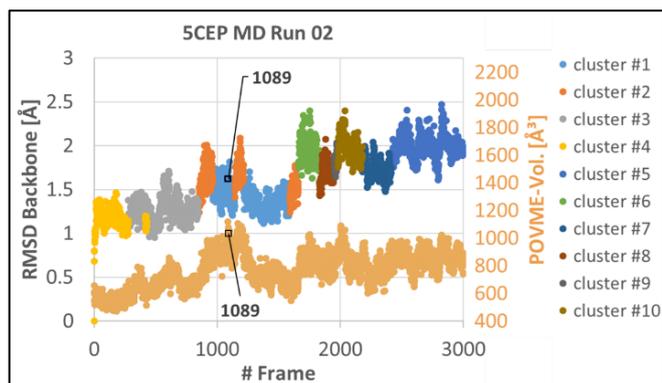


Abb. 5-27 – Verlauf des RMSD Backbone der zweiten **5CEP**-MD-Trajektorie mit farbiger Darstellung der Cluster. Auswahl des Frames 1089 mit vergleichsweise großem Bindetaschenvolumen.

Das Clustering der **5CEQ**-MD-Trajektorien ergab – wie aus den vorherigen Analysen wegen der hohen Stabilität der RMSD-Verläufe zu erwarten war – ein bis zwei sehr große Cluster pro Trajektorie neben vielen kleinen Clustern (s. Abb. 5-29).

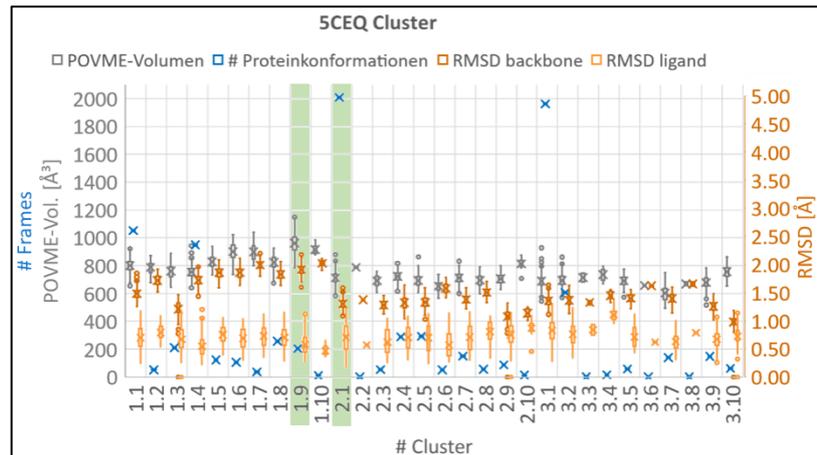


Abb. 5-29 – Ergebnisse des Clustering der Frames der jeweiligen **5CEQ**-MD-Trajektorien. Cluster 1.1 bis 1.10 gehören zu *Run 01* etc.; Cluster, aus denen je ein Frame ausgewählt wurde, sind grün hinterlegt.

Um eine möglichst von der Ausgangsstruktur abweichende Proteinkonformation zu erhalten, wurde zum einen eine Konformation aus der ersten MD-Trajektorie ausgewählt, die im Vergleich zu den beiden anderen sehr stabilen Simulationen Proteinkonformationen mit einer stärkeren Abweichung von der Ausgangsstruktur in der zweiten Hälfte der Simulationszeit ergab (vgl. dazu auch Abb. 5-12, S. 134).

Zum Ende der Simulationszeit ergab sich hierbei eine Erhöhung des Bindetaschenvolumens (s. Abb. 5-30). Das durchschnittliche Bindetaschenvolumen der Proteinkonformationen des Clusters #9 beträgt $961,0 \text{ \AA}^3$ und ist damit das größte Volumen aller Cluster der drei **5CEQ**-Simulationen (s. Abb. 5-29). Frame 2899 mit einem Volumen von 951 \AA^3 (statt 1147 \AA^3 , s.o.; s. Abb. 5-30) wurde ausgewählt. Dieser weist

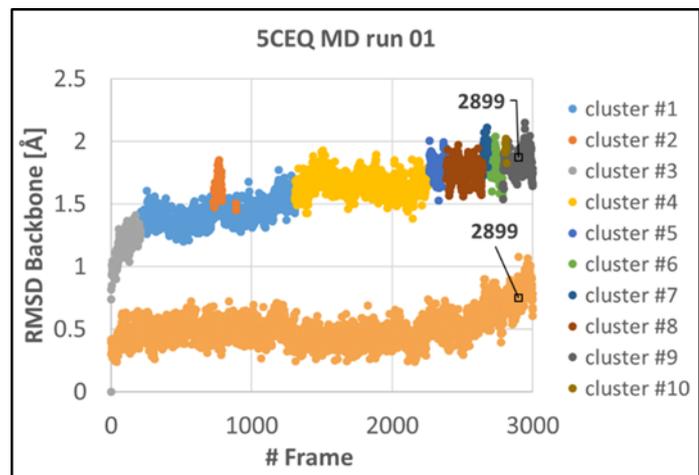


Abb. 5-30 – Verlauf des RMSD Backbone und des Bindetaschenvolumens der ersten **5CEQ**-MD-Trajektorie. Auswahl des Frames 2899 mit vergleichsweise erhöhtem Bindetaschenvolumen.

im Vergleich zur **5CEQ-PL**-Struktur (746 \AA^3) ein deutlich größeres Volumen auf.

Im Gegensatz zu diesem von der Norm stark abweichenden Bindetasche wurde ein weiterer Frame aus dem sehr stabilen Cluster 2.1 (2007 Frames) ausgewählt. Die Größe des Clusters deutet auf eine besonders hohe Stabilität der Clustermitglieder hin.

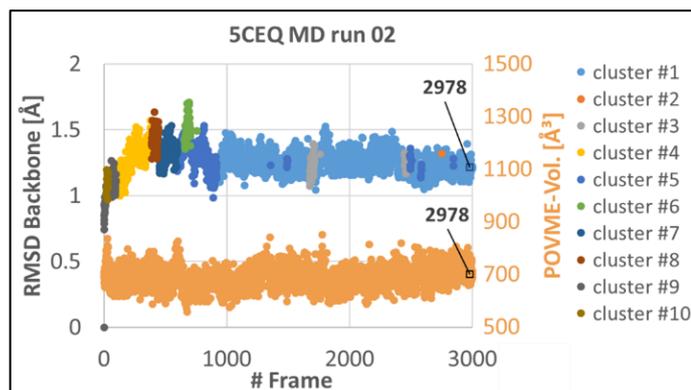


Abb. 5-31 – Verlauf des RMSD Backbone der zweiten **5CEQ**-MD-Trajektorie mit farbiger Darstellung der Cluster. Auswahl von Frame 2978 als Mitglied des größten Clusters.

Hier sollte eine Proteinkonformation mit leicht erhöhtem Bindetaschenvolumen ausgewählt werden. Der ausgewählte Frame 2978 (s. Abb. 5-31) besitzt ein vergleichbares

Bindetaschenvolumen (723 \AA^3 statt leicht erhöht, s.o.) wie die **5CEQ-PL**-Struktur (746 \AA^3). Von allen für die Dockingstudien ausgewählten Proteinkonformationen ist dies die Proteinkonformation, die der verwendeten Ausgangsstruktur am ähnlichsten ist ($\text{RMSD} = 1,215 \text{ \AA}$). Die berechneten Eigenschaften für die ausgewählten **5CEQ**-basierten Strukturen werden in Tab. 5-12 aufgeführt und die Bindetaschen in Abb. 5-32 dargestellt.

Tab. 5-12 – Berechnete Eigenschaften der **5CEQ-PL**-Struktur und der zwei ausgewählten Frames aus den **5CEQ**-MD-Trajektorien. PB = Protein-Backbone; Vol.: Bindetaschenvolumen (POVME), Dist. = Abstand, L = Ligand

	RMSD PB	Vol.	Dist. F192 - G132	Dist. M190 - A135	RMSD L
5CEQ-PL-Struktur	0 Å	746 Å	12,79 Å	15,46 Å	0 Å
5CEQ Run 01 Frame 2899	1,872 Å	951 Å	12,32 Å	16,25 Å	0,765 Å
5CEQ Run 02 Frame 2978	1,215 Å	723 Å	12,88 Å	15,59 Å	0,379 Å

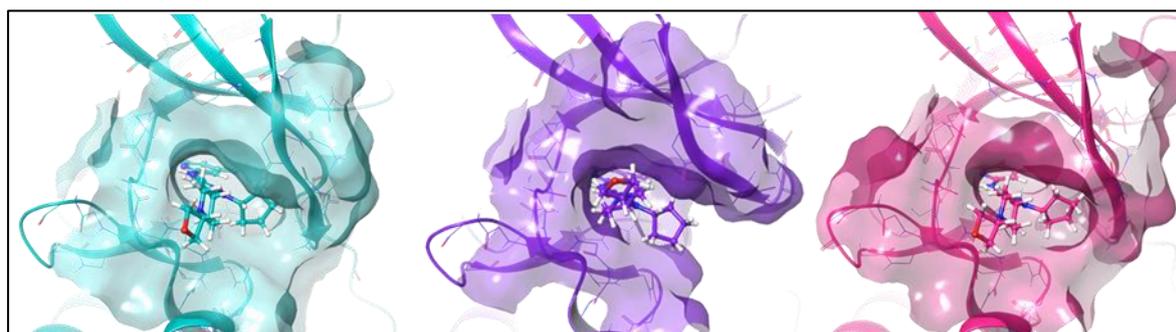


Abb. 5-32 – Bindetaschen der drei ausgewählten DLK-Proteinkonformationen auf Basis von **5CEQ**. Von links nach rechts: **5CEQ-PL**-Struktur (cyan), Frame 2899 *Run 01* (violett), Frame 2978 *Run 02* (magenta).

5.2 Analyse und Vorbereitung der Liganden

5.2.1 Analyse bekannter DLK-Inhibitoren

835 Verbindungen mit bekannter DLK-Hemmaktivität aus der Literatur wurden in dieser Arbeit verwendet, von denen definierte K_i -, K_d - bzw. IC_{50} -Werte bekannt sind und die in ihrer Struktur, einschließlich der Stereokonfiguration, eindeutig festgelegt sind. 805 dieser 835 Strukturen stammen entweder aus Veröffentlichungen von PATEL ET AL. [70, 166] oder aus diversen Patentschriften [170–172, 174, 175] dieser Forschungsgruppe und sind strukturell nach dem gleichen Prinzip aufgebaut (s. S. 148 ff.). Die restlichen 30 [125, 164, 165] Strukturen weichen in ihrer Struktur von diesen ab und lassen sich teilweise in unterschiedliche strukturelle Gruppen einteilen (s. S. 152 ff.). Sie sind tendenziell weniger aktiv und selektiv als die 805 erstgenannten Strukturen und wurden entweder als Hemmstoffe für andere Kinasen (Rezeptor-Tyrosinkinasen, Cyclin-abhängige Kinasen, andere Serin/Threonin-Kinasen, wie z.B. Aurora-Kinasen oder Raf-Kinasen) entwickelt oder gelten als unspezifische Proteinkinase-Inhibitoren. Dennoch sollten auch diese Strukturen in dieser Arbeit berücksichtigt werden, da sie die strukturelle Diversität der bekannten DLK-Inhibitoren erhöhen und den Aktivitätsbereich erweitern.

SAR

805 der 835 analysierten DLK-Inhibitoren gehören zum gleichen Chemotyp wie GNE-3511 (**7**) [166] (cokristallisierender Ligand in der PDB **5CEO**). Aufbau

und Wechselwirkungen dieses Liganden mit der DLK-

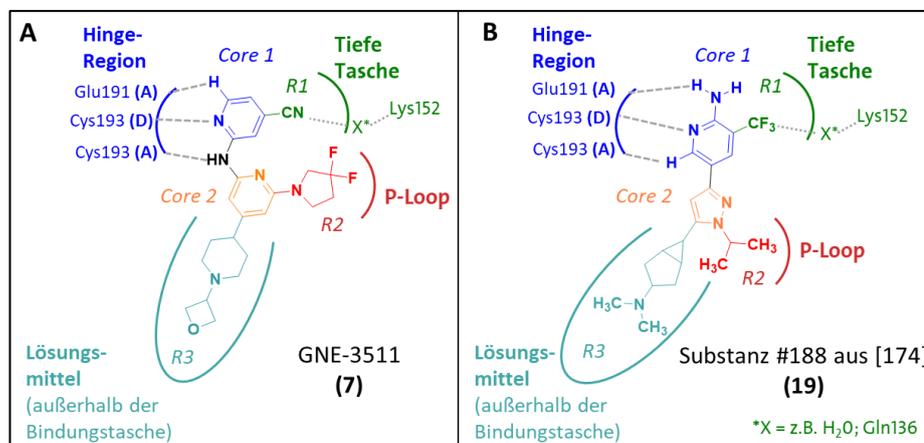


Abb. 5-33 – SAR selektiver DLK-Inhibitoren. **A** – Am Beispiel des GNE-3511 (**7**) (Ligand in **5CEO**): Die beiden aromatischen Ringe sind über eine Aminofunktion miteinander verknüpft. **B** – Am Beispiel der Substanz Nr. 188 (**19**) aus der Patentschrift WO 2015/091889 [174]; die aromatischen Ringe sind direkt miteinander verknüpft.

Bindetasche wurden bereits geschildert (s. Kap. 5.1.1, S. 125, Abb. 5-5). Ergänzend wird hier ein weiterer, strukturell verwandter Ligand, Verbindung Nr. 188 (**19**) aus der Patentschrift WO2015/091889 [174], dargestellt (s. Abb. 5-33). Mithilfe dieser beiden Strukturen wird im Folgenden der Aufbau der einzelnen Strukturmodule Core 1, Core 2 und der drei Substituenten R1-3 der hier analysierten 805 Verbindungen erläutert.

Core 1 (Hinge-Anker): Insgesamt werden alle 805 selektiven DLK-Inhibitoren von lediglich 13 verschiedenen N-Heterocyclen an dieser Position repräsentiert (s. Abb. 5-34). Dies ist am häufigsten 2-Aminopyridin (Stickstoffatom in *meta*-Position zur Core-2-Verknüpfungsstelle, s. Abb. 5-34 A), Pyridin (Stickstoffatom in *ortho*-Position zur Core-2-Verknüpfungsstelle, s. Abb. 5-34 B), 2-Aminopyrimidin (s. Abb. 5-34 C) und 7-Azaindol (s. Abb. 5-34 G). Die Aminofunktion des 2-Aminopyridins (s. Abb. 5-34 A) kann in einem Cyclus eingebunden sein (s. Abb. 5-34 G, H und I).

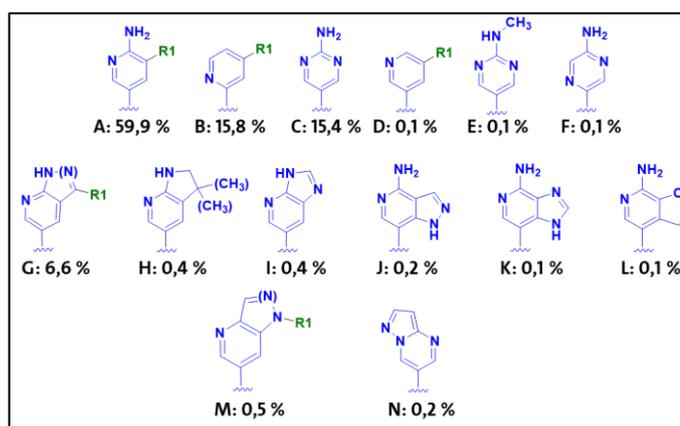


Abb. 5-34 – Vorkommen der verschiedenen N-Heterocyclen in Core 1 (s. auch Abb. 5-33, S. 148) in den 805 strukturell ähnlichen, selektiven DLK-Inhibitoren. Die Aromaten G-L können als Derivate des 2-Aminopyridins (C) angesehen werden. Die anellierte Fünfferringe können auch als Rest (z.B. R1 zur tiefen Tasche) interpretiert werden.

Core 2 (Verteiler): Core 2 ist entweder – so wie im GNE-3511 (**7**) – über eine Aminofunktion (zu 15,8 % der hier analysierten Verbindungen, s. Abb. 5-33 A) oder – wie in (**19**) – direkt mit Core 1 verbunden (C-C-Verknüpfung, 84,2 % der Fälle, s. Abb. 5-33 B). Core 2 wird am häufigsten von einem Pyrazol (s. Abb. 5-35 A und D), Pyrimidin (Stickstoffatome in *ortho*- und *para*-Position zur Core-1-Verknüpfungsstelle, s. Abb. 5-35 B), Pyridin (Stickstoffatom in *ortho*-Position zur Core-1-Verknüpfungsstelle, s. Abb. 5-35 C), oder einem Purinderivat (s. Abb. 5-35 F-H) repräsentiert wird. Im letzteren Ringsystem wird der hydrophile Rest (R3, s. S. 148, Abb. 5-33) mit integriert und dadurch in seiner Beweglichkeit eingeschränkt („Rigidisierung“). Die hydrophoben Kontakte zur Seitenkette des Val131 (oberhalb) und zur Methylengruppe des Gly196-Backbones (unterhalb) halten Core 2 als Sandwich-Komplex in Position.

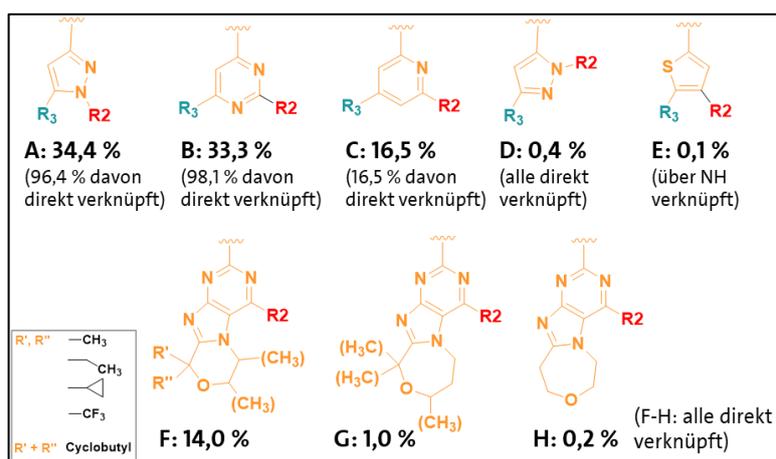


Abb. 5-35 – Vorkommen der verschiedenen Aromaten an der Position Core 2 (siehe dazu Abb. 5-33, S. 148) in den 805 strukturell ähnlichen, selektiven DLK-Inhibitoren. Die Aromaten A, B, C und F sind am häufigsten vertreten. Die Aromaten F, G und H enthalten den hydrophilen Rest R3 in zyklischer Form (Rigidisierung des R3).

Die hydrophoben Kontakte zur Seitenkette des Val131 (oberhalb) und zur Methylengruppe des Gly196-Backbones (unterhalb) halten Core 2 als Sandwich-Komplex in Position.

R1: Der Substituent R1 an Core 1, in *meta*-Position zur Core-2-Verknüpfungsstelle

(vgl. Abb. 5-33), stellt einen kleinen Rest dar (s. Abb. 5-36), welcher in eine tief liegende Tasche, begrenzt durch

Met190, Lys152 und Seitenketten des P-Loops (z.B. Val139),

hineinragt. Hier werden vorrangig vdW- und elektrostatische Wechselwirkungen (s. S. 148, Abb. 5-33) ausgebildet. Am Häufigsten handelt es sich um eine Difluoromethoxy-, Trifluormethyl- oder eine Nitrilfunktion (s. Abb. 5-36 A, B, D). Im Falle eines Aminopyrimidins als Core 1 (vgl. Abb. 5-34 C, E und F, S. 149) fehlt R1.

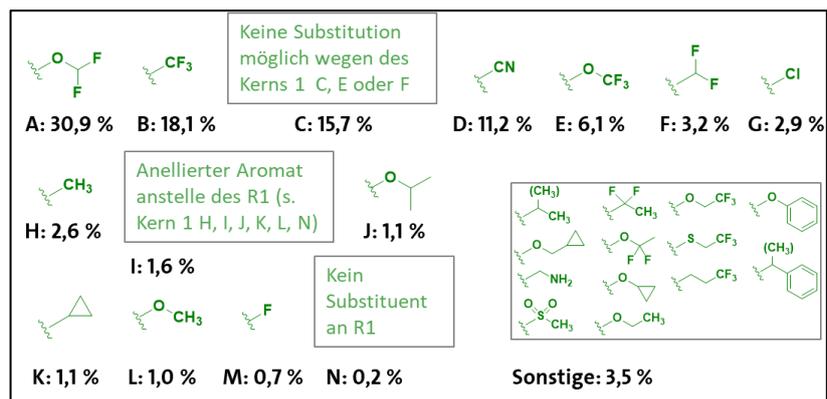


Abb. 5-36 – Vorkommen des Substituenten R1 an Core 1 in *meta*-Position zur Core-2-Verknüpfungsstelle (vgl. dazu Abb. 5-33, S. 148) in den 805 strukturell ähnlichen, selektiven DLK-Inhibitoren.

R2: Die Variabilität der Substituenten, welche über hydrophobe Wechselwirkungen mit dem P-Loop interagieren, ist deutlich höher als in den bisher beschriebenen Struktureinheiten. In Abb. 5-37 sind die häufigsten Strukturen aufgeführt, welche zusammen 80 % der R2-Substituenten abdecken. Sie sind zum größten Teil lipophil. Am häufigsten sind Isopropyl- (s. Abb. 5-37 A), 2-Azabicyclo[2.1.1]hexyl- (s. Abb. 5-37 B),

Cyclopropyl- (s. Abb. 5-37 C) und 3,3-Difluoropyrrolidylsubstituenten (s. Abb. 5-37 D) vertreten. Die Grundgerüste der Substituenten an dieser Position bestehen häufig aus unterschiedlich großen stickstoff-, sauerstoffhaltigen oder reinen Kohlenstoffzyklen ($n = 3-7$), welche häufig mit Fluoratom substituiert sind. Auch polyzyklische Verbindungen sind vertreten.

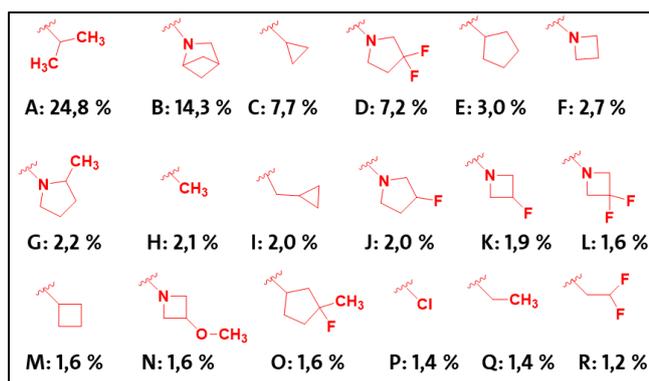


Abb. 5-37 – Die häufigsten Struktureinheiten für den Substituenten R2, welcher in Kontakt mit dem P-Loop der DLK steht (vgl. Abb. 5-33, S. 148), in den 805 strukturell ähnlichen, selektiven DLK-Inhibitoren.

R3: Die Substituenten, welche in das Lösungsmittel hinausragen und daher zum größten Teil hydrophil sind, sind tendenziell größer als die Substituenten R1 und R2 (s. Abb. 5-38). Ebenso wie in R2 ist die Variabilität der Struktureinheiten von R3 sehr hoch. Wie schon für den Core 2 beschrieben, ist der hydrophile Rest R3 sehr häufig zusammen mit Core 2 in einem anellierten trizyklischen aromatischen System integriert (s. dazu Abb. 5-35 F-H, S. 149).

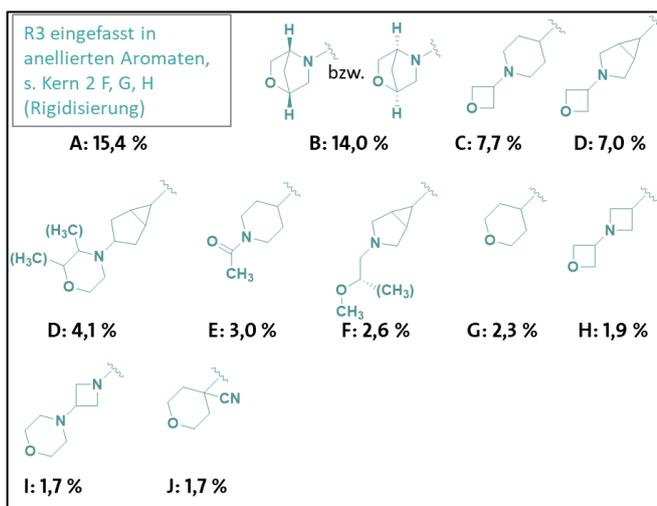


Abb. 5-38 – Die häufigsten Struktureinheiten für den Substituenten R3, welcher zum Lösungsmittel herausragt (vgl. Abb. 5-33, S. 148), in den 805 strukturell ähnlichen, selektiven DLK-Inhibitoren.

Auffällig ist das Vorkommen von gleichen Grundbausteinen in R3 (s. Abb. 5-39⁴³). Sehr häufig sind substituierte Bicyclen unter diesen Grundbausteinen vertreten (s. Abb. 5-39 A, C, F). Desweiteren sind sehr häufig am Stickstoff substituierte Piperidiny- und Piperazinygruppen (s. Abb. 5-39 B und G) sowie am Stickstoff oder dem Stickstoffatom gegenüberliegenden Kohlenstoff substituierte Azetidin-Einheiten als Grundbaustein zu finden.

Diese Grundbausteine sind im Allgemeinen mit kleinen Resten substituiert, wie z.B. mit Oxetan- bzw. Azetidin-Einheiten, Acetylgruppen, Halogen(alkan)en oder heteroatomhaltigen Spiroverbindungen. Die Bicyklen (z.B. Abb. 5-39 A, C und F) liegen alle in einer solchen Stereokonformation vor, dass der zum Lösungsmittel gerichtete Rest auf der anderen Seite der von der Cyclopropyleinheit gebildeten Ebene liegt als der Core-2-haltige Rest (s. Abb. 5-40). Nur bei einer der untersuch-

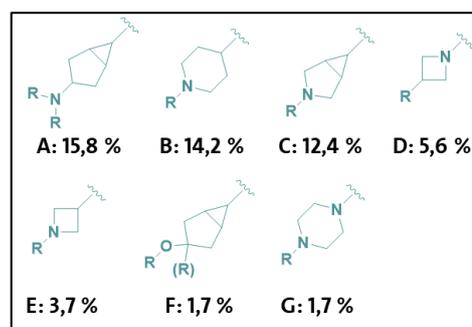


Abb. 5-39 – Häufig vorkommende Grundbausteine der R3-Substituenten in den 805 strukturell ähnlichen, selektiven DLK-Inhibitoren.

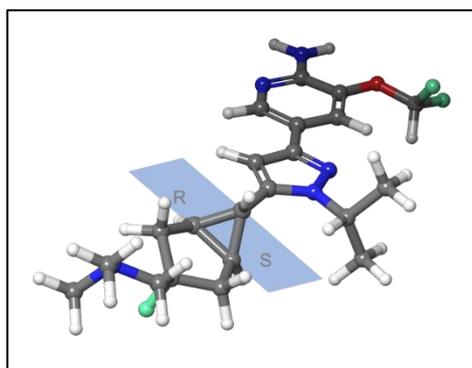


Abb. 5-40 – Substanz Nr. 178 aus der Patentschrift WO 2015/091889 [174]; die Substituenten am Dreiering zeigen in unterschiedliche Richtungen bezüglich der vom Dreiering gebildeten Ebene.

⁴³ Hier wurden die Grundbausteine der an Core 2 anellierten Strukturen (s. Abb. 5-35, F-H) nicht extra mit aufgeführt.

ten Verbindungen, Substanz Nr. 283 aus der Patentschrift WO 2015/091889 [174], trifft dies nicht zu. Hier liegen beide Reste auf der gleichen Seite der Ebene.

Die Strukturen und Aktivitätsdaten zur DLK-Hemmung der übrigen 30 Verbindungen wurden aus den Veröffentlichungen von KARAMAN ET AL. [164], FERRARIS ET AL. [125] und TAO ET AL. [165] entnommen. Sie lassen sich nach ihrem Aufbau in verschiedene Gruppen einteilen. Die Ausrichtung und Einfärbung der im Folgenden dargestellten Strukturen erfolgte in Analogie zu dem Bindungsmodell in Abb. 5-33, allerdings handelt es sich um einen vermuteten Bindungsmodus. Im Gegensatz zu den 805 vorher beschriebenen Strukturen, deren Bindungsmodus alle aus den DLK-Röntgenkristallstrukturen **5CEO**, **5CEP** und **5CEQ** abgeleitet werden können, kann der Bindungsmodus dieser unspezifischen DLK-Inhibitoren nur aus Röntgenkristallstrukturen anderer Kinasen abgeleitet werden (s. Kap. 7.10, S. 281, Tab. 7-6). Es ist nicht sicher, ob die Bindungsmodi so auch für die DLK zutreffen.

14 der 30 Verbindungen haben einen Staurosporin-ähnlichen Aufbau (s. Abb. 5-41). Sie binden mit der Amidgruppe des γ -Lactams an die Hinge-Region der ATP-Bindetasche von diversen Kinasen. CEP-701 (**20**) hemmt vor allem die MKK7, während CEP-1347 (**1**), CEP-5214 (**31**) und (+)K-252a (**21**) die MLK-1 und MLK-2 hemmen. Für die Strukturen aus der Veröffentlichung von TAO ET AL. [165] (s. Abb. 5-41, Verbindungen B) und die Strukturen CEP-1347 (**1**),

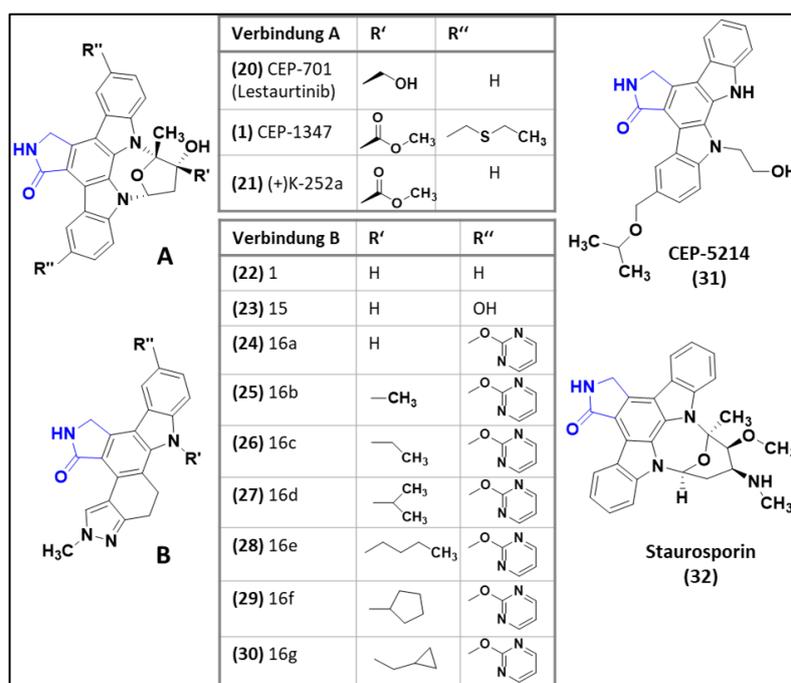


Abb. 5-41 – Indolocarbazole [125, 164] und Hexahydroindazolo[5,4-a]pyrrolo[3,4-c]carbazole und deren 8-Pyrimidinyloxy-Analoga [165]. Die Strukturen sind so ausgerichtet, wie sie in einer Kinasen-Bindetasche vorliegen würden (vgl. dazu Abb. 5-33, S.148 u. Kap. 7.10). Die Amidgruppe (blau) bildet Wasserstoffbrückenbindungen zur Hinge-Region aus. CEP5214 und (+)K-252a (**21**) sind IC₅₀-Werte für die DLK-Hemmung hinterlegt – bei allen anderen hier vorgestellten Strukturen handelt es sich um K_i- oder K_d-Werte.

Sunitinib (**2**) und SU-14813 (**33**) (s. Abb. 5-42) binden mutmaßlich mit dem Oxindol-Ring an die Hinge-Region der DLK-Bindetasche. Sie verfügen über einen kleinen Rest (Fluor bzw. fluoriertes anelliertes Benzen), welcher in die tiefe Tasche hineinragen könnte. Ein möglicherweise zum Lösungsmittel orientierter, hydrophiler Substituent (ähnlich wie in den zuvor beschriebenen, 805 strukturell ähnlichen DLK-Inhibitoren)

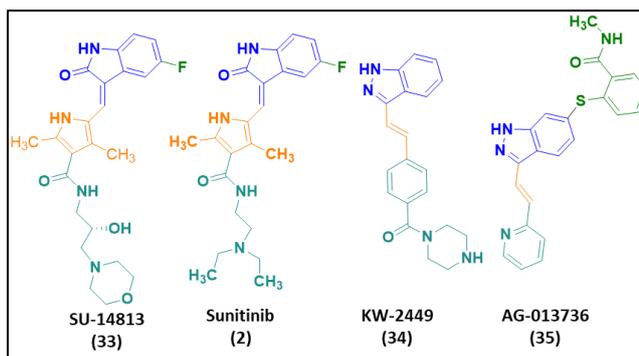


Abb. 5-42 – Oxindol- [164] und Indazol- [125] basierte Strukturen mit bekannter DLK-Aktivität. Der bicyclische Aromat (blau) bindet an die Hinge-Region. Die Ausrichtung und Einfärbung der Strukturen erfolgten nach dem Schema entsprechend zur Abb. 5-33, S. 148.

ist über einen dimethylierten Pyrrolring mit dem Hinge-interagierenden Aromaten verbunden.

Die Verbindungen KW-2449 (**34**) und AG-013736 (**35**) binden mit einem N-Heterocyclus als Core 1 – hier Indazol – an die Hinge-Region. Dieser ist über eine Vinylgruppe mit einem zum Lösungsmittel orientierten Rest verbunden.

Desweiteren gibt es DLK-hemmende Verbindungen, die mit einem Chinolinring an die Hinge-Region von Kinasen binden (s. Abb. 5-43). Sie enthalten – ähnlich wie die 805 selektiven DLK-Inhibitoren – einen zum Lösungsmittel gerichteten (cyan, R3) und einen mit dem P-Loop interagierenden Substituenten (rot, R2). HKI-272 (**36**) bindet darüber hinaus mit R2 kovalent an das Schwefelatom einer Cysteinseitenkette in der Bindetasche. Bei XL-880 (**37**) handelt es sich um einen Typ-II-Inhibitor von c-Met und VEGFR-2 (beides Tyrosinkinase) [517–519]. Dieser Inhibitortyp enthält einen meist sehr großen Substituenten (hier schwarz dargestellt), welcher in eine zusätzliche hydrophobe Tasche im DFG-out-Zustand der Kinase hineinragt (s. Kap. 3.2, S. 20). Der Bindungsmodus von Typ-II-Inhibitoren wurde in der DLK noch nicht kristallographisch erfasst. Eine andere Ausrichtung in der Bindetasche der DLK als hier dargestellt ist möglich, z.B. in einer nicht-Typ-II-typischen Position.

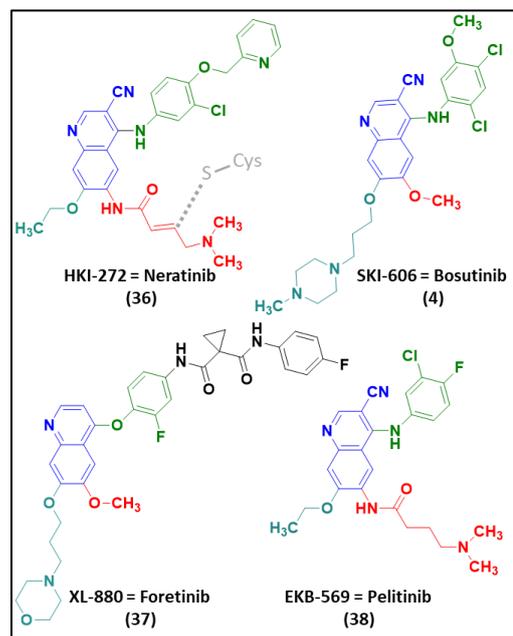


Abb. 5-43 – Chinolinderivate [125, 164] mit DLK-hemmender Aktivität. Die Ausrichtung der Strukturen, wie sie in der Bindetasche vorliegen würden, und die Einfärbung der Strukturen erfolgte nach dem Schema entsprechend zur Abb. 5-33, S. 148. Unter Addition einer Cystein-Thiolgruppe an die Doppelbindung des HKI-272 (**36**) kann eine kovalente Verknüpfung an die Bindetasche erfolgen.

Die übrigen acht Verbindungen sind strukturell sehr unterschiedlich (s. Abb. 5-44). Bei dem Hinge-interagierenden Aromaten (Core 1, vgl. Abb. 5-33) kann es sich um einen Pyrazol-, Imidazol-, oder Triazolring handeln. Desweiteren kommen Chinolin-2-on und Pyrimidin⁴⁴ an dieser Position vor. Das 2-Aminopyrimidin in Crizotinib (**3**) ist – als einziger Aromat an Core-1-Position dieser 30 hier beschriebenen Strukturen – auch in den selektiven DLK-Inhibitoren vertreten (vgl. dazu Abb. 5-34 A). Ansonsten stellen die Hinge-interagierenden Aromaten in den übrigen 29 Strukturen eine Verschiedenartigkeit zu den Core-1-Aromaten der 805 selektiven DLK-Inhibitoren dar.

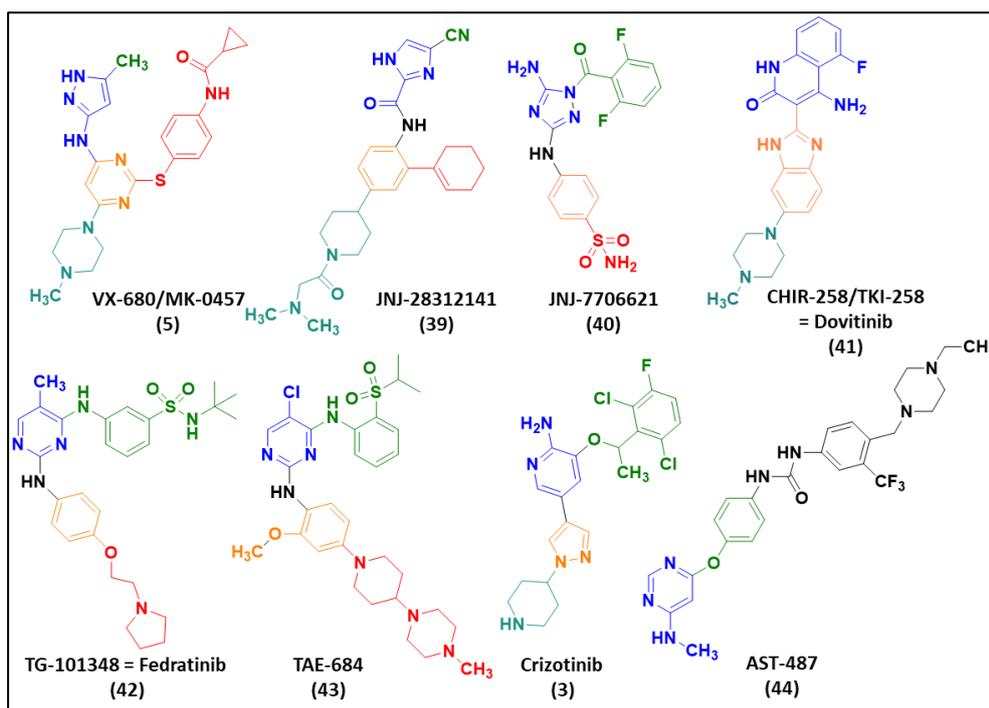


Abb. 5-44 – Aufbau verschiedenartiger Strukturen mit DLK-hemmender Aktivität [125, 164]. Die Ausrichtung der Strukturen, wie sie in der Bindetasche vorliegen würden, und die Einfärbung der Strukturen erfolgten nach dem Schema entsprechend zur Abb. 5-33, S. 148.

Auf Grundlage von Molecular-Modeling-Experimenten wird für AST-487 (**44**) (RET-, FLT3-, c-ABL-Kinaseinhibitor [520]) ein Typ-II-inhibitorischer Mechanismus postuliert [521], allerdings ist der Bindungsmodus in einer Kinase, anders als für XL-880 (**37**), nicht kristallographisch belegt. Anders als andere Typ-II-Inhibitoren sind AST-487 (**44**) und XL-880 (**37**) weniger selektiv [81, 83]. Es könnte sein, dass der N-Heterozyklus (s. Abb. 5-44, blau) mit der Hinge-Region interagiert und der Rest, wie in XL-880 (**37**) (vgl. Abb. 5-43) nach hinten in die durch DFG-out-Konformation einer inaktiven Kinase offenen hinteren Tasche hineinragt⁴⁵.

⁴⁴ Dieses Pyrimidin (AST-487 (**44**), Abb. 5-44) ist räumlich allerdings anders zur Hinge-Region ausgerichtet als die Pyrimidine an dieser Position in den zuerst beschriebenen, selektiven DLK-Inhibitoren (vgl. Abb. 5-34 C und E).

⁴⁵ Vgl. auch [522] und PDB **3KQ7** mit einem ähnlich großen Typ-II-Inhibitor und ähnlicher kinase-interagierender Einheit.

Für alle anderen in dieser Arbeit beschriebenen unspezifischen DLK-Inhibitoren ist ein Typ-I-Inhibitionsmechanismus in anderen Kinasen beschrieben worden und auch für die DLK am nächsten liegendsten.

Die strukturelle Variabilität der 805 strukturell ähnlichen Kinaseinhibitoren ist in Core 1 und 2 sowie auch im Substituenten R1 eher gering. Die Vielzahl der Strukturen wird zum größten Teil durch die Variabilität der Substituenten erreicht, welche mit dem P-Loop interagieren bzw. in das Lösungsmittel herausragen. Durch die 30 weiteren Verbindungen mit DLK-hemmender Aktivität wird die strukturelle Diversität des Datensatzes erhöht. Die Aromaten an den Positionen für Core 1 und 2 sind zum größten Teil andere als in den vorher beschriebenen 805 DLK-Inhibitoren.

5.2.2 Ligandendatensätze

Liganden für QSAR

Etwa 830 Strukturen⁴⁶ aus den in Kapitel 5.2.1 beschriebenen Patentschriften [170–172, 174] (alle außer WO 2016/142310 [175], da zu dem Zeitpunkt noch nicht veröffentlicht) und den Veröffentlichungen von PATEL ET AL. [70, 166] sowie 18 der 30 unspezifischen DLK-Inhibitoren – nämlich solche mit bekanntem K_d -Wert [125, 164] – wurden einem OPLS3-LIGPREP unterzogen (s. Kap. 7.11.1, Tab. 7-7, S. 282: „DLK-Inh. für QSAR“). Es wurden 1909 Liganden (*Multiple States*⁴⁷) generiert. Nachträglich wurden Liganden unwahrscheinlicher Protonierung gelöscht (s. Kap. 7.11.2). Es verblieben 1295 Strukturen (inkl. verschiedener Protonierungszustände und Tautomere), die dem SP-Docking und zum Teil einem Induced-Fit-Docking zur Erstellung des Docking-Alignments unterzogen wurden (s. Kap. 5.3.1). Erst danach wurden Ligandendatensätze zusammengestellt, die für die Erstellung der QSAR-Modelle eingesetzt wurden (s. Kap. 5.3.1, S. 166 f.).

Ligandendatensätze zur Optimierung und Kontrolle des Virtual-Screening-Workflows

In dieser Arbeit wurden Testdatensätze bestehend aus Liganden bekannter DLK-Aktivität (im Folgenden: Aktive) aus der Literatur (vgl. Kap. 5.2.1) und sogenannten Decoy-Molekülen eingesetzt. Wie in Kap. 4.4.8 (S. 78) erläutert, wird mit solchen Testdatensätzen die Fähigkeit des Virtual-Screening-Workflows überprüft, Aktive von mutmaßlich inaktiven Liganden zu trennen.

⁴⁶ Dies waren noch mehr als die in Kap. 5.2.1 beschriebenen 805 Strukturen, da hier noch nicht genügend aussortiert wurde bzgl. definierter Stereochemie, DLK-Aktivität und redundanten Strukturen.

⁴⁷ *Multiple States*: Alle Protonierungszustände und Tautomere eines Liganden werden mitgezählt.

Im Laufe der Arbeit wurden drei solcher Testdatensätze generiert, welche sich in ihrer Zusammensetzung und in der Art der Decoygenerierung unterscheiden und an unterschiedlichen Schnittstellen der Screening-Vorbereitung und -Durchführung eingesetzt wurden:

1.) Ligandenliste a (Lig. a)

Die Decoys wurden auf Basis von 206 Aktiven, ausgewählt aus den besten Dockinglösungen für das QSAR-Alignment mit möglichst gleichmäßiger pK_i -Werte-Verteilung (s. Tab. 5-13), generiert (s. Kap. 7.11.2, S. 283). Nach der Decoygenerierung wurde der Datensatz mit den Original-Liganden aus den PDBs **5CEO**, **5CEP** und **5CEQ**, den 30 unspezifischen DLK-Inhibitoren (s. S. 152 ff.) und ANP⁴⁸ aufgefüllt.

Insgesamt sind es 237 Aktive⁴⁹ und 10.292 Decoys. Mit dieser Ligandenliste wurden nach OPLS3- bzw. OPLS2005-LIGPREP (s. Kap. 7.11.1, Tab. 7-7, „Lig. a (Aktive) bzw. (Decoys)“) Testdockings in acht vorselektierte Proteinkonformationen (s. Kap. 5.4.1.3) durchgeführt.

Tab. 5-13 – Verteilung der pK_i -Werte der Aktiven zur Decoygenerierung der **Lig. a**

pK_i -Wertebereich	Anzahl
5,0 - 5,9	20
6,0 - 6,9	37
7,0 - 7,9	39
8,0 - 8,9	44
9,0 - 9,9	65
> 10	1

2.) Ligandenliste b (Lig. b)

Die Aktiven wurden aus allen zu dem Zeitpunkt bekannten selektiven DLK-Inhibitoren (aus den beiden PATEL-Veröffentlichungen und den Patentschriften bis auf WO 2016/142310 [175]) durch systematisches Abzählen so ausgewählt, dass die pK_i -Werte gleichmäßig über den Datensatz verteilt vorliegen (s. Tab. 5-14 u. Kap. 7.11.2). Der Datensatz wurde mit den 30 unspezifischen DLK-Inhibitoren und den drei originalen Liganden aus **5CEO**, **5CEP** und **5CEQ** aufgefüllt. Insgesamt besteht die Liste aus 170 Strukturen und 12.387 Decoys. Es wurde ein OPLS3-LIGPREP durchgeführt. Ein wichtiger Unterschied zur **Lig. a** ist, dass auch für die Decoys alle Protonierungszustände und Tautomere in dem definierten pH-Bereich gebildet wurden (s. Kap. 7.11.1, Tab. 7-7, „Lig. b (Aktive + Decoys)“). Der Verwendungszweck ist der gleiche wie für **Lig. a**.

Tab. 5-14 – Verteilung der pK_i -Werte der ausgewählten aktiven Verbindungen für die **Lig. b**

pK_i/pK_d -Wertebereich	Anzahl
5,0 - 5,9	26
6,0 - 6,9	32
7,0 - 7,9	36
8,0 - 8,9	31
9,0 - 9,9	31
ohne pK_i (stattdessen IC_{50})	12
unbekannt	2

Insgesamt besteht die Liste aus 170 Strukturen und 12.387 Decoys. Es wurde ein OPLS3-LIGPREP durchgeführt. Ein wichtiger Unterschied zur **Lig. a** ist, dass auch für die Decoys alle Protonierungszustände und Tautomere in dem definierten pH-Bereich gebildet wurden (s. Kap. 7.11.1, Tab. 7-7, „Lig. b (Aktive + Decoys)“). Der Verwendungszweck ist der gleiche wie für **Lig. a**.

3.) Ligandenliste c (Lig. c)

Ausgangspunkt waren alle in Kapitel 5.2.1 genannten Strukturen aus den benannten Literaturstellen mit definiertem K_i - oder K_d -Wert (nicht: IC_{50} -Wert) für die DLK. Dabei handelte es sich

⁴⁸ ANP = Adenylyl-imidodiphosphat: Analogon des ATP (Adenosintri-phosphat), wobei der Sauerstoff, welcher das β - mit dem γ -Phosphat verbindet, durch ein Stickstoffatom ersetzt ist.

⁴⁹ AST-487 und VX-680 kommen versehentlich doppelt vor. Es handelt sich demnach um 235 echt unterschiedliche Aktive.

um 820 Strukturen. Im Gegensatz zur **Lig. a** und **b** sind hier auch die Strukturen der Patentschrift WO 2016/142310 eingegangen. Bei der Ligandenvorbereitung mit LIGPREP wurden bewusst keine geladenen Verbindungen erzeugt (s. Kap. 7.11.1, Tab. 7-7, „820 DLK-Inh. aus (P)“), um danach ein hierarchisches Clustering (vgl. Kap. 4.7.2) von eindeutig festgelegten, echt unterschiedlichen Strukturen – auf Basis ihrer linearen binären Fingerprints (s. Kap. 7.11.2, S. 284) – durchzuführen. Die daraus hervorgegangenen Clusterrepräsentanten sind bezüglich ihrer Fingerprints möglichst unähnlich zueinander und wurden für die **Lig. c** als Aktive ausgewählt. Alle Liganden aus den in dieser Arbeit untersuchten Röntgenkristallstrukturen **5CEO**, **5CEP** und **5CEQ** und alle noch nicht vorhandenen unspezifischen DLK-Inhibitoren wurden ergänzt. Dies ergab eine Liste von 216 Aktiven (s. Anhang B, S. 343 ff, Tab. B 1). 15.849 Decoys wurden basierend auf diesen Strukturen generiert. Die Aktiven dieser **Lig. c** wurde als Referenzdatensatz für Aktive im zentralen virtuellen Screening der ZINC-Strukturen zum Monitoring der Anreicherungen verwendet. Desweiteren wurde mit den Aktiven der **Lig. c** mit bekannten K_i - bzw. K_d -Werten die Korrelation zwischen Score- und Aktivitätswerten überprüft (s. Kap. 5.4.2). Diese Liste wurde als Testdatensatz zur Untersuchung der Anreicherung von Aktiven und der Aktivität-Score-Korrelation nach Anwendung verschiedener Scoringmethoden und -kombinationen zur Optimierung des Postprocessings verwendet (s. Kap. 5.5.2).

5.2.3 Filterprozess der ZINC-Clean-Drug-like-Datenbank

Wie in Kap. 4.3, S. 55 erläutert, wurde als Ausgangsdatenbank zur Vorbereitung der Substanzbibliothek für das virtuelle Screening das *Clean-Druglike-Subset* der ZINC-Datenbank ausgewählt, welches 5.142.887 Verbindungen enthielt. Mithilfe von MONA wurde diese Menge auf Basis eigener Filterkriterien weiter reduziert.

Tab. 5-15 – Eigene Filterkriterien zur weiteren Eingrenzung des Ligandendatensatzes für das virtuelle Screening

Molekulare Eigenschaft	Grenzen
#HBA	1 - 10
#HBD	1 - 5
#Heteroatome	2 - 100
#RB	0 - 8
#Ringe	1 - 100
Berechneter Log P	-4 bis +6
MW	150 - 600
Gesamtladung	0 - 3
#Siliciumatome	0

Die Besonderheiten in diesen Filterkriterien (s. Tab. 5-15) liegen zum einen in der Vorgabe, dass mindestens ein Wasserstoffbrückendonator (HBD) und -akzeptor (HBA) vorhanden sein sollen. Dies ist der Tatsache geschuldet, dass die DLK-Inhibitoren über diese Funktionen an die Hinge-Region binden (s. Kap. 5.1.1, S. 125, Abb. 5-5). Auffällig ist zudem die hohe Anzahl an Heteroatomen in DLK-Inhibitoren. Dies liegt auch daran, dass die HBD und HBA zumeist Heteroatome sind. Daher sollten die Liganden in der Screening-Datenbank mindestens zwei Heteroatome enthalten.

Darüber hinaus enthält der Großteil an bekannten Kinaseinhibitoren (aromatische) Ringe [88]. Unter den literaturbekannten DLK-Inhibitoren gibt es keine Struktur ohne einen Ring. Daher sollten die Liganden in der Ausgangsdatenbank mindestens einen Ring enthalten. Nach Anwendung dieser eigenen Filterkriterien auf die ZINC-Teilmenge blieben 3.912.599 Strukturen übrig, die einer Ligandenpräparation mit LIGPREP unterzogen werden sollten.

Die Auswirkung des Filters auf die Verteilung der Eigenschaften in der Substanzbibliothek wird im Folgenden im Vergleich zu den Eigenschaften der 835 DLK-Inhibitoren gezeigt.

Die Verteilung des Molekulargewichts ist durch Anwendung des Filters nahezu gleichgeblieben (s. Abb. 5-45). Der Schwerpunkt der Verbindungen des gefilterten Datensatzes (s. Abb. 5-45, graue Balken) liegt ebenso wie für den ungefilterten Datensatz zwischen 325 und 350 g/mol, während die bekannten, DLK-hemmenden Verbindungen (s. Abb. 5-45, rote Balken) im Mittel größer sind. Hier liegt der Schwerpunkt etwa zwischen

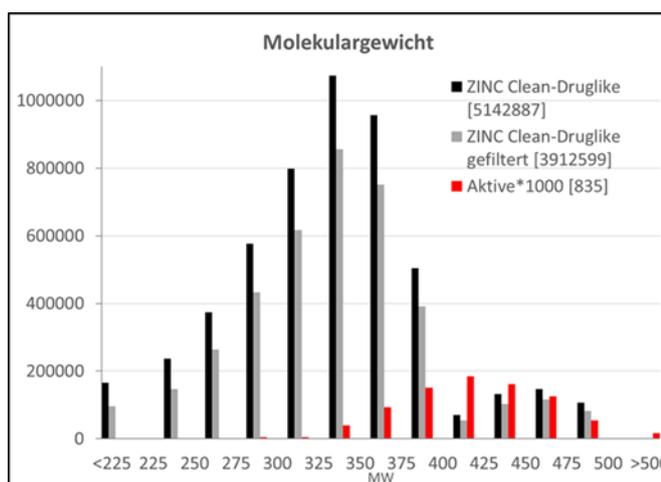


Abb. 5-45 – Verteilung des Molekulargewichts vor (schwarz) und nach (grau) Anwendung der Filterkriterien auf den ZINC-Datensatz. Die 835 bekannten DLK-Inhibitoren (rot) sind 1000x vergrößert dargestellt.

400 und 450 g/mol mit einem maximalen Molekulargewicht von 633,67 g/mol und einem Minimum von 292,34 g/mol. Insbesondere einige Strukturen mit abweichendem Grundgerüst von den sehr selektiven DLK-Inhibitoren (vgl. Kap. 5.2.1) haben das größte Molekulargewicht, dazu zählen in absteigender Reihenfolge: XL-880 (633,67 g/mol), TAE-684 (616,23 g/mol), CEP-1347 (615,77 g/mol), HKI-272 (558,06 g/mol), SKI-606 (Bosutinib, 532,47 g/mol), AST-487 (531,58 g/mol) und TG-101348 (525,69 g/mol). Erst dann folgen einige der sehr selektiven DLK-Inhibitoren. Das im Durchschnitt geringere Molekulargewicht des gefilterten ZINC-Datensatzes im Vergleich zu den bekannten DLK-Inhibitoren trägt zum einen zu einer tendenziell besseren peroralen Bioverfügbarkeit bei, darüber hinaus weisen kleinere Moleküle ein höheres Potenzial zur nachträglichen Optimierung auf (*lead-like*-Charakter, vgl. Kap. 4.3, S. 53), da durch das Anhängen funktioneller Gruppen zur Verbesserung der Interaktion mit dem Target oder zur Verbesserung pharmakokinetischer Eigenschaften das Molekulargewicht mit großer Wahrscheinlichkeit noch unter 500 g/mol gehalten werden kann.

Die logP-Wert-Verteilung der gefilterten ZINC-Datenbank gleicht etwa der Verteilung der Aktiven (vgl. Abb. 5-46). Für eine gute orale Bioverfügbarkeit ist ein logP-Wert von 5 als oberer Grenzwert definiert (s. Kap. 4.3, S. 55, Tab. 4-2). Dies wird von 99,4% des gefilterten ZINC-Datensatzes erfüllt.

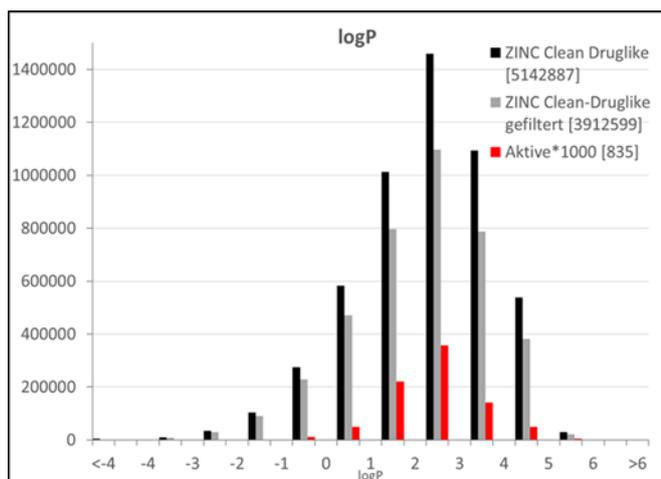


Abb. 5-46 – Verteilung des logP vor (schwarz) und nach (grau) Anwendung der Filterkriterien auf den ZINC-Datensatz. Die 835 bekannten DLK-Inhibitoren (rot) sind 1000x vergrößert dargestellt.

Die Verteilung der Anzahl an Wasserstoffbrückendonoren (s. Abb. 5-47) hat sich durch Anwendung des Filters im Vergleich zu den anderen molekularen Eigenschaften am drastischsten verändert, denn durch die Filterregeln wurde festgelegt, dass mindestens ein Donor vorhanden sein muss – im Hinblick auf die Interaktion mit der Hinge-Region, für die i.d.R. zwei Donor-Funktionen erforderlich sind (vgl. Abb. 5-33, S. 148). Durch diese Vorgabe wurden mehr als eine Million Verbindungen aus dem ursprünglichen Datensatz entfernt. Unter den bekannten DLK-Inhibitoren gibt es drei Strukturen (45-47), die keinen typischen Donor besitzen (s. Abb. 5-48). Es ist anzunehmen, dass die Wasserstoffbrückenbindungen von der Hinge-Region zu den zur Hinge-Region ausgerichteten Wasserstoffatomen des Aromaten („Core 1“, vgl. Abb. 5-33, S. 148) ausgebildet werden. Diese werden nicht als Donoren im herkömmlichen Sinne bezeichnet, da eine Do-

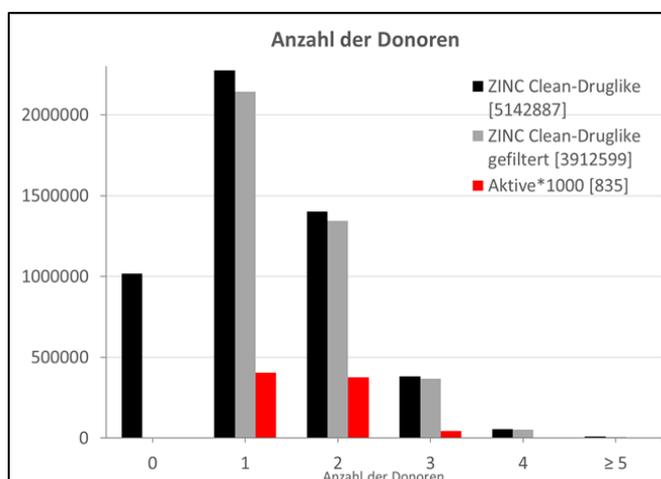


Abb. 5-47 – Verteilung der Anzahl an Donoren vor (schwarz) und nach (grau) Anwendung der Filterkriterien auf den ZINC-Datensatz. Die 835 bekannten DLK-Inhibitoren (rot) sind 1000x vergrößert dargestellt.

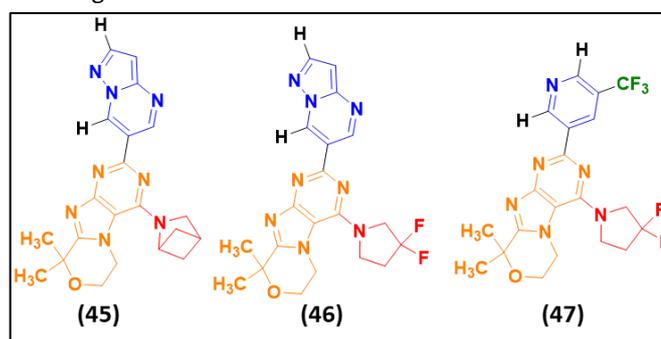


Abb. 5-48 – Strukturen ohne Donor unter den bekannten DLK-Inhibitoren. Die Donorfunktion wird vermutlich durch die hier schwarz dargestellten Wasserstoffatome am aromatischen Ring (Core 1) übernommen.

Region ausgerichteteten Wasserstoffatomen des Aromaten („Core 1“, vgl. Abb. 5-33, S. 148) ausgebildet werden. Diese werden nicht als Donoren im herkömmlichen Sinne bezeichnet, da eine Do-

norfunktion üblicherweise von einem an stickstoff- oder sauerstoffgebundenen Wasserstoffatom besteht; hier fehlt das Heteroatom als Bindungspartner. Auch unter den DLK-Inhibitoren mit nur einer Donorfunktion ist meist ein aromatisches Wasserstoffatom als weiterer „Donor“ an der Wasserstoffbrückenbindung beteiligt. Um diese Möglichkeit der Interaktion zu berücksichtigen, wurden Verbindungen mit nur einer offiziellen Donorfunktion im Datensatz belassen. Schließlich handelt es sich hierbei um 404 DLK-Inhibitoren (48 %).

Für die Interaktion mit der Hinge-Region ist mindestens eine Akzeptorfunktion erforderlich. Die bekannten DLK-Inhibitoren besitzen im Mittel 4-5 Akzeptoren (s. Abb. 5-49). Es sind immer mindestens zwei Akzeptoren vorhanden. Dennoch wurde für die Filterung der ZINC-Datenbank auch das Vorhandensein nur einer einzigen Akzeptorfunktion zugelassen, da dies ein ausreichendes Kriterium für die Hinge-Interaktion ist. Nur die Verbindungen ohne Akzeptorfunktion wurden entfernt. Dabei handelte es sich lediglich um 40.508 Strukturen.

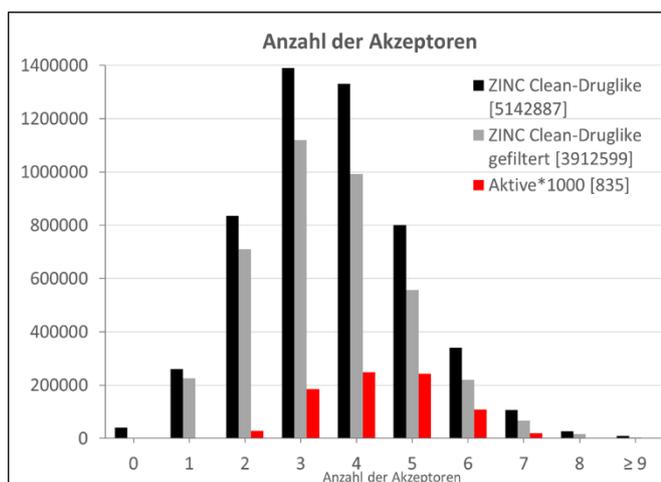


Abb. 5-49 – Verteilung der Anzahl an Akzeptoren vor (schwarz) und nach (grau) Anwendung der Filterregeln. Die 835 bekannten DLK-Inhibitoren (rot) sind 1000x vergrößert dargestellt.

Strukturen.

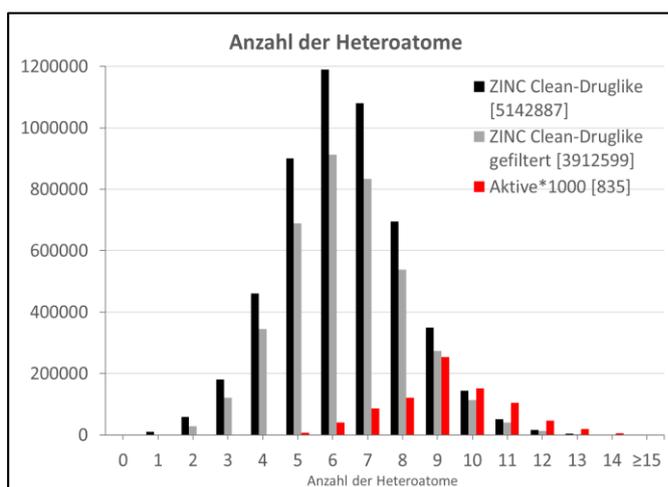


Abb. 5-50 – Verteilung der Anzahl an Heteroatomen vor (schwarz) und nach (grau) Anwendung der Filterregeln. Die 835 bekannten DLK-Inhibitoren (rot) sind 1000x vergrößert dargestellt.

wasserstoffatom besitzen, welches als Donor fungiert.

Die bekannten DLK-Inhibitoren enthalten im Mittel eine höhere Anzahl an Heteroatomen als die ZINC-Verbindungen (s. Abb. 5-50). Mit der Vorgabe, dass mindestens ein Akzeptor und ein Donor vorhanden sein müssen, befinden sich im Umkehrschluss mindestens zwei Heteroatome in jedem Molekül. Verbindungen mit zwei Heteroatomen könnten eine Akzeptor- und eine Donorfunktion beinhalten sowie ein aromatisches Was-

Wie schon in Kapitel 4.3 (S. 53) erläutert, geht eine zu hohe Anzahl an rotierbaren Bindungen (#RB) mit einem höheren Molekulargewicht und somit einer geringeren Bioverfügbarkeit einher. Darüber hinaus erhöht sich durch die Erweiterung des Torsionsraums der Rechenaufwand beim Docking. Da ohnehin ein geringeres Molekulargewicht für eine mögliche Weiterentwicklung von Hits erwünscht war und die bekannten DLK-Inhibitoren mit tendenziell höherem Molekulargewicht maximal sieben RB enthielten, wurde #RB auf acht begrenzt. Dadurch wurden 35.836 Verbindungen entfernt (s. Abb. 5-51).

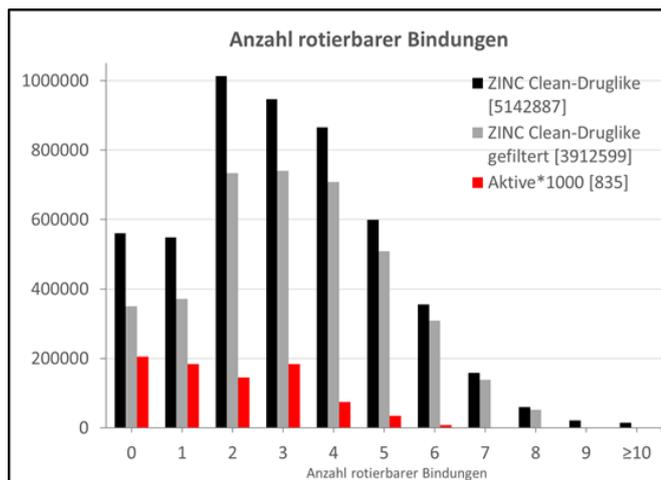


Abb. 5-51 – Verteilung der Anzahl an rotierbaren Bindungen (RB) vor (schwarz) und nach (grau) Anwendung der Filterregeln. Die 835 bekannten DLK-Inhibitoren (rot) sind 1000x vergrößert dargestellt.

Die bekannten DLK-Inhibitoren besitzen bis zu vier Stereozentren (s. Abb. 5-52). Nur vergleichsweise wenige Strukturen der ZINC-Datenbank besitzen fünf oder mehr Stereozentren. Aufgrund dieser Beobachtung wurde die Zahl der Stereozentren nicht in den Filterregeln berücksichtigt.

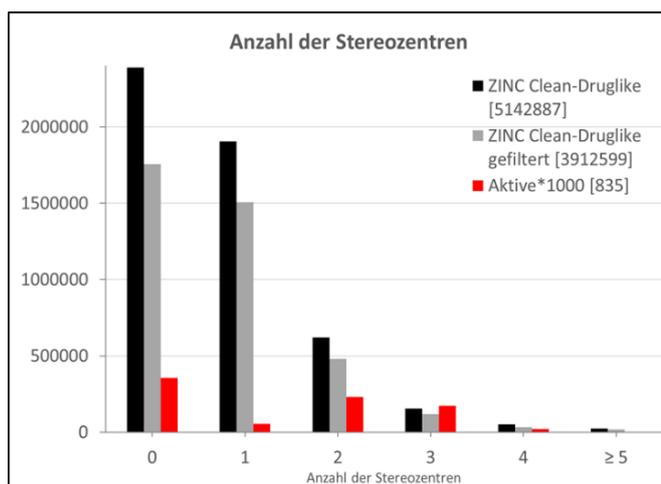


Abb. 5-52 – Verteilung der Anzahl an Stereozentren vor (schwarz) und nach (grau) Anwendung der Filterregeln. Die 835 bekannten DLK-Inhibitoren (rot) sind 1000x vergrößert dargestellt.

Die Anzahl der Ringe in den bekannten DLK-Inhibitoren beträgt mindestens zwei (s. Abb. 5-53, oben). Nur vier der Aktiven enthalten lediglich zwei Ringstrukturen, welche darüber hinaus identisch sind. Die staurosporinähnlichen Strukturen (s. S. 152, Abb. 5-41) zählen zu den Verbindungen, welche eine Vielzahl von Ringen enthalten.

Strukturen ohne Ringstruktur wurden aus der ZINC-Datenbank entfernt, da diese für gewöhnlich nicht mit einer Kinaseinhibition in Verbindung gebracht werden. Für die Anzahl der aromatischen Ringe (s. Abb. 5-53, unten) wurde keine weitere Filterregel getroffen, obwohl hier die Eliminierung aller Strukturen ohne oder mit nur einem aromatischen Ring in Betracht gezogen wurde, da es in diesen

beiden Kategorien keine bekannten DLK-Inhibitoren gibt. Die Diversität der ZINC-Strukturen sollte aber in dem Fall nicht so weit eingeschränkt werden.

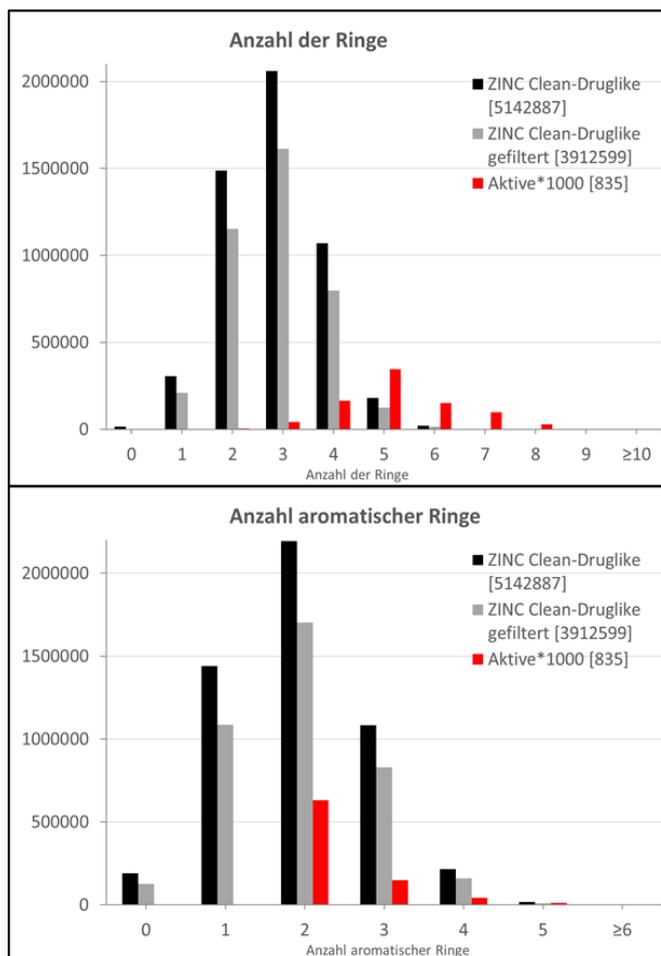


Abb. 5-53 – Verteilung der Anzahl der Ringe (oben) bzw. der aromatischen Ringe (unten) vor (schwarz) und nach (grau) Anwendung der Filterregeln. Die 835 bekannten DLK-Inhibitoren (rot) sind 1000x vergrößert dargestellt.

5.3 3D-QSAR

Mit DLK-Inhibitoren bekannter Aktivität (K_i - bzw. K_d -Wert) wurden im Rahmen dieser Arbeit 3D-QSAR-Modelle erstellt. Zur Beurteilung der Modelle wurden die statistischen Kennzahlen (s. Kap. 4.8.2) genutzt, um das qualitativ beste QSAR-Modell auszuwählen und dieses im Postprocessing zur Nachbewertung der Dockingposen als ligandbasierten Ansatz zu verwenden (s. Kap. 5.5.1-5.5.3).

5.3.1 Alignment durch Docking

Die wichtigste Voraussetzung zur Erstellung von 3D-QSAR-Modellen ist die Übereinanderlagerung der Liganden im Raum, idealerweise so, wie sie tatsächlich in der biologisch aktiven Form am Wirkort vorliegen (vgl. Kap. 4.8.1). Die aus der Ligandenpräparation bekannter DLK-Inhibitoren (s. Kap. 5.2.2, Liganden für QSAR) erhaltenen 1295 unterschiedlichen Strukturen (*Multiple States*) von 847 DLK-Inhibitoren wurden mit dem GLIDE-Modul der SCHRÖDINGER-Software in die vorbereitete **5CEO**- und **5CEQ-PL**-Struktur sowie in die drei aus den MD-Simulationen ausgewählten Proteinkonformationen (s. Kap. 5.1.3, S. 139, Tab. 5-8) mit dem SP (*Standard Precision*)-Modus im OPLS3-Kraftfeld gedockt (s. Kap. 7.5.2, Tab. 7-3, „Docking für QSAR-Alignment“). Höchstens fünf Posen pro Protonierungszustand eines Liganden wurden zugelassen. Damit die Posen nach dem Docking in die verschiedenen Proteinkonformationen deckungsgleich sind, wurden die Bindetaschen zuvor räumlich aufeinandergelegt (s. Kap. 7.5.1). Für das Docking wurde festgelegt, dass mindestens eine der drei Wasserstoffbrückenbindungen zur Hinge-Region (vgl. Kap. 5.2.1, S. 148, Abb. 5-33) ausgebildet werden muss (Constraints).

Mit vier der fünf Proteinkonformationen wurden jeweils mehr als 5000 Dockingposen erzeugt, während mit der **5CEN**-MD-Struktur 3699 Posen erzeugt wurden. Die Dockingposen (*Multiple States*) wurden für jede Proteinkonformation nach Dockingscore sortiert und die beste Pose (*Single State*⁵⁰) beibehalten (*Glide Ensemble Merge*). Diese Posen wurden auf ihre Korrektheit⁵¹ überprüft. Das Docking in die **5CEO**- und **5CEQ-PL**-Struktur ergab die größte Anzahl korrekter Posen und im Mittel die negativsten Dockingscores, d.h. die beste Bewertung der Posen (s. Tab. 5-16). Über alle Proteinkonformationen hinweg konnten 828 (97,76 %) der 847 Liganden in der erwarteten Pose gedockt werden (s. Tab. 5-16).

⁵⁰ *Single State* = ein definierter Protonierungs- bzw. Tautomerisierungszustand einer chemischen Struktur wird berücksichtigt; *Multiple States* = unterschiedliche Protonierungs- bzw. Tautomerisierungszustände einer sonst identischen chemischen Struktur werden berücksichtigt.

⁵¹ „Korrekte Pose“ wird definiert als die in Übereinstimmung mit der in Kap. 5.2.1 beschriebenen Ausrichtung der Liganden in der DLK-Bindetasche.

Tab. 5-16 – Ergebnisse des Dockings in die 5 Proteinkonformationen als Vorbereitung des Alignments für QSAR. #Posen gesamt: alle Zustände werden mitgezählt. # Posen single: Nur ein Zustand eines Liganden wird in der Zählung berücksichtigt. „unsp.“ = unspezifische DLK-Inhibitoren (ohne AST-487). \bar{x} DS = Mittelwert des Dockingscores über alle Posen.

Proteinkonformation	#Posen gesamt // #Posen single	#korrekte Pos- sen single	#korrekte Pos- sen single unsp.	\bar{x} DS	VK DS [%]
5CEO-PL	5714//829	788	6	-8,867	-12,086
5CEQ-PL	5715//831	797	5	-8,640	-11,136
5CEO Run 02 Frame 2000	5857//831	723	6	-7,825	-13,043
5CEQ Run 03 Frame 3000	5321//800	716	3	-7,738	-16,224
5CEN Run 02 Frame 2661	3699//633	418	2	-3,856	-24,103
Insgesamt	26306// 836	828	7	-7,653	-25,101

Von den 18 unspezifischen DLK-Inhibitoren konnten lediglich 7 in der erwarteten Position gedockt werden (s. Tab. 5-17). JNJ-770661 konnte nur in **5CEN Run 02** Frame 2661 in der erwarteten Pose gedockt werden.

Anhand der visuellen Betrachtung im Vergleich zur Pose des **5CEO**-Liganden GNE-3511 in der Bindetasche

der Röntgenkristallstruktur **5CEO** und anhand der Docking- und GlideScores wurde die bestbewertete Pose mit dem korrekten Bindungsmodus für jeden Liganden beibehalten. Protonierungen am basischen, gesättigten N-Heterozyklus (Piperazin, Piperidin, vor allem zu finden am Substituenten R3) wurden gegenüber einer neutralen Struktur bevorzugt ausgewählt, wenn beide Posen korrekt waren. Alle neutralen Strukturen, die eine Möglichkeit zur Protonierung am gesättigten N-Heterozyklus (Piperazin, Piperidin) hatten, wurden nachträglich manuell im MAESTRO WORKSPACE mit dem 3D-BUILDER protoniert und unter Selektion aller Atome minimiert. Struktu-

ren mit Protonierungen an den aromatischen Stickstoffen (z.B. Pyridin, Pyrimidin) wurden nicht übernommen oder ggf. manuell deprotoniert und minimiert. Diese „einfachen Kraftfeldminimierungen“ mit dem 3D BUILDER finden unter Anwendung von Constraints statt, sodass die Pose immer noch beibehalten wird (s. Abb. 5-54). Die Minimierung wird beendet, wenn die maximale Abstandsänderung zwischen den Atomen weniger als 0,1 Å beträgt.

Tab. 5-17 – Unspezifische DLK-Inhibitoren mit korrekter Pose

Verbindung	pK _d (DLK)
JNJ-770661	5,284
CHIR-258/TKI-258	5,678
JNJ-28312141	6,482
VX-680	6,721
Crizotinib	6,770
Sunitinib	7,000
KW-2449	7,174

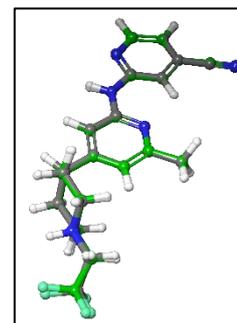


Abb. 5-54 – Pose vor (grau) und nach (grün) manueller Protonierung und Minimierung

Die Posen aus dem Frame 2661 der zweiten **5CEN**-MD-Trajektorie wurden allerdings nicht weiterverwendet, weil sich der Gatekeeper Met190 im Vergleich zur **5CEO-PL**-Struktur näher an der Bindetasche befand und die Posen somit nach vorne versetzt auftraten (s. Abb. 5-55 A). Die Liganden sitzen somit nicht so tief in der Ta-

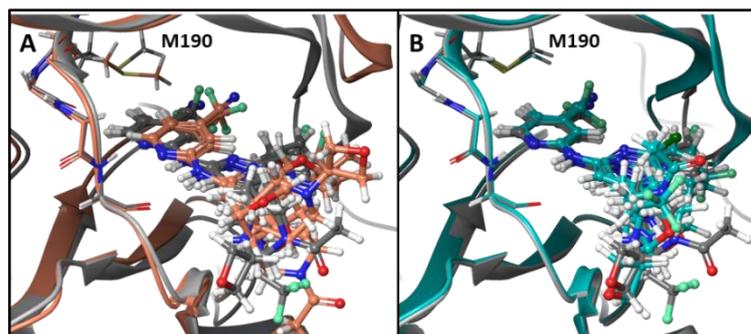


Abb. 5-55 – **A** – Vergleich von Dockingposen in **5CEO-PL** (grau) und in **5CEN Run 02** Frame 2661 (orange). Letztere ragen weiter aus der Bindetasche heraus und werden nicht mit den **5CEO**-Dockingposen zur Deckung gebracht. **B** – Vergleich von Dockingposen in der **5CEO-PL** (grau) und in **5CEQ-PL** (cyan). Die Posen werden gut zur Deckung gebracht (Alignment möglich).

sche und interagieren weniger ausgeprägt über die kinasetypischen Interaktionsmuster mit der Zielstruktur. Dies ist auch an den im Mittel deutlich weniger negativen Dockingscores in der **5CEN**-Konformation zu erkennen (s. Tab. 5-16).

Für ein qualitativ hochwertiges QSAR-Modell ist eine strukturelle Variation sowie eine breite Verteilung der pK_i -Werte im Ligandenset von Vorteil. Diese beiden Kriterien sind vor allem für das Trainingsset von großer Bedeutung. Daher wurde Wert darauf gelegt, dass auch einige der wenigen unspezifischen DLK-Inhibitoren (s. Kap. 5.2.1, S. 152 ff.) für die Erstellung der QSAR-Modelle ausgewählt werden, da sie zum einen strukturell von den selektiven DLK-Inhibitoren aus den Veröffentlichungen von PATEL ET AL. und den Patentschriften (s. Kap. 5.2.1, S. 148 ff.) abweichen und zum anderen tendenziell eine schwächere DLK-Hemmaktivität besitzen, von denen es unter den selektiven DLK-Inhibitoren nur wenige gibt.

Es wurden daher zusätzliche Induced-Fit-Docking-Experimente (s. Kap. 7.5.2, Tab. 7-3, „*Induced-Fit-Docking*“) mit den 18 unspezifischen DLK-Inhibitoren durchgeführt. Hier wurden lediglich die Proteinkonformationen **5CEN** Frame 2661 *Run 02* und die **5CEO-PL**-Struktur als Targets verwendet. Es wurden zwei unterschiedliche Einstellungen für das Trimmen von Seitenketten verwendet, sodass insgesamt vier Dockingexperimente durchgeführt wurden. Zum einen wurden die zu trimmenden Seitenketten durch die Software automatisch auf Basis des B-Faktors ausgewählt (*default*). Zum anderen wurden manuell fünf Aminosäuren der Bindetasche zum Trimmen der Seitenketten ausgewählt: Met190, Lys152, Asp254 und Ser253 für die **5CEO**-MD-Ausgangsstruktur und zusätzlich Glu200 für **5CEN Run 02** Frame 2661. Im Ergebnis konnten zwei weitere korrekte Posen erhalten werden:

JNJ-7706621 [5CEN Frame 2661 Run 02 (default)]

SU-14813 [5CEO-PL-Struktur (manuell)]

Unter den restlichen Induced-Fit-Dockingergebnissen konnten keine weiteren Posen gefunden werden, die korrekt waren oder nicht schon korrekt im SP-Modus gedockt waren. Mithilfe des Induced-Fit-Dockings konnte die Position des JNJ-7706621 im Vergleich zur SP-Dockingpose verbessert werden (s. Abb. 5-56).

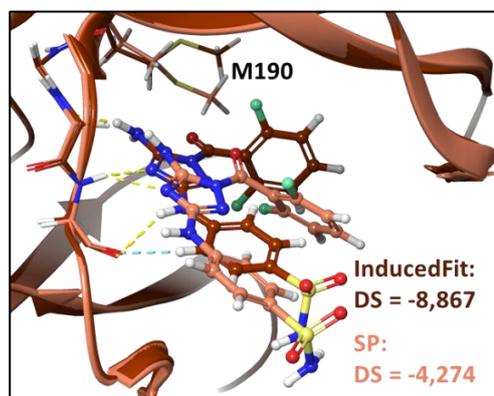


Abb. 5-56 – Vergleich der SP-Pose (hellorange) und der Induced-Fit-Pose (braun) des JNJ-7706621. Die Konformation des Met190 beeinflusst das Ausmaß des Hineinragens in die Bindetasche

Auswahl von Ligandenkonformeren für QSAR

827 der 847 DLK-Inhibitoren⁵² wurden mit dem Ziel einer möglichst gleichmäßigen pK_i - bzw. pK_d -Verteilung in drei Ligandendatensätze aufgeteilt. Zu diesem Zweck wurden sie zunächst in 10 Klassen hinsichtlich ihrer DLK-hemmenden Aktivität kategorisiert (s. Abb. 5-57). Nur sehr wenige Liganden besitzen eine niedrige ($pK_i = 5-6$) bzw. sehr hohe ($pK_i = 10-11$) DLK-Aktivität. Die DLK-Inhibitoren mit pK_i - bzw. pK_d -Werten zwischen 5,1 und 6,1 (Bin 1 und 2) sowie zwischen 9,1 und 10,1 (Bin 9 und 10) wurden für alle drei Ligandendatensätze ausgewählt. Aus den übrigen 6 Bins wurden zufällig jeweils etwa 25 Liganden ausgewählt. Die Anzahl an Liganden mit pK_i - bzw. pK_d -Werten zwischen 7 und 9 konnte so reduziert werden (s. Abb. 5-58).

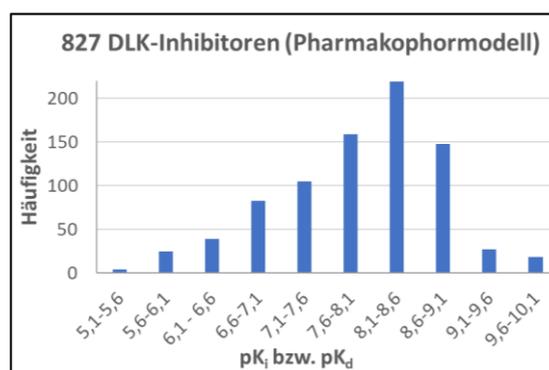


Abb. 5-57 – pK_i - bzw. pK_d -Verteilung der 827 DLK-Inhibitoren

⁵² Diese Auswahl erfolgte aufgrund eines Nebenprojekts mit nur 827 Strukturen und kann als zufällig angesehen werden.

Zwei der Ligandendatensätze (im Folgenden: Selektion 1 und 3) bestehen aus 219 Liganden, während Selektion 2 aus 218 Liganden besteht. Die Liganden wurden aus den zuvor aufbereiteten Dockingposen (s. S. 163 ff.) ausgewählt. Hier wurden bevorzugt – wenn vorhanden – die Posen aus der **5CEO-PL**-Struktur ausgewählt, da die Wahrscheinlichkeit eines passenden Alignments größer ist, wenn die Liganden aus der gleichen Proteinkonformation stammen. Die hier nicht gefundenen Posen wurden aus den übrigen drei Proteinkonformationen (alle bis auf **5CEN** Run02 Frame 2661) ergänzt.

Fehlende Liganden – solche, welche nicht in korrekter Pose gedockt werden konnten – wurden mit Ligandenkonformeren aus einem Pharmakophormodell (PK-Modell) ergänzt (s. Kap. 7.7, S. 272, PK-Modelle erstellt von Dr. Thomas Lemcke). Dies betraf die in Tab. 5-18 aufgeführten Liganden.

Die beiden Staurosporin-Verbindungen Staurosporin und CEP-701 wurden anders als erwartet durch das PK-Modell ausgerichtet (s. Abb. 5-59, B+D). Normalerweise interagiert der γ -Lactamring wie in Abb. 5-59 A und C dargestellt mit der Hinge-Region. Mit dem PK-Modell erhielt man nur die in Abb. 5-59 B und D dargestellten Ausrichtungen, die zwar als nicht optimal bewertet wurden, allerdings mangels Alternativen dennoch für die QSAR-Alignments verwendet wurden.

Im Gegensatz zu den Staurosporinderivaten passt die PK-Modell-Ausrichtung der selektiven DLK-Inhibitoren (s. Tab. 5-18, folgende Seite) wie eine Dockingpose in die Bindetasche des **5CEO-PL** (s. Abb. 5-59 E). In dieser

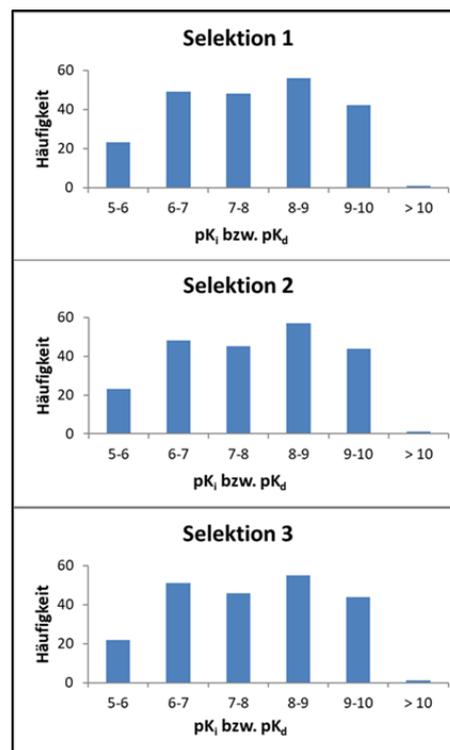


Abb. 5-58 – pK_i- bzw. pK_d-Verteilung der erstellten Ligandendatensätze für die Erstellung der QSAR-Modelle

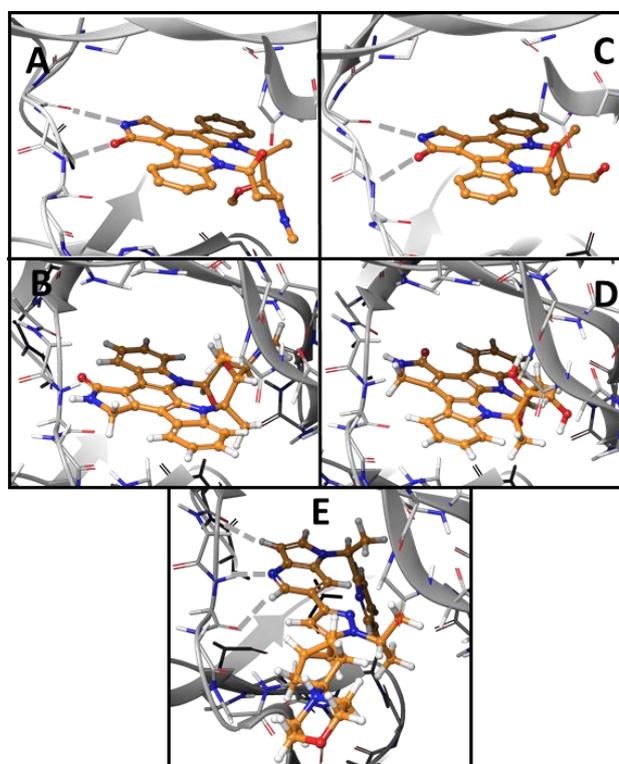


Abb. 5-59 – A, C – Bindungsmodus des Staurosporins (A) in CSK und des CEP-701 (C) in PRK1 in einer Röntgenkristallstruktur (s. Tab. 7-6, S. 281). B, D – PK-Modell-Ausrichtung des Staurosporins (B) und des CEP-701 (D) dargestellt in der **5CEO-PL**-Bindetasche. E – PK-Modell-Ausrichtung der Verbindung Nr. 24 aus LYSSIKATOS ET AL (2014) [174]) dargestellt in der **5CEO-PL**-Bindetasche.

Ausrichtung können die Verbindungen die erwarteten Wasserstoffbrückenbindungen mit der Hinge-Region eingehen.

Die Pose des in allen Selektionen verwendeten JNJ-7706621 ($pK_d = 5,284$) stammt aus dem zusätzlich ausgeführten Induced-Fit-Docking. Alle übrigen Strukturen, darunter KW-2449 und CHIR-258/TKI-258 von den unspezifischen DLK-Inhibitoren, stammen aus den SP-Dockingergebnissen.

Beispielhaft ist das Alignment der 218 DLK-Inhibitoren aus Selektion 2 in Abb. 5-60 zusammen mit der **5CEO-PL**-Bindetasche dargestellt. Der Hinge-interagierende Teil der Liganden liegt sehr geordnet vor, während der lösungsmittelorientierte Molekülteil (R3) sterisch weniger gehindert wird und eine Vielzahl von Positionen einnehmen kann. Ebenso ist eine gewisse räumliche Variabilität des Substituenten R2 zu sehen, welcher mit dem P-Loop interagiert.

Tab. 5-18 – Ligandenkonformere aus dem Pharmakophor-Alignment, welche in die QSAR-Datensätze aufgenommen wurden. Sel. = Selektion. [174] = Patent WO 2015/091889 (LYSSIKATOS ET AL. 2014)

Ligand	$pK_i // pK_d$	Sel.
Staurosporin	5,824	1, 2, 3
CEP-701	6,076	1, 2, 3
Nr. 24 aus [174]	9,469	1, 2, 3
Nr. 26 aus [174]	7,121	3
Nr. 27 aus [174]	9,854	1, 2, 3
Nr. 31 aus [174]	7,100	1
Nr. 33 aus [174]	6,910	2
Nr. 34 aus [174]	8,378	2

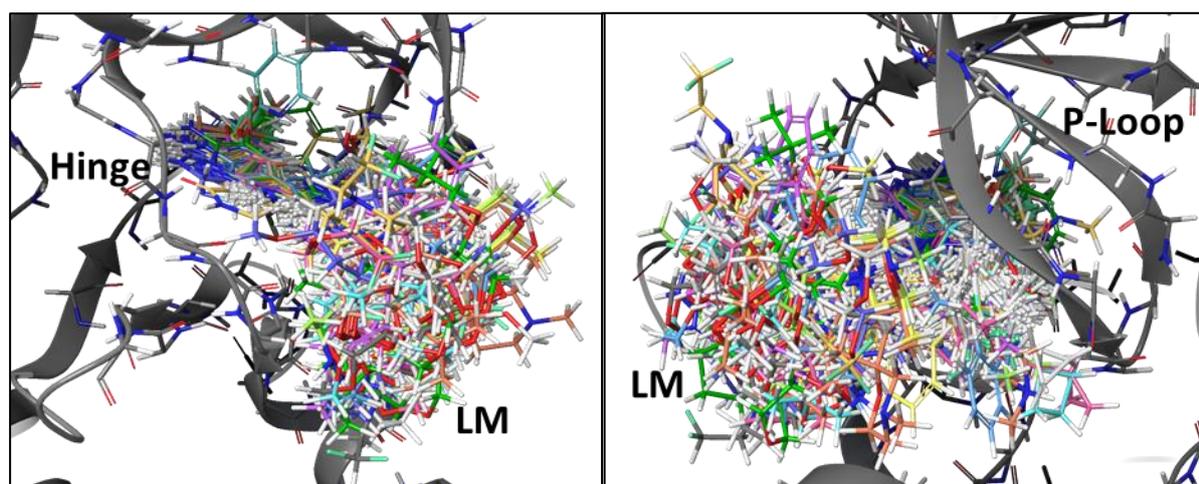


Abb. 5-60 – Darstellung des Alignments von 218 DLK-Inhibitoren (Selektion 2), welches vorrangig aus Dockingposen besteht (hier dargestellt mit der Bindetasche der **5CEO-PL**-Struktur), aus 2 unterschiedlichen Blickwinkeln. LM = Lösungsmittlexponierter Bereich.

5.3.2 Test- und Trainingsdatensatz und Berechnung der QSAR-Modelle

Die drei Selektionen wurden jeweils zufällig in Test- und Trainingsdatensätze aufgeteilt. 66 %, 70 % bzw. 80 % des jeweiligen Datensatzes wurde als Trainingsdatensatz verwendet, während die übrigen 34 %, 30 % bzw. 20 % entsprechend als Testdatensatz fungierten.

Pro Selektion und festem Trainingsdatensatzanteil wurden drei feldbasierte QSAR-Modelle (3 unterschiedliche *random seeds*) mit der Software SCHRÖDINGER generiert (s. Kap. 7.8), d.h. die Zusammensetzung des Trainingsdatensatzes wurde jeweils drei Mal variiert (im Folgenden: *Split 1-3*). Die maximale Anzahl an Komponenten (#*k*) wurde mit 10 festgelegt, sodass 10 Modelle für einen definierten Trainingsdatensatz erzeugt wurden. Es wurde eine LOO-Kreuzvalidierung durchgeführt (*leave-one-out*, s. Kap. 4.8.2, S. 102). Das Modell (#*k*) mit dem höchsten R^2_{cv} wurde ausgewählt.

Wie in der Übersicht (s. Tab. 5-19) dargestellt, wurden unterschiedliche Kombinationen der Felder für die Berechnung der QSAR-Modelle verwendet. Bei der GAUSSIAN-Methode standen 5 bzw. im EXTENDED GAUSSIAN 6 Felder zur Auswahl:

Sterisch (S)

Elektrostatisch (E)

Hydrophob (H)

Wasserstoffbrückenakzeptor (A)

Wasserstoffbrückendonator (D)

Aromatische Ringe (R, nur in der EXTENDED GAUSSIAN-Einstellung)

Zusätzlich zu den CoMSIA-ähnlichen GAUSSIAN-Modellen wurden mit der 80 %-Trainingsdatensatz-Einstellung auch CoMFA-ähnliche Modelle (FORCE FIELD-Modelle) für die Zweifeldkombination und für je ein Feld berechnet.

Tab. 5-19 – Übersicht über die erstellten QSAR-Modelle mit den drei Selektionen. **S** – sterisch, **E** – elektrostatisch, **H** – hydrophob, **A** – Wasserstoffbrückenakzeptoren, **D** – Wasserstoffbrückendonoren, **R** – aromatische Ringe. Kombi. = Kombinationsmöglichkeiten.

	66 % Trainingsset	70 % Trainingsset	80 % Trainingsset
GAUSSIAN	5 Felder (SEHAD)	5 Felder (SEHAD)	6 Felder (SEHADR)
	4 Felder (5 Kombi.)	4 Felder (5 Kombi.)	5 Felder (SEHAD)
	3 Felder (10 Kombi.)		4 Felder (5 Kombi.)
	2 Felder (10 Kombi.)		
	1 Feld (S, E, H, A, D)		
FORCE FIELD			2 Felder (SE)
			1 Feld (S, E)

5.3.3 Ergebnisse und Auswahl des besten QSAR-Modells

Die Modelle wurden nach den Kriterien $R^2_{cv} > 0,5$ und $Q^2 > 0,6$ beurteilt (s. Kap. 4.8.2). Drei der auf Basis der ersten Selektion für den Trainingsdatensatz von 66 % berechneten GAUSSIAN-Modelle erfüllen diese beiden Kriterien (s. Tab. 5-20). Sie wurden in Split 3 mit der 4-Feld-Kombination SEHD ($R^2_{cv} = 0,593$, $Q^2 = 0,628$) sowie mit den 3-Feld-Kombinationen SEH ($R^2_{cv} = 0,587$, $Q^2 = 0,625$) und SHD ($R^2_{cv} = 0,543$, $Q^2 = 0,615$) berechnet.

Tab. 5-20 – Statistische Parameter der GAUSSIAN-Modelle der Selektion 1 für einen Trainingsdatensatz von 66 % und drei Randomisierungen (Split 1-3). Modelle mit den Kriterien $R^2_{cv} > 0,5$ und $Q^2 > 0,6$ sind rot eingerahmt.

Selektion 1, Gaussian, 66 % Trainingsdatensatz												
Felder	Split 1				Split 2				Split 3			
	#k	R^2	R^2_{cv}	Q^2	#k	R^2	R^2_{cv}	Q^2	#k	R^2	R^2_{cv}	Q^2
SEHAD	6	0,836	0,634	0,493	3	0,736	0,608	0,477	4	0,763	0,603	0,589
SEHA	5	0,827	0,639	0,478	3	0,742	0,615	0,477	4	0,764	0,619	0,585
SEHD	5	0,803	0,636	0,481	3	0,719	0,603	0,468	4	0,741	0,593	0,628
SEAD	5	0,792	0,589	0,560	4	0,750	0,586	0,525	5	0,777	0,564	0,546
SHAD	6	0,823	0,624	0,457	3	0,722	0,596	0,479	4	0,751	0,585	0,592
EHAD	6	0,843	0,623	0,369	4	0,757	0,577	0,345	4	0,754	0,560	0,485
SEH	6	0,799	0,623	0,457	3	0,718	0,601	0,450	5	0,768	0,587	0,625
SEA	6	0,816	0,607	0,495	4	0,771	0,596	0,466	5	0,779	0,600	0,541
SED	5	0,753	0,579	0,532	4	0,704	0,562	0,513	3	0,658	0,525	0,531
SHA	6	0,840	0,629	0,403	3	0,721	0,594	0,472	4	0,748	0,597	0,579
SHD	5	0,760	0,602	0,439	4	0,711	0,588	0,451	4	0,700	0,543	0,615
SAD	5	0,755	0,527	0,443	4	0,694	0,527	0,528	3	0,682	0,520	0,470
EHA	6	0,842	0,607	0,334	4	0,774	0,577	0,351	3	0,728	0,568	0,492
EHD	5	0,797	0,613	0,361	5	0,762	0,575	0,280	5	0,756	0,524	0,561
EAD	4	0,707	0,500	0,326	4	0,743	0,559	0,216	3	0,675	0,511	0,386
HAD	6	0,816	0,609	0,258	4	0,719	0,542	0,342	4	0,726	0,528	0,490
SE	5	0,750	0,584	0,503	4	0,721	0,570	0,475	5	0,722	0,554	0,596
SH	5	0,762	0,585	0,360	4	0,732	0,559	0,464	1	0,693	0,537	0,594
SA	7	0,816	0,566	0,362	4	0,696	0,525	0,504	4	0,727	0,549	0,482
SD	4	0,664	0,499	0,439	4	0,641	0,508	0,497	3	0,601	0,480	0,459
EH	6	0,815	0,608	0,307	4	0,748	0,573	0,293	4	0,716	0,537	0,560
EA	5	0,750	0,493	0,327	4	0,738	0,552	0,216	3	0,671	0,506	0,336
ED	3	0,615	0,482	0,438	4	0,694	0,526	0,072	4	0,668	0,466	0,468
HA	7	0,808	0,569	0,310	4	0,735	0,526	0,349	4	0,722	0,534	0,428
HD	5	0,737	0,555	0,246	5	0,693	0,519	0,237	4	0,635	0,447	0,543
AD	5	0,641	0,372	-0,054	5	0,668	0,391	0,142	4	0,666	0,386	0,188
S	5	0,647	0,472	0,336	4	0,610	0,472	0,471	4	0,604	0,454	0,508
E	4	0,696	0,470	0,188	5	0,741	0,540	0,077	5	0,697	0,462	0,435
H	6	0,736	0,522	0,136	5	0,679	0,497	0,237	3	0,583	0,415	0,521
A	4	0,575	0,352	-0,090	3	0,550	0,363	0,098	3	0,565	0,360	0,125
D	5	0,368	0,210	0,089	4	0,404	0,297	0,121	2	0,275	0,175	0,262

Mit Selektion 2 und 66 % Trainingsdatensatz (s. Anhang D, S. 356, Tab. D 1) konnte ein GAUSSIAN-Modell generiert werden, welches den Kriterien genügt. Dies wurde mit fünf Feldern (SEHAD) berechnet und ergab einen R^2_{cv} von 0,565 und ein Q^2 von 0,602. Mit Selektion 3 konnte für einen Trainingsdatensatz von 66 % kein GAUSSIAN-Modell berechnet werden, welches den festgelegten Anforderungen entsprach (s. Anhang D, S. 356, Tab. D 2). Mit einem Anteil von 66 % als Trainingsdatensatz wurden demnach insgesamt vier GAUSSIAN-Modelle mit der geforderten Qualität erhalten. Die sterischen und hydrophoben Felder sind in all diesen Modellen zu finden, was vermuten lässt, dass diese in der Bindetasche von großer Bedeutung sind.

Mit einem Anteil von 70 % als Trainingsdatensatz und der Kombination aller 5 bzw. 4 Felder konnte kein GAUSSIAN-Modell mit den gewünschten R^2_{cv} - und Q^2 -Werten erhalten werden (s. Tab. 5-21). Das beste Modell wurde mit der Selektion 3, Split 3 mit 5 Feldern erzielt ($R^2_{cv} = 0,840$ und $Q^2 = 0,588$). Die Ergebnisse der Aktivitätsvorhersage für die Liganden mit diesem Modell sind in Tab. C 3 (Anhang C, S. 354 f.) dargestellt. Es handelt sich hierbei um das beste Modell für die Selektion 3, wenn man alle generierten QSAR-Modelle für Selektion 3 betrachtet.

Tab. 5-21 – Statistische Parameter der GAUSSIAN-Modelle der Selektionen 1-3 für einen Trainingsdatensatz von 70 % und drei Randomisierungen (Split 1, 2 und 3). Es wurde kein Modell mit den Kriterien $R^2_{cv} > 0,5$ und $Q^2 > 0,6$ erhalten. Das beste Modell (schwarz eingerahmt) wurde mit Selektion 3 in Split 3 mit dem 5-Felder-Modell erhalten.

Selektionen 1 – 3, Gaussian, 70 % Trainingsdatensatz												
Selektion 1	Split 1				Split 2				Split 3			
	Felder	#k	R^2	R^2_{cv}	Q^2	#k	R^2	R^2_{cv}	Q^2	#k	R^2	R^2_{cv}
SEHAD	4	0,748	0,590	0,546	4	0,773	0,636	0,363	5	0,796	0,603	0,552
SEHA	4	0,743	0,591	0,556	4	0,769	0,642	0,372	5	0,789	0,607	0,554
SEHD	5	0,761	0,574	0,547	3	0,722	0,619	0,311	4	0,747	0,590	0,553
SEAD	3	0,716	0,589	0,515	5	0,789	0,628	0,389	4	0,715	0,556	0,542
SHAD	4	0,748	0,581	0,538	4	0,759	0,628	0,328	3	0,710	0,576	0,510
EHAD	5	0,761	0,521	0,467	5	0,782	0,577	0,340	4	0,751	0,567	0,429
Selektion 2	Split 1				Split 2				Split 3			
Felder	#k	R^2	R^2_{cv}	Q^2	#k	R^2	R^2_{cv}	Q^2	#k	R^2	R^2_{cv}	Q^2
SEHAD	5	0,821	0,636	0,446	6	0,853	0,627	0,419	6	0,818	0,611	0,558
SEHA	5	0,817	0,632	0,494	6	0,835	0,643	0,405	5	0,790	0,614	0,517
SEHD	2	0,669	0,584	0,484	6	0,813	0,600	0,416	6	0,791	0,563	0,570
SEAD	4	0,767	0,599	0,443	3	0,703	0,590	0,410	4	0,730	0,576	0,433
SHAD	5	0,813	0,629	0,455	7	0,843	0,621	0,424	6	0,806	0,608	0,559
EHAD	6	0,861	0,668	0,389	6	0,833	0,606	0,278	6	0,838	0,615	0,499
Selektion 3	Split 1				Split 2				Split 3			
Felder	#k	R^2	R^2_{cv}	Q^2	#k	R^2	R^2_{cv}	Q^2	#k	R^2	R^2_{cv}	Q^2
SEHAD	3	0,746	0,625	0,527	3	0,767	0,663	0,438	6	0,840	0,604	0,588
SEHA	4	0,766	0,622	0,535	4	0,795	0,667	0,397	6	0,836	0,593	0,581
SEHD	3	0,714	0,601	0,522	3	0,696	0,588	0,464	4	0,764	0,596	0,583
SEAD	4	0,767	0,599	0,443	3	0,703	0,590	0,410	4	0,730	0,576	0,433
SHAD	5	0,813	0,629	0,455	7	0,843	0,621	0,424	5	0,784	0,607	0,561
EHAD	4	0,762	0,566	0,433	4	0,777	0,634	0,291	4	0,774	0,571	0,445

Mit der Auswahl von 80 % als Trainingsdatensatz wurden mit den drei Selektionen insgesamt 15 GAUSSIAN-Modelle mit $R^2_{cv} > 0,5$ und $Q^2 > 0,6$ generiert, von denen 13 Modelle mit Selektion 2 erzeugt wurden (s. Tab. 5-22). Ein sehr gutes Modell unter diesen 13 Modellen mit einem guten Kompromiss aus hohem R^2_{cv} , Q^2 und R^2 ist z.B. das 5-Felder-Modell (SEHAD) des Splits 2 ($R^2_{cv} = 0,626$ und $Q^2 = 0,648$). Die Aufteilung der Selektion 2 in Test- und Trainingsdatensatz dieses Splits sowie die Aktivitätsvorhersage für die Liganden mit diesem SEHAD-Modell sind in Tab. C 2 (Anhang C, S. 349 ff.) dargestellt.

Tab. 5-22 – Statistische Parameter der GAUSSIAN-Modelle der Selektionen 1-3 für einen Trainingsdatensatz von 80 % und drei Randomisierungen (Split 1, 2 und 3). Es wurden 15 Modelle mit den Kriterien $R^2_{cv} > 0,5$ und $Q^2 > 0,6$ erhalten (rot eingerahmt). Ein Modell mit besonders guten Werten für die statistischen Kennzahlen ist schwarz eingerahmt dargestellt.

Datensätze 1 - 3, Gaussian, 80 % Trainingsdatensatz												
Selektion 1												
Felder	Split 1				Split 2				Split 3			
	#k	R^2	R^2_{cv}	Q^2	#k	R^2	R^2_{cv}	Q^2	#k	R^2	R^2_{cv}	Q^2
SEHADR	5	0,799	0,636	0,603	5	0,783	0,600	0,567	3	0,757	0,633	0,310
SEHAD	4	0,751	0,618	0,579	4	0,730	0,587	0,539	3	0,748	0,626	0,346
SEHA	4	0,763	0,623	0,588	4	0,759	0,604	0,600	5	0,810	0,637	0,316
SEHD	4	0,722	0,593	0,585	4	0,707	0,576	0,540	3	0,729	0,609	0,343
SEAD	5	0,782	0,624	0,503	5	0,761	0,559	0,543	6	0,831	0,606	0,243
SHAD	4	0,738	0,608	0,555	4	0,715	0,575	0,537	3	0,739	0,610	0,319
EHAD	5	0,760	0,548	0,473	6	0,785	0,548	0,472	4	0,767	0,597	0,131
Selektion 2												
Felder	Split1				Split2				Split3			
	#k	R^2	R^2_{cv}	Q^2	#k	R^2	R^2_{cv}	Q^2	#k	R^2	R^2_{cv}	Q^2
SEHADR	5	0,797	0,615	0,648	5	0,805	0,630	0,649	7	0,849	0,648	0,477
SEHAD	5	0,789	0,608	0,666	6	0,818	0,626	0,648	8	0,854	0,646	0,530
SEHA	5	0,779	0,592	0,718	5	0,799	0,620	0,624	7	0,819	0,642	0,509
SEHD	4	0,730	0,601	0,630	6	0,791	0,602	0,653	6	0,787	0,611	0,515
SEAD	5	0,756	0,569	0,651	5	0,773	0,581	0,590	5	0,761	0,582	0,591
SHAD	5	0,784	0,600	0,664	6	0,810	0,610	0,661	8	0,841	0,639	0,507
EHAD	5	0,783	0,600	0,634	5	0,790	0,609	0,634	7	0,847	0,625	0,455
Selektion 3												
Felder	Split1				Split2				Split3			
	#k	R^2	R^2_{cv}	Q^2	#k	R^2	R^2_{cv}	Q^2	#k	R^2	R^2_{cv}	Q^2
SEHADR	5	0,827	0,667	0,452	5	0,817	0,645	0,493	9	0,811	0,650	0,410
SEHAD	5	0,822	0,675	0,452	9	0,838	0,641	0,484	7	0,810	0,640	0,402
SEHA	4	0,797	0,671	0,419	6	0,811	0,636	0,574	5	0,802	0,646	0,432
SEHD	5	0,791	0,631	0,495	6	0,788	0,620	0,529	4	0,736	0,589	0,471
SEAD	5	0,809	0,654	0,383	9	0,868	0,623	0,349	6	0,821	0,619	0,354
SHAD	5	0,810	0,665	0,407	5	0,615	0,784	0,564	5	0,791	0,622	0,387
EHAD	5	0,815	0,627	0,405	5	0,792	0,609	0,398	5	0,805	0,616	0,222

Mit den CoMFA-ähnlichen FORCE FIELD-Modellen (80 % Trainingsdatensatz) konnte kein Modell erhalten werden, welches den Anforderungen genügte (s. Tab. 5-23). Insbesondere das Bestimmtheitsmaß R^2 ist im Vergleich zu den GAUSSIAN-Modellen sehr gering ($R^2 < 0,7$). Das beste Modell wurde mit dem 2-Feld-Modell (SE) der Selektion 1, Split 3 erhalten ($R^2_{cv} = 0,493$ und $Q^2 = 0,644$).

Tab. 5-23 – Statistische Parameter der FORCE FIELD-Modelle der Selektionen 1-3 für einen Trainingsdatensatz von 80 % und drei Randomisierungen (Split 1, 2 und 3). Es wurde kein Modell mit den Kriterien $R^2_{cv} > 0,5$ und $Q^2 > 0,6$ erhalten. Das beste Modell (schwarz eingerahmt) wurde mit Selektion 1, Split 3 und zwei Feldern (SE) erhalten.

Selektionen 1 - 3, Force Field, 80 % Trainingsdatensatz												
Selektion 1	Split 1				Split 2				Split 3			
Felder	#k	R^2	R^2_{cv}	Q^2	#k	R^2	R^2_{cv}	Q^2	#k	R^2	R^2_{cv}	Q^2
SE	4	0,626	0,460	0,169	5	0,620	0,481	-1,058	5	0,648	0,493	0,644
S	5	0,573	0,396	-0,232	5	0,562	0,436	-0,960	5	0,580	0,401	-0,141
E	2	0,357	0,299	0,440	3	0,407	0,314	0,276	4	0,460	0,379	0,142
Selektion 2	Split 1				Split 2				Split 3			
Felder	#k	R^2	R^2_{cv}	Q^2	#k	R^2	R^2_{cv}	Q^2	#k	R^2	R^2_{cv}	Q^2
SE	5	0,560	0,363	0,422	4	0,545	0,390	0,255	5	0,577	0,436	0,204
S	3	0,416	0,252	0,218	5	0,537	0,325	-1,167	4	0,519	0,379	-0,481
E	3	0,338	0,245	0,492	4	0,359	0,262	0,452	4	0,370	0,276	0,208
Selektion 3	Split 1				Split 2				Split 3			
Felder	#k	R^2	R^2_{cv}	Q^2	#k	R^2	R^2_{cv}	Q^2	#k	R^2	R^2_{cv}	Q^2
SE	6	0,683	0,494	0,282	6	0,650	0,501	0,037	6	0,635	0,453	-0,178
S	5	0,592	0,414	0,118	5	0,495	0,308	-1,700	5	0,539	0,319	-0,503
E	3	0,425	0,343	0,319	3	0,472	0,395	-0,003	3	0,426	0,347	0,283

Abschließend lässt sich festhalten, dass nur mit der GAUSSIAN-Methode QSAR-Modelle der geforderten Qualität erhalten werden konnten. Dies könnte dadurch begründet sein, dass insbesondere die hydrophoben Felder, aber auch die Akzeptor- und Donorfelder wichtige Beiträge zur Beschreibung der Wechselwirkungsfelder der DLK-Inhibitoren liefern, während die FORCE FIELD-Modelle lediglich sterische und elektrostatische Felder berücksichtigen. Desweiteren zeigte sich, dass mit der Selektion 3 kein QSAR-Modell erhalten werden konnte, das den Anforderungen genügte. Mit einem Anteil von 70 % als Trainingsdatensatz ließen sich insbesondere in Split 3 die für diese Selektion besten QSAR-Modelle generieren (s. Anhang C, S. 354 f., Tab. C 3).

Die Aufteilung der jeweiligen Liganden in Test- und Trainingsdatensatz der Selektionen 1 und 2 wird für den Trainingsdatensatz 80 % für alle Splits im Anhang (C, S. 347 ff.) dargestellt. Die berechneten Aktivitätswerte werden jeweils für das beste Modell (s. Anhang C, Tab. C 1 u. Tab. C 2, rot markiert) aufgeführt.

Die statistischen Kennzahlen des besten QSAR-Modells für jede Selektion (im Folgenden als *Sel1*-, *Sel2*- bzw. *Sel3*-Modell bezeichnet) sind in Tab. 5-24 dargestellt. Eine gute Modellierung der Trainingsdaten gelingt mit dem *Sel3*-Modell (höchster R^2 -Wert und geringste Standardabweichung), während das *Sel1*-Modell mit der geringsten Komponentenanzahl $\#k$ auskommt und die höchste statistische Signifikanz (F) und Robustheit (R^2_{cv} , interne Validierung) aufweist. Das *Sel2*-Modell wiederum zeichnet sich im Vergleich zu den anderen beiden Modellen durch die beste Modellierung der Testdaten (höchster Q^2 -Wert und niedrigster RMSE) aus.

Tab. 5-24 – Statistische Analyse des besten GAUSSIAN-QSAR-Modells pro Selektion (Sel.). Tr. = Trainingsdatensatz. Fehlender Wert „-“ aus ursprünglichen Berechnungen nicht verfügbar → zweiter Wert aus neuen Berechnungen (s. Kap. 7.8).

	Sel. 1, 80 % Tr. SEHADR, Split 1	Sel. 2, 80 % Tr. SEHAD, Split 2	Sel. 3, 70 % Tr. SEHAD, Split 3
$\#k$	5	6	6
R^2	0,799	0,818	0,840
SD	0,577	0,548	0,518
F	135,5	125,9	128,5
R^2_{cv}	0,636	0,626	0,604
Q^2	0,603 // 0,601	0,648 // 0,642	0,588 // 0,611
RMSE	- // 0,80	0,75 // 0,76	- // 0,81
Beiträge der Felder			
S	36,5 %	37,4 %	37,8 %
E	8,0 %	7,7 %	10,3 %
H	22,5 %	29,0 %	21,1 %
A	17,3 %	18,0 %	18,0 %
D	8,4 %	7,9 %	12,8 %
R	7,4 %	-	-

Bezüglich der Verteilung der Felder werden die sterischen (S) und hydrophoben (H) Felder in jedem der drei Modelle am stärksten gewichtet, während die elektrostatischen (E) Felder den geringsten Beitrag liefern (s. Tab. 5-24). Der Beitrag der Akzeptorfelder ist etwa doppelt so hoch wie derjenige der Donor-Felder. Der Anteil der Feldbeiträge durch aromatische Ringe im 6-Felder-Modell der *Sel1* fällt ebenso gering aus wie die elektrostatischen und Donor-Felder.

Für alle drei Modelle ist eine gute Korrelation ($R^2 > 0,75$) zwischen den experimentellen DLK-inhibitorischen Aktivitätsdaten aus der Literatur und den mit dem Modell berechneten Aktivitätswerten sowohl für die Trainings- als auch für die Testdaten sichtbar (s. Abb. 5-61).

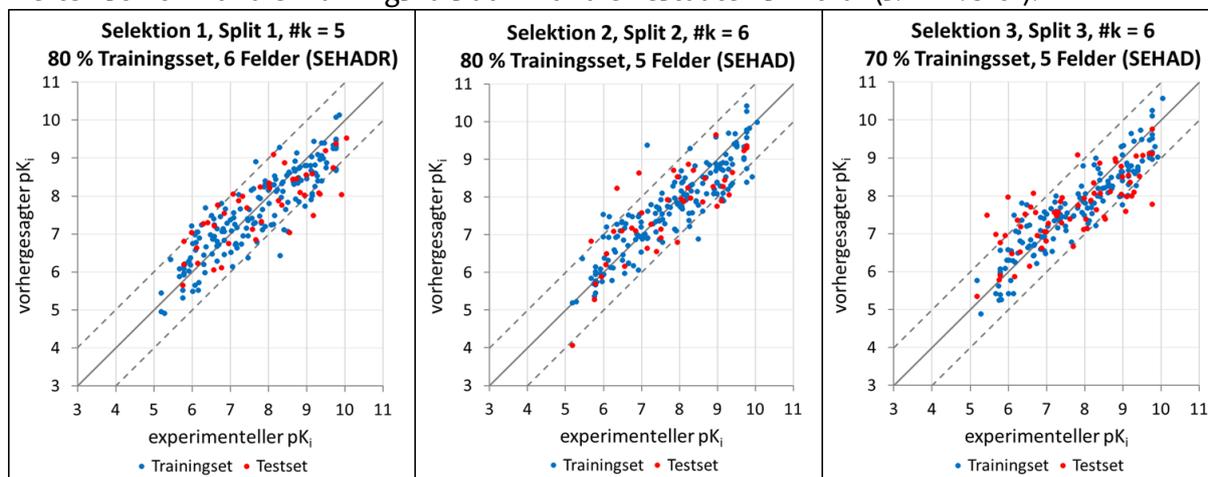


Abb. 5-61 – Korrelation der experimentellen (in der Literatur veröffentlichten) DLK-inhibitorischen Aktivitäten (pK_i bzw. pK_d) mit den vorhergesagten Aktivitäten. Dargestellt ist das – nach statistischen Kriterien – bestmögliche QSAR-Modell für jede Selektion.

Im *Sel1-Modell* ist eine starke Abweichung von $> 1,5$ pK_i-Einheiten als „Unterbewertung“ bei vier Verbindungen zu beobachten, während man eine solch große Abweichung hin zu einer geringeren Vorhersage-Aktivität im *Sel2-Modell* nur bei einer Verbindung beobachtet. In diesem Modell wiederum werden einige Substanzen (vier Verbindungen um $> 1,5$ pK_i-Einheiten) überbewertet, d.h. mit einer höheren Aktivität vorhergesagt als experimentell bestimmt wurde. Im *Sel3-Modell* sind Über- und Unterbewertung gleichermaßen zu beobachten.

Ligand	Akt.	Sel 1	Δ 1	Sel 2	Δ 2	Sel 3	Δ 3
F CEP-701	6,076	5,642	0,434	6,196	-0,120	5,754	0,322
K CHIR-258 TKI-258	5,678	6,078	-0,400	6,828	-1,150	6,973	-1,295
K JNJ-7706621	5,284	4,898	0,386	5,211	0,073	4,882	0,402
CH 019	8,292	9,280	-0,988	9,289	-0,997	8,979	-0,687
CH 023	9,495	9,189	0,306	9,444	0,051	9,473	0,022
CH 024	6,854	6,830	0,024	7,102	-0,248	7,077	-0,223
CH 030	9,770	10,074	-0,304	10,408	-0,638	10,239	-0,469
CH 031	9,770	9,322	0,448	9,295	0,475	9,606	0,164
CH 033	9,770	9,252	0,518	9,572	0,198	9,108	0,662
CH 036	9,770	9,295	0,475	9,714	0,056	9,141	0,629
CH 039	9,770	9,488	0,282	9,776	-0,006	10,095	-0,325
CH 040	9,770	9,380	0,390	9,354	0,416	9,751	0,019
CH 067	9,398	8,386	1,012	8,356	1,042	8,754	0,644
CH 068	9,523	8,865	0,658	9,321	0,202	9,053	0,470
CH 071	9,444	8,989	0,455	9,392	0,052	8,514	0,930
CH 074	9,699	9,248	0,451	9,225	0,474	9,117	0,582
CH 092	9,155	8,577	0,578	8,708	0,447	8,533	0,622
CH 097	9,222	8,805	0,417	9,117	0,105	9,637	-0,415
CH 113	5,796	6,799	-1,003	6,137	-0,341	6,757	-0,961
CH 115	10,046	9,523	0,523	9,976	0,070	10,562	-0,516
CH 121	9,699	9,330	0,369	9,325	0,374	9,465	0,234
CH 122	9,398	9,387	0,011	8,856	0,542	8,648	0,750
CH 123	9,699	8,745	0,954	9,030	0,669	9,071	0,628
CH 131	9,357	8,041	1,316	8,300	1,057	8,256	1,101
CH 140	9,770	9,458	0,312	10,269	-0,499	9,509	0,261
CH 145	5,444	6,323	-0,879	6,353	-0,909	7,486	-2,042
CH 146	5,678	5,868	-0,190	5,832	-0,154	5,417	0,261
CH 147	5,921	6,021	-0,100	5,972	-0,051	6,073	-0,152
EL 040	9,770	8,706	1,064	8,387	1,383	7,769	2,001
EL 051	9,770	8,667	1,103	8,740	1,030	8,962	0,808
EL 054	9,319	8,078	1,241	8,048	1,271	8,580	0,739
EL 078	6,000	7,035	-1,035	6,941	-0,941	6,564	-0,564
EL 092	6,027	6,736	-0,709	6,369	-0,342	6,290	-0,263
EL 095	5,921	6,278	-0,357	6,089	-0,168	6,939	-1,018
EL 114	6,018	5,481	0,537	5,932	0,086	5,413	0,605
EL 133	9,398	9,061	0,337	8,651	0,747	8,708	0,690
EL 155	6,523	6,472	-0,051	6,466	0,057	6,627	-0,104
EL 177	9,174	7,481	1,693	7,898	1,276	8,331	0,843
EL 179	6,000	7,207	-1,207	7,529	-1,529	7,956	-1,956
CD 001	5,192	4,949	0,243	4,057	1,135	5,333	-0,141
CD 002	5,770	5,311	0,459	5,272	0,498	5,780	-0,010
CD 003	5,192	5,434	-0,242	5,193	-0,001	5,766	-0,574
CD_004	5,762	5,514	0,248	5,353	0,409	5,238	0,524

Ligand	Akt.	Sel 1	Δ 1	Sel 2	Δ 2	Sel 3	Δ 3
CD 005	6,081	6,970	-0,889	7,200	-1,119	6,547	-0,466
CD 008	6,658	6,272	0,386	6,222	0,436	6,409	0,249
CD_017	9,000	8,790	0,210	8,767	0,233	8,497	0,503
CD_024	5,793	5,967	-0,174	5,454	0,339	5,379	0,414
CD 025	5,963	6,370	-0,407	5,875	0,088	6,297	-0,334
CD 030	6,092	6,574	-0,482	6,483	-0,391	6,471	-0,379
CD 045	5,757	5,648	0,109	5,691	0,066	5,615	0,142
CD 046	5,793	6,158	-0,365	5,938	-0,145	5,995	-0,202
CD 048	5,793	5,949	-0,156	5,662	0,131	5,931	-0,138
CD 064	6,569	6,870	-0,301	7,176	-0,607	6,673	-0,104
CD 067	6,187	5,714	0,473	6,124	0,063	6,243	-0,056
CD 068	5,793	6,193	-0,400	5,771	0,022	6,056	-0,263
CD 078	5,793	5,913	-0,120	5,699	0,094	5,885	-0,092
CD 109	6,319	6,189	0,130	6,528	-0,209	6,509	-0,190
CD 113	6,131	6,622	-0,491	6,314	-0,183	6,274	-0,143
CD 114	6,161	7,050	-0,889	7,480	-1,319	7,477	-1,316
CD 116	5,793	6,075	-0,282	6,012	-0,219	6,549	-0,756
CD 136	9,292	8,397	0,895	8,735	0,557	8,093	1,199
CD 138	7,006	7,371	-0,365	7,573	-0,567	7,514	-0,508
CD 157	9,509	8,738	0,771	9,365	0,144	8,916	0,593
CD 228	9,114	7,907	1,207	7,893	1,221	7,972	1,142
CD 355	9,241	8,595	0,646	8,889	0,352	7,988	1,253
CD_462	8,241	8,164	0,077	8,868	-0,627	8,331	-0,090
LL 024	9,469	9,391	0,078	9,676	-0,207	9,724	-0,255
LL 027	9,854	10,126	-0,272	9,819	0,035	9,291	0,563
LL 064	9,117	8,665	0,452	8,728	0,389	8,797	0,320
LL 075	9,131	9,025	0,106	7,953	1,178	7,996	1,135
LL 078	9,284	8,520	0,764	8,190	1,094	8,643	0,641
LL 080	9,222	8,252	0,970	8,535	0,687	8,783	0,439
LL 083	9,921	8,027	1,894	8,529	1,392	9,013	0,908
LL 087	9,194	9,437	-0,243	8,441	0,753	8,353	0,841
LL 101	9,357	8,961	0,396	9,013	0,344	9,039	0,318
LL 102	9,102	8,798	0,304	8,883	0,219	8,862	0,240
LL_132	6,588	7,232	-0,644	7,289	-0,701	7,695	-1,107
LL 133	6,493	6,970	-0,477	7,310	-0,817	7,758	-1,265
LL 154	9,155	8,626	0,529	8,907	0,248	8,503	0,652
LL 237	8,959	8,803	0,156	9,648	-0,689	9,291	-0,332
LL 257	9,155	8,064	1,091	8,100	1,055	8,295	0,860
PC 026	9,301	9,088	0,213	9,689	-0,388	9,434	-0,133
PC 027	9,155	8,000	1,155	8,263	0,892	8,661	0,494
PC 030	9,699	8,726	0,973	8,988	0,711	8,997	0,702
PH 009	6,357	7,685	-1,328	7,269	-0,912	7,805	-1,448
PH 010	6,260	7,236	-0,976	6,111	0,149	6,479	-0,219
K Staurosporin	5,824	6,202	-0,378	5,739	0,085	5,246	0,578
\bar{x} Diff.]			0,556		0,510		0,582

Tab. 5-25 – Vergleich der experimentellen (Akt.) und berechneten (Sel 1 bis 3) Aktivitäten für die in allen 3 Selektionen vorkommenden Verbindungen mit den Sel-Modellen. (Literaturquellen für Akt. im Ligandennamen ersichtlich, z.B. „CH“ s. exp. Teil, Kap. 7.11, S. 281; Nummerierung wie in der jeweiligen Patentschrift). Δ x = Differenz zwischen Akt. und Sel x. \bar{x} |Diff.] = Mittelwert des Betrags der Abweichungen Δ x.

Schaut man auf die mit dem Modell vorhergesagten Aktivitäten der Verbindungen, die in allen drei Selektionen vorkommen, und bildet die Differenz zwischen berechneter und experimenteller DLK-inhibitorischer Aktivität, so erkennt man, dass mit allen Modellen gute Vorhersagen für folgende Grundstrukturen getroffen werden (s. S. 176, Tab. 5-25, rot umrandet): zum einen für die über eine Aminofunktion verbundenen Pyridine, welche am Core 1 mit einer Cyanofunktion substituiert sind, sowie für die Bipyrimidinamin-Grundstruktur und dessen Difluoromethoxyderivat (s. Abb. 5-62 A). Relativ große Abweichungen vom experimentellen Wert (s. S. 176, Tab. 5-25, schwarz umrandet) erhalten einige wenige der trifluormethylsubstituierten Aminopyridine sowie einige der wenigen Pyrrolopyridine (beide Verbindungsklassen jeweils mit einem Pyrazolring als Core 2, s. Abb. 5-62 B).

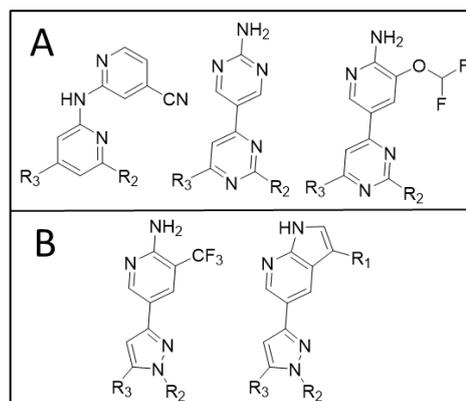


Abb. 5-62 – A – Grundstrukturen, für die gut mit den experimentellen Aktivitätsdaten übereinstimmende Aktivitäten mit den Sel-Modellen berechnet werden konnten. **B** – Grundstrukturen mit weniger guten Übereinstimmungen zwischen berechnetem und wahren Wert.

Mit dem *Sel3-Modell* erhält man für drei Verbindungen Abweichungen von ca. 2 Aktivitätseinheiten (1,956 bis 2,042). Eine solche große Abweichung (1,894) ist auch für eine Verbindung mit dem *Sel1-Modell* zu sehen, während die größte Abweichung mit dem *Sel2-Modell* 1,529 beträgt. Mit dem *Sel2-Modell* erhält man im Mittel die geringste Abweichung ($\bar{x} = 0,510$) vom experimentellem Aktivitätswert im Vergleich zu den anderen beiden Modellen. Darüber hinaus werden die Aktivitätswerte von den Staurosporinen (Staurosporin und CEP-701) sowie für JNJ-7706621 als unselektive DLK-Inhibitoren mit untypischen Grundstrukturen vom *Sel2-Modell* am besten vorhergesagt (s. Tab. 5-25, unselektive DLK-Inhibitoren blau umrandet).

Aufgrund der erkennbaren Vorteile wurde das *Sel2-Modell* ausgewählt, um die Posen des virtuellen Screenings im Rahmen des Postprocessings (s. Kap. 5.5.1-5.5.3) zusätzlich zu bewerten.

Mit Konturdiagrammen auf Basis von QSAR-Modellen können Volumenbereiche sichtbar gemacht werden, in denen Strukturänderungen des Liganden zu einer höheren oder niedrigeren Aktivität der Verbindung führen würden. In Abb. 5-63 werden Volumenbereiche auf Basis des ausgewählten *Sel2*-Modells dargestellt, in denen zum einen die Auswirkungen von sterischen Veränderungen des Moleküls und zum anderen von hydrophoben funktionellen Gruppen auf die Bindungsaffinität erkennbar sind⁵³. Die günstigen sterischen Bereiche 1-3 (grün) werden allesamt von

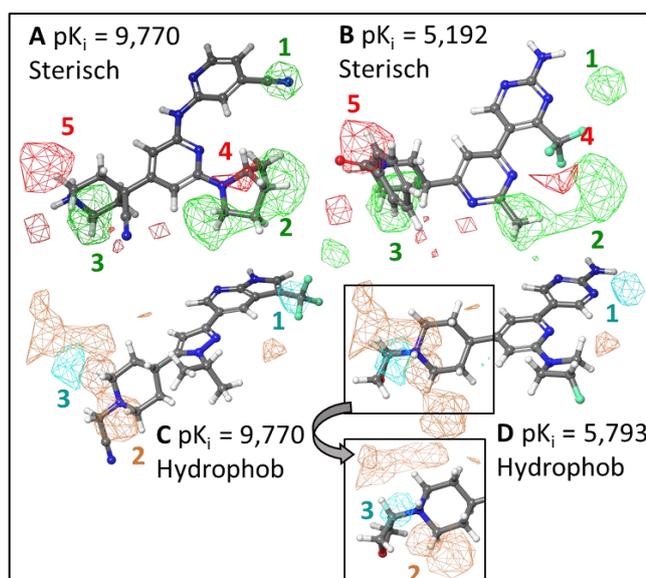


Abb. 5-63 – Konturdiagramme (*Sel2*-Modell). **A+B** – Für eine höhere Affinität günstige (grün) bzw. ungünstige (rot) sterische Felder. **C+D** – Für eine höhere Affinität günstige (cyan) bzw. ungünstige (orange) hydrophobe Felder. Pfeil: veränderte perspektivische Darstellung des Substituenten R3 zur verbesserten Sichtbarkeit der Felder.

dem hoch aktiven DLK-Inhibitor ausgefüllt (s. Abb. 5-63 A): Diese Bereiche entsprechen etwa den Bereichen für die Substituenten R1, R2 und R3 (s. Kap. 5.2.1). Die ungünstigen Bereiche 4 und 5 (rot) werden nicht tangiert. Bereich 5 wird allerdings von dem weniger wirksamen DLK-Inhibitor (s. Abb. 5-63 B) besetzt. Die vorteilhaften Bereiche werden hingegen weniger gut besetzt, da R2 (zum P-Loop gerichteter Substituent) kleiner ist als beim hoch aktiven Inhibitor, und da R1 (hintere Tasche, Bereich 1) fehlt.

Der günstige hydrophobe Bereich 1 (s. Abb. 5-63 C, cyan) wird von einem hoch aktiven Inhibitor mit einer Trifluormethylgruppe besetzt. Beim weniger aktiven Inhibitor (s. Abb. 5-63 D) ist Bereich 1 nicht (hydrophob) besetzt. Der hydrophob günstige Bereich 3 wird zum Teil vom basischen tertiären Amin (hydrophil) des weniger aktiven Inhibitors besetzt, was ungünstig für die Bindungsaffinität ist. Im Gegensatz dazu ragt beim stark aktiven Inhibitor das tertiäre Amin (hydrophil) in einen hydrophob ungünstigen Bereich 2 hinein, was sich günstig auf die Bindungsaffinität auswirkt.

Die dargestellten Konturdiagramme ermöglichen somit eine zusätzliche optische Beurteilung von Ligandbindungsgeometrien.

⁵³ Hier werden nur die sterischen und hydrophoben Felder dargestellt, da sie zusammen den Großteil der Wechselwirkungsfelder ausmachen (s. dazu Tab. 5-24, S. 174).

5.4 Strukturbasiertes virtuelles Screening

Der erste Teil dieses Kapitels beschreibt die Experimente zur Testung der verschiedenen Einstellungs- und Durchführungsoptionen der im Screening eingesetzten Dockingsoftware GLIDE (SCHRÖDINGER). Zunächst wurde überprüft, ob bzw. wie gut die Liganden aus den Röntgenkristallstrukturen **5CEO**, **5CEP** und **5CEQ** in die jeweils eigene Proteinkonformation (Redocking) und in die jeweiligen anderen Proteinkonformationen (Crossdocking) gedockt werden können. Die Proteinkonformationen wurden mit unterschiedlichen PROTPREP-Einstellungen vorbereitet, um herauszufinden, welche Einstellungen besonders gut geeignet sind, um die korrekte Pose zu reproduzieren. Hierbei wurden unterschiedliche Wassermoleküle in der Proteinstruktur berücksichtigt sowie auch die in Kap. 5.1.3 ausgewählten Proteinkonformationen aus den MD-Trajektorien verwendet. Anschließend wurde die Fähigkeit der Proteinkonformationen dahingehend untersucht, wie viele der unspezifischen DLK-Inhibitoren (s. S. 152 ff.) erfasst werden können. Im nächsten Schritt erfolgte die Analyse der Anreicherungsqualität mithilfe von Testdatensätzen im Größenbereich von etwa 10.000 Molekülen (Aktive und Decoys). Enrichmentplots stellten hierbei die Basis für die Auswahl dar. Schließlich wurde mit den auf diese Weise ausgewählten Proteinkonformationen mit besonders guten Dockingergebnissen für aktive Liganden und mit den optimierten Einstellungen der Screeningoptionen das eigentliche virtuelle Screening der ZINC-Datenbank durchgeführt (Kap. 5.4.2).

5.4.1 Vorbereitende Untersuchungen

5.4.1.1 Re- und Crossdocking des 5CEO-, 5CEP- und 5CEQ-Liganden

Zunächst wurden die Liganden aus den Röntgenkristallstrukturen **5CEO**, **5CEP** und **5CEQ** in ihre eigene Proteinkonformation (Redocking) sowie in die jeweils anderen drei Proteinkonformationen (Crossdocking) gedockt. Hier wurden die möglichen Einstellungen des „GLIDE DOCKING“-Tools von SCHRÖDINGER getestet.

Proteinpräparation

Die PDB **5CEN** und die drei kokristallisierten Proteinstrukturen der PDBs **5CEO**, **5CEP** und **5CEQ** wurden mit dem PROTPREP-Tool vorbereitet (s. Tab. 5-26 u. Kap. 7.3). Der fehlende Activation Loop in **5CEO**, **5CEP** und **5CEQ** wurde nicht vervollständigt, da er nicht Teil der im Docking wichtigen Bindetasche ist. Bei der Proteinvorbereitung wurde bei der Auswahl von Wassermolekülen und des Kraftfelds variiert (s. Tab. 5-26).

Tab. 5-26 – Übersicht über die vorbereiteten Proteinkonformationen aus den PDBs **5CEN**, **5CEO**, **5CEP** und **5CEQ**. ow = ohne Wassermoleküle; Ref. = Refinement; PP = Preprocessing; WBB = Wasserstoffbrückenbindungen

Bezeichnung	Wassermoleküle	Minimierung im Kraftfeld	Activation Loop
5CEN_opls2005_ow	Keine (vor Ref. gelöscht)	OPLS2005	vorhanden
5CEN_opls3_ow	Keine (vor Ref. gelöscht)	OPLS3	
5CEO_opls2005_w	PP: H ₂ O > 5 Å zu Heterogruppen gelöscht. Nach Ref.: Wassermoleküle ohne WBB zu Nicht-H ₂ O gelöscht	OPLS2005	nicht modelliert
5CEP_opls2005_w			
5CEQ_opls2005_w			
5CEO_opls2005_ow	Keine (im PP gelöscht)	OPLS3	
5CEP_opls2005_ow			
5CEQ_opls2005_ow			
5CEO_opls3_ow			
5CEP_opls3_ow			
5CEQ_opls3_ow			
5CEQ_opls3_ow			

Ligandenpräparation

Vor dem Docking wurden die Liganden aus den PDBs **5CEO**, **5CEP** und **5CEQ** mit LIGPREP vorbereitet (s. Kap. 7.11.1, Tab. 7-7, „(X) aus **5CEX_opls2005_w** für Re- und Crossdocking“). Von den Liganden des **5CEO** und **5CEQ** wurde sowohl eine neutrale als auch eine am Piperidin-Stickstoff des Substituenten R3 protonierte Form mit möglichst geringer Energie ausgewählt, während für den Liganden des **5CEP** nur die am Piperidin-Stickstoff des R3 protonierte Form ausgewählt wurde. Hier wurde kein neutraler Ligand automatisch durch LIGPREP erzeugt.

Erzeugung der Grids und Durchführung des Dockings

In den ATP-Bindetaschen der in Tab. 5-26 aufgeführten Proteinkonformationen wurden Grids (s. Kap. 4.4.1 u. Kap. 7.5.1) ohne jegliche Constraints generiert. Die vorbereiteten Liganden wurden mit dem „GLIDE DOCKING“-Tool (SCHRÖDINGER) in alle Proteinkonformationen gedockt (s. Kap. 7.5.2, Tab. 7-3, „Re- und Crossdocking“). Um die unterschiedlichen Präzisionsmodi der Dockingsoftware zu vergleichen, wurde zum einen der SP-Modus und zum anderen der XP-Modus von GLIDE verwendet. Für die im OPLS2005-Kraftfeld vorbereiteten Proteinkonformationen (s. Tab. 5-26) wurde auch das OPLS2005-Kraftfeld beim Docking ausgewählt, für die OPLS3-vorbereiteten Proteinkonformationen das OPLS3-Kraftfeld. Es wurden zehn Posen pro Ligand zugelassen.

Dockingergebnisse

In Tab. 5-27 sind die Dockingergebnisse zusammengefasst.

Tab. 5-27 –Re- und Crossdockingergebnisse im SP- und XP-Modus. Dargestellt ist jeweils das Ergebnis für die Pose mit dem besten (d.h. kleinsten) RMSD-Wert (**RM.**) im Vergleich zum Liganden der Röntgenkristallstruktur. Zu jedem RMSD-Wert ist der entsprechende Dockingscore (**DS**) aufgeführt. Abkürzung für die Liganden (linke Spalte): **5ceo_n** = **5CEO**-Ligand nach LIGPREP, neutral; **5ceo_p** = **5CEO**-Ligand nach LIGPREP, protoniert. **f** = falsche Pose; **f*** = nur die Substituenten R2 und R3 sind in ihrer Position vertauscht (d.h. Core 2 liegt gedreht vor, s. Abb. 5-64). ****** = Pose leicht nach vorne versetzt, ansonsten korrekte Ausrichtung der einzelnen Molekülteile.

SP	5CEO						5CEP						5CEQ						5CEN			
	opls2005_w		opls2005_ow		opls3_ow		opls2005_w		opls2005_ow		opls3_ow		opls2005_w		opls2005_ow		opls3_ow		opls2005_ow		opls3_ow	
	RM.	DS	RM.	DS	RM.	DS	RM.	DS	RM.	DS	RM.	DS	RM.	DS	RM.	DS	RM.	DS	RM.	DS	RM.	DS
5ceo_n	0,18	-9,54	1,70	-9,34	0,34	-9,41	3,36 f	-6,39	2,33 **	-6,72	2,32 **	-6,90	1,88	-8,16	1,20	-7,68	1,28	-7,81	5,75 f	-2,82	6,49 f	-3,63
5ceo_p	0,43	-10,88	6,30 f	-6,01	0,61	-10,75	3,50 f	-5,68	2,39 **	-8,17	2,39 **	-8,47	5,88 f*	-9,34	1,17	-8,99	1,32	-9,37	10,01 f	-4,94	6,54 f	-5,19
5cep_p	0,68	-9,07	0,62	-9,77	2,57 f	-6,48	0,65	-9,97	0,61	-9,49	0,27	-9,33	0,46	-11,00	0,50	-9,90	0,39	-10,06	1,30	-7,37	3,29 f	-6,20
5ceq_n	0,49	-8,51	3,01 f	-4,97	2,75 f	-4,69	0,73	-8,65	0,70	-8,16	0,73	-8,42	0,54	-9,81	0,45	-8,80	0,40	-8,39	4,54 f	-3,46	5,29 f	-4,56
5ceq_p	0,47	-9,41	0,75	-9,23	0,61	-9,49	0,87	-9,52	1,19	-8,94	0,85	-8,46	0,36	-11,45	0,85	-8,99	0,64	-9,41	4,44 f	-5,30	5,06 f	-5,04
XP	RM.	DS	RM.	DS	RM.	DS	RM.	DS	RM.	DS	RM.	DS	RM.	DS	RM.	DS	RM.	DS	RM.	DS	RM.	DS
5ceo_n	1,93	-6,06	1,69	-8,27	0,48	-8,19	5,28 f*	-8,51	5,26 f*	-8,64	5,29 f*	-8,12	1,89	-7,09	1,36	-8,84	1,53	-7,13	6,53 f	-2,66	6,58 f	-1,88
5ceo_p	1,07	-9,87	1,13	-9,06	0,42	-9,05	6,13 f	-5,77	2,82 **	-1,48	4,06 f	-4,81	5,87 f*	-9,10	1,19	-7,65	1,35	-7,79	6,64 f	-0,03	5,93 f	-3,22
5cep_p	0,77	-7,73	3,39 f*	-10,34	0,60	-8,21	0,48	-9,75	0,44	-10,28	0,38	-9,73	0,35	-9,98	0,51	-9,62	0,41	-8,91	4,15 f	-3,27	4,73 f	-2,56
5ceq_n	0,87	-9,13	0,90	-8,51	5,11 f*	-8,4	0,93	-9,09	0,77	-9,14	0,93	-8,80	0,55	-9,49	0,45	-8,28	0,28	-7,83	5,49 f	-4,14	5,56 f	-2,61
5ceq_p	0,78	-9,96	0,80	-7,06	0,70	-8,74	1,17	-8,18	1,32	-9,72	1,10	-9,77	0,35	-10,50	1,07	-9,56	0,67	-9,30	4,65 f	-4,13	4,98 f	-2,22

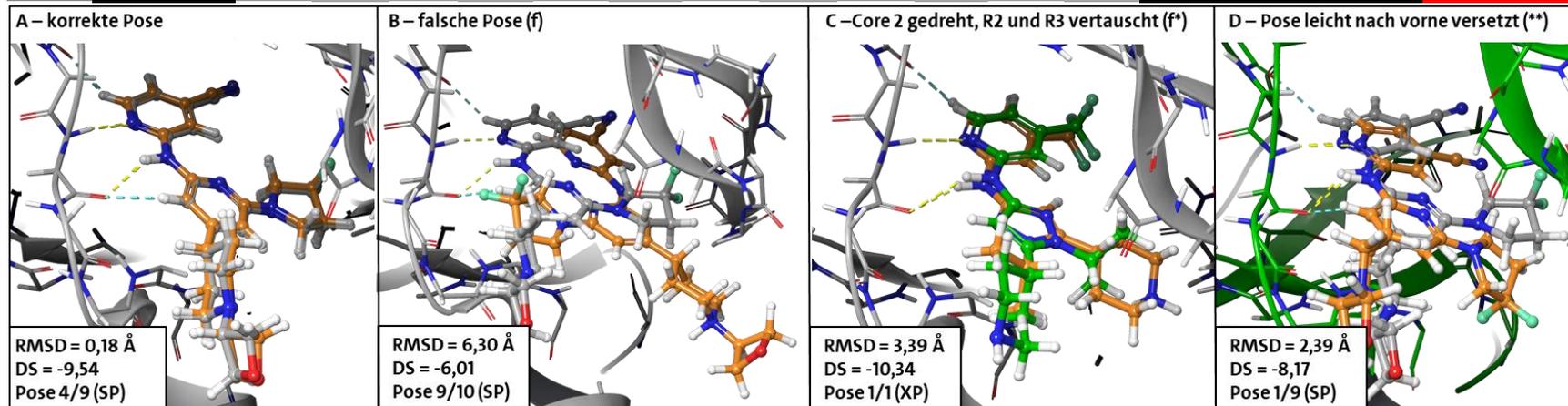


Abb. 5-64 – Beispielhafte Szenarien bei der Posenbewertung. **A** – korrekte Pose der gedockten Struktur **5ceo_n** (orange) im Vergleich zum **5CEO**-Referenzliganden (grau); **B** – falsche Pose (orange), keine Ausbildung der charakteristischen Interaktionen zur Bindetasche möglich. **C** – Core 2 liegt falsch herum, daher sind R2 und R3 in ihrer Lage vertauscht. Dennoch gute Bewertung (Dockingscore). **D** – Pose sitzt nicht ganz passgenau in der Tasche, dennoch sind die Interaktionen zur Hinge-Region noch möglich.

Die Abweichung der gefundenen Dockingpose zur Originalposition des gebundenen Liganden wurde mithilfe des RMSD-Werts beurteilt (Berechnung s. Kap. 7.5.3). Bei einem RMSD-Wert von $< 2,0 \text{ \AA}$ kann meist eine richtige Ausrichtung des Liganden in der Bindetasche angenommen werden. Eine korrekte Pose (Bsp. s. Abb. 5-64 A) wurde in den dargestellten Ergebnissen (bester RMSD-Wert pro Ligand, Tab. 5-27) noch bis zu einem RMSD-Wert von $1,93 \text{ \AA}$ erhalten, während der schlechteste Dockingscore einer korrekten Pose bei $-6,06$ lag. Dennoch können Dockingposen mit besseren Dockingscores falsch ausgerichtet sein: z.B. erhält eine Pose, in der die räumliche Position der Substituenten R2 und R3 durch Rotation des Cores 2 vertauscht ist, dennoch gute Dockingscores, da die wichtigen Interaktionen, wie z.B. die Wasserstoffbrückenbindungen des Cores 1 zur Hinge-Region und der Sandwichkomplex durch Ala150 (oben) und Leu243 (unten) dennoch ausgebildet werden können (s. Abb. 5-64 C). Eine solche Pose kann man weniger am Dockingscore, sondern eher am erhöhten RMSD-Wert (hier $3,39\text{-}5,88 \text{ \AA}$) erkennen. Auch eine leicht nach vorne versetzte Pose (s. Abb. 5-64 D) hat einen erhöhten RMSD-Wert ($2,32\text{-}2,82 \text{ \AA}$), aber einen immer noch guten Dockingscore, da die wichtigen Interaktionen ausgebildet werden können.

Zuerst lässt sich u.a. anhand des Farbmusters in Tab. 5-27 erkennen, dass das Docking im SP- und XP-Modus zu ähnlichen Ergebnissen führt:

- RMSD-Werte: Der Mittelwert der besten RMSD-Werte liegt für den SP-Modus bei $2,08 \text{ \AA}$ und beim XP-Modus bei $2,36 \text{ \AA}$. Somit sind die Posen des SP-Dockings ähnlicher zur Originalposition des gebundenen Liganden als die Posen des XP-Dockings.
- Scorewerte: Die hier aufgeführten SP-Posen mit dem jeweils besten RMSD-Wert haben im Mittel einen negativeren Dockingscore ($-7,97$) als die XP-Posen ($-7,42$). Somit sind die Scorewerte im SP-Modus etwas besser.
- Anzahl korrekter Posen: Fasst man all die beschriebenen Posenszenarien mit einem RMSD $> 1,93 \text{ \AA}$ (s. Abb. 5-64 B-D) als falsche Posen zusammen, so werden mit dem SP-Modus 20 und mit dem XP-Modus 19 falsche Posen erzeugt. Die Anzahl der korrekten Posen ist somit vergleichbar.

Die Ligandengröße beeinflusst das Dockingergebnis. Sie nimmt ab von **5CEO** $>$ **5CEQ** $>$ **5CEP**. Der **5CEO**-Ligand ergab wegen seines größten Platzbedarfs die höchsten RMSD-Werte (im Mittel über die in Tab. 5-27 dargestellten RMSD-Werte: $3,22 \text{ \AA}$), während der mittlere RMSD-Wert für den **5CEQ**-Liganden bei $1,71 \text{ \AA}$ und für den **5CEP**-Liganden bei $1,25 \text{ \AA}$ liegt. Der **5CEP**-Ligand konnte wegen seiner geringen Größe in fast jede Proteinkonformation gedockt werden (s. Tab. 5-27, blau umrandet).

Das Apoenzym **5CEN** wurde als die Proteinkonformation mit der schlechtesten Aufnahmefähigkeit für Liganden identifiziert. Bis auf eine Pose (**5CEP**-Ligand in **5CEN_opls2005_ow**) kann in den **5CEN**-Proteinkonformationen keine korrekte Pose erzeugt werden (s. S. 181, Tab. 5-27, rot umrandet, mittlerer RMSD: 5,40 Å, mittlerer Dockingscore: -3,76). Die Bindetasche des **5CEP** ist für den **5CEO**-Liganden zu klein – hier war kein Docking erfolgreich. **5CEO** und **5CEQ** stellten einen größeren Raum zur Verfügung. Besonders gute Dockingergebnisse wurden mit den Proteinkonformationen **5CEO_opls2005_w** (mittlerer RMSD = 0,77 Å), **5CEQ_opls2005_ow** (0,88 Å) und **5CEQ_opls3_ow** (0,83 Å) erhalten (s. S. 181, Tab. 5-27, schwarz umrandet).

Tendenziell wurden bessere Ergebnisse mit OPLS3 als mit OPLS2005 erzielt. Die Anwesenheit von Wassermolekülen führt insbesondere in **5CEO** zu besseren RMSD- und Dockingscorewerten. Aufgrund der genannten Vorteile des SP-Modus und des wesentlichen Nachteils des XP-Modus der acht- bis zehnmals längeren Zeit für den Dockingprozess⁵⁴ wurde die Verwendung des SP-Modus als die für ein Screening geeignetere Methode angesehen.

5.4.1.2 VSW des **5CEO**-, **5CEP**- und **5CEQ**-Liganden sowie der unspezifischen DLK-Inhibitoren und ANP

Das „VIRTUAL SCREENING WORKFLOW“ (VSW)-Tool der Software SCHRÖDINGER hat gegenüber dem „GLIDE DOCKING“-Tool den Vorteil, dass in mehrere Grids gleichzeitig gedockt werden kann und dass die Dockingmodi HTvS, SP und XP hintereinandergeschaltet werden können. Die Optionen dieses Tools sollten mithilfe von Re- und Crossdockingexperimenten näher untersucht werden. Dazu wurden zum einen die Liganden aus den Röntgenkristallstrukturen **5CEO**, **5CEP** und **5CEO** verwendet. Desweiteren wurde überprüft, wie gut die 30 unspezifischen DLK-Inhibitoren, die von den selektiven DLK-Inhibitoren strukturell stark abweichen (vgl. Kap. 5.2.1), und ANP in die Proteinkonformationen gedockt werden. Gute Dockingergebnisse wären ein Zeichen dafür, dass die betreffende Proteinkonformation nicht nur eine Art von Inhibitortyp aufnehmen kann. Die Zahl der eingesetzten Proteinkonformationen wurde durch weitere Variationen in der Protein-vorbereitung sowie durch die zusätzlichen Proteinkonformationen aus den MD-Trajektorien (s. Kap. 5.1.3) im Vergleich zum vorherigen Dockingexperiment erhöht: es sollte geklärt werden, welchen Einfluss eine unterschiedliche Proteinpräparation, wie z.B. das Vorhandensein von Wassermolekülen, sowie eine diverse Auswahl von Bindetaschenkonformationen auf die Dockingergebnisse hat.

⁵⁴ Bsp.: Docking des **5CEP**-Liganden in die **5CEP_opls2005_w**-Struktur → SP: 13 Sek., XP: 108 Sek.

Ligandenpräparation

Mit dem OPLS3-Kraftfeld wurden die Liganden minimiert und alle möglichen Protonierungszustände und Tautomere im pH-Bereich 5-9 erstellt (s. Kap. 7.11.1, Tab. 7-7, „(X) aus **5CEX_opls2005_w**; ANP, (KFT) für VSW-Exp.“).

Proteinpräparation

Da die Anwendung des OPLS3-Kraftfelds in den vorherigen Versuchen bessere Ergebnisse erzielte, wurden die verwendeten Proteinkonformationen nur noch mit dem neueren OPLS3-Kraftfeld vorbereitet und nicht mehr beide Varianten untersucht. Mit unterschiedlichen Einstellungen im PROTPREP-Tool wurden die vier Röntgenkristallstrukturen **5CEO**, **5CEP**, **5CEQ** und **5CEN** vorbereitet. Es wurden insgesamt 44 Varianten erstellt (s. Anhang E, S. 357, Tab. E 1). Insbesondere wurden folgende Variationen bei der Protein Vorbereitung eingesetzt:

- 1) Alle in der Röntgenkristallstruktur vorhandenen Wassermoleküle wurden vor der Minimierung gelöscht (IDs 3, 8, 13, 18 in Anhang E, S. 357, Tab. E 1).
- 2) Alle Wassermoleküle aus der Röntgenkristallstruktur wurden für die Minimierung beibehalten. Dies führt zu minimalen Unterschieden in der Proteinkonformation im Vergleich zu 1). Nach Abschluss der Minimierung wurden
 - a) alle Wassermoleküle beibehalten⁵⁵ (IDs 6, 11, 16).
 - b) einzelne definierte Wassermoleküle mit einer potenziellen Rolle als *Structural Waters* in der Bindetasche beibehalten (für **5CEO**⁵⁶, IDs 6.1-6.11) oder
 - c) alle Wassermoleküle nachträglich gelöscht (IDs 2, 7, 12, 17).

Der unterschiedliche Umgang mit Wassermolekülen führt zu verschiedenen Szenarien: zum einen werden durch das Entfernen von proteingebundenen Wassermolekülen Stellen freigelegt, die vom Liganden eingenommen werden können. An diesen Stellen sind starke Kontakte durch Wasserstoffbrückenbindungen zum Protein möglich. Zum anderen können gebundene Wassermoleküle wichtige Vermittler zwischen Ligand und Protein als Bestandteil eines Wasserstoffbrückenbindungsnetzwerks sein. Die Auswahl einzelner definierter Wassermoleküle

⁵⁵ Für **5CEN** als Apoenzym nicht möglich, da auch Wassermoleküle so in der Bindetasche liegen, dass kein Ligand gedockt werden könnte.

⁵⁶ Die Auswahl definierter Wassermoleküle des **5CEO** erfolgte erst im Anschluss der ersten Evaluation der Dockingergebnisse mit den IDs 1-33, wird aber in dieser Arbeit zusammengefasst dargestellt.

(**Wats 1-4**) erfolgte nach der Analyse der konservierten Wassermoleküle in der Bindetasche (s. Kap. 5.1.1). Darüber hinaus wurden für eine Proteinkonformation (ID 6.8) vier weitere Wassermoleküle (s. Abb. 5-65) im Bereich hinter der Bindetasche beibehalten.

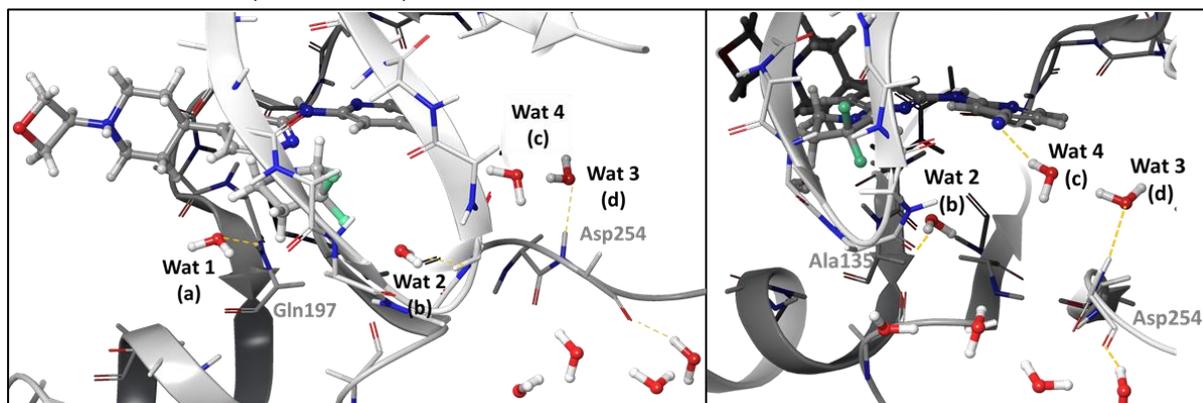


Abb. 5-65 – 5CEO mit acht Wassermolekülen, die in der Proteinkonformation ID 6.8 verwendet wurden, aus unterschiedlichen Perspektiven. Darüber hinaus wurden verschiedene Kombinationen aus den Wassermolekülen **Wat 1-4** (in den Proteinkonformationsbezeichnungen: Buchstaben in Klammern verwendet) in den IDs 6.1-6.7 und 6.9-6.11 verwendet.

3) Wassermoleküle mit einer Entfernung von mehr als 5 Å von Liganden wurden im *Preprocessing* gelöscht. Bei der Optimierung wurden alle Wassermoleküle gelöscht, die weniger als eine Wasserstoffbrückenbindung zu Nicht-Wassermolekülen ausbildeten. Nach Abschluss der Vorbereitung wurden die verbliebenen Wassermoleküle entweder beibehalten (IDs: 4, 9, 14) oder gelöscht (IDs 5, 10, 15). Genauso wurde beim 5CEN verfahren, nur, dass im *Preprocessing* alle Wassermoleküle beibehalten wurden. Nach der Minimierung wurden alle Wassermoleküle gelöscht (ID 1).

In dieser Bearbeitung hat man also im Vergleich zu 1) – alle Wassermoleküle vor der Minimierung gelöscht – und 2) – alle Wassermoleküle vor der Minimierung beibehalten – einige Wassermoleküle mit einem Maximalabstand vom Protein bzw. zu anderen Wassermolekülen für die Minimierung beibehalten. Dies führt wieder zu minimal unterschiedlichen Proteinkonformationen im Vergleich zu 1) und 2). Darüber hinaus spielen hier die Optionen An- und Abwesenheit von Wassermolekülen in der Bindetasche eine Rolle.

4) Die Proteinkonformationen wie in 2) bis zum Abschluss des Optimierungsschrittes, aber ohne die anschließende Minimierung, wurden verwendet. Entweder wurden die Wassermoleküle beibehalten (IDs 20, 22, 24) oder entfernt (IDs 19, 21, 23, 25). Diese Proteinkonformationen wurden verwendet, um einen Vergleich zwischen minimierten und nicht-minimierten Strukturen im Docking zu haben.

5) Die aus den MD-Trajektorien auf Basis ihres Bindetaschenvolumens und ihrer von den üblichen Röntgenkristallstrukturen abweichenden Konformationen – insbesondere in Bezug auf

den P-Loop – wurden verwendet (s. Kap. 5.1.3). Sie wurden mit Wassermolekülen in der Binde-tasche in einem Abstand von 6 Å um den Liganden herum im OPLS3-Kraftfeld minimiert (IDs 26-33). Alle Wassermoleküle wurden anschließend gelöscht. Dies sind die einzigen **5CEO**-, **5CEP**- und **5CEQ**-Konformationen in diesem Dockingexperiment, welche einen vervollständig-ten Activation Loop besitzen.

Erzeugung der Grids und Durchführung des Dockings

Als Constraints wurden die Atome der drei Aminosäuren der Hinge-Region definiert, welche die charakteristischen, kinasetypischen Wasserstoffbrückenbindungen zum Liganden ausbilden (s. Kap. 7.5.1): Glu191 O (Backbone), Cys192 H des Amidstickstoffs im Backbone und Cys193 O (Backbone, vgl. Kap. 5.2.1, S. 148, Abb. 5-33). Zur Durchführung des Dockings wurde das „VIRTUAL SCREENING WORKFLOW“-Tool der Software SCHRÖDINGER genutzt. Hierbei wurde festgelegt, dass mindes-tens eine Wasserstoffbrückenbindung der drei definierten Constraints ausgebildet werden muss. Es wurde mit allen Dockingmodi (HTvS, SP, XP)⁵⁷ unter Anwendung des OPLS2005-Kraft-felds hierarchisch hintereinander flexibel gedockt und eine Pose pro Ligandenzustand zugelas-sen. Diese Option wurde getestet, weil es für das virtuelle Screening der ZINC-Datenbank auf-grund der großen Menge von Millionen virtuellen Molekülen von großer Bedeutung sein würde, dass nur eine Pose pro Ligandenzustand erzeugt wird und diese – falls aktiv – korrekt wiederge-geben bzw. gut bewertet wird. Die Pose pro Ligand unter den verschiedenen Ligandenzuständen mit dem besten Dockingscore wurde in den folgenden Auswertungen berücksichtigt.

Auf diese Weise wurden folgende Dockingexperimente mit unterschiedlichen Liganden bzw. Targetauswahlen durchgeführt (s. s. Kap. 7.5.2, Tab. 7-3, „VSW Re- und Crossdocking“):

- a) Die vorbereiteten Liganden aus **5CEO**, **5CEP** und **5CEQ** sowie die 30 unspezifischen DLK-Inhibi-toren (s. Kap. 5.2.1, S. 152 ff.) und ANP wurden in die Proteinkonformationen der IDs 1-33 (noch nicht: IDs 6.1-6.11) gedockt.
- b) Zusätzlich wurden alle hier genannten Liganden in die 11 Proteinkonformationen der IDs 6.1-6.11 (s. Anhang E, S. 357, Tab. E 1) gedockt, welche Variationen der Proteinkonformation ID 6 (**5CEO_opls3_minwat**) hinsichtlich der Auswahl bestimmter Wassermoleküle in der Bindeta-sche (s. S. 184, Punkt 2 bzw. Tab. E 1, Anhang E, S. 357) darstellen.

⁵⁷ Für alle Dockingmodi war zunächst das OPLS2005-Kraftfeld hinterlegt. Unter Edit/Preferences/Force Field (SCHRÖDINGER MAESTRO) kann dies geändert werden.

Dockingergebnisse

Das XP-Docking erfolgte nach dem HTvS- und SP-Docking als letzte Instanz, somit können Liganden schon vorher ausgesiebt worden sein.

Zu a) Docking in die IDs 1-33:

In Tab. 5-28 ist der beste Dockingscore pro Ligand nach dem letzten Schritt des XP-Dockings der vorbereiteten Röntgenkristallstrukturliganden des **5CEO**, **5CEP** und **5CEQ** in die 33 verschiedenen Proteinkonformationen (IDs 1-33, s. Anhang E, Tab. E 1) aufgeführt. Die Proteinkonformationen IDs 19-25 (s. Tab. 5-28, grau hinterlegt) sind in der Proteinvorbereitung nicht-minimierte Strukturen. Sie wurden lediglich als Vergleich verwendet, falls wesentliche Unterschiede hinsichtlich des Aufnahmevermögens für Liganden oder Korrektheit der Posen feststellbar wären. Bei deutlich besseren Ergebnissen der nicht-minimierten Strukturen hätte man diese weiterverwendet – das ist hier aber nicht der Fall.

Von den übrigen Proteinstrukturen ergaben 13 IDs für jeden der drei Liganden richtige Dockingposen mit Dockingscore-Werten besser als -6 (s. Tab. 5-28, blau umrandet). ID 6 sticht mit den besten Dockingscores hervor. Dies ist eine Konformation, die mit allen ursprünglich vorhandenen Wassermolekülen minimiert wurde, welche für das Docking beibehalten wurden. Aufgrund dieses Ergebnisses wurde diese Proteinkonformation ausgewählt und hinsichtlich der Wassermolekülkombinationen (s. Anhang E, Tab. E 1, IDs 6.1-6.11) modifiziert, um weitere Testdockings (s. S. 189 f.) durchzuführen.

In vier der **5CEN**-Strukturen (IDs 1-3, ID 19) war ein Docking kaum möglich. Dies könnte an der kleinen Bindetasche des **5CEN** liegen, welche im Gegensatz zu den anderen drei

Tab. 5-28 – Dockingscores der XP-Posen des jeweils bestbewerteten Ligandenzustands nach dem VSW in die erstellten Proteinkonformationen. 0 = keine Dockingpose. f = falsche Pose, f* = Core 2 gedreht, R2 und R3 vertauscht, siehe Abb. 5-64, S. 181. **5ceol** = **5CEO**-Ligand etc. Hinter der ID ist die PDB aufgeführt, aus der die Proteinkonformation hervorgeht. Unzureichende Dockingergebnisse sind rot umrandet. Gute Ergebnisse sind blau markiert. Die grau hinterlegten Strukturen sind nicht minimiert.

	5ceol	5cepl	5ceql
1 5CEN	0	0	-3,396 f
2 5CEN	-3,421 f	0	-3,309 f
3 5CEN	-3,325 f	0	-4,043 f
4 5CEO	-10,171	-3,890 f*	-9,842
5 5CEO	-8,665	0	-8,971
6 5CEO	-10,752	-9,942	-10,505
7 5CEO	-8,810	0	-8,898
8 5CEO	-9,469	0	-8,582
9 5CEP	-8,343 f*	-9,702	-9,550
10 5CEP	-8,664 f*	-10,118	-9,893
11 5CEP	0	-8,379	-10,596
12 5CEP	-8,344 f*	-10,137	-9,193
13 5CEP	-8,783 f*	-9,910	-9,264
14 5CEQ	0	-10,225 f*	-10,217
15 5CEQ	-7,700	-9,899 f*	-9,253
16 5CEQ	0	-11,479	-12,503
17 5CEQ	-8,775 f*	-10,176 f*	-9,600
18 5CEQ	-7,750 f*	-9,661	-9,959
19 5CEN	-2,530 f	0	-5,037 f
20 5CEO	0	-10,046	-10,729
21 5CEO	-10,759	-10,164 f	-9,877
22 5CEP	0	-8,959	-11,067
23 5CEP	0	-10,152	-9,928
24 5CEQ	0	-10,436	-9,236
25 5CEQ	-8,702	-10,203	-9,947
26 5CEN	-7,128	-7,373	-6,164
27 5CEN	0	-7,552	-6,257 f*
28 5CEO	-8,360	-8,320	-8,792
29 5CEO	-5,890	0	0
30 5CEP	-6,089	-8,656	-7,301
31 5CEP	-5,412 f*	0	-6,127
32 5CEQ	-7,924 f*	-8,752	-7,993
33 5CEQ	-8,901	-8,868	-9,335

PDBs **5CEO**, **5CEP** und **5CEQ** nicht an einen Liganden angepasst ist. Mit den beiden modifizierten **5CEN**-Proteinkonformationen aus den MD-Trajektorien (ID 26 u. 27) wurden etwas bessere Dockingergebnisse erzielt.

Die Auswahl von Proteinkonformationen sollte einen guten Kompromiss darstellen zwischen Aufnahmevermögen für hochselektive DLK-Inhibitoren und für DLK-Inhibitoren mit von den üblichen Inhibitoren abweichenden Grundstrukturen. Für die Auswertung des Crossdockings der 30 unspezifischen DLK-Inhibitoren und ANP in die Proteinkonformationen der IDs 1-33 wurde zunächst gezählt, wie viele der Liganden überhaupt in eine Proteinkonformation (#Lig) gedockt werden konnten und wie viele dieser Posen in der zu erwartenden Ausrichtung (#Korr, s. Kap. 5.2.1) auftraten (s. Tab. 5-29). Hier sind die tatsächlichen Bindekonformationen in der DLK nicht bekannt, sondern wurden aus anderen Kinasen abgeleitet (s. Kap. 7.10). Die zuvor auf Basis der Dockingergebnisse des **5CEO**-, **5CEP**- und **5CEQ**-Liganden getroffene Vorauswahl von 13 IDs (s. Tab. 5-29, blau) wurde anhand dieser Ergebnisse weiter eingegrenzt auf die sechs IDs 7, 13, 18, 26, 30 und 33.

ID 7 (basierend auf **5CEO**) entspricht ID 6 ohne Wassermoleküle und erzielte eine höhere Anzahl korrekter Posen (s. Tab. 5-29, grün umrandet) als ID 6. ID 6 wird im Folgenden eine weitere Rolle spielen, indem die Kombi-

nation der Wassermoleküle verändert wird und daraus unterschiedliche Grids gebildet werden. Die **5CEP**-IDs 9, 10, 12 und 13 hatten sehr ähnliche Dockingscores im XP-Docking der Röntgenkristallstruktur-Liganden (s. S. 187, Tab. 5-28), aber anhand der Dockingergebnisse der unspezifischen DLK-Inhibitoren wurde nur ID 13 ausgewählt. Zusätzlich wurde ID 30 als weiterer **5CEP**-Repräsentant mit vervollständigtem Loop (MD-Simulation) ausgewählt. Von den **5CEQ**-Strukturen ID 15, 17 und 18 wurde letztlich ID 18 gewählt. Hier waren die Ergebnisse aber recht ähnlich. ID 33 als

Tab. 5-29– Ergebnisse des VSW der 30 unspezifischen DLK-Inhibitoren und ANP mit den IDs 1-33. #Lig = Anzahl gedockter Liganden, #Korr = Anzahl gedockter Liganden in erwarteter Pose (korrekte Pose). Auswahl = grün umrandet

ID	HTvs		SP		XP	
	#Lig	#Korr	#Lig	#Korr	#Lig	#Korr
1 5CEN	13	3	12	3	4	2
2 5CEN	11	2	11	2	4	2
3 5CEN	15	3	14	2	5	2
4 5CEO	16	4	16	3	13	4
5 5CEO	16	3	16	5	11	7
6 5CEO	18	3	18	3	15	4
7 5CEO	17	6	17	5	10	5
8 5CEO	19	3	19	6	13	5
9 5CEP	12	2	11	2	10	2
10 5CEP	17	3	16	5	16	4
11 5CEP	10	1	9	2	6	1
12 5CEP	18	3	18	2	15	3
13 5CEP	20	4	20	4	19	4
14 5CEQ	13	2	13	3	9	2
15 5CEQ	19	4	19	3	16	3
16 5CEQ	4	3	4	3	3	3
17 5CEQ	17	5	17	4	14	3
18 5CEQ	19	4	18	3	16	4
19 5CEN	10	1	10	2	5	2
20 5CEO	18	1	18	3	14	4
21 5CEO	17	2	17	5	11	7
22 5CEP	10	1	10	2	8	2
23 5CEP	15	3	15	3	13	2
24 5CEQ	4	2	4	2	3	2
25 5CEQ	18	4	18	3	15	4
26 5CEN	22	4	22	7	21	11
27 5CEN	21	4	21	5	15	5
28 5CEO	17	3	17	4	14	3
29 5CEO	15	5	14	7	12	6
30 5CEP	17	4	17	6	17	7
31 5CEP	15	3	15	8	13	5
32 5CEQ	22	3	22	5	21	4
33 5CEQ	21	2	21	2	16	2
Auswahl	28	6	27	9	27	13

5CEQ-MD-Konformation mit vervollständigtem Loop wurde beibehalten, obgleich nur zwei un-spezifische DLK-Inhibitoren korrekt gedockt werden konnten. Dafür waren die Ergebnisse des Crossdockings zuvor (s. S. 187, s. Tab. 5-28) für ID 33 besser als für ID 32. Ebenso wie im Docking zuvor mit den X-Ray-Liganden wurden nicht zufriedenstellende Ergebnisse mit den **5CEN**-IDs 1-3 erhalten. Unerwartet positiv waren die Ergebnisse für die **5CEN**-MD-Konformation ID 26. Diese wurde daher weiterverwendet.

In den meisten Proteinkonformationen bleibt die Anzahl gedockter Liganden (#Lig, unabhängig davon, ob in der zu erwartenden Ausrichtung gedockt wurde) im SP-Docking gegenüber dem HTvS-Docking unverändert. In 16 Proteinkonformationen werden mit dem SP-Modus mehr, in 8 Fällen weniger korrekte Posen (#Korr) als mit dem HTvS-Modus erzeugt. In den übrigen 9 Proteinkonformationen bleibt die Zahl korrekter Posen (#Korr) gleich. Beim Übergang vom SP- zum XP-Modus nimmt die Zahl gedockter Liganden (#Lig) immer ab (nur in ID 30 gleichbleibend). In 11 Proteinkonformationen nimmt die Zahl korrekter Posen (#Korr) ab, in 10 Konformationen zu, in den übrigen 12 sind sie gleichbleibend. Dies lässt darauf schließen, dass vom Übergang vom HTvS- zum SP-Modus die Zahl korrekter Posen tendenziell zunimmt, während die Gesamtzahl der gedockten Liganden gleichbleibt. Hingegen werden im XP-Modus im Vergleich zu SP falsche Posen tendenziell aussortiert. Nach Vereinigung der Dockingposen aus den sechs ausgewählten Proteinkonformationen und Beibehaltung der Pose mit dem besten Dockingscore, stellt man eine Erhöhung der Anzahl korrekter Posen von HTvS < SP < XP fest.

Mit den sechs Proteinkonformationen (ID 7, 13, 18, 26, 30, 33) wird eine sehr hohe Wiederfindungsrate erzielt: nach dem HTvS werden insgesamt 28 und nach dem SP- und XP-Docking 27 der 31 Liganden (30 un-spezifische DLK-Inhibitoren und ANP) erfasst. Im SP- und XP-Modus konnten lediglich vier Liganden – alles Staurosporinderivate – nicht gedockt werden: Staurosporin (**32**), CEP-701 (**20**), (+)K-252a (**21**) und CEP-1347 (**1**) (Strukturen s. Kap. 5.2.1, S. 152, Abb. 5-41).

Zu b) Docking in die IDs 6.1-6.11:

Das Crossdockingexperiment wurde für die 34 Liganden – drei Röntgenkristallstrukturliganden, die 30 un-spezifischen DLK-Inhibitoren und ANP – mit den 11 zusätzlich generierten, nur in der Auswahl der bei den Wassermolekülen variierenden Proteinkonformationen der **5CEO**-ID 6 durchgeführt (s. Anhang E, Tab. E 1, ID 6.1-6.11).

Tab. 5-30 – Dockingscores des jeweils bestbewerteten Ligandenzustands der Röntgenkristallstrukturliganden nach dem VSW in den 11 erstellten **5CEO**-Proteinkonformationen ID 6.1-6.11. 0 = keine Dockingpose erhalten. **5ceol** = **5CEO**-Ligand etc. f = falsche Pose; f* = nur die Substituenten R2 und R3 sind in ihrer Position vertauscht (d.h. Core 2 liegt gedreht vor, s. Abb. 5-64 C, S. 181). Gute Ergebnisse sind blau hinterlegt.

HTvS	6.1	6.2	6.3	6.4	6.5	6.6	6.7	6.8	6.9	6.10	6.11
5ceol	-3,975 f	-3,979 f	-4,016 f	0	0	-4,273 f	-4,251 f	-4,095 f	-4,171 f	0	0
5cepL	-9,924 f*	-9,781 f*	-9,956 f*	0	-9,654 f*	-9,579 f*	-9,470 f*	-9,454 f*	-9,990 f*	0	-10,130 f*
5ceqL	-9,901	-9,983	-9,661	-9,573	-9,091	-9,730	-9,714	-10,082	-9,278	-9,328	-9,360
SP	6.1	6.2	6.3	6.4	6.5	6.6	6.7	6.8	6.9	6.10	6.11
5ceol	-8,057	-10,159	-9,886	0	0	-10,298	-10,242	-10,169	-9,462	0	0
5cepL	-9,835	-9,929	-9,813	0	-10,120	-10,019	-9,817	-9,815	-10,029	0	-9,581
5ceqL	-10,071	-10,286	-9,978	-10,122	-9,706	-10,384	-10,379	-10,318	-9,836	-9,584	-9,683
XP	6.1	6.2	6.3	6.4	6.5	6.6	6.7	6.8	6.9	6.10	6.11
5ceol	-9,768	-10,010	-10,363	0	0	-10,311	-10,332	-10,296	-10,048	0	0
5cepL	-10,323 f*	-8,473	-10,360 f*	0	-9,769	-9,274	-9,706	-9,330	-10,213 f*	0	-9,935 f*
5ceqL	-10,245	-9,529 f*	-9,684 f*	-10,188	-10,111	-10,280	-10,222	-10,743	-9,246	-9,957	-9,056

Betrachtet man die nach Dockingscore besten Ergebnisse für die Röntgenkristallstrukturliganden alleine (s. Tab. 5-30), stellt man gute Ergebnisse für die IDs 6.1-6.3 sowie 6.6-6.9 fest (blau markiert). Die Ergebnisse des SP-Dockings sind deutlich besser als die des HTvS-Dockings, während die Posen im XP-Docking nicht mehr verbessert werden können. Hier liegen die Substituenten R2 und R3 bei einigen Posen sogar vertauscht vor (vgl. auch Abb. 5-64 C, S. 181). Ein Vorteil des XP-Dockings gegenüber dem SP-Docking wird nicht sichtbar.

Die Zahl der gedockten Liganden und der korrekten Posen pro Dockingschritt für die 30 unspezifischen DLK-Inhibitoren und ANP sind in Tab. 5-31 dargestellt. Auch hier ist die Zahl der gedockten Liganden im HTvS- und SP-Output nahezu gleich, während die Wiederfindung im XP reduziert ist, aber die Zahl korrekter Posen gleichbleibt oder steigt.

Letztendlich wurden die IDs 6.3 (Wassermoleküle b, c, d in der Bindetasche, vgl. Abb. 5-65, S. 185) und 6.11 (Wassermoleküle b und d in der Bindetasche, vgl. Abb. 5-65, S. 185) wegen des hohen Dockingvermögens im HTvS-, SP- und XP-Modus für die unspezifischen DLK-Inhibitoren und ANP ausgewählt.

Die anfänglichen 44 möglichen Proteinkonformationen wurden durch diese Vorversuche (a-c) auf acht geeignete Proteinkonformationen reduziert.

Tab. 5-31 – Ergebnisse des VSW der 30 unspezifischen DLK-Inhibitoren und ANP mit den **5CEO**-IDs 6.1-6.11. #Lig = Anzahl gedockter Liganden, #Korr = Anzahl gedockter Liganden in erwarteter Pose (korrekte Pose). Auswahl = grün umrandet

ID	HTvS		SP		XP	
	# Lig	# Korr	# Lig	# Korr	# Lig	# Korr
6.1	17	2	17	2	13	3
6.2	19	3	22	3	9	5
6.3	18	2	18	3	14	5
6.4	16	3	16	3	8	4
6.5	15	2	15	4	10	5
6.6	16	3	16	3	13	4
6.7	18	2	15	3	12	4
6.8	15	3	15	3	8	4
6.9	17	4	17	4	13	5
6.10	15	2	13	4	10	5
6.11	18	2	18	4	15	6

Mit diesen acht Vertretern (s. Tab. 5-32) wurden im Folgenden Virtual-Screening-Workflows an definierten Testdatensätzen (Aktive plus Decoys) durchgeführt, um zusätzliche Validierungsmetriken (Anreicherung von Aktiven; Korrelation zwischen pK_i -Wert und Dockingscore) bei der engeren Auswahl von geeigneten Proteinkonformationen für das Ensemble-Docking zu berücksichtigen.

Tab. 5-32 – Auswahl von acht Proteinkonformationen auf Basis von Re- und Crossdockingexperimenten der 5CEO-, 5CEP- und 5CEQ-Liganden sowie der unspezifischen DLK-Inhibitoren und ANP.

ID	Interne Bezeichnung	H ₂ O	Activation Loop	Vol [Å ³]	Dist. [Å] F192-G132	Dist. [Å] M190-A135
6.3	5CEO_opls3_minwat_bcd	mit allen vorhandenen Wassermolekülen minimiert, anschließend alle Wassermoleküle bis auf Wat 2, 3 und 4 entfernt	Nicht modelliert	806	15,43	13,32
7	5CEO_opls3_minwat_nw	wie ID 6.3 nach Entfernung aller Wassermoleküle		806	15,43	13,32
6.11	5CEO_opls3_minwat_bd	wie ID 6.3 ohne Wat 4		806	15,43	13,32
13	5CEP_opls3_ow	Minimierung nach Entfernung aller Wassermoleküle		717	15,35	12,67
18	5CEQ_opls3_ow	Minimierung nach Entfernung aller Wassermoleküle		725	15,26	12,80
26	5CEN_run01_fr2006_opls3_6A	mit Wassermolekülen im Radius von 6 Å in der Bindetasche minimiert, danach alle gelöscht	vorhanden	1128	12,10	16,99
30	5CEP_run02_fr1089_opls3_6A	mit Wassermolekülen im Radius von 6 Å in der Bindetasche minimiert, danach alle gelöscht	modelliert mit PRIME (s. Kap. 5.1.2)	1159	14,06	18,28
33	5CEQ_run02_fr2978_opls3_6A	mit Wassermolekülen im Radius von 6 Å in der Bindetasche minimiert, danach alle gelöscht		723	12,88	15,59

5.4.1.3 VSW mit Testdatensätzen aus bekannten DLK-Inhibitoren und Decoys

In die acht ausgewählten Proteinkonformationen (s. Tab. 5-32) wurden mithilfe des „VIRTUAL SCREENING WORKFLOW“-Tools der Software SCHRÖDINGER zwei Testdatensätze, bestehend aus bekannten DLK-Inhibitoren (im Folgenden: Aktive) und Decoys, in einer Größenordnung von jeweils etwa 10.000 Molekülen gedockt. Diese Untersuchungen zeigen, welche Proteinkonformationen und welche Screeningoptionen (s. Tab. 5-33) in der Lage sind, die Aktiven gegenüber den Decoys am besten anzureichern und die beste Korrelation zwischen den Dockingscore-Werten und den aus der Literatur bekannten pK_i - bzw. pK_d -Werten der Aktiven zu erzeugen. Dieses Wissen soll für die Durchführung des eigentlichen virtuellen Screenings der vorbereiteten ZINC-Datenbank (s. Kap. 5.4.2) genutzt werden.

Mit **Ligandenliste a (Lig. a)**, s. Kap. 5.2.2, S. 156 u. Kap. 7.11.2, S. 283), bestehend aus 237 Aktiven und 10.292 Decoys, wurde jeweils, nach OPLS3- und nach OPLS2005-LIGPREP (s. Kap. 7.11.1) in die acht ausgewählten Proteinkonformationen (s. S. 191, Tab. 5-32) im HTvS-Modus gedockt. Im HTvS-Modus war nur die Auswahl des OPLS2005-Kraftfelds möglich. Die besten 10.000 Posen des HTvS-Dockings wurden anschließend entweder unter Verwendung des OPLS3- oder des OPLS2005-Kraftfelds im SP-Modus gedockt (s. Tab. 5-33). Wiederum die besten 10.000 Posen aus dem SP-Docking wurden im OPLS3- bzw. OPLS2005-XP-Modus gedockt. Eine weitere Variation in den durchgeführten Virtual-Screening-Workflows war die unterschiedliche Vorgabe der Erfüllung der drei festgelegten Constraints (s. S. 186). So sollten entweder kein, mindestens ein oder mindestens zwei Constraints erfüllt, d.h. Wasserstoffbrückenbindungen zur Hinge-Region ausgebildet werden müssen.

Tab. 5-33 – Übersicht über die Einstellungen der durchgeführten Virtual-Screening-Workflows. Lig = Ligandenliste.

Workflow	Lig	LIGPREP	HTvS	SP	XP	# Constraints
OPLS2005-A-1	a	OPLS2005	OPLS2005	OPLS2005	OPLS2005	≥ 1
OPLS3-A-0		OPLS3	OPLS2005	OPLS3	OPLS3	0
OPLS3-A-1		OPLS3	OPLS2005	OPLS3	OPLS3	≥ 1
OPLS3-A-2		OPLS3	OPLS2005	OPLS3	OPLS3	≥ 2
OPLS3-B-1	b	OPLS3	OPLS2005	OPLS3	OPLS3	≥ 1

Als Vergleich und zur Überprüfung der Ergebnisse wurde der OPLS3-Workflow unter der Vorgabe, mindestens einen Constraint zu erfüllen, mit einer zweiten Ligandenliste – **Ligandenliste b (Lig. b)** – durchgeführt. Diese unterscheidet sich in der Zusammensetzung der Aktiven (s. Kap. 7.11.2) sowie in der Vorgehensweise des LIGPREPS von der **Lig. a** (s. Kap. 7.11.1, Tab. 7-7). Für **Lig. b** wurde im LIGPREP die Bildung aller Protonierungszustände und Tautomere der Decoys in dem definierten pH-Bereich zugelassen, während für die Decoys der **Lig. a** nur jeweils ein Protonierungszustand generiert wurde. **Lig. b** besteht (vor LIGPREP) aus 170 Aktiven und 12.387 Decoys.

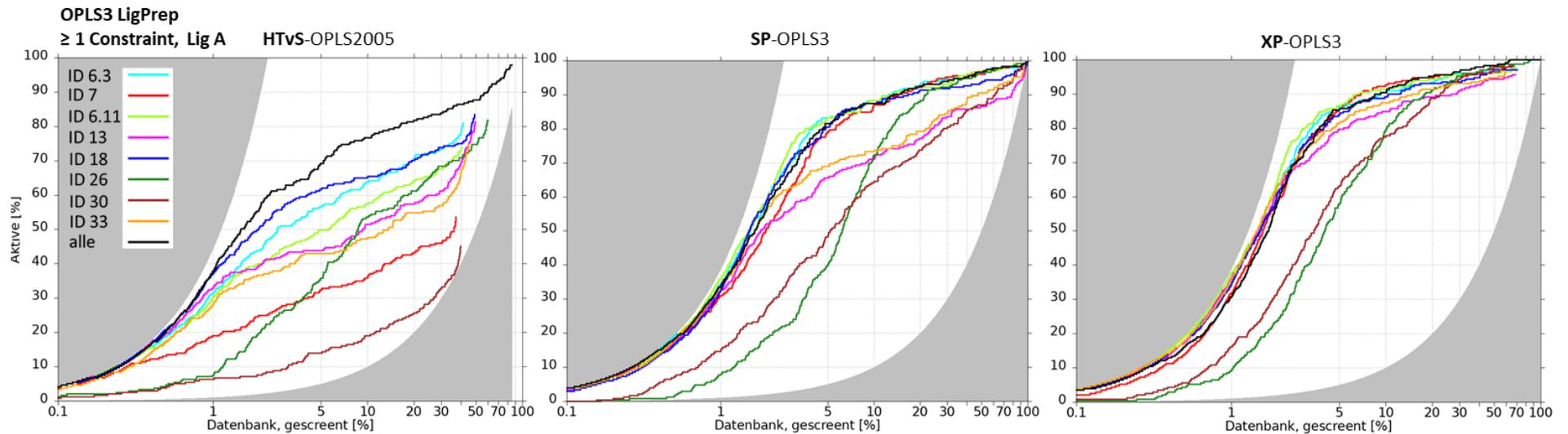


Abb. 5-66 – Enrichmentplots des OPLS3-A-1-Workflows für die drei verschiedenen Dockingmodi. Graue Kurve oben: perfekte Anreicherung; graue Kurve unten: „Anreicherung“ der Aktiven bei einer rein zufälligen Verteilung der Aktiven in der Input-Datenbank.

HTvS	Anreicherung EF x%				# gedockte Liganden		
	1%	5%	10%	50%	#A	#L	#A/#L [%]
alle	37,72	13,83	7,68	1,76	232	9006	2,576
ID 6.3	31,43	11,30	6,37	1,62	192	4377	4,387
ID 6.11	30,18	9,95	5,74	1,48	175	4244	4,123
ID 7	18,86	6,49	3,63	1,07	127	3906	3,251
ID 13	33,11	8,77	5,15	1,63	193	5243	3,681
ID 18	37,30	12,31	6,50	1,67	198	5173	3,828
ID 26	7,54	7,17	5,36	1,49	194	6285	3,087
ID 30	6,29	2,78	1,90	0,90	107	4180	2,560
ID 33	28,50	8,60	4,77	1,49	177	4744	3,731
6IDs	38,56	13,83	7,64	1,75	226	8091	2,793

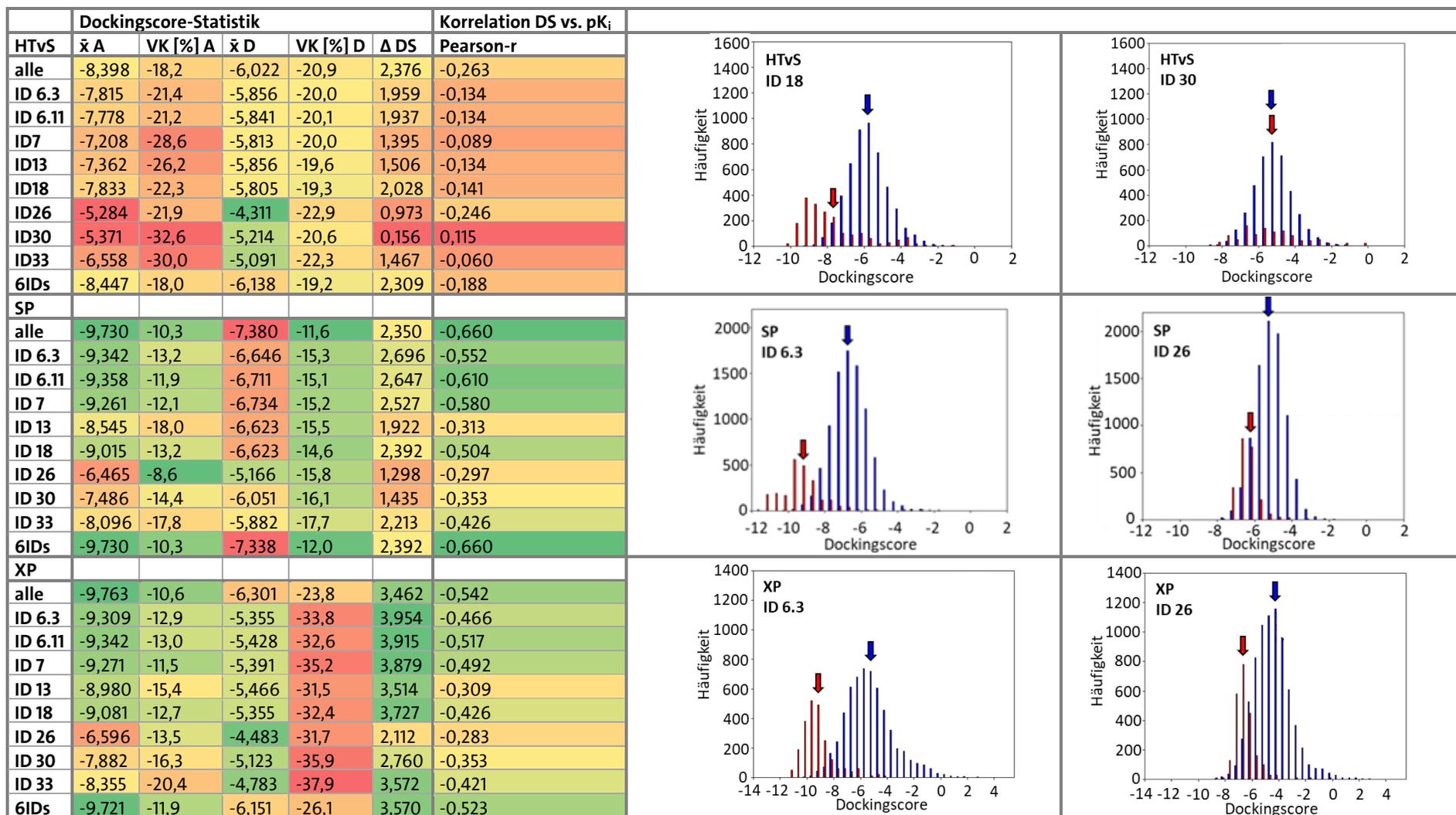
Tab. 5-34 – Enrichmentfaktoren und Anzahl gedockter Liganden des HTvS im OPLS3-A-1-Workflow. A = Aktive, L = Liganden (Decoys + Aktive). 6IDs: Merging aus ID 6.3, 6.11, 7, 13, 18 und 33. #A/#L (Input-Set) = 2,251 %.

SP	Anreicherung EF x%				# gedockte Liganden		
	1%	5%	10%	50%	#A	#L	#A/#L [%]
alle	34,13	16,27	8,75	1,93	231	9002	2,566
ID 6.3	34,55	16,61	8,83	1,93	231	8846	2,611
ID 6.11	36,26	16,61	8,83	1,94	231	8825	2,618
ID 7	30,71	15,84	8,70	1,92	231	8821	2,619
ID 13	32,42	13,08	7,11	1,73	231	8891	2,598
ID 18	32,85	16,10	8,75	1,87	231	8915	2,591
ID 26	8,11	8,00	7,24	1,93	231	8990	2,570
ID 30	15,36	9,73	6,46	1,74	231	8861	2,607
ID 33	34,13	13,77	7,32	1,81	230	8828	2,605
6IDs	34,55	16,27	8,75	1,94	231	8985	2,571

Tab. 5-35 – Enrichmentfaktoren und Anzahl gedockter Liganden des SP-Dockings im OPLS3-A-1-Workflow. A = Aktive, L = Liganden (Decoys + Aktive). 6IDs: Merging aus ID 6.3, 6.11, 7, 13, 18 und 33.

XP	Anreicherung EF x%				# gedockte Liganden		
	1%	5%	10%	50%	#A	#L	#A/#L [%]
alle	30,83	17,11	9,04	1,97	231	8876	2,603
ID 6.3	35,97	17,20	9,04	1,95	226	6145	3,678
ID 6.11	37,68	17,28	9,13	1,92	225	6214	3,621
ID 7	31,69	17,11	9,26	1,96	226	5848	3,865
ID 13	36,40	15,90	8,48	1,89	221	6177	3,578
ID 18	34,69	16,76	8,91	1,94	224	6331	3,538
ID 26	9,42	11,58	7,96	1,97	231	7824	2,952
ID 30	16,27	12,62	7,79	1,93	227	6054	3,750
ID 33	35,97	16,33	8,74	1,90	223	5299	4,208
6IDs	32,55	17,11	9,21	1,97	231	8392	2,753

Tab. 5-36 – Enrichmentfaktoren und Anzahl gedockter Liganden des XP-Dockings im OPLS3-A-1-Workflow. A = Aktive, L = Liganden (Decoys + Aktive). 6IDs: Merging aus ID 6.3, 6.11, 7, 13, 18 und 33.



Tab. 5-37 – Mittelwert und Variationskoeffizient der Dockingscores der Aktiven (A) bzw. der Decoys (D) sowie die Differenz Δ DS = \bar{x} D - \bar{x} A aus dem OPLS3-A-1-Workflow. Pearson-Korrelation zwischen Dockingscore und pK_i -Wert.

Abb. 5-67 – Häufigkeitsverteilung der Dockingscores der Decoys (blau) und der Aktiven (rot, 10 fach vergrößert). Oben: HTVs, Mitte: SP, unten: XP. Es sind jeweils die Ergebnisse der Protein-konformation mit der besten Trennung der Aktiven von den Decoys (links) und der geringsten Trennleistung (rechts) dargestellt. Roter Pfeil: Dockingscore-Mittelwert der Aktiven; blauer Pfeil: Dockingscore-Mittelwert der Decoys.

Die Auswertung der Dockingergebnisse nach jedem Einzelnen der drei hierarchischen Docking-schritte *HTvS-SP-XP* erfolgte nach Zusammenführung der Dockingergebnisse verschiedener States eines Liganden (*Multiple States*) aus einer Proteinkonformation zugunsten der besten Pose (*Single State*) nach Sortierung anhand des Dockingscores mit dem *Glide Ensemble Merge* der SCHRÖDINGER-Software. Diese *Merging*-Methode wurde für jede der acht Proteinkonformationen einzeln (ID 6.3, ID 6.11, ID 7 ...) sowie für alle Proteinkonformationen („alle“) durchgeführt. Desweiteren wurden die Dockingposen aus sechs der acht Proteinkonformationen auf diese Weise kombiniert. Hierbei handelt es sich um IDs, die besonders gute Einzeldockingergebnisse lieferten: ID 6.3, 7, 6.11, 13, 18 und 33 (zusammengefasst bezeichnet als „6IDs“).

Eine Analyse der Ergebnisse des *OPLS3-A-1-Workflows* (s. S. 193, Abb. 5-66 u. Tab. 5-34 bis Tab. 5-36) ergab nämlich, dass in allen drei Dockingmodi zwei der acht Proteinkonformationen, nämlich ID 26 und 30, zu einer signifikant schlechteren Anreicherung der Aktiven führen, die Aktiven im Mittel schlechtere Dockingscores erhalten sowie eine generell geringere Korrelation zwischen pK_i - und Dockingscore-Wert zu verzeichnen ist als mit den übrigen sechs IDs. Sowohl mit den einzelnen sechs IDs 6.3, 6.11, 7, 13, 18 und 33 als auch mit der Kombination aus diesen („6 IDs“) können die Aktiven besonders gut früh angereichert werden. Auch die Trennung der Aktiven von den Decoys mithilfe des Dockingscores gelingt mit den „6 IDs“ besser als mit den IDs 26 und 30 (s. S. 194, Tab. 5-37 und Abb. 5-67).

Schaut man auf die Anreicherung der Aktiven im *OPLS3-A-1-Workflow* (s. S. 193, s. Abb. 5-66, Tab. 5-34 bis Tab. 5-36), so verbessert sich die Anreicherung durch Anwendung des SP-Modus im Vergleich zum HTvS-Modus insbesondere im Bereich von 5-10 % der gescreenten Datenbank (EF5%, EF10%). Mit dem XP-Modus kann die Anreicherung nicht mehr deutlich gesteigert werden.

Die Trennung von Aktiven und Decoys verbessert sich vom HTvS über SP bis zum XP-Modus. Die Aktiven haben insbesondere im SP- und XP-Modus im Mittel signifikant bessere Scorewerte als die Decoys. Die Korrelation zwischen den experimentellen pK_i -Werten der Aktiven aus der Literatur und dem Dockingscore (s. Kap. 4.4.8, S. 80) gelingt im SP-Modus am besten, während sie im HTvS-Modus im Prinzip nicht vorhanden ist (s. auch Abb. 5-68: geringste und stärkste Korrelation im Vergleich). In den IDs 6.3, 6.11, 7 und 18 ist der Dockingscore besonders gut dazu in der Lage, die Aktiven in die nach Aktivität absteigende Reihenfolge zu bringen.

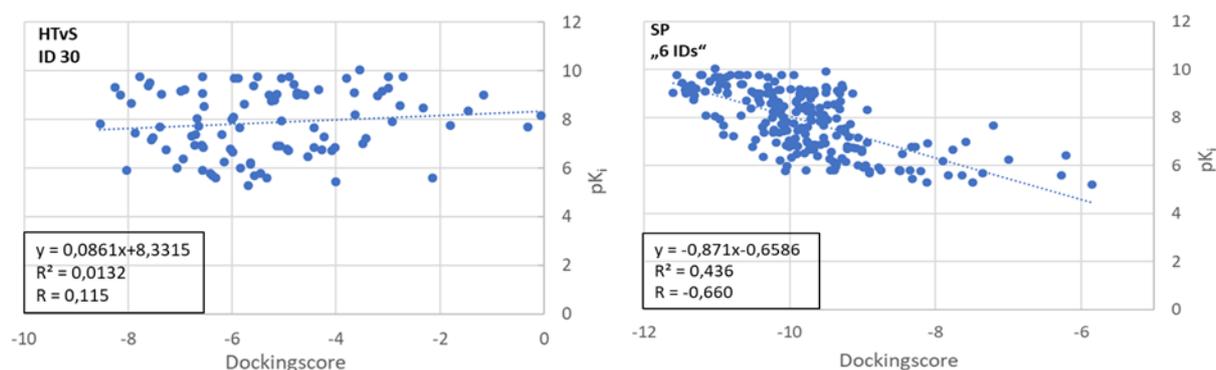


Abb. 5-68 – Korrelation zwischen Dockingscore und pK_i -Wert. Links: geringste/keine Korrelation (HTvS-Ergebnisse in ID 30, 97 Aktive mit definiertem pK_i - bzw. pK_d -Wert); rechts: stärkste Korrelation (SP-Ergebnisse der „6 IDs“, 219 Aktive mit definiertem pK_i - bzw. pK_d -Wert).

Die Durchführung des *OPLS2005-A-1-Workflows* sowie des *OPLS3-B-1-Workflows* (Ergebnisse s. Anhang F, S. 358 ff.) ergab nahezu identische Ergebnisse bezüglich der Anreicherung, Aufnahme-fähigkeit und Trennleistung der Proteinkonformationen im Vergleich zum *OPLS3-A-1-Workflow*.

Der Einsatz unterschiedlicher Constraints hat einen größeren Einfluss auf Anreicherung, Aktive-Decoy-Verhältnis und pK_i -Score-Korrelation als der Einsatz eines unterschiedlichen Kraftfelds oder einer in der Zusammensetzung veränderten Ligandenliste (s. Tab. 5-38 u. Anhang F, S. 358 ff.).

Unverändert blieben wiederum die im Allgemeinen schlechteren Kennzahlen in den IDs 26 und 30 im Vergleich zu allen anderen IDs. Die Qualität der Dockingergebnisse mit den IDs 26 und 30

kann also weder durch Veränderung des Kraftfelds, Veränderung der Ligandenzusammensetzung oder Veränderung der Constraints-Vorgaben auf das Niveau der übrigen IDs angehoben werden. Daher werden die Docking-ergebnisse aus diesen beiden IDs in Tab. 5-38 nicht mehr berücksichtigt, da sie im weiteren Verlauf der Arbeit nicht mehr weiter eingesetzt werden sollten.

Beim Docking ohne Constraints ist die frühe Anreicherung im SP- und XP-Modus im Vergleich zum Docking mit mindestens einem

Tab. 5-38 – Ergebnisse der verschiedenen Workflows mit den sechs ausgewählten IDs 6.3, 6.11, 7, 13, 18 und 33. EF 1% = Enrichmentfaktor 1%, #A/#L = Verhältnis gedockter Aktiver zu der Gesamtzahl der gedockten Liganden, ΔDS = Differenz des Dockingscoremittelwerts der Aktiven und des Dockingscoremittelwerts der Decoys. #A korr. = Anzahl korrekter Posen der Aktiven.

"6 IDs"	EF1%	#A/#L [%]	ΔDS	pearson-r	#A korr./#A
OPLS2005-A-1					
HTvS	39,40	2,816	2,451	-0,343	115/226 = 50,9 %
SP	33,88	2,558	2,395	-0,646	174/231 = 75,3 %
XP	33,15	2,736	3,630	-0,602	147/231 = 63,6 %
OPLS3-A-0					
HTvS	37,72	2,250	1,841	-0,288	106/234 = 45,3 %
SP	36,53	2,343	2,197	-0,634	178/230 = 77,4 %
XP	35,81	2,343	3,183	-0,559	143/230 = 62,2 %
OPLS3-A-1					
HTvS	38,56	2,793	2,309	-0,188	115/226 = 50,9 %
SP	34,55	2,571	2,392	-0,660	174/231 = 75,3 %
XP	32,55	2,753	3,570	-0,523	147/231 = 63,6 %
OPLS3-A-2					
HTvS	42,33	12,084	2,669	-0,265	111/196 = 56,6 %
SP	10,39	9,816	2,537	-0,547	165/203 = 81,3 %
XP	10,38	14,027	3,423	-0,440	126/201 = 62,7 %
OPLS3-B-1					
HTvS	44,57	1,555	2,273	-0,393	87/157 = 55,4 %
SP	38,86	1,880	2,095	-0,547	106/145 = 73,1 %
XP	36,75	1,959	3,025	-0,422	99/145 = 68,3 %

Constraint leicht erhöht, während der Anteil der gedockten Aktiven zur Gesamtzahl an gedockten Liganden sowie die Trennleistung zwischen Aktiven und Decoys leicht erniedrigt ist. Die pK_i -Score-Korrelation bleibt nahezu unverändert.

Beim Docking mit mindestens zwei Constraints ist die frühe Anreicherung im SP- und XP-Modus erheblich reduziert im Vergleich zu den anderen Constraints-Dockingmodi. Allerdings ist der erhebliche Vorteil des *OPLS3-A-2-Workflows*, dass die frühe Anreicherung im HTvS-Modus sehr hoch ist und schon in diesem ersten Schritt eine große Anzahl an Decoys aussortiert werden. Das Verhältnis gedockter Aktiver zur Gesamtzahl an gedockten Liganden ist stark erhöht im Vergleich zu den anderen Workflows.

Die Ergebnisse legen nahe, dass zum einen die Proteinkonformationen ID 26 und 30 eine unzureichende Qualität der Dockingergebnisse hervorbringen und daher für das Screening der ZINC-Datenbank nicht verwendet werden sollten. Zum anderen erschien es für den ersten Schritt, dem HTvS, sinnvoll, mit der Vorgabe zu docken, dass mindestens zwei der drei Wasserstoffbrückenbindungen zur Hinge-Region (Constraints) ausgebildet werden sollten, weil auf diese Weise in den hier durchgeführten Testdockings viele mutmaßlich inaktive Verbindungen aussortiert werden konnten. Dies ist aber nur unter der Voraussetzung sinnvoll, dass man den gesamten Output des HTvS dann für das SP-Docking und nicht nur die besten Dockingposen weiterverwendet, da ja die frühe Anreicherung von Aktiven unter Verwendung von mindestens zwei Constraints unzureichend ist. Im weiteren Verlauf erscheint es sinnvoll, im SP-Modus eine Workflow-Einstellung zu wählen, die eine hohe frühe Anreicherung hat und eine gute Korrelation zwischen pK_i -Wert und Dockingscore herstellt. Hier sollten daher die Einstellungen des *OPLS3-A-1-Workflows* verwendet werden.

Der Kosten-Nutzen-Aufwand des XP-Dockings wird als negativ bewertet. Zwar kommt es zu einer besseren Trennung der Aktiven von den Decoys, welches an der Differenz der Dockingscoremittelwerte der Aktiven und Decoys festgemacht werden kann. Allerdings ist der Variationskoeffizient der Decoy-Dockingscorewerte in allen durchgeführten Workflows im XP-Modus weitaus größer als im SP-Modus, d.h. die Decoys können sich dennoch in Hochscoringbereichen aufhalten. Darüber hinaus ist die frühe Anreicherung (EF1%) der zusammengefassten Dockingergebnisse (alle IDs bzw. 6 IDs) im SP-Modus höher als im XP-Modus. Ebenso ist die Korrelation zwischen pK_i -Wert und Dockingscore im SP-Modus etwas besser, sodass die Wahrscheinlichkeit höher ist, mit einem guten Scorewert auch eine aktivere Verbindung zu finden. Desweiteren ist die Zahl korrekter Posen mit dem SP-Modus immer höher als mit dem XP-Modus (s. S. 196, Tab.

5-38). Ein wesentlicher Nachteil des XP-Dockings ist die 8-10 Mal längere Berechnungszeit im Vergleich zum SP-Modus: für eine kleine Ligandenzahl ist dies sicherlich zu vernachlässigen (13 sek. vs. 108 sek. für ein Dockingbeispiel eines Liganden in dieser Arbeit). Für 100.000 Liganden würde der zeitliche Unterschied allerdings schon Monate betragen. Der XP-Modus sollte daher beim ZINC-Screening nicht verwendet werden; es sollte stattdessen nach Alternativen gesucht werden, die Anreicherung der Aktiven nach dem SP-Docking weiter zu verbessern (s. Kap. 5.5).

5.4.2 ZINC-Screening

Der eigentliche Screening-Prozess der vorbereiteten ZINC-Datenbank wird im Folgenden geschildert. Die Auswahl der Proteinkonformationen und Constraints erfolgte auf Basis der Ergebnisse der Vorversuche (s. Kap. 5.4.1). Da die gewählte Screening-Abfolge HTvS (mind. 2 Constraints) und anschließend SP (mind. 1 Constraint) so vorher nicht durchgeführt wurde, wurde die **Ligandenliste c (Lig. c)**⁵⁸, bestehend aus 216 Aktiven und etwa 16.000 Decoys (s. Kap. 5.2.2 u. 7.11.2) zur Steuerung des Gesamtprozesses mitgeführt. Ziel des Screenings ist eine große Reduktion der Decoys und eine möglichst geringe Aussortierung der Aktiven. Diese sollen möglichst früh angereichert, d.h. möglichst gut bewertet werden und in der Scoringliste an vorderen Plätzen erscheinen, sowie mithilfe des Dockingscores signifikant von den Decoys getrennt werden. Korrekte Posen der Aktiven wären wünschenswert und würden auf eine gute Dockingqualität schließen lassen.

Proteinpräparation und Grid-Generierung

Als Zielstrukturen für das Docking wurden die, wie in Kap. 5.4.1 beschrieben, ausgewählten sechs Proteinkonformationen der DLK eingesetzt. Die Grid-Generierung erfolgte wie in Kap. 5.4.1.2 beschrieben. Die Proteinstrukturen berücksichtigen unterschiedlich große Bindetaschen mit und ohne Wassermoleküle. Darüber hinaus ist die Verwendung der kinasetypischen Constraints der Hinge-Region (vgl. Kap. 5.2.1) möglich.

Ligandenpräparation

Als Screeningsubstanzen dienten die in Kapitel 5.2.3 beschriebenen 3.912.599 ZINC-Strukturen⁵⁹ (*Single States*) mit den gewünschten Eigenschaften (tendenziell *lead-like*-Charakter, mindestens eine Donorfunktion als Kinaseinhibitor-typische Eigenschaft). Es wurde eine Ligandenpräparation im OPLS2005-Kraftfeld durchgeführt (s. Kap. 7.11.1, Tab. 7-7, „ZINC-Strukturen“). Dies führte

⁵⁸ Es wurde in dieser Arbeit in der Auswahl der Ligandenlisten (**a**, **b**, **c**) variiert, um die Entscheidungen nicht nur von einem einzigen Datensatz abhängig zu machen.

⁵⁹ gezählt mit MONA

zu 5.941.050 Strukturen⁶⁰ (*Multiple States*). Zu diesem vorbereiteten Datensatz wurden die 216 Aktiven (*Single States*) der **Lig. c**, nach dem gleichen OPLS2005-LIGPREP, hinzugefügt (394 *Multiple States*).

Ebenso wurde die **Lig. c**, bestehend aus 216 Aktiven und 15.851 Decoys (*Single States*) vorbereitet (s. Kap. 7.11.1, Tab. 7-7, „*Lig. c (Aktive + Decoys)*“), sodass 394 Aktive und 29.347 Decoys (*Multiple States*) generiert wurden. Diese Liste wurde zur Inprozess-Kontrolle parallel im verwendeten Workflow mitgeführt.

Durchführung des Screenings und Ergebnisse

a) HTvS

Die vorbereiteten ZINC-Strukturen und Aktive sowie die vorbereitete **Lig. c** wurden jeweils im HTvS-Modus in die sechs ausgewählten Proteinkonformationen unter Anwendung des OPLS2005-Kraftfelds und unter der Voraussetzung gedockt, dass mindestens zwei der drei Constraints erfüllt werden (s. Abb. 5-69).

Nach Durchführung des HTvS-Dockings wurden die Dockingposen des Outputs aus den sechs verschiedenen Proteinkonformationen mithilfe des *Glide Ensemble Merge* zu einer gemeinsamen Ligandenliste zusammengefügt, indem die unterschiedlichen States eines Liganden zugunsten des *Single States* mit dem besten Dockingscore eliminiert wurden (s. Kap. 7.11.2, S. 285). Die 811.661 Substanzen (davon 172 Aktive) wurden einem OPLS3-LIGPREP (s. Kap.

7.11.1, Tab. 7-7, „*ZINC / Lig. c aus GEM des HTvS*“) unterworfen, um für das SP-Docking wieder alle *Multiple States* bereitzustellen. Es ergaben sich 1.467.205 ZINC-Strukturen und 328 Aktive (*Multiple States*). Die Zusammenführung der HTvS-Dockingposen und das erneute LIGPREP wurde so

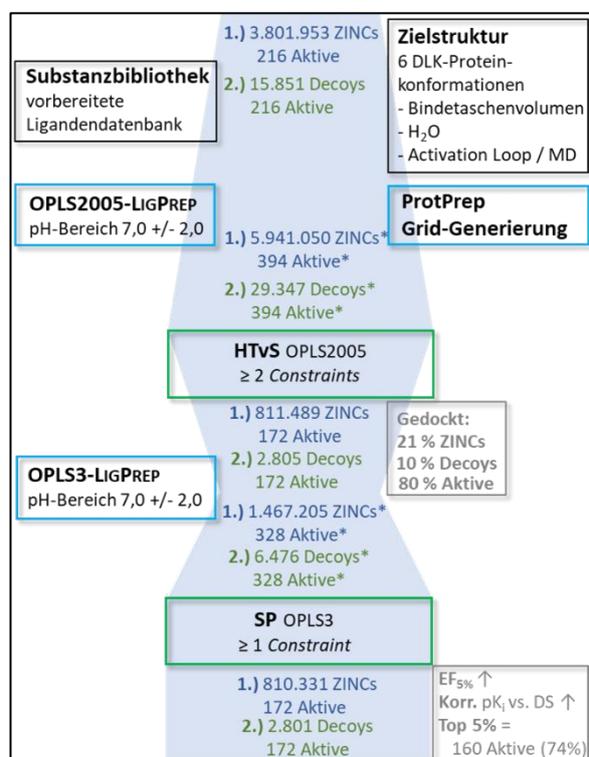


Abb. 5-69 – Durchführung des strukturbasierten virtuellen Screenings mit 1.) der ZINC- und Aktiven-Ligandenliste und mit 2.) der Decoy- und Aktiven-Ligandenliste. (*) markiert die Zahl der Strukturen nach LIGPREP (*Multiple States*). Ohne (*) = chemisch echt unterschiedliche Strukturen (*Single States*).

⁶⁰ Nach *Glide Ensemble Merge* mit SCHRÖDINGER zur erneuten Zählung nach LIGPREP waren es 3.801.953 *Single States*.

durchgeführt, da potenziell aktive Liganden nicht wegen der Berücksichtigung nur eines *Single States* im SP-Modus aussortiert werden sollten, während ein anderer *Single State* desselben Liganden, welcher das HTvS nicht überlebt hat, im SP-Modus erfolgreich gedockt werden könnte.

b) SP

Um das SP-Docking in die sechs Proteinkonformationen effizient durchzuführen, wurde dieser Datensatz in vier Subgruppen (s. Kap. 7.11.2, S. 285) aufgeteilt und die Dockingergebnisse nachträglich für jede Proteinkonformation wieder zusammengefügt. Das SP-Docking wurde aufgrund der oben beschriebenen Vorversuche mit der Vorgabe durchgeführt, dass mindestens eines der drei Constraints erfüllt werden muss.

Ergebnisse

Die Auswertung aller Dockingergebnisse (s. Abb. 5-70 u. Abb. 5-71, Tab. 5-39 bis Tab. 5-41, S. 201 f.) erfolgte, wie auch in den Testscreenings zuvor, nach Zusammenführung der verschiedenen Posen (*Multiple States* eines Liganden) aus einer Proteinkonformation zugunsten der besten Pose (*Single State*), bewertet anhand des Dockingscores.

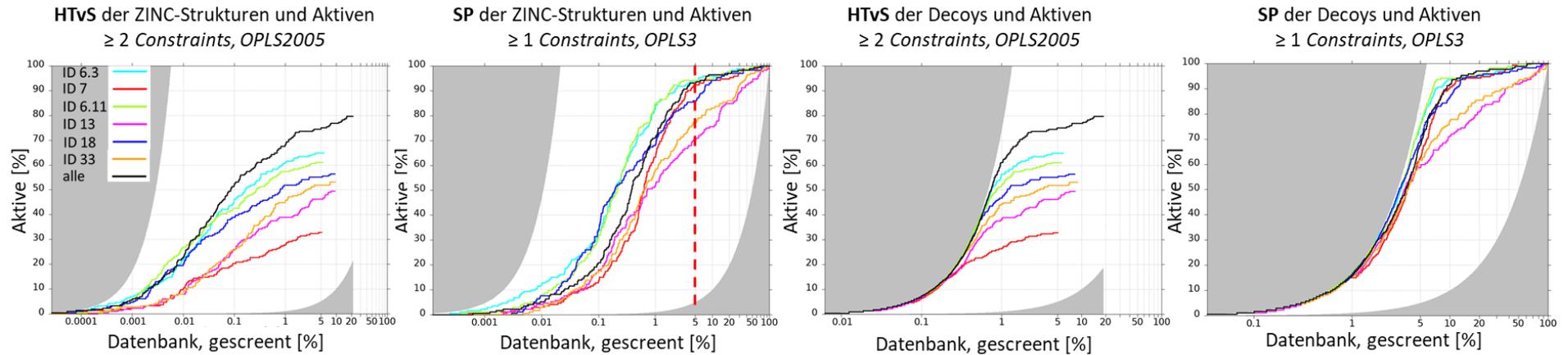


Abb. 5-70 – Enrichmentplots des ZINC-Actives-Screenings (HTvS und SP, links) sowie des Decoy-Actives-Screenings (HTvS und SP, rechts).

HTvS	Anreicherung EF x%				# gedockte Liganden			
	1 %	5 %	10 %	50 %	#A	#A korr	#L	#A/#L [%]
alle	68,06	14,91	7,69	1,59	172	123	811661	0,021
ID 6.3	61,11	12,96	6,48	1,30	140	87	216991	0,065
ID 6.11	57,41	12,22	6,11	1,22	132	86	207022	0,064
ID 7	27,78	6,57	3,29	0,66	71	34	198425	0,036
ID 13	38,89	9,26	4,95	0,99	107	67	356259	0,030
ID 18	51,85	11,02	5,65	1,13	122	78	355339	0,034
ID 33	44,91	10,37	5,32	1,06	115	76	364681	0,032
SP								
alle	71,51	18,60	9,65	1,98	172	134	810503	0,021
ID 6.3	83,13	18,72	9,59	1,98	172	131	809462	0,021
ID 6.11	84,88	18,84	9,48	1,99	172	134	809281	0,021
ID 7	63,37	18,49	9,42	1,98	172	132	808949	0,021
ID 13	51,74	14,07	7,56	1,87	172	130	809918	0,021
ID 18	68,60	17,21	9,42	1,97	172	125	810140	0,021
ID 33	58,14	15,47	8,31	1,94	172	133	809308	0,021

Tab. 5-39 – Enrichmentfaktoren (EF) und Anzahl gedockter Liganden (HTvS und SP) der ZINC- und Aktiven. A = Aktive, A korr = korrekte Posen der Aktiven, L = Liganden (ZINC+Aktive). Alle = Zusammenschluss aus allen Proteinkonformationen (Single States).

HTvS	Anreicherung EF x%				# gedockte Liganden			
	1 %	5 %	10 %	50 %	#A	#A korr	#L	#A/#L [%]
alle	60,98	14,99	7,68	1,59	172	124	2977	5,78
ID 6.3	56,36	12,95	6,48	1,30	140	87	948	14,77
ID 6.11	51,74	12,21	6,11	1,22	132	86	890	14,83
ID 7	26,33	6,57	3,29	0,66	71	34	812	8,74
ID 13	38,34	9,25	4,95	0,99	107	67	1334	8,02
ID 18	46,66	11,01	5,65	1,13	122	79	1320	9,24
ID 33	44,35	10,36	5,32	1,06	115	76	1423	8,08
SP								
alle	16,15	13,59	9,18	1,98	172	141	2973	5,79
ID 6.3	17,31	15,10	9,35	1,99	172	134	2966	5,80
ID 6.11	17,31	15,22	9,41	1,99	172	132	2964	5,80
ID 7	14,92	12,48	9,04	1,99	172	136	2961	5,81
ID 13	16,15	11,73	7,14	1,84	172	127	2965	5,80
ID 18	16,73	14,06	8,65	1,95	172	131	2970	5,79
ID 33	15,58	12,31	7,72	1,86	172	127	2971	5,79

Tab. 5-40 – Enrichmentfaktoren (EF) und Anzahl gedockter Liganden (HTvS und SP) der Decoys und Aktiven. A = Aktive, A korr = korrekte Posen der Aktiven, L = Liganden (Decoys+Aktive). Alle = Zusammenschluss aus allen Proteinkonformationen (Single States).

1.)	Dockingscore-Statistik ZINC-Actives-Screening					Korrelation DS vs. pK _i
HTvS	\bar{x} A	VK A [%]	\bar{x} Z	VK [%] Z	Δ DS	pearson
alle	-9,079	-12,672	-5,677	-28,323	3,402	-0,318
ID 6.3	-8,820	-11,677	-5,628	-28,214	3,192	-0,156
ID 6.11	-8,794	-12,521	-5,557	-28,545	3,237	-0,248
ID7	-8,421	-16,173	-5,483	-29,252	2,938	-0,146
ID13	-8,362	-15,107	-5,745	-27,162	2,617	-0,045
ID18	-8,804	-12,213	-5,657	-27,236	3,147	-0,210
ID33	-7,939	-13,733	-5,128	-30,884	2,812	-0,221
SP						
alle	-9,843	-7,087	-7,428	-12,840	2,416	-0,442
ID 6.3	-9,580	-8,224	-6,770	-15,383	2,810	-0,343
ID 6.11	-9,607	-7,780	-6,814	-15,149	2,793	-0,438
ID 7	-9,403	-8,405	-6,801	-15,825	2,603	-0,408
ID 13	-8,885	-12,749	-6,862	-14,968	2,024	-0,315
ID 18	-9,268	-8,741	-6,760	-14,804	2,509	-0,511
ID 33	-8,412	-12,176	-6,202	-16,794	2,210	-0,320

2.)	Dockingscore-Statistik Decoy-Actives-Screening					Korrelation DS vs. pK _i
HTvS	\bar{x} A	VK A [%]	\bar{x} D	VK [%] D	Δ DS	pearson
alle	-9,079	-12,672	-5,673	-27,679	3,406	-0,318
ID 6.3	-8,820	-11,677	-5,591	-27,821	3,229	-0,156
ID 6.11	-8,794	-12,521	-5,502	-29,153	3,292	-0,248
ID7	-8,421	-16,173	-5,455	-29,679	2,966	-0,146
ID13	-8,362	-15,107	-5,641	-27,944	2,722	-0,045
ID18	-8,804	-12,218	-5,629	-26,322	3,175	-0,210
ID33	-7,939	-13,733	-5,058	-30,796	2,881	-0,221
SP						
alle	-9,843	-7,107	-7,552	-11,688	2,291	-0,471
ID 6.3	-9,573	-8,694	-6,854	-14,749	2,719	-0,356
ID 6.11	-9,616	-7,605	-6,901	-14,454	2,716	-0,450
ID 7	-9,418	-8,177	-6,941	-14,849	2,477	-0,425
ID 13	-8,866	-12,904	-6,928	-14,373	1,937	-0,331
ID 18	-9,275	-8,997	-6,853	-13,931	2,422	-0,519
ID 33	-8,337	-13,891	-6,251	-16,447	2,086	-0,259

Tab. 5-41 – Dockingscorestatistik und Korrelation zwischen Dockingscore (DS) und pK_i-Wert der Aktiven des ZINC-Actives-Screenings (1.) oben und des Decoy-Actives-Screenings (2.) unten. A = Aktive, Z = ZINC, D = Decoys. Δ DS = \bar{x} D - \bar{x} A bzw. \bar{x} Z - \bar{x} A.

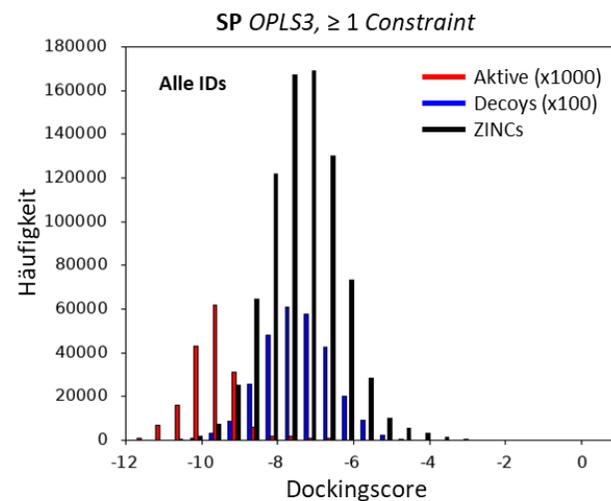
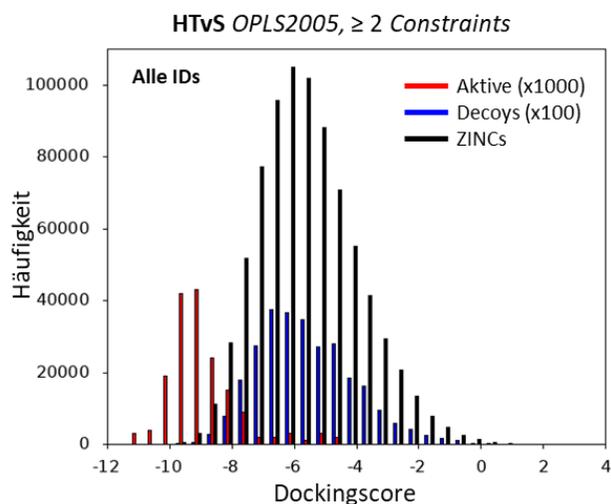


Abb. 5-71 – Häufigkeitsverteilung der Dockingscores der ZINC (schwarz) und Aktiven (rot) aus dem ZINC-Actives-Screening sowie der Decoys (blau) aus dem Decoy-Actives-Screening.

Bei Verwendung der **Ligandenliste c** wurden im HTvS 90 % der Decoys aussortiert (s. Abb. 5-70 u. Tab. 5-40). Dies bestätigt das Ergebnis des Vorversuchs mit **Ligandenliste a** (s. S. 193 f.), in dem ebenso eine hohe Anzahl an Decoys aussortiert werden konnte. Von den rund 3,8 Millionen eingesetzten ZINC-Strukturen (gezählt als *Single States*) wurden in diesem Schritt 2,99 Millionen ZINC-Strukturen (79 %) ausgesiebt, während von den 216 Aktiven noch 172 Aktive gedockt werden konnten (Verlust von lediglich 20 %, s. S. 199, Abb. 5-70 u. S. 201, Tab. 5-39). Es wurde eine gute Balance zwischen großer Reduktion der ZINC-Datenbank und einer möglichst großen übrig gebliebenen Anzahl an Aktiven erreicht. Der Scorewert der Posen spielt an dieser Stelle noch keine Rolle, da alle Dockingergebnisse weiterverwendet wurden. Es ist also nur entscheidend, ob eine Dockingpose erstellt wurde. Es fällt auf, dass im HTvS nicht jede Proteinkonformation jeden Liganden aufnehmen konnte (s. S. 201, Tab. 5-39 u. Tab. 5-40).

Das SP-Docking führte zu keiner weiteren Reduktion der Ligandenzahl (-0,14 % ZINC, -0,14 % Decoys). Positiv zu bewerten ist, dass keine Aktiven mehr in diesem Schritt verloren gingen. Die frühe Anreicherung der Aktiven im *Decoy-Actives-Screening*, insbesondere bei 1 % der gescreenteten Datenbank, kann als nahezu perfekt bewertet werden (s. S. 201, Abb. 5-70, rechts). Die Korrelation zwischen Dockingscore und pK_i - bzw. pK_d -Wert der Aktiven wird durch das SP-Docking im Vergleich zum HTvS erhöht (s. S. 202, Tab. 5-41). Im Gegensatz zum HTvS wurde im SP-Modus nahezu jeder Ligand in jede Proteinkonformation gedockt (s. S. 201, Tab. 5-39 u. Tab. 5-40). Dies hängt auch damit zusammen, dass lediglich mindestens 1 statt 2 Constraints erfüllt werden musste. Darüber hinaus ist die Anzahl korrekter Posen im SP-Modus größer als im HTvS-Modus. Der mittlere Dockingscore der Decoys ist im Vergleich zu dem der Aktiven deutlich niedriger (HTvS und SP, s. S. 202, Tab. 5-41), während er im Vergleich zu dem Dockingscore-Mittelwert der ZINC-Strukturen nur geringfügig kleiner ist. Sowohl im HTvS- wie auch im SP-Modus ist der Dockingscore-Mittelwert der ZINC-Strukturen deutlich niedriger als der Dockingscore-Mittelwert der Aktiven. Dieser Effekt ist zu beobachten, weil im ZINC-Datensatz viel mehr Inaktive als Aktive vorhanden sind.

Das SP-Docking führte im Mittel zu besseren Dockingscores der Aktiven, aber auch der ZINC-Strukturen und der Decoys. Dadurch ist die Trennung im SP-Modus zwischen Aktiven und Decoys im Vergleich zum HTvS-Modus etwas schlechter. Die Dockingscorewerte im SP-Modus streuen allerdings deutlich weniger stark um den jeweiligen Dockingscoremittelwert (s. Variationskoeffizient VK, S. 202, Tab. 5-41). Im Rahmen des Postprocessings soll die Trennung zwischen Aktiven und Decoys über das Scoring sowie die frühe Anreicherung der Aktiven weiter gesteigert werden.

5.5 Postprocessing

Ziel des Postprocessings ist die Verringerung der Anzahl der anzuschauenden Dockingposen für die Auswahl der zu testenden Substanzen. Die SP-Dockingposen aus dem virtuellen Screening sollten so bearbeitet werden, dass

- a) aktive Verbindungen möglichst weit vorne in der jeweiligen Rangliste platziert werden
- und b) der Zusammenschluss aller sechs Ranglisten auf einem angeglichenen Scoringniveau stattfinden kann (Normalisierung der Scorewerte, s. Kap. 4.4.7, S. 76 f.).

Dazu wurden zunächst weitere Scoringfunktionen neben dem SCHRÖDINGER Dockingscore mithilfe eines Testdatensatzes von Aktiven und Decoys hinsichtlich ihrer Fähigkeit untersucht, Aktive anzureichern und mit den DLK-Aktivitätsdaten zu korrelieren. Ausgewählte Scoringfunktionen mit akzeptablen Anreicherungs- und Korrelationswerten wurden anschließend eingesetzt, um den Output des *Decoy-Actives-Screenings* (s. Kap. 5.4.2) zu bewerten und die Anreicherung der Aktiven zu beurteilen. Der Einfluss von unterschiedlichen Kombinationen der Scorewerte auf die Anreicherung wurde untersucht und schließlich ein Consensuscore gebildet, welcher letztendlich für die Analyse des ZINC-Screenings eingesetzt wurde. Desweiteren wurden Fingerprint-Berechnungen und ein darauf basierendes Clustering nach Ligandenähnlichkeit durchgeführt und eine Hit-Auswahl aus den Clusterrepräsentanten vorgenommen.

5.5.1 Einfluss verschiedener Scoringfunktionen auf die Anreicherung und die Korrelation mit Aktivitätswerten

Die Bewertung mit weiteren Scoringfunktionen im Vergleich zum SCHRÖDINGER Dockingscore fand zunächst an den Dockingposen des Testworkflows *OPLS3-A-1* nach jedem einzelnen Dockingschritt (HTvS, SP, XP) statt (s. S. 192, Tab. 5-33). In diesem Workflow wurden noch acht statt sechs Proteinkonformationen eingesetzt. Die Auswertung im Folgenden erfolgte an den Dockingposen der letztendlich im ZINC-Screening verwendeten sechs Proteinkonformationen.

Nach jedem Docking erhält man den Docking- [284, 301] (s. Kap. 4.4.1, S. 68) und GlideScore [284, 301] (s. Kap. 4.4.1, S. 68, Formel (4)) für jede Pose. Die Posen in jeder der sechs Ranglisten wurden zusätzlich mit folgenden Methoden bzw. Scoringfunktionen zusätzlich neu bewertet:

- ChemScore [308] (s. Kap. 4.4.1, S. 67 u. Kap. 7.9.1)
- PLP-Score [327] (s. Kap. 4.4.4 u. Kap. 7.9.1)
- FlexX-Score [523] (s. Kap. 4.4.5 u. 7.9.1)
- Freie Bindungsenergie (PRIME MM/GBSA) [320, 321] (s. Kap. 4.4.3 u. Kap. 7.9.3)

Dann wurden die sechs Ranglisten nach dem entsprechenden Scorewert neu sortiert und alle Ranglisten so zusammengefügt, dass die Pose (*Single State*) unter allen States (*Multiple States*) mit dem besten Scorewert beibehalten wurde. Die entsprechende Korrelation zur bekannten Bioaktivität und Enrichmentfaktoren wurden berechnet.

Folgende Scoringmethoden wurden zusätzlich zur Bewertung der Posen aus dem SP-Docking verwendet:

- HYDE-Score [311–313] (s. Kap. 4.4.2 u. Kap. 7.9.2)
- Predicted pK_i des *SeI2*-QSAR-Modells [470] (s. Kap. 5.3.3 u. Kap. 7.9.4)
- IFP-Tanimotokoeffizient [434, 524, 525] (s. Kap. 4.7.1, S. 90 u. Kap. 7.9.5)

Bis auf den HYDE-Score konnten alle Scoringmethoden alle Posen der Aktiven und Decoys bewerten. Insbesondere besonders große Decoys konnten nicht bewertet werden und erhielten keinen Score (s. Abb. 5-72). Dies betraf nach Zusammenfassung der SP-Ergebnisse aus den sechs Proteinkonformationen aber lediglich 26 von 8.754 Decoys. Da es sich bei den Posen ohne HYDE-Score um sehr große Moleküle handelt, die zum Teil weit aus der Bindetasche herausragen⁶¹, und weil

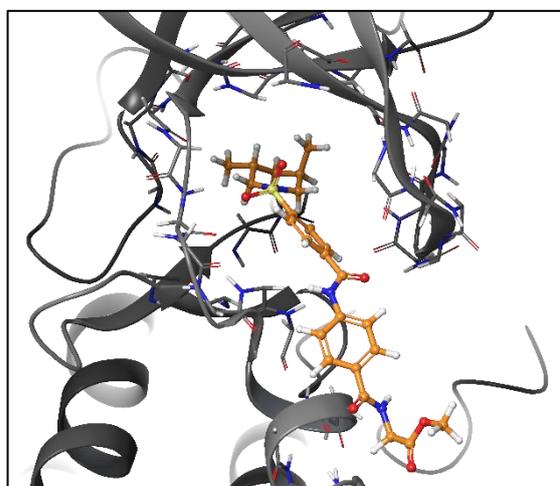


Abb. 5-72 – Beispiel eines Decoys, der mit HYDE nicht bewertet werden konnte (SP-Pose in ID 6.3).

alle Aktiven mit HYDE bewertet werden konnten, wurde dieser Tatsache keine weitere Bedeutung gegeben.

Die Ergebnisse des Rescorings (s. Tab. 5-42 u. Abb. 5-73, S. 206 f.) des Testdatensatzes lassen sich folgendermaßen zusammenfassen:

⁶¹ Für die Hits ist in dieser Arbeit ein geringes Molekulargewicht erwünscht, da sie durch Anhängen funktioneller Gruppen optimiert werden können.

- Die Anreicherung der Aktiven und die Korrelation der jeweiligen Scorewerte mit den literaturbekannten Aktivitätsdaten sind bei den SP-Dockingergebnissen besser als bei den HTVS- und XP-Ergebnissen. Dies bestätigte ein weiteres Mal, dass sich die Bewertung der riesigen Anzahl an ZINC-Dockingposen mit dem XP-Modus angesichts der unverhältnismäßig langen Berechnungsdauer nicht lohnen würde.
- Die in GLIDE verwendeten Docking- und GlideScore führen jeweils schon zu sehr guten Ergebnissen bezüglich der frühen Anreicherung und der Korrelation zu den experimentellen Aktivitätswerten im Vergleich zu den anderen Scorewerten.
- Bei dem PLP- und FlexX-Score sind die Anreicherung der Aktiven und die Korrelation mit den experimentellen pK_i -Werten im Vergleich zu den übrigen hier untersuchten Scoringmethoden schlechter.
- Mit der Anwendung des *predicted pK_i* des QSAR-Modells wird die beste Korrelation (Pearson-r = 0,676, SP-Modus) erhalten. Mit dem HYDE-Score erzielt man die beste frühe Anreicherung von Aktiven. Auch mit der PRIME MM/GBSA-Methode und dem IFP-Tanimotokoeffizienten erhält man akzeptable Ergebnisse. Docking-, Glide- und ChemScore liegen mit ihren Korrelationskoeffizienten eng beieinander, was auf die Ähnlichkeit ihrer zugrundeliegenden Scoringfunktionen zurückzuführen ist. Mit ChemScore wird allerdings eine geringere Anreicherung erzielt.

Tab. 5-42 – Enrichment- und Pearson-Korrelationsfaktor nach Anwendung verschiedener Scoringfunktionen auf die Dockingergebnisse des *OPLS3-A-1-Workflows*.

DS = Dockingscore, GS = GlideScore, CS = ChemScore, dG Bind = Freie Bindungsenergie (PRIME MM/GBSA), Pred. pK_i = vorhergesagter pK_i -Wert mit dem in dieser Arbeit erstellten QSAR-Modell (s. Kap. 5.3.3). IFP = IFP-Tanimotokoeffizient. Korrelationsfaktoren: pearson = Pearson-r. 6 IDs = Dockingergebnisse aus den sechs für das ZINC-Screening ausgewählten Proteinstruktur-IDs wurden vereint.

HTvS	EF1%	EF5%	EF10%	EF50%	pearson
DS	32,71	13,52	7,87	1,81	-0,188
GS	34,47	15,29	8,40	1,82	-0,202
CS	20,17	11,75	7,52	1,85	-0,353
PLP	9,65	6,72	4,60	1,81	-0,254
FlexX	8,33	9,37	6,95	1,81	-0,089
dG Bind	13,26	8,49	5,04	1,59	-0,344
SP	EF1%	EF5%	EF10%	EF50%	pearson
DS	34,57	16,34	8,78	1,95	-0,660
GS	35,44	16,77	8,91	1,97	-0,659
CS	27,01	12,54	7,53	1,99	-0,658
PLP	6,86	4,50	3,89	1,86	-0,322
FlexX	17,15	10,29	7,70	1,93	-0,252
dG Bind	19,45	8,64	5,58	1,69	-0,554
HYDE	37,49	15,32	8,22	1,81	-0,576
Pred. pK_i	20,31	8,56	5,62	1,71	0,676
IFP	31,12	13,83	7,74	1,84	0,581
XP	EF1%	EF5%	EF10%	EF50%	pearson
DS	30,27	17,13	9,17	1,97	-0,523
GS	32,87	17,56	9,30	1,98	-0,566
CS	22,31	11,75	7,40	1,99	-0,665
PLP	1,72	3,80	3,68	1,86	-0,258
FlexX	14,16	10,20	7,53	1,97	-0,218
dG Bind	17,30	7,70	5,67	1,64	-0,536

Basierend auf diesen Ergebnissen wurden der Dockingscore, die „Freie Bindungsenergie“ (PRIME MM/GBSA), der HYDE-Score, der mit dem QSAR-Modell vorhergesagte pK_i -Wert und der IFP-Tanimotokoeffizient für Scorekombinationsuntersuchungen und deren Einfluss auf die Anreicherung von Aktiven und die Korrelation zwischen Score und pK_i -Wert anhand der SP-Ergebnisse des *Decoy-Actives-Screenings* (Ligandenliste c, vgl. Kap. 5.4.2) ausgewählt. Der GlideScore wurde wegen seiner Ähnlichkeit zum Dockingscore und vergleichbarer Ergebnisse nicht weiterverwendet.

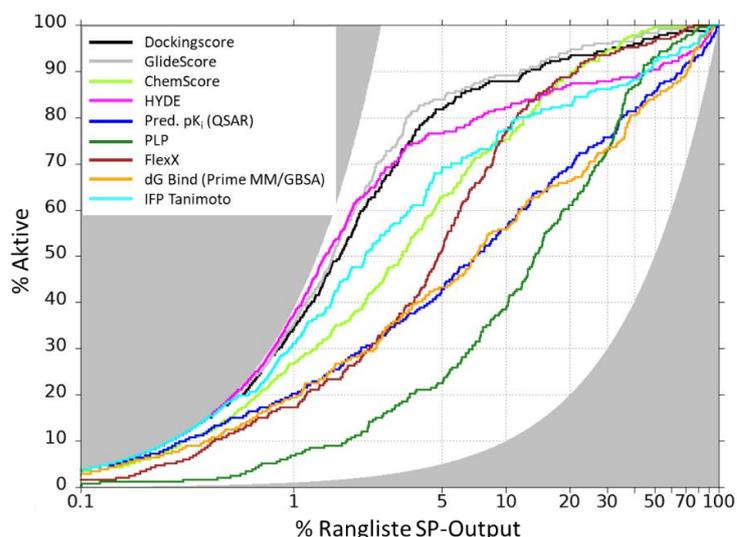


Abb. 5-73 – Enrichmentplot nach Anwendung verschiedener Scoringmethoden (nicht-normalisiert) auf die SP-Dockingergebnisse des *OPLS3-A-1-Workflows* (6IDs merged).

Der GlideScore wurde wegen seiner Ähnlichkeit zum Dockingscore und vergleichbarer Ergebnisse nicht weiterverwendet.

5.5.2 Normalisierung der Scorewerte und Kombination ausgewählter Scoringmethoden

Die Verteilung der Dockingscore-Werte unterscheidet sich zwischen den einzelnen Proteinkonformationen (s. Tab. 5-43). Eine Zusammenführung der Ranglisten auf Basis des „Roh-Dockingscores“ würde zu einer Liste führen, in der die besten Posen tendenziell aus ID 6.3, 6.11 und 7 stammen, da hier die Posen insgesamt besser bewertet werden als in ID 13, 18 und 33.

Tab. 5-43 – Dockingscorestatistik (*Decoy-Actives-Screening*, SP-Docking, vgl. Kap. 5.4.2)

ID	D max.	D min.	$\bar{x}(D)$	$s(D)$
6.3	-11,530	-2,482	-7,012	1,186
6.11	-11,504	-2,866	-7,058	1,171
7	-11,475	-1,290	-7,084	1,171
13	-10,712	-2,851	-7,041	1,103
18	-10,682	-2,720	-6,993	1,104
33	-10,043	-0,703	-6,372	1,145

Die Posen aus ID 13, 18 und 33 würde man nach Zusammenfassung benachteiligen. Eine Normalisierung der Rangpositionsnummern macht die Scorewerte aus verschiedenen Protein-IDs besser miteinander vergleichbar und die Benachteiligung von Posen aus „ungünstigen“ Bindetaschen wird aufgehoben.

Desweiteren kann die kombinierte Anwendung verschiedener Scoringmethoden die Anreicherung von Aktiven verbessern (s. Kap. 4.4.6). Alle SP-Posen (*Multiple States*) eines Liganden in einer ID-Rangliste (des *Decoy-Actives-Screenings* der Ligandenliste c, s. Kap. 5.4.2) wurden zunächst zugunsten der Pose (*Single State*) mit dem besten Dockingscore zusammengefasst, um nicht alle möglichen Liganden-States mit jeder Scoringmethode bewerten zu müssen. Diese sechs „Dockingscore-merge“-Listen wurden dann mit den ausgewählten Scoringmethoden (s. Kap. 5.5.1) neu bewertet:

- Dockingscore (**D**)
- HYDE-Score (**H**)
- Freie Bindungsenergie mit PRIME MM/GBSA (**M**)
- *Ligand Strain Energy* (**L**)
- IFP-Tanimotokoeffizient (**I**)
- Vorhergesagter pK_i-Wert mit dem QSAR-Modell (**Q**)

Der Parameter *Ligand Strain Energy* aus der PRIME MM/GBSA-Berechnung (s. Kap. 4.4.3) wurde hinzugenommen, da auffiel, dass die Aktiven besonders geringe Werte für diesen Parameter besaßen.

Der Zusammenschluss der sechs Ranglisten erfolgte – bei Vorhandensein gleicher Liganden (*single* oder *Multiple States*) in mindestens zwei Ranglisten – zugunsten der Pose (*Single State*) mit der besten Bewertung. Die so zusammengefasste Liste wurde nach dem entsprechenden Scorewert sortiert, der für den Zusammenschluss ausgewählt wurde. Folgende Varianten wurden durchgespielt:

- a - Zusammenschluss zugunsten des Liganden mit dem besten **Scorewert**, z.B. Dockingscore („**D**“)
- b - Zusammenschluss zugunsten des Liganden mit dem besten **normalisierten Scorewert**, z.B. normalisierter Dockingscore („**n_D**“, s. Kap. 4.4.7, S. 77, Formel (16) und Kap. 7.9.6)
- c - Zusammenschluss zugunsten des Liganden mit der niedrigsten **Rangpositionsnummer** einer nach einem Scorewert sortierten Liste (z.B. „**r_D**“ für Rangpositionsnummern einer nach Dockingscore sortierten Liste). Nachteilig an dieser Methode ist das Vorkommen der gleichen Rangpositionsnummer für verschiedene Liganden nach Zusammenschluss, in dem Fall erfolgt die Reihenfolge dieser Liganden „zufällig“.
- d - Zusammenschluss zugunsten des Liganden mit der niedrigsten **normalisierten Rangpositionsnummer** (z.B. „**nr_D**“ für normalisierte Rangpositionsnummern einer nach Dockingscore sortierten Liste). Auch hier können gleiche normalisierte Rangpositionsnummern für verschiedene Liganden vorkommen und die Reihenfolge dieser Liganden mit gleicher Bewertung erfolgt zufällig.
- e - Zusammenschluss zugunsten des Liganden mit der niedrigsten **normalisierten, mittleren Rangpositionsnummer** auf Basis der verschiedenen Scoringmethoden (z.B. „**nmr_DH**“ → kombinierter Score im Sinne des Consensuscores, s. Kap. 4.4.6, S. 75, Formel (15) und anschließender Normalisierung, s. Kap. 4.4.7, S. 77, Formel (16))

Die Anreicherung der Aktiven und die Korrelation der jeweiligen Scorewerte mit den experimentellen pK_i -Werten der Aktiven wurden nach Zusammenschluss der sechs Ranglisten bewertet.

Ergebnisse des Zusammenschlusses ohne Kombination verschiedener Scorewerte (a–d)

Die Verteilung des „Roh-Scorewerts“ unterscheidet sich, wie zuvor beschrieben (s. S. 207, Tab. 5-43), je nach Proteinkonformation. Für die Dockingergebnisse der Proteinkonformation ID 6.3 liegt der beste Dockingscorewert bei -11,530 und der schlechteste bei -2,482 ($\bar{x} = -7,012$), während die Werte für ID 33 bei -10,043 bzw. -0,703 ($\bar{x} = -6,372$) liegen. Durch Verwendung der normalisierten Rangpositionsnummern werden diese Unterschiede nivelliert: die beste normalisierte Rangpositionsnummer nach Dockingscore (**nr_D**) liegt sowohl für ID 6.3 als auch für ID 33 (sowie für alle anderen IDs) bei -1,731 und der schlechteste bei 1,731 ($\bar{x} = 0,000$). Die Bewertung mit der normalisierten Rangpositionsnummer (z.B. **nr_D**) führte somit zu einem vergleichbaren Score-Niveau aller Dockingposen in den verschiedenen Protein-IDs.

Unter den sechs verschiedenen Scoringmethoden lässt sich die höchste frühe Anreicherung der Aktiven mit dem Dockingscore (**D und n_D**) als Zusammenschlusskriterium erzielen (s. Tab. 5-44). Mit der *Ligand Strain Energy* ließ sich keine Anreicherung von Aktiven erreichen. So lässt sich eine Reihenfolge bezüglich der frühen Anreicherung von **D > H > I >> M > (=) Q > L** beobachten. Die Korrelation zwischen pK_i -Wert der Aktiven und dem jeweiligen Scorewert nimmt von **Q > H > D > I > M > L** ab. Weder die Normalisierung der Scorewerte (**n_D** etc.) noch die Verwendung der (normierten oder nicht-normierten) Rangpositionsnummern (**nr_D, r_D** etc.) führte zu nennenswerten Verbesserungen der Anreicherung bzw. Korrelationskoeffizienten.

Tab. 5-44 – Enrichmentfaktoren (EF) und Pearson-Korrelationskoeffizienten nach Zusammenschluss der sechs Ranglisten auf Grundlage des jeweiligen Scores. (n = normalisiert; r = Verwendung der Rangpositionsnr.).

	EF1%	EF5%	EF10%	EF50%	pearson
D	16,13	13,57	9,16	1,98	-0,471
H	14,98	11,60	7,60	1,94	-0,542
M	8,64	5,45	4,23	1,59	-0,434
L	0,00	0,58	1,39	1,41	-0,130
I	15,56	8,93	6,84	1,91	0,445
Q	8,64	3,71	3,36	1,62	0,599
n_D	16,13	13,22	9,05	1,96	-0,500
n_H	14,98	11,60	7,48	1,94	-0,551
n_M	9,22	5,34	4,12	1,60	-0,408
n_L	0,00	0,46	1,39	1,42	-0,124
n_I	16,13	9,74	6,79	1,91	0,459
n_Q	9,79	5,45	4,06	1,67	0,598
r_D	15,56	12,88	9,11	1,96	-0,388
r_H	13,83	11,72	7,37	1,94	-0,426
r_M	9,79	4,76	4,00	1,60	-0,401
r_L	0,00	0,70	1,45	1,39	-0,143
r_I	13,83	8,93	6,73	1,91	-0,413
r_Q	9,22	5,68	4,12	1,70	-0,508
nr_D	15,56	12,88	9,11	1,96	-0,388
nr_H	13,25	11,83	7,37	1,94	-0,424
nr_M	9,79	4,76	4,00	1,60	-0,401
nr_L	0,00	0,58	1,45	1,39	-0,143
nr_I	13,83	8,82	6,73	1,91	-0,413
nr_Q	8,64	5,22	4,12	1,67	-0,508

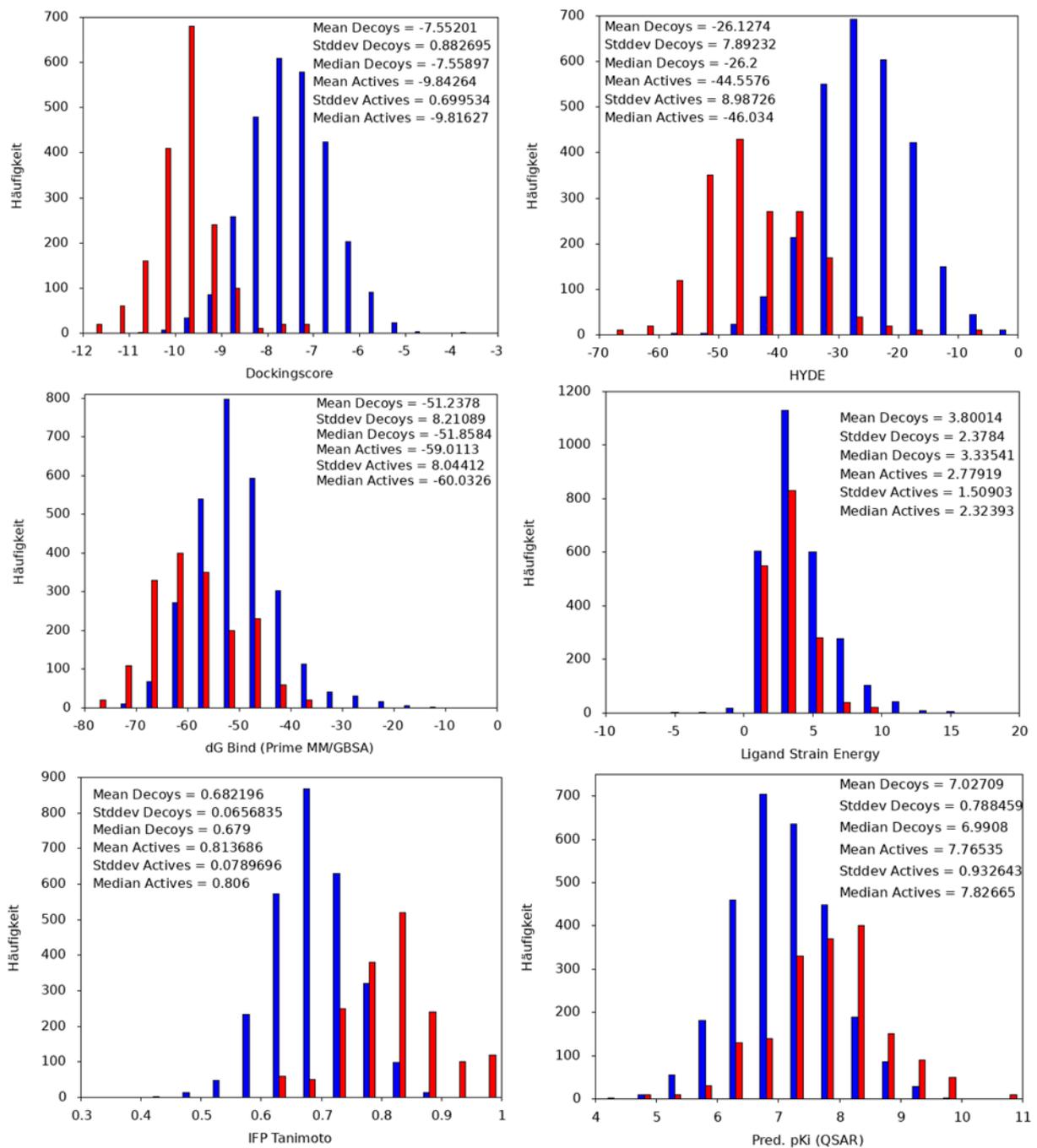


Abb. 5-74 – Häufigkeitsverteilung der Aktiven (rot, 10 fach vergrößert dargestellt) und Decoys (blau) nach Zusammenschluss der sechs Ranglisten und anschließender Sortierung der Liganden nach dem entsprechenden Score

Die Trennung zwischen Aktiven und Decoys in der zusammengefassten Liste ist besonders gut mithilfe von **D**, **H** und **I** möglich (s. Abb. 5-74). **Q** und **M** führen zu einer mäßigen Trennung, während **L** keine nennenswerte Unterscheidung ermöglicht.

Ergebnisse des Zusammenschlusses nach Kombination verschiedener Scorewerte (e)

Je mehr einzelne Scorewerte kombiniert wurden, desto höher wurde die Wahrscheinlichkeit, eine Verbesserung der Anreicherung und Korrelation zu erzielen (s. Tab. 5-45).

Tab. 5-45 – Enrichment- und Pearson-Korrelationsfaktoren nach Zusammenschluss der sechs Ranglisten auf Grundlage der normalisierten, mittleren Rangpositionsnummer (e). Die Kombination **nmr_DHMQ** wurde abschließend als Zusammenschluss- und Rangkriterium für das ZINC-Screening ausgewählt. Grau hinterlegt sind die Scorekombinationen, welche die *Ligand Strain Energy* (L) enthalten und tendenziell schlechtere Ergebnisse hervorrufen.

norm_mean_rank	EF1%	EF5%	EF10%	EF50%	Pearson-r
nmr_DH	17,28	14,50	9,05	1,96	-0,441
nmr_DM	16,13	10,09	6,44	1,87	-0,463
nmr_DL	16,13	11,02	6,79	1,91	-0,276
nmr_DI	16,71	14,15	8,87	1,96	-0,459
nmr_DQ	17,28	11,48	7,42	1,93	-0,582
nmr_HM	16,13	9,63	6,38	1,81	-0,503
nmr_HL	14,40	9,63	6,15	1,86	-0,314
nmr_HI	17,28	13,46	8,41	1,94	-0,504
nmr_HQ	15,56	11,60	7,08	1,84	-0,589
nmr_ML	10,37	5,80	4,12	1,51	-0,343
nmr_MI	17,28	9,98	6,67	1,85	-0,503
nmr_MQ	13,25	7,89	5,68	1,70	-0,541
nmr_LI	12,68	9,28	5,63	1,89	-0,353
nmr_LQ	12,68	7,54	5,16	1,69	-0,531
nmr_IQ	15,56	9,63	6,38	1,80	-0,560
nmr_DHM	17,28	12,18	7,42	1,96	-0,514
nmr_DHL	16,13	12,76	7,89	1,96	-0,381
nmr_DHI	17,28	15,20	9,16	1,96	-0,507
nmr_DHQ	17,28	13,57	8,35	1,94	-0,593
nmr_DML	16,71	10,67	6,50	1,87	-0,396
nmr_DMI	17,28	12,53	7,60	1,96	-0,526
nmr_DMQ	17,28	11,25	7,13	1,88	-0,571
nmr_DLI	16,71	12,53	8,12	1,96	-0,416
nmr_DLQ	16,71	11,72	7,31	1,98	-0,548
nmr_DIQ	16,71	12,99	7,77	1,95	-0,587
nmr_HML	16,13	10,44	6,21	1,80	-0,423
nmr_HMI	17,28	12,88	7,54	1,95	-0,563
nmr_HMQ	17,28	11,25	7,19	1,84	-0,589
nmr_HLI	16,71	11,83	7,54	1,96	-0,460
nmr_HLQ	17,28	11,02	7,19	1,88	-0,553
nmr_MIQ	16,71	10,67	7,02	1,78	-0,577
nmr_MLQ	16,13	9,98	5,97	1,67	-0,518
nmr_LIQ	16,13	10,09	6,50	1,87	-0,572
nmr_DHML	16,71	12,06	7,19	1,94	-0,451
nmr_DHMI	17,28	13,69	8,00	1,95	-0,558
nmr_DHMQ	17,28	12,53	7,83	1,95	-0,596
nmr_DHLI	16,71	13,46	8,87	1,96	-0,479
nmr_DHLQ	17,28	13,57	8,24	1,96	-0,568
nmr_DHIQ	16,71	13,92	8,24	1,95	-0,608
nmr_DMML	17,28	12,30	7,25	1,94	-0,473
nmr_DMLQ	16,71	11,72	7,02	1,87	-0,538
nmr_DMIQ	17,28	12,53	7,54	1,96	-0,594
nmr_HMLI	16,71	12,18	7,42	1,93	-0,503
nmr_HMLQ	17,28	11,83	6,96	1,85	-0,549
nmr_HMIQ	17,28	13,11	7,71	1,94	-0,622
nmr_MLIQ	17,28	11,14	6,73	1,84	-0,563
nmr_HLIQ	17,28	12,76	7,66	1,96	-0,608
nmr_DLIQ	17,28	12,76	8,00	1,99	-0,589
nmr_DHMLI	16,71	13,22	8,12	1,96	-0,512
nmr_DHMLQ	17,28	13,11	7,71	1,96	-0,561
nmr_DHMIQ	17,28	13,69	8,00	1,95	-0,619
nmr_DHLIQ	17,28	14,15	8,53	1,98	-0,610
nmr_DMMLIQ	17,28	12,64	7,83	1,98	-0,577
nmr_HMLIQ	17,28	12,88	7,66	1,95	-0,597
nmr_DHMLIQ	17,28	13,57	8,00	1,96	-0,600
„Perfekt“	17,28	17,28	17,28	17,28	-1

Allerdings ist auch hier zu beobachten, dass die *Ligand Strain Energy* (L) als Kombinationspartner die Ergebnisse verschlechtert (s. Tab. 5-45, grau hinterlegt). Die besten Ergebnisse der jeweiligen Kategorie (2er-, 3er-, 4er-, 5er- und 6er- Kombination) sind:

- DH (Anreicherung) bzw. DQ und HQ (Korrelation)
- DHQ
- DHMQ, DMIQ und HMIQ
- DHMIQ
- DHMLIQ (einzige 6er-Kombination, allerdings schlechter als DHMIQ)

Besonders gute Kombinationspartner sind demnach **D**, **H** und **Q**. Basierend auf diesen Ergebnissen folgte der Ausschluss der *Ligand Strain Energy* (L) als Scorewert für den Zusammenschluss der Ranglisten. Zwar ist dies ein nützlicher Parameter, da überaus hohe Werte für L (d.h. stark beanspruchte Liganden in der Bindetasche) ausgeschlossen werden können, allerdings spielt er

für die Anreicherung keine Rolle. Für die Hit-Auswahl kann er als zusätzliches Bewertungskriterium herangezogen werden.

Der IFP-Tanimotokoeffizient (**I**) hat einen sehr positiven Einfluss, sowohl auf die Anreicherung als auch auf die Korrelation zwischen pK_i -Wert der Aktiven und dem Scorewert (**I** bzw. kombinierter Score mit **I**). Als Referenzligand zur Berechnung des IFP-Tanimotokoeffizienten wurde der jeweilige Originalligand der Proteinkonformation verwendet (z.B. der **5CEQ**-Ligand für ID 6.3, 6.11 und 7, s. Kap. 7.9.5). Demnach werden die Interaktionen der Aktiven bzw. Decoys zum Rezeptor mit den Interaktionen des Referenzliganden zum Rezeptor verglichen. Da viele der hier eingesetzten Aktiven sehr ähnlich zu den Referenzliganden (**5CEO**-, **5CEP**- und **5CEQ**-Liganden) sind, ist eine hohe Übereinstimmung der Interaktionen sehr wahrscheinlich. Würde man den IFP-Tanimotokoeffizienten als Zusammenschluss- und Rangkriterium berücksichtigen, könnte dies ungewollt zu einer starken Anreicherung von Strukturen führen, die den bereits bekannten DLK-Inhibitoren sehr ähnlich sind. Interessante Strukturen mit neuen Interaktionsmustern würden eventuell benachteiligt werden. Deshalb wurde **I** nicht als Zusammenschluss- und Rangkriterium verwendet. Als Beurteilungskriterium soll er später bei der Hit-Auswahl herangezogen werden.

Da der Maximalwert von 17,28 für $EF_{1\%}$ mit vielen Scoringkombinationen erreicht werden konnte (s. Tab. 5-45), wurden zur weiteren Differenzierung die besonders guten Scoringkombinationen **nmr_DHQM**, **nmr_DH**, **nmr_DQ**, **nmr_HQ** und **nmr_DHQ** auf die jeweils besten 5 % der sechs Ranglisten des *ZINC-Actives-Screenings* (SP) angewendet und die Listen anschließend zusammengeführt. Die Verwendung der Zweier- bzw. Dreierkombinationen ergab hierbei keinen deutlichen Vorteil für die frühe Anreicherung gegenüber **nmr_DHQM**. Insbesondere in der sehr frühen Anreicherung ($EF_{0,5\%}$) schnitt **nmr_DHQM** am besten ab (s. Abb. 5-75).

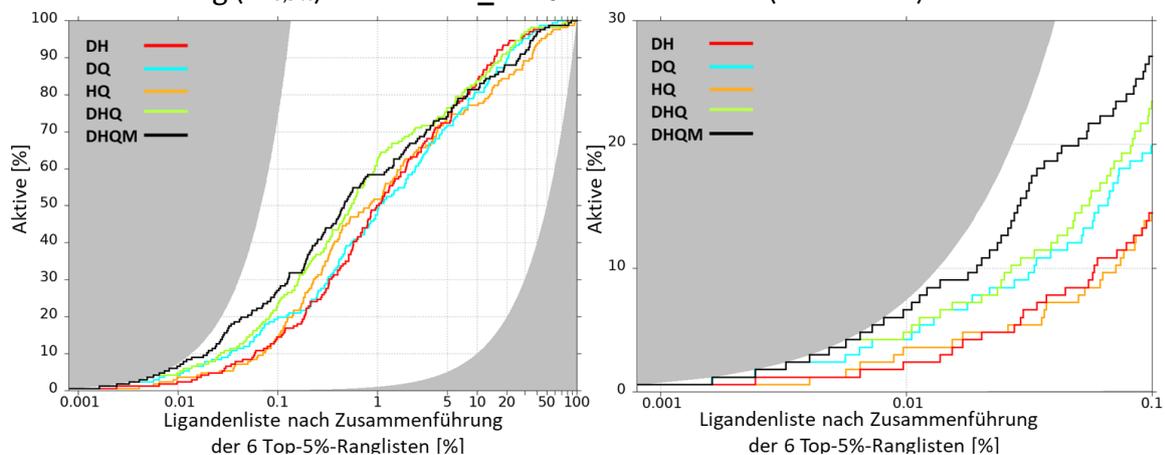


Abb. 5-75 – Links: Anreicherung der Aktiven nach Sortierung und Zusammenführung der 6 Top-5%-Ranglisten des *ZINC-Actives-Screenings* unter Beibehaltung der besten Pose und anschließender Sortierung unter Verwendung der kombinierten Scorewerte **nmr_DH**, **nmr_DQ**, **nmr_HQ**, **nmr_DHQ** und **nmr_DHQM**. Rechts: Zoom auf die sehr frühe Anreicherung bis 0,1 %.

Der kombinierte Score **nmr_DHMQ** wurde letztendlich als Consensuscore für das ZINC-Screening ausgewählt. Er zeichnet sich durch eine besonders hohe frühe Anreicherung (s. Abb. 5-76 A, schwarze Kurve), hohe Korrelationswerte sowie durch eine gute Trennung der Aktiven (Mittelwert Consensuscore = -2,09) von den Decoys (-0,61) aus (s. Abb. 5-76 B). Darüber hinaus besteht er aus vier voneinander unabhängigen Bewertungsmethoden, wodurch ein Bias in Richtung einer bestimmten Interaktionsbewertung reduziert werden kann.

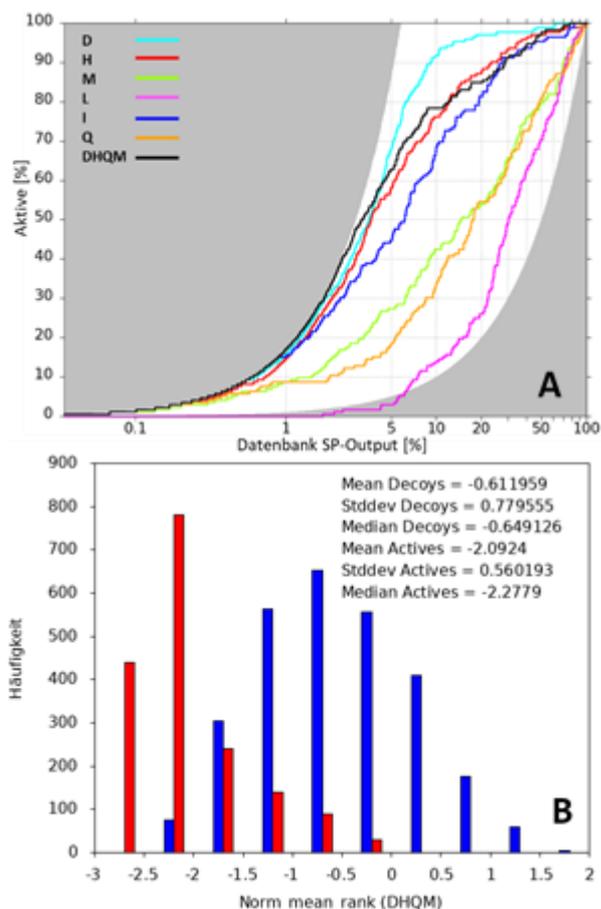


Abb. 5-76– A – Anreicherung der Aktiven nach Zusammenschluss der Ranglisten und anschließender Sortierung unter Verwendung der Einzelscorewerte (**D**, **H**, **M**, **L**, **I** und **Q**) sowie des kombinierten **nmr_DHQM**-Consensuscores. **B** – Häufigkeitsverteilung der Aktiven (rot, 10x vergrößert) und der Decoys (blau) nach Zusammenschluss der sechs Ranglisten und anschließender Sortierung der Liganden unter Verwendung des **nmr_DHQM**-Consensuscores.

Die Consensuscore-Verteilung der Posen in den einzelnen Proteinkonformationen ist nahezu gleich (s. Tab. 5-46).

Tab. 5-46 – Consensuscore-Statistik (*Decoy-Actives-Screening*, SP-Docking) in den einzelnen Proteinkonformationen

	ID 6.3	ID 6.11	ID 7	ID 13	ID 18	ID 33
nmr_DHMQ min.	-2,591	-2,573	-2,660	-2,642	-2,563	-2,632
nmr_DHMQ max.	2,378	2,327	2,505	2,408	2,488	2,378
$\bar{x}(\text{nmr_DHMQ})$	$1,124 \cdot 10^{-16}$	$-3,121 \cdot 10^{-15}$	$-9,569 \cdot 10^{-17}$	$-3,709 \cdot 10^{-15}$	$3,470 \cdot 10^{-15}$	$-6,413 \cdot 10^{-16}$
s(nmr_DHMQ)	1	1	1	1	1	1

5.5.3 Postprocessing des ZINC-Screenings

Die Anreicherungskurven, welche aus den SP-Dockingergebnissen der einzelnen Proteinkonformationen hervorgehen, zeigen einen steilen Anstieg der kumulierten Anzahl der gedockten Aktiven zwischen 0,1 und 10 % der gescreenten Datenbank (s. S. 201, Abb. 5-70). Die jeweils besten Posen (Top 1 %, 5 % und 10 %) jeder ID-Rangliste – unter Entfernung mehrfach vorkommender Liganden bzw. von *Multiple States* zugunsten der Pose (*Single State*) mit dem besten Dockingscore – wurden zusammengefügt (s. Tab. 5-47 und Abb. 5-77). Dies ergab 30.737 (Top 1 %), 123.889 (Top 5 %) bzw. 216.491 (Top 10 %) ZINC-Strukturen. Erfasst man die besten 5 % statt 1 % jeder ID-Rangliste, so werden insgesamt 166 statt 157 Aktive (+11) wiedergefunden. Berücksichtigt man die besten 10 % jeder ID-Rangliste, so werden 168 Aktive (+2 im Vergleich zu Top 5%) wiedergefunden. Der Zusammenschluss der jeweils besten 5 % der Dockingergebnisse aus den sechs Proteinkonformationen erschien daher als ein guter Kompromiss zwischen einer hohen Wiederfindungsrate von Aktiven (76,9 % der anfänglich 216 Aktiven) sowie einer gut zu verarbeitenden Ligandenzahl für die nachfolgenden Rescoring-Operationen (123.889 ZINC-Verbindungen = 3,26 % der Startmenge von 3,8 Millionen ZINC-Strukturen).

Tab. 5-47 – Anzahl gedockter Aktiver und ZINCs in den ersten 1%, 5% und 10% der Dockingergebnisse pro Proteinkonformation. Ensemble = Ergebnis des Zusammenschlusses der Top x%-Dockingergebnisse aus den sechs Proteinkonformationen. A = Aktive, Z = ZINCs

ZINC-Actives-Screening												
	Top 1 %		Top 5 %				Top 10 %					
	# Aktive	# ZINCs	# Aktive	# ZINCs	# Aktive	# ZINCs	# Aktive	# ZINCs				
ID 6.3	143	7952	161	40312	165	80781						
ID 6.11	146	7947	162	40302	163	80765						
ID 7	109	7980	159	40288	162	80733						
ID 13	89	8010	121	40375	130	80862						
ID 18	118	7983	148	40359	162	80852						
ID 33	100	7993	133	40332	143	80788						
Ensemble	157	30737	166	123889	168	216491						
Dockingscore-Statistik												
	\bar{x} A	VK A [%]	\bar{x} Z	VK Z [%]	\bar{x} A	VK A [%]	\bar{x} Z	VK Z [%]	\bar{x} A	VK A [%]	\bar{x} Z	VK Z [%]
Ensemble	-9,98	-5,36	-9,35	-4,37	-9,92	-5,76	-8,80	-5,22	-9,90	-6,03	-8,53	-5,88

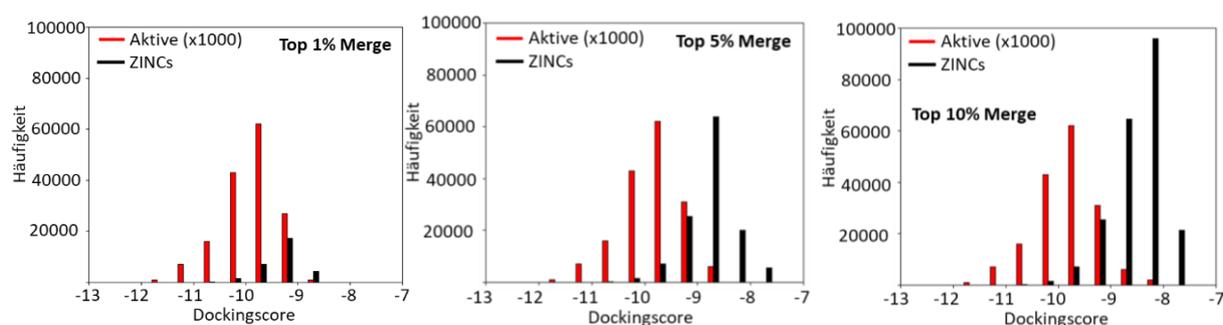


Abb. 5-77 – Häufigkeitsverteilung der Dockingscores der ZINCs (schwarz) und Aktiven (rot, 1000x vergrößert) aus dem ZINC-Actives-Screening nach Zusammenführung der besten 1 %, 5 % bzw. 10 % der Dockingergebnisse aus den sechs Proteinkonformationen

Die Top 5 % jeder Rangliste – im Folgenden als Top-5%-Rangliste bezeichnet – beinhalten etwa je 40.000 Substanzen (*Single State*, s. Tab. 5-48). Nach Zusammenführen zu einer Gesamtrangliste werden nicht, wie man erwarten könnte, etwa 240.000 Substanzen (6 x 40.000) erhalten, sondern lediglich 123.889 Substanzen. Das bedeutet, dass viele der Proteinkonformationen gleiche Liganden docken. Wenn man nur die Top 5 % der ID 6.3 berücksichtigen würde, würden etwa 80.000 andere Strukturen fehlen, da sie in ID 6.3 nicht gedockt bzw. nicht so gut bewertet wurden. Hier wird der Vorteil des Ensemble-Dockings deutlich, da die bestbewerteten Substanzen aus jeder Protein-ID berücksichtigt werden und die Diversität der Strukturen erhöht wird, indem mehrfach vorkommende Strukturen eliminiert werden.

Für die Posen der sechs Top-5%-Ranglisten wurden nun zusätzlich zum Dockingscore der HYDE-Score („H“), die Freie Bindungsenergie (PRIME MM/GBSA, „M“) sowie der vorhergesagte pK_i-Wert mit dem *Sel2*-QSAR-Modell (s. Kap. 5.3.3, „Q“) berechnet. Mithilfe dieser vier Scorewerte wurde der in Kap. 5.5.2 ausgewählte Consensuscore **nmr_DHQM** berechnet. Die sechs Top-5%-Ranglisten wurden zu einer Liste zusammengeführt. Hier wurde nur der Ligand (*Single State*) aus einer der sechs Proteinkonformationen mit dem besten Consensuscore beibehalten und die so zusammengefasste Liste nach dem Consensuscore sortiert. Zum Vergleich mit dem **nmr_DHQM** wurden die sechs Top-5%-Ranglisten – nach gleichem Prinzip – jeweils nach einem der vier Einzelscores (**D, H, Q, M**) sortiert und zusammengeführt.

259 ZINC-Strukturen (*Single State*) hatten nach Zusammenführung der sechs Ranglisten keinen HYDE-Score. Jeder Aktive erhielt in mindestens einer der sechs Top-5%-Ranglisten einen HYDE-Score. Die ZINC-Strukturen ohne HYDE-Score wurden bei der Berechnung des Consensuscores nicht berücksichtigt. Die Auswertung bezieht sich bei Verwendung des HYDE- und des Consensuscores daher immer auf 123.630 statt auf 123.889 ZINC-Strukturen (-259, *Single State*).

Tab. 5-48 – Übersicht über Anzahl der gedockten ZINC-Strukturen (Z) und Aktiven (A) (*Single State*)

ZINC-Actives-Screening Top 5 %		
ID	# Z	# A
6.3	40312	161
6.11	40302	162
7	40288	159
13	40375	121
18	40359	162
33	40332	133
Ensemble	123889	166

Die Verwendung des Consensuscores **nmr_DHQM** führt zu einer deutlichen Verbesserung der frühen Anreicherung der Aktiven in jeder der sechs Proteinkonformationen (s. Tab. 5-49 u. Abb. 5-78) im Vergleich zu den Einzelscorewerten (**D**, **H**, **Q**, **M**, s. Anhang G, S. 366, Tab. G 1 u. Abb. G 1). Die Zusammenführung der sechs Ranglisten unter Verwendung des Consensuscores zum Proteinensemble führt zu einer nochmaligen Steigerung der frühen Anreicherung (EF1% = 58,43). Selbst gegenüber den schon sehr guten Einzelscorewerten für Dockingscore und HYDE-Score konnte die frühe Anreicherung (EF1%) etwa um den Faktor 2 gesteigert werden. Da bei der Substanzauswahl für die Testung nicht hunderttausende Substanzen angeschaut und verglichen werden können, ist diese nahezu perfekte Anreicherung von Aktiven in der frühen Phase besonders wertvoll. Die Mehrzahl der Aktiven (91 von 166) besitzt einen Consensuscore besser als -2,5 (s. Abb. 5-79). Nur 696 der 123.796 ZINC-Strukturen (0,56 %) liegen in diesem Bereich.

Tab. 5-49 – Enrichment- und Pearson-Korrelationsfaktoren nach Sortierung und Zusammenführung der 6 Top-5%-Ranglisten des ZINC-Screenings (SP) nach Consensuscore (**nmr_DHQM**)

nmr_DHQM	EF1%	EF5%	EF10%	EF50%	Pearson-r
ID 6.3	44,04	13,66	7,76	1,94	-0,419
ID 6.11	43,76	13,95	7,96	1,94	-0,525
ID 7	32,64	11,19	7,17	1,87	-0,496
ID 13	31,09	11,93	6,89	1,78	-0,300
ID 18	44,79	14,82	7,96	1,82	-0,465
ID 33	38,55	13,03	7,80	1,94	-0,439
alle	58,43	14,82	8,13	1,96	-0,529

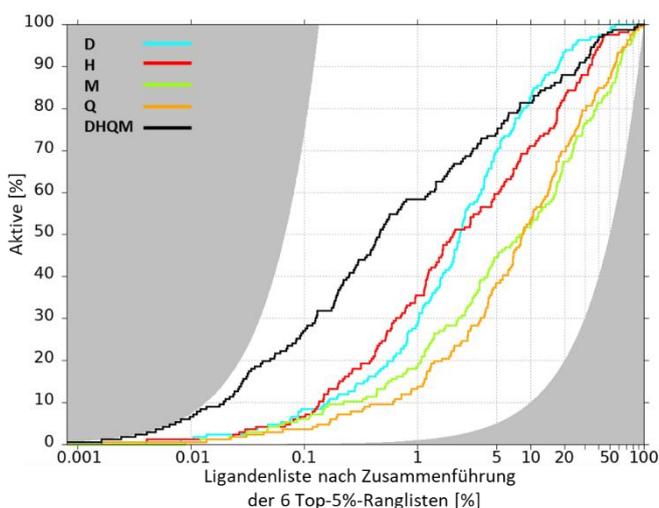


Abb. 5-78 – Anreicherung der Aktiven nach Sortierung und Zusammenführung der 6 Top-5%-Ranglisten des ZINC-Screenings (SP) unter Beibehaltung der besten Pose und anschließender Sortierung unter Verwendung der Einzelscorewerte (**D**, **H**, **M**, **Q**) sowie des kombinierten **nmr_DHQM**-Consensuscores.

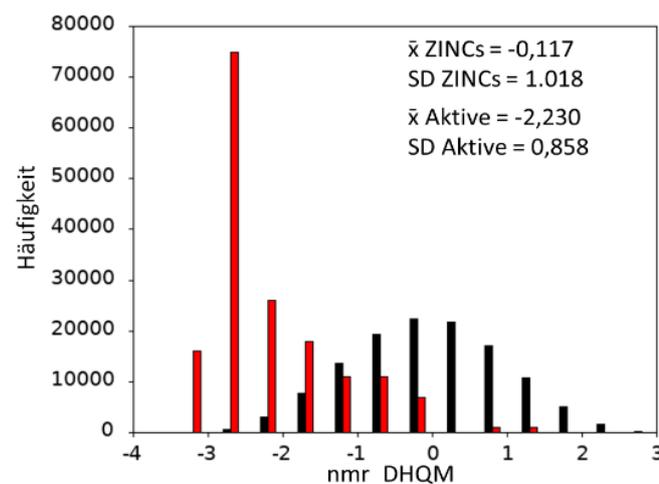


Abb. 5-79 – Häufigkeitsverteilung der Aktiven (rot, 1000-fach vergrößert dargestellt) und ZINC-Cs (schwarz) der nach Consensuscore zusammengefassten Liste (entspr. Tab. 5-49).

Interessant ist darüber hinaus der Vergleich mit den Ergebnissen des *Decoy-Actives-Screenings* (**Ligandenliste c**, s. Kap. 5.4.2): Auch dort betrug die Reihenfolge für die Korrelation des Scorewertes mit der bekannten Aktivität **Q > H > D > M**. Die frühe Anreicherung gelang im Testscreening

in der absteigenden Reihenfolge **D > H > M > (=) Q**. Auch im *ZINC-Actives-Screening* war die frühe Anreicherung mit dem Docking- und HYDE-Score besser als mit den übrigen beiden Scorewerten. Die frühe Anreicherung von Aktiven im *Decoy-Actives-Screening* war zwar unter Verwendung des Consensuscores vergleichsweise mit am höchsten, allerdings bestand keine große Differenz zum Enrichmentfaktor unter Verwendung des Dockingscores (**nmr_DHQM**: $EF_{1\%} = 17,28$, **D**: $EF_{1\%} = 16,13$). Dies lag wie bereits beschrieben daran, dass die Anreicherung von $EF_{1\%} = 17,28$ nicht weiter gesteigert werden konnte, da dies schon der perfekten Anreicherung entsprach. Man konnte nur im Enrichmentplot (s. S. 213, Abb. 5-76 A) bereits erkennen, dass die Anreicherungskurve für **nmr_DHQM** über einen minimal größeren Bereich deckungsgleich mit der perfekten Anreicherungskurve war (bis 1,95 % der Datenbank) als die Anreicherungskurve für den Dockingscore alleine (bis 0,54 %).

5.5.4 Leader-Clustering mit RDKit-Fingerprints

Um die Anzahl an ZINC-Strukturen weiter zu reduzieren und möglichst unterschiedliche Strukturen auszuwählen, ohne dass wichtige strukturelle Informationen verloren gehen, sollten ähnliche ZINC-Strukturen zu Gruppen zusammengefasst werden.

Für die 123.796 Strukturen (123.630 ZINC und 166 Aktive) der auf Basis des Consensuscores erstellten und sortierten Gesamtträngliste wurden RDKit-Fingerprints (s. Kap. 4.7.1, S. 92 u. Kap. 7.9.7) berechnet und mit dem Leader-Algorithmus (s. Kap. 4.7.2, S. 94 u. Kap. 7.9.8) geclustert („Clustering OMW“). Die Liste der Fingerprints lag dafür nach dem Consensuscore sortiert vor, sodass der Clusterrepräsentant der Ligand mit der besten Bewertung in dem Cluster wurde.

Die ZINC-Strukturen haben im Durchschnitt ein geringeres Molekulargewicht als die hier eingesetzten Aktiven (s. Abb. 5-80). Ein weiteres Clustering („Clustering $MW < 350$ “) wurde auf gleiche o.g. Weise nur mit den Liganden mit geringem Molekulargewicht ($MW < 350$, 88.060 ZINC Strukturen und 21 Aktive) durchgeführt, da bei der Substanzauswahl auf die Größe der Liganden besonders geachtet werden sollte, denn kleine

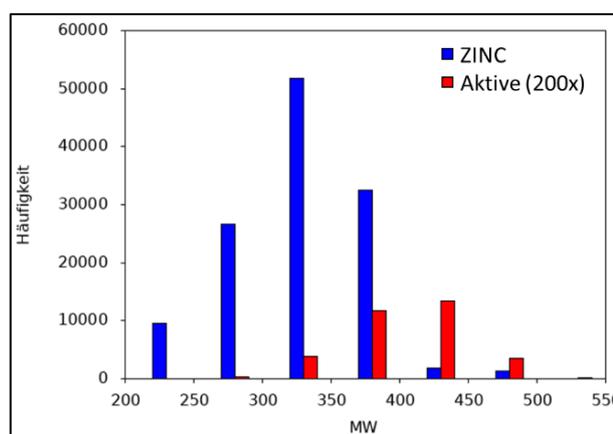


Abb. 5-80 – Häufigkeitsverteilung der 123.630 ZINC-Strukturen (blau) und der 166 Aktiven (rot, 200 fach vergrößert) nach Molekulargewicht

Liganden bieten mehr Möglichkeiten für strukturelle Veränderungen bei einer Optimierung der Wirksamkeit und Bioverfügbarkeit.

Beide Clusterings – OMW und MW<350 – wurden mit den Optionen $t = 0,2$ (80 % Übereinstimmung pro Cluster) bzw. $t = 0,3$ (70 % Übereinstimmung pro Cluster) durchgeführt.

Im Vergleich zum Clustering mit $t = 0,2$ nimmt die Clustergröße bei $t = 0,3$ erwartungsgemäß zu und die Zahl der Einzelcluster ab (s. Tab. 5-50). Ebenso nimmt die Gesamtzahl der Cluster ab (49.065 mit $t = 0,3$ vs. 72.633 mit $t = 0,2$).

Die Diversität der Clusterrepräsentanten ist bei $t = 0,3$ größer als bei $t = 0,2$. Nachteilig für das Clustering mit $t = 0,3$ statt mit $0,2$ ist allerdings, dass Liganden einem Clusterrepräsentanten zugeordnet werden könnten, die in der Bindetasche eine deutlich andere Pose als der Clusterrepräsentant einnehmen und daher vom Experten bei manuellem Clustering möglicherweise nicht dem gleichen Cluster zugeordnet worden wären. Es ist also dadurch ein größeres Risiko vorhanden, dass interessante Moleküle durch die größeren Cluster verloren gehen.

Tab. 5-50 – Ergebnisse des Leader-Clusterings der RDKit-Fingerprints aller 123.796 Strukturen (OMW) bzw. der Strukturen mit Molekulargewicht < 350 (MW<350). # Rep. = Anzahl der Cluster(repräsentanten).

OMW				MW<350			
$t = 0,2$		$t = 0,3$		$t = 0,2$		$t = 0,3$	
Clustergröße	Anzahl	Clustergröße	Anzahl	Clustergröße	Anzahl	Clustergröße	Anzahl
50 - 237	25	50 - 321	98	50 - 226	9	50 - 305	58
30 - 49	61	30 - 49	163	30 - 49	36	30 - 49	97
10 - 29	711	10 - 29	1386	10 - 29	430	10 - 29	898
9	167	9	294	9	131	9	193
8	227	8	390	8	164	8	282
7	364	7	538	7	244	7	409
6	613	6	767	6	422	6	527
5	938	5	1250	5	634	5	840
4	1877	4	2067	4	1331	4	1514
3	3647	3	3859	3	2589	3	2887
2	12579	2	9359	2	9171	2	7235
1	51424	1	28894	1	38207	1	22851
# Rep.	72633	# Rep.	49065	# Rep.	53368	# Rep.	37791

Die Anzahl der zu evaluierenden Posen konnte für $t = 0,2$ auf etwa 60 % und für $t = 0,3$ auf etwa 40 % reduziert werden (s. Tab. 5-50). Für $t = 0,2$ sind etwa 70 % der Cluster Einzelcluster, für $t = 0,3$ sind es etwa 60 %. Mehr als die Hälfte der Substanzen konnten also keiner ähnlichen Substanz zugeordnet werden. Um eine hohe Anzahl an großen Clustern und weniger Clusterrepräsentanten zu erhalten, wurden die Clusterergebnisse für $t = 0,3$ statt für $t = 0,2$ für die Substanzauswahl verwendet.

Die nach Consensuscore sortierten ersten 1000 ZINC-Clusterrepräsentanten sollten jeweils bei der Substanzauswahl (s. Kap. 5.5.5) genauer betrachtet werden. Sie repräsentieren 11.700 (OMW) bzw. 8.616 (MW<350) Substanzen. Dies macht noch einmal den positiven Effekt des Clusterings deutlich.

Am Beispiel des Clusters #15 ($t = 0,3$, OMW, Abb. 5-81) lässt sich erkennen, dass die Ähnlichkeit der Strukturen untereinander sehr hoch ist. Stereoisomere (hier: **(48)** und **(50)**) können mit dieser Methode gruppiert werden. Core-1-Einheit (blau) und der Substituent R1 (grün) sind identisch. Die Consensuscore-Werte dieses Clusters sind alle sehr gut. Der Clusterrepräsentant **(48)** dieses Clusters ist einer der ausgewählten Strukturen für die Testung. Der Vorteil des Clusterings wird an dieser Stelle deutlich: Die Entscheidung wurde anhand des Clusterrepräsentanten **(48)** getroffen und die übrigen drei Strukturen **(49-51)** mussten nicht betrachtet werden. Die Clusterbildung hat darüber hinaus den Vorteil, dass nach erfolgreicher Testung schnell weitere ähnliche Clustervertreter identifiziert und getestet werden könnten.

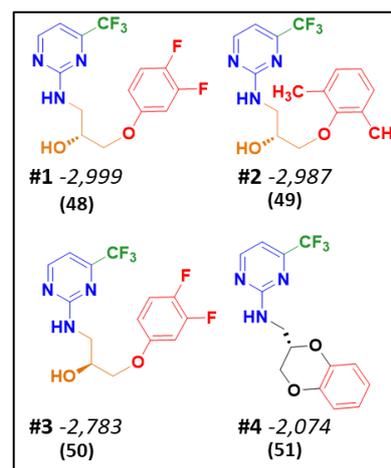


Abb. 5-81 – Clustermitglieder des Clusters #15 ($t = 0,3$, OMW) mit entspr. Consensuscore (kursiv); Ausrichtung und Färbung der Strukturen entspricht der Dockingpose (vgl. Kap. 5.2.1).

5.5.5 Substanzauswahl

Bei der Auswahl der Substanzen für die biologische Testung wurden jeweils die Dockingposen der besten 1000 ZINC-Clusterrepräsentanten des obengenannten Leader-Clusterings ($t = 0,3$) „OMW“ und „MW“ in der DLK-Bindetasche einzeln visuell beurteilt.

Strukturen wurden positiv bewertet, wenn sie möglichst viele Wasserstoffbrücken zur Hinge-Region ausbildeten, möglichst tief in der Bindetasche platziert und nicht zu stark deformiert waren. Den letzteren Punkt konnte man auch häufig anhand einer erhöhten *Ligand Strain Energy* erkennen. Die Strukturen sollten nicht zu viele Hydroxylgruppen besitzen, da diese häufig zu einer Überbewertung unspezifischer bzw. ungerichteter Wasserstoffbrückenbindungen durch die Scoringfunktionen führten. Zudem wurde es als positiv angesehen, wenn eine H-Donorgruppe anstelle eines Wassermoleküls in der Bindetasche platziert war. Dies war z.B. in der Proteinkonformation ID 7 im Vergleich zu den IDs 6.3 und 6.11 zu erkennen. Darüber hinaus wurde darauf geachtet, möglichst keine Strukturen auszuwählen, die gleiche Grundstrukturen (insbesondere Core 1) wie bereits bekannte aktive DLK-Inhibitoren (s. Kap. 5.2.1) besaßen. Bei Clusterrepräsentanten mit einer interessanten, neuartigen Grundstruktur, aber mit unerwünschten Strukturelementen (z.B. zu vielen Hydroxylgruppen) wurden weitere Vertreter des Clusters angeschaut, um eventuell geeignete Alternativen zu finden. **C03 (78)** wurde so als einziger Nicht-Clusterrepräsentant ausgewählt.

Zur weiteren Beurteilung der Pose wurden auch verschiedene numerische Parameter (s. Tab. 5-51) herangezogen:

- *Consensus*score: Der Consensuscore **nmr_DHQM** als zentraler Bewertungsscore liegt bei den zu betrachtenden Posen zwischen -3,054 (bester Wert) und -2,069. Auch die vier Einzelscorewerte (**D, H, Q, M**) wurden zu jeder Posenbeurteilung herangezogen.
- *Ligand Strain Energy (L)*: L befindet sich bei den Aktiven hauptsächlich in einem Bereich von etwa 0 bis 5 kcal/mol. Die *Ligand Strain Energy* sollte einen möglichst kleinen Wert annehmen als Indikator dafür, dass der gedockte Ligand eine nicht zu stark deformierte Konformation einnimmt (s. Kap. 4.4.3, S. 73).
- *Glide Ligand Efficiency (LE)*: Im Mittel liegt die LE (s. Kap. 4.4.1, S. 69) der 166 Aktiven bei -0,339. Die LE der ZINC-Strukturen ist im Mittel sogar negativer (d.h. besser). Dies liegt an dem im Durchschnitt geringeren Molekulargewicht (s. Tab. 5-51). Bei der Auswahl wurde auf einen niedrigen LE-Wert Wert gelegt.
- *IFP-Tanimotokoeffizient* (s. Kap. 4.7.1, S. 93) ($t = 0-1$): vergleicht die Interaktionen des Rezeptor-ZINC-Komplexes mit einem Rezeptor-Referenzliganden-Komplex (**5CEO, 5CEP** bzw. **5CEQ**, s. Kap. 7.9.5). Je näher der Wert an 1, desto ähnlicher sind die Interaktionen. Dieser Parameter sollte kein Entscheidungskriterium sein, stellt aber einen interessanten Parameter für den Vergleich mit den Interaktionen der Aktiven dar.

Tab. 5-51 – Parameter zur Beurteilung bei der Ligandenauswahl. Angegeben ist der Mittelwert des jeweiligen Scorewerts und die Standardabweichung in Klammern dahinter. Zusätzlich wird das Molekulargewicht angegeben. ZINC_s bzw. Aktive = 123.630 ZINC-Strukturen bzw. 166 Aktive der Gesamttrangliste vor dem Clustering. Top1000 OMW bzw. MW<350: Die nach Consensuscore besten 1000 Clusterrepräsentanten des OMW- bzw. MW<350-Clusterings.

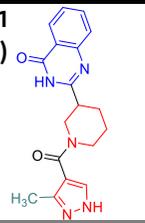
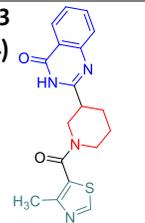
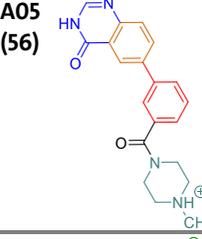
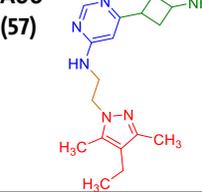
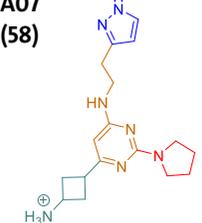
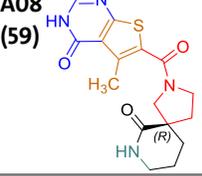
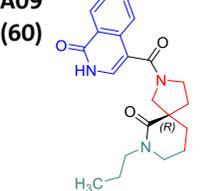
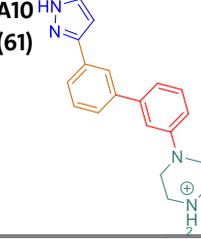
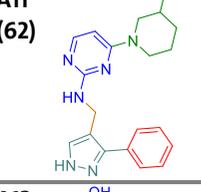
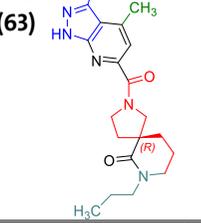
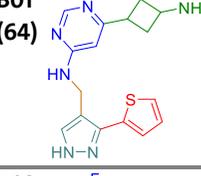
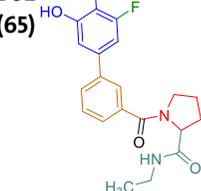
Scores	ZINC _s	Aktive	Top1000 OMW	Top1000 MW<350
nmr_DHQM	-0,117 (1,018)	-2,230 (0,858)	-2,525 (0,173)	-2,340 (0,212)
D	-8,749 (0,453)	-9,704 (0,629)	-9,555 (0,515)	-9,484 (0,489)
H	-25,373 (11,305)	-41,454 (8,866)	-40,900 (5,680)	-39,804 (5,997)
Q	6,477 (0,911)	7,575 (0,930)	7,564 (0,579)	7,419 (0,548)
M	-48,193 (8,889)	-57,438 (8,448)	-59,770 (4,410)	-57,578 (4,133)
L	7,255 (4,731)	4,022 (2,198)	6,580 (3,493)	5,774 (3,252)
LE	-0,387 (0,064)	-0,339 (0,037)	-0,381 (0,051)	-0,414 (0,043)
IFP	0,622 (0,083)	0,748 (0,098)	0,654 (0,085)	0,657 (0,086)
MW	322,819 (47,376)	399,950 (40,667)	352,794 (40,795)	318,483 (24,710)

Die 1000 Clusterrepräsentanten wurden jeweils in zwei Durchgängen bewertet:

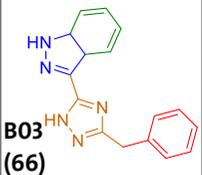
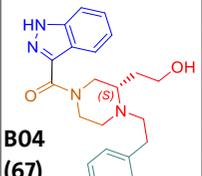
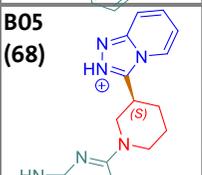
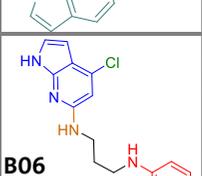
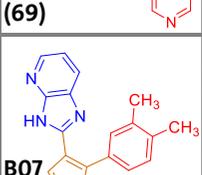
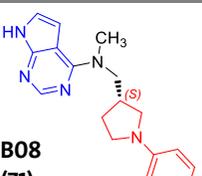
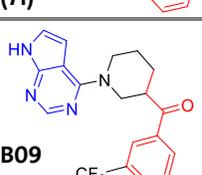
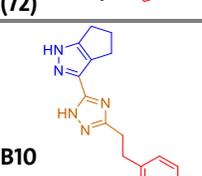
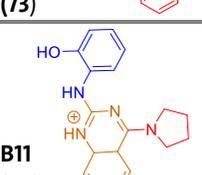
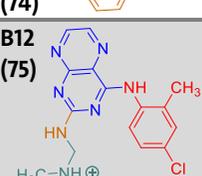
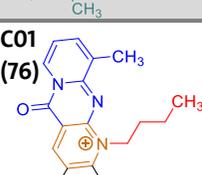
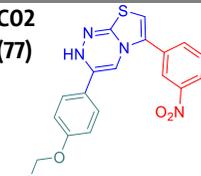
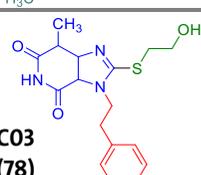
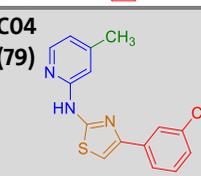
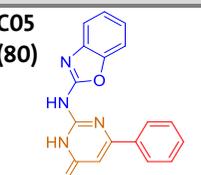
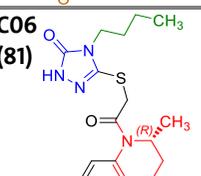
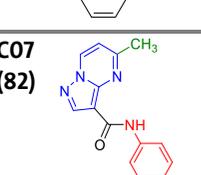
Nach der ersten Durchsicht der beiden Top-1000-Clusterrepräsentanten-Datensätze wurden zunächst 214 Strukturen (OMW) und 216 Strukturen (MW<350) nach den o.g. Kriterien ausgewählt. 90 Strukturen kamen in beiden dieser Selektionen vor. Das Zusammenführen ergab 340 unterschiedliche Strukturen. Diese wurden im zweiten Durchgang manuell in Gruppen unterteilt, bei

denen gleiche Grundkörper, die mit der Hinge-Region interagieren (vgl. Kap. 5.2.1, Core 1), zusammengefasst wurden. Dies führte letztendlich zur Auswahl von 67 diversen Strukturen; 58 dieser Substanzen waren lieferbar und wurden bei der Firma AMBINTER c/o Greenpharma, Orléans, Frankreich, bestellt. Substanzen, die Stereozentren enthalten, sind als Racemate geliefert worden.

Die ausgewählten Strukturen (**52-118**) sind in Tab. 5-52 dargestellt (Ausrichtung wie in der Binde-tasche, vgl. dazu Kap. 5.2.1, Ausrichtung und Färbung der Strukturelemente).

Struktur	Ambinter-Nr.	ID	IFP
	ZINC-Nr.	Cons.	L
	MW	D	Emodel
	Cluster-Nr.	H	LE
	Clustergr.	Q	Cl.M.
	# IDs docked	M	Assay
 A01 (52)	Amb20127492	ID 13	0,632
	ZINC12116228	-2,371	2,585
	337,381	-9,507	-78,010
	388	-36,853	-0,380
	2	7,274	MW<350
A02 (53)	Amb20072553	ID 6.3	0,734
ZINC19695540	-2,817	6,477	
397,423	-10,786	-85,960	
101	-48,590	-0,385	
10	7,532	OMW	
6	-58,870	1, 2, 3	
 A03 (54)	Amb20127497	ID 13	0,632
	ZINC12145143	-2,602	3,128
	354,431	-9,880	-77,842
315	-36,748	-0,395	
2	7,312	OMW	
1	-59,366	1, 2, 3	
 A04 (55)	Amb20072924	ID 6.3	0,642
	ZINC72375192	-2,817	5,592
	394,496	-9,583	-80,688
	100	-42,484	-0,342
36	7,955	OMW	
6	-60,196	1, 2, 3	
 A05 (56)	Amb24135001	ID 6.11	0,677
	ZINC299756482	-2,407	2,704
	349,412	-9,402	-80,620
	704	-38,482	-0,362
	11	7,377	beide
6	-56,403	1, 2, 3	
 A06 (57)	Amb23414001	ID 6.11	0,672
	ZINC218761118	-2,666	2,974
	315,442	-10,045	-81,531
	231	-41,930	-0,437
	4	8,490	beide
	2	-53,762	1, 2, 3
 A07 (58)	Amb23411659	ID 18	0,792
	ZINC244811558	-2,319	5,948
	328,441	-8,959	-74,172
	970	-51,973	-0,373
	3	7,935	beide
5	-55,293	1, 2, 3	
 A08 (59)	Amb22262541	ID 18	0,534
	ZINC91596126	-2,436	5,651
	346,408	-10,223	-68,253
	629	-44,268	-0,426
	104	7,248	beide
	2	-55,581	1, 2
 A09 (60)	Amb22262056	ID 18	0,607
	ZINC91416964	-2,733	5,569
	367,447	-10,557	-75,828
	166	-41,451	-0,391
	37	7,519	OMW
	6	-60,177	1, 2
 A10 (61)	Amb11087780	ID 6.11	0,651
	ZINC23398672	-2,729	2,477
	305,403	-9,540	-82,668
	172	-44,118	-0,415
	2	7,423	beide
6	-61,718	1, 2	
 A11 (62)	Amb17628857	ID 7	0,766
	ZINC67473135	-2,566	8,056
	350,424	-10,218	-93,177
	369	-52,186	-0,393
5	7,999	OMW	
2	-53,617	1, 2, 3	
 A12 (63)	Amb22255713	ID 18	0,597
	ZINC91906982	-2,802	3,945
	371,439	-10,100	-78,333
	111	-41,158	-0,374
	13	7,960	OMW
6	-59,940	1, 2, 3	
 B01 (64)	Amb23412553	ID 6.11	0,597
	ZINC218586710	-2,139	2,575
	327,433	-9,935	-90,284
	817	-30,259	-0,432
	3	7,247	MW<350
	3	-59,435	1, 2, 3
 B02 (65)	Amb32528767	ID 13	0,623
	ZINC426506578	-2,680	3,279
	374,385	-9,595	-71,456
	218	-40,082	-0,355
	10	7,323	OMW
6	-60,841	1, 2	

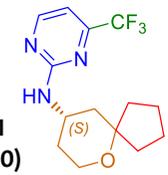
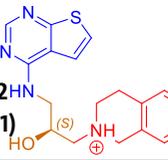
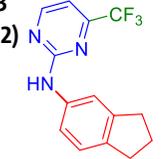
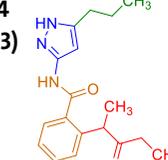
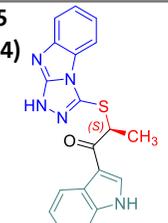
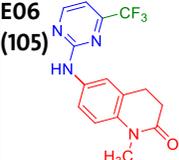
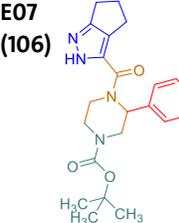
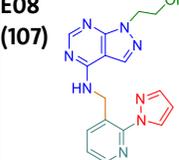
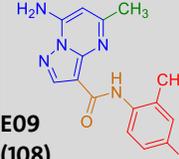
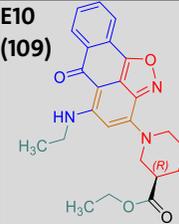
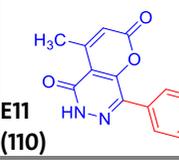
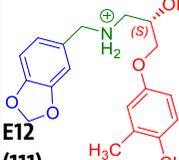
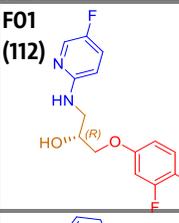
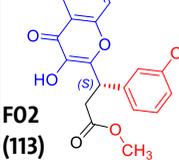
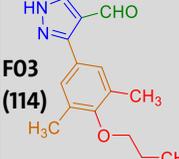
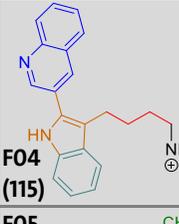
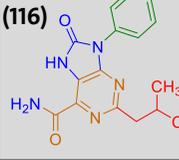
Tab. 5-52 – Übersicht über die 67 ausgewählten Substanzen. Clustergr. = Clustergröße. # IDs docked = Anzahl der Proteinkonformationen, in die diese Struktur gedockt werden konnte. ID = Proteinkonformation des Komplexes mit dem besten Consensuscore. D = Dockingscore, H = HYDE-Score, Q = vorhergesagter pK_i -Wert mit dem *Sel2*-QSAR-Modell. M = Freie Bindungsenergie (PRIME MM/GBSA). IFP = IFP-Tanimotokoeffizient. L = *Ligand Strain Energy*. LE = *Glide Ligand Efficiency*. Cl. M. = Clustermethode, aus der dieser Clusterrepräsentant stammt. Grau hinterlegt = nicht lieferbar. Assay: Getestet in ADP-Glo™ Kinase Assay (1), MST (2) bzw. KINOMEscan® (3)

Struktur	Ambinter-Nr.	ID	IFP
	ZINC-Nr.	Cons.	L
	MW	D	Emodel
	Cluster-Nr.	H	LE
	Clustergr.	Q	Cl.M.
	# IDs docked	M	Assay
	Amb17645970	ID 6.3	0,758
	ZINC67755917	-2,125	4,719
	275,314	-9,725	-67,303
	849	-37,859	-0,463
	3	7,560	MW<150
	3	-50,412	1, 2, 3
	Amb19633365	ID 18	0,545
	ZINC72434492	-2,888	11,884
	379,481	-10,583	-73,697
	58	-57,980	-0,378
	216	8,269	OMW
	3	-58,859	1, 2, 3
	Amb19630507	ID 6.3	0,508
	ZINC72162424	-2,084	0,903
	318,382	-9,240	-70,487
	965	-36,612	-0,385
	1	6,982	MW<150
	6	-56,299	1, 2, 3
	Amb19629172	ID 13	0,600
	ZINC72154829	-2,320	4,622
	301,779	-9,862	-69,944
	966	-40,713	-0,470
	9	6,904	beide
	3	-56,333	1, 2, 3
	Amb19622687	ID 6.3	0,672
	ZINC72409488	-2,368	3,900
	289,341	-9,442	-77,544
	819	-35,866	-0,429
	3	7,970	beide
	5	-54,612	1, 2, 3
	Amb19619951	ID 6.3	0,632
	ZINC77302737	-2,321	3,516
	307,399	-10,235	-67,165
	962	-46,509	-0,445
	10	7,846	beide
	5	-51,179	1, 2, 3
	Amb19583232	ID 7	0,677
	ZINC71524623	-2,643	3,000
	374,3643	-10,491	-82,003
	257	-37018	-0,389
	12	8,352	OMW
	1	-56,572	1, 2, 3
	Amb17647787	ID 7	0,719
	ZINC65452750	-2,339	3,146
	279,3453	-9,603	-77,228
	903	-35,686	-0,457
	7	7,924	beide
	6	-56,479	1, 2, 3
	Amb17624605	ID 6.3	0,710
	ZINC41148123	-2,828	1,1373
	306,3672	-9,704	-76,506
	35	-44,642	-0,4219
	32	7,445	beide
	5	-66,766	1, 2
	Amb1436169	ID 7	0,821
	ZINC13010895	-2,333	3,927
	358,855	-9,114	-77,544
	924	-41,457	-0,365
	44	8,292	MW<350
	2	-56,910	-
	Amb1317229	ID 6.11	0,931
	ZINC100707544	-2,433	6,711
	341,436	-10,486	-79,118
	636	-37,334	-0,437
	59	6,994	beide
	5	-63,491	1, 2
	Amb17758836	ID 7	0,845
	ZINC2271003	-2,652	6,821
	380,4254	-10,371	-80,255
	248	-35,738	-0,384
	2	8,103	OMW
	1	-60,186	1, 2
	Amb19750310	ID 7	0,652
	ZINC2239690	-2,036	2,725
	346,4082	-9,719	-82,307
	395	-25,439	-0,405
	92	8,154	OMW*
	2	-58,489	1, 2
	Amb19920479	ID 6.11	0,810
	ZINC21364351	-2,858	1,254
	301,799	-9,622	-71,346
	78	-43,949	-0,481
	43	7,862	beide
	5	-61,566	-
	Amb22568376	ID 6.11	0,759
	ZINC8974410	-2,459	1,592
	304,308	-9,649	-75,206
	578	-35,676	-0,420
	2	8,105	beide
	4	-55,032	1, 2, 3
	Amb19244962	ID 33	0,672
	ZINC9313834	-2,830	4,612
	360,4786	-8,975	-70,968
	90	-47,531	-0,359
	11	7,279	OMW
	1	-63,848	1, 2
	Amb19318321	ID 6.11	0,656
	ZINC65614203	-2,481	1,735
	336,353	-10,021	-79,257
	528	-42,404	-0,401
	24	6,997	beide
	6	-58,173	1, 2, 3

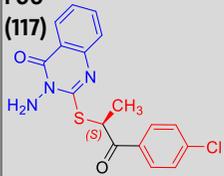
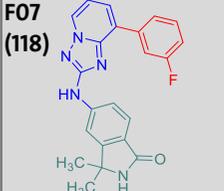
Tab. 5-52 – Fortsetzung. C03 (78): OMW* = nicht Clusterrepräsentant

Struktur	Ambinter-Nr.	ID	IFP
	ZINC-Nr.	Cons.	L
	MW	D	Emodel
	Cluster-Nr.	H	LE
	Clustergr.	Q	Cl.M.
	# IDs docked	M	Assay
C08 (83) 	Amb19481421	ID 6.3	0,902
	ZINC69709602	-2,999	3,400
	349,258	-10,180	-71,708
	15	-56,673	-0,424
	4	8,418	beide
	5	-63,275	1, 2
C09 (84) 	Amb15768964	ID 7	0,839
	ZINC58324957	-2,636	1,029
	305,262	-10,131	-74,132
	264	-44,500	-0,460
	1	8,6676	beide
	6	-52,921	1, 2, 3
C10 (85) 	Amb15748262	ID 6.11	0,645
	ZINC46083122	-2,560	4,309
	329,362	-9,448	-76,621
	387	-39,236	-0,378
	3	7,389	beide
	5	-60,317	1, 2
C11 (86) 	Amb15720636	ID 33	0,672
	ZINC55722169	-2,692	3,369
	347,413	-9,324	-73,074
	204	-41,637	-0,373
	21	6,932	beide
	4	-62,752	1, 2
C12 (87) 	Amb14423572	ID 6.3	0,774
	ZINC67329728	-2,239	2,731
	312,394	-9,680	-66,282
	609	-36,740	-0,440
	32	7,199	MW<350
	6	-54,332	1, 2, 3
D01 (88) 	Amb13996882	ID 6.3	0,823
	ZINC24054653	-2,438	1,892
	308,266	-8,978	-66,322
	623	-41,665	-0,408
	3	7,808	beide
	4	-58,887	1, 2
D02 (89) 	Amb12708884	ID 6.11	0,850
	ZINC65066112	-2,426	5,926
	364,426	-9,475	-85,192
	653	-33,049	-0,364
	4	7,821	OMW
	3	-60,831	1, 2, 3
D03 (90) 	Amb28898896	ID 7	0,750
	ZINC65586446	-2,426	1,618
	293,368	-9,456	-68,929
	652	-43,493	-0,430
	12	7,918	beide
	5	-55,376	1, 2, 3
D04 (91) 	Amb28920486	ID 6.3	0,836
	ZINC408515154	-2,569	1,171
	308,262	-10,252	-72,584
	366	-43,174	-0,466
	3	7,316	beide
	6	-55,582	1, 2, 3
D05 (92) 	Amb11163250	ID 6.11	0,712
	ZINC46019604	-2,982	5,824
	307,355	-10,089	-88,229
	20	-45,287	-0,439
	1	8,393	beide
	4	-62,717	1, 2
D06 (93) 	Amb28633276	ID 33	0,652
	ZINC97164980	-2,099	5,525
	325,414	-8,685	-70,559
	923	-47,686	-0,362
	3	6,909	MW<150
	1	-51,258	1, 2
D07 (94) 	Amb28850444	ID 6.11	0,586
	ZINC72290889	-2,920	5,626
	324,349	-10,345	-85,644
	38	-44,675	-0,470
	1	7,758	beide
	2	-62,429	1, 2
D08 (95) 	Amb28870343	ID 6.3	0,790
	ZINC75105269	-2,464	4,730
	339,3969	-9,626	-84,890
	559	-39,993	-0,385
	5	6,974	beide
	6	-61,122	1, 2
D09 (96) 	Amb28872154	ID 13	0,780
	ZINC366785354	-2,763	1,851
	341,437	-9,282	-77,469
	141	-43,585	-0,371
	5	8,313	beide
	5	-61,256	1, 2
D10 (97) 	Amb19198453	ID 13	0,765
	ZINC24842714	-2,767	2,528
	347,416	-9,549	-79,284
	53	-39,897	-0,367
	15	7,471	beide
	6	-64,474	1, 2, 3
D11 (98) 	Amb19301478	ID 18	0,613
	ZINC65529816	-2,933	4,329
	397,476	-10,169	-81,722
	33	-55,743	-0,339
	13	7,580	OMW
	4	-68,212	1, 2
D12 (99) 	Amb22115677	ID 33	0,661
	ZINC83243930	-2,194	2,699
	299,3719	-8,893	-59,076
	693	-38,698	-0,404
	5	7,055	MW<150
	6	-52,546	1, 2

Tab. 5-52 – Fortsetzung.

Struktur	Ambinter-Nr.	ID	IFP
	ZINC-Nr.	Cons.	L
	MW	D	Emodel
	Cluster-Nr.	H	LE
	Clustergr.	Q	Cl.M.
	# IDs docked	M	Assay
 E01 (100)	Amb22165028	ID 13	0,769
	ZINC98020513	-2,452	4,3573
	301,310	-9,801	-62,003
	591	-45,457	-0,467
	7	7,624	beide
 E02 (101)	Amb10640381	ID 18	0,709
	ZINC40511142	-2,280	3,158
	341,456	-9,350	-61,327
	538	-45,626	-0,390
	2	7,094	MW<350
 E03 (102)	Amb8297022	ID 33	0,635
	ZINC32908926	-2,430	1,167
	279,264	-9,245	-58,587
	642	-36,671	-0,462
	2	8,149	beide
 E04 (103)	Amb28953511	ID 33	0,719
	ZINC97010861	-2,289	4,379
	314,387	-8,621	-74,203
	526	-39,259	-0,375
	2	6,920	MW<350
 E05 (104)	Amb89839	ID 13	0,673
	ZINC4266839	-2,422	2,377
	361,427	-10,381	-85,566
	662	-29,368	-0,399
	4	7,524	OMW
 E06 (105)	Amb24233823	ID 33	0,651
	ZINC171564710	-2,729	1,861
	322,289	-9,459	-63,117
	173	-43,313	-0,411
	3	7,865	beide
 E07 (106)	Amb28549399	ID 18	0,610
	ZINC225741390	-2,811	4,247
	414,4783	-9,477	-73,302
	104	-44,637	-0,316
	28	8,095	OMW
 E08 (107)	Amb28609278	ID 6.11	0,638
	ZINC97200001	-2,344	4,179
	336,3574	-11,180	-89,617
	886	-42,036	-0,447
	1	6,783	beide
 E09 (108)	Amb20260021	ID 18	0,685
	ZINC101299475	-2,371	2,348
	295,344	-9,989	-69,214
	812	-47,610	-0,454
	3	7,108	beide
 E10 (109)	Amb16693130	ID 13	0,737
	ZINC6737979	-2,629	5,8257
	419,479	-9,476	-80,500
	274	-33,869	-0,306
	53	7,651	OMW
 E11 (110)	Amb2761361	ID 3	0,820
	ZINC8651696	-2,263	1,890
	254,244	-9,831	-68,721
	569	-32,217	-0,517
	1	7,593	MW<350
 F01 (112)	Amb2739643	ID 7	0,627
	ZINC119211	-2,626	3,282
	330,402	-9,165	-73,048
	277	-40,131	-0,382
	4	9,361	OMW
 F02 (113)	Amb24241426	ID 6.11	0,845
	ZINC252480847	-2,546	4,201
	298,263	-9,523	-59,494
	408	-43,472	-0,453
	5	7,447	beide
 F03 (114)	Amb24050664	ID 7	0,683
	ZINC253407448	-2,502	8,574
	340,329	-9,967	-75,481
	492	-35,427	-0,399
	20	7,697	beide
 F04 (115)	Amb23201549	ID 13	0,604
	ZINC113703678	-2,497	2,782
	258,319	-9,424	-70,674
	498	-42,056	-0,496
	8	7,093	beide
 F05 (116)	Amb10607110	ID 6.11	0,644
	ZINC6270008	-2,489	5,880
	316,426	-9,708	-76,329
	515	-45,388	-0,404
	12	7,125	beide
 F06 (111)	Amb20420583	ID 11	0,774
	ZINC4175878	-2,665	8,525
	325,370	-10,219	-78,501
	232	-41,049	-0,426
	19	7,697	beide

Tab. 5-52 – Fortsetzung.

Struktur	Ambinter-Nr.	ID	IFP
	ZINC-Nr.	Cons.	L
	MW	D	Emodel
	Cluster-Nr.	H	LE
	Clustergr.	Q	Cl.M.
	# IDs docked	M	Assay
F06 (117) 	Amb22036719	ID 13	0,673
	ZINC4603034	-2,741	1,850
	359,835	-10,040	-78,123
	160	-36,473	-0,418
	19	7,486	OMW
	5	-62,725	-
F07 (118) 	Amb22495683	ID 7	0,793
	ZINC95616928	-2,489	3,206
	387,416	-9,786	-82,186
	513	-40,074	-0,337
	1	7,943	OMW
	3	-55,514	-

Tab. 5-52 – Fortsetzung.

5.6 Biologische Prüfung

Die 58 bei AMBINTER gekauften Hit-Substanzen (s. S. 222 ff., Tab. 5-52) wurden zunächst mit dem ADP-Glo™ Kinase Assay auf ihre DLK-aktivitätshemmende Wirkung sowie mit der *Microscale Thermophoresis* (MST) auf die Bindung an die DLK überprüft. Anschließend wurden ausgewählte Substanzen extern mit der KINOMEScan®-Methode von der Firma Eurofins DiscoverX Corporation in San Diego geprüft. In dieser Methode wird die humane DLK in voller Länge verwendet.

Für die Durchführung des ADP-Glo™ Kinase Assays und der MST wurden zwei Proteinlösungen verwendet, die möglichst hohe DLK-Konzentrationen und möglichst wenige Verunreinigungen enthielten (s. Kap. 7.15.1). Die Konzentration der beiden im Folgenden eingesetzten MBP-TEV-DLK₁₄₀₋₅₁₇-Lösungen wurden semiquantitativ bestimmt (s. Kap. 7.15.1). Die MBP-TEV-DLK₁₄₀₋₅₁₇-Konzentration für Fraktion e bzw. g beträgt 0,135 bzw. 0,119 µg/µl.

5.6.1 ADP-Glo™ Kinase Assay

(1.) Bestimmung der Aktivität der DLK-Fraktionen

Zur Abschätzung der im Assay einzusetzenden Volumina der MBP-TEV-DLK₁₄₀₋₅₁₇-Lösungen e und g (s. Kap. 7.15.1) wurde der ADP-Glo™ Kinase Assay gegen eine Lösung bekannter DLK-Konzentration (= Positivkontrolle, 1 µg MBP-TEV-DLK₁₄₀₋₅₁₇ in 7,5 µl, semiquantitativ durch Gelelektrophorese bestimmt [526]) durchgeführt (s. Kap. 7.16). Die Volumina der MBP-TEV-DLK₁₄₀₋₅₁₇-Lösungen wurden so gewählt, dass sie ca. 1 µg MBP-TEV-DLK₁₄₀₋₅₁₇ enthalten sollten (s. Kap. 7.15.1, S. 289, Tab. 7-13). Von der Positivkontrolle wurden 7,5 µl eingesetzt.

Aus dieser Messung ging hervor, dass insbesondere die Fraktion e eine nahezu gleiche Lumineszenz wie die Positivkontrolle ergab (s. Abb. 5-82). Fraktion e wurde daher für die folgenden Experimente mit dem ADP-Glo™ Kinase Assay verwendet. Fraktion g (wenig Verunreinigung, s. Kap. 7.15.1, S. 288, Abb. 7-7) wurde für die MST-Versuche (s. Kap. 5.6.2) ausgewählt.

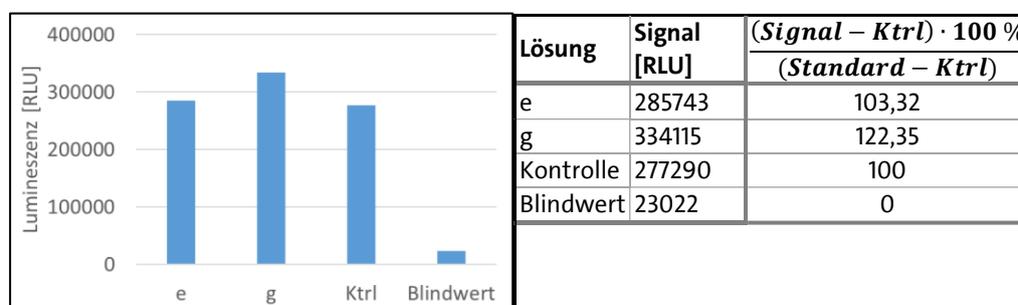


Abb. 5-82 – Lumineszenz der Fraktionen e und g im Vergleich zur Positivkontrolle (hier: Ktrl).

(2.) Optimierung der Substratkonzentration (β -Casein)

Um die für den Assay optimale β -Casein-Konzentration zu bestimmen, wurde wie vom Hersteller PROMEGA empfohlen [503], eine serielle Verdünnung von β -Casein eingesetzt (s. Kap. 7.16, S. 291, Tab. 7-16). Zusätzlich wurden die Messungen mit den verschiedenen β -Caseinkonzentrationen ohne Kinase durchgeführt. Um den DMSO-Gehalt zu berücksichtigen, der durch Zusatz der Inhibitorlösungen (in 100 % DMSO gelöst) bei den später folgenden Hit-Testungen im Reaktionsansatz enthalten sein würde, wurde eine DMSO-Puffer-Mischung (1:3) verwendet, welche jedem Well mit 2,5 μ l zugesetzt wurde. Dies entspricht einem Anteil von 5 % DMSO in der Endlösung, welches der maximal tolerierbaren Konzentration in diesem Assay entspricht (s. Kap. 4.10.2, S. 111, Tab. 4-3).

Die optimale Substratkonzentration erhält man bei der größten Differenz in der Lumineszenz zwischen den Lösungen mit stattfindender Kinase und den Lösungen ohne Kinase [503]. Die Differenz zwischen den Lumineszenzwerten mit und

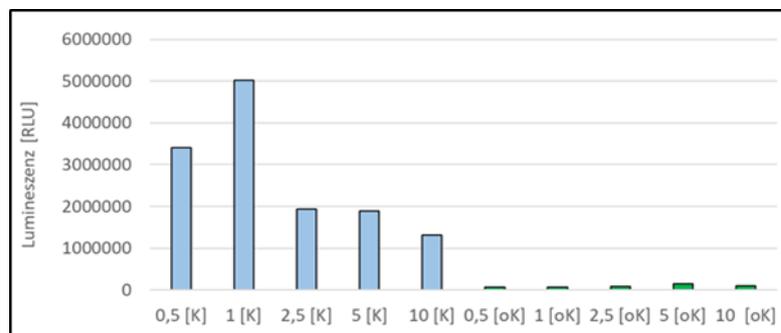


Abb. 5-83 – Ergebnis der Lumineszenzmessung mit unterschiedlichen β -Casein-Konzentrationen der zugesetzten Substratlösung; [K] = mit Kinase; [oK] = ohne Kinase, Bezeichnungen entsprechend Tab. 7-16, S. 291.

ohne MBP-TEV-DLK₁₄₀₋₅₁₇ war bei einer β -Caseinkonzentration von 1 μ g/ μ l in der zugesetzten Lösung am größten (s. Abb. 5-83). Dies entspricht einer β -Caseinkonzentration von 0,02 μ g/ μ l in der Endlösung im 12,5 μ l-Ansatz der Kinasereaktion. Diese Substratkonzentration wurde für alle weiteren Durchführungen verwendet.

(3.) Optimierung der DLK-Konzentration

Es ist ein Vorteil des ADP-Glo™ Kinase Assays, dass selbst sehr kleine ADP-Mengen mit einem guten Signal-Rausch-Verhältnis detektiert werden. [503] Daher reicht es aus, eine kleine DLK-Menge einzusetzen, sodass nur eine geringe ATP-zu-ADP-Konversion stattfindet.

HENNEK ET AL. [504] setzten für die Standardisierung von optimalen Kinaseaktivitäten für die Erstellung von Inhibitorprofilen mit dem ADP-Glo™ Kinase Assay eine Kinasemenge ein, die einer 10 bis 30 %igen Konversion von ATP zu ADP entsprach, mit einem Signal-Rausch-Verhältnis von mindestens 10 und einer ATP-Konzentration von 10 μ M. In dieser Arbeit wurde entsprechend untersucht, mit welcher MBP-TEV-DLK₁₄₀₋₅₁₇-Menge man eine 10-30 %ige ATP-zu-ADP-Konversion erreicht.

a) Dazu wurde zunächst eine Kalibrierkurve mit ATP-zu-ADP-Mischungen ohne Einsatz der Kinase, ausgehend von 10 μM ATP im „Reaktionsansatz“ erstellt (s. Kap. 7.16, S. 291, Tab. 7-17). Eine ATP-ADP-Mischung von 90:10 führte zu einem Lichtsignal von 803.086 RLU (Berechnung mit der durch lineare Regression durch die Messpunkte erhaltenen Geradengleichung, s. Abb. 5-84). Ein ATP-ADP-Verhältnis von 30:70 erzeugt, berechnet mit der Geradengleichung in Abb. 5-84, 1.947.426 RLU.

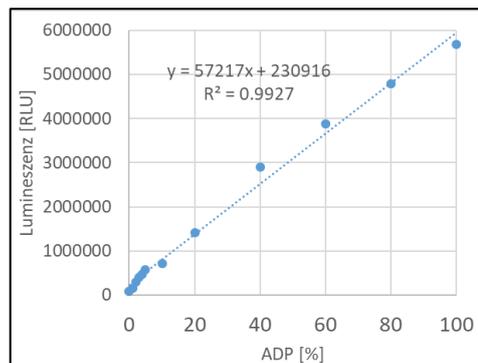


Abb. 5-84 – Abhängigkeit des Lumineszenzsignals vom ADP-ATP-Verhältnis

b) Für die Bestimmung der optimalen MBP-TEV-DLK₁₄₀₋₅₁₇-Konzentration wurde eine MBP-TEV-DLK₁₄₀₋₅₁₇-Verdünnungsreihe hergestellt (s. Kap. 7.16, S. 291). Die übrigen Bedingungen für die Kinasereaktion wurden wie zuvor (2.) beschrieben beibehalten und die optimale β -Caseinkonzentration von 0,02 $\mu\text{g}/\mu\text{l}$ im Endvolumen des Reaktionsansatzes eingesetzt.

Die optimale MBP-TEV-DLK₁₄₀₋₅₁₇-Menge erhält man aus der Kinase-Titrationskurve an der Stelle, wo das Lichtsignal im linearen Bereich liegt [504] (s. Abb. 5-85). Der lineare Bereich befindet sich hier etwa Enzymkonzentrationen zwischen 0,0058 und 0,0461 $\mu\text{g}/\mu\text{l}$. Bei sehr hohen Konzentrationen (0,0692 und 0,0921 $\mu\text{g}/\mu\text{l}$) könnte das verringerte Signal aus einer unvollständig ablaufenden Kinasereaktion resultieren, da im Verhältnis weniger

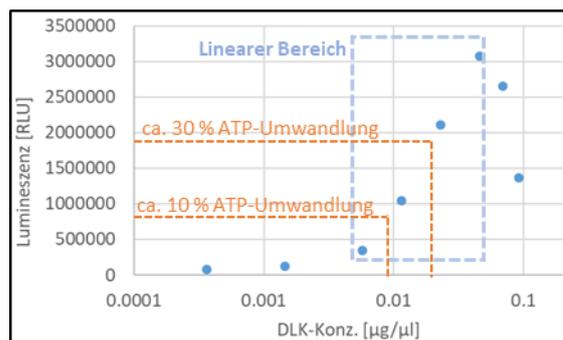


Abb. 5-85 – Abhängigkeit der Lumineszenz von der MBP-TEV-DLK₁₄₀₋₅₁₇-Konzentration. Messung 30 min. nach ADP-Detektionsreagenz-Zusatz. Gestricheltes Rechteck umschließt den linearen Bereich. Orange gestrichelte Linien stammen aus der Bestimmung mit der ATP-ADP-Konversionskurve.

Puffer zugesetzt wurde und demnach eine geringere und ungenügende Magnesiumkonzentration in der Lösung für die Kinasereaktion zur Verfügung stand.

Das Lichtsignal einer 10 %igen ATP-Umwandlung von 803.086 RLU, wie zuvor in a) bestimmt, wurde mit einer Kinasekonzentration von etwa 0,009 $\mu\text{g}/\mu\text{l}$ erhalten. Das Signal liegt im unteren Abschnitt des linearen Bereichs (s. Abb. 5-85). Eine 30 %ige ATP-Umwandlung (1.947.426 RLU) erfolgte etwa bei einer Konzentration von 0,02 $\mu\text{g}/\mu\text{l}$. Das Signal liegt im mittleren Abschnitt des linearen Bereichs der Konzentrationskurve (s. Abb. 5-85). Die von HENNEK ET AL. [504] vorgeschlagene, optimale Kinasekonzentration für Inhibitortests würde demnach

etwa zwischen 0,009 µg/µl und 0,02 µg/µl liegen. Die Konzentration für die geplanten Versuche wurde auf 0,0098 µg/µl⁶² festgelegt, um möglichst ressourcensparend zu arbeiten.

(4.) Aktivitätsbestimmung mit Referenzsubstanzen und den virtuellen Hits

- a) Es wurden vier Referenzsubstanzen mit bekannter DLK-inhibitorischer Aktivität (s. Kap. 4.10, S. 109, Abb. 4-11) – Bosutinib **(4)**, Sunitinib **(2)**, Crizotinib **(3)** und GNE-3511 **(7)** – daraufhin überprüft, ob sie die DLK-Aktivität in dem Assay sichtbar hemmen. In diesem Zusammenhang wurde eine Verkürzung in den ersten beiden Inkubationszeiten im

Tab. 5-53 – Bedingungen für die Durchführung des ADP-Glo™ Kinase Assays. Rot: Veränderung im Vergleich zu den in den zuvor durchgeführten Assays.

Substanz	Konzentration im 12,5 µl Reaktionsansatz
DLK	0,009 µg/µl
β-Casein	0,02 µg/µl
ATP	10 µM
Testsubstanz	500 µM bzw. 125 µM (GNE-3511)
Reagenzienzusatz	Inkubationszeit danach
Inhibitorlösung	keine, sofortiger ATP-β-Casein-Zusatz (vorher: 30 min.)
ATP-β-Casein-Lösung	10 min., dann ADP-Glo-Reagenzzusatz (vorher: 60 min.)
ADP-Glo-Reagenz	40 min., dann ADP-Glo-Depletionsreagenzzusatz
ADP-Glo-Depletionsreagenz	30 min., dann Messung

Vergleich zu den Assays zuvor (s. Tab. 5-53) vorgenommen, um das darauffolgende Screening zeiteffizienter durchzuführen. Die erste Inkubationszeit nach Zugabe der Inhibitorlösung wurde gestrichen, während die zweite Inkubationszeit nach Zusatz der ATP-β-Casein-Lösung auf 10 min. statt 60 min. reduziert wurde (s. Tab. 5-53).

Während Bosutinib und Crizotinib zu einer sichtbaren Hemmung der DLK-Aktivität führten, wurde eine nicht erwartungsgemäße Zunahme der Lumineszenz mit Sunitinib und GNE-3511 beobachtet (s. Abb. 5-86). Dies könnte an einer Verstärkung des Lumineszenzsignals durch die Referenzsubstanz selbst liegen. Die Sunitiniblösung war

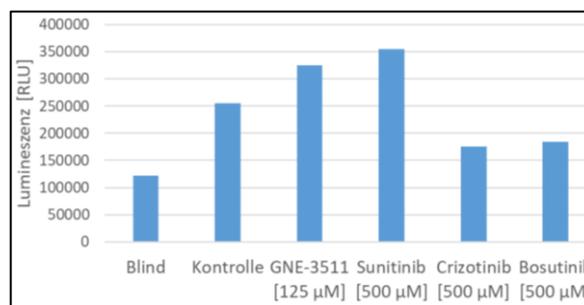


Abb. 5-86 – Lumineszenz ohne Inkubationszeit zwischen DLK und Inhibitor sowie einer lediglich 10-minütigen Inkubationszeit nach Zugabe der ATP-β-Casein-Lösung.

gelb gefärbt. GNE-3511 ist eine gelbe Festsubstanz. Sunitinib und GNE-3511 sind zwar nach Zugabe von 100 % DMSO zur Herstellung der 10 mM-(Sunitinib) bzw. 2,5 mM (GNE-3511)-Ausgangslösung (s. Kap. 7.16, S. 292) löslich, allerdings existieren für GNE-3511 kaum Daten zur Wasserlöslichkeit (s. Tab. 5-54). Der Zusatz des Puffers auf wässriger Basis bzw. die Zusätze der Enzymlösung und Reagenzien im Well könnten zur Ausfällung des GNE-3511 geführt haben (in den Wells nicht sichtbar / überprüfbar).

⁶² Dies entspricht 0,90 µl der Fraktion e im 12,5 µl Reaktionsansatz.

Tab. 5-54 – Löslichkeit, Absorptionsmaximum und Feststofffarbe der Referenzsubstanzen. Informationsquellen in eckigen Klammern (Firmenwebseiten).

Referenz-substanz	Löslichkeit in DMSO	Löslichkeit in Wasser	Max. Absorption bei Wellenlänge...	Farbe des Feststoffs
GNE-3511	31,25 mg/ml [Med-ChemExpress] 20 mg/ml [Cayman]	„keine Daten“ [SBD Cayman] Nicht löslich in Wasser [MedKoo Biosciences, Inc.]	275, 320, 358 nm [Biomol, Cayman]	gelb
Sunitinib-malat	≥ 15 mg/ml [Med-ChemExpress] 5 mg/ml [Cayman]	unlöslich [Selleckchem] Sehr schlecht löslich [LC Labs] 0,25 mg/ml in 1:3 DMSO:PBS-Puffer (pH 7,2)-Lösung [Cayman] 12,5 mg/ml (Ultraschall, pH 3) [Med-ChemExpress]	267, 427 nm [Cayman]	gelb
Crizotinib	12,5 mg/ml (Ultraschall) [MedChemExpress] 0,5 mg/ml [Cayman]	< 0,1 mg/ml (Ultraschall) [MedChemExpress] 10-20 µM, Erhöhung oder Verringerung möglich durch Puffer, andere Zusätze [Biomol]	206, 269, 323 nm [Cayman]	weiß
Bosutinib	≥ 46 mg/ml [Med-ChemExpress] 20 mg/ml [Cayman]	1 mg/ml in PBS-Puffer (pH 7,2) [Cayman] Unlöslich [SelleckChem]	269, 344 nm [Cayman]	blassgelb

Es ist auch möglich, dass das Enzym im ADP-Glo-Reagenz durch die Referenzsubstanz gehemmt wird, welches für die Zerstörung des überschüssigen ATP benötigt wird. Dann wäre zu viel ATP vorhanden, welches zu einem falsch hohen Lumineszenzsignal führen würde.

Als Referenzsubstanz für das Screening der 58 virtuellen Hits wurde daher Crizotinib (500 µM im Reaktionsansatz) eingesetzt, weil mit dieser Substanz die Hemmung der DLK am stärksten sichtbar war.

- b) Das Screening der 58 Hit-Substanzen (s. S. 222 ff., Tab. 5-52) erfolgte mit den optimierten Parametern und einer Substanzkonzentration von 500 µM im Reaktionsansatz (Durchführung s. Kap. 7.16, S. 292). Weil die einzelnen Wells nicht gleichzeitig im Luminometer gemessen werden und die Lumineszenz mit der Zeit abnimmt, wurde die Kinaselösung ohne Inhibitor erst nach den Lösungen mit Crizotinib bzw. den Testsubstanzen gemessen, um auszuschließen, dass die Lumineszenz wegen der längeren Zeit bis zur Messung erniedrigt ist, was zu einem falsch positiven Ergebnis führen würde. Jeweils zwei Reaktionsansätze mit drei Testsubstanzen, Crizotinib und einer Kontrolllösung (= DLK ohne Inhibitor) wurden in einem Durchgang gemessen. Eine DLK-inhibitorisch aktive Testsubstanz sollte in beiden Ansätzen eine verringerte Lumineszenz im Vergleich zu der Kontrolllösung aufweisen und es sollten immer auch beide Crizotinib-Ansätze eine verringerte Lumineszenz im Vergleich zu beiden Kontrollansätzen aufweisen. Durch Rechenfehler wurden für die drei Testsubstanzen **E05**, **E07** und **E08** versehentlich abweichende, leicht erniedrigte Konzentrationen bei der Messung eingesetzt (s. Tab. 5-55).

Tab. 5-55 – Abweichende Konzentrationen von 3 Testsubstanzen.

Test-substanz	Soll-Konz. in 100 % DMSO	Ist-Konz. in 100 % DMSO	Soll-Konz. in DMSO-Puffer-Mischung	Ist-Konz. in DMSO-Puffer-Mischung	Soll-Konz. im Reaktionsansatz (Assay)	Ist-Konz. im Reaktionsansatz (Assay)
E05	10 mM	8,42 mM	2500 µM	2105,35 µM	500 µM	421,07 µM
E07		6,63 mM		1658,70 µM		331,74 µM
E08		8,84 mM		2210,75 µM		442,15 µM

Die Ergebnisse des Screenings werden in Abb. 5-87 dargestellt:



Abb. 5-87 – Ergebnisse des Screenings der 58 Testsubstanzen mit dem ADP-Glo™ Kinase Assay im Vergleich zu Crizotinib als DLK-inhibitorische Substanz. „Blitz“ = Signal übersteigt die auf y-Achse maximal darstellbare Lumineszenz.

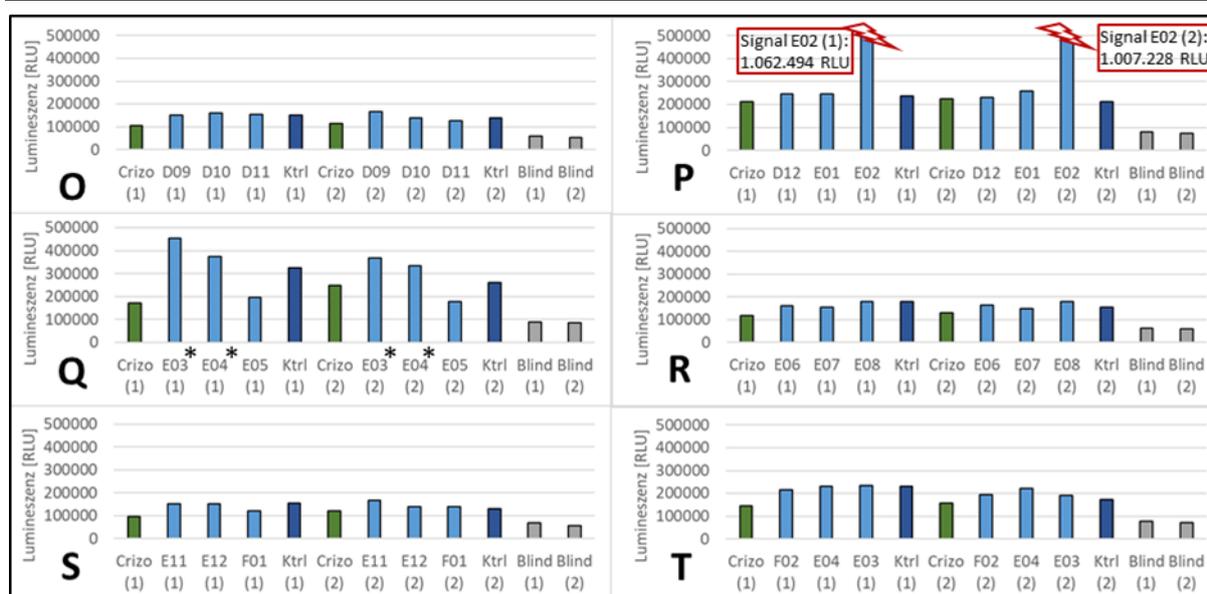


Abb. 5-87 – Fortsetzung. *Die Messdaten von **E03** und **E04** in Testreihe Q sind nicht zu berücksichtigen, weil dort zu hohe Testsubstanzkonzentrationen eingesetzt wurden.

15 der 20 Testreihen waren gut auswertbar, da jeweils beide Crizotinibsignale unter den Kontrollsignalen lagen. Dies sind die Testreihen: A, C, E, G, H, I, K, L, M, N, O, Q, R, S, T. Da in Testreihe A eine andere Messreihenfolge vorliegt, wird diese Testreihe später gesondert behandelt. In den übrigen Testreihen zeigten folgende Testsubstanzen in einem der beiden Signale ein um mehr als 10 % erniedrigtes Messsignal im Vergleich zum nachfolgenden Kontrollsignal (Testreihe in Klammern dahinter): **B03** (E), **B10** (H), **D01** (L), **D03** (M), **D07** (N), **E05** (Q), **E06** (R), **E07** (R), **F01** (S). Für **E05** (Q) wurde im Folgenden eine Verdünnungsreihe zur Ermittlung des IC₅₀-Werts erstellt, da beide Messsignale gleichermaßen besonders stark erniedrigt waren im Vergleich zu beiden Kontrollsignalen. Für **B03** (E), **B10** (H), **D03** (M), **D07** (N), **E07** (R) und **F01** (S) wurden Wiederholungsmessungen durchgeführt, während **D01** (L) und **E06** (R) nicht weiterverfolgt wurden, da sie im Mittel nur geringe Abweichungen zu den Kontrollsignalen zeigten; rückblickend hätte die Auswahl von **D07** jedoch der von **E06** vorgezogen werden müssen (s. Tab. 5-56).

Tab. 5-56 – Prozentuale Abweichungen der Messsignale der Testsubstanzen vom Kontrollsignal: **1)** Abweichung des Testsubstanz-Messsignals (1) zum Kontrollsignal (1); **2)** Abweichung des Testsubstanz-Messsignals (2) zum Kontrollsignal (2)

	B03	B10	D01	D03	D07	E06	E05	E07	F01
1)	-10,49	-17,81	-21,87	-10,00	-10,11	-10,68	-39,30	-13,88	-21,8
2)	-5,57	-27,92	22,89	-95,00	8,33	6,82	-31,76	-3,04	5,39
Mittelwert	-8,03	-22,87	0,51	-52,50	-0,89	-1,93	-35,53	-8,46	-8,21

Von Testreihe A wurden zum einen die Testsubstanz **A01** (A) wiederholt gemessen, da es sich um ein falsch erhöhtes Signal handeln könnte und beide **A01**-Proben vor dem Kontrollsignal gemessen wurden. Aus dieser Testreihe wurde auch die Messung von **A03** (A) wiederholt, da beide Signale deutlich unter dem Kontrollsignal lagen, aber zum Schluss dieser Testreihe gemessen wurden – daher könnte es sich hierbei um ein falsch niedriges Signal handeln. Bei 5 der 20 Testreihen lagen ein oder beide Crizotinib-Messsignale über einem oder beide Kontrollsignalen. Dies waren die Testreihen B, D, F, J, P. Die Testsubstanzen aus Testreihe D und J, nämlich **A10** (D), **A11** (D), **A12** (D), **C06** (J), **C07** (J) und **C08** (J) wurden wiederholt gemessen. Aus Testreihe B wurde **A04** (B) erneut gemessen, da sich die beiden **A04**-Signale sehr stark voneinander unterschieden und ein Signal (**A04** (2)) unter beiden Kontrollwerten lag. Die übrigen Substanzen aus den Testreihen B, F und P wurden nicht wiederholt gemessen, da keiner der Testsubstanzsignale ein um mehr als 10 % erniedrigtes Messsignal im Vergleich zu einem der Kontrollsignale aufwies und niemals beide Testsubstanzsignale unter einem der Kontrollmesswerte lag, sodass von einer nicht vorhandenen Aktivität ausgegangen wurde.

c) Das Re-Testing der 12 Testsubstanzen im zweiten Durchgang erfolgte in etwas veränderter Form: Es wurden nicht zwei Lösungen einer Testsubstanz ((1) + (2)) in einem Durchgang gemessen, sondern nur jeweils drei Testsubstanzen ein Mal pro Durchführung. Die Lösungs- und Reagenzienvolumina sowie die Inkubationszeiten wurden auf die gleiche Weise wie zuvor gewählt. Die Ergebnisse werden Abb. 5-88 dargestellt.

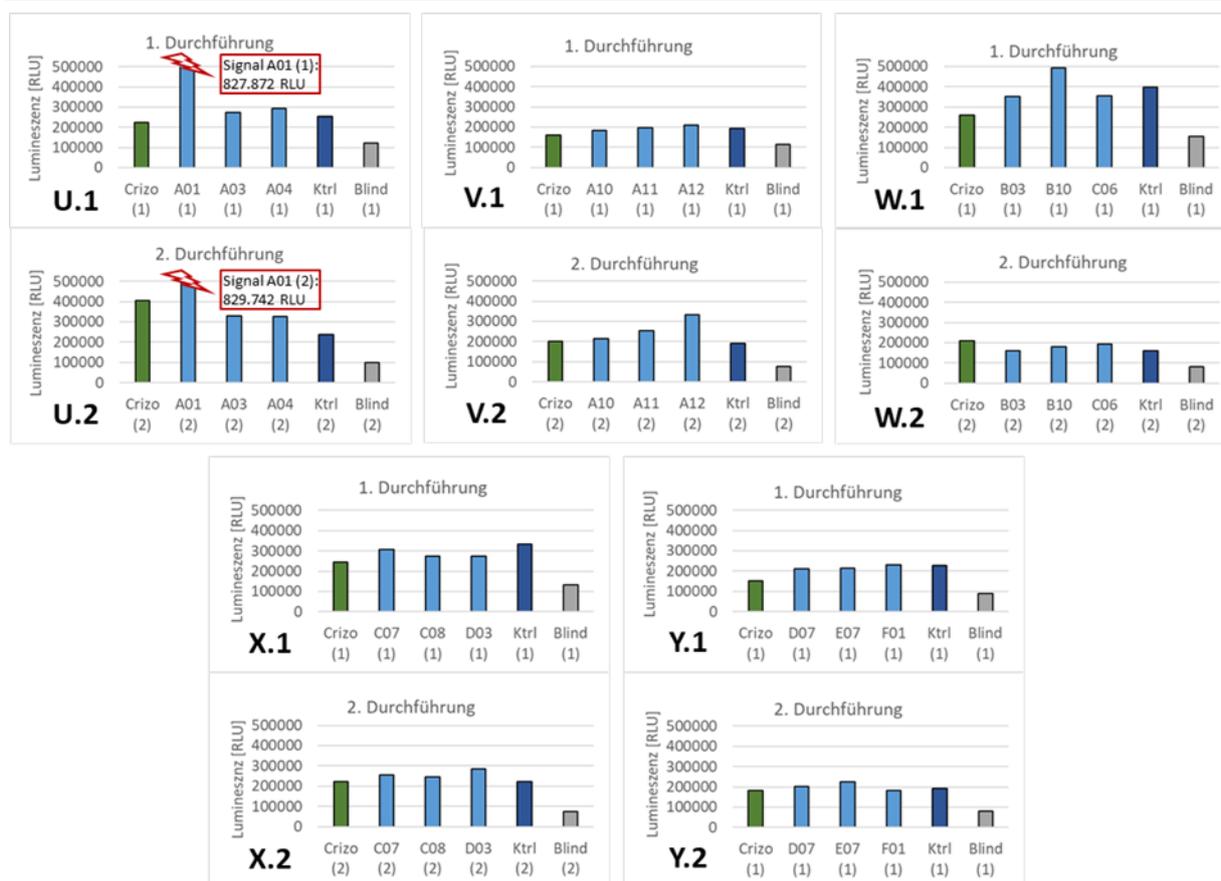


Abb. 5-88 – Ergebnisse des zweiten Screenings von 12 ausgewählten Testsubstanzen mit dem ADP-Glo™ Kinase Assay im Vergleich zu Crizotinib als DLK-inhibitorische Substanz.

Das Re-Testing der 12 Testsubstanzen ergab keine sichere Verifizierung eines Hits (s. Abb. 5-88). Die Testreihen U.2, V.2, W.2, und X.2 sind nicht auswertbar, weil auch Crizotinib hier keine verminderte Lumineszenz im Vergleich zur Kontrolle zeigte. In den auswertbaren Testreihen ist eine schwache Hemmung durch **A10**, **B03**, **C06**, **C07**, **C08**, **D03**, **D07**, **E07** und **F01** (in Y.2, aber nicht in Y.1) zu sehen. Die Ergebnisse dieser Testsubstanzen lassen aber abschließend keine deutliche Zuordnung als Hit zu.

- d) **E05** war die einzige Substanz, für die verlässlich eine Lumineszenzminderung gemessen werden konnte. Daher wurde eine **E05**- und vergleichend eine Crizotinib-Verdünnungsreihe zur IC_{50} -Bestimmung im ADP-Glo™ Kinase Assay erstellt (s. Kap. 7.16, S. 293, Tab. 7-18). Es wurden zwei Modelle zur Anpassung der Messdaten verwendet (s. Kap. 4.10.2, S. 113) und eine ANOVA-Analyse zum Vergleich der beiden Modelle durchgeführt. Liegt der p-Wert der ANOVA-Analyse über 0,05 so wird die Nullhypothese nicht abgelehnt, die besagt, dass das 5-PL-Modell dem 4-PL-Modell nicht überlegen ist.

Mit einem Signifikanzwert von über 80 % ist das 5-PL- dem 4-PL-Modell (Crizotinib) nicht überlegen (s. Tab. 5-57). Die beiden Regressionsmodelle für Crizotinib ergeben einen sehr ähnlichen Verlauf (s. Abb. 5-89) und ähnliche IC_{50} -Werte von 294,62 μM (4-PL-Modell) bzw. 329,44 μM (5-PL-Modell). Die Parameter c und d haben einen hohen Signifikanzwert von $< 0,5\%$.

Tab. 5-57 – Bestimmung des IC_{50} -Werts für E05 und Crizotinib mit dem 4-PL- und 5-PL-Modell.

E05 4-PL-Modell				Crizotinib 4-PL-Modell							
	Wert	Std.Fehler	p-Wert		Wert	Std.Fehler	p-Wert				
b	1,181	1,119	0,3688	b	5,188	3,483	0,2331				
c = min.	105190	63188	0,1946	c = min.	125830	12516	0,0021				
d = max.	189540	9462,7	0,0003	d = max.	173200	3751,3	0,00002				
e = IC_{50}	123,70	201,5	0,5827	e = IC_{50}	294,62	50,868	0,0102				
E05 5-PL-Modell				Crizotinib 5-PL-Modell							
	Wert	Std.Fehler	p-Wert		Wert	Std.Fehler	p-Wert				
b	5,030	1,057	0,6811	b	7,460	16,292	0,6920				
c = min.	-79423	651160	0,9141	c	113540	102080	0,3818				
d = max.	188650	8496,2	0,0020	d	173220	4453,2	0,0007				
e	20,192	16,022	0,3347	e	219,78	192,72	0,3723				
f	0,019	0,048	0,7314	f	0,226	1,3705	0,8842				
IC_{50}	28511,306			IC_{50}	329,430						
ANOVA-Analyse				ANOVA-Analyse							
	Freiheitsgrade	RSS	Df	F-Wert	P-Wert		Freiheitsgrade	RSS	Df	F-Wert	P-Wert
4-PL	3	389260658				4-PL	3	163064100			
5-PL	2	284724949	1	0,7343	0,4818	5-PL	2	156943448	1	0,0780	0,8063

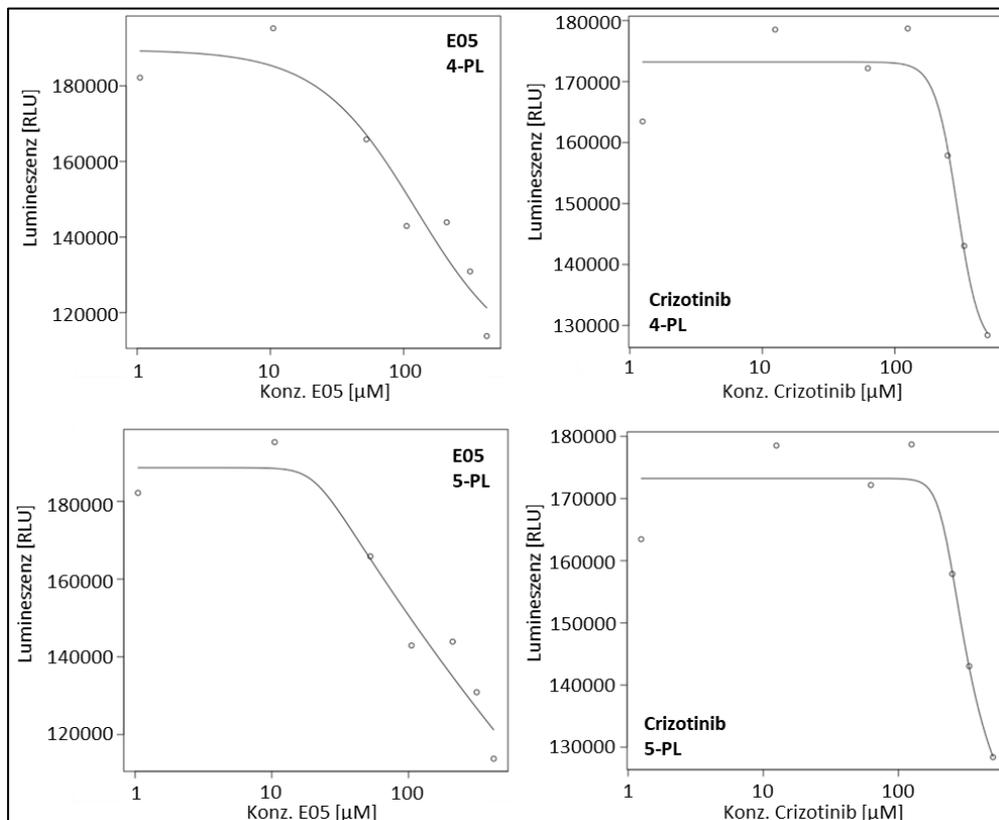


Abb. 5-89 – Bestimmung des IC_{50} -Werts für E05 und Crizotinib mit je zwei Regressionsmodellen (4-PL und 5-PL)

Die beiden Modelle ergeben für **E05** stark unterschiedliche IC_{50} -Werte (4-PL: 123,70 μM , 5-PL: 28511,31 μM , s. Tab. 5-57). Dies ist auch an den Kurvenverläufen (s. Abb. 5-89) zu erkennen: während sich bei der 4-PL-Kurve ein sigmoidaler Verlauf andeutet (sichtbare Annäherung an die c-Asymptote), ist dies bei der 5-PL-Kurve nicht zu erkennen; hier befindet sich die Kurve im Abstieg ohne sichtbare Annäherung an eine Asymptote. Nach ANOVA-Analyse ist das 5-PL-Modell dem 4-PL-Modell nicht überlegen (Signifikanzniveau 48,18 %, s. Tab. 5-57).

Sowohl für Crizotinib als auch für **E05** kann demnach das 4-PL- statt des 5-PL-Modells verwendet werden. Die relativ schlechten statistischen Daten für die einzelnen Parameter des 4-PL-Modells könnten auf hohe Zufallsschwankungen in den Daten zurückzuführen sein. Darüber hinaus wurde eine vollständige Hemmung der DLK, welche zu einem Plateau bzw. zu einem nicht weiter abnehmenden Lumineszenzsignal führen würde, nicht erreicht. Dies könnte die Anpassung an ein Modell für einen sigmoidalen Kurvenverlauf erschweren. Eine höhere Inhibitorkonzentration konnte mit der jeweiligen Testsubstanz-Stammlösung nicht erreicht werden.

Trotz problematischer Anpassung der Messdaten mit einem Kurvenmodell kann sowohl für Crizotinib als auch für **E05** eine konzentrationsabhängige Reduktion der Lumineszenz beobachtet werden. Vergleicht man die beiden 4-PL-Modelle miteinander, ist **E05** in dieser Bestimmung stärker aktiv als Crizotinib ($K_i = 0,170 \mu\text{M}$ [125]).

5.6.2 Microscale Thermophoresis

Für die Messungen mit der MST wurde die MBP-TEV-DLK₁₄₀₋₅₁₇ (Fraktion g, s. Kap. 5.6.1, S. 227) zuvor nach dem vom Hersteller vorgegebenen Protokoll [512] fluoreszenzmarkiert (s. Kap. 7.17, S. 293 f.). Das nach dem Labelingprozess aufgefangene Eluat (im Folgenden als DLK-Eluat bezeichnet) enthält nach den dort vorgenommenen Verdünnungsschritten rechnerisch eine MBP-TEV-DLK₁₄₀₋₅₁₇-Konzentration von 227,3 nM. Typischerweise wird für die MST eine Proteinkonzentration von 5-100 nM eingesetzt.

(1.) Optimierung der Enzymkonzentration:

Zur Identifikation der optimalen Enzymkonzentration für die MST-Messung wurden drei Verdünnungen des DLK-Eluats – 1,5625 %, 3,125 % und 6,25 % – gemessen. Als Puffer für die Verdünnung wurde der MST-optimierte Puffer eingesetzt (s. Kap. 7.17, S. 294, Tab. 7-19). Anhand der Fluoreszenzintensität sollte die optimale DLK-Konzentration für die weiteren Tests ausgewählt werden. Als optimal gilt eine Fluoreszenzintensität zwischen 200 und 1500 Fluorocounts [511]. Des Weiteren ist ein gleichmäßiger Peak im Kapillarscan erstrebenswert.

Die Fluoreszenzzeitspuren (s. Abb. 5-90) zeigten einen gleichmäßigen Verlauf ohne Sprünge oder Unebenheiten. Die Peaks des Kapillarscans sind gleichmäßig und symmetrisch (s. Abb. 5-91).

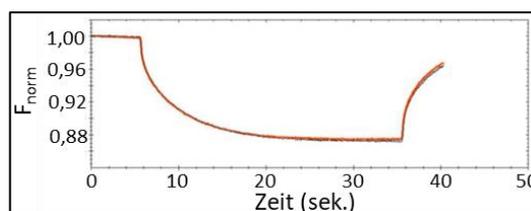
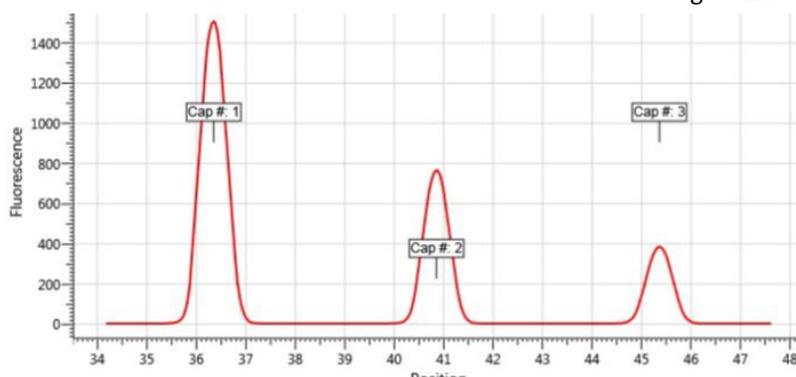


Abb. 5-90 – Normalisierte Fluoreszenzzeitspuren der drei unterschiedlich konzentrierten DLK-Lösungen. LED-Power 50 %, MST-Power 40 %.



Cap #	DLK-Eluat-Anteil	Fluoreszenzintensität
1	6,25 %	1499,51
2	3,125 %	764,48
3	1,5625 %	385,13

Abb. 5-91 – Kapillarscan der drei unterschiedlich konzentrierten DLK-Eluat-Puffer-Mischungen. Cap #1: 6,25 % DLK-Eluat-Anteil. Cap #2: 3,125 % DLK-Eluat-Anteil. Cap #3: 1,5625 % DLK-Eluat-Anteil. Angegeben sind die absoluten Fluoreszenzintensitäten.

Als optimale DLK-Eluat-Konzentration wurde die Probe mit einem Anteil von 3,125 % ausgewählt, da deren Signal mit 764,48 Fluorocounts im angestrebten, optimalen Bereich zwischen 200 und 1500 liegt (s. Abb. 5-91). Nach dieser Verdünnung handelt es sich um eine „rechnerische“ MBP-

TEV-DLK₁₄₀₋₅₁₇-Konzentration von 7,1 nM. Diese Konzentration liegt in dem vorgesehenen Bereich von 5-100 nM für die MST.

(2.) *Präzision des Messsignals:*

In einem weiteren Versuch zur Vorbereitung und Optimierung des MST-Assays wurde untersucht, wie groß die Abweichungen der Messsignale zwischen einzelnen Kapillaren mit Lösungen gleicher Zusammensetzung sind. Nach Herstellervorgaben sollten die Fluoreszenzintensitäten einer Testlösung bei Vergleichsmessungen nicht stärker als +/- 10 % vom Mittelwert abweichen. [511]

Hierzu wurde eine DLK-Puffer-Mischung aus 5,5 µl DLK-Eluat und 82,5 µl MST-optimiertem Puffer hergestellt. 10 µl der vorbereiteten DLK-Puffer-Mischung, 9 µl MST-optimierter Puffer und 1 µl DMSO (Lösungsmittel für die Inhibitoren) wurde in jeweils acht Kapillaren gefüllt. Der Anteil des DLK-Eluats betrug 3,125 %. Zwei weitere Kapillaren wurden als Blindproben (19 µl Puffer und 1 µl DMSO, ohne DLK) vermessen.

Alle Fluoreszenzintensitäten lagen innerhalb der geforderten Grenzen von $\bar{x} \pm 10\%$ (s. Abb. 5-92). Die größten Abweichungen zum Mittelwert betragen -4,08 % bzw. +2,38 %. Die Verwendbarkeit des Testsystems ist diesbezüglich gegeben.

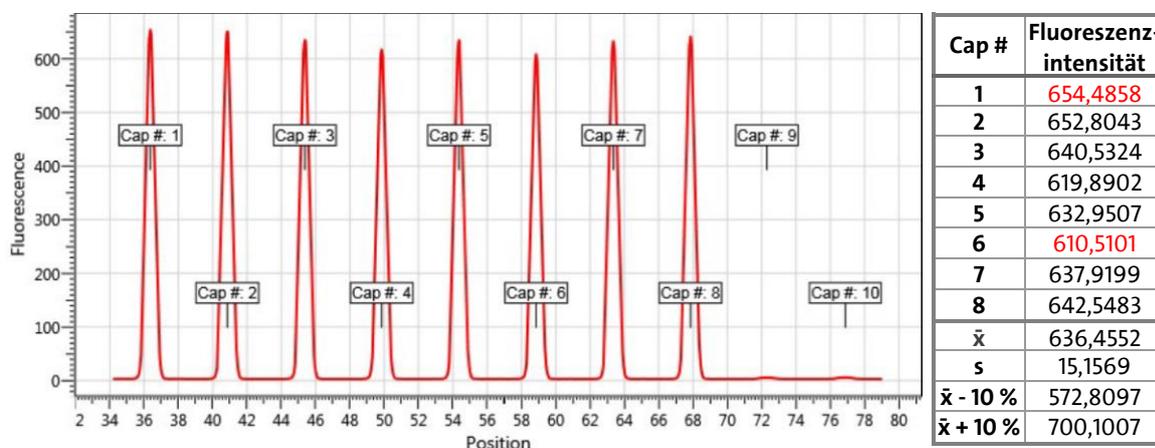


Abb. 5-92 – Fluoreszenzintensitäten der gleichartig befüllten Kapillaren 1-8 und der Blindproben (Kapillaren 9 und 10). Min. und Max. sind rot dargestellt (Tabelle rechts); die erlaubten Abweichungen von $\bar{x} \pm 10\%$ werden nicht überschritten.

(3.) Messung einer Verdünnungsreihe von Crizotinib

Mit dem in (1.) bestimmten optimalen DLK-Eluatanteil von 3,125 % wurde eine Verdünnungsreihe von Crizotinib zur Bestimmung des K_d -Werts gemessen. Der Fluoreszenzsprung erfolgte nach Anschalten des IR-Lasers umso stärker, je verdünnter die Crizotiniblösung war (s. Abb. 5-93).

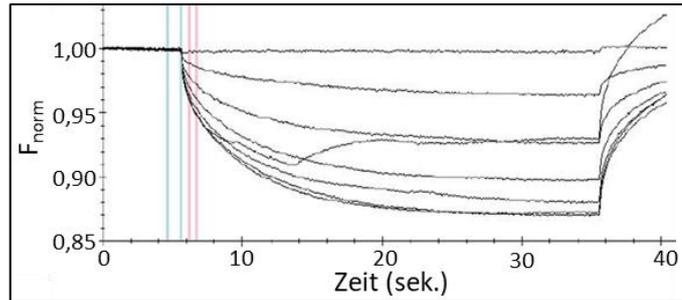
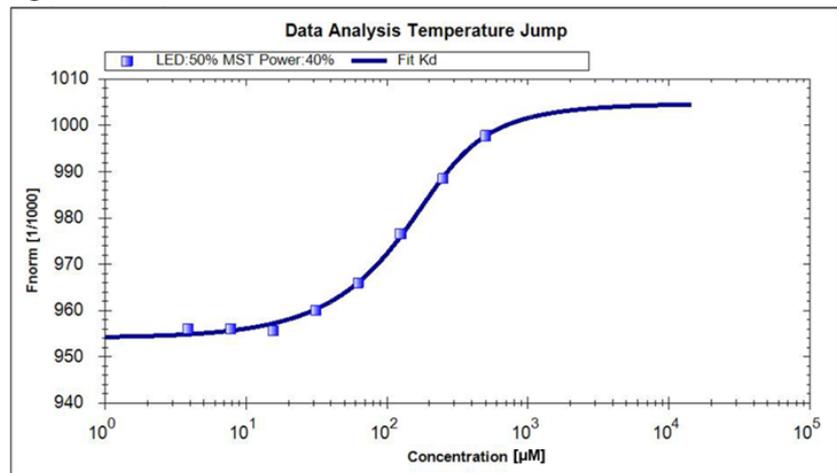


Abb. 5-93 – Normalisierte Fluoreszenzzeitkurve der Crizotinibverdünnungsreihe mit DLK; LED-Power: 50 %, MST-Power: 40 %.

Durch die Bindung des Crizotinibs an die DLK verändert sich somit die Beweglichkeit des Protein-Liganden-Komplexes konzentrationsabhängig. Nach Auswertung mit dem K_d -Modell der Software MO.AFFINITY ANALYSIS (s. Kap. 4.10.3, S. 117) ergibt sich ein K_d -Wert von $53,2 \pm 2,96 \mu\text{M}$ (s. Abb. 5-94).



Fitting for Kd Formular

Fitted Parameter	Fitted Value
Dissociation Constant	53.2 +/- 2.96
Fluo.Conc	195
Bound	1004.70
Unbound	954.07
Amplitude	50.63

Abb. 5-94 – Graphische Darstellung von F_{norm} gegenüber der Crizotinibkonzentration. Das Fitting der Messpunkte (s. Kap. 4.10.3) führt zu einer sigmoide verlaufenden Kurve, von der die Bindungskonstante K_d abgeleitet werden kann.

(4.) Testung der virtuellen Hits:

a) Für das Screening der 58 Testsubstanzen wurde eine Substanzkonzentration von $500 \mu\text{M}^{63}$ in der zu messenden Lösung eingesetzt (s. Kap. 7.17, S. 295).

⁶³ Drei Substanzen (E05, E07, E08) wurden mit einer abweichenden Konzentration eingesetzt (s. Tab. 5-55, S. 232).

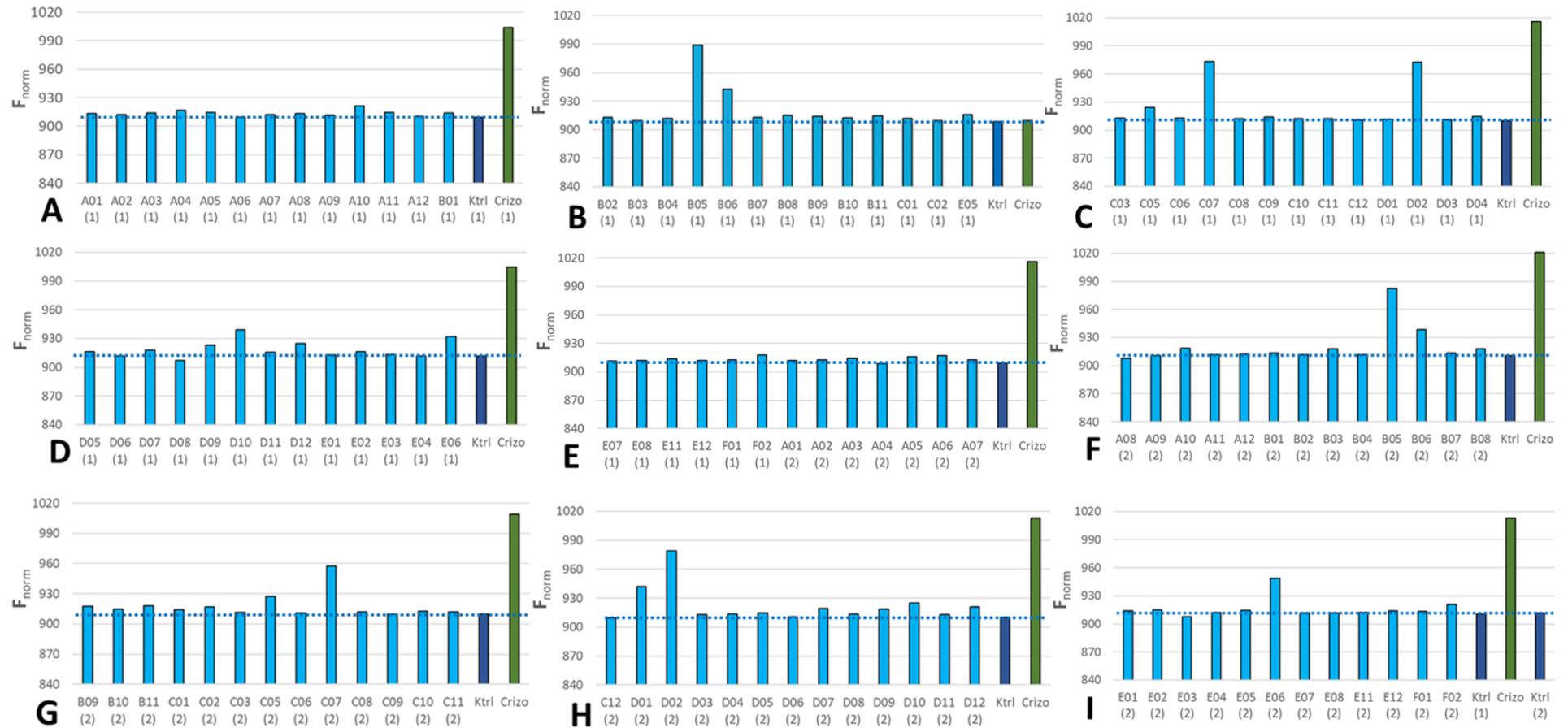


Abb. 5-95 – Ergebnisse des Screenings der 58 Testsubstanzen (hellblau) mit der *Microscale Thermophoresis* (MST) im Vergleich zu Crizotinib (grün). Ktrl = Kontrolllösung (DLK-Puffer-DMSO-Mischung ohne Test- bzw. Referenzsubstanz).

Tab. 5-58 – Betrag der prozentualen Abweichung des Messsignals zum Kontrollwert der gleichen Testreihe (in der Testreihe I wurde nur der erste von beiden Kontrollwerten verwendet). Die fettgedruckten Testsubstanzen zeigten in beiden Messungen eine Abweichung von > 1,00 % zur Kontrolllösung ohne Inhibitor.

A01(1)	0,45	A10(1)	1,30	B07(1)	0,50	C06(1)	0,30	D03(1)	0,11	D12(1)	1,49	E11(1)	0,56
A01(2)	0,35	A10(2)	0,87	B07(2)	0,28	C06(2)	0,13	D03(2)	0,28	D12(2)	1,20	E11(2)	0,19
A02(1)	0,27	A11(1)	0,53	B08(1)	0,75	C07(1)	6,93	D04(1)	0,48	E01(1)	0,17	E12(1)	0,33
A02(2)	0,43	A11(2)	0,08	B08(2)	0,80	C07(2)	5,25	D04(2)	0,40	E01(2)	0,39	E12(2)	0,41
A03(1)	0,48	A12(1)	0,10	B09(1)	0,60	C08(1)	0,20	D05(1)	0,52	E02(1)	0,54	F01(1)	0,43
A03(2)	0,58	A12(2)	0,14	B09(2)	0,89	C08(2)	0,24	D05(2)	0,49	E02(2)	0,50	F01(2)	0,30
A04(1)	0,80	B01(1)	0,48	B10(1)	0,43	C09(1)	0,40	D06(1)	0,003	E03(1)	0,22	F02(1)	0,94
A04(2)	0,06	B01(2)	0,32	B10(2)	0,53	C09(2)	0,02	D06(2)	0,08	E03(2)	0,28	F02(2)	1,11
A05(1)	0,53	B02(1)	0,46	B11(1)	0,70	C10(1)	0,24	D07(1)	0,73	E04(1)	0,005		
A05(2)	0,79	B02(2)	0,13	B11(2)	0,90	C10(2)	0,34	D07(2)	0,97	E04(2)	0,21		
A06(1)	0,01	B03(1)	0,10	C01(1)	0,38	C11(1)	0,19	D08(1)	0,48	E05(1)	0,81		
A06(2)	0,88	B03(2)	0,79	C01(2)	0,49	C11(2)	0,25	D08(2)	0,35	E05(2)	0,47		
A07(1)	0,31	B04(1)	0,36	C02(1)	0,08	C12(1)	0,01	D09(1)	1,25	E06(1)	2,27		
A07(2)	0,39	B04(2)	0,10	C02(2)	0,79	C12(2)	0,09	D09(2)	0,95	E06(2)	4,21		
A08(1)	0,41	B05(1)	8,86	C03(1)	0,29	D01(1)	0,13	D10(1)	3,02	E07(1)	0,29		
A08(2)	0,36	B05(2)	7,86	C03(2)	0,22	D01(2)	3,51	D10(2)	1,64	E07(2)	0,12		
A09(1)	0,26	B06(1)	3,73	C05(1)	1,53	D02(1)	6,86	D11(1)	0,45	E08(1)	0,35		
A09(2)	0,007	B06(2)	3,06	C05(2)	1,96	D02(2)	7,54	D11(2)	0,30	E08(2)	0,12		

Die Messsignale der Kontroll- und Referenz-Lösungen waren gut reproduzierbar (s. Abb. 5-95). Lediglich in der Testreihe B konnte kein von der Kontrolllösung abweichendes Crizotinib-Messsignal beobachtet werden. Dennoch wurde diese Testreihe mit in die Bewertung einbezogen, weil die zweite Messung der Testsubstanzen **B02** bis **C02** (s. Testreihe F und G, Abb. 5-95) sowie **E05** (s. Testreihe I, Abb. 5-95) zu fast identischen Ergebnissen wie in Testreihe B führte. Crizotinib führte im Mittel zu einer Abweichung von der Kontrolllösung ohne Inhibitor von 9,97 % ($s = 3,75$), ohne den offensichtlichen Ausreißer in Testreihe B 11,20 % ($s = 0,67$). Die Kontroll-Messsignale schwankten lediglich maximal um +0,151 % bzw. -0,177 % um den Mittelwert (F_{norm} min. 908,533, max. 911,519, $\bar{x} = 910,148$). Acht Testsubstanzen hatten in beiden Messungen eine Abweichung von mehr als 1,00 % von dem jeweiligen Kontrollwert der gleichen Testreihe (s. Tab. 5-58): **B05**, **B06**, **C05**, **C07**, **D02**, **D10**, **D12** und **E06**.

- b) Für die besten sieben dieser Substanzen (= alle außer **D12**) wurden Verdünnungsreihen zur Bestimmung eines K_d -Werts erstellt und gemessen. Zusätzlich wurde dies für **E05** durchgeführt, da diese Substanz im ADP-Glo™ Kinase Assay als Hit erschien und man in diesem MST-Screening zumindest eine geringfügige Interaktion beobachten konnte (Abweichung im Mittel von 0,64 % vom Kontrollmesssignal). Die Verdünnungsreihen wurden ebenso für die Referenzsubstanzen Crizotinib, Sunitinib, GNE-3511 und Bosutinib erstellt. Es wurde eine Dreifachbestimmung mit der ersten Messung (Messung 1) im Zeitraum vom 25.10.-22.11.19 und den Messungen 2 und 3 im Zeitraum vom 02.12.-03.12.19 durchgeführt.

Tab. 5-59 – Auswertung der Messdaten mit dem K_d -Modell. SER: Standard Error of Regression. K_d -Konf.: 68%iges Konfidenzintervall des K_d -Werts. S/N: Signal-Rausch-Verhältnis.

Substanz	K_d [μ M]	K_d Konf. [μ M]	SER	Amplitude	Unbound	Bound	S/N
Crizotinib (1)	46,88	$\pm 19,29$	5,36	72,25	928,19	1000,44	15,96
Crizotinib (2)	49,02	$\pm 19,52$	5,25	73,22	929,65	1002,86	16,50
Crizotinib (3)	50,22	$\pm 19,35$	4,99	72,01	929,12	1001,128	17,08
Bosutinib (1)	-	-	-	-	-	-	-
Bosutinib (2)	64,96	$\pm 59,20$	1,26	7,85	930,09	937,94	7,36
Bosutinib (3)	49995,00	$\pm 1020562,97$	0,35	714,22	931,83	1646,05	2427,50
Sunitinib (1)	40,98	$\pm 40,88$	11,66	64,85	931,69	996,54	6,58
Sunitinib (2)	63,64	$\pm 18,93$	3,69	70,23	933,29	1003,51	22,52
Sunitinib (3)	89,08	$\pm 35,43$	4,60	69,14	934,87	1004,01	17,79
GNE-3511 (1)	1,04	$\pm 1,58$	2,46	8,49	933,03	941,51	4,08
GNE-3511 (2)	19,94	$\pm 10,63$	1,22	18,31	935,89	954,20	17,82
GNE-3511 (3)	0,28	$\pm 1,11$	1,82	10,35	928,74	939,09	6,72
B05 (1)	261,19	$\pm 36,71$	1,09	73,31	935,36	1008,67	79,29
B05 (2)	251,43	$\pm 29,01$	0,92	73,10	935,02	1008,13	94,24
B05 (3)	360,94	$\pm 77,49$	1,51	84,13	935,58	1019,71	65,85
B06 (1)	410,37	$\pm 134,48$	1,94	48,57	934,87	983,44	48,12
B06 (2)	142,86	$\pm 79,80$	2,65	32,75	934,70	967,46	32,75
B06 (3)	34,37	$\pm 13,48$	2,50	35,65	928,54	964,19	16,90
C05 (1)	-	-	-	-	-	-	-
C05 (2)	2027,66	$\pm 14222,00$	0,62	9,18	934,38	943,56	17,39
C05 (3)	94,80	$\pm 107,56$	1,42	7,61	932,37	939,98	6,32
C07 (1)	137,81	$\pm 168,54$	7,28	40,47	933,48	973,95	6,58
C07 (2)	50,96	$\pm 50,14$	5,22	29,54	931,11	960,66	6,69
C07 (3)	88,91	$\pm 104,75$	5,57	28,27	933,34	961,61	6,00
D02 (1)	754,97	$\pm 533,50$	1,68	59,83	933,49	993,32	42,22
D02 (2)	296,45	$\pm 126,44$	1,84	44,24	931,43	975,67	28,50
D02 (3)	158,88	$\pm 31,60$	1,00	36,19	930,74	966,94	42,83
D10 (1)	-	-	-	-	-	-	-
D10 (2)	978,99	$\pm 1128,98$	0,40	12,03	931,39	943,42	35,96
D10 (3)	-	-	-	-	-	-	-
E05 (1)	31,69	$\pm 44,67$	0,94	3,75	932,43	936,18	4,73
E05 (2)	44,81	$\pm 47,22$	0,87	6,58	931,24	935,82	6,23
E05 (3)	81,89	$\pm 76,37$	1,01	6,37	931,26	937,62	7,46
E06 (1)	524,71	$\pm 65,72$	0,31	41,47	933,99	975,47	159,02
E06 (2)	442,49	$\pm 106,98$	0,76	44,78	932,02	976,80	69,83
E06 (3)	2117,96	$\pm 1446,37$	0,40	65,37	932,18	997,55	191,71

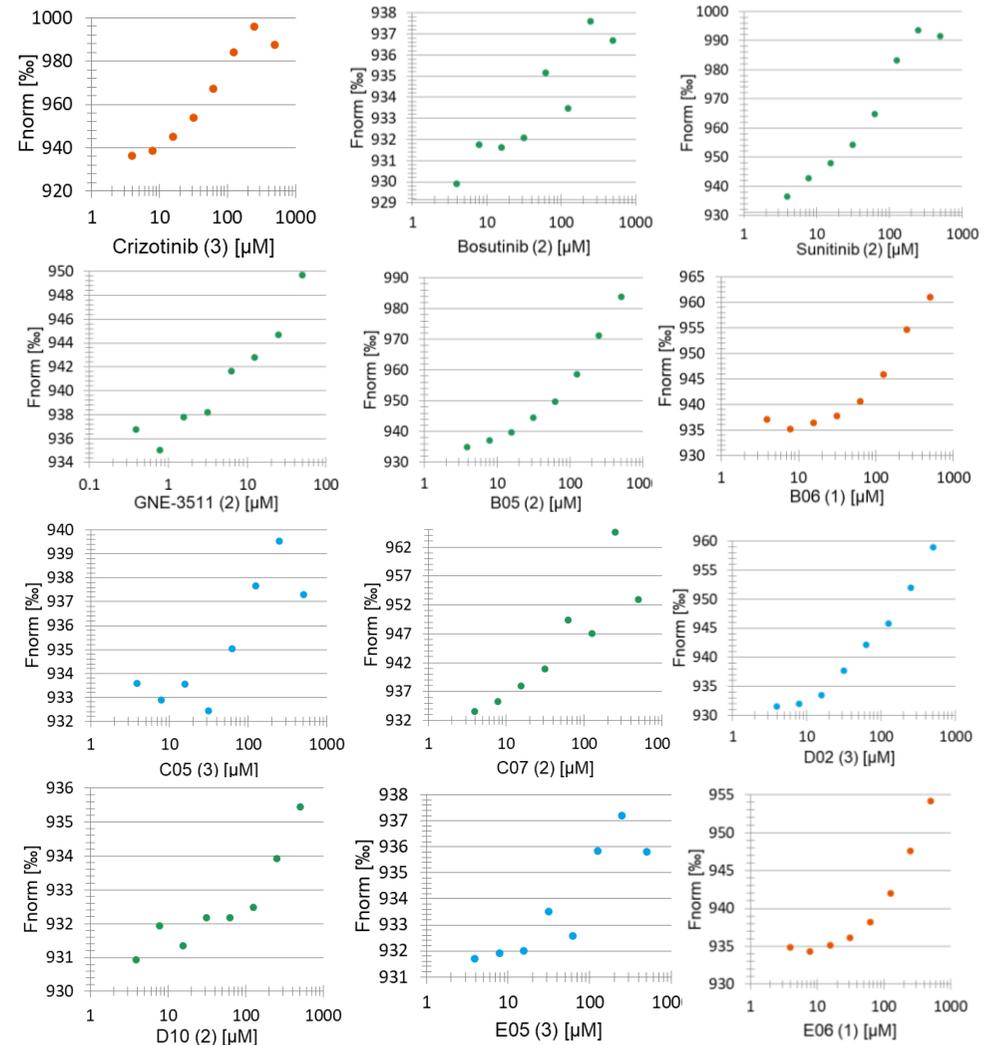


Abb. 5-96 – Messdaten einer ausgewählten Testreihe pro Substanz. Dargestellt ist jeweils das Modell mit geringstem Anteil des 68%- K_d -Konfidenzintervalls am K_d -Wert.

Nur für wenige Substanzen gelang eine gute Reproduktion des K_d -Werts innerhalb der Dreifachmessung (s. Tab. 5-59). Für Crizotinib, Sunitinib und **B05** ergaben sich relativ gut reproduzierbare Daten.

In Abb. 5-96 sind die Messdaten derjenigen Testreihe pro Substanz mit dem geringstem Anteil des 68%- K_d -Konfidenzintervalls am K_d -Wert dargestellt. Das K_d -Konfidenzintervall der MO.AFFINITY ANALYSIS-Software (K_d Konf, s. Tab. 5-59) gibt den Bereich an, in dem der K_d -Wert mit einer 68 %igen Wahrscheinlichkeit liegt. Ein besonders geringer Wert des Konfidenzintervalls spricht für eine größere Verlässlichkeit des K_d -Werts. In Tab. 5-59 sind diejenigen Werte grau markiert, die für die Mittelwertbildung (s. Tab. 5-60) berücksichtigt wurden. Nicht berücksichtigt wurden demnach Ausreißer, welche ein sehr großes Konfidenzintervall besaßen. Die Referenzsubstanzen lassen sich demnach in folgende Reihenfolge nach ihrem ermittelten K_d -Wert bringen:

Tab. 5-60 – K_d -Mittelwerte.

Substanz	$\bar{x} K_d [\mu M]$
GNE-3511	0,66
Crizotinib	48,71
E05	52,80
Sunitinib	64,57
Bosutinib	64,96
C07	92,56
C05	94,80
B06	195,87
D02	227,66
B05	291,19
E06	483,60
D10	978,99

GNE-3511 ($K_d = 0,66 \mu M$) > **Crizotinib** ($K_d = 48,71 \mu M$) > **Sunitinib** ($K_d = 64,57 \mu M$) \approx **Bosutinib** ($K_i = 64,96 \mu M$).

Nach ihrem literaturbekannten pK_i -Wert für die DLK wäre die Reihenfolge:

GNE-3511 ($K_i < 0,501 nM$) > **Sunitinib** ($K_i = 0,100 \mu M$) > **Crizotinib** ($K_i = 0,170 \mu M$) > **Bosutinib** ($K_i = 0,676 \mu M$).

Lediglich Sunitinib und Crizotinib ließen sich durch die hier betrachtete Auswertung nicht in die korrekte Reihenfolge bringen. Allerdings sind sowohl die Werte aus der Literatur als auch die mit der MST gemessenen Werte sehr ähnlich untereinander, sodass die Abweichung möglicherweise auf die allgemein eingeschränkte Reproduzierbarkeit biologischer Daten beruht.

Die Testsubstanzen ließen sich nach dieser Auswertung in folgende absteigende Reihenfolge ihrer Bindungsaffinität bringen: **E05** > **C07** > **C05** > **B06** > **D02** > **B05** > **E06** > **D10**. Dass **E05** mit der K_d -Bestimmung als die am stärksten bindende Substanz interpretiert werden kann, diese Substanz allerdings im MST-Screening mit der Einzelmessung (s. Abb. 5-95, S. 241) nicht als bindende Substanz eingestuft worden wäre, deutet darauf hin, dass die Einzelmessungen allein die Bindungsstärke nicht ausreichend erfassen können.

Aufgrund der unzureichenden Genauigkeit und Reproduzierbarkeit der IC_{50} -Messungen im ADP-Glo™ Kinase Assay und der K_d -Messungen mit der MST wurden ausgewählte Testsubstanzen zusätzlich extern mit der KINOMEScan®-Methode (s. Kap. 4.10.4 u. 5.6.3) untersucht.

5.6.3 KINOMEScan®

Um die Ergebnisse des ADP-Glo™ Kinase Assays (s. Kap. 5.6.1) und der MST (s. Kap. 5.6.2) zu überprüfen und um die Bindung der Hit-Substanzen in der ATP-Bindetasche der vollständigen, humanen DLK zu untersuchen, wurden 36 der 58 virtuellen Hits sowie die Referenzsubstanzen Bosutinib, Crizotinib, GNE-3511 und Sunitinib zunächst im scanELECT-Assay mit der KINOMEScan®-Methode untersucht. Die Durchführung erfolgte durch die Firma Eurofins DiscoverX Corporation in San Diego.

Die Auswahl der 36 Testsubstanzen erfolgte zum einen auf Basis der Vorergebnisse: So wurden der im ADP-Glo™ Kinase Assay gefundene Hit **E05** sowie die nicht vollends bestätigten Hits mit schwach reduzierten Messsignalen (**B03**, **C07**, **D03** und **E07**, s. S. 232 ff.) mit in die engere Auswahl genommen. Darüber hinaus wurden die Hits aus der MST ausgewählt: **C07**, **C05**, **D02**, **B05**, **B06**, **E06** und **D10**. In Tab. 5-52 (S. 222 ff.) ist hinterlegt, welche der 58 Substanzen für den KINOMEScan®-Assay (scanELECT-Assay) ausgewählt wurden (Überschrift: „Assay“ → 3). Bei der weiteren Auswahl wurden wiederum die in Kap. 5.5.5 beschriebenen Parameter verglichen sowie die Interaktionen zur Bindestelle, insbesondere zur Hinge-Region berücksichtigt. Dies führte z.B. dazu, dass Substanzen, deren Dockingposen keine starke Interaktion mit der Hinge-Region zeigten (s. Abb.

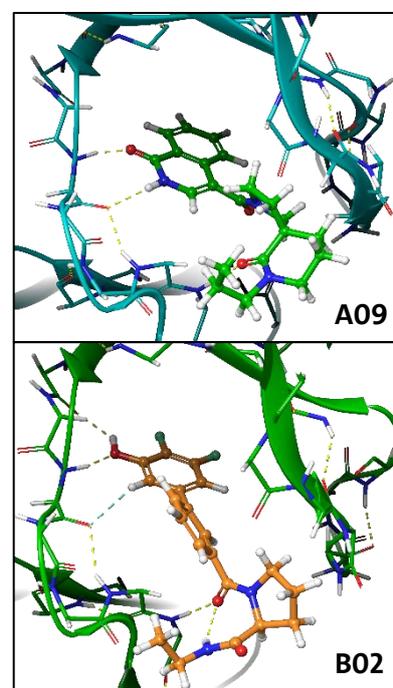


Abb. 5-97 – Liganden mit weniger starken Interaktionen zur Hinge-Region wurden aussortiert.

5-97, **A09**) bzw. für die Hinge-Interaktion eher untypische Hydroxylgruppen trugen (s. Abb. 5-97, **B02**) nicht ausgewählt wurden. Darüber hinaus wurden zueinander ähnliche Substanzen herausgefiltert: Der Core 1 wie in der Substanz **C09 (84)** (s. Abb. 5-98, blau) erschien bei der Auswahl der 58 Substanzen interessant und acht Substanzen wiesen dieses Strukturelement auf. Für den KINOMEScan® wurden die Substanzen **C08**, **C10** und **D01** mit dem gleichen Core 1 entfernt, sodass noch fünf Substanzen mit diesem Core 1 übrigblieben. Diese sind: **C09 (84)**, **D04**, **E01**, **E03** und **E06** (positiver MST-Test).

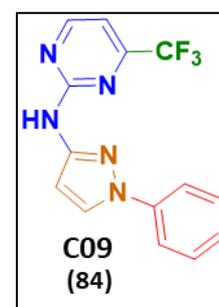


Abb. 5-98 – C09 (84)

Mit den in dieser Arbeit hergestellten 10 mM-Lösungen der Test- und Referenzsubstanzen in 100 % DMSO (außer GNE-3511: 1 mM-Lösung in 100 % DMSO) konnte der Assay maximal mit einer 10 µM-Konzentration der Substanz (bzw. 1 µM für GNE-3511) durchgeführt werden. Wiederum ist zu beachten, dass die Konzentration der Testsubstanzen **E05**, **E07** und **E08** davon abweicht (s. Kap. 5.6.1, Tab. 5-55, S.232).

Die Ergebnisse dieser Einpunktmessungen sind in Tab. 5-61 aufgeführt. Die Referenzsubstanzen zeigten erwartungsgemäß eine vollständige Bindung an die DLK (0 % Ctrl). Sechs Testsubstanzen konnten zu mehr als 25 % im Vergleich zu einer Kontrollsubstanz an die ATP-Bindetasche der DLK binden (% Ctrl < 75 %): **A04**, **C07**, **C12**, **D02**, **D04** und **E06**. Drei von diesen Substanzen zeigten bereits im MST ein Bindungssignal: **C07**, **D02** und **E06**.

E05 konnte als Hit in dieser hier eingesetzten Konzentration von 8,4 µM nicht bestätigt werden. Man beachte, dass auch im ADP-Glo™ Kinase Assay **E05**-Konzentrationen kleiner als 10 µM nicht zu einer Lumineszenzabnahme führten (s. S. 236, Abb. 5-89). Eine Abnahme der Lumineszenz bei Messung der **E05**-Verdünnungsreihe erfolgte hier ab einer Konzentration von 26,32 µM. Allerdings sind diese Vergleiche schwierig, da die eingesetzten DLK-Konzentrationen unbekannt und die Funktionsweise der beiden Assays unterschiedlich sind. Darüber hinaus ist die Beobachtung einer Hemmung im ADP-Glo™ Kinase Assay auch von der eingesetzten ATP-Konzentration abhängig.

Bei sechs Testsubstanzen mit einem % Ctrl-Signal < 75 % wurden zusätzlich mit dem KdELECT®-Assay

von Eurofins DiscoverX K_d -Werte bestimmt. Da hierbei eine Startkonzentration von 30 µM

Tab. 5-61 – Ergebnisse des scanELECT-Assays [% Ctrl-Wert]. Im Vergleich dazu sind die Hits aus dem ADP-Glo™ Kinase Assay und aus der MST markiert.

Substanz	Konz. [µM]	% Ctrl	ADP-Glo	MST
A01	10,0	96		
A02	10,0	92		
A03	10,0	100		
A04	10,0	66		
A05	10,0	85		
A06	10,0	100		
A07	10,0	88		
A11	10,0	98		
A12	10,0	87		
B01	10,0	92		
B03	10,0	81		
B04	10,0	90		
B05	10,0	95		x
B06	10,0	97		x
B07	10,0	84		
B08	10,0	86		
B09	10,0	86		
B10	10,0	86		
C05	10,0	92		x
C07	10,0	41		x
C09	10,0	100		
C12	10,0	49		
D02	10,0	71		x
D03	10,0	100		
D04	10,0	30		
D10	10,0	100		x
E01	10,0	100		
E02	10,0	100		
E03	10,0	79		
E04	10,0	100		
E05	8,4	100	x	
E06	10,0	57		x
E07	6,6	92		
E08	8,8	100		
E11	10,0	99		
F01	10,0	100		
Bosutinib	10,0	0		
Crizotinib	10,0	0		
Sunitinib	10,0	0,95		
GNE-3511	1,0	0		

mit den vorliegenden Testsubstanzlösungen möglich war, wurden zusätzlich auch die Testsubstanzen **B06**, **C05** und **E05** vermessen, auch wenn sie bei der 10 μM -scanELECT-Messung kein nennenswertes Signal zeigten. Für die K_d -Bestimmung wurde eine neue Lösung für **E05** (10 μM) eingesetzt, da die Konzentration von **E05** in den vorherigen Experimenten im Vergleich zu den anderen Testsubstanzen zu gering konzentriert war.

Die Lösungen wurden für die Aufnahme der 11-Punkte-Verdünnungsreihe von 30 μM ausgehend in 1:3-Verdünnungsschritten hergestellt. Für jede Substanz wurde eine Doppelbestimmung durchgeführt. Vier der neun Substanzen zeigten K_d -Werte von unter 30 μM (s. Tab. 5-62, Messpunkte s. Abb. 5-99).

Tab. 5-62 – Ergebnisse des KdELECT®-Assays

Substanz	K_d (1) [μM]	K_d (2) [μM]	K_d (gesamt) [μM]
A04	19	29	24
B06	> 30	> 30	> 30
C05	> 30	> 30	> 30
C07	11	11	11
C12	> 30	> 30	> 30
D02	> 30	> 30	> 30
D04	2,6	2,9	2.8
E05	> 30	> 30	30
E06	2,7	2,3	2.5

Die Bindungsaffinität zur ATP-Bindetasche der humanen DLK fällt innerhalb der Hitreihe von **E06** > **D04** > **C07** > **A04** ab.

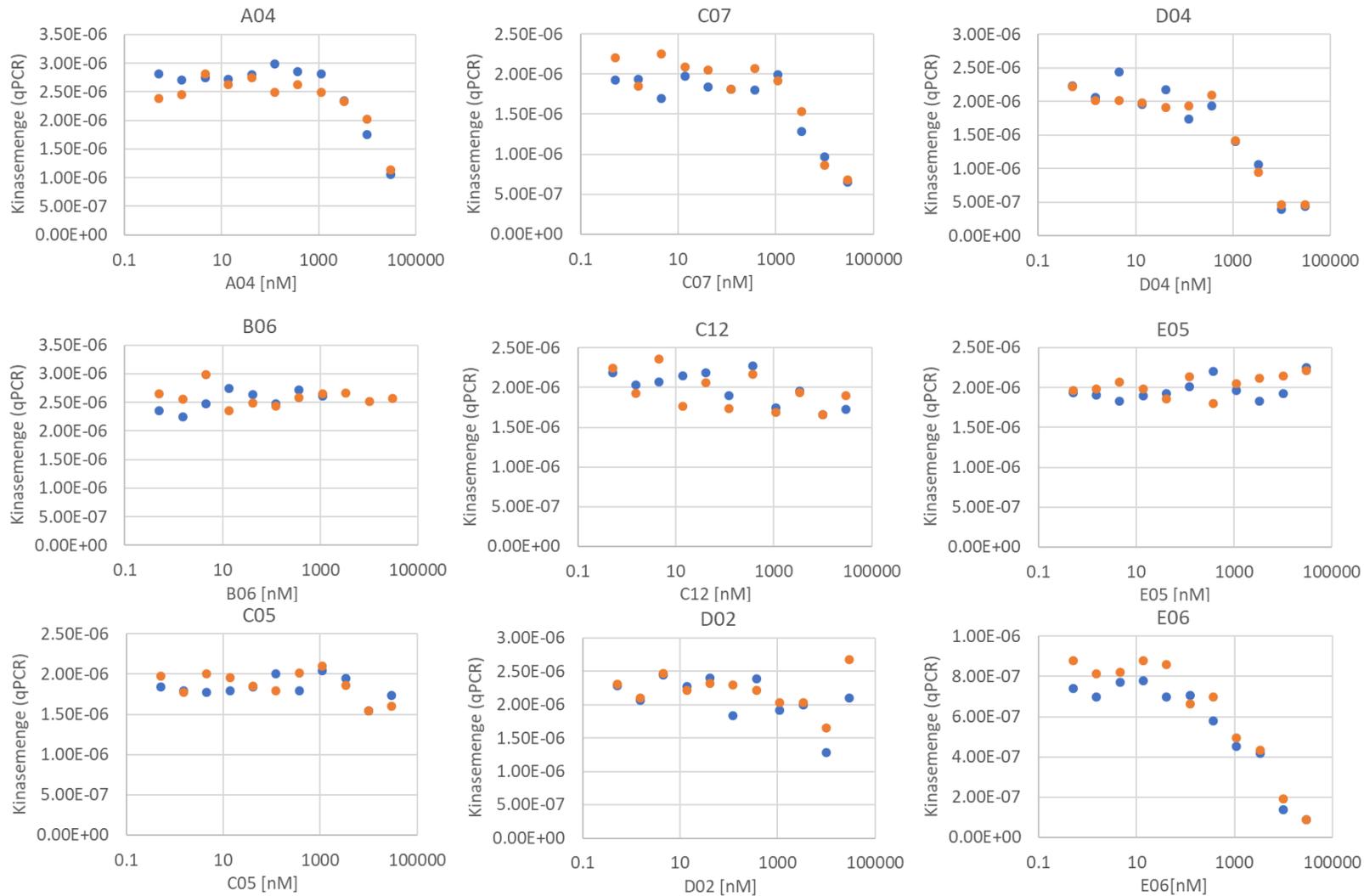


Abb. 5-99 – Messdaten der KdELECT®-Assays zur Bestimmung des K_d -Werts. **A04**, **C07**, **D04** und **E06** zeigen eine konzentrationsabhängige Erniedrigung der mittels qPCR gemessenen Kinasemenge.

6 Diskussion

Die mithilfe der biologischen Testung identifizierten Hits sollen im Folgenden genauer diskutiert werden.

Die Ergebnisse der biologischen Testungen der virtuellen 58 Hits waren sehr heterogen. Zum einen brachte die MST noch weitere Hits hervor als der ADP-Glo™ Kinase Assay, zum anderen bestätigte der KINOMEScan®-Assay nur wenige der MST-Hits und brachte weitere, andere Hits hervor. In diesem Kapitel liegt der Fokus auf den Hits der KINOMEScan®-Messungen sowie auf der Substanz **E05**, welche im ADP-Glo™ Kinase Assay als Hit identifiziert und für die in der MST eine Bindungskurve erstellt werden konnte.

Die Ergebnisse für diese fünf Substanzen waren in den drei Assays sehr unterschiedlich (s. Tab. 6-1). Mit der KINOMEScan®-Methode wurden für vier Substanzen K_d -Werte

Tab. 6-1 – K_d bzw. IC_{50} der Hits aus den biologischen Prüfungen. n.b. = nicht bestimmt.

Hit	KINOMEScan® (KdELECT®)	ADP-Glo™ Kinase Assay	MST
E06	$K_d = 2,5 \mu\text{M}$	n.b.	$K_d = 483,60 \mu\text{M}$
D04	$K_d = 2,8 \mu\text{M}$	n.b.	n.b.
C07	$K_d = 11 \mu\text{M}$	n.b.	$K_d = 69,93 \mu\text{M}$
A04	$K_d = 24 \mu\text{M}$	n.b.	n.b.
E05	$K_d > 30 \mu\text{M}$	$IC_{50} = 123,70 \mu\text{M}$ (4-PL)	$K_d = 52,80 \mu\text{M}$

unter $30 \mu\text{M}$ ermittelt (**A04**, **C07**, **D04** und **E06**). Aufgrund der Messmethode kann vermutet werden, dass sie in der ATP-Bindetasche der DLK binden. **E05** hat hingegen einen K_d -Wert über $30 \mu\text{M}$. Daraus könnte gefolgert werden, dass die Aktivität tatsächlich deutlich geringer ist oder dass **E05** gar nicht in der ATP-Bindetasche bindet.

E05 konnte nur im ADP-Glo™ Kinase Assay und im MST-Assay (K_d -Wertbestimmung) als Hit identifiziert werden. Zwei der vier KINOMEScan®-Hits wurden auch mit der MST als Hit erfasst. Nach ihrem K_d -Wert im KINOMEScan® können die fünf Hits in folgende Reihenfolge mit abnehmender Bindungsaffinität zur ATP-Bindetasche der DLK gebracht werden: **E06** > **D04** > **C07** > **A04** (> **E05**, kein Hit im KINOMEScan®).

Tab. 6-2 – Scorewerte und weitere Parameter, die zur Hitauswahl herangezogen wurden. **Cons.** = Consensuscore nmr_DHQM, **D** = Dockingscore, **H** = HYDE-Score, **Q** = vorhergesagter pK_i-Wert mit dem *Sel2*-QSAR-Modell. **M** = Freie Bindungsenergie (PRIME MM/GBSA). **I** = IFP-Tanimotokoeffizient. **L** = *Ligand Strain Energy*. **ID** = Proteinkonformation des Komplexes mit dem besten Consensuscore. **#IDs docked** = Anzahl der Proteinkonformationen, in die diese Struktur SP-gedockt werden konnte.

Hit	Cons.	D	H	Q	M	I	L	MW	ID	#IDs docked	K _d (KINOMEScan®) [µM]
E06	-2,729	-9,459	-43,313	7,865	-56,026	0,651	1,861	322,29	33	6	2,5
D04	-2,569	-10,252	-43,174	7,316	-55,582	0,836	1,171	308,26	6.3	6	2,8
C07	-2,481	-10,021	-42,404	6,997	-58,173	0,656	1,735	336,35	6.11	6	11
A04	-2,817	-9,583	-42,484	7,955	-60,196	0,642	5,592	394,50	6.3	6	24
E05	-2,422	-10,381	-29,368	7,524	-60,936	0,673	2,377	361,43	13	6	> 30

Betrachtet man die Scorewerte und weitere Parameter, die zur Hitauswahl geführt haben (s. Tab. 6-2), gibt es folgende Besonderheiten:

- HYDE-Score (**H**): Von allen Scorewerten ist der HYDE-Score am besten dazu in der Lage, die Hits in die gleiche Reihenfolge wie mit dem K_d-Wert der KINOMEScan®-Messungen zu bringen. Das zeigt, dass es sinnvoll wäre, den HYDE-Score bei zukünftigen Untersuchungen stärker zu gewichten. Der HYDE-Score für **E05** ist deutlich schlechter als für die anderen vier Verbindungen; dies passt ebenso zu den KINOMEScan®-Ergebnissen.
- IFP-Tanimotokoeffizient (**I**): Die Interaktionen mit der Bindetasche ähneln bei **D04** am ehesten den Interaktionen des gedockten **5CEO**-Liganden (IFP = 0,836).
- *Ligand Strain Energy* (**L**) und Molekulargewicht (**MW**): **D04** steht im Vergleich zu den anderen Hits in der Bindetasche am wenigsten unter „Spannung“ (**L** = 1,171), dies könnte auch mit seinem geringen Molekulargewicht zusammenhängen. Das vergleichsweise geringe Molekulargewicht von **D04** ist eine gute Voraussetzung für eine Aktivitätsoptimierung durch Anhängen funktioneller Gruppen. **A04** hingegen wird am stärksten defomiert und ist das größte Molekül.

Die Reinheit von **A04**, **C07**, **D04**, **E05** und **E06** sowie von Crizotinib wurde mittels HPLC untersucht (s. Kap. 7.18). Alle Substanzen besaßen bei einer Wellenlänge von 267 nm eine über HPLC ermittelte Reinheit von mindestens 88 % (s. Kap. 7.18 u. Anhang I). Die Substanzen unterschieden sich recht stark in ihrer Reinheit (s. Tab. 6-3). Insbesondere **E05** weist eine geringe Reinheit auf. **E05** lag demnach in den Messungen vergleichsweise niedriger konzentriert vor als

Tab. 6-3 – Ermittelte Reinheit der Substanzen (Mittelwert aus Doppelbestimmung)

Substanz	Gehalt
A04	94,05 %
C07	98,49 %
D04	93,84 %
E05	88,09 %
E06	95,86 %
Crizotinib	99,95 %

die anderen hier analysierten Substanzen. Nimmt man dann noch hinzu, dass nur das S-Steereoisomer als virtueller Hit ausgewählt wurde, bei den Messungen aber das Racemat vorlag, so kann vermutet werden, dass die DLK-Hemmaktivität bzw. die Bindungsaffinität von **E05** an die DLK stärker sein könnte als in den Ergebnissen dieser Arbeit dargestellt.

Im Folgenden werden die fünf Hits strukturell und hinsichtlich ihrer Scorewerte im Vergleich zu anderen Clustermitgliedern genauer untersucht.

E06

E06 (105) ($K_d = 2,5 \mu\text{M}$) hat die stärkste Bindungsaffinität unter den in dieser Arbeit gefundenen Hits. Der Ligand hat gewisse Ähnlichkeiten im Aufbau zu den bereits bekannten DLK-Inhibitoren: auch hier sind Core 1 und Core 2 über eine Aminofunktion verknüpft, es gibt einen kleinen Substituenten R1 (Trifluormethylgruppe, typisch für DLK-Inhibitoren) am Core 1 in *meta*-Position zur Amino-Verknüpfungsstelle und einen Substituenten R2 an Core 2, welcher mit dem P-Loop interagiert (s. Abb. 6-1).

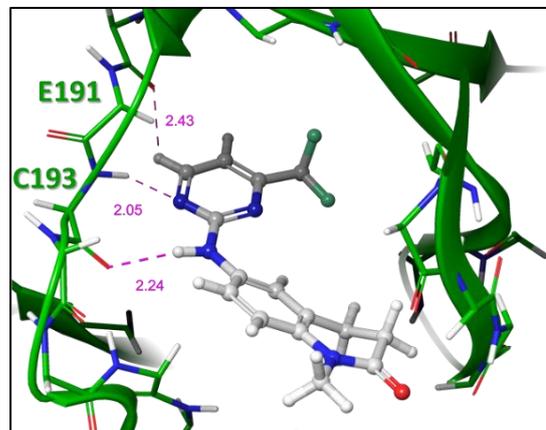


Abb. 6-1 – SP-Dockingpose von **E06** in ID 33

Anders als in den bekannten DLK-Inhibitoren ist R2 Teil des Core-2-Ringsystems. Bei den bekannten DLK-Inhibitoren gibt es als Core 1 auch Pyrimidine, allerdings nicht wie in **E06 (105)** mit der Verknüpfungsstelle über die 2. Position zu Core 2 (s. Abb. 6-2). Anders als bei den bekannten DLK-Inhibitoren fehlt der zum Lösungsmittel orientierte Substituent R3 am Core 2-Ringsystem. Core 2 in **E06 (105)** ist ein Carbozyklus, während die literaturbekannten DLK-Inhibitoren immer einen stickstoff- oder schwefelhaltigen Core 2 enthalten.

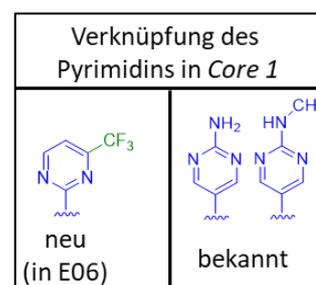


Abb. 6-2 – Unterschied zwischen der Verknüpfung des Core-1-Pyrimidins in **E06 (105)** und in literaturbekannten DLK-Inhibitoren

Unabhängig von der Cluster-Methode („OMW“ bzw. „MW<350“) sind zwei weitere Substanzen im **E06**-Cluster enthalten (s. Abb. 6-3 u. Tab. 6-4). Verbindung #2 (**119**) unterscheidet sich strukturell nur durch den Rest R1 von **E06 (105)**, während sich Verbindung #3 (**120**) im anellierten Ring, welcher mit dem P-Loop interagiert, unterscheidet. Die Posen von #2 (**119**) und #3 (**120**) sind mit der Pose von

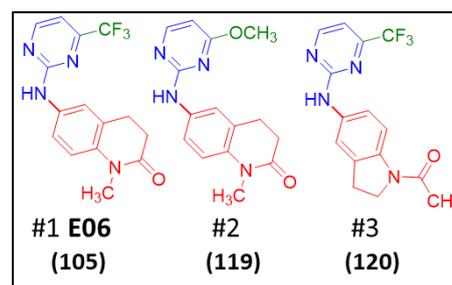


Abb. 6-3 – Clustermitglieder des Clusters #173 (**E06 (105)** als Repräsentant)

E06 (105) vergleichbar. Die vier Scorewerte **D**, **H**, **Q** und **M** der **E06**-Dockingpose sind alle besser als für die anderen beiden Ligandenposen. Ein deutlicher Unterschied ist im HYDE-Score zu sehen (s. Tab. 6-4).

Tab. 6-4 – Übersicht über Mitglieder des Clusters #173 (**E06 (105)** als Repräsentant). **Cons.** = Consensuscore nmr_DHQM, **D** = Dockingscore, **H** = HYDE-Score, **Q** = vorhergesagter pK_i-Wert mit dem *Sel2*-QSAR-Modell. **M** = Freie Bindungsenergie (PRIME MM/GBSA). **L** = *Ligand Strain Energy*. **ID** = Proteinkonformation des Komplexes mit dem besten Consensuscore.

Struktur	Cons.	D	H	Q	M	L	MW	ID
#1 E06 (ZINC171564710)	-2,729	-9,459	-43,313	7,865	-56,026	1,861	322,29	33
#2 ZINC171564668	-1,859	-9,047	-30,692	6,677	-53,915	1,575	284,32	33
#3 ZINC24054470	-1,775	-9,011	-31,023	6,545	-53,651	1,237	322,29	33

Die vergleichsweise deutlich geringeren Consensus-Scorewerte von #2 (**119**) und #3 (**120**) lassen auf eine viel geringere Bindungsaffinität als **E06 (105)** schließen, obwohl nur geringe strukturelle Unterschiede bestehen. Durch Drehen der Indolstruktur in Verbindung #3 (**120**) um die Einfachbindung der Core-1-Core-2-Verknüpfungsstelle ließe sich eine zu #1 (**105**) und #2 (**119**) ähnliche Ausrichtung der hydrophoberen Ringregion zum P-Loop erzeugen.

D04

D04 (91) ($K_d = 2,8 \mu\text{M}$) ist **E06 (105)** strukturell sehr ähnlich, obwohl beide Substanzen zu unterschiedlichen Clustern zugeordnet wurden. Core 1, Core 2 und R1 sind identisch zu **E06 (105)**, während R2 kein anelliertes System ist, sondern aus einem an Core 2 einfach gebundenen Azetidinon besteht (s. Abb. 6-4).

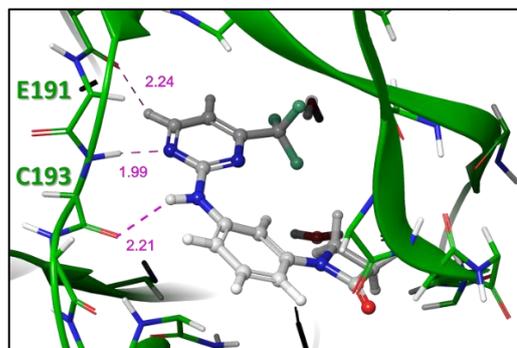


Abb. 6-4 – SP-Dockingpose von **D04 (91)** in ID 6.3

D04 (91) repräsentiert ein Cluster mit zwei weiteren Mitgliedern (**121**) und (**122**) (identisch in „OMW“- und „MW<350“-Clustering). Diese sind in Core 1, Core 2 und R1 identisch zu **D04 (91)** (s. Abb. 6-5). Sie unterscheiden sich lediglich in dem zum P-Loop gerichteten Rest R2, welcher darüber hinaus in den beiden Vertretern in *para*- statt in *meta*-Position zur Core-1-Verknüpfungs-

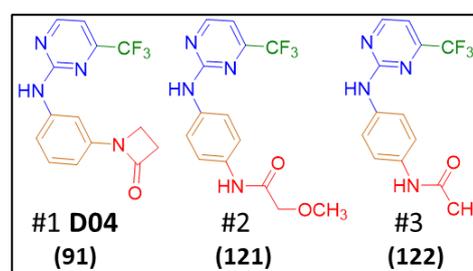


Abb. 6-5 Clustermitglieder des Clusters #366 (**D04 (91)** als Repräsentant)

stelle an Core 2 gebunden ist. Die Dockingpose von **D04 (91)** wird mit allen Scorewerten besser bewertet als die Posen der übrigen beiden Clustermitglieder (s. Tab. 6-5). Die Clustermitglieder liegen bezüglich ihrer Struktureinheiten so wie **D04 (91)** in der Bindetasche.

Tab. 6-5 – Übersicht über Mitglieder des Clusters #366 (**D04 (91)** als Repräsentant). **Cons.** = Consensuscore nmr_DHQM, **D** = Dockingscore, **H** = HYDE-Score, **Q** = vorhergesagter pK_i-Wert mit dem *Sel2*-QSAR-Modell. **M** = Freie Bindungsenergie (PRIME MM/GBSA). **L** = *Ligand Strain Energy*. **ID** = Proteinkonformation des Komplexes mit dem besten Consensuscore.

Struktur	Cons.	D	H	Q	M	L	MW	ID
#1 D04 (ZINC408515154)	-2,569	-10,252	-43,174	7,316	-55,582	1,171	308,26	6.3
#2 ZINC44724969	-2,333	-9,642	-39,445	7,168	-54,956	2,989	326,28	13
#3 ZINC24054009	-2,139	-9,944	-37,329	7,218	-52,44	1,983	296,25	13

C07

C07 (82) ($K_d = 11 \mu\text{M}$) interagiert über ein Pyrazolo[1,5-*a*]pyrimidin als Core 1 mit der Hinge-Region über drei Wasserstoffbrückenbindungen (s. Abb. 6-6). Eine Methylgruppe (R1) ragt in die tiefe Tasche. Über eine Amidgruppe mit Core 1 verbunden interagiert ein Phenylrest mit dem P-Loop. Dieser Phenylrest ist in *meta*-Position zur Amidgruppe mit einem Imidazolidinonring substituiert, welcher nach außen zum Lösungsmittel gerichtet ist.

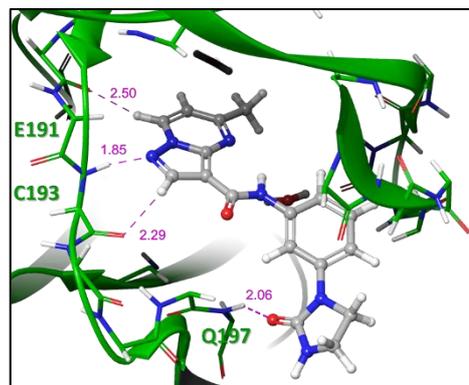


Abb. 6-6 – SP-Dockingpose von **C07 (82)** in ID 6.11

Zwar kommt dieser hier beschriebene Core 1 auch in zwei der literaturbekannten DLK-Inhibitoren vor (s. Abb. 6-7), allerdings ist der Core 1 in den bekannten DLK-Inhibitoren (**123**) und (**124**) über den Pyrimidinring mit Core 2 verknüpft, während der Core 1 des **C07 (82)** am Pyrazolring verknüpft ist. Darüber hinaus gibt es keinen Core 2 in **C07 (82)**, welcher in den bekannten DLK-Inhibitoren immer als Aromat vorliegt. In **C07 (82)** könnte man eventuell die Amidgruppe als Core 2 bezeichnen.

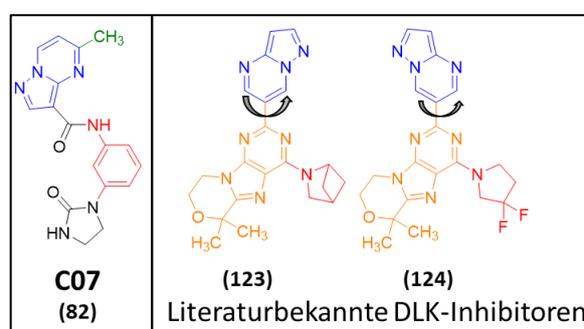


Abb. 6-7 – Unterschied zwischen der Core 1-Verknüpfung in **C07 (82)** und den bekannten DLK-Inhibitoren Nr. 38 (**123**) und 83 (**124**) aus Patent WO 2016/142310 [175]. Es ist nicht belegt, wie Core 1 zur Hinge-Region ausgerichtet ist (Rotation möglich), allerdings lässt sich nicht die Ausrichtung des Core 1 zur Hinge-Region wie in **C07 (82)** erreichen.

C07 (82) repräsentiert ein Cluster von 24 („OMW-Clustering“) bzw. 20 Vertretern („MW<350-Clustering“, alle auch im Cluster des OMW-Clustering enthalten). Alle bis auf eine Struktur dieser weiteren Vertreter tragen anstelle des Wasserstoffatoms, welcher die hintere Wasserstoffbrückenbindung zu Glu191 ausbildet, eine Methylgruppe (s. Abb. 6-8).

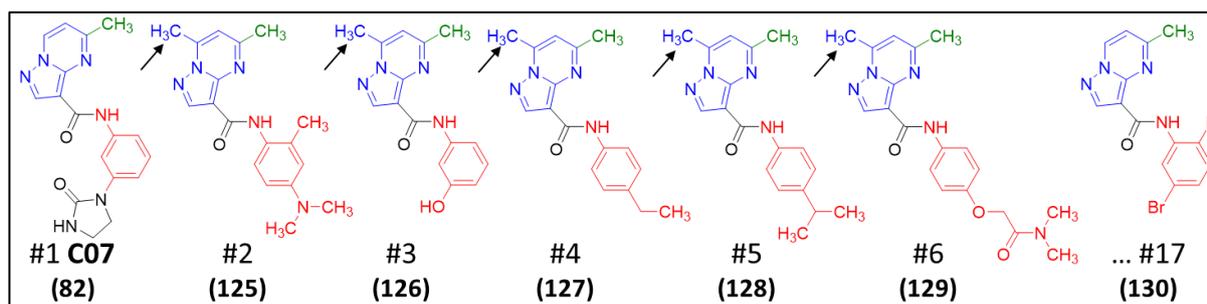
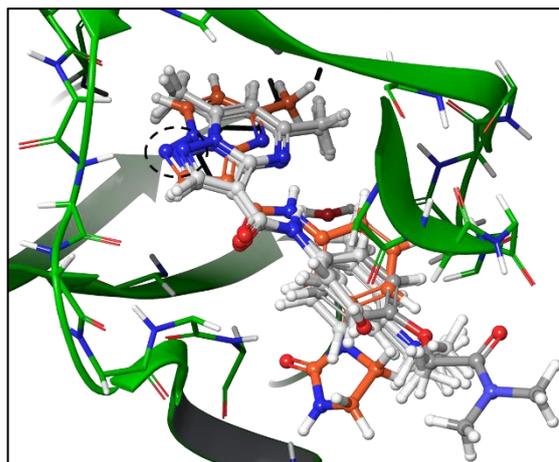


Abb. 6-8 – Die sechs bestbewerteten Strukturen des Clusters #528 (Nummerierung s. Tab. 6-6) sowie eine weitere Struktur #17 (**130**) des Clusters, die wie **C07 (82)** auch keine Methylgruppe in Richtung der Hinge-Region trägt.

Durch die Methylgruppe kommt es zu einer geringfügigen Verschiebung der Pose weg von der Hinge-Region (s. Abb. 6-9), wodurch es zu einem geringfügig größeren Abstand des wasserstoffbrückenbildenden Liganden-Stickstoffatoms zum Hinge-C193(N) kommt (**C07 (82)**: 2,85 Å zwischen den schweren Atomen; Clustervertreter #2 bis #6 im Mittel 3,09 Å).



Diese methylierten Strukturen wären nicht für eine biologische Prüfung ausgewählt worden, da die Methylgruppe, anders als ein aromatisches

Abb. 6-9 – Pose des **C07 (82)** (orange) im Vergleich zu den Clustervertretern #2 (**125**) bis #6 (**129**) (grau). Das N-Atom (s. gestrichelter Kreis) von #2-6 (**125-129**) liegt weiter von der Hinge-Region weg als das N-Atom von **C07 (82)**.

Wasserstoffatom, keine Wasserstoffbrückenbindung zur Hinge-Region ausbildet. Die tendenziell schlechteren Scorewerte der an dieser Stelle methylierten Verbindungen im Vergleich zu **C07 (82)** könnten damit erklärt werden (s. Tab. 6-6).

Interessanterweise ist die *Ligand Strain Energy* nur für **C07 (82)** und die weitere, nicht an dieser Stelle methylierten Verbindung #17 (**130**) viel geringer (d.h. besser) als bei den übrigen Substanzen (s. Tab. 6-6), was auf eine weniger stark deformierte Ligandenpose von **C07 (82)** und #17 (**130**) in der Bindetasche schließen lässt.

Tab. 6-6 – Übersicht über die nach Consensuscore besten 6 Mitglieder des Clusters #528 (**C07 (82)** als Repräsentant) und einer weiteren, wie **C07 (82)** nicht in Hinge-Richtung methylierten Struktur #17 (**130**). **Cons.** = Consensuscore nmr_DHQM, **D** = Dockingscore, **H** = HYDE-Score, **Q** = vorhergesagter pK_i-Wert mit dem *Sel2*-QSAR-Modell. **M** = Freie Bindungsenergie (PRIME MM/GBSA). **L** = *Ligand Strain Energy*. **ID** = Proteinkonformation des Komplexes mit dem besten Consensuscore.

Struktur	Cons.	D	H	Q	M	L	MW	ID
#1 C07 (ZINC65614203)	-2,481	-10,021	-42,404	6,997	-58,173	1,735	336,35	6.11
#2 ZINC40111636	-2,315	-9,279	-40,831	7,737	-54,238	3,534	323,40	18
#3 ZINC820987	-1,845	-9,726	-32,415	6,779	-54,562	4,196	282,30	13
#4 ZINC21922690	-1,757	-8,877	-39,865	7,213	-53,552	3,110	294,36	18
#5 ZINC25659286	-1,702	-8,847	-37,242	7,153	-54,757	3,077	308,38	18
#6 ZINC32795790	-1,546	-8,776	-42,717	6,556	-57,728	3,678	367,41	18
#17 ZINC75136817	-0,512	-8,612	-31,897	6,643	-52,026	1,380	349,16	6.11

A04

In **A04 (55)** ($K_d = 24 \mu\text{M}$) sind zwei aromatische Ringe (Core 1 und Core 2) über eine Aminofunktion verbunden, ein kleiner Rest (R1, hier ein Methylrest) ragt in die tiefe Tasche hinein, während ein Pyrrolidinrest (R2) mit dem P-Loop interagiert (s. Abb. 6-10). Die für DLK-Inhibitoren typischen drei Wasserstoffbrückenbindungen zur Hinge-Region, ausgehend von Core 1 und der Brücken-Aminofunktion, werden ausgebildet. Anders als bei den bekannten DLK-Inhibitoren ist der zum Lösungsmittel orientierte Rest nicht in

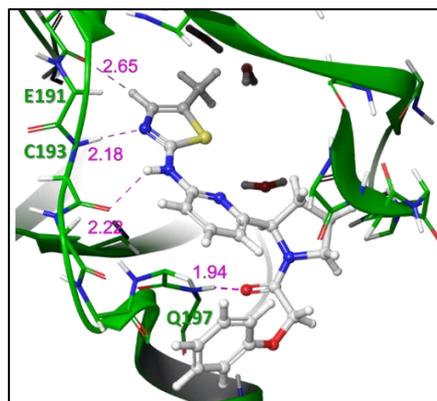


Abb. 6-10 – SP-Dockingpose von **A04 (55)** in ID 6.3

meta-Position zur Amino-Verknüpfungsstelle am Core 2-Rest gebunden, sondern ist an R2 geknüpft. Ein Thiazolring als Core 1 war auch im ersten HTS-Hit von PATEL ET AL. enthalten, während er in den darauffolgend entwickelten Strukturen nicht mehr vorkommt. Der Pyridinring als Core 2 ist in 16,5 % der bekannten DLK-Inhibitoren vorhanden. Der R3-Substituent, welcher zum Lösungsmittel herausragt, ist nicht wie gewöhnlich mit Core 2, sondern mit R2 verknüpft.

A04 (55) ist Repräsentant eines Clusters von 36 Strukturen („OMW-Clustering“). Die gemeinsame Grundstruktur (**131**) von 16 dieser Verbindungen – an Position 12 in der nach Consensuscore sortierten Liste des Clusters – ist in Abb. 6-11 dargestellt. Weiteren 16 Strukturen fehlt an dieser Grundstruktur lediglich die Methylgruppe (R1).

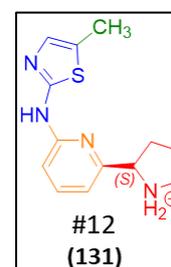


Abb. 6-11 – Struktur aus Cluster #100 mit dem geringsten MW

Die im Consensuscore besten sechs Strukturen (inkl. **A04 (55)**) und deren Scorewerte sind in Tab. 6-7 aufgeführt. Man sieht, dass **A04 (55)** vergleichsweise groß ist und vor allem einen guten vorhergesagten pK_i -Wert durch das QSAR-Modell hat. Struktur #2 (**132**) unterscheidet sich nur durch eine Methylgruppe (R1) von Struktur #5 (**135**) (s. Abb. 6-12).

Tab. 6-7 – Übersicht über die nach Consensuscore besten 6 Mitglieder des Clusters #100 (**A04 (55)** als Repräsentant). **Cons.** = Consensuscore nmr_DHQM, **D** = Dockingscore, **H** = HYDE-Score, **Q** = vorhergesagter pK_i -Wert mit dem *Sel2*-QSAR-Modell. **M** = Freie Bindungsenergie (PRIME MM/GBSA). **L** = Ligand Strain Energy. **ID** = Proteinkonformation des Komplexes mit dem besten Consensuscore.

Struktur	Cons.	D	H	Q	M	L	MW	ID
#1 A04 (ZINC72375192)	-2,817	-9,583	-42,484	7,955	-60,196	5,592	394,50	6.3
#2 ZINC72383068	-2,755	-10,297	-41,969	7,558	-58,551	2,516	318,42	6.3
#3 ZINC72375431	-2,558	-9,259	-48,955	7,254	-59,109	8,288	396,47	6.3
#4 ZINC257314029	-2,492	-9,635	-49,359	6,878	-61,536	9,343	382,49	6.11
#5 ZINC72383052	-2,450	-9,923	-37,786	7,282	-56,340	2,384	304,40	6.3
#6 ZINC72375230	-2,439	-9,565	-40,988	7,059	-57,755	3,545	332,43	6.3

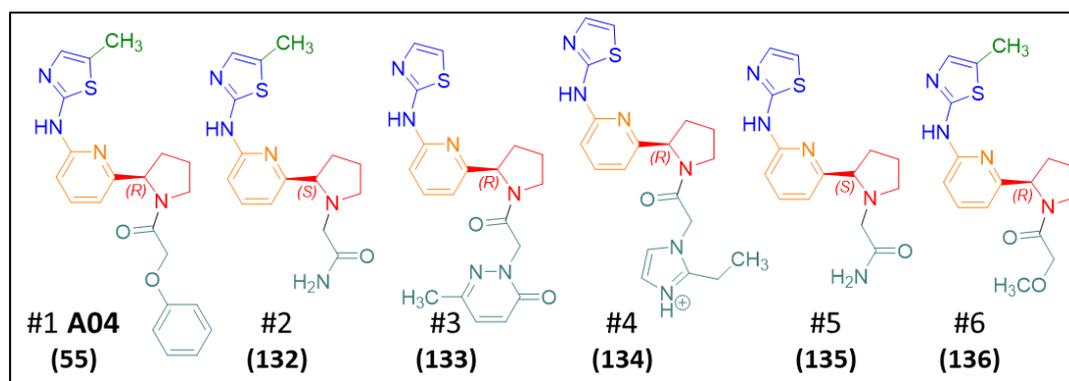


Abb. 6-12 – Die sechs bestbewerteten Strukturen des Clusters #100 (Nummerierung s. Tab. 6-7).

Die Konfiguration des Stereozentrums variiert bei den Verbindungen zwischen der *R*- und *S*-Konfiguration (s. Abb. 6-12). Verbindung #2 (**132**) und #3 (**133**) des Clusters haben weniger gute HYDE-Scores – hier könnte die Stereochemie (*S*- statt *R*-konfiguriert) für die Bindungsaffinität eine Rolle spielen. Bei diesen beiden Verbindungen ist zusätzlich eine Protonierung des tertiärenamins in der Pyrrolidinteilstruktur möglich, während eine solche Protonierung in den übrigen hier dargestellten Strukturen nicht möglich ist, da es sich an dieser Stelle um ein Amid handelt. Verbindung #4 (**134**) ist an einem zum Lösungsmittel gerichteten Imidazolring protoniert. Diese Struktur hat den besten HYDE-Score unter diesen Verbindungen.

Alle Vertreter dieser Gruppe wären für eine biologische Prüfung interessant. Man könnte herausfinden, welchen Beitrag die Stereokonfiguration oder ein protonierbarer Pyrrolidinrest, welcher zum Lösungsmittel herausragt, auf die Bindungsaffinität haben. Auch die Bedeutung der An- oder Abwesenheit der Methylgruppe (R1) auf die Bindungsaffinität wäre interessant. Mutmaßlich erhöht die Methylgruppe die Bindungsaffinität.

Das *S*-Enantiomer (ZINC72375191) von **A04 (55)** taucht in dem Cluster nicht auf, wird also nicht von den ersten 1.000 Clusterrepräsentanten wiedergespiegelt. Die Struktur steht an Position 6.928 der Consensuscore-sortierten Top-5%-Rangliste. Während die Werte für **D**, **Q** und **M** mit dem *R*-Enantiomer von **A04 (55)** vergleichbar sind, weicht vor allem **H** (HYDE-Score) stark ab und reduziert dadurch den Consensuscore (s. Tab. 6-8). Ebenso ist die *Ligand Strain Energy* vergleichsweise erhöht, die Pose des *S*-Enantiomers steht also stärker unter Spannung als das *R*-Enantiomer. Das *S*-Enantiomer wurde in der gleichen Ausrichtung wie das *R*-Enantiomer gedockt (gleiche Wasserstoffbrückenbindungen zur Hinge-Region, R2 in Richtung P-Loop), sitzt allerdings nicht genauso tief in der Bindetasche, sondern etwas nach vorne versetzt.

Tab. 6-8 – Vergleich der Scorewerte und weiterer Parameter der beiden Stereoisomere von **A04 (55)**

Struktur	Cons.	D	H	Q	M	L	MW	ID
ZINC72375192 (<i>R</i> -Enantiomer)	-2,817	-9,583	-42,484	7,955	-60,196	5,592	394,50	6.3
ZINC72375191 (<i>S</i> -Enantiomer)	-1,771	-9,363	-17,316	7,332	-62,997	13,429	394,50	33

E05

E05 (104) ($K_d > 30 \mu\text{M}$) besteht aus einem stickstoffhaltigen Trizyklus als Core 1, welcher in der Dockingpose mit der Hinge-Region über drei Wasserstoffbrückenbindungen interagiert (s. Abb. 6-13). Über eine mit dem Core 1 verknüpfte schwefelhaltige Seitenkette ragt ein Indolring zum Lösungsmittel nach außen.

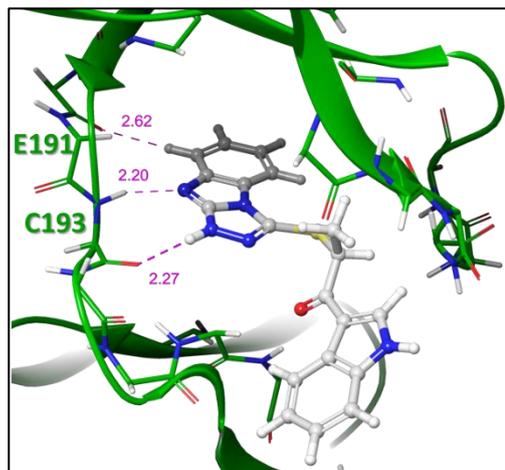


Abb. 6-13 – SP-Dockingpose von **E05 (104)** in ID 13

Drei weitere Mitglieder sind in dem Cluster #662, welches von **E05 (104)** repräsentiert wird, enthalten (s.

Abb. 6-14). Verbindung #3 (**138**) ist das Enantiomer von **E05 (104)**. An dieser Stelle sei zu erwähnen, dass nach Rückfrage bei der Firma AMBINTER **E05 (104)** als Racemat vorliegt. Dies wird vermutlich auch für **A04 (55)** der Fall sein. Daher wäre es interessant zu wissen, ob das *S*-Enantiomer von **E05 (104)** und das *R*-Enantiomer von **A04 (55)** eine höhere

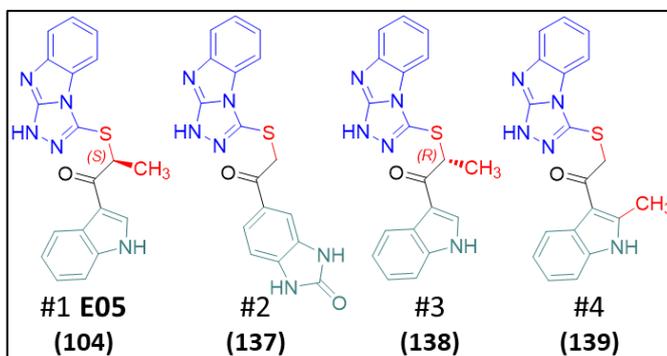


Abb. 6-14 – Clustermitglieder des Clusters #662 (**E05 (104)** als Repräsentant)

Bindungsaffinität haben als das Racemat bzw. als das entsprechende andere Enantiomer. Diese Frage konnte im Rahmen dieser Arbeit nicht mehr geklärt werden.

Die Scorewerte der ersten drei Verbindungen des Clusters sind sehr ähnlich, während die Scorewerte für Verbindung #4 (**139**), insbesondere **D** und **H**, deutlich schlechter sind (s. Tab. 6-9). Die Posen sind grundsätzlich ähnlich: der mit der Hinge-Region interagierende Trizyklus der Verbindungen #2 (**137**) bis #4 (**139**) ist nahezu deckungsgleich mit dem des **E05 (104)**. Der zum Lösungsmittel hinausragende Aromat der Verbindungen zeigt eine höhere Flexibilität.

Tab. 6-9 – Übersicht über Mitglieder des Clusters #662 (**E05 (104)** als Repräsentant). **Cons.** = Consensuscore nmr_DHQM, **D** = Dockingscore, **H** = HYDE-Score, **Q** = vorhergesagter pK_f -Wert mit dem *Sel2*-QSAR-Modell. **M** = Freie Bindungsenergie (PRIME MM/GBSA). **L** = *Ligand Strain Energy*. **ID** = Proteinkonformation des Komplexes mit dem besten Consensuscore.

Struktur	Cons.	D	H	Q	M	L	MW	ID
#1 E05 (ZINC4266839)	-2,422	-10,381	-29,368	7,524	-60,936	2,377	361,43	13
#2 ZINC7983247	-2,260	-9,734	-29,294	7,200	-60,731	3,285	364,39	13
#3 ZINC4266843	-2,102	-9,552	-29,64	7,425	-58,860	3,075	361,43	18
#4 ZINC6506473	-1,456	-8,984	-22,306	7,439	-57,908	2,157	361,43	13

MD-Simulation der Hit-Strukturen

Alle fünf Hit-Strukturen wurden – sowohl ohne Veränderung der SP-Posenstruktur als auch nach *OPLS3-LIGPREP* (s. Kap. 7.11.1, Tab. 7-7, „5 Hits (...)“) – in eine vorbereitete **5CEO**-Struktur mit dem XP-Modus gedockt (s. Kap. 7.5.2, Tab. 7-3, „XP 5 Hits“). Die jeweils beste Pose (bester Dockingscore bzw. Ausrichtung in der Bindetasche so wie in der SP-Pose des ZINC-Screenings) wurde ausgewählt. Mit diesen Protein-Ligand-Komplexen wurden jeweils drei MD-Simulationen über 30 ns durchgeführt (s. Kap. 7.4). Alle Hit-Strukturen blieben während der gesamten Simulationszeit in der ATP-Bindetasche. Die stärkste Bewegung war in den zum P-Loop bzw. zum Lösungsmittel gerichteten Substituenten zu sehen. Core 1 blieb immer mit der Hinge-Region verbunden.

Die mittleren RMSD-Werte des Protein-Backbones und des Liganden wurden mit denen der **5CEO**-, **5CEP**- und **5CEQ**-Trajektorien, welche die natürlich gebundenen DLK-Inhibitoren hoher Bindungsaffinität enthalten (s. Kap. 5.1.3), verglichen: Während sich die Backbone-RMSD-Werte zwischen den DLK-Inhibitor-Protein-Komplexen und den Hit-Protein-Komplexen nicht deutlich voneinander unterscheiden, kann die experimentell ermittelte Bindungsaffinität mit dem Liganden-RMSD sehr gut abgebildet werden (s. Tab. 6-10 u. Abb. 6-15).

Tab. 6-10 – RMSD-Mittelwerte [\AA] des Protein-Backbones (PB) und des Liganden (L) der bekannten DLK-Inhibitor-Protein-Komplexe und der Hit-Protein-Komplexe im Vergleich zu den Bindungsaffinitäten. ges = Gesamt.

	5CEN	5CEO	5CEP	5CEQ	E06	D04	C07	A04	E05
K_i bzw. K_d	-	< 0,0005	0,042	0,145	2,5	2,8	11	24	> 30
\bar{x} RMSD PB (1)	1,423	1,868	1,624	1,576	1,400	1,466	1,314	1,407	1,737
\bar{x} RMSD PB (2)	1,306	1,352	1,595	1,261	1,464	1,336	1,437	1,483	1,701
\bar{x} RMSD PB (3)	1,462	1,590	1,643	1,297	1,274	1,506	1,579	1,323	1,653
\bar{x} RMSD PB ges	1,397	1,603	1,621	1,378	1,380	1,436	1,443	1,404	1,697
\bar{x} RMSD L (1)	-	0,468	0,668	0,621	0,951	1,456	0,766	2,835	1,878
\bar{x} RMSD L (2)	-	0,693	0,644	0,678	0,964	0,625	1,286	1,198	2,636
\bar{x} RMSD L (3)	-	0,536	0,637	0,766	0,974	0,627	2,085	2,365	1,625
\bar{x} RMSD L ges	-	0,565	0,65	0,688	0,963	0,903	1,379	2,133	2,046

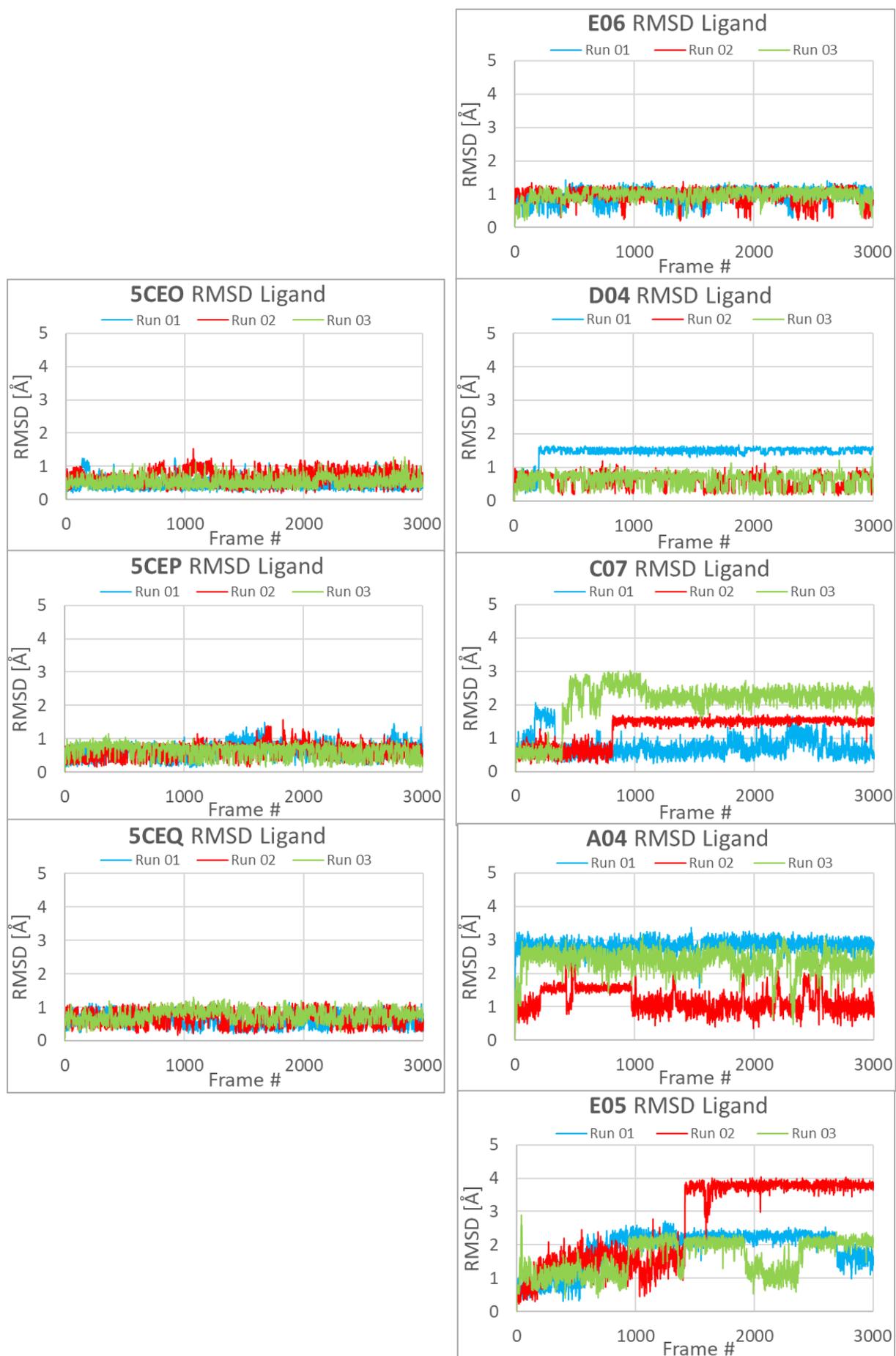


Abb. 6-15 – Ligand-RMSD-Plots der DLK-Inhibitor-Proteinkomplexe (links) und der Hit-5CEO-Komplexe (rechts).

Während die mittleren Liganden-RMSD-Werte der Röntgenkristallstrukturliganden im Bereich von etwa 0,6 Å liegen und somit die Liganden wenig beweglich in der Bindetasche sind, liegen die besten Hits (**E06** und **D04**) bei etwa 0,9 Å, der moderate Hit **C07** bei 1,4 Å und die beiden Hits **A04** und **E05** mit der geringsten Bindungsaffinität bei > 2,0 Å.

Die Spannweite des Ligand-RMSD ist bei den bekannten DLK-Inhibitoren geringer als bei den Hits. Auffällig ist, dass alle drei Simulationen der bekannten DLK-Inhibitoren gleichermaßen geringe Ligand-RMSD-Schwankungen aufweisen und keine Trajektorie hinsichtlich der Ligandenbeweglichkeit stark von den anderen abweicht. Bei den Hits hingegen stellt man mit abnehmender Bindungsaffinität stärkere Schwankungen des Ligand-RMSD innerhalb einer Trajektorie fest sowie stärkere Unterschiede bezüglich des Ligand-RMSD-Verlaufs zwischen den drei Trajektorien.

MD-Simulationen könnten demnach eine gute Methode sein, um einige der o.g. Clustervertreter in Komplex mit der ATP-Bindetasche der DLK hinsichtlich ihrer Beweglichkeit und Posenveränderung zu überprüfen und nur solche mit geringem Liganden-RMSD für eine biologische Prüfung auszuwählen. Allerdings gehört die MD-Simulation zu den rechenintensiven Methoden, daher ist eine Vorauswahl einer kleinen Anzahl interessanter Vertreter sinnvoll.

Im Folgenden werden die Positionsveränderungen beschrieben, welche für die Ligand-RMSD-Sprünge in den Plots von **D04**, **C07**, **A04** und **E05** verantwortlich sind:

Die erste Trajektorie (*Run 01*, blau) von **D04** weicht von den übrigen beiden ab: Hier dreht sich der zum P-Loop gerichtete Substituent R2 um 180° um die Einfachbindung zum Phenylring (s. Abb. 6-16) und bleibt in dieser Position bis zum Ende der Simulation bestehen.

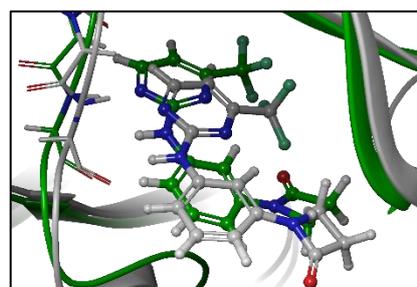


Abb. 6-16 – Frame 10 (grau) und 1000 (grün) von **D04** *Run 01*.

Der RMSD-Sprung in *Run 02* von **C07** bei 813 ns kommt durch Drehung des zum Lösungsmittel gerichteten Imidazolidinonrings um 180° um die Einfachbindung zum Phenylring zustande, während die starke Ligand-RMSD-Abweichung in *Run 03* durch eine Drehung der Einfachbindung zwischen Amid-Stickstoff und dem Phenylring entsteht, sodass der Imidazolidinonring näher an den P-Loop rückt (s. Abb. 6-17).

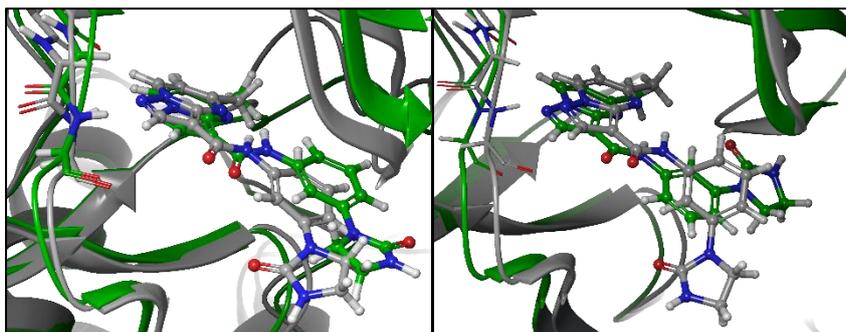


Abb. 6-17 – Links: Frame 10 (grau) und 1000 (grün) von **C07 Run 02**. Rechts: Frame 10 (grau) und 2000 (grün) von **C07 Run 03**.

In allen **A04**-Trajektorien kommt die starke Veränderung des Ligand-RMSD ebenso durch den nach außen orientierten Rest zustande (s. Abb. 6-18) Andere signifikante Positionsveränderungen sind hier nicht zu sehen.

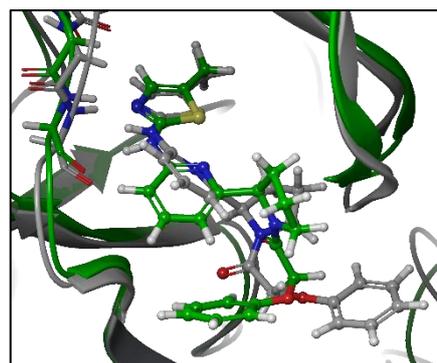


Abb. 6-18 – Frame 1 (grau) und 1000 (grün) von **A04 Run 01**.

Die Simulationen von **E05** weisen eine Veränderung der Position des Indolrings auf: in *Run 01* bewegt er sich weg von dem P-Loop hinaus zum Lösungsmittel (s. Abb. 6-19, links), in *Run 02* klappt er in die Bindetasche hinein (s. Abb. 6-19, rechts) und in *Run 03* wechselt die Indolring-Position wie in *Run 01*.

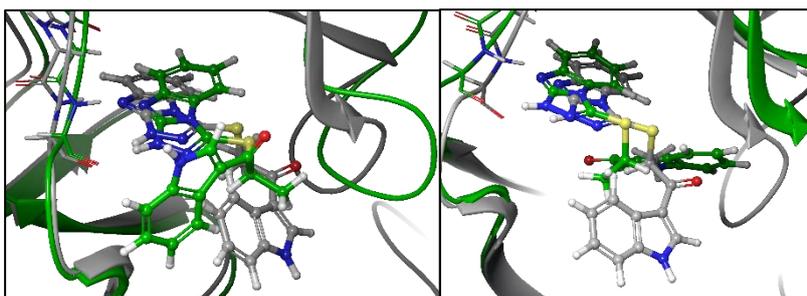


Abb. 6-19 – Frame 1 (grau) und 2000 (grün) von **E05 Run 01** (links) und *Run 02* (rechts).

7 Experimenteller Teil

7.1 Alignment von Proteinstrukturen und Berechnung von RMSD-Werten

Proteinüberlagerungen wurden mit dem „PROTEIN STRUCTURE ALIGNMENT“-Tool der SCHRÖDINGER SOFTWARE SUITE durchgeführt (mit „PrSA“ angegeben). Die zu überlagernden Proteine wurden in den Workspace geladen, selektiert und (wenn nicht anders angegeben) mit der Option *all residues* auf den ersten Eintrag der Liste überlagert. Bei dieser Option werden alle Atome des Proteins berücksichtigt. Für das Alignment von Bindetaschen wurde auch das „BINDING SITE ALIGNMENT“-Tool der SCHRÖDINGER SOFTWARE SUITE angewendet, dafür wurde der Backbone definierter Bindestellenamino­säuren ausgewählt (mit „BiSA“ angegeben).

Zur Berechnung von RMSD-Werten wurde nach vorausgegangenem Alignment das SUPERPOSITION-Tool der SCHRÖDINGER SOFTWARE SUITE nach Selektion der gewünschten Einträge und Inklusion des an erster Stelle stehenden Referenzeintrags verwendet. Die Regionen des Proteins (entsprechende Aminosäuren werden in den jeweiligen Kapiteln separat angegeben), zu denen man den RMSD-Wert bestimmen möchte, wurden über die ASL-Schaltfläche ausgewählt.

7.2 Berechnung von Volumina und Oberflächen mit POVME

Berechnungen von Volumina und Oberflächen der Bindetaschen der Röntgenkristallstrukturen bzw. der MD-Frames (s. Kap. 5.1.1 u. 5.1.3) wurden mit POVME 3.0 [484–486] durchgeführt. Dazu wurde die jeweilige Röntgenkristallstruktur zunächst in MAESTRO (2016-3) geladen, die Wassermoleküle und Liganden wurden bis auf den Liganden von **5CEO** gelöscht. Alle Proteine wurden über **5CEO** gelagert (s. dazu Kap. 7.1, „PrSA“). Alle Einträge wurden als PDBs exportiert, der Ligand in **5CEO** wurde ebenfalls einzeln exportiert. Die **5CEN**-, **5CEO**-, **5CEP**- und **5CEQ**-Proteine und der **5CEO**-Ligand und die insgesamt 12 MD-Trajektorien (3x für jeden PDB-Code, s. Kap. 5.1.3 u. 7.4) wurden in VMD [487] (Version 1.9.2) geladen. Die Frames der Trajektorien wurden auf die jeweilige MD-Ausgangsstruktur (Frame 0) gelagert („RMSD VISUALIZER“-TOOL, VMD). Die Trajektorie wurde daraufhin nur als Protein (ohne Wassermoleküle und ohne Ligand) exportiert und wieder hineingeladen und auch hier der **5CEO**-Ligand als Bezugspunkt für alle Trajektorien genutzt. Die x-, y- und z-Koordinaten des N9⁶⁴ des **5CEO**-Liganden wurden als zentraler Punkt für die *InclusionSphere* (dt. Einschlussbereich, vgl. Kap. 4.9.1) zur Berechnung des Volumens aller Bindetaschen verwendet.

⁶⁴ N9 ist das Stickstoffatom des Amins zwischen den beiden Phenylgruppen.

Folgendes Eingabeskript (exemplarisch) wurde für die POVME-Berechnungen eingesetzt:

```
GridSpacing 1.0
InclusionSphere 7.804320 -1.605597 -24.207811 10.0
PDBFileName 5cen_run01_waterremoved_aligned.pdb
AbstandCutoff 1.0
NumProcessors 8
OutputFileNamePrefix ./povme_5cen_run01
```

Der Gitterabstand (*GridSpacing*) zwischen den Punkten der Punktwolke wurde auf 1,0 Å definiert. Das Zentrum des Einschlussbereichs wurde mit den entsprechenden xyz-Koordinaten festgelegt und ein Radius der Punktwolke von 10.0 Å definiert. Punkte, die innerhalb eines maximalen Abstands von 1,0 Å von jedem Proteinatom lagen, wurden entfernt (*AbstandCutoff*). Nur bei den Röntgenkristallstrukturen wurde zusätzlich eine *ExclusionSphere* (dt. Ausschlussbereich) von 5,0 Å neben der Hinge-Region (außerhalb der Bindetasche) eingesetzt, um Punkte in diesem Bereich zu entfernen (s. auch S. 104, Abb. 4-9). Für die Bindetaschenvolumenberechnung der MD-Proteinkonformationen konnte diese *ExclusionSphere* nicht eingesetzt werden, da aufgrund der strukturellen Diversität der zahlreichen Proteinkonformationen einer Trajektorie mitunter auch Punkte innerhalb der Bindetasche entfernt würden und man ohnehin nicht alle Punkte außerhalb der Bindetasche bei jeder Proteinkonformation mit einer *ExclusionSphere* erfassen würde. Anders als für die vier Röntgenkristallstrukturen erschien die visuelle Beurteilung jeder einzelnen Punktwolke hier aufgrund der großen Menge an Proteinkonformationen der MD-Trajektorien nicht sinnvoll.

7.3 Proteinvorbereitungen der Röntgenkristallstrukturen mit ProtPrep

Die Vorbereitung der Proteinstrukturen wurde mit dem PROTPREP-Tool der SCHRÖDINGER SOFTWARE SUITE durchgeführt. Es werden grundsätzlich vier aufeinanderfolgende Operationen durchgeführt (Erläuterungen s. Kap. 4.2):

(1) Preprocess → **(2) Review and Modify** → **(3) Refine (H-bond assignment)** → **(4) Refine (Restrained minimization)**

Die Standardvorgehensweise wird im Folgenden aufgeführt:

PROTEIN PREPARATION WIZARD	
IMPORT AND PROCESS:	
<input checked="" type="checkbox"/>	Assign bond orders
<input checked="" type="checkbox"/>	Use CCD database
<input checked="" type="checkbox"/>	Add hydrogens
<input checked="" type="checkbox"/>	Create zero-order bonds to metals
<input checked="" type="checkbox"/>	Create disulfide bonds
(a) <input checked="" type="checkbox"/>	Delete waters beyond x Å from het groups) Wassermoleküle mit def. Abstand zu Het-Gruppen können gelöscht werden
<input checked="" type="checkbox"/>	Generate het states using Epik: pH: 7.0 +/- 2.0
Click: Preprocess	→ generiert neuen Eintrag: PDBX_preprocessed (1)
Alternative Positions: Entscheidung über die zu verwendende Konformation	
REVIEW AND MODIFY:	
Click: Analyze Workspace;	(b) Wassermoleküle können selektiv ausgewählt und gelöscht werden. (2)

<p>REFINE:</p> <p>H-bond assignment: ✓ Sample water orientations ✓ Use PROPKA pH: 7.0 Click: Optimize → generiert neuen Eintrag: PDBX_optimized (3)</p> <p>(c) (Remove waters with less than x H-bonds to non-waters) Wassermoleküle können je nach WBBs zum Protein gelöscht werden</p> <p>Restrained minimization: Converge heavy atoms to RMSD: 0.30 Å Force field: OPLS3 bzw. OPLS2005 (im Endnamen angegeben) Click: Minimize → generiert neuen Eintrag: PDBX_minimized (4)</p>
--

Die vorbereiteten Strukturen wurden nach dem letzten Schritt als MAE-Dateien abgespeichert (*PDBX_kraftfeld_wasserstatus.mae*). Alle erhaltenen Proteinstrukturen wurden insbesondere im Bereich der ATP-Bindetasche auch visuell auf Plausibilität und Richtigkeit hin überprüft.

Alle Strukturen wurden auf **5CEN aligned** (PrSA, s. Kap. 7.1). Folgende Proteinvorbereitungen wurden durchgeführt (s. Tab. 7-1):

Tab. 7-1 – Durchgeführte Proteinvorbereitungen; *Alternate Positions: nur solche aufgeführt, bei denen nicht die erste, sondern die zweite vorgeschlagene Konformation ausgewählt wurde. Erläuterung der Optionen für Wasserbehandlung (a, b, c): siehe „Protein Preparation Wizard“-Skript S. 263 f.

Proteinstruktur	Startstruktur	Wasserbehandlung	Besonderheiten	Version
5CEN_opls3_ow	5CEN	alle vor <i>Refine</i> gelöscht (b)	Alternate Position*: Phe270 (2)	2016-2
5CEN_opls2005_ow				
5CEO_opls3_ow	5CEO			
5CEO_opls2005_ow				
5CEP_opls3_ow	5CEP	alle vor <i>Refine</i> gelöscht (a)		
5CEP_opls2005_ow				
5CEQ_opls3_ow	5CEQ		Alternate Position*: Asn241 (2)	
5CEQ_opls2005_ow				
5CEN_opls3_minwat	5CEN	alle behalten		
5CEO_opls3_minwat	5CEO			
5CEP_opls3_minwat	5CEP			
5CEQ_opls3_minwat	5CEQ			
5CEN_opls3_w	5CEN	H ₂ O ohne WBB zu Nicht-H ₂ O (c) vor <i>Minimierung</i> gelöscht		
5CEO_opls3_w	5CEO	H ₂ O mit > 5 Å Entfernung von Het-Gruppen (a) vor <i>Refine</i> und ohne WBB zu Nicht- H ₂ O (c) vor <i>Minimierung</i> gelöscht		
5CEO_opls2005_w				
5CEP_opls3_w	5CEP			
5CEP_opls2005_w				
5CEQ_opls3_w	5CEQ			
5CEQ_opls2005_w				
5CEO_min_waterrem_prime_opls3 (= 5CEO-PL-Struktur)	5CEO	alle vor <i>Refine</i> gelöscht (a)	Loop-Modelling im <i>Pre-process</i> : ✓ Fill in missing side chains using Prime ✓ Fill in missing loops using Prime	2016-1
5CEP_min_waterrem_prime_opls3 (= 5CEP-PL-Struktur)	5CEP			
5CEQ_min_waterrem_prime_opls3 (= 5CEQ-PL-Struktur)	5CEQ			
MD-Frames für VSW, z.B. 5cen_run01_fr2006_opls3_6A	aus MD	H ₂ O und Ionen > 6 Å vom Liganden vor <i>Minimierung</i> gelöscht	nur <i>Minimierung</i> durchgeführt	2016-4
MD-Frames für Docking (QSAR-Alignment), z.B. 5cen_run02_frame2661_opls3_ow	aus MD	alle vor <i>Minimierung</i> gelöscht	nur <i>Minimierung</i> durchgeführt	2016-2

Die mit „prime“ gekennzeichneten Strukturen (**PL-Strukturen**, z.B. **5CEX_min_waterrem_prime_opls3**) enthalten einen mit PRIME generierten Loop. Dieser wurde im „REFINE LOOPS“-Tool (MAESTRO 2016-1, PRIME 4.3) zusätzlich in seiner Struktur verfeinert und minimiert:

REFINE LOOPS

```

Solvation model: VSGB; Force field: OPLS3
Find loops in Workspace structure: All
„loop 1“ auswählen und Residues selbst ändern: Res1: 254 (= Asp254) und Res2: 273 (= Thr273)
Options: Environment: Default
        Seed: Constant: 0
        Method: Serial loop sampling: Ultra Extended (100-200 subjobs)
        Subjobs: Distribute over maximum available processors
        Output for single loop feature: Maximum number of structures to return: 5
        ✓ Refine side chain within: 7.50 Å
        Minimum distance: 0.70 of vdW radii

```

Pro PDB wurden 5 Dateien mit verschiedenen Loopkonformationen generiert:

5ceo_opls3_looprefinement1.pdb bis 5ceo_opls3_looprefinement5.pdb (entsprechend für **5CEP** und **5CEQ**)

7.4 MD-Simulationen

Vorbereitung der MD-Startstruktur mit dem „SYSTEM SETUP“-Tool und den aus den Röntgenkristallstrukturen abgeleiteten Proteinstrukturen („bekannt“) **5CEN_opls3_ow**, **5ceo_opls3_looprefinement1.pdb** (= **5CEO-PL**-Struktur), **5cep_opls3_looprefinement2.pdb** (= **5CEP-PL**-Struktur) und **5ceq_opls3_looprefinement3.pdb** (= **5CEQ-PL**-Struktur) mit der Version MAESTRO 2016-2 [527] sowie mit den Hit-**5CEO-PL**-Komplexen („Hits“) mit der Version MAESTRO 2020-1 [528]:

SYSTEM SETUP

```

Solvation: Solvent Model: TIP3P
            Boundary conditions: Box shape: Orthorhombic
            Size calculation method: Buffer
            Distances (Å): a: 10.0, b: 10.0, c: 10.0
            Click: Minimize Volume
            Force field: OPLS3 („bekannt“) OPLS3e („Hits“)

Ions:
    Ion placement: Click: Recalculate → Berechnet automatisch, welches Ion und wie viele benötigt werden
    Für 5CEN: 1 Na+-Ion, für 5CEO: kein Zusatz von Ionen, für 5CEP: 1 Cl--Ionen, für 5CEQ: 3 Cl--Ionen
    Für die Hit-5CEO-PL-Komplexe: 1 Na+-Ion

```

Durchführung der MD mit DESMOND [361] v4.6 in der MAESTRO-GUI (2016-2 [529]) („bekannt“) bzw. mit Desmond v.4.9 in der MAESTRO-GUI (2020-1 [362]) („Hits“):

MOLECULAR DYNAMICS

```

Model System: Load from Workspace (dort ist die entsprechende Datei aus SYSTEM SETUP hinterlegt)
Simulation:
    Simulation time (ns): total 30.0;           elapsed: 0.0
    Recording interval (ps): trajectory 10.0;   energy: 1.2
    Approximate number of frames: 3001;         Ensemble class: NPT
    Temp: 300.0                                 Pressure (bar): 1.01325
    ✓ Relax model system before simulation
    Relaxation protocol: desmond_npt_relax_langevin.msgj (s. Kap. 7.4.1, S. 266 f.)
Advanced Options:
    Integration:
        RESPA integrator
        Time step (fs): bonded 2.0 near 2.00 far 6.00
    Ensemble:
        Thermostat method: Langevin; Relaxation time (ps): 1.0
        Number of groups: 1 (Temp: 300 K)
        Barostat method: Langevin; Relaxation time (ps): 2.0
        Coupling style: Isotropic
    Coulombic:
        Short range method: Cutoff
        Cutoff radius (Å): 9.0

```

```

Misc:
      Seed: Random
      ✓ Randomize velocities
        Starting time (ps): 0.0; Interval (ps): 0.0

```

- Write: Es wurde ein Skript gespeichert, welches dann über die Kommandozeile des Linux-Terminals gestartet werden konnte.
- Dreifache Durchführung pro Setup-Datei mit Random-Einstellung.

Analyse der MD-Trajektorien („bekannt“) mit SIMULATION INTERACTION DIAGRAM und SIMULATION EVENT ANALYSIS der SCHRÖDINGER SOFTWARE SUITE (MAESTRO 2016-2):

SIMULATION INTERACTION DIAGRAM

```

Load: entsprechende Trajektorie hereinladen (Endung: ...-out.cms)
SID Analysis Setup:
  Protein: Auto; Ligand: None (5CEN)/ Auto (5CEO, 5CEP, 5CEQ)
  Reference Structure: Frame 0
  Analysis type: Click: Select All

```

SIMULATION EVENT ANALYSIS

```

Structure (entspr. Trajektorien-Datei importiert)
Measurements: Pick atoms for measurements
  Phe192 und Gly132 (jeweils die C-alpha-Atome), Met190 und Ala135 (jeweils die C-alpha-Atome)

```

Die RMSD-Daten (Backbone, Ligand) der Hit-5CEO-PL-Komplexe („Hits“) wurden nach dem Laden der jeweiligen Trajektorie in den MAESTRO-Workspace (2020-1) nach Selektion der Backbone-Atome bzw. des Liganden mit *Plot* → *Descriptors (Selected Atoms)* → *RMSD* generiert und mit *Export Results* gespeichert.

7.4.1 Relaxationsprotokoll

Als *Relaxation Protocol* für alle MD-Simulationen wurde das folgende Langevin-Protokoll eingesetzt (Datei: `desmond_npt_relax_langevin.msgj`):

```

# Desmond standard NPT relaxation protocol
# All times are in the unit of ps.
# Energy is in the unit of kcal/mol.
task {
  task = "desmond:auto"
  set_family = {
    desmond = {
      checkpoint.write_last_step = no
    }
  }
}

simulate {
  annealing = off
  time = 100
  timestep = [0.001 0.001 0.003 ]
  temperature = 10.0
  ensemble = {
    class = "NVT"
    method = "Brownie"
    brownie = {
      delta_max = 0.1
    }
  }
  restrain = {
    atom = "solute_heavy_atom"
    force_constant = 50.0
  }
}

simulate {
  title = "Langevin NVT, T = 10 K, small timesteps, and restraints on solute heavy atoms"
  effect_if = [{"==" "-gpu" "@*.*.jlaunch_opt[-1]"} 'ensemble.method = Langevin']
  annealing = off

```

```

time           = 12
timestep       = [0.001 0.001 0.003]
temperature    = 10.0
restrain       = { atom = solute_heavy_atom force_constant = 50.0 }
ensemble = {
  class = NVT
  method = Langevin
  thermostat.tau = 0.1
}

randomize_velocity.interval = 1.0
eneseq.interval             = 0.3
trajectory.center            = solute
}

simulate {
  title           = "Langevin NPT, T = 10 K, and restraints on solute heavy atoms"
  effect_if       = [{"==" "-gpu" "@*.*.jlaunch_opt[-1]"} 'ensemble.method = Langevin']
  annealing       = off
  time            = 12
  temperature     = 10.0
  restrain        = retain

  ensemble = {
    barostat.tau = 50.0
    class = NPT
    method = Langevin
    thermostat.tau = 0.1
  }

  randomize_velocity.interval = 1.0
  eneseq.interval             = 0.3
  trajectory.center            = solute
}

solvate_pocket {
  should_skip = true
}

simulate {
  title           = "Langevin NPT and restraints on solute heavy atoms"
  effect_if       = [{"@*.*.annealing"} 'annealing = off temperature = "@*.*.temperature[0][0]"'
                    [{"==" "-gpu" "@*.*.jlaunch_opt[-1]"} 'ensemble.method = Langevin']
  time            = 12
  restrain        = retain
  ensemble = {
    barostat.tau = 50.0
    class = NPT
    method = Langevin
    thermostat.tau = 0.1
  }
  randomize_velocity.interval = 1.0
  eneseq.interval             = 0.3
  trajectory.center            = solute
}

simulate {
  title           = "Langevin NPT and no restraints"
  effect_if       = [{"@*.*.annealing"} 'annealing = off temperature = "@*.*.temperature[0][0]"'
                    [{"==" "-gpu" "@*.*.jlaunch_opt[-1]"} 'ensemble.method = Langevin']
  time            = 24
  ensemble = {
    barostat.tau = 2.0
    class = NPT
    method = Langevin
    thermostat.tau = 0.1
  }
  eneseq.interval = 0.3
  trajectory.center = solute
}

```

7.4.2 Clustering der MD-Frames

Zum RMSD-basierten Clustering der Frames aus den Trajektorien (s. Kap. 5.1.3) wurde das Tool RMSD BASED CLUSTERING OF FRAMES FROM DESMOND TRAJECTORY der MAESTRO-GUI (2016-3 [530]) mit den folgenden Einstellungen verwendet:

```

RMSD BASED CLUSTERING OF FRAMES FROM DESMOND TRAJECTORY
Use structures from File: (entspr. Trajektoriendatei „...-out.cms“ ausgewählt)
RMSD Matrix: Use ASL: Backbone
Frequency: 1
Clustering: Hierarchical cluster linkage method: average
            Use 10 as Number of clusters
  
```

7.5 Glide-Docking

7.5.1 Grid-Generierung

Für die Erstellung der Docking-Grids wurde der „RECEPTOR GRID GENERATION“-Panel des GLIDE-Moduls der SCHRÖDINGER SOFTWARE SUITE verwendet. Das Standardprotokoll zur Grid-Erstellung lautet:

```

RECEPTOR GRID GENERATION
Receptor: Define receptor: Bezugsligand auswählen
Van der Waals radius scaling: (Default-Einstellungen)
  Scaling factor: 1.0; Partial charge cutoff: 0.25
  Use ASL: Backbone
Advanced Settings:
  ✓Include halogens as acceptors
  (✓Include aromatic H as donors) ausgewählt, wenn angegeben
  Force Field: OPLS3 oder OPLS2005 Auswahl des Kraftfelds
Site:
Enclosing box
  Center: Centroid of Workspace ligand (selected in the Receptor tab)
  Size: (a) Dock ligands with length <= 30 Å
        oder: (b) Dock ligands similar in size to the Workspace ligand
Constraints:
H-bond/Metal:
  (a) Glu191 Carbonylsauerstoffatom des Backbones (Akzeptorfunktion)
  (b) Cys193 H-Atom des Backbone-NH (Donorfunktion)
  (c) Cys193 Carbonylsauerstoffatom des Backbones (Akzeptorfunktion)
  (d) Lys152 H-Atom des Amins in der Seitenkette
Rotatable Groups: Allow rotation: Auswahl von Aminosäuren möglich, wie angegeben
Excluded Volumes: keine
  
```

Folgende Grids wurden generiert (s. Tab. 7-2):

Tab. 7-2 – Generierte Grids. Constr = Constraints, KF = Kraftfeld; Rot. = erlaubte Rotationen; Optionen für Größe (des Grids) **a**, **b** und Constraints **a-d**: s. Skript RECEPTOR GRID GENERATION, S. 268, Unterpunkt „Site“ bzw. „Constraints“.

Proteinstruktur	KF	Bezugsligand	Constr	Größe	Besonderheiten	Einsatz
5CEO-PL-Struktur	Glide 7.1 / OPLS3	5CEO-Ligand aus 5CEO-PL-Struktur	(a)-(d)	(a)	Rot.: S253; Alignment auf Scen_run02_frame2661_opls3_ow (PrSA, s. Kap. 7.1; all residues), dann auf 5CEO-PL-Struktur (BiSA der Aminosäuren M190, C193, G196, Q197 und K152; s. Kap. 7.1)	Docking für QSAR-Alignment (s. Kap. 5.3.1)
5CEQ-PL-Struktur						
Scen_run02_frame2661_opls3_ow						
5ceo_run02_frame2000_opls3_ow						
5ceq_run03_frame3000_opls3_ow						
5CEN_opls2005_ow	Glide 7.0 / OPLS 2005	5CEO-Ligand gebundener Ligand	Keine	(b)	-	Re- und Crossdocking (s. Kap. 5.4.1.1)
5CEN_opls3_ow						
5CEO_opls2005_w						
5CEP_opls2005_w						
5CEQ_opls2005_w						
5CEO_opls2005_ow						
5CEP_opls2005_ow						
5CEQ_opls2005_ow						
5CEO_opls3_ow						
Protein-IDs 1-44 aus Tab. E 1 (Anhang E)	Glide 7.3 / OPLS3	Ligand aus 5CEO_opls3_w	(a)-(c)	(b)	Rot.: S133, T175, C193, T245, S253 ✓Include aromatic H as donors	VSW-Experimente (s. Kap. 5.4.1.2, 5.4.1.3)
5CEO-PL-Struktur	Glide 8.0 / OPLS3e	gebundener Ligand	(a)-(c)	(b)	Alignment auf Protein-ID 7 (PrSA, s. Kap. 7.1), Rot.: alle ✓Include aromatic H as donors	XP-Docking der Hits u. MD (s. Kap. 6)

7.5.2 Docking

In den folgenden Dockingprotokollen wird die Verwendung von Default-Einstellungen nicht extra aufgeführt. Für das Docking mit dem „GLIDE DOCKING (GD)“-Panel der SCHRÖDINGER SOFTWARE SUITE wurden folgende Einstellungen verwendet:

<p>GLIDE DOCKING <u>Settings:</u> Precision: HTvS, SP oder XP Auswahl des Dockingmodus (Advanced Settings: Count aromatic H and halogen H-bonds: ✓Include halogens as acceptors) Auswahl wenn angegeben Force Field: OPLS3 oder OPLS2005 <u>Constraints:</u> Auswahl vordefinierter Constraints <u>Output:</u> Anzahl der Posen pro Ligand festlegen (#Pos)</p>

Die Induced-Fit-Dockings wurden mit dem *Induced-Fit-Modus* (IF) der SCHRÖDINGER SOFTWARE SUITE durchgeführt:

<p>INDUCED FIT DOCKING Protocol: Standard, Force Field: OPLS3 <u>Receptor:</u> Box center: Centroid of Workspace ligand: Ligand der 5CEO-PL-Struktur als Referenzligand angeklickt Box size: Dock ligands with length <= 30 Å <u>Constraints:</u> Hbond/Metal: Cys193H, Cys193O, Glu191O <u>Glide Docking:</u> ✓Trim side chains: Automatic (based on B-factor) oder manual: M190, K152, D254, S253 (5CEO-PL-Struktur) sowie zusätzlich E200 in Scen_run02_frame2661_minopls3_waterrem Receptor van der Waals scaling: 0.70, Ligand van der Waals scaling: 0.50 Maximum number of poses: 20 <u>Prime Refinement:</u> Refine residues within 6.5 Å of ligand poses ✓Optimize side chains <u>Glide Redocking:</u> Redock into structures within 30.0 kcal/mol of the best structure, and within the top 5 structures overall Precision: SP</p>

Für das Docking mit dem „VIRTUAL SCREENING WORKFLOW (VSW)“-Panel der SCHRÖDINGER SOFTWARE SUITE wurde folgendes Protokoll verwendet:

VIRTUAL SCREENING WORKFLOW	
<u>Filtering/Preparation:</u>	deaktiviert (Liganden sind schon separat mit LigPrep bearbeitet worden)
<u>Receptors:</u>	Constraints: alle 3 voreingestellten Constraints ausgewählt: Must match at least 0, 1 oder 2
✓ Dock with Glide HTVS: Options:	
Dock flexibly	
Amide bonds:	Penalize non-planar conformation
✓ Perform post-docking minimization	✓ Use selected constraints for each grid
Generate up to 1 pose per compound state	
After docking, keep: 100 % of best compounds	
Retain: all good scoring states oder only best scoring state (macht keinen Unterschied bei Produktion nur einer Dockingpose pro Ligandenzustand)	
UND/ODER:✓ Dock with Glide SP:	gleiche Optionen wie beim HTVS ausgewählt
UND/ODER:✓ Dock with Glide XP:	gleiche Optionen wie beim HTVS ausgewählt

Folgende Dockingexperimente wurden durchgeführt (s. Tab. 7-3):

Tab. 7-3 – Durchgeführte Dockingexperimente. #Pos = Anzahl Posen pro Ligand. Lig. = Ligandenliste

Docking	Tool / Modus	Kraftfeld	Constraints	#Pos	Liganden (s. Kap. 7.11.1)	Grids (s. Kap. 7.5.1)	Besonderheiten
Docking für QSAR-Alignment	GD (Glide 7.1) / SP	OPLS3	C193H ≥ 1+ C193O/E191O ≥ 1	5	1295 DLK-Inh. (<i>Multiple States</i>)	Grids für „Docking für QSAR-Alignment“	Halogene = Akzeptoren
InducedFit-Docking	IF (Glide 7.2) / SP	OPLS3	alle	20	18 un spez. DLK-Inh. der 1295 DLK-Inh. (<i>Multiple States</i>)	Grids von 5CEO-PL-Struktur und Scen_run02_frame2661_minopls3_waterrem	
Re- und Crossdocking	GD (Glide 7.0) / SP u. XP	je nach Kraftfeld der PROT-PREP	keine	10	3 X-Ray-Liganden	Grids für „Re- und Crossdocking“	
VSW Re- und Crossdocking	VSW (Glide 7.3) / HTvS, SP, XP in einem Job	OPLS2005	≥ 1 der 3 Constraints	1	La) 3 X-Ray-Liganden Lb) 30 un spez. DLK-Inh. und ANP	Ga) Grids der Protein-IDs 1-33 Gb) Grids der Protein-IDs 6.1-6.11	ins. 4 Jobs (La+Ga, La+Gb, Lb+Ga, Lb+Gb)
VSW Actives / Decoys Lig. a / b	VSW (Glide 7.3) / HTvS, SP, XP (getrennte Jobs)	HTvS: OPLS2005 SP/XP: OPLS3 oder OPLS2005	0, ≥ 1 oder ≥ 2	1	HTvS: Lig. a nach OPLS2005- oder OPLS3-LIGPREP, Lig. b nach OPLS3-LIGPREP SP: HTvS-Output XP: SP-Output	Grids der Protein-IDs 6.3, 7, 6.11, 13, 18, 26, 30 und 33	insg. 15 Jobs
VSW ZINC / Actives bzw. Actives / Decoys Lig. c	VSW (Glide 7.4) / HTvS, SP, XP (getrennte Jobs)	HTvS: OPLS2005 SP/XP: OPLS3	HTvS: ≥ 2, SP: ≥ 1	1	HTvS: a) ZINC u. Aktive der Lig. c (LIGPREP s. Kap. 7.11.1, Tab. 7-7) b) Aktive u. Decoys der Lig. c (LIGPREP s. Kap. 7.11.1, Tab. 7-7) SP: Glide Ensemble Merge des HTvS-Output	Grids der Protein-IDs 6.3, 7, 6.11, 13, 18 und 33	
XP 5 Hits	VSW (Glide 8.6) / XP	OPLS3	≥ 1 der 3 Constraints	5	5 Hits (A04, C07, D04, E05 und E06)	Grid der 5CEO-PL-Struktur	

In „VSW Actives/Decoys Lig. a/b“ und „VSW ZINC / Actives bzw. Actives / Decoys Lig. c“ erfolgte die Durchführung von drei Jobs (HTvS, SP und XP) durch Herausschreiben (*Write...*) der eingestellten Optionen über die MAESTRO-GUI in ein Verzeichnis und Starten dieses Skripts. In den .inp-Files musste die Input-Datei (Output des jeweiligen Vor-Dockings) für den SP- und XP-Job angepasst werden. Der VSW wurde dann mit einem Shellskript gestartet, der beispielhaft dargestellt wird:

```
#!/bin/bash
cd /Verzeichnis/vsw_htvs_8targets_ligandset_b_opls3ligprep_withoutconstraints
$SCHRODINGER/vsw vsw_htvs_8targets_ligandset_b_opls3ligprep_withoutconstraints.inp -WAIT
-host_glide localhost:32 -adjust -NJOBS 32 -TMPLAUNCHDIR -ATTACHED

cd /Verzeichnis/vsw_sp_opls3_8targets_ligandset_b_opls3ligprep_withoutconstraints
$SCHRODINGER/vsw -WAIT vsw_sp_opls3_8targets_ligandset_b_opls3ligprep_withoutconstraints.inp
-WAIT -host_glide localhost:32 -adjust -NJOBS 32 -TMPLAUNCHDIR -ATTACHED

cd /Verzeichnis/vsw_xp_opls3_8targets_ligandset_b_opls3ligprep_withoutconstraints
$SCHRODINGER/vsw -WAIT vsw_xp_opls3_8targets_ligandset_b_opls3ligprep_withoutconstraints.inp
-WAIT -host_glide localhost:32 -adjust -NJOBS 32 -TMPLAUNCHDIR -ATTACHED
```

7.5.3 Berechnung der RMSD-Werte von Dockingposen

Die RMSD-Werte zwischen Docking- und Originalpose wurden mit der Option „Core: Use core pattern comparison: Use for RMSD calculations only“ im „GLIDE Docking“-Panel nur für die neutralen Liganden berechnet (Parameter: *glide rmsd*). Für die protonierten Liganden wurde der RMSD nachträglich berechnet im Vergleich zum Referenzliganden der vorbereiteten Proteinstrukturen **5CEX_opls2005_w** (s. Kap. 7.3) mit folgendem SCHRÖDINGER-Skript:

```
>> $SCHRODINGER/run rmsd.py -use_neutral_scaffold ligand1 ligand2
```

„Neutral Scaffold“ entspricht den schweren Atomen (*heavy atoms*) des Moleküls.

7.6 Berechnung der Validierungsmetriken und Diagrammerstellung

Zur Erstellung der Anreicherungskurven wurden mithilfe eines KNIME-Workflows die entsprechenden Parameter aus den Liganden-Score-Ranglisten berechnet (s. Abb. 7-1 und Kap. 4.4.8, S. 79, Formel (17)). Ebenso wurde die Darstellung der Häufigkeitsverteilung von Aktiven, Decoys bzw. ZINC-Substanzen mit einem KNIME-Workflow vorbereitet, indem die Aktiven, Decoys bzw. ZINC-Substanzen getrennt abgespeichert wurden. Die so erzeugten Dateien wurden mit GNU PLOT v.4.6 verarbeitet, um die entsprechenden Anreicherungskurven und Histogramme graphisch darzustellen. Die Enrichmentfaktoren wurden basierend auf der in Abb. 7-1 erstellten Textdatei mithilfe weiterer Nodes berechnet ($EF = \frac{\%aktives}{\%screened}$). Der Pearson-r zwischen Scorewert und pK_i -Wert von Aktiven wurde mit dem Node „Linear Correlation (deprecated)“ berechnet, während die Rangkorrelationskoeffizienten mit dem Node „Rank Correlation (deprecated)“ berechnet wurden.

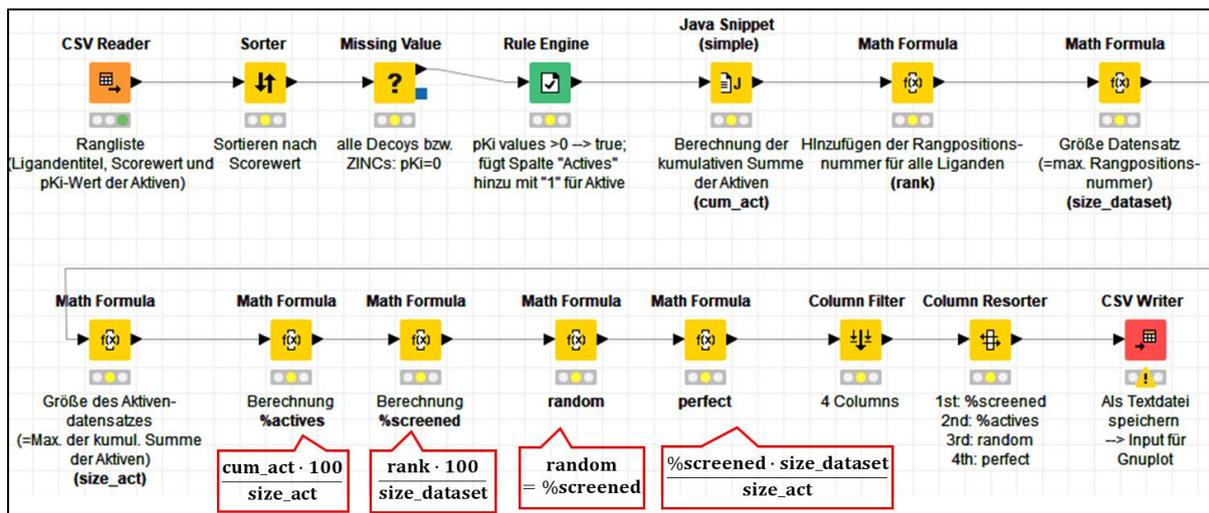


Abb. 7-1 – KNIME-Workflow zur Berechnung der Parameter %screened, %actives, random, perfect zur Erstellung der Anreicherungskurve mit GNUPLOT

7.7 Erstellung von Pharmakophormodellen⁶⁵

Der Ligand GNE-3511 aus der vorbereiteten Proteinstruktur **5CEO_opls2005_w** (s. Kap. 7.3) wurde als Referenzligand zur Pharmakophor-Hypothesen-Generierung verwendet werden.

Die aus der Ligandenpräparation gewonnenen 1909 Strukturen (*Multiple States*, s. Kap. 7.11.1, „DLK-Inh. für QSAR“) wurden auf 825 Liganden (*Single States*) reduziert, wobei bei der Auswahl dieser Strukturen darauf geachtet wurde, dass niemals die aromatischen Ringe (meist in Core 1 und Core 2) protoniert vorlagen. Es wurde, wenn möglich, immer der Ligand ausgewählt, der in der basischen Seitenkette protoniert war. Doppelt geladene Liganden wurden nicht ausgewählt. Für die Liganden aus den drei Röntgenkristallstrukturen **5CEO**, **5CEP** und **5CEQ** wurden mit EPIK v3.6 (MAESTRO 2016-2) Protonierungen in den möglichen Kombinationen im pH-Bereich 7,0 +/- 2,0 durchgeführt. Nur die Protonierung in der basischen Seitenkette wurde beibehalten. Diese drei Liganden wurden einer kurzen Minimierung unterzogen⁶⁶.

⁶⁵ Die Pharmakophormodelle wurden von Dr. Thomas Lemcke erstellt und sein Vorgehen im Folgenden beschrieben.

⁶⁶ Minimierung: im BUILD-Menü des MAESTRO-Workspace wird „minimize selected atoms“ ausgewählt.

Mit dem „MANAGE HYPOTHESIS“-Tool (MAESTRO 2016-2) wurden alle möglichen *Sites* des Referenzliganden dargestellt (s. Abb. 7-2). Es gibt:

- 4 ACCEPTOR SITES (A): die beiden aromatischen Ringe, die Cyanogruppe, der Sauerstoff der zum Lösungsmittel gerichteten Seitenkette
- 1 DONOR SITE (D): sekundäres Amin zwischen den beiden aromatischen Ringen
- 3 HYDROPHOBE SITES (H): an dem P-Loop zugewandten Pyrrolidinring
- 1 POSITIVE SITE: am protonierten tertiären Amin
- 2 RING SITES: die beiden aromatischen Ringe

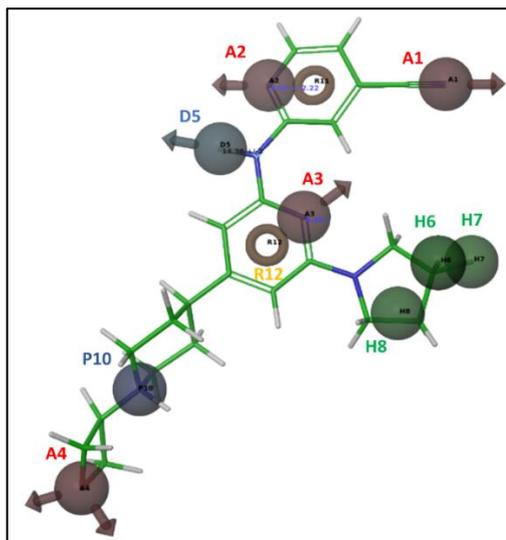


Abb. 7-2 – Mögliche Pharmakophore des **5ceo_minimized_ligand_cryst.**

4 Acceptor Sites (rot mit Pfeil), 1 Donor Site (blau mit Pfeil), 3 Hydrophobe Sites (grün), 1 Positive Site (blau), 2 Ring Sites (rote Ringe)

Aus diesen Sites wurden verschiedene Hypothesen

(z.B. ADHPRR, s. Tab. 7-4) generiert. Das Screening nach Konformeren wurde in MANAGE PHARMACOPHORE HYPOTHESES und der Einstellung ADVANCED PHARMACOPHORE SCREENING (MAESTRO 2016-2, PHASE 4.7) durchgeführt:

Alle 828 Liganden wurden selektiert. Während der Suche wurden Konformere generiert. Die Anzahl der *Sites*, die von den Liganden abgedeckt werden sollten, wurde variiert (z.B. 5/5 *Matches*). Alle anderen Einstellungen blieben unverändert. In zwei Hypothesen wurden Bereiche ausgewählt, die nicht von den Konformeren besetzt werden durften (sog. *excluded volumes*, s. Abb. 7-3), da in dem Screening mit den Hypothesen ohne *excluded volumes* einige Liganden eine Position einnahmen, in der sie in einer hypothetischen DLK-Bindetasche Kollisionen mit dem Protein verursachen würden.

Schließlich wurden die Datensätze zur QSAR-Erstellung mit den folgenden Strukturen (s. Tab. 7-4) aus diesem Pharmakophoren-Alignment ergänzt, da für diese keine geeigneten Dockinglösungen vorlagen:

Tab. 7-4 – Ligandenkonformere aus dem Pharmakophor-Alignment, die in die Selektion aufgenommen wurden. „Ferraris“ bzw. „Karaman“: Liganden aus der Veröffentlichung von FERRARIS ET AL. [125] bzw. KARAMAN ET AL. [164] gescreent. 4/6 = 4 der 6 festgelegten Sites mussten erfüllt werden. ex = ausgeschlossene Bereiche festgelegt (s. Abb. 7-3), die von den Konformeren nicht eingenommen werden durften.

Ligand	Hypothese	Sites	Vorhanden in Selektion Nr.
Staurosporine	ADHPRR_ex1 4/6/exclude Karaman	A2, D5, H8, P10, R11, R12, excluded volume (Abb. 7-3 A)	1, 2, 3
CEP-701	ADHPRR 4/6 Ferraris	A2, D5, H8, P10, R11, R12	1, 2, 3
pat_wo2015091889_024	ADHPRR_ex2 4/6/exclude	A2, D5, H8, P10, R11, R12, excluded volume (Abb. 7-3 B)	1, 2, 3
pat_wo2015091889_026	ADHPRR_ex2 4/6/exclude		3
pat_wo2015091889_027	ADHPRR_ex2 4/6/exclude		1, 2, 3
pat_wo2015091889_031	ADHPRR_ex1 4/6/exclude	A2, D5, H8, P10, R11, R12, excluded volume (Abb. 7-3 A)	1
pat_wo2015091889_033	ADHPRR_ex1 4/6/exclude		2
pat_wo2015091889_034	ADHPRR_ex1 4/6/exclude		2

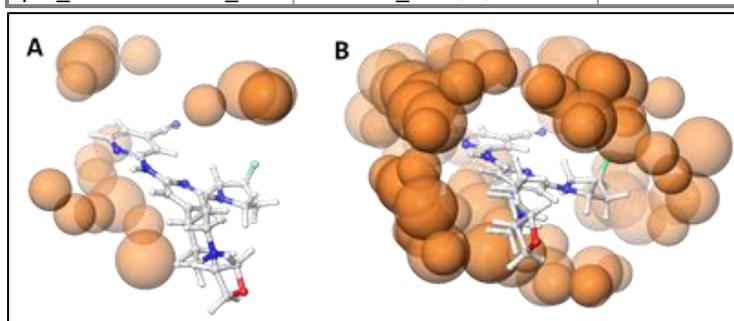


Abb. 7-3 – Excluded volumes (ausgeschlossene Bereiche) in der Pharmakophorhypothese ADHPRR-ex1 (A) bzw. ADHPRR_ex2 (B)

Diese wurden mit dem „SHAPE SCREENING“-Tool der SCHRÖDINGER-Software auf einen der Liganden des Docking-Alignments der jeweiligen Selektion *aligned*.

7.8 Erstellung von QSAR-Modellen

Die Erstellung der QSAR-Modelle erfolgte mit dem „FIELD-BASED QSAR“-Panel der MAESTRO-GUI der SCHRÖDINGER-Software (Version 2016-3):

FIELD-BASED QSAR
BUILD... Field Style: Feldkombination ausgewählt (z.B. Gaussian Recommended für das SEHAD-Modell, Extended Gaussian für das SEHAD-Modell) Maximum PLS factors: 10 Grid spacing: 1.0 Å Extend grid by: 3.0 Å beyond training set limits <input checked="" type="checkbox"/> Ignore force fields within: 2.0 Å of any training set atom Truncate steric force fields at: 30.0 kcal/mol Truncate electrostatic force fields at: 30.0 kcal/mol Eliminate variables with StdDev < 0.01 <input checked="" type="checkbox"/> Eliminate variables with t-value < 2.0 Number of ligands to leave out for cross-validation: 1

Für einen festgelegten Anteil des Trainingsdatensatzes (z.B. 70 %) wurde jede Selektion zufällig in Test- und Trainingsdatensatz aufgeteilt. Dies geschah mit der Option „Apply Random Seed“ im „FIELD-BASED QSAR“-Panel:

Random training set: 70 % → Apply Random Seed: 14658 (Run 01); 79263 (Run 02); 18487 (Run 03)

```

ligand
  read SP_posefile_ID7.mol2 $(lig) n # Multi-MOL2-Datei mit dem Rezeptor an erster
                                   Stelle und darauffolgend die Ligandenposen
end
docking
  analyze
    seloutp output_plp_ID7.txt a
    output "$(lig)"
    plp fix #hier kann statt plp auch score eingesetzt werden
            (= FlexX-Score)
    seloutp screen
end_for
quit y

```

Das Skript *rescoring_script.bat* wird als Kommandozeile ausgeführt mit Angabe der Rezeptor-Moleküldatei (hier: ID7.mol2):

```
>> leadit -b rescoring_onemol2file.bat -a '$(protein)=ID7.mol2'
```

7.9.2 HYDE-Score

Das Rescoring der Liganden mit dem HYDE-Score erfolgte mit dem Protein-Komplex-Scoring-Tool HYDE-SCORER (HYDE 3.2.5 Version 1.3.0) über folgende Kommandozeile, beispielhaft für Protein-ID 7:

```
>> hydescorer --target ID7_20A.fxx --input ligands_ID7.mol2 --output-file hydescoring_ligands_ID7_scoring_output.log
```

ID7_20A.fxx: Zur Erstellung dieser Projektdatei wurde die Proteinkonformation und der dazugehörige Referenzligand⁶⁷ im MOL2-Format in die Leadit-GUI (Version 2.1.8) geladen. Der maximal einstellbare Radius von 20 Å um den Referenzliganden wurde ausgewählt zur Markierung der Bindetasche und die so erstellte Projektdatei gespeichert.
 ligands_ID7.mol2: zu bewertende Ligandendatei (in diesem Fall Dockingposen nach Glide-Docking)
 hydescoring_ligands_ID7_scoring_output.log: selbst benannte Output-Datei, in der die Ergebnisse des HYDE-Scorings gespeichert werden

7.9.3 Prime MM/GBSA

Für die Berechnung der *Freien Bindungsenergien* und der *Ligand Strain Energies* von Protein-Ligand-Komplexen (s. Kap. 5.5.1 bis 5.5.3) mit PRIME MM/GBSA wurde im „PRIME MM-GBSA“-Panel (PRIME MMGBSA v3.0) der SCHRÖDINGER MAESTRO-GUI (2017-1 in Kap. 5.5.1 und 2017-3 in Kap. 5.5.2 u. 5.5.3) mit folgenden Einstellungen ein Shell-Skript geschrieben (*Write ...*):

```

PRIME MM-GBSA
Structures:
  Take complexes from separated ligand and protein structures: entspr. Ligandendatei (aus dem GLIDE-
  Docking) sowie die dazugehörige Proteinkonformation ausgewählt
Options:
  Solvation model: VSGB; Force Field: OPLS3
Protein flexibility: Flexible residue distances are defined using all ligands processed
  Distance from ligand (Å): 0.0 (d.h. keine Flexibilität der Aminosäuren)
  Sampling method: Minimize
Write ...

```

⁶⁷ Das ist der entsprechende natürlich gebundene Ligand der Röntgenkristallstruktur bzw. der MD-Proteinkonformation nach PROTPREP.

Der Inhalt des Shell-Skripts (*mmgbsa.sh*) lautete (exemplarisch für vsw-SP_OUT_7_pv.maegz als Rezeptor-Liganden-Input-Datei):

```
#!/bin/bash

"${SCHRODINGER}/prime_mmgbsa" -OVERWRITE -prime_opt "OPLS_VERSION=OPLS3" vsw-SP_OUT_7_pv.maegz
-HOST "localhost:40" -NJOBS 40 -TMPLAUNCHDIR -ATTACHED
```

Es wurde im Terminal aktiviert und gestartet mit:

```
>> chmod +x mmgbsa.sh
>> ./mmgbsa.sh
```

7.9.4 Rescoring mit dem QSAR-Modell

Das in Kap 5.3.3 ausgewählte 5-Felder-Modell auf Basis des Ligandenalignments der Selektion 2 musste zunächst auf die Proteinkonformationen der im ZINC-Screening verwendeten IDs (s. Anhang E, Tab. E 1) zur Deckung gebracht werden, da für die Erstellung des Docking-Alignments andere Proteinkonformationen in einer davon abweichenden Ausrichtung verwendet wurden. Dazu wurden die Liganden des Alignments mit der **5CEO-PL**-Struktur, welche für das Docking der meisten Liganden für QSAR verwendet wurde (s. Kap. 5.3.1), in der

Tab. 7-5 – Statistische Daten des *SeI2*-QSAR-Modells vor dem neuen Alignment (s. Kap. 5.3.3) und nach Alignment auf die Proteinkonformation ID 6.3

	Alte Orientierung	Neue Orientierung
#k	6	6
R²	0,818	0,816
SD	0,548	0,551
F	125,9	124,5
R²_{cv}	0,626	0,622
Q²	0,642	0,642
RMSE	0,76	0,76
Beiträge der Felder		
S	37,4 %	37,7 %
E	7,7 %	7,7 %
H	29,0 %	28,7 %
A	18,0 %	18,4 %
D	7,9 %	7,5 %

SCHRÖDINGER MAESTRO GUI zusammengefügt (Rechtsklick auf die markierten Ligandeneinträge → MERGE). Dieser Protein-Multiple-Liganden-Komplex wurde auf Proteinkonformation ID 6.3 *aligned* (s. Kap. 7.1, „PrSA“) und die Liganden wurden anschließend wieder ohne die **5CEO-PL**-Struktur gespeichert. Auf Basis dieser neu ausgerichteten Liganden wurde das QSAR-Modell mit den gleichen Einstellungen wie für das *SeI2*-QSAR-Modell neu gerechnet. Die statistischen Daten des neu gerechneten Modells gleichen in etwa dem alten *SeI2*-Modell (s. Tab. 7-5, Ergebnisse der Berechnung mit SCHRÖDINGER Version 2020-1) und wurde daher für die Bewertung der SP-Dockingposen im Rahmen des Postprocessings eingesetzt.

Das „Rescoring“ von Dockingposen mit diesem QSAR-Modell wurde mit folgender Kommandozeile (exemplarisch für SP-Dockingergebnisse in Protein-ID 7) im Terminal gestartet (MAESTRO 2017-3, PHASE V.5.2):

```

$SCHRODINGER/phase_fqsar ID7_SP-Posen.maegz ID7_SP-Posen-qsar_out.maegz r_user_pKi -test -imod
/Verzeichnis/model_sel2_run2_neworientation_sehad_80%.fqsar -osum ID7_qsar_totaloutput_sum
-opred ID7_qsar_totaloutput_pred

```

Input-Datei: ID7_SP-Posen.maegz

Output-Datei: ID7_SP-Posen-qsar_out.maegz

Aktivität: r_user_pKi

Angewandtes QSAR-Modell: model_sel2_run2_neworientation_sehad_80%.fqsar

Output-Datei (allg. Daten): ID7_qsar_totaloutput

Output-Datei mit vorhergesagten Aktivitäten: ID7_qsar_totaloutput_pred

7.9.5 IFP-Tanimotokoeffizient

Zur Berechnung des IFP-Tanimotokoeffizienten für Dockingposen wurde der „INTERACTION FINGERPRINTS“-Panel in der SCHRÖDINGER MAESTRO GUI verwendet. Der Rezeptor (1. Eintrag) und alle Posen wurden selektiert und die Fingerprints berechnet:

INTERACTION FINGERPRINTS

Fingerprint Generation: Docked poses auswählen
Interactions to include:
alle selektieren: Any Contact
Backbone Interaction
Sidechain Interaction
Polar Residues
Hydrophobic Residues
Hydrogen Bond Acceptor
Hydrogen Bond Donor
Aromatic Residue
Charged Residue

Click: Generate Fingerprints

Danach wurden die Ähnlichkeiten (als Tanimotokoeffizient) zu einem ausgewählten Referenzliganden anhand der berechneten *Interaction Fingerprints* berechnet. Als Referenzligand wurde immer der gedockte DLK-Inhibitor verwendet, welcher ursprünglich gebundener Inhibitor der Röntgenkristallstruktur war, aus der die zum Docking verwendete Proteinkonformation hervorgeht. Für Proteinkonformation ID 6.3 (hervorgehend aus der **5CEO**-Röntgenkristallstruktur) ist dies beispielsweise der gedockte **5CEO**-Ligand.

INTERACTION FINGERPRINTS

Similarity: Similarity settings
Similarity metric: Tanimoto
Similarity Calculation: Click Calculate Similarity
 Sort selected entries by similarity

Dadurch wurden die Liganden nach ihrer Ähnlichkeit zum Bezugsliganden sortiert.

Wichtig: Zum Zeitpunkt der Bewertung der Ranglisten mit dem IFP-Tanimotokoeffizienten wurden die Ranglisten NACH Dockingscore-Merge bewertet. Da im Postprocessing die anderen Scorewerte auf alle Dockingposen (OHNE vorheriges Merging) angewendet wurden, wurde der

IFP-Tanimotokoeffizient bei Zusammenstellung dieser Arbeit erneut für alle Dockingposen OHNE Merging, d.h. für ein größeres Substanzkollektiv, berechnet (MAESTRO 2020-1), und diese Ergebnisse für einen besseren Vergleich mit den anderen Scorewerten in dieser Arbeit dargestellt. Die IFP-Tanimotokoeffizienten in Tab. 5-52 (S. 222 ff.) stammen aus den ursprünglichen Berechnungen (MAESTRO 2017-3), welche nur für die Top1000-Clusterrepräsentanten (OMW- und MW<350) für die Auswahl der zu testenden Substanzen berechnet wurden.

7.9.6 Normalisierung der Scorewerte und Kombination von Scorewerten

Die Vergabe von Rangpositionsnummern (z.B. r_D) erfolgte nach Sortierung nach dem ent-

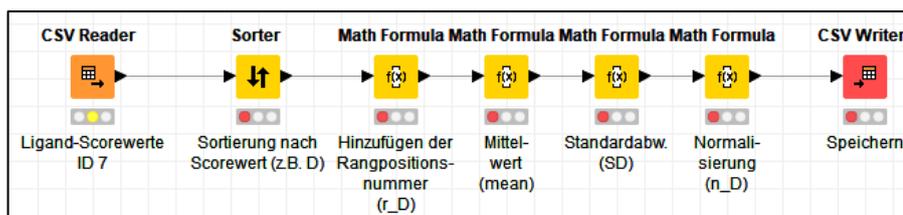


Abb. 7-4 – KNIME-Workflow zur Verarbeitung der Scorewerte

sprechenden Scorewert (z.B. nach Dockingscore, D) mit einem entsprechenden KNIME-Workflow (s. Abb. 7-4). Die Normalisierung der Scorewerte (z.B. n_D) erfolgte nach Formel (16) (s. Kap. 4.4.7, S. 77). Analog wurden die Rangpositionsnummern normalisiert (z.B. nr_D). Nach Mittelwertbildung verschiedener Rangpositionsnummern (r_D , r_H , etc.) wurde eine mittlere Rangpositionsnummer berechnet und anschließend normalisiert (z.B. nmr_{DH}).

Die Zusammenführung der sechs Ranglisten unter Beibehaltung der Ligandenpose mit dem besten Scorewert erfolgte ebenso mithilfe eines KNIME-Workflows (s. Abb. 7-5).

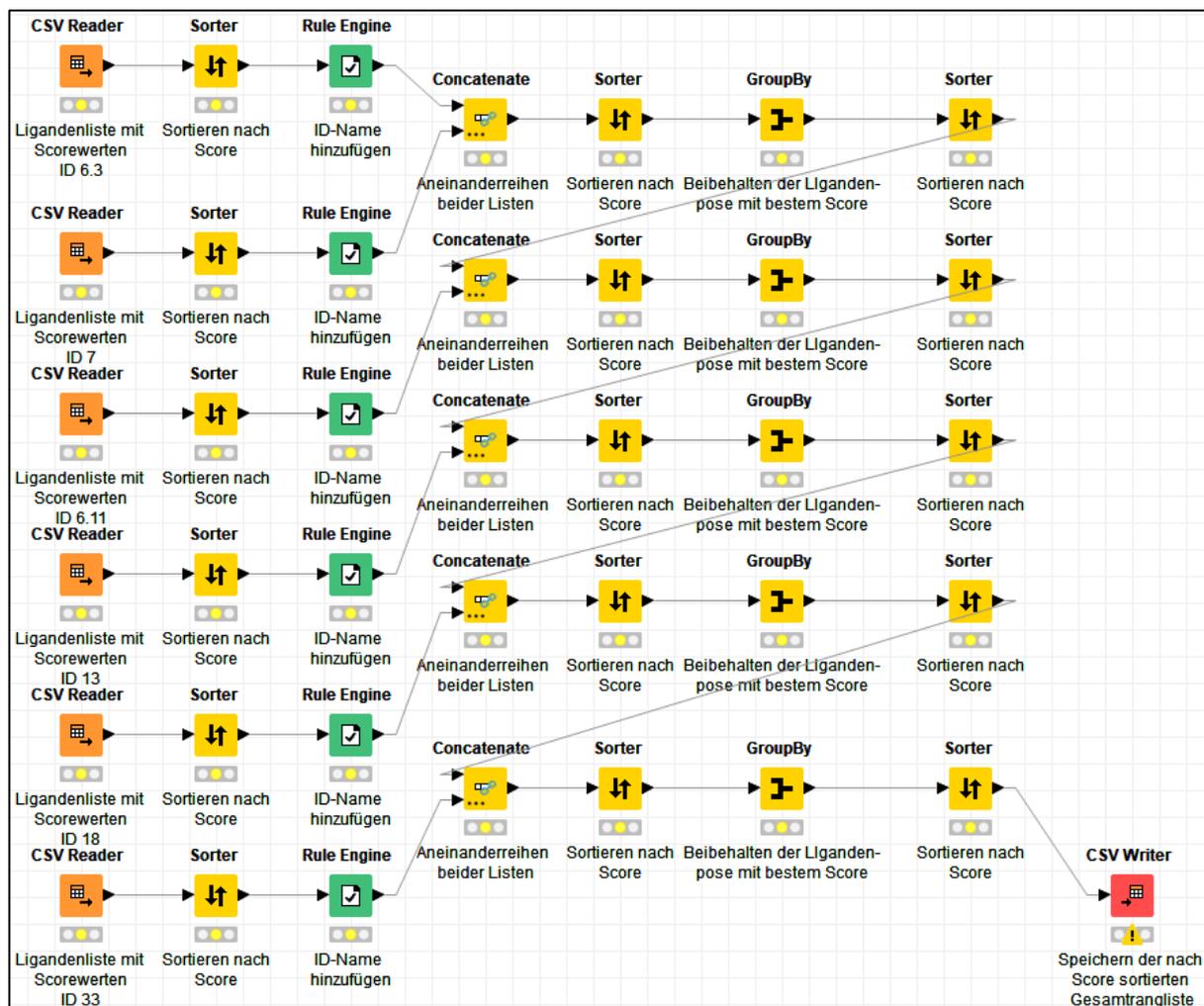


Abb. 7-5 – KNIME-Workflow für das Zusammenführen der sechs Ranglisten nach einem ausgewählten Scorewert

7.9.7 Berechnung der RDKit-Fingerprints

Die Berechnung der 2048 Bit langen RDKit-Fingerprints (s. Kap. 5.5.4) erfolgte mithilfe des „RDKit FINGERPRINT“-Nodes in KNIME ausgehend von der Moleküldatei im MAE-Format (s. Abb. 7-6).

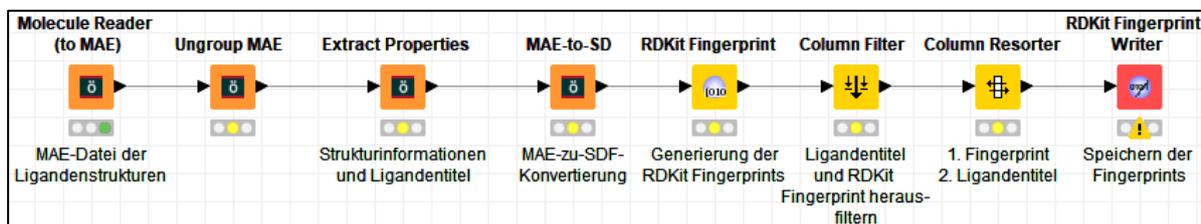


Abb. 7-6 – KNIME-Workflow zur Generierung der RDKit-Fingerprints

7.9.8 Leader-Clustering

Das Leader-Clustering (s. Kap. 5.5.4) erfolgte mit einem CPU-Skript von LIAO ET AL. [531] unter Verwendung der generierten RDKit-Fingerprints:

```
>> cpu_leader -t 0.3 input_fingerprint_file.fps > output_file.out
```

→ Durchführung für t = 0,2 und t = 0,3

7.10 Ableiten von Bindungsmodi nichtselektiver DLK-Inhibitoren

Die in Kap. 5.2.1, S. 152 ff. dargestellten Ausrichtungen und Einfärbungen der nichtselektiven DLK-Inhibitoren erfolgte – bis auf für AST-487⁶⁸ – auf Basis der Positionierung in Röntgenkristallstrukturen anderer Kinasen, wie in Tab. 7-6 dargestellt.

Tab. 7-6 – Röntgenkristallstrukturen (PDBs) für die Ableitung der Bindungsmodi der in dieser Arbeit verwendeten unselektiven DLK-Inhibitoren; in 1BYG und 4OTG dargestellt in Abb. 5-59, S. 167.

Ligand	PDB	Kinase
Staurosporin	1BYG	Humane C-terminale Src Kinase (CSK)
CEP-701	4OTG	Protein kinase C related kinase (PRK1)
(+)K-252a	1R0P	c-Met (Rezeptortyrosinkinase)
alle übrigen Staurosporinderivate: s. Staurosporin	-	-
Sunitinib	4AGD	VEGFR-2 (Rezeptortyrosinkinase)
SU-14813 : s. Sunitinib-Bindungsmodus	-	-
AG-013736	4AG8	VEGFR-2 (Rezeptortyrosinkinase)
KW-2449 : siehe AG-013736-Bindungsmodus	-	-
HKI-272 (Neratinib)	2JIV	EGFR (T790M, Rezeptortyrosinkinase)
SKI-606 (Bosutinib)	3UE4	Abl (Tyrosinkinase)
XL-880 (Foretinib)	3LQ8	c-Met (Rezeptortyrosinkinase)
EKB-569 (Pelitinib)	5VCW	Humane Myt1 (Serin/Threonin-Kinase)
VX-680/MK-0457 (Tozasertib)	2XYN	Humane Abl 2 (Tyrosinkinase)
JNJ-28312141	3KRJ	cFMS (Tyrosinkinase)
JNJ-7706621	4QMU	MST3 (Serin/Threonin-Kinase)
CHIR-258/TKI-258 (Dovitinib)	5OWQ	Humane STK10 (Serin/Threonin-Kinase)
TG-101348 (Fedratinib)	6VNE	JAK2 (Tyrosinkinase)
Crizotinib	3ZBF	Humane ROS1 (Rezeptortyrosinkinase)

7.11 Erfassung und Bearbeitung von Liganden

Die Liganden mit bekannter DLK-Aktivität und eindeutiger Strukturformel (s. Kap. 5.2.1) wurden aus den dort benannten Literaturstellen entnommen, mit dem Programm CHEMDRAW PRO 11.0 [254] gezeichnet, als SDF-Dateien in den MAESTRO WORKSPACE (MAESTRO 2016-1) überführt und dort mit dem 2D-SKETCHER überprüft und gegebenenfalls korrigiert. Je nach Literaturquelle werden folgende Abkürzungen für die Strukturen im Folgenden verwendet:

Karaman et al. [164]	(K)
Ferraris et al. [125]	(F)
Tao et al. [165]	(T)
Patel, Cohen et al. [166]	(PC)
Patel, Harris et al. [70]	(PH)
Patentschrift WO2013/174780 von Cohen, Huestis et al. [170]	(CH)
Patentschrift WO2014/111496 von Estrada, Liu et al. [171]	(EL)
Patentschrift WO2014/177060 von Chen, Dong et al. [172]	(CD)
Patentschrift WO2015/091889 von Lyssikatos, Liu et al. [174]	(LL)
Patentschrift WO2016/142310 von Estrada, Huestis et al. [175]	(EH)
Liganden aus den Röntgenkristallstrukturen 5CEO , 5CEP und 5CEQ [70]	(X)

Die DLK-Inhibitoren von Patel und den Patenten (PC bis EH) werden als (P) zusammengefasst dargestellt. Die 30 un-spezifischen DLK-Inhibitoren aus (K), (F) und (T) werden als (KFT) zusammengefasst dargestellt.

⁶⁸ Der Bindungsmodus von AST-487 in einer Kinase ist kristallographisch nicht belegt.

7.11.1 LigPrep

Die Ligandenvorbereitungen wurden nach folgendem Protokoll mit dem LIGPREP-Tool der SCHRÖDINGER SOFTWARE SUITE durchgeführt:

LIGPREP	
FORCE FIELD: OPLS3 oder OPLS2005 Auswahl des Kraftfelds	
IONIZATION:	(a) ✓Generate possible states at target pH: 7.0 +/- 2.0 Using ✓Epik
	oder (b) ✓Do not change
	oder (c) ✓Neutralize Using ✓Epik
✓Desalt	
	(a) ✓Generate tautomers oder (b) keine Tautomerisierung
STEREISOOMERS: (a) ✓Generate all combinations oder (b) ✓Determine chiralities from 3D structure	
GENERATE AT MOST: x per ligand max. Anzahl festlegen (# Stereo.)	

Folgende Ligandenvorbereitungen wurden durchgeführt (s. Tab. 7-7):

Tab. 7-7 – Ligandenpräparationen in dieser Arbeit. KF = Kraftfeld, Ion. = Ionisierung, Taut. = Tautomerisierung, Stereo. = Stereoisomere, #Stereo. = max. zugelassene Anzahl zu generierender Stereoisomere; Inh. = Inhibitoren; GEM = *Glide Ensemble Merge*, Lig. = Ligandenliste. Erläuterungen für die Optionen **a**, **b**, **c** (Ion.), **a** und **b** (Taut. bzw. Stereo.) im o.g. Skript (LIGPREP, s.o.). Erläuterungen zu **(EH)**, **(P)**, **(PC)**, **(PH)**, **(K)**, **(F)**, **(KFT)**, **(X)** s. S. 281.

Liganden	LIGPREP-Version/KF	Ion.	Taut.	Stereo.	#Stereo.	Besonderheiten
DLK-Inh. für QSAR (alle außer (EH))	LIGPREP v.3.8 für (K) und (F) mit K_d , v.3.7 (Rest) / OPLS3	(a)	(a)	(a) für (PC) und (PH) Rest: (b)	32	(PC) und (PH) : nur korrekte Stereochemie beibehalten
Lig. a (Aktive) OPLS3	LIGPREP v.3.8 für (K) und (F) mit K_d , LIGPREP v.4.0 (X) , v.3.7 (Rest) / OPLS3	(a)	(a)	(b)	32	
Lig. a (Decoys) OPLS3	LIGPREP v.3.9 / OPLS3	(b)	(b)	(b)	1	
Lig. a (Aktive+Decoys) OPLS2005	LIGPREP v.4.0 / OPLS2005	(b)	(b)	(b)	32	Input sind die Lig. a-OPLS3-LIGPREP-Ergebnisse
Lig. b (Aktive+Decoys)	LIGPREP v.4.0 / OPLS3	(a)	(a)	(b)	32	
820 DLK-Inh. aus (P)	LIGPREP v.4.1 / OPLS3	(c)	(b)	(b)	1	für Fingerprint-Generierung und Ähnlichkeitsanalyse
216 Aktive	LIGPREP v.4.1 / OPLS2005	(c)	(b)	(b)	1	Vorbereitung für Decoygenerierung (Lig. c)
Lig. c (Aktive + Decoys)	LIGPREP v.4.1 / OPLS3 und OPLS2005	(a)	(a)	(b)	32	
(X) aus 5CEO_opls2005_w, 5CEP_opls2005_w und 5CEQ_opls2005_w für Re- und Crossdocking (s. Kap. 5.4.1.1)	LIGPREP v.3.7 / OPLS2005	(a)	(a)	(a)	32	
(X) aus 5CEO_opls2005_w, 5CEP_opls2005_w und 5CEQ_opls2005_w; ANP, (KFT) für VSW-Exp. (s. Kap. 5.4.1.2)	LIGPREP v.3.8 für (K) und (F) mit K_d , LigPrep v.4.0 (X) , v.3.7 (ANP) / OPLS3	(a)	(a)	(b)	32	
ZINC-Strukturen	LIGPREP v.4.1 / OPLS2005	(a)	(a)	(b)	32	für HTvS
ZINC / Lig. c aus GEM des HTvS	LIGPREP v.4.1 / OPLS2005	(a)	(a)	(b)	32	für SP
5 Hits (A04 , C07 , D04 , E05 , E06)	LIGPREP v.5.3 / OPLS3e	(a)	(a)	(b)	32	für XP-Docking und anschließende MD

7.11.2 Bearbeitung der Ligandenlisten vor bzw. nach LigPrep

DLK-Inhibitoren für QSAR: Nach LIGPREP wurden die 1909 dreidimensionalen Strukturen auf ihre Korrektheit überprüft und ggf. eliminiert. Liganden mit unwahrscheinlicher Protonierung wurden gelöscht: Dabei handelte es sich um mehrfach protonierte Strukturen oder um States mit

einer erhöhten *Ionization Penalty*, insbesondere um Protonierungen an einem aromatischen Stickstoff. Es verblieben 1295 Strukturen (*Multiple States*) für das Docking zur Erstellung des Alignments für QSAR (s. Kap. 5.3.1).

Ligandenliste a (Lig. a): Es wurden 206 Aktive aus den besten Dockinglösungen für QSAR mit gleichmäßiger pK_i -Verteilung ausgewählt. Diese wurden mithilfe von MAESTRO in einer SMILES-Datei gespeichert, welche zur Decoygenerierung genutzt wurde. Dazu wurde die Ligandendatei auf der Webseite dude.docking.org/generate hochgeladen. Es wurde immer der niedrigste Protonierungszustand eines Aktiven beibehalten und die übrigen gelöscht. Die Decoy-Datei wurde in MONA v.2.1 geladen. Es handelte sich um 10.292 Strukturen, da acht Duplikate auftraten. Die Duplikate wurden aus der SMILES-Datei entfernt. Die Liste aus Aktiven und Decoys wurde mit (**KFT**) (s. S. 281 und S. 152 ff., 30 Verbindungen), ANP und (**X**) (s. S. 281) aufgefüllt bzw. – wenn schon vorhanden – ersetzt. Danach handelte es sich um 204 selektive DLK-Inhibitoren, 30 unspezifische DLK-Inhibitoren⁶⁹ und ANP (insgesamt 235 Aktive) sowie 10.292 Decoys (insgesamt: 10.527 *Single States*).

Ligandenliste b (Lig. b): Für die Erstellung der **Lig. b** wurden alle bis dahin veröffentlichten Strukturen (s. Kap. 5.2.1, alle außer (EH)⁷⁰) mit eindeutig festgelegter Stereochemie zusammengetragen (712 Strukturen). Zunächst wurden die Liganden aus (**P**) (s. S. 281) so ausgewählt, dass der Aktivitätsbereich des pK_i -Wertebereichs von 5-10 gleichmäßig abgedeckt ist. Es sollten etwa 150 Strukturen ausgewählt werden. Dazu wurden die Liganden (**P**) nach ihrem pK_i -Wert sortiert und die Auswahl folgendermaßen vorgenommen:

pK_i 5-6: alle	(21 Strukturen)
pK_i 6-7: jeder 5., beim 3. begonnen	(25 Strukturen)
pK_i 7-8: jeder 9., beim 4. begonnen	(31 Strukturen)
pK_i 8-9: jeder 10., beim 2. begonnen	(32 Strukturen)
pK_i 9-10: jeder 2., beim 2. begonnen	(31 Strukturen)

Wegen unklarer Stereoisomerie mussten 2 Liganden nachträglich gelöscht werden. Desweiteren wurden alle 30 Liganden aus (**KFT**) (s. S. 281) hinzugefügt sowie die drei Liganden (**X**) aus den Röntgenkristallstrukturen ergänzt, von denen einer bereits in der Auswahl vorhanden war. Insgesamt handelt es sich um 170 Aktive.

⁶⁹ Versehentlich kommen die DLK-unspezifischen Verbindungen AST-487 und VX-680 in dieser Ligandenliste doppelt vor, da sie jeweils aus zwei Literaturstellen hervorgehen. Es handelt sich um 235 „echt unterschiedliche“ Aktive in **Lig. a**.

⁷⁰ Die Strukturen dieser Patentschrift wurden erst im Laufe dieser Arbeit bekannt und wurden erst in **Lig. c** verarbeitet.

Aus den generierten States nach LIGPREP wurde jeweils nur ein Protonierungszustand beibehalten. Nach den gleichen Regeln wie für die Ligandenvorbereitung für QSAR (s. S. 282 f.) wurden die Strukturen mit unwahrscheinlicheren States gelöscht. Mit diesen 170 DLK-Inhibitoren wurde die Decoygenerierung – wie für **Lig. a** – durchgeführt, nur dass nicht der niedrigste Protonierungszustand jedes Aktiven und dessen Decoys beibehalten wurde, sondern zunächst alle generierten Decoys beibehalten wurden. Diese wurden anschließend mit der Einstellung *Topology, Tautomer and Charge equal* in MONA v.2.1 geladen. Auf diese Weise wurden alle Duplikate entfernt sowie verschiedene Protonierungszustände und Tautomere zusammengefasst. Es handelte sich um 12.387 Decoys. Anschließend wurden die Aktiven und Decoys einem OPLS3-LIGPREP unterzogen (insgesamt: 12.557 *Single States*).

Ligandenliste c (Lig. c): Von den 835 in Kapitel 5.2.1 beschriebenen DLK-Inhibitoren wurden solche ohne definierten K_i - bzw. K_d -Wert entfernt. Die restlichen 820 Strukturen wurden zunächst mit LIGPREP vorbereitet und nur der neutrale Zustand generiert. Danach wurden die Strukturen in CANVAS importiert. Die linearen binären Fingerprints wurden mit CANVAS für alle Strukturen berechnet (*fp_linear_1*). Auf Basis dieser Fingerprints wurde ein hierarchisches Clustering mit CANVAS mit der Linkage-Methode „Average“ auf Basis des Tanimotokoeffizienten durchgeführt. Eine gewünschte Cluster-Anzahl von 200 wurde vorgegeben. Nachträglich wurde eine Struktur wegen unbekannter Konfiguration gelöscht. Die Liste wurde mit (**KFT**) (s. S. 281) mit vorhandenen IC_{50} -Werten ergänzt, sodass alle 30 unspezifischen DLK-Inhibitoren in der Liste enthalten waren und – wenn nicht vorhanden – mit (**X**) ergänzt, sodass sich eine Liste von 216 Aktiven ergab. Für die Decoygenerierung wurden zur Vereinheitlichung der Protonierungszustände alle 216 Liganden über OPLS2005-LIGPREP neutralisiert, der Output als SMILES-Datei abgespeichert und für die Decoygenerierung hochgeladen (wie auch für **Lig. a** und **b**). Nach Entfernung aller Duplikate und Beibehaltung nur eines Protonierungszustands bzw. Tautomers mithilfe von MONA v.2.1 handelte es sich um 15.851 Decoys. Die Aktiven und Decoys der **Lig. c** wurden sowohl im OPLS3- als auch im OPLS2005-Kraftfeld präpariert (insgesamt: 16.067 *Single States*).

ZINC-Strukturen: Ein Shell-Skript wurde mit den Voreinstellungen „Clean“, „Druglike“, „In-Stock“, „3D-Bibliothek“ von der Webseite <http://zinc15.docking.org/tranches/home/#> heruntergeladen und im Terminal ausgeführt:

```
>> csh ZINC-CleanD-sdf.gz.curl
```

In der Folge werden die sdf.gz-Moleküldateien in Unterordner heruntergeladen. Diese wurden zu sdf-Dateien entpackt und alle gemeinsam in MONA v.2.1 (*Topology, Charges, Tautomers equal*)

geladen. Es wurden 5.142.887 Strukturen gezählt. Nach Anwendung der in Kap. 5.2.3, S. 157, Tab. 5-15 genannten Filterkriterien mithilfe von MONA v.2.1 verblieben 3.912.599 Strukturen, die vor dem HTvS mit einem OPLS2005-LIGPREP vorbereitet wurden. Die LIGPREP-Output-Datei wurde für das virtuelle Screening mit der OPLS2005-LIGPREP-Output-Datei der 216 Aktiven der **Lig. c** mithilfe eines SCHRÖDINGER-Befehls vereinigt:

```
>> $SCHRODINGER/utilities/structcat Datei1 Datei2 -o output.mae
```

Mithilfe des *Glide Ensemble Merge* wurden die HTvS-Dockingposen aus den sechs Proteinkonformationen zusammengefasst:

```
>> $SCHRODINGER/utilities/glide_ensemble_merge HTvS_OUT_ID6.3.maegz HTvS_OUT_ID7.ma-
egz HTvS_OUT_ID6.11.maegz HTvS_OUT_ID13.maegz HTvS_OUT_ID18.maegz HTvS_OUT_ID33.maegz
```

HTvS_OUT_IDx.maegz = Output-Datei des HTvS-Dockings in die Proteinkonformation ID x (z.B. 6.3)

Mit dieser zusammengefassten Datei wurde ein OPLS3-LIGPREP durchgeführt und diese Ergebnisse in vier etwa gleich große Subgruppen für das SP-Docking aufgeteilt:

```
>> $SCHRODINGER/utilities/maesubset -n 1:366883 LigPrepHTVOutputDatei.mae > subset1.mae
>> $SCHRODINGER/utilities/maesubset -n 366884:673767 LigPrepHTVOutputDatei.mae > subset2.mae
>> $SCHRODINGER/utilities/maesubset -n 673768:1040651 LigPrepHTVOutputDatei.mae > subset3.mae
>> $SCHRODINGER/utilities/maesubset -n 1040652:1467533 LigPrepHTVOutputDatei.mae > subset4.mae
```

Nach Durchführung des SP-Dockings wurden die SP-Dockingposen aus den vier Subsets einer Protein-ID zunächst zu einer MAE-Datei zusammengefasst:

```
>> $SCHRODINGER/utilities/glide_ensemble_merge SP-out_ID7_subset1.maegz SP-out_ID7_subset2.ma-
egz SP-out_ID7_subset3.maegz SP-out_ID7_subset3.maegz SP-out_ID7_subset4.maegz
```

Aus dieser nach Dockingscore sortierten Gesamtdatei einer Protein-ID wurde für den weiteren Verlauf die besten 5 % der jeweiligen Liste verwendet, z.B. für ID 7 mit 808.949 Liganden:

```
>> $SCHRODINGER/utilities/maesubset -n 1:40447 GlideEnsembleMergeDatei_SP-out_ID7.maegz > SP-
out_ID7_top5.mae
```

7.11.3 Konvertierung von Moleküldateiformaten

Folgende SCHRÖDINGER-Skripte wurden für die Konvertierung von Dateiformaten verwendet (auch umkehrbar):

MAE- zu MOL2-Format:

```
>> $SCHRODINGER/utilities/mol2converter t -imae input.mae -omol2 output.mol2
```

MAE- zu SDF-Format:

```
>> $SCHRODINGER/utilities/sdconvert -imae input.mae -osd output.sdf
```

7.12 Berechnung von physikochemischen bzw. molekularen Eigenschaften und Erstellung von Histogrammen

Die in Kapitel 5.2.3 analysierten molekularen Eigenschaften wurden mit der Software MONA v.2.1 berechnet und die Histogramme zunächst mit MONA v.2.1 erstellt. Die Häufigkeitswerte daraus wurden in MICROSOFT EXCEL überführt, um die Histogramme verschiedener Ligandengruppen nebeneinander darstellen zu können.

7.13 Darstellung von dreidimensionalen Strukturen und Erstellen von Abbildungen und Diagrammen

Die dreidimensionalen Darstellungen von Proteinen und Molekülen wurden mit MAESTRO (SCHRÖDINGER SUITE, Maestro Version 12.3.013, MMshare Version 4.9.013, Release 2020-1, Schrödinger LLC, New York, Platform Windows-x64) erstellt. Für die Darstellungen in Abb. 4-9, Abb. 7-2 und Abb. 7-3 wurde eine Vorgängerversion dieser Software verwendet. Die in dieser Arbeit verwendeten DLK-Röntgenkristallstrukturdaten stammen aus den PDB-Codes **5CEN**, **5CEO**, **5CEP** und **5CEQ**, veröffentlicht von PATEL ET AL. [70].

MICROSOFT EXCEL, GNUPLOT und NT ANALYSIS v1.5.41 wurden für die Erstellung der Diagramme in dieser Arbeit verwendet. Die Abbildungen wurden zum größten Teil mit MICROSOFT POWER POINT bearbeitet bzw. erstellt.

7.14 Kauf von Testsubstanzen

Die 58 Testsubstanzen wurden von AMBINTER, c/o Greenpharma, Orléans, Frankreich erworben. Die Reinheit aller Substanzen wurde von dem Unternehmen mit > 90 % angegeben. Bei den Substanzen mit Stereozentren handelte es sich nach Lieferantenaussage um Racemate.

7.15 Charakterisierung der verwendeten DLK-Lösungen

Die in den Assays verwendete MBP-TEV-DLK₁₄₀₋₅₁₇ aus der Arbeit von Dr. Jorge Enrique Duque Escobar liegt in einem Puffer gelöst vor, dessen Zusammensetzung in Tab. 7-8 aufgeführt ist.

Tab. 7-8 – Zusammensetzung des Puffers, in der die MBP-TEV-DLK₁₄₀₋₅₁₇ gelöst vorliegt.

Pufferbestandteile	Konzentration
Tris-HCl pH = 8,0	20 mM
NaCl	1 M
EDTA	1 mM
DTT	1 mM
PMSF	1 mM
Glycerol	10 %
dH ₂ O	

7.15.1 Durchführung der SDS-PAGE

Das Trenngel (s. Tab. 7-9) wurde zu bis zu 2/3 zwischen zwei Glasplatten in die Gießkammer gegossen. 1 ml Isopropanol wurden zur Glättung der oberen Kante des entstehenden Gels hinzugegeben. Nach Erstarren des Trenngels (20 min.) wurde das Isopropanol wieder entfernt. Die Lösung für das Sammelgel (s. Tab. 7-9) wurde bis zum Rand zwischen die Glasplatten über das Trenngel pipettiert und sofort danach ein Kamm zur Bildung der Probestaschen in die Lösung des Sammelgels gesetzt.

Tab. 7-9 – Zusammensetzung des Trenngels und des Sammelgels sowie der verwendeten Tris-Puffer.

Trenngel (pH 8,8)		Tris-Puffer 1,5 M (pH 8,8)		Sammelgel (pH 6,8)		Tris-Puffer 0,5 M (pH 6,8)	
dH ₂ O	2 ml	Tris HCl pH 8,8	1,5 M	dH ₂ O	1,8 ml	Tris HCl pH 6,8	0,5 M
Acrylamid 30 %	1,65 ml	SDS	0,4 % (w/v)	Acrylamid 30 %	400 µl	SDS	0,4 % (w/v)
Tris-Puffer 1,5 M (pH 8,8)	1,25 ml	dH ₂ O	ad 200 ml	Tris-Puffer 0,5 M (pH 6,8)	0,75 ml	dH ₂ O	ad 100 ml
SDS 10 %	50 µl			SDS 10 %	30 µl		
TEMED	5 µl			TEMED	3 µl		
APS 10 %	50 µl			APS 10 %	30 µl		

Nach Erstarrung des Sammelgels (20 min.) wurde das Gel in den beiden Glasplatten in die Elektrophorese-Kammer gesetzt. Die Kammer wurde mit dem Laufpuffer 1x (s. Tab. 7-10) befüllt und der Kamm aus den Probestaschen vor Zugabe der Proteinproben entfernt.

Die Proteinproben wurden **Tab. 7-10** – Zusammensetzung des Laufpuffers 1x und 5x sowie des Loading dyes.

in den entsprechenden Mengen in Vials pipettiert und – falls nötig – mit Wasser auf 20 µl Endvolumen gebracht. Zu jedem Vial

Laufpuffer 5x		Laufpuffer 1x		Loading dye	
dH ₂ O	2 ml	Laufpuffer 5x	400 ml	0,5 M Tris (pH 6,8)	2,5 ml
Tris HCl	250 mM	dH ₂ O	ad 2 l	20 % SDS	2,0 ml
Glycin	1,9 M			1 % Bromophenolblau	0,2 ml
SDS	0,5 %			Glycerol	5,0 ml
dH ₂ O	ad 2 l			Aliquotiert zu 250 µl, vor Gebrauch werden 15 µl Mercaptoethanol zugegeben	

wurden 10 µl Loading Dye (s. Tab. 7-10) gegeben und für 5 min. bei 95 °C im Thermomixer (Eppendorf) denaturiert. Die Proben und 5,5 µl der Ladder-X-Lösung als Größenmarker wurden jeweils in eine Probestasche pipettiert.

An die Elektrophorese-Kammer wurde eine Spannung von 80 V für 15-20 min. bis zum Übergang der Proben vom Sammelgel ins Trenngel angelegt, danach wurde die Spannung für etwa 1,5 bis 2 Stunden auf 100 V gehalten. Nach der erfolgten Elektrophorese wurde das Gel zwei Mal für je 5 min. mit Wasser (dH₂O) gewaschen. Für 15 min. wurde das Gel dann in einer Coomassie-Lösung (Zusammensetzung s. Tab. 7-11) geschwenkt. Nach Entsorgung dieser Färbelösung wurde dann Entfärberlösung zugegeben und eine Stunde inkubiert. Die Entfärberlösung wurde durch frische Entfärberlösung ersetzt und für eine

weitere Stunde inkubiert. Die so hergestellten Gele wurden abfotografiert und die Bandenintensitäten mit der Software IMAGE LAB 5.2.1 (BIORAD LABORATORIES, München, Deutschland) ausgewertet.

Tab. 7-11 – Zusammensetzung der Coomassie-Färbelösung und der Entfärber-Lösung.

Coomassie-Färbelösung		Entfärber-Lösung	
Coomassie-Brillant-Blau G-250	0,25 % (w/v)	Isopropanol	25 % (v/v)
Methanol	40 % (v/v)	Eisessig	10 % (v/v)
Eisessig	10 % (v/v)	Methanol	10 % (v/v)
dH ₂ O	ad 1 l	dH ₂ O	ad 1 l

Semiquantitative Bestimmung der DLK-Fraktionen

Zur Vorbereitung der eigenen Untersuchungen wurde die MBP-TEV-DLK₁₄₀₋₅₁₇-Menge in der für die Assays zu verwendenden Proteinlösungen semiquantitativ mittels SDS-PAGE bestimmt. Dazu wurden zunächst 16 Proteinlösungen (Fraktionen a-p nach der Aufreinigung bei der DLK-Herstellung von Dr. Jorge Enrique Duque Escobar [498]) mit jeweils 20 µl eine SDS-PAGE durchgeführt (s.o. und Tab. 7-12), um anhand der Bandenintensität eine Vorauswahl von Fraktionen vorzunehmen, die eine besonders intensive Bande bei 94 kDa (für MBP-TEV-DLK₁₄₀₋₅₁₇) aufweisen und wenig verunreinigt sind.

Tab. 7-12 – Pipettierschema für die SDS-PAGE zur groben Bestimmung der Anwesenheit einer ausreichenden MBP-TEV-DLK₁₄₀₋₅₁₇-Menge; BSA = bovines Serumalbumin; LD = Loading dye

Bezeichnung	Zusammensetzung
Ladder X	5,5 µl Ladder X
a bis p	20 µl MBP-TEV-DLK ₁₄₀₋₅₁₇ -Lösung + 10 µl LD
BSA 1 µg	0,5 µl BSA-Lsg. (2 mg/ml) + 19,5 µl dH ₂ O + 10 µl LD

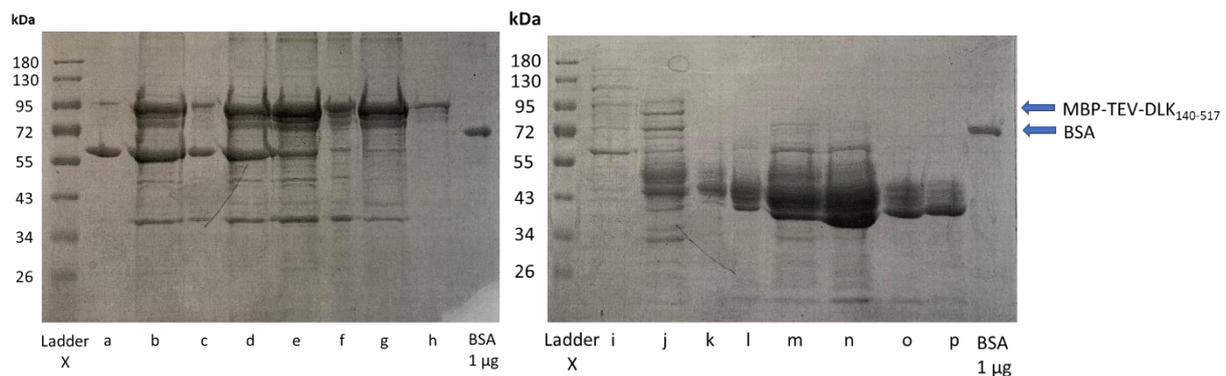


Abb. 7-7 – SDS-PAGE von 16 Proteinlösungen (20 µl) im Vergleich zu einer BSA-Lösung (absolute Menge: 1 µg BSA).

Die SDS-PAGE der Fraktionen e und g (s. Abb. 7-7) zeigten eine starke MBP-TEV-DLK₁₄₀₋₅₁₇-Bande (94 kDa) sowie vergleichsweise wenig intensive Verunreinigungsbanden und die MBP-TEV-DLK₁₄₀₋₅₁₇-Konzentration für Fraktion e und g war nach Voruntersuchungen nahe an der zu erwarteten Konzentration von 1 µg in 7,5 µl Lösung des Proteins. Beide Fraktionen wurden zur genaueren Bestimmung der Konzentration mit einer MBP-TEV-DLK₁₄₀₋₅₁₇-Positivkontrolle (1 µg in 7,5 µl

Lösung) verglichen. Dazu wurde eine SDS-PAGE mit einer Verdünnungsreihe der beiden Fraktionen e und g und 7,5 µl der Positivkontrolle (entspricht 1 µg) durchgeführt. (s. Tab. 7-13).

Tab. 7-13 – Pipettierschema für die SDS-PAGE zur Bestimmung der MBP-TEV-DLK₁₄₀₋₅₁₇-Menge der Fraktionen e und g im Vergleich zu einer Positivkontrolle (1 µg in 7,5 µl). LD = Loading Dye.

Bezeichnung	Zusammensetzung
Ladder X	5,5 µl Ladder X
7,5 µl Positivkontrolle	7,5 µl MBP-TEV-DLK ₁₄₀₋₅₁₇ -Lösung [1µg/7,5µl] + 12,5 µl dH ₂ O + 10 µl LD
2 µl Fraktion e	2 µl MBP-TEV-DLK ₁₄₀₋₅₁₇ -Lösung Fraktion e + 18 µl dH ₂ O + 10 µl LD
4,8 µl Fraktion e	4,8 µl MBP-TEV-DLK ₁₄₀₋₅₁₇ -Lösung Fraktion e + 15,2 µl dH ₂ O + 10 µl LD
7,5 µl Fraktion e	7,5 µl MBP-TEV-DLK ₁₄₀₋₅₁₇ -Lösung Fraktion e + 12,5 µl dH ₂ O + 10 µl LD
7,5 µl Positivkontrolle	7,5 µl MBP-TEV-DLK ₁₄₀₋₅₁₇ -Lösung [1µg/7,5µl] + 12,5 µl dH ₂ O + 10 µl LD
2 µl Fraktion g	2 µl MBP-TEV-DLK ₁₄₀₋₅₁₇ -Lösung Fraktion g + 18 µl dH ₂ O + 10 µl LD
4,8 µl Fraktion g	4,8 µl MBP-TEV-DLK ₁₄₀₋₅₁₇ -Lösung Fraktion g + 15,2 µl dH ₂ O + 10 µl LD
7,5 µl Fraktion g	7,5 µl MBP-TEV-DLK ₁₄₀₋₅₁₇ -Lösung Fraktion g + 12,5 µl dH ₂ O + 10 µl LD

Die Auswertung (s. Abb. 7-8) ergab, dass ein Volumen von 7,38 µl für Fraktion e und 8,41 µl für Fraktion g jeweils 1 µg MBP-TEV-DLK₁₄₀₋₅₁₇ enthält.

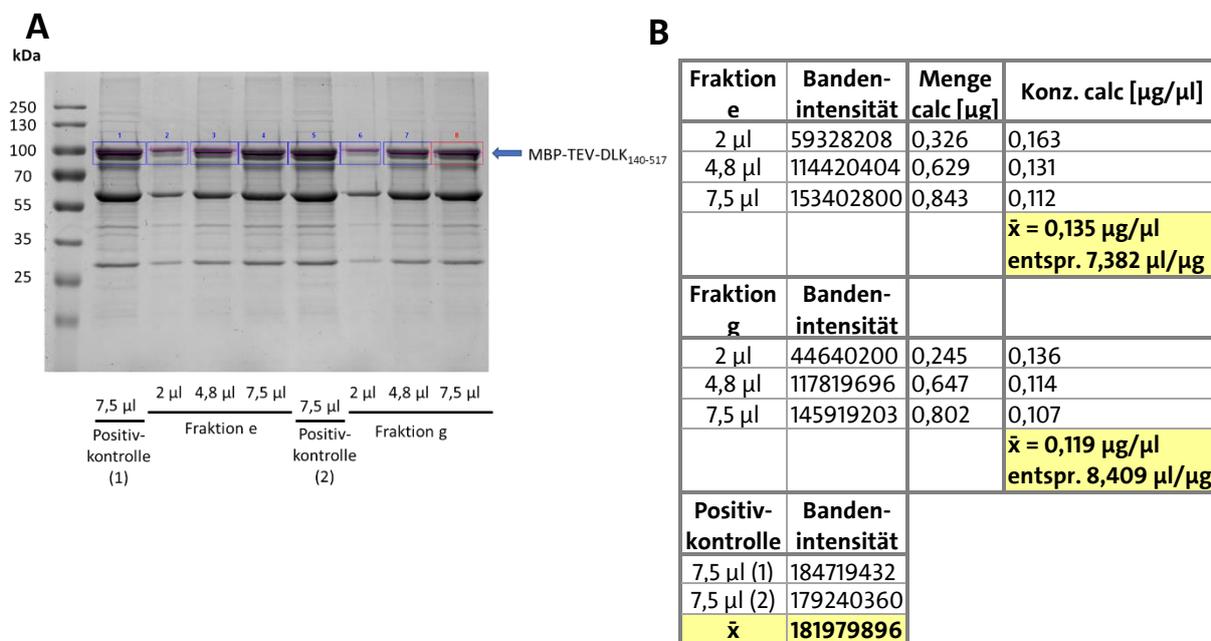


Abb. 7-8 – Ergebnisse der SDS-PAGE der Fraktion-e- und g-Verdünnungsreihe im Vergleich zu einer Standardlösung. **A** – Gel mit den gefärbten Banden. **B** – Bandenintensitäten und Berechnung der MBP-TEV-DLK₁₄₀₋₅₁₇-Menge [µg] in Fraktion e und g ($\text{Bandenintensität}_{\text{Fraktion}}/181979896$).

7.16 Durchführung des ADP-Glo™ Kinase Assays

Die MBP-TEV-DLK₁₄₀₋₅₁₇-Lösung „Fraktion e“ (s. Kap. 7.15.1, im Folgenden: „DLK-Lösung“) wurde für die Durchführung der ADP-Glo™ Kinase Assays verwendet. Die Messungen erfolgten in 96-Well-Mikrotiterplatten. Die Messung der Lumineszenz erfolgte für alle Durchführungen mit dem Luminometer von BERTHOLD CENTRO L 960 (MIKROWIN-Software, Measurement Time 1 sec, Shake Time 0,1 sec.). Die Inkubation erfolgte unter Lichtausschluss durch Abdeckung mit Aluminiumfolie und unter ständigem Schwenken auf einem Minishaker bei Raumtemperatur.

Die β -Casein-Lösungen wurden aus β -Casein (Festschubstanz) und dem Reaktionspuffer hergestellt. Verdünnungen von ATP bzw. ADP erfolgte immer ausgehend von einer 10 mM-Lösung (ADP-Glo-Kit) mit dem Reaktionspuffer (s. Tab. 7-14). Verdünnungen der DLK-Lösungen erfolgten ebenso mit dem Reaktionspuffer.

Tab. 7-14 – Zusammensetzung des Reaktionspuffers (PROMEGA Handbuch).

Reaktionspuffer	
Tris pH 7,5	40 mM
MgCl ₂	40 % (v/v)
BSA	0,1 mg/ml

(1.) Bestimmung der Aktivität der DLK-Lösungen

Zu den jeweiligen DLK-Lösungen wurde der Reaktionspuffer zugesetzt, sodass sich ein Gesamtvolumen von 22 μ l ergab (s. Tab. 7-15). Zu jedem Well wurden 3 μ l einer Mischung aus ATP (2,5 μ l einer 10 μ M-Lösung) und β -Casein-Lösung (0,5 μ l einer 1 μ g/ μ l-Lösung) gegeben, sodass sich eine Endkonzentration von 1 μ M ATP und 0,02 μ g/ μ l β -Casein in 25 μ l Lösung (= Reaktionsansatz) ergab. Nach einer Inkubationszeit (Ablauf der Kinasereaktion) von 1 h wurden 25 μ l ADP-Glo-Reagenz zu jedem Well zugesetzt und weitere 40 min. inkubiert. Dann erfolgte der Zusatz von 50 μ l ADP-Glo-Detektionsreagenz zu jedem Well (Gesamtvolumen im Well: 100 μ l). Nach einer Inkubationszeit von 30 min. unter Lichtausschluss erfolgte die Messung der Lichtintensitäten mit dem Luminometer.

Tab. 7-15 – Eingesetzte Mengen der MBP-TEV-DLK₁₄₀₋₅₁₇-Fraktionen für den ADP-Glo™ Kinase Assay

Position (Well)	DLK-Fraktion	Menge DLK [μ l]	Menge Reaktionspuffer [μ l]
C	e	7,38	14,62
D	g	8,40	13,60
G	Positivkontrolle	7,50	14,50
H	- (Blindwert)	-	22

(2.) Optimierung der Substratkonzentration (β -Casein)

In diesem Versuch wurden die Volumina im Vergleich zum Vorversuch (1.) halbiert, um DLK-Lösung und Reagenzienmengen einzusparen, sodass das Volumen für die Kinasereaktion 12,5 μ l betrug. Die in Tab. 7-16 aufgeführten DLK- und Reaktionspuffervolumina wurden gemischt bzw. eingesetzt sowie jeweils mit 2,5 μ l einer DMSO-Puffer(1:3)-Mischung versetzt. Der ATP- β -Casein-Zusatz erfolgte – mit variierender β -Casein-Konzentration (s. Tab. 7-16, grau) – nach einer 30 min. Inkubationszeit von DLK-Lösung, Reaktionspuffer und DMSO-Puffer-Mischung, um die Inhibitor-Wechselwirkungszeit zu simulieren. Die eingesetzte ATP-Konzentration betrug in jedem Well 10 μ M im Endvolumen dieses Reaktionsansatzes. Dann wurde der Ansatz 60 min. bei Raumtemperatur inkubiert (Ablauf der Kinasereaktion). Die nachfolgende Reagenzienzugabe und Messung des Lumineszenzsignals entsprach dem in (1.) beschriebenen Protokoll, nur dass 12,5 statt 25 μ l des ADP-Glo-Reagenz und 25 statt 50 μ l des ADP-Glo-Detektionsreagenz zugegeben wurde.

Tab. 7-16 – Ansatz Kinasereaktion zur Optimierung der Substratkonzentration. [K] = mit Kinase, [oK] = ohne Kinase.

Bez.	DLK-Lösung (Fraktion e) [μ l]	Reaktionspuffer [μ l]	DMSO-Puffer-Mischung (1:3) [μ l]	ATP- β -Casein-Mischung* [μ l]	*Konz. der zugesetzten β -Casein-Lösung [μ g/ μ l]
0,5 [K]	3,69	4,81	2,50	1,50	0,5
1 [K]	3,69	4,81	2,50	1,50	1
2,5 [K]	3,69	4,81	2,50	1,50	2,5
5 [K]	3,69	4,81	2,50	1,50	5
10 [K]	3,69	4,81	2,50	1,50	10
0,5 [oK]	-	8,50	2,50	1,50	0,5
1 [oK]	-	8,50	2,50	1,50	1
2,5 [oK]	-	8,50	2,50	1,50	2,5
5 [oK]	-	8,50	2,50	1,50	5
10 [oK]	-	8,50	2,50	1,50	10

(3.) Optimierung der DLK-Konzentration

a) Für die Erstellung der ATP-zu-ADP-Konversionskurve wurden ATP und ADP in unterschiedlichen Verhältnissen gemischt (s. Tab. 7-17). Für eine 10 %-Konversion (d.h. 10 % ADP und 90 % ATP) wurden 9 μ l einer 100 μ M ATP-Lösung mit 1 μ l einer 100 μ M ADP-Lösung gemischt. Dazu wurden 2 μ l Reaktionspuffer gegeben (statt β -Casein-Lösung). Jeder Ansatz im Well enthielt 8,5 μ l Reaktionspuffer, 2,5 μ l DMSO-Puffer-Mischung (1:3, 5% DMSO im Ansatz) und 1,5 μ l der entsprechenden ATP-ADP-Puffer-Mischung (in diesem Beispiel: 9 μ M ATP und 1 μ M ADP in diesem 12,5 μ l-Ansatz). Danach folgte das übliche Protokoll bis zur Messung der Lumineszenz nach Ablauf der Kinasereaktion (siehe (1.)).

Tab. 7-17 – Eingesetzte Mischungsverhältnisse von ADP und ATP im Assay für die Erstellung einer ATP-ADP-Konversionskurve

% Konversion	Konz. ADP [μ M] im 12,5 μ l-Ansatz	Konz. ATP [μ M] im 12,5 μ l-Ansatz
0	0	10
1	0,1	9,9
2	0,2	9,8
3	0,3	9,7
4	0,4	9,6
5	0,5	9,5
10	1	9
20	2	8
40	4	6
60	6	4
80	8	2
100	10	0

b) Für die Bestimmung der optimalen DLK-Konzentration wurde eine 8-Punkte-Verdünnungsreihe hergestellt mit Konzentrationen von 0,0004 bis 0,0921 μ g/ μ l MBP-TEV-DLK₁₄₀₋₅₁₇ im 12,5 μ l-Reaktionsansatz. Die übrigen Bedingungen für die Kinasereaktion wurden wie zuvor (2.) beschrieben beibehalten: 8,5 μ l der jeweiligen DLK-Puffer-Mischung, 10 μ M ATP, 0,02 μ g/ μ l β -Casein und 5 % DMSO in einem 12,5 μ l Ansatz. Die Inkubationszeiten und Reagenzienzugabe erfolgte wie zuvor in Absatz (2.) beschrieben.

(4.) Aktivitätsbestimmung mit Referenzsubstanzen und den virtuellen Hits

a) Die Herstellung der Referenzsubstanzlösungen erfolgte durch Lösen der Substanzen in 100 % DMSO zu einer 10 mM-Lösung (Bosutinib, Sunitinib als Sunitinibmalat und Crizotinib, Einwaage bzw. Lösungsmittelzugabe s. Anhang H, Tab. H 2) bzw. 2,5 mM-Lösung (GNE-3511, Lösungsmittelzugabe s. Anhang, Tab. H 1) und anschließender Verdünnung mit dem Reaktionspuffer zu einer 2500 μ M- bzw. 625 μ M-Lösung (DMSO:Puffer = 1:3). Nach Zusatz von je 2,5 μ l dieser Inhibitor-Lösung zu 8,5 μ l DLK-Puffer-Ansatz (DLK-Konz. = 0,0098 μ g/ μ l) pro Well erfolgte die ATP- β -Casein-Zugabe sofort ohne Inkubationszeit. Bereits nach 10 min. Inkubationszeit wurden dann 12,5 μ l ADP-Glo-Reagenz zugegeben. Dann wurde das übliche Protokoll bis zur Messung (s. (2.)) fortgesetzt. Zusätzlich wurde eine Blindprobe (ohne Kinase, d.h. 8,5 μ l Reaktionspuffer) und eine Probe ohne Inhibitor (Kontrolle, d.h. statt Inhibitorlösung eine DMSO:Puffer (1:3) Mischung zugesetzt) gemessen.

b) Die Durchführung des Screenings der 58 Hit-Substanzen aus dem virtuellen Screening mit den bislang optimierten Parametern erfolgte nach dem in Abb. 7-9 dargestellten Schema.

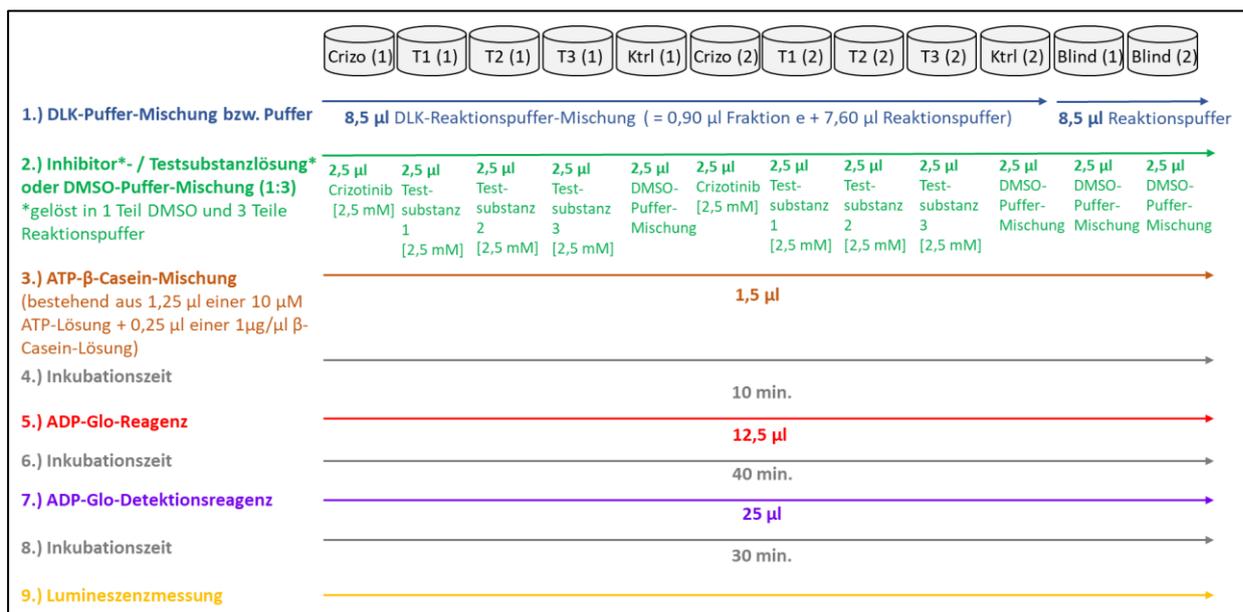


Abb. 7-9 – Pipettierschema in den einzelnen Wells 1-12 mit Reihenfolge der zugegebenen Lösungen sowie den Inkubationszeiten. Die Messung der Lumineszenz erfolgte in der hier angegebenen Reihenfolge (Testreihe C bis T).

Die Testsubstanzen wurden wie zuvor das Crizotinib zunächst in 100 % DMSO gelöst⁷¹ und mit dem Reaktionspuffer 1:3 verdünnt, sodass eine 2500 μ M-Lösung resultierte. Nach Zugabe von 2,5 μ l dieser Lösung zur DLK-Puffer-Mischung und nachfolgender ATP- β -Caseinzugabe ergab sich eine Konzentration von 500 μ M im Reaktionsansatz. Durch Rechenfehler ergaben sich für die

⁷¹ Vorlage der Substanz (etwa 1 mg) und Zusatz des benötigten Lösungsmittelvolumens zur Erreichung einer 10 mM-Konzentration (s. Anhang H, Tab. H 1)

drei Testsubstanzen **E05**, **E07** und **E08** versehentlich abweichende Konzentrationen (s. Tab. 5-55, S. 232). **E03** und **E04** wurden in Testreihe T (s. Kap. 5.6.1, S. 232, Abb. 5-87) in korrekter Konzentration eingesetzt, nicht in Testreihe Q. Jeweils zwei Reaktionsansätze mit drei Testsubstanzen und Crizotinib wurden in einem Durchgang gemessen (1 Testreihe). Die ersten beiden Testreihen A und B erfolgten in einer anders als in Abb. 7-9 dargestellten Pipettier- und Messreihenfolge (s. Kap. 5.6.1, Abb. 5-87). Die Reihenfolge ist in den Abbildungen der Messergebnisse entsprechend dargestellt (Messung: links nach rechts).

c) Beim Re-Testing der 12 Testsubstanzen (s. Kap. 5.6.1, S. 235, Abb. 5-88) wurde je eine Lösung pro Durchführung in insgesamt zwei Durchführungen gemessen. Die Lösungen- und Reagenzienvolumina sowie die Inkubationszeiten wurden auf die gleiche Weise wie zuvor bei den 58 Testsubstanzen gewählt.

d) Für die IC_{50} -Bestimmung von **E05** bzw. Crizotinib wurde die 8,42 mM-**E05**-Lösung (100 % DMSO) bzw. die 10 mM Crizotiniblösung (100 % DMSO) mit 100 % DMSO in 1:1-Verdünnungsschritten verdünnt (7 Konzentrationen). Die so hergestellten Lösungen wurden jeweils 1:3 mit Reaktionspuffer verdünnt. 2,5 μ l dieser Lösungen wurden im Reaktionsansatz eingesetzt (Konzentrationen im Reaktionsansatz s. Tab. 7-18).

Tab. 7-18 – Konzentration von **E05** bzw. Crizotinib im 12,5 μ l-Reaktionsansatz für die IC_{50} -Bestimmung

E05 [μM]	Crizotinib [μM]
421,07	500,00
210,54	250,00
105,27	125,00
52,63	62,50
26,32	31,25
13,16	15,63
6,58	7,81

Zuerst wurde die Kontrolllösung (DLK ohne **E05** bzw. Crizotinib), dann die **E05**- bzw. Crizotinib-Lösungen in aufsteigender Substanzkonzentration und zum Schluss die Blindlösung (ohne DLK und ohne Test-/Referenzsubstanz) gemessen. Die eingesetzten Enzym-, Puffer-, ATP-, β -Casein- und Reagenzlösungen sowie die Inkubationszeiten entsprechen den zuvor im Substanzscreening eingesetzten Mengen, Konzentrationen und Zeiten (s. Abb. 7-9). Die Messpunkte wurden mit dem 4-PL- und 5-PL-Modell (s. Kap. 4.10.2, S. 113) mit dem „drc“-Paket der Software R (Version 3.6.0) [489] zur Berechnung des IC_{50} -Werts gefittet.

7.17 Vorbereitung und Durchführung der MST (Microscale Thermophoresis)

Labeling

Für die MST-Messungen wurde die MBP-TEV-DLK₁₄₀₋₅₁₇-Lösung „Fraktion g“ (im Folgenden: DLK-Lösung) mit dem Monolith NT™ Protein Labeling Kit RED-NHS nach Protokoll des Herstellers [512] fluoreszenzmarkiert: Die Aufschlammung (unbekannter Zusammensetzung) in der im Kit

enthaltenen Spin-Säule A wurde durch Umdrehen suspendiert, der Boden der Säule abgedreht und die Kappe der Säule entfernt. Die Säule wurde in ein 2 ml-Eppendorf-Reaktionsgefäß gesetzt und bei 1500 xg 1min. zentrifugiert (Zentrifuge EPPENDORF 5424R), um die überschüssige Flüssigkeit zu entfernen. Für den Pufferaustausch wurden 3 ml destilliertes Wasser zu dem im Kit enthaltenen Vial mit dem Labeling-Puffer-Salz gegeben. 300 µl dieses Labeling-Puffers wurden zur Spin-Säule A gegeben und die überschüssige Flüssigkeit wie zuvor beschrieben abzentrifugiert. Dieser Schritt des Waschens mit dem Labeling-Puffer wurde noch drei Mal wiederholt.

Dann wurden 100 µl der DLK-Lösung zur Säule gegeben, ohne das Resin in der Säule dabei zu berühren. Die Säule mit der Probe wurde in ein neues 2 ml-Eppendorf-Reaktionsgefäß gegeben und bei 1500 xg 2min. zentrifugiert.

30 µl DMSO wurden zum festen Fluoreszenzfarbstoff gegeben (Farbstoffkonzentration 435 µM). Der Farbstoff wurde gründlich gemischt durch Vortexen, bis alles gelöst war. Die Farbstoffkonzentration wurde mit dem Labeling-Puffer angepasst, sodass sie ein 2-3faches der DLK-Lösung betrug. Hierzu wurden 3,50 µl der Farbstofflösung und 96,50 µl des Labeling-Puffers eingesetzt. Die umgepufferte Proteinlösung und die vorbereitete Farbstofflösung wurden 1:1 gemischt (200 µl finales Volumen) und 30 min. bei Raumtemperatur unter Lichtausschluss inkubiert.

Nicht umgesetzt, freier Farbstoff wurde entfernt. Zur Vorbereitung dazu wurde zunächst die obere Kappe der Säule B wurde entfernt und die Säulenaufbewahrungslösung abgegossen. Die untere Kappe wurde entfernt und die Säule B mit dem im Kit enthaltenen Adapter in ein 15 ml Falcon-Zentrifugenröhrchen gesetzt. Die Säule B wurde drei Mal mit 3 ml des MST-optimierten Puffers (s. Tab. 7-19) gewaschen. Dies geschah durch einen einfachen Durchfluss durch Schwerkraft ohne Zentrifugieren. Die gesamte Farbstoff-DLK-Probe (200 µl) wurde nun auf die Säule B gegeben. Nach vollständigem Durchdringen der Säulenmatrix mit der Probe wurde mit 300 µl MST-optimiertem Puffer nachgespült. Die Säule wurde in ein neues 15 ml Falcon-Zentrifugenröhrchen zum Sammeln der aufgereinigten Lösung gesetzt. 600 µl des MST-optimierten Puffers wurden zur Säule gegeben und das Eluat aufgefangen. Hierbei wurden die ersten Mikroliter verworfen. Das aufgefangene Eluat (im Folgenden: DLK-Eluat) konnte nun für die Thermophorese eingesetzt werden.

Tab. 7-19 – Zusammensetzung des MST-optimierten Puffers

MST-optimierter Puffer		
Tris-HCl pH 7,4	50 mM	0,7880 g
NaCl	150 mM	0,8766 g
MgCl₂	10 mM	0,0952 g
Tween-20	0,05 %	0,05 g
dH₂O		ad 100 ml

MST-Messungen

Für die MST-Messungen wurde das DLK-Eluat mit dem MST-optimierten Puffer (Zusammensetzung s. Tab. 7-19) verdünnt. Alle Lösungen wurden in Vials vorbereitet und in Premiumkapillaren (Nanotemper, etwa 4 µl Volumen) durch Kapillarwirkung gefüllt. Die Messung erfolgte mit dem Nanotemper Monolith NT.115 über die Software NT CONTROL v2.1.33 und die Darstellung der Messergebnisse (MST-Kurven, Capillary-Scan) mit der Software NT ANALYSIS v1.5.41. Für die Berechnung von K_d -Werten wurde die Software MO.AFFINITY ANALYSIS v2.3 genutzt. Die Messungen erfolgten immer mit einer LED-Power von 50 % und einer MST-Power von 40 %.

Zur Erstellung einer vorbereitenden Verdünnungsreihe (s. Kap. 5.6.2, S. 240, Punkt (3.)) wurde eine Crizotinib-Lösung (10 mM in 100 % DMSO) 1:10 mit dem MST-optimierten Puffer verdünnt. Diese Lösung wurde dann schrittweise mit dem MST-optimierten Puffer 1:1 verdünnt. Mit diesen Crizotiniblösungen wurde die entsprechend zuvor optimierte DLK-Puffer-Mischung (6 µl DLK + 90 µl Puffer) 1:1 gemischt, sodass eine Crizotinibkonzentrationsreihe von 500 µM bis 3,91 µM in der Endlösung erstellt wurde.

Bei der Hit-Testung (s. Kap. 5.6.2, S. 240, Punkt (4.) a) wurden pro Testreihe jeweils 16 Kapillare gemessen, deren Proben wie folgt vorbereitet wurden: 13 Testlösungen und eine crizotinibhaltige Lösung enthalten 10 µl der 6,25 %igen DLK-Eluat-Puffer-Mischung, 9 µl Puffer und 1 µl einer 10 mM Test- bzw. Referenzsubstanzlösung in 100 % DMSO. Die Kontrollprobe enthält 10 µl der 6,25 %igen DLK-Eluat-Puffer-Mischung, 9 µl Puffer und 1 µl DMSO. Die Blindprobe enthält 19 µl Puffer und 1 µl DMSO.

Die Herstellung der Lösungen für die Verdünnungsreihe der 1. Messung (s. Kap. 5.6.2, S. 242, Punkt (4.) b) erfolgte nach dem Schema in Abb. 7-10: Zuerst wurden 18 µl des MST-optimierten

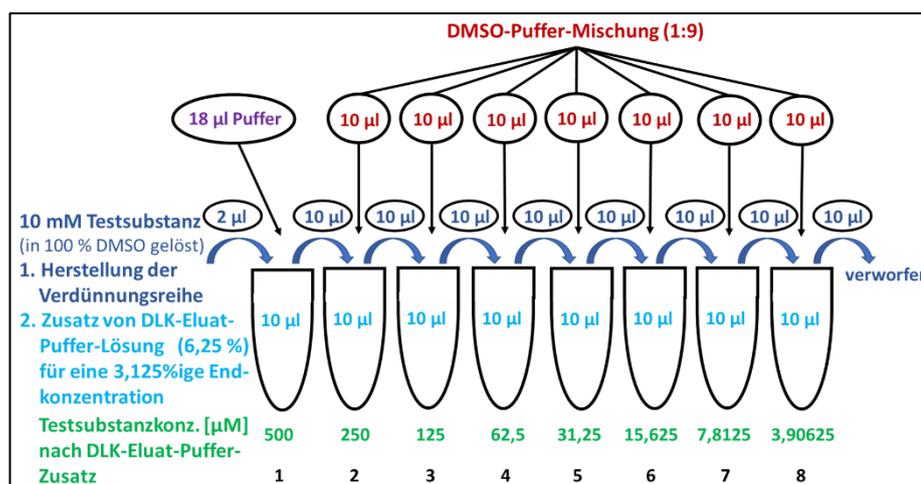


Abb. 7-10 – Vorbereitung der Lösungen zur Messung einer Verdünnungsreihe

Puffer und 2 µl der Testsubstanzlösung (10 mM in DMSO) im ersten Vial gemischt. In den übrigen 7 Vials wurden je 10 µl einer DMSO-Puffer-Mischung (1:9) vorgelegt. 10 µl der Lösung aus Vial 1

wurden in Vial 2 überführt und gemischt. Wiederum 10 µl dieses Vials wurden in Vial 3 übertragen etc. Am Ende wurden 10 µl aus Vial 8 entnommen und verworfen. Zu jedem Vial wurden dann 10 µl der vorbereiteten DLK-Eluat-Puffer-Mischung (6,25 %) gegeben. Vial 9 und Vial 10 (hier nicht abgebildet) enthielten eine Kontroll- und eine Blindlösung (wie zuvor, s. Punkt (4.) a, Hit-Testung).

Die Herstellung der Lösungen für die 2. und 3. Messung erfolgte nach dem gleichen Schema, allerdings mit jeweils der Hälfte an Volumina. Nur GNE-3511 wurde in allen drei Messungen 10fach verdünnter eingesetzt, d.h. ab einer Konzentration von 50 µM statt 500 µM getestet. Crizotinib wurde vollständig in drei aufeinanderfolgenden Messungen nach diesem Schema gemessen (und nicht nach Schema der 1. Messung), die dritte Messung erfolgte aber mit Standardkapillaren statt – wie sonst für alle Lösungen in dieser Arbeit – mit Premiumkapillaren. Wichtig: **E05** wurde versehentlich wie die anderen Substanzen mit einer Konzentrationsreihe ab 500 µM zur Messung in die Software eingetragen, obwohl die Konzentration tatsächlich abweicht (ab 421,07 µM, s. S. 232, Tab. 5-55).

7.18 Hochleistungs-Flüssigchromatografie (HPLC)

Für die HPLC-Analytik der Substanzen **A04**, **C07**, **D04**, **E05** und **E06** wurde die HPLC-Anlage AGILENT 1220 INFINITY LC der Firma AGILENT (Diodenarray-Detektor), Waldbronn, mit der Säule AGILENT POROSHELL 120 SB-C18 der Firma Agilent, Waldbronn, verwendet. Für die Eingabe der Parameter und zur Auswertung wurde die Software LC1 (AGILENT OPEN LAB CDS) der Firma AGILENT, Waldbronn, verwendet.

<p>Methode 1:</p> <p>Fließmittel: 0,1 % Phosphorsäure : Acetonitril (60:40)</p> <p>Flussrate: 1 ml/min, isokratisch</p> <p>Injektionsvolumen: 20 µl</p> <p>Methode 2:</p> <p>Fließmittel: 0,1 % Phosphorsäure : Acetonitril (80:20)</p> <p>Flussrate: 1 ml/min, isokratisch</p> <p>Injektionsvolumen: 10 µl</p>

Die Detektion erfolgte bei 220, 267 und 310 nm. Zudem wurden 3D-Chromatogramme (190-400 nm) aufgenommen.

Probelösung: 10 µl der Substanzlösungen (10 mM-Lösung in 100 % DMSO) wurden mit 90 µl einer Acetonitril-Wasser-Mischung (50:50) gemischt. Die Analytik der Substanzen **A04** und **C07** wurden mit beiden Methoden (1 + 2) durchgeführt, allerdings nur die Ergebnisse der Methode 2 für die Gehaltsbestimmung verwendet, da die Substanzpeaks bei Verwendung der Methode 1 schon

sehr früh auftraten (2,5-3 min.) und mit den Lösungsmittelpeaks zusammenfielen. Die Analytik von Crizotinib wurde nur mit Methode 2 durchgeführt, da in Vorversuchen mit einer vergleichbaren Säule (AGILENT POROSHELL 120 EC-C18) unter Verwendung der Methode 1 der Substanzpeak sehr früh (1-2 min.) auftrat. Die Analytik von **D04**, **E05** und **E06** wurde nur mit Methode 1 durchgeführt, da in Vorversuchen mit der o.g. EC-C18-Säule kein Substanzpeak mit Methode 2 detektiert wurde.

Blank: Eine Blank-Lösung (100 µl DMSO 100 % + 900 µl Acetonitril-Wasser-Mischung 50:50) wurde mit beiden Methoden analysiert.

Jede vorbereitete Lösung wurde zwei Mal injiziert und analysiert. Die Auswertung zur Gehaltsbestimmung erfolgte mit der 100 %-Methode anhand der Daten der 267 nm-Detektion, da die Absorption durch die Testsubstanzen bei dieser Wellenlänge im Verhältnis zu der Absorption des Lösungsmittels höher ist im Vergleich zur Absorption bei 220 nm (vgl. 3D-Chromatogramme, Anhang I). Mit der Wellenlänge 330 nm wiederum werden manche Verunreinigungen nicht erfasst, die bei 267 nm erfasst werden. Die frühen Peaks in den Testsubstanzen-Chromatogrammen, welche auch in den Blank-Chromatogrammen auftauchten, wurden nicht für die Reinheitsberechnung der Substanzen berücksichtigt.

8 Literaturverzeichnis

- [1] Hu, F. B., Satija, A., Manson, J. E., Curbing the Diabetes Pandemic: The Need for Global Policy Solutions. *JAMA* 2015, *313*, 2319–2320.
- [2] Saeedi, P., Petersohn, I., Salpea, P., Malanda, B., Karuranga, S., Unwin, N., Colagiuri, S., Guariguata, L., Motala, A. A., Ogurtsova, K., Shaw, J. E., Bright, D., Williams, R., Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: Results from the International Diabetes Federation Diabetes Atlas, 9th edition. *Diabetes Res. Clin. Pract.* 2019, *157*, 107843.
- [3] Raghavan, S., Vassy, J. L., Ho, Y.-L., Song, R. J., Gagnon, D. R., Cho, K., Wilson, P. W. F., Phillips, L. S., Diabetes Mellitus-Related All-Cause and Cardiovascular Mortality in a National Cohort of Adults. *J. Am. Heart Assoc.* 2019, *8*, e011295.
- [4] International Diabetes Federation, IDF Diabetes Atlas, 10. Edition, Brüssel, Belgien: International Diabetes Federation 2021.
- [5] Rangel, É. B., Rodrigues, C. O., Sá, J. R. de, Micro- and Macrovascular Complications in Diabetes Mellitus: Preclinical and Clinical Studies. *J. Diabetes Res.* 2019, 2161085.
- [6] Shah, B. R., Hux, J. E., Quantifying the risk of infectious diseases for people with diabetes. *Diabetes Care* 2003, *26*, 510–513.
- [7] Giovannucci, E., Harlan, D. M., Archer, M. C., Bergenstal, R. M., Gapstur, S. M., Habel, L. A., Pollak, M., Regensteiner, J. G., Yee, D., Diabetes and cancer: a consensus report. *Diabetes Care* 2010, *33*, 1674–1685.
- [8] World Health Organization, *Classification of diabetes mellitus*, <https://www.who.int/publications/i/item/classification-of-diabetes-mellitus> (accessed August 23, 2021).
- [9] Adolf, C., Freiboth, I., Seißler, J., Lechner, A., Aktuelle Konzepte in der Diagnostik und Therapie des Diabetes mellitus. *Der Radiologe* 2015, *55*, 295–298.
- [10] Kaiser, D., Oetjen, E., Something old, something new and something very old: drugs for treating type 2 diabetes. *Br. J. Pharmacol.* 2014, *171*, 2940–2950.
- [11] Stumvoll, M., Goldstein, B. J., van Haeften, T. W., Type 2 diabetes: principles of pathogenesis and therapy. *Lancet* 2005, *365*, 1333–1346.
- [12] Zaccardi, F., Webb, D. R., Yates, T., Davies, M. J., Pathophysiology of type 1 and type 2 diabetes mellitus: a 90-year perspective. *Postgrad. Med. J.* 2016, *92*, 63–69.

- [13] Virally, M., Blicklé, J.-F., Girard, J., Halimi, S., Simon, D., Guillausseau, P.-J., Type 2 diabetes mellitus: epidemiology, pathophysiology, unmet needs and therapeutical perspectives. *Diabetes Metab.* 2007, 33, 231–244.
- [14] Clodi, M., Abrahamian, H., Brath, H., Brix, J., Drexel, H., Fasching, P., Föger, B., Francesconi, C., Fröhlich-Reiterer, E., Harreiter, J., Hofer, S. E., Hoppichler, F., Huber, J., Kaser, S., Kautzky-Willer, A., Lechleitner, M., Ludvik, B., Luger, A., Mader, J. K., Paulweber, B., Pieber, T., Prager, R., Rami-Merhar, B., Resl, M., Riedl, M., Roden, M., Saely, C. H., Schelkshorn, C., Schernthanner, G., Sourij, H., Stechemesser, L., Stingl, H., Toplak, H., Wascher, T. C., Weitgasser, R., Winhofer-Stöckl, Y., Zlamal-Fortunat, S., Antihyperglykämische Therapie bei Diabetes mellitus Typ 2 (Update 2019). *Wiener klinische Wochenschrift* 2019, 131, 27–38.
- [15] Philippe, J., Raccach, D., Treating type 2 diabetes: how safe are current therapeutic agents? *Int. J. Clin. Pract.* 2009, 63, 321–332.
- [16] Ahrén, B., GLP-1 for type 2 diabetes. *Exp. Cell Res.* 2011, 317, 1239–1245.
- [17] Fonseca, V. A., Handelsman, Y., Staels, B., Colesevelam lowers glucose and lipid levels in type 2 diabetes: the clinical evidence. *Diabetes Obes. Metab.* 2010, 12, 384–392.
- [18] Kahn, S. E., Cooper, M. E., Del Prato, S., Pathophysiology and treatment of type 2 diabetes: perspectives on the past, present, and future. *Lancet* 2014, 383, 1068–1083.
- [19] Schmitz, O., Brock, B., Rungby, J., Amylin agonists: a novel approach in the treatment of diabetes. *Diabetes* 2004, 53 Suppl 3, S233-8.
- [20] Tsoutsouki, J., Wunna, W., Chowdhury, A., Chowdhury, T. A., Advances in the management of diabetes: therapies for type 2 diabetes. *Postgrad. Med. J.* 2020, 96, 610–618.
- [21] BfArM, *Pioglitazon - Europäische Arzneimittelagentur empfiehlt neue Kontraindikationen und Warnhinweise für pioglitazonhaltige Arzneimittel aufgrund eines leicht erhöhten Blasenkrebsrisikos*, <https://www.bfarm.de/SharedDocs/Risikoinformationen/Pharmakovigilanz/DE/RI/2011/RI-pioglitazon.html> (accessed March 21, 2022).
- [22] Garofalo, C., Borrelli, S., Liberti, M. E., Andreucci, M., Conte, G., Minutolo, R., Provenzano, M., Nicola, L. de, SGLT2 Inhibitors: Nephroprotective Efficacy and Side Effects. *Medicina* 2019, 55.
- [23] Bailey, C. J., Day, C., The future of new drugs for diabetes management. *Diabetes Res. Clin. Pract.* 2019, 155, 107785.
- [24] Perreault, L., Skyler, J. S., Rosenstock, J., Novel therapies with precision mechanisms for type 2 diabetes mellitus. *Nat. Rev. Endocrinol.* 2021, 17, 364–377.

- [25] Matschinsky, F. M., Assessing the potential of glucokinase activators in diabetes therapy. *Nat. Rev. Drug Discov.* 2009, *8*, 399–416.
- [26] Araki, T., Hirayama, M., Hiroi, S., Kaku, K., GPR40-induced insulin secretion by the novel agonist TAK-875: first clinical findings in patients with type 2 diabetes. *Diabetes Obes. Metab.* 2012, *14*, 271–278.
- [27] Krug, A. W., Vaddady, P., Railkar, R. A., Musser, B. J., Cote, J., Ederveen, A., Krefetz, D. G., DeNoia, E., Free, A. L., Morrow, L., Chakravarthy, M. V., Kauh, E., Tatosian, D. A., Kothare, P. A., Leveraging a Clinical Phase Ib Proof-of-Concept Study for the GPR40 Agonist MK-8666 in Patients With Type 2 Diabetes for Model-Informed Phase II Dose Selection. *Clin. Transl. Sci.* 2017, *10*, 404–411.
- [28] Frias, J. P., Nauck, M. A., Van, J., Kutner, M. E., Cui, X., Benson, C., Urva, S., Gimeno, R. E., Miličević, Z., Robins, D., Haupt, A., Efficacy and safety of LY3298176, a novel dual GIP and GLP-1 receptor agonist, in patients with type 2 diabetes: a randomised, placebo-controlled and active comparator-controlled phase 2 trial. *Lancet* 2018, *392*, 2180–2193.
- [29] Tillner, J., Posch, M. G., Wagner, F., Teichert, L., Hijazi, Y., Einig, C., Keil, S., Haack, T., Wagner, M., Bossart, M., Larsen, P. J., A novel dual glucagon-like peptide and glucagon receptor agonist SAR425899: Results of randomized, placebo-controlled first-in-human and first-in-patient trials. *Diabetes Obes. Metab.* 2019, *21*, 120–128.
- [30] OPKO Health, Inc., *A Phase 2, Double-blind Dose Escalation Regimen of Once-Weekly OPK-88003 in Subjects With Type 2 Diabetes: NCT03406377, DPO-203*, <https://clinicaltrials.gov/ct2/show/NCT03406377> (accessed March 22, 2022).
- [31] Vajda, E. G., Logan, D., Lassetter, K., Armas, D., Plotkin, D. J., Pipkin, J. D., Li, Y.-X., Zhou, R., Klein, D., Wei, X., Dilzer, S., Zhi, L., Marschke, K. B., Pharmacokinetics and pharmacodynamics of single and multiple doses of the glucagon receptor antagonist LGD-6972 in healthy subjects and subjects with type 2 diabetes mellitus. *Diabetes Obes. Metab.* 2017, *19*, 24–32.
- [32] Morgan, E. S., Tai, L.-J., Pham, N. C., Overman, J. K., Watts, L. M., Smith, A., Jung, S. W., Gajdošík, M., Krššák, M., Krebs, M., Geary, R. S., Baker, B. F., Bhanot, S., Antisense Inhibition of Glucagon Receptor by IONIS-GCGRx Improves Type 2 Diabetes Without Increase in Hepatic Glycogen Content in Patients With Type 2 Diabetes on Stable Metformin Therapy. *Diabetes Care* 2019, *42*, 585–593.
- [33] Henriksen, K., Broekhuizen, K., Boon, W. M. I. de, Karsdal, M. A., Bihlet, A. R., Christiansen, C., Dillingh, M. R., Kam, M. de, Kumar, R., Burggraaf, J., Kamerling, I. M. C., Safety, tolerability

- and pharmacokinetic characterisation of DACRA KBP-042 in healthy male subjects. *Br. J. Clin. Pharmacol.* 2021, *87*, 4786–4796.
- [34] Köhn, M., Turn and Face the Strange: A New View on Phosphatases. *ACS Cent. Sci.* 2020, *6*, 467–477.
- [35] Balakumar, P., Rose, M., Ganti, S. S., Krishan, P., Singh, M., PPAR dual agonists: are they opening Pandora's Box? *Pharmacol. Res.* 2007, *56*, 91–98.
- [36] Deeks, E. D., Chiglitazar: First Approval. *Drugs* 2022, *82*, 87–92.
- [37] Lee, M.-K., Kim, S. G., Watkins, E., Moon, M. K., Rhee, S. Y., Frias, J. P., Chung, C. H., Lee, S.-H., Block, B., Cha, B. S., Park, H. K., Kim, B. J., Greenway, F., A novel non-PPARgamma insulin sensitizer: MLR-1023 clinical proof-of-concept in type 2 diabetes mellitus. *J. Diabetes Complicat.* 2020, *34*, 107555.
- [38] Gaich, G., Chien, J. Y., Fu, H., Glass, L. C., Deeg, M. A., Holland, W. L., Kharitonov, A., Bumol, T., Schilske, H. K., Moller, D. E., The effects of LY2405319, an FGF21 analog, in obese human subjects with type 2 diabetes. *Cell Metab.* 2013, *18*, 333–340.
- [39] Pfizer, A Phase 1B, Randomized, Double-Blind, Placebo-Controlled Trial To Assess The Efficacy And Safety Of 4-Week Administration Of Multiple Oral Doses Of PF-04620110 In Type 2 Diabetes Mellitus Subjects With Insufficient Glycemic Control On Metformin: NCT01298518, B0961007, <https://clinicaltrials.gov/ct2/show/NCT01298518> (accessed March 22, 2022).
- [40] Alonso, R., Cuevas, A., Mata, P., Lomitapide: a review of its clinical use, efficacy, and tolerability. *Core Evid.* 2019, *14*, 19–30.
- [41] Finkel, R. S., Finanger, E., Vandenborne, K., Sweeney, H. L., Tennekoon, G., Shieh, P. B., Willcocks, R., Walter, G., Rooney, W. D., Forbes, S. C., Triplett, W. T., Yum, S. W., Mancini, M., MacDougall, J., Fretzen, A., Bista, P., Nichols, A., Donovan, J. M., Disease-modifying effects of edasalonexent, an NF- κ B inhibitor, in young boys with Duchenne muscular dystrophy: Results of the MoveDMD phase 2 and open label extension trial. *Neuromuscul. Disord.* 2021, *31*, 385–396.
- [42] Tregellas, J. R., Wylie, K. P., Alpha7 Nicotinic Receptors as Therapeutic Targets in Schizophrenia. *Nicotine Tob. Res.* 2019, *21*, 349–356.
- [43] Allas, S., Delale, T., Ngo, N., Julien, M., Sahakian, P., Ritter, J., Aribat, T., van der Lely, A. J., Safety, tolerability, pharmacokinetics and pharmacodynamics of AZP-531, a first-in-class analogue of unacylated ghrelin, in healthy and overweight/obese subjects and subjects with type 2 diabetes. *Diabetes Obes. Metab.* 2016, *18*, 868–874.

- [44] Clemmons, D. R., Miller, S., Mamputu, J.-C., Safety and metabolic effects of tesamorelin, a growth hormone-releasing factor analogue, in patients with type 2 diabetes: A randomized, placebo-controlled trial. *PLoS ONE* 2017, *12*, e0179538.
- [45] Dhillon, S., Tesamorelin: a review of its use in the management of HIV-associated lipodystrophy. *Drugs* 2011, *71*, 1071–1091.
- [46] Heymsfield, S. B., Coleman, L. A., Miller, R., Rooks, D. S., Laurent, D., Petricoul, O., Praestgaard, J., Swan, T., Wade, T., Perry, R. G., Goodpaster, B. H., Roubenoff, R., Effect of Bimagrumab vs Placebo on Body Fat Mass Among Adults With Type 2 Diabetes and Obesity: A Phase 2 Randomized Clinical Trial. *JAMA Netw. Open* 2021, *4*, e2033457.
- [47] Pirags, V., Lebovitz, H., Fouquieray, P., Imeglimin, a novel glimin oral antidiabetic, exhibits a good efficacy and safety profile in type 2 diabetic patients. *Diabetes Obes. Metab.* 2012, *14*, 852–858.
- [48] Fischer, E. H., Krebs, E. G., Conversion of phosphorylase b to phosphorylase a in muscle extracts. *J. Biol. Chem.* 1955, *216*, 121–132.
- [49] Sutherland, E. W., Wosilait, W. D., Inactivation and activation of liver phosphorylase. *Nature* 1955, *175*, 169–170.
- [50] Ebina, Y., Araki, E., Taira, M., Shimada, F., Mori, M., Craik, C. S., Siddle, K., Pierce, S. B., Roth, R. A., Rutter, W. J., Replacement of lysine residue 1030 in the putative ATP-binding region of the insulin receptor abolishes insulin- and antibody-stimulated glucose uptake and receptor kinase activity. *Proc. Natl. Acad. Sci. U.S.A.* 1987, *84*, 704–708.
- [51] Kaneto, H., Xu, G., Fujii, N., Kim, S., Bonner-Weir, S., Weir, G. C., Involvement of c-Jun N-terminal kinase in oxidative stress-mediated suppression of insulin gene expression. *J. Biol. Chem.* 2002, *277*, 30010–30018.
- [52] Aguirre, V., Uchida, T., Yenush, L., Davis, R., White, M. F., The c-Jun NH(2)-terminal kinase promotes insulin resistance during association with insulin receptor substrate-1 and phosphorylation of Ser(307). *J. Biol. Chem.* 2000, *275*, 9047–9054.
- [53] Hirosumi, J., Tuncman, G., Chang, L., Görgün, C. Z., Uysal, K. T., Maeda, K., Karin, M., Hottel, G. S., A central role for JNK in obesity and insulin resistance. *Nature* 2002, *420*, 333–336.
- [54] Hunter, T., Oncoprotein Networks. *Cell* 1997, *88*, 333–346.
- [55] Mackay, H. J., Twelves, C. J., Targeting the protein kinase C family: are we there yet? *Nat. Rev. Cancer* 2007, *7*, 554–562.

- [56] Brazil, D. P., Hemmings, B. A., Ten years of protein kinase B signalling: a hard Akt to follow. *Trends Biochem. Sci.* 2001, *26*, 657–664.
- [57] Elphick, L. M., Lee, S. E., Child, E. S., Prasad, A., Pignocchi, C., Thibaudeau, S., Anderson, A. A., Bonnac, L., Gouverneur, V., Mann, D. J., A Quantitative Comparison of Wild-Type and Gatekeeper Mutant Cdk2 for Chemical Genetic Studies with ATP Analogues. *ChemBioChem* 2009, *10*, 1519–1526.
- [58] Liu, Q., Sabnis, Y., Zhao, Z., Zhang, T., Buhrlage, S. J., Jones, L. H., Gray, N. S., Developing irreversible inhibitors of the protein kinase cysteinome. *Chem. Biol.* 2013, *20*, 146–159.
- [59] Rubin, G. M., Yandell, M. D., Wortman, J. R., Gabor Miklos, G. L., Nelson, C. R., Hariharan, I. K., Fortini, M. E., Li, P. W., Apweiler, R., Fleischmann, W., Cherry, J. M., Henikoff, S., Skupski, M. P., Misra, S., Ashburner, M., Birney, E., Boguski, M. S., Brody, T., Brokstein, P., Celniker, S. E., Chervitz, S. A., Coates, D., Cravchik, A., Gabrielian, A., Galle, R. F., Gelbart, W. M., George, R. A., Goldstein, L. S., Gong, F., Guan, P., Harris, N. L., Hay, B. A., Hoskins, R. A., Li, J., Li, Z., Hynes, R. O., Jones, S. J., Kuehl, P. M., Lemaître, B., Littleton, J. T., Morrison, D. K., Mungall, C., O'Farrell, P. H., Pickeral, O. K., Shue, C., Vosshall, L. B., Zhang, J., Zhao, Q., Zheng, X. H., Lewis, S., Comparative genomics of the eukaryotes. *Science* 2000, *287*, 2204–2215.
- [60] Cheng, H.-C., Qi, R. Z., Paudel, H., Zhu, H.-J., Regulation and function of protein kinases and phosphatases. *Enzyme Res.* 2011, *2011*, 794089.
- [61] Pinna, L. A., Ruzzene, M., How do protein kinases recognize their substrates? *Biochim. Biophys. Acta* 1996, *1314*, 191–225.
- [62] Heim, M. H., The Jak-STAT pathway: specific signal transduction from the cell membrane to the nucleus. *Eur. J. Clin. Invest.* 1996, *26*, 1–12.
- [63] Fey, D., Croucher, D. R., Kolch, W., Kholodenko, B. N., Crosstalk and signaling switches in mitogen-activated protein kinase cascades. *Front. Physiol.* 2012, *3*, 355.
- [64] Chico, L. K., van Eldik, L. J., Watterson, D. M., Targeting protein kinases in central nervous system disorders. *Nat. Rev. Drug Discov.* 2009, *8*, 892–909.
- [65] Arslan, M. A., Kutuk, O., Basaga, H., Protein kinases as drug targets in cancer. *Curr. Cancer Drug Targets* 2006, *6*, 623–634.
- [66] Felice, F. G. de, Ferreira, S. T., Inflammation, defective insulin signaling, and mitochondrial dysfunction as common molecular denominators connecting type 2 diabetes to Alzheimer disease. *Diabetes* 2014, *63*, 2262–2272.
- [67] Slawson, C., Copeland, R. J., Hart, G. W., O-GlcNAc signaling: a metabolic link between diabetes and cancer? *Trends Biochem. Sci.* 2010, *35*, 547–555.

- [68] Manning, G., Whyte, D. B., Martinez, R., Hunter, T., Sudarsanam, S., The protein kinase complement of the human genome. *Science* 2002, *298*, 1912–1934.
- [69] Hanks, S. K., Genomic analysis of the eukaryotic protein kinase superfamily: a perspective. *Genome Biol.* 2003, *4*, 111.
- [70] Patel, S., Harris, S. F., Gibbons, P., Deshmukh, G., Gustafson, A., Kellar, T., Lin, H., Liu, X., Liu, Y., Liu, Y., Ma, C., Scearce-Levie, K., Ghosh, A. S., Shin, Y. G., Solanoy, H., Wang, J., Wang, B., Yin, J., Siu, M., Lewcock, J. W., Scaffold-Hopping and Structure-Based Discovery of Potent, Selective, And Brain Penetrant N-(1H-Pyrazol-3-yl)pyridin-2-amine Inhibitors of Dual Leucine Zipper Kinase (DLK, MAP3K12). *J. Med. Chem.* 2015, *58*, 8182–8199.
- [71] Huse, M., Kuriyan, J., The Conformational Plasticity of Protein Kinases. *Cell* 2002, *109*, 275–282.
- [72] Tong, M., Seeliger, M. A., Targeting conformational plasticity of protein kinases. *ACS Chem. Biol.* 2015, *10*, 190–200.
- [73] Liu, Y., Gray, N. S., Rational design of inhibitors that bind to inactive kinase conformations. *Nat. Chem. Biol.* 2006, *2*, 358.
- [74] Adams, J. A., Kinetic and Catalytic Mechanisms of Protein Kinases. *Chem. Rev.* 2001, *101*, 2271–2290.
- [75] Matte, A., Tari, L. W., Delbaere, L. T. J., How do kinases transfer phosphoryl groups? *Structure* 1998, *6*, 413–419.
- [76] Backes, A., Zech, B., Felber, B., Klebl, B., Müller, G., Small-molecule inhibitors binding to protein kinases. Part I: exceptions from the traditional pharmacophore approach of type I inhibition. *Expert Opin. Drug Discov.* 2008, *3*, 1409–1425.
- [77] Gujral, M., Trafny, E. A., Xuong, N. H., Adams, J. A., Eyck, L. F. ten, Taylor, S. S., Sowadski, J. M., cAMP-dependent protein kinase: crystallographic insights into substrate recognition and phosphotransfer. *Protein Sci.* 1994, *3*, 176–187.
- [78] Roskoski, R., Classification of small molecule protein kinase inhibitors based upon the structures of their drug-enzyme complexes. *Pharmacol. Res.* 2016, *103*, 26–48.
- [79] Wu, P., Nielsen, T. E., Clausen, M. H., FDA-approved small-molecule kinase inhibitors. *Trends Pharmacol. Sci.* 2015, *36*, 422–439.
- [80] Pargellis, C., Tong, L., Churchill, L., Cirillo, P. F., Gilmore, T., Graham, A. G., Grob, P. M., Hickey, E. R., Moss, N., Pav, S., Regan, J., Inhibition of p38 MAP kinase by utilizing a novel allosteric binding site. *Nat. Struct. Biol.* 2002, *9*, 268–272.

- [81] Vijayan, R. S. K., He, P., Modi, V., Duong-Ly, K. C., Ma, H., Peterson, J. R., Dunbrack, R. L., Levy, R. M., Conformational analysis of the DFG-out kinase motif and biochemical profiling of structurally validated type II inhibitors. *J. Med. Chem.* 2015, *58*, 466–479.
- [82] Angell, R. M., Angell, T. D., Bamborough, P., Bamford, M. J., Chung, C., Cockerill, S. G., Flack, S. S., Jones, K. L., Laine, D. I., Longstaff, T., Ludbrook, S., Pearson, R., Smith, K. J., Smees, P. A., Somers, D. O., Walker, A. L., Biphenyl amide p38 kinase inhibitors 4: DFG-in and DFG-out binding modes. *Bioorg Med Chem Lett* 2008, *18*, 4433–4437.
- [83] Davis, M. I., Hunt, J. P., Herrgard, S., Ciceri, P., Wodicka, L. M., Pallares, G., Hocker, M., Treiber, D. K., Zarrinkar, P. P., Comprehensive analysis of kinase inhibitor selectivity. *Nat. Biotechnol.* 2011, *29*, 1046–1051.
- [84] Roskoski, R., Properties of FDA-approved small molecule protein kinase inhibitors: A 2021 update. *Pharmacol. Res.* 2021, *165*, 105463.
- [85] Myers, S. H., Brunton, V. G., Unciti-Broceta, A., AXL Inhibitors in Cancer: A Medicinal Chemistry Perspective. *J. Med. Chem.* 2016, *59*, 3593–3608.
- [86] Savage, D. G., Antman, K. H., Imatinib mesylate - a new oral targeted therapy. *N. Engl. J. Med.* 2002, *346*, 683–693.
- [87] Carles, F., Bourg, S., Meyer, C., Bonnet, P., PKIDB: A Curated, Annotated and Updated Database of Protein Kinase Inhibitors in Clinical Trials. *Molecules* 2018, *23*, 908.
- [88] Bournez, C., Carles, F., Peyrat, G., Aci-Sèche, S., Bourg, S., Meyer, C., Bonnet, P., Comparative Assessment of Protein Kinase Inhibitors in Public Databases and in PKIDB. *Molecules* 2020, *25*.
- [89] Blair, H. A., Belumosudil: First Approval. *Drugs* 2021.
- [90] Schultze, S. M., Hemmings, B. A., Niessen, M., Tschopp, O., PI3K/AKT, MAPK and AMPK signalling: protein kinases in glucose homeostasis. *Expert Rev. Mol. Med.* 2012, *14*, e1.
- [91] Bullock, W. H., Magnuson, S. R., Choi, S., Gunn, D. E., Rudolph, J., Prospects for kinase activity modulators in the treatment of diabetes and diabetic complications. *Curr. Top. Med. Chem.* 2002, *2*, 915–938.
- [92] Nandipati, K. C., Subramanian, S., Agrawal, D. K., Protein kinases: mechanisms and downstream targets in inflammation-mediated obesity and insulin resistance. *Mol. Cell Biochem.* 2017, *426*, 27–45.
- [93] Luca, C. de, Olefsky, J. M., Inflammation and insulin resistance. *FEBS Lett.* 2008, *582*, 97–105.

- [94] Liu, K., Xu, L., Szalkowski, D., Li, Z., Ding, V., Kwei, G., Huskey, S., Moller, D. E., Heck, J. V., Zhang, B. B., Jones, A. B., Discovery of a potent, highly selective, and orally efficacious small-molecule activator of the insulin receptor. *J. Med. Chem.* 2000, *43*, 3487–3494.
- [95] Huang, X.-F., Chen, J.-Z., Obesity, the PI3K/Akt signal pathway and colon cancer. *Obes. Rev.* 2009, *10*, 610–616.
- [96] Lawlor, M. A., Alessi, D. R., PKB/Akt: a key mediator of cell proliferation, survival and insulin responses? *J. Cell Sci.* 2001, *114*, 2903–2910.
- [97] Duronio, V., The life of a cell: apoptosis regulation by the PI3K/PKB pathway. *Biochem. J.* 2008, *415*, 333–344.
- [98] Nikoulina, S. E., Ciaraldi, T. P., Mudaliar, S., Mohideen, P., Carter, L., Henry, R. R., Potential role of glycogen synthase kinase-3 in skeletal muscle insulin resistance of type 2 diabetes. *Diabetes* 2000, *49*, 263–271.
- [99] Cross, D. A., Alessi, D. R., Cohen, P., Andjelkovich, M., Hemmings, B. A., Inhibition of glycogen synthase kinase-3 by insulin mediated by protein kinase B. *Nature* 1995, *378*, 785–789.
- [100] Eldar-Finkelman, H., Krebs, E. G., Phosphorylation of insulin receptor substrate 1 by glycogen synthase kinase 3 impairs insulin action. *Proc. Natl. Acad. Sci. U.S.A.* 1997, *94*, 9660–9664.
- [101] AMO Pharma Limited, *A Randomized, Double-Blind Study to Evaluate the Efficacy and Safety of Tideglusib Versus Placebo for the Treatment of Children and Adolescents With Congenital Myotonic Dystrophy (REACH CDM): NCT03692312, AMO-02-MD-2-003*, <https://clinicaltrials.gov/ct2/show/NCT03692312> (accessed September 24, 2021).
- [102] Lovestone, S., Boada, M., Dubois, B., Hüll, M., Rinne, J. O., Huppertz, H.-J., Calero, M., Andrés, M. V., Gómez-Carrillo, B., León, T., Del Ser, T., A phase II trial of tideglusib in Alzheimer's disease. *J. Alzheimer's Dis.* 2015, *45*, 75–88.
- [103] Mackenzie, R. W., Elliott, B. T., Akt/PKB activation and insulin signaling: a novel insulin signaling pathway in the treatment of type 2 diabetes. *Diabet. Metab. Synd. Obes.* 2014, *7*, 55–64.
- [104] Schmitz-Peiffer, C., Biden, T. J., Protein kinase C function in muscle, liver, and beta-cells and its therapeutic implications for type 2 diabetes. *Diabetes* 2008, *57*, 1774–1783.
- [105] Wellen, K. E., Hotamisligil, G. S., Inflammation, stress, and diabetes. *J. Clin. Investig.* 2005, *115*, 1111–1119.

- [106] Gao, Z., Hwang, D., Bataille, F., Lefevre, M., York, D., Quon, M. J., Ye, J., Serine phosphorylation of insulin receptor substrate 1 by inhibitor kappa B kinase complex. *J. Biol. Chem.* 2002, 277, 48115–48121.
- [107] Fea, K. de, Roth, R. A., Modulation of insulin receptor substrate-1 tyrosine phosphorylation and function by mitogen-activated protein kinase. *J. Biol. Chem.* 1997, 272, 31400–31406.
- [108] Hemi, R., Yochananov, Y., Barhod, E., Kasher-Meron, M., Karasik, A., Tirosh, A., Kanety, H., p38 mitogen-activated protein kinase-dependent transactivation of ErbB receptor family: a novel common mechanism for stress-induced IRS-1 serine phosphorylation and insulin resistance. *Diabetes* 2011, 60, 1134–1145.
- [109] Li, G., Barrett, E. J., Barrett, M. O., Cao, W., Liu, Z., Tumor necrosis factor-alpha induces insulin resistance in endothelial cells via a p38 mitogen-activated protein kinase-dependent pathway. *Endocrinology* 2007, 148, 3356–3363.
- [110] Holzman, L. B., Merritt, S. E., Fan, G., Identification, molecular cloning, and characterization of dual leucine zipper bearing kinase. A novel serine/threonine protein kinase that defines a second subfamily of mixed lineage kinases. *J. Biol. Chem.* 1994, 269, 30808–30817.
- [111] Gallo, K. A., Johnson, G. L., Mixed-lineage kinase control of JNK and p38 MAPK pathways. *Nat. Rev. Mol. Cell Biol.* 2002, 3, 663–672.
- [112] Yu, Y., Coiled-coils: stability, specificity, and drug delivery potential. *Adv. Drug Deliv. Rev.* 2002, 54, 1113–1129.
- [113] Vidal, S., Khush, R. S., Leulier, F., Tzou, P., Nakamura, M., Lemaitre, B., Mutations in the *Drosophila* dTAK1 gene reveal a conserved function for MAPKKKs in the control of rel/NFκB-dependent innate immune responses. *Genes Dev.* 2001, 15, 1900–1912.
- [114] Wang, L. H., Besirli, C. G., Johnson, E. M., Mixed-lineage kinases: a target for the prevention of neurodegeneration. *Annu. Rev. Pharmacol. Toxicol.* 2004, 44, 451–474.
- [115] Qi, M., Elion, E. A., MAP kinase pathways. *J. Cell. Sci.* 2005, 118, 3569–3572.
- [116] Kim, E. K., Choi, E.-J., Pathological roles of MAPK signaling pathways in human diseases. *Biochim. Biophys. Acta* 2010, 1802, 396–405.
- [117] Hao, Y., Frey, E., Yoon, C., Wong, H., Nestorovski, D., Holzman, L. B., Giger, R. J., DiAntonio, A., Collins, C., An evolutionarily conserved mechanism for cAMP elicited axonal regeneration involves direct activation of the dual leucine zipper kinase DLK. *eLife* 2016, e14048.
- [118] Huntwork-Rodriguez, S., Wang, B., Watkins, T., Ghosh, A. S., Pozniak, C. D., Bustos, D., Newton, K., Kirkpatrick, D. S., Lewcock, J. W., JNK-mediated phosphorylation of DLK suppresses its ubiquitination to promote neuronal apoptosis. *J. Cell Biol.* 2013, 202, 747–763.

- [119] Sengupta Ghosh, A., Wang, B., Pozniak, C. D., Chen, M., Watts, R. J., Lewcock, J. W., DLK induces developmental neuronal degeneration via selective regulation of proapoptotic JNK activity. *J. Cell. Biol.* 2011, *194*, 751–764.
- [120] Geden, M. J., Deshmukh, M., Axon degeneration: context defines distinct pathways. *Curr. Opin. Neurobiol.* 2016, *39*, 108–115.
- [121] Wallbach, M., Duque Escobar, J., Babaeikelishomi, R., Stahnke, M.-J., Blume, R., Schröder, S., Kruegel, J., Maedler, K., Kluth, O., Kehlenbach, R. H., Miosge, N., Oetjen, E., Distinct functions of the dual leucine zipper kinase depending on its subcellular localization. *Cell. Signal.* 2016, *28*, 272–283.
- [122] Wu, C.-C., Wu, H.-J., Wang, C.-H., Lin, C.-H., Hsu, S.-C., Chen, Y.-R., Hsiao, M., Schuyler, S. C., Lu, F. L., Ma, N., Lu, J., Akt suppresses DLK for maintaining self-renewal of mouse embryonic stem cells. *Cell cycle* 2015, *14*, 1207–1217.
- [123] Morrison, D. K., MAP kinase pathways. *Cold Spring Harb. Perspect. Biol.* 2012, *4*.
- [124] Miller, B. R., Press, C., Daniels, R. W., Sasaki, Y., Milbrandt, J., DiAntonio, A., A dual leucine kinase-dependent axon self-destruction program promotes Wallerian degeneration. *Nat. Neurosci.* 2009, *12*, 387–389.
- [125] Ferraris, D., Yang, Z., Welsbie, D., Dual leucine zipper kinase as a therapeutic target for neurodegenerative conditions. *Future Med. Chem.* 2013, *5*, 1923–1934.
- [126] Couture, J. P., Blouin, R., DLK (Dual Leucine Zipper-Bearing Kinase). In: Choi S. (Hg.) *Encyclopedia of Signaling Molecules*. Springer NY, USA. 2012, 522–526.
- [127] Collins, C. A., Wairkar, Y. P., Johnson, S. L., DiAntonio, A., Highwire restrains synaptic growth by attenuating a MAP kinase signal. *Neuron* 2006, *51*, 57–69.
- [128] Nakata, K., Abrams, B., Grill, B., Goncharov, A., Huang, X., Chisholm, A. D., Jin, Y., Regulation of a DLK-1 and p38 MAP kinase pathway by the ubiquitin ligase RPM-1 is required for presynaptic development. *Cell* 2005, *120*, 407–420.
- [129] Mata, M., Merritt, S. E., Fan, G., Yu, G. G., Holzman, L. B., Characterization of dual leucine zipper-bearing kinase, a mixed lineage kinase present in synaptic terminals whose phosphorylation state is regulated by membrane depolarization via calcineurin. *J. Biol. Chem.* 1996, *271*, 16888–16896.
- [130] Hirai, S., Kawaguchi, A., Suenaga, J., Ono, M., Cui, D. F., Ohno, S., Expression of MUK/DLK/ZPK, an activator of the JNK pathway, in the nervous systems of the developing mouse embryo. *Gene Expr. Patterns* 2005, *5*, 517–523.

- [131] Couture, J.-P., Daviau, A., Fradette, J., Blouin, R., The mixed-lineage kinase DLK is a key regulator of 3T3-L1 adipocyte differentiation. *PLoS ONE* 2009, 4, e4743.
- [132] Oetjen, E., Lechleiter, A., Blume, R., Nihalani, D., Holzman, L., Knepel, W., Inhibition of membrane depolarisation-induced transcriptional activity of cyclic AMP response element binding protein (CREB) by the dual-leucine-zipper-bearing kinase in a pancreatic islet beta cell line. *Diabetologia* 2006, 49, 332–342.
- [133] Robitaille, H., Simard-Bisson, C., Larouche, D., Tanguay, R. M., Blouin, R., Germain, L., The small heat-shock protein Hsp27 undergoes ERK-dependent phosphorylation and redistribution to the cytoskeleton in response to dual leucine zipper-bearing kinase expression. *J. Invest. Dermatol.* 2010, 130, 74–85.
- [134] Nihalani, D., Meyer, D., Pajni, S., Holzman, L. B., Mixed lineage kinase-dependent JNK activation is governed by interactions of scaffold protein JIP with MAPK module components. *EMBO J.* 2001, 20, 3447–3458.
- [135] Nihalani, D., Wong, H. N., Holzman, L. B., Recruitment of JNK to JIP1 and JNK-dependent JIP1 phosphorylation regulates JNK module dynamics and activation. *J. Biol. Chem.* 2003, 278, 28694–28702.
- [136] Nihalani, D., Wong, H., Verma, R., Holzman, L. B., Src family kinases directly regulate JIP1 module dynamics and activation. *Mol. Cell. Biol.* 2007, 27, 2431–2441.
- [137] Nihalani, D., Merritt, S., Holzman, L. B., Identification of structural and functional domains in mixed lineage kinase dual leucine zipper-bearing kinase required for complex formation and stress-activated protein kinase activation. *J. Biol. Chem.* 2000, 275, 7273–7279.
- [138] Hirai, S., Cui, D. F., Miyata, T., Ogawa, M., Kiyonari, H., Suda, Y., Aizawa, S., Banba, Y., Ohno, S., The c-Jun N-terminal kinase activator dual leucine zipper kinase regulates axon growth and neuronal migration in the developing cerebral cortex. *J. Neurosci.* 2006, 26, 11992–12002.
- [139] Pozniak, C. D., Sengupta Ghosh, A., Gogineni, A., Hanson, J. E., Lee, S.-H., Larson, J. L., Solano, H., Bustos, D., Li, H., Ngu, H., Jubb, A. M., Ayalon, G., Wu, J., Scarce-Levie, K., Zhou, Q., Weimer, R. M., Kirkpatrick, D. S., Lewcock, J. W., Dual leucine zipper kinase is required for excitotoxicity-induced neuronal degeneration. *J. Exp. Med.* 2013.
- [140] Watkins, T. A., Wang, B., Huntwork-Rodriguez, S., Yang, J., Jiang, Z., Eastham-Anderson, J., Modrusan, Z., Kaminker, J. S., Tessier-Lavigne, M., Lewcock, J. W., DLK initiates a transcriptional program that couples apoptotic and regenerative responses to axonal injury. *Proc. Natl. Acad. Sci. U.S.A.* 2013, 110, 4039–4044.

- [141] Le Pichon, C. E., Meilandt, W. J., Dominguez, S., Solanoy, H., Lin, H., Ngu, H., Gogineni, A., Sengupta Ghosh, A., Jiang, Z., Lee, S.-H., Maloney, J., Gandham, V. D., Pozniak, C. D., Wang, B., Lee, S., Siu, M., Patel, S., Modrusan, Z., Liu, X., Rudhard, Y., Baca, M., Gustafson, A., Kaminer, J., Carano, R. A. D., Huang, E. J., Foreman, O., Weimer, R., Scearce-Levie, K., Lewcock, J. W., Loss of dual leucine zipper kinase signaling is protective in animal models of neurodegenerative disease. *Sci. Transl. Med.* 2017, *9*.
- [142] Siu, M., Sengupta Ghosh, A., Lewcock, J. W., Dual Leucine Zipper Kinase Inhibitors for the Treatment of Neurodegeneration. *J. Med. Chem.* 2018, *61*, 8078–8087.
- [143] Welsbie, D. S., Yang, Z., Ge, Y., Mitchell, K. L., Zhou, X., Martin, S. E., Berlinicke, C. A., Hackler, L., Fuller, J., Fu, J., Cao, L., Han, B., Auld, D., Xue, T., Hirai, S., Germain, L., Simard-Bisson, C., Blouin, R., Nguyen, J. V., Davis, C. O., Enke, R. A., Boye, S. L., Merbs, S. L., Marsh-Armstrong, N., Hauswirth, W. W., DiAntonio, A., Nickells, R. W., Inglese, J., Hanes, J., Yau, K.-W., Quigley, H. A., Zack, D. J., Functional genomic screening identifies dual leucine zipper kinase as a key mediator of retinal ganglion cell death. *Proc. Natl. Acad. Sci. U.S.A.* 2013, *110*, 4045–4050.
- [144] Oetjen, E., Regulation of Beta-Cell Function and Mass by the Dual Leucine Zipper Kinase. *Arch. Pharm.* 2016, *349*, 410–413.
- [145] Altarejos, J. Y., Montminy, M., CREB and the CRTC co-activators: sensors for hormonal and metabolic signals. *Nat. Rev. Mol. Cell Biol.* 2011, *12*, 141–151.
- [146] Phu, D. T., Wallbach, M., Depatie, C., Fu, A., Screatton, R. A., Oetjen, E., Regulation of the CREB coactivator TORC by the dual leucine zipper kinase at different levels. *Cell. Signal.* 2011, *23*, 344–353.
- [147] Jhala, U. S., Canettieri, G., Screatton, R. A., Kulkarni, R. N., Krajewski, S., Reed, J., Walker, J., Lin, X., White, M., Montminy, M., cAMP promotes pancreatic beta-cell survival via CREB-mediated induction of IRS2. *Genes Dev.* 2003, *17*, 1575–1580.
- [148] Jambal, P., Masterson, S., Nesterova, A., Bouchard, R., Bergman, B., Hutton, J. C., Boxer, L. M., Reusch, J. E.-B., Pugazhenti, S., Cytokine-mediated down-regulation of the transcription factor cAMP-response element-binding protein in pancreatic beta-cells. *J. Biol. Chem.* 2003, *278*, 23055–23065.
- [149] Saldeen, J., Cytokines induce both necrosis and apoptosis via a common Bcl-2-inhibitable pathway in rat insulin-producing cells. *Endocrinology* 2000, *141*, 2003–2010.
- [150] Stahnke, M.-J., Dickel, C., Schröder, S., Kaiser, D., Blume, R., Stein, R., Pouponnot, C., Oetjen, E., Inhibition of human insulin gene transcription and MafA transcriptional activity by the dual leucine zipper kinase. *Cellular Signalling* 2014, *26*, 1792–1799.

- [151] Wang, H., Brun, T., Kataoka, K., Sharma, A. J., Wollheim, C. B., MAFA controls genes implicated in insulin biosynthesis and secretion. *Diabetologia* 2007, 50, 348–358.
- [152] Oetjen, E., Lemcke, T., Dual leucine zipper kinase (MAP3K12) modulators: a patent review (2010–2015). *Expert Opin. Ther. Pat.* 2016, 26, 607–616.
- [153] Yin, C., Huang, G.-F., Sun, X.-C., Guo, Z., Zhang, J. H., DLK silencing attenuated neuron apoptosis through JIP3/MA2K7/JNK pathway in early brain injury after SAH in rats. *Neurobiol. Dis.* 2017, 103, 133–143.
- [154] Holland, S. M., Collura, K. M., Ketschek, A., Noma, K., Ferguson, T. A., Jin, Y., Gallo, G., Thomas, G. M., Palmitoylation controls DLK localization, interactions and activity to ensure effective axonal injury signaling. *Proc. Natl. Acad. Sci. U.S.A.* 2016, 113, 763–768.
- [155] Maroney, A. C., Glicksman, M. A., Basma, A. N., Walton, K. M., Knight, E., Murphy, C. A., Bartlett, B. A., Finn, J. P., Angeles, T., Matsuda, Y., Neff, N. T., Dionne, C. A., Motoneuron Apoptosis Is Blocked by CEP-1347 (KT 7515), a Novel Inhibitor of the JNK Signaling Pathway. *J. Neurosci.* 1998, 18, 104–111.
- [156] Kaneko, M., Saito, Y., Saito, H., Matsumoto, T., Matsuda, Y., Vaught, J. L., Dionne, C. A., Angeles, T. S., Glicksman, M. A., Neff, N. T., Rotella, D. P., Kauer, J. C., Mallamo, J. P., Hudkins, R. L., Murakata, C., Neurotrophic 3,9-bis(alkylthio)methyl-and-bis(alkoxymethyl)-K-252a derivatives. *J. Med. Chem.* 1997, 40, 1863–1869.
- [157] Maroney, A. C., Finn, J. P., Connors, T. J., Durkin, J. T., Angeles, T., Gessner, G., Xu, Z., Meyer, S. L., Savage, M. J., Greene, L. A., Scott, R. W., Vaught, J. L., Cep-1347 (KT7515), a semisynthetic inhibitor of the mixed lineage kinase family. *J. Biol. Chem.* 2001, 276, 25302–25308.
- [158] Kase, H., Iwahashi, K., Matsuda, Y., K-252a, a potent inhibitor of protein kinase C from microbial origin. *J. Antibiot.* 1986, 39, 1059–1065.
- [159] Nakano, H., Omura, S., Chemical biology of natural indolocarbazole products: 30 years since the discovery of staurosporine. *J. Antibiot.* 2009, 62, 17–26.
- [160] Kase, H., Iwahashi, K., Nakanishi, S., Matsuda, Y., Yamada, K., Takahashi, M., Murakata, C., Sato, A., Kaneko, M., K-252 compounds, novel and potent inhibitors of protein kinase C and cyclic nucleotide-dependent protein kinases. *Biochem. Biophys. Res. Comm.* 1987, 142, 436–440.
- [161] Knüsel, B., Hefti, F., K-252 compounds: modulators of neurotrophin signal transduction. *J. Neurochem.* 1992, 59, 1987–1996.
- [162] Ruggeri, B., Singh, J., Gingrich, D., Angeles, T., Albom, M., Yang, S., Chang, H., Robinson, C., Hunter, K., Dobrzanski, P., Jones-Bolin, S., Pritchard, S., Aimone, L., Klein-Szanto, A., Herbert,

- J.-M., Bono, F., Schaeffer, P., Casellas, P., Bourie, B., Pili, R., Isaacs, J., Ator, M., Hudkins, R., Vaught, J., Mallamo, J., Dionne, C., CEP-7055: a novel, orally active pan inhibitor of vascular endothelial growth factor receptor tyrosine kinases with potent antiangiogenic activity and antitumor efficacy in preclinical models. *Cancer Res.* 2003, *63*, 5978–5991.
- [163] Hudkins, R. L., Diebold, J. L., Tao, M., Josef, K. A., Park, C. H., Angeles, T. S., Aimone, L. D., Husten, J., Ator, M. A., Meyer, S. L., Holskin, B. P., Durkin, J. T., Fedorov, A. A., Fedorov, E. V., Almo, S. C., Mathiasen, J. R., Bozyczko-Coyne, D., Saporito, M. S., Scott, R. W., Mallamo, J. P., Mixed-Lineage Kinase 1 and Mixed-Lineage Kinase 3 Subtype-Selective Dihydronaphthyl[3,4-a]pyrrolo[3,4-c]carbazole-5-ones: Optimization, Mixed-Lineage Kinase 1 Crystallography, and Oral in Vivo Activity in 1-Methyl-4-phenyltetrahydropyridine Models. *J. Med. Chem.* 2008, *51*, 5680–5689.
- [164] Karaman, M. W., Herrgard, S., Treiber, D. K., Gallant, P., Atteridge, C. E., Campbell, B. T., Chan, K. W., Ciceri, P., Davis, M. I., Edeen, P. T., Faraoni, R., Floyd, M., Hunt, J. P., Lockhart, D. J., Milanov, Z. V., Morrison, M. J., Pallares, G., Patel, H. K., Pritchard, S., Wodicka, L. M., Zarrinkar, P. P., A quantitative analysis of kinase inhibitor selectivity. *Nat. Biotechnol.* 2008, *26*, 127–132.
- [165] Tao, M., Park, C. H., Josef, K., Hudkins, R. L., Regioselective synthesis of 2-methyl-2,5,6,11,12,13-hexahydro 4H indazolo[5,4-a]pyrrolo[3,4-c]carbazole-4-ones. *J. Heterocyclic Chem.* 2009, *46*, 1185–1189.
- [166] Patel, S., Cohen, F., Dean, B. J., La Torre, K. de, Deshmukh, G., Estrada, A. A., Ghosh, A. S., Gibbons, P., Gustafson, A., Huestis, M. P., Le Pichon, C. E., Lin, H., Liu, W., Liu, X., Liu, Y., Ly, C. Q., Lyssikatos, J. P., Ma, C., Scarce-Levie, K., Shin, Y. G., Solanoy, H., Stark, K. L., Wang, J., Wang, B., Zhao, X., Lewcock, J. W., Siu, M., Discovery of Dual Leucine Zipper Kinase (DLK, MAP3K12) Inhibitors with Activity in Neurodegeneration Models. *J. Med. Chem.* 2014, *58*, 401–418.
- [167] Patel, S., Meilandt, W. J., Erickson, R. I., Chen, J., Deshmukh, G., Estrada, A. A., Fuji, R. N., Gibbons, P., Gustafson, A., Harris, S. F., Imperio, J., Liu, W., Liu, X., Liu, Y., Lyssikatos, J. P., Ma, C., Yin, J., Lewcock, J. W., Siu, M., Selective Inhibitors of Dual Leucine Zipper Kinase (DLK, MAP3K12) with Activity in a Model of Alzheimer's Disease. *J. Med. Chem.* 2017, *60*, 8083–8102.
- [168] Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., Bourne, P. E., The Protein Data Bank. *Nucleic Acids Res.* 2000, *28*, 235–242.
- [169] RCSB Protein Data Bank, <http://www.rcsb.org/> (accessed October 28, 2016).

- [170] Cohen, F., Huestis, M., Ly, C., Patel, S., Siu, M., Zhao, X., *Substituted dipyridylamines and uses thereof* 2013.
- [171] Estrada, A., Liu, W., Patel, S., Siu, M., *3-Substituted pyrazoles and use as DLK inhibitors* 2014.
- [172] Chen, K., Dong, L., Estrada, A., Gibbons, P., Huestis, M., Kellar, T., Liu, W., Lyssikatos, J. P., Ma, C., Olivero, A., Patel, S., Shore, D., Siu, M., *Biheteroaryl compounds and uses thereof* 2014.
- [173] Cohen, F., Patel, S., *C-Linked heterocycloalkyl substituted pyrimidines and their uses* 2014.
- [174] Lyssikatos, J. P., Liu, W., Siu, M., Estrada, A., Patel, S., Liang, G., Huestis, M., Chen, K., *Pyrazole derivatives and uses thereof as inhibitors of DLK* 2014.
- [175] Estrada, A., Huestis, M., Kellar, T., Patel, S., Shore, D., Siu, M., *Tricyclic DLK inhibitors and uses thereof* 2016.
- [176] Soth, M. J., Jones, P., Ray, J., Liu, G., Le, K., Cross, J., *Inhibitors of dual leucine zipper (DLK) kinase for the treatment of disease* 2017.
- [177] Soth, M. J., Liu, G., Le, K., Cross, J., Jones, P., *Bicyclo[1.1.1]pentane Inhibitors of Dual Leucine Zipper (DLK) Kinase for the Treatment of Disease* 2017.
- [178] Genentech, Inc., *A Phase I, Double-Blind, Randomized, Placebo-Controlled, Multicenter, Single- and Multiple-Ascending-Dose Study to Determine Initial Safety, Tolerability, and Pharmacokinetics of GDC-0134 in Patients With Amyotrophic Lateral Sclerosis: NCT02655614, GN29823*, <https://clinicaltrials.gov/ct2/show/NCT02655614> (accessed December 9, 2021).
- [179] Genentech, Inc., Quotient Clinical, *A Phase I Open-Label Study to Determine the Relative Bioavailability of GDC-0134 and to Investigate the Effect of Food and Proton Pump Inhibitor on Pharmacokinetics of GDC-0134 in Healthy Female Subjects of Non-childbearing Potential: NCT03237741, GP39778*, <https://clinicaltrials.gov/ct2/show/NCT03237741> (accessed December 9, 2021).
- [180] Fischer, D., Breitenbach, J., *Die Pharmaindustrie*, Springer, Berlin, Heidelberg 2020.
- [181] Schmidt, M. F., *Chemische Biologie*, Springer, Berlin, Heidelberg 2020.
- [182] Katsila, T., Spyroulias, G. A., Patrinos, G. P., Matsoukas, M.-T., Computational approaches in target identification and drug discovery. *Comput. Struct. Biotechnol. J.* 2016, 14, 177–184.
- [183] Klebe, G., *Wirkstoffdesign: Entwurf und Wirkung von Arzneistoffen*, Spektrum Akademischer Verlag, Heidelberg 2009.
- [184] Feher, M., Schmidt, J. M., Property distributions: differences between drugs, natural products, and molecules from combinatorial chemistry. *J. Chem. Inf. Comput. Sci.* 2003, 43, 218–227.

- [185] Ertl, P., Cheminformatics analysis of organic substituents: identification of the most common substituents, calculation of substituent properties, and automatic identification of drug-like bioisosteric groups. *J. Chem. Inf. Comput. Sci.* 2003, *43*, 374–380.
- [186] Bohacek, R. S., McMartin, C., Guida, W. C., The art and practice of structure-based drug design: A molecular modeling perspective. *Med. Res. Rev.* 1996, *16*, 3–50.
- [187] Mayr, L. M., Fuerst, P., The future of high-throughput screening. *J. Biomol. Screen.* 2008, *13*, 443–448.
- [188] McDonald, P. R., Roy, A., Chaguturu, R., A roadmap for achieving self-sustainability of academic high throughput screening core facilities. *Drug Discov. World* 2011.
- [189] Sliwoski, G., Kothiwale, S., Meiler, J., Lowe, E. W., Computational Methods in Drug Discovery. *Pharmacol. Rev.* 2014, *66*, 334–395.
- [190] Vénien-Bryan, C., Li, Z., Vuillard, L., Boutin, J. A., Cryo-electron microscopy and X-ray crystallography: complementary approaches to structural biology and drug discovery. *Acta Crystallogr. F Struct. Biol. Commun.* 2017, *73*, 174–183.
- [191] Bera, I., Payghan, P. V., Use of Molecular Dynamics Simulations in Structure-Based Drug Discovery. *Curr. Pharm. Des.* 2019, *25*, 3339–3349.
- [192] Stumpfe, D., Bajorath, J., Similarity searching. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* 2011, *1*, 260–282.
- [193] Leach, A. R., Gillet, V. J., *An introduction to chemoinformatics*, Springer, Dordrecht 2007.
- [194] Maia, E. H. B., Assis, L. C., Oliveira, T. A. de, Da Silva, A. M., Taranto, A. G., Structure-Based Virtual Screening: From Classical to Artificial Intelligence. *Front. Chem.* 2020, *8*, 343.
- [195] Lionta, E., Spyrou, G., Vassilatis, D., Cournia, Z., Structure-Based Virtual Screening for Drug Discovery: Principles, Applications and Recent Advances. *Curr. Top. Med. Chem.* 2014, *14*, 1923–1938.
- [196] Lyne, P. D., Structure-based virtual screening: an overview. *Drug Discov. Today* 2002, *7*, 1047–1055.
- [197] Coupez, B., Lewis, R. A., Docking and scoring--theoretically easy, practically impossible? *Curr. Med. Chem.* 2006, *13*, 2995–3003.
- [198] Perola, E., Minimizing false positives in kinase virtual screens. *Proteins* 2006, *64*, 422–435.
- [199] Brunskole Svegelj, M., Turk, S., Brus, B., Lanisnik Rizner, T., Stojan, J., Gobec, S., Novel inhibitors of trihydroxynaphthalene reductase with antifungal activity identified by ligand-based and structure-based virtual screening. *J. Chem. Inf. Model.* 2011, *51*, 1716–1724.

- [200] Lim, S. V., Rahman, M. B. A., Tejo, B. A., Structure-based and ligand-based virtual screening of novel methyltransferase inhibitors of the dengue virus. *BMC Bioinform.* 2011, *12 Suppl 13*, S24.
- [201] Yu, M., Gu, Q., Xu, J., Discovering new PI3K α inhibitors with a strategy of combining ligand-based and structure-based virtual screening. *J. Comput. Aided Mol. Des.* 2018, *32*, 347–361.
- [202] Perez-Pineiro, R., Burgos, A., Jones, D. C., Andrew, L. C., Rodriguez, H., Suarez, M., Fairlamb, A. H., Wishart, D. S., Development of a novel virtual screening cascade protocol to identify potential trypanothione reductase inhibitors. *J. Med. Chem.* 2009, *52*, 1670–1680.
- [203] Banegas-Luna, A.-J., Cerón-Carrasco, J. P., Pérez-Sánchez, H., A review of ligand-based virtual screening web tools and screening algorithms in large molecular databases in the age of big data. *Future Med. Chem.* 2018, *10*, 2641–2658.
- [204] Wang, Z., Sun, H., Shen, C., Hu, X., Gao, J., Li, D., Cao, D., Hou, T., Combined strategies in structure-based virtual screening. *Phys. Chem. Chem. Phys.* 2020, *22*, 3149–3159.
- [205] Sterling, T., Irwin, J. J., ZINC 15 – Ligand Discovery for Everyone. *J. Chem. Inf. Model.* 2015, *55*, 2324–2337.
- [206] Leach, A. R., *Molecular modelling: Principles and applications*, Pearson/Prentice Hall, Harlow 2009.
- [207] Tavanti, F., Pedone, A., Menziani, M. C., Computational Insight into the Effect of Natural Compounds on the Destabilization of Preformed Amyloid- β (1–40) Fibrils. *Molecules* 2018, *23*.
- [208] Weiner, P. K., Kollman, P. A., AMBER: Assisted model building with energy refinement. A general program for modeling molecules and their interactions. *J. Comp. Chem.* 1981, *2*, 287–303.
- [209] Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W., Kollman, P. A., A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* 1995, *117*, 5179–5197.
- [210] Jorgensen, W. L., Maxwell, D. S., Tirado-Rives, J., Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* 1996, *118*, 11225–11236.

- [211] Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., Karplus, M., CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comp. Chem.* 1983, *4*, 187–217.
- [212] Harder, E., Damm, W., Maple, J., Wu, C., Reboul, M., Xiang, J. Y., Wang, L., Lupyan, D., Dahlgren, M. K., Knight, J. L., Kaus, J. W., Cerutti, D. S., Krilov, G., Jorgensen, W. L., Abel, R., Friesner, R. A., OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins. *J. Chem. Theory Comput.* 2016, *12*, 281–296.
- [213] Banks, J. L., Beard, H. S., Cao, Y., Cho, A. E., Damm, W., Farid, R., Felts, A. K., Halgren, T. A., Mainz, D. T., Maple, J. R., Murphy, R., Philipp, D. M., Repasky, M. P., Zhang, L. Y., Berne, B. J., Friesner, R. A., Gallicchio, E., Levy, R. M., Integrated Modeling Program, Applied Chemical Theory (IMPACT). *J. Comp. Chem.* 2005, *26*, 1752–1780.
- [214] Jorgensen, W. L., Madura, J. D., Swenson, C. J., Optimized intermolecular potential functions for liquid hydrocarbons. *J. Am. Chem. Soc.* 1984, *106*, 6638–6646.
- [215] Kaminski, G., Duffy, E. M., Matsui, T., Jorgensen, W. L., Free Energies of Hydration and Pure Liquid Properties of Hydrocarbons from the OPLS All-Atom Model. *J. Phys. Chem.* 1994, *98*, 13077–13082.
- [216] Weiner, S. J., Kollman, P. A., Nguyen, D. T., Case, D. A., An all atom force field for simulations of proteins and nucleic acids. *J. Comp. Chem.* 1986, *7*, 230–252.
- [217] Shivakumar, D., Harder, E., Damm, W., Friesner, R. A., Sherman, W., Improving the Prediction of Absolute Solvation Free Energies Using the Next Generation OPLS Force Field. *J. Chem. Theory Comput.* 2012, *8*, 2553–2558.
- [218] Roos, K., Wu, C., Damm, W., Reboul, M., Stevenson, J. M., Lu, C., Dahlgren, M. K., Mondal, S., Chen, W., Wang, L., Abel, R., Friesner, R. A., Harder, E. D., OPLS3e: Extending Force Field Coverage for Drug-Like Small Molecules. *J. Chem. Theory Comput.* 2019, *15*, 1863–1874.
- [219] Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., Klein, M. L., Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 1983, *79*, 926–935.
- [220] Sastry, G. M., Adzhigirey, M., Day, T., Annabhimoju, R., Sherman, W., Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *J. Comput. Aided Mol. Des.* 2013, *27*, 221–234.
- [221] Mackay, J. P., Landsberg, M. J., Whitten, A. E., Bond, C. S., Whaddaya Know: A Guide to Uncertainty and Subjectivity in Structural Biology. *Trends Biochem. Sci.* 2017, *42*, 155–167.

- [222] Gohlke, H., Klebe, G., Approaches to the Description and Prediction of the Binding Affinity of Small-Molecule Ligands to Macromolecular Receptors. *Angew. Chem. Int. Ed.* 2002, *41*, 2644–2676.
- [223] Kontoyianni, M., McClellan, L. M., Sokol, G. S., Evaluation of Docking Performance: Comparative Data on Docking Algorithms. *J. Med. Chem.* 2004, *47*, 558–565.
- [224] Anderson, A. C., The Process of Structure-Based Drug Design. *Chem. Biol.* 2003, *10*, 787–797.
- [225] Jones, G., Willett, P., Glen, R. C., Leach, A. R., Taylor, R., Development and validation of a genetic algorithm for flexible docking¹. *J. Mol. Biol.* 1997, *267*, 727–748.
- [226] Alvarez, J. C., Shoichet, B., *Virtual screening in drug discovery*, CRC Press, Taylor & Francis Group, Boca Raton, Florida 2005.
- [227] Kim, M. O., Nichols, S. E., Wang, Y., McCammon, J. A., Effects of histidine protonation and rotameric states on virtual screening of M. tuberculosis RmlC. *J. Comput. Aided Mol. Des.* 2013, *27*, 235–246.
- [228] Schrödinger Release 2015-2, *Protein Preparation Guide*, Schrödinger, LLC, New York, NY 2015.
- [229] Schrödinger Release 2016-3, *Protein Preparation Wizard*, Schrödinger, LLC, New York, NY 2016.
- [230] Schrödinger Release 2016-3, *Protein Preparation Wizard: Manual*, Schrödinger, LLC, New York, NY 2016.
- [231] Andreasen, M., Skeby, K. K., Zhang, S., Nielsen, E. H., Klausen, L. H., Frahm, H., Christiansen, G., Skrydstrup, T., Dong, M., Schiøtt, B., Otzen, D., The importance of being capped: Terminal capping of an amyloidogenic peptide affects fibrillation propensity and fibril morphology. *Biochemistry* 2014, *53*, 6968–6980.
- [232] Lau, Y. H., Andrade, P. de, Quah, S.-T., Rossmann, M., Laraia, L., Sköld, N., Sum, T. J., Rowling, P. J. E., Joseph, T. L., Verma, C., Hyvönen, M., Itzhaki, L. S., Venkitaraman, A. R., Brown, C. J., Lane, D. P., Spring, D. R., Functionalised staple linkages for modulating the cellular activity of stapled peptides - Supporting information. *Chem. Sci.* 2014, *5*, 1804–1809.
- [233] Schrödinger Release 2016-3, *Prime*, Schrödinger, LLC, New York, NY 2016.
- [234] Jacobson, M. P., Pincus, D. L., Rapp, C. S., Day, T. J. F., Honig, B., Shaw, D. E., Friesner, R. A., A hierarchical approach to all-atom protein loop prediction. *Proteins* 2004, *55*, 351–367.

- [235] Goldfeld, D. A., Friesner, R. A., The protein local optimization program and G-protein-coupled receptors: loop restoration and applications to homology modeling. *Meth. Enzymol.* 2013, 522, 1–20.
- [236] Schrödinger Release 2016-3, *Prime: User Manual*, Schrödinger, LLC, New York, NY 2016.
- [237] Gallicchio, E., Zhang, L. Y., Levy, R. M., The SGB/NP hydration free energy model based on the surface generalized born solvent reaction field and novel nonpolar hydration free energy estimators. *J. Comput. Chem.* 2002, 23, 517–529.
- [238] Ghosh, A., Rapp, C. S., Friesner, R. A., Generalized Born Model Based on a Surface Integral Formulation. *J. Phys. Chem. B* 1998, 102, 10983–10990.
- [239] Waszkowycz, B., Clark, D. E., Gancia, E., Outstanding challenges in protein–ligand docking and structure-based virtual screening. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* 2011, 1, 229–259.
- [240] van de Waterbeemd, H., Smith, D. A., Beaumont, K., Walker, D. K., Property-based design: optimization of drug absorption and pharmacokinetics. *J. Med. Chem.* 2001, 44, 1313–1333.
- [241] Lipinski, C. A., Lombardo, F., Dominy, B. W., Feeney, P. J., Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* 1997, 23, 3–25.
- [242] *ClogP: Daylight Chemical Information Software*, Daylight Chemical Information Systems, Inc., 27401 Los Altos, Suite #370, Mission Viejo, CA 92691.
- [243] Leo, A., Jow, P. Y., Silipo, C., Hansch, C., Calculation of hydrophobic constant (log P) from pi and f constants. *J. Med. Chem.* 1975, 18, 865–868.
- [244] Gasteiger, J., *Handbook of Chemoinformatics: From Data to Knowledge*, Wiley-VCH, Weinheim 2003.
- [245] Veber, D. F., Johnson, SR, Cheng, H. Y., Smith, B. R., Ward, K. W., Kopple, K. D., Molecular properties that influence the oral bioavailability of drug candidates. *J. Med. Chem.* 2002, 45, 2615–2623.
- [246] Ruan, K., Zhang, R., Ding, F., Li, L., Zhao, Z., Lin, J., A User-Friendly Task Editor Environment for Large-scale Virtual Screening Application. *in: Proceedings of the Ninth Conference on Grid and Cloud Computing* 2010, 93–97.
- [247] Kitchen, D. B., Decornez, H., Furr, J. R., Bajorath, J., Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat. Rev. Drug Discov.* 2004, 3, 935–949.
- [248] Ashenden, S. K., Screening Library Design. *Meth. Enzymol.* 2018, 610, 73–96.

- [249] Oprea, T. I., Davis, A. M., Teague, S. J., Leeson, P. D., Is There a Difference between Leads and Drugs?: A Historical Perspective. *J. Chem. Inf. Comput. Sci.* 2001, *41*, 1308–1315.
- [250] Ntie-Kang, F., Onguéné, P. A., Scharfe, M., Owono Owono, L. C., Megnassan, E., Mbaze, L. M., Sippl, W., Efange, S. M. N., ConMedNP: a natural product library from Central African medicinal plants for drug discovery. *RSC Adv.* 2014, *4*, 409–419.
- [251] Branson, K. M., Smith, B. J., The Role of Virtual Screening in Computer Aided Structure-Based Drug Design. *Aust. J. Chem.* 2004, *57*, 1029.
- [252] Manoharan, P., Ghoshal, N., Fragment-based virtual screening approach and molecular dynamics simulation studies for identification of BACE1 inhibitor leads. *J. Biomol. Struct. Dyn.* 2018, *36*, 1878–1892.
- [253] Weininger, D., SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J. Chem. Inf. Model.* 1988, *28*, 31–36.
- [254] PerkinElmer Inc., ChemDraw Pro 11.0: www.cambridgesoft.com 1998-2016.
- [255] Schrödinger Release 2017-3, *Maestro Version 11.3.016: MMshare Version 3.9.016*, Schrödinger LLC, New York, NY 2017.
- [256] Schrödinger Release 2017-1, *Ligprep*, Schrödinger, LLC, New York, NY 2017.
- [257] Schrödinger Release 2020-1 Documentation, *User Manual*, Schrödinger, LLC, New York, NY 2020.
- [258] Shelley, J. C., Cholleti, A., Frye, L. L., Greenwood, J. R., Timlin, M. R., Uchimaya, M., Epik: a software program for pK(a) prediction and protonation state generation for drug-like molecules. *J. Comput. Aided Mol. Des.* 2007, *21*, 681–691.
- [259] Greenwood, J. R., Calkins, D., Sullivan, A. P., Shelley, J. C., Towards the comprehensive, rapid, and accurate prediction of the favorable tautomeric states of drug-like molecules in aqueous solution. *J. Comput. Aided Mol. Des.* 2010, *24*, 591–604.
- [260] Irwin, J. J., Shoichet, B. K., ZINC - A free database of commercially available compounds for virtual screening. *J. Chem. Inf. Model.* 2005, *45*, 177–182.
- [261] Irwin, J. J., Sterling, T., Mysinger, M. M., Bolstad, E. S., Coleman, R. G., ZINC: A Free Tool to Discover Chemistry for Biology. *J. Chem. Inf. Model.* 2012, *52*, 1757–1768.
- [262] ZINC, <http://zinc15.docking.org/> (accessed November 24, 2016).
- [263] Hilbig, M., Rarey, M., MONA 2: A Light Cheminformatics Platform for Interactive Compound Library Processing. *J. Chem. Inf. Model.* 2015, *55*, 2071–2078.
- [264] Hilbig, M., Urbaczek, S., Groth, I., Heuser, S., Rarey, M., MONA - Interactive manipulation of molecule collections. *J. Cheminformatics* 2013, *5*, 38.

- [265] Urbaczek, S., Kolodzik, A., Fischer, J. R., Lippert, T., Heuser, S., Groth, I., Schulz-Gasch, T., Raray, M., NAOMI: On the Almost Trivial Task of Reading Molecules from Different File formats. *J. Chem. Inf. Model.* 2011, *51*, 3199–3207.
- [266] Kirchmair, J., Markt, P., Distinto, S., Wolber, G., Langer, T., Evaluation of the performance of 3D virtual screening protocols: RMSD comparisons, enrichment assessments, and decoy selection—What can we learn from earlier mistakes? *J. Comput. Aided Mol. Des.* 2008, *22*, 213–228.
- [267] Good, A., Oprea, T., Optimization of CAMD techniques 3. Virtual screening enrichment studies: a help or hindrance in tool selection? *J. Comput. Aided Mol. Des.* 2008, *22*, 169–178.
- [268] Mysinger, M. M., Carchia, M., Irwin, J. J., Shoichet, B. K., Directory of Useful Decoys, Enhanced (DUD-E): Better Ligands and Decoys for Better Benchmarking. *J. Med. Chem.* 2012, *55*, 6582–6594.
- [269] Chaput, L., Martinez-Sanz, J., Saettel, N., Mouawad, L., Benchmark of four popular virtual screening programs: construction of the active/decoy dataset remains a major determinant of measured performance. *J. Cheminformatics* 2016, *8*, 56.
- [270] Gaulton, A., Bellis, L. J., Bento, A. P., Chambers, J., Davies, M., Hersey, A., Light, Y., McGlinchey, S., Michalovich, D., Al-Lazikani, B., Overington, J. P., ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Res.* 2012, *40*, D1100-7.
- [271] Rogers, D., Hahn, M., Extended-Connectivity Fingerprints. *J. Chem. Inf. Model.* 2010, *50*, 742–754.
- [272] Saikia, S., Bordoloi, M., Molecular Docking: Challenges, Advances and its Use in Drug Discovery Perspective. *Current Drug Targets*, 20(5), 501-521. *Curr. Drug Targets* 2019, *20*, 501–521.
- [273] Sherman, W., Day, T., Jacobson, M. P., Friesner, R. A., Farid, R., Novel Procedure for Modeling Ligand/Receptor Induced Fit Effects. *J. Med. Chem.* 2006, *49*, 534–553.
- [274] Nabuurs, S. B., Wagener, M., Vlieg, J. de, A Flexible Approach to Induced Fit Docking. *J. Med. Chem.* 2007, *50*, 6507–6518.
- [275] Clark, A. J., Tiwary, P., Borrelli, K., Feng, S., Miller, E. B., Abel, R., Friesner, R. A., Berne, B. J., Prediction of Protein-Ligand Binding Poses via a Combination of Induced Fit Docking and Metadynamics Simulations. *J. Chem. Theory Comput.* 2016, *12*, 2990–2998.
- [276] Barreca, M. L., Iraci, N., Luca, L. de, Chimirri, A., Induced-fit docking approach provides insight into the binding mode and mechanism of action of HIV-1 integrase inhibitors. *ChemMedChem* 2009, *4*, 1446–1456.

- [277] Gabb, H. A., Jackson, R. M., Sternberg, M. J., Modelling protein docking using shape complementarity, electrostatics and biochemical information. *J. Mol. Biol.* 1997, 272, 106–120.
- [278] Claussen, H., Buning, C., Rarey, M., Lengauer, T., FlexE: efficient molecular docking considering protein structure variations. *J. Mol. Biol.* 2001, 308, 377–395.
- [279] Yoon, S., Welsh, W. J., Identification of a minimal subset of receptor conformations for improved multiple conformation docking and two-step scoring. *J. Chem. Inf. Comput. Sci.* 2004, 44, 88–96.
- [280] Craig, I. R., Essex, J. W., Spiegel, K., Ensemble Docking into Multiple Crystallographically Derived Protein Structures: An Evaluation Based on the Statistical Analysis of Enrichments. *J. Chem. Inf. Model.* 2010, 50, 511–524.
- [281] Campbell, A. J., Lamb, M. L., Joseph-McCarthy, D., Ensemble-Based Docking Using Biased Molecular Dynamics. *J. Chem. Inf. Model.* 2014, 54, 2127–2138.
- [282] Sousa, S. F., Fernandes, P. A., Ramos, M. J., Protein-ligand docking: current status and future challenges. *Proteins* 2006, 65, 15–26.
- [283] Neves, M. A. C., Totrov, M., Abagyan, R., Docking and scoring with ICM: the benchmarking results and strategies for improvement. *J. Comput. Aided Mol. Des.* 2012, 26, 675–686.
- [284] Friesner, R. A., Banks, J. L., Murphy, R. B., Halgren, T. A., Klicic, J. J., Mainz, D. T., Repasky, M. P., Knoll, E. H., Shelley, M., Perry, J. K., Shaw, D. E., Francis, P., Shenkin, P. S., Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *J. Med. Chem.* 2004, 47, 1739–1749.
- [285] Brooijmans, N., Kuntz, I. D., Molecular recognition and docking algorithms. *Annu. Rev. Biophys. Biomol. Struct.* 2003, 32, 335–373.
- [286] Kramer, B., Rarey, M., Lengauer, T., Evaluation of the FLEXX incremental construction algorithm for protein-ligand docking. *Proteins* 1999, 37, 228–241.
- [287] Verdonk, M. L., Cole, J. C., Hartshorn, M. J., Murray, C. W., Taylor, R. D., Improved protein-ligand docking using GOLD. *Proteins* 2003, 52, 609–623.
- [288] Morris, G. M., Goodsell, D. S., Halliday, R. S., Huey, R., Hart, W. E., Belew, R. K., Olson, A. J., Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comp. Chem.* 1998, 19, 1639–1662.
- [289] Jain, A. N., Scoring functions for protein-ligand docking. *Curr. Protein Pept. Sci.* 2006, 7, 407–420.

- [290] Huang, S.-Y., Grinter, S. Z., Zou, X., Scoring functions and their evaluation methods for protein-ligand docking: recent advances and future directions. *Phys. Chem. Chem. Phys.* 2010, *12*, 12899–12908.
- [291] Ferreira, L. G., dos Santos, R. N., Oliva, G., Andricopulo, A. D., Molecular docking and structure-based drug design strategies. *Molecules* 2015, *20*, 13384–13421.
- [292] Salmaso, V., Moro, S., Bridging Molecular Docking to Molecular Dynamics in Exploring Ligand-Protein Recognition Process: An Overview. *Front. pharmacol.* 2018, *9*, 923.
- [293] Böhm, H. J., The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional structure. *J. Comput. Aided Mol. Des.* 1994, *8*, 243–256.
- [294] Rarey, M., Kramer, B., Lengauer, T., Time-efficient docking of flexible ligands into active sites of proteins. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* 1995, *3*, 300–308.
- [295] Gohlke, H., Hendlich, M., Klebe, G., Knowledge-based scoring function to predict protein-ligand interactions. *J. Mol. Biol.* 2000, *295*, 337–356.
- [296] Velec, H. F., Gohlke, H., Klebe, G., DrugScore(CSD)-knowledge-based scoring function derived from small molecule crystal data with superior recognition rate of near-native ligand poses and better affinity prediction. *J. Med. Chem.* 2005, *48*, 6296–6303.
- [297] Muegge, I., Martin, Y. C., A general and fast scoring function for protein-ligand interactions: a simplified potential approach. *J. Med. Chem.* 1999, *42*, 791–804.
- [298] Muegge, I., PMF scoring revisited. *J. Med. Chem.* 2006, *49*, 5895–5902.
- [299] Charifson, P. S., Corkery, J. J., Murcko, M. A., Walters, W. P., Consensus scoring: A method for obtaining improved hit rates from docking databases of three-dimensional structures into proteins. *J. Med. Chem.* 1999, *42*, 5100–5109.
- [300] Wang, R., Wang, S., How Does Consensus Scoring Work for Virtual Library Screening? An Idealized Computer Experiment. *J. Chem. Inf. Comput. Sci.* 2001, *41*, 1422–1426.
- [301] Halgren, T. A., Murphy, R. B., Friesner, R. A., Beard, H. S., Frye, L. L., Pollard, W. T., Banks, J. L., Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening. *J. Med. Chem.* 2004, *47*, 1750–1759.
- [302] Friesner, R. A., Murphy, R. B., Repasky, M. P., Frye, L. L., Greenwood, J. R., Halgren, T. A., Sanschagrin, P. C., Mainz, D. T., Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein–Ligand Complexes. *J. Med. Chem.* 2006, *49*, 6177–6196.
- [303] Patrick, G. L., *An introduction to medicinal chemistry*, Oxford University Press, Oxford 2013.

- [304] Schrödinger Release 2020-1, *Glide: User Manual*, Schrödinger, LLC, New York, NY 2020.
- [305] Schrödinger, *Docking and Scoring*, <https://www.schrodinger.com/science-articles/docking-and-scoring> (accessed November 10, 2020).
- [306] Repasky, M. P., Murphy, R. B., Banks, J. L., Greenwood, J. R., Tubert-Brohman, I., Bhat, S., Friesner, R. A., Docking performance of the glide program as evaluated on the Astex and DUD datasets: a complete set of glide SP results and selected results for a new scoring function integrating WaterMap and glide. *J. Comput. Aided Mol. Des.* 2012, 26, 787–799.
- [307] Ramírez, D., Caballero, J., Is It Reliable to Use Common Molecular Docking Methods for Comparing the Binding Affinities of Enantiomer Pairs for Their Protein Target? *Int. J. Mol. Sci.* 2016, 17.
- [308] Eldridge, M. D., Murray, C. W., Auton, T. R., Paolini, G. V., Mee, R. P., Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J. Comput. Aided Mol. Des.* 1997, 11, 425–445.
- [309] Sherman, W., Beard, H. S., Farid, R., Use of an Induced Fit Receptor Structure in Virtual Screening. *Chem. Biol. Drug Des.* 2006, 67, 83–84.
- [310] Sun, Z., Liu, Q., Qu, G., Feng, Y., Reetz, M. T., Utility of B-Factors in Protein Science: Interpreting Rigidity, Flexibility, and Internal Motion and Engineering Thermostability. *Chem. Rev.* 2019, 119, 1626–1665.
- [311] Reulecke, I., Lange, G., Albrecht, J., Klein, R., Rarey, M., Towards an Integrated Description of Hydrogen Bonding and Dehydration: Decreasing False Positives in Virtual Screening with the HYDE Scoring Function. *ChemMedChem* 2008, 3, 885–897.
- [312] Schneider, N., Lange, G., Hindle, S., Klein, R., Rarey, M., A consistent description of Hydrogen bond and Dehydration energies in protein–ligand complexes: methods behind the HYDE scoring function. *J. Comput. Aided Mol. Des.* 2013, 27, 15–29.
- [313] Schneider, N., Hindle, S., Lange, G., Klein, R., Albrecht, J., Briem, H., Beyer, K., Claußen, H., Gastreich, M., Lemmen, C., Rarey, M., Substantial improvements in large-scale redocking and screening using the novel HYDE scoring function. *J. Comput. Aided Mol. Des.* 2012, 26, 701–723.
- [314] Huang, N., Shoichet, B. K., Irwin, J. J., Benchmarking Sets for Molecular Docking. *J. Med. Chem.* 2006, 49, 6789–6801.
- [315] Liebeschuetz, J. W., Cole, J. C., Korb, O., Pose prediction and virtual screening performance of GOLD scoring functions in a standardized test. *J. Comput. Aided Mol. Des.* 2012, 26, 737–748.

- [316] BiosolveIT GmbH, *LeadIT 2.1.5*, St. Augustin, Germany 2013.
- [317] Schrödinger Release 2017-3, *Prime: User Manual*, Schrödinger, LLC, New York, NY 2017.
- [318] Li, J., Abel, R., Zhu, K., Cao, Y., Zhao, S., Friesner, R. A., The VSGB 2.0 model: A next generation energy model for high resolution protein structure modeling. *Proteins* 2011, *79*, 2794–2812.
- [319] Sirin, S., Kumar, R., Martinez, C., Karmilowicz, M. J., Ghosh, P., Abramov, Y. A., van Martin, Sherman, W., A Computational Approach to Enzyme Design: Predicting ω -Aminotransferase Catalytic Activity Using Docking and MM-GBSA Scoring. *J. Chem. Inf. Model.* 2014, *54*, 2334–2346.
- [320] Srinivasan, J., Cheatham, T. E., Cieplak, P., Kollman, P. A., Case, D. A., Continuum Solvent Studies of the Stability of DNA, RNA, and Phosphoramidate–DNA Helices. *J. Am. Chem. Soc.* 1998, *120*, 9401–9409.
- [321] Kollman, P. A., Massova, I., Reyes, C., Kuhn, B., Huo, S., Chong, L., Lee, M., Lee, T., Duan, Y., Wang, W., Donini, O., Cieplak, P., Srinivasan, J., Case, D. A., Cheatham, T. E., Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc. Chem. Res.* 2000, *33*, 889–897.
- [322] Srinivasan, J., Miller, J., Kollman, P. A., Case, D. A., Continuum solvent studies of the stability of RNA hairpin loops and helices. *J. Biomol. Struct. Dyn.* 1998, *16*, 671–682.
- [323] Tripathi, S. K., Soundarya, R. N., Singh, P., Singh, S. K., Comparative analysis of various electrostatic potentials on docking precision against cyclin-dependent kinase 2 protein: a multiple docking approach. *Chem. Biol. Drug Des.* 2015, *85*, 107–118.
- [324] Lyne, P. D., Lamb, M. L., Saeh, J. C., Accurate prediction of the relative potencies of members of a series of kinase inhibitors using molecular docking and MM-GBSA scoring. *J. Med. Chem.* 2006, *49*, 4805–4808.
- [325] Adasme-Carreño, F., Muñoz-Gutierrez, C., Caballero, J., Alzate-Morales, J. H., Performance of the MM/GBSA scoring using a binding site hydrogen bond network-based frame selection: the protein kinase case. *Phys. Chem. Chem. Phys.* 2014, *16*, 14047–14058.
- [326] Verkhivker, G. M., Bouzida, D., Gehlhaar, D. K., Rejto, P. A., Arthurs, S., Colson, A. B., Freer, S. T., Larson, V., Luty, B. A., Marrone, T., Rose, P. W., Deciphering common failures in molecular docking of ligand-protein complexes. *J. Comput. Aided Mol. Des.* 2000, *14*, 731–751.
- [327] Gehlhaar, D. K., Verkhivker, G. M., Rejto, P. A., Sherman, C. J., Fogel, D. R., Fogel, L. J., Freer, S. T., Molecular recognition of the inhibitor AG-1343 by HIV-1 protease: conformationally flexible docking by evolutionary programming. *Chem. Biol.* 1995, *2*, 317–324.

- [328] Yang, J.-M., Chen, C.-C., GEMDOCK: a generic evolutionary method for molecular docking. *Proteins* 2004, 55, 288–304.
- [329] Rarey, M., Kramer, B., Lengauer, T., Klebe, G., A fast flexible docking method using an incremental construction algorithm. *J. Mol. Biol.* 1996, 261, 470–489.
- [330] Teramoto, R., Fukunishi, H., Consensus Scoring with Feature Selection for Structure-Based Virtual Screening. *J. Chem. Inf. Model.* 2008, 48, 288–295.
- [331] Berthold, M. R., Cebron, N., Dill, F., Gabriel, T. R., Kötter, T., Meinl, T., Ohl, P., Sieb, C., Thiel, K., Wiswedel, B., KNIME: The Konstanz Information Miner. In: Preisach, C, Burkhardt H., Schmidt-Thieme L., Decker R. (Hg.) *Data analysis, machine learning and applications. Studies in Classification, Data Analysis, and Knowledge Organization. Springer Berlin Heidelberg.* 2008, 319–326.
- [332] Sastry, G. M., Inakollu, V. S. S., Sherman, W., Boosting Virtual Screening Enrichments with Data Fusion: Coalescing Hits from Two-Dimensional Fingerprints, Shape, and Docking. *J. Chem. Inf. Model.* 2013, 53, 1531–1542.
- [333] Casey, F. P., Pihan, E., Shields, D. C., Discovery of small molecule inhibitors of protein-protein interactions using combined ligand and target score normalization. *J. Chem. Inf. Model.* 2009, 49, 2708–2717.
- [334] Jain, A. N., Bias, reporting, and sharing: computational evaluations of docking methods. *J. Comput. Aided Mol. Des.* 2008, 22, 201–212.
- [335] Hevener, K. E., Zhao, W., Ball, D. M., Babaoglu, K., Qi, J., White, S. W., Lee, R. E., Validation of molecular docking programs for virtual screening against dihydropteroate synthase. *J. Chem. Inf. Model.* 2009, 49, 444–460.
- [336] Pearlman, D. A., Charifson, P. S., Improved Scoring of Ligand–Protein Interactions Using OWFEG Free Energy Grids. *J. Med. Chem.* 2001, 44, 502–511.
- [337] Williams, T., Kelley, C., *gnuplot 4.6: An Interactive Plotting Program* 2013.
- [338] Afifi, K., Al-Sadek, A. F., Improving classical scoring functions using random forest: The non-additivity of free energy terms' contributions in binding. *Chem. Biol. Drug Des.* 2018, 92, 1429–1434.
- [339] Alder, B. J., Wainwright, T. E., Phase Transition for a Hard Sphere System. *J. Chem. Phys.* 1957, 27, 1208–1209.
- [340] van Gunsteren, W. F., Berendsen, H. J. C., Moleküldynamik-Computersimulationen; Methodik, Anwendungen und Perspektiven in der Chemie. *Angew. Chem.* 1990, 102, 1020–1055.

- [341] Verlet, L., Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Phys. Rev.* 1967, *159*, 98–103.
- [342] Swope, W. C., Andersen, H. C., Berens, P. H., Wilson, K. R., A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters. *J. Chem. Phys.* 1982, *76*, 637–649.
- [343] Hockney, R. W., The potential calculation and some applications. *Meth. Comput. Phys.* 1970, 135–211.
- [344] Jensen, F., *Introduction to computational chemistry*, Wiley, Chichester 2009.
- [345] Bray, S. A., Senapathi, T., Barnett, C. B., Grüning, B. A., Intuitive, reproducible high-throughput molecular dynamics in Galaxy: a tutorial. *Journal of Cheminformatics* 2020, *12*, 54.
- [346] Xu, Y., Meng, X., Molecular Simulation Elaborating the Mechanism of 1 β -Hydroxy Alantolactone Inhibiting Ubiquitin-Conjugating Enzyme UbcH5s. *Sci. Rep.* 2020, *10*.
- [347] Khalili-Araghi, F., Gumbart, J., Wen, P.-C., Sotomayor, M., Tajkhorshid, E., Schulten, K., Molecular dynamics simulations of membrane channels and transporters. *Curr. Opin. Struct. Biol.* 2009, *19*, 128–137.
- [348] Amir, M., Mohammad, T., Kumar, V., Alajmi, M. F., Rehman, M. T., Hussain, A., Alam, P., Dohare, R., Islam, A., Ahmad, F., Hassan, M. I., Structural Analysis and Conformational Dynamics of STN1 Gene Mutations Involved in Coat Plus Syndrome. *Front. Mol. Biosci.* 2019, *6*, 41.
- [349] Zhu, J., Lv, Y., Han, X., Xu, D., Han, W., Understanding the differences of the ligand binding/unbinding pathways between phosphorylated and non-phosphorylated ARH1 using molecular dynamics simulations. *Sci. Rep.* 2017, *7*, 12439.
- [350] Wacker, D., Wang, S., McCorvy, J. D., Betz, R. M., Venkatakrisnan, A. J., Levit, A., Lansu, K., Schools, Z. L., Che, T., Nichols, D. E., Shoichet, B. K., Dror, R. O., Roth, B. L., Crystal Structure of an LSD-Bound Human Serotonin Receptor. *Cell* 2017, *168*, 377-389.e12.
- [351] Pandey, R. K., Narula, A., Naskar, M., Srivastava, S., Verma, P., Malik, R., Shah, P., Prajapati, V. K., Exploring dual inhibitory role of febrifugine analogues against Plasmodium utilizing structure-based virtual screening and molecular dynamic simulation. *J. Biomol. Struct. Dyn.* 2017, *35*, 791–804.
- [352] Choudhary, M. I., Shaikh, M., Tul-Wahab, A., Ur-Rahman, A., In silico identification of potential inhibitors of key SARS-CoV-2 3CL hydrolase (Mpro) via molecular docking, MMGBSA predictive binding energy calculations, and molecular dynamics simulation. *PLoS ONE* 2020, *15*, e0235030.

- [353] Samykannu, G., Vijayababu, P., Antonyraj, C. B., Narayanan, S., Basheer Ahamed, S. I., Investigations of binding mode insight in Salmonella typhi type-III secretion system tip protein (SipD): A molecular docking and MD simulation study. *Inform. Med. Unlocked* 2017, 9, 166–172.
- [354] Strahs, D., Weinstein, H., Comparative modeling and molecular dynamics studies of the delta, kappa and mu opioid receptors. *Protein Eng.* 1997, 10, 1019–1038.
- [355] Vijayan, R. S. K., Trivedi, N., Roy, S. N., Bera, I., Manoharan, P., Payghan, P. V., Bhattacharyya, D., Ghoshal, N., Modeling the closed and open state conformations of the GABA(A) ion channel - plausible structural insights for channel gating. *J. Chem. Inf. Model.* 2012, 52, 2958–2969.
- [356] Offutt, T. L., Swift, R. V., Amaro, R. E., Enhancing Virtual Screening Performance of Protein Kinases with Molecular Dynamics Simulations. *J. Chem. Inf. Model.* 2016, 56, 1923–1935.
- [357] Dror, R. O., Pan, A. C., Arlow, D. H., Borhani, D. W., Maragakis, P., Shan, Y., Xu, H., Shaw, D. E., Pathway and mechanism of drug binding to G-protein-coupled receptors. *Proc. Natl. Acad. Sci. U.S.A.* 2011.
- [358] Bisignano, P., Doerr, S., Harvey, M. J., Favia, A. D., Cavalli, A., Fabritiis, G. de, Kinetic characterization of fragment binding in AmpC β -lactamase by high-throughput molecular simulations. *J. Chem. Inf. Model.* 2014, 54, 362–366.
- [359] Aqvist, J., Luzhkov, V. B., Brandsdal, B. O., Ligand binding affinities from MD simulations. *Acc. Chem. Res.* 2002, 35, 358–365.
- [360] Kollman, P., Free energy calculations: Applications to chemical and biochemical phenomena. *Chem. Rev.* 1993, 93, 2395–2417.
- [361] Bowers, K. J., Chow, E., Xu, H., Dror, R. O., Eastwood, M. P., Gregersen, B. A., Klepeis, J. L., Kolossvary, I., Moraes, M. A., Sacerdoti, F. D., Salmon, J. K., Shan, Y., Shaw, D. E., *Scalable Algorithms for Molecular Dynamics Simulations on Commodity Clusters*. Proceedings of the ACM/IEEE Conference on Supercomputing (SC06), Tampa, Florida 2006.
- [362] Schrödinger Release 2020-1, *Desmond Molecular Dynamics System*, D. E. Shaw Research, New York, NY 2020.
- [363] Braun, E., Gilmer, J., Mayes, H. B., Mobley, D. L., Monroe, J. I., Prasad, S., Zuckerman, D. M., Best Practices for Foundations in Molecular Simulations Article v1.0. *LiveCoMS* 2019, 1, 5957.
- [364] Chintha, C., Carlesso, A., Gorman, A. M., Samali, A., Eriksson, L. A., Molecular modeling provides a structural basis for PERK inhibitor selectivity towards RIPK1. *RCS Advances* 2020, 10, 367–375.

- [365] Ivanova, L., Tammiku-Taul, J., García-Sosa, A. T., Sidorova, Y., Saarma, M., Karelson, M., Molecular Dynamics Simulations of the Interactions between Glial Cell Line-Derived Neurotrophic Factor Family Receptor GFR α 1 and Small-Molecule Ligands. *ACS omega* 2018, 3, 11407–11414.
- [366] Tomassi, S., Montalban, F. F., Russo, R., Novellino, E., Messere, A., Di Maro, S., Investigation of the Stereochemical-Dependent DNA and RNA Binding of Arginine-Based Nucleopeptides. *Symmetry* 2019, 11, 567.
- [367] Steinmetz, A., Yvorra, T., Retailleau, P., Lantz, O., Schmidt, F., Datasets and analyses of molecular dynamics simulations of covalent binary and ternary complexes of MHC class I-related molecule/T-cell receptor (MR1/TCR) agonists to understand complex formation and conditions of fluorescent labelling. *Data Brief*. 2021, 34, 106704.
- [368] Patodia, S., Molecular Dynamics Simulation of Proteins: A Brief Overview. *J. Phys. Chem. Biophys.* 2014, 4, 1–4.
- [369] MacKerell, A. D., Bashford, D., Bellott, M., Dunbrack, R. L., Evanseck, J. D., Field, M. J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F. T., Mattos, C., Michnick, S., Ngo, T., Nguyen, D. T., Prodhom, B., Reiher, W. E., Roux, B., Schlenkrich, M., Smith, J. C., Stote, R., Straub, J., Watanabe, M., Wiórkiewicz-Kuczera, J., Yin, D., Karplus, M., All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* 1998, 102, 3586–3616.
- [370] Kollman, P., Dixon, R., Cornell, W., Fox, T., Chipot, C., Pohorille, A., The development/application of a ‘minimalist’ organic/biochemical molecular mechanic force field using a combination of ab initio calculations and experimental data. In: *van Gunsteren, W.F., Weiner, P.K., Wilkinson, A.J. (Hg.) Computer Simulation of Biomolecular Systems: Theoretical and Experimental Applications, Vol. 3. Springer, Dordrecht*. 1997, 83–96.
- [371] Scott, W. R. P., Hunenberger, P. H., Tironi, I. G., Mark, A. E., Billeter, S. R., Fennel, J., Torda, A. E., Huber, T., Kruger, P., van Gunsteren, W. F., The GROMOS biomolecular simulation program package. *J. Phys. Chem. A* 1999, 103, 3596–3607.
- [372] Halgren, T. A., MMFF VII. Characterization of MMFF94, MMFF94s, and other widely available force fields for conformational energies and for intermolecular-interaction energies and geometries. *J. Comp. Chem.* 1999, 20, 730–748.
- [373] Tuckerman, M., Berne, B. J., Martyna, G. J., Reversible multiple time scale molecular dynamics. *J. Chem. Phys.* 1992, 97, 1990–2001.

-
- [374] Darden, T., York, D., Pedersen, L., Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems. *J. Chem. Phys.* 1993, *98*, 10089–10092.
- [375] Ewald, P. P., Die Berechnung optischer und elektrostatischer Gitterpotentiale. *Ann. Phys.* 1921, *369*, 253–287.
- [376] Schlick, T., *Molecular Modeling and Simulation: An Interdisciplinary Guide: An Interdisciplinary Guide*, Springer, New York 2010.
- [377] Williams, D. E., Accelerated convergence of crystal-lattice potential sums. *Acta Cryst. A* 1971, *27*, 452–455.
- [378] Ryckaert, J.-P., Ciccotti, G., Berendsen, H. J. C., Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* 1977, *23*, 327–341.
- [379] Tuckerman, M. E., Martyna, G. J., Understanding Modern Molecular Dynamics: Techniques and Applications. *J. Phys. Chem. B* 2000, *104*, 159–178.
- [380] Basconi, J. E., Shirts, M. R., Effects of Temperature Control Algorithms on Transport Properties and Kinetics in Molecular Dynamics Simulations. *J. Chem. Theory Comput.* 2013, *9*, 2887–2899.
- [381] Scherer, P. O. J., *Computational Physics: Simulation of Classical and Quantum Systems*, Springer, Berlin, Heidelberg 2010.
- [382] Luckhurst, G. R., Veracini, C. A., *Molecular Dynamics of Liquid Crystals*, Springer, Dordrecht 2012.
- [383] Grest, G. S., Kremer, K., Molecular dynamics simulation for polymers in the presence of a heat bath. *Phys. Rev. A* 1986, *33*, 3628–3631.
- [384] Lemons, D. S., Gythiel, A., Paul Langevin’s 1908 paper “On the Theory of Brownian Motion” [“Sur la théorie du mouvement brownien,” *C. R. Acad. Sci. (Paris)* 146, 530–533 (1908)]. *Am. J. Phys.* 1997, *65*, 1079–1081.
- [385] Izaguirre, J. A., Catarello, D. P., Wozniak, J. M., Skeel, R. D., Langevin stabilization of molecular dynamics. *J. Chem. Phys.* 2001, *114*, 2090–2098.
- [386] Munakata, T., Rosinberg, M. L., Entropy production and fluctuation theorems for Langevin processes under continuous non-Markovian feedback control. *Phys. Rev. Lett.* 2014, *112*, 180601.
- [387] Shang, X., Kröger, M., Time Correlation Functions of Equilibrium and Nonequilibrium Langevin Dynamics: Derivations and Numerics Using Random Numbers. *SIAM Rev.* 2020, *62*, 901–935.

- [388] Mishin, Y., Hickman, J., Energy spectrum of a Langevin oscillator. *Phys. Rev. E* 2016, *94*, 62151.
- [389] Tropper, M. M., Ergodic and quasideterministic properties of finite-dimensional stochastic systems. *J. Stat. Phys.* 1977, *17*, 491–509.
- [390] Yang, L., Ahmed, A., Sandler, S. I., Comparison of two simulation methods to compute solvation free energies and partition coefficients. *J. Comp. Chem.* 2013, *34*, 284–293.
- [391] Medved, A., Davis, R., Vasquez, P. A., Understanding Fluid Dynamics from Langevin and Fokker–Planck Equations. *Fluids* 2020, *5*, 40.
- [392] Ruiz-Franco, J., Rovigatti, L., Zaccarelli, E., On the effect of the thermostat in non-equilibrium molecular dynamics simulations. *Eur. Phys. J. E* 2018, *41*, 80.
- [393] Feller, S. E., Zhang, Y., Pastor, R. W., Brooks, B. R., Constant pressure molecular dynamics simulation: The Langevin piston method. *J. Chem. Phys.* 1995, *103*, 4613–4621.
- [394] Andersen, H. C., Molecular dynamics simulations at constant pressure and/or temperature. *J. Chem. Phys.* 1980, *72*, 2384–2393.
- [395] D. E. Shaw Research, *Desmond Users Guide: Release 3.4.0 / 0.7* 2012.
- [396] Schrödinger Release 2020-1, *Desmond: User Manual*, Schrödinger LLC, New York, NY 2020.
- [397] Botón-Fernández, M., Martín-Vide, C., Santander-Jiménez, S., Vega-Rodríguez, M. A., *Algorithms for Computational Biology*, Springer International Publishing, Cham 2016.
- [398] Aggarwal, R., Koes, D. R., *Learning RMSD to Improve Protein-Ligand Scoring and Pose Selection*, ChemRxiv Preprint 2020.
- [399] Yusuf, D., Davis, A. M., Kleywegt, G. J., Schmitt, S., An alternative method for the evaluation of docking performance: RSR vs RMSD. *J. Chem. Inf. Model.* 2008, *48*, 1411–1422.
- [400] Sargsyan, K., Grauffel, C., Lim, C., How Molecular Size Impacts RMSD Applications in Molecular Dynamics Simulations. *J. Chem. Theory Comput.* 2017, *13*, 1518–1524.
- [401] Lee, H. S., Qi, Y., Im, W., Effects of N-glycosylation on protein conformation and dynamics: Protein Data Bank analysis and molecular dynamics simulation study. *Sci. Rep.* 2015, *5*, 8926.
- [402] Sharma, R. D., Lynn, A. M., Sharma, P. K., Rajnee, Jawaid, S., High temperature unfolding of Bacillus anthracis amidase-03 by molecular dynamics simulations. *Bioinformatics* 2009, *3*, 430–434.
- [403] Kaushik, A. C., Sahi, S., Molecular modeling and molecular dynamics simulation-based structural analysis of GPR3. *Netw. Model. Anal. Health Inform. Bioinform.* 2017, *6*.

- [404] Palazzolo, L., Parravicini, C., Laurenzi, T., Guerrini, U., Indiveri, C., Gianazza, E., Eberini, I., In silico Description of LAT1 Transport Mechanism at an Atomistic Level. *Front. Chem.* 2018, 6, 350.
- [405] Rao, F., Karplus, M., Protein dynamics investigated by inherent structure analysis. *Proc. Natl. Acad. Sci. U.S.A.* 2010, 107, 9152–9157.
- [406] Schrödinger Release 2016-3, *Maestro: User Manual*, Schrödinger, LLC, New York, NY 2016.
- [407] Auffinger, P., Louise-May, S., Westhof, E., Multiple Molecular Dynamics Simulations of the Anticodon Loop of tRNA^{Asp} in Aqueous Solution with Counterions. *J. Am. Chem. Soc.* 1995, 117, 6720–6726.
- [408] Wang, H., Gao, Z., Song, P., Hu, B., Wang, J., Cheng, M., Molecular dynamics simulation and QM/MM calculation reveal the selectivity mechanism of type I 1/2 kinase inhibitors: the effect of intramolecular H-bonds and conformational restriction for improved selectivity. *Phys. Chem. Chem. Phys.* 2019, 21, 24147–24164.
- [409] Han, J., Kamber, M., Pei, J., *Data Mining: Concepts and Techniques: Concepts and Techniques*, Elsevier Inc., Waltham, Massachusetts, USA 2011.
- [410] Amaro, R. E., Baron, R., McCammon, J. A., An improved relaxed complex scheme for receptor flexibility in computer-aided drug design. *J. Comput. Aided Mol. Des.* 2008, 22, 693–705.
- [411] Alexander, N., Woetzel, N., Meiler, J., bcl::Cluster: A method for clustering biological molecules coupled with visualization in the Pymol Molecular Graphics System. *IEEE Int. Conf. Comput. Adv. Bio. Med. Sci.* 2011, 2011, 13–18.
- [412] Makeneni, S., Thieker, D. F., Woods, R. J., Applying Pose Clustering and MD Simulations To Eliminate False Positives in Molecular Docking. *J. Chem. Inf. Model.* 2018, 58, 605–614.
- [413] Backhaus, K., Erichson, B., Plinke, W., Weiber, R., *Multivariate Analysemethoden: Eine anwendungsorientierte Einführung*, Springer Gabler, Berlin, Heidelberg 2016.
- [414] Bero, S. A., Muda, A. K., Choo, Y. H., Muda, N. A., Pratama, S. F., Similarity Measure for Molecular Structure: A Brief Review. *J. Phys.: Conf. Ser.* 2017, 892, 12015.
- [415] Bajusz, D., Rácz, A., Héberger, K., Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations? *J. Cheminformatics* 2015, 7, 20.
- [416] Willett, P., The Calculation of Molecular Structural Similarity: Principles and Practice. *Mol. Inform.* 2014, 33, 403–413.
- [417] Rácz, A., Bajusz, D., Héberger, K., Life beyond the Tanimoto coefficient: similarity measures for interaction fingerprints. *J. Cheminformatics* 2018, 10, 48.

- [418] Holliday, J. D., Salim, N., Whittle, M., Willett, P., Analysis and display of the size dependence of chemical similarity coefficients. *J. Chem. Inf. Comput. Sci.* 2003, *43*, 819–828.
- [419] Flower, D. R., On the Properties of Bit String-Based Measures of Chemical Similarity. *J. Chem. Inf. Comput. Sci.* 1998, *38*, 379–386.
- [420] Gimeno, A., Ojeda-Montes, M. J., Tomás-Hernández, S., Cereto-Massagué, A., Beltrán-Debón, R., Mulero, M., Pujadas, G., Garcia-Vallvé, S., The Light and Dark Sides of Virtual Screening: What Is There to Know? *Int. J. Mol. Sci.* 2019, *20*.
- [421] Śmieja, M., Warszycki, D., Average Information Content Maximization--A New Approach for Fingerprint Hybridization and Reduction. *PLoS ONE* 2016, *11*, e0146666.
- [422] Willett, P., Barnard, J. M., Downs, G. M., Chemical Similarity Searching. *J. Chem. Inf. Comput. Sci.* 1998, *38*, 983–996.
- [423] Abdoun, N., El Assad, S., Manh Hoang, T., Deforges, O., Assaf, R., Khalil, M., Designing Two Secure Keyed Hash Functions Based on Sponge Construction and the Chaotic Neural Network. *Entropy* 2020, *22*.
- [424] Hert, J., Willett, P., Wilton, D. J., Acklin, P., Azzaoui, K., Jacoby, E., Schuffenhauer, A., Comparison of topological descriptors for similarity-based virtual screening using multiple bioactive reference structures. *Org. Biomol. Chem.* 2004, *2*, 3256–3266.
- [425] Klein, K., Kriege, N., Mutzel, P., CT-index: Fingerprint-based graph indexing combining cycles and trees. *IEEE 27th International Conference on Data Engineering* 2011, 1115–1126.
- [426] He, W., Peng, X., Meng, X., Liu, X., Optical hierarchical authentication based on interference and hash function. *Appl. Opt.* 2012, *51*, 7750–7757.
- [427] Gao, K., Nguyen, D. D., Sresht, V., Mathiowetz, A. M., Tu, M., Wei, G.-W., Are 2D fingerprints still valuable for drug discovery? *Phys. Chem. Chem. Phys.* 2020, *22*, 8373–8390.
- [428] Schrödinger Release 2016-3, *Canvas*, Schrödinger, LLC, New York, NY 2016.
- [429] Sastry, M., Lowrie, J. F., Dixon, S. L., Sherman, W., Large-scale systematic analysis of 2D fingerprint methods and parameters to improve virtual screening enrichments. *J. Chem. Inf. Model.* 2010, *50*, 771–784.
- [430] Cereto-Massagué, A., Ojeda, M. J., Valls, C., Mulero, M., Garcia-Vallvé, S., Pujadas, G., Molecular fingerprint similarity search in virtual screening. *Methods* 2015, *71*, 58–63.
- [431] Huang, G., Li, J., Zhao, C., Computational Prediction and Analysis of Associations between Small Molecules and Binding-Associated S-Nitrosylation Sites. *Molecules* 2018, *23*.
- [432] Landrum, G., *RDKit: Open-source cheminformatics* 2018.

- [433] James, C. A., Weininger, D., Daylight theory manual 1995, Daylight Chemical Information Systems, Inc., Irvine, CA.
- [434] Deng, Z., Chuaqui, C., Singh, J., Structural interaction fingerprint (SIFt): a novel method for analyzing three-dimensional protein-ligand binding interactions. *J. Med. Chem.* 2004, *47*, 337–344.
- [435] Liu, W., Johnson, D. E., Clustering and its application in multi-target prediction. *Curr. Opin. Drug Discov. Devel.* 2009, *12*, 98–107.
- [436] Pham, D. T., Afify, A. A., Clustering techniques and their applications in engineering. *Proc. Inst. Mech. Eng. C* 2007, *221*, 1445–1459.
- [437] Yim, O., Ramdeen, K. T., Hierarchical Cluster Analysis: Comparison of Three Linkage Measures and Application to Psychological Data. *Quant. Meth. Psych.* 2015, *11*, 8–21.
- [438] Chopade, N., Sheetlani, J., Recent Trends in Incremental Clustering: A Review. *IOSR JCE* 2017, *19*, 19–24.
- [439] Thu Thuy, H. T., Anh, D. T., Ngoc Chau, V. T., Incremental Clustering for Time Series Data Based on an Improved Leader Algorithm. *IEEE-RIVF International Conference on Computing and Communication Technologies* 2019, 1–6.
- [440] Hartigan, J. A., *Clustering algorithms*, Wiley, New York 1975.
- [441] Hodes, L., Clustering a large number of compounds. 1. Establishing the method on an initial sample. *J. Chem. Inf. Comput. Sci.* 1989, *29*, 66–71.
- [442] Patibandla, R. S. M. L., Veeranjanyulu, N., Performance Analysis of Partition and Evolutionary Clustering Methods on Various Cluster Validation Criteria. *Arab. J. Sci. Eng.* 2018, *43*, 4379–4390.
- [443] Honkela, T. (Ed.), *Artificial neural networks and machine learning - ICANN 2011: 21st International Conference on Artificial Neural Networks, Espoo, Finland, June 14-17, 2011; proceedings*, Springer, Heidelberg 2011.
- [444] Singh, N., Shah, P., Dwivedi, H., Mishra, S., Tripathi, R., Sahasrabudde, A. A., Siddiqi, M. I., Integrated machine learning, molecular docking and 3D-QSAR based approach for identification of potential inhibitors of trypanosomal N-myristoyltransferase. *Mol. Biosyst.* 2016, *12*, 3711–3723.
- [445] Hansch, C., Fujita, T., ρ - σ - π Analysis. A Method for the Correlation of Biological Activity and Chemical Structure. *J. Am. Chem. Soc.* 1964, *86*, 1616–1626.
- [446] Free, S. M., Wilson, J. W., A Mathematical Contribution to Structure-Activity Studies. *J. Med. Chem.* 1964, *7*, 395–399.

- [447] Cramer, R. D., Patterson, D. E., Bunce, J. D., Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. *J. Am. Chem. Soc.* 1988, *110*, 5959–5967.
- [448] Klebe, G., Abraham, U., Mietzner, T., Molecular similarity indices in a comparative Analysis (CoMSIA) of drug molecules to correlate and predict their biological activity. *J. Med. Chem.* 1994, *37*, 4130–4146.
- [449] Acharya, C., Coop, A., Polli, J. E., MacKerell, A. D., Recent advances in ligand-based drug design: relevance and utility of the conformationally sampled pharmacophore approach. *Curr. Comput. Aided Drug Des.* 2011, *7*, 10–22.
- [450] Kunick, C., Lauenroth, K., Wieking, K., Xie, X., Schultz, C., Gussio, R., Zaharevitz, D. W., Leost, M., Meijer, L., Weber, A., Joergensen, F. S., Lemcke, T., Evaluation and Comparison of 3D-QSAR CoMSIA Models for CDK1, CDK5, and GSK-3 Inhibition by Paullones. *J. Med. Chem.* 2004, *47*, 22–36.
- [451] Cherkasov, A., Muratov, E. N., Fourches, D., Varnek, A., Baskin, I. I., Cronin, M., Dearden, J., Gramatica, P., Martin, Y. C., Todeschini, R., Consonni, V., Kuz'min, V. E., Cramer, R., Benigni, R., Yang, C., Rathman, J., Terfloth, L., Gasteiger, J., Richard, A., Tropsha, A., QSAR Modeling: Where Have You Been? Where Are You Going To? *J. Med. Chem.* 2013, *57*, 4977–5010.
- [452] Bajorath, J., Selected Concepts and Investigations in Compound Classification, Molecular Descriptor Analysis, and Virtual Screening. *J. Chem. Inf. Comput. Sci.* 2001, *41*, 233–245.
- [453] Bajot, F., The Use of Qsar and Computational Methods in Drug Design. In: Puzyn, T., Leszczynski, J., Cronin, M. (Hg.) *Recent advances in QSAR studies. Challenges and Advances in Computational Chemistry and Physics, Vol. 8.* Springer, Dordrecht. 2010, *8*, 261–282.
- [454] Consonni, V., Todeschini, R., Molecular Descriptors. In: Puzyn, T., Leszczynski, J., Cronin, M. (Hg.) *Recent advances in QSAR studies. Challenges and Advances in Computational Chemistry and Physics, Vol. 8.* Springer, Dordrecht. 2010, 29–102.
- [455] Taft, R. W., Linear Free Energy Relationships from Rates of Esterification and Hydrolysis of Aliphatic and Ortho-substituted Benzoate Esters. *J. Am. Chem. Soc.* 1952, *74*, 2729–2732.
- [456] Roy, K., Das, R. N., A review on principles, theory and practices of 2D-QSAR. *Curr. Drug Metab.* 2014, *15*, 346–379.
- [457] Kubinyi, H., QSAR and 3D QSAR in drug design Part 1: methodology. *Drug Discov. Today* 1997, *2*, 457–467.

- [458] Pastor, M., Cruciani, G., McLay, I., Pickett, S., Clementi, S., GRid-INdependent descriptors (GRIND): a novel class of alignment-independent three-dimensional molecular descriptors. *J. Med. Chem.* 2000, *43*, 3233–3243.
- [459] Todeschini, R., Gramatica, P., The Whim Theory: New 3D Molecular Descriptors for Qsar in Environmental Modelling. *SAR QSAR Environ. Res.* 1997, *7*, 89–115.
- [460] Livingstone, D. J., Manallack, D. T., Neural Networks in 3D QSAR. *QSAR Comb. Sci.* 2003, *22*, 510–518.
- [461] Karelson, M., Lobanov, V. S., Katritzky, A. R., Quantum-Chemical Descriptors in QSAR/QSPR Studies. *Chem. Rev.* 1996, *96*, 1027–1044.
- [462] Golbraikh, A., Shen, M., Xiao, Z., Xiao, Y.-D., Lee, K.-H., Tropsha, A., Rational selection of training and test sets for the development of validated QSAR models. *J. Comput. Aided Mol. Des.* 2003, *17*, 241–253.
- [463] Gramatica, P., Principles of QSAR Modeling. *Int. J. Quant. Struct.-Prop. Relat.* 2020, *5*, 61–97.
- [464] Sippl, W., 3D-QSAR – Applications, Recent Advances, and Limitations. In: *Puzyn, T., Leszczynski, J., Cronin, M. (Hg.) Recent advances in QSAR studies. Challenges and Advances in Computational Chemistry and Physics, Vol. 8. Springer, Dordrecht.* 2010, 103–125.
- [465] Kearsley, S. K., Smith, G. M., An alternative method for the alignment of molecular structures: Maximizing electrostatic and steric overlap. *Tetrahedron Comput. Methodol.* 1990, *3*, 615–633.
- [466] Mittal, R. R., McKinnon, R. A., Sorich, M. J., Effect of steric molecular field settings on CoMFA predictivity. *J. Mol. Model.* 2008, *14*, 59–67.
- [467] Saxena, A. K., Prathipati, P., Comparison of MLR, PLS and GA-MLR in QSAR analysis. *SAR QSAR Environ. Res.* 2003, *14*, 433–445.
- [468] Zhao, M., Wang, L., Zheng, L., Zhang, M., Qiu, C., Zhang, Y., Du, D., Niu, B., 2D-QSAR and 3D-QSAR Analyses for EGFR Inhibitors. *Biomed Res. Int.* 2017, *2017*, 4649191.
- [469] Sutherland, J. J., O'Brien, L. A., Weaver, D. F., A comparison of methods for modeling quantitative structure-activity relationships. *J. Med. Chem.* 2004, *47*, 5541–5554.
- [470] Dixon, S. L., Smondyrev, A. M., Knoll, E. H., Rao, S. N., Shaw, D. E., Friesner, R. A., PHASE: a new engine for pharmacophore perception, 3D QSAR model development, and 3D database screening: 1. Methodology and preliminary results. *J. Comput. Aided Mol. Des.* 2006, *20*, 647–671.
- [471] Schrödinger Release 2016-3, *Phase*, Schrödinger, LLC, New York, NY 2016.

- [472] Dixon, S. L., Smondyrev, A. M., Rao, S. N., PHASE: a novel approach to pharmacophore modeling and 3D database searching. *Chem. Biol. Drug Des.* 2006, *67*, 370–372.
- [473] Ghose, A. K., Viswanadhan, V. N., Wendoloski, J. J., Prediction of Hydrophobic (Lipophilic) Properties of Small Organic Molecules Using Fragmental Methods: An Analysis of ALOGP and CLOGP Methods. *J. Phys. Chem. A* 1998, *102*, 3762–3772.
- [474] Schrödinger Release 2020-1, *Field-Based QSAR: User Manual*, Schrödinger, LLC, New York, NY 2020.
- [475] Tobias, R. D., *An Introduction to Partial Least Squares Regression*, SAS Institute, Cary, NC 1995.
- [476] Stanton, D. T., QSAR and QSPR model interpretation using partial least squares (PLS) analysis. *Curr. Comput. Aided Drug Des.* 2012, *8*, 107–127.
- [477] Alexander, D. L. J., Tropsha, A., Winkler, D. A., Beware of R²: Simple, Unambiguous Assessment of the Prediction Accuracy of QSAR and QSPR Models. *J. Chem. Inf. Model.* 2015, *55*, 1316–1322.
- [478] Sahu, N. K., Shahi, S., Sharma, M. C., Kohli, D. V., QSAR studies on imidazopyridazine derivatives as PfPK7 inhibitors. *Mol. Simul.* 2011, *37*, 752–765.
- [479] Gramatica, P., Principles of QSAR models validation: Internal and external. *QSAR Comb. Sci.* 2007, *26*, 694–701.
- [480] Tropsha, A., Best Practices for QSAR Model Development, Validation, and Exploitation. *Mol. Inform.* 2010, *29*, 476–488.
- [481] González, M. P., Terán, C., Saíz-Urra, L., Teijeira, M., Variable selection methods in QSAR: an overview. *Curr. Top. Med. Chem.* 2008, *8*, 1606–1627.
- [482] Golbraikh, A., Tropsha, A., Beware of q²! *J. Mol. Graph. Model.* 2002, *20*, 269–276.
- [483] Verma, J., Khedkar, V. M., Coutinho, E. C., 3D-QSAR in drug design - a review. *Curr. Top. Med. Chem.* 2010, *10*, 95–115.
- [484] Durrant, J. D., de Oliveira, César Augusto F., McCammon, J. A., POVME: An algorithm for measuring binding-pocket volumes. *J. Mol. Graph. Model.* 2011, *29*, 773–776.
- [485] Wagner, J. R., Sørensen, J., Hensley, N., Wong, C., Zhu, C., Perison, T., Amaro, R. E., POVME 3.0: Software for Mapping Binding Pocket Flexibility. *J. Chem. Theory Comput.* 2017, *13*, 4584–4592.
- [486] Durrant, J. D., Votapka, L., Sørensen, J., Amaro, R. E., POVME 2.0: An Enhanced Tool for Determining Pocket Shape and Volume Characteristics. *J. Chem. Theory Comput.* 2014, *10*, 5047–5056.

- [487] Humphrey, W., Dalke, A., Schulten, K., VMD: Visual molecular dynamics. *J. Mol. Graph. Model.* 1996, 14, 33–38.
- [488] Berthold, M. R., Cebron, N., Dill, F., Di Fatta, G., Gabriel, T. R., Georg, F., Meinl, T., Ohl, P., Sieb, C., Wiswedel, B., KNIME: the Konstanz Information Miner. *In: Workshop on Multi-Agent Systems and Simulation (MAS&S), 4th Annual Industrial Simulation Conference (ISC), 05.-07.06.2006, Palermo, Italien.* 2006, 58–61.
- [489] R Core Team, *R: A language and environment for statistical computing.: R Foundation for Statistical Computing.* URL: <https://www.R-project.org> (R version 3.6.0), Wien, Österreich 2019.
- [490] Ritz, C., Streibig, J. C., Bioassay Analysis Using R. *J. Stat. Softw.* 2005, 12, 1–22.
- [491] Glickman, J. F., Assay Development for Protein Kinase Enzymes (01.05.2012, Update 01.10.2012). *In: Markossian, S., Grossman, A., Brimacombe, K., Arkin, M., Auld, D., Austin, C., Baell, J., Chung, T.D.Y., Coussens, N.P., Dahlin, J.L., Devanarayan, V., Foley, T.L., Glicksman, M., Haas, J.V., Hall, M.D., Hoare, S., Inglese, J., Iversen, P.W., Kales, S.C., Lal-Nag, M., Li, Z., McGee, J., McManus, O., Riss, T., Saradjian, P., Sittampalam, G.S., Tarselli, M., Trask, O.J. Jr., Wang, Y., Weidner, J.R., Wildey, M.J., Wilson, K., Xia, M., Xu, X. (Hg.) Assay Guidance Manual [Internet]. Bethesda (MD): Eli Lilly & Company and the National Center for Advancing Translational Sciences; 2004.*
- [492] Müller-Esterl, W., *Biochemie*, Springer, Berlin, Heidelberg 2018.
- [493] Ma, H., Deacon, S., Horiuchi, K., The challenge of selecting protein kinase assays for lead discovery optimization. *Expert Opin. Drug Discov.* 2008, 3, 607–621.
- [494] Li, H., Totoritis, R. D., Lor, L. A., Schwartz, B., Caprioli, P., Jurewicz, A. J., Zhang, G., Evaluation of an antibody-free ADP detection assay: ADP-Glo. *Assay Drug Dev. Technol.* 2009, 7, 598–605.
- [495] Wienken, C. J., Baaske, P., Rothbauer, U., Braun, D., Duhr, S., Protein-binding assays in biological liquids using microscale thermophoresis. *Nat. Commun.* 2010, 1, 100.
- [496] Fabian, M. A., Biggs, W., Treiber, D. K., Atteridge, C. E., Azimioara, M. D., Benedetti, M. G., Carter, T. A., Ciceri, P., Edeen, P. T., Floyd, M., Ford, J. M., Galvin, M., Gerlach, J. L., Grotzfeld, R. M., Herrgard, S., Insko, D. E., Insko, M. A., Lai, A. G., Lelias, J.-M., Mehta, S. A., Milanov, Z. V., Velasco, A. M., Wodicka, L. M., Patel, H. K., Zarrinkar, P. P., Lockhart, D. J., A small molecule-kinase interaction map for clinical kinase inhibitors. *Nat. Biotechnol.* 2005, 23, 329–336.

- [497] Discoverx.com, *DiscoverX KINOMEScan Kinase Screening & Profiling Services* -, <https://www.discoverx.com/services/drug-discovery-development-services/kinase-profiling/kinomescan> (accessed January 26, 2021).
- [498] Duque Escobar, J. E., *Role of the dual leucine zipper kinase DLK in the development of diabetes mellitus type 2*. Dissertation, Hamburg 2017.
- [499] Discoverx.com, *KINOMEScan® DLK*, <https://www.discoverx.com/kinase-data-sheets/dlk> (accessed January 26, 2021).
- [500] Kim, S., Chen, J., Cheng, T., Gindulyte, A., He, J., He, S., Li, Q., Shoemaker, B. A., Thiessen, P. A., Yu, B., Zaslavsky, L., Zhang, J., Bolton, E. E., PubChem in 2021: new data content and improved web interfaces. *Nucleic Acids Res.* 2021, 49, D1388-D1395.
- [501] Laemmli, U. K., Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. *Nature* 1970, 227, 680–685.
- [502] Kashem, M. A., Nelson, R. M., Yingling, J. D., Pullen, S. S., Prokopowicz, A. S., Jones, J. W., Wolak, J. P., Rogers, G. R., Morelock, M. M., Snow, R. J., Homon, C. A., Jakes, S., Three mechanistically distinct kinase assays compared: Measurement of intrinsic ATPase activity identified the most comprehensive set of ITK inhibitors. *J. Biomol. Screen.* 2007, 12, 70–83.
- [503] Promega Corporation, Technical Manual: ADP-Glo™ Kinase Assay 2015, Madison, USA.
- [504] Hennek, J., Alves, J., Yao, E., Goueli, S. A., Zegzouti, H., Bioluminescent kinase strips: A novel approach to targeted and flexible kinase inhibitor profiling. *Anal. Biochem.* 2016, 495, 9–20.
- [505] Hill, A., The possible effects of the aggregation of the molecules of haemoglobin on its dissociation curves. *J. Physiol.* 1910, 40, iv–vii.
- [506] Gottschalk, P. G., Dunn, J. R., The five-parameter logistic: A characterization and comparison with the four-parameter logistic. *Anal. Biochem.* 2005, 343, 54–65.
- [507] Findlay, J. W. A., Dillard, R. F., Appropriate calibration curve fitting in ligand binding assays. *AAPS J.* 2007, 9, E260-7.
- [508] NanoTemper Technologies GmbH, User Manual Monolith® NT.115, http://www.isbg.fr/IMG/pdf/manual_nt115_new.pdf (accessed October 18, 2021).
- [509] Rainard, J. M., Pandarakalam, G. C., McElroy, S. P., Using Microscale Thermophoresis to Characterize Hits from High-Throughput Screening: A European Lead Factory Perspective. *SLAS Discov.* 2018, 23, 225–241.
- [510] Gudim, I., Lofstad, M., Hammerstad, M., Hersleth, H.-P., Measurement of FNR-NrdI Interaction by Microscale Thermophoresis (MST). *Bio Protoc.* 2017, 7.

- [511] NanoTemper Technologies GmbH, MST Starting Guide - Monolith NT.115, https://physiology.case.edu/media/eq_manuals/eq_manual_nano-temper_mst_starting_guide.pdf (accessed October 18, 2021).
- [512] NanoTemper Technologies GmbH, *Monolith NT™ Protein Labeling Kit RED-NHS: User Manual*.
- [513] Jerabek-Willemsen, M., André, T., Wanner, R., Roth, H. M., Duhr, S., Baaske, P., Breitsprecher, D., MicroScale Thermophoresis: Interaction analysis and beyond. *J. Mol. Struct.* 2014, *1077*, 101–113.
- [514] DiscoverX, KINOMEScan™ Profiling Service: Primary Screen Report (scanELECT, DLK) 2019.
- [515] DiscoverX, KINOMEScan™ Profiling Service: Kd Report 2020.
- [516] Feixas, F., Lindert, S., Sinko, W., McCammon, J. A., Exploring the role of receptor flexibility in structure-based drug discovery. *Biophys. Chem.* 2014, *186*, 31–45.
- [517] Collie, G. W., Koh, C. M., O'Neill, D. J., Stubbs, C. J., Khurana, P., Eddershaw, A., Snijder, A., Mauritzson, F., Barlind, L., Dale, I. L., Shaw, J., Phillips, C., Hennessy, E. J., Cheung, T., Narvaez, A. J., Structural and Molecular Insight into Resistance Mechanisms of First Generation cMET Inhibitors. *ACS Med. Chem. Lett.* 2019, *10*, 1322–1327.
- [518] Qian, F., Engst, S., Yamaguchi, K., Yu, P., Won, K.-A., Mock, L., Lou, T., Tan, J., Li, C., Tam, D., Loughheed, J., Yakes, F. M., Bentzien, F., Xu, W., Zaks, T., Wooster, R., Greshock, J., Joly, A. H., Inhibition of tumor cell growth, invasion, and metastasis by EXEL-2880 (XL880, GSK1363089), a novel inhibitor of HGF and VEGF receptor tyrosine kinases. *Cancer Res.* 2009, *69*, 8009–8016.
- [519] Modi, S. J., Kulkarni, V. M., Vascular Endothelial Growth Factor Receptor (VEGFR-2)/KDR Inhibitors: Medicinal Chemistry Perspective. *Med. Drug Discov.* 2019, *2*, 100009.
- [520] Akeno-Stuart, N., Croyle, M., Knauf, J. A., Malaguarnera, R., Vitagliano, D., Santoro, M., Stephan, C., Grosios, K., Wartmann, M., Cozens, R., Caravatti, G., Fabbro, D., Lane, H. A., Fagin, J. A., The RET kinase inhibitor NVP-AST487 blocks growth and calcitonin gene expression through distinct mechanisms in medullary thyroid cancer cells. *Cancer Res.* 2007, *67*, 6956–6964.
- [521] Wodicka, L. M., Ciceri, P., Davis, M. I., Hunt, J. P., Floyd, M., Salerno, S., Hua, X. H., Ford, J. M., Armstrong, R. C., Zarrinkar, P. P., Treiber, D. K., Activation state-dependent binding of small molecule kinase inhibitors: structural insights from biochemistry. *Chem. Biol.* 2010, *17*, 1241–1249.

- [522] Meng, S., Wu, H., Wang, J., Qiu, Q., Systematic Analysis of Tyrosine Kinase Inhibitor Response to RET Gatekeeper Mutations in Thyroid Cancer. *Mol. Inform.* 2016, *35*, 495–505.
- [523] Rarey, M., Kramer, B., Lengauer, T., Multiple automatic base selection: Protein-ligand docking based on incremental construction without manual intervention. *J. Comput. Aided Mol. Des.* 1997, *11*, 369–384.
- [524] Chuaqui, C., Deng, Z., Singh, J., Interaction profiles of protein kinase-inhibitor complexes and their application to virtual screening. *J. Med. Chem.* 2005, *48*, 121–133.
- [525] Singh, J., Deng, Z., Narale, G., Chuaqui, C., Structural interaction fingerprints: a new approach to organizing, mining, analyzing, and designing protein-small molecule complexes. *Chem. Biol. Drug Des.* 2006, *67*, 5–12.
- [526] Duque Escobar, J. E. Persönliche Kommunikation 2017, Universitätsklinikum Eppendorf, Arbeitskreis Prof. Dr. Oetjen, Hamburg.
- [527] Schrödinger Release 2016-2, *Maestro*, Schrödinger, LLC, New York, NY 2016.
- [528] Schrödinger Release 2020-1, *Maestro*, Schrödinger, LLC, New York, NY 2020.
- [529] Schrödinger Release 2016-2, *Desmond Molecular Dynamics System*, D. E. Shaw Research, New York, NY 2016.
- [530] Schrödinger Release 2016-3, *Desmond Molecular Dynamics System: D. E. Shaw Research, New York, NY, 2016*, Schrödinger, New York, NY 2016.
- [531] Liao, Q., Wang, J., Watson, I. A., Accelerating Two Algorithms for Large-Scale Compound Selection on GPUs. *J. Chem. Inf. Model.* 2011, *51*, 1017–1024.

Anhang

A. Übersicht über die verwendeten Knime-Nodes und deren Funktion

1. Einlese-Nodes:

1a) CSV-Reader: Einlesen von CSV-Dateien (in dieser Arbeit: Tabellen mit Ligandendaten)

1b) Molecule-Reader (to MAE): Einlesen von Moleküldateien im MAESTRO-, SD-, Mol-, Mol2-, SMILES- oder PDB-Format und Konvertierung in das MAESTRO-Format (*Node* aus MAESTRO-Plug-In); wurde in dieser Arbeit zum Einlesen von Liganden zur Generierung der RDKit-Fingerprints verwendet.

2. Datenverarbeitungs-Nodes:

2a) Column-Filter, Column-Resorter, Sorter, Column-Rename, String Replacer: Der *Column-Filter* lässt nur die gewünschten Spalten durch. Mit dem *Column-Resorter* ordnet man die Spalten nebeneinander in einer bestimmten Reihenfolge an, während man im *Sorter-Node* eine Spalte auswählt, nach der die gesamte Tabelle auf- oder absteigend sortiert werden soll. Mit *Column-Rename* werden Spaltenüberschriften umbenannt. Mit dem *String-Replacer-Node* wird nach einem Pattern in allen Zellen einer definierten Spalte gesucht, der mit einem anderen, definierten String ersetzt werden soll.

2b) Joiner, Concatenate: Mit dem *Joiner* können zwei Tabellen vereint werden, indem man ein Merkmal auswählt (Spalte, z.B. Ligandename), in dem die zu vereinigenden Tabellen übereinstimmen sollen. Die Daten aus den beiden Tabellen zu einer bestimmten gemeinsamen Eigenschaft werden dann in Spalten nebeneinander ausgegeben. Dies ist z.B. hilfreich im Zusammenfügen verschiedener Scorewerte zu einem Ligandeneintrag. Im Gegensatz zum *Joiner* hängt der *Concatenate-Node* zwei Tabellen aneinander. In dieser Arbeit wurde dieser *Node* für das Aneinanderhängen von Ranglisten aus verschiedenen Grids genutzt, um in weiteren Teilschritten gleiche Ligandeneinträge zu erkennen und herauszufiltern.

2c) RowID, Row-Filter, Rule-based Row-Filter, Rule Engine: Mit dem *RowID-Node* kann man die RowID (eine Nummerierung der Tabelleneinträge) zu einer Spalte konvertieren, mit deren Inhalt man alle gewöhnlichen Operationen ausführen kann. Mit dem *Row-Filter* können Reihen (z.B. Ligandeneinträge mit bestimmten Eigenschaften, wie „schlechten“ Score-Werten oder fehlenden Daten) gelöscht oder gefiltert werden. Ähnlich zum Row-Filter können mit dem *Rule-based Row-Filter* Einschluss- und/oder Ausschlussregeln festgelegt werden, anhand derer jede Zeile einer Input-Tabelle abgeglichen wird. Mit dem *Rule-Engine-Node* werden benutzerdefinierte Regeln mit dem Tabelleninhalt abgeglichen und bei Übereinstimmung eine definierte Ausgabe in einer neuen Zeile hinzugefügt. Dieser *Node* wurde vor allem dazu verwendet, eingelesene Ranglisten mit der dazugehörigen Grid-Nummerierung zu versehen und um Ligandeneinträge mit pK_i -Werten > 0 als „aktiv“ (z.B. *true*) zu bezeichnen und im Folgenden als Aktive zu erkennen.

2d) Math Formula, Java Snippet (Simple), Missing Value: Der *Math-Formula-Node* kann für alle mathematischen Operationen verwendet werden, in dieser Arbeit z.B. zur Berechnung der Gesamtzahl des Datensatzes, der Aktiven, den Anteil der Aktiven, zur Berechnung des Consensuscores und für die Normierung der Scorewerte. Mit dem *Java-Snippet (Simple)-Node* können beliebige Java-Codes ausgeführt werden. Dieser *Node* wurde in dieser Arbeit für mathematische Operationen genutzt, z.B. um die kumulative Summe der Aktiven entlang der Spalteneinträge zu berechnen. Mit dem *Missing-Value-Node* kann definiert werden, wie man fehlende Werte in Tabellenzellen behandelt. In dieser Arbeit wurden mit diesem *Node* fehlende pK_i -Werte (z.B. bei Decoys) auf „0“ gesetzt, um sie von den Aktiven zu unterscheiden.

2e) GroupBy: Der *GroupBy-Node* ermöglicht die Gruppierung von Tabellenreihen bei Übereinstimmung des Tabelleninhalts einer definierten Spalte (z.B. Ligandename). Dieser Node wurde in dieser Arbeit als zentrales Element für die Zusammenfassung von Ranglisten aus verschiedenen Grids eingesetzt, um bei doppelt vorkommenden Liganden nach dem Zusammenfügen der Listen nur denjenigen mit dem besten Scorewert zu behalten. Dies bedingt eine vorherige Sortierung der Liste nach dem gewünschten Scorewert mit dem *Sorter-Node*.

2f) Extract Properties, MAE-to-SD (beide aus dem *MAESTRO-Plug-In*): Der *Extract-Properties-Node* wird um Filtern der erforderlichen Informationen aus einer Moleküldatei (z.B. mae-Datei) verwendet. Der *MAE-to-SD-Node* dient der Konvertierung vom MAE- zum SD-Molekülformat. Die Konvertierung ist nötig, um den *RDKit-Fingerprint-Node* zu verwenden.

2g) RDKit Fingerprint: Zur Berechnung von bitbasierten Hash-Fingerprints. Es stehen verschiedene Fingerprint-Typen zur Auswahl. In dieser Arbeit wurde „RDKit“ ausgewählt für die Generierung Daylight-ähnlicher topologischer Fingerprints im Postprocessing.

2h) Linear / Rank Correlation: Zur Berechnung der linearen Korrelation (Pearson-Korrelationskoeffizient) bzw. der Rangkorrelationskoeffizienten (Auswahl möglich zwischen Spearman's Rho, Kendall's Tau-a bzw. -b und Goodman and Kruskal's Gamma).

3. Ausgabe-Nodes:

3a) Table Viewer: Mit dem *Table-Viewer* des *MAESTRO-Plug-Ins* lassen sich die verarbeiteten Tabellen nach den gewünschten Teilschritten anzeigen. Dieser *Node* ist besonders nützlich bei der Erstellung von neuen Workflows um zu überprüfen, ob die *Nodes* mit den eingegebenen Einstellungen die Operationen wie erwartet ausführen und der Workflow zum gewünschten Ergebnis führt.

3b) CSV-Writer: zur Abspeicherung von Daten / einer Datentabelle in einem gewünschten Verzeichnis als .csv- oder .txt-File.

3c) RDKit Fingerprint Writer: Dieser Node speichert eine Fingerprint-Datei (FPS-Format) ab. In dieser Arbeit wurden so die mit dem *RDKit-Fingerprint-Node* generierten *RDKit-Fingerprints* der Liganden aus dem Postprocessing abgespeichert, um sie im weiteren Verlauf für das CPU-Leader-Clustering zu verwenden.

B. Zusammensetzung der Ligandenliste c

s. Tab. B 1, S. 343 ff.

F, T, K, PC, CH, PH, EL, LL, CD = Quellenangaben für die Strukturen, siehe Kap. 7.11, S. 281.

C. QSAR-Selektionen

s. Tab. C 1, Tab. C 2 u. Tab. C 3, S. 347 ff.

„+“ = Training, „-“ = Test

p = protoniert am Stickstoffatom des Rests R3, **pm** = manuell protoniert und minimiert

p* = zwei Protonierungsmöglichkeiten an R3 → nur Stickstoffatom protoniert, das die meisten Substituenten hat bzw. näher an Core 2 liegt

Sp = Split, **Pred** = vorhergesagter pK_i -Wert mit dem QSAR-Modell der rot markierten Selektion

F, T, K, PC, CH, PH, EL, LL, CD = Quellenangaben für die Strukturen, siehe Kap. 7.11, S. 281.

Struktur	Bez.	pIC ₅₀
s. S. 152, Abb. 5-41	F_CEP-1347	6,946
	F_CEP-5214	6,924
	F_(+)-K252a	6,444
	T_1	6,928
	T_15	7,149
	T_16a	8,699
	T_16b	8,301
	T_16c	8,699
	T_16d	9,000
	T_16e	8,155
T_16f	8,347	
T_16g	8,301	
Struktur	Bez.	pK _d
s. S. 153, Abb. 5-42	F_AG-013736	5,770
	F_KW-2449	7,174
	K_SU-14813	6,252
	K_Sunitinib	7,000
s. S. 153, Abb. 5-43	K_EKB-569	5,585
	F_HKI-272	6,432
	F_SKI-606	6,174
F_XL-880	7,658	
s. S. 154, Abb. 5-44	F_AST-487	5,602
	K_CHIR-258/TKI-258	5,678
	F_Crizotinib	6,770
	F_JNJ-28312141	6,482
	K_JNJ-7706621	5,284
	F_TAE-684	7,377
F_TG-101348	5,201	
F_VX-680/MK-0457	6,721	
s. S. 152, Abb. 5-41	F_CEP-701	6,076
	K_Staurosporin	5,824

Struktur	R3	R2	Bez.	pK _i
			PC_021	7,602
			PC_030	9,699
			CH_028	8,854
			CH_031	9,770
			CH_032	6,658
			CH_033	9,770
			CH_046	7,886
			CH_047	7,398
			CH_063	8,066
			CH_075	7,886
			CH_077	6,721
			CH_079	7,456
			CH_080	7,237
			CH_087	7,292
			CH_089	8,097
			CH_090	8,721
			CH_103	7,796
			CH_104	7,900
			CH_108	7,032
			CH_112	6,658
		CH_124	6,149	
			PC_015	8,699
			PC_018	6,780
			PH_010	6,260
			PH_015	6,876
			PH_016	7,319

Struktur	R3	R2	Bez.	pK _i
			PH_007	8,155
			PH_019	6,245
			CH_027	7,000
			CH_048	7,678
			CH_117	9,027
			CH_122	9,398
			CH_137	9,770
			CH_144	7,215
			CH_145	5,444
			CH_146	5,678
			EL_039	6,337
			EL_071	6,569
			EL_105	7,018
			EL_111	6,745
			EL_123	7,569
		EL_136	7,620	

Tab. B 1 – Liganden der Ligandenliste c; Bez. = Bezeichnung. Abkürzungen F, T, K, PC, CH, PH, EL, LL, CD = Quellenangaben für die Strukturen, siehe Kap. 7.11, S. 281.

Struktur	R2	R3	Bez.	pK _i
			EL_004	6,108
			EL_034	6,796
			EL_060	7,569
			EL_076	6,824
			EL_104	7,041
			EL_107	8,620
			EL_134	8,569
			EL_152	9,770
			EL_174	8,796
			EL_188	8,553
			LL_053	8,326
			LL_056	8,318
			LL_147	8,244
			LL_201	8,886
			LL_084	8,775
			LL_085	7,321
			LL_131	8,456
			LL_160	8,620
			LL_216	8,180
			LL_267	6,658
			EL_001	7,469
			EL_038	7,237
			EL_054	9,319
			EL_070	8,638

Tab. B 1 – Fortsetzung

Struktur	R2	R3	Bez.	pK _i
			EL_009	7,119
			EL_095	5,921
			EL_143	7,131
			EL_106	6,886
			EL_132	7,602
			LL_025	7,151
			LL_031	7,100
			LL_132	6,588
			EL_114	6,018

Struktur	R2	R3	Bez.	pK _i
			CD_005	6,081
			CD_014	6,367
			CD_030	6,092
			CD_037	6,658
			CD_039	7,155
			CD_047	7,155
			CD_064	6,569
			CD_065	7,398
			CD_069	6,959
			CD_083	6,770
			CD_086	6,678
			CD_095	6,167
			CD_099	7,367
			CD_109	6,319
			CD_110	7,699
			CD_117	6,149
			CD_087	6,678
			CD_071	7,046
			CD_108	6,495
			CD_116	5,793
			CD_114	6,161

Struktur	R2	R3	Bez.	pK _i
			CD_013	6,699
			CD_040	8,398
			CD_072	7,398
			CD_076	8,097
			CD_092	8,699
			CD_100	9,000
			CD_126	8,523
			CD_140	8,303
			CD_148	7,830
			CD_159	8,638
			CD_163	8,302
			CD_165	6,674
			CD_172	7,064
			CD_173	7,173
			CD_185	7,772
			CD_195	8,228
			CD_202	7,745
			CD_203	6,369
			CD_208	8,180
			CD_221	7,770
			CD_228	9,114
			CD_266	7,917
			CD_268	6,854
			CD_270	8,077
			CD_275	7,220
			CD_279	8,349
			CD_287	7,213

Struktur	R2	R3	Bez.	pK _i
			CD_297	8,203
			CD_312	7,726
			CD_349	7,551
			CD_355	9,241
			CD_373	7,413
			CD_384	7,762
			CD_402	8,509
			CD_403	7,219
			CD_404	8,194
			CD_463	6,733
			CD_477	8,034
			CD_493	8,208
			CD_494	8,099
				CD_214
			CD_214	7,182
			CD_079	7,222
			CD_138	7,006
			CD_179	7,943
			CD_189	7,690
			CD_191	8,374
			CD_200	7,943
			CD_252	7,101
			CD_261	7,496
			CD_272	8,903
			CD_288	6,493

Struktur	R2	Bez.	pK _i
		EH_015	6,444
		EH_016	6,495
		EH_018	6,396
		EH_029	5,185
		EH_031	6,373
		EH_064	6,573
		EH_076	7,011
		EH_085	6,243
		EH_091	5,854
		EH_095	6,322
		EH_113	6,283
		EH_122	6,513
		EH_123	7,221
		EH_032	7,883
		EH_074	7,587
		EH_087	7,854
		EH_129	8,785

Tab. B 1 – Fortsetzung

Struktur	R2	Bez.	pK _i
		EH_131	7,217
		EH_038	6,553
		EH_083	6,304
		EH_048	6,003
		EH_078	6,253
		EH_090	8,223
		EH_096	7,377
		EH_115	6,646

Struktur	R2	Bez.	pK _i
		EH_126	6,578
		EH_132	6,551
		EH_133	6,350
		EH_139	6,827
		EH_153	6,583
		EH_102	6,804
		EH_108	6,622

Struktur	R2	Bez.	pK _i
		EH_136	6,163
		EH_148	7,245
		EH_127	6,824
		5CEO-Ligand	> 9,301
		5CEP-Ligand	6.839
		5CEQ-Ligand	7,377

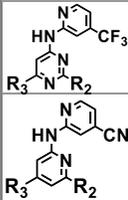
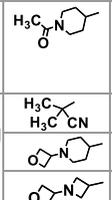
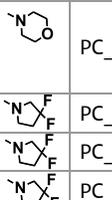
Tab. B 1 – Fortsetzung

Bez.	Sp1	Sp2	Sp3	pK _{i/d}	Pred	Bez.	Sp1	Sp2	Sp3	pK _i	Pred	Bez.	Sp1	Sp2	Sp3	pK _i	Pred	Bez.	Sp1	Sp2	Sp3	pK _i	Pred
K_CHIR-258/TKI-258p	+	+	+	5,678	6,078	CD_082	+	+	+	7,469	6,369	CD_297	+	+	+	8,203	8,118	CH_068p	+	-	+	9,523	8,865
K_JNJ-7706621	+	-	+	5,284	4,898	CD_084p	+	+	+	7,292	6,733	CD_344	+	+	+	7,232	8,033	CH_071p	+	+	+	9,444	8,989
F_CEP-701	+	+	+	6,076	5,642	CD_095	+	-	+	6,167	5,506	CD_34	+	-	+	7,551	8,149	CH_072p	+	+	+	6,824	7,131
K_Staurosporin	+	+	+	5,824	6,202	CD_101	-	+	-	7,585	7,115	CD_355	+	+	+	9,241	8,595	CH_074	+	-	+	9,699	9,248
CD_001	+	+	+	5,192	4,949	CD_102	+	+	+	7,620	7,269	CD_367	+	+	+	7,268	7,660	CH_075	+	+	+	7,886	8,232
CD_002	+	+	+	5,770	5,311	CD_109	+	+	+	6,319	6,189	CD_381	+	+	-	8,020	8,247	CH_078	+	-	+	8,509	7,065
CD_003	+	+	+	5,192	5,434	CD_110	+	+	+	7,699	6,779	CD_388	-	+	+	6,845	7,544	CH_092p	-	+	+	9,155	8,577
CD_004	+	+	+	5,762	5,514	CD_113p	-	+	+	6,131	6,622	CD_392	-	+	+	7,437	7,682	CH_095p	-	+	+	8,137	9,090
CD_005	+	-	-	6,081	6,970	CD_114p	+	+	-	6,161	7,050	CD_40	-	+	+	7,219	7,871	CH_097p	+	+	+	9,222	8,805
CD_008	+	-	+	6,658	6,272	CD_115p	+	+	-	6,310	6,439	CD_427	+	+	+	8,638	8,429	CH_107	+	+	+	6,921	7,284
CD_016p	+	+	+	9,00	8,454	CD_116p	+	+	+	5,793	6,075	CD_42	-	+	+	8,015	8,331	CH_113	-	+	+	5,796	6,799
CD_017p	+	+	+	9,00	8,790	CD_117	-	+	+	6,149	6,216	CD_438	+	+	+	7,921	8,304	CH_115p	-	+	+	10,046	9,523
CD_021p	+	+	+	8,155	8,290	CD_128	+	+	+	8,097	8,294	CD_461	-	+	+	8,943	8,018	CH_116	-	+	+	6,668	7,750
CD_024	+	+	+	5,793	5,967	CD_133	+	+	+	7,097	7,419	CD_46	+	+	+	8,241	8,164	CH_121p	+	+	+	9,699	9,330
CD_025	+	-	+	5,963	6,370	CD_136p	+	-	+	9,292	8,397	CD_477	-	-	+	8,034	8,201	CH_122	+	-	+	9,398	9,387
CD_030	+	+	+	6,092	6,574	CD_137p	+	-	+	7,921	8,385	CD_49	+	+	+	8,334	7,955	CH_123	-	+	+	9,699	8,745
CD_033	+	+	+	5,971	6,182	CD_138	+	+	+	7,006	7,371	CH_019	+	+	+	8,292	9,280	CH_128	+	-	+	6,347	5,975
CD_036p	+	+	+	6,658	7,140	CD_139	+	+	+	7,788	7,925	CH_023	-	+	-	9,495	9,189	CH_129	-	+	-	6,292	7,267
CD_038p	+	+	+	6,538	6,854	CD_154p	+	-	+	8,164	7,476	CH_02	+	+	+	6,854	6,830	CH_131p	-	+	+	9,357	8,041
CD_042	-	+	+	6,770	6,101	CD_157	+	+	+	9,509	8,738	CH_02	+	+	+	7,745	7,960	CH_135p	+	+	+	7,678	8,905
CD_044	+	+	+	7,155	6,710	CD_166	+	-	+	7,335	6,951	CH_03	+	+	-	9,770	10,074	CH_140p	+	+	-	9,770	9,458
CD_045p	-	+	+	5,757	5,648	CD_194	+	+	+	7,924	7,185	CH_031	+	+	+	9,770	9,322	CH_145	+	+	-	5,444	6,323
CD_046p	+	-	-	5,793	6,158	CD_201p	+	+	+	6,793	7,061	CH_032	+	+	+	6,658	6,673	CH_146p	+	-	+	5,678	5,868
CD_047	+	+	+	7,155	6,738	CD_203	+	+	+	6,369	6,650	CH_033	+	-	+	9,770	9,252	CH_147p	+	+	-	5,921	6,021
CD_048p	+	+	+	5,793	5,949	CD_214	+	-	+	7,182	7,168	CH_03	+	+	+	9,770	9,295	EL_004	+	+	+	6,108	6,772
CD_051	+	+	+	6,69	6,710	CD_228p	+	+	-	9,114	7,907	CH_03	+	+	+	9,770	9,488	EL_006	+	-	-	6,854	7,056
CD_052	+	-	+	7,066	6,136	CD_231p	+	+	+	7,613	8,191	CH_04	-	+	+	9,770	9,380	EL_008	+	+	+	8,000	8,142
CD_064	+	+	+	6,569	6,870	CD_254	+	+	+	7,770	7,467	CH_04	+	+	+	8,194	8,440	EL_009	+	-	+	7,119	7,251
CD_067	+	+	+	6,187	5,714	CD_261p	+	+	+	7,496	7,636	CH_04	+	+	-	8,167	8,529	EL_040p	+	+	+	9,770	8,706
CD_068p	-	+	+	5,793	6,193	CD_270	+	+	+	8,077	7,238	CH_052	+	+	+	8,699	8,651	EL_051	+	+	+	9,770	8,667
CD_069	-	+	-	6,959	6,749	CD_295	+	+	+	7,187	6,626	CH_05	-	-	+	8,824	8,089	EL_054p	-	+	+	9,319	8,078
CD_078p	+	+	+	5,793	5,913	CD_296p	+	+	+	8,018	7,793	CH_06	+	+	-	9,398	8,386	EL_058p	+	+	+	6,194	6,886

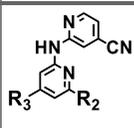
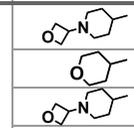
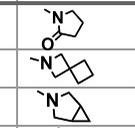
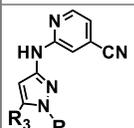
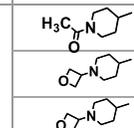
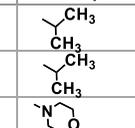
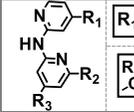
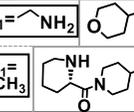
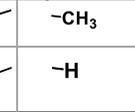
Tab. C 1 – QSAR Selektion 1. Bez. = Bezeichnung. Pred = durch das QSAR-Modell. Abkürzungen F, T, K, PC, CH, PH, EL, LL, CD = Quellenangaben für die Strukturen, siehe Kap. 7.11, S. 281.

Bez.	Sp1	Sp2	Sp3	pK _i	Pred	Bez.	Sp1	Sp2	Sp3	pK _i	Pred	Bez.	Sp1	Sp2	Sp3	pK _i	Pred
EL_059p	+	+	+	6,509	7,483	EL_177p	-	+	+	9,174	7,481	LL_205	+	+	+	8,538	8,613
EL_068p	+	+	+	7,824	7,282	EL_178p	+	+	+	6,745	7,453	LL_208	+	+	+	8,824	8,750
EL_071p	+	-	+	6,569	7,398	EL_179p	+	+	+	6,000	7,207	LL_212p	+	+	-	8,357	8,326
EL_075p	+	+	-	7,018	6,511	LL_001p	-	+	-	8,642	8,436	LL_215p	+	+	+	8,854	8,411
EL_078p	-	+	+	6,000	7,035	LL_006	+	+	+	7,842	7,637	LL_224	-	+	+	8,420	8,863
EL_081p	-	+	+	7,076	8,045	LL_010p	+	+	-	8,484	8,531	LL_225	+	+	+	8,377	8,356
EL_083p	-	-	+	6,409	7,298	LL_020p	-	-	-	8,358	7,754	LL_229	+	+	+	8,40	7,109
EL_084p	-	-	+	6,585	7,200	LL_022p	+	+	-	8,804	8,722	LL_232	+	-	+	8,959	7,722
EL_088p	+	+	+	6,585	6,935	LL_024p	+	+	+	9,469	9,391	LL_233	-	+	+	8,745	8,453
EL_089p	+	+	+	6,854	6,44	LL_027p	+	-	-	9,854	10,126	LL_237	+	+	+	8,959	8,803
EL_092p	+	+	+	6,027	6,736	LL_031p	+	+	+	7,100	7,551	LL_238	+	+	+	8,795	8,477
EL_095p	+	+	+	5,921	6,278	LL_050p	+	-	+	8,548	8,621	LL_239	+	+	+	8,770	8,337
EL_101	+	+	-	9,000	8,531	LL_061p	+	+	+	8,516	7,821	LL_240	+	-	+	8,678	8,308
EL_102p	+	+	+	7,469	7,343	LL_064p	+	+	+	9,117	8,665	LL_242	+	+	+	8,678	8,346
EL_104	+	+	+	7,041	7,075	LL_075p	+	+	+	9,131	9,025	LL_257	+	-	+	9,155	8,06
EL_108	+	+	+	8,310	6,42	LL_076p	+	+	+	8,664	8,364	LL_264	+	-	-	8,745	9,130
EL_114p	+	-	+	6,018	5,481	LL_078p	+	+	+	9,284	8,520	LL_267	+	+	-	6,658	6,80
EL_118p	+	-	-	7,469	6,947	LL_080p	+	+	+	9,222	8,252	LL_268	+	+	+	6,656	6,69
EL_119p	-	-	+	8,260	7,865	LL_083p	-	+	+	9,921	8,027	PC_015	+	+	+	8,69	8,918
EL_123p	+	+	+	7,569	7,415	LL_086p	+	+	+	8,796	7,849	PC_018	+	-	-	6,780	7,796
EL_126p	+	+	-	8,721	7,90	LL_087p	+	+	-	9,194	9,437	PC_026	+	+	-	9,301	9,08
EL_133p	+	+	+	9,398	9,061	LL_095p	+	+	+	8,672	8,709	PC_027	+	+	-	9,155	8,00
EL_139p	+	+	+	7,432	7,141	LL_100p	-	+	+	8,987	8,550	PC_030	+	+	+	9,69	8,726
EL_143p	+	+	-	7,131	6,712	LL_101p	+	+	+	9,357	8,961	PH_00	+	+	+	6,357	7,685
EL_144p	+	-	-	7,658	7,289	LL_102p	+	-	+	9,102	8,798	PH_010	+	+	+	6,260	7,236
EL_145p	+	+	-	8,131	7,245	LL_114	-	+	+	7,821	7,325	PH_015	+	+	+	6,876	7,648
EL_146p	+	-	-	7,854	7,100	LL_127p	+	+	+	8,450	7,645	PH_016	-	+	-	7,319	7,981
EL_155p	+	+	+	6,523	6,472	LL_132p	+	+	-	6,588	7,232						
EL_158p	+	+	+	7,456	7,399	LL_133p	+	+	-	6,493	6,970						
EL_161p	-	+	+	7,678	6,84	LL_144p	-	+	+	8,553	7,038						
EL_171p	-	-	-	7,796	8,226	LL_149p	+	+	+	8,387	8,119						
EL_172p	-	+	+	6,569	6,04	LL_154p	+	-	+	9,155	8,626						

Tab. C 1 – Fortsetzung.

Struktur	R3	R2	Bez.	Sp1	Sp2	Sp3	pK _{i/d}	Pred
s. S. 154, Abb. 5-44			K_CHIR-258_TKI-258p	+	-	+	5,678	6,828
			K_JNJ-7706621	+	+	+	5,284	5,211
s. S. 152, Abb. 5-41			K_Staurosporin	+	+	-	5,824	5,739
			F_CEP-701	-	-	+	6,076	6,196
s. S. 153, Abb. 5-42			F_KW-2449p	+	+	+	7,174	6,903
			PC_019	+	-	+	6,366	8,231
			PC_024	+	-	+	7,854	8,708
			PC_026p	+	+	+	9,301	9,689
			PC_027p	+	-	-	9,155	8,263
			PC_030p	+	+	+	9,699	8,988
			CH_024p	+	-	-	6,854	7,102
			CH_030pm	-	+	+	9,770	10,408
			CH_031p	+	-	-	9,770	9,295
			CH_033p	+	+	+	9,770	9,572
			CH_036pm	-	+	+	9,770	9,714
			CH_039p	+	+	+	9,770	9,776
			CH_040p	+	-	+	9,770	9,354
			CH_041	-	-	+	7,252	7,270
			CH_046	+	+	-	7,886	8,386
			CH_060	-	+	+	7,553	7,541
			CH_063	+	+	+	8,066	7,863
			CH_065	+	-	+	7,523	7,144
			CH_067	+	+	+	9,398	8,356
			CH_074	-	-	+	9,699	9,225
			CH_078	+	+	+	8,509	6,878

Tab. C 2 – QSAR Selektion 2.

Struktur	R3	R2	Bez.	Sp1	Sp2	Sp3	pK _i	Pred
			CH_079p	-	+	+	7,456	7,701
			CH_082	+	+	+	7,000	7,333
			CH_092p	-	+	+	9,155	8,708
			CH_097p	+	+	+	9,222	9,117
			CH_098p	+	+	+	7,375	7,554
			CH_105p	+	+	-	8,155	8,502
			CH_108	-	+	+	7,032	6,882
			CH_111p	+	+	+	7,051	6,978
			CH_115p	+	+	+	10,046	9,976
			CH_116	+	+	+	6,668	6,236
			CH_121p	+	+	-	9,699	9,325
			CH_123	+	+	+	9,699	9,030
			CH_124p	+	+	+	6,149	6,919
			CH_129	+	+	+	6,292	6,589
			CH_131p	+	+	-	9,357	8,300
			PH_009	+	+	+	6,357	7,269
			PH_010p	+	+	+	6,260	6,111
			PH_014p	+	+	+	6,227	5,782
			CH_019p	-	+	-	8,292	9,289
			CH_023p	+	+	+	9,495	9,444
			CH_117p	+	+	+	9,027	8,821
			CH_120p	+	+	+	6,770	7,467
			CH_122	+	+	+	9,398	8,856
			CH_140p	+	+	-	9,770	10,269
			CH_113	+	+	+	5,796	6,137
			CH_147pm	+	+	+	5,921	5,972

Struktur	R3	R2	Bez.	Sp1	Sp2	Sp3	pK _i	Pred
			CH_025p	+	+	+	9,076	8,280
			CH_045	+	+	+	7,319	7,284
			CH_055	+	+	-	8,051	8,682
			CH_068p	+	+	+	9,523	9,321
			CH_071p	-	+	+	9,444	9,392
			EL_053p	+	+	+	7,252	7,791
			EL_107p	+	+	+	8,620	7,619
			EL_109p	+	+	+	9,000	8,076
			EL_122p	+	+	+	8,102	8,398
			EL_124p	+	+	+	7,569	7,407
			EL_130	+	+	+	8,252	7,371
			EL_167p	+	+	-	7,569	6,796
			EL_177p	+	+	+	9,174	7,898
			EL_179p	+	+	+	6,000	7,529
			LL_030p	+	+	+	8,445	7,858
			LL_037p	+	+	-	8,625	8,533
			LL_045p	+	+	+	8,481	8,135
			LL_090p	+	-	-	8,225	8,240
			LL_102p	+	+	+	9,102	8,883
			LL_103p	+	-	+	8,893	8,272
			LL_141p	-	+	+	8,070	7,748
			LL_148p	+	+	+	8,051	8,155
			LL_235p*	+	+	+	8,420	8,815
			LL_266p*	+	+	-	8,959	8,873
		LL_294	+	-	+	8,051	7,934	

Tab. C 2 – Fortsetzung.

Struktur	R3	R2	Bez.	Sp1	Sp2	Sp3	pK _i	Pred
			EL_039	-	+	-	6,337	6,916
			EL_078p	+	+	+	6,000	6,941
			EL_116p	-	+	+	7,959	7,807
			EL_118p	+	+	+	7,469	6,871
			EL_120p	+	-	+	7,523	6,904
			EL_136p	-	+	+	7,620	7,280
			LL_048p	-	+	-	8,509	7,776
			LL_057p	+	+	-	8,726	8,908
			LL_064p	+	+	+	9,117	8,728
			LL_101p	+	+	+	9,357	9,013
			LL_109p	-	+	+	8,799	8,855
			LL_214p	+	+	+	8,824	8,625
			LL_220p*	+	-	+	8,678	8,480
			LL_237p*	+	-	+	8,959	9,648
			LL_244p*	+	+	+	8,959	8,693
			LL_256p	-	+	-	8,959	9,589
			LL_295p	+	+	+	8,620	8,358
			EL_003	+	+	+	6,854	6,871
			EL_137p	-	-	+	8,131	7,871
			EL_151p	+	+	+	8,260	7,554
			EL_164p	+	+	+	7,886	7,972
			LL_004	+	+	-	8,346	7,572
			LL_075p	+	+	+	9,131	7,953
			LL_078p	+	+	+	9,284	8,190
			LL_080p	+	+	-	9,222	8,535

Struktur	R3	R2	Bez.	Sp1	Sp2	Sp3	pK _i	Pred
			LL_083p	+	+	+	9,921	8,529
			LL_087p	+	-	+	9,194	8,441
			LL_105	+	+	+	7,738	7,548
			LL_127p	+	+	+	8,450	7,894
			LL_130p	+	+	+	8,317	8,149
			LL_132p	+	+	+	6,588	7,289
			LL_133p	+	+	+	6,493	7,310
			LL_154p	+	+	+	9,155	8,907
			LL_159p	+	+	+	8,721	8,084
			LL_160p	+	-	+	8,620	7,865
			LL_183p	-	+	+	8,347	8,328
			LL_257p	+	+	+	9,155	8,100
			LL_268p	+	+	+	6,656	6,709
				EL_090	+	+	+	6,921
			EL_149p	+	-	+	7,959	6,792
			EL_092	+	+	-	6,027	6,369
			EL_155p	+	+	+	6,523	6,466
			EL_178p	-	-	+	6,745	7,163
			EL_002	-	+	+	6,959	6,950
			EL_169p	+	+	+	7,387	7,062
			EL_098	+	+	+	6,444	7,090
			CD_002	+	-	+	5,770	5,272
			CD_004	+	+	+	5,762	5,353
			CD_005	+	+	+	6,081	7,200
			CD_007	+	+	+	7,000	6,529
			CD_008	+	+	+	6,658	6,222

Tab. C 2 – Fortsetzung.

Struktur	R3	R2	Bez.	Sp1	Sp2	Sp3	pK _i	Pred
			CD_024	+	+	-	5,793	5,454
			CD_025	-	-	+	5,963	5,875
			CD_030	+	-	+	6,092	6,483
			CD_033	+	+	-	5,971	5,744
			CD_036p	+	+	-	6,658	7,519
			CD_038p	+	+	+	6,538	6,845
			CD_039	+	+	+	7,155	6,888
			CD_044	+	-	+	7,155	6,627
			CD_045p	+	+	+	5,757	5,691
			CD_046p	+	+	+	5,793	5,938
			CD_048p	-	+	+	5,793	5,662
			CD_049	+	-	+	6,268	7,086
			CD_053	+	-	+	6,569	6,151
			CD_054	-	+	+	6,796	6,167
			CD_064	-	+	+	6,569	7,176
			CD_065	-	-	+	7,398	6,547
			CD_067	-	+	+	6,187	6,124
			CD_068p	-	+	+	5,793	5,771
			CD_078p	+	-	+	5,793	5,699
			CD_109	-	+	+	6,319	6,528
			CD_290	+	+	-	6,928	6,060
			CD_440	+	+	+	8,272	7,217
			CD_017p	+	+	+	9,000	8,767
			CD_028	-	-	+	9,000	7,745
			CD_029	+	+	+	8,301	8,299
			CD_072	+	+	+	7,398	7,739
			CD_118p	+	+	-	9,000	8,363

Struktur	R3	R2	Bez.	Sp1	Sp2	Sp3	pK _i	Pred
			CD_119	+	+	+	8,699	8,508
			CD_125p	-	+	+	6,959	7,530
			CD_133	+	+	+	7,097	7,009
			CD_136p	-	+	+	9,292	8,735
			CD_140	+	-	+	8,303	7,963
			CD_148	+	+	+	7,830	8,288
			CD_156p	+	+	-	7,155	9,365
			CD_157 dpm	+	+	-	9,509	9,365
			CD_173	+	+	+	7,173	7,614
			CD_174	+	+	+	8,063	8,371
			CD_193	+	+	+	8,143	8,051
			CD_195	+	+	+	8,228	7,965
			CD_203	+	+	+	6,369	6,959
			CD_206p	+	+	-	7,386	7,342
			CD_208	-	+	+	8,180	7,925
			CD_221	+	+	+	7,770	7,720
			CD_228p	-	-	+	9,114	7,893
			CD_266	+	+	+	7,917	7,099
			CD_308	+	+	+	8,123	8,051
			CD_314	+	+	-	7,526	7,921
		CD_319p	+	+	+	6,493	7,340	
		CD_321	+	+	+	8,921	8,190	
		CD_332p	+	-	-	6,935	8,633	
		CD_349	+	+	+	7,551	8,098	
		CD_355p	-	+	+	9,241	8,889	
		CD_357p	+	-	+	7,947	8,535	

Tab. C 2 – Fortsetzung.

Struktur	R3	R2	Bez.	Sp1	Sp2	Sp3	pK _i	Pred
			CD_381p	+	+	+	8,020	7,738
			CD_446	-	+	+	7,742	7,907
			CD_455	+	+	-	8,086	7,680
			CD_462	+	-	+	8,241	8,868
			CD_485	-	+	+	7,886	7,144
			CD_015p	+	+	+	9,000	8,792
			CD_023	+	-	-	7,699	7,923
			CD_123p	+	+	+	7,699	7,752
			CD_012	+	+	-	8,000	8,527
			CD_277	+	+	+	7,479	7,666
			CD_138	+	-	-	7,006	7,573
			CD_179p	+	+	+	7,943	7,845
			CD_272p	-	+	+	8,903	8,084
			CD_113p	+	+	+	6,131	6,314
			CD_114p	+	+	-	6,161	7,480
			EL_114p	+	+	+	6,018	5,932
			EL_172p	+	+	+	6,569	5,971
			EL_095p	+	+	+	5,921	6,089
			CH_145	-	+	+	5,444	6,353
			CH_146p	+	+	-	5,678	5,832

Struktur	R3	R2	Bez.	Sp1	Sp2	Sp3	pK _i	Pred
			EL_010	+	+	+	8,194	8,051
			EL_040p	+	+	+	9,770	8,387
			EL_051	+	+	+	9,770	8,740
			EL_054p	+	-	+	9,319	8,048
			EL_101p	+	+	+	9,000	8,382
			EL_133p	+	-	+	9,398	8,651
			EL_035	+	+	+	7,000	7,195
			EL_106	+	+	+	6,886	7,011
			EL_057p	+	+	-	7,469	7,362
			EL_059p	+	+	-	6,509	6,318
			LL_024p	+	+	+	9,469	9,676
			LL_027p	+	+	+	9,854	9,819
			LL_034p	-	-	+	8,378	8,700
			LL_033p	+	+	+	6,910	6,983
			CD_001	-	-	-	5,192	4,057
			CD_003	+	+	+	5,192	5,193
			CD_108	-	-	-	6,495	7,097
			CD_116p	+	+	+	5,793	6,012

Tab. C 2 – Fortsetzung.

Bez.	Sp1	Sp2	Sp3	pK _{i,d}	Pred	Bez.	Sp1	Sp2	Sp3	pK _i	Pred	Bez.	Sp1	Sp2	Sp3	pK _i	Pred	Bez.	Sp1	Sp2	Sp3	pK _i	Pred
K_CHIR-258_TKI-258p	+	-	-	5,678	6,973	CD_072	+	+	-	7,398	7,281	CD_210	+	-	-	8.074	7.128	CH_027p	+	+	-	7,000	7,051
K_JNJ-7706621	+	+	+	5,284	4,882	CD_074	-	+	+	7,301	7,143	CD_228p	-	+	-	9.114	7.972	CH_030p	+	-	+	9,770	10,239
F_CEP-701	+	+	+	6,076	5,754	CD_078p	-	+	-	5,793	5,885	CD_258	+	+	+	7.499	7.190	CH_031p	+	+	+	9,770	9,606
K_Staurosporin	+	-	+	5,824	5,246	CD_086p	+	+	+	6,678	6,420	CD_272p	-	+	+	8.903	8.516	CH_033p	+	+	-	9,770	9,108
CD_001	-	+	-	5,192	5,333	CD_087	+	+	+	6,678	6,679	CD_275	+	+	-	7.220	7.553	CH_036p	+	-	+	9,770	9,141
CD_002	+	-	-	5,770	5,780	CD_095	+	-	-	6,167	5,857	CD_284	-	+	-	7.264	7.464	CH_039p	-	+	+	9,770	10,095
CD_003	-	+	+	5,192	5,766	CD_102	+	-	-	7,620	7,386	CD_295	+	-	+	7.187	6.860	CH_040p	+	+	-	9,770	9,751
CD_004	-	+	+	5,762	5,238	CD_108	+	-	+	6,495	7,195	CD_296	+	-	+	8.018	8.050	CH_060	-	+	+	7,553	7,538
CD_005	+	+	+	6,081	6,547	CD_109	-	-	-	6,319	6,509	CD_321	+	+	+	8.921	7.969	CH_065	+	+	+	7,523	7,081
CD_007	-	+	-	7,000	6,802	CD_110	+	-	-	7,699	6,660	CD_330p	+	+	+	8.131	7.340	CH_067	+	+	+	9,398	8,754
CD_008	-	-	+	6,658	6,409	CD_113p	-	+	+	6,131	6,274	CD_332p	-	-	+	6.935	7.393	CH_068p	+	+	-	9,523	9,053
CD_014pm	+	+	+	6,367	6,751	CD_114p	-	+	+	6,161	7,477	CD_341p	-	-	+	8.350	8.718	CH_069	+	+	-	8,398	8,850
CD_015	+	-	+	9,000	9,078	CD_115p	+	+	+	6,310	6,803	CD_344	+	+	+	7.232	7.828	CH_071p	+	+	-	9,444	8,514
CD_017p	-	+	-	9,000	8,497	CD_116p	+	+	+	5,793	6,549	CD_347p	+	+	+	7.983	7.828	CH_074	-	+	-	9,699	9,117
CD_023	-	+	+	7,699	7,456	CD_117	+	+	+	6,149	5,406	CD_350	+	+	+	7.373	7.493	CH_077	+	+	-	6,721	7,559
CD_024	+	+	+	5,793	5,379	CD_118p	+	+	+	9,000	8,394	CD_355p	+	+	-	9.241	7.988	CH_085	+	-	-	7,824	9,076
CD_025	-	+	+	5,963	6,297	CD_119	+	-	+	8,699	8,486	CD_356	+	+	-	8.033	7.373	CH_087	+	-	+	7,292	7,471
CD_026p	+	-	-	8,000	7,934	CD_125p	+	-	+	6,959	6,916	CD_366	+	-	+	7.407	7.456	CH_092p	-	+	-	9,155	8,533
CD_028	-	+	+	9,000	8,624	CD_131	+	+	+	8,000	7,483	CD_373	-	-	-	7.413	7.939	CH_097p	+	+	+	9,222	9,637
CD_030	+	-	-	6,092	6,471	CD_136p	+	+	-	9,292	8,093	CD_374	+	-	-	7.367	7.846	CH_098p	+	+	+	7,375	7,686
CD_039	-	+	+	7,155	6,532	CD_138	+	+	+	7,006	7,514	CD_382p	+	-	+	7.963	7.838	CH_100	+	+	+	7,959	8,783
CD_045p	+	+	+	5,757	5,615	CD_142	+	+	+	8,492	7,859	CD_425p	+	+	+	6.959	7.567	CH_109	+	+	+	6,854	7,416
CD_046p	-	-	+	5,793	5,995	CD_147	-	+	+	6,975	6,461	CD_438	-	+	+	7.921	8.245	CH_111p	+	-	-	7,051	7,257
CD_048p	+	+	+	5,793	5,931	CD_148	+	-	+	7,830	8,052	CD_455	+	+	+	8.086	7.675	CH_113	+	-	-	5,796	6,757
CD_049	+	+	+	6,268	7,029	CD_154p	+	+	-	8,164	7,878	CD_462	+	-	-	8.241	8.331	CH_115p	+	+	+	10,046	10,562
CD_051	-	-	+	6,699	6,717	CD_157	+	+	+	9,509	8,916	CD_468	+	+	-	8.547	7.387	CH_117p	+	+	+	9,027	9,325
CD_053	+	+	-	6,569	6,134	CD_163	-	+	+	8,302	7,201	CD_478	+	+	+	8.476	8.185	CH_121p	+	+	+	9,699	9,465
CD_055	+	+	-	6,886	6,612	CD_179p	-	+	+	7,943	7,548	CH_019p	+	+	+	8.292	8.979	CH_122	+	-	+	9,398	8,648
CD_057	+	+	+	6,921	6,908	CD_183	-	+	-	7,983	7,101	CH_020	+	+	+	6.959	6.546	CH_123	-	+	+	9,699	9,071
CD_064	+	+	+	6,569	6,673	CD_199p	-	+	+	8,028	7,798	CH_023	-	+	+	9.495	9.473	CH_131p	-	-	+	9,357	8,256
CD_067	+	-	+	6,187	6,243	CD_208	+	+	+	8,180	7,810	CH_024	+	+	+	6.854	7.077	CH_140p	+	+	+	9,770	9,509
CD_068p	+	+	+	5,793	6,056	CD_209	+	+	+	7,309	7,370	CH_025p	+	-	-	9.076	7.582	CH_144p	-	+	+	7,215	7,331

Tab. C 3 – QSAR Selektion 3.

Bez.	Sp1	Sp2	Sp3	pK _i	Pred
CH_145	+	+	-	5,444	7,486
CH_146p	-	+	+	5,678	5,417
CH_147pm	+	+	+	5,921	6,073
EL_003	+	+	+	6,854	6,920
EL_007p	-	-	-	6,328	7,172
EL_040p	-	+	-	9,770	7,769
EL_049p	+	+	+	7,721	7,466
EL_051	+	-	+	9,770	8,962
EL_052	+	-	+	7,252	7,583
EL_053p	+	+	+	7,252	7,988
EL_054p	+	-	+	9,319	8,580
EL_058p	+	+	+	6,194	7,061
EL_066p	+	+	-	6,959	7,052
EL_071p	+	+	+	6,569	7,216
EL_078p	+	+	+	6,000	6,564
EL_084p	+	-	+	6,585	6,834
EL_089p	-	-	+	6,854	6,583
EL_092p	-	+	+	6,027	6,290
EL_093p	-	-	-	6,770	6,946
EL_095p	+	-	-	5,921	6,939
EL_098p	+	+	-	6,444	7,516
EL_101p	+	+	+	9,000	7,997
EL_104	+	-	+	7,041	7,190
EL_107p	+	+	+	8,620	8,078
EL_109p	+	+	+	9,000	7,751
EL_110p	+	+	+	8,347	7,984
EL_111p	+	+	+	6,745	6,209
EL_114p	+	+	+	6,018	5,413
EL_119p	+	+	-	8,260	8,059
EL_133p	-	+	+	9,398	8,708
EL_138p	+	+	+	7,509	7,436
EL_141p	-	+	+	8,097	7,261

Bez.	Sp1	Sp2	Sp3	pK _i	Pred
EL_150p	+	+	+	7,658	6,765
EL_155p	-	+	+	6,523	6,627
EL_156p	+	+	+	7,495	7,791
EL_164p	+	+	+	7,886	7,938
EL_177p	-	+	+	9,174	8,331
EL_179p	+	-	-	6,000	7,956
LL_006	+	+	-	7,842	7,745
LL_024p	+	-	+	9,469	9,724
LL_026p	+	+	+	7,121	7,661
LL_027p	+	+	+	9,854	9,291
LL_037p	+	-	+	8,625	8,307
LL_043pm	+	-	-	8,516	7,433
LL_052	+	-	+	8,401	8,403
LL_057p	-	+	+	8,726	8,244
LL_064p	+	+	+	9,117	8,797
LL_069p	-	-	+	8,527	8,198
LL_075p	+	+	+	9,131	7,996
LL_078p	+	+	+	9,284	8,643
LL_080p	-	-	+	9,222	8,783
LL_083p	+	-	+	9,921	9,013
LL_085p	-	+	-	7,321	7,568
LL_087p	+	-	-	9,194	8,353
LL_098p	+	-	-	8,311	7,632
LL_101p	+	+	-	9,357	9,039
LL_102p	+	-	+	9,102	8,862
LL_103p	+	+	+	8,893	8,496
LL_109p	+	-	-	8,799	8,976
LL_110	-	+	-	8,410	8,060
LL_114	-	+	+	7,821	7,663
LL_132p	-	+	-	6,588	7,695
LL_133p	-	-	+	6,493	7,758
LL_135p	+	+	+	8,254	7,799

Bez.	Sp1	Sp2	Sp3	pK _i	Pred
LL_154p	+	+	+	9,155	8,503
LL_159p	+	+	+	8,721	8,189
LL_160p	+	+	+	8,620	8,374
LL_166	-	-	+	8,174	7,858
LL_212p	+	+	+	8,357	8,191
LL_214p	-	+	-	8,824	8,895
LL_216p	-	+	+	8,180	7,678
LL_220p*	+	+	-	8,678	8,109
LL_224p*	-	+	+	8,420	8,582
LL_229	+	+	+	8,409	7,969
LL_237p*	+	+	+	8,959	9,291
LL_244p*	+	-	-	8,959	8,772
LL_256p	-	+	+	8,959	8,640
LL_257p*	+	+	+	9,155	8,295
LL_266p*	+	-	-	8,959	8,031
LL_267p	+	+	-	6,658	8,054
LL_272	-	-	+	8,252	8,809
LL_287	+	+	+	8,201	7,525
LL_295pm	-	+	+	8,620	8,248
PC_019	+	+	+	6,366	7,353
PC_026p	-	+	+	9,301	9,434
PC_027p	+	-	+	9,155	8,661
PC_030p	+	-	+	9,699	8,997
PH_009	+	+	+	6,357	7,805
PH_010p	+	+	+	6,260	6,479
PH_011p	-	+	+	7,377	7,397
PH_019p	-	+	-	6,245	7,353

Tab. C 3 – Fortsetzung.

D. Statistische Parameter weiterer QSAR-Modelle

Tab. D 1 – Statistische Parameter der GAUSSIAN-Modelle der **Selektion 2** für einen Trainingsdatensatz von 66 % und drei Randomisierungen (Split 1-3). Ein Modell genügt den Kriterien $R_{cv}^2 > 0,5$ und $Q^2 > 0,6$ (rot eingerahmt).

Selektion 2, Gaussian, 66 % Trainingsdatensatz												
Felder	Split 1				Split 2				Split 3			
	#k	R^2	R_{cv}^2	Q^2	#k	R^2	R_{cv}^2	Q^2	#k	R^2	R_{cv}^2	Q^2
SEHAD	5	0,822	0,658	0,410	5	0,795	0,565	0,602	5	0,826	0,648	0,268
SEHA	5	0,820	0,660	0,383	4	0,761	0,563	0,555	4	0,803	0,644	0,285
SEHD	5	0,797	0,634	0,373	4	0,737	0,537	0,547	5	0,799	0,637	0,276
SEAD	3	0,765	0,613	0,356	5	0,773	0,546	0,559	4	0,720	0,596	0,241
SHAD	5	0,811	0,644	0,431	5	0,782	0,557	0,574	5	0,824	0,647	0,258
EHAD	5	0,817	0,629	0,319	5	0,808	0,582	0,555	4	0,806	0,638	0,360
SEH	5	0,796	0,642	0,343	6	0,769	0,527	0,569	4	0,772	0,635	0,269
SEA	5	0,812	0,640	0,252	4	0,745	0,542	0,571	2	0,686	0,587	0,211
SED	4	0,744	0,594	0,373	4	0,672	0,504	0,521	3	0,684	0,583	0,239
SHA	5	0,808	0,651	0,407	4	0,753	0,566	0,545	4	0,794	0,642	0,302
SHD	5	0,791	0,608	0,394	4	0,703	0,517	0,499	5	0,786	0,629	0,258
SAD	4	0,714	0,558	0,314	5	0,726	0,478	0,521	3	0,704	0,580	0,190
EHA	5	0,820	0,644	0,245	5	0,799	0,583	0,515	5	0,835	0,626	0,336
EHD	5	0,803	0,614	0,272	4	0,726	0,553	0,567	5	0,807	0,628	0,370
EAD	4	0,762	0,559	0,135	5	0,786	0,581	0,420	5	0,794	0,567	0,297
HAD	5	0,792	0,601	0,279	6	0,800	0,569	0,535	5	0,825	0,636	0,297
SE	5	0,778	0,631	0,217	4	0,487	0,665	0,526	2	0,657	0,567	0,235
SH	6	0,792	0,624	0,308	5	0,715	0,500	0,571	4	0,755	0,635	0,267
SA	5	0,749	0,589	0,282	4	0,694	0,489	0,553	3	0,706	0,571	0,165
SD	5	0,698	0,515	0,354	3	0,584	0,424	0,451	3	0,660	0,563	0,175
EH	4	0,788	0,637	0,191	4	0,752	0,545	0,544	5	0,801	0,625	0,407
EA	4	0,746	0,570	0,081	4	0,751	0,545	0,430	4	0,756	0,553	0,249
ED	3	0,655	0,520	0,308	5	0,716	0,486	0,492	4	0,706	0,516	0,194
HA	5	0,799	0,626	0,258	5	0,776	0,572	0,490	5	0,813	0,633	0,261
HD	6	0,755	0,576	0,195	5	0,696	0,504	0,465	6	0,790	0,614	0,316
AD	4	0,629	0,427	0,010	5	0,676	0,415	0,258	4	0,684	0,490	-0,053
S	6	0,695	0,530	0,244	4	0,587	0,412	0,464	5	0,661	0,518	0,112
E	5	0,767	0,559	0,035	6	0,725	0,469	0,449	5	0,722	0,520	0,254
H	5	0,738	0,581	0,134	4	0,665	0,489	0,411	6	0,771	0,582	0,241
A	4	0,632	0,452	-0,080	4	0,641	0,370	0,285	3	0,609	0,421	0,057
D	4	0,333	0,167	0,201	4	0,385	0,385	0,095	4	0,436	0,321	0,119

Tab. D 2 – Statistische Parameter der GAUSSIAN-Modelle der **Selektion 3** für einen Trainingsdatensatz von 66 % und drei Randomisierungen (Split 1-3), Keines der Modelle genügt den Kriterien $R_{cv}^2 > 0,5$ und $Q^2 > 0,6$.

Selektion 3, Gaussian, 66 % Trainingsdatensatz												
Felder	Split 1				Split 2				Split 3			
	#k	R^2	R_{cv}^2	Q^2	#k	R^2	R_{cv}^2	Q^2	#k	R^2	R_{cv}^2	Q^2
SEHAD	5	0,877	0,759	0,124	3	0,762	0,647	0,410	6	0,855	0,659	0,514
SEHA	5	0,883	0,772	0,122	4	0,796	0,640	0,392	6	0,848	0,673	0,483
SEHD	5	0,860	0,748	0,093	4	0,757	0,637	0,382	6	0,811	0,625	0,522
SEAD	6	0,878	0,741	0,154	3	0,733	0,608	0,434	5	0,809	0,614	0,490
SHAD	5	0,872	0,750	0,115	3	0,758	0,641	0,395	5	0,823	0,640	0,523
EHAD	5	0,854	0,697	0,082	4	0,765	0,601	0,241	5	0,809	0,586	0,446
SEH	5	0,864	0,765	0,089	4	0,744	0,611	0,451	4	0,760	0,643	0,443
SEA	6	0,880	0,754	0,112	4	0,756	0,603	0,405	6	0,840	0,635	0,478
SED	7	0,858	0,716	0,131	4	0,720	0,590	0,379	2	0,670	0,581	0,488
SHA	5	0,874	0,762	0,112	4	0,776	0,625	0,389	6	0,834	0,647	0,519
SHD	5	0,844	0,726	0,067	4	0,739	0,617	0,370	2	0,690	0,595	0,471
SAD	5	0,836	0,685	0,117	4	0,753	0,565	0,229	3	0,682	0,555	0,464
EHA	5	0,853	0,704	0,030	4	0,771	0,613	0,271	5	0,820	0,600	0,400
EHD	5	0,832	0,689	0,021	3	0,709	0,606	0,295	7	0,818	0,596	0,464
EAD	5	0,802	0,621	0,154	3	0,692	0,492	0,274	5	0,773	0,505	0,453
HAD	5	0,848	0,692	0,007	4	0,740	0,581	0,206	5	0,783	0,549	0,448
SE	7	0,865	0,745	0,171	3	0,671	0,576	0,467	4	0,731	0,590	0,496
SH	5	0,840	0,727	0,053	4	0,718	0,582	0,423	5	0,738	0,581	0,472
SA	5	0,834	0,691	0,119	4	0,714	0,553	0,377	3	0,681	0,566	0,455
SD	4	0,737	0,629	0,062	4	0,654	0,513	0,326	4	0,654	0,507	0,453
EH	6	0,860	0,708	-0,054	4	0,746	0,597	0,319	6	0,812	0,594	0,447
EA	5	0,793	0,591	0,165	4	0,795	0,569	0,149	5	0,782	0,531	0,354
ED	4	0,734	0,601	0,276	5	0,752	0,536	0,163	5	0,709	0,517	0,493
HA	6	0,863	0,708	-0,103	4	0,752	0,601	0,242	5	0,801	0,574	0,402
HD	6	0,832	0,686	-0,152	5	0,736	0,557	0,225	6	0,743	0,513	0,435
AD	5	0,718	0,520	-0,193	3	0,563	0,367	0,104	3	0,548	0,336	-0,332
S	5	0,736	0,609	0,044	4	0,614	0,468	0,360	2	0,591	0,512	0,415
E	4	0,744	0,588	0,191	4	0,715	0,555	0,262	6	0,737	0,492	0,437
H	6	0,829	0,689	-0,105	5	0,719	0,548	0,286	5	0,713	0,514	0,411
A	5	0,708	0,485	-0,170	3	0,558	0,392	0,076	4	0,609	0,362	0,293
D	4	0,400	0,277	0,036	5	0,412	0,236	-0,009	4	0,322	0,147	-0,301

E. Proteinkonformationen für Docking- und VSW-Experimente

Tab. E 1 – Übersicht über die Protein-IDs zur Erstellung von Grids in den Docking- und VSW-Experimenten; H₂O = Wassermoleküle; PROTREP siehe auch exp. Teil, Kap. 7.3.

ID	Bezeichnung	Charakteristika
1	5CEN_opls3_w_nw	5CEN nach PROTREP: Minimierung mit H ₂ O, die mind. 1 WBB zu Nicht-H ₂ O eingehen; alle H ₂ O nachträglich entfernt (= nw)
2	5CEN_opls3_minwat_nw	5CEN nach PROTREP (Minimierung mit allen H ₂ O; alle H ₂ O nachträglich entfernt)
3	5CEN_opls3_ow	5CEN nach PROTREP (Minimierung nach Entfernung aller H ₂ O)
4	5CEO_opls3_w	5CEO nach PROTREP (Minimierung mit H ₂ O, die mind. 1 WBB zu Nicht-H ₂ O eingehen UND die höchstens 5 Å von Het-Gruppen entfernt liegen)
5	5CEO_opls3_w_nw	= ID 4 nach Entfernung aller H ₂ O
6	5CEO_opls3_minwat	5CEO nach PROTREP (Minimierung mit allen H ₂ O)
6.1	5CEO_opls3_minwat_c	= ID 6 nach Entfernung aller H ₂ O bis auf eins (c) in der Bindetasche (s. Kap. 5.4.1.2, S. 185, Abb. 5-65)
6.2	5CEO_opls3_minwat_cd	= ID 6 nach Entfernung aller Wassermoleküle bis auf 2 definierte H ₂ O in der Bindetasche (c, d)
6.3	5CEO_opls3_minwat_bcd	= ID 6 nach Entfernung aller H ₂ O bis auf 3 definierte H ₂ O (b, c, d) in der Bindetasche
6.4	5CEO_opls3_minwat_acd	= ID 6 nach Entfernung aller H ₂ O bis auf 3 definierte H ₂ O (a, c, d) in der Bindetasche
6.5	5CEO_opls3_minwat_abd	= ID 6 nach Entfernung aller H ₂ O bis auf 3 definierte H ₂ O (a, b, d) in der Bindetasche
6.6	5CEO_opls3_minwat_abd	= ID 6 nach Entfernung aller H ₂ O bis auf 3 definierte H ₂ O (a, b, c) in der Bindetasche
6.7	5CEO_opls3_minwat_abcd	= ID 6 nach Entfernung aller H ₂ O bis auf 4 definierte H ₂ O (a, b, c, d) in der Bindetasche
6.8	5CEO_opls3_minwat_8wat	= ID 6 mit nur 8 definierten H ₂ O in (der Nähe) der Bindetasche
6.9	5CEO_opls3_minwat_b	= ID 6 nach Entfernung aller H ₂ O bis auf eins (b) in der Bindetasche
6.10	5CEO_opls3_minwat_d	= ID 6 nach Entfernung aller H ₂ O bis auf eins (d) in der Bindetasche
6.11	5CEO_opls3_minwat_bd	= ID 6 nach Entfernung aller H ₂ O bis auf 2 definierte H ₂ O (b, d)
7	5CEO_opls3_minwat_nw	= ID 6 nach Entfernung aller H ₂ O
8	5CEO_opls3_ow	5CEO nach PROTREP (Minimierung nach Entfernung aller H ₂ O)
9	5CEP_opls3_w	5CEP nach PROTREP (Minimierung mit H ₂ O, die mind. 1 WBB zu Nicht- H ₂ O eingehen UND die höchstens 5 Å von Het-Gruppen entfernt liegen)
10	5CEP_opls3_w_nw	= ID 9 nach Entfernung aller H ₂ O (= nw)
11	5CEP_opls3_minwat	5CEP nach PROTREP (Minimierung mit allen H ₂ O)
12	5CEP_opls3_minwat_nw	= ID 11 nach Entfernung aller H ₂ O (= nw)
13	5CEP_opls3_ow	5CEP nach PROTREP (Minimierung nach Entfernung aller H ₂ O)
14	5CEQ_opls3_w	5CEQ nach PROTREP (Minimierung mit H ₂ O, die mind. 1 WBB zu Nicht- H ₂ O eingehen UND die höchstens 5 Å von Het-Gruppen entfernt liegen)
15	5CEQ_opls3_w_nw	= ID 14 nach Entfernung aller Wassermoleküle
16	5CEQ_opls3_minwat	5CEQ nach PROTREP (Minimierung mit allen Wassermolekülen)
17	5CEQ_opls3_minwat_nw	= ID 16 nach Entfernung aller H ₂ O
18	5CEQ_opls3_ow	5CEQ nach PROTREP (Minimierung nach Entfernung aller H ₂ O)
19	5CEN_waters_opt_nw	= 5CEN_waters_opt nach PROTREP mit allen H ₂ O, aber <u>ohne</u> die anschließende Minimierung. Alle H ₂ O nachträglich entfernt.
20	5CEO_waters_opt	= 5CEO_waters_opt nach PROTREP mit allen H ₂ O, aber <u>ohne</u> die anschließende Minimierung
21	5CEO_waters_opt_nw	= ID 20 nach Entfernung aller Wassermoleküle
22	5CEP_waters_opt	= 5CEP_waters_opt nach PROTREP mit allen H ₂ O, aber <u>ohne</u> die anschließende Minimierung
23	5CEP_waters_opt_nw	= ID 22 nach Entfernung aller H ₂ O
24	5CEQ_waters_opt	= 5CEQ_waters_opt nach PROTREP mit allen H ₂ O, aber <u>ohne</u> die anschließende Minimierung
25	5CEQ_waters_opt_nw	= ID 24 nach Entfernung aller H ₂ O
26	5CEN_run01_fr2006_opls3_6A	Frame 2006 aus der 1. 5CEN-MD-Trajektorie nach PROTREP
27	5CEN_run03_fr1702_opls3_6A	Frame 1702 aus der 1. 5CEN-MD-Trajektorie nach PROTREP
28	5CEO_run01_fr2106_opls3_6A	Frame 2106 aus der 1. 5CEO-MD-Trajektorie nach PROTREP
29	5CEO_run02_fr1469_opls3_6A	Frame 1469 aus der 2. 5CEO-MD-Trajektorie nach PROTREP
30	5CEP_run01_fr2492_opls3_6A	Frame 2492 aus der 1. 5CEP-MD-Trajektorie nach PROTREP
31	5CEP_run02_fr1089_opls3_6A	Frame 1089 aus der 2. 5CEP-MD-Trajektorie nach Minimierung mit PROTREP
32	5CEQ_run01_fr2899_opls3_6A	Frame 2899 aus der 1. 5CEQ-MD-Trajektorie nach Minimierung mit PROTREP
33	5CEQ_run02_fr2978_opls3_6A	Frame 2978 aus der 2. 5CEN-MD-Trajektorie nach Minimierung mit PROTREP

F. Ergebnisse des OPLS2005-A-1-, OPLS3-B-1- und OPLS3A-2-Workflows

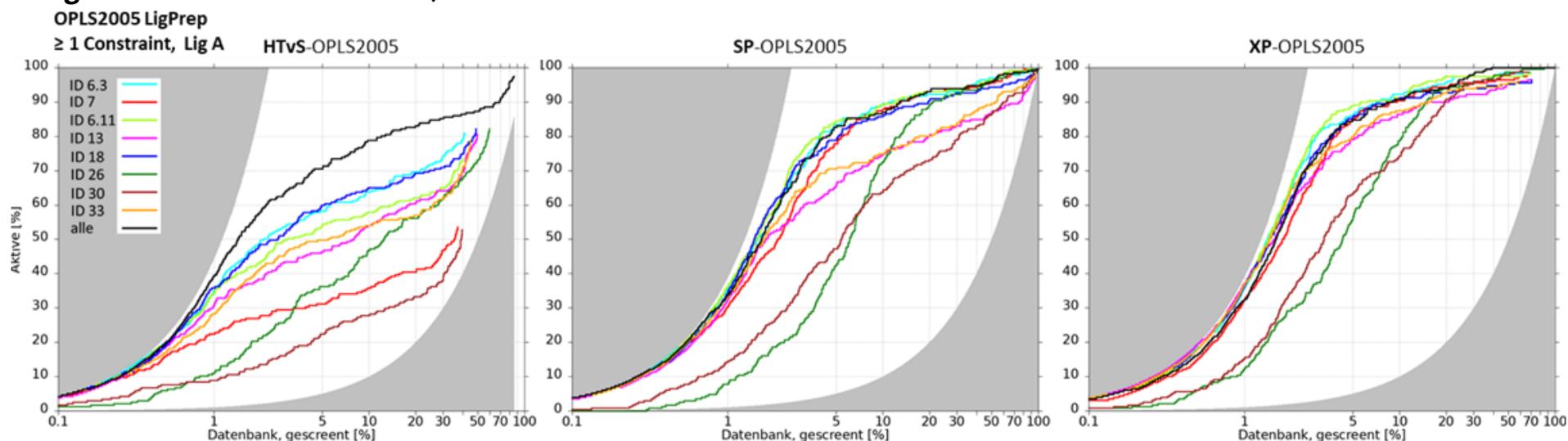


Abb. F 1 – Enrichmentplots des *OPLS2005-A-1-Workflows* für die drei verschiedenen Dockingmodi. Graue Kurve oben: perfekte Anreicherung; graue Kurve unten: „Anreicherung“ der Aktiven bei einer rein zufälligen Verteilung der Aktiven in der Input-Datenbank.

HTvs	Anreicherung EF x%				# gedockte Liganden		
	1%	5%	10%	50%	#A	#L	#A/#L
alle	39,40	14,08	7,89	1,75	231	9007	2,565
ID 6.3	34,79	11,63	6,37	1,62	192	4324	4,440
ID 6.11	34,37	10,87	5,78	1,47	174	4216	4,127
ID 7	22,21	6,24	3,59	1,07	127	3918	3,241
ID 13	31,01	9,27	5,40	1,61	191	5217	3,661
ID 18	35,62	11,97	6,50	1,65	195	5139	3,795
ID 26	10,90	7,25	4,68	1,46	195	6260	3,115
ID 30	8,80	4,47	2,78	1,05	125	4197	2,978
ID 33	28,08	9,95	5,40	1,58	187	4739	3,946
6IDs	39,40	14,16	7,85	1,76	226	8025	2,816

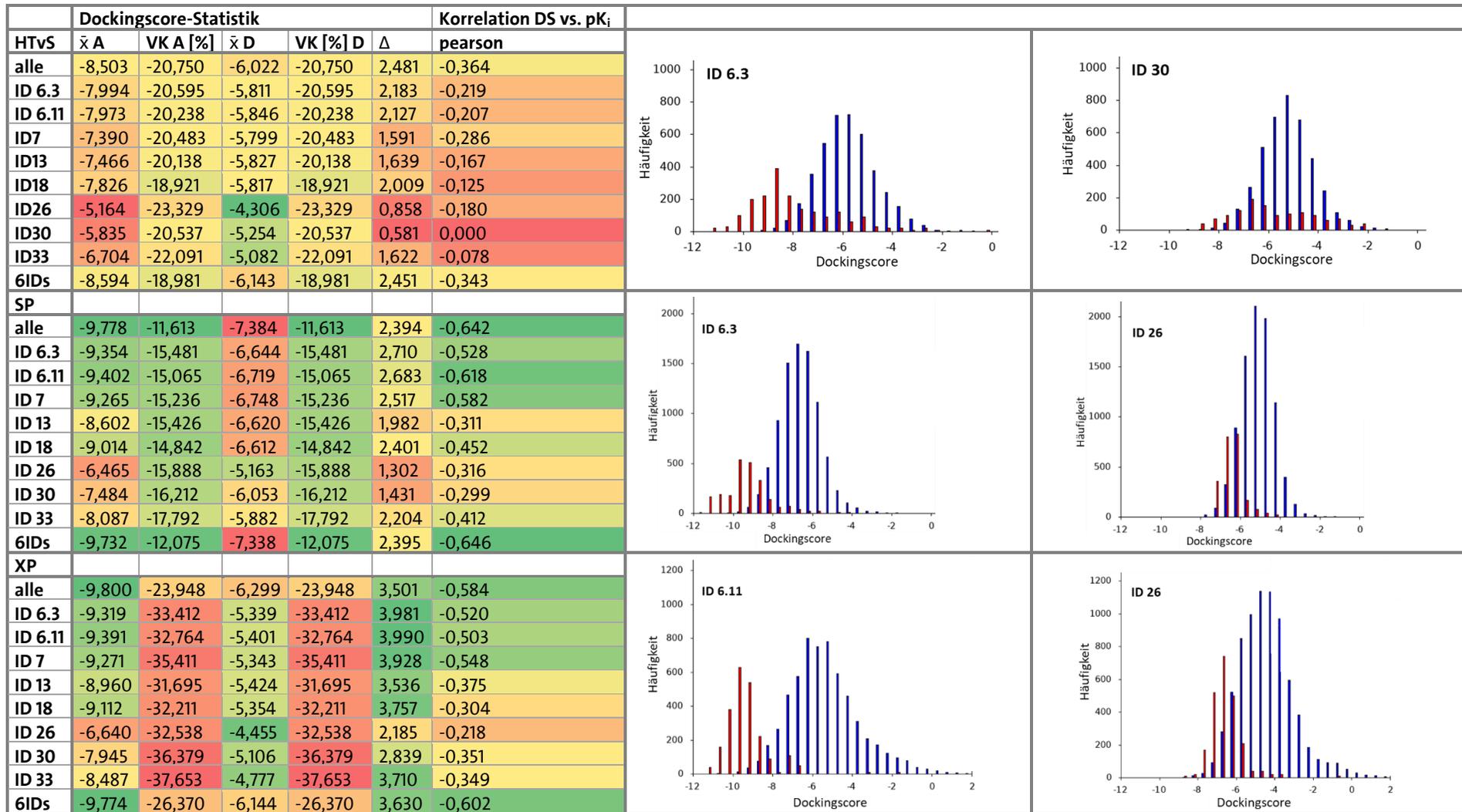
Tab. F 1 – Enrichmentfaktoren und Anzahl gedockter Liganden des HTvs im *OPLS2005-A-1-Workflow*. A = Aktive, L = Liganden (Decoys + Aktive). 6IDs: Merging aus ID 6.3, 6.11, 7, 13, 18 und 33.

SP	Anreicherung EF x%				# gedockte Liganden		
	1%	5%	10%	50%	#A	#L	#A/#L
alle	33,85	16,51	8,70	1,92	230	9004	2,554
ID 6.3	35,56	16,69	8,96	1,94	230	8834	2,604
ID 6.11	36,42	16,86	8,91	1,93	230	8809	2,611
ID 7	30,42	15,56	8,78	1,91	230	8811	2,610
ID 13	32,56	13,14	7,49	1,73	230	8903	2,583
ID 18	34,71	15,73	8,61	1,89	230	8920	2,578
ID 26	8,14	8,47	7,23	1,92	230	8989	2,559
ID 30	14,57	9,42	6,40	1,73	230	8885	2,589
ID 33	33,42	14,09	7,49	1,79	230	8822	2,607
6IDs	33,88	16,61	8,65	1,93	230	8990	2,558

Tab. F 2 – Enrichmentfaktoren und Anzahl gedockter Liganden des SP-Dockings im *OPLS2005-A-1-Workflow*. A = Aktive, L = Liganden (Decoys + Aktive). 6IDs: Merging aus ID 6.3, 6.11, 7, 13, 18 und 33.

XP	Anreicherung EF x%				# gedockte Liganden		
	1%	5%	10%	50%	#A	#L	#A/#L
alle	32,26	16,93	9,08	2,00	230	8885	2,589
ID 6.3	25,10	16,73	9,07	1,97	227	6198	3,662
ID 6.11	36,57	17,79	9,12	1,95	226	6316	3,578
ID 7	32,69	17,01	9,04	1,93	224	5897	3,799
ID 13	37,00	15,80	8,65	1,91	222	6208	3,576
ID 18	36,57	17,10	9,12	1,90	222	6351	3,496
ID 26	12,48	11,28	7,86	1,97	230	7838	2,934
ID 30	15,49	12,67	7,43	1,96	227	6084	3,731
ID 33	37,00	15,97	8,73	1,91	221	5295	4,174
6IDs	33,15	17,03	9,17	2,00	230	8405	2,736

Tab. F 3 – Enrichmentfaktoren und Anzahl gedockter Liganden des XP-Dockings im *OPLS2005-A-1-Workflow*. A = Aktive, L = Liganden (Decoys + Aktive). 6IDs: Merging aus ID 6.3, 6.11, 7, 13, 18 und 33.



Tab. F 4 – Mittelwert und Variationskoeffizient der Dockingscores der Aktiven (A) bzw. der Decoys (D) sowie die Differenz $\Delta = \bar{x} D - \bar{x} A$ aus dem *OPLS2005-A-1-Workflow*. Pearson-Korrelation zwischen Dockingscore und pK_i -Wert.

Abb. F 2 – Häufigkeitsverteilung der Dockingscores der Decoys (blau) und der Aktiven (rot, 10 fach vergrößert). Oben: HTVs, Mitte: SP, unten: XP. Es sind jeweils die Ergebnisse der Proteinkonformation mit der besten Trennung der Aktiven von den Decoys (links) und der geringsten Trennleistung (rechts) dargestellt.

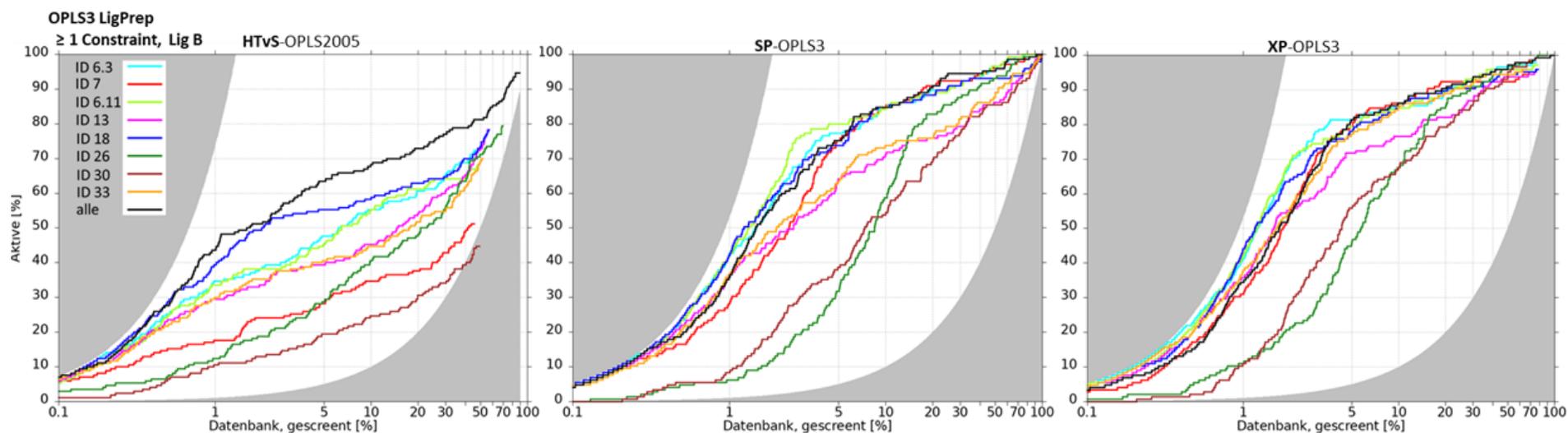


Abb. F 3 – Enrichmentplots des *OPLS3-B-1-Workflows* für die drei verschiedenen Dockingmodi. Graue Kurve oben: perfekte Anreicherung; graue Kurve unten: „Anreicherung“ der Aktiven bei einer rein zufälligen Verteilung der Aktiven in der Input-Datenbank.

HTvs	Anreicherung EF x%				# gedockte Ligan-		
	1%	5%	10%	50%	#A	#L	#A/#L
alle	43,37	12,70	6,82	1,62	161**	11192	1,439
ID 6.3	34,58	9,52	5,53	1,45	123	6136	2,005
ID 6.11	33,41	9,05	5,53	1,36	116	6112	1,898
ID7	17,58	5,64	3,47	1,02	87	5735	1,517
ID13	29,30	7,88	4,53	1,46	133	7066	1,882
ID18	39,27	11,05	5,88	1,47	133	7122	1,867
ID26	12,31	5,76	4,06	1,42	135	8720	1,548
ID30	10,55	3,88	2,47	0,89	76	6206	1,225
ID33	29,89	8,11	4,47	1,38	119	6508	1,829
6IDs	43,96	12,82	6,88	1,60	157	10097	1,555

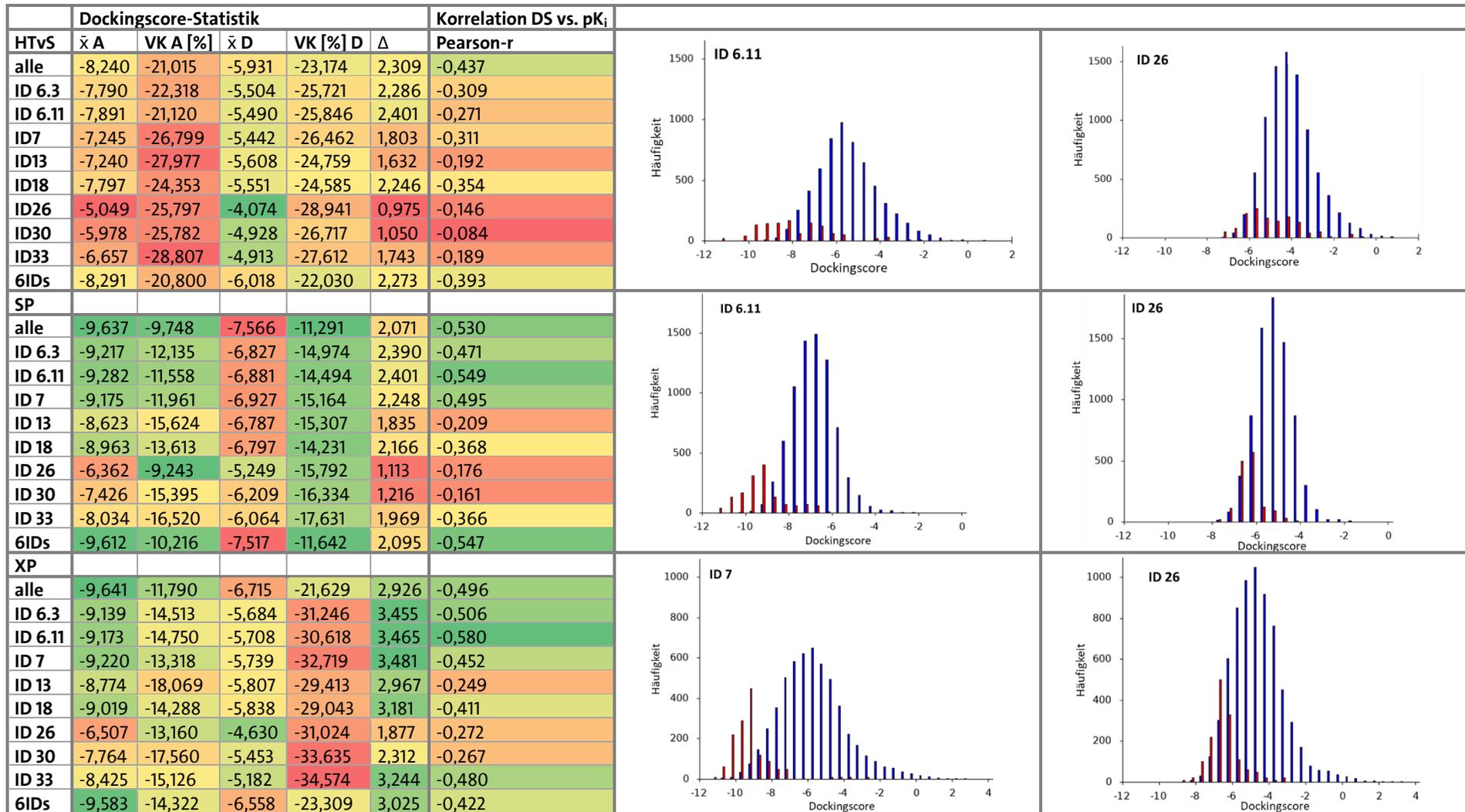
Tab. F 5 – Enrichmentfaktoren und Anzahl gedockter Liganden des HTvs im *OPLS3-B-1-Workflow*. A = Aktive, L = Liganden (Decoys + Aktive). 6IDs: Merging aus ID 6.3, 6.11, 7, 13, 18 und 33. ** 145 Aktive in der *Final-Dockout*-Datei.

SP	Anreicherung EF x%				# gedockte Liganden		
	1%	5%	10%	50%	#A	#L	#A/#L
alle	36,17	15,03	8,48	1,90	145	7718	1,879
ID 6.3	42,31	15,44	8,41	1,92	145	7624	1,902
ID 6.11	42,31	16,00	8,41	1,93	145	7607	1,906
ID 7	28,66	15,03	8,48	1,93	145	7598	1,908
ID 13	36,17	12,82	7,10	1,72	145	7665	1,892
ID 18	42,31	14,75	8,48	1,86	145	7674	1,889
ID 26	6,14	6,34	5,86	1,86	145	7709	1,881
ID 30	8,87	7,72	5,45	1,71	145	7649	1,896
ID 33	36,85	12,69	7,38	1,75	144	7611	1,892
6IDs	38,94	15,01	8,41	1,89	145	7711	1,880

Tab. F 6 – Enrichmentfaktoren und Anzahl gedockter Liganden des SP-Dockings im *OPLS3-B-1-Workflow*. A = Aktive, L = Liganden (Decoys + Aktive). 6IDs: Merging aus ID 6.3, 6.11, 7, 13, 18 und 33.

XP	Anreicherung EF x%				# gedockte Liganden		
	1%	5%	10%	50%	#A	#L	#A/#L
alle	34,80	16,27	8,62	1,92	145	7674	1,889
ID 6.3	40,94	16,27	8,48	1,93	142	5877	2,416
ID 6.11	42,31	15,86	8,48	1,92	141	5991	2,354
ID 7	31,39	15,86	8,62	1,85	138	5626	2,453
ID 13	35,49	14,34	7,65	1,88	139	5954	2,335
ID 18	43,67	15,58	8,48	1,89	139	6051	2,297
ID 26	11,60	9,24	6,76	1,92	145	6998	2,072
ID 30	10,92	11,17	6,76	1,88	143	5781	2,474
ID 33	37,53	15,44	8,48	1,89	139	5162	2,693
6IDs	37,57	16,52	8,62	1,94	145	7401	1,959

Tab. F 7 – Enrichmentfaktoren und Anzahl gedockter Liganden des XP-Dockings im *OPLS3-B-1-Workflow*. A = Aktive, L = Liganden (Decoys + Aktive). 6IDs: Merging aus ID 6.3, 6.11, 7, 13, 18 und 33.



Tab. F 8 – Mittelwert und Variationskoeffizient der Dockingscores der Aktiven (A) bzw. der Decoys (D) sowie die Differenz $\Delta = \bar{x} D - \bar{x} A$ aus dem *OPLS3-B-7-Workflow*. Pearson-Korrelation zwischen Dockingscore und pK_i -Wert.

Abb. F 4 – Häufigkeitsverteilung der Dockingscores der Decoys (blau) und der Aktiven (rot, 10 fach vergrößert). Oben: HTVs, Mitte: SP, unten: XP. Es sind jeweils die Ergebnisse der Protein-konformation mit der besten Trennung der Aktiven von den Decoys (links) und der geringsten Trennleistung (rechts) dargestellt.

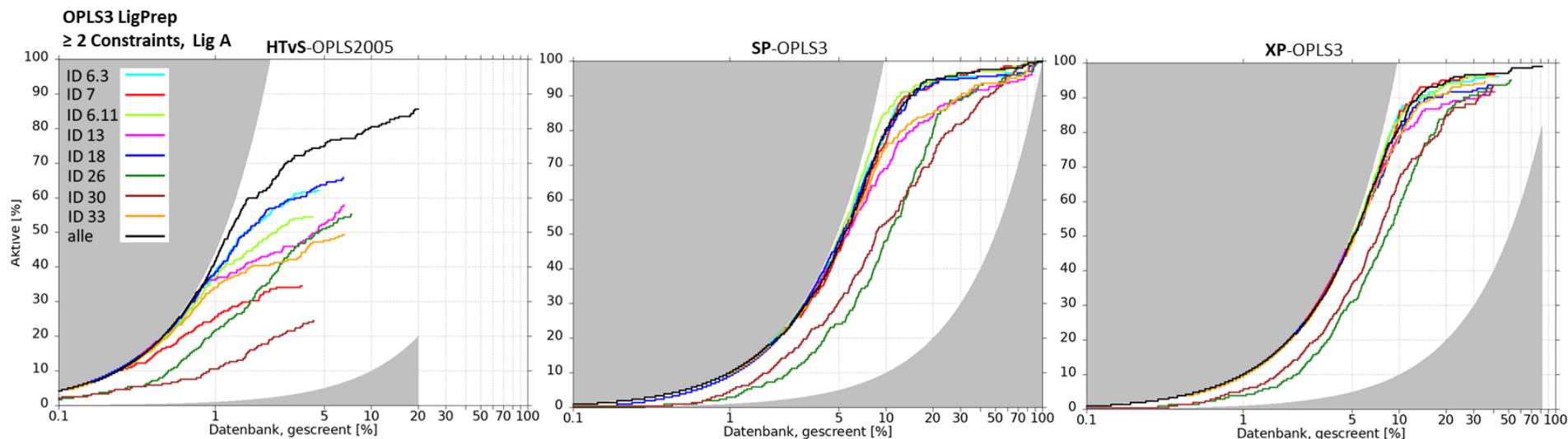


Abb. F 5 – Enrichmentplots des *OPLS3-A-2-Workflow*s für die drei verschiedenen Dockingmodi. Graue Kurve oben: perfekte Anreicherung; graue Kurve unten: „Anreicherung“ der Aktiven bei einer rein zufälligen Verteilung der Aktiven in der Input-Datenbank.

HTvS	Anreicherung EF x%				# gedockte Liganden		
	1%	5%	10%	50%	#A	#L	#A/#L
alle	41,91	15,01	8,06	1,71	203	2102	9,657
ID 6.3	38,98	12,41	6,20	1,24	147	486	30,247
ID 6.11	36,88	10,89	5,44	1,09	129	443	29,120
ID7	25,57	6,92	3,46	0,69	82	379	21,636
ID13	36,46	10,45	5,78	1,16	137	706	19,405
ID18	38,56	12,73	6,58	1,32	156	696	22,414
ID26	21,79	10,20	5,53	1,11	131	782	16,752
ID30	10,48	4,89	2,45	0,49	58	450	12,889
ID33	33,95	9,53	4,94	0,99	117	702	16,667
6IDs	42,33	15,09	7,93	1,65	196	1622	12,084

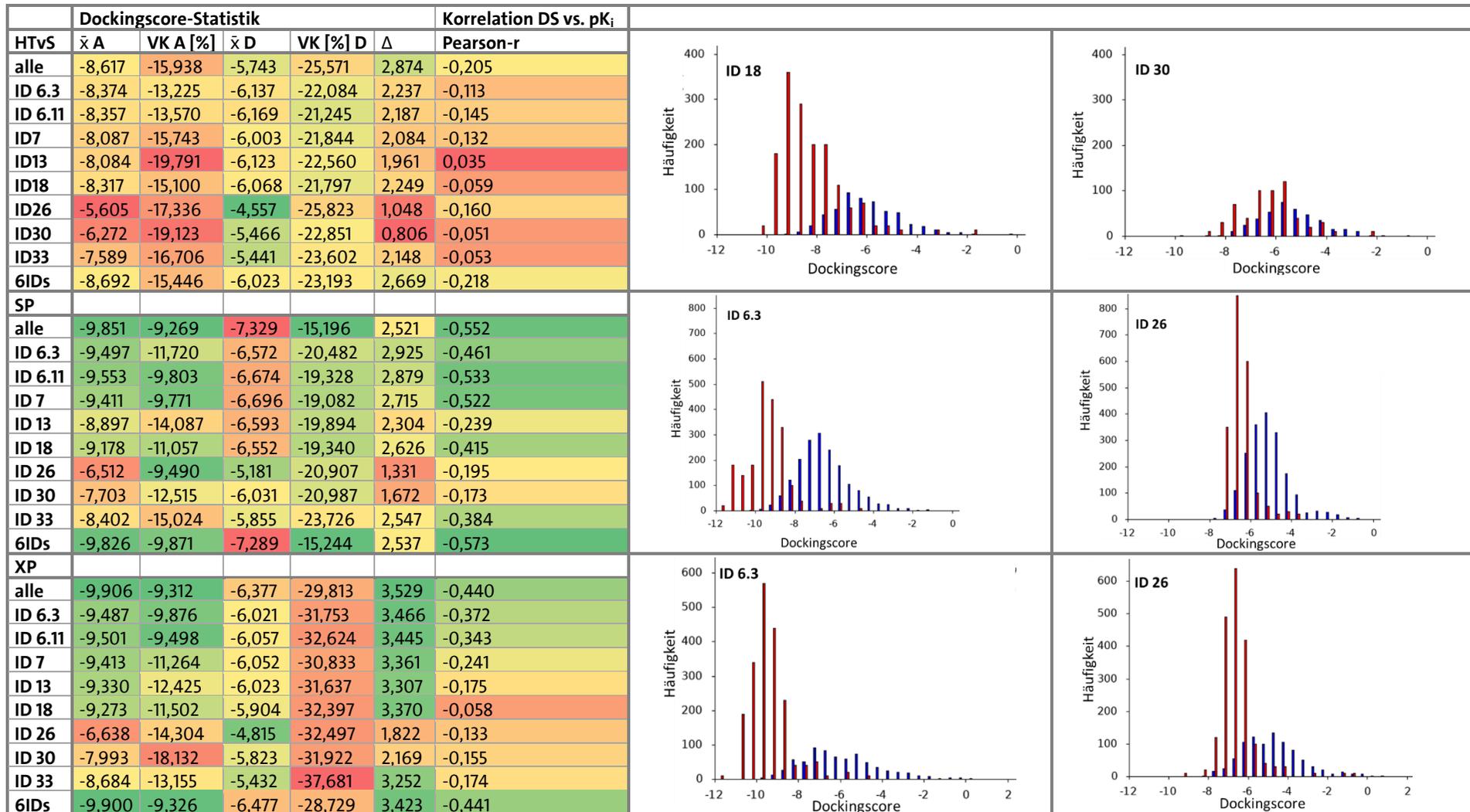
Tab. F 9 – Enrichmentfaktoren und Anzahl gedockter Liganden des HTvS im *OPLS3-A-2-Workflow*. A = Aktive, L = Liganden (Decoys + Aktive). 6IDs: Merging aus ID 6.3, 6.11, 7, 13, 18 und 33.

SP	Anreicherung EF x%				# gedockte Liganden		
	1%	5%	10%	50%	#A	#L	#A/#L
alle	10,35	9,28	8,05	1,95	203	2100	9,667
ID 6.3	10,35	9,57	7,95	1,91	202	1940	10,412
ID 6.11	10,35	9,87	8,49	1,94	202	1963	10,290
ID 7	10,35	8,99	7,70	1,95	202	1968	10,264
ID 13	10,35	9,48	6,87	1,85	201	1930	10,415
ID 18	9,41	9,67	7,95	1,91	202	1948	10,370
ID 26	2,82	4,79	4,81	1,87	202	2083	9,698
ID 30	5,18	6,15	5,35	1,82	203	2001	10,145
ID 33	9,88	9,38	7,51	1,87	201	1812	11,093
6IDs	10,39	9,51	8,08	1,94	203	2068	9,816

Tab. F 10 – Enrichmentfaktoren und Anzahl gedockter Liganden des SP-Dockings im *OPLS3-A-2-Workflow*. A = Aktive, L = Liganden (Decoys + Aktive). 6IDs: Merging aus ID 6.3, 6.11, 7, 13, 18 und 33.

XP	Anreicherung EF x%				# gedockte Liganden		
	1%	5%	10%	50%	#A	#L	#A/#L
alle	10,34	9,95	8,13	1,94	201	1725	11,652
ID 6.3	10,34	9,95	8,62	1,92	195	903	21,595
ID 6.11	10,34	10,05	8,37	1,92	195	854	22,834
ID 7	10,34	9,85	8,57	1,94	197	854	23,068
ID 13	10,34	9,75	7,93	1,83	186	852	21,831
ID 18	10,34	9,75	7,83	1,87	190	860	22,093
ID 26	3,94	6,21	5,86	1,88	193	1092	17,674
ID 30	5,91	7,29	6,70	1,87	190	850	22,353
ID 33	9,85	9,66	7,88	1,89	192	744	25,806
6IDs	10,38	10,09	8,32	1,97	201	1433	14,027

Tab. F 11 – Enrichmentfaktoren und Anzahl gedockter Liganden des XP-Dockings im *OPLS3-A-2-Workflow*. A = Aktive, L = Liganden (Decoys + Aktive). 6IDs: Merging aus ID 6.3, 6.11, 7, 13, 18 und 33.



Tab. F 12 – Mittelwert und Variationskoeffizient der Dockingscores der Aktiven (A) bzw. der Decoys (D) sowie die Differenz $\Delta = \bar{x} D - \bar{x} A$ aus dem OPLS3-A-2-Workflow. Pearson-Korrelation zwischen Dockingscore und pK_i -Wert.

Abb. F 6 – Häufigkeitsverteilung der Dockingscores der Decoys (blau) und der Aktiven (rot, 10 fach vergrößert). Oben: HTvS, Mitte: SP, unten: XP. Es sind jeweils die Ergebnisse der Proteinkonformation mit der besten Trennung der Aktiven von den Decoys (links) und der geringsten Trennleistung (rechts) dargestellt.

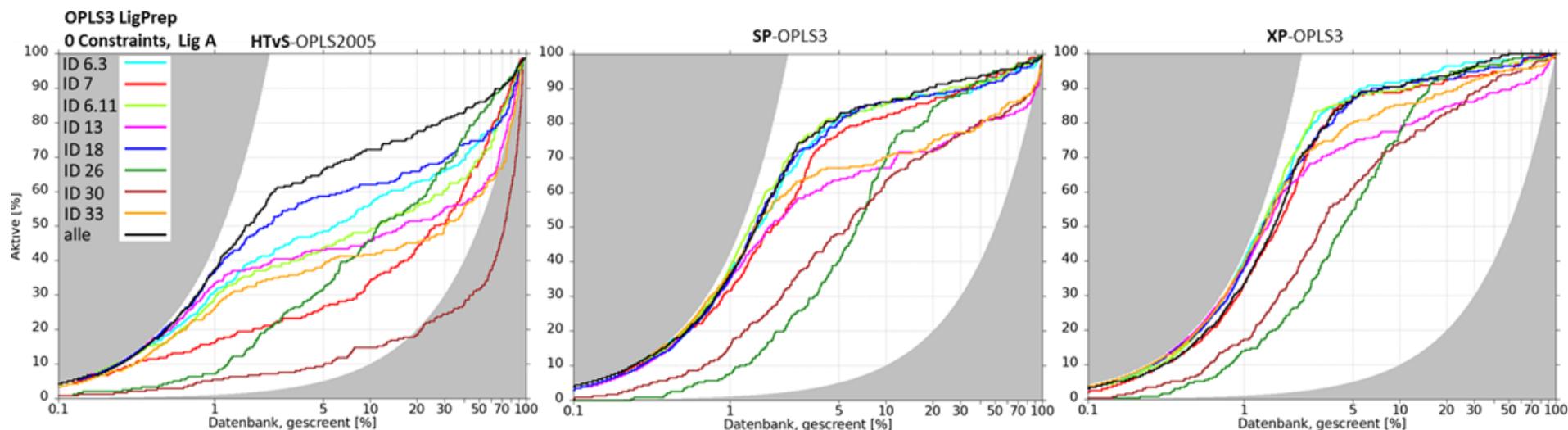


Abb. F 7 – Enrichmentplots des *OPLS3-A-0-Workflows* für die drei verschiedenen Dockingmodi. Graue Kurve oben: perfekte Anreicherung; graue Kurve unten: „Anreicherung“ der Aktiven bei einer rein zufälligen Verteilung der Aktiven in der Input-Datenbank.

HTVS	Anreicherung EF x%				# gedockte Liganden		
	1%	5%	10%	50%	#A	#L	#A/#L
alle	37,30	13,24	7,21	1,72	234**	10432	2,243
ID 6.3	31,01	9,69	5,61	1,49	234	9993	2,342
ID 6.11	29,34	8,77	4,85	1,38	231	9870	2,34
ID7	16,35	5,40	3,42	1,38	234	9907	2,362
ID13	33,53	8,60	4,60	1,21	234	10154	2,305
ID18	36,88	11,72	6,20	1,49	234	10151	2,305
ID26	7,12	6,32	4,56	1,65	234	10376	2,255
ID30	5,45	1,94	1,48	0,63	226	10040	2,251
ID33	26,40	7,76	4,18	1,17	233	10069	2,314
6IDs	37,72	13,32	7,21	1,71	234	10398	2,250

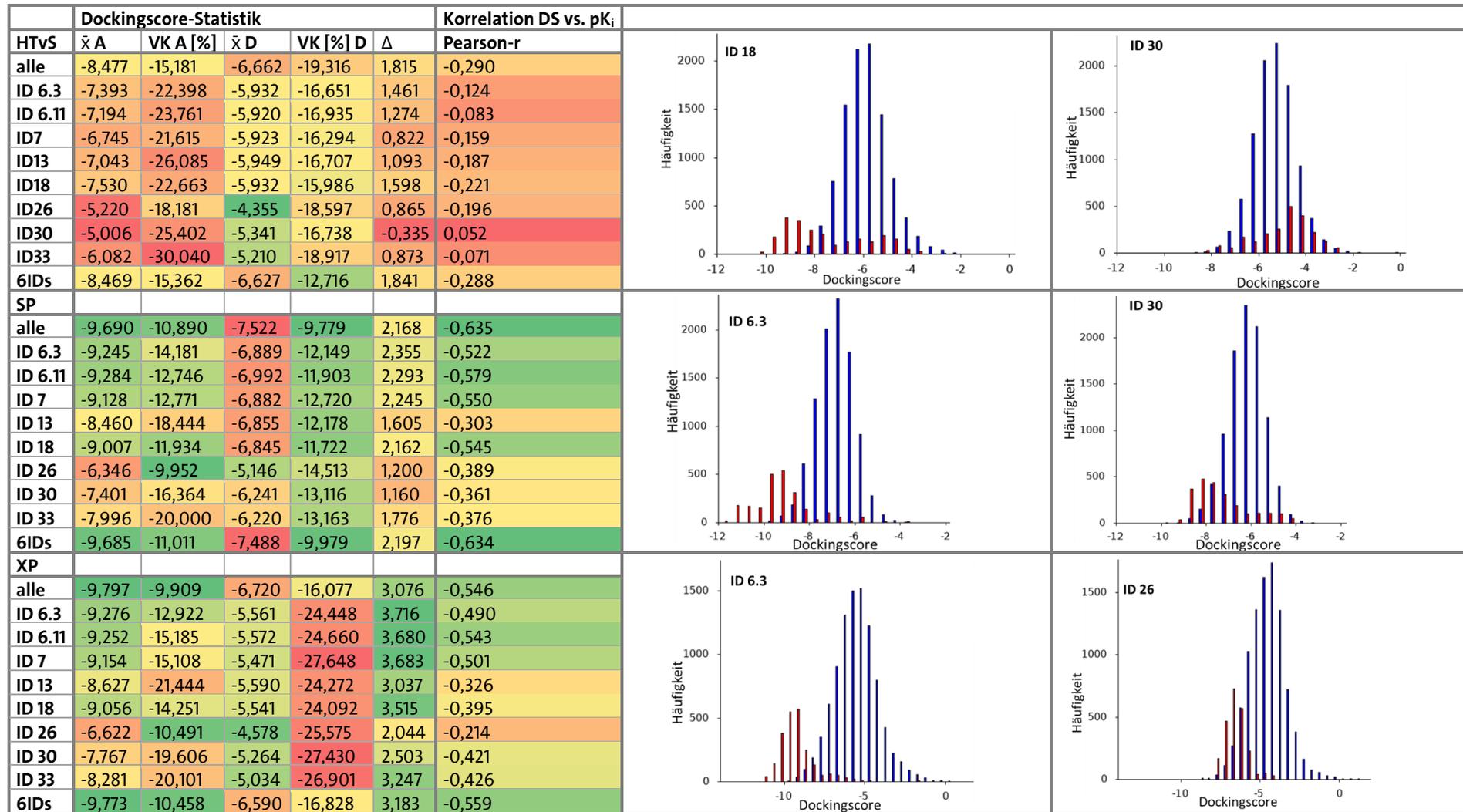
Tab. F 13 – Enrichmentfaktoren und Anzahl gedockter Liganden des HTVS im *OPLS3-A-0-Workflow*. A = Aktive, L = Liganden (Decoys + Aktive). 6IDs: Merging aus ID 6.3, 6.11, 7, 13, 18 und 33. ** 231 Aktive in der *Final-Dockout*-Datei.

SP	Anreicherung EF x%				# gedockte Liganden		
	1%	5%	10%	50%	#A	#L	#A/#L
alle	36,07	16,45	8,61	1,90	230	9817	2,343
ID 6.3	37,36	16,28	8,57	1,86	230	9816	2,343
ID 6.11	38,65	16,36	8,57	1,85	230	9814	2,344
ID 7	31,35	15,50	8,18	1,87	230	9810	2,345
ID 13	35,21	12,73	6,71	1,63	230	9815	2,343
ID 18	35,21	16,10	8,61	1,84	230	9817	2,343
ID 26	7,73	8,05	6,97	1,90	230	9815	2,343
ID 30	15,89	9,61	6,36	1,64	230	9807	2,345
ID 33	36,93	13,42	7,01	1,65	230	9816	2,343
6IDs	36,53	16,60	8,61	1,89	230	9817	2,343

Tab. F 14 – Enrichmentfaktoren und Anzahl gedockter Liganden des SP-Dockings im *OPLS3-A-0-Workflow*. A = Aktive, L = Liganden (Decoys + Aktive). 6IDs: Merging aus ID 6.3, 6.11, 7, 13, 18 und 33. c

XP	Anreicherung EF x%				# gedockte Liganden		
	1%	5%	10%	50%	#A	#L	#A/#L
alle	33,63	17,47	9,04	2,00	230	9817	2,343
ID 6.3	40,53	17,56	9,17	1,97	230	9817	2,343
ID 6.11	39,23	17,39	8,95	1,94	230	9817	2,343
ID 7	33,63	17,56	8,87	1,90	230	9817	2,343
ID 13	38,37	14,87	7,78	1,79	230	9817	2,343
ID 18	38,37	17,39	9,04	1,93	230	9817	2,343
ID 26	13,80	10,95	7,74	1,97	230	9817	2,343
ID 30	17,25	12,17	7,43	1,88	230	9816	2,343
ID 33	39,66	16,00	8,52	1,90	230	9817	2,343
6IDs	35,81	17,71	9,13	1,98	230	9817	2,343

Tab. F 15 – Enrichmentfaktoren und Anzahl gedockter Liganden des XP-Dockings im *OPLS3-A-0-Workflow*. A = Aktive, L = Liganden (Decoys + Aktive). 6IDs: Merging aus ID 6.3, 6.11, 7, 13, 18 und 33. c



Tab. F 16 – Mittelwert und Variationskoeffizient der Dockingscores der Aktiven (A) bzw. der Decoys (D) sowie die Differenz $\Delta = \bar{x} D - \bar{x} A$ aus dem *OPLS3-A-O-Workflow*. Pearson-Korrelation zwischen Dockingscore und pK_i -Wert.

Abb. F 8 – Häufigkeitsverteilung der Dockingscores der Decoys (blau) und der Aktiven (rot, 10 fach vergrößert). Oben: HTvS, Mitte: SP, unten: XP. Es sind jeweils die Ergebnisse der Proteinkonformation mit der besten Trennung der Aktiven von den Decoys (links) und der geringsten Trennleistung (rechts) dargestellt.

G. Ergebnisse der Bewertung des ZINC-Actives-Screenings mit den Einzelscorewerten

D	EF1%	EF5%	EF10%	EF50%	Pearson-r
ID 6.3	26,07	11,92	7,70	1,96	-0,214
ID 6.11	22,20	11,60	7,84	1,96	-0,321
ID 7	11,31	5,53	4,21	1,80	-0,304
ID 13	14,05	9,75	5,95	1,82	-0,207
ID 18	27,64	12,29	7,16	1,91	-0,321
ID 33	14,27	7,82	5,71	1,82	-0,312
alle	28,91	13,98	8,31	1,96	-0,351

H	EF1%	EF5%	EF10%	EF50%	Pearson-r
ID 6.3	14,27	6,95	4,78	1,76	-0,257
ID 6.11	9,25	6,79	4,81	1,85	-0,416
ID 7	10,67	8,17	5,22	1,81	-0,334
ID 13	8,40	4,87	4,03	1,50	-0,143
ID 18	12,22	6,39	4,15	1,54	-0,257
ID 33	30,99	9,69	5,91	1,73	-0,214
alle	35,54	11,93	7,11	1,95	-0,462

Q	EF1%	EF5%	EF10%	EF50%	Pearson-r
ID 6.3	21,10	8,82	5,84	1,75	0,563
ID 6.11	19,12	7,65	5,31	1,74	0,597
ID 7	3,77	2,39	2,26	1,48	0,604
ID 13	19,83	7,60	5,12	1,70	0,460
ID 18	14,83	7,16	4,80	1,76	0,552
ID 33	23,29	9,47	6,69	1,88	0,446
alle	13,25	7,71	5,36	1,75	0,586

M	EF1%	EF5%	EF10%	EF50%	Pearson-r
ID 6.3	19,24	7,20	4,47	1,59	-0,333
ID 6.11	12,95	6,42	4,32	1,62	-0,318
ID 7	10,05	6,04	4,21	1,60	-0,328
ID 13	14,05	6,45	4,05	1,42	-0,088
ID 18	16,85	8,38	5,13	1,77	-0,296
ID 33	11,27	4,81	3,31	1,56	-0,262
alle	18,67	8,92	5,24	1,67	-0,416

Tab. G 1 – Enrichment- und Pearson-Korrelationsfaktoren nach Sortierung und Zusammenführung der 6 Top-5%-Ranglisten des ZINC-Screenings (SP) nach dem jeweiligen Scorewert. D = Dockingscore, H = HYDE-Score, Q = vorhergesagter pK_i-Wert (QSAR), M = Freie Bindungsenergie (PRIME MM/GBSA)

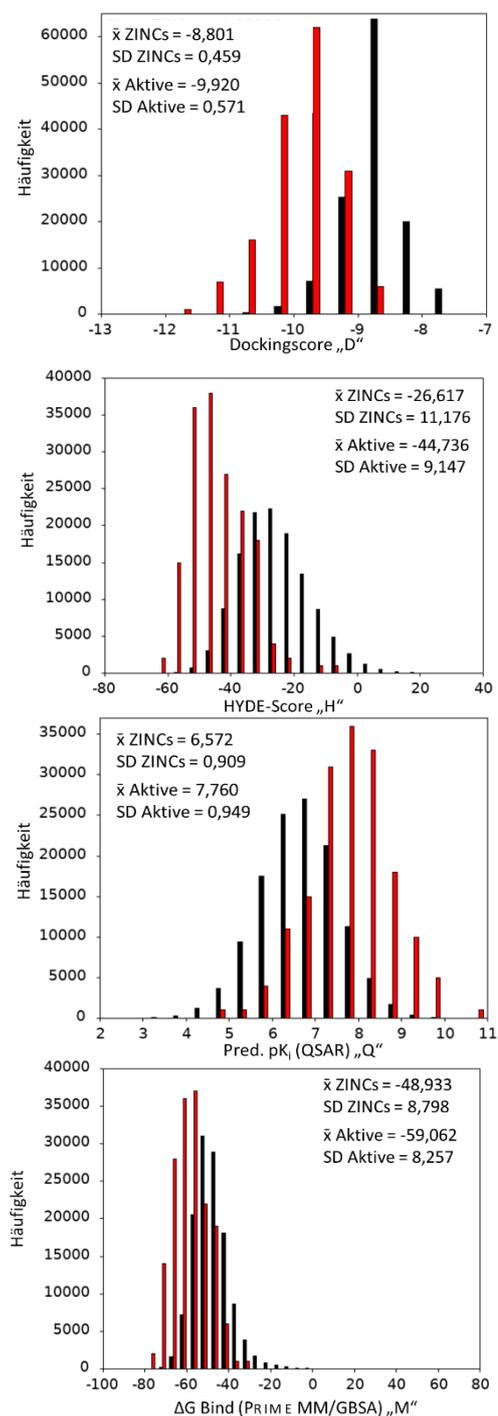


Abb. G 1 – Häufigkeitsverteilung der Activen (rot, 1000 fach vergrößert dargestellt) und ZINCs (schwarz) je nach Scorewert (entspr. Tab. G 1).

H. Referenz- und Testsubstanzlösungen

Tab. H 1 – DMSO-Zusatz zur Herstellung der Testsubstanzlösungen (A01 bis F02) und der GNE-3511-Lösung. Menge [mg] = deklarierte Menge lt. Hersteller. MW-Daten von www.ambinter.com.

Substanz	MW	Menge [mg]	DMSO [μ l] Soll	DMSO [μ l] Ist
A01	337,376	1	296,41	296,4
A02	396,410	1	252,26	252,2
A03	354,426	1	282,15	282,0
A04	394,490	1	253,49	253,5
A05	348,398	1,1	315,73	315,6
A06	314,429	1,1	349,84	349,8
A07	327,427	1,2	366,49	366,4
A08	346,404	1,2	346,42	346,4
A09	367,441	1,1	299,37	299,3
A10	304,389	1,1	361,38	361,4
A11	350,418	1	285,37	285,4
A12	371,434	1	269,23	269,2
B01	326,419	1,2	367,63	367,6
B02	374,381	1,2	320,53	320,4
B03	275,308	1	363,23	363,2
B04	378,467	1	264,22	264,2
B05	318,376	1,2	376,91	376,8
B06	301,774	1,2	397,65	397,6
B07	289,335	1	345,62	345,6
B08	307,393	1	325,32	325,2
B09	374,360	1,2	320,55	320,4
B10	279,340	1	357,99	360,0
B11	306,362	1,1	359,05	359,0
C01	341,431	1,1	322,17	322,0
C02	380,420	1,1	289,15	289,1
C03	346,404	1,1	317,55	317,4
C05	304,303	1,1	361,48	361,4
C06	360,474	1,2	332,90	332,8
C07	336,348	1,1	327,04	327,0
C08	349,256	1,2	343,59	343,4
C09	305,258	1,2	393,11	393,0
C10	329,355	1,1	333,99	334,0
C11	347,409	1,2	345,41	345,4
C12	312,389	1,2	384,14	384,0
D01	308,262	1,1	356,84	356,8
D02	364,421	1,1	301,85	301,8
D03	293,363	1,1	374,96	374,8
D04	308,259	1,1	356,84	356,8
D05	307,350	1,1	357,90	357,8
D06	325,408	1,2	368,77	368,6
D07	324,346	1,2	369,98	369,8
D08	339,392	1,1	324,11	324,0
D09	340,423	1,2	352,50	352,4
D10	347,410	1,1	316,63	316,6
D11	397,469	1,2	301,91	301,8
D12	299,368	1,2	400,84	400,8
E01	301,307	1,2	398,26	398,2
E02	340,443	1,1	323,11	323,0
E03	279,260	1,1	393,90	393,8
E04	314,382	1,1	349,89	349,8
E05	361,420	1,1	304,36	361,4
E06	322,285	1,1	341,31	341,2
E07	414,473	1,1	265,40	400,0
E08	336,351	1,1	327,04	336,2
E12	329,390	1	303,59	303,6
F01	298,260	1,2	402,33	402,2
F02	340,327	1,4	411,37	411,2
GNE-3511	440,489	1	227,02	227,0 → 1,25 μ l dieser Lösung + 3,75 μ l DMSO

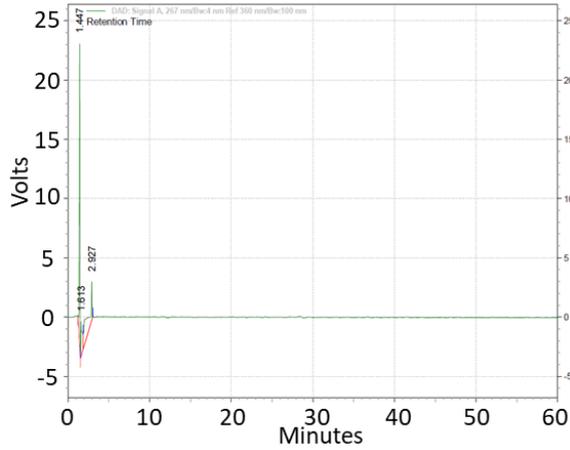
Tab. H 2 – Einwaage und DMSO-Zusatz zur Herstellung von Test- und Referenzlösungen. MW = Molekulargewicht: Daten von www.ambinter.com.

Substanz	MW	Einwaage [mg]	DMSO [μ l] Soll	DMSO [μ l] Ist
E11	254,241	1,82	715,86	715,8
Crizotinib	450,337	2,13	472,98	473,0
Sunitinibmalat	532,561	1	187,77	187,8
Bosutinib	530,446	5	94,26	94,2 → 1 μ l dieser Lösung + 9 μ l DMSO

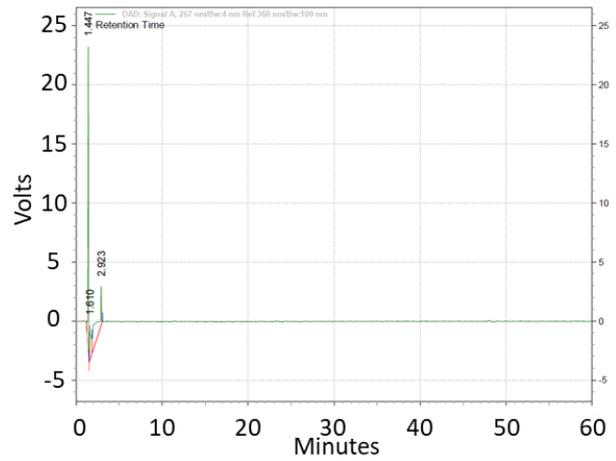
I. Chromatogramme der HPLC-Analytik

Blank (100 μ l DMSO + 900 μ l Acetonitril-Wasser-Mischung 50:50)

- Blank Methode 1 (s. Kap. 7.18), zweifache Messung der gleichen Lösung, 267 nm

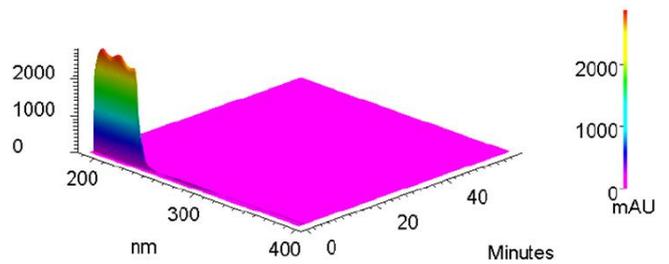
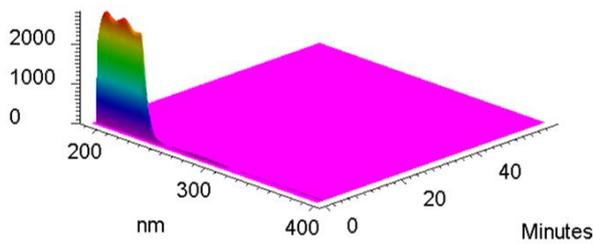


Retention Time	Area	Area [%]
1,447	229537	46,76
1,613	63354	12,91
2,927	198022	40,34

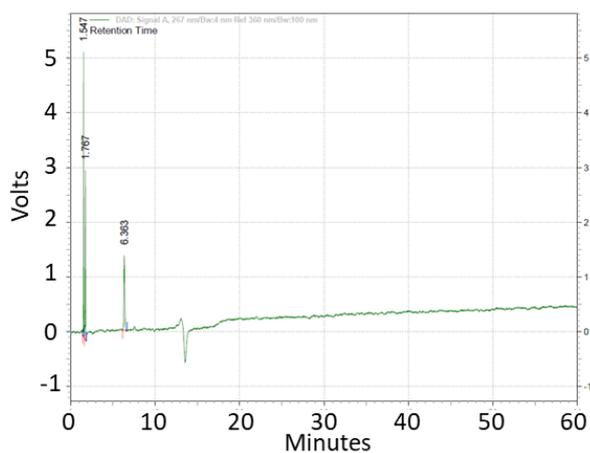


Retention Time	Area	Area [%]
1,447	226138	46,53
1,610	64975	13,37
2,923	194912	40,10

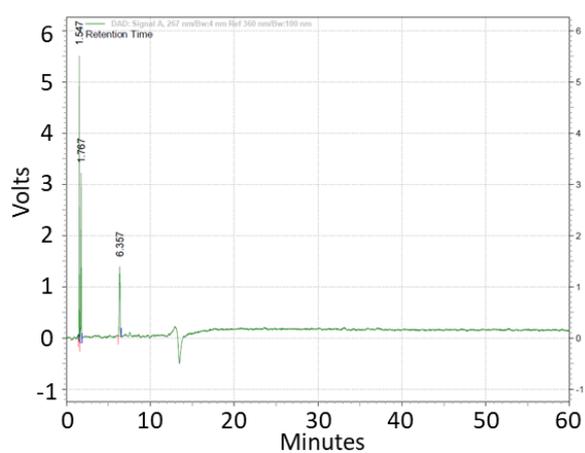
3D-Chromatogramme (190-400 nm):



- **Blank Methode 2** (s. Kap. 7.18), zweifache Messung der gleichen Lösung, 267 nm

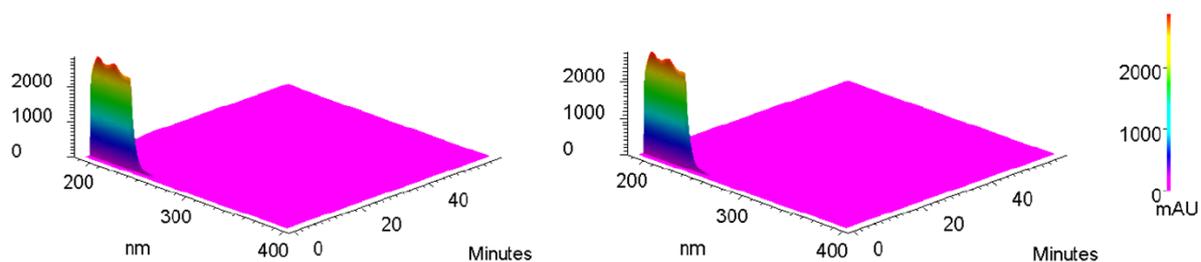


Retention Time	Area	Area [%]
1,547	22201	36,39
1,767	18671	30,60
6,363	20138	33,01



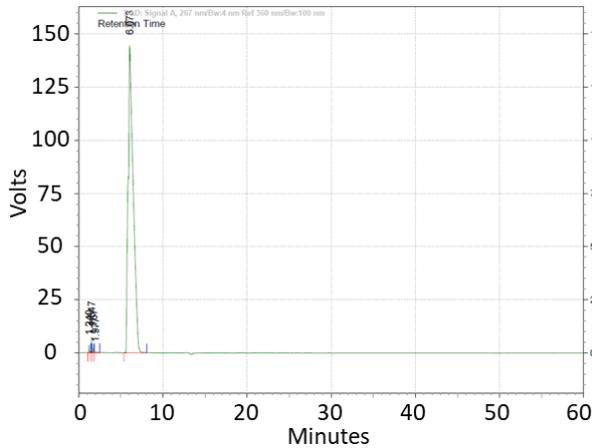
Retention Time	Area	Area [%]
1.547	22915	36,24
1.767	20734	32,79
6.357	19579	30,97

3D-Chromatogramme (190-400 nm):

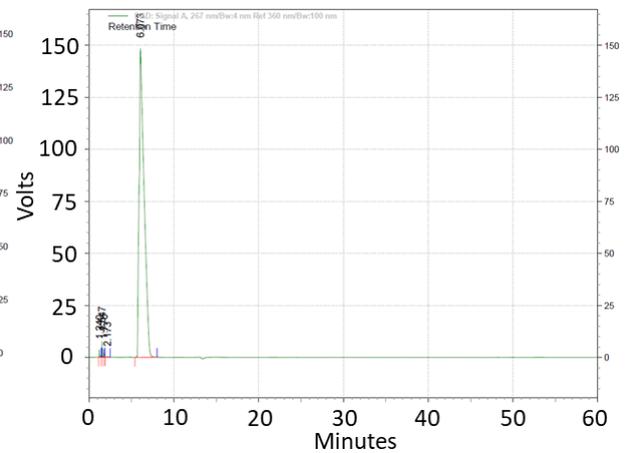


Crizotinib (10 µl 10 mM-Lösung in DMSO + 90 µl ACN:H₂O 50:50)

- **Crizotinib Methode 2** (s. Kap. 7.18), zweifache Messung der gleichen Lösung, 267 nm

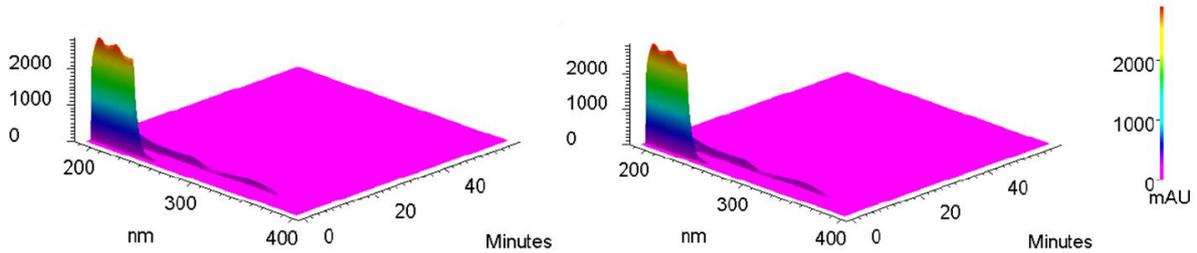


Retention Time	Area	Area [%]
1,240	19064	n.b.
1,547	32309	n.b.
1,767	23689	n.b.
1,977	8086	0,06
6,073	12641745	99,94



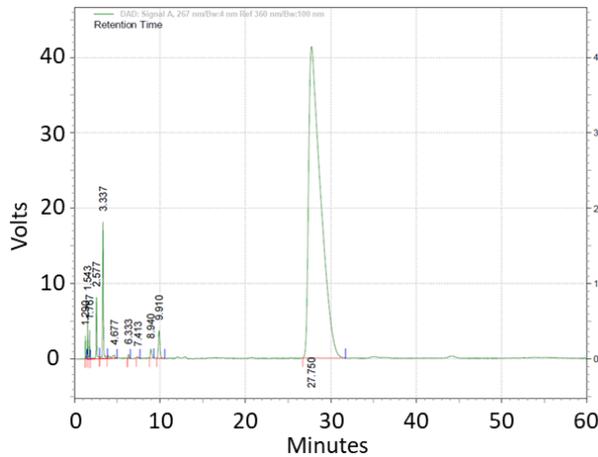
Retention Time	Area	Area [%]
1,240	19368	n.b.
1,547	32186	n.b.
1,770	23288	n.b.
2,173	6330	0,05
6,073	12670884	99,95

3D-Chromatogramme (190-400 nm):

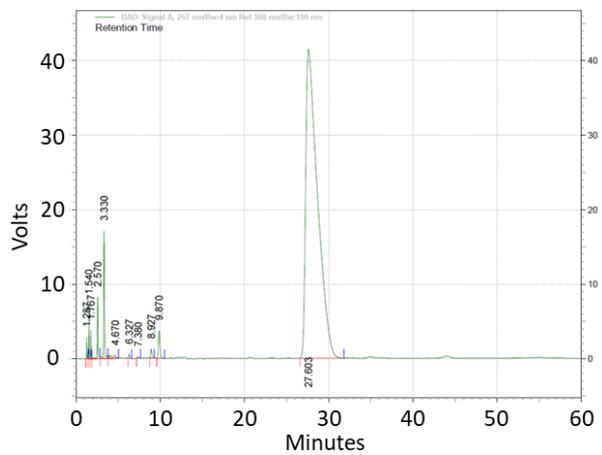


A04 (10 µl 10 mM-Lösung in DMSO + 90 µl ACN:H₂O 50:50)

- A04 Methode 2 (s. Kap. 7.18), zweifache Messung der gleichen Lösung, 267 nm

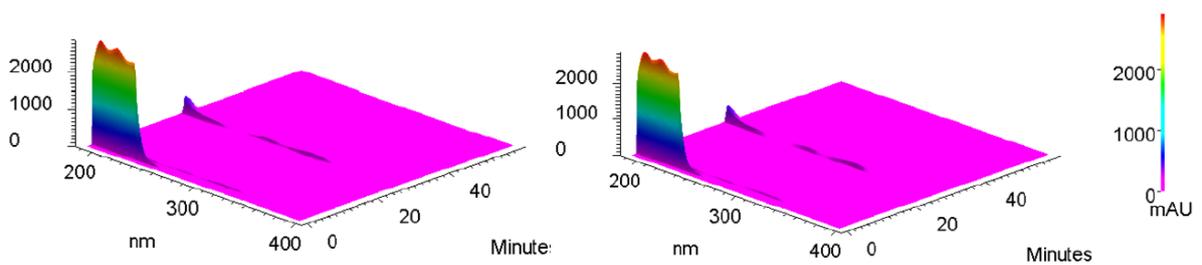


Retention Time	Area	Area [%]
1,290	15897	n,b,
1,543	39370	n,b,
1,767	23707	n,b,
2,577	104649	1,17
3,337	258442	2,89
4,677	30859	0,34
6,333	7262	0,08
7,413	4280	0,05
8,940	28916	0,32
9,910	96777	1,08
27,750	8418533	94,06



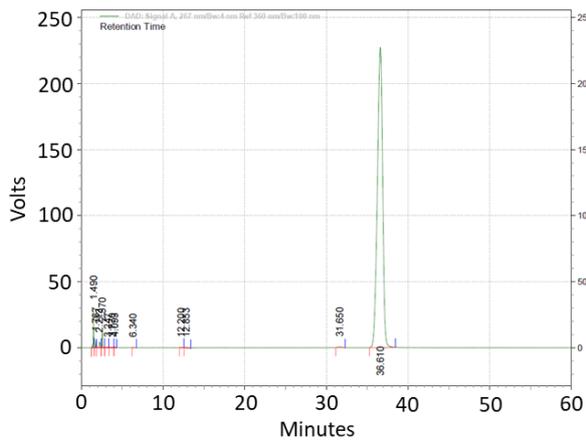
Retention Time	Area	Area [%]
1,287	15859	n.b.
1,540	38508	n.b.
1,767	23136	n.b.
2,570	104688	1,17
3,330	259043	2,89
4,670	32716	0,36
6,327	8380	0,09
7,380	4784	0,05
8,927	29878	0,33
9,870	96536	1,08
27,603	8441585	94,03

3D- Chromatogramme (190-400 nm):

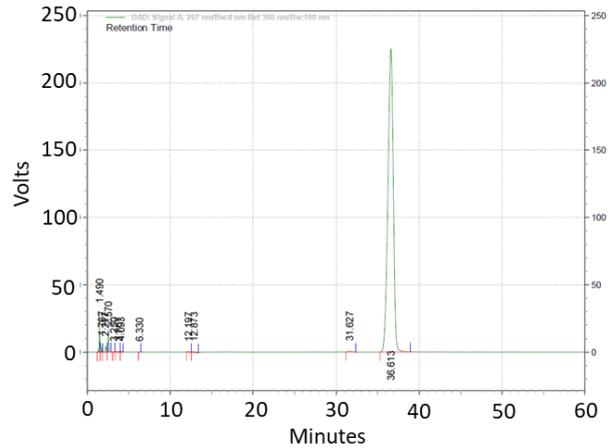


C07 (10 µl 10 mM-Lösung in DMSO + 90 µl ACN:H₂O 50:50)

- C07 Methode 2 (s. Kap. 7.18), zweifache Messung der gleichen Lösung, 267 nm

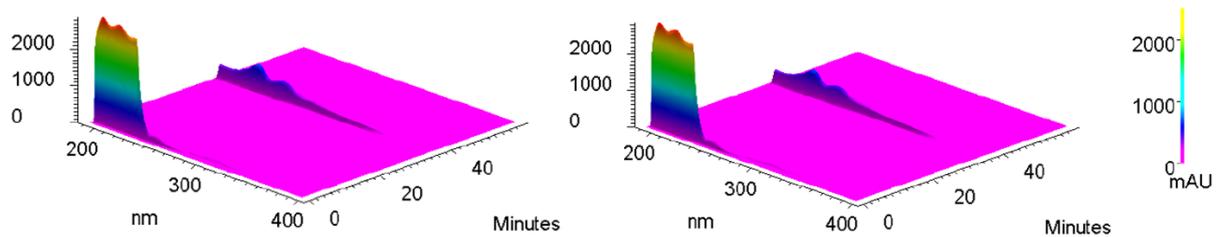


Retention Time	Area	Area [%]
1,490	145833	n,b
1,767	28178	n,b
2,213	71114	0,33
2,570	132526	0,62
3,247	12307	0,06
3,770	17961	0,08
4,093	7895	0,04
6,340	5076	0,02
12,200	10145	0,05
12,853	35766	0,17
31,650	39563	0,18
36,610	21181902	98,46



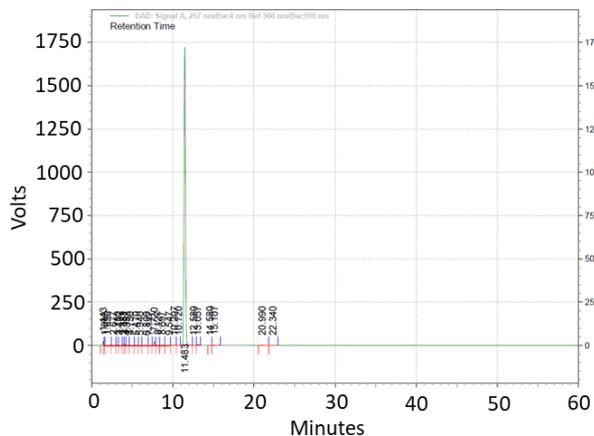
Retention Time	Area	Area [%]
1,490	146659	n,b
1,767	27751	n,b
2,217	69891	0,32
2,570	130088	0,60
3,250	7322	0,03
3,767	16817	0,08
4,093	8369	0,04
6,330	4287	0,02
12,197	12440	0,06
12,873	34625	0,16
31,627	38290	0,18
36,613	21268250	98,51

3D-Chromatogramme (190-400 nm):

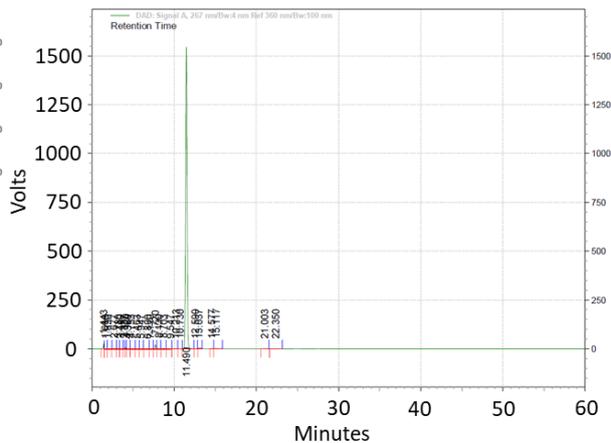


D04 (10 µl 10 mM-Lösung in DMSO + 90 µl ACN:H₂O 50:50)

- **D04 Methode 1** (s. Kap. 7.18), zweifache Messung der gleichen Lösung, 267 nm

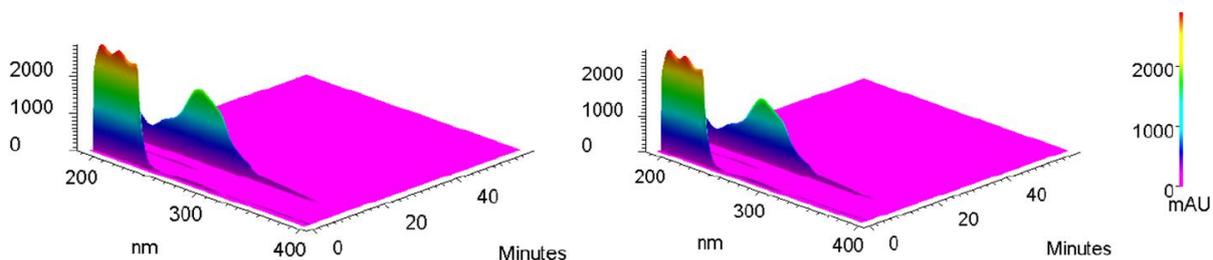


Retention Time	Area	Area [%]
1,443	296507	n.b.
1,613	49663	n.b.
1,890	279374	n.b.
2,677	276717	0,57
3,180	132811	0,28
3,437	173794	0,36
3,883	96120	0,20
4,117	132380	0,27
4,353	146059	0,30
4,750	206931	0,43
5,470	146173	0,30
5,940	117066	0,24
6,800	210301	0,44
7,117	100295	0,21
7,720	478421	0,99
8,127	124847	0,26
8,697	152732	0,32
9,547	109580	0,23
10,207	110313	0,23
10,720	56789	0,12
11,483	45268427	93,79
12,580	34928	0,07
13,057	11008	0,02
14,580	13306	0,03
15,107	71902	0,15
20,990	63629	0,13
22,340	29096	0,06



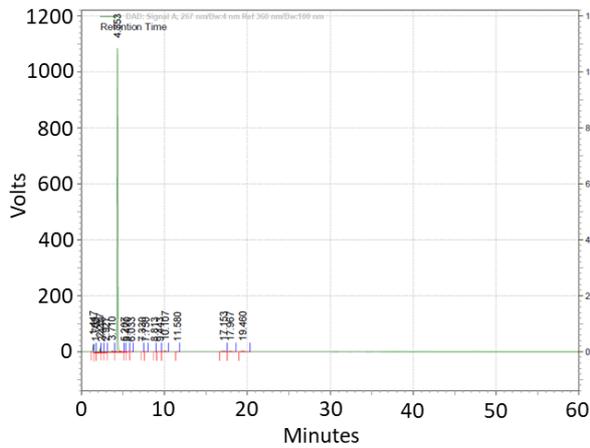
Retention Time	Area	Area [%]
1,443	270686	n.b.
1,610	78569	n.b.
1,890	208232	n.b.
2,677	237379	0,55
3,180	148209	0,34
3,440	130702	0,30
3,887	88964	0,21
4,120	117782	0,27
4,360	134808	0,31
4,753	177664	0,41
5,463	129126	0,30
5,947	127724	0,30
6,800	157444	0,37
7,110	97261	0,23
7,720	412855	0,96
8,130	109210	0,25
8,703	132467	0,31
9,547	89718	0,21
10,213	95800	0,22
10,730	47039	0,11
11,490	40446265	93,89
12,590	26122	0,06
13,037	9964	0,02
14,577	11514	0,03
15,117	65238	0,15
21,003	56472	0,13
22,350	29097	0,07

3D-Chromatogramme (190-400 nm):

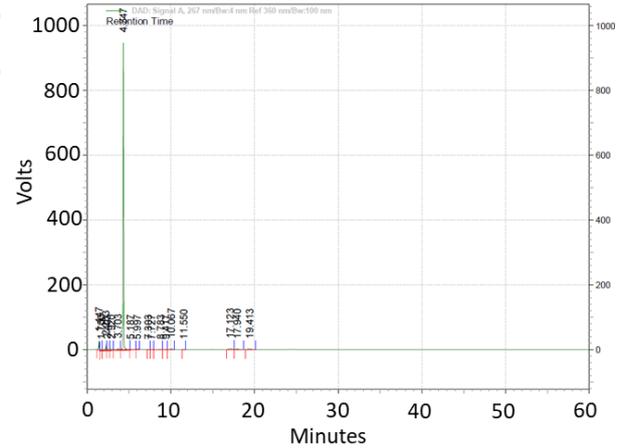


E05 (10 µl 10 mM-Lösung in DMSO + 90 µl ACN:H₂O 50:50)

- E05 Methode 1 (s. Kap. 7.18), zweifache Messung der gleichen Lösung, 267 nm:

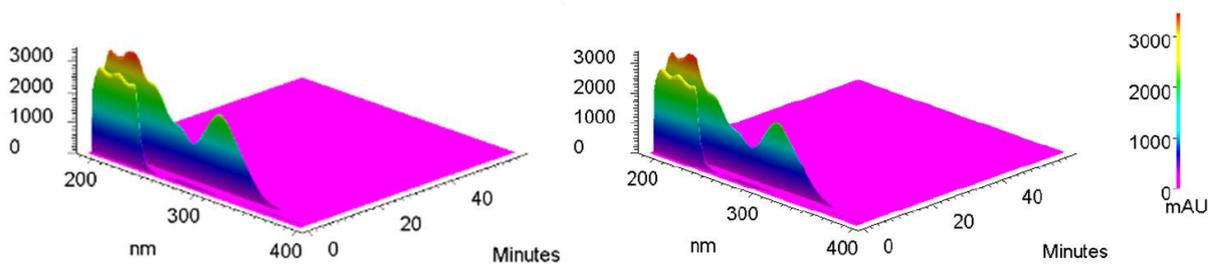


Retention Time	Area	Area [%]
1,447	295577	n.b.
1,703	70028	n.b.
2,257	303092	2,39
2,410	154107	1,22
2,927	165653	1,31
3,710	301465	2,38
4,353	11197201	88,39
5,207	35670	0,28
5,470	41874	0,33
6,033	15133	0,12
7,330	3742	0,03
7,750	4538	0,04
8,813	6000	0,05
9,443	4340	0,03
10,107	76419	0,60
11,580	6057	0,05
17,153	150243	1,19
17,967	75323	0,59
19,460	126495	1,00



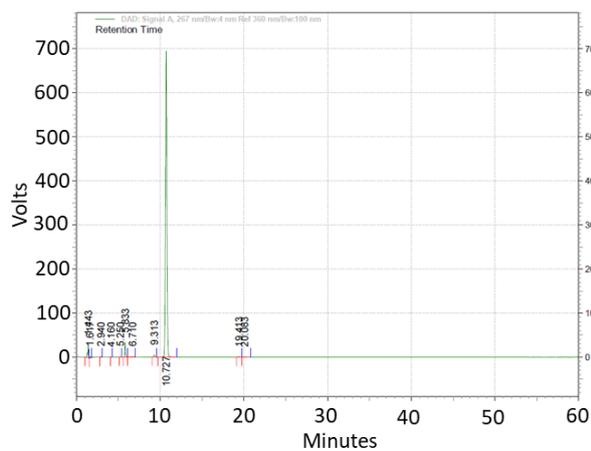
Retention Time	Area	Area [%]
1,447	274570	n.b.
1,700	67276	n.b.
2,253	282383	2,53
2,403	149890	1,34
2,920	159106	1,42
3,703	281144	2,52
4,347	9802256	87,79
5,187	77658	0,70
5,997	12158	0,11
7,303	3972	0,04
7,727	3161	0,03
8,783	10331	0,09
9,417	4811	0,04
10,067	68412	0,61
11,550	4960	0,04
17,123	129671	1,16
17,940	66953	0,60
19,413	108960	0,98

3D-Chromatogramme (190-400 nm):

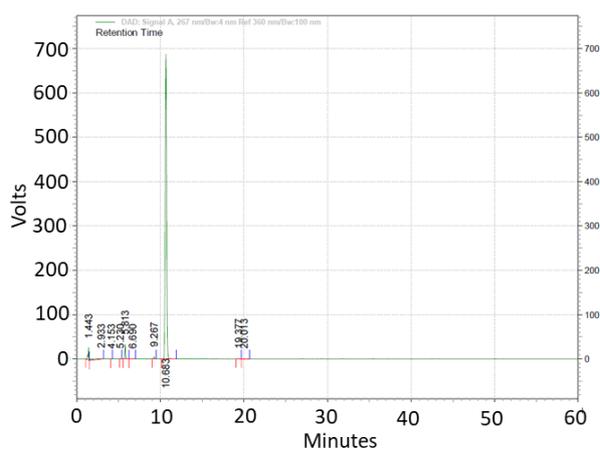


E06 (10 µl 10 mM-Lösung in DMSO + 90 µl ACN:H₂O 50:50)

- E06 Methode 1 (s. Kap. 7.18), zweifache Messung der gleichen Lösung, 267 nm

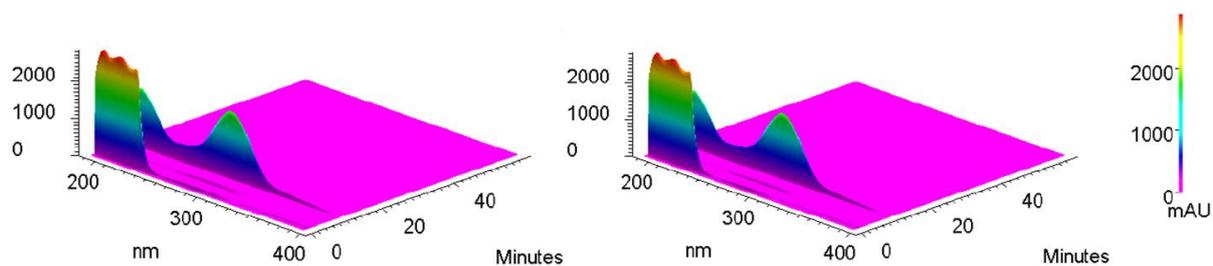


Retention Time	Area	Area [%]
1,443	393112	n.b.
1,617	44844	n.b.
2,940	9140	0,05
4,160	4079	0,02
5,250	7992	0,05
5,833	407404	2,32
6,710	19904	0,11
9,313	97465	0,55
10,727	17015062	96,69
19,413	23035	0,13
20,083	14060	0,08



Retention Time	Area	Area [%]
1,443	340450	n.b.
2,933	314811	1,80
4,153	3794	0,02
5,230	7947	0,05
5,813	401606	2,29
6,690	14881	0,09
9,267	92803	0,53
10,683	16632336	95,03
19,377	19263	0,11
20,013	14840	0,08

3D-Chromatogramme (190-400 nm):



J. Gefahrstoffverzeichnis

Tab. J1 – Verwendete Gefahrstoffe in dieser Arbeit und deren Kennzeichnung nach GHS (*Globally Harmonized System*) mit H (*Hazard*)- und P (*Precautionary*)-Sätzen

Gefahrstoff	GHS-Kennzeichnung	H-Sätze	P-Sätze
Acetonitril	GHS02, GHS07	H225, H302+H312+H332, H319	P210, P280, P301+P312, P303+P361+P353, P304+P340+P312, P305+P351+P338
Acrylamid/Bis-Acrylamid 37.5:1 30 %	GHS07, GHS08	H302+H332, H315, H317, H319, H340, H350, H361f, H372	P280, P301+P312, P302+P352, P304+P340+P312, P305+P351+P338, P308+P313
Ammoniumperoxodisulfat (APS)	GHS03, GHS07, GHS08	H272, H302, H315, H317, H319, H334, H335	P210, P280, P301+P312, P302+P352, P304+P340+P312, P305+P351+P338
Bosutinib	GHS07	H319, H413	P305+P351+P338
Crizotinib	GHS07, GHS08, GHS09	H317, H319, H341, H400	P201, P273, P280, P302+P352, P305+P351+P338, P308+P313
1,4-Dithiothreitol (DTT)	GHS05, GHS07	H302, H315, H318	P264, P270, P280, P301+P312, P302+P352, P305+P351+P338
Eisessig	GHS02, GHS05	H226, H314	P210, P233, P240, P280, P303+P361+P353
Ethylendiamintetraessigsäure (EDTA)	GHS07	H319	P305+P351+P338
GNE-3511	GHS07	H302, H315, H319, H335	P261, P264, P270, P271, P280, P301+P312, P330, P302+P352, P321, P304+P340, P305+P351+P338, P332+P313, P403+P233, P405, P501
2-Mercaptoethanol	GHS05, GHS06, GHS08, GHS09	H301+H331, H310, H315, H318, H361d, H373, H410	P273, P280, P301+P310, P302+P352+P310, P304+P340+P311, P305+P351+P338
Methanol	GHS02, GHS06, GHS08	H225, H301+H311+H331, H370	P210, P280, P301+P310+P330, P302+P352+P312, P304+P340+P311
Natriumdodecylsulfat (SDS)	GHS02, GHS05, GHS07	H228, H302+H332, H315, H318, H335, H412	P210, P261, P264, P280, P305+P351+P338+P310, P370+P378
Phenylmethylsulfonylfluorid (PMSF)	GHS05, GHS06	H301, H314	P260, P280, P301+P310+P330, P303+P361+P353, P305+P351+P338+P310
Phosphorsäure	GHS05, GHS07	H290, H302, H314	P234, P270, P280, P301+P312, P303+P361+P353, P305+P351+P338
Sunitinibmalat	GHS08	H360, H372	P202, P260, P264, P270, P280, P308+P313
N,N,N',N'-Tetramethylethyldiamin (TEMED)	GHS02, GHS05, GHS06	H225, H301+H331, H314	P210, P233, P280, P303+P361+P353, P304+P340+P310, P305+P351+P338

Danksagung

Zuallererst gilt mein größter Dank Herrn Dr. Thomas Lemcke und Frau Prof. Dr. Elke Oetjen für die Überlassung des Promotionsthemas und für die Betreuung dieser Arbeit. Sie haben den Prozess mit allen Höhen und Tiefen begleitet, gaben mir das nötige Vertrauen und ermöglichten die Anfertigung meiner Dissertation. Dr. Thomas Lemcke ermöglichte mir parallel die Weiterbildung zur Fachapothekerin für pharmazeutische Analytik. Die interessanten Diskussionen und Gespräche, nicht nur im fachlichen Bereich, werden mir immer in Erinnerung bleiben. Ihm gilt mein ganz besonderer Dank für die lange, intensive Betreuung und Unterstützung, fachlich wie moralisch, zur Fertigstellung dieser Arbeit.

Herrn Prof. Dr. Andrew Torda danke ich für die Übernahme des Zweitgutachtens.

Desweiteren danke ich Dr. Jorge Enrique Duque Escobar, Sabine Schröder und Christine Bostedt für die Unterstützung, den Austausch und die gemeinsamen Labortage. Danke an Dr. Moritz Münsterberg für den E-Mail-Austausch bezüglich unserer Arbeiten. Mein Dank gilt dem Arbeitskreis von Frau Prof. Dr. Elke Oetjen und Herrn Prof. Dr. Wolfgang Maison für die zahlreichen fachlichen Diskussionen.

Vielen Dank für das Zurverfügungstellen von – mal mehr, mal weniger – ruhigen Plätzen für das Schreiben: Andrea und Uwe Painer, Arnd Ohlenbusch.

Danke an meine Familie und Freunde. Meinen Eltern danke ich für die bedingungslose Unterstützung während meines Pharmaziestudiums, meinen Schwiegereltern für die klaren Worte während meiner Promotionszeit. Ganz besonders möchte ich meiner Schwester Dorena Ceylan und meiner guten Freundin Antje Piegler für die moralische Unterstützung danken.

Zu guter Letzt danke ich dir, Sven, dass du immer an meiner Seite bist und mir den Rücken freigehalten hast. In guten wie in schlechten Zeiten.

Eidesstattliche Versicherung

Hiermit versichere ich an Eides statt, die vorliegende Dissertation selbst verfasst und keine anderen als die angegebenen Hilfsmittel benutzt zu haben. Die eingereichte schriftliche Fassung entspricht der auf dem elektronischen Speichermedium. Ich versichere, dass diese Dissertation nicht in einem früheren Promotionsverfahren eingereicht wurde.

Datum

Unterschrift