

DISSERTATION

zur Erlangung des akademischen Grades eines Doktors der Naturwissenschaften
(Dr. rer. nat.) an der Universität Hamburg im Fachbereich Chemie

Molecular-Modelling-Untersuchungen zu strukturellen Voraussetzungen, Bindungsmodus und Design von Inhibitoren der plasmodialen Glykogen Synthase Kinase-3

vorgelegt von
Sebastian Kruggel
aus Kiel

Hamburg 2011

1. Gutachter: Prof. Dr. Hans-Jürgen Duchstein
2. Gutachter: Prof. Dr. Andrew Torda

Tag der mündlichen Prüfung: 9.9.2011

Jonathan und Carl gewidmet

Die vorliegende Arbeit wurde in der Zeit von Januar 2008 bis Juli 2011 am Institut für Pharmazie im Fachbereich Chemie der Universität Hamburg angefertigt. Sie entstand unter der Leitung von

Herrn Professor Dr. Hans-Jürgen Duchstein,

dem ich an dieser Stelle für die freundliche Aufnahme in seinen Arbeitskreis danken möchte. Mein besonderer Dank gilt außerdem

Herrn Dr. Thomas Lemcke

für die Überlassung des interessanten Themas, für seine engagierte Betreuung und Unterstützung sowie für zahlreiche wertvolle Anregungen während der Durchführung dieser Arbeit.

Danken möchte ich zudem

Herrn Professor Dr. Andrew Torda

für die Übernahme des Koreferats.

Veröffentlichungen

Teile der vorliegenden Arbeit wurden vorab in Form von Tagungsbeiträgen und Artikeln an folgender Stelle veröffentlicht:

Posterbeiträge auf Tagungen

- S. KRUGGEL, S. VOGEL, T. LEMCKE: Generation and evaluation of a homology model of *Pf*GSK-3, 22nd Darmstädter Molecular Modelling Workshop, Erlangen (D), April **2008**
- S. KRUGGEL, T. LEMCKE: Comparative investigation of the ATP-binding site of human and plasmodial glycogen synthase kinase-3, Euro-QSAR 2008, Uppsala (Schweden), September **2008**
- T. LEMCKE, S. KRUGGEL: The search for the right pose: a straightforward workflow for GSK-3 inhibitors, DPhG Jahrestagung 2010, Braunschweig (D), Oktober **2010**
- S. KRUGGEL, T. LEMCKE: Binding mode prediction of *Pf*GSK-3 inhibitors with a thieno[2,3-*b*]pyridine scaffold, DPhG Jahrestagung 2010, Braunschweig (D), Oktober **2010**
- S. KRUGGEL, T. LEMCKE: 3D-QSAR of thieno[2,3-*b*]pyridines as inhibitors of *Pf*GSK-3, Joint Meeting of the Austrian and German pharmaceutical societies, Innsbruck (A), September **2011**

Veröffentlichungen

- S. KRUGGEL, T. LEMCKE: Generation and evaluation of a homology model of *Pf*GSK-3. *Arch. Pharm. Chem. Life Sci.* **2009**, *342*, 327–332
- S. KRUGGEL, T. LEMCKE: Comparative investigation of the ATP-binding site of human and plasmodial glycogen synthase kinase-3. *QSAR Comb. Sci.* **2009**, *28*, 885–890

Inhaltsverzeichnis

Veröffentlichungen	vii
Abkürzungsverzeichnis	xv
(Englische) Fachausdrücke	xix
1. Einleitung und Zielsetzung	1
1.1. Malaria	1
1.1.1. Strategien zur Kontrolle der Malaria	3
1.1.2. Eine Impfung gegen Malaria?	3
1.1.3. Pharmakotherapie mit etablierten Wirkstoffklassen	4
1.1.3.1. Derivate des Artemisinins	4
1.1.3.2. Kombinationspartner in der ACT	5
1.1.3.3. Andere Wirkstoffe	6
1.1.4. Wirkstoffe in der Pipeline	7
1.2. Proteinkinasen	9
1.2.1. Das humane Kinom	10
1.2.2. Die dreidimensionale Gestalt von Proteinkinasen	11
1.2.3. Reaktionsmechanismus der Übertragung des Phosphatrestes	12
1.2.4. Topologie der ATP-Bindetasche	13
1.2.5. Proteinkinaseinhibitoren	15
1.2.6. Die Glykogen Synthase Kinase-3	17
1.2.6.1. Physiologische und pathophysiologische Relevanz	17
1.2.6.2. Regulation und Spezifität der <i>HsGSK-3β</i>	18
1.2.6.3. Inhibitoren der <i>HsGSK-3β</i>	19
1.3. Plasmodiale Kinasen und die <i>PfGSK-3</i>	21
1.3.1. Das plasmodiale Kinom	21
1.3.2. Orphan Kinases: Das Beispiel der <i>PfPK7</i>	22
1.3.3. Plasmodiale CMGC und die <i>PfGSK-3</i>	22
1.4. Zielsetzung	24

2. Homologie-Modelle der <i>PfGSK-3</i> und deren Untersuchung mit MIFs	27
2.1. Homologie-Modelle als Voraussetzung für strukturbasiertes Design	27
2.2. Untersuchungen mit molekularen Interaktionsfeldern	28
2.3. Material und Methoden	29
2.3.1. Homology Modelling	29
2.3.2. Das Verfahren des Comparative Modelling	29
2.3.2.1. Verwendete Software	31
2.3.2.2. Prinzip des <i>RCS</i>	32
2.3.3. Vorgehen bei der Erstellung der Modelle	33
2.3.4. Experimentelle Grundlagen für das Homology Modelling	33
2.3.4.1. Die Sequenz der <i>PfGSK-3</i>	33
2.3.4.2. Die Kristallstrukturen der Vorlagen	34
2.3.5. Berechnung molekularer Interaktionsfelder mit GRID	36
2.3.6. Die Hauptkomponentenanalyse zur Auswertung von MIFs	39
2.3.6.1. Allgemeines zum Verfahren	39
2.3.6.2. Die PCA in Kombination mit MIFs	40
2.3.6.3. PCA <i>vs.</i> CPCA	42
2.3.7. MIFs zum Vergleich von <i>HsGSK-3β</i> und <i>PfGSK-3</i>	44
2.4. Ergebnisse	46
2.4.1. Homologie-Modelle	46
2.4.1.1. Alignment	47
2.4.1.2. Modelle	48
2.4.1.3. Topologie der Bindetasche	53
2.4.1.4. MD-Simulationen	54
2.4.1.5. Fazit Homologie-Modelle	56
2.4.2. Analyse der molekularen Interaktionsfelder	57
2.4.2.1. Visueller Vergleich der Felder	57
2.4.2.2. Gegenüberstellung wichtiger Unterschiede	64
2.4.2.3. Paarweise statistische Analyse mittels PCA	66
2.4.2.4. Statistische Analyse mittels CPCA	71
2.4.2.5. Fazit MIFs	73
2.5. Experimenteller Teil: Computational Details	74
2.5.1. Homology Modelling	74
2.5.2. Berechnung Molekularer Interaktionsfelder	75
3. Der Bindungsmodus von Inhibitoren der <i>PfGSK-3</i>	79
3.1. Grundlagen des Dockings	79
3.1.1. Erkundung möglicher Bindungsmodi: Docking	79

3.1.2.	Bewertung von Dockingergebnissen: Scoring	81
3.1.3.	Der rmsd als Bewertungsparameter	82
3.2.	Material und Methoden: Verbindungen und Software	83
3.2.1.	Inhibitoren der <i>Pf</i> GSK-3	83
3.2.1.1.	Bekannte Struktur-Wirkungsbeziehungen	84
3.2.1.2.	Biologische Daten	85
3.2.2.	Vorgehensweise	86
3.2.2.1.	Orientierendes Docking für Synthesevorschläge	86
3.2.2.2.	Der optimierter Arbeitsablauf DoReCCoM	89
3.2.2.3.	Datensätze für die Evaluierung von DoReCCoM	92
3.2.2.4.	Datensätze für die Anwendung von DoReCCoM	95
3.2.3.	Software	95
3.2.3.1.	FlexX/E	95
3.2.3.2.	AutoDock	97
3.2.3.3.	FRED	98
3.2.3.4.	DrugScore(X)	99
3.2.3.5.	SZYBKI	99
3.3.	Ergebnisse 1: Orientierendes Docking für Synthesevorschläge	100
3.3.1.	Mögliche Bindungsmodi	100
3.3.2.	Bewertung der Hypothesen	107
3.3.2.1.	Bindungsmodi der Klasse 1	107
3.3.2.2.	Bindungsmodi der Klasse 2	110
3.3.2.3.	Zwischenfazit	112
3.3.3.	Resultierende Synthesevorschläge	113
3.4.	Präparative Umsetzung der Synthesevorschläge	117
3.4.1.	2-(Thiophen-2-ylcarbonyl)thieno[2,3- <i>b</i>]pyridin-5-carbonitrile	117
3.4.2.	Biologische Daten	118
3.4.3.	Ergebnis der Synthesen	119
3.4.4.	Experimenteller Teil	119
3.4.4.1.	Allgemeines	119
3.4.4.2.	Synthesen und analytische Daten	120
3.5.	Ergebnisse 2: Entwicklung eines optimierten Workflows	125
3.5.1.	Evaluierung des Verfahrens DoReCCoM	125
3.5.1.1.	Ergebnisse der einzelnen Dockingprogramme	125
3.5.1.2.	Clustering	128
3.5.1.3.	Komplexminimierung	129
3.5.1.4.	Bewertung von DoReCCoM	130

3.5.2.	Anwendung von DoReCCoM auf Thienopyridine in der <i>Pf</i> GSK-3	131
3.5.2.1.	Bindungsmodi der Klasse 1a	131
3.5.2.2.	Bindungsmodi der Klasse 1b	133
3.5.2.3.	Bindungsmodi der Klasse 2	133
3.6.	Fazit DoReCCoM	134
3.7.	Experimenteller Teil: Computational Details	136
3.7.1.	Erstes Docking mit FlexE und AutoDock	136
3.7.2.	Rescoring mit DrugScore(X)	138
3.7.3.	Entwicklung eines optimierten Workflows	139
4.	3D-QSAR-Untersuchungen möglicher Bindungsmodi	141
4.1.	Quantitative Struktur-Wirkungsbeziehungen	141
4.1.1.	Grundlagen	142
4.1.2.	3D-QSAR: CoMFA und CoMSIA	144
4.1.2.1.	Prinzip vergleichender Feldanalysen	144
4.1.2.2.	Voraussetzungen für 3D-QSAR-Untersuchungen	145
4.1.2.3.	Überlagerung	146
4.1.3.	Multivariate Analyse mittels PLS	147
4.1.3.1.	Partial Least Square Regression	147
4.1.3.2.	Validierung von PLS-Modellen	149
4.2.	Material und Methoden	151
4.2.1.	Rationale	151
4.2.2.	Der Datensatz	151
4.2.2.1.	Aufteilung in Trainings- und Testsatz	152
4.2.2.2.	Eigenschaften des Datensatzes	153
4.2.3.	Methodik: Überlagerungen und 3D-CoMSIA	153
4.3.	Ergebnisse	154
4.3.1.	Überlagerungen und deren Analyse mittels SAMPLS	154
4.3.1.1.	Überlagerung nach Modus A und Modus H	154
4.3.1.2.	Überlagerung basierend auf Modus A	155
4.3.1.3.	Überlagerung basierend auf Modus B	156
4.3.2.	Analyse resultierender 3D-QSAR-Modelle	157
4.3.2.1.	Modelle basierend auf Modus A	157
4.3.2.2.	Das Modell AcDoSt	159
4.3.2.3.	Das Modell AcEl	165
4.3.2.4.	Modelle basierend auf Modus B	168
4.3.2.5.	Das Modell AcDoSt	170
4.3.2.6.	Das Modell AcEl	173

4.4. Fazit 3D-QSAR	177
4.5. Experimenteller Teil: Computational Details	178
5. Diskussion und Ausblick	181
5.1. Homologie-Modelle und MIFs	181
5.1.1. Die Sequenzen als Ansatzpunkt für Selektivität	181
5.1.2. Homologie-Modelle der <i>Pf</i> GSK-3	182
5.1.3. Möglichkeiten und Grenzen der Analyse mit MIFs	184
5.1.3.1. Visuelle Analyse molekularer Interaktionsfelder	184
5.1.3.2. Analyse mittels (C)PCA	186
5.2. Der Bindungsmodus von Thieno[2,3- <i>b</i>]pyridinen	187
5.2.1. Docking mit DoReCCoM	187
5.2.2. Argumente für mögliche Bindungsmodi	191
5.2.3. 3D-QSAR-Untersuchungen	196
5.2.4. Die Quartärstruktur als Erklärungsansatz für Selektivität	198
5.3. Ausblick	200
5.3.1. Optimierung eingesetzter Verfahren	200
5.3.2. Weiterführende Untersuchungen	200
6. Zusammenfassung	203
7. Summary	205
Literatur	207
Anhang	227
A. Ergänzende Daten	227
A.1. Alignments	227
A.2. Abbildungen	230
B. Tabellen	231
B.1. Validierung der Vorlagen und Homologie-Modelle	231
B.2. Übersicht der Strukturen und biologischen Daten	233
B.3. Docking	238
B.4. QSAR	243
B.4.1. Ergebnisse der SAMPLS-Berechnungen	243
B.4.2. Vorhersagen näher untersuchter Modelle	245
C. Verwendete Gefahrstoffe und KMR-Substanzen	247

D. Danksagung	249
E. Lebenslauf	251
F. Eidesstattliche Erklärung	253

Abkürzungsverzeichnis

AAV	Allgemeine Arbeitsvorschrift
Abb.	Abbildung
AcDoSt	Kombination aus sterischem, Wasserstoffbrückenakzeptor- und -donorfeld in CoMSIA-Berechnungen
AcDoEI	Kombination aus elektrostatischem, Wasserstoffbrückenakzeptor- und -donorfeld in CoMSIA-Berechnungen
AcEI	Kombination von Wasserstoffbrückenakzeptor- und elektrostatischem Feld in CoMSIA-Berechnungen
ACN	Acetonitril
ACT	Artemisinin-basierte Kombinationstherapie (engl.: Artemisinin-based Combination Therapy)
AD	AutoDock
ADME	Vorgänge im Rahmen der Pharmakokinetik eines Arzneistoffes: Absorption, Distribution, Metabolisierung, Exkretion
ADT	AutoDockTools
AGC	Gruppe von Proteinkinasen, die PKA, PKG und PKC umfasst
Ala	Alanin
anel.	anelliert
aPK	atypische Proteinkinase
Arg	Arginin
Asn	Asparagin
Asp	Aspartat
ATP	Adenosin Triphosphat
BDNF	Brain-Derived Neurotrophic Factor
BLAST	Basic Local Alignment Search Tool
BUW	Block Unscaled Weights
bzw.	beziehungsweise
C_α	Alpha-Kohlenstoffatom
CAMK	Calcium/Calmodulin-abhängige Kinasen
CDK	Cyclin-abhängige Kinase
CK1	Casein Kinase 1

CSD	Cambridge Structural Database
CLK	CDK-ähnliche Kinase
CMGC	Gruppe von Proteinkinasen, die CDK, MAPK, GSK und CLK umfasst
CoMFA	Comparative Molecular Field Analysis
CoMSIA	Comparative Molecular Similarity Indices Analysis
Cys	Cystein
d	als Abkürzung für Signalmultiplizitäten: Dublett
DHA	Dihydroartemisinin
DHFR	Dihydrofolatreduktase
DHPS	Dihydropteroatsynthetase
DMF	Dimethylformamid
DMSO	Dimethylsulfoxid
DTG	Deutschen Gesellschaft für Tropenmedizin und Internationale Gesundheit
DXR	1-Deoxy-D-Xylulose-5-phosphat-Reduktoisomerase
(e)br	als Abkürzung für Signalmultiplizitäten: (extrem) breites Signal
EGFR	Epidermal growth factor receptor
EMA	Europäische Arzneimittelagentur (engl.: European Medicines Agency)
ERK	Extracellular signal-Regulated Kinase
engl.	Übersetzung eines deutschen Begriffes ins Englische oder <i>vice versa</i>
ePK	eukaryotische Proteinkinase
ff.	und folgende
FRED	Fast Rigid Exhaustive Docking
Gln	Glutamin
Glu	Glutamat
Gly	Glycin
ggf.	gegebenenfalls
GOLPE	Generating Optimal Linear PLS Estimations
GSK	Glykogen Synthase Kinase
GUI	Graphical User Interface
HsGSK-3β	Glykogen Synthase Kinase-3 β von <i>Homo sapiens</i>
HIV	Humanes Immundefizienz-Virus
His	Histidin
HTS	High Throughput Screening
IC₅₀	mittlere inhibitorische Konzentration
Ile	Isoleucin
IPN	Chemoprävention der Malaria für Hochrisikopatienten (engl.: Intermittent Preventive Treatment)
IRS	Behandlung von Hausinnenwänden zur Bekämpfung der <i>Anopheles</i> -Mücken

	(engl.: Indoor Residual Spraying)
ITN	insektizidbehandelten Moskitonetzen (engl.: Insecticide-Treated mosquito Nets)
IUPAC	International Union of Pure and Applied Chemistry
Leu	Leucin
LOO	Leave-one-out-Kreuzvalidierung
LSO	Leave-some-out-Kreuzvalidierung
Lys	Lysin
m	als Abkürzung für Signalmultiplizitäten: Multiplett
MAPK	Mitogen-aktiviertes Protein (engl.: Mitogen-Activated Protein) Kinase
MD	Moleküldynamik
MEK1/2	MAP/ERK Kinase 1/2; auch: MAP2K1/2 (MAP Kinase Kinase 1/2)
Met	Methionin
MLR	multiple lineare Regression
MM	Molekülmechanik
MMFF	Merck Molecular Force Field
μM	mikromolar, gängige Einheit für IC_{50} -Werte (SI: 10^{-6} mol/L)
Met	Methionin
MIFs	Molekulare Interaktionsfelder (engl.: Molecular Interaction Fields)
(C)PCA	(Consensus) Principal Component Analysis (engl.: (hierarchische) Hauptkomponentenanalyse)
PC_x	<i>x</i> te Hauptkomponente einer PCA
PDB	Protein Data Base
pdf	probability density function
PfAPI	Inzidenz der durch <i>Plasmodium falciparum</i> verursachten Malaria (engl.: <i>Plasmodium falciparum</i> Annual Parasite Rate)
PfCK1	Casein Kinase 1 von <i>Plasmodium falciparum</i>
PfGSK-3	Glykogen Synthase Kinase-3 von <i>Plasmodium falciparum</i>
PfPK5	Protein Kinase 5 von <i>Plasmodium falciparum</i> , homolog zu humanen CDK
PfPK7	Protein Kinase 7 von <i>Plasmodium falciparum</i> , orphan kinase
PfPR	Risiko einer durch <i>Plasmodium falciparum</i> verursachten Malaria (engl.: <i>Plasmodium falciparum</i> Parasite Rate)
Phe	Phenylalanin
pIC₅₀	negativer dekadischer Logarithmus des IC_{50}
PKA	Proteinkinase A
PKC	Proteinkinase C
PKG	Proteinkinase G
PLS	Partial Least Square (regression); auch: Projection to Latent Structures

PRESS	Predictive Sum of Squares
Pro	Prolin
PSP	Progressiver Supranukleäre Paralyse
q	als Abkürzung für Signalmultiplizitäten: Quartett
QSAR	Quantitative Structure-Activity Relationship
QSPR	Quantitative Structure-Property Relationship
RCS	Rank-by-rank based Consensus Score
RDT	diagnostischer Schnelltest (engl.: Rapid Diagnostic Test)
rmsd	mittlere quadratische Abweichung (engl.: root mean square deviation)
RGC	Rezeptor-Guanylat-Cyclasen
RMS	Standardabweichung, auch <i>s</i> (engl.: Root Mean Square error)
ROCS	Rapid Overlay of Chemical Structures
RSS	Summe der Abweichungsquadrate (engl.: Residual Sum of Squares)
s	als Abkürzung für Signalmultiplizitäten: Singulett
S.	Seite
SAMPLS	SAMple-distance Partial Least Squares
SAS	dem Lösungsmittel zugängliche Oberfläche (engl.: Solvent Accesible Surface)
Ser	Serin
SS	Summe der quadratischen Abweichungen vom Mittelwert (engl.: Sum of Squares)
SsGSK-3	Glykogen Synthase Kinase von <i>Sus scrofa</i>
STE	Homologe der Sterile 7-, Sterile 11-, Sterile 20-Kinasen
t	als Abkürzung für Signalmultiplizitäten: Triplet
Tab.	Tabelle
TCAMS	Tres Cantos Antimalarial Compound Set
Thr	Threonin
TK	Tyrosin-Kinasen
TKL	TK-ähnliche Kinasen
TZDZ	Thiadiazolidinone
Tyr	Tyrosin
Val	Valin
vgl.	vergleiche
WHO	World Health Organization
XAccum	Aufsummierter Prozentualer Anteil der Varianz einer <i>X</i> -Matrix, die durch die latenten Variablen (Hauptkomponenten) einer PCA erklärt wird
XVarExp	Prozentualer Anteil der Varianz, die durch eine einzelne Komponente einer PCA erklärt wird

(Englische) Fachausdrücke

Da der überwiegende Anteil der Fachliteratur in englischer Sprache verfasst ist, haben sich zahlreiche feststehende englische Fachausdrücke etabliert, die in der Literatur so verbreitet sind, dass eine Übersetzung unangemessen oder weniger zutreffend erschien. Sie werden deshalb an dieser Stelle übersetzt, wenn notwendig erläutert und dann im laufenden Text ohne weitere Erklärung verwendet. Zudem werden in der Art eines Glossars kurz besonders wichtige, häufig verwendete Fachausdrücke erläutert.

Da außerdem die Programme in der Regel in englischer Sprache dokumentiert sind, werden die entsprechenden Einstellungen und Befehle aus der Dokumentation nicht übersetzt, um Missverständnissen vorzubeugen. Solche Begriffe sind dann der Übersichtlichkeit halber in der abweichenden Schrift <term> wiedergegeben.

Alignment	räumliche Überlagerung von Molekülen, im Falle von Proteinen häufig bezogen auf ihre Aminosäuresequenz (vgl. auch Sequenz-Alignment)
Apoenzym	Enzym ohne Cofaktor (bzw. Ligand)
Backbone	als Rückgrat eines Makromoleküls bezeichnete Kette von sich wiederholenden Einzelbausteinen; im engeren Sinn: bei Proteinen die Folge von Peptidbindungen
Clustering	Gruppieren einzelner Elemente eine Menge nach ähnlichen Eigenschaften
Consensus Scoring	Zusammenfassung mehrerer Bewertungssysteme zu einer übergreifenden Einschätzung
Crossdocking	Docking eines Liganden in einen nicht kokristallisierten Rezeptor
Docking	Einpassen eines Liganden in eine Bindetasche (vgl. IUPAC-Definition S. 79)
Deskriptor	Eigenschaft einer Struktur, die (z.B. für QSAR-Berechnungen) in eine mathematisch fassbare Größe überführt werden kann

Drug Design	Arzneistoffentwicklung; weit gefasster Begriff, der nach IUPAC neben der Leitstruktursuche und -optimierung auch Pharmakokinetik und Toxikologie umfasst
Enrichment	Anreicherung; im engeren Sinn: die Anreicherung aktiver Verbindungen im Rahmen eines virtuellen Screenings im Vergleich zur Wahrscheinlichkeit eines zufälligen Auffindens der Substanz
Gatekeeper	Aminosäure, die in der ATP-Bindetasche von Proteinkinase für den Zugang zur hydrophoben Untertasche (der Back Pocket) eine entscheidende Rolle spielt (vgl. S. 13)
Glycine-rich Loop	Bereich, der die ATP-Bindetasche von Proteinkinase begrenzt und der besonders häufig die Aminosäure Glycin enthält
Hinge region	engl.: Scharnier; Verbindung von C- und N-terminaler Domäne in Proteinkinase
Homology Modelling	Berechnung von Homologie-Modellen (auch Comparative Modelling, vgl. S. 29 ff.)
Holoenzym	Enzym mit Cofaktor (bzw. Ligand)
Induced Fit	Anpassen eines Rezeptors an einen Liganden durch dessen Bindung
Loading Plot	Wichtige grafische Darstellung der Ergebnisse einer Hauptkomponentenanalyse (PCA), vgl. S. 39
Overfitting	Zu genaue Anpassung eines Modelles auf einen speziellen Datensatz, die dazu führt, dass die allgemeine Anwendbarkeit des Modelles verloren geht
Pose	engl.: Pose; im Zusammenhang mit Docking: eine durch das eingesetzte Dockingprogramm berechnete mögliche Position des Liganden im Rezeptor
rmsd	vgl. Abkürzungen; im engeren Sinn: Wert [\AA] für die mittlere quadratische Abweichung der Distanz der (schweren) Atome zweier überlagerter Moleküle zueinander
Redocking	Docking eines Liganden in den kokristallisierten Rezeptor
Score Plot	vgl. Loading Plot bzw. S. 39
Scoring	Bewertung; im engeren Sinn eine Funktion zur vergleichenden Abschätzung der Stärke einer Interaktion zwischen Ligand und Protein
Scrambling	Zufälliges Vermischen von Daten (Randomisierung)

(Virtual / High Throughput) Screening	Auswahlverfahren; im engeren Sinn: das Testen einer großen Anzahl von Molekülen auf ihre Aktivität an einer Zielstruktur hin. Geschieht dies am Rechner (<i>in silico</i>), wird das Verfahren als Virtual Screening bezeichnet. Werden große Substanzbibliotheken real mit biochemischen Assays getestet, spricht man von High Throughput Screening (HTS).
Target	Im Allgemeinen: Zielstruktur eines Wirkstoffes. Im Zusammenhang mit MIFs: Struktur, für die Interaktionsenergien berechnet werden sollen.
United Atom Model	In der Molekülmechanik häufig angewandtes Modell, bei dem zur Vereinfachung und damit zur Beschleunigung der Berechnung (unpolare) Wasserstoffe und schwere Atome virtuell zu einem Atom zusammengefasst werden
United Protein Structure	von FlexE konstruierte, virtuelle Proteinstruktur

1. Einleitung und Zielsetzung

1.1. Malaria

Malaria stellt neben Tuberkulose und HIV eine der drei weltweit verbreitetsten und folgenschwersten Infektionskrankheiten dar: 2009 infizierten sich schätzungsweise 225 Millionen Menschen, davon kamen 781 000 um [1]. Besonders betroffen sind Entwicklungsländer, auf 30 afrikanische und fünf asiatische Länder entfallen zusammengenommen 98 % aller Todesfälle (vgl. Abb. 1.1). Außerdem leiden insbesondere Kinder unter der Verbreitung der Krankheit: Etwa 85 % der Todesopfer sind Kinder unter fünf Jahren [2]. Die Infektion erfolgt durch die Protozoen *Plasmodium vivax*, *Plasmodium ovale*, *Plasmodium malariae* und *Plasmodium falciparum*, die ihrerseits durch den Stich der weiblichen *Anopheles*-Mücke auf den Menschen übertragen

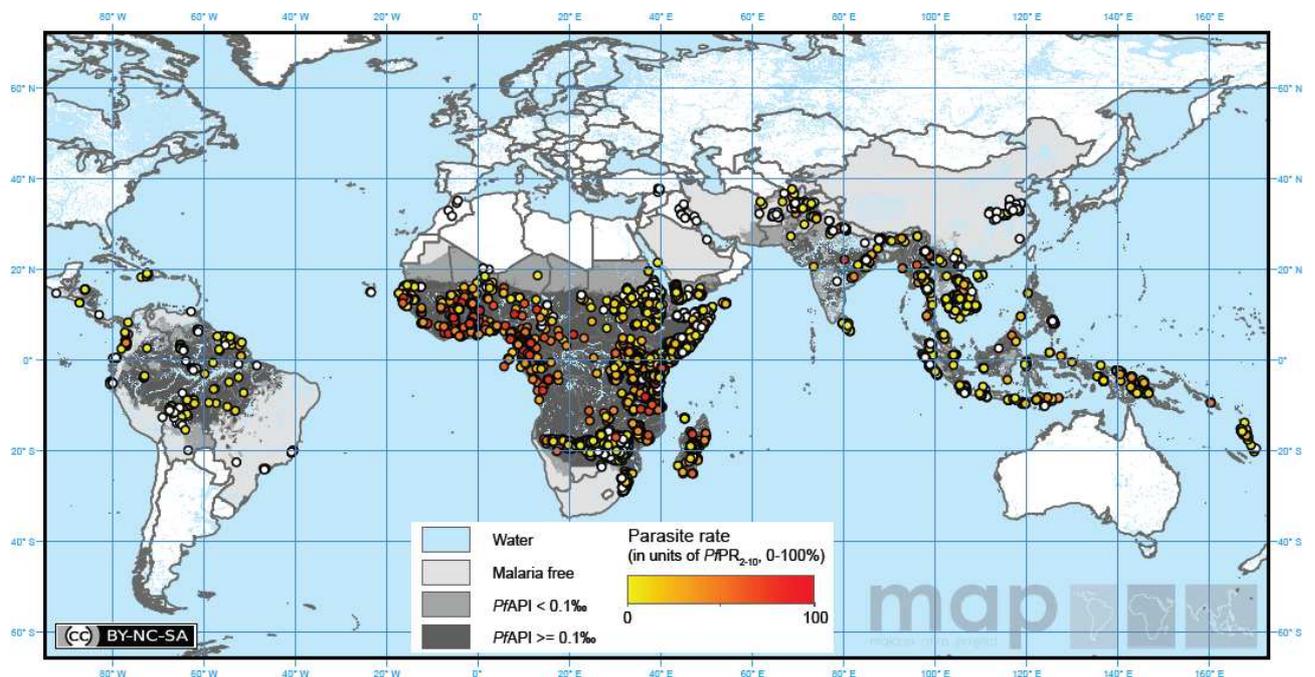


Abbildung 1.1. | Weltweite Verteilung der Inzidenz und des Risikos einer durch *Plasmodium falciparum* verursachten Malaria nach HAY *et al.* [3] und GUERRA *et al.* [4]. Die unterschiedlichen Grautöne zeigen in der Legende bezeichnete Bereiche der jährlichen Inzidenz an (PfAPI: *P. falciparum* annual parasite incidence). Die zitierten Veröffentlichungen beziehen 7953 Studien aus den Jahren 1985 bis 2008 ein, aus denen altersstandardisiert auf 2 bis 10 Jahre das Risiko einer Erkrankung berechnet wurde. Alle Studien sind als Punkte farbkodiert nach Höhe dieses Risikos in der Karte eingetragen (PfPR: *P. falciparum* parasite rate).

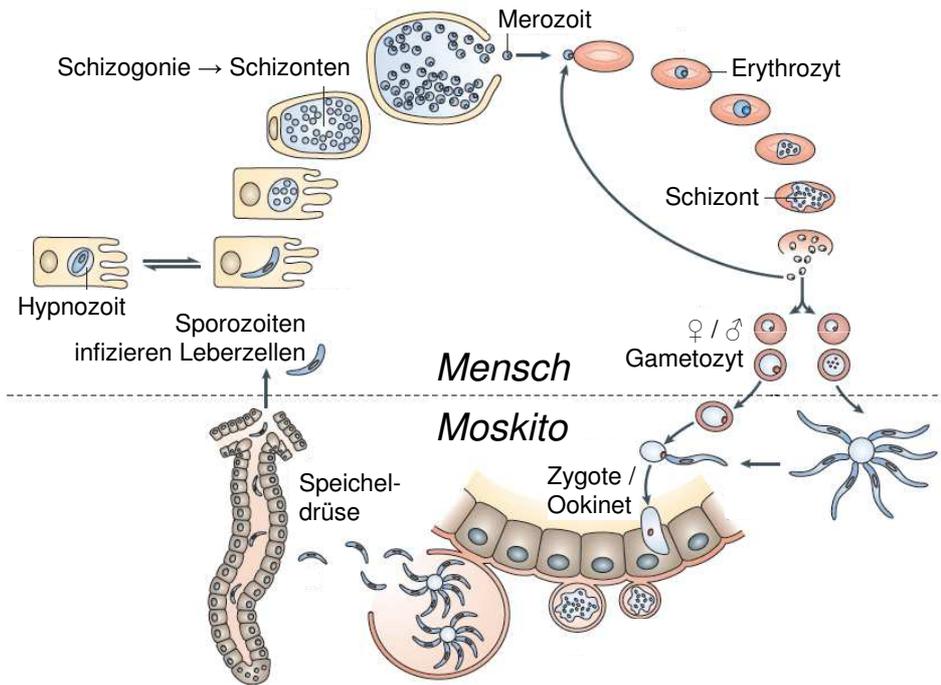


Abbildung 1.2. | Lebenszyklus der Plasmodien (modifiziert nach [6]). Nach der Übertragung der Parasiten durch den Stich des Moskitos findet im menschlichen Organismus die im Text beschriebene ungeschlechtliche Vermehrung statt. Nimmt die *Anopheles*-Mücke bei einer Blutmahlzeit Erreger im Stadium der Gametozyten auf, kann nach dem Wirtswechsel im Mosquito die geschlechtliche Vermehrung über die Stadien der Zygote, Ookinete und Oocyste zu Sporozoiten erfolgen. Diese können dann erneut auf den Menschen übertragen werden.

werden. *P. falciparum* und *P. ovale* sind gemeinsam für 95 % aller Infektionen verantwortlich, wobei nahezu alle schweren Verläufe durch *P. falciparum* verursacht werden. Die durch *P. falciparum* verursachte Malaria wird als Malaria tropica bezeichnet. *P. ovale* und *P. vivax* sind Erreger der Malaria tertiana, die ihren Namen wegen der klinisch kennzeichnenden Fieberanfälle an jedem dritten Tag trägt. Eine Infektion durch *P. malariae* wird als Malaria quartana bezeichnet, die Fieberanfälle kehren hier in der Regel am vierten Tag wieder [5].

Der Lebenszyklus der Parasiten ist in Abbildung 1.2 dargestellt. Er läuft über unterschiedliche Entwicklungsstufen teils in der *Anopheles*-Mücke und teils im befallenen Organismus ab, dort wiederum in Gewebe und Blut (bei *P. vivax* und *P. ovale* auch in der Leber). Nach Übertragung der Erreger als Sporozoiten dringen diese in Leberparenchymzellen ein, wo sie sich durch ungeschlechtliche Vielteilung (Schizogonie) vermehren. Die dabei entstehenden zahlreichen Merozoiten werden in die Blutbahn freigesetzt, wo sie in Erythrozyten eindringen können. In den Erythrozyten besteht wiederum die Möglichkeit der Schizogonie. Die Entwicklung der Plasmodien verläuft hier synchron, so dass es gleichzeitig zur Lyse vieler Erythrozyten kommt, die zur Freisetzung zahlreicher Merozoiten und folglich zum klinischen Erscheinungsbild der Fieberschübe führt. Von den freigesetzten Merozoiten können sich einige zu weiblichen und männlichen Gametozyten weiterentwickeln, die von der Mücke bei einer Blutmahlzeit aufge-

nommen werden können. In der Mücke erfolgt die geschlechtliche Vermehrung, an deren Ende wiederum die Sporozoiten stehen, die beim nächsten Stich auf den Menschen übertragen werden können [7].

P. vivax und *P. ovale* können in den Leberzellen als Hypnozoiten persistieren. Werden diese Stadien nach Wochen bis Jahren wieder aktiviert, kommt es zu den für diese Erreger charakteristischen Rezidiven [8]. Bei diesen Erregern müssen in der Therapie demnach Arzneistoffe eingesetzt werden, die auch in den genannten Entwicklungsstadien wirksam sind.

1.1.1. Strategien zur Kontrolle der Malaria

Die eingangs zitierten Fall- und Todeszahlen belegen eindrucksvoll, dass Malaria eine weltweite Bedrohung der öffentlichen Gesundheit darstellt. Daraus ergibt sich die dringende Notwendigkeit der Bekämpfung dieser Erkrankung, auch wenn die World Health Organization (WHO) in ihrem aktuellen Jahresbericht einen leichten Rückgang der Fall- und Todeszahlen im Vergleich zum Vorjahr verzeichnet. Diese Tendenz wird in erster Linie auf Maßnahmen der Prävention zurückgeführt. Als wichtigste Strategien sieht die WHO die folgenden drei Maßnahmen an [1]:

- Einsatz von insektizidbehandelten Moskitonetzen, insbesondere als Bettnetze (engl.: insecticide-treated mosquito nets, ITNs).
- Behandlung von Hausinnenwänden zur Bekämpfung der Mücken (engl.: indoor residual spraying, IRS).
- Chemoprävention für Hochrisikopatienten, insbesondere für Schwangere und Kleinkinder (engl.: intermittent preventive treatment, IPT).

Zudem wird die Wichtigkeit einer schnellen und eindeutigen Diagnose mit geeigneten Testsystemen (engl.: rapid diagnostic test, RDT) hervorgehoben. Diese erlaubt den gezielten Einsatz der Pharmakotherapie und vermeidet unnötige, unter Umständen Resistenzen fördernde Anwendungen [1]. Bei nachgewiesener Infektion kann die von Erreger und Resistenzlage abhängige Pharmakotherapie gewählt werden, deren Bestandteile in Abschnitt 1.1.3 vorgestellt werden.

1.1.2. Eine Impfung gegen Malaria?

Im Mai 2009 ging mit RTS,S der erste Impfstoff gegen Malaria in eine klinische Prüfung der Phase III [9]. Bei RTS,S handelt es sich um ein Fusionsprotein aus dem Oberflächenantigen des Hepatitis B Virus in Kombination mit dem von *P. falciparum* im Stadium der Sporozoiten und Leberschizonten exprimierten Oberflächenantigen Circumsporozoit. Der Impfstoff wird mit einem proprietären Adjuvans kombiniert,¹ das für die Wirkung des Impfstoffes entscheidend ist.

¹Die Entwicklung wird von GlaxoSmithKline mit Unterstützung der PATH Malaria Vaccine Initiative (MVI) durchgeführt. Diese zählt wie auch Medicines for Malaria Venture (MMV) zu den sogenannten public-private

Durch die Verabreichung wird eine Immunantwort in Form von Antikörperbildung und T-Zell-Aktivierung erreicht. In Studien der Phase II konnten beachtliche Erfolge erzielt werden [10, 11], so dass die Hoffnung berechtigt scheint, dass 2012 ein Impfstoff zugelassen werden wird, der zu einem Rückgang der klinischen Malaria-Anfälle führen kann. Die Effektivität der Impfung liegt dabei in den genannten Studien bei 49 bzw. 65.2 % – diese Zahlen machen deutlich, dass die Impfung einen großen Fortschritt darstellen würde, eine Behandlung der Erkrankung jedoch unabhängig davon unerlässlich bleiben wird.

1.1.3. Pharmakotherapie mit etablierten Wirkstoffklassen

Die gängigen Arzneistoffe lassen sich auf unterschiedliche Weise nach Struktur bzw. Wirkungsmechanismus einteilen, in der folgenden kurzen Darstellung wichtiger Gruppen wird dabei der Einteilung nach SCHLITZER *et al.* gefolgt [12]. Gegen alle in der Therapie eingesetzten Arzneistoffklassen wurden mittlerweile Resistenzen dokumentiert. Wegen dieser dramatischen Entwicklung konstatiert die WHO aktuell, dass die Artemisinin-basierte Kombinationstherapie (ACT) die einzige effektive Behandlung einer unkomplizierten Malariainfektion darstellt [13]. Es werden im Folgenden deshalb neben Artemisinin diejenigen Wirkstoffklassen vorgestellt, die im Rahmen der ACT eine Bedeutung haben. Die in der WHO-Empfehlung genannten Wirkstoffe sind als Vertreter ihrer Gruppe in Abbildung 1.3 dargestellt.

1.1.3.1. Derivate des Artemisinins

Aus dem einjährigen Beifuß *Artemisia annua* L. konnte 1972 das Sesquiterpenlacton Artemisinin (**1**) extrahiert werden [14]. Es zeichnet sich durch seine Endoperoxidstruktur aus, in der Folge entwickelte Derivate sind Dihydroartemisinin (DHA, aktiver Metabolit, **2**), Artemether (**3**), Artemotil (**4**) und Artesunat (**5**). Artemisinin-Derivate interagieren in der Nahrungsvakuole des Parasiten über freie Radikale (Strukturmerkmal der Substanzen: Endoperoxid) mit Häm und hemmen zudem die Nukleinsäure- und Proteinbiosynthese der Plasmodien [15]. Artemisinin-derivate wirken auf alle Entwicklungsformen der Plasmodien im Blut, wodurch eine rasche Linderung der Fieberanfälle möglich wird. Zusätzlich werden Gametozyten angegriffen, wodurch die Transmission erschwert wird.

JAMBOU *et al.* beschrieben schon 2005 *in vitro* eine Resistenzentwicklung von *P. falciparum*-Stämmen gegen Artemether und assoziierten diese mit der Punktmutation S769N der PfATPase6 als mögliches Target der Artemisinine [16]. Weitere Berichte von Resistenzentwicklungen gegen diese wichtige Arzneistoffklasse [17, 18] geben Anlass zur Therapieempfehlung der WHO, die eine Monotherapie zur Vermeidung weiterer Resistenzen strikt ablehnt.

partnerships (PPP), die in den späten 1990er Jahren gegründet wurden. Diese Organisationen betreiben mithilfe der Unterstützung von Stiftungen Forschung an Projekten, die kostenintensiv sind ohne dabei entsprechende ökonomische Perspektiven für die forschende Arzneimittellindustrie zu bieten.

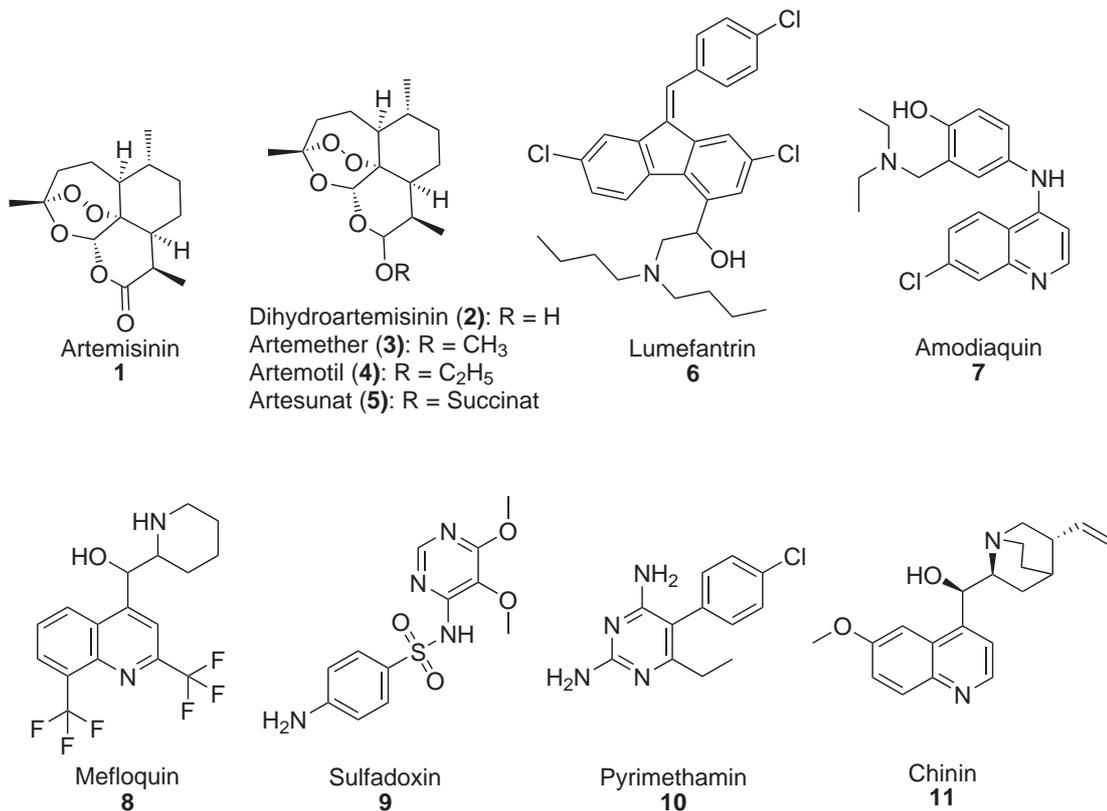


Abbildung 1.3. | Artemisinin-Derivate und ausgewählte Kombinationspartner in der ACT.

1.1.3.2. Kombinationspartner in der ACT

Die von der WHO empfohlenen Mittel der ersten Wahl in der Therapie einer durch *P. falciparum* verursachten unkomplizierten Malaria sind die folgenden Kombinationen:

1. Artemether (**3**) + Lumefantrin (**6**)
2. Artesunat (**5**) + Amodiaquin (**7**)
3. Artesunat (**5**) + Mefloquin (**8**)
4. Artesunat (**5**) + Sulfadoxin/Pyrimethamin (**9/10**)

Bei Therapieversagen der genannten Kombinationen können laut WHO als Therapie der zweiten Wahl Artesunat oder Chinin (**11**) in Kombination mit den Antibiotika Tetracyclin, Doxycyclin oder Clindamycin eingesetzt werden [13].

1.1.3.2.1. Lumefantrin und Mefloquin: Arylamino-Alkohole Der Wirkungsmechanismus der Arylamino-Alkohole ist nicht vollständig geklärt. Als entscheidend wird die Inhibition der Endozytose im Zusammenhang mit einer Beeinflussung des Häm-Stoffwechsels angenommen [19]. Zu der Substanzklasse zählt neben Lumefantrin (**6**) und Mefloquin (**8**) der Prototyp der Arylamino-Alkohole und älteste Antimalaria-Wirkstoff Chinin (**11**). Reines Chinin wird bereits seit 1820 eingesetzt [20], für die Anwendung von Zubereitungen aus Chinarinde gibt es Berichte aus dem frühen 17. Jahrhundert [21]. Chinin wird bei Infektionen mit sensiblen Erregern im-

mer noch erfolgreich verwendet und war Ausgangspunkt der Entwicklung von Lumefantrin (**6**), Mefloquin (**8**) und weiteren Vertretern der Substanzklasse wie z. B. Halofantrin. Als Monotherapie sollten die Substanzen aufgrund der Resistenzentwicklung jedoch nicht mehr angewendet werden.

1.1.3.2.2. Amodiaquin: 4-Aminochinoline Zu diesen Substanzen zählt neben Amodiaquin z. B. Chloroquin (**12**, vgl. Abb. 1.4), das zur Bekämpfung der durch *P. vivax* verursachten Malaria tertiana nach wie vor Mittel der ersten Wahl ist. Die Wirkung beruht auf der Bildung eines Komplexes mit Ferriprotoporphyrin IX, einem für den Parasiten toxischen Abbauprodukt des Hämoglobins. Durch die Komplexbildung kann die gewöhnlich zur Entgiftung ablaufende Polymerisation zum nicht mehr toxischen Hemazoin nicht mehr stattfinden [22]. Resistenzen gegen diese Stoffklasse beruhen vor allem auf einer Mutation des Chloroquine Resistance Transporters (CRT), die zu vermehrter Ausfuhr der Wirkstoffe aus der Verdauungsvakuole des Parasiten führt [23, 24].

1.1.3.2.3. Sulfadoxin/Pyrimethamin: Inhibitoren der DHFR und der DHPS

Die Biosynthese des Tetrahydrofolats wird durch die Inhibition der beiden Enzyme Dihydropteroatsynthetase (DHPS) und Dihydrofolatreduktase (DHFR) unterbunden. Sulfonamide wie Sulfadoxin können als kompetitive Inhibitoren der 4-Aminobenzoessäure an der DHPS eingesetzt werden – ein Prinzip, das auf die frühe Entwicklung antibakterieller Wirkstoffe zurückgeht, auf die Entdeckung der antibakteriellen Aktivität des Sulfachrysoidins 1932 durch DOMAGK [25]. In Kombination mit einem Inhibitor der DHFR (neben Pyrimethamin z. B. Proguanil (**13**), üblicherweise in Kombination mit Dapson (**14**), vgl. Abb. 1.4) wird ein synergistischer Effekt erreicht, der die Kombination lange zu einer wichtigen Option in der Pharmakotherapie der Malaria machte. Mutationen der Zielenzyme [26, 27] führten dazu, dass die Kombination heute nur noch als Bestandteil der ACT eingesetzt werden sollte.

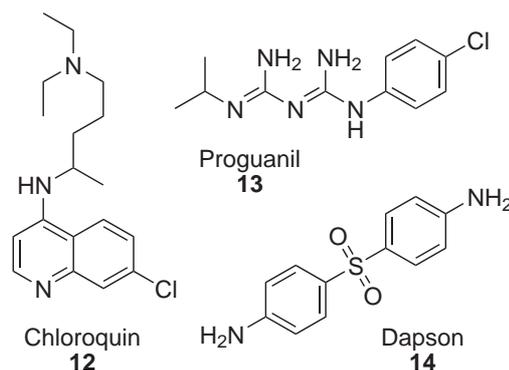


Abbildung 1.4. | Weitere Beispiele für 4-Aminochinoline bzw. DHFR- und DHPS-Inhibitoren.

1.1.3.3. Andere Wirkstoffe

Die im Folgenden kurz vorgestellten Wirkstoffgruppen sind zwar nicht Bestandteil der ACT, haben jedoch durch ihr spezielles Einsatzgebiet einen so hohen Stellenwert, dass sie hier nicht unerwähnt bleiben sollen.

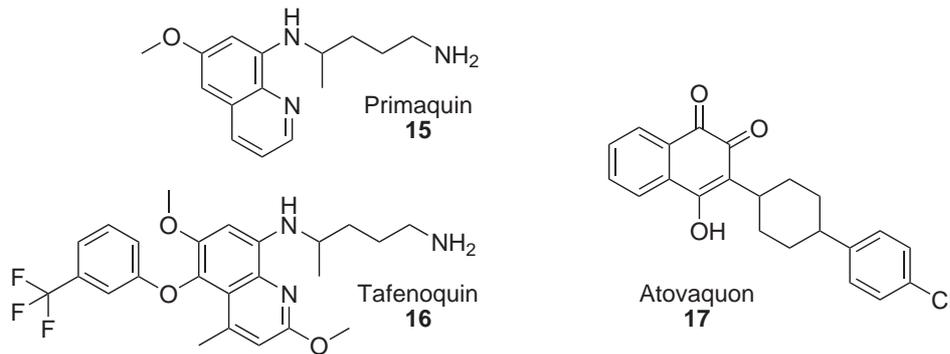


Abbildung 1.5. | Beispielsubstanzen aus der Gruppe der 8-Aminoquinoline und Atovaquon als Wirkstoff, der die Atemkette der Erreger als Angriffspunkt hat.

1.1.3.3.1. 8-Aminoquinoline Diese Gruppe (z. B. Primaquin (15), aktuelle Weiterentwicklung: Tafenoquin (16), vgl. Abb. 1.5) wirkt auch auf Entwicklungsstadien der Erreger, die in der Leber durchlaufen werden. Besonderheit ist daher die Anwendung zur Eradikationstherapie bei schweren Infektionen mit *P. vivax*, derzeit von der WHO in Kombination mit Chloroquin (12) bei sensitiven Stämmen als erste Wahl empfohlen [13].

1.1.3.3.2. Angriffspunkt Atemkette Atovaquon (17, vgl. Abb. 1.5) konkurriert mit Ubichinon um dessen Bindestelle am Cytochrom bc₁-Komplex und blockiert auf diese Weise im Erreger den Elektronentransport im Rahmen der Atemkette [28]. Es hat sich ein ausgesprochen starker Synergismus von Atovaquon und dem DHFR-Inhibitoren Proguanil gezeigt, das Kombinationspräparat Atovaquon plus Proguanil wird insbesondere in der Malariaphylaxe eingesetzt. Die Kombination hat sich als prophylaktisch wirksam, aktiv gegen Blutschizonten und transmissionsblockierend erwiesen [29]. Aus diesem Grund gehört die Kombination neben Mefloquin und Doxycyclin zu den aktuell von der Deutschen Gesellschaft für Tropenmedizin und Internationale Gesundheit (DTG) empfohlenen Medikamenten zur Malariaphylaxe (vgl. Abb. A.1 im Anhang auf S. 230) [30].

1.1.4. Wirkstoffe in der Pipeline

Die besorgniserregende Entwicklung von Resistenzen auch gegen Artemisinin [17, 18] macht die Notwendigkeit der Entwicklung neuer Wirkstoffe zur Bekämpfung der Malaria deutlich. Bei vielen in der Entwicklung befindlichen Arzneistoffen handelt es sich um Weiterentwicklungen von Substanzen aus einer der genannten Klassen, für die daher kein neuer Wirkmechanismus angenommen werden kann. Beispiele für diese Strukturen zeigt Abbildung 1.6, eine umfassende Übersicht geben beispielsweise WELLS *et al.* [6]. An dieser Stelle sollen exemplarisch wichtige Entwicklungen vorgestellt werden.

So wurde die Kombination aus Artesunat (5) und der Neuentwicklung Pyronaridin (18) im ersten Quartal 2010 zur Zulassung bei der Europäischen Arzneimittelagentur (European Medi-

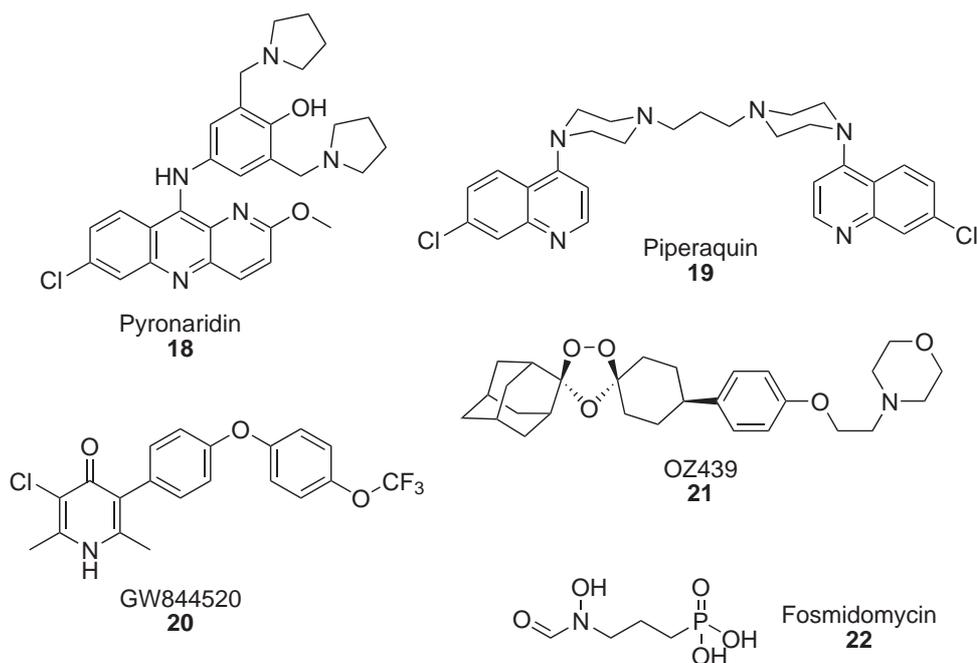


Abbildung 1.6. | Antiplasmodiale Wirkstoffe in der Entwicklung.

eines Agency, EMA) eingereicht, die Kombination aus Dihydroartemisinin (**2**) und Piperaquin (**19**, gemeinsam mit Pyronaridin und weiteren Neuentwicklungen in Abb. 1.6) bereits im Juli 2009. Piperaquin und Pyronaridin können beide zu den 4-Aminochinolinen gezählt werden. Auch wenn Pyronaridin als Grundkörper ein Azaacridin enthält, sind die Substanzen strukturell so ähnlich, dass der gleiche Wirkmechanismus angenommen wird [31].

2009 wurden die klinischen Prüfungen der Phase I für den Kandidaten GSK932121 aus der Substanzklasse der 4-Pyridone aufgenommen [31]. Die Substanzen greifen wie das in Abschnitt 1.1.3.3.2 erwähnte Atovaquon in die Atemkette ein. Die genaue Struktur wird nicht offengelegt, es handelt sich aber prinzipiell um ein Derivat des wegen Kardiotoxizität aus der Entwicklung ausgeschiedenen GW844520 (**20**) [20, 32].

Weiterentwicklungen basierend auf dem wirksamen Prinzip der Artemisinin-Derivate, der Endoperoxid-Struktur, führten neben partialsynthetischen Abwandlungen (z. B. Artemison) vor allem zu den Ozoniden. Das vollsynthetische Ozonid OZ439 (**21**) hat erfolgreich die klinische Prüfung der Phase I absolviert und wird derzeit in Phase II getestet [31, 33].

Substanzklassen mit völlig neuem Wirkmechanismus befinden sich derzeit mehrheitlich in sehr frühen Phasen der Wirkstofffindung. In diesem Zusammenhang ist eine Veröffentlichung von Interesse, die wichtige Anstöße in der Entwicklung neuer Substanzklassen geben könnte. GAMO *et al.* veröffentlichten 2010 die Ergebnisse eines Hochdurchsatzscreenings bei Glaxo-SmithKline, bei dem aus einer Anzahl von knapp 2 000 000 Verbindungen 13 533 als potentielle Leitstrukturen bei der Entwicklung von Wirkstoffen gegen Malaria gelten können, da sie das Wachstum der Parasiten *in vitro* in zwei von drei durchgeführten Tests um mehr als 80 % hemmten [34]. Die Verbindungen sind der Öffentlichkeit als Tres Cantos Antimalarial Compound Set

(TCAMS) zugänglich gemacht worden. Die Qualität des Datensatzes wurde kritisch hinterfragt: So kommen EKINS *et al.* zu dem Schluss, dass schon die Anwendung von einfachen Filtern in der Art der *rule of five*² oder Ausschluss problematischer reaktiver Gruppen den Datensatz um bis zu 76 % reduziert hätten [36]. Andererseits bleibt abzuwarten, was die Forschungsgemeinschaft aus diesem vielversprechenden und umfangreichen Material weiterentwickeln kann.

In weiter fortgeschrittenen Stadien der Entwicklung befinden sich Substanzen, die in den Isoprenoid-Stoffwechsel des Erregers eingreifen, der sich an entscheidenden Punkten von dem humanen unterscheidet. Hier ist das Fosmidomycin (**22**) zu nennen, das die 1-Deoxy-D-xylulose-5-phosphat-Reduktoisomerase (DXR) der Plasmodien hemmt. Die DXR ist ein Schlüsselenzym in der Isoprenoid-Synthese der Erreger, im menschlichen Organismus kommt sie nicht vor und ist daher als Target für antiplasmodiale Wirkstoffe geeignet [37]. Obgleich Kombinationen von Fosmidomycin mit Clindamycin [38, 39] und Artesunat [40] sich als klinisch effektiv erwiesen haben, gibt es bislang kein zugelassenes Präparat.

Eine Strategie bei der Suche nach neuen Targets zur Bekämpfung von *P. falciparum* ist es, die Kenntnisse aus der Wirkstoffforschung homologer Strukturen in anderen Organismen auf die des Erregers zu übertragen [6]. Dieses Vorgehen wurde durch die Aufklärung des plasmodialen Genoms [41] befördert, die einen Überblick über solche homologe Strukturen des Erregers gab. Das Vorgehen wurde beispielsweise bei den Histon Deacetylasen [42] und den Kinasen [43] angewandt. Da sich diese Arbeit mit Inhibitoren eines Enzyms der zuletzt genannten Familie beschäftigt, soll im Folgenden eine kurze Einführung zu Proteinkinasen im Allgemeinen und der Glykogen Synthase Kinase-3 im Speziellen gegeben werden. Des Weiteren werden in Abschnitt 1.3 die plasmodialen Kinasen und insbesondere die plasmodiale Glykogen Synthase Kinase-3 (*PfGSK-3*) vorgestellt.

1.2. Proteinkinasen

Grundlegende Vorgänge in der Zelle wie z. B. die Zellteilung und der Energiestoffwechsel hängen maßgeblich vom Phosphorylierungsgrad beteiligter Proteine ab. Die Wichtigkeit der reversiblen Phosphorylierung für zelluläre Vorgänge wurde bereits 1959 beschrieben [44]. Sie wird heute als essentiell für die „Regulation praktisch aller Aspekte im Leben einer Zelle“ eingeschätzt [45]. Die Phosphorylierung kann an den Alkoholfunktionen von Serin-, Threonin- oder Tyrosinresten geschehen. Enzyme, die diese Reaktion katalysieren, werden als Proteinkinasen bezeichnet, genauer substratabhängig als Serin/Threonin-Kinasen oder aber als Tyrosin-Kinasen [46]. Prinzipiell spielt auch die (De-) Phosphorylierung von Histidin-, Arginin-, Lysin- und Aspartatresten eine Rolle, über die genauen Mechanismen ist jedoch bislang deutlich weniger bekannt als über die Vorgänge an den erstgenannten Aminosäuren [47], so dass im Folgenden mit Proteinkina-

²Einfache Möglichkeit zur ersten Abschätzung der Bioverfügbarkeit potentieller Arzneistoffe nach C. A. LIPINSKI [35].

sen wie allgemein üblich Tyrosin- und Serin/Threonin-Kinasen gemeint sind. Proteinkinasen spielen als Targets in der aktuellen Wirkstoffforschung eine herausragende Rolle [48, 49].

1.2.1. Das humane Kinom

Die Gesamtheit der für Proteinkinasen kodierenden Genabschnitte im menschlichen Organismus wird in Analogie zum Genom als Kinom bezeichnet. Es wurde 2002 von MANNING *et al.* katalogisiert [50], die Autoren identifizierten 518 für Proteinkinasen kodierende Gene und strukturieren diese in Erweiterung des Modells nach HANKS *et al.* [51]. Der Großteil gehört der von HANKS beschriebenen übergreifenden Familie der eukaryotischen Proteinkinasen (ePK) an. Sie wurden von MANNING in die großen Gruppen AGC (enthält Familien der PKA, PKG, PKC), CAMK (Calcium/Calmodulin-abhängige Kinasen), CK1 (Casein Kinase 1), CMGC (CDK, MAPK, GSK und CDK-ähnliche), STE (Homologe der Sterile 7-, Sterile 11-, Sterile 20-Kinasen), TK (Tyrosin-Kinasen), TKL (Tyrosin-Kinasen-ähnliche), RGC (Rezeptor-Guanylat-Cyclasen) und „andere“ eingeteilt, die ihrerseits in Familien und Unterfamilien unterteilt werden können (vgl.

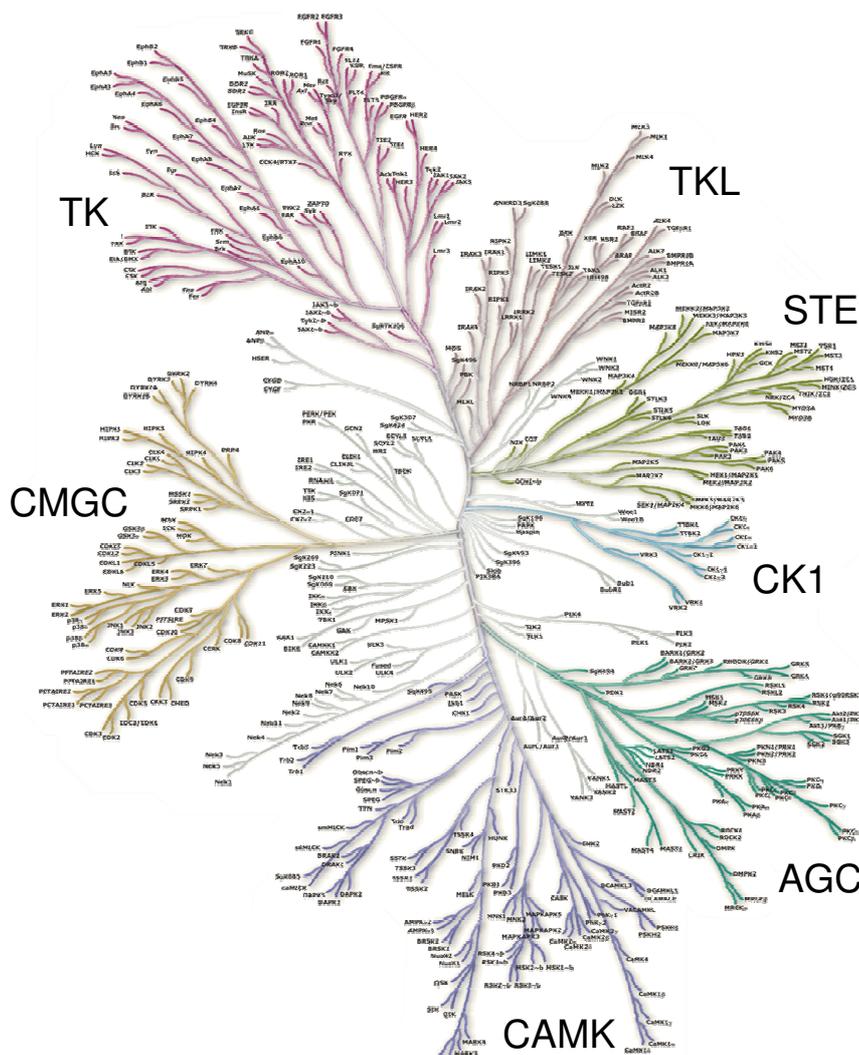


Abbildung 1.7. | Phylogenetischer Baum der 518 humanen Proteinkinasen nach MANNING *et al.* [50]. Die sieben Hauptgruppen sind farblich abgesetzt, nahe der Mitte befinden sich die restlichen Kinasen in grau. Jede ist an der Spitze eines Astes positioniert, je ähnlicher die Enzyme im Rahmen der Klassifizierung eingeschätzt werden, desto näher liegen sie im Diagramm beieinander. Es ist gut zu erkennen, dass die meisten der Kinasen in die Hauptgruppen fallen, in denen die Ähnlichkeit zueinander groß ist.

Abb. 1.7). Zusätzlich wurden 13 Familien sogenannter atypischer Proteinkinasen (aPK) identifiziert. Die Klassifizierung erfolgte zunächst sequenzbasiert (in erster Linie bezogen auf den Bereich der Kinasedomäne). Zusätzlich wurde der Vergleich bekannter biologischer Funktionen und ähnliche Klassifizierungen in anderen Kinomen berücksichtigt [50], so dass bei Zugehörigkeit zu einer Gruppe nahe Verwandtschaft der betreffenden Kinasen zu erwarten ist.

1.2.2. Die dreidimensionale Gestalt von Proteinkinasen

Die katalytische Domäne von Proteinkinasen zeichnet sich durch eine gemeinsame Tertiärstruktur aus, die in Abbildung 1.8 dargestellt ist. Charakteristisch sind die beiden kompakten Unter-Domänen, die über die als Hinge Region (engl.: Scharnier; magenta in Abb. 1.8) bezeichneten Aminosäuren gegeneinander beweglich sind. Während im C-terminalen Bereich α -Helices vorherrschen, ist die N-terminale Domäne von β -Faltblättern bestimmt. Eine Ausnahme bildet die in der N-terminalen Domäne liegende α -Helix C (cyan in Abb. 1.8).

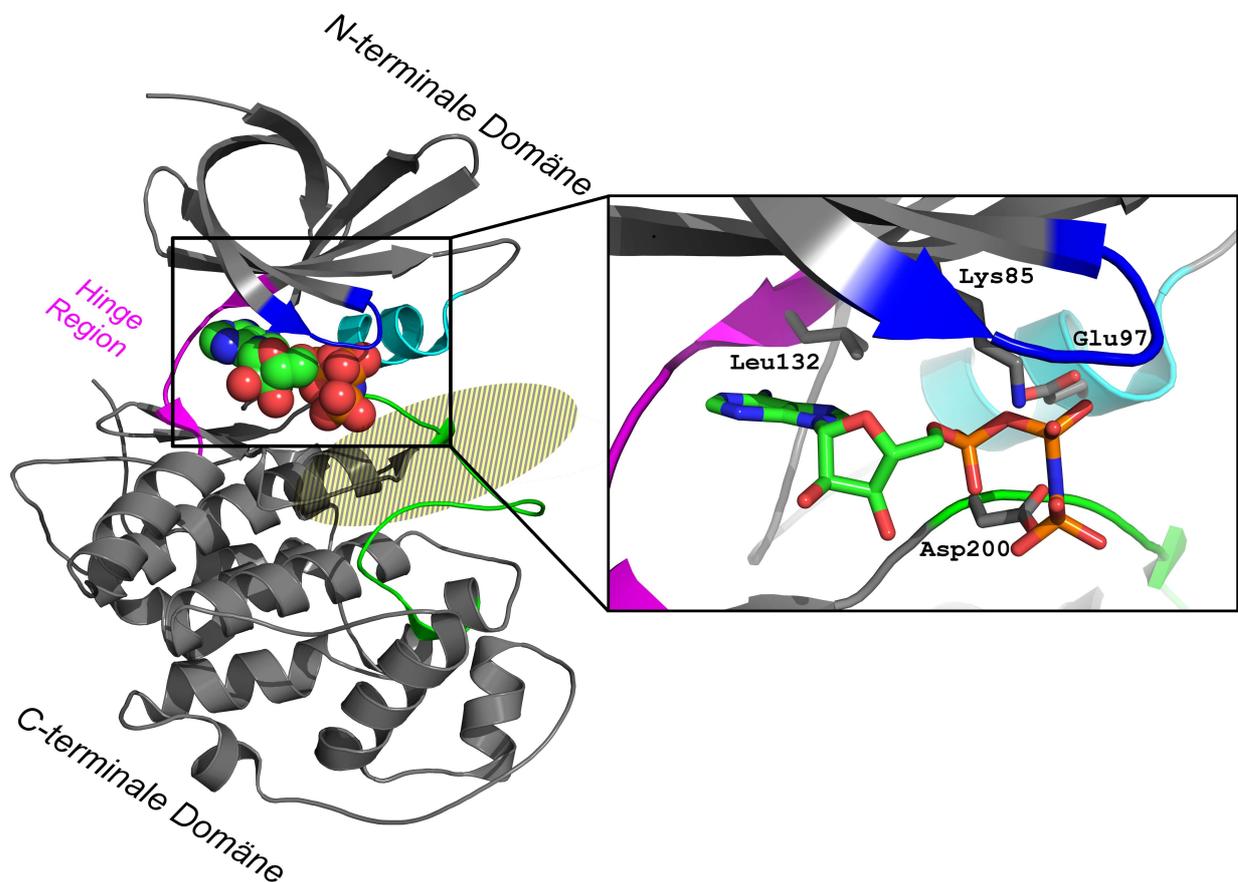


Abbildung 1.8. | Dreidimensionale Struktur von Proteinkinasen am Beispiel der humanen Glykogen Synthase Kinase-3 (*HsGSK-3 β* , PDB-Eintrag 1J1B). Farblich hervorgehoben sind Hinge Region (magenta), Activation Loop (grün), Glycine-rich Loop (blau), α -Helix C (cyan) und der Bereich der substratbindenden Region (gelb). Der Bereich der ATP-Bindestasche ist vergrößert dargestellt, die bezeichneten Aminosäuren sind der Gatekeeper Leu132, das Asp200 des DFG-Motivs und die beiden Aminosäuren Lys85 und Glu97, durch deren Salzbrücke die α -Helix C stabilisiert wird.

Zwischen den beiden Domänen bildet sich eine Furche aus, in der ATP mit dem Purinkörper zur Hinge Region hin bindet, während die Phosphatreste zur Seite in die Richtung des sogenannten Activation Loops (grün in Abb. 1.8) zeigen. Dieser kann bei vielen Kinasen eine Stellung einnehmen, in der er Teile der ATP-Bindetasche verdeckt und auf diese Weise das Enzym inaktiviert. Er beginnt stets mit dem charakteristischen DFG-Motiv (Asparagin, Phenylalanin, Glycin), das auch im Reaktionsmechanismus eine entscheidende Rolle spielt (vgl. S. 13) und deshalb in ePK hochkonserviert vorliegt. Die letzten Aminosäuren des Activation Loops sind in der Regel Alanin, Prolin und Glutaminsäure (APE-Motiv).

Die Position des Activation Loops wird üblicherweise einem der beiden Grenzzustände aktiv bzw. inaktiv zugeordnet. Im ersten Fall ist die Bindetasche frei für ATP und das für den Mechanismus der Übertragung des Phosphatrestes notwendige Aspartat des DFG-Motivs zeigt in die ATP-Bindtasche hinein – die entsprechende Anordnung des DFG-Motivs wird daher als DFG-in bezeichnet. Im inaktiven Zustand ist das Aspartat in der Regel aus der Tasche hinausgewandt, statt dessen nimmt das Phenylalanin des DFG-Motivs einen Teil der Bindetasche ein – diese inaktive Konformation des Activation Loops ist also in der Regel mit einer Anordnung des DFG-Motivs verbunden, die in Entsprechung zur erstgenannten als DFG-out bezeichnet wird. Für die inaktive Form von Kinasen, in der der Activation Loop Teile der Bindetasche verdeckt, wurden allerdings ebenso Strukturen mit DFG-in und Zwischenformen (DFG-out-like) beschrieben [52]. Für den Activation Loop gibt es ebenfalls Übergangsformen. Zwischen diesen liegt ein Gleichgewicht vor, das maßgeblich durch den Phosphorylierungsgrad des Activation Loops bestimmt wird [53].

Neben der Beweglichkeit der C- und N-terminalen Domänen zueinander und der des Activation Loops ist hinsichtlich der Flexibilität noch der sogenannten Glycine-rich Loop (blau in Abb. 1.8) zu nennen. Er kann sich in der Art eines Deckels auf ATP legen und auf diese Weise die Bindung zum Nukleotid oder einem kompetitiven Inhibitor verstärken.

1.2.3. Reaktionsmechanismus der Übertragung des Phosphatrestes

Die katalysierte Reaktion besteht in der Übertragung des γ -Phosphates von ATP auf die Hydroxylgruppe einer Seitenkette (Serin, Threonin, Tyrosin) des Substrates. Dabei spielen verschiedene Aminosäuren eine zentrale Rolle.³ Die Phosphatgruppen werden von zwei Magnesium-Ionen, die über das Asp184 bzw. Asn171 koordiniert werden, und das Lys72, das seinerseits vom Glu91 aus der α -Helix C fixiert wird, in Position gebracht. Die Hydroxylgruppe des Substrates greift nun nucleophil das γ -Phosphat des ATP an. Von entscheidender Bedeutung ist dabei das in Proteinkinasen hochkonservierte Asp166, das die Reaktanden optimal positioniert, einen Übergangszustand stabilisiert und das Proton in einem späten Stadium der Reaktion übernimmt [54–56]. Der Mechanismus ist in Abbildung 1.9 dargestellt.

³Viele Untersuchungen zum Reaktionsmechanismus wurden am Beispiel der Proteinkinase A (PKA) durchgeführt. Nummerierungen in diesem Abschnitt beziehen sich daher auf die PKA.

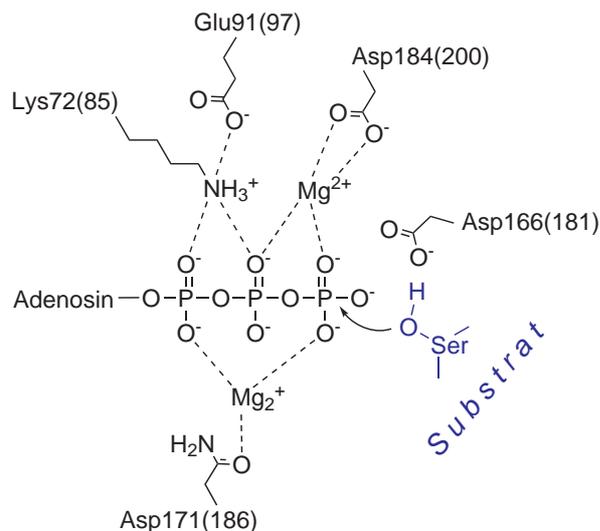


Abbildung 1.9. | Reaktionsmechanismus der Phosphorylierung nach [57]. Die Nummerierungen beziehen sich auf die PKA, in Klammern ist die entsprechende Nummerierung in der *HsGSK-3β* angegeben.

1.2.4. Topologie der ATP-Bindetasche

Wie in Abbildung 1.8 bereits zu erkennen ist, befindet sich die ATP-Bindetasche in Form einer Furche zwischen der N- und der C-terminalen Domäne einer ePK. Sie wird in der Literatur [58, 59] meist in verschiedene Bereiche eingeteilt, die zweidimensional in Abbildung 1.10 am Beispiel der Bindung des ATP in der *HsGSK-3β* dargestellt sind.

Hinge Region Die Hinge Region stellt die Verbindung zwischen den beiden Domänen dar. Vom Backbone dieser Aminosäuren (im Beispiel der *HsGSK-3β*: Asp133 und Val135) gehen nach dem Muster Donor-Akzeptor-Donor Wasserstoffbrückenbindungen aus, die die Position des ATP und diejenige ATP-kompetitiver Inhibitoren bestimmen. Der am weitesten außen gelegene Kontakt vom Carbonyl des Backbones der Hinge Region zum C2 des Purinkörpers wird nicht von allen Autoren als Wasserstoffbrückenbindung anerkannt. Der Charakter dieser Interaktion wurde am Beispiel von Kinaseinhibitoren von PIERCE *et al.* [60, 61] untersucht.

Back Pocket An die Hinge Region schließt sich ins Innere des Proteins eine hydrophobe Untertasche an – die Back Pocket (orange in Abb. 1.10), deren Zugänglichkeit vom sogenannten Gatekeeper bestimmt wird und die in ihrer Größe zwischen unterschiedlichen ePK variiert [59]. Die Selektivität verschiedener Inhibitoren wird durch Kontakte zum Gatekeeper möglich (z. B. [62–64]).

Adenine Region Dieser Bereich bildet das hydrophobe Zentrum der Bindetasche (blau in Abb. 1.10). Der Purinkörper des ATP wird sandwichartig zwischen Glycine-rich Loop und dem Grund der Tasche eingebettet und kann zahlreiche Van-der-Waals-Kontakte sowie die oben beschriebenen Wasserstoffbrückenbindungen zur Hinge Region ausbilden.

Front Pocket Diese Region wird von TRAXLER als zweite hydrophobe Region beschrieben, die in der Art einer Kerbe im Oberflächenprofil zum Lösungsmittel hin verläuft (rot in Abb.

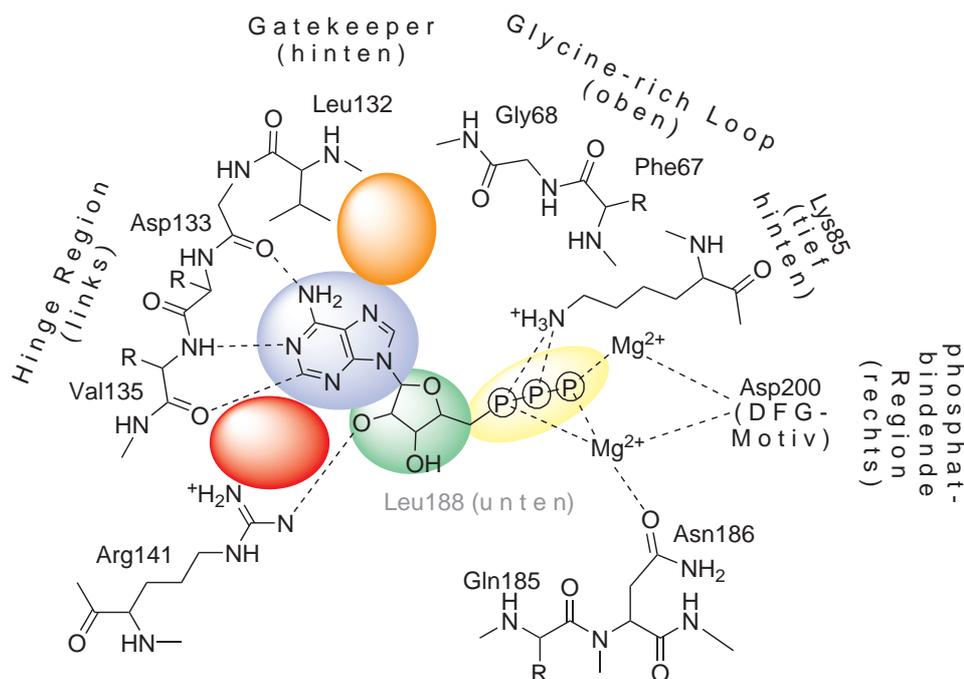


Abbildung 1.10. | Topologie der ATP-Bindetasche mit den im Text beschriebenen farbigen gekennzeichneten Regionen (orange: Back Pocket, blau: Adenine Region, rot: Front Pocket, grün: Ribose Pocket, gelb: Phosphate Pocket). Die Nummerierung der Aminosäuren bezieht sich auf die *HsGSK-3β*.

1.10) und im EGFR von Gly772 und Leu649 bestimmt wird [58]. Da dieser Bereich nicht hochkonserviert ist, kann er zur Verbesserung der Selektivität, aber auch zur Modulation der ADME-Eigenschaften eines Inhibitors genutzt werden [59]. In vielen Kinasen fehlt das Analogon zum Gly772, insbesondere in denen der CDK-Familie [58]. Dadurch ist der Charakter des Bereiches weniger der einer Tasche, der Übergang zum Lösungsmittel und damit zu einem hydrophilen Bereich ist unmittelbarer. Dies ist auch bei der GSK-3 der Fall. Der Bereich wird in der vorliegenden Arbeit daher als lösungsmittelzugewandte Region bezeichnet und ist weniger als hydrophobe Tasche zu verstehen.

Ribose Pocket Dieser Bereich öffnet sich neben der Front Pocket zum Lösungsmittel hin (grün in Abb. 1.10). Er ist vorwiegend von hydrophilem Charakter und im Vergleich zur Front Pocket in ePK konserviert.

Phosphate Pocket Die phosphatbindende Region (gelb in Abb. 1.10) ist in ePK ebenfalls hochkonserviert und hat ausgeprägt hydrophilen Charakter. Sie grenzt an den äußeren Rand des Glycine-rich Loops und damit an einen der flexibelsten Teile des Proteins [59]. Sie ist aus diesen Gründen kein bevorzugter Angriffspunkt im Design selektiver Inhibitoren von ePK.

1.2.5. Proteinkinaseinhibitoren

Die prominente Rolle von Proteinkinaseinhibitoren in der Wirkstoffforschung führte in den letzten Jahren zu einer Vielzahl an diesbezüglichen Veröffentlichungen und zu einer Vielfalt von neuen Substanzen mit inhibitorischen Eigenschaften an ePK. Sie können gemäß ihrer Interaktion mit dem Protein in verschiedene Typen eingeteilt werden [53, 65]. Diese werden im Folgenden kurz charakterisiert, Beispiele für experimentell aufgeklärte dreidimensionale Strukturen sind anhand ihrer Kennung in der Protein Database (PDB, [66, 67]) aufgeführt.

Typ-I-Inhibitoren

Inhibitoren, die die klassischen Wasserstoffbrückenbindungen zur Hinge Region ausbilden (vgl. Abb. 1.10 auf S. 14) und sich durch einen Heterozyklus auszeichnen, der die Position des Purinkörpers im ATP einnimmt, werden als Typ-I-Inhibitoren bezeichnet. Sie binden in der Regel an die aktive Form einer Kinase, die aufgrund der allen ePK gemeinsamen, für die Katalyse essentiellen Strukturmerkmale eine Unterscheidung zwischen verschiedenen Proteinkinasen und damit eine Selektivität der Inhibitoren erschwert. Beispiele für Typ-I-Inhibitoren sind Gefitinib (**23**, PDB-Eintrag 2IT0 [68]) und Sunitinib (**24**, PDB-Eintrag 3MIY [69]) als Wirkstoffe zugelassener Arzneimittel oder Staurosporin (**25**, z. B. PDB-Eintrag 1Q3D [70]) als Paradebeispiel eines unselektiven Proteinkinaseinhibitors (vgl. Abb. 1.11).

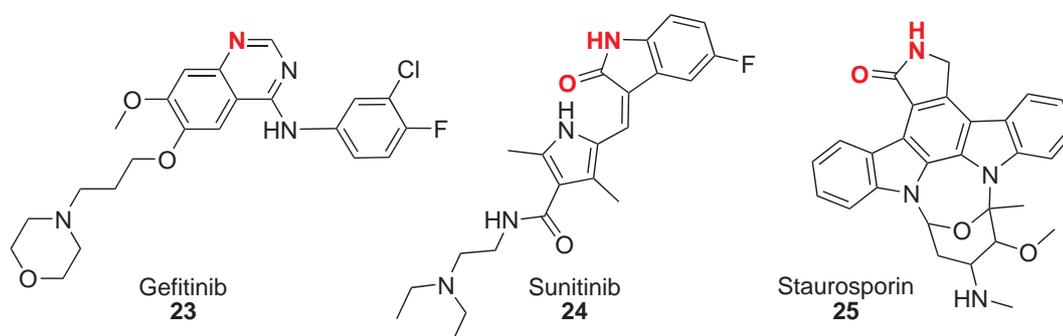


Abbildung 1.11. | Beispiele für Inhibitoren vom Typ I. Über die rot markierten Heteroatome werden Wasserstoffbrückenbindungen zur Hinge Region ausgebildet.

Typ-II-Inhibitoren

Inhibitoren vom Typ II binden an die inaktive Form einer ePK. Diese zeichnet sich in der Regel durch die oben beschriebene DFG-out-Konformation aus. Durch die Umstrukturierung wird ein Bereich der Bindetasche zugänglich, der an die Adenine Region angrenzt und in der Literatur als allosterische Tasche bezeichnet wird. Der erste Vertreter vom Typ II und der erste zugelassene Kinaseinhibitor überhaupt war Imatinib (**26**, PDB-Eintrag 1IEP [71]), der Protein-Ligand-Komplex ist in Abbildung 1.12 dargestellt. Die allosterische Tasche unterscheidet sich zwischen ePK stärker, wodurch eine Selektivität der Inhibitoren besser zu erreichen ist [72].

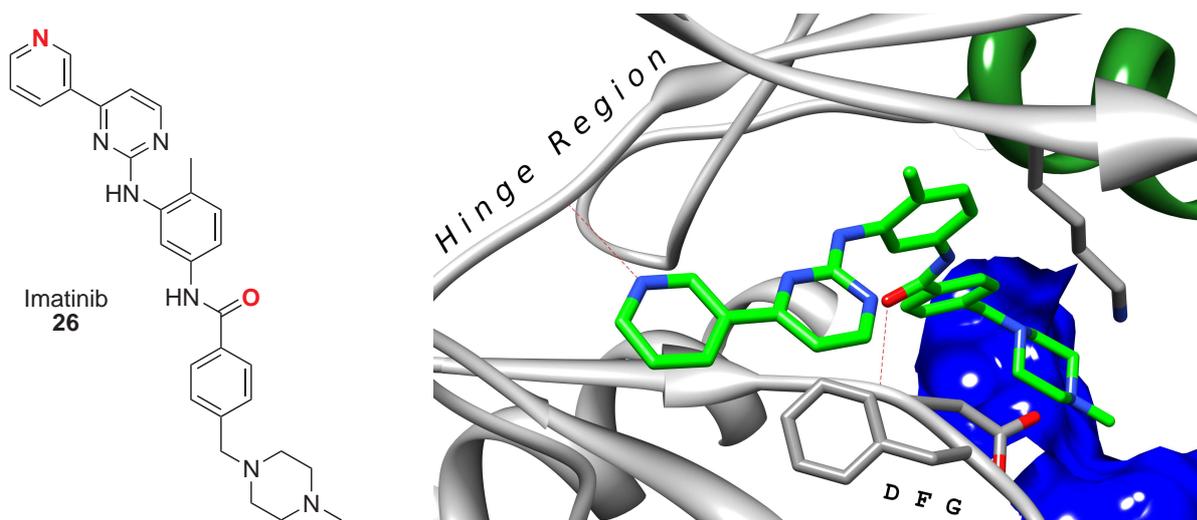


Abbildung 1.12. | Imatinib (**26**) als Beispiel eines Proteinkinaseinhibitors vom Typ II, in der Strukturformel rot gekennzeichnete Atome bilden wichtige Wasserstoffbrücken aus. Inhibitoren dieses Typs haben einen Wasserstoffbrückenkontakt zur Hinge Region und besetzen zusätzlich die sogenannte allosterische Tasche, die sich durch die konformativen Änderungen in der Bindetasche in der DFG-out-Position auftut (blaues Oberflächenprofil). In der Abbildung ist gut zu erkennen, dass die Seitenkette des Phenylalanins des DFG-Motivs in die Tasche hineinzeigt und eine Wasserstoffbrücke zum Backbone des Aspartats ausgebildet wird. Im Hintergrund sind das für den Reaktionsmechanismus wichtige Lysin und die grün gekennzeichnete α -Helix C zu erkennen – beide liegen in der inaktiven Form weiter vom Zentrum der Bindetasche entfernt als in der aktiven (vgl. Abb. 1.8).

ZUCCOTTO *et al.* definieren zusätzlich einen Typ I^{1/2}, der sich durch Interaktionen mit Aminosäuren der allosterischen Tasche auszeichnet, ohne dass das Protein jedoch in der inaktiven DFG-out-Konformation vorliegt [73]. Inhibitoren dieser Gruppe bilden zusätzlich lipophile Interaktionen mit der hydrophoben Back Pocket aus. Eine noch differenziertere Unterteilung der Back Pocket und davon abgeleitete Typisierung von Bindungsmodi diskutieren LIAO *et al.* [52].

Typ-III-Inhibitoren

Nach RABILLER *et al.* binden die Vertreter des Typs III nur in der allosterischen Tasche. Die Autoren definieren zudem einen Typ IV, bei dem die Inhibitoren einige Ångström von der eigentlichen ATP-Bindetasche entfernt binden und konstatieren, dass die Typen III und IV allgemein als allosterische Inhibitoren bezeichnet werden [53].

Ein Beispiel für den Typ III sind die PDB-Einträge 1S9I und 1S9J [74], in denen ein nicht kompetitiver Inhibitor der MEK2 bzw. MEK1 gemeinsam mit ATP kokristallisiert ist. Abbildung 1.13 zeigt exemplarisch die in 1S9I kokristallisierte Verbindung **27**. Weitere Beispiele allosterischer Inhibitoren werden übersichtsartig von BOGOYEVITCH *et al.* beschrieben [75]. Sie stehen derzeit im besonderen Fokus der Entwicklung, da ihr Potential hinsichtlich der Möglichkeiten einer Selektivität aufgrund der weniger konservierten allosterischen Bindetasche hoch eingeschätzt wird [49].

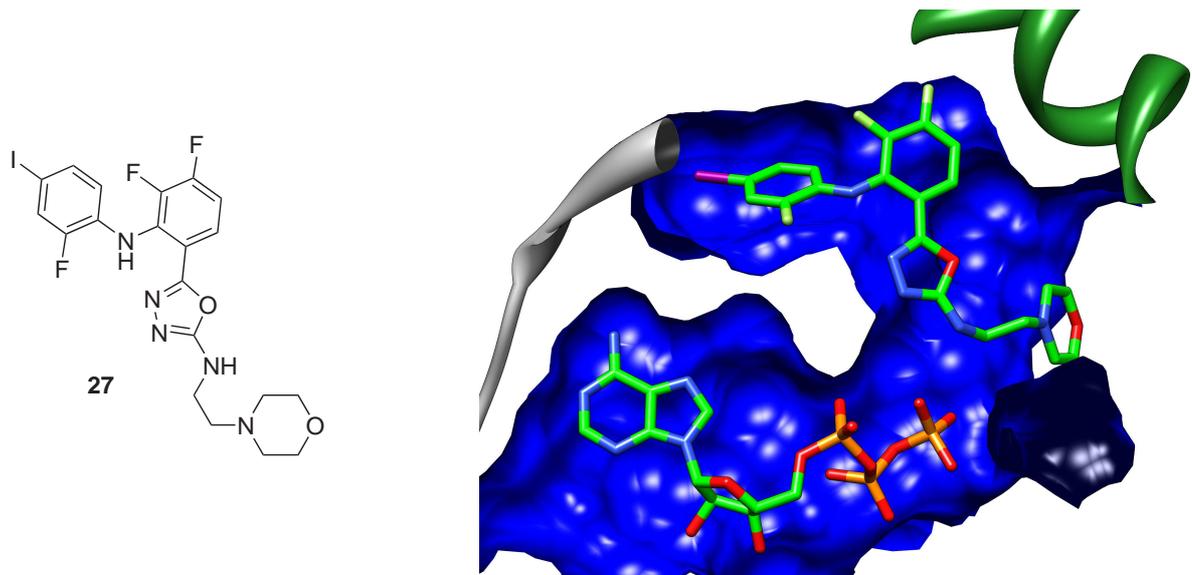


Abbildung 1.13. | **27** als Beispiel eines Proteinkinaseinhibitors vom Typ III. ATP bindet in der aus Abbildung 1.10 bekannten Position unter Ausbildung der Wasserstoffbrückenbindungen zur in der Abbildung grau dargestellten Hinge Region. Gleichzeitig sitzt **27** tief in der Back Pocket und in Richtung der in grün dargestellten α -Helix C.

1.2.6. Die Glykogen Synthase Kinase-3

1.2.6.1. Physiologische und pathophysiologische Relevanz

Die Glykogen Synthase Kinase 3 (GSK-3) ist eine Serin/Threonin-Kinase, die als Regulator der insulinabhängigen Glykogensynthese entdeckt wurde [76]. Sie ist Mitglied der CMGC-Familie der ePK und somit nahe verwandt mit Cyclin-abhängigen Proteinkinasen (CDK), Mitogen-aktivierten Proteinkinasen (MAPK) und CDK-ähnliche Kinasen (CLK, vgl. Abb. 1.14). Die GSK-3 phosphoryliert die Glykogen Synthase, wodurch diese inaktiviert und der Glukosespiegel aufrechterhalten werden kann, da weniger Glukose zur Glykogensynthese verwendet wird. Es existieren die beiden Isoformen GSK-3 α und GSK-3 β , die innerhalb der Proteinkinasedomäne als fast identisch anzusehen sind (98%) und ubiquitär exprimiert werden [77, 78]. Die genaue Abgrenzung der Isoformen der humanen GSK-3 hinsichtlich ihrer Funktion ist bislang nicht vollständig verstanden [79].⁴

Die *HsGSK-3 β* ist an zahlreichen Signaltransduktionswegen beteiligt, darunter die Insulin-, Brain-derived Neurotrophic Factor (BDNF)-, Hedgehog- und Wnt/ β -Catenin-Signalkaskade. Es sind zudem diverse Substrate bekannt, neben Proteinen aus den genannten Signalkaskaden sind dies insbesondere das Tau-Protein und eine Reihe an Transkriptionsfaktoren (Überblick z. B. in [77, 79]). Die *HsGSK-3 β* wurde zunächst vorrangig als potentielles Target in der Therapie des Diabetes mellitus Typ II oder neurodegenerativer Erkrankungen beforscht [81]. Insbeson-

⁴Die *PfGSK-3*, die als Target im Mittelpunkt dieser Arbeit steht, ist näher mit der Isoform GSK-3 β als mit der Isoform GSK-3 α verwandt [80]. Es wird daher im Folgenden die Abkürzung *HsGSK-3 β* für das humane Enzym verwendet, auch wenn viele der beschriebenen Sachverhalte für die Isoform GSK-3 α ähnlich anzunehmen sind.

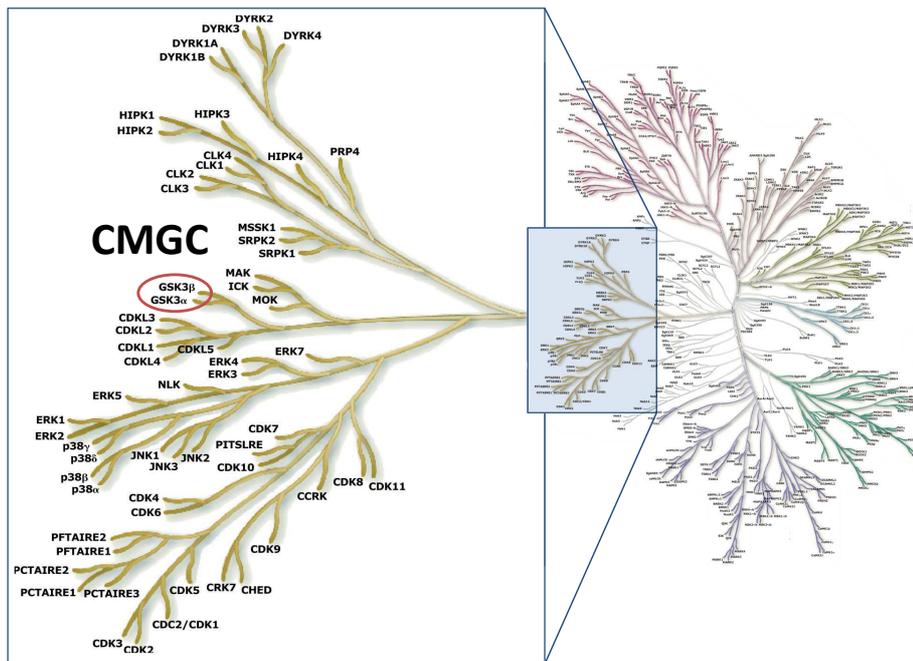


Abbildung 1.14. | Humane Proteinkinasen der Gruppe CMGC nach [50]. Die GSK-3 ist rot gekennzeichnet.

dere der potentielle Einsatz von Inhibitoren der *HsGSK-3β* bei der Alzheimer Demenz erschien dabei aussichtsreich, da ein hyperphosphoryliertes Tau-Protein zu den charakteristischen pathophysiologischen Erscheinungen im Zusammenhang mit dieser neurodegenerativen Erkrankung zählt [82]. Mit der wachsenden Kenntnis der vielfältigen Rolle der *HsGSK-3β* im menschlichen Organismus wurden zusätzlich verschiedene Krebserkrankungen, Osteoporose und Herzhypertrophie als mögliche therapeutische Einsatzgebiete von Inhibitoren der *HsGSK-3β* diskutiert. Einen aktuellen Überblick geben PHUKAN *et al.* [83].

1.2.6.2. Regulation und Spezifität der *HsGSK-3β*

Eine Besonderheit der *HsGSK-3β* besteht darin, dass das Enzym konstitutionell aktiv vorliegt. Die Regulation erfolgt vor allem über Phosphorylierung an Ser9 im N-terminalen Ende bzw. Autophosphorylierung des Tyr216 im Activation Loop [78]. Während die Phosphorylierung am Ser9 zu einer Inaktivierung führt, bewirkt die Autophosphorylierung des Tyr216 eine etwa 5-fach verstärkte Aktivität. Auch in der weniger aktiven Form ohne Phosphotyrosin positioniert sich der Activation Loop der *HsGSK-3β* nicht in der charakteristischen geschlossenen Konformation einer inaktiven Kinase, nur die Seitenkette des Tyr216 ist um 120° in Richtung der Aminosäuren Arg220 und Arg223 gedreht, während der Backbone weiterhin in der aktiven Form bleibt [84]. Es wurde von der *HsGSK-3β* bislang keine Kristallstruktur in der inaktiven Form mit geschlossenem Activation Loop und DFG-out-Konformation veröffentlicht.

Eine weitere Besonderheit, die die *HsGSK-3β* von den nahe verwandten CDKs unterscheidet, ist ihre Spezifität für Substrate, die bereits phosphoryliert sind. Diese bereits vorhandene Phosphorylierung des Substrates muss von dem zu phosphorylierenden Serin oder Threonin aus gerechnet an einem vier Aminosäuren in C-terminaler Richtung liegenden Serin oder Threonin

vorliegen und führt zu einer Steigerung der Aktivität auf das 100- bis 1000-fache im Vergleich zum nicht phosphorylierten Substrat [85]. Der Mechanismus wird als Primed Phosphorylation bezeichnet und ist im Zusammenhang mit der Aktivierung zu sehen. Während andere Kinasen zur Aktivierung typischerweise an mehreren Positionen des Activation Loops phosphoryliert werden müssen, ist dies für die *HsGSK-3 β* nicht nötig – der phosphorylierte Rest des Substrates kann in der Oxyanion-Tasche nahe den Resten Lys205, Arg180 und Arg96 einen ähnlichen Effekt auf den Activation Loop ausüben, so dass die mehrfache Phosphorylierung des Activation Loops nicht notwendig ist [77].

Dieser Mechanismus ist auch Teil der Inaktivierung durch Phosphorylierung am N-terminalen Ser9. Das resultierende Phosphoserin ist dabei Resultat einer Primed Phosphorylation und kann in der Art eines Pseudosubstrates intramolekular die Substratbindestelle blockieren und die *HsGSK-3 β* auf diese Weise inaktivieren [79].

1.2.6.3. Inhibitoren der *HsGSK-3 β*

Wegen ihrer vielversprechenden Eigenschaften als potentiell Target ist die *HsGSK-3 β* auch röntgenkristallografisch bereits von einer Anzahl verschiedener Arbeitsgruppen untersucht worden, was eine Reihe von veröffentlichten Protein-Ligand-Komplexen der *HsGSK-3 β* zum Ergebnis hatte (Auflistung siehe Tabelle 2.2 auf S. 35).

Die Inhibitoren lassen sich gemäß der Einteilung aus Abschnitt 1.2.5 in der Regel als Typ-I-Inhibitoren einordnen, die als ATP-kompetitive Inhibitoren wirken. Sie bilden die für diese Gruppe charakteristischen mehrfachen Wasserstoffbrückenbindungen zur Hinge Region aus, und es liegt im Bereich der Adenine Region ein Heteroaromat vor, der lipophile Interaktionen mit Aminosäuren oberhalb der Adenine Region und solchen am Taschenrund ausbilden kann.

Beispielhaft ist in Abbildung 1.15 der Bindungsmodus des Inhibitors Alsterpaullon (**28**) ab-

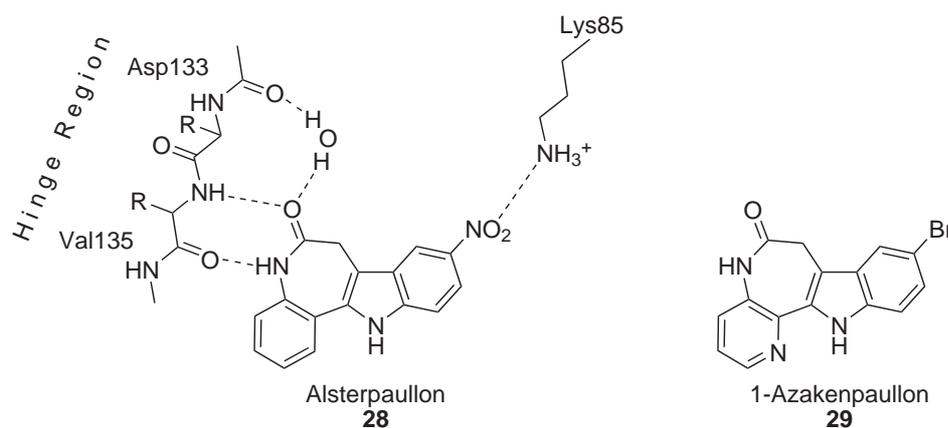


Abbildung 1.15. | Paullone als Beispiele für Typ-I-Proteinkinaseinhibitoren in der *HsGSK-3 β* . Am Beispiel von Alsterpaullon (**28**) werden typische Merkmale des Bindungsmodus von Inhibitoren dieses Typus deutlich: Wasserstoffbrückenbindung zur Hinge Region und Heterozyklus in der Adenine Region. Rechts ist 1-Azakenpaullon (**29**) als selektiver Inhibitor der *HsGSK-3 β* abgebildet.

gebildet, wie er im Komplex mit der *HsGSK-3 β* vorliegt (PDB-Eintrag 1Q3W [70]). Während das kokristallisierte Alsterpaullon als Inhibitor der CDK (insbesondere CDK1/Cyclin B) entwickelt wurde [86] und Aktivität an *HsGSK-3 β* und CDK1 zeigt, konnte mit 1-Azakenpaullon (**29**) eine Verbindung entwickelt werden, deren IC_{50} an der *HsGSK-3 β* bei $0.018 \mu M$ liegt und die damit eine 111-fache Selektivität gegenüber CDK1/Cyclin B zeigt (IC_{50} an CDK1/Cyclin B: $2 \mu M$, [87]).

Die ersten nicht ATP-kompetitiven Inhibitoren der *HsGSK-3 β* wurden von MARTINEZ *et al.* beschrieben. Es handelt sich um Thiadiazolidinone, in Abbildung 1.16 ist das aktivste Derivat dieser Reihe, TZDZ-8 (**30**), mit einem IC_{50} von $2 \mu M$ dargestellt. Es liegt bislang keine Röntgenstruktur mit einem TDZD vor, die Autoren postulierten zunächst eine Bindung in der Oxyanion-Bindetasche im äußeren Bereich des Activation Loop nahe Lys205 und Arg96. Die genannten Aminosäuren könnten dabei Wasserstoffbrücken zu den Carbonylsauerstoffen am Heterozyklus ausbilden [88]. Weiterführende Untersuchungen der gleichen Arbeitsgruppe führten daneben zu einer alternativen Hypothese, bei der die TDZD doch innerhalb der ATP-Bindetasche liegen [89]. **30** wurde im Tierversuch erfolgreich zur Reduktion der mit Alzheimer Demenz assoziierten Tau-Hyperphosphorylierung eingesetzt [90]. Außerdem laufen derzeit Phase II Studien zum Einsatz von **30** bei progressiver supranukleärer Paralyse (PSP), einer Erkrankung, die wie die Alzheimer Demenz zur Gruppe der Tauopathien zählt [91].

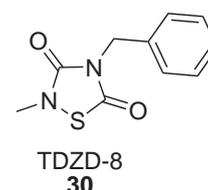


Abbildung 1.16. | Thiadiazolidinon TZDZ-8 (**30**, auch NP12).

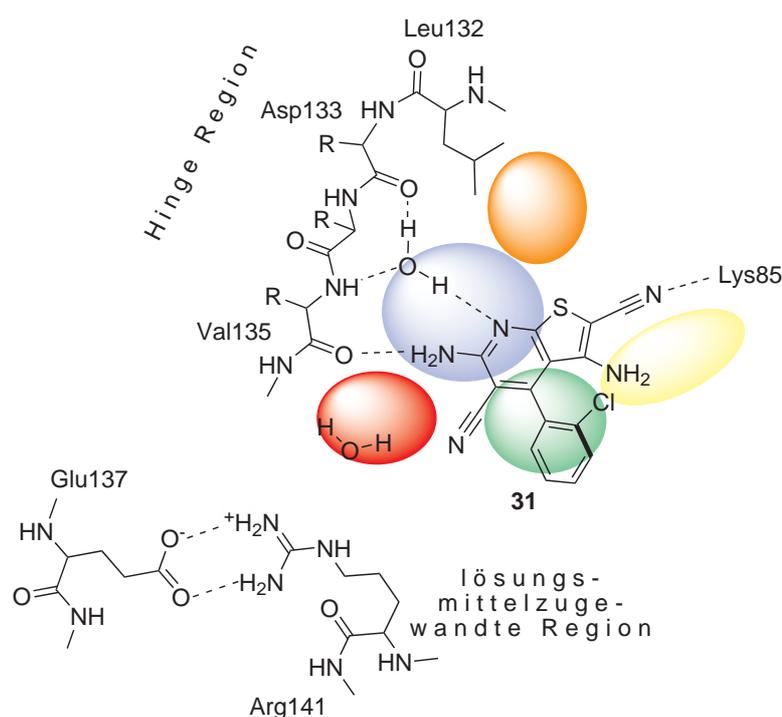


Abbildung 1.17. | Thienopyridin **31** in der Bindetasche der *HsGSK-3 β* . Die farbig gekennzeichneten Regionen entsprechen der auf Seite 14 vorgestellten Einteilung nach [58, 59]. Die Back Pocket (orange) ist trotz des vergleichsweise kleinen Gatekeepers Leu132 nicht sehr groß. Durch die Positionierung des Wassers im vorderen Bereich der Bindetasche wird deutlich, dass der lösungsmittelzugewandte Bereich (rot) weniger stark ausgeprägten Charakter einer lipophilen Tasche hat. Das Wasser in diesem Bereich besetzt eine Position, an der auch der Stickstoff des Arg141 liegen kann (z. B. in der A-Kette von 1J1B). Allerdings gibt es mit Glu137 einen Partner zur Ausbildung einer Salzbrücke – diese Interaktion liegt in kubra vor und ermöglicht das Wasser im Bereich der Front Pocket. Die gelb gekennzeichnete phosphatbindende Region wird von **31** kaum besetzt.

Für die vorliegende Arbeit von besonderem Interesse ist der Bindungsmodus der Verbindung **31** in der *HsGSK-3 β* . Es existiert eine bislang nicht in der PDB abgelegte Röntgenkristallstruktur des Komplexes von **31** mit der *HsGSK-3 β* , die im Rahmen der Arbeiten von BRANDT [92] entstand.⁵ Die Verhältnisse in der Bindetasche sind in Abbildung 1.17 dargestellt. Der Thieno[2,3-*b*]pyridin-Grundkörper nimmt einen Teil der für Typ-I-Inhibitoren charakteristischen Position im Bereich der Adenine Region ein. Es werden Wasserstoffbrückenbindungen zum Lys85 und zur Hinge Region ausgebildet. Bemerkenswert ist dabei, dass nur eine einzige direkte Wasserstoffbrücke vorliegt und diese den äußersten der drei üblicherweise bestehenden Kontakte darstellt (vgl. Abb. 1.10 auf S. 14). Dafür wird zusätzlich eine indirekte Wasserstoffbrücke vom Stickstoff des Pyridins über das in der Abbildung dargestellte Wasser ausgebildet. Das Wasser führt allerdings dazu, dass der Heteroaromat vergleichsweise weit aus der Tasche herausgeschoben wird und der 4-(2-Chlorphenyl)-Rest zum Teil in der Ribose Pocket liegt.

1.3. Plasmodiale Kinasen und die *PfGSK-3*

1.3.1. Das plasmodiale Kinom

Auf die Aufklärung des plasmodialen Genoms im Jahr 2002 [41] und dessen Überführung in eine öffentlich zugängliche Datenbank (PlasmoDB; <http://PlasmoDB.org> [93]) folgte 2004 die Publikation der Gesamtheit von Genen, die für plasmodiale Proteinkinasen kodieren – des plasmodialen Kinoms [94, 95]. Im Vergleich zum humanen Kinom ist das plasmodiale kleiner, zur genauen Anzahl plasmodialer Proteinkinasen gibt es unterschiedliche Angaben. ANAMIKA *et al.* nennen 99 für Proteinkinasen kodierende Sequenzen, die sie als homolog zu anderen ePK einschätzen [95]. WARD *et al.* ordnen 65 der plasmodialen Kinasen bekannten Familien der ePK zu (AGC, CMGC, CK1, TKL und CAMK; Einordnung nach HANKS [96, 97], vgl. S. 10), während sie für die Familien TK und STE keine plasmodiale Entsprechung finden [94]. Zusätzlich identifizieren sie 20 Mitglieder der neuen, für Apicomplexa spezifischen Familie der FIKK (benannt nach einem konservierten Phe-Ile-Lys-Lys-Motiv) und weitere plasmodiale Kinasen, die keiner der Gruppen der ePK zuzuordnen sind (sog. Orphan Kinases). Diese sind als potentielle Angriffspunkte antiplasmodialer Wirkstoffe besonders reizvoll, da weniger Probleme hinsichtlich der Selektivität der zu entwickelnden Inhibitoren zu erwarten sind [43] (Prinzip der selektiven Toxizität von Antiinfektiva nach EHRlich). Als Vertreter der Orphan Kinases soll beispielhaft die Protein Kinase 7 von *P. falciparum* (*PfPK7*) vorgestellt werden, da hier zudem Kristallstrukturen vorliegen, die von Interesse sind hinsichtlich der strukturellen Besonderheiten von plasmodialen Proteinkinasen. Einen umfassenden Überblick geben DOERIG *et al.* [98, 99].

⁵Persönliche Mitteilung durch A. Oberholzer, Universität Bern und L. Meijer, Station Biologique, Roscoff. Die Struktur wird im Folgenden anstelle einer PDB-Kennung als **kubra** bezeichnet.

1.3.2. Orphan Kinases: Das Beispiel der *PfPK7*

Da eine Inaktivierung der *PfPK7* die asexuelle Vermehrung der Plasmodien beeinträchtigt [100], wurde sie als vielversprechendes Target näher untersucht. MERCKX *et al.* gelang es 2008, mehrere Kristallstrukturen des Enzyms mit verschiedenen Liganden aufzuklären [101]. Dabei konnten sie mehrere Sequenzeinschübe identifizieren, wie sie typischerweise bei plasmodialen Enzymen vorliegen [102].

Außerdem werden von MERCKX *et al.* vier ATP-kompetitive Inhibitoren vorgestellt, die die *PfPK7* im niedrigen mikromolaren Bereich inhibieren. Der Inhibitor K510 (**32**) bindet dabei mit seiner Nitrilfunktion über ein konstitutiv gebundenes Wasser in der Back Pocket (PDB-Eintrag 2PMN). Diese ist in der *PfPK7* zugänglich und groß genug für diese Interaktion, obwohl in der *PfPK7* mit Tyr117 ein großen Gatekeeper vorliegt. Mit **32** strukturell vergleichbare Imidazopyridazine wurden auch von der Arbeitsgruppe von BRYANS bearbeitet. Das am stärksten aktive Derivat **33** (vgl. Abb. 1.18) erreichte dabei nach Optimierung der Verbindungen einen IC_{50} von $0.13 \mu M$ [103].

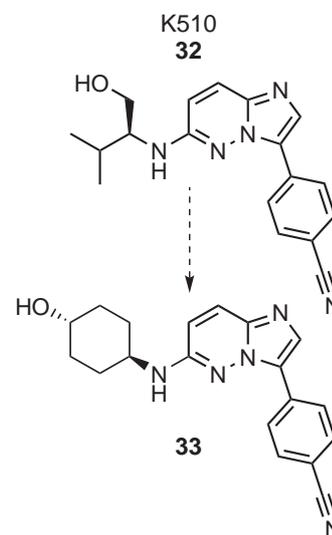


Abbildung 1.18. | Optimierung von Inhibitoren der *PfPK7*.

1.3.3. Plasmodiale CMGC und die *PfGSK-3*

Auch wenn die Idee vielversprechend erscheint, auf ein Target des Erregers abzielen, das keine Entsprechung im menschlichen Organismus hat, wurde ebenso versucht, Kinasen zu adressieren, für die homologe Enzyme im Menschen existieren. Innerhalb der Gruppe CMGC wurden beispielsweise Versuche unternommen, mit selektiven Inhibitoren die plasmodialen Analoga zu humanen CDKs zu inhibieren, das dabei untersuchte Target war die Protein Kinase 5 von *P. falciparum* (*PfPK5*). Allerdings zeigten die getesteten Inhibitoren Staurosporin (**25**), Hymenialdisin (**34**), Purvalanol (**35a/b**) und Indirubin-3'-monoxim (**36**, vgl. Abb. 1.19) am humanen Enzym stärkere Aktivität als am plasmodialen [104, 105].

Ebenfalls aus der Gruppe CMGC stammt die *PfGSK-3*, deren Klonierung und Expression 2004 von DROUCHEAU *et al.* beschrieben wurde [80]. Die Homologie zum humanen Enzym wird von den Autoren anhand der Sequenz einerseits und durch Antikörperreaktionen andererseits belegt, wobei die *PfGSK-3* der Isoform *HsGSK-3 β* näher steht als der Isoform *HsGSK-3 α* (Sequenzen siehe 2.3.4.1/2.4.1.1 auf S. 33/47). DROUCHEAU *et al.* konnten das Protein in Erythrozyten nachweisen, die mit *P. falciparum* infiziert waren.

Aufgrund der vielfältigen Bedeutungen der *HsGSK-3 β* im humanen Organismus wurden diverse Funktionen des Enzyms im Parasiten vorgeschlagen. Sie reichen von einer Beteiligung im Glukosestoffwechsel über eine Rolle im Zellzyklus bis hin zur Differenzierung der verschie-

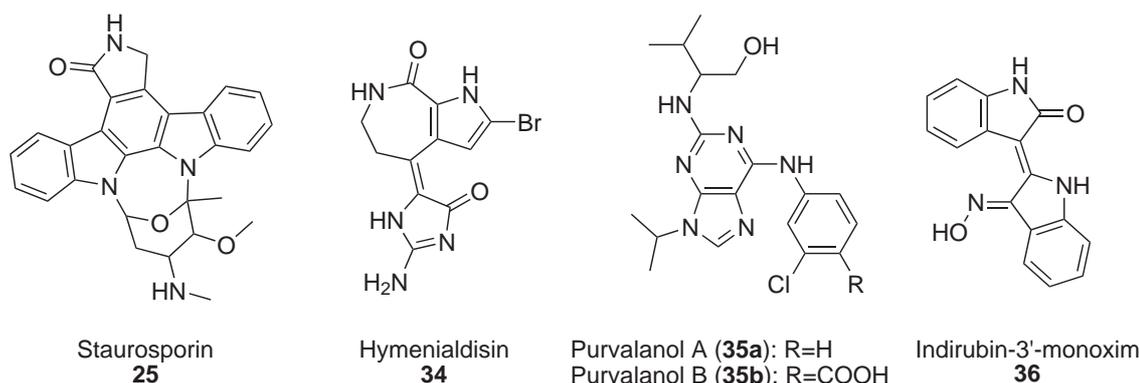


Abbildung 1.19. | Inhibitoren der PfPK5.

denen Entwicklungsstadien von *P. falciparum*. DROUCHEAU *et al.* favorisieren jedoch einen Zusammenhang mit der Regulation des zirkadianen Rhythmus. Für ein homologes Enzym der GSK-3 in der Fruchtfliege wurde eine solche Rolle beschrieben [106]. Zudem gibt es mit der Casein Kinase 1 von *P. falciparum* (PfCK-1) ein weiteres Enzym, für das in der Fruchtfliege ein homologes beschrieben wurde, das eine Rolle im zirkadianen Rhythmus spielt [107]. Das Vorhandensein zweier Kinasen, für die homologe Enzyme in der Fruchtfliege existieren, die eine Rolle im zirkadianen Rhythmus spielen, führt DROUCHEAU *et al.* zu der Annahme einer homologen Signalkaskade bei *P. falciparum*, in der die PfGSK-3 eine Rolle spielen könnte [80].

In der genannten Veröffentlichung wurden auch einige bekannte Inhibitoren der HsGSK-3 β auf ihre Aktivität an der PfGSK-3 hin geprüft (Beispiele in Tab. 1.1). Dabei wurden mit vier getesteten Paullonen (**28**, **37-39**) Substanzen identifiziert, die am humanen Enzym etwa 100- bis 300-mal stärker wirksam sind als am plasmodialen. Die starken (aber nicht selektiven) GSK-3-Inhibitoren Indirubin-3'-monoxim (**36**) und Hymenialdisin (**34**) waren in der Lage, die Proliferation von Plasmodien in *in vitro*-Kulturen zu hemmen. Diese Ergebnisse machen die PfGSK-3 zu einem potentiellen Target auf der Suche nach Wirkstoffen gegen *P. falciparum*.

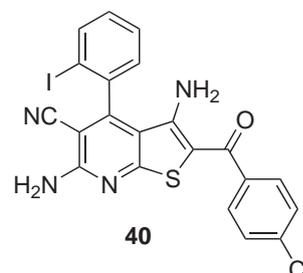
In den Laboratorien von L. Meijer (Station Biologique de Roscoff, Frankreich) konnten Ver-

Tabelle 1.1. | Bekannte Inhibitoren der GSK-3 β und ihre Aktivitäten an der PfGSK-3 nach [80], angegeben ist jeweils der IC₅₀ in μ M. Die Werte der GSK-3 β wurden am Enzym der Ratte vermessen.

	Verbindung	R ¹	R ²	PfGSK-3	GSK-3 β
	Purvalanol A (35a)			>100	13.0
	Staurosporin (25)			0.01	0.015
	Indirubin-3'-monoxim (36)			0.065	0.022
	Paullon (37)	H	H	18.00	0.62
	Kenpaullon (38)	Br	H	8.00	0.023
	Gwennpaullon (39)	Br	H ₂ N(CH ₂) ₃ O	1.5	0.04
	Alsterpaullon (28)	NO ₂	H	0.32	0.004
	Hymenialdisin (34)			0.023	0.01

bindungen mit Thieno[2,3-*b*]pyridin als Grundkörper bei einem Hochdurchsatzscreening als Inhibitoren der *Pf*GSK-3 identifiziert werden. Daraufhin wurde in einem zweiten Screening ein Datensatz von 427 unterschiedlich substituierten Thieno[2,3-*b*]pyridinen getestet. Dabei konnte die Verbindung **40** mit einem IC₅₀ von 0.43 μM als potentester Inhibitor der *Pf*GSK-3 identifiziert werden, der gleichzeitig Selektivität gegenüber der *Hs*GSK-3β zeigte [92].

Diese Verbindung war der Ausgangspunkt für die Dissertation von W. BRANDT [92], der die Synthese umfangreicher Variationen der Verbindung **40** gelang. Unter diesen waren ergänzend zu **40** weitere Verbindungen, die an der *Pf*GSK-3 Aktivität im niedrigen mikromolaren bzw. submikromolaren Bereich zeigten. Die Verbindungen waren teilweise selektiv an der *Pf*GSK-3 wirksam. Diese Ergebnisse gaben den Anstoß zur vorliegenden Arbeit, die in Zusammenarbeit mit dem Arbeitskreis von Conrad Kunick entstand. Sie sollte die präparativen Arbeiten von Wiebke Brandt mit Methoden des Molecular Modelling unterstützen bzw. weiterführen und auf diese Weise zum Verständnis der Wirkung von Thieno[2,3-*b*]pyridinen als Inhibitoren der *Pf*GSK-3 beitragen.



1.4. Zielsetzung

Am Anfang der Arzneistoffentwicklung steht in der Regel eine Leitstruktur, die z. B. ein Naturstoff sein kann, für den eine bestimmte physiologische Wirkung entdeckt wurde oder aber eine Hitstruktur, die im Rahmen eines virtuellen oder eines Hochdurchsatz-Screenings entdeckt wurde. Im Rahmen der darauf folgenden Leitstrukturoptimierung sind die in Abbildung 1.20

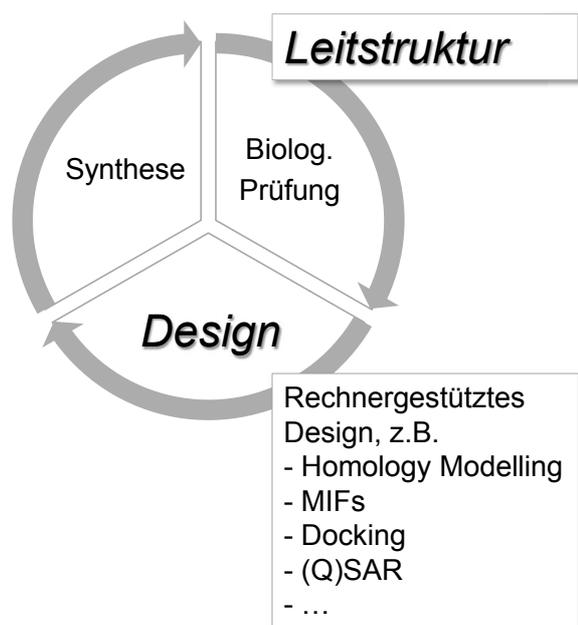


Abbildung 1.20. | Der zyklische Prozess der Leitstrukturoptimierung. Ausgehend von einer Leitstruktur werden üblicherweise mehrere Zyklen durchlaufen, in denen jeweils Veränderungen an der Substanz vorgeschlagen werden (Design), diese synthetisch umgesetzt werden (Synthese) und die biologische Testung eine Bewertung des Vorschlags ermöglicht. Das Design kann rechnergestützt verlaufen, in der Abbildung (in Anlehnung an [108]) sind die in dieser Arbeit dazu beschrittenen Wege beispielhaft genannt. Andere computergestützte Techniken wären z. B. pharmakophorbasiertes oder fragmentbasiertes Design.

vereinfachend dargestellten Phasen der Arzneistoffentwicklung wichtige Stationen eines iterativen Prozesses. Dieser wird von einer Leitstruktur ausgehend in der Regel vielfach durchlaufen, bis die Verbesserung der biologischen, physikochemischen und ggf. toxikologischen Eigenschaften des Kandidaten in die präklinische Entwicklung münden kann [108]. Die Anwendung computergestützter Methoden ist heute etablierter Bestandteil dieses Zyklus, sie steht methodisch im Zentrum dieser Arbeit.

Das allgemeine, übergeordnete Ziel war die Weiterentwicklung potenter und selektiver Inhibitoren der *PfGSK-3* durch eine Verbesserung des Verständnisses der vorliegenden chemischen und biologischen Daten [80, 92]. Die dazu notwendigen Schritte waren die im Folgenden einzeln aufgeführten Untersuchungen.

Homologie-Modelle der *PfGSK-3*

Da die dreidimensionale Struktur der *PfGSK-3* bislang experimentell nicht ermittelt wurde, stand am Anfang der Arbeit das Erstellen von Homologie-Modellen des Enzyms als Grundlage für strukturbasierte Untersuchungen der *PfGSK-3*.

Unterschiede in der ATP-Bindetaschen von *HsGSK-3 β* und *PfGSK-3*

Die Homologie-Modelle sollten eingehend untersucht werden, um bereits unabhängig von den zu entwickelnden Inhibitoren mögliche selektivitätsbestimmende Unterschiede zu dem humanen Enzym zu erkennen.

Bindungsmodus der Thieno[2,3-*b*]pyridine

Von zentraler Bedeutung für die Weiterentwicklung und die Erklärung von Struktur-Wirkungsbeziehungen der untersuchten Verbindungsklasse ist das Verständnis ihres Bindungsmodus. Dessen Untersuchung bildet daher den Mittelpunkt dieser Arbeit.

Ableitung von Synthesevorschlägen und deren Umsetzung

Im Sinne des in Abbildung 1.20 dargestellten Zyklus sollten aus den Ergebnissen der genannten Untersuchungen Synthesevorschläge abgeleitet werden. Deren Realisierung und die Ergebnisse der darauf folgenden biologischen Testungen sollten eine Hilfestellung bei der Bewertung der Hypothesen zum Bindungsmodus geben.

(Quantitative) Struktur-Wirkungsbeziehungen

Zur Interpretation der erarbeiteten potentiellen Bindungsmodi sollten QSAR-Berechnungen durchgeführt werden. Durch die Anwendung dieses ligandbasierten Verfahrens sollte einerseits die Möglichkeit genutzt werden, den Informationsgehalt des bereits vorliegenden Datensatzes zu nutzen. Andererseits sollte durch die Berücksichtigung der Ergebnisse aus den vorgenannten strukturbasierten Untersuchungen ein klares Bild über das Bindungsverhalten der Thieno[2,3-*b*]pyridine in der *PfGSK-3* erhalten werden.

2. Homologie-Modelle der *Pf*GSK-3 und deren Untersuchung mit molekularen Interaktionsfeldern

2.1. Homologie-Modelle als Voraussetzung für das strukturbasierte Design von *Pf*GSK-3-Inhibitoren

Die dreidimensionale Struktur der *Pf*GSK-3 ist bislang nicht aufgeklärt. Als Ausgangspunkt für einen strukturbasierten Ansatz in der Entwicklung von Inhibitoren der *Pf*GSK-3 steht deshalb die Berechnung von Homologie-Modellen des Proteins und deren Untersuchung im Mittelpunkt dieses ersten Kapitels. Es wird zunächst die Erstellung der Modelle und im Anschluss deren Untersuchung mit Molekularen Interaktionsfeldern (MIFs) beschrieben.

DROUCHEAU *et al.* [80] erwähnen ein Homologie-Modell, das über den vollautomatischen Webserver SWISS-MODEL erstellt wurde. Sie beschreiben das Modell als „nahezu identisch“ mit der *Hs*GSK-3 β . Insbesondere die hochkonservierte Region der Bindetasche sowie die substratbindende Region werden dabei hervorgehoben. Unterschiede hingegen werden hinsichtlich des Bereiches um Ser9 konstatiert – eine interessante Feststellung, da die Phosphorylierung dieser Aminosäure bei der *Hs*GSK-3 β eine Rolle bei der Autoinhibition spielt (vgl. S. 17 ff.). Das Modell beruht auf der Kristallstruktur eines humanen Apoenzyms (PDB-Eintrag 1H8F, [109]).

Untersuchungen zur Eignung von Holo- und Apoenzymen sowie Homologie-Modellen als Zielstrukturen von Dockings zeigten, dass Holoenzyme besser geeignet sind als Apoenzyme oder Homologie-Modelle, um hohe Enrichments im Rahmen von Virtual Screenings zu erreichen und Bindungsmodi zu reproduzieren [110]. Homologie-Modelle schnitten im Vergleich am schlechtesten ab, allerdings wurden diese teilweise auf Apoenzymen basierend konstruiert. Wenn die Sequenzen von Vorlage und Modell hohe Übereinstimmungen aufwiesen und ein Holoenzym als Vorlage gewählt wurde, erreichten Homologie-Modelle in der Studie gute Ergebnisse. Kürzlich untersuchten BORDOGNA *et al.* die Eignung von Homologie-Modellen für Dockings. Die Autoren kommen zu dem Schluss, dass es einen nachweislichen Zusammenhang zwischen der Qualität der Dockinglösungen und der der Homologie-Modelle gibt – und dass diese Qualität

durch die Kennzahlen üblicher Evaluierungswerkzeuge für Homologie-Modelle gut beschrieben wird [111].

Vor diesem Hintergrund wird die Auswahl von 18HF als Vorlage für Modelle der PfGSK-3 nicht als optimal eingeschätzt. Zudem ist durch umfangreiche Evaluierung einer Vielzahl erstellter Modelle und eine daraus resultierende Auswahl eine Verbesserung aller auf dem ausgewählten Homologie-Modell basierenden weiteren Untersuchungen zu erwarten. Darüber hinaus weist die HsGSK-3 β im Bereich des Glycine-rich Loops hohe Flexibilität auf, die bei der Konstruktion der Modelle berücksichtigt werden muss. Deshalb soll der erste Schritt im rationalen Design selektiver Inhibitoren der PfGSK-3 die Entwicklung von Homologie-Modellen sein, die sich folgendermaßen auszeichnen:

- Es werden ausgewählte Holoenzyme als Vorlagen herangezogen.
- Es wird eine Vielzahl an Modellen erzeugt. Diese werden einer umfangreichen Evaluierung unterzogen, die zur Auswahl qualitativ hochwertiger Homologie-Modelle führt, woraus sich ein Vorteil für das weitere strukturbasierte Design ergibt.
- Durch die Konstruktion basierend auf mehreren unterschiedlichen Vorlagen soll die Flexibilität des Enzyms abgebildet werden.

Bei der Konstruktion und Evaluierung der Modelle stellen sich Fragen, die über die Bedeutung für das Erstellen der Modelle hinaus im Hinblick auf die Entwicklung selektiver Inhibitoren von Interesse sind:

- Welche Unterschiede in den Sequenzen von HsGSK-3 β und PfGSK-3 gibt es und inwieweit lassen sich diese in der Entwicklung eines selektiven Inhibitors nutzen?
- Kann die Information von in der Vorlage gebundenen Liganden beim Erstellen der Modelle sinnvoll genutzt werden?

2.2. Untersuchungen mit molekularen Interaktionsfeldern

Molekulare Interaktionsfelder (MIFs) beschreiben die räumliche Variation der Interaktionsenergie zwischen einem Molekül und einer Sonde in dessen Umfeld [112]. P. GOODFORD entwickelte in den frühen 1980er Jahren das erste und bis heute meistgenutzte Programm für dieses Problem: GRID [113, 114]. MIFs können für Moleküle berechnet werden, deren dreidimensionale Gestalt entweder experimentell ermittelt wurde oder für die zum Zweck der Berechnungen ein entsprechender Vorschlag postuliert wird – z. B. ein Homologie-Modell. Die Felder können wichtige Anregungen an unterschiedlichen Stationen des Drug Designs geben, einen umfassenden Überblick gibt [115].

In der Literatur finden sich zahlreiche Beispiele für die Anwendung von GRID. In Kombination mit den Verfahren der Hauptkomponentenanalyse und der hierarchischen Hauptkomponentenanalyse wurde das Programm häufig verwendet, um die Problemstellung der Selektivität von Inhibitoren durch Betrachtung ihrer Zielstrukturen zu untersuchen, so z. B. [116–129]. Von besonderem Interesse im Blick auf die vorliegende Arbeit ist die letztgenannte Veröffentlichung von VULPETTI *et al.* [129]. Darin werden unter anderem mit GRID/CPCA die *HsGSK-3 β* und die CDK2 verglichen. Ziel ist das bessere Verständnis für die Selektivität bekannter Inhibitoren der CDK2 und deren weitere Verbesserung. Die Autoren können mit den angewandten Methoden den aus der Aminosäuresequenz bereits zu vermutenden Unterschied im Bereich des Gatekeepers genauer beschreiben. Während die *HsGSK-3 β* hier mit Leu132 eine vergleichsweise kleine Aminosäure hat, befindet sich an der entsprechenden Position der CDK2 ein Phenylalanin, das sterisch wesentlich anspruchsvoller ist. Im vorliegenden Projekt gibt es diesbezüglich eine Parallele, da der Gatekeeper von *HsGSK-3 β* und *PfGSK-3* (Met157) sich ebenfalls unterscheidet. Des Weiteren kommen VULPETTI *et al.* durch Unterschiede der MIFs zwischen den Proteinen im Bereich der zuckerbindenden Region und nahe dem Lys85 der *HsGSK-3 β* zu Synthesevorschlägen, die die Selektivität der vorgeschlagenen Inhibitoren an der CDK2 deutlich verbessern. Die Autoren finden zudem mithilfe der MIFs retrospektiv Erklärungen für die biologischen Daten anderer publizierter Verbindungen.

Diese Erfolge ließen die Methode hinsichtlich der Anwendung auf das Problem der Selektivität zwischen *HsGSK-3 β* und *PfGSK-3* aussichtsreich erscheinen.

2.3. Material und Methoden

2.3.1. Homology Modelling: Konstruktion eines virtuellen Proteins

Die Anzahl an experimentell bestimmten dreidimensionalen Proteinstrukturen nimmt stetig zu. So sind derzeit 68 651 solcher Strukturen über die RCSB PDB (www.pdb.org, [67]) öffentlich abrufbar. Dennoch klafft eine große Lücke zu den 530 264 Einträgen in der ebenfalls frei zugänglichen Sequenzdatenbank Swiss-Prot (www.expasy.org/sprot/, [130]). Durch das Erstellen von Proteinmodellen basierend auf der dreidimensionalen Struktur eines homologen Proteins und der Aminosäuresequenz der zu modellierenden Struktur werden strukturbasierte Methoden anwendbar, was die große Bedeutung des Homology Modelling im Drug Design erklärt [131].

2.3.2. Das Verfahren des Comparative Modelling

Unter Homology Modelling (auch Comparative Modelling) wird allgemein die Vorhersage dreidimensionaler Proteinstrukturen aufgrund der Übereinstimmung zu einem homologen Protein, dessen Struktur bereits experimentell aufgeklärt wurde, verstanden [132]. Das Vorgehen liegt

darin begründet, dass Proteine mit ähnlichen Sequenzen auch ähnliche dreidimensionale Strukturen aufweisen sollten, insbesondere hinsichtlich des Backbones, der für die Gestalt maßgeblich verantwortlich ist. Ausgehend von einer Aminosäuresequenz des zu modellierenden Proteins wird dabei prinzipiell in vier Schritten vorgegangen [133]:

1. Auswahl einer Vorlage

Die Auswahl der dreidimensionalen Vorlage (Template) ist eine wichtige Voraussetzung für ein sinnvolles Modell. Die Methode der Wahl zur Auffindung dieser Vorlagen ist BLAST: *Basic Local Alignment Search Tool* [134, 135]. BLAST vergleicht fragliche Sequenzen von Nukleotiden oder Proteinen mit in Datenbanken hinterlegten bekannten Sequenzen und berechnet statistische Übereinstimmungen auf der Basis lokaler Überlagerung.

Liegt eine Sequenzübereinstimmung von mehr als 50 % vor, so kann von einer hohen Qualität des daraus abgeleiteten Modells ausgegangen werden, liegt die Identität zwischen 20 und 50 %, wird die Ähnlichkeit zwischen Vorlage und Zielprotein als mittelmäßig eingeordnet, bei weniger als 20 % wird die Strukturvorhersage auf der Basis einer Ähnlichkeit problematisch [136].

2. Sequenzalignment

Das Sequenzalignment ist für die Berechnung eines Homologie-Modells von herausragender Bedeutung, da hier festgelegt wird, welche Passagen des zu konstruierenden Proteins an welche der vorhandene Muster angepasst werden sollen und welche Passagen *de novo* entworfen werden müssen.

Ist die Auswahl der Vorlage über BLAST erfolgt, so liegt hier bereits ein Sequenzalignment vor. Für das bei der Erstellung des Modells eingesetzte Sequenzalignment wird allerdings häufig der langsamere aber genauere Algorithmus nach NEEDLEMAN und WUNSCH verwendet [137].

3. Konstruktion von Modellen, Comparative Modelling

Wenn Template und Alignment vorliegen, besteht der eigentliche Prozess des Comparative Modellings darin, Ähnlichkeiten in der Sequenz in geometrische Ähnlichkeiten zu überführen und dabei die räumlichen Gegebenheiten des Modells zu optimieren [138].

4. Evaluierung

Bei der Evaluierung der berechneten Modelle werden im Wesentlichen zwei Fragen gestellt:

- a) Ist das Modell an sich als Protein sinnvoll, ist es also realistisch hinsichtlich sterischer (und damit häufig energetischer) Eigenschaften aufgebaut?
- b) Wie stark gleicht das Modell der Vorlage?

Dazu können die in dieser Arbeit verwendeten unter 2.3.2.1 dargestellten Programme eingesetzt werden. Sie überprüfen grundsätzlich die Beschaffenheit des Modells im Vergleich zu Daten, die aus experimentell bestimmten Proteinen abgeleitet wurden.

Eine zusätzliche Art der Evaluierung ist die Durchführung von Moleküldynamik- (MD-) Simulationen. Wenn die konstruierten Modelle sich dabei als stabil erweisen und energetische Verläufe innerhalb enger Grenzen liegen, spricht dies für die Plausibilität des Modells.

2.3.2.1. Verwendete Software

2.3.2.1.1. MODELLER [138, 139] ist ein komplexes Programm zum Homology Modelling. Das Prinzip von MODELLER bei der Erstellung von Homologie-Modellen liegt in der Berechnung und Optimierung einer *probability density function* (pdf), die die Erfüllung sterischer Gegebenheiten beschreibt. Diese werden durch die Vorlage definiert und sollen vom Modell möglichst gut nachempfunden werden. Dazu wird einerseits das Sequenzalignment herangezogen und andererseits bekannte Geometrien, die in Proteinen vorliegen, deren dreidimensionale Struktur experimentell aufgeklärt wurde (z. B. mögliche Konformationen für eine bestimmte Aminosäure oder häufige und demnach wahrscheinliche Konformationen eines äquivalenten Restes in einem homologen Protein).

MODELLER bezieht die C_{α} - C_{α} -Distanzen, N-O-Distanzen der Hauptkette und die Seitenketten-Torsionswinkel mit in die Berechnungen ein. Das automatisierte Verfahren optimiert die erhaltenen Modelle durch kurze Moleküldynamik-Berechnungen (simulated annealing) und eine abschließende Energieminimierung (conjugate gradient).

2.3.2.1.2. PROCHECK Basierend auf der Berechnung der Bindungswinkel und -längen bietet PROCHECK [140] eine umfassende räumliche Analyse von Proteinen. Die Ausgabe erfolgt sowohl in Form von Textdateien, die komplette Auflistungen der Reste und korrespondierender geometrischer Eigenschaften enthalten, als auch in Form von Grafiken, die aus den tabellari-schen Daten errechnete Übersichten darstellen. Ein zentrales Merkmal ist die Verteilung der Torsionswinkel Φ und Ψ in der Hauptkette, die grafisch im Ramachandran Plot gegeneinander aufgetragen werden [141]. Bestimmte Kombinationen der Winkel gelten dabei (jeweils engl.) als favoured (begünstigt, auch als Kernregion bezeichnet), allowed (möglich), generously allowed (grundsätzlich möglich) und disfavoured (nicht günstig).

Auch für die Seitenketten wird eine komplette Analyse der Torsionswinkel vorgenommen, außerdem werden Bindungslängen und zu dichte Kontakte nicht kovalent gebundener Atome ausgegeben. PROCHECK errechnet als zusammenfassende Größe den *G-score*, der eine Aussage über die stereochemische Ungewöhnlichkeit des Proteins macht und einen möglichst hohen Wert annehmen sollte.

2.3.2.1.3. ProSa Die Evaluierungssoftware ProSa wurde von SIPPL entwickelt, um Probleme bei der Interpretation experimenteller Daten zur Erstellung von Proteinstrukturen zu erkennen [142, 143]. Das Programm berechnet energetische Zustände für einzelne Aminosäuren einer Proteinstruktur, die als Ergebnis der Interpretation experimenteller Daten oder aber als theoretisches Modell entstanden sein können, um deren Qualität zu bewerten. ProSa weist dazu wissenschaftsbasiert jedem C_α eine Energie zu, die dann in einem für bestimmte Proteingrößen charakteristischen Gesamtwert (*z-score*, möglichst nahe an der Vorlage, bei Modellen grundsätzlich zu hoch) zusammengefasst oder als Profil über einen Sequenzbereich dargestellt werden kann. Auf diese Weise können Hinweise auf Probleme hinsichtlich der Faltung des Proteins in genau definierten Regionen – solchen mit (zu) hoher Energie – aus der grafischen Darstellung abgeleitet werden.

2.3.2.1.4. GROMACS [144–146] ist eine umfangreiches Programmpaket zur Durchführung von MD-Simulationen. Es ermöglicht also die Simulation von Wechselwirkungen vieler Teilchen eines Systems auf molekülmechanischer (kraftfeldbasierter) Basis durch die Lösung der Newtonschen Bewegungsgleichung und macht auf diese Weise thermodynamische Berechnungen sowie die Untersuchung von Bewegungen innerhalb eines Systems möglich.

GROMACS wurde insbesondere zur Simulation von Bio- (Makro-) Molekülen in wässriger Lösung oder in Membranen entwickelt und kann verschiedene Kraftfelder zur Berechnung heranziehen (z. B. GROMOS, OPLS oder AMBER). Neben einer Vielzahl an möglichen Varianten der Berechnung an sich sind zahlreiche Werkzeuge zur Analyse der MD-Simulation Teil des Paketes.

2.3.2.2. Prinzip des Rank-by-rank based consensus scorings (RCS)

Um eine möglichst vielschichtige Bewertung der Modelle zu erreichen, wurden diese mit den im vorangegangenen Abschnitt aufgeführten Analysewerkzeugen untersucht. Dabei kam es zu unterschiedlichen Beurteilungen der Modelle, weshalb die erhaltenen Werte zu einer einzigen übergreifenden Bewertung kombiniert werden sollten. Den Modellen wurde dazu nach N Kriterien ein Rang R zugewiesen. Eine Gesamtplatzierung im Sinne eines Rank-by-rank based consensus scorings (*RCS*) wurde nach Gleichung 2.1 berechnet.

$$RCS = \frac{\sum_{i=1}^N R_i}{N} \quad (2.1)$$

Die N Kriterien sind in Tabelle 2.1 aufgeführt. Die Parameter bad contacts (*bc*), *G-factor* und *Ramachandran* ergaben für viele Modelle gleiche Ergebnisse, so dass hier eine weitere Unterteilung vorgenommen wurde, wie sie aus der Tabelle hervorgeht. Wurden beispielsweise für mehrere Modelle die gleiche Anzahl an Aminosäuren in nicht erlaubten Regionen des Rama-

chandran Plots festgestellt, so wurden die auf diese Weise nicht unterscheidbaren Modelle nach ihrer möglichst großen Anzahl an Aminosäuren in der Kernregion sortiert. Entsprechend wurde mit den Kriterien *bc* und *G-factor* verfahren. Nach skriptgesteuerter Zusammenführung der Ergebnisse wurde der *RCS* innerhalb von KNIME [147] ermittelt.

Tabelle 2.1. | Einzelbeiträge zum *RCS*.

Kriterium <i>i</i>	Genauere Bewertung bzw. weitere Aufteilung
1. Ramachandran	1. wenige Aminosäuren in nicht erlaubten Regionen 2. viele Aminosäuren in Kernregion
2. G-factor	1. hoher G-factor (gesamt) 2. hoher G-factor (dihedral) 3. hoher G-factor (covalent)
3. bad contacts (<i>bc</i>)	1. wenige <i>bc</i> in der Bindetasche 2. wenige <i>bc</i> insgesamt
4. MODELLER	niedriger Wert der probability density function
5. ProSa	niedriger z-score

2.3.3. Vorgehen bei der Erstellung der Modelle

Bei der Erstellung der Modelle wurde in mehreren Schritten vorgegangen:

- Recherche bezüglich der Sequenz der *PfGSK-3* (\Rightarrow 2.3.4.1)
- Auswahl der Vorlagen (\Rightarrow 2.3.4.2; 11 Vorlagen identifiziert)
- Sequenzalignment (\Rightarrow 2.4.1.1; paarweise Sequenzalignments)
- Konstruktion der Modelle (\Rightarrow (2.3.2.1.1/2.4.1.2; für jede Vorlage 200 Modelle generiert, um den konformativen Raum der Seitenketten gut abzubilden)
- Validierung der Modelle (\Rightarrow 2.3.2.2/2.4.1.2 mittels *RCS*)
- MD-Simulation der besten Modelle zur Bestätigung von deren Qualität (\Rightarrow 2.4.1.4)

2.3.4. Experimentelle Grundlagen für das Homology Modelling

2.3.4.1. Die Sequenz der *PfGSK-3*

Die Aminosäuresequenz der *PfGSK-3* wird in der Literatur in zwei Varianten beschrieben:

1. Der Eintrag *Plasmodium falciparum 3D7* / *Gene ID: pfc0525c* der *Plasmodium falciparum* genome database PlasmoDB [148]. Dieser entspricht dem UniProt-Eintrag 077344 mit einer Länge von 440 Aminosäuren.

2. Der ältere PIR-PSD-Eintrag T18457 [149], der auch von DROUCHEAU *et al.* referenziert wird [80] und 452 Aminosäuren lang ist.

Der Unterschied der beiden Sequenzen liegt in den ersten Aminosäuren begründet. Ab Position 20 nach T18457 bzw. Position 8 nach 077344 sind sie identisch. Für die Sequenzalignments wurde die kürzere Version ausgewählt, da für den ersten Bereich der Sequenz keine Vorlage existiert und er somit bei der Erstellung der Homologie-Modelle nicht berücksichtigt werden kann (vgl. S. 47 ff.).

2.3.4.2. Die Kristallstrukturen der Vorlagen

Tabelle 2.2 gibt einen Überblick über die derzeit (Stand Januar 2011) in der PDB abgelegten Strukturen der *HsGSK-3 β* , die als Vorlage für Homologie-Modelle infrage kommen. Zusätzlich sind im Folgenden aufgeführte Strukturen in der PDB vorhanden, die aber aus den genannten Gründen als Vorlagen nicht infrage kommen.

- 106K, 2UW9, 106L, 2JDO, 2JDR, 2XH5 und 2X39 liegen im Komplex mit Protein Kinase B (PKB) vor. Für die *PfGSK-3* ist über ein Kokristallisat nichts bekannt, weshalb die genannten Strukturen nicht als repräsentativ für die dreidimensionale Gestalt der *PfGSK-3* angenommen werden können.
- 109U und 1GNG liegen im Komplex mit dem Polypeptid Axin vor. Es gilt somit das gleiche für sie wie für die eben genannten Strukturen.¹
- 2JLD, 3M1S und 3PUP sind mit einem Ruthenium-Komplex als Ligand kokristallisiert. Ein sehr spezieller kokristallisierter Ligand eines Holoenzym kann bei Dockings zu Problemen führen, da die Bindetasche zu sehr auf einen Einzelfall präformiert ist [111].
- 1I09 und 1H8F liegen als Apoenzyme vor (vgl. S. 27 bzw. ebenfalls [111]).

Als Vorlage für die Homologie-Modelle wurden mehrere Kristallstrukturen gewählt, da unterschiedliche Zustände in der Bindetasche repräsentiert werden sollten. Diese Herangehensweise wurde als sinnvoll im Hinblick auf verbesserte Ergebnisse bei nachfolgenden Dockings zur Vorhersage von Bindungsmodi beschrieben [160]. Mit unterschiedlichen Zuständen sind dabei nicht nur unterschiedliche Konformationen einzelner Aminosäuren gemeint, vielmehr nimmt der gesamte Glycine-rich Loop unterschiedliche Positionen ein. Diese spiegeln sich unter anderem im Torsionswinkel des Phe67 wider, der deshalb in der Tabelle mit aufgeführt ist. Für die Beschreibung der lösungsmittelzugewandten Region ist von Interesse, ob eine Salzbrücke zwischen Arg141 und Glu137 vorliegt. Es sind daher die Entfernungen zwischen den zentralen Kohlenstoffatomen der jeweiligen geladenen funktionellen Gruppe angegeben – bei Werten um 4 Å kann

¹Die starke Bindung an Axin ist gerade einer der experimentell ermittelbaren Unterschiede zwischen *HsGSK-3 β* und *PfGSK-3* – die *PfGSK-3* bindet Axin nicht [80].

Tabelle 2.2. | Kristallstrukturen der *HsGSK-3 β* . Im unteren Teil der Tabelle sind Strukturen aufgeführt, die erst nach Abschluss der Arbeiten an den Homologie-Modellen veröffentlicht wurden. Die Wertepaare bei dem Torsionswinkel des Phe67 und bei der Entfernung R141–E137 geben die Werte für die A- und B-Kette der Struktur an.

PDB-Eintrag	Ligand	Cofaktor	Ketten	fortl.	Auflösung	Ⓟ–Tyr216	Phe67 ^a	R141–E137	Literatur
1J1B	ANP	–	A, B	ja	1.80	nein	-65°/-45°	5.6/3.7	AOKI <i>et al.</i> [150]
1J1C	ADP	Mg ²⁺	A, B	ja	2.10	nein	-53°/-50°	5.5/3.5	ebd.
1PYX	ANP	Mg ²⁺	A, B	nein	2.40	nein	-144°/-140°	3.5/3.5	BERTRAND <i>et al.</i> [70]
1Q3D	STU	–	A, B	nein	2.20	nein	-59°/-82°	4.0/3.9	ebd.
1Q3W	ATU	–	A, B	nein	2.30	nein	-81°/-78°	4.1/4.0	ebd.
1Q41	IXM	–	A, B	nein	2.10	nein	-60°/-72°	4.0/4.0	ebd.
1Q4L	679	–	A, B	nein	2.77	nein	-66°/-45°	4.4/4.5	ebd.
1Q5K	TMU	–	A, B	nein	1.94	nein	-77°/-100°	4.2/4.1	BHAT <i>et al.</i> [151]
1R0E	DFN	Citrat	A, B	nein	2.25	nein	+46°/+46°	4.3/4.0 ^b	ALLARD <i>et al.</i> ^c
1UV5	BRW	Cl ⁻ , Co ²⁺ , PO ₄ ³⁻	A	ja ^d	2.80	nein	-66°/-	4.3/-	MEIJER <i>et al.</i> [152]
205K	HBM	–	A	ja ^d	3.20	nein	-122°/-	8.8/-	SHIN <i>et al.</i> [153]
20W3	BIM	–	A, B	nein/ja ^e	2.80	ja	+50°/+53°	4.3/4.7	ZHANG <i>et al.</i> [154]
kubra	DRG	PO ₄ ³⁻	A	ja	2.75	ja	-61°/-	3.3/-	pers. Mitteilung ^f
3DU8	553	–	A, B	nein	2.20	nein	+59.6°/+55°	4.2/4.0	MENICHINCHERI <i>et al.</i> [155]
3F7Z	34O	–	A, B	nein	2.40	ja	+40°/+44°	4.3/5.1	SAITOH <i>et al.</i> [156]
3F88	3HT	2HT	A, B	nein	2.60	ja	+47°/+47°	4.2/4.1	ebd.
3GB2	G3B	–	A	nein	2.40	ja	+56°/-	4.0/-	SAITOH <i>et al.</i> [157]
3I4B	Z48	–	A, B	(ja)	2.30	nein	-63°/-59°	3.8/3.8	ARONOV <i>et al.</i> [158]
3L1S	Z92	PO ₄ ³⁻	A, B	nein	2.90	nein	+59°/+60°	3.9/3.9	ARNOST <i>et al.</i> [159]

^aTorsionswinkel des Phe67, bei negativen Werten ist der Phenylrest aus der Tasche hinausgerichtet, bei positiven ragt er hinein; ^bIn 1R0E liegt für beide Ketten eine alternative Position der Seitenkette des Arg141 vor, in der keine Salzbrücke vorliegt (Distanzen 5.5 und 6.1 Å); ^cUnveröffentlichter Artikel mit dem Titel *From genetics to therapeutics: the Wnt pathway and osteoporosis*;

^dEs sind zwar in der Originaldatei fehlende Aminosäuren angegeben, diese befinden sich jedoch am Anfang bzw. Ende der Sequenz; ^eKette B ist fortlaufend, in Kette A fehlen Aminosäuren;

^fpersönliche Mitteilung durch A. Oberholzer, Universität Bern und L. Meijer, Station Biologique, Roscoff; fortl.: fortlaufend; Auflösung in Å; Ⓟ–Tyr216: Liegt das Tyr216 als Phosphotyrosin vor?

eine solche Interaktion angenommen werden,² in den wenigeren Fällen mit einer Distanz von mehr als 4.5 Å nicht.

Von den zur Verfügung stehenden ersten 13 Proteinen der Tabelle 2.2³ wurden bis auf 1J1C und 1PYX alle Strukturen für die weiteren Berechnungen berücksichtigt. 1J1C wurde aufgrund der Ähnlichkeit zu 1J1B ausgeschlossen. Da 1J1B höher aufgelöst ist und zudem kein kokristallisiertes Magnesium in der Bindetasche liegt, fiel die Wahl auf 1J1B. Damit lag bereits eine Struktur mit kokristallisiertem ATP-Analogen vor, so dass 1PYX mit der niedrigeren Auflösung und ebenfalls mit dem im weiteren Ablauf schlecht zu handhabenden Magnesium nicht als Vorlage herangezogen wurde.

In der Mehrzahl der Fälle liegen die Proteine als Dimer jeweils zwei Ketten vor, von denen eine ausgewählt wurde. Die beiden Ketten unterscheiden sich in der Regel wenig. Ausnahmen sind 1Q3D, 1Q41, 1Q5K und 20W3, bei denen die B-Kette im Gegensatz zur A-Kette weniger fehlende Aminosäuren aufweist bzw. fortlaufend aufgelöst ist. Deshalb wurde als Vorlage in diesen Fällen die B-Kette ausgewählt.

Ebenso wurde von 1J1B die B-Kette gewählt. In dieser ist im Gegensatz zur A-Kette die Seitenkette des Arg141 in Richtung des Glu137 orientiert, wodurch die schon erwähnte Salzbrücke möglich wird. Diese Interaktion liegt in der Mehrzahl der Kristallstrukturen vor, weshalb auch bei 1J1B die B-Kette verwendet wurde.

2.3.5. Berechnung molekularer Interaktionsfelder mit GRID

Zur Berechnung der im einleitenden Teil 2.2 erwähnten MIFs wurde das Programm GRID verwendet. Das Prinzip von GRID besteht in der Berechnung von Interaktionsenergien zwischen einer Sonde mit definierten physikochemischen Eigenschaften und einem Molekül (in diesem Zusammenhang im Englischen als Target bezeichnet⁴), das im vorliegenden Fall ein Protein ist, grundsätzlich aber auch ein Ligand sein kann, wodurch viele Anwendungsgebiete von GRID im ligandbasierten Drug Design möglich werden.

Durch Verschieben der Sonde entlang eines imaginären Gitters entsteht ein Datenfeld von Interaktionsenergien, das grafisch dargestellt werden kann und dann Positionen bevorzugter Interaktionen zwischen Sonde und Zielmolekül anzeigt. Die Datenmatrix kann andererseits zur statistischen Analyse herangezogen werden (siehe Abschnitt 2.3.6.2).

²Eine Entfernung von 4 Å für die Zentralatome der delokalisierten Ladung ist für Salzbrücken innerhalb von Proteinen eine gängige Definition [161], während bei Ligand-Protein-Interaktionen in Analogie zur Wasserstoffbrückenbindung die Entfernung zwischen den Heteroatomen gemessen wird, die dann mit 2.5-3.2 Å kürzer ist [162].

³Die als *kubra* bezeichnete Struktur stand erst im Spätsommer 2008 zur Verfügung, als die Arbeiten zum Homology Modelling bereits vorläufig abgeschlossen und die Veröffentlichung [163] eingereicht war – sie ist deshalb nicht Teil des Artikels. Wegen ihrer großen Bedeutung für das Projekt wurde sie nachträglich bearbeitet und in den Arbeitsablauf des *RCS* einbezogen.

⁴Wegen der Überschneidung mit der Begrifflichkeit eines Targets als Angriffspunkt eines Wirkstoffes werden im Folgenden die Begriffe Sonde und Zielmolekül anstelle von *probe* und *target* verwendet.

Dabei unterscheidet sich GRID gegenüber anderen Methoden zur Berechnung von MIFs insbesondere in den folgenden Punkten:

- Die Sonden sind oft anisometrisch, also räumlich gerichtet.
- Das Zielmolekül reagiert auf die Sonde. Zur Berechnung der Interaktionsenergien wird nicht nur die Sonde gedreht. Auch für Reste des Zielmoleküls, die zur optimalen Ausbildung von Wasserstoffbrückenbindungen variiert werden müssen, wird nach der optimalen Ausrichtung gesucht.
- Für beide Strukturen wird angenommen, dass sie in Wasser gelöst vorliegen. Diese Voraussetzung wird in die Berechnung der Interaktionsenergien einbezogen, indem im freien Lösungsmittel eine Dielektrizitätskonstante von 80 angenommen wird, die bis in das Innere eines Makromoleküls hin auf 4 abnimmt.
- Die Vielfalt und die sorgfältige Parametrisierung der unterschiedlichen Sonden erlauben eine sehr feine Anpassung an das vorliegende Problem. So existieren z. B. verschiedene Wasserstoffbrückendonor- und -akzeptorsonden, die sich jeweils in ihrer Partiaalladung, Polarisierbarkeit, Anzahl der effektiv zugänglichen Elektronen, etc. unterscheiden.

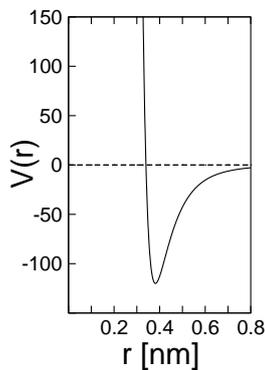


Abbildung 2.1. | Der charakteristische Verlauf des Lennard-Jones-Potentials am Beispiel des Argon (Parameter für Ar nach [164]).

Die mit GRID berechneten Interaktionsfelder beziehen sich auf *nichtkovalente* Wechselwirkungsenergien. Diese werden durch eine empirische Energiefunktion beschrieben, die sich für jeden Gitterpunkt aus der Summe mehrerer Komponenten zusammensetzt, die jeweils für alle (in definierter Reichweite liegenden) wechselwirkenden Atome berechnet werden. Mit den Beiträgen ΣE_{lj} für das Lennard-Jones-Potential, ΣE_{el} für den elektrostatischen bzw. ΣE_{hb} für den über Wasserstoffbrückenbindungen vermittelten Anteil und $\Sigma E_{entropy}$ für den hydrophoben Effekt ergibt sich die Interaktionsenergie E_{xyz} an einem Punkt des Gitters mit den Koordinaten x , y und z nach der in GRID verwendeten zusammengesetzten empirische Energiefunktion gemäß Gleichung 2.2, deren Summanden im Folgenden kurz vorgestellt werden.

$$E_{xyz} = \Sigma E_{lj} + \Sigma E_{el} + \Sigma E_{hb} + E_{entropy} \quad (2.2)$$

Lennard-Jones-Potential: Das Lennard-Jones-Potential [165] beschreibt ganz allgemein widerstrebende anziehende (insbesondere Van-der-Waals- und Dipol-Dipol-Wechselwirkungen) und abstoßende Kräfte (Pauli-Repulsion durch Abstoßung zwischen Elektronen mit gleichem Spin) zwischen zwei Teilchen. Der prinzipielle Verlauf ist in Abbildung 2.1 dargestellt. Die

Wechselwirkung ungeladener Teilchen führt in Abhängigkeit ihres Abstandes r zueinander zu einem Potential V , das sich nach Gleichung 2.3 ergibt.

$$V(r) = \frac{A}{r^{12}} - \frac{B}{r^6} \quad (2.3)$$

Darin ist der erste Teil der repulsive, der mit kleiner werdendem Abstand d schnell steigt, während der zweite Term den anziehenden Anteil beschreibt. A und B werden aus dem Van der Waals-Radius rad , der Polarisierbarkeit α_i und der effektiven Anzahl der Elektronen N_{eff} berechnet. Diese Werte sind in Tabelle 2.3 auf Seite 45 für die in dieser Untersuchung berechneten Sonden aufgelistet. Bei Abständen über 8 Å liegen die Energien für E_{lj} typischerweise unterhalb 0.01 kcal/mol und werden deshalb nicht mehr berechnet.

Elektrostatische Funktion: Die ladungsbedingten Interaktionspotentiale nehmen bei weitem nicht so schnell mit der Entfernung ab wie das Lennard-Jones-Potential, weshalb hier kein fester Grenzwert existiert, über den hinaus keine Energien ermittelt werden. Die Berechnung erfolgt prinzipiell nach Gleichung 2.4, gemäß dem Coulombschen Gesetz.

$$E_{el} = \frac{q_p q_t}{4\pi\epsilon r_{pt}} \quad (2.4)$$

E_{el} ist damit proportional zur Ladung der Sonde q_p und der des Zielmoleküls q_t . Umgekehrte Proportionalität besteht zur Entfernung der Teilchen voneinander r_{pt} sowie zur Permittivität des Mediums ϵ (auch Dielektizitätskonstante). ϵ kommt dabei eine kritische Rolle zu, es ergibt sich als Produkt aus der Permittivität des Vakuums ϵ_0 und der relativen Permittivität des entsprechenden Mediums ϵ_r . GRID berücksichtigt bei der Berechnung von E_{el} nach Gleichung 2.5 für q_p und q_t deren Tiefe s_p und s_t in den jeweiligen Medien mit den relativen Permittivitäten ϵ_s im Lösungsmittel und ϵ_t in der Zielstruktur, in K werden geometrische und Naturkonstanten zusammengefasst [112, 113, 166].

$$E_{el} = \frac{q_p q_t}{K \epsilon_t} \left[\frac{1}{r_{pt}} + \frac{(\epsilon_t - \epsilon_s)/(\epsilon_t + \epsilon_s)}{\sqrt{r_{pt}^2 + 4s_p s_t}} \right] \quad (2.5)$$

Wasserstoffbrücken: Die Berechnung des Potentials für Wasserstoffbrückenbindungen in GRID orientiert sich an BROOKS *et al.* [166]. Dabei wird eine Gleichung vom Typ 2.3 angepasst, indem die Größen A und B Atomtypen der beteiligten Atome berücksichtigen und zudem ein Faktor einbezogen wird, über den der Winkel berücksichtigt wird, den Donor (bzw. Akzeptor), Wasserstoff und Sonde zueinander bilden.

Entropische Effekte: Durch das Binden eines Liganden können fixierte Wassermoleküle in das Lösungsmittel freigesetzt werden oder zuvor flexible Anteile der an der Bindung beteilig-

ten Moleküle können in ihrer Position festgesetzt werden – beides hat Auswirkungen auf die Entropie des Systems. Diese kann in GRID durch die Berechnung eines „hydrophoben Effektes“ berücksichtigt werden. Er wird als inverse Interaktionsenergie eines Wassermoleküls berechnet, da ein solches keine Kontakte mehr zum Zielmolekül ausbilden kann, wenn die hydrophobe Sonde DRY bindet [167].

2.3.6. Die Hauptkomponentenanalyse zur Auswertung von MIFs

2.3.6.1. Allgemeines zum Verfahren

Die Hauptkomponentenanalyse (engl.: Principal Component Analysis, PCA) wurde als Verfahren von PEARSON entwickelt, der bereits 1901 die grundlegende Idee beschrieb, eine Punkteschar im kartesischen Koordinatensystem durch approximierte Linien bzw. Flächen darzustellen [168]. Die PCA etablierte sich im vergangenen Jahrhundert als wichtige Anwendung der multivariaten Statistik zur Strukturierung großer Datensätze, da sie das Überführen einer hohen Anzahl von Variablen in wenige aussagekräftige latente Variablen – die namensgebenden Hauptkomponenten – ermöglicht. Sie kann einerseits rein explorativ zur Extraktion wesentlicher Information und deren übersichtlicher Darstellung dienen. Andererseits können durch die berechneten Modelle Aussagen über wahrscheinliche Ausreißer und nicht relevante Daten gemacht werden, die als störendes Rauschen aus dem Datensatz entfernt werden können. Ferner besteht die Möglichkeit, nicht im Datensatz enthaltene Beobachtungen auf ein Modell anzuwenden und auf diese Weise die Zugehörigkeit der besagten externen Beobachtungen zu bestimmten Gruppen des Modelldatensatzes zu prognostizieren.

Eine kurze Einführung in die Methode, zugrunde liegende Literatur und die Anwendung insbesondere in der Chemie geben WOLD *et al.* [169], weitere verwandte multivariate Methoden und deren Anwendungen samt Literatur finden sich beispielsweise bei ERIKSSON *et al.* [170].

Ausgangspunkt für die PCA ist eine Datenmatrix X , die aus n Reihen (Experimente, Substanzen, Individuen... \Rightarrow *Objekte*) und den korrespondierenden k Spalten (Beobachtungen, Messungen, individuelle Werte... \Rightarrow *Variablen*) besteht und die eine bestimmte Gesamtvarianz aufweist. Es ergibt sich eine Anzahl von $n * k$ Werten, die als Punkte im k -dimensionalen Raum verstanden werden können. Im einfachsten und anschaulichen Fall ist dies eine Punkteschar im dreidimensionalen kartesischen Koordinatensystem, deren Dimensionalität durch Projektion auf eine Fläche (grundlegende Idee von PEARSON, s.o.) reduziert wird. Dadurch wird ein möglichst hoher Anteil der Varianz innerhalb der Matrix durch wenige aussagekräftige Variablen erklärt.

Die Matrix X kann nun in das Produkt zweier Matrices T und P plus der verbleibenden Residuenmatrix E zerlegt werden. Das Vorgehen ist schematisch in Abbildung 2.2 dargestellt. Die niedrigdimensionalen Matrices T (Scores) und P (Loadings) beschreiben die zugrunde liegende komplexe Matrix X . Ein grundlegendes Problem kann dadurch entstehen, dass die Datensätze stark unterschiedliche Varianzen und Mittelwerte haben, was im Ergebnis zur Überbewertung

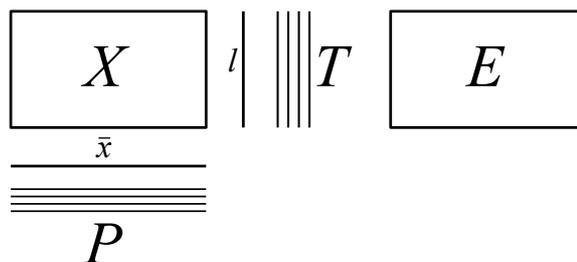


Abbildung 2.2. | Schematische Darstellung einer PCA nach [170].

der betreffenden Variablen im Modell führt. Dies kann eine Skalierung⁵ notwendig machen, außerdem wird der Mittelwert zur besseren Interpretierbarkeit häufig subtrahiert, um die Daten auf einen gemeinsamen Ursprung bezogen darzustellen.

Formal lässt sich der Vorgang demnach für beliebig viele Dimensionen folgendermaßen darstellen:

$$X = l * \bar{x} + T * P + E \quad (2.6)$$

Darin wird die Matrix X durch drei verschiedene Terme beschrieben:

- $l * \bar{x}$ beschreibt die Mittelwerte der Datenpunkte und resultiert aus der ggf. notwendigen Vorbereitung der Daten für die Berechnung.
- $T * P$ modelliert durch das Produkt zweier Vektoren die Matrix X .
- E ist eine Matrix der verbleibenden Residuen und bildet damit das Rauschen ab.

2.3.6.2. Die PCA in Kombination mit MIFs

Die Anwendung der Hauptkomponentenanalyse auf Datensätze, die durch Berechnung von MIFs mit GRID entstanden, wurde zunächst von CRUCIANI *et al.* [171] durchgeführt und systematisch von PASTOR *et al.* [172] beschrieben. Er entwickelte gemeinsam mit M. BARONI in der Folge GOLPE (Generating Optimal Linear PLS Estimations) [173, 174], ein Programm zur Weiterverarbeitung der insbesondere von GRID berechneten Ergebnisse aber auch anderer Datenmatrices mit multivariaten Methoden.

Das Verfahren stellt somit eine Alternative dar zur ebenfalls sehr verbreiteten CoMFA- [175] bzw. CoMSIA- [176] Analyse zur Berechnung von dreidimensionalen Struktur-Wirkungsbeziehungen (3D-QSAR, vgl. Kapitel 4). CoMFA und CoMSIA werden in der Regel auf kleine Moleküle angewandt, um deren Aktivität als Ligand an einem bestimmten Protein zu untersuchen (ligandbasiertes Vorgehen).

Um bei der Suche nach selektiven Inhibitoren von bisher bestehenden Verbindungen unabhängig zu werden, kann von den Zielstrukturen anstelle der Inhibitoren ausgegangen wer-

⁵In der Regel wird jede Variable mit dem Kehrwert ihrer Standardabweichung multipliziert, wodurch alle Variablen auf den Umfang ihrer Varianz normiert werden. Dieses Vorgehen wird als Unitvariable Scaling bezeichnet.

den (strukturbasiertes Vorgehen). An diese Zielstrukturen (Enzyme, Rezeptoren, Ionenkanäle, Transporter) müssen alle Liganden binden und die Darstellung und der Vergleich möglicher Interaktionen mit MIFs ist der an dieser Stelle verfolgte Ansatz zur Erklärung von Selektivität. Im Rahmen der Methode werden in der Regel die folgenden Schritte vorgenommen: Am Beginn steht die Auswahl geeigneter dreidimensionaler Strukturen der Proteine, die überlagert werden. Es folgt die Berechnung der MIFs mit unterschiedlichen Sonden für die fragliche Region (in der Regel für die Bindetasche). Die Daten werden dann für eine PCA aufbereitet und mithilfe dieser statistischen Methode strukturiert und dargestellt.

Die Aufbereitung der Daten aus den MIFs für die PCA ist in Abbildung 2.3 grafisch nach PASTOR *et al.* [172] und FOX [177] dargestellt. Zunächst werden MIFs in einem definierten Bereich mit den Koordinaten $X_1/Y_1/Z_1$ bis $X_m/Y_n/Z_o$ berechnet. Die Matrix wird für s Sonden und zwei unterschiedliche Zielstrukturen berechnet. Die Berechnungen resultieren in zwei linearen Vektoren, die jeweils in s Vektoren zerschnitten wird. Werden diese Vektoren beider Proteine untereinander geschrieben, ist das Ergebnis eine X -Matrix, in der jede der $m * n * o = k$ Spalten die Interaktionen der Zielstrukturen an einer definierten $x/y/z$ -Koordinate beschreibt. Diese zweidimensionale Matrix liegt der nach Abschnitt 2.3.6.1 zu berechnenden PCA zugrunde.

Wie im vorangehenden Abschnitt erläutert, wird nun im Rahmen der PCA eine Aufteilung der k -dimensionalen Matrix in eine Score-Matrix T , eine Loadings-Matrix P und einen Fehlermatrix E vorgenommen (vgl. Abb. 2.2). Die Betrachtung der ersten Komponenten der Matrices

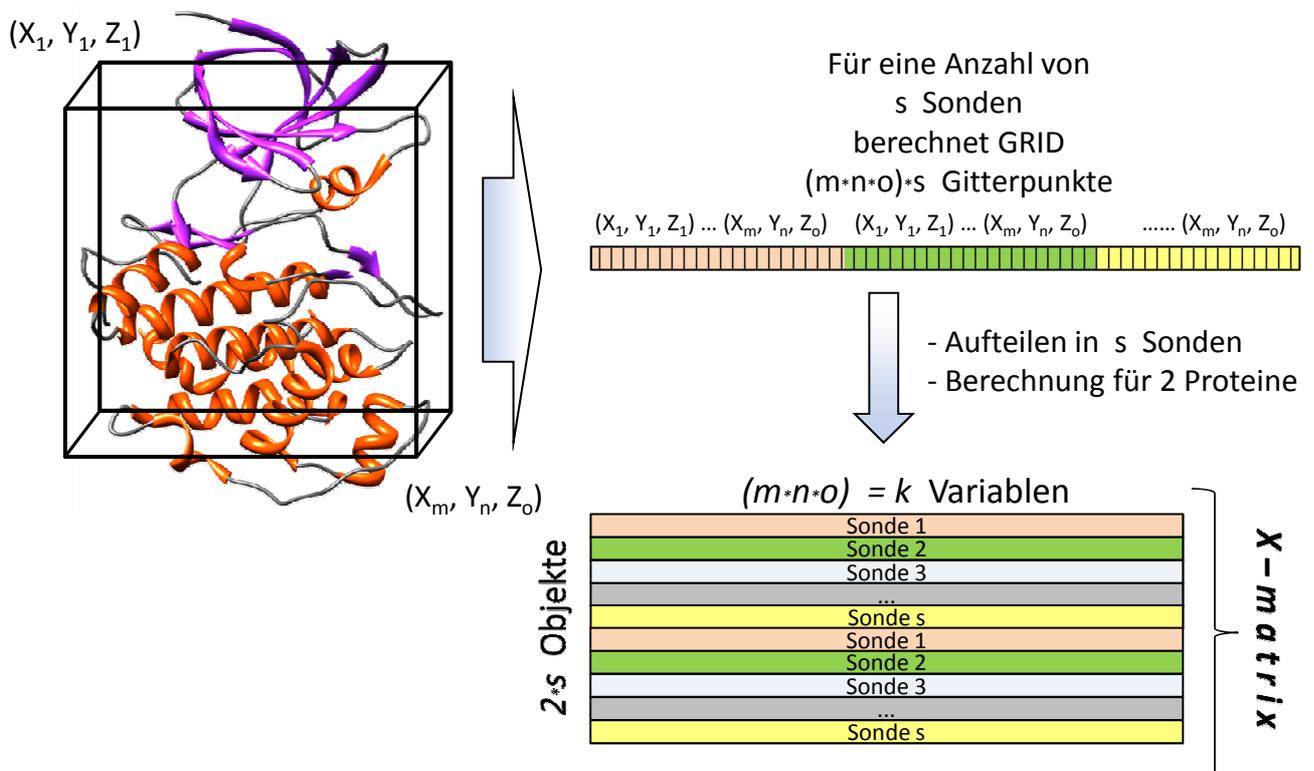


Abbildung 2.3. | Aufbau einer X -Matrix aus GRID-MIFs zur PCA nach [172, 177].

T und P erlaubt die Erklärung eines hohen Anteils der Varianz durch wenige Variablen. Die grafische Darstellung dieser Komponenten kann eine übersichtliche Darstellung der komplexen Matrix ermöglichen. Die Objekte sind in diesem Kontext die $2 * s$ Kombinationen aus Sonde und Zielmolekül. Score Plots sind demnach die Auftragungen von Komponenten aus T gegeneinander und erlauben einen Überblick über Zusammenhänge zwischen den Objekten (z. B. das Verhältnis der Interaktionsfelder der beiden Proteine mit einer Sonde s). Die korrespondierenden Loading Plots fassen in Analogie dazu die Zusammenhänge zwischen den k Variablen zusammen (Komponenten von P). Da jeder Punkt im Diagramm eine Koordinate wiedergibt, ist hier zur Interpretation insbesondere von Interesse, welche Variablen weit vom Ursprung entfernt liegen und somit einen großen Anteil zur Erklärung der Varianz des Datensatzes beitragen. Die Graphen stehen in direkter Beziehung zueinander: Die im Loading Plot abgebildeten Variablen, die in einer bestimmten Region des Plots liegen, sind diejenigen, die zur Position der entsprechenden Objekte im Score Plot führen.

Mit der beschriebenen Methodik verbinden sich einige Probleme: Zunächst werden Sonden mit schwacher Interaktionsenergie tendenziell vernachlässigt, obwohl sie wichtige Bestandteile der Rezeptor-Ligand-Wechselwirkung sein können. Das betrifft insbesondere lipophile Sonden, für die keine Wasserstoffbrücken berechnet werden, wodurch der Betrag der berechneten Interaktionsenergie in der Regel deutlich geringer ist als derjenige von Wasserstoffbrückenbindungspartnern. In der durchgeführten Analyse sind dies die C3- und die DRY-Sonde. Die erstgenannte repräsentiert einen Methylrest, die letztgenannte beschreibt hydrophobe Energien (Details vgl. Tab. 2.3).

Der zweite Nachteil besteht darin, dass zwar ein Modell berechnet wird, das eine Aussage darüber macht, ob bestimmte Koordinaten für die Selektivität interessant sind oder nicht – die Aussage aufgrund welcher Sonde dies der Fall ist, ist jedoch nicht möglich, da die Variablen jeweils die Information verschiedener Sonden enthalten. Ein dritter und entscheidender Nachteil besteht darin, dass die Untersuchung von mehr als zwei Zielmolekülen in einer Analyse unübersichtlich bis unmöglich ist, da sich die Anzahl der Objekte abhängig von der Anzahl p der Zielstrukturen bei s Sonden auf $s * p$ vervielfältigt, wodurch das Erkennen von Anhäufungen erschwert wird.

2.3.6.3. PCA vs. CPCA

Um die Nachteile der Kombination MIFs/PCA auszuräumen, wendeten KASTENHOLZ *et al.* das Prinzip der Consensus PCA (CPCA oder hierarchische Hauptkomponentenanalyse) [118] auf die MIFs an.⁶ Dabei wird die den Berechnungen zugrunde liegende Matrix etwas anders organisiert: Die Aufteilung der s Sonden zu einzelnen Objekten (Reihen in der X -Matrix) entfällt. Stattdessen können mehrere Zielmoleküle verwendet werden, deren aus der GRID-Rechnung resultierende lineare Vektoren dann untereinander eingereiht werden, so dass die

⁶Das Verfahren an sich wurde bereits 1987 von WOLD *et al.* beschrieben [178].

Anzahl der Objekte gleich der Anzahl p an untersuchten Molekülen ist (vgl. Abb. 2.4). Die Sonden bilden Blöcke innerhalb eines Objektes. Dadurch wird bereits der erste Nachteil der PCA aufgehoben: Im Score Plot können übersichtlich die Gruppierungen verschiedener Zielstrukturen auseinandergelassen werden, sofern sich diese hinsichtlich ihrer dreidimensionalen Struktur und der resultierenden MIFs klar voneinander abgrenzen lassen.

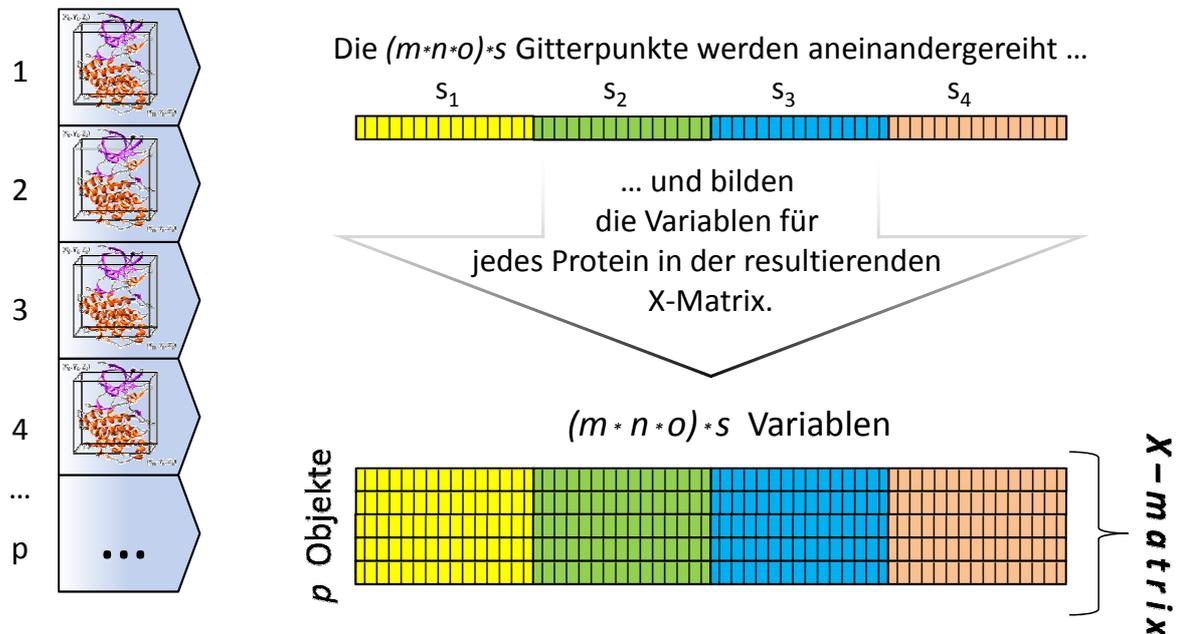


Abbildung 2.4. | Aufbau einer X -Matrix aus GRID-MIFs zur CPCA nach [118]. Im Unterschied zur Aufbereitung für die PCA nach Abb. 2.3 werden die Zeilenvektoren der Proteine nicht zerteilt. Dadurch repräsentieren unterschiedliche Blöcke die eingesetzten Sonden und eine Betrachtung von Koordinate und isolierter Sonde ist möglich. Außerdem bleibt die Anzahl der Objekte überschaubar, auch wenn mehr als $p = 2$ Proteine in die Analyse eingehen.

Ein weiterer wichtiger Vorteil ergibt sich durch die Umorganisation der Matrix: Da jede Variable nun nur noch die Information *einer* Sonde wiedergibt, kann gruppenweise eine Skalierung vorgenommen werden, so dass einzelne Sonden (insbesondere lipophile Interaktionen) nicht mehr benachteiligt werden. Dies geschieht in einer Block Unscaled Weights (BUW) genannten Prozedur, in der die Varianzen innerhalb eines Blockes unverändert bleiben, die Blöcke zueinander jedoch gewichtet werden.

Durch die Aufteilung in Blöcke ergeben sich zwei Ebenen der Analyse. Im Ergebnis entstehen die analogen Matrices T , P und E wie bei einer Analyse, die durch Kombination aller Blöcke durchgeführt wird. Allerdings wird zusätzlich bei der CPCA Information über die Bedeutung einzelner Blöcke gewonnen – im Falle der Analyse von MIFs also Information über die Rolle einzelner Sonden.

2.3.7. MIFs zum Vergleich von HsGSK-3 β und PfGSK-3

Als Zielmoleküle für die Berechnung der MIFs wurden einerseits alle 11 Kristallstrukturen der HsGSK-3 β herangezogen, die auch als Vorlage für Homologie-Modelle der PfGSK-3 ausgewählt wurden (vgl. Tab. 2.2), andererseits die Homologie-Modelle, die sich für jede Vorlage im Rahmen des RCS als bestes Modell erwiesen hatten. Nach dreidimensionaler Überlagerung der Vorlagen (vgl. exp. Teil S. 75) erfolgte die Auswahl der Sonden.

Im ersten Schritt der visuellen Analyse (Abschnitt 2.4.2.1 ab S. 57) sollte einerseits ein übergreifender Eindruck möglicher Interaktionen mit den Proteinen gewährleistet werden. Andererseits sollten mögliche Interaktionen der Zielstrukturen mit solchen Sonden untersucht werden, die Eigenschaften von Verbindungen repräsentieren, für die biologische Daten vorliegen. Entsprechende funktionelle Gruppen (z. B. 2- und 5-Cyano- oder 6-Amino-Substitution am Thienopyridin der Verbindungen der Klasse 1, vgl. S. 83) wurden mit bei den zu berechnenden Sonden aufgenommen. Auf diese Weise fiel die Wahl auf die ersten 10 der in Tabelle 2.3 aufgeführten Sonden. Sie repräsentieren hydrophile bzw. hydrophobe Sonden (insbesondere OH2 bzw. DRY, I, C3), Wasserstoffbrückenakzeptoren bzw. -donoren (OC1, N:#, O bzw. N2, N1+) sowie eine Multiatomsonde, die als Wasserstoffbrückendonator und -akzeptor fungieren kann (AR.CONH2). Weiter Einzelheiten zu den Parametern der einzelnen Sonden finden sich in der Tabelle. Zudem werden die Eigenschaften der Sonden insbesondere im Bezug auf die Anwendung in dieser Untersuchung den korrespondierenden Abschnitten ab Seite 57 jeweils vorangestellt.

Die visuelle Analyse bestand darin, zunächst vergleichend die Felder von humaner und plasmodialer GSK-3 zu betrachten. Dazu wurden die Vertreter pfgsk_1j1b als plasmodiales Enzym und 1J1B als entsprechende Vorlage ausgewählt. Es zeigten sich einige auffällige Unterschiede und in der Folge wurde überprüft, inwieweit diese sich in den anderen Homologie-Modell/Vorlage-Paaren wiederfinden (in der Regel nicht grafisch dargestellt). Zusätzlich wurde ein Vergleich der Minima pro Fläche angestellt. Diese Information wurde mit BIOCUBE [179] aus den Ergebnisdateien von GRID extrahiert und für alle Homologie-Modell/Vorlage-Paare gemeinsam betrachtet. Wenn es in Bereichen, in denen für einen Einzelfall Unterschiede im Profil zwischen HsGSK-3 β und PfGSK-3 festgestellt wurden, zudem eine entsprechende Anhäufung von Minima pro Fläche gab, wurde dies als Bestätigung eines Unterschiedes gewertet. Die Ergebnisse dieser Untersuchungen stellen den ersten Teil der Analyse mit MIFs dar und werden ab Seite 57 beschrieben.

Der zweite Schritt war die statistische Analyse der Daten mittels (C)PCA. In Entsprechung zum ersten Schritt wurde zunächst paarweise eine PCA für Vorlage und Modell berechnet (Abschnitt 2.4.2.3 ab S. 66) und dann eine CPCA zur Analyse der Varianz innerhalb der Gesamtheit aller 22 Proteine durchgeführt (Abschnitt 2.4.2.4 ab S. 71).

Für die paarweise PCA wurde die Multiatomsonde AR.CONH2 ausgespart, da die Ergebnisse der visuellen Analyse nicht zur Differenzierung der Proteine beitragen konnte. Stattdessen wurde die OH-Sonde eingeführt, die ebenfalls Donor- und Akzeptoreigenschaften annehmen kann.

Tabelle 2.3. | Parameter der ausgewählten Sonden. Die Beschreibungen der aufgelisteten Sonden geben jeweils den originalen GRID-Text wieder. Parameter zur Berechnung des Lennard-Jones-Potentials sind Van-der-Waals-Radius (rad in Å), effektive Anzahl der zugänglichen Elektronen (N_{eff}) und die Polarisierbarkeit (α_i in Å³). Die Ladung ist als Vielfaches der Elementarladung e angegeben. Neben der Anzahl der Wasserstoffbrücken (HB), die als Donor (HB_{don}) bzw. Akzeptor (HB_{acc}) ausgebildet werden können, sind zusätzlich der H-bond energy parameter ($HB E_{min}$) in kcal/mol und der H-bond distance parameter ($HB r_{min}$) in Å aufgelistet. Für die DRY-Sonde werden entsprechend der OH2-Sonde Wasserstoffbrücken berechnet, diese gehen dann allerdings als positiver Wert in die Berechnung der Interaktionsenergie ein.

Sonde	Beschreibung	Lennard-Jones				Wasserstoffbrücken			
		rad	N_{eff}	α_i	Ladung	HB_{don}	HB_{acc}	$HB E_{min}$	$HB r_{min}$
OH2	water	1.70	7	1.2	0	2	2	-4.5	1.40
DRY	the hydrophobic probe	1.65	7	1.2	0	2	2	-4.5	1.40
I	organic iodine atom	2.15	10	6.31	0	<i>Sonde bildet keine HB aus</i>			
C3	methyl CH3 group	1.95	8	2.17	0	<i>Sonde bildet keine HB aus</i>			
OC1	aro/ali ether oxygen	1.60	6	1	-0.15	0	1	-2.0	1.50
N:#	sp N with lone pair	1.65	6	2.9	0	0	1	-5.0	1.55
O	sp2 carbonyl oxygen	1.60	6	0.84	-0.25	0	2	-5.0	1.40
N2	neutral flat NH2 eg amide	1.70	8	1.7	0	2	0	-5.3	1.50
N1+	sp3 amine NH cation	1.65	7	1.40	0.66	1	0	-3.0	1.60
AR.CONH2	aromatic amide	1.74	6	1.46	0	2	2	-3.25	1.47
OH	phenol or carboxy OH	1.65	7	1.50	-0.15	1	1	-4.50	1.40
N1	neutral flat amide	1.65	7	1.40	-0.08	1	0	-4.00	1.60

Weiterhin wurde auf den Einsatz der I-Sonde verzichtet, da in der visuellen Analyse die Felder große Ähnlichkeit mit der C3-Sonde aufwiesen. Letztere wurde ausgewählt, um die in der visuellen Analyse beobachteten Differenzen zwischen humanem und plasmodialen Enzym in die paarweisen PCA zu repräsentieren. Als zehnte Sonde wurde die N1-Sonde mit aufgenommen. Sie wurde auch in der Untersuchung von NAUMANN *et al.* im Rahmen einer CPCA eingesetzt, mit deren Hilfe den Autoren die Charakterisierung verschiedener Kinasen und deren Einordnung in sogenannte *Target Family Landscapes* gelang [120].

Für die CPCA wurden dann insgesamt 9 Sonden ausgewählt. Es sollten insbesondere Sonden in die Analyse einbezogen werden, die bei der visuellen Analyse Unterschiede erkennen ließen (I, OC1, N2, N1+). Für die Berechnung wurden zusätzlich Sonden gewählt, die auch in der paarweisen PCA vergleichsweise deutliche Unterschiede aufgezeigt hatten, wenn auch in der Regel nur in der zweiten Hauptkomponente (OH, N:#, OH2, O). Aus der paarweisen Analyse per PCA ging hervor, dass die DRY-Sonde keinen Beitrag zur Differenzierung zwischen *HsGSK-3 β* und *PfGSK-3* mit MIFs leisten konnte. Dies trifft weniger deutlich auf die C3-Sonde ebenso zu, wobei in entsprechenden Score Plots immerhin eine Unterscheidung zwischen dem humanen und dem plasmodialen Enzym möglich war. Die C3-Sonde wurde folglich zu den in der CPCA eingesetzten Sonden aufgenommen, um einen klassischen Vertreter für lipophile Interaktionen mit in die Analyse zu integrieren.

In den Abbildungen des Ergebnisteils sind die *HsGSK-3 β* (1J1B) und deren MIFs grün dargestellt, die *PfGSK-3* (pfgsk_1j1b) und deren MIFs magenta. Die Bezeichnung der Aminosäuren erfolgt grundsätzlich in der Form <Aminosäure_*PfGSK-3*/Aminosäure_*HsGSK-3 β* >, da in der Regel die Überlagerung beider Enzyme gezeigt wird.

2.4. Ergebnisse

2.4.1. Homologie-Modelle

Bei der BLAST-Suche nach geeigneten Vorlagen wurden alle in Abschnitt 2.3.4.2 genannten Strukturen der *HsGSK-3 β* als mögliche Proteine mit kaum unterscheidbaren Kennzahlen gefunden. Kristallstrukturen der CDK2 und der GSK-3 anderer Spezies erzielten auch gute Übereinstimmungen hinsichtlich der Sequenzhomologie. Da die Werte der verschiedenen *HsGSK-3 β* -Strukturen jedoch deutlich besser ausfielen, wurden diese für das Homology Modelling ausgewählt. Aus den in Abschnitt 2.3.4.2 erläuterten Gründen kamen dabei einige Strukturen als Vorlage nicht infrage, so dass die dort genannten 11 Strukturen ausgewählt wurden (vgl. oberer Teil von Tab. 2.2 auf S. 35). Es wird zunächst das resultierende Sequenzalignment genauer betrachtet (vgl. Abb. 2.5).

```

ruler      58                               85                               105
pfc0525c  SEDEDEERTIDNEINRSPNKSYPKLNIGNGSFGVVYEAICIDTSEQVAIKKVLQDPQYKNRELMIMKLNHNHINIY
P49841    SKVTTVVATPGQGPDRPQEVSYTDTKVIGNGSFGVVYQAKLCDSGELVAIKKVLQDKRFKNRELQIMRKL DHCNIIVR
* :      *  :  : *  :  **  : :***** : *  : *  *****  : :***** ** : : *  *  * :

ruler      135                               157                               166                               206
pfc0525c  LKDYYYTESFKKNEKNIFLNVM EYIPQTVHKYMKYYSRNNQALPMFLVKLYSYQLCRALSYIHSKFICHRDLKIPQN
P49841    LRYFFYSSGEKKDE--VYLNVL DYPVETVYRVARHYSRAKQTL PVIYVKLYMYQLFRSLAYIHSFGICHRIKIPQN
* : : : : . .  * : *  : : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * :

ruler      212                               224                               250
pfc0525c  LLIDPRHTLKL CDFGS AKNLLAGQRSVSYICSRFYRAPELMLGSTNYTTHIDLWLSLGCIIAEMILGYPIFSGQSSV
P49841    LLDPDTAVLKL CDFGS AKQLVRGEPNVSYICSRYYRAPELIFGATDYTSSIDVWSAGCVLAELLGQPIFPDGSV
* * : * * *  *  * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * :

ruler      289                               300                               350
pfc0525c  DQLVRIIQVLGTPTEDQLKEMNP NYADIKFPDVKSKDLR KVFVKGT PDEAINLITQFLKYEPLKRLNPIEALADPFF
P49841    DQLVEIIKVLGTPTREQIREMNP NYTEFKFPQIKAH PWTKVFRPRT PPEAIALCSRLLEYTP TARLTPLEACAHSFF
* * * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * :

ruler      366                               400
pfc0525c  DELRDPCKL PKYIDKLP ELFNFCKEEIQ EMSMECRRKI I PKNVYEEF
P49841    DELRDPN VKLPNGRD -TPALFNFTT QELSSNPPLAT -ILIPPHARIQA
* * * * : * * : * * * * * : * : * : * : * : * : * : * : * : * : * : * : * :

```

Abbildung 2.5. | Sequenzalignment von *Pf*GSK-3 (pfc0525c, oben) und *Hs*GSK-3 β (P49841, unten). Die gelb markierten Aminosäuren bilden die ATP-Bindetasche. Fett markiert ist der Gatekeeper Met157/Leu132, die Ähnlichkeit der Aminosäuren ist abnehmend mit * > : > . > \square gekennzeichnet.

2.4.1.1. Alignment

Die Übereinstimmung zwischen der kompletten Sequenz der *Hs*GSK-3 β (Swiss-Prot Eintrag P49841, 420 Aminosäuren) und der *Pf*GSK-3 (PlasmoDB Eintrag pfc0525c, 440 Aminosäuren) beträgt 44.3%, die Ähnlichkeit⁷ 60.5%. Da die Kristallstrukturen der Vorlagen nicht den gesamten Sequenzbereich beschreiben, wurden Modelle nur für den Abschnitt Ser58 bis Phe413 der *Pf*GSK-3 (356 Aminosäuren, entsprechend Ser35 bis Ala388 der *Hs*GSK-3 β) erstellt. In diesem Bereich liegt eine Identität von 53.1% vor, eine Ähnlichkeit von 69.7%. In Abbildung 2.5 ist das Sequenzalignment dargestellt, Aminosäuren der Bindetasche⁸ sind markiert. Von diesen 35 Aminosäuren unterscheiden sich nur 6, es liegt damit eine Identität von 82.9% vor.⁹ In den Kristallstrukturen der Vorlagen sind unterschiedliche Aminosäuren nicht aufgelöst. Das komplette Alignment, in dem diese Lücken abgebildet sind, ist im Anhang auf Seite 227 ff. beschrieben. Aus dem Alignment ist ersichtlich, dass sich die beiden Enzyme in der ATP-Bindetasche in 6 Aminosäuren unterscheiden, die im oberen Teil der Tabelle 2.4 aufgeführt

⁷An dieser Stelle wird die evolutionäre Ähnlichkeit beschrieben, definiert als positiver Wert innerhalb der verwendeten BLOSUM62-Matrix. Sie ist in der Veröffentlichung [163] für den gesamten Sequenzbereich mit 61.4% angegeben. Dieser Wert konnte bei der Durchsicht der Ergebnisse für die Dissertation nicht reproduziert werden, es werden von EMBOSS mit den im experimentellen Teil aufgeführten Einstellungen die im Fließtext angegebenen 60.5% ermittelt.

⁸Die Definition der Bindetasche wurde auf die *Pf*GSK-3 bezogen vorgenommen. Es wurde dazu das ATP-Analogon ANP aus 1J1B in das korrespondierende Modell *pf*gsk_1j1b_128 projiziert und alle Aminosäuren im Umkreis von 6.5 Å als zur Bindetasche gehörig definiert.

⁹In der Veröffentlichung [163] ist die Identität in der Bindetasche fälschlicherweise mit 85.7% angegeben.

sind. Dabei ist insbesondere der unterschiedliche Gatekeeper von Interesse. Durch den Wechsel von Leu132 (human) zu Met157 (plasmodial) wird die Zugänglichkeit zur Back Pocket verändert. Als zweiter deutlicher Unterschied ist das in der PfGSK-3 durch Gln162 ersetzte Glu137 der HsGSK-3 β zu nennen. Wie aus Tabelle 2.2 hervorgeht, bildet dieses Glutamat in vielen Kristallstrukturen der HsGSK-3 β eine Salzbrücke zum Arg141 aus, die so in der PfGSK-3 nicht möglich ist.

Betrachtet man auch weiter entfernte Aminosäuren (bis zu 10 Å), dann gibt es weitere Unterschiede (Tabelle 2.4, unterer Teil). Dabei ist der Wechsel von Leu81 zu Gln104 und Tyr140 zu His165 insofern interessant, als dass jeweils polare Seitenketten entstehen. Diese erlauben in der PfGSK-3 Wasserstoffbrücken zu potentiellen Liganden, die in der HsGSK-3 β nicht möglich wären. Durch die Veränderung von Gln72 zu Glu95 kommt ein negativ geladenes Zentrum dazu, durch His165 statt Tyr140 ein positiv geladenes, was die elektrostatischen Oberflächeneigenschaften des Proteins verändert.

Die Verhältnisse in der Bindetasche sind schematisch in Abbildung 2.6 dargestellt. Insgesamt stellen sich anhand der Sequenz bereits Unterschiede dar, die erste Anhaltspunkte bei der Suche nach Erklärungen für die Selektivität von Inhibitoren der PfGSK-3 bieten.

2.4.1.2. Modelle

Der in Abschnitt 2.3.3 dargestellte Ablauf des Homology Modelling führte zu 4400 Modellen, für die ein RCS berechnet wurde. Dabei wurde einerseits das beste Modell unter allen ermittelt, andererseits wurde für jeden der 22 Abläufe (11 Vorlagen, jeweils mit und ohne Ligand) ein Topmodell bestimmt. Die Ergebnisse der unterschiedlichen Evaluierungsprogramme, die zur Berechnung des RCS herangezogen wurden, sind für die jeweiligen besten Modelle in Tabelle B.2 im Anhang auf Seite 232 aufgeführt.

Das beste Modell unter allen 4400 ist `pfgsk_1j1b_128`.¹⁰ Der Ramachandran Plot für dieses Modell, derjenige der zugrunde liegenden Vorlage 1J1B und derjenige des besten Modells, das basierend auf 1J1B unter Berücksichtigung des Liganden generiert wurde, sind in Abbildung 2.7 dargestellt. Nur die beiden Aminosäuren Cys243 und Phe246 haben in `pfgsk_1j1b_128` ungünstige Kombinationen von Φ - und Ψ -Winkel. Bei beiden liegt die Kombination in der Vorlage begründet. Das Lys623 war in der Vorlage 1J1B an der Grenze zur ungünstigen Region positioniert und weist im Modell eine bessere Winkelkombination auf. Arg205 und Tyr378 sind in `pfgsk_1j1b_128` nur im generell möglichen Bereich eingestuft, während ihre Entsprechungen in der humanen Vorlage im möglichen Bereich liegen – allerdings so knapp an der Grenze zum möglichen Bereich, dass dies nicht als Mangel gewertet wird. Insgesamt bestätigt der Ramachandran Plot also die gute stereochemische Qualität von `pfgsk_1j1b_128`. Hinsichtlich des in Abbildung 2.7(c) dargestellten Ramachandran Plots zeigt auch `pfgsk_1j1b_ANP_27` keine Auffälligkeiten. Wie aus Tabelle B.2 im Anhang auf Seite 232 zu ersehen ist, führen jedoch die

¹⁰Die Bezeichnung der Modelle erfolgt nach dem Muster `pfgsk_<Vorlage>[_lig]_<laufende Nummer #>`.

Tabelle 2.4. | Sich unterscheidende Aminosäuren im Bereich der ATP-Bindetasche von *HsGSK-3 β* und *PfGSK-3*. Im hochkonservierten Bereich 6.5 Å um den Liganden sind nur zwei Aminosäuren mit abweichenden Eigenschaften vorhanden (+), die anderen sind als ähnlich (-) einzuordnen. Erweitert man den Bereich auf 10 Å, fallen die im unteren Teil der Tabelle aufgeführten Reste auf. Dabei ist zu bedenken, dass sie vergleichsweise weit vom Liganden entfernt sind. In den eingeklammerten Fällen zeigen die Aminosäuren mit den Seitenketten vom Liganden weg, aus der Bindetasche hinaus.

<i>HsGSK-3β</i>	<i>PfGSK-3</i>	Unterschiedlichkeit
Val110	Ile133	[-] beide lipophil
Leu132	Met157	+ -S-CH ₃ , Gatekeeper
Asp133	Glu158	- homologe Aminosäure
Val135	Ile160	[-] beide lipophil
Glu137	Gln162	+ Verlust saurer Funktion
Arg141	Lys166	- beide basisch
Lys60	Asn83	+ Verlust basischer Funktion
Val61	Ile84	- beide lipophil
Gln72	Glu95	+ neue saure Funktion
Leu81	Gln104	+ neue ungeladene polare Seitenkette
Arg111	Tyr134	[+] Wechsel basisch zu aromatisch
Leu130	Val155	- beide lipophil
Tyr140	His165	+ neue polare geladene Aminosäure
Val142	Tyr167	[+] Wechsel Alkyl zu Aryl
Arg144	Lys169	- beide basisch

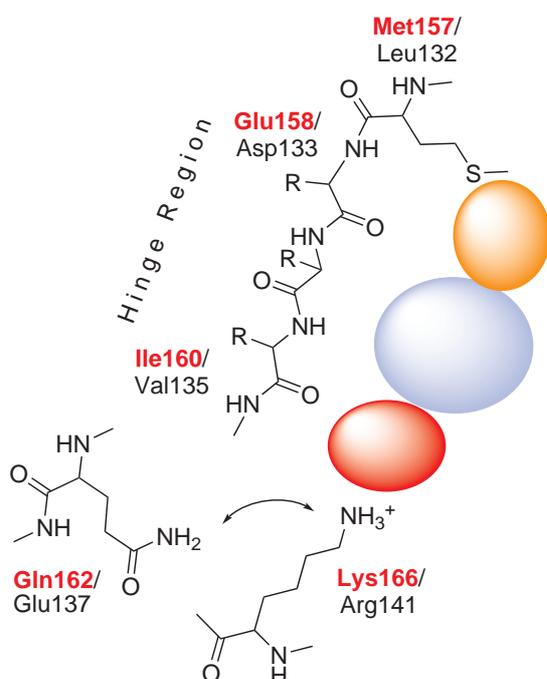


Abbildung 2.6. | Sich unterscheidende Aminosäuren innerhalb der ATP-Bindetasche. Zur Orientierung ist die Farbkodierung der Bindetaschenbereiche an der Hinge Region nach TRAXLER [58] in Entsprechung zu den Abbildungen auf Seite 14 und 20 angegeben. Die rot bezeichneten Aminosäuren sind in der *PfGSK-3* vertreten und unterscheiden sich von denjenigen der *HsGSK-3 β* . Besonders interessant ist der unterschiedliche Gatekeeper (Met157/Leu132) und die Abweichungen im Bereich der lösungsmittelzugewandten Region bzw. der Front Pocket. Das längere und flexiblere Methionin der *PfGSK-3* verändert die Größe und Zugänglichkeit der Back Pocket. Durch den Wechsel von Glu137 (human) zu Gln162 (plasmodial) kann eine Salzbrücke nicht ausgebildet werden, das Lys166 ist freier beweglich. In der zucker- und in der phosphatbindenden Region liegen in der Sequenz keine Unterschiede vor. Im Glycine-rich Loop gibt es den Unterschied Ile133/Val110, der wegen der physikochemischen Ähnlichkeit der Aminosäuren nicht in der Abbildung dargestellt ist.

vergleichsweise schlechten Ergebnisse in weiteren Evaluierungsparametern (insbesondere die hohe Anzahl an zu dichten Kontakten und der schlechte MODELLER pdf) zur schlechteren Bewertung im Rahmen des *RCS*.

Das Modell `pfgsk_1j1b_128` beruht auf der Vorlage mit kokristallisiertem Substratanalogen ANP, bei der Entstehung wurde der Ligand jedoch nicht berücksichtigt. Die entsprechenden Modelle `pfgks_1j1b_ANP_<#>` und überhaupt alle Modelle, bei deren Erstellung die Information des Liganden berücksichtigt wird, werden im Rahmen des *RCS* deutlich schlechter bewertet als die übrigen. Dieser Zusammenhang wird in Abbildung 2.8 sehr deutlich. Hier ist außerdem der wichtige Einfluss der Vorlage auf die Bewertung des Modells zu erkennen: Während z. B. auf 205K basierende Modelle im *RCS* insgesamt schlecht abschneiden, erreichen auf 1Q41 und 1J1B basierende die besten Platzierungen.

Die Qualität der Modelle sollte nicht nur untereinander vergleichend betrachtet werden, sinnvoll ist insbesondere auch der Bezug auf die zugrunde liegenden Vorlagen. Bei den mit PROCHECK ermittelten Parametern bezüglich *bad contacts* und *Ramachandran* schneiden die Modelle besser ab als die Vorlagen – ein nicht ungewöhnlicher Befund, da es sich bei den Vorlagen um experimentell ermittelte Strukturen handelt, während die Modelle zum Abschluss ihrer Konstruktion einer Optimierung unterzogen werden. Hinsichtlich des *G-factor* (PROCHECK) und *z-score* (ProSa) ergeben die Vorlagen etwas bessere Werte, die Modelle jedoch immer noch unauffällige, so dass das Ergebnis des Abgleiches mit den Vorlagen als unauffällig gelten kann. Die genauen Einzelwerte für die Kristallstrukturen sind im Anhang in Tabelle B.1 auf Seite 231 aufgeführt.

Tabelle 2.5. | Rangfolge der Vorlagen gemäß *RCS*.

PDB	<i>RCS</i>
1Q41	2.75
1J1B	3.50
1R0E	4.00
1Q4L	5.25
20W3	5.25
1Q5K	5.75
1Q3D	6.00
1Q3W	6.50
kubra	7.75
205K	9.25
1UV5	10.00

Werden die Vorlagen vergleichend mit dem *RCS* bewertet,¹¹ so zeigen sich Entsprechungen zu Abbildung 2.8. Die resultierende Rangfolge der Vorlagen ist Tabelle 2.5 zu entnehmen. Auch unter den Vorlagen belegen 1Q41 und 1J1B die ersten Plätze, während 205K und 1UV5 am Ende stehen. Während für die Vorlagen allerdings immer nur eine einzige Struktur bewertet wird, entstehen durch die 400 Homologie-Modelle je Vorlage Modelle von sehr unterschiedlicher Qualität. Selbst wenn man nur die besseren Homologie-Modelle ohne Berücksichtigung der Liganden betrachtet, so ist das schlechteste auf 1J1B basierende Modell etwa so gut bewertet wie das beste auf 205K basierende. Das Generieren der großen Anzahl ermöglicht demnach die Auswahl eines guten Modells hinsichtlich der Kriterien, die in den *RCS* einbezogen werden.

Der Verlauf der ProSa-Energiepotentiale für Topmodelle und korrespondierende Vorlagen ist in Abbildung 2.9 dargestellt. Er zeigt zum einen die Ähnlichkeit der Verläufe für Modelle und

¹¹Die durchschnittliche Platzierung wurde ohne den MODELLER-Score mit den anderen 4 in Tabelle 2.1 genannten Kriterien durchgeführt.

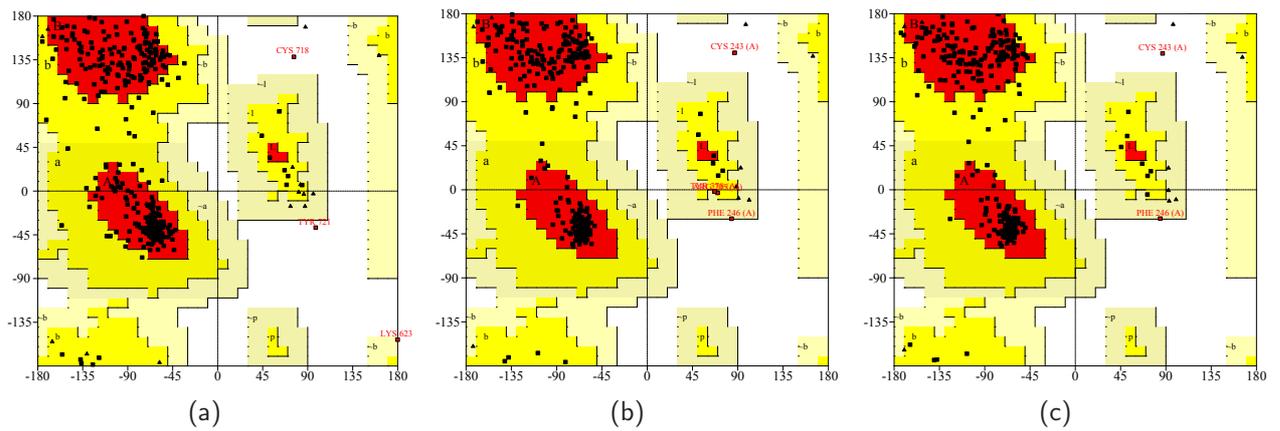


Abbildung 2.7. | Ramachandran Plots der Vorlage 1J1B (a) und der resultierenden Modelle pfgsk_1j1b_128 (b) bzw. pfgsk_1j1b_ANP_27 (c). Farbkodierung der Bereiche: rot – günstig (favoured); gelb – erlaubt (allowed); hellgrün – generell möglich (generously allowed); weiß – disfavored (ungünstig).

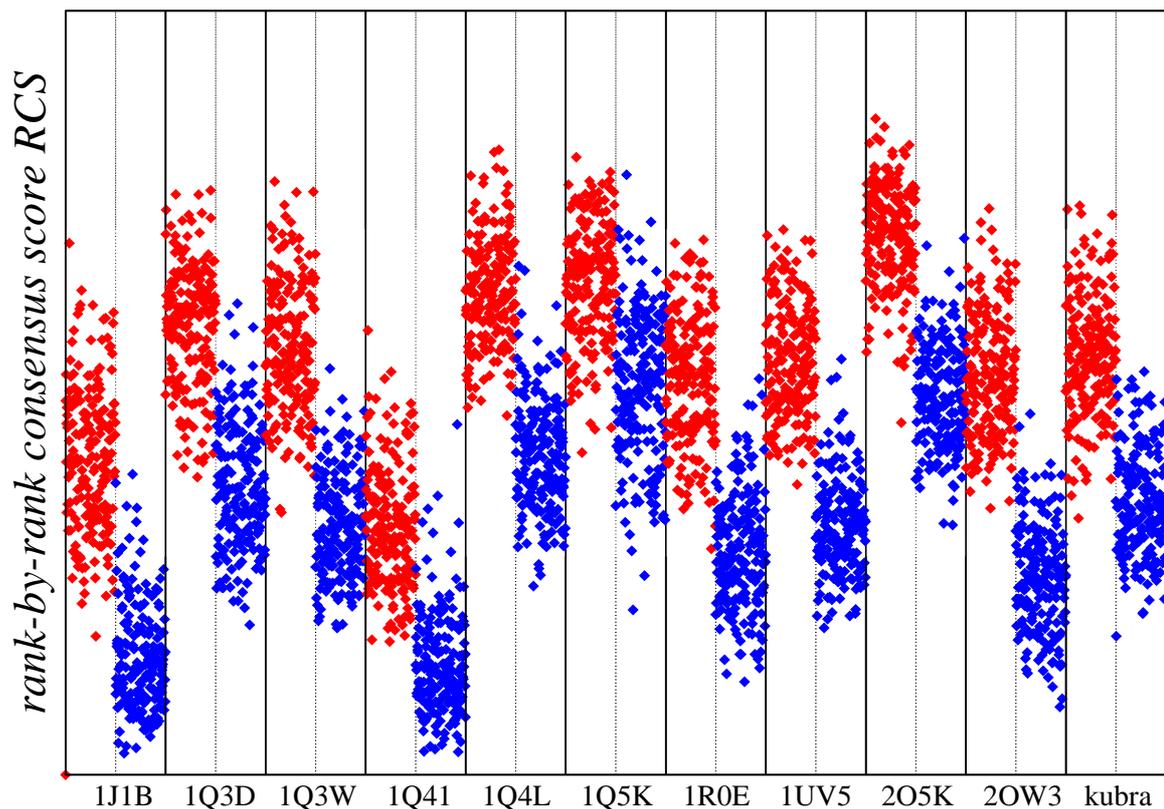


Abbildung 2.8. | Vergleich des *RCS* aller Modelle. Als rote Punkte dargestellt sind Modelle, die unter Berücksichtigung der Ligandinformation entstanden sind, die blau dargestellten wurden ohne diese zusätzliche Bedingung (vgl. exp. Teil S. 74 ff.) erstellt. Aus der Berechnung des *RCS* nach Gleichung 2.1 ergibt sich, dass kleine Werte günstig sind (einen vorderen Platz in der durchschnittlichen Bewertung bedeuten), während hohe Zahlen eine hintere Platzierung für das Modell anzeigen.

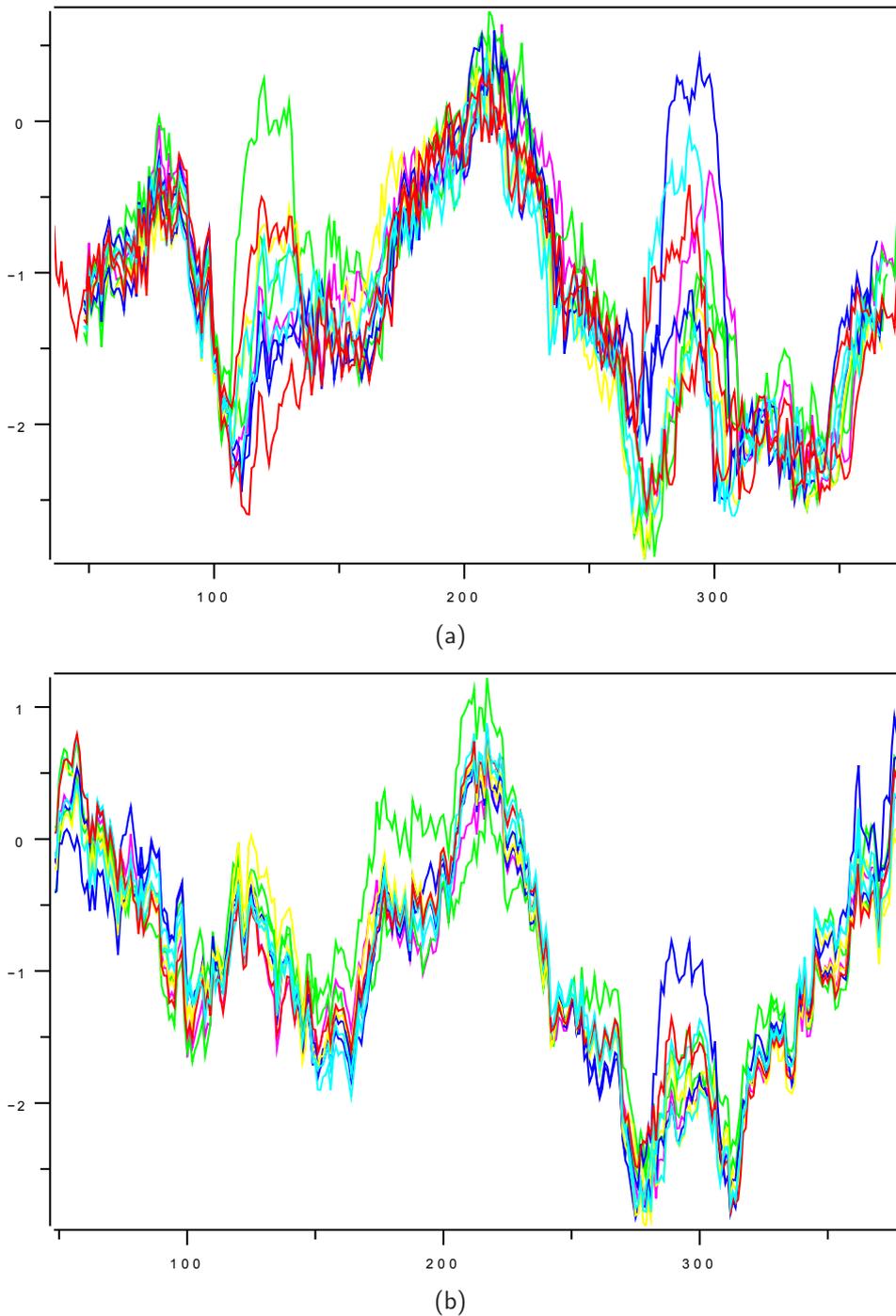


Abbildung 2.9. | ProSa-Energieverläufe von Vorlagen (a) und Modellen (b). Bei den Vorlagen fällt die grüne Kurve von 20W3 im Bereich der Aminosäuren 115-135 auf, das entsprechende Modell zeigt nicht diesen höherenergetischen Bereich. Ansonsten stimmen die Profile von Vorlagen und Modellen gut überein. Als Unterschied ist noch die höhere Variabilität der Vorlagen insbesondere im Bereich der Aminosäuren 280-300 zu nennen.

Vorlagen. Zum anderen ist die Variabilität der Vorlagen untereinander größer als diejenige der Modelle. In den Bereichen der Aminosäuren 210 bis 220 (rot in Abb. 2.10) haben Vorlagen und Modelle ein vergleichsweise hohes Potential, das abhängig von der einzelnen Struktur über 0 steigen kann. Da dies allerdings auch in den Vorlagen der Fall ist, wird es für die Modelle nicht als Nachteil gewertet. Es handelt sich um den vorderen Bereich des Activation Loops. Im Bereich der Aminosäuren 275 bis 300 (blau in Abb. 2.10) gibt es jeweils innerhalb der Gruppen eine höhere Variabilität. Diese Beobachtung überrascht nicht, da es sich um einen Bereich handelt, der C-terminal dem Lösungsmittel zugewandt liegt, weit von der Bindetasche entfernt ist und daher eine höhere Variabilität erlaubt.

Ebenso gibt es einen an der Hinge Region ansetzenden Bereich (grün in Abb. 2.10, Aminosäuren 115 bis 135), für den ProSa unter den Vorlagen für 20W3 ungewöhnlich hohe Werte berechnet. Offenbar wird die Spitze des Loops als höherenergetischer Bereich eingestuft (insbesondere für 20W3). Hier befinden sich bei einigen Vorlagen nicht aufgelöste Aminosäuren, die den ungünstigen Verlauf erklären können. Das korrespondierende Modell liegt hingegen innerhalb des Profils, das den anderen *PfGSK-3* gemeinsam ist. Insgesamt bestätigt der Verlauf der einzelnen ProSa-Diagramme die gute Qualität der Modelle.

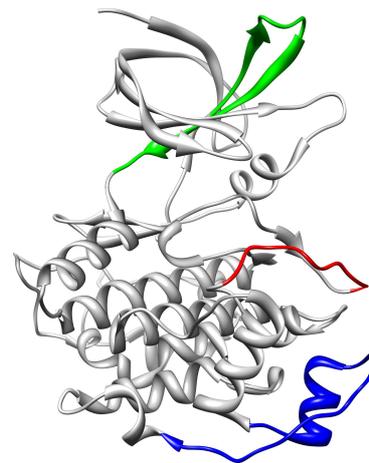


Abbildung 2.10. | Auffällige Bereiche der ProSa-Energieverläufe.

2.4.1.3. Topologie der Bindetasche

Die Dimensionen der Bindetasche können durch Entfernungen zwischen charakteristischen Aminosäuren in der Bindetasche beschrieben werden. Tabelle 2.6 zeigt solche Entfernungen, die für die Breite (Gln210–Met157 / Gln185–Leu132), Tiefe (Lys166–Lys108 / Arg141–Lys85) und Höhe (Leu213–Val93 / Leu188–Val70) stehen. Die Höhe der Bindetasche unterscheidet sich kaum zwischen dem humanen und dem plasmodialen Enzym und schwankt wenig (s_{rel} von 5.86%). Breite und Tiefe sind bei der *HsGSK-3 β* im Mittel um 0.86 und 0.48 Å größer. Die Werte schwanken stärker als die Höhe, insbesondere bei der *HsGSK-3 β* (s_{rel} vgl. Tab. 2.6).

Zwei interessantere Unterschiede zwischen Modell und Vorlage werden bei Betrachtung der Oberflächen der beiden Proteine deutlich. Einerseits wird durch den größeren Gatekeeper Methionin in der *PfGSK-3* die Back Pocket zwischen Met157/Leu132 und Lys108/Lys85 verkleinert bzw. der Zugang erschwert (gut zu erkennen bei *pfgsk_1q3w*, *pfgsk_1q41*, *pfgsk_1q41*, *pfgsk_1q5k*, *pfgsk_1uv5*, *pfgsk_2o5k*, *pfgsk_kubra*). Andererseits weisen mehrere der Modelle eine Wölbung auf, die oberhalb des Gatekeepers Met157 zwischen den Aminosäuren Ala106 und Val155 liegt und die in den humanen Strukturen nicht zu beobachten ist. Dieses Volumen ist besonders deshalb interessant, da in ihm ein Interaktionsfeld liegt, das in der *HsGSK-3 β* weniger ausgeprägt ist als in der *PfGSK-3* (vgl. S. 64). Der Unterschied ist bei den Modellen *pfgsk_1q3d* und *pfgsk_1q3w* besonders gut zu erkennen. Die Oberflächen von *HsGSK-3 β*

Tabelle 2.6. | Ausmaße der Bindetasche von *HsGSK-3 β* vs. *PfGSK-3* in Form charakteristischer Entfernungen d [Å] von Aminosäurepaaren. Es wurden jeweils die durch *RCS* ermittelten besten Homologie-Modelle pro Vorlage vermessen.

<i>PfGSK-3</i>	Gln210– Met157	Lys166– Lys108	Leu213– Val93	<i>HsGSK-3β</i>	Gln185– Leu132	Arg141– Lys85	Leu188– Val70
pfgsk_1j1b	13.03	14.03	9.55	1J1B	15.41	16.35	9.75
pfgsk_1q3d	16.26	17.78	10.66	1Q3D	16.06	19.47	10.70
pfgsk_1q3w	13.49	15.55	11.01	1Q3W	15.51	17.25	10.81
pfgsk_1q41	15.88	17.63	9.76	1Q41	16.15	17.56	9.65
pfgsk_1q4l	13.45	17.80	9.92	1Q4L	12.87	16.51	9.96
pfgsk_1q5k	13.83	15.59	9.74	1Q5K	15.66	17.17	9.68
pfgsk_1r0e	15.61	17.04	10.41	1R0E	16.12	19.29	10.41
pfgsk_1uv5	16.17	19.49	9.78	1UV5	16.58	17.79	9.65
pfgsk_2o5k	16.19	13.65	9.46	2O5K	16.58	10.71	9.22
pfgsk_2ow3	16.09	20.01	11.19	2OW3	17.23	21.25	11.20
pfgsk_kubra	15.23	20.17	9.92	kubra	16.54	18.47	9.96
\bar{d}	15.02	16.86	10.15		15.88	17.34	10.10
s_{rel} [%]	8.60	13.37	5.86		7.14	15.34	6.01

und *PfGSK-3* sind am Beispiel von 1Q3W und dem korrespondierenden Homologie-Modell in Abbildung 2.11 dargestellt.

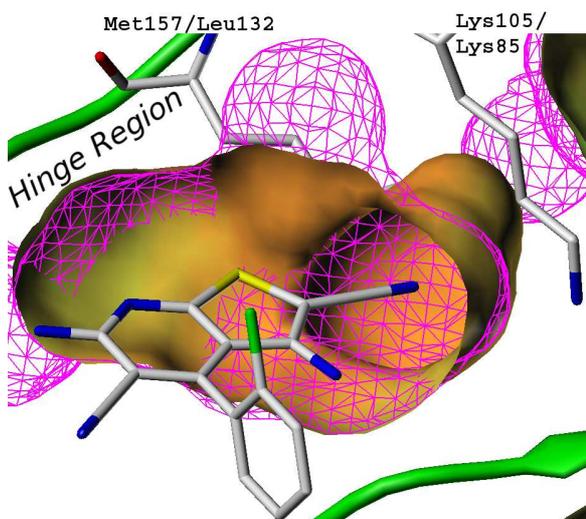


Abbildung 2.11. | Oberflächenprofil der Bindetaschen von *HsGSK-3 β* (farbkodiert nach Taschentiefe) und *PfGSK-3* (magentafarbenes Netz) am Beispiel von 1Q3W bzw. pfgsk_1q3w. Zur Orientierung ist 31 dargestellt, dessen 2-Cyanosubstituent in Richtung des Lys85 der *HsGSK-3 β* zeigt. Direkt daneben tut sich in der *HsGSK-3 β* die Back Pocket auf, die in der *PfGSK-3* durch den größeren Gatekeeper in einigen Modellen stärker verdeckt wird als in der *HsGSK-3 β* . Das Oberflächenprofil oberhalb des Gatekeepers Met157/Leu132 in der Nähe der Aminosäuren Ala106 und Val155 weist in der *PfGSK-3* die im Text beschriebene Ausbuchtung auf, die in 1Q3W gut zu erkennen ist.

2.4.1.4. MD-Simulationen

Zur Absicherung der Qualität der Modelle wurden MD-Simulationen über 2 ns bei 300 K durchgeführt und die Energieverläufe sowie die Stabilität der Modelle im Sinne von rmsd-Fluktuationen begutachtet. Es wurde das gemäß *RCS* am besten bewertete Protein von jeder der 22 Modellvarianten in die Berechnung einbezogen. Die Proteine zeigten ähnliche Verläufe, in Abbildung 2.12 ist beispielhaft derjenige des Modells pfgsk_1j1b_128 abgebildet.

Stellvertretend für die Vorlagen wurde eine entsprechende Simulation mit 1J1B durchgeführt. Die dabei beobachteten rmsd-Fluktuationen und Energieverläufe waren den bei den

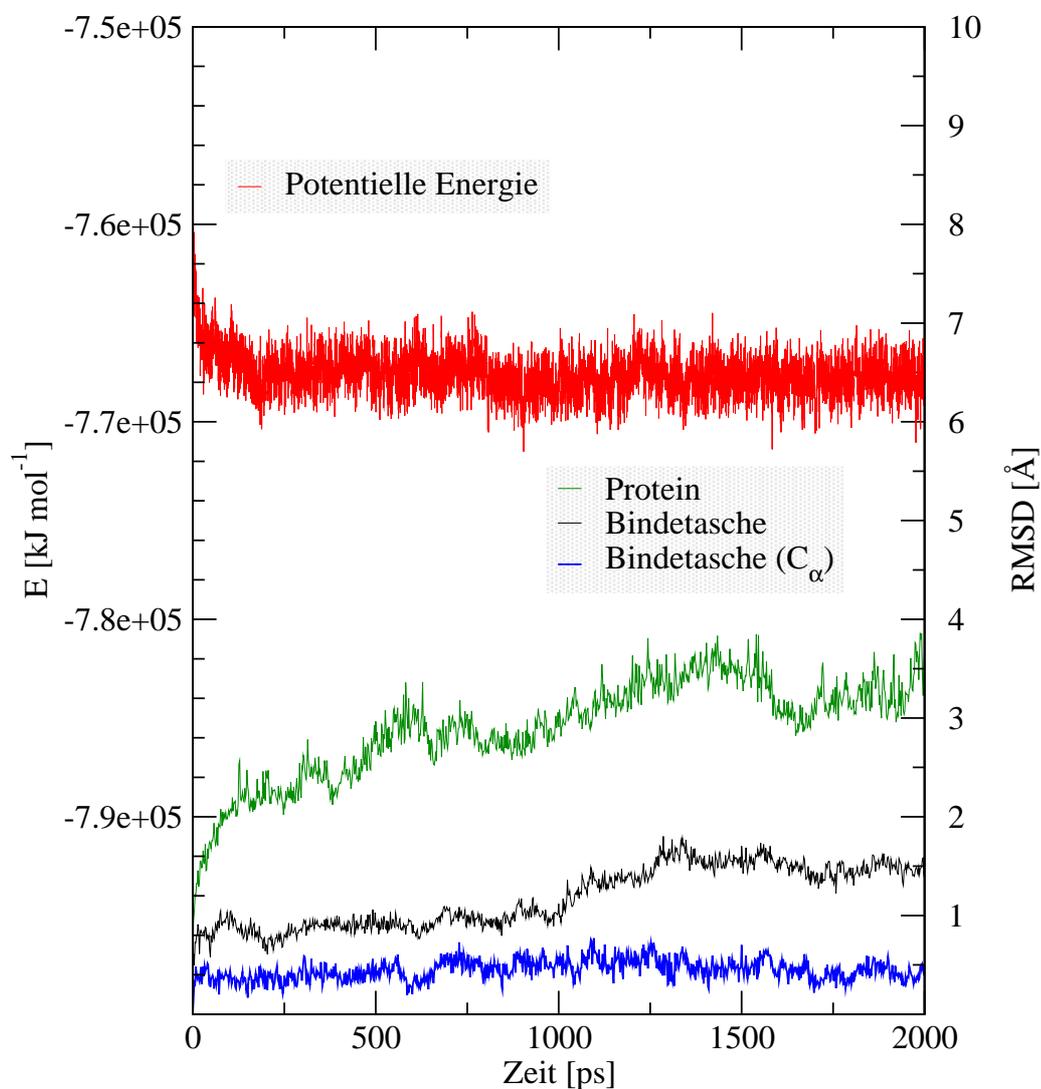


Abbildung 2.12. | Moleküldynamik von pfgsk_1j1b_128. Die rmsd-Werte wurden nach Überlagerung auf die C_α-Atome der Ausgangskonformation berechnet. Sie zeigen, dass es keine größeren Fluktuationen gibt. Für das gesamte Protein betragen die Abweichungen zur Ausgangskonformation etwa 3.5 Å. Für die Bindetasche stellvertretend die Aminosäuren Lys108, Gln158, Met157, Ile160, Lys166 und Gln210 betrachtet. Dieser Bereich ist sehr stabil, die Schwankungen sind mit Abweichungen unter 0.5 Å für die C_α-Atome praktisch zu vernachlässigen. Auch die Energiewerte sprechen für das Modell, nach 250 ps erreicht die potentielle Energie E ein stabiles Plateau.

Homologie-Modellen beobachteten ähnlich, was deren gute Qualität bestätigt. Bei Modellen, die unter Berücksichtigung von Ligandinformationen entstanden, liegt das Plateau der potentiellen Energie, das im Äquilibrium erreicht wird, tendenziell über demjenigen des entsprechenden Modells, das ohne diese Information konstruiert wurde. Bei den Modellen `pfgsk_2o5k_HBM` und `pfgsk_1uv5_BRW` wurde der MD-Lauf sogar aufgrund sterischer Hinderungen innerhalb des Proteins abgebrochen. Die zu dichten Kontakte von Atomen innerhalb des Proteins scheinen Folge des Erstellungsprotokoll zu sein, das zusätzliche Einschränkungen bei der Konstruktion des Modells macht. Der Abbruch der MD-Simulation aufgrund sterischer Hinderungen und die Energieverläufe bestätigen die schlechtere *RCS*-Bewertung der Modelle, die unter Berücksichtigung von Ligandinformationen entstanden.

2.4.1.5. Fazit Homologie-Modelle

Durch die Anwendung des *RCS* konnte eine Reihe hochwertiger Homologie-Modelle für weitere strukturbasierte Untersuchungen zur Verfügung gestellt werden. Ihre gute Qualität wurde durch die Ergebnisse der verschiedenen Evaluationsprogramme und durch MD-Simulationen bestätigt. Die zusätzlichen sterischen Einschränkungen in Form von festgesetzten Distanzen hingegen, die bei der Konstruktion der Modelle unter Einbeziehung von Ligandinformation gesetzt wurden, führten offenbar zu problematischen Proteingeometrien. Sie stellten sich im Rahmen der Evaluierung als unterlegen gegenüber den übrigen heraus. Möglicherweise kam dies dadurch zustande, dass die starre Einschränkung im Bereich der Bindetasche zu weniger vorteilhaften Geometrien auch in angrenzenden Bereichen führte, da eine Optimierung durch die Einschränkungen nur unvollständig möglich war. Dies könnte zu den schlechteren Werten im Rahmen der Evaluierung und zu den abgebrochenen MD-Simulationen geführt haben.

Die betreffenden Strukturen wurden daher als weniger geeignet betrachtet, um weitere Schlüsse zu begründen. Alle nachfolgenden Untersuchungen wurden mit den 11 ohne Ligandinformation erstellten Modellen durchgeführt. Ihre bei der Evaluation ermittelten Kennzahlen sind im oberen Teil von Tabelle B.2 auf Seite 232 im Anhang aufgeführt. Sie wurden ausgewählt, um verschiedene konformative Zustände der PfGSK-3 in den folgenden Untersuchungen zu repräsentieren.¹²

¹²Da für jede Vorlage ein Modell ausgewählt wurde, wird dieses im Folgenden ohne die laufende Nummerierung mit `pfgsk_<Vorlage>` bezeichnet.

2.4.2. Analyse der molekularen Interaktionsfelder

2.4.2.1. Visueller Vergleich der Felder

Die Energieminima aller berechneten Interaktionsfelder und deren Koordinaten für 1J1B und das entsprechende plasmoidale Modell sind in Tabelle 2.7 dargestellt. Ohne eine absolute Vergleichbarkeit zu gewährleisten, geben sie einen Eindruck davon, ob insgesamt große Wechselwirkungsenergien zu erwarten sind (z. B. OH2-Sonde) oder ob der Beitrag einer Interaktion vom GRID-Kraftfeld als weniger ausschlaggebend bewertet wird (z. B. DRY-Sonde). Dieser Vergleich macht keine Aussage über die Wichtigkeit der entsprechenden Interaktion für eine Bindung. Auch wenn das berechnete Interaktionspotential der DRY-Sonde klein ist, können entropische Beiträge ausschlaggebend für die Aktivität eines Liganden sein. Die Werte sind vielmehr angegeben, um Gemeinsamkeiten zwischen Sonden aufzuzeigen und Unterschiede zwischen den Proteinen schon anhand der Energieminima deutlich zu machen. Sie geben zudem einen Eindruck davon, wie die im Einzelfall angepasste Auswahl der Energielevel einzuordnen ist, bei denen die Interaktionsfelder in diesem Kapitel visualisiert wurden.

Tabelle 2.7. | Energieminima der Felder von 1J1B und pfgsk_1j1b. Bei den Koordinaten der Energieminima fällt zunächst auf, dass nur im Falle der N:#-Sonde die Position stärkster Interaktionsenergie bei *HsGSK-3 β* und *PfGSK-3* gleich ist. Darüber hinaus fallen beieinanderliegende Koordinaten für Sonden mit ähnlichen Eigenschaften ins Auge (z. B. OC1 und OES in *HsGSK-3 β* , I und C3 in *PfGSK-3*).

Sonde	<i>HsGSK-3β</i> (1J1B)				<i>PfGSK-3</i> (pfgsk_1j1b)			
	E_{min}	x	y	z	E_{min}	x	y	z
OH2	-20.6	-14.00	-11.50	-2.50	-20.3	7.00	6.50	-0.50
DRY	-3.1	3.75	0.25	4.25	-2.8	6.75	9.50	9.75
I	-13.5	0.50	11.50	-2.50	-13.6	-5.75	5.50	-8.00
C3	-5.0	-1.00	2.75	1.75	-5.0	-5.50	5.25	-7.70
N:#	-12.8	6.00	6.00	-3.00	-13.4	6.25	6.25	-2.50
O	-11.8	-2.00	11.00	9.00	-11.9	-7.00	1.50	10.75
OC1	-6.9	-2.00	11.00	8.75	-6.6	-11.50	4.75	7.50
OES	-6.2	-2.00	11.00	9.00	-5.9	-11.50	4.75	7.50
N2	-17.0	6.00	6.25	2.75	-16.9	7.00	7.00	-0.50
N1+	-20.3	3.50	2.25	2.50	-20.0	-4.50	8.75	-8.50
AR.CONH2	-18.6	-14.25	11.25	-2.50	-18.6	7.00	6.75	-0.50

2.4.2.1.1. Die OH2-Sonde Mit dieser Sonde werden Interaktionsenergien mit Wassermolekülen berechnet. Sie repräsentiert Wasser als ein Sauerstoffatom mit den in Tabelle 2.3 aufgeführten Eigenschaften. Die Felder werden bei einem Energielevel von -14 kcal/mol betrachtet (vgl. Abb. 2.13), wobei sich Übereinstimmungen vor allem im vorderen Bereich A der Bindetasche und weiter hinten (B und rechts davon) erkennen lassen.

Deutliche Abweichungen sind das in der Hinge Region der *PfGSK-3* ausgeprägte Feld im Bereich D und die Verschiebung des weit außen liegenden Bereiches C. Letztere ist jedoch

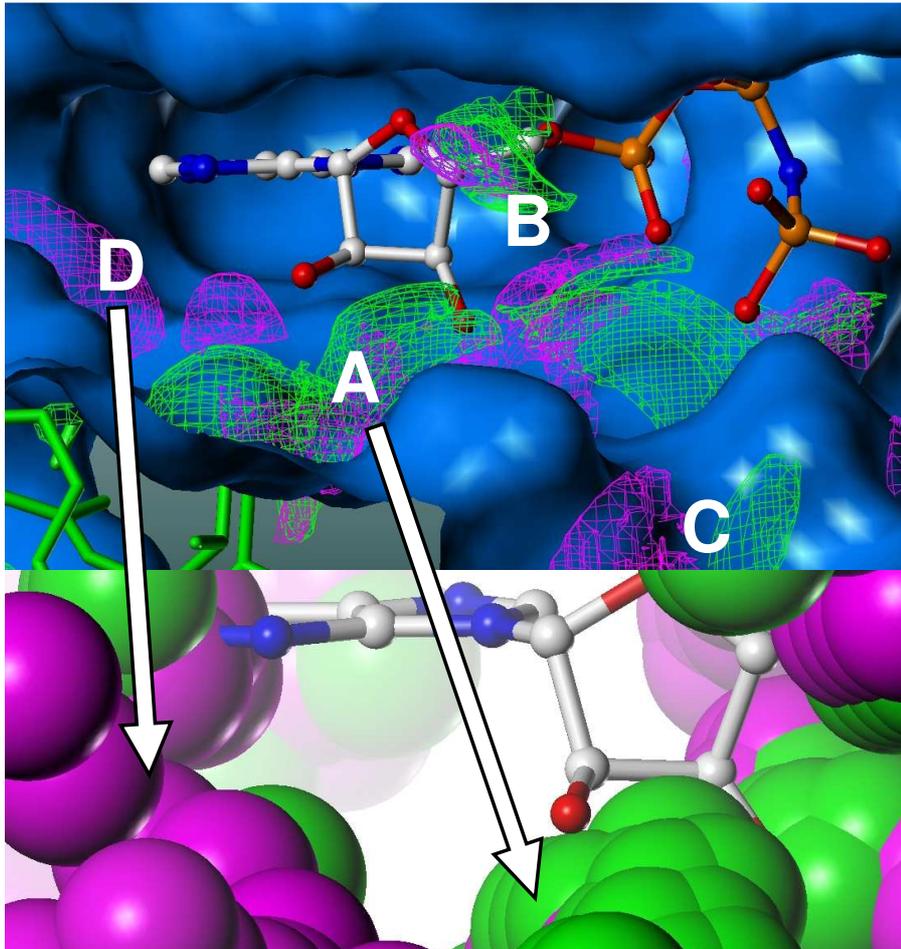


Abbildung 2.13. | Felder der OH2-Sonde. Dargestellt ist ANP in der aus 1J1B bekannte Position mit MIFs der OH2-Sonde. Der Bereich C liegt weit vom ANP entfernt, klare Übereinstimmungen sind bei B und weiter hinten zu erkennen. Der Abstand des Feldes bei D vom Backbone-Carbonyl des Pro161 beträgt 2.97 Å. Der untere Teil der Abbildung zeigt die Häufung von Minima pro Fläche im Bereich der Hinge Region der PfGSK-3 und bei der HsGSK-3 β mehrere Minima im Bereich A.

aufgrund der großen Entfernung zum Liganden (5.78 Å zum nächsten Phosphat-Sauerstoff des dargestellten ANP) weniger interessant. Zudem werden in der HsGSK-3 β im Bereich der zuckerbindenden Region stärkere Interaktionsenergien berechnet als in der PfGSK-3. Die genannten Unterschiede bestätigen sich auch bei der Betrachtung der Minima pro Fläche: In der PfGSK-3 befindet sich im Bereich D eine Anhäufung von Minima, für die HsGSK-3 β ist dies vor allem im Bereich A zu beobachten.

Das starke OH2-Feld der PfGSK-3 im Bereich D ist auf die Carbonyl-Sauerstoffe von Pro161 und Ile160 zurückzuführen, es wäre für die HsGSK-3 β ebenso zu erwarten und zeigt sich auch bei höheren Energielevels. Der Unterschied ist also überraschend, zumal es in diesem Bereich keine Seitenketten der Aminosäuren der Hinge Region gibt, die direkt in die Bindetasche hineinragen. Die Atome des Backbones bilden vielmehr übereinstimmend das konservierte Wasserstoffbrückenmuster mit Inhibitoren aus. Unterschiede in der Sequenz finden sich allerdings im nahegelegenen lösungsmittelzugewandten Bereich. So können das 2.90/3.04 Å entfernte

Lys166/Arg141 bzw. das Gln162/Glu136 einen Erklärungsansatz bieten (vgl. 0C1-Sonde S. 61).

Der Vergleich der weiteren Homologie-Modell/Vorlage-Paare zeigt im Bereich D bei 1Q4L, 20W3, 1Q3D, 1Q5K und 1Q3W einen vergleichbaren Unterschied, wobei das Feld bei 1Q3D, 1Q5K und 1Q3W noch weiter in die Tasche hinein zum Gatekeeper hin verschoben liegt. Die Distanz zwischen Feld und der Methylgruppe des Met157 beträgt bei pfgsk_1q5k nur 2.35 Å.

2.4.2.1.2. Die DRY-Sonde Die DRY-Sonde berechnet mögliche hydrophobe Interaktionen (den hydrophoben Effekt) der Sonde mit dem Zielmolekül in der Art eines „inversen Wassermoleküls“ [167].

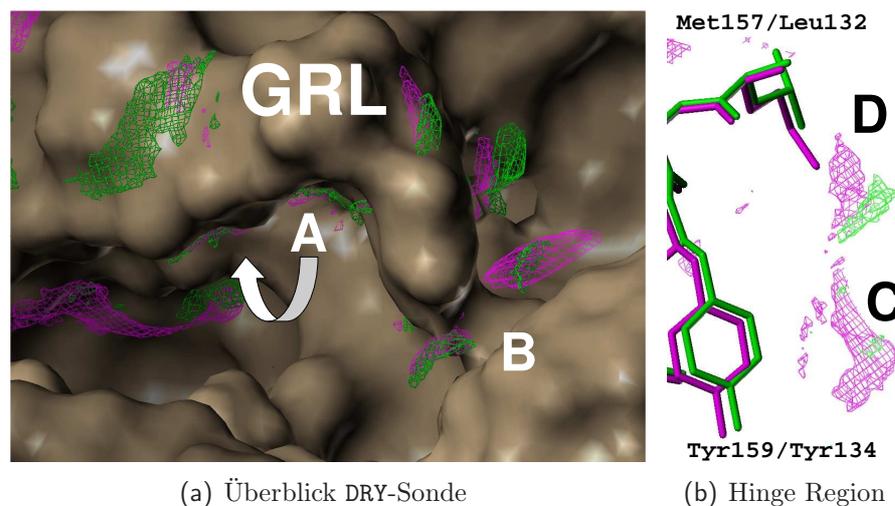


Abbildung 2.14. | Felder der DRY-Sonde. In (a) ist die Oberfläche der ATP-Bindetasche der *HsGSK-3β* mit den umgebenden DRY-Feldern dargestellt. Direkt unterhalb des Glycine-rich Loops (GRL) sind praktisch identische Felder (A) und auch um die Spitze herum nahe Phe90/Phe67 ist die Ähnlichkeit groß. Weiter Richtung Hinge Region zeigen sich allerdings die im Text beschriebenen und in (b) dargestellten Unterschiede.

Die Darstellung erfolgt bei einem Energielevel von -0.8 kcal/mol, die Werte liegen im Vergleich zur OH2-Sonde deutlich höher, die Interaktionsenergien sind also weniger stark ausgeprägt. Gemeinsamkeiten (vgl. Abb. 2.14(a)) fallen zunächst insbesondere unterhalb des Glycine-rich Loops nahe Gly91/Gly68, Gly88/Gly65 (A) und um das Phe90/Phe67 bzw. Gly227/Gly202 (B) auf. Auch der flächige Bereich, der z. B. vom Purin des ANP besetzt wird, ist bei *HsGSK-3β* und *PfGSK-3* zu erkennen (vgl. Abb. 2.14(b)), hier zeigt sich allerdings ein Unterschied in der Ausprägung: Das magenta dargestellte Feld der *PfGSK-3* ist stärker als das grüne der *HsGSK-3β* (Areal C in Abb. 2.14(b)). Auch der vom Gatekeeper bestimmte Bereich D ist in der *PfGSK-3* zum Met157 hin verschoben. Diese Verschiebung scheint zunächst widersprüchlich, da bei der Untersuchung der Oberflächenprofile von Modellen und Vorlagen beobachtet wurde, dass das Met157 der *PfGSK-3* die hydrophobe Tasche eher verschließt als das kürzere Leu132 der *HsGSK-3β*. Offenbar führt jedoch die höhere Flexibilität des Methionins dazu, dass mögliche Interaktionen der Sonde mit der *PfGSK-3* im Vergleich zur *HsGSK-3β* im tieferen Bereich der Bindetasche berechnet werden.

Beim Vergleich weiterer Homologie-Modelle bestätigen 1Q3D und 205K das maßgeblich vom Ile85/Ile62 bestimmte stärkere DRY-Feld in der *PfGSK-3* im Bereich C, das von planaren Strukturen wie dem Purinkörper des ATP besetzt werden kann. Die Verschiebung des Interaktionsfeldes der *PfGSK-3* gegenüber der *HsGSK-3 β* beim Gatekeeper kann nur bei 1Q3D beobachtet werden (D in Abb. 2.14(b)). Für die anderen untersuchten Strukturen werden praktisch identische DRY Felder berechnet.

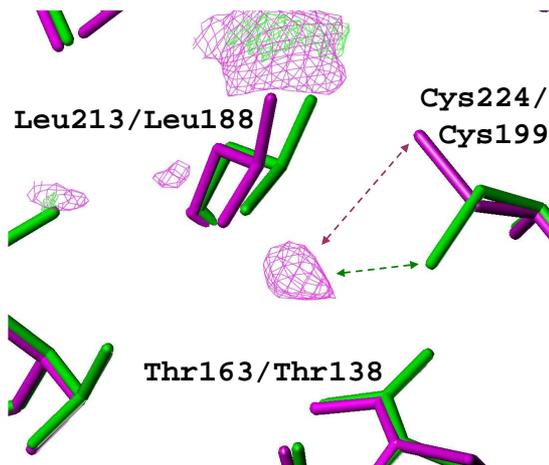


Abbildung 2.15. | Felder der I-Sonde. Der untere Bereich der Bindetasche wird unter anderem von den abgebildeten Aminosäuren begrenzt. Die große, gut polarisierbare I-Sonde findet hier in der *HsGSK-3 β* weniger Platz, während in der *PfGSK-3* das zentrale Feld zu sehen ist. Es liegt nur 2.4 Å vom Cys199 der *HsGSK-3 β* entfernt, während mit 4.2 Å genug Platz bei der *PfGSK-3* bleibt - die Sonde selbst hat einen Radius von 2.15 Å.

2.4.2.1.3. Die I-Sonde Die I-Sonde bildet die Eigenschaften von organisch gebundenem Iod ab und zeichnet sich durch ihre ausgeprägte Polarisierbarkeit aus. Sie zeigt mit einem Minimum von -13.5 kcal/mol eine vergleichsweise starke Interaktion an – dieses Minimum liegt unter dem aller untersuchten Sauerstoff-Sonden (vgl. Tabelle 2.7 auf S. 57). Die Felder werden bei einem Energielevel von -8 kcal/mol betrachtet. Dabei fällt das I-Feld im Zentrum eines Dreiecks zwischen Thr163/Thr138, Leu213/Leu188 und Cys224/Cys199 auf, das am Boden der Bindetasche in der *PfGSK-3*, nicht aber in der *HsGSK-3 β* vorhanden ist (vgl. Abb. 2.15). Die Auswertung mit BIOCU-BE unterstreicht die dargestellten Verhältnisse.

Der Vergleich fällt für 1Q3W, 1Q41, 1R0E, 1UV5 und 20W3 und die abgeleiteten Modelle ähnlich aus. Bei *pfgsk_1q5k* ist das Feld zwar aus der Tasche hinaus zum Asp211 hin verschoben, es ist aber auch hier im plasmodialen Enzym ausgeprägter als im humanen.

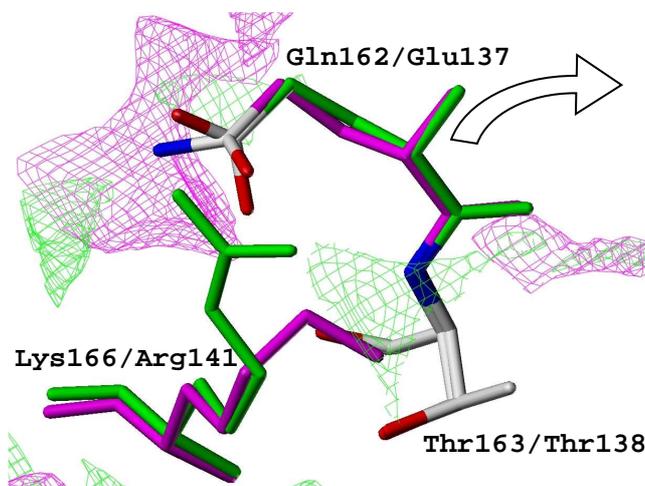
2.4.2.1.4. Die C3-Sonde Diese Sonde repräsentiert eine Methylgruppe, sie ist also nicht zur Ausbildung von Wasserstoffbrückenbindungen befähigt und ihre Parameter gleichen bis auf die geringere Polarisierbarkeit denen der I-Sonde. Die berechneten Interaktionsenergien wiederum sind denen der DRY-Sonde ähnlich und es gibt erwartungsgemäß viele Parallelen mit den in den letzten Abschnitten betrachteten Feldern.

So findet sich erneut der im letzten Abschnitt beschriebene interessanter Unterschied zwischen 1J1B und *pfgsk_1j1b* im unteren Bereich der Bindetasche nahe Leu213/Leu188 und Cys224/Cys199. Das plasmodiale Homologie-Modell zeigt hier ein ausgeprägtes C3-Feld, das es in dieser Form in der humanen Vorlage nicht gibt. Da die Positionierung derjenigen des in Abbildung 2.15 im vorangegangenen Abschnitt dargestellten I-Feldes gleicht, ist es nicht gesondert dargestellt.

Während bei 1Q41 und 1UV5 bzw. deren korrespondierenden Modellen der gleiche, bei 1J1B beobachtete deutliche Unterschied zwischen *Hs*GSK-3 β und *Pf*GSK-3 besteht, ist dieser bei 1Q3W erst bei höherem Energielevel zu beobachten. Bei den Homologie-Modell/Vorlage-Paaren beruhend auf 1Q4L und 1Q5K sind die Felder ähnlich intensiv, während sie im Fall von 1Q3D und 205K praktisch nicht zu beobachten sind. Bei *pfgsk_1r0e* ist das Feld stark zum Leu213 verschoben (Distanz zum Cys224: 3.6 Å) und 20W3 ist die einzige Struktur, in der das berechnete Feld für die *Hs*GSK-3 β stärker ist als für die *Pf*GSK-3.

2.4.2.1.5. Die OC1-Sonde Diese Sonde steht für einen Wasserstoffbrückenakzeptor. Sie hat die Eigenschaften eines gemischten Aryl-Alkyl-Ethers (schwache negative Partialladung, schwächer polarisierbar als bisher betrachtete Sonden).

Abbildung 2.16. | OC1-Feld an der Hinge Region. Die dargestellten Aminosäuren geben möglicherweise die im Text erläuterte Erklärung für den Unterschied im Bereich der Hinge Region, die im dargestellten Bereich ansetzt und sich in Richtung des Blockpfeils mit Pro161-Ile160 etc. fortsetzt.



Die Felder werden bei einem Energielevel von -4.5 kcal/mol betrachtet. Es fällt vor allem der in Abbildung 2.16 dargestellte Unterschied auf, der im vorderen Bereich der Hinge Region bei Ile160/Val135 (erste Aminosäure der Hinge Region in Richtung des Blockpfeils) bzw. Thr163/Thr138 beobachtet wird. Die Abbildung gibt einen Erklärungsansatz für diesen möglicherweise wichtigen Unterschied zwischen *Hs*GSK-3 β und *Pf*GSK-3. Dadurch dass in der *Hs*GSK-3 β ein Glutamat an Position 137 vorliegt, wird mit dem naheliegende Arg141 eine Salzbrücke möglich (Distanzen der verschiedenen Vorlagen vgl. Tab. 2.2 auf S. 35), die in der *Pf*GSK-3 nicht entstehen kann, da das Amid des Gln162 dazu nicht in der Lage ist. Auch wenn eine Wasserstoffbrücke nach wie vor möglich ist, ist das Lys166 weniger stark fixiert und wird dadurch als Wasserstoffbrückendonator frei. Es kann durch den großen konformativen Spielraum der Seitenkette in die Tasche hinein einen weiteren Bereich abdecken, indem es mit einem entsprechenden Akzeptor wechselwirken kann – und so möglicherweise das bereits bei den Sonden O und N:# auffällige Feld begründet. Eine ähnliche Verschiebung des OC1-Feldes an der Hinge Region wie in Abbildung 2.16 wird bei den auf 1Q4L, 1UV5 und 1Q41 beruhenden Homologie-Modellen der *Pf*GSK-3 beobachtet.

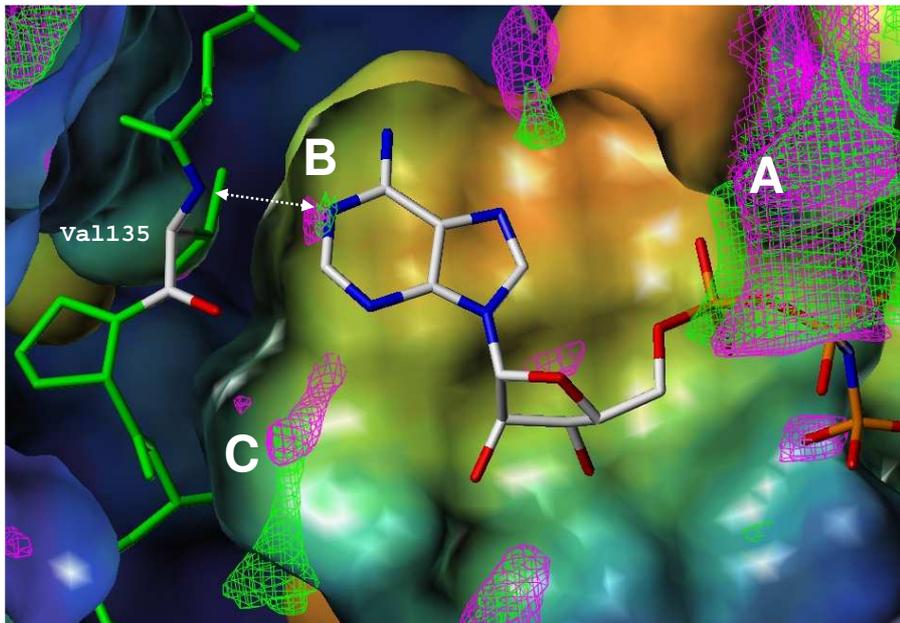


Abbildung 2.17. | Felder der N: #-Sonde. Im Bereich A nahe dem Glycine-rich Loop fallen für die MIFs der N: #-Sonde wenige Unterschiede auf. Auch die wichtige Wasserstoffbrücke des Stickstoffs im Purin des ATP zum Val135 kann als Interaktionsfeld (B) für beide Proteine lokalisiert werden (Entfernung N/Purin zu NH/Val135: 3.08 Å) – auch wenn die Sondenparameter auf einen *sp*-hybridisierten Stickstoff mit einem einsamen Elektronenpaar eingestellt sind, ist hier grundsätzlich die Bedeutung eines Wasserstoffbrückenakzeptors zu erkennen. Die Aufsicht auf die Bindetasche gibt bei C allerdings den Hinweis auf einen wesentlichen Unterschied: Das Feld liegt für die PfGSK-3 weiter ins Innere der Bindetasche verschoben.

2.4.2.1.6. Die N: #-Sonde Ein *sp*-hybridisierter Stickstoff mit freiem Elektronenpaar wird durch die N: #-Sonde vertreten. Die Ergebnisse dieser Sonde sind insofern besonders interessant, als dass eine Nitrilgruppe als Substitution an C2 bzw. C5 charakteristisch für die in dieser Arbeit untersuchten Thienopyridine ist.

Gemeinsamkeiten der bei -9 kcal/mol betrachteten MIFs zeigen sich insbesondere im Bereich des Glycine-rich Loops und zum Teil in Wechselwirkung mit der Hinge Region, sie sind in Abbildung 2.17 zu erkennen (Bereiche A und B). Man sieht jedoch ebenso einen sehr interessanten Unterschied im vorderen Bereich der Hinge Region: Das N: #-Feld für die PfGSK-3 ist in die Tasche hinein verschoben. Es wird also im Falle der HsGSK-3β eher vom Arg141 am Rand bestimmt, während das entsprechende Lys166 auf weiter innerhalb der Tasche gelegene Bereiche Einfluss hat. Prinzipiell ist eine Wasserstoffbrücke der Sonde mit der Hydroxy-Gruppe des Thr163 am Grund der Tasche denkbar (unterhalb der Oberfläche bei C), insbesondere wenn hier durch eine Änderung des Torsionswinkels der Seitenkette die Geometrie angepasst würde. Ein Unterschied in der Ausrichtung dieser Aminosäure, der durch die unterschiedlich starken Felder naheliegt, könnte sich deutlich auf den Bindungsmodus eines Liganden auswirken.

Das Homologie-Modell/Vorlage-Paar *kubra* bestätigt den Unterschied beim Thr163. Andere Differenzen (auch diejenigen der bis hier beschriebenen anderen Sonden) wurden in diesem Paar nicht gefunden. Die im Liganden enthaltenen Nitrilgruppen finden kein Gegenstück in Form von

entsprechenden Feldern der Sonde N:#. Beim Vergleich anderer Paare ergibt sich nur bei 1UV5 bzw. 1Q4L eine so klare Verschiebung des Feldes.

2.4.2.1.7. Die O-Sonde Eine weitere Sonde mit ausgeprägtem Wasserstoffbrückenakzeptor-Charakter ist die O-Sonde, die das physikochemische Verhalten eines Carbonylsauerstoffes imitiert. Im Unterschied zur N:#-Sonde trägt sie eine negative Partialladung und ist befähigt, zwei Wasserstoffbrücken auszubilden.

Es wird wiederum ein an der Hinge Region entlang führendes Feld beobachtet, das sich bei den plasmodialen Homologie-Modellen tiefer in die Tasche hineinzieht, als es bei der humanen Vorlage der Fall ist (vgl. OH2-Sonde, Abschnitt 2.4.2.1.1). Dieser Unterschied wird von den Strukturen 1UV5, 1R0E und 1Q5K untermauert. Die Strukturen 1Q3D, 1Q4L, 205K, 20W3 sowie *kubra* und die entsprechenden abgeleiteten Homologie-Modelle haben im Bereich der Hinge Region ähnliche O-Felder.

2.4.2.1.8. Die N2-Sonde Die Felder dieser Sonde geben Interaktionsenergien mit einem neutralen, planaren Stickstoff wieder, wie er in einem Amid oder als Substituent eines Aromaten vorliegt. Es werden also Interaktionsmöglichkeiten mit einem Wasserstoffbrückendonator beschrieben. Die Minima für 1J1B und *pfgsk_1j1b* werden mit -17.0 bzw. -16.9 kcal/mol berechnet – es handelt sich damit um vergleichsweise starke Wechselwirkungen, die Felder werden bei -11 kcal/mol betrachtet. Dabei zeigt sich bei *pfgsk_1j1b* gegenüber der humanen Struktur ein deutlicher ausgeprägtes Feld, das von den beiden Backbone-Carbonyl-Sauerstoffen von Ile160 und Pro161 in der Hinge Region bestimmt wird (vgl. Abb. 2.18).

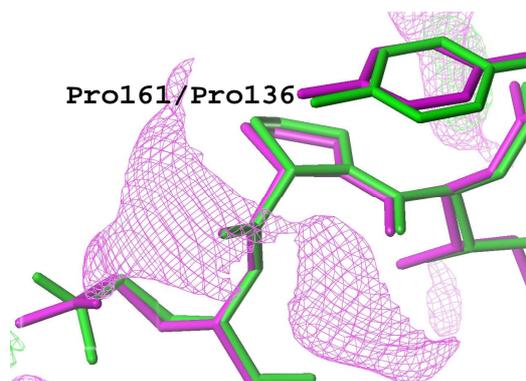


Abbildung 2.18. | N2-Feld nahe Pro161/Pro136 bzw. Ile160/Val135. Der dargestellte Teil des N2-Feldes der *PfGSK-3* deutet in der Hinge Region auf deutlich stärkere Wechselwirkungsenergien hin, als diese bei der *HsGSK-3 β* berechnet wurden. Das Feld entspricht in Lage und Ausdehnung dem in Abbildung 2.13 auf Seite 58 dargestellten OH2-Feld.

Eine entsprechende Abweichung der Felder zwischen Homologie-Modell und Vorlage kann bei 1UV5, 20W3, 1Q3W und 1Q4L beobachtet werden. Für die letzten beiden wird zusätzlich ein stärkeres N2-Feld der *PfGSK-3* gegenüber der *HsGSK-3 β* durch das Backbone-Carbonyl des Glu158/Asp133 berechnet, während dieses Verhältnis sich bei 20W3 und *pfgsk_2ow3* umkehrt (ebenso bei 1Q41).

Die Strukturen 1Q3D, 1Q5K und 1R0E und die abgeleiteten Homologie-Modelle zeigen beim gewählten Energielevel von -11 kcal/mol kein starkes Feld nahe der Hinge Region und bei 205K entsprechen sich die Felder von *HsGSK-3 β* und *PfGSK-3* nahe Pro161/Pro136, während das N2-Feld bei Glu158/Asp133 in der *HsGSK-3 β* stärker ausfällt als bei der *PfGSK-3*.

2.4.2.1.9. Die N1+-Sonde Mit dieser Sonde wird das Volumen in der Bindetasche auf Interaktionsmöglichkeiten mit einem protonierten Stickstoff abgetastet. Es zeigt sich ein Unterschied zwischen *PfGSK-3* und *HsGSK-3 β* im hinteren Bereich der Tasche nahe dem Gatekeeper (siehe Abb. 2.19).

Das Feld kann in der *PfGSK-3* in vergleichbarer Ausprägung in den Modellen basierend auf 1Q3D, 1Q41, 1Q4L, 1UV5, 205K und 20W3 festgestellt werden, während für die Vorlagen hier bei einem Energielevel von -8.5 kcal/mol keine Interaktionsmöglichkeit berechnet wird. Bei 1Q3W und 1Q5K ist dies auch der Fall, allerdings weniger ausgeprägt bzw. erst bei höheren Energielevels. Das Feld ist insofern besonders bemerkenswert, als dass es in einer Ausbuchtung der Bindetasche liegt, die bei mehreren Homologie-Modellen stärker ausgeprägt ist als bei den entsprechenden Vorlagen der *HsGSK-3 β* (vgl. Oberflächenprofile von *HsGSK-3 β* und *PfGSK-3* auf S. 54).

2.4.2.1.10. Die AR.CONH2-Sonde Die Multiatom-Sonde AR.CONH2 hat die Eigenschaften eines aromatischen Amids. Es können also Wasserstoffbrücken als Donor- oder auch als Akzeptor ausgebildet werden. Insgesamt werden vergleichsweise starke Interaktionsenergien berechnet (vgl. Tab. 2.7 auf S. 57).

Es bestätigt sich wiederum im Bereich der Hinge Region das ausgeprägtere Feld der *PfGSK-3* (vgl. OH2-Sonde, Abschnitt 2.4.2.1.1). Zudem ist im Falle von 1J1B und pfgsk_1j1b im Bereich des Glycine-rich Loops das Feld der *HsGSK-3 β* stärker ausgeprägt. Da sich diese Differenzen in den weiteren Modell/Vorlage-Paaren teilweise umkehren, werden sie nicht als interpretierbar eingestuft.

2.4.2.2. Gegenüberstellung wichtiger Unterschiede der MIFs von *HsGSK-3 β* und *PfGSK-3*

Die wichtigsten Unterschiede zwischen den GRID-Feldern der *HsGSK-3 β* und denen der abgeleiteten Homologie-Modelle sind in Tabelle 2.8 zusammengefasst. Die auffälligsten Differenzen betreffen die Hinge Region (Pro161-Ile160, OH2- und N2-Sonde) bzw. den lösungsmittelzuge-

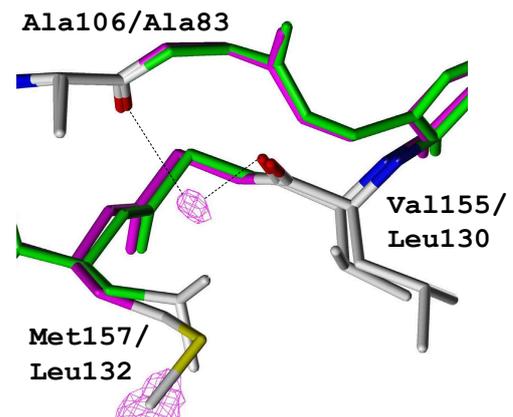


Abbildung 2.19. | Feld der N1+-Sonde. Die Entfernung des Feldes zum Carbonyl des Ala106/Ala83 bzw. Val1155/Leu130 beträgt 3.2 bzw. 3.1 Å.

wandten Bereich nahe der Hinge Region (N:#-, O- und OC1-Sonde). Dieser Befund überrascht insofern, als dass gerade hier keine entscheidenden Unterschiede in der Sequenz vorzufinden sind und der Abschnitt darüber hinaus als wenig flexibel gilt. Allerdings ergibt sich über die geschilderte Möglichkeit des Lys166 der *PfGSK-3*, in die Tasche hineinzuschwingen, ein Erklärungsansatz für die Differenzen (vgl. OC1-Sonde, Abb. 2.16).

Tabelle 2.8. | Wichtige unterschiedliche MIFs der Homologie-Modell/Vorlage-Paare, die bei 1J1B beobachtet wurden. Die genannte Felder sind in der *PfGSK-3* stärker ausgeprägt oder aber in die angegebene Richtung verschoben. Die AR.CONH2-Sonde ist wegen des uneinheitlichen Bildes nicht mit aufgeführt.

Sonde	Feld im Bereich	Bestätigung durch Homologie-Modell/Vorlage-Paar
OH2	Pro161-Ile160	1Q3D, 1Q3W, 1Q4L, 1Q5K, 20W3
DRY	Purin des ANP	1Q3A, 1Q5K
I	Leu213/Cys224/Thr168	1Q3W, 1Q41, 1R0E, 1UV5, 20W3, (1Q5K)
C3	⇒ I-Sonde	1Q4L, 1UV5, (1Q3W, 1R0E)
N:#	⇒ OH2-Sonde, tiefer in die Tasche hinein, mehr Thr168 bzw. Lys166	1Q4L, 1UV5, kubra
OC1	⇒ N:#-Sonde;	1Q3W, 1R0E, 20W3 bzw. 1Q4L, 1Q41, 1UV5
O	⇒ N:#-Sonde	1Q5K, 1R0E, 1UV5
N2	⇒ OH2-Sonde	1Q3W, 1Q4L, 1UV5, 20W3
N1+	Ala106/Val155/Met157	1Q3D, 1Q41, 1Q4L, 1UV5, 205K, 20W3, (1Q3W, 1Q5K)

Ein weiterer in mehreren Paaren festgestellter Unterschied besteht in den Feldern der Sonden I und C3 am Taschengrund. Hier unterscheidet sich das Profil der *PfGSK-3* von demjenigen der *HsGSK-3 β* offenbar so deutlich, dass eine Interaktion mit den genannten Sonden im plasmodialen Enzym bevorzugt erscheint. Unterschiede nahe dem Gatekeeper im Bereich der Back Pocket werden vor allem durch die N1+-Sonde (oberhalb bei Ala106/Ala83 bzw. Val155/Leu130) und durch die DRY-Sonde (in Richtung des Lys108) detektiert.

Bemerkenswert ist, dass die für das Projekt besonders interessante Kristallstruktur *kubra* bis auf die Unterschiede der N:#-Sonde keine der anderen Differenzen stützt. Allerdings muss dabei beachtet werden, dass es sich bei dem kokristallisierten Liganden **31** gerade um einen *unselektiven* Inhibitor handelt. Möglicherweise werden durch dessen Bindung Unterschiede zwischen *HsGSK-3 β* und *PfGSK-3* nivelliert, die bei der Interaktion eines selektiven Liganden wichtig sind (z. B. Seitenkettenkonformationen oder Position des Glycine-rich Loops).

Die unübersichtlichen bzw. vielschichtigen oder wenigstens schwer interpretierbaren und teils sogar widersprüchlichen Daten sollten, wie im folgenden Kapitel beschrieben, mithilfe der Hauptkomponentenanalyse klarer strukturiert werden, um eventuell verborgene Muster zu erkennen.

2.4.2.3. Paarweise statistische Analyse mittels PCA

Es wurde für alle Homologie-Modell/Vorlage-Paare eine PCA durchgeführt, in der die Varianz der MIFs von *HsGSK-3 β* und *PfGSK-3* vergleichend untersucht wurde. Die beiden Score Plots für die Berechnung der Felder ohne (MOVE = 0) und mit (MOVE = 1) Berücksichtigung der Flexibilität von Seitenketten sind für das Beispiel *pfgsk_1j1b* und *1J1B* in Abbildung 2.20 dargestellt. Die weiteren Paare ähneln den abgebildeten Score und Loading Plots stark. Es fällt zunächst auf, dass zwischen den beiden Optionen keine grundsätzlichen Unterschiede festzustellen sind. Außerdem befinden sich die Sonden der *PfGSK-3* alle in der oberen Hälfte des Score Plots, diejenigen der *HsGSK-3 β* in der unteren. Da die x-Achse die erste Komponente wiedergibt, unterscheidet diese also nicht über Unterschiede zwischen den Proteinen sondern über Unterschiede zwischen den Sonden. Die zweite Komponente hingegen, die auf der y-Achse aufgetragen wird, trennt die Felder von Sonden der *PfGSK-3* und solchen der *HsGSK-3 β* .

Des Weiteren ist die Lage der beiden Sonden *DRY* und *C3* auffällig. Sie liegen eng beieinander und zudem nahe der x-Achse. Sie leisten also der PCA zufolge jeweils nur einen kleinen Beitrag zur Erklärung der Unterschiede zwischen *HsGSK-3 β* und *PfGSK-3*. Diese Beobachtung leitet sich auch aus den vergleichsweise kleinen absoluten Beträgen der lipophilen Sonden ab (vgl. Tab. 2.7), die der Wichtigkeit dieses Beitrags zur Protein-Ligand-Interaktion nicht gerecht werden. Andererseits deutet das direkte Übereinanderliegen der *DRY*-Sonde in den Score Plots in der hier aufgefundenen Ausprägung tatsächlich darauf hin, dass in den MIFs der Proteine keine zur Differenzierung verwertbare Information enthalten ist.

Klarer zu unterscheiden sind die Sonden *O*, *OC1*, *N1*, und *N1+*, den wichtigsten Beitrag zur Differenzierung zwischen *HsGSK-3 β* und *PfGSK-3* leisten *OH*, *N2* und *OH2*. Plasmodiale und humane MIFs sind bei diesen Sonden im Score Plot am weitesten voneinander entfernt. Allerdings gelingt diese Unterscheidung erst über die zweite Hauptkomponente. Diese erklärt nur 20-30 % der Varianz in der Matrix gegenüber 40-50 %, die von der ersten Komponente erklärt werden (Schwankungen abhängig vom Homologie-Modell/Vorlage-Paar). Diese Beobachtung wurde bereits bei verschiedenen PCA/GRID-Studien gemacht (z. B. [127, 171]).

Die Loading Plots sind in Abbildung 2.21 wiederum für *pfgsk_1j1b* und *1J1B* dargestellt. Ihre Interpretation fällt analog zu den Score Plots aus: Für die Selektivität interessante Punkte sind bei extremen Werten hinsichtlich der PC2 zu suchen, da diese humanes und plasmodiales Enzym diskriminiert. Da die Punkte im Loadings Plot Koordinaten mit Wechselwirkungsenergien verschiedener Sonden repräsentieren (vgl. Organisation der Matrix für die PCA auf S. 41), können diese für die Erklärung der Varianz wichtige Areale in GOLPE visualisiert werden, um entsprechende Bereiche zu untersuchen. Auf diese Weise sind in Abbildung 2.22 diejenigen Punkte im Loading Plot abgebildet, deren Werte der PC2 0.02 überschreiten bzw. -0.02 unterschreiten.

Dabei zeigt sich, dass es Übereinstimmungen mit den in Abschnitt 2.4.2.1 dargestellten visuellen Vergleichen von Homologie-Modell/Vorlage-Paaren gibt. So gibt es am Grund der Tasche

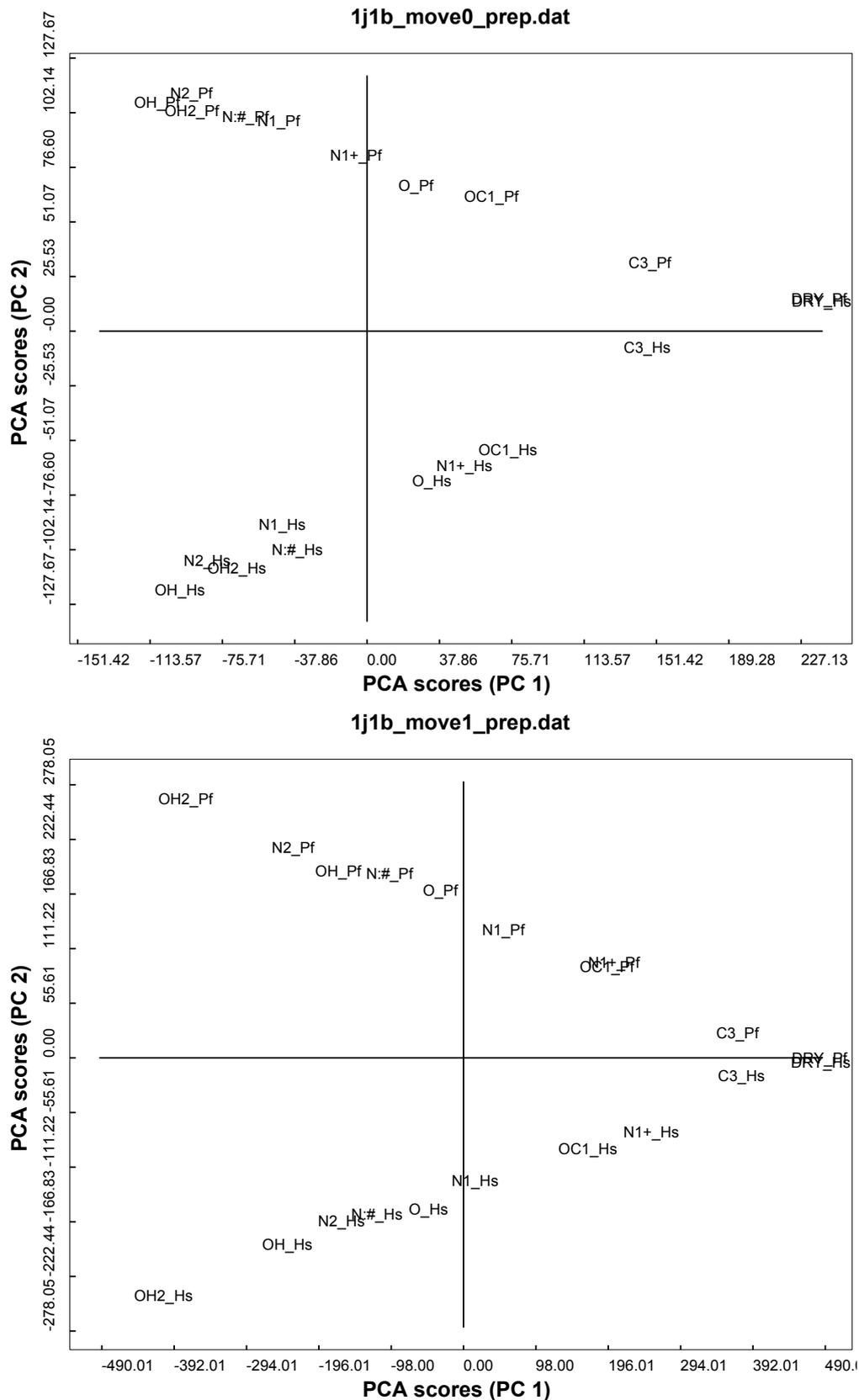


Abbildung 2.20. | Score Plots von pfgsk_1j1b und 1J1B mit MOVE=0 und MOVE=1. Die MIFs der humanen Struktur sind mit <Sonde>_Hs gekennzeichnet, diejenigen der plasmodialen mit <Sonde>_Pf.

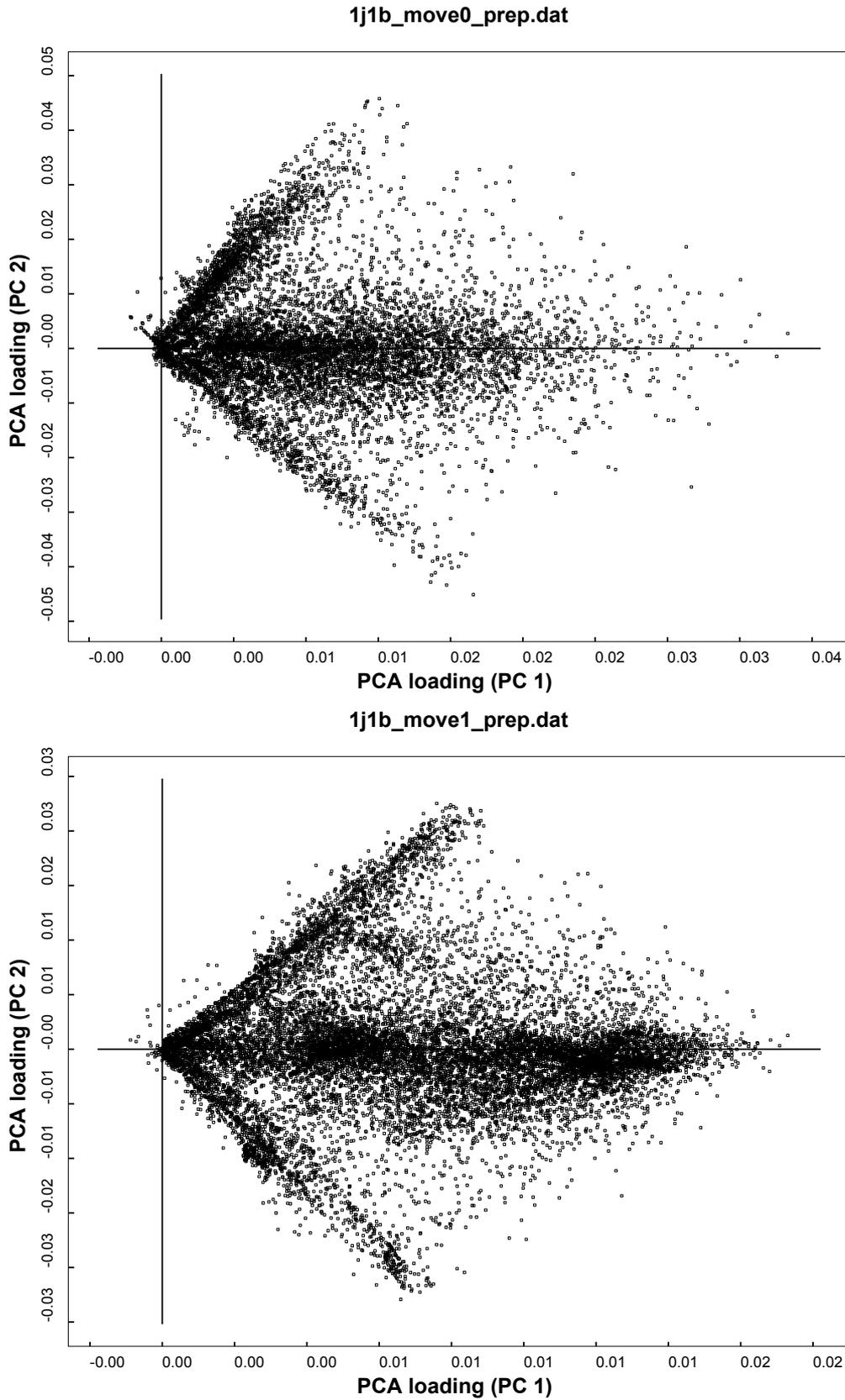


Abbildung 2.21. | Loading Plots von pfgsk_1j1b und 1J1B mit MOVE=0 und MOVE=1.

zwischen Thr163/Thr138, Leu213/Leu188 und Cys224/Cys199 einen in der Abbildung cyan dargestellten Bereich, der von den eingesetzten Sonden bevorzugte Interaktionsenergien mit der *Pf*GSK-3 anzeigt. Diese Beobachtung entspricht der visuellen Analyse des I- und C3-Feldes. Auch der favorisierte Bereich der Sonden im plasmodialen Enzym nahe dem Gatekeeper, der bei der Betrachtung der DRY-Sonde auffiel, wird von der PCA bestätigt. Im Bereich des Ausganges der Bindetasche nahe Lys166/Arg141 werden durch die PCA Bereiche lokalisiert, für die eine Interaktion der eingesetzten Sonden mit der *Hs*GSK-3 β favorisiert wird. Auch hier kann indirekt eine Beobachtung bestätigt werden, die bei der visuellen Analyse von verschiedenen Sonden bestätigt wurde: Während Felder der *Hs*GSK-3 β weiter außerhalb im Bereich der Front Pocket lokalisiert sind, reichen die Felder verschiedener Sonden (OH2, N2, N:#, OC1) weiter in die Tasche hinein. Das gelbe Feld in Abbildung 2.22 bestätigt die erstgenannte Beobachtung, die Differenzen nahe der Hinge Region als favorisierte Bereiche der *Pf*GSK-3 hingegen sind in der Projektion der Loadings nicht zu erkennen.

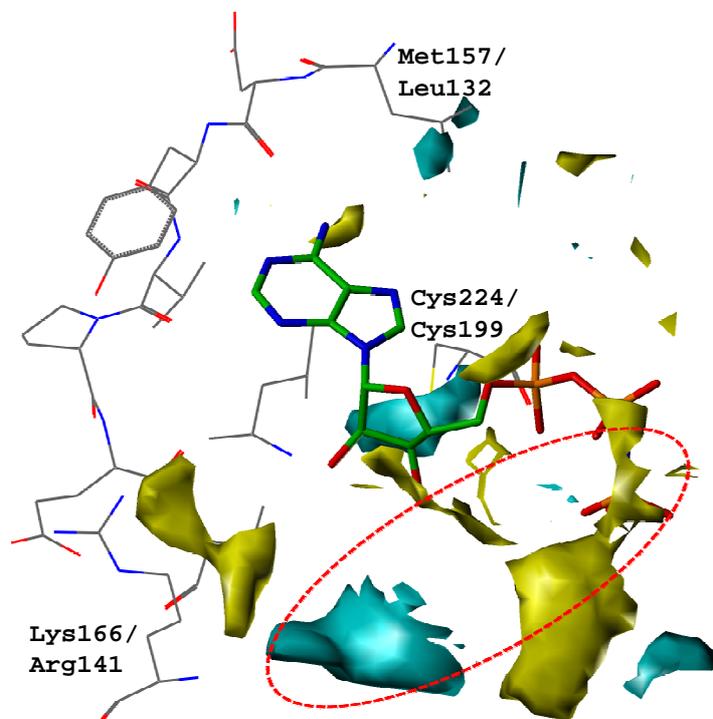


Abbildung 2.22. | Projizierte Loadings in Entsprechung zu Abbildung 2.21. Zur Orientierung ist 1J1B mit ANP dargestellt, die Felder sind entsprechend ihren Werten für die PCA2 gefärbt (>0.02 gelb \Rightarrow favorisierter Bereich der *Hs*GSK-3 β ; <-0.02 cyan \Rightarrow favorisierter Bereich der *Pf*GSK-3). Zusätzlich zu den im Text genannten Bereichen nahe den bezeichneten Aminosäuren gibt es Abweichungen im Bereich der Phosphatregion und weiter zum Taschenausgang hin (rote Ellipse), die beim visuellen Vergleich nicht auffällig erschienen.

Das in Abbildung 2.22 gelb gefärbte Areal im roten Kreis nahe dem γ -Phosphat vom ANP direkt beim Lys208/Lys183 (Aminosäuren nicht abgebildet) wurde in den vorangegangenen Untersuchungen nicht als Anhaltspunkt für Selektivität für die *Hs*GSK-3 β gefunden. Auch das in Richtung des Lys166/Arg141 positionierte cyanfarbene Feld beim Gln210/Gln185 (Aminosäuren sind wie Lys208/Lys183 der Übersichtlichkeit halber nicht dargestellt) wurde beim visuellen Vergleich nicht als Selektivitätsmerkmal identifiziert. Beide Felder sind in der Abbildung rot eingekreist und liegen relativ weit von der Kernregion der Bindetasche entfernt, wenn diese durch den Purinkörper des Adenin definiert wird (10.1 \AA vom C_D des Gln210 in 1J1B zum nächstgelegenen Stickstoff des ANP (N9), 10.7 \AA vom Stickstoff des Lys183

zum selben Stickstoff). Bei den Überlegungen, ob diese Region an einer Bindung beteiligt sein kann, muss diese Entfernung in Betracht gezogen werden.

Die Interpretation der Loadings wird dadurch erschwert, dass wie in der Einleitung beschrieben ein Nachteil der PCA gegenüber der CPCA darin liegt, dass die Felder zwar lokalisiert werden können, durch die Konstruktion der Matrix bedingt jedoch einer Koordinate die Werte mehrerer Sonden zugeordnet sind. Dadurch ist es nicht möglich festzustellen, durch welche Sonde die Unterscheidung im Einzelnen zustande gekommen ist.

Die Aussagekraft der PCA wird dadurch bestärkt, dass PC1 und PC2 gemeinsam die in Tabelle 2.23(a) dargestellten hohen Prozentsätze der Varianz erklären. Dabei fällt auf, dass es einen deutlichen Unterschied zwischen den Werten ohne und mit Berücksichtigung der Seitenkettenflexibilität bei Berechnung der MIFs gibt: Im zweiten Fall liegt der Wert höher als im ersten. Die Verteilung ist in Abbildung 2.23 dargestellt.¹³ Der hohe Wert bedeutet, dass die Projektion auf wenige latente Variablen besonders gut möglich ist. Dass dies unter Berücksichtigung der Seitenkettenflexibilität noch besser gelingt also ohne, spricht für die Verwendung dieser Option in Kombination mit der PCA, da die Reduktion auf wenige aussagekräftige Variablen gerade eines der vorrangigen Ziele des Verfahrens ist.

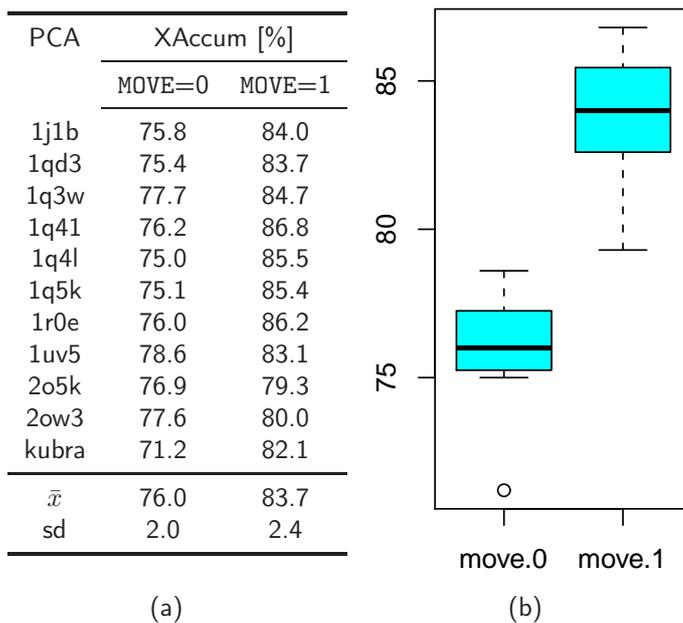


Abbildung 2.23. | (a) Gegenüberstellung des Anteils der Varianz, der durch erste und zweite Hauptkomponente (PC1/PC2) gemeinsam erklärt wird (XAccum). Unter Berücksichtigung der Seitenkettenflexibilität liegt der Wert deutlich über dem Anteil, der mit MOVE = 0 erfasst wird (signifikanter Unterschied mit $p \ll 0.001$). Der Sachverhalt wird durch den Boxplot (b) grafisch verdeutlicht.

¹³Der dargestellte Boxplot gibt einen Eindruck von der Unterschiedlichkeit durch Darstellung von Median und Quartilen. Der als Ausreißer gekennzeichnete Wert bei MOVE = 0 weicht mehr als das 1.5-fache des Interquartilsabstandes ab.

2.4.2.4. Statistische Analyse mittels CPCA

Die entscheidende Darstellung zur Abgrenzung verschiedener Proteine gegeneinander im Rahmen einer CPCA-Studie ist der Score Plot. Hier zeigen sich Gruppierungen von ähnlichen Objekten innerhalb der Matrix X , also die Nähe der Zielmoleküle zueinander hinsichtlich der berechneten MIFs. Abbildung 2.24 auf Seite 72 bildet die Ergebnisse der CPCA mit den Sonden I, 0C1, N2, N1+, 0H, N:#, 0H2, C3 und 0 ab (Auswahl vgl. S. 46).

Dabei fällt auf den ersten Blick die paarweise Verteilung der Scores von Modell und Vorlage auf. Offenbar sind die MIFs der Modelle so nahe an denen der Vorlage, dass der Unterschied zwischen den verschiedenen Kristallstrukturen größer ist, als der zwischen dem humanen und dem plasmodialen Enzym. Diese Beobachtung wurde auch in einem verwandten Fall von AFZELIUS *et al.* [126] beschrieben. Untersucht wurden die Unterschiede zwischen MIFs verschiedener Isoformen des Cytochrom P450. Für diese lagen einerseits Kristallstrukturen vor, zusätzlich wurden Homologie-Modelle erstellt und vergleichend mittels CPCA untersucht. Die Autoren führten ein Cross Homology Modelling durch, bei dem die Vorlagen für die Isoformen im Rahmen der Entwicklung des Modells ausgetauscht wurden, um die Auswirkungen auf das resultierende Modell zu untersuchen. Sie kommen zu dem Schluss, dass die Eigenschaften der Modelle stärker von der dreidimensionalen Vorlage bestimmt wird, die dem Modell zugrunde liegt, als von der Sequenz, die zum eigentlichen Zielmolekül führen soll. Modelle, die mit einer weiter entfernten Isoform als dreidimensionale Vorlage konstruiert wurden, lagen bei der Analyse mittels CPCA näher an der Vorlage als an der in diesem Fall bekannten experimentellen Struktur des eigentlichen Zielproteins.

In der vorliegenden Arbeit wurden verschiedene Parameter des Verfahrens variiert, um trotz der Nähe von Modell und Vorlage zu einer besseren Auftrennung zu gelangen. Die Auswahl der Sonden wurde variiert, die Auflösung bei der Berechnung der MIFs sowie der in der Analyse berücksichtigte Bereich und auch die Auswahl der Proteine, die in die Analyse aufgenommen wurden.

So bildet sich beispielsweise im Score Plot (Abb. 2.24) eine Untergruppierung von 1Q41, 1J1B und 1Q5K, für die bei Betrachtung der ersten beiden Hauptkomponenten gegeneinander eine Trennung zwischen humanen und plasmodialen Strukturen gelingt, auch wenn die beiden Gruppen nahe beieinander liegen. Führt man die Analyse jedoch nur mit den drei genannten Paaren durch, ergibt sich eine Verteilung in Entsprechung zu Abbildung 2.24, bei der Vorlage und Modell wiederum beieinander liegen. Die Gruppierung bei der Analyse mit allen Proteinen ist also als Folge der Abgrenzung gegenüber den übrigen Strukturen zu interpretieren, weniger als Detektion latenter Variablen, die *Pf*GSK-3 und *Hs*GSK-3 β diskriminieren. Eine Interpretation der CPCA im Sinne der Projektion von Feldern in die Bindetasche schien unter diesen Umständen nicht sinnvoll und wurde nicht vorgenommen.

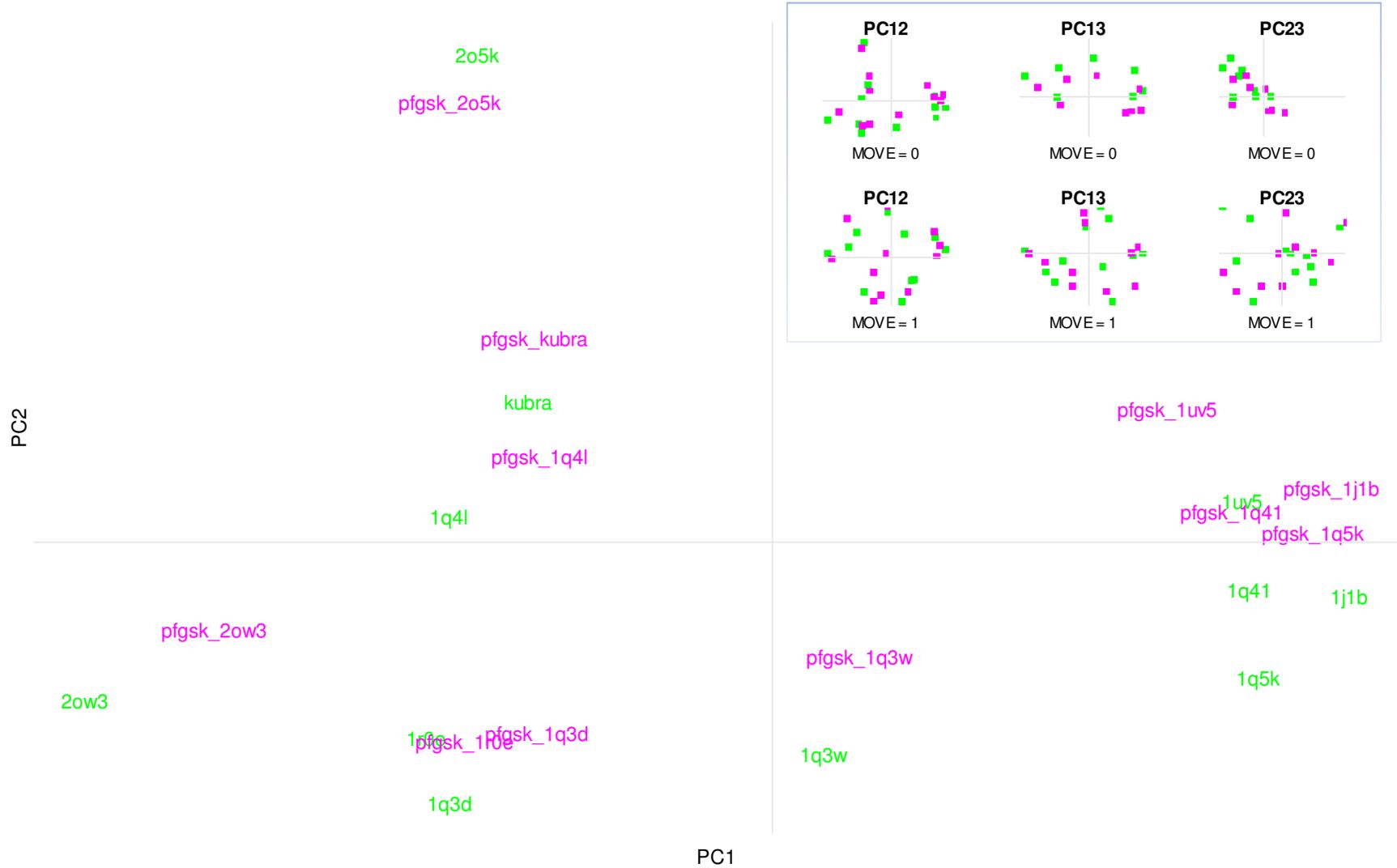


Abbildung 2.24. | Score Plot der CPCA (*HsGSK-3β*: grün, *PfgGSK-3*: magenta). In der großen Darstellung ist das Ergebnis bei Berechnung der MIFs mit MOVE = 0 abgebildet, in dem kleinen Kasten sind jeweils die Ergebnisse mit MOVE = 0 und 1 für alle Kombinationen der ersten drei Hauptkomponenten gegenübergestellt. Es ist zu erkennen, dass mit keiner der Kombinationen eine Auftrennung in humane und plasmodiale Strukturen möglich ist.

Experimentelle Daten (Selektivität von Inhibitoren bzw. Substratspezifität) zeigen dennoch, dass sowohl in der geschilderten Literatur als auch im hier bearbeiteten Projekt de facto Unterschiede zwischen den untersuchten Enzymen existieren. Für diesen widersprüchlichen Befund sind unterschiedliche Erklärungen plausibel:

- Die Modelle bilden nicht die Verhältnisse im plasmodialen Protein ab. Vorstellbar ist, dass durch die starke Orientierung an der Vorlage wichtige Eigenschaften des modellierten Proteins nicht berücksichtigt werden. Gerade bei sehr flexiblen Bereichen ist die Ausrichtung an der Vorlage möglicherweise irreführend, da eine bestimmte Geometrie unter vielen möglichen gewählt wird, nur weil sie der Vorlage ähnelt.
- Die MIFs sind keine geeignete Methode, die Unterschiede zwischen den Spezies zu erklären.

2.4.2.5. Fazit MIFs

Bei der visuellen Analyse der Felder wurden die in Abschnitt 2.4.2.2 zusammengefassten Unterschiede zwischen humaner und plasmodialer GSK-3 bei verschiedenen Homologie-Modellen und Vorlagen entdeckt. Diese Unterschiede konnten in der paarweisen PCA von Modell und Vorlage in ihrer Lokalisation bestätigt werden. In der übergreifenden Analyse der Strukturen per CPCA hingegen konnte kein Modell entwickelt werden, das in der Lage war, zwischen humanen und plasmodialen Strukturen zu unterscheiden. Mögliche Ursachen für diesen Befund sind auf Seiten des Verfahrens oder aber in den Ausgangsstrukturen zu suchen.

Die Ausgangsstrukturen an sich können nach der im ersten Teil des Kapitels beschriebenen Validierung der Homologie-Modelle als gute dreidimensionale Strukturen gelten. Es ist unabhängig davon möglich, dass ihre Flexibilität trotz der Auswahl mehrerer Vorlagen und der Berücksichtigung von Flexibilität von Seitenketten im Rahmen der GRID-Berechnungen nicht ausreichend repräsentiert wird.

Mit der Methode der CPCA war es nicht möglich, ein mathematisches Modell zu entwickeln, das systematische Unterschiede zwischen *HsGSK-3 β* und *PfGSK-3* aufzeigte. Trotzdem scheint es sinnvoll, die entwickelten Homologie-Modelle als Grundlage für weiterführende Untersuchungen zu verwenden und dabei die Ergebnisse der Untersuchungen mit MIFs als Anhaltspunkte zur Erklärung der Unterschiede zwischen *HsGSK-3 β* und *PfGSK-3* zu berücksichtigen.

2.5. Experimenteller Teil: Computational Details

2.5.1. Homology Modelling

Auswahl der Vorlagen Die BLAST-Suche wurde mit den Sequenzen der PDB als Suchraum und den Standardparametern (BLOSUM62-Matrix, `Gap costs`: 11 existence, 1 extension) durchgeführt.

Sequenzalignment Das paarweise Alignment für die Modelle wurde mit dem in MODELLER (Version: 9v2) implementierten Algorithmus nach Needleman und Wunsch [137] berechnet (MODELLER-Routine `align2d.py`), für Abbildung 2.5 wurde ClustalW bzw. ClustalX verwendet [180, 181]. Werte der Identität und Ähnlichkeit wurden mit EMBOSS [182] ermittelt (für *HsGSK-3 β* : P49841; für *PfGSK-3*: pfc0525c bzw. beide gekürzt nach orientierenden Alignments auf Sequenzabschnitt Ser58 bis Phe413 der *PfGSK-3*, vgl. S. 47; `gap penalty`: 25.0, `extend penalty`: 0.5, Matrix: EBLOSUM62).

Homology Modelling und Evaluierung Die Routine `automodel` von MODELLER führt alle notwendigen Schritte eines Comparative Modellings durch, sie wurde für das Generieren der Modelle verwendet. Zusätzlich wurde für die unter Berücksichtigung der Liganden konstruierten Modelle `automodel.nonstd_restraints` aktiviert. Dieses Modul legt zusätzliche Einschränkung beim Erstellen der Modelle fest. Es werden alle C_{α} -Atome des Proteins fixiert, deren Entfernung zu HETATM-Atomen (in die Bindetasche projiziertem Liganden) unter 10 Å liegen, außerdem werden direkte Kontakte des Proteins zum Liganden aus der Vorlage übernommen, indem auch Atome von Seitenketten des Proteins fixiert werden, wenn sie im Modell vorhanden sind und ihre Entfernung 2.3 Å oder weniger beträgt.

Die zur Evaluierung verwendeten Versionsnummern der Programme waren ProSa 2003 v4.0 (`winsize` = 25 in Abb. 2.10) und PROCHECK v3.5.4 (für die Homologie-Modelle gibt es keine Auflösung, es wurde für alle Modelle 2 Å eingesetzt).

Topologie der Bindetasche Das in Abbildung 2.11 dargestellte Oberflächenprofil wurde mit dem MOPAC-Modul in Sybyl [183] berechnet (Fast Connolly-Oberfläche mit `dot density` = 6 und `probe radius` = 1.4 Å).

Moleküldynamik Die MD-Simulationen wurden mit GROMACS 3.2.1 durchgeführt, es wurde das OPLSAA-Kraftfeld [184, 185] in Kombination mit dem spc- (simple point charge) Wassermodell verwendet. Es wurde eine dodekaeder-förmige Box definiert (9 Å über das Protein hinaus, Größe ca. 510 nm³, ca. 15 000 Wassermoleküle), in der das System nach Ladungsausgleich durch Cl⁻/Na⁺ zunächst äquilibriert wurde. Dies wurde durch initiale Steepest-Descent-Minimierung (100 Schritte, initiale Schrittweite 0.01 nm, Konvergenzkriterium 1000 kJ/mol * nm)

Tabelle 2.9. | Koordinaten der berechneten GRID-Felder. Die Angabe der Koordinaten für das zu errechnende Raster erfolgt in GRID in Form zweier Punkte: Die eine Koordinate (unten, vorne, links) beschreibt in Kombination mit der zweiten (oben, hinten, rechts) einen Quader, in den die Flächen für das Raster geschichtet werden.

	unten	oben
x	-19	14.5
y	-1.5	14.5
z	-9	18

und anschließende Position-Restraint-Simulation für 30 ps bei fixiertem Protein erreicht. Die eigentliche Simulation erfolgte für 2 ns. Um eine konstante Temperatur von etwa 300 K zu gewährleisten, wurde der Berendsen-Thermostat [186] eingesetzt, zur Lösung des Problems starker Bindungsschwankungen LINear Constraint Solver LINCS [187]. Die Distanz zum periodischen wiederkehrenden Bild der Proteine war zu jedem Zeitpunkt der Simulationen $\gg 10 \text{ \AA}$, so dass ein nicht erwünschter Kontakt über die Grenze der Box hinweg ausgeschlossen werden konnte.

2.5.2. Berechnung Molekularer Interaktionsfelder

Verwendete Versionen der Programme: GRID 22 bzw. GOLPE 4.6.0.

Überlagerung der Proteine Das Alignment wurden mit Sybyl 8.1 [188] durchgeführt. Es wurden das BIOPOLYMER-Modul verwendet (Überlagerung auf alle C_α -Atome). Für die visuelle Analyse wurden dabei für alle Proteine 1J1B als Referenz eingesetzt. Bei der visuellen Analyse konnten die kleinen Verschiebungen, die bei den Homologie-Modell/Vorlage-Paaren beobachtet wurden, bei der Betrachtung berücksichtigt werden. Für die statistische Analyse wurde das Alignment leicht variiert: Nach Überlagerung der Vorlagen auf 1J1B wurden die Homologie-Modelle paarweise auf ihre Vorlage bezogen.

Einstellungen von GRID für die Berechnungen zu 2.4.2.1 Für die visuelle Analyse wurden die Felder mit hoher Auflösung gerechnet ($NPLA = 4$, entspricht einer Fläche alle 0.25 \AA). Es wurde grundsätzlich unter Berücksichtigung der Flexibilität von Seitenketten gerechnet ($MOVE = 1$). Der berechnete Bereich wurde manuell an die überlagerten Proteine angepasst und durch die in Tabelle 2.9 aufgeführten Koordinaten definiert ($33.5 * 16 * 27 \text{ \AA}$). Für alle weiteren Einstellungen wurden Standardwerte gesetzt.

Neutralisation der Proteinstrukturen Es wurden für die dargestellten Ergebnisse in Abschnitt 2.4.2.1 und 2.4.2.3 keine Gegenionen zum Ladungsausgleich in das System eingeführt. Hintergrund war die Überlegung, dass durch das Einführen einer unterschiedlichen Anzahl von Gegenionen bei den verschiedenen Röntgenstrukturen der *HsGSK-3 β* Unterschiede künstlich generiert werden, die nur darin begründet sind, dass zugrunde liegende Röntgenstrukturen verschieden gut aufgelöst sind und dadurch in einigen der Strukturen Seitenketten oder ganze

Reste fehlen – die Sequenz der *HsGSK-3 β* und damit die Abfolge von Aminosäuren im betrachteten Bereich ist schließlich identisch. Das Hinzufügen der Gegenionen würde dazu führen, dass identische Strukturen unterschiedlich behandelt werden, da schlecht aufgelöste Bereiche in einer Struktur nicht erfasst werden, in der nächsten dann aber zum Einsetzen eines Gegenions führten.

Zudem wurden in Vorversuchen zu Abschnitt 2.4.2.4 vergleichende Untersuchungen mit und ohne Neutralisation des Systems unternommen. Diese zeigten, dass ein Unterschied zwischen den Ergebnissen mit und ohne Neutralisation praktisch nicht vorhanden war. Für die Berechnungen der GRID-Felder für Abschnitt 2.4.2.4 wurde dennoch eine Neutralisation mit den in GRID integrierten Hilfsprogrammen MINIM und FILMAP vorgenommen, um Vergleichbarkeit zu den einleitend zitierten Studien zu gewährleisten, bei denen dieses Vorgehen teilweise gewählt wurde.

Einstellungen von GRID für die Berechnungen zu 2.4.2.3 und 2.4.2.4 Während beim visuellen Vergleich eine höhere Auflösung die Betrachtung erleichtert, kann bei einer multivariaten statistischen Analyse eine zu hohe Anzahl an nicht zwangsläufig informativen Datenpunkten die Auswertung deutlich erschweren oder gar unmöglich machen. Es wurden deshalb 0.5 Å (NPLA = 2) für die paarweise PCA und 1 Å für die CPCA (NPLA = 1, ausdrückliche Empfehlung in [177]) gewählt. Die Flexibilität der Seitenketten wurde vergleichend freigegeben bzw. nicht freigegeben (MOVE = 1 bzw. 0).

Vorbereitung der Daten auf die (C)PCA in GOLPE Für die Berechnung der (hierarchischen) Hauptkomponentenanalyse wurden die im Folgenden aufgeführten Vorbereitungen des Datensatzes vorgenommen.

- Begrenzung des Maximums auf 0 kcal/mol, um ausschließlich anziehende Interaktionen (negative Energien) auszuwerten.
- Ausschluss von Variablen (Koordinaten) mit einer Standardabweichung von weniger als 0.02 kcal/mol zur Reduktion redundanter Information (Verbesserung des Signal-Rausch-Verhältnisses).
- Auswahl der Region um den Liganden: Da GRID immer Felder in Form eines Quaders berechnet, gehen viele Werte mit in die Berechnung ein, die räumlich weit von dem Liganden entfernt sind oder innerhalb des Proteins liegen. In GOLPE besteht die Möglichkeit, eine Region um ein Molekül auszuschneiden, durch das die Bindetasche definiert wird. Dazu wurde ANP in 1J1B gewählt, Datenpunkte im Umkreis von 3 Å werden berücksichtigt.
- Bei den paarweise berechneten PCAs nach 2.4.2.3: Kein automatisches Skalieren der Daten. Da die Werte alle Ergebnisse der Berechnung von MIFs darstellen, also aus einer Quelle stammen und zudem in der gleichen Einheit [kcal/mol] vorliegen, ist eine Skalierung

nicht sinnvoll und kann sogar zu einer Verzerrung der Verhältnisse durch Überbewertung unwichtiger Regionen führen.

- Bei der CPCA nach 2.4.2.4: Hier wurde eine blockweise Skalierung durchgeführt (Block Unscaled Weight, BUW), die die Gleichgewichtung der Blöcke zueinander erlaubt, ohne Varianzen innerhalb des einzelnen Blocks zu verändern (nach Organisation der Daten für die CPCA beschreibt ein Block die Wechselwirkungen einer Sonde, vgl. S. 43).

Zudem wurde in den dargestellten Ergebnissen der Radius um ANP als Ligand auf 4 Å erweitert. In den Versuchen zur Optimierung der CPCA wurden sowohl größere Bereiche (Radius 6 Å) als auch kleinere Ausschnitte (Purinkörper des ANP plus 2 Å; **36** und **28** plus 2.5 Å) gewählt.

3. Der Bindungsmodus von Thieno[2,3-*b*]pyridinen als Inhibitoren der *Pf*GSK-3

3.1. Grundlagen des Dockings

Die IUPAC bezeichnet Docking im Rahmen der Definition von *Docking Studies* als computer-gestützte Erkundung möglicher Bindungsmodi.

Docking studies are computational techniques for the exploration of the possible binding modes of a substrate to a given receptor, enzyme or other binding site.

Es wird also virtuell nach Möglichkeiten gesucht, auf welche Weise ein kleines Molekül (der Ligand) an einen Rezeptor, ein Enzym oder eine Bindetasche im Allgemeinen binden kann [189]. Das Verfahren bildet mittlerweile einen festen Bestandteil im Prozess der Wirkstoffsuche. Bei genauer Betrachtung lässt sich diese Aufgabe in zwei Teilprobleme aufspalten. Zunächst müssen Protein-Ligand-Komplexe erzeugt werden und diese müssen dann bewertet werden, um den richtigen Kandidaten als Lösung auszuwählen. Die beiden Schritte werden in der Literatur häufig als das Docking- und das Scoring-Problem bezeichnet [190]. Grundsätzliche Prinzipien zu deren Lösung werden in den beiden folgenden Abschnitten erläutert.

3.1.1. Erkundung möglicher Bindungsmodi: Docking

Die Suche läuft ab, indem *in silico* (am Computer, in Entsprechung zu *in vitro* bzw. *in vivo*) die betreffenden Liganden in eine Bindetasche eingepasst werden. Der erste Algorithmus, der dazu eingesetzt wurde und dann der ersten Version des Programmes DOCK zugrunde lag, wurde bereits 1982 von KUNTZ *et al.* veröffentlicht [191]. Die heute eingesetzten Algorithmen lassen sich z. B. nach RESTER [192] im Wesentlichen in vier Kategorien einteilen, die nicht als exklusiv zu verstehen sind:

1. **fragmentbasiert**: Der Ligand wird inkrementell in der Bindetasche aufgebaut. Er wird also in Fragmente aufgespalten, aus denen er in der Bindetasche sukzessive wieder zusammengesetzt wird (z. B. FlexX [193]).

2. **genetisch:** Es werden Prinzipien der biologischen Evolution in den Algorithmus integriert. Eigenschaften des Liganden (Torsionswinkel, Interaktionen mit dem Protein, etc.) werden als Gene verstanden, die zu einem bestimmten Phänotyp führen, dessen Fitness (Abschätzung der Interaktionsenergie) berechnet werden kann. Sie entscheidet darüber, welche Individuen der Population ihre Merkmale in die nächste Generation weitergeben (Selektion). Neben dem Weitergeben vorteilhafter Eigenschaften der Elterngeneration in die Folgegenerationen entstehen neue Gene (und damit Veränderungen im Phänotyp) durch Cross Over und Mutation (z. B. Autodock [194]).
3. **shape based:** Die Suche wird maßgeblich von der dreidimensionalen Gestalt der Bindetasche geleitet, deren Komplementarität mit dem Liganden die generierten Protein-Ligand-Komplexe bestimmt (z. B. FRED [195]).
4. **kraftfeldbasiert:** Durch Miteinbeziehen von Moleküldynamik- oder Monte Carlo-Simulationen wird die Flexibilität des Proteins mit einbezogen (z. B. Glide [196, 197]).

Es konnte bereits früh nachgewiesen werden [198], dass das Docking besser gelingt, wenn der Ligand flexibel behandelt wird. Diese Anforderung wird heute von praktisch allen gängigen Ansätzen erfüllt. Die Flexibilität des Rezeptors hingegen wurde zunächst nicht berücksichtigt. 2008 stellten COZZINI *et al.* fest, dass dies im Allgemeinen im Rahmen des strukturbasierten Designs immer noch der Fall ist. Sie zeigten gleichzeitig die Entwicklungen und Möglichkeiten auf, die es mittlerweile auf dem Gebiet des Dockings unter Berücksichtigung der Flexibilität seitens des Rezeptors gibt [199]. Das Problem ist allerdings heute noch eine Herausforderung, wie B-RAO *et al.* 2009 darlegten [200].

Die Anwendungen des Dockings sind vielfältig. LEACH *et al.* sehen in einem Übersichtsartikel [201] zu Möglichkeiten und Problemen des Verfahrens die derzeitige Hauptanwendung im Zusammenhang mit

- der Vorhersage des Bindungsmodus von aktiven Liganden,
- der Identifikation neuer Liganden bei virtuellen Screenings oder
- der Vorhersage von Bindungsaffinitäten bei nahe verwandten aktiven Substanzen einer Serie.

Dabei sehen die Autoren die Erfolge vor allem im ersten Punkt und geben an, dass eine Vorhersage kristallstrukturnaher Bindungsmodi mit einem geeigneten Dockingprogramm in etwa 70% der Fälle möglich sei. Diese Zahl erscheint hoch, bedenkt man, dass z. B. TUCCINARDI *et al.* in einer Studie mit 700 hoch aufgelösten Kristallstrukturen von Proteinkinasen und 17 Dockingprogrammen zum Schluss kommen, dass die Wahrscheinlichkeit, einen unbekanntem Liganden erfolgreich zu docken, nur bei etwa 30-37% liegt [202].

Diese Diskrepanz macht ein bereits angedeutetes Problem beim Docking deutlich. Es muss nicht nur möglich sein, die korrekte Geometrie des Protein-Ligand-Komplexes¹ zu generieren (das Docking-Problem). Erfolgreich ist das Docking erst, wenn es gelingt, durch die Reihung der in der Regel zahlreichen möglichen Lösungen eine „richtige“, also kristallstrukturnahe, zu identifizieren – das Scoring-Problem.

3.1.2. Bewertung von Dockingergebnissen: Scoring

Die Bewertung der erhaltenen Posen ist ein zentrales Problem im Rahmen des Dockings. Prinzipiell ist es das Ziel, aus einer vorgegebenen Bindungsgeometrie die Bindungsaffinität abzuschätzen. Dabei kann weiter unterschieden werden zwischen dem Fall, dass es sich um ein einziges Molekül handelt (Unterscheiden mehrerer alternativer Lösungen für einen Liganden: Ranking), und dem noch schwierigeren Fall, dass verschiedene Moleküle verglichen werden (Erstellen einer Abstufung z. B. im Rahmen eines virtuellen Screenings: Priorisierung).

Das grundsätzliche Problem bei der Berechnung der Bindungsaffinität aus einem Protein-Ligand-Komplex besteht darin, dass die einzigen vom thermodynamischen Standpunkt her korrekten Methoden zur Vorhersage der relativen freien Energie hierzu explizite Solvensmoleküle und die Flexibilität von Ligand und Rezeptor berücksichtigen müssen [162]. Wird eine Gruppe von Molekülen (ein Ensemble aus Ligand und Protein gemeinsam mit Solvensmolekülen) betrachtet, lässt sich aus deren Verteilung über mögliche Zustände und die Aufsummierung der entsprechenden Energiebeiträge eine Zustandssumme bilden, aus der die freie Enthalpie ΔG abgeschätzt werden kann [203]. Auf diese Weise kann der entropische Anteil der freien Enthalpie erfasst werden.

Ein solches Vorgehen ist jedoch so rechenintensiv, dass es realistischerweise nicht im Rahmen eines Dockings durchgeführt werden kann, wo schon für einen Liganden häufig dreistellige Anzahlen an unterschiedlichen Posen erzeugt und gegeneinander abgewogen werden müssen. Noch weniger ist es denkbar für die Suche in Datenbanken, bei denen häufig eine 6- oder 7-stellige Anzahl an Molekülen untersucht wird.

Diese Problematik führte zur Entwicklung vereinfachender Ansätze für das Scoring, die die Anforderungen an Geschwindigkeit erfüllen und häufig in die folgenden Kategorien eingeordnet werden (z. B. REXTER [192]):

1. **kraftfeldbasiert**: Anwendung klassischer Molekülmechanik.
2. **empirisch**: Entsprechung zu 1. unter Einbeziehung experimenteller Daten zur Anpassung an einen Trainingsatz.

¹Die Position des Liganden wird im Englischen in diesem Zusammenhang als *pose* bezeichnet. Der Begriff wird wegen seiner Geläufigkeit als Terminus technicus auch in dieser Arbeit verwendet.

3. **wissensbasiert:** Ableitung von Potentialen aus einer Wahrscheinlichkeitsverteilung bestimmter Merkmale (in der Regel paarweise Entfernungen von schweren Atomen) eines bekannten Datensatzes von Protein-Ligand-Komplexen.

Vereinfachende Annahmen führen dazu, dass das Ergebnis eine Annäherung an die Bindungsaffinität darstellt und nicht als absoluter Wert missverstanden werden darf. Es handelt sich vielmehr um eine für die jeweilige Scoringfunktion spezifische Abstufung der Stärke der Interaktion des Liganden mit dem Protein. Schon der Vergleich zwischen verschiedenen Verbindungen kann problematisch werden, da Beiträge für Interaktionen aufaddiert werden (siehe Gleichung 3.1 auf S. 97 als Beispiel), so dass große Liganden unter Umständen automatisch höhere Werte erreichen als kleinere, nur weil sie aufgrund ihrer Größe eine höhere Anzahl möglicher Interaktionspunkte haben.

Die Entwicklung von Scoringfunktionen ist Gegenstand aktueller Forschung und es gibt derzeit eine schwer überschaubare Anzahl an Beispielen für jede der genannten Kategorien. Sie zeichnen sich durch Vorteile in Teilaspekten aus, entweder hinsichtlich der Möglichkeiten bezüglich Ranking bzw. Priorisierung, oder aber hinsichtlich des Einsatzes für bestimmte Strukturen. Insgesamt kann das Scoring-Problem derzeit jedoch nicht als gelöst angesehen werden. Es werden immer wieder Studien veröffentlicht, die verschiedene Verfahren vergleichen und zusammenfassend zu dem Schluss kommen, dass es derzeit keine Überlegenheit einer einzigen Scoringfunktion in jeder Hinsicht gibt (z. B. [204–206]).

3.1.3. Der rmsd als Bewertungsparameter

Vielfach wird bei der Beurteilung der Qualität eines Dockings der rmsd (engl.: root-mean-square deviation, mittlere quadratische Abweichung in der Regel der schweren Atome) des Liganden zu einer experimentell aufgeklärten Struktur als Maß herangezogen. Dabei werden Lösungen mit einem rmsd von weniger als 2 Å im Allgemeinen als gut angesehen. Der rmsd als Kennzahl für die Qualität des Dockings ist mit mehreren Problemen behaftet, die z. B. von YUSUF *et al.* dargelegt werden [207]. Zum einen wird nur der Ligand betrachtet, die Umgebung – das Protein, mit dem der Ligand interagiert – geht nicht in den rmsd mit ein.

So kann bei einem Crossdocking die Pose exakt der Position des Liganden in der Kristallstruktur entsprechen, also einen sehr kleinen rmsd aufweisen. Wenn die relative Lage zum Protein jedoch so ungünstig ist, dass entscheidende Interaktionen nicht ausgebildet werden können oder Kollisionen mit dem Protein vorliegen, ist die Lösung dennoch als schlecht einzuschätzen und der kleine Wert täuscht ein gutes Ergebnis vor. Dieser Fall kann eintreten, wenn sich das im Docking eingesetzte Protein deutlich von demjenigen unterscheidet, mit dem der Ligand kokristallisiert wurde.

Andererseits kann ein Ligand in der für ihn berechneten Pose entscheidende Interaktionen einer Teilstruktur gut ausbilden, ein flexibler Rest jedoch anders vorhergesagt werden, als er

in der Kristallstruktur vorliegt. Die betreffende Lösung wäre eine gute, der rmsd jedoch unter Umständen deutlich größer als 2 Å. Dieser Fall ist insbesondere bei einem Redocking vorstellbar, der rmsd wäre dann irreführend hoch.

Das zweite grundsätzliche Problem ist, dass beim Vergleich mit der Kristallstruktur wiederum der Bezug zu einem Modell hergestellt wird, da die Koordinaten einer Kristallstruktur nur die Interpretation einer Elektronendichtekarte darstellen. Es wird also durch den rmsd die Güte der Anpassung an dieses Modell beschrieben, nicht an die tatsächlich vorliegende Situation.

Aus diesen Gründen schlagen YUSUF *et al.* den real space R-factor (RSR) als Kennzahl für die Qualität einer Dockinglösung vor [207]. Ein anderer Versuch, die geschilderten Probleme des rmsd zu relativieren, wurde von KROEMER *et al.* mit dem interaction-based accuracy classification scheme (IBAC) vorgeschlagen [208]. Die genannten Alternativen zum rmsd konnten sich allerdings ebenso wenig durchsetzen wie andere Versuche – der rmsd wird nach wie vor de facto als Standardkennzahl in der Literatur verwendet. Aus Gründen der Vergleichbarkeit wurde er auch in dieser Arbeit als Kennzahl herangezogen.

3.2. Material und Methoden: Verbindungen und Software

3.2.1. Inhibitoren der PfGSK-3

In Tabelle B.3 auf Seite 234 im Anhang sind alle Verbindungen aufgeführt, für die im Laufe der hier beschriebenen Untersuchungen gemessene Aktivitäten an PfGSK-3 und HsGSK-3 β zur Verfügung standen.² Die Verbindungen lassen sich nach ihrer Substitution am zugrunde liegenden Thieno[2,3-*b*]pyridin³ an den Positionen 5 und 6 in die Klasse 1 und 2 einteilen, wobei eine weitere Aufteilung der ersten Klasse in 1a und 1b gemäß ihrer Substitution an C2 sinnvoll erscheint (siehe Abb. 3.1).

Die Klasse 1 zeichnet sich durch eine 6-Amino-5-cyano-Substitution am Grundkörper aus. Dabei ist die Position 2 in der Klasse 1a mit einer Nitrilgruppe substituiert, während in den Verbindungen der Klasse 1b das Thienopyridin über eine Keto-funktion mit einem (zumeist weiter substituierten) aromatischen Kern verbunden ist. Die Klasse 2 ist zwischen den Positionen 5 und 6 mit einem gesättigten Ring anelliert (Cyclopentan bis Cy-

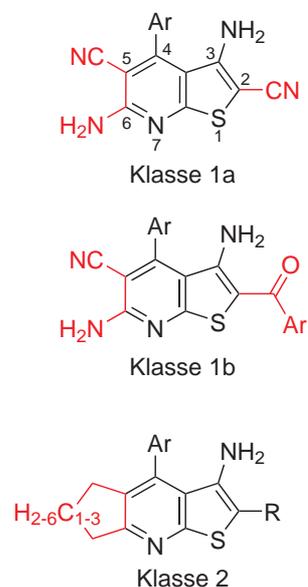


Abbildung 3.1. | Aufteilung der Thienopyridine in die Klassen 1a, 1b und 2.

²Der Übersichtlichkeit halber sind *alle* Verbindungen dieses Datensatzes (in der Regel mit Thieno[2,3-*b*]pyridin als Grundkörper) mit laufenden Nummern innerhalb der Tabelle – und nicht in der Reihenfolge ihres Erscheinens – versehen und werden entsprechend referenziert.

³Da es sich bei dem Grundkörper fast aller Verbindungen des Datensatzes um ein Thieno[2,3-*b*]pyridin handelt, wird dieser Heteroaromat im Folgenden vereinfachend als Thienopyridin bezeichnet.

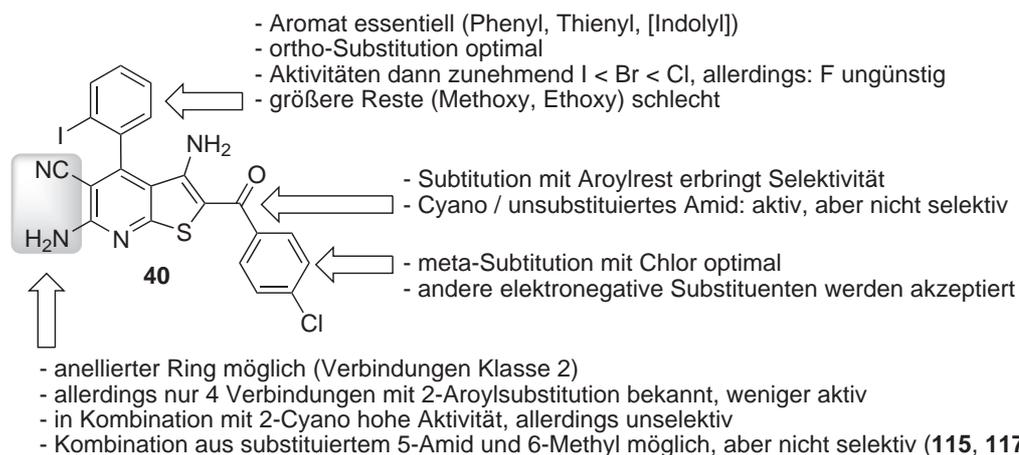


Abbildung 3.2. | Bekannte Struktur-Wirkungsbeziehungen der untersuchten Thienopyridine.

cloheptan). Als Substituent an C2 kann in Entsprechung zur Klasse 1 ein Nitril vorliegen, als kleiner Rest kommt ebenso ein Amid vor. Es gibt auch Verbindungen der Klasse 2 mit dem Substitutionsmuster der Klasse 1b an C2. Eine Aufteilung in die Klassen 2a und 2b ist in Entsprechung zur Klasse 1 demnach möglich. Sie wurde jedoch nicht durchgeführt, da dann mit 8 (Klasse 2a) und 4 (Klasse 2b) Verbindungen, für die diskrete Aktivitäten an der PfGSK-3 gemessen wurden, extrem kleine Gruppierungen entstanden wären, die nicht als repräsentativ eingeschätzt wurden.

3.2.1.1. Bekannte Struktur-Wirkungsbeziehungen

Im Rahmen der Synthese der Verbindungen durch BRANDT [92] wurden bereits wichtige Aspekte der Struktur-Wirkungsbeziehungen der Verbindungsklasse der Thienopyridine diskutiert. Dabei war die Aufmerksamkeit auf die Klasse 1 gerichtet, da die im Rahmen der zitierten Arbeit synthetisierten Derivate der Klasse 2 (**129-131**) inaktiv waren. Wichtige Punkte der in der oben zitierten Arbeit entwickelten Struktur-Wirkungsbeziehungen sollen hier im Folgenden rekapituliert werden, da sie als Ausgangspunkt für diese Arbeit von zentraler Bedeutung sind. Sie sind zusammenfassend in Abbildung 3.2 dargestellt.

Die erste wichtige Feststellung ist der Zusammenhang zwischen einer 2-Aroylsubstitution des Grundkörpers der Verbindungen der Klasse 1b und der erwünschten Selektivität am plasmoidalen Enzym. Alle Verbindungen, die der Klasse 1b angehören und Aktivität an der PfGSK-3 zeigen, sind auch selektiv. Erstaunlich ist in diesem Zusammenhang, dass es zahlreiche Beispiele für Verbindungen gibt, die strukturell nahe verwandt sind mit der Hitstruktur **40** aus dem HTS und dennoch weder an der PfGSK-3 noch an der HsGSK-3 β aktiv sind (z. B. **82**, **86**, **88-100**). Für den Aroylsubstituenten wurde eine Substitution in meta-Stellung mit Chlor als günstigste Variante ermittelt. Für die resultierende Verbindung **77** wurde zunächst ein IC₅₀ von 0.48 μ M an der PfGSK-3 gemessen, bei einer Wiederholmessung mit 0.06 μ M sogar das beste Ergebnis im Rahmen der Serie.

Die Thienopyridin-2-carbonitrile der Klasse 1a hingegen sind zwar aktiv, so lange der Substituent an Position 2 des 4-Arylrestes nicht zu groß ist (Iod, Methoxy und Ethoxy werden nicht toleriert), aber nicht selektiv für die PfGSK-3. Der Arylrest der Klasse 1b spielt offenbar eine entscheidende Rolle für die Selektivität der Substanzen an der PfGSK-3.

Ein weiterer wichtiger Befund ist, dass die Aktivität in Abhängigkeit vom Substituenten des 4-Arylrestes in Position 2 in der Reihe $I < Br < Cl$ zunimmt (**57** < **58** < **31**; **68** < **75** < **78**; **66** < **74** < **77**). Die Aktivität scheint mit abnehmender Größe des Substituenten zu steigen. Andererseits ist das Fluorderivat **84** inaktiv. Zudem gibt es aus dem Screeningdatensatz zwei Beispiele für nicht aktive Substanzen der Klasse 1b, die am 4-Arylrest unsubstituiert sind (**106** und **97**). Das einzige unsubstituierte Derivat der Klasse 1a (**56**) hingegen ist mit einem IC_{50} von $0.4 \mu M$ an der PfGSK-3 eine vergleichsweise aktive Substanz.

Dass die fluorsubstituierte Verbindung keine Wirksamkeit zeigt, deutet auf hydrophobe Wechselwirkungen durch die offenbar optimale Chlorsubstitution hin. Diese kann das Fluor mit seinem kleinen Radius und seiner hohen Elektronegativität nicht ausbilden – die Wirkungslosigkeit der unsubstituierten Derivate der Klasse 1b (**106** und **97**) wäre die folgerichtige Konsequenz. Die Wirksamkeit von **56** bleibt allerdings widersprüchlich und kann nur so erklärt werden, dass die 2-Cyanosubstitution eine so entscheidende Rolle für die Interaktion mit dem Protein spielt, dass die Substitution des 4-Arylrestes vergleichsweise wenig ins Gewicht fällt.

Die beschriebenen Struktur-Wirkungsbeziehungen sind Ergebnis der Arbeit [92] und wurden an dieser Stelle als Ausgangspunkt für weitere Untersuchungen zusammengefasst.

3.2.1.2. Biologische Daten

Die biologischen Aktivitäten der Verbindungen wurden in den Laboratorien von L. Meijer (Station Biologique de Roscoff, Frankreich) ermittelt. Es wurde dazu ein Kinase-Assay unter Verwendung von radioaktivem ATP eingesetzt, um die inhibitorische Wirkung der Verbindungen an PfGSK-3 und Säuger-GSK-3 zu testen. Gemessen wurden Aktivitäten des jeweiligen Enzyms bei 8 bis 9 unterschiedlichen Konzentrationen der Inhibitoren, daraus wurde der IC_{50} als die Konzentration abgeleitet, bei der 50% der maximalen Inhibition erreicht sind. Die vermessenen Konzentrationen reichten von 0.005 bis $10 \mu M$, zum Teil bis $100 \mu M$ [92].

Als Säuger-GSK-3 wurde nicht die HsGSK-3 β eingesetzt (UniProt: P49841), sondern diejenige vom Schwein, *Sus scrofa* (UniProt: B2LUN4). Es kann davon ausgegangen werden, dass die so erhaltenen Ergebnisse die Werte für das humane Enzym gut repräsentieren, da sich die GSK-3 des Menschen von der des Schweines nur in 6 von 420 Aminosäuren unterscheidet. Damit ergibt sich eine Identität von 98.6% und eine Ähnlichkeit von 99.0% (Alignment und Werte: PSI-BLAST über NCBI-GUI [135, 209], vgl. Abb. 3.3; Ähnlichkeit entsprechend zu S. 47 evolutionäre Ähnlichkeit). Die sich unterscheidenden Aminosäuren liegen nur in zwei der sechs Fälle in dem Bereich, der durch das Homologie-Modell beschrieben wird und keine der beiden liegt in der Bindetasche (Val40/Met40 bzw. Thr59/Ala59 in HsGSK-3 β /SsGSK-3).

3. Der Bindungsmodus von Inhibitoren der PfGSK-3

HsGSK	1	<u>MSGRPRTTSFAESCKPVQQPSAFGSMKVS</u> RDKDGSKVTTV <u>VATPGQGPDRPQEVSYTDTK</u>	60
SsGSK	1	MSGRPRTTSFAESCKPVQQPSAFGSMKVSRDKDGSKVTTM <u>VATPGQGPDRPQEVSYTDAK</u>	60
HsGSK	61	<u>VIGNGSFGVVYQAKL</u> CDSGELVAIKKVLQDKRFKNRELQIMRKL ⁴⁰ DHCNIVRLRYFFYS <u>SSG</u>	120
SsGSK	61	VIGNGSFGVVYQAKLCDSGELVAIKKVLQDKRFKNRELQIMRKL ⁴⁰ DHCNIVRLRYFFYS <u>SSG</u>	120
HsGSK	121	<u>EKKDEVYLN</u> LVLDYVPETVYRVARHYSRAKQTL ⁵⁹ PVIYVKLYMYQLFRSLAYIHSFGI <u>CHR</u>	180
SsGSK	121	EKKDEVYLNLVLDYVPETVYRVARHYSRAKQTL ⁵⁹ PVIYVKLYMYQLFRSLAYIHSFGI <u>CHR</u>	180
HsGSK	181	<u>DIKPQN</u> LLLLDPDTAVLKL ¹³ CDFGSAKQLVRGEPNVSYICSRYR <u>APELIFGATDYTSSIDV</u>	240
SsGSK	181	DIKPQNLLLLDPDTAVLKL ¹³ CDFGSAKQLVRGEPNVSYICSRYR <u>APELIFGATDYTSSIDV</u>	240
HsGSK	241	<u>WSAGCV</u> LAELLLGQPIFFPGDSGVDQLVEI ¹⁹ IKVLGTP <u>TREQIREMNPNYTEFKFPQIKAHP</u>	300
SsGSK	241	WSAGCVLAELLLGQPIFFPGDSGVDQLVEI ¹⁹ IKVLGTP <u>TREQIREMNPNYTEFKFPQIKAHP</u>	300
HsGSK	301	<u>WTKVFR</u> PRTPEAIALCSRLLEYTP <u>TARLTPLEACAHSFFDEL</u> RDPNVKLPNGRDT <u>PALF</u>	360
SsGSK	301	WTKVFRPRTPEAIALCSRLLEYTP <u>TARLTPLEACAHSFFDEL</u> RDPNVKLPNGRDT <u>PALF</u>	360
HsGSK	361	<u>NFTTQEL</u> SSNPPLATILIPPHARIQAAASTP <u>TNATAASDANTGDRGQTNN</u> AASASASNST	420
SsGSK	361	NFTTQELSSNPPLATILIPPHARIQAAASTP <u>SNATAASDTNAGDRGQTNN</u> TASASASNST	420

Abbildung 3.3. | Sequenzalignment von HsGSK-3 β und SsGSK-3. Nur der unterstrichene Bereich wird durch das Homologie-Modell beschrieben. In diesen fallen die beiden Unterschiede Val40/Met40 bzw. Thr59/Ala59 in HsGSK-3 β /SsGSK-3. Beide befinden sich relativ weit (etwa 13 bzw. 19 Å) von der ATP-Bindetasche entfernt, so dass durch den Unterschied kein wesentlicher Einfluss auf die Affinität ATP-kompetitiver Inhibitoren zu erwarten ist.

3.2.2. Vorgehensweise

Nach der Konstruktion der Homologie-Modelle (vgl. Kapitel 2) war der nächste Schritt auf dem Weg zu optimierten Inhibitoren der PfGSK-3 die Entwicklung einer Hypothese zum Bindungsmodus der bearbeiteten Inhibitoren, um rationale Synthesevorschläge für Variationen am Grundkörper machen zu können. Dazu wurde ein erstes orientierendes Docking mit einer Auswahl an Verbindungen durchgeführt, das als Ideengenerator für die darauf folgende präparative Arbeit zu verstehen ist, die im Rahmen der Promotion geplant war (Ergebnisse vgl. S. 100 ff.). Dieses orientierende Docking wurde aus organisatorischen Gründen zügig durchgeführt, da erste Vorstellungen über den Bindungsmodus und daraus abgeleitete Synthesevorschläge zu Jahresbeginn 2009 vorliegen sollten. Chronologisch folgte darauf die Synthese der Verbindungen **101** bis **104**, die in Abschnitt 3.4 beschrieben ist. Im Anschluss daran wurde mit dem dann erweiterten Datensatz ein zweites Docking durchgeführt, wobei eine Evaluierung des angewendeten Verfahrens mit Inhibitoren der HsGSK-3 β vorangestellt wurde (Ergebnisse vgl. S. 125 ff.).

3.2.2.1. Orientierendes Docking für Synthesevorschläge

Aus den in Abschnitt 3.2.1 vorgestellten Verbindungen wurden für das erste Docking einige repräsentative Vertreter ausgewählt. Die Auswahl sollte die folgenden Anforderungen erfüllen:

- Die Strukturen sollten an der PfGSK-3 aktiv sein, da es vorrangiges Ziel war, Hypothesen

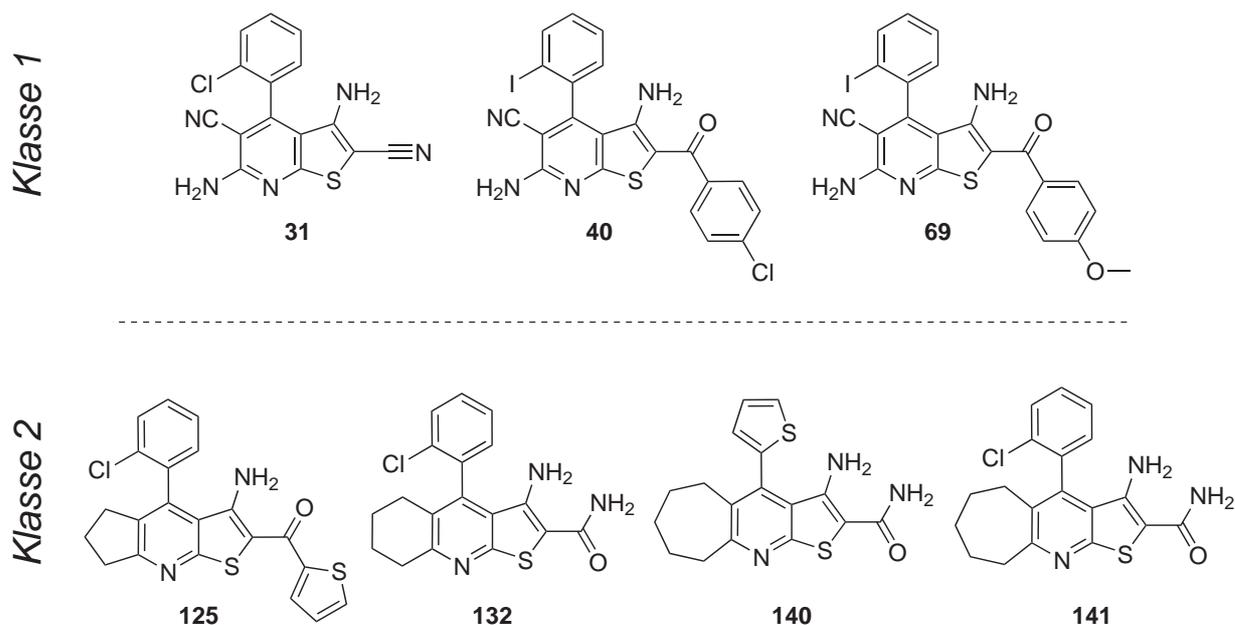


Abbildung 3.4. | Für das orientierende Docking ausgewählte Verbindungen. In der oberen Reihe sind diejenigen der Klasse 1 (Substitution an C5 und C6: Nitril bzw. Amino), in der unteren die der Klasse 2 (anellierter gesättigter fünf- bis siebengliederiger Ring) abgebildet.

für den Bindungsmodus zu entwerfen und daraus resultierende Synthesevorschläge zu entwickeln.

- Es sollten sowohl selektive als auch nicht selektive Verbindungen gedockt werden, um schon an dieser Stelle der Arbeit mögliche Unterschiede beobachten zu können.
- Im Datensatz enthaltene Grundmuster sollten repräsentiert werden. Das bedeutet insbesondere, dass
 - sowohl Vertreter der Klasse 1 mit 6-Amino-5-cyano-Substitution als auch solche der Klasse 2 mit ungesättigtem Anellanden an dieser Position ausgewählt wurden.
 - die Klassen 1a und 1b mit unterschiedlicher Substitution am C2 des Thienopyridins vertreten sein sollten.

Nach diesen Kriterien wurden die in Abbildung 3.4 dargestellten Verbindungen ausgewählt. Um die Kombination der Substitutionsmuster an C5 bzw. C6 und an C2 in dem ersten Docking zu untersuchen, wurden die Verbindungen **31** und **40** als Vertreter der Klasse 1 bzw. **125** und **141** für die Klasse 2 gewählt. Durch den Austausch von 2-Chlorphenyl in **141** gegen Thiophen an C4 in **140** kommt es zu Wirkungsverlust an der *HsGSK-3 β* , weshalb diese Variation interessant erschien und ebenfalls mit aufgenommen wurde. Als zusätzliche Verbindung der Klasse 2 sollte **132** gedockt werden, um die Auswirkungen der ansteigenden Ringgröße vom fünf- über den sechs- zum siebengliedrigen Ring zu beobachten und außerdem, um ein an Position 2 benzoylsubstituiertes Derivat der Klasse 2 zu untersuchen. Für die Klasse 1 wurde zusätzlich **69**

Verbindung	IC ₅₀ an PfGSK-3	IC ₅₀ an HsGSK-3β
31	0.13	0.4
40	0.45	>100
69	3.2	>10
125	30	>100
132	3	3
140	1.1	>100
141	3.8	9

Tabelle 3.1. | IC₅₀-Werte der Verbindungen aus Abbildung 3.4, alle Angaben in μM. Im oberen Teil der Tabelle sind Verbindungen der Klasse 1 aufgeführt, im unteren diejenigen der Klasse 2.

gedockt, um den darin verwirklichten Austausch eines Halogens gegen Methoxy zu untersuchen. Die IC₅₀-Werte der sieben Verbindungen sind in Tabelle 3.1 dargestellt.

Das Docking wurde mit FlexE und AutoDock durchgeführt. Da FlexE durch das Einbinden verschiedener Moleküle in eine United Protein Structure die Flexibilität des Proteins berücksichtigt, sollte dies auch in AutoDock geschehen. Es wurden dazu die Seitenketten der Aminosäuren Lys108, Met157, Thr163, Lys166 und Gln210 als flexibel behandelt. Die Aminosäuren wurden ausgewählt, da ihre Seitenketten in die Bindetasche hineinragen und sie deshalb Kontakte mit einem Inhibitor ausbilden oder aber zu sterischen Hinderungen führen können.

Repräsentanten der verschiedenen Bindungsmodi wurden in einer anschließenden Komplexminimierung optimiert und es wurde neben der Bewertung des Komplexes mit der im jeweiligen Dockingprogramm implementierten Scoringfunktion (in diesem Zusammenhang auch als generische Scoringfunktion bezeichnet) ein Vergleich mit der wissensbasierten Scoringfunktion DrugScore angestellt.

Die Minimierung war insbesondere bei den mit FlexE erhaltenen Ergebnissen zwingend notwendig, denn diese bestehen nicht aus einem regulären Protein-Ligand-Komplex. Die als Rezeptor für die Berechnung des Scores von FlexE herangezogene Proteinstruktur stellt vielmehr ein theoretisches Konstrukt dar. Beim Generieren dieser United Protein Structure müssen Grenzen und damit Kompromisse für das Zusammenfassen ähnlicher Untereinheiten zu einem Repräsentanten gewählt werden (Details vgl. exp. Teil S. 136). Daher ist die United Protein Structure als Kombination dieser Repräsentanten ohne Nachbehandlung nicht unbedingt eine biologisch oder biochemisch realistische Struktur, wie sie für eine Minimierung notwendiger Ausgangspunkt wäre. Es wurde deshalb jeweils ein Protein ausgewählt, das die gleiche Pose erlaubt und dieses in Kombination mit dem mit FlexE gedockten Liganden minimiert.

Für das in FlexE verwendete Ensemble wurden die in Kapitel 2 gemäß *RCS* bestplatzierten 11 Proteine verwendet, für das Docking mit AutoDock wurde eine Auswahl von 5 Proteinen getroffen, da die Ergebnisse in absehbarer Zeit visuell ausgewertet werden sollten und eine überschaubare Lösungsmenge notwendig war. Es wurden mit *pfgsk_1j1b*, *pfgsk_1q3d*, *pfgsk_1q41*, *pfgsk_2ow3* und *pfgsk_kubra* möglichst diverse Proteine ausgewählt (vgl. Distanzen innerhalb der Bindetasche, S. 54; Vorbereitung der Proteine und Liganden auf das Docking siehe exp. Teil S. 136 ff.).

Es wurden für die beiden Klassen mehrere mögliche Bindungsmodi identifiziert und eine Arbeitshypothese aufgestellt, die als Grundlage für Modifikationsvorschläge am Grundkörper dienen sollte. Diese Synthesevorschläge wurden ebenfalls wie oben beschrieben gedockt und analog zum Docking der bekannten Substanzen bewertet. Als mögliche Vorschläge wurden diejenigen in Betracht gezogen, die

- in der modifizierten Form im gleichen Modus gedockt wurden wie die Ausgangsverbindungen,
- eine ähnlich gute Bewertung im Vergleich zur Ausgangsverbindung erhielten und
- vom Kontext der bekannten Struktur-Wirkungsbeziehungen interessant erschienen, da die Modifikation aktive und selektive Verbindungen erwarten ließ.

3.2.2.2. Der optimierter Arbeitsablauf DoReCCoM

Die Ergebnisse aus dem ersten Docking erlaubten noch keine Entscheidung für einen Bindungsmodus (siehe Ergebnisteil 1 auf S. 100 ff.), so dass in einem zweiten Docking mit dem erweiterten Datensatz⁴ ein modifizierter, stark erweiterter Arbeitsablauf durchgeführt wurde, der eine zuverlässigere Aussage erlauben sollte.

Dabei war die zentrale Idee, in einem ersten Schritt durch die Kombination verschiedener Dockingalgorithmen mit hoher Wahrscheinlichkeit eine kristallstrukturnahe Posen zu erzeugen (vgl. Docking-Problem, Abschnitt 3.1.1 auf S. 79 ff.). Durch Kombination mit einer geeigneten Scoringfunktion sollte diese dann besser aufzufinden sein, als es mit den generischen Scorings der einzelnen Programme möglich ist. Zusätzlich sollte die Flexibilität der Proteine in das Verfahren einbezogen werden – einerseits durch die Betrachtung mehrerer Kristallstrukturen, andererseits durch eine Komplexminimierung als letzten Schritt des Verfahrens, der in der Art eines Induced Fits eine Anpassung der Bindetasche an den Liganden erlauben sollte.

Ein wichtiger Unterschied zum ersten Durchlauf sollte darin liegen, dass zunächst eine Evaluierung des gewählten Verfahrens mit bekannten Kristallstrukturen der *HsGSK-3 β* und deren Liganden im Sinne eines klassischen Crossdockings durchgeführt wurde, um die Ergebnisse besser beurteilen zu können.

Der Arbeitsablauf wurde aus fünf wesentlichen Schritten konzipiert: Docking, Rescoring, Clustering, Complex Minimization, Rescoring (DoReCCoM, vgl. Abb. 3.5).

Docking Es wurde neben den beiden bereits im ersten Durchlauf gewählten Programmen FlexX und AutoDock zusätzlich mit FRED gearbeitet. Dadurch wurden drei unterschiedliche Prinzipien des Aufbaus eines Protein-Ligand-Komplexes gewählt (FlexX: inkrementell, AutoDock: genetischer Algorithmus, FRED: rigider Aufbau in der Bindetasche, shape

⁴Neu waren dabei zum einen die selbst synthetisierten Verbindungen **101** bis **104**, deren Darstellung im Syntheseteil beschrieben ist. Zum anderen wurden im Frühjahr 2009 die ergänzenden Daten zu weiteren Substanzen bekannt, die in Tabelle B.3 auf Seite 234 mit einem ^x versehen sind.

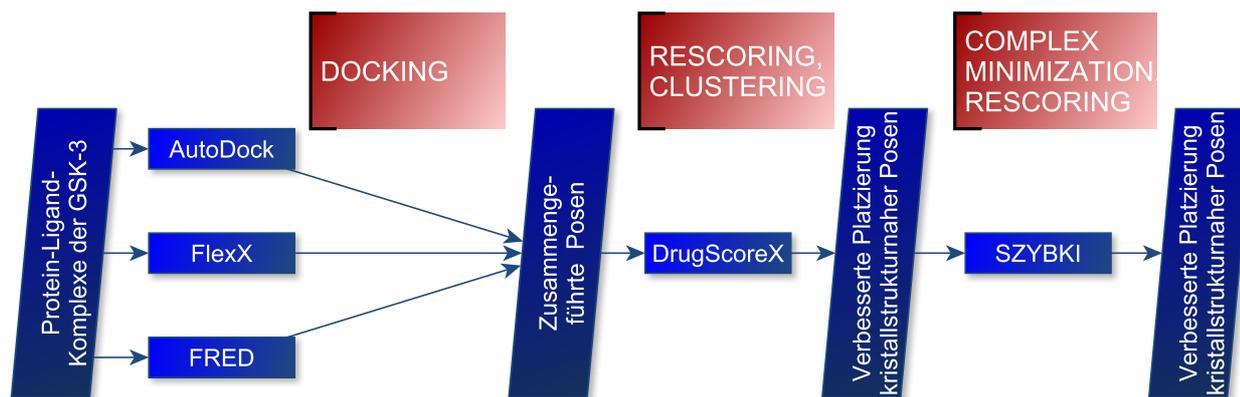


Abbildung 3.5. | Arbeitsablauf DoReCCoM.

based), so dass die Wahrscheinlichkeit hoch ist, Schwächen eines einzelnen Programmes durch Stärken eines anderen auszugleichen. Das Docking wurde für alle ausgewählten Liganden in die entsprechenden Proteine durchgeführt (Crossdocking, siehe *Datensatz für die Validierung* auf S. 92 ff.).

Rescoring Es wurde mit DrugScoreX eine wissenschaftsbasierte Scoringfunktion gewählt, mit der alle Lösungen der verschiedenen Programme bewertet wurden. Die Potentiale von DrugScoreX basieren gegenüber der Vorgängerversion auf einem erweiterten Datensatz an Protein-Ligand-Komplexen, so dass hier eine Verbesserung zu erwarten war.

Clustering Durch das Zusammenführen aller Lösungen der drei unterschiedlichen Dockingprogramme ergab sich für jede Kombination von Ligand und Protein eine Vielzahl von Lösungen. Sie wurden für jedes einzelne im Rahmen des Crossdocking zu berücksichtigende Protein-Ligand-Paar rmsd-basiert in Cluster eingeteilt. Der gemäß DrugScoreX beste Komplex jedes Clusters wurde als Repräsentant in den nächsten Schritt übernommen.

Komplexminimierung Um die Flexibilität des Proteins nicht nur durch die Verwendung unterschiedlicher Proteinkonformationen in das Verfahren zu integrieren, wurde als nächstes eine Komplexminimierung durchgeführt, um die Anpassung des Proteins an die Position des Liganden zu ermöglichen (Induced-Fit-Effekt).

Rescoring Die optimierten Komplexe wurden erneut mit DrugScoreX bewertet. Zur Auswertung wurden die Lösungen aus verschiedenen Proteinen für jeden Liganden vereinigt und anhand ihres DrugScoreX-Wertes sortiert. Dieses Vorgehen hat sich beim Docking in Homologie-Modelle in mehreren Studien als vorteilhaft erwiesen [210, 211].

3.2.2.2.1. Docking Die Probleme bei der Bewertung der durch FlexE erhaltenen Komplexe in Kombination mit den heterogenen Ergebnissen führte zu der Entscheidung, für DoReCCoM FlexX anstelle von FlexE einzusetzen. Bei den mit AutoDock durchgeführten Docking-

Experimenten wurde auf eine Freigabe von Seitenketten des Rezeptors und eine dadurch mögliche Repräsentation der Flexibilität verzichtet. In Vorversuchen konnten die Ergebnisse von MISHRA *et al.* [212] nicht bestätigt werden. Die Autoren hatten in einer Studie zu Virtual Screening und Induced-Fit-Mechanismen bei der *HsGSK-3 β* Vorteile der Verwendung dieser Option beschrieben, die sich in den durchgeführten Vorversuchen mit dem unter 3.2.2.1 beschriebenen Datensatz nicht bestätigten. Möglicherweise liegt dies in dem Fokus der genannten Untersuchung auf die Anreicherung aktiver Verbindungen in virtuellen Screening begründet. Der hier verfolgte Ansatz stellt hingegen den Bindungsmodus in den Vordergrund, was nicht notwendigerweise eine bessere Anreicherung zur Folge haben muss. Die Verwendung von FRED als drittes Docking-Werkzeug wurde wegen des abweichenden Docking-Prinzips gewählt, der die beiden anderen Ansätze sinnvoll ergänzt.

3.2.2.2.2. Einführung eines Pharmakophores Als ATP-kompetitive Proteinkinaseinhibitoren vom Typ I bilden alle in der Evaluierung eingesetzten Liganden einen Wasserstoffbrückenkontakt zu Hinge Region aus (vgl. Abb. 3.6 auf S. 93 bzw. allg. Teil S. 15 ff.). Es ist daher davon auszugehen, dass dieser auch bei Inhibitoren der *PfGSK-3* vorliegt.

Während von ATP die klassische Anordnung Donor-Akzeptor-Donor bekannt ist, wenn der Wasserstoff des C2 als Donor einer Wasserstoffbrücke zum Val135 betrachtet wird, und zahlreiche Kinaseinhibitoren mehr als einen Kontakt zur Hinge Region ausbilden, kann in anderen Beispielen durchaus mit weniger Kontakten eine Bindung zustande kommen. Besonders zu nennen ist in diesem Zusammenhang der Bindungsmodus von **31**. Hier wird die zentrale Bindung zur Hinge Region über Wasser vermittelt, so dass eine einzige direkte Wasserstoffbrückenbindung zum Carbonyl des Val135 ausreicht (vgl. Abb. 1.17 auf S. 20).

Es sollte deshalb in dem zweiten Docking mindestens eine der folgenden Wasserstoffbrücken zwischen Ligand und Protein ausgebildet werden:

1. Carbonyl des Asp133 \Rightarrow Wasserstoffbrückendonor des Liganden
2. Backbone-NH des Val135 \Rightarrow Wasserstoffbrückenakzeptor des Liganden
3. Carbonyl des Val135 \Rightarrow Wasserstoffbrückendonor des Liganden

In FlexX ist eine solche Definition eines Pharmakophors, bestehend z. B. aus Wasserstoffbrücken, als logische Verknüpfung möglich (mindestens eine von dreien) und wurde gesetzt. In AutoDock besteht die Möglichkeit überhaupt nicht und in FRED müssen die Wasserstoffbrücken in der verwendeten Version als zwingend definiert werden. Lösungen ohne den vorgeschriebenen Kontakt werden dann ausgeschlossen. Da wie oben geschildert keiner der drei Kontakte als absolut zwingend gelten kann, wurde deshalb in FRED auf diese Bedingung verzichtet.

3.2.2.2.3. Komplexminimierung Es wurde einer Minimierung der Vorzug gegenüber einer wesentlich aufwändigeren MD-Simulation gegeben, deren Kombination mit Dockings einen wertvollen Beitrag bei der Suche nach einem Bindungsmodus darstellen kann (Übersicht in [213], erfolgreiche Anwendung z. B. von NABUURS *et al.* mit Fleksy [214]). Allerdings sind MD-Simulationen herausfordernd hinsichtlich der Interpretation und können mitunter Unschärfe in das System einführen [215]. Da andererseits durch die Kombination verschiedener Dockings das Problem des Auffindens der richtigen Pose lösbar ist, sollte der Versuch unternommen werden, kleine Veränderungen durch eine Minimierung einzuführen, die für eine verbessertes Scoring ausreichen.

3.2.2.3. Datensätze für die Evaluierung von DoReCCoM

Für die Validierung wurden die in Tabelle 3.2 aufgeführten Proteine aus der PDB verwendet. Es handelt sich dabei um eine Auswahl der in Tabelle 2.2 auf Seite 35 aufgeführten Kristallstrukturen der *HsGSK-3 β* . Der Informationsgewinn durch eine immer weiter ansteigende Anzahl ähnlicher Proteine wurde als gering eingeschätzt, da ähnliche Liganden die Bindetasche auf ähnliche Weise präformieren – aus diesem Grund wurden nicht alle in der genannten Tabelle aufgeführten Proteine verwendet.

Es wurden die kokristallisierten Liganden eingesetzt (vgl. Abb. 3.6), wobei ANP (**41**), STU (**25**) und BIM (**45**) nicht in die Auswertung mit einbezogen wurden, da die Übertragbarkeit

Tabelle 3.2. | In der Evaluierung eingesetzte Kristallstrukturen. Die grau unterlegten Liganden wurden aus den im Text genannten Gründen nicht mit in die Analyse einbezogen.

PDB-Eintrag	Ligand	Auflösung [Å]	Literatur
1J1B	ANP (41)	1.8	AOKI <i>et al.</i> [150]
1Q3D	STU (25)	2.2	BERTRAND <i>et al.</i> [70]
1Q3W	ATU (28)	2.3	ebd.
1Q41	IXM (36)	2.1	ebd.
1Q4L	UN1 (679) (42)	2.77	ebd.
1R0E	DFN (43)	2.25	ALLARD <i>et al.</i> ^a
205K	HBM (44)	3.2	SHIN <i>et al.</i> [153]
20W3	BIM (45)	2.8	ZHANG <i>et al.</i> [154]
3DU8	UN2 (553) (46)	2.2	MENICHINCHERI <i>et al.</i> [155]
3F7Z	UN3 (3FO) (47)	2.4	SAITOH <i>et al.</i> [156]
3F88	HT3 (48)	2.6	ebd.
(kubra)	DRG (31)	2.75	pers. Mitteilung ^b

^aUnveröffentlichter Artikel mit dem Titel *From genetics to therapeutics: the Wnt pathway and osteoporosis*; ^bA. Oberholzer, Universität Bern und L. Meijer, Station Biologique, Roscoff (Sommer 2009);

(RCSB) PDB: Research Collaboratory for Structural Bioinformatics Protein Data Base [66, 67]; Trivialname bzw. abweichende Bezeichnung für ATU: Alsterpaulon, für STU: Staurosporin; UN1-3 sind eigene Bezeichnungen für die jeweiligen Liganden, der Name in der Originaldatei ist in Klammern angegeben.

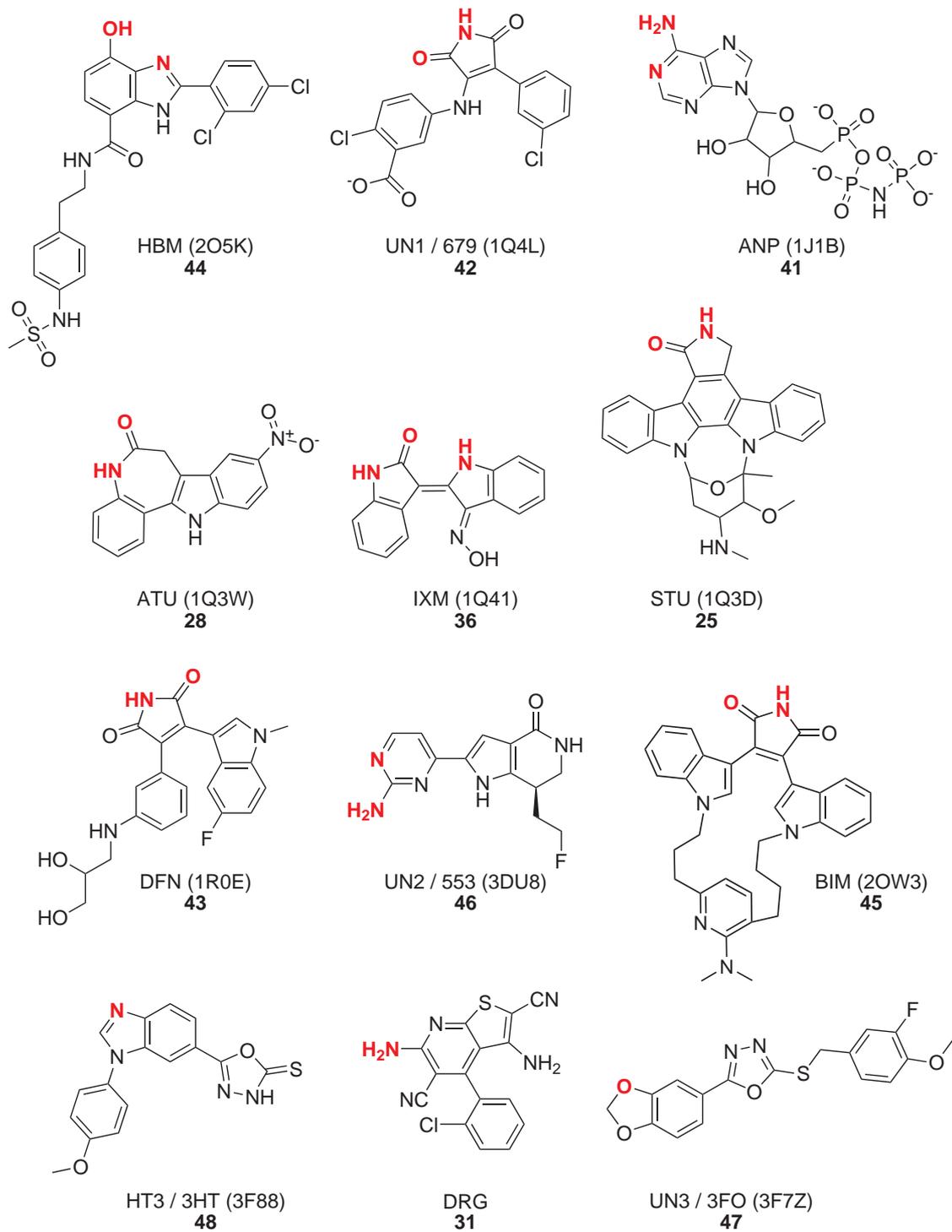


Abbildung 3.6. | Kokristallisierte Liganden zur Evaluierung von DoReCCoM. Die Verbindungen bilden mit den rot markierten Atomen Wasserstoffbrücken zur Hinge Region der *HsGSK-3β* aus.

der Ergebnisse dieser Liganden auf die Thienopyridine des Projektdatensatzes aufgrund ihrer starken Unterschiedlichkeit fraglich schien. Im Falle des ATP-Analogons ANP (**41**) besteht der Unterschied in der mehrfachen negativen Ladung der Triphosphat-Untereinheit des Moleküls und dem ausgedehnten konformativen Spielraum dieser Teilstruktur. Staurosporin (**25**) und BIM (**45**) sind beide von ihrer Größe her nicht vergleichbar mit den Zielstrukturen und wurden deshalb nicht miteinbezogen. Zudem ist **25** aufgrund des ausgedehnten starren Grundgerüsts und **45** wegen des großen ungesättigten Ringes nicht vergleichbar mit den Thienopyridinen. Die dazugehörigen Proteine wurden dennoch mit in die Untersuchungen einbezogen, da gerade die Fragestellung, wie gut der Bindungsmodus eines Liganden in einem *nicht* kokristallisierten und dadurch präformierten Protein gefunden wird, für die Anwendung des Verfahrens auf neue Verbindungen entscheidend ist.

Mit 3DU8, 3F7Z und 3F88 wurden Kristallstrukturen in die Evaluierung des Verfahrens einbezogen, für die kein Homologie-Modell der PfGSK-3 konstruiert wurde. Diese Strukturen waren noch nicht veröffentlicht, als die Homologie-Modelle berechnet wurden. Da die kokristallisierten Liganden UN2 (**46**), UN3 (**47**) und HT3 (**48**) allerdings vergleichbar mit den Thienopyridinen schienen, wurden sie mit für die Evaluierung verwendet.

Bei der Verbindung **48** gibt es eine Besonderheit: Das in Abbildung 3.6 dargestellte Molekül ist eines von zweien, die in der Originaldatei als Liganden geführt werden (PDB-Eintrag 3F88, [156]). Die beiden Verbindungen sind in Abbildung 3.7 dargestellt. Sie sind Spaltprodukte aus der zur Verbindung UN3 (**47**) analoge Substanz 20x (**49**), die in den entsprechenden Arbeiten synthetisiert und in den biologischen Tests eingesetzt wurde. Die Autoren stellen jedoch eine Spaltung der C–S-Bindung von **49** in die getrennten Moleküle **50** und **48** im Rahmen des experimentellen Ablaufes der Aufnahme dieser Röntgenkristallstruktur fest. Der in der Abbildung bezeichnete Abstand zwischen Schwefel und Kohlenstoff in der Kristallstruktur von 3.4 Å belegt dies und es ist davon auszugehen, dass hier nicht ein Molekül vorliegt, sondern tatsächlich zwei einzelne kokristallisiert wurden. Es wurde für das Docking demzufolge nur **48** verwendet, das die Interaktionen zur Hinge Region ausbildet und daher als Ligand anzusehen ist. Theoretisch kann das 1,3,4-Oxadiazol-2-thiol bzw. -thion in **48** in zwei tautomeren Formen dargestellt werden. Aufgrund der vorzuziehenden Aromatisierung, den Autoren SAITOH *et al.* [156] folgend und nicht zuletzt in Analogie zu **47** wurde die Thiolform gewählt.

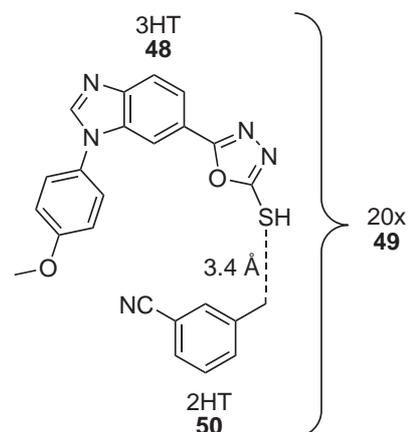


Abbildung 3.7. | Besonderheit von **48**. Es wurde die Struktur 20x (**49**) bei der Kristallisation eingesetzt. In der Röntgenkristallstruktur liegen nach Spaltung der C–S-Bindung die beiden einzelnen Liganden **48** und **50** vor [156].

3.2.2.4. Datensätze für die Anwendung von DoReCCoM

Es wurden wie beim ersten Durchlauf repräsentative Verbindungen aus dem vollständigen Datensatz ausgewählt, sie sind in Tabelle 3.3 aufgeführt. Dabei war eines der Probleme des ersten Dockings, dass nur ein einziger Vertreter der Klasse 1a untersucht wurde, es wurden nun die drei gelisteten Verbindungen ausgewählt. Auch die Variationsbreite der Klasse 1b wurde durch die 9 verschiedenen Verbindungen gegenüber den zwei zuvor betrachteten deutlich erweitert.

3.2.3. Software

3.2.3.1. FlexX/E

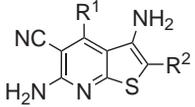
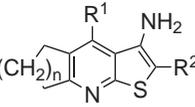
Ein etabliertes Docking-Programm ist FlexX [193, 216, 217]. FlexX verfolgt bei der Suche nach optimalen Ligand-Protein-Komplexen einen inkrementellen Ansatz. Nach Auswahl und Positionierung eines rigiden Basisfragments wird der Ligand in der Bindetasche schrittweise aufgebaut, wodurch dessen konformationelle Flexibilität berücksichtigt wird. Es stehen in FlexX prinzipiell neben der manuellen Positionierung des Basisfragmentes drei automatisierte Möglichkeiten zur Auswahl.

- `single interaction scan (sis)`
Beim `sis` wird das Basisfragment basierend auf einem einzigen Kontakt zum Rezeptor positioniert. Verschiedene Konformationen des Basisfragmentes in dieser Position mit dem besten `contact score` werden ausgewählt.
- `line fragments`
Diese Variante sollte laut Handbuch nur gewählt werden, wenn die anderen Verfahren keine Lösungen bieten. Da die hier nicht der Fall ist, wird er nicht weiter behandelt.
- `triangle hashing techniques based (tri)`
Dieser Algorithmus stellt das Standardverfahren in FlexX dar. Das einmal ausgewählte Basisfragment wird dabei so in der Bindetasche platziert, dass drei Kontakte des Liganden zustande kommen.

Dabei wird auf jeder Stufe für den hinzugefügten Teil des Liganden nach möglichen Interaktionen mit dem Protein gesucht, diese werden dann bewertet (hinsichtlich der Distanzen und ggf. der Winkel), wobei die jeweils besten Erweiterungen des Basisfragmentes letztlich zur Pose des ganzen wiederaufgebauten Liganden in der Bindetasche führen.

FlexE [218] ist ein Modul von FlexX, in dem durch Kombination mehrerer Proteinstrukturen zu einer United Protein Structure (dem namensgebenden **Ensemble**) die Flexibilität der Zielstruktur imitiert werden soll. Dabei wird die Bindetasche des Proteins in Bereiche unterteilt, die entweder sehr ähnlich sind und deshalb zusammengefasst werden können zu einem einzigen

Tabelle 3.3. | Verbindungen im zweiten Docking. Die Klassen 1a, 1b und 2 sind durch Linien abgegrenzt. In den Spalten *PfgGSK-3* und *HsGSK-3 β* sind die IC₅₀-Werte des jeweiligen Enzyms in μ M angegeben.

Verbindung	Struktur	n	R ¹	R ²	<i>PfgGSK</i>	<i>SsGSK</i>
56			Phenyl	CN	0.4	2.5
31			2-Chlor-phenyl	CN	0.13	0.4
62			2-Methyl-phenyl	CN	0.51	2
40			2-Iod-phenyl	4-Chlor-benzoyl	1.6	>10
69			2-Iod-phenyl	4-Methoxy-benzoyl	1.7	>10
71			2-Iod-phenyl	4-Methyl-benzoyl	3.2	>10
72			2-Iod-phenyl	Benzoyl	2.4	>10
77			2-Chlor-phenyl	3-Chlor-benzoyl	0.48	>10
78			2-Chlor-phenyl	3,4-Dichlor-benzoyl	2.7	>10
80			2-Chlor-phenyl	3-Methoxy-benzoyl	1.7	>10
101			2-Iod-phenyl	2-Thienyl-carbonyl	11.5	>100
102			2-Iod-phenyl	5-Cyano-thiophen-2-yl-carbonyl	68	>100
121		1	2-Thienyl	Benzoyl	43	>100
137		2	2-Thienyl	Benzoyl	1.8	>100
138		2	2-Thienyl	CN	0.8	2
140		3	2-Thienyl	Carboxamid	1.1	>100

repräsentativen Baustein der Bindetasche oder solche, die sich deutlich unterscheiden und deshalb als unterschiedliche mögliche Varianten einer Region (sog. Instances) bei dem eigentlichen Docking als Repräsentanten der betreffenden Region berücksichtigt werden. Durch Kombinationen dieser Bausteine kann es dazu kommen, dass Repräsentanten unterschiedlicher Bereiche zu einer hypothetischen Struktur zusammengesetzt werden, die in den Ausgangsstrukturen nicht vorlag. So werden zusätzlich zu den bestehenden Proteinen auch Strukturen generiert, die neue Geometrien der Bindetasche und damit andere Bindungsmodi ermöglichen.

Die von FlexX verwendete Scoringfunktion beruht auf der von BÖHM *et al.* beschriebenen empirischen Berechnung der freien Bindungsenergie von Protein-Ligand-Komplexen, die in der Art eines Kraftfeldes erfolgt [219]. In die von BÖHM beschriebene Funktion gehen die Energieanteile der Wasserstoffbrückenbindungen (ΔG_{hb}), der ionischen Interaktionen (ΔG_{ion}) und der durch lipophile Kontaktfläche bedingte Anteil der Bindungsenergie ($\Delta G_{lipo} |A_{lipo}|$) ein. Zudem wird die Anzahl N_{rot} der durch die Bindung immobilisierten rotierbaren Bindungen berücksichtigt. Abweichungen von geometrischen Idealwerten für die gerichteten Energien werden durch eine skalierende Funktion mit einbezogen ($f(\Delta R, \Delta \alpha)$). Nach BÖHM wird zusätzlich ΔG_0 als nicht direkt mit spezifischen Interaktionen zu korrelierender Anteil berücksichtigt, der entropischen Charakter hat. Die Gesamtenergie setzt sich nach BÖHM *et al.* demzufolge nach Gleichung 3.1 zusammen:

$$\begin{aligned} \Delta G_{binding} = \Delta G_0 &+ \Delta G_{hb} \Sigma_{h-bonds} f(\Delta R, \Delta \alpha) \\ &+ \Delta G_{ion} \Sigma_{ionic} f(\Delta R, \Delta \alpha) \\ &+ \Delta G_{lipo} |A_{lipo}| \\ &+ \Delta G_{rot} N_{rot} \end{aligned} \quad (3.1)$$

Der ΔG_{lipo} -Term wird in der von FlexX eingesetzten Form dieser Gleichung als Summe paarweiser Atom-Atom-Kontakte berechnet. Zusätzlich wird mit ΔG_{aro} ein Anteil für die Wechselwirkung von π -Systemen eingeführt, der wie ΔG_{hb} und ΔG_{ion} als geometrisch ausgerichtete Größe in die Gleichung eingeht [193].

3.2.3.2. AutoDock

In AutoDock [194, 220] wird für das Auffinden einer günstigen Konformation und Position des Liganden in der Bindetasche ein genetischer Algorithmus verwendet. Mögliche Posen werden dabei als Individuen einer Population verstanden. Als Ausgangspunkt des Dockings wird diese durch zufällige Zusammenstellung der Gene (Translation, Orientierung, rotierbare Bindungen) der Individuen generiert. Ein Generationswechsel umfasst fünf Schritte: Im ersten werden die genetisch kodierten Eigenschaften der Individuen in einen Protein-Ligand-Komplex überführt und dessen Fitness wird ermittelt. Die Fitness wird mittels einer empirischen Funktion zur Bestimmung der Bindungsenergie berechnet, die einen wichtigen Bestandteil von AutoDock

darstellt. Anschließend erfolgt die Selektion, bei der abhängig von der Fitness festgelegt wird, wie viele Nachfahren eine Verbindung haben wird. Im dritten und vierten Schritt eines Generationswechsels finden Cross Over und Mutation statt. Es werden paarweise zwischen zufällig ausgewählten Individuen kodierte Eigenschaften ausgetauscht. Die Mutation wird durch zufällige Veränderung des im Cross Over erhaltenen Gens erhalten. Als letzter Schritt gibt es die Möglichkeit, eine Anzahl von günstigsten Posen einer Generation festzulegen, die automatisch in die nächste Generation übertragen werden (Elitismus). Endpunkte der Berechnungen sind entweder die Anzahl an Generationen oder die der Energieberechnungen [194]. Größen wie die Anzahl der Generationen, die Populationsgröße und die Mutations- bzw. Cross-Over-Rate bestimmen damit den Verlauf des Dockings. Die Durchführung und insbesondere die Analyse des Dockings werden durch die Verwendung von AutoDockTools (ADT) [221, 222] wesentlich verbessert und erleichtert.

Bei der Vorbereitung der Liganden können Torsionswinkel freigegeben und damit einer Variation durch den genetischen Algorithmus unterworfen werden oder nicht. Es besteht zudem in AutoDock die Möglichkeit, die Flexibilität des Rezeptors in die Berechnungen aufzunehmen, indem die Seitenketten ausgewählter Aminosäuren in den Prozess integriert werden [222].

Das Scoring erfolgt innerhalb von AutoDock über die bereits erwähnte Abschätzung der freien Bindungsenergie mithilfe einer empirischen Funktion, die prinzipiell aufgebaut ist wie Gleichung 3.1. Ein Unterschied in der Vorgehensweise im Vergleich zu FlexX ist, dass die Energien basierend auf einem zuvor mit AutoGrid berechneten Gitter erhalten werden. An jedem Punkt dieses Gitters werden Interaktionen verschiedener Atomtypen des Liganden mit dem Protein bereits vor dem eigentlichen Docking berechnet. Für eine konkrete Pose setzt sich die Interaktionsenergie dann aus den Teilbeiträgen der Atome an den Gitterpunkten zusammen. Außerdem unterscheiden sich die Terme zur Berechnung der entropischen Anteile der Bindungsenergie (vgl. [194]). Ab Version 4 ist eine verbesserte Scoringfunktion in AutoDock implementiert, die sich durch die größere Anzahl an zugrundeliegenden Protein-Ligand-Komplexen und durch die verbesserte Berechnung der Desolvatationsenergie von der in Version 3 implementierten Fassung unterscheidet [220].

3.2.3.3. FRED

FRED [195, 223] ist die Abkürzung für **F**ast **R**igid **E**xhaustive **D**ocking. Ein grundsätzlicher Unterschied im Vorgehen von FRED gegenüber den im Vorangegangenen beschriebenen Programmen wird aus diesem Titel bereits klar: Es werden fixierte (rigid), zuvor generierte Konformere in der Bindetasche positioniert, während bei AutoDock Torsionswinkel eines zuvor konstruierten ganzen Moleküls freigegeben werden und bei FlexX ein inkrementeller Ansatz verfolgt wird. Die Flexibilität des Liganden wird in FRED berücksichtigt, indem für jeden Liganden verschiedene Konformere aus einer zuvor erstellten Bibliothek gedockt werden. Das erschöpfende (engl.: exhaustive) Docking besteht aus mehreren Schritten, die für jeden rigide behandelten

Eintrag einer Bibliothek durchgeführt werden: (1) Es werden mögliche Positionen des Liganden in der Bindetasche durch Translation und Rotation generiert. Die Bindetasche ist dabei durch ein Shape Potential Grid mit innerer und äußerer Kontur definiert. (2) Möglichkeiten mit Überschneidungen zum Protein (Atomen außerhalb der äußeren Kontur) werden zurückgewiesen. (3) Lösungen ohne Atome innerhalb der inneren Kontur werden ebenfalls zurückgewiesen. (4) Es können optional festgelegte zusätzliche Bedingungen überprüft werden. (5) Die verbleibenden Möglichkeiten können mit unterschiedlichen Scoringfunktionen bewertet werden. Für die aus dem exhaustive Docking übernommenen ersten 100 Lösungen aus allen Versuchen wird eine Optimierung unter Beibehaltung der Geometrie des Liganden durchgeführt (solid body optimization), gefolgt von einem erneuten Scoring, das auch als Consensus Scoring der verschiedenen implementierten Scoringfunktionen durchgeführt werden kann. Optional kann eine Energieminimierung als letzter Schritt durchgeführt werden.

Für die Bewertung der Posen stehen eine Reihe unterschiedlicher Funktionen zur Verfügung. Zum Vergleich mit den anderen Programmen wurde hier die in FRED standardmäßig voreingestellte Funktion Chemgauss3 verwendet. Die Funktion berechnet die Komplementarität von Ligand und Bindetasche als angepasste Gausskurve, sie erfasst Wasserstoffbrückenbindungen und berücksichtigt die Desolvatation von Ligand und Proteins [223].

3.2.3.4. DrugScore(X)

DrugScore [224, 225] unterscheidet sich grundsätzlich durch einen wissensbasierten Ansatz von den in FlexX, AutoDock und FRED eingesetzten Scoringfunktionen. DrugScore bewertet Protein-Ligand Komplexe basierend auf distanzabhängigen Paarpotentialen und aufgrund von Einteilchenpotentialen, die aus der lösungsmittelzugänglichen Oberfläche ermittelt wurden. Die beiden älteren Versionen DrugScore^{PDB} und DrugScore^{CSD} unterscheiden sich hinsichtlich der experimentellen Kristallstrukturen, aus denen die Potentiale abgeleitet wurden: Während bei DrugScore^{PDB} Protein-Ligand Komplexe der PDB als Grundlage dienten, basieren die Potentiale von DrugScore^{CSD} auf einer deutlich höheren Anzahl kleiner kristallisierter Moleküle der Cambridge Structural Database (CSD, [226]). Eine Erweiterung des zugrunde liegenden Datensatzes und daraus resultierende neue Potentiale liegen in DrugScoreX vor. Während der Bearbeitung dieses Kapitels gab es eine Umstellung auf diese Version (Details siehe exp. Teil S. 138 ff.).

3.2.3.5. SZYBKI

Zur Komplexminimierung der Protein-Ligand-Komplexe wurde das Programm SZYBKI [227] verwendet. Es wurde das MMFF94s-Kraftfeld gewählt [228], das gegenüber der ansonsten sehr ähnlichen Variante MMFF94 optimierte Parameter für mesomoriestabilisierte planare (static) Stickstoffatome enthält.

Es wurden standardmäßig alle Aminosäuren im Bereich von 6.5 Å um den Liganden zur Minimierung freigegeben. Dabei wurde unterschieden zwischen der ausschließlichen Freigabe der Seitenketten und der Freigabe aller Atome in dem genannten Bereich. Der erstgenannte Ansatz erwies sich in Vorversuchen als überlegen und wurde angewandt.

3.3. Ergebnisse 1: Orientierendes Docking für Synthesevorschläge

Bei der Auswertung stand zunächst die visuelle Begutachtung der Posen im Vordergrund. Diese wurde für jede der ausgewählten Verbindungen (vgl. Abb. 3.4 auf S. 87) einzeln durchgeführt, indem die errechneten Posen betrachtet und auf übereinstimmende Ergebnisse in verschiedenen Proteinen (AutoDock) bzw. gleiche Lösungen der Docking-Programme überprüft wurden. In AutoDock wurden die von ADT im Rahmen der Analyse des Dockings gebildeten Cluster betrachtet, in FlexE wurden die ersten 100 Lösungen skriptgesteuert geclustert und in die visuelle Analyse einbezogen. Posen, die für mehrere Verbindungen mit guten Scores bei den unterschiedlichen Docking-Programmen abschnitten, wurden als vorläufige Hypothesen für den Bindungsmodus angesehen. Sie werden anhand von Beispielen im Folgenden erläutert.

3.3.1. Mögliche Bindungsmodi

Verbindungen der Klasse 1

In Tabelle 3.4 sind zur Übersicht wichtige Eckdaten der Dockingergebnisse von AutoDock und FlexE für die ausgewählten Moleküle der Klasse 1 aufgelistet. Für AutoDock sind die Ergebnisse der besten und der schlechtesten Lösung angegeben, zusätzlich das Protein, in dem diese Lösung erhalten wurde. Für FlexE sind die Anzahl der gefundenen Lösungen und die jeweils erzielten Scores der besten und der 100sten Pose (als schlechteste betrachtete) aufgeführt. Die Scores liegen mit -6.88 bis -10.10 kcal/mol für die beste Lösung in einem Bereich, der für aktive Inhibitoren vernünftig erscheint.⁵ Die schlechtesten Lösungen liegen mit Werten zwischen -3.64 und -6.96 kcal/mol deutlich niedriger und zeigen die vergleichbare Spreizung zwischen guten und schlechten Lösungen bei beiden Dockingprogrammen.⁶ Man erkennt zudem, dass für die beiden größeren Moleküle **40** und **69** gegenüber der kleineren Verbindung **31** die höheren Scores

⁵Eine freie Bindungsenthalpie ΔG von -10.10 kcal/mol entspricht nach

$$\Delta G = RT \ln K_i \quad (3.2)$$

einer Bindungskonstante K_i von knapp 1 $\mu\text{mol/L}$ (bei $T = 310 \text{ K}$, mit Gaskonstante $R = 8.13 \text{ J / mol K}$, 1 kcal = 4.1819 kJ). Auch wenn die experimentellen Ergebnisse als IC_{50} -Werte ermittelt wurden, kann dieses Ergebnis als vergleichbar zu den experimentell erhaltenen Werten eingeordnet werden.

⁶Nach Gleichung 3.2 entspricht ein Score von -6 kcal/mol bei gleichen Bedingungen einer Bindungskonstante von 60 $\mu\text{mol/L}$.

Tabelle 3.4. | Ergebnisse der Klasse 1 im ersten Docking. Aufgeführt sind die Scores der auf Position # gesetzten Lösung. Dabei ist der Score des angegebenen Programmes eine Abschätzung der freien Bindungsenthalpie, zur Vergleichbarkeit angegeben in kcal/mol. Während bei FlexE die Anzahl n der Lösungen (wie viele unterschiedliche Posen gefunden wurden) eine Aussage über den Verlauf des Dockings macht, kann die Anzahl bei AutoDock festgelegt werden und ist daher nicht aufgeführt.

	Docking mit AutoDock				Docking mit FlexE		
	#1	in Protein	#50	in Protein	n	#1	#100
31	-6.88	pfgsk_1j1b	-5.81	pfgsk_1q3d	481	-8.46	-3.64
40	-10.10	pfgsk_1q41	-6.95	pfgsk_2ow3	532	-10.06	-6.39
69	-10.10	pfgsk_1j1b	-6.96	pfgsk_1q3d	410	-10.10	-6.75

berechnet werden. Dies ist möglicherweise auf die größere Anzahl schwerer Atome und daraus resultierend die größere Anzahl an Interaktionsmöglichkeiten zurückzuführen (zur Problematik der Vergleichbarkeit siehe Abschnitt 3.1.2 auf S. 81 bzw. 3.3.2 auf S. 107).

Die Interpretation des Bindungsmodus von Klasse 1, insbesondere der Klasse 1a, geschieht vor dem Hintergrund der Röntgenstruktur des Komplexes von **31** in der *HsGSK-3 β* . Die Verhältnisse in der humanen Bindetasche sind in Abbildung 3.8 schematisch dargestellt, die Begrifflichkeit zur Orientierung wird bei der Beschreibung der verschiedenen Bindungsmodi wie dort aufgeführt verwendet. Aufgrund der Ähnlichkeit von *PfGSK-3* und *HsGSK-3 β* einerseits und der nicht vorhandenen Selektivität von Verbindungen der Klasse 1a andererseits ist der analoge Bindungsmodus allein durch die Kristallstruktur wahrscheinlich, er wird als Modus A bezeichnet.

Ein wichtiger Kontakt besteht in der Wasserstoffbrücke über das Amin des Liganden an C6 zum Carbonyl des Ile160. Die Nitrilgruppierung an C2 (bzw. das Carbonyl in Verbindungen der Klasse 1b) kann einen Kontakt zum Lys108 tief in der Bindetasche ausbilden. Der Chlor- bzw. Iodphenylrest an Position 4 zeigt in diesem Modus aus der Tasche hinaus, während zum Inneren der Tasche hin noch genügend Raum für ein Wassermolekül ist. Dieses bildet den wichtigen überbrückten Kontakt zur Hinge Region aus.

Posen der Verbindung **31** Der **Modus A** wird sowohl von FlexE als auch von AutoDock für **31** berechnet. AutoDock berechnet ihn beim Docking in pfgsk_2ow3 und pfgsk_1q3d, wobei die entsprechenden Lösungen im ersten Fall sowohl die häufigsten als auch die am besten bewerteten sind. Im Falle von pfgsk_1q3d hingegen wird der Modus A nur beim flexiblen Docking auch an Platz 1 gesetzt (Platz 33 bei rigidem Rezeptor). Die von FlexE am besten bewertete Lösung entspricht dem in Abbildung 3.11 auf Seite 105 dargestellten und als Modus C weiter unten beschriebenen Modus.

Bei den Dockings mit AutoDock werden häufig zwei Bindungsmodi erzeugt (**Modus Xa/b**, vgl. Abb. 3.9), die an dieser Stelle wegen ihrer Häufigkeit dargestellt werden sollen, auch wenn sie letztlich nicht bei den Hypothesen zum Bindungsmodus mit aufgenommen wurden. AutoDock generiert viele Varianten von Posen, bei denen die Liganden mit dem Schwefel S1 und dem

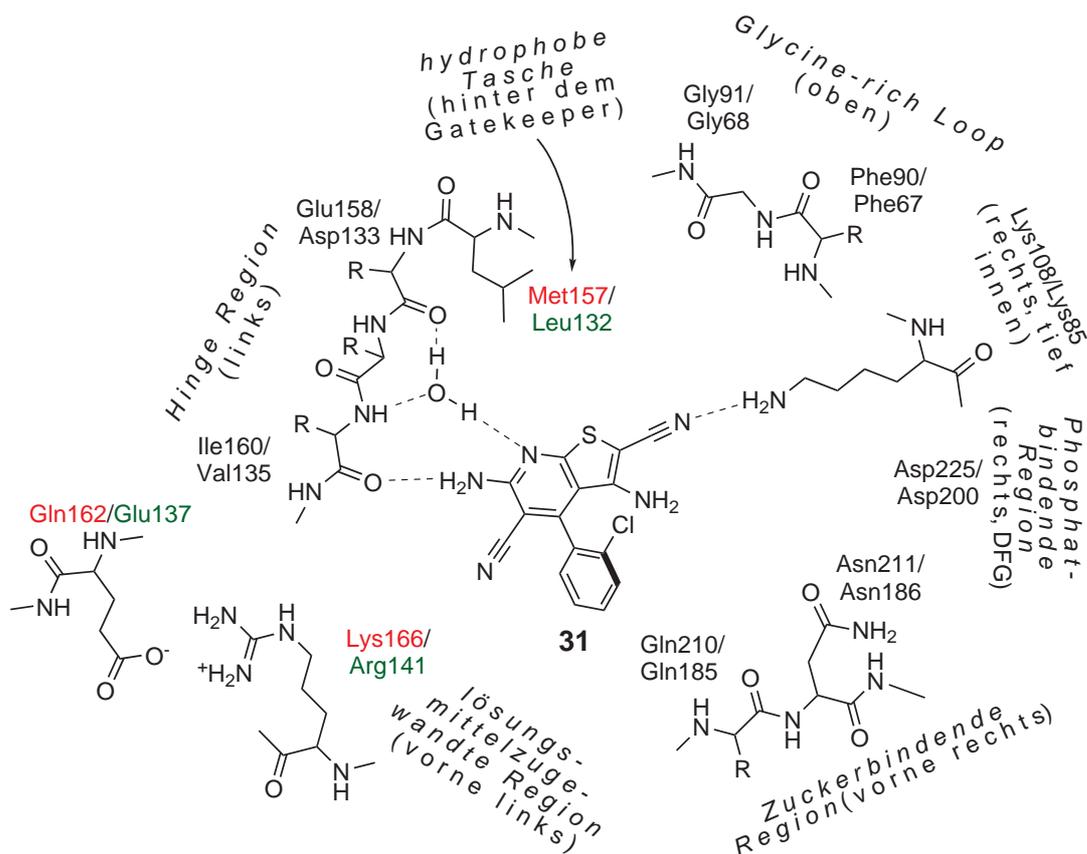


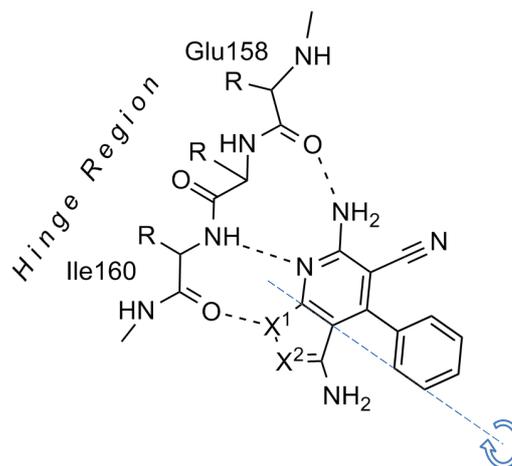
Abbildung 3.8. | Modus A: **31** in der Bindetasche der *HsGSK-3β* (Aminosäuren: *PfGSK-3/HsGSK-3β*). Die wichtige Interaktion des C2-Substituenten zum Lys108/85 tief in der Tasche sowie die überbrückte Wasserstoffbrückenbindung kennzeichnen Modus A. Das am Tascheneingang befindliche Arg141 des humanen Enzyms bildet in der Kristallstruktur eine Salzbrücke zum Glu137 aus – an dieser Position ist in der *PfGSK-3* Gln162, so dass zwar eine Wasserstoffbrückenbindung möglich ist, jedoch keine Salzbrücke.

Stickstoff N7 des Thienopyridins an der Hinge Region entlang parallel positioniert sind. Es kommen beide denkbaren Orientierungen vor, mit S1 in die Tasche hinein und aus der Tasche hinaus.

Dabei werden allerdings für die Wasserstoffbrückenbindungen der 6-Aminosubstitution in der Regel unrealistische Winkel erreicht. Optimalerweise sollten die Atome =O...HN- in einer Linie liegen. Tatsächlich weichen die Winkel mit 110-120° deutlich von den erwarteten 180° ab. Zudem verhindert der Schwefel S1 die Ausbildung von Wasserstoffbrückenbindungen zur Hinge Region. Er liegt vor dem Carbonyl des Ile160 und müsste demnach als Donor fungieren, damit eine Wasserstoffbrücke zustande kommen kann. Da dies nicht möglich ist, wird eine der typischen Wasserstoffbrückenbindungen von Kinaseinhibitoren verhindert. Dieser Bindungsmodus wird auch beim Docking der Verbindungen der Klasse 2 als Lösung angeboten, er wird aus den oben genannten Gründen allerdings auch bei der Klasse 2 nicht in Betracht gezogen.

Interessanterweise wurde dieser Modus auch von *CHIOUA et al.* für Inhibitoren der *HsGSK-3β* aufgrund einer Dockingstudie mit AutoDock postuliert [229]. Allerdings handelt es sich bei den

Abbildung 3.9. | Modus Xa. Die an der Bindung C3a-C7a gespiegelte Geometrie wird ebenfalls berechnet (Modus Xb). Allerdings ist für die Thienopyridine keine Interaktion des Schwefels an Position X¹ mit der Hinge Region möglich (für die in [229] beschriebenen 1*H*-Pyrazolo[3,4-*b*]pyridine mit Stickstoff an X¹ – und X² – ist dies der Fall), außerdem sind die Winkel der Wasserstoffbrückenbindung zwischen Aminogruppe und Glu158 ungünstig.



dort untersuchten Verbindungen um Grundkörper vom Typ der 1*H*-Pyrazolo[3,4-*b*]pyridine, die genau die oben beschriebenen Wasserstoffbrückenbindungen zur Hinge Region ausbilden können, wie es für die hier untersuchten Thienopyridine wegen des Schwefels an S1 *nicht* möglich ist (vgl. Abb. 3.9). In der zitierten Studie wurden zwei Verbindungen gedockt, weiterführende Angaben über die Untersuchungen sind der Veröffentlichung nicht zu entnehmen. Die Ergebnisse sind daher nicht auf den vorliegenden Datensatz übertragbar und können nicht als Argument für die Modi Xa/b gelten.

Posen der Verbindung 40 Die Dockings der Klasse 1b Inhibitors **40** mit FlexE ergeben insgesamt weniger plausible Lösungen. Es befindet sich zwar auf Rang 44 eine Entsprechung zu Modus A, die höchstgelistete Pose und viele weitere liegen jedoch relativ weit außerhalb der Bindetasche. In einer häufig vertretenen Pose befindet sich der an Position 4 des Grundkörpers sitzende Arylkern an der Stelle, an der der Purinkörper im ATP verschiedener Kristallstrukturen der *Hs*GSK-3 β positioniert ist. Dies erscheint insofern weniger vorteilhaft, als dass das starre planare Thienopyridin dann weiter außerhalb positioniert ist und nicht mehr in der Lage ist, zwischen Taschengrund und Glycine-rich Loop lipophile Interaktionen auszubilden.

Eine überaus interessante Pose ist jedoch auf Platz 63 zu finden: Hier wird über das Carbonyl, das in den Verbindungen **40** und **69** anstelle des Nitrils von **31** sitzt, eine Wasserstoffbrücke zum Ile160 in der Hinge Region ausgebildet. Der 4-(2-Iodphenyl)körper nimmt etwa die Position der Ribose des ATP ein und in der tieferen Tasche kann ein Kontakt des Nitrils an C5 zum Lys108 ausgebildet werden. Die Aminosubstitution an C6 zeigt zwischen dem Gatekeeper Met157 und dem Lys108 in die Richtung der Back Pocket. Die Anordnung der Stickstoffe von Lys108, Nitril an C5 und Aminosubstituent an C6 ist damit ungewöhnlich: Sie liegen alle mit 3.6 bis 3.9 Å in ähnlicher Distanz voneinander entfernt in Form eines Dreiecks. Auch wenn die Nähe der 6-Aminosubstitution zum Lys108 überrascht, scheint die Passgenauigkeit der Aminogruppe zwischen Met157 und Lys108 so gut, dass die Geometrie auch bei Dockings mit AutoDock mehrfach für **40** berechnet wird, insbesondere bei den Proteinen pfgsk_1j1b und pfgsk_1q41. Hinter dem Aromaten ist in Richtung der Hinge Region genug Platz für ein

Wassermolekül, das ähnlich wie in Modus A eine indirekte Wasserstoffbrücke des Liganden zum Protein vermitteln könnte. Allerdings steht mit dem Schwefel des Thienopyridins an dieser Position nur ein schwacher Wasserstoffbrückenakzeptor zur Verfügung. Der Bindungsmodus, der im Folgenden als **Modus B** bezeichnet wird, ist in Abbildung 3.10 dargestellt und erläutert.

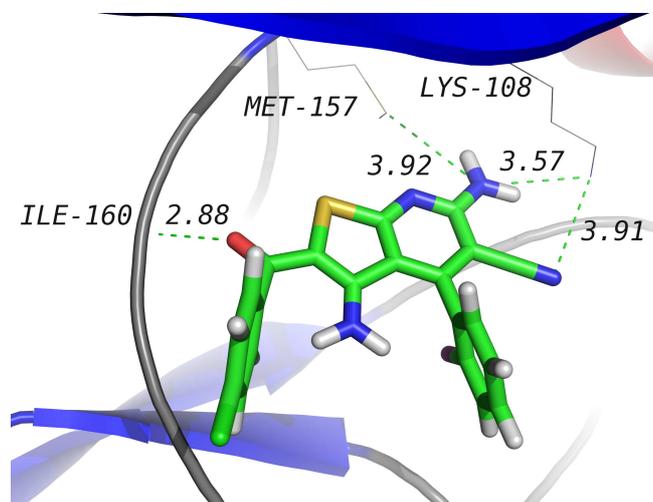


Abbildung 3.10. | Modus B am Beispiel von **40**. Die entscheidende Interaktion besteht zum Ile160 in der Hinge Region. Die Geometrie des Kontaktes zu Lys108 ist nicht optimal und auch die Entfernung ist zu groß für eine klassische Wasserstoffbrückenbindung, allerdings ist es vorstellbar, dass die flexible Seitenkette des Lys108 sich in eine vorteilhaftere Position bewegen kann. Die Aminsubstitution an C6 liegt genau zwischen Met157 und Lys108 und zeigt in Richtung der Back Pocket.

Posen der Verbindung **69** Die Verbindung **69** ist bis auf die Substitution am Benzoylrest mit **40** identisch. Erwartungsgemäß finden sich starke Übereinstimmungen zu den Dockingergebnissen von **40**: Der Modus A wird, wenn auch erst auf Rang 92, im FlexE-Docking ebenso vorgefunden wie der zuletzt diskutierte Modus B. Zusätzlich wurde ein weiterer Modus vorgefunden, bei dem der Ligand besonders tief in der Tasche sitzt. Er wird im Folgenden als **Modus C** bezeichnet und ist schematisch in Abbildung 3.11 dargestellt. In dieser Positionierung liegt der Stickstoff des 5-Cyanosubstituenten dort, wo in der Kristallstruktur von **31** mit der *HsGSK-3 β* Wasser kokristallisiert ist (vgl. Abb. 3.8 auf S. 102). Ein Verdrängen dieses Wassers würde durch die Stärkung des entropischen Anteils der Bindungsenthalpie zu höherer Affinität beitragen. Die Pose wird sowohl für **40** (FlexE) als auch für **69** (AutoDock in pfgsk_1q41) berechnet.

Nahe verwandt mit dem Modus C ist die gespiegelte Fassung **Modus D**, bei der das Molekül weiter außerhalb der Tasche liegt, aber nach wie vor den Kontakt zur Hinge Region ausbildet, der möglicherweise das Wasser verdrängen könnte. Auch wenn diese Lösungen relativ weit außerhalb der Tasche liegen, wurde der Modus D dennoch als Hypothese mit aufgenommen, da er sich in guter Übereinstimmung bei Dockings mit FlexE und AutoDock wiederfindet.

Verbindungen der Klasse 2

Erwartungsgemäß unterscheiden sich die Dockingergebnisse der Verbindungen der Klasse 2 wegen der unterschiedlichen Substitution am Grundkörper deutlich. Das zeigt sich bereits in der tabellarischen Übersicht der Ergebnisse (siehe Tab. 3.5). Insgesamt werden im Vergleich zum Docking der Verbindungen der Klasse 1 weniger Lösungen von FlexE mit schlechteren Scores gefunden. In AutoDock liegen die Werte für die berechneten Bindungsenergien der Vertreter der

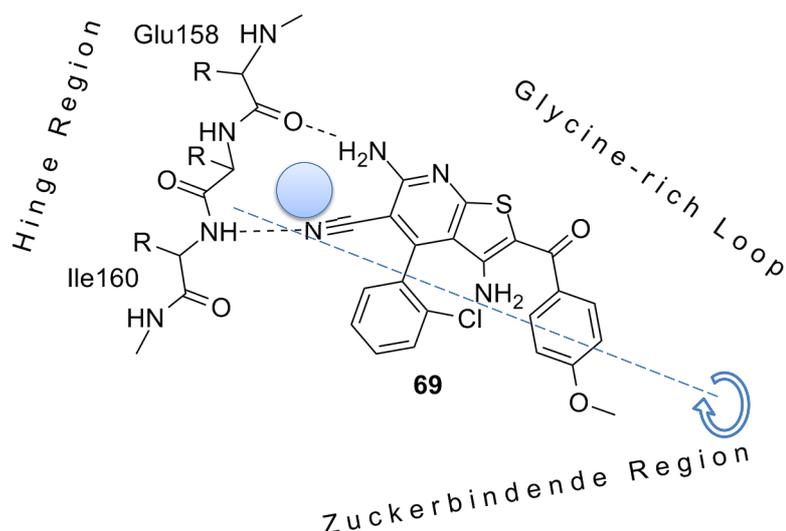


Abbildung 3.11. | Modus C und D am Beispiel von **69**. Die entscheidende Interaktion findet direkt zum Stickstoff des Ile160 in der Hinge Region statt. Es könnte das in der Kristallstruktur vorhandene Wasser (hellblauer Kreis) verdrängt werden, das sich etwa im Bereich der Cyanogruppe befindet. Es werden von **69** im abgebildeten Modus C teilweise Interaktionen der Carbonylgruppe an C2 mit Backbone-Stickstoffen im Glycine-rich Loops ausgebildet. Wird das Molekül an der gestrichelten Achse gespiegelt, ergibt sich Modus D (nicht abgebildet), der sich ebenfalls den gleichen Kontakt der Cyanogruppe auszeichnet, während die 5-Aminogruppe nun mit dem Carbonyl des Ile160 eine Wasserstoffbrückenbindung ausbilden kann. Die Verbindung kann mit ihrem Carbonyl-Sauerstoff eine Bindung zum Glycine-rich Loop (Modus C) bzw. zur zuckerbindenden Region (Modus D) ausbilden.

Klasse 2 hingegen in einem ähnlichen Bereich wie für Klasse 1. Hinsichtlich der beobachteten Bindungsmodi sind die im Folgenden beschriebenen Unterschiede zum Docking der Klasse 1 festzustellen.

Es findet sich bei allen vier Verbindungen zwar ein Bindungsmodus, der durch die Wasserstoffbrückenbindung des Carbonyls zum Lys108 gewisse Ähnlichkeit zu Modus A hat, das Thienopyridingerüst ist jedoch tiefer in die Tasche hinein verschoben, so dass der lipophile Ring diese praktisch ausfüllt. Die Geometrie wird als **Modus E** bezeichnet und ist in Abbildung 3.12 am Beispiel von **125** dargestellt.

Posen der Verbindung **125** Neben Modus E werden beim Docking von **125** mit FlexE insbesondere Lösungen produziert, die relativ weit außerhalb der Tasche liegen. Wie in Tabelle

Tabelle 3.5. | Übersicht der Ergebnisse für Klasse 2 in Analogie zu Tabelle 3.4 (Scores [kcal/mol]).

	Docking mit AutoDock				Docking mit FlexE		
	#1	in Protein	#50	in Protein	n	#1	#100
125	-9,48	pfgsk_1q41	-7,08	pfgsk_2ow3	284	-5,42	-4,12
132	-9,09	pfgsk_1j1b	-6,68	pfgsk_2ow3	384	-6,46	-4,67
140	-9,22	pfgsk_1j1b	-6,94	pfgsk_1q3d	195	-5,31	-3,82
141	-9,21	pfgsk_1j1b	-7,00	pfgsk_1q3d	495	-7,45	-5,34

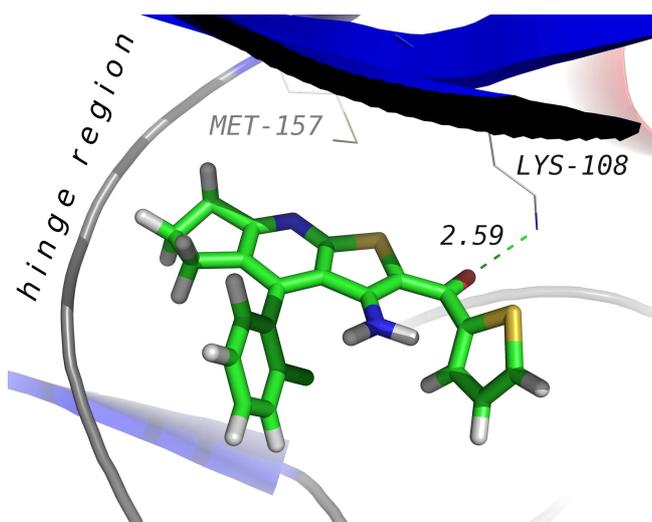


Abbildung 3.12. | Modus E der Klasse 2 am Beispiel von **125**. In dieser Variante des Modus A ist der lipophile Ring auf die Hinge Region ausgerichtet, der entscheidende Kontakt wird zum Lys108 ausgebildet. Im Unterschied zum Modus A sind die Verbindungen so weit in die Tasche hinein verschoben, dass keine über Wasser vermittelte Wasserstoffbrückenbindung mehr möglich ist.

3.5 dargestellt, fallen die FlexE Scores als Folge der dadurch weniger vorhandenen Kontakte insgesamt für die Klasse 2 im Vergleich zu den Werten der Klasse 1 und im Vergleich zu den durch AutoDock berechneten schlechter aus.

Posen der Verbindungen **132**, **140** und **141** Die Verbindungen **132**, **140** und **141** haben gegenüber den bislang betrachteten Verbindungen eine grundsätzliche strukturelle Abwandlung: Sie besitzen statt des Arylcarbonylkörpers an C2 ein Amid. Die Veränderung von Nitril (z. B. **31**) zu Arylcarbonyl (z. B. **40**) erbrachte im vorliegenden Datensatz die Selektivität der Verbindungen für die PfGSK-3, die Veränderung zum Amid hingegen scheint keine einheitliche Konsequenz zu haben – während **132** und **141** nicht selektiv sind, ist dies bei **140** der Fall (vgl. Tab. 3.1 auf S. 88). Der Unterschied scheint demnach bei diesen Verbindungen über die

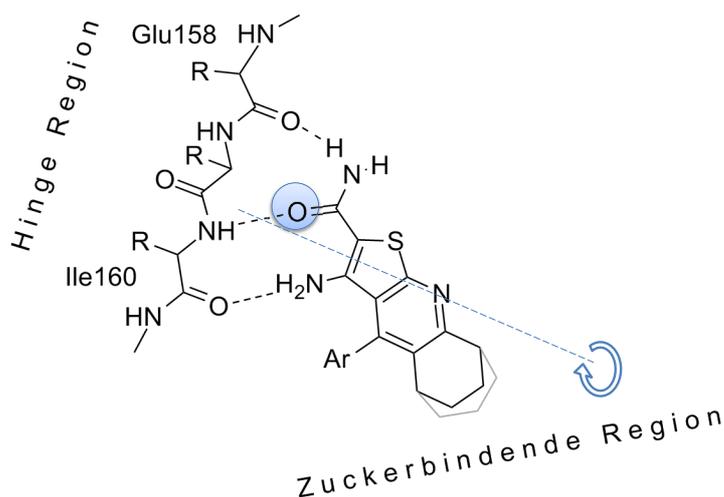


Abbildung 3.13. | Modus F und G der Verbindungen **132**, **140** und **141**. Im abgebildeten Modus F wird befindet ist der 4-Arylrest in Richtung der lösungsmittelzugewandten Region ausgerichtet, während der lipophile Ring im Bereich der zuckerbindenden Region liegt. Die an der hellblauen Linie gespiegelte Version wird als Modus G bezeichnet, hier können ebenfalls die drei eingezeichneten Wasserstoffbrückenbindungen ausgebildet werden, der anellierte lipophile Ring zeigt tief in die Tasche hinein in Richtung des Lys108. In beiden Fällen würde wieder das in der HsGSK-3 β beobachtete Wasser verdrängt werden.

Arylsubstitution am C4 zustande zu kommen. Die Dockings sollten helfen, diese Unterschiede zu erklären.

Vorherrschend waren für die drei genannten Verbindungen zwei Posen, die mit der Hinge Region die klassischen drei Wasserstoffbrückenbindungen nach dem Muster Donor-Akzeptor-Donor ausbilden, wie sie typischerweise in Kristallstrukturen von Proteinkinasen mit ATP vorliegen. Dabei wurden beide möglichen Orientierungen des Grundkörpers beobachtet, in Abbildung 3.13 ist schematisch die als **Modus F** bezeichnete Version mit aus der Tasche hinausweisendem 4-Arylrest gezeigt – **Modus G** ist die gespiegelte Form, bei der der 4-Arylrest in die Tasche hineinzeigt. Daneben wurden wieder häufig die bereits bei Klasse 1 als Modus Xa/b diskutierten und verworfenen Hypothesen beobachtet. Insgesamt konnte aus den Ergebnissen dieses Dockings keine Erklärung für die drastischen Unterschiede hinsichtlich der Selektivität bei Veränderung des 4-Arylrestes (**140** \Rightarrow **141**) abgeleitet werden.

3.3.2. Bewertung der Hypothesen

3.3.2.1. Bindungsmodi der Klasse 1

Um möglichst gezielte Synthesvorschläge machen zu können, sollte als nächstes eine Rangfolge der Hypothesen erstellt werden, an denen sich die Vorschläge orientieren sollten. Dazu wurden AutoDock-Lösungen mit dem Scoring von FlexX bewertet und andersherum, zusätzlich wurde DrugScore^{CSD} herangezogen (Rescoring, vollständige Ergebnisse im oberen Teil von Tab. 3.6). Dieses Bewertungsschema wurde für repräsentative Beispiele der jeweiligen Modi durchgeführt, wobei jeweils (a) der generierte Komplex und (b) der minimierte Komplex bewertet wurde.

Bei der Beurteilung der Scores sind zwei Unterschiede zu beachten: Einerseits sind die absoluten Ergebnisse nicht vergleichbar, da sie mit unterschiedlichen Energiefunktionen berechnet wurden. Dies ist bei den Werten von DrugScore gegenüber den anderen beiden Methoden besonders augenfällig, gilt aber für die Gegenüberstellung der Werte von FlexX und AutoDock ebenso, auch wenn hier in beiden Fällen Bindungsenergien in kcal/mol angegeben werden. Andererseits haben verschiedene Verbindungen innerhalb *einer* Bewertungsweise aufgrund unterschiedlicher Eigenschaften (Molmasse bzw. Anzahl schwerer Atome und resultierend Anzahl an Interaktionspunkten bzw. Oberfläche) andere Voraussetzungen bei der Berechnung des Scores. Diesen Unterschied gibt es bei der Bewertung von verschiedenen Posen *einer* Verbindung naturgemäß nicht.

Um dennoch zu einer vergleichenden Bewertung der Posen zu kommen, wurden die Ergebnisse der unterschiedlichen Scoringfunktionen s für die 6 Bewertungsmethoden (FlexX, AutoDock, DrugScore, jeweils vor und nach Minimierung) auf den Bereich zwischen dem schlechtesten Wert s_{worst} und dem besten Wert s_{best} normiert (vgl. Tab. 3.6, unterer Teil). Der normierte Score s_{norm} wurde nach Gleichung 3.3 aus dem tatsächlichen Score s_{act} und dem schlechtesten

3. Der Bindungsmodus von Inhibitoren der PfGSK-3

s_{worst} bzw. besten s_{best} berechnet.

$$s_{norm} = \frac{s_{act} - s_{worst}}{s_{best} - s_{worst}} \quad (3.3)$$

Auf diese Weise werden Werte zwischen 0 und 1 erhalten, die die Positionierung im Bereich zwischen dem schlechtesten und dem besten Wert angeben. Dadurch kann ein Überblick erhalten werden, wie die Modi für verschiedene Verbindungen von den unterschiedlichen Scoringfunktionen relativ zueinander eingeordnet werden, das Ergebnis ist grafisch in Abbildung 3.14 dargestellt. Die hohen absoluten Werte, die sich bei der Bewertung von nicht minimierten FlexE-Lösungen ergaben, wurden nicht in die Normierung einbezogen. Hier wurde die unminimierten FlexE-Lösungen aus der Berechnung ausgeschlossen und $s_{worst} = 0$ gesetzt. Der extrem schlechte Wert von s_{worst} führt ansonsten dazu, dass außer bei der schlechtesten Lösung zwei

Tabelle 3.6. | (Re-)Scoring der Klasse 1. Im unteren Teil der Tabelle sind die nach Gleichung 3.3 normierten Scores aufgeführt.

Hypothese	Komplex	Original Pose			Staged Minimization		
		DS	FlexX	AD4	DS	FlexX	AD4
Modus A	40_kubra_FlexE	189663	14211	48905033	-143149	-14.40	-35.94
	31_kubra_AD4	-102832	-9.54	-29.85	-147581	-12.80	-31.87
Modus B	40_1j1b_AD4	-110570	-16.60	-36.19	-172641	-25.82	-45.49
	40_1q41_AD4	-149191	-22.06	-42.23	-197092	-28.55	-46.16
	69_1j1b_AD4	-159910	-27.28	-39.03	-173910	-18.12	-42.62
Modus C	40_1q41_AD	-43308	-15.04	-38.60	-146014	-18.01	-40.43
	40_1q41_FlexE	137052	-0.21	28157	-116986	-21.44	-43.17
	69_1q41_FlexE	70172	-0.58	363.83	-156058	-23.81	-43.53
Modus D	69_1q3d_ADflex	-127526	-9.02	-27.60	-149703	-11.11	-26.67
	69_1q3d_FlexE	-6886	-5.21	1158	-156475	-8.58	-28.62
Modus A	40_kubra_FlexE	-/-	-/-	-/-	0.33	0.29	0.48
	31_kubra_AD4	0.64	0.35	0.71	0.38	0.21	0.27
Modus B	40_1j1b_AD4	0.69	0.61	0.86	0.69	0.86	0.97
	40_1q41_AD	0.93	0.81	1.00	1.00	1.00	1.00
	69_1j1b_AD4	1.00	1.00	0.92	0.71	0.48	0.82
Modus C	40_1q41_AD	0.27	0.55	0.91	0.36	0.47	0.71
	40_1q41_FlexE	-/-	0.01	-/-	0.00	0.64	0.85
	69_1q41_FlexE	-/-	0.02	-/-	0.49	0.76	0.87
Modus D	69_1q3d_ADflex	0.80	0.33	0.65	0.41	0.13	0.00
	69_1q3d_FlexE	0.04	0.19	-/-	0.49	0.00	0.10

DS: DrugScore; AD4: AutoDock4; Bezeichnung der Posen: <Verbindung>_<Vorlage Homologie-Modell>_<Docking-Programm> (Abkürzungen in Tabelle 3.7 entsprechend).

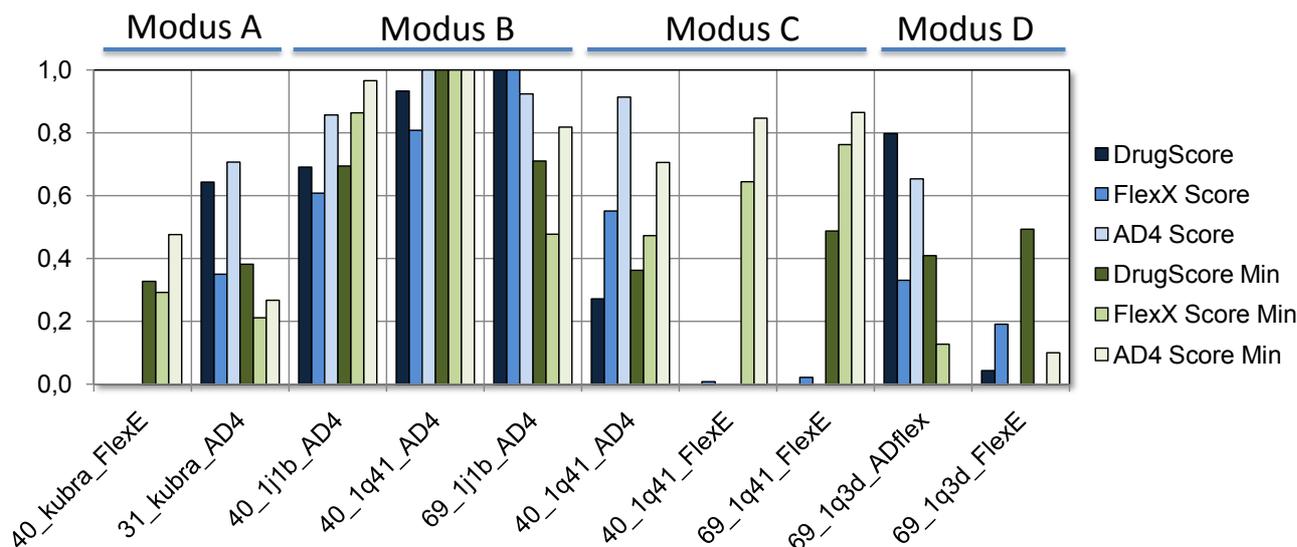


Abbildung 3.14. | Normiertes Rescoring der Verbindungen der Klasse 1. Die blauen Balken stehen für Scores vor der Minimierung, grüne für diejenigen nach der Minimierung (Min). Ein normierter Score von 1 bedeutet, dass für diesen Komplex die beste Bewertung im jeweiligen Scoringsschema berechnet wurde. Die Bezeichnung der Beispiele für die verschiedenen Posen erfolgt nach <Verbindung>_<Vorlage Homologie-Modell>_<Docking-Programm> (AD4: AutoDock4; ADflex: AutoDock4 mit flexiblen Seitenketten). Es wird sehr deutlich, dass die mit FlexE erzielten Lösungen vor der Minimierung keine konkurrenzfähigen Bewertungen erzielen. Hier wurde anstelle der United Protein Structure ein Protein eingesetzt, das den gleichen Modus erlaubt. Zudem ist klar zu erkennen, dass unterschiedliche Scoringfunktionen die Modi sehr unterschiedlich einstufen. Modus C wird beispielsweise nach der Minimierung von AutoDock und FlexX gut, von DrugScore hingegen schlecht bewertet.

extrem große Beträge als Differenzen in Zähler *und* Nenner von $\frac{s_{act}-s_{worst}}{s_{best}-s_{worst}}$ stehen. Dies hat zur Folge, dass nicht unterscheidbare gute Werte nahe 1 für alle Lösungen außer der schlechtesten berechnet werden und die Vergleichbarkeit auch der nicht mit FlexE gedockten nicht minimierten Lösungen verloren geht.

Modus B wird bei diesem Vorgehen für die eingesetzten Moleküle der Klasse 1 als beste Hypothese bewertet. Die besten Scores der 6 Bewertungsmethoden (3 Funktionen, jeweils vor und nach Minimierung) werden von einer Pose nach dieser Hypothese erzielt. Modus C folgt dahinter, wenn die minimierten Komplexe betrachtet werden und hier wiederum insbesondere die Scoringfunktionen von AutoDock und FlexX. Hinsichtlich des DrugScore-Wertes sind die Modi weniger klar voneinander abgrenzbar. Betrachtet man wiederum die minimierten Ergebnisse (dunkelgrüne Balken in Abb. 3.14), erzielt Modus D Ergebnisse, die denen von Modus A leicht überlegen sind. Allerdings ist hier zu bemerken, dass für die Modi A und D nur unterschiedliche Verbindungen erhalten wurden, so dass ein direkter Vergleich schwer fällt. Für **31** wird der Modus B nicht erzeugt. Da die Modi Xa/b als Alternativen nicht in Betracht gezogen werden, wird für **31** Modus A angenommen.

Insgesamt werden bei der Betrachtung einige problematische Punkte deutlich, die eine klare Entscheidung für einen einzigen Bindungsmodus zu diesem Zeitpunkt nicht erlauben:

- Es werden für keinen Modus Beispiele jeder Verbindung erhalten, so dass nicht der direkte Vergleich der Scoringwerte einer Verbindung in allen unterschiedlichen Modi möglich ist.
- Es liegen Daten für wenige Verbindungen vor, so gibt es nur einen Vertreter der Klasse 1a und die beiden Vertreter der Klasse 1b sind sich so ähnlich, dass sich die Frage stellt, wie weit die Ergebnisse auf den gesamten Datensatz übertragen werden können.
- Die Bewertung der Komplexe, die durch FlexE entstanden sind, bereitet Probleme hinsichtlich der Vergleichbarkeit. Die Tatsache, dass FlexE-Posen in Kombination mit einem der ursprünglichen Proteine extrem schlechte Werte beim Scoring mit verschiedenen Strukturen ergeben, deutet darauf hin, dass die United Protein Structure weit von nativen Proteinen entfernt ist. Die Konstruktion derselben war mit den im experimentellen Teil beschriebenen Einstellungen möglich, führte offenbar aber nicht immer zu Strukturen, die den Modellen so ähnlich waren, dass durch die durchgeführte Komplexminimierung die Überführung in einen günstigen Protein-Ligand-Komplex gelang.
- Die Verwendung mehrerer Proteine beim Docking mit AutoDock ist ein Vorteil im Hinblick auf die Möglichkeit, dadurch Flexibilität des Proteins zu imitieren. Allerdings bleiben im gewählten Ansatz zwei Probleme bestehen: Einerseits ist nicht klar, welches Ergebnis (aus welchem Protein) vorzuziehen ist, da die verschiedenen Scoringfunktionen widersprüchliche Ergebnisse liefern. Andererseits wurden nicht alle Homologie-Modelle berücksichtigt, da aus organisatorischen Gründen eine zeitnahe Auswertung gefordert war. Mit allen Proteinen wäre eine gründliche visuelle Kontrolle der zahlreichen Posen und eine Nutzung von DrugScore^{CSD} im Online-Verfahren problematisch gewesen, so so dass der Vorteil nicht voll ausgenutzt werden konnte.

Diese Punkte sollten durch die Entwicklung eines optimierten, stark erweiterten Arbeitsablaufes verbessert werden. Aus den vorliegenden Daten wurde für die Synthesevorschläge als Arbeitshypothese für Verbindungen der Klasse 1b der Modus B als wahrscheinlicher Bindungsmodus angenommen, da er wie oben dargestellt Vorteile gegenüber den anderen Modi aufweist. Für die Klasse 1a wurde Modus A angenommen.

3.3.2.2. Bindungsmodi der Klasse 2

Die beim (Re-)Scoring der Posen von Komplexen mit Verbindungen der Klasse 2 berechneten Werte sind in Tabelle 3.7 aufgeführt. Zum Vergleich der Hypothesen wurden in Entsprechung zum vorangegangenen Abschnitt normierte Werte nach Gleichung 3.3 berechnet (vgl. Tab. 3.7, unterer Teil), die in Abbildung 3.15 dargestellt sind.

Es ist klar zu erkennen, dass die verschiedenen Scoringfunktionen den Modus G als schlechtesten einordnen. Differenzierter sind die Unterschiede zwischen den anderen beiden Modi: Während die Verbindungen mit anelliertem gesättigten Siebenring **140** und **141** bessere Bewertun-

Tabelle 3.7. | (Re-)Scoring der Klasse 2 in Entsprechung zu Tabelle 3.6.

Hypothese	Komplex	Original Pose			Staged Minimization		
		DS	FlexX	AD4	DS	FlexX	AD4
Modus E	141_2ow3_AD4	-66692	-2,65	-26,99	-140308	-10,89	-26,68
	125_1j1b_AD4	-161131	-7,75	-36,75	-179414	-18,60	-38,16
	132_2ow3_AD4	-134745	-13,13	-35,53	-199375	-15,04	-34,52
	140_1q3d_AD4	-88842	-8,42	-30,34	-170986	-11,55	-28,94
Modus F	125_1q3d_AD4	-124754	-9,70	-35,67	-177265	-17,57	-34,08
	140_1j1b_AD4	-106684	-14,66	-38,57	-185672	-15,44	-35,23
	141_1j1b_AD4	-113873	-11,23	-37,20	-153883	-17,96	-34,83
Modus G	140_1q3d_AD4	-67692	-8,77	-31,88	-137579	-12,17	-29,60
	132_1j1b_AD4	-66178	-1,67	-25,81	-118177	-13,08	-29,92
	141_1q3d_AD4	-79122	-10,47	-32,23	-165576	-16,27	-35,21
Modus E	141_2ow3_AD4	0,01	0,08	0,09	0,27	0,00	0,00
	125_1j1b_AD4	1,00	0,47	0,86	0,75	1,00	1,00
	132_2ow3_AD4	0,72	0,88	0,76	1,00	0,54	0,68
	140_1q3d_AD4	0,24	0,52	0,36	0,65	0,09	0,20
Modus F	125_1q3d_AD4	0,62	0,62	0,77	0,73	0,87	0,64
	140_1j1b_AD4	0,43	1,00	1,00	0,83	0,59	0,74
	141_1j1b_AD4	0,50	0,74	0,89	0,44	0,92	0,71
Modus G	140_1q3d_AD4	0,02	0,55	0,48	0,24	0,17	0,25
	132_1j1b_AD4	0,00	0,00	0,00	0,00	0,28	0,28
	141_1q3d_AD4	0,14	0,68	0,50	0,58	0,70	0,74

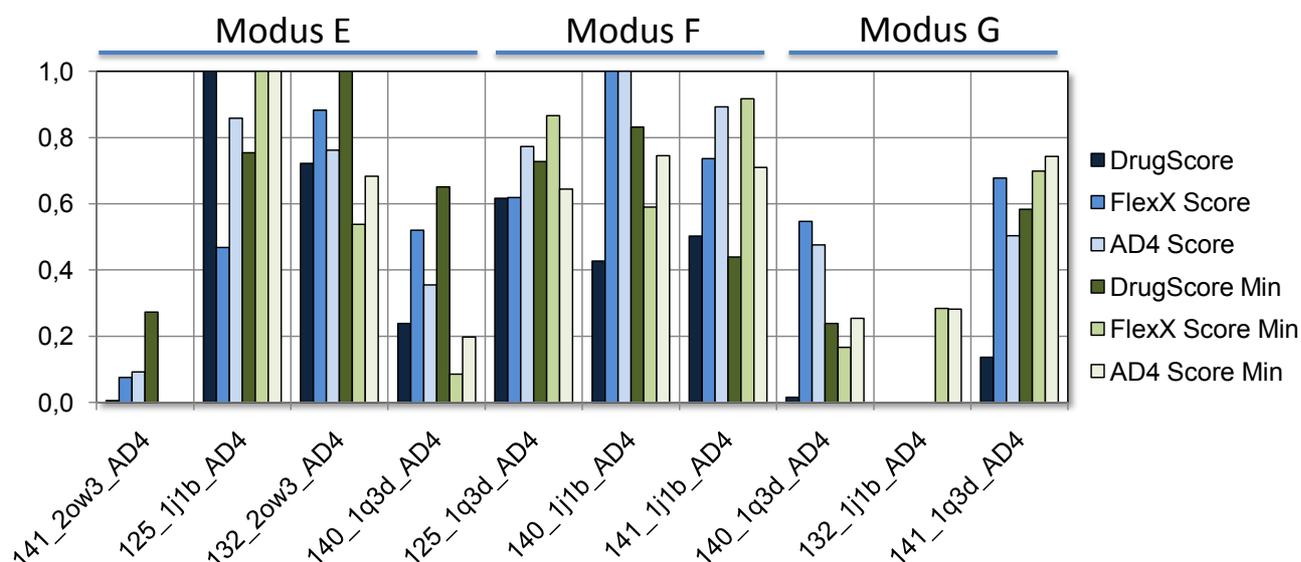


Abbildung 3.15. | Normiertes Rescoring der Verbindungen der Klasse 2. Die Auftragung der normierten Scores entspricht Abbildung 3.14. Vergleichsweise klar fällt der Modus G gegenüber den anderen beiden Modi ab. Auffällig ist außerdem die gleichmäßigere Verteilung der Verbindungen über die Modi, so ist jede der eingesetzten vier Verbindungen in mindestens zwei Modi vertreten. Für Modus E gibt es Beispiele von jeder der eingesetzten Verbindungen.

gen im Modus F erzielen, werden die beiden anderen Moleküle mit den kleineren Ringsystemen besser im Modus E bewertet. Der Grund ist in Abbildung 3.12 auf Seite 106 gut zu erkennen: Für die beiden Verbindungen mit den großen Ringsystemen ist nicht ausreichend Platz zur Hinge Region, während die Verbindungen **125** und **132** mit ihrem anellierten Cyclopentan bzw. -hexan klein genug sind. Wird der Modus G aufgrund des Scorings abgelehnt, bleiben bei der Abwägung zwischen den Modi E und F die im vorangegangenen Abschnitt genannten Einschränkungen zu bedenken. An diesem Punkt der Arbeit kann als Argument für Modus E die Nähe zur experimentellen Kristallstruktur festgehalten werden, außerdem gibt es für diesen Modus Beispiele von jeder der eingesetzten Verbindungen. Modus E wird daher zunächst als Arbeitshypothese für den Bindungsmodus von Verbindungen der Klasse 2 angenommen.

3.3.2.3. Zwischenfazit

Die ersten Dockings erlauben noch keine Festlegung auf einen Bindungsmodus. Auch eine Abgrenzung zwischen selektiven und unselektiven Verbindungen innerhalb der Klasse 2 ist nicht möglich. Es können allerdings einige interessante Beobachtungen festgehalten werden:

- Für Verbindungen der Klassen 1 und 2 werden grundsätzlich verschiedene Dockingergebnisse erhalten.
- Innerhalb der Klasse 1 gibt es Unterschiede zwischen der Verbindung **31** als Vertreter der Klasse 1a und den beiden Vertretern der Klasse 1b (**40** und **69**).
- Aus diesen Unterschieden werden Arbeitshypothesen für Bindungsmodi abgeleitet (vgl. Abb. 3.16):
 - Klasse 1a \Rightarrow Modus A
 - Klasse 1b \Rightarrow Modus B
 - Klasse 2 \Rightarrow Modus E
- Die einzige Hypothese, die einen gleichen oder ähnlichen Bindungsmodus beider Klassen erlaubt, wäre eine der Kristallstruktur nahe Kombination des Modus A mit einer zum Taschenausgang hin verschobenen (allerdings in den Ergebnissen nicht beobachteten) Variante von E.
- Die kleine Anzahl von Verbindungen und die extremen Schwankungen hinsichtlich der Bewertung der Hypothesen durch die verschiedenen Scoring-Funktionen müssen einschränkend festgehalten werden.
- Für ein späteres Docking konnten daraus Hinweise zur besseren Durchführung abgeleitet werden:

- Vom Einsatz von FlexE sollte abgesehen werden, wenn Protein-Ligand-Komplexe aufgearbeitet werden sollen.
- Es sollten mehr Verbindungen insbesondere der Klasse 1a eingesetzt werden, um die Unterschiede zwischen 1a und 1b näher zu untersuchen.
- Grundsätzlich sollte ein Test mit bekannten Röntgenstrukturen vor dem eigentlichen Verfahren durchgeführt werden, um die Aussagekraft der Ergebnisse besser einordnen zu können.

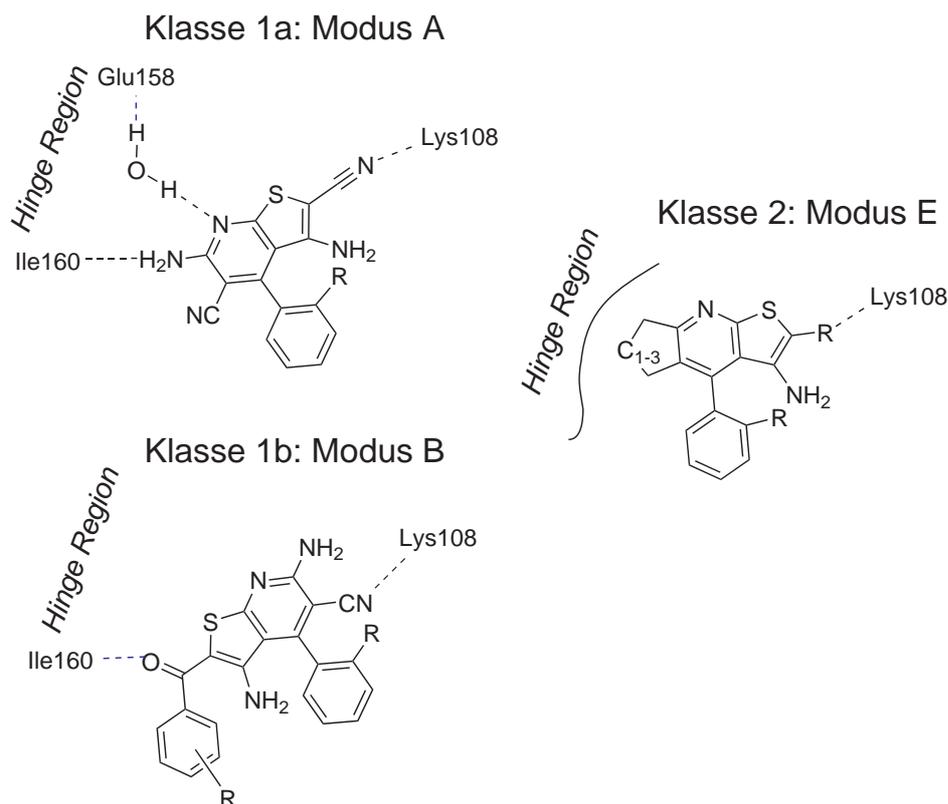
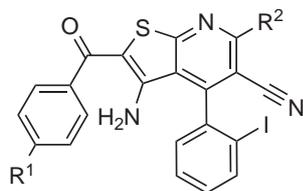


Abbildung 3.16. | Übersicht der Arbeitshypothesen zum Bindungsmodus.

3.3.3. Resultierende Synthesevorschläge

Bei den resultierenden Synthesevorschlägen wurden zwei grundsätzliche Vorgehensweisen verfolgt: Entweder sollte eine als entscheidend angesehene Interaktion eines Bindungsmodus blockiert werden – die resultierende Verbindung müsste dann deutlich weniger Affinität zeigen und die Hypothese des entsprechenden Bindungsmodus würde bestärkt. Oder aber die Veränderung in der Struktur sollte neue oder bessere Interaktionen erlauben – wenn die entsprechende Verbindung tatsächlich bessere biologische Daten zeigen sollte, wäre dies wiederum ein Hinweis auf die Richtigkeit der Hypothese.

Die Vorschläge wurden in Entsprechung zu Abschnitt 3.2.2.1 gedockt, primäre Fragestellung war dabei: Finden sich unter den Posen solche, die im gleichen Bindungsmodus orientiert sind,



R ¹	R ²	Name
Cl	CH ₃	40 _mod00
Cl	NHCH ₃	40 _mod01a
Cl	N(CH ₃) ₂	40 _mod01b
Cl	OH	40 _mod02a
Cl	OCH ₃	40 _mod02b
OCH ₃	CH ₃	69 _mod00
OCH ₃	NHCH ₃	69 _mod01a
OCH ₃	N(CH ₃) ₂	69 _mod01b
OCH ₃	OH	69 _mod02a
OCH ₃	OCH ₃	69 _mod02b

Abbildung 3.17. | Modifikationen (1) des Amins am C6.

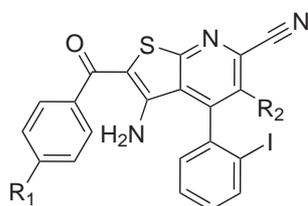
werden Hypothesen zum Bindungsmodus also bestärkt oder nicht? Über die im Folgenden beschriebenen Vorschläge wurde Rücksprache mit Prof. Kunick gehalten und eine Entscheidung für präparative Arbeiten getroffen. Dabei wurden im Rahmen dieser Arbeit nur Variationen der Klasse 1 erwogen.

Grundsätzlich wurden die Varianten ausgehend von den Verbindungen **40** und **69** entwickelt. Modifikation (1) betrifft das Amin (bzw. Amidin) an C2, die konstruierten und gedockten Verbindungen sind in Abbildung 3.17 dargestellt. Da das Amin in den Modi A und D Teil einer Wasserstoffbrückenbindung ist und in Modus B sehr eng am Protein in der tiefer gelegenen Bindetasche zwischen Met157 und Lys108 liegt, sollte eine zweifache Methylierung des Amins die aufgeführten Modi praktisch unmöglich machen. Eine einfache Methylierung sollte möglich sein, allerdings im Falle von Modus A und D zu Affinitätsverlust führen, während bei Modus B günstigenfalls eine Ausrichtung in die Back Pocket möglich wäre, die sich vorteilhaft auf die Aktivität auswirken könnte. Ein Austausch der Amino- gegen eine Hydroxylgruppe sollte die beschriebenen Modi weiterhin ermöglichen, wohingegen Methoxy die für die Methylierung des Amins geschilderten Folgen hat. Der Ersatz des Amins durch Methyl führt schließlich zu einer sterisch ähnlichen Gestalt des Moleküls bei Wegfall der Möglichkeit einer Wasserstoffbrückenbindung – dieses Derivat wäre als Knock-Out Variante einzustufen, ebenso wie die 6-Dimethylsubstitution.

Generell befinden sich unter den Dockingergebnissen der Verbindungen der Modifikation (1) sowohl Posen im Modus A als auch im Modus B. Dabei liegen die hydroxylierten- und die monomethylierten Varianten bei vergleichbaren Scores sehr ähnlich zu den Ausgangsverbindungen, während für das dimethylierte Molekül erwartungsgemäß keine Entsprechung zu den Hypothesen gefunden wird. Ein Syntheserversuch dieser Verbindungen wird als sinnvoll angesehen.

In der Modifikation (2) sollten die Substituenten an C5 und C6 ausgetauscht und das Amin dann entsprechend der ersten Modifikation variiert werden (siehe Abbildung 3.18). Der Effekt könnte zur Optimierung der Bindungswinkel im Falle von Modus B beitragen und bei Vorliegen

von Modus D für die nicht methylierten Moleküle eventuelle Vorteile in der Interaktion mit der Hinge Region erbringen.

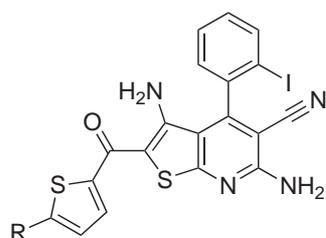


R ¹	R ²	Name
Cl	NH ₂	40 _mod03a
Cl	NHCH ₃	40 _mod03b
Cl	N(CH ₃) ₂	40 _mod03c
Cl	OH	40 _mod04a
Cl	OCH ₃	40 _mod04b
OCH ₃	NH ₂	69 _mod03a
OCH ₃	NHCH ₃	69 _mod03b
OCH ₃	N(CH ₃) ₂	69 _mod03c
OCH ₃	OH	69 _mod04a
OCH ₃	OCH ₃	69 _mod04b

Abbildung 3.18. | Modifikationen (2): Austausch der Funktionalität an C5 und C6.

Für diese Varianten werden im Falle der Modifikationen von **40** beim unmethylierten und monomethylierten Amin von AutoDock Posen im Modus B gefunden, während das hydroxylierte Derivat nicht in den beschriebenen Modi gedockt wird. Das dimethylierte **40**_mod03c wird von FlexE in Modus A gedockt – da durch den Tausch der Funktionalitäten an C5 und C6 das Amin aus der Tasche hinauszeigt, wird diese Orientierung im Gegensatz zu den Gegebenheiten bei der Modifikation (1) möglich. FlexE findet bei den Verbindungen mod3a mit unsubstituiertem Amin jeweils den Modus C. Für **69**_mod3b bis **69**_mod4b werden von FlexE keine akzeptablen Lösungen innerhalb der Bindetasche berechnet. Die Darstellung und Testung der Modifikationen (2) mit Amin an C5 könnte folglich den Bindungsmodus bestätigen.

Die Modifikation (3) sieht den Austausch des über ein Carbonyl an den Grundkörper gebundenen aromatischen Kerns an C2 vor, die Verbindungen sind in Abbildung 3.19 dargestellt. Da im vorliegenden Datensatz zwar Verbindungen mit Thiophen an dieser Position existieren, diese jedoch in der Substitution an C5 und C6 sämtlich der Klasse 2 zuzuordnen sind, ist die Modifikation bereits ohne Betrachtung der Hypothesen zum Bindungsmodus von Interesse. Darüber hinaus würde dieser Teil des Moleküls bei Vorliegen von Modus B in eine Region zeigen, die



R	Name
H	40 _mod13a
Cl	40 _mod13b
OCH ₃	40 _mod13c
OC ₂ H ₅	40 _mod13d
CN	40 _mod13e
CF ₃	40 _mod13f

Abbildung 3.19. | Modifikationen (3): Thiophen als Arylcarbonylsubstituent an C2.

hinsichtlich der Selektivität Potential bietet, so dass sie unter Annahme dieser Hypothese zu selektiven Inhibitoren führen sollte.

Die Verbindungen werden mit AutoDock bei guten Scores überwiegend in Modus B und teilweise in Modus A gedockt, während FlexE vorrangig Posen in den Modi C und D findet. Diese Modifikation wird folglich ebenfalls als interessant eingestuft.

Die Modifikationen (4) zielen auf eine Veränderung der Verbindung zwischen den Aromaten ab (vgl. Abbildung 3.20). Durch den Ersatz des Carbonyl durch eine Methylengruppe wird die Möglichkeit einer Wasserstoffbrückenbindung unterbunden, so dass diese Modifikation ohne ihre zentrale Interaktion in jedem vorgeschlagenen Modus mit dem Protein nicht mehr binden sollte. In der zweiten Kombination dieser Gruppe wird durch das Oxim eine Kombination von Donor und Akzeptor möglich. Auch wenn eine vergleichbare Interaktion beim Indirubin (**36**) im Komplex mit der *Hs*GSK-3 β nicht beobachtet wird, wurde die Modifikation gedockt, um zu untersuchen, inwieweit eine solche Interaktion als möglich vorhergesagt wird.

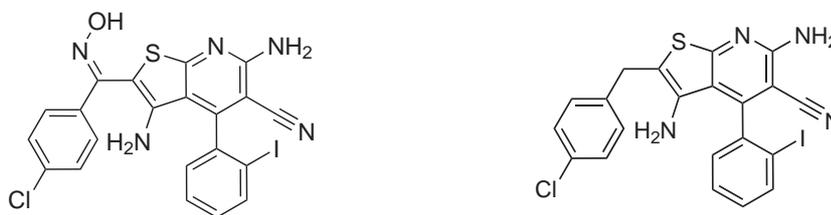


Abbildung 3.20. | Modifikationen (4): Oxim anstelle von Carbonyl (**40_modOx**, links) und Reduktion zur Methylengruppierung (**40_modMet**, rechts).

Da beim Oxim EZ-Isomerie vorliegt, wurden beide möglichen Verbindungen konstruiert und gedockt. Für beide Isomere wird allerdings von FlexE keine Lösung im Modus B gefunden, von AutoDock nur mit einem Protein (*pfgsk_1q41*) und bei schlechten Scores und ungewöhnlicher Konformation des Oxims – die Modifikation wird als weniger gut bewertet. Die Methylengruppierung ergibt zwar Posen, die nahe am Modus B liegen. Die Scores sind erwartungsgemäß vergleichsweise niedrig, die Synthese dieser Verbindung und eine anschließende Testung könnte zeigen, dass der Carbonylgruppe tatsächlich eine zentrale Rolle bei der Bindung zur PfGSK-3 zukommt.

3.4. Präparative Umsetzung der Synthesevorschläge

Im Rahmen der präparativen Arbeiten in Braunschweig sollte eine der im vorangehenden Abschnitt beschriebenen Modifikationen (1) bis (4) dargestellt werden. Die dabei dargestellten Substanzen wurden in Analogie zum vorliegenden Datensatz auf ihre Aktivität an *Pf*GSK-3 und *Hs*GSK-3 β getestet. Die Ergebnisse sollten einerseits bei der Bewertung der im ersten Docking erhaltenen Hypothesen zum Bindungsmodus helfen und andererseits den Datensatz erweitern. Es wurden die Modifikationen (3) mit Chlor und Iod als Halogen am 4-Arylrest ausgewählt (vgl. Abb. 3.21). Durch die Darstellung dieser beiden Halogensubstituenten sollte die Konsistenz mit beschriebenen Struktur-Wirkungsbeziehungen untersucht werden. Dabei sollte der Carbonylthiophensubstituent an C2 unsubstituiert (**40_mod13a** in Abb. 3.19 auf S. 115) und mit 5-Cyano-Substitution am Thiophen (**40_mod13e** in Abb. 3.19 auf S. 115) eingeführt werden.

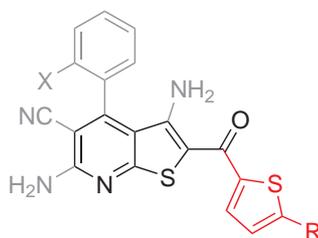
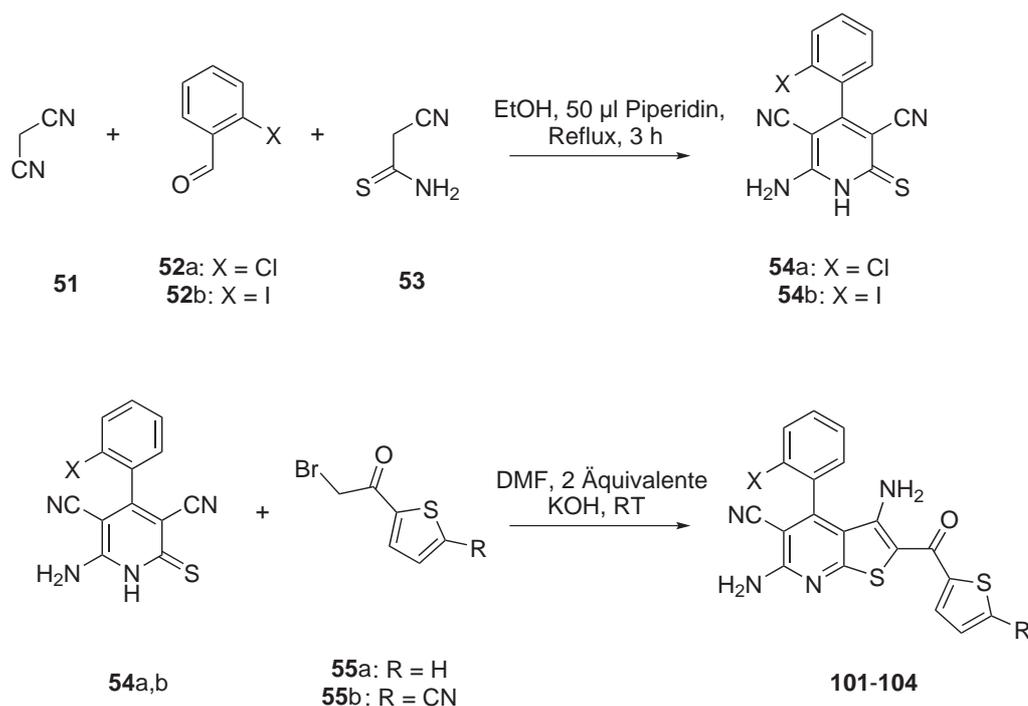


Abbildung 3.21. | Umgesetzte Varianten der Modifikation (3) von 2-(Thiophen-2-ylcarbonyl)thieno[2,3-*b*]pyridin-5-carbonitrilen. Bei der Modifikation (3) wird der Arylrest bioisoster gegen Thiophenylcarbonyl ausgetauscht – ein Substitutionsmuster, das in Kombination mit der 6-Amino-5-cyano-substitution bislang nicht beschrieben ist. Durch den Wechsel zum fünfgliedrigen Aromaten ergeben sich für weitere Substituenten andere Winkel als im Phenylrest. Die Modifikation ist darüber hinaus besonders interessant, da die variierte Region des Moleküls im Falle des für die Klasse 1b favorisierten Bindungsmodus B im lösungsmittelzugewandten Bereich liegt, der verschiedenen MIFs und der Sequenz zufolge für die Selektivität des Liganden von Interesse sein könnte. Es wurden für X die Halogene Chlor und Iod am 4-Arylrest des Grundkörpers und –H bzw. Cyano für R am Thiophen dargestellt.

3.4.1. Synthese von 2-(Thiophen-2-ylcarbonyl)-thieno[2,3-*b*]pyridin-5-carbonitrilen

Für die Synthese der Zielverbindungen **101-104** wird zunächst in Anlehnung an literaturbekannte Vorschriften [92, 230] eine Eintopfreaktion mit Malonodinitril (**51**), dem betreffenden Arylaldehyd (**52a,b**) und 2-Cyanothioacetamid (**53**) durchgeführt. Die erhaltenen 6-Amino-4-aryl-2-thioxo-1,2-dihydro-3,5-pyridindicarbonitrile (**54a,b**) werden in einem zweiten Schritt nach SHARANIN *et al.* [231] mit den betreffenden 2-Brom-1-(thiophen-2-yl)ethanonen (**55a,b**) zu **101-104** umgesetzt (siehe Schema 3.1).

Schema 3.1 | Zweistufige Synthese von **101-104**.

Die Reaktionen verliefen glatt mit Ausbeuten zwischen 15.1 und 55.3% (vgl. exp. Teil). Die Verbindungen **101-104** wurden in die Laboratorien von L. Meijer (CNRS, Station Biologique de Roscoff) getestet. Die Ergebnisse werden im folgenden Abschnitt dargestellt.

	X	R
101	I	H
102	I	CN
103	Cl	H
104	Cl	CN

3.4.2. Biologische Daten

Die Ergebnisse der biologischen Tests sind in Tabelle 3.22 aufgeführt. Die Verbindungen zeigen Aktivität an der PfGSK-3, die in allen Fällen stärker ist als die an der HsGSK-3 β . Die Werte liegen allerdings nur im mittleren mikromolaren Bereich. Die Verbindungen sind damit im Vergleich zu **40** (1.6 μ M) weniger stark wirksam.

Hinsichtlich der Struktur-Wirkungsbeziehungen geben die Ergebnisse ein widersprüchliches Bild ab. So erbringt der Wechsel von I zu Cl in einem Fall (ohne Cyanosubstitution, **101** \Rightarrow **103**) eine Verringerung der Aktivität, im anderen ein geringes Ansteigen auf niedrigem Niveau (mit Cyanosubstitution, **102** \Rightarrow **104**). Diese letztgenannte Veränderung ist in Übereinstimmung mit den im bislang vorliegenden Datensatz beobachteten Tendenzen, während die Verbesserung durch den Iodsubstituenten den in Abschnitt 3.2.1.1 auf Seite 84 geschilderten Beobachtungen widerspricht.

Die Cyanosubstitution am Thiophen wiederum bedeutet bei der 2-iod-4-arylsubstituierten Verbindung eine Verschlechterung des IC₅₀ um den Faktor 6 (**101** \Rightarrow **102**), während es bei dem 2-chlor-4-arylsubstituierten Derivat zu einer marginale Verbesserung kommt, wiederum

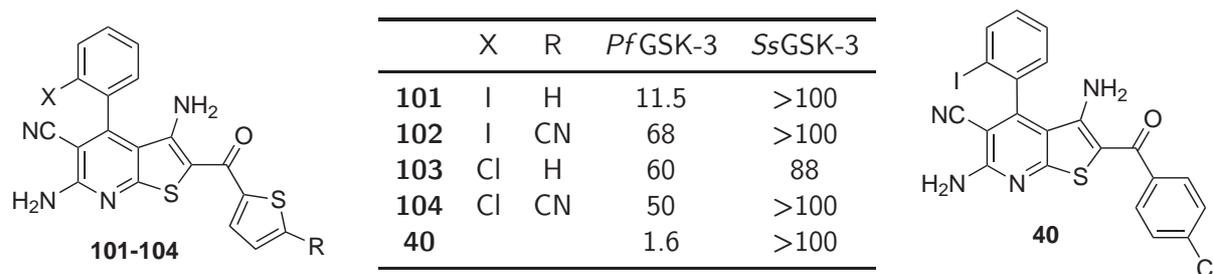


Abbildung 3.22. | Biologische Daten der Verbindungen **101-104** und **40** (Angaben des IC_{50} an *Pf*GSK-3 und *Ss*GSK-3 in μ M). Es wurden jeweils Dreifachmessungen durchgeführt, die Standardabweichungen der Aktivitäten bewegte sich in den Einzelmessungen zwischen 0.55 und 12.05 % (zum Assay siehe Abschnitt 3.2.1.2 auf S. 85).

auf niedrigem Niveau (**103** \Rightarrow **104**). Die letztgenannte Entwicklung ist in Übereinstimmung zu den bekannten Struktur-Wirkungsbeziehungen: Die 3-Cyanosubstitution des Benzoylrestes in **83** führt zu einem IC_{50} von 0.5μ M. Dieser stellt eine Verbesserung dar gegenüber Derivaten, die am 2-Benzoylrest unsubstituiert sind (z. B. **72**, **76**). **83** ist ebenso aktiv an der *Pf*GSK-3 wie die beste Verbindung der Klasse 1b aus der gesamten betrachteten Serie (**77** mit IC_{50} von 0.48μ M).

3.4.3. Ergebnis der Synthesen

Es konnten mit **101-104** vier bislang nicht beschriebene Derivate der Thienopyridine dargestellt werden, die als Verbindungen der Klasse 1b selektiv (**103** präferentiell) die *Pf*GSK-3 inhibieren. Die im vorigen Abschnitt beschriebenen Verhältnisse zu den bekannten Struktur-Wirkungsbeziehungen lassen allerdings keine klare Bestätigung eines Bindungsmodus zu, da sie sowohl in sich als auch zu den bekannten Zusammenhängen (vgl. Abschnitt 3.2.1.1 auf S. 84) teilweise im Widerspruch stehen. Der Informationsgewinn durch die Substanzen ist somit in erster Linie in der Erweiterung des Datensatzes zu sehen.

3.4.4. Experimenteller Teil

3.4.4.1. Allgemeines

Dünnschichtchromatographie:

DC-Mikrokarten Polygram[®]Sil G/UV₂₅₄, 40 x 80 mm, Fa. Macherey-Nagel; Detektion unter UV-Licht bei 254 und 366 nm.

Elementaranalysen:

CE Instruments FlashEA[®]1112 Elemental Analyser, Fa. Thermo Quest.

Analytische Hochleistungs-Flüssigchromatographie (HPLC):

Merck/Hitachi La Chrom Elite: Pump L-2130, Autosampler L-2200, Diode Array Detector L-2450, OrganizerBox L-2000, Fa. Merck/Hitachi, Darmstadt.

Trennsäule Merck LiChroCART 125-4, LiChrosphere 100, RP 18, 5 μ m, Fa. Merck/Hitachi, Darmstadt.

Probenvorbereitung: 1 mg Substanz wird in 1 ml DMSO gelöst, die erhaltene Lösung 4:6 mit DMSO verdünnt.

Injektionsvolumen: 10 μ l; Flussrate 1.000 ml/min; isokratisch; Detektion: 254/280 nm; AUC-%-Methode; Gesamtlaufzeit 15 min; Nettoretentionszeit (t_s) und Totzeit (t_m) bezogen auf DMSO; Elutionsmittel: ACN:H₂O - Gemisch 55:45.

IR-Spektren:

Thermo Nicolet FT-IR 200 (vermessen als KBr-Pressling), Fa. Thermo Nicolet, Madison, WI, USA.

Kernresonanzspektrometrie:

¹H-NMR (DMSO-*d*₆, 600 MHz) und ¹³C-NMR (DMSO-*d*₆, 150.9 MHz):

Bruker Avance DRX-400 bzw. Bruker Avance II-600 (NMR-Laboratorien der Chemischen Institute der TU Braunschweig); interner Standard: Tetramethylsilan (TMS); Angaben der Verschiebung in ppm nach der δ_{TMS} -Skala; Ermittlung der Protonenverhältnisse durch Integration; Abkürzung für Signalmultiplizitäten: s = Singulett, d = Dublett, t = Triplett, q = Quartett, m = Multiplett, (e)br = (extrem) breites Signal.

Schmelzpunkte:

Ia 9100, Fa. Barnstead, Southend-on-Sea, Essex, Großbritannien.

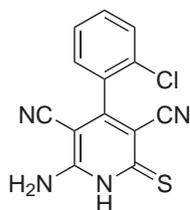
3.4.4.2. Synthesen und analytische Daten

3.4.4.2.1. Allgemeine Arbeitsvorschrift 1 (AAV1) zur Synthese der 6-Amino-4-aryl-2-thioxo-1,2-dihydropyridin-3,5-dicarbonitrile 54a,b

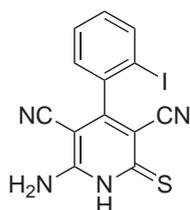
Der entsprechende aromatische Aldehyd (1.0 mmol), 2-Cyanothioacetamid (1.0 mmol) und Malonodinitril (1.0 mmol) werden mit 50 μ l Piperidin in Ethanol (4 ml) 3 Stunden refluxiert. Der Fortgang der Reaktion wird dünnschichtchromatographisch überwacht. Nach Beendigung wird das Lösungsmittel am Rotationsverdampfer im Vakuum eingeeengt und einige Tropfen Essigsäure, 10 ml Wasser und 2 ml Dichlormethan hinzugefügt. Der Ansatz wird gründlich gemischt und für 12 Stunden bei 0-7 °C gelagert. Der entstandene Niederschlag wird abgesaugt und direkt in AAV2 eingesetzt.

3.4.4.2.2. Allgemeine Arbeitsvorschrift 2 (AAV2) zur Synthese der 2-(Thiophen-2-ylcarbonyl)-thieno[2,3-*b*]pyridine 101-104

Ein 6-Amino-4-aryl-2-thioxo-1,2-dihydropyridin-3,5-dicarbonitril (0.4 mmol) wird in 0.4 ml DMF gelöst. 224 μ l einer 10 %igen wässrigen KOH Lösung (0.4 mmol) werden hinzugegeben. Nach einer Minute wird das ggf. substituierte 2-Brom-1-(thiophen-2-yl)ethanon (0.4 mmol) zugesetzt. Bei Raumtemperatur wird 30 Minuten gerührt, dann werden weitere 224 μ l der 10 %igen wässrigen KOH Lösung (0.4 mmol) hinzugetropft und der Ansatz wird bis zu 4 Stunden weiter gerührt, wobei die Beendigung der Reaktion aufgrund dünnschichtchromatographischer Überwachung festgelegt wird und in Abhängigkeit von den beteiligten Reaktanden deutlich vorher erreicht sein kann. Der Ansatz wird dann in Wasser gegeben, der ausgefallene Niederschlag wird abgesaugt und aus Ethanol oder Ethylacetat umkristallisiert.

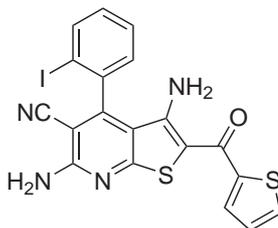
6-Amino-4-(2-chlorphenyl)-2-thioxo-1,2-dihydro-pyridin-3,5-dicarbonitril (54a)

Durchführung nach AAV1 aus 2-Chlorbenzaldehyd (702.9 mg, 5.000 mmol), 2-Cyanothioacetamid (500.7 mg, 5.000 mmol) und Malonodinitril (330.3 mg, 5.000 mmol) mit 250 μ l Piperidin in 20 ml Ethanol. Nach Aufarbeitung gemäß AAV1 werden 793.4 mg (Ausbeute: 55.3 %) erhalten und direkt weiter eingesetzt.

6-Amino-4-(2-iodphenyl)-2-thioxo-1,2-dihydropyridin-3,5-dicarbonitril (54b)

Durchführung nach AAV 1 aus 2-Iodbenzaldehyd (465.5 mg, 2.001 mmol), 2-Cyanothioacetamid (202.7 mg, 2.024 mmol) und Malonodinitril (132.1 mg, 2.001 mmol) mit 100 μ l Piperidin in 8 ml Ethanol. Nach Aufarbeitung gemäß AAV1 werden 279.5 mg (Ausbeute: 37.0 %) erhalten und direkt weiter eingesetzt.

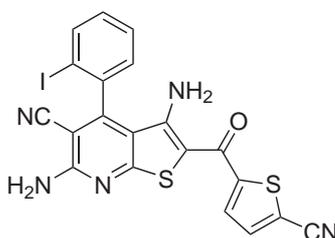
3,6-Diamino-4-(2-iodphenyl)-2-(thiophen-2-ylcarbonyl)-thieno[2,3-*b*]pyridin-5-carbonitril (101)



Durchführung nach AAV2 aus 6-Amino-4-(2-iodphenyl)-2-thioxo-1,2-dihydro-pyridin-3,5-dicarbonitril (177.0 mg, 0.470 mmol), Kaliumhydroxid-Lösung (10 %, zweimal 258 μ l, 0.920 mmol) und 2-Brom-1-(thiophen-2-yl)ethanon (96.4 mg, 0.470 mmol). Nach Umkristallisation aus Ethanol erhält man 68.5 mg (29.0 %) dunkelgelbes Pulver.

Schmp.: 268-271 °C; IR (KBr): 3448 cm^{-1} , 3292 cm^{-1} und 3210 cm^{-1} (NH), 2213 cm^{-1} (C \equiv N), 1622 cm^{-1} (C=O); $^1\text{H-NMR}$ (DMSO- d_6 , 600 MHz): δ (ppm) 6.63 (ebr s, 2H, NH $_2$)⁷, 7.25 (dd, 1 H, $J=4.9, 1.1$ Hz, ArH (Thiophen)), 7.40 (td, 1 H, $J=7.7, 1.7$ Hz, ArH (Iodphenyl)), 7.60 (dd, 1 H, $J=7.6, 1.6$ Hz, ArH (Iodphenyl)), 7.69 (td, 1 H, $J=7.7, 1.5$ Hz, ArH (Iodphenyl)), 7.82 (br s, 2H, NH $_2$)⁸, 7.87 (dd, 1 H, $J=3.9, 1.0$ Hz, ArH (Thiophen)), 7.93 (dd, 1 H, $J=4.9, 1.1$ Hz, ArH (Thiophen)), 8.14 (dd, 1 H, $J=8.09, 0.94$ Hz, ArH (Iodphenyl)); $^{13}\text{C-NMR}$ (DMSO- d_6 , 150.9 MHz): δ (ppm) 90.6, 97.3, 97.9, 111.9, 114.7, 128.3, 128.7, 129.3, 129.8, 131.9, 132.5, 138.0, 139.4, 145.1, 151.5, 155.0, 159.3, 165.8, 177.7; C $_{19}$ H $_{11}$ IN $_4$ OS $_2$ [502.35]; ber. C 45.43, H 2.21, N 11.15, S 12.77; gef. C 45.41, H 2.25, N 11.42, S 12.52; HPLC: 98.3 % bei 254 nm und 98.6 % bei 280 nm, $t_s = 5.73$ min, $t_m = 1.02$ min (ACN/H $_2$ O 55:45), λ_{max} : 335 nm.

3,6-Diamino-2-[(5-cyanothiophen-2-yl)carbonyl]-4-(2-iodphenyl)-thieno[2,3-*b*]pyridin-5-carbonitril (102)



Durchführung nach AAV2 aus 6-Amino-4-(2-iodphenyl)-2-thioxo-1,2-dihydro-pyridin-3,5-dicarbonitril (189.1 mg, 0.500 mmol), Kaliumhydroxid-Lösung (10 %, zweimal 274 μ l, 0.977

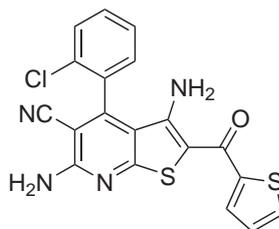
⁷Das Signal für die Aminogruppe an C3 ist extrem breit. Während es bei den Verbindungen **101** und **103** mit unsubstituiertem Thiophen noch bei Verschiebung von 5.25 bis 7.25 integrierbar ist, kann es bei den Verbindungen **102** und **104** nicht detektiert werden.

⁸Dieses Signal kann der Aminogruppe an C6 zugeordnet werden, da es bei den Verbindungen **54a** und **54b** bei ähnlichen Verschiebungen detektiert wurde.

mmol) und 5-(Bromacetyl)thiophen-2-carbonitril (115.0 mg, 0.500 mmol). Nach Umkristallisation aus Ethylacetat erhält man 60.8 mg (23.1 %) rötliches Pulver.

Schmp.: 311-315 °C; IR (KBr): 3441 cm^{-1} , 3346 cm^{-1} und 3230 cm^{-1} (NH), 2219 cm^{-1} ($\text{C}\equiv\text{N}$), 1631 cm^{-1} ($\text{C}=\text{O}$); $^1\text{H-NMR}$ ($\text{DMSO-}d_6$, 600 MHz): δ (ppm) 7.41 (td, 1 H, $J=7.6$, 1.6 Hz, ArH (Iodphenyl)), 7.60 (dd, 1 H, $J=7.6$, 1.6 Hz, ArH (Iodphenyl)), 7.70 (td, 1 H, $J=7.5$, 1.1 Hz, ArH (Iodphenyl)), 7.90 (dd, 1 H, $J=4.1$ Hz, ArH (Thiophen)), 7.95 (br s, 2H, NH_2), 8.05 (dd, 1 H, $J=4.1$ Hz, ArH (Thiophen)), 8.15 (dd, 1 H $J=8.1$, 0.9 Hz, ArH (Iodphenyl))⁹; $^{13}\text{C-NMR}$ ($\text{DMSO-}d_6$, 150.9 MHz): δ (ppm) 90.9, 97.1, 97.8, 111.4, 112.3, 113.8, 114.5, 128.7, 129.1, 129.5, 132.0, 137.7, 139.5, 139.7, 151.9, 153.0, 155.4, 159.6, 166.3, 175.4; $\text{C}_{20}\text{H}_{10}\text{IN}_5\text{OS}_2$ [527.36]; ber. C 45.55, H 1.91, N 13.28, S 12.16; gef. C 45.71, H 2.07, N 13.33, S 11.77; HPLC: 99.1 % bei 254 nm und 99.0 % bei 280 nm, $t_s = 6.53$ min, $t_m = 1.02$ min (ACN/ H_2O 55:45), λ_{max} : 344 nm.

3,6-Diamino-4-(2-chlorphenyl)-2-(thiophen-2-ylcarbonyl)-thieno[2,3-*b*]pyridin-5-carbonitril (103)

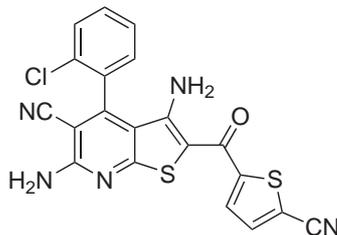


Durchführung nach AAV2 aus 6-Amino-4-(2-chlorphenyl)-2-thioxo-1,2-dihydro-pyridin-3,5-dicarbonitril (229.4 mg, 0.800 mmol), Kaliumhydroxid-Lösung (10 %, zweimal 450 μl , 1.604 mmol) und 2-Brom-1-(thiophen-2-yl)-ethanon (164.1 mg, 0.800 mmol). Nach Umkristallisation aus Ethylacetat erhält man 49.6 mg (15.1 %) rötliches Pulver.

Schmp.: 280-285 °C; IR (KBr): 3458 cm^{-1} , 3306 cm^{-1} , 3252 cm^{-1} und 3144 cm^{-1} (NH), 2216 cm^{-1} ($\text{C}\equiv\text{N}$), 1642 cm^{-1} ($\text{C}=\text{O}$); $^1\text{H-NMR}$ ($\text{DMSO-}d_6$, 600 MHz): δ (ppm) 6.70 (ebr s, 2H, NH_2), 7.26 (dd, 1 H, $J=4.9$, 3.8 Hz, ArH (Thiophen)), 7.62 - 7.72 (m, 3 H, ArH (Chlorphenyl)), 7.80 (dd, 1 H, $J=8.1$, 1.1 Hz, ArH (Chlorphenyl)), 7.82 (br s, 2H, NH_2), 7.87 (dd, 1 H, $J=3.8$, 0.9 Hz, ArH (Thiophen)), 7.93 (dd, 1 H, $J=5.1$, 1.1 Hz, ArH (Thiophen)); $^{13}\text{C-NMR}$ ($\text{DMSO-}d_6$, 150.9 MHz): δ (ppm) 90.8, 98.1, 112.2, 114.7, 128.4, 128.4, 129.9 (2C), 130.3, 131.0, 131.7, 132.3, 132.5, 145.1, 150.4, 151.6, 159.3, 165.9, 177.8; $\text{C}_{19}\text{H}_{11}\text{ClN}_4\text{OS}_2$ [410.9]; ber. C 55.54, H 2.70, N 13.64, S 15.61; gef.: C 55.45, H 2.76, N 13.42, S 15.48; HPLC: 97.7 % bei 254 nm und 98.3 % bei 280 nm, $t_s = 5.13$ min, $t_m = 1.03$ min (ACN/ H_2O 55:45), λ_{max} : 335 nm.

⁹Fehlendes Signal für NH_2 : siehe $^1\text{H-NMR}$ von **101**.

3,6-Diamino-4-(2-chlorphenyl)-2-[(5-cyanothiophen-2-yl)-carbonyl]thieno[2,3-*b*]pyridin-5-carbonitril (**104**)



Durchführung nach AAV2 aus 6-Amino-4-(2-chlorphenyl)-2-thioxo-1,2-dihydro-pyridin-3,5-dicarbonitril (563.4 mg, 1.965 mmol), Kaliumhydroxid-Lösung (10 %, zweimal 1.097 ml, 3.910 mmol) und 5-(Bromacetyl)thiophen-2-carbonitril (451.0 mg, 1.960 mmol). Nach Umkristallisation aus Ethylacetat erhält man 346.0 mg (40.5 %) rötliches Pulver.

Schmp.: 303-305 °C; IR (KBr): 3472 cm⁻¹, 3448 cm⁻¹, 3307 cm⁻¹, 3251 cm⁻¹ und 3138 (NH), 2218 cm⁻¹ (C≡N), 1642 cm⁻¹ (C=O); ¹H-NMR (DMSO-*d*₆, 600 MHz): δ (ppm) 7.63 - 7.73 (m, 3H, ArH (Chlorphenyl)), 7.81 (d, 1 H, *J*=8.1 Hz, ArH (Chlorphenyl)), 7.90 (d, 1 H, ArH (Thiophen)), 7.95 (br s, 2 H, NH₂), 8.05 (d, 1H, ArH (Thiophen))¹⁰; ¹³C-NMR (DMSO-*d*₆, 150.9 MHz): δ (ppm) 91.1, 98.0, 111.6, 112.4, 113.8, 114.6, 128.5, 129.1, 129.9, 130.4, 130.9, 131.5, 132.4, 139.7, 150.8, 151.9, 153.1, 159.5, 166.4, 175.5; C₂₀H₁₀ClN₅OS₂ [435.9]; ber. C 55.11, H 2.31, N 16.07, S 14.71; gef. C 55.01, H 2.34, N 15.95, S 14.74; HPLC: 97.8 % bei 254 nm und 97.7 % bei 280 nm, *t*_s = 5.79 min, *t*_m = 1.03 min (ACN/H₂O 55:45), λ_{max}: 344 nm.

¹⁰Fehlendes Signal für NH₂: siehe ¹H-NMR von **101**.

3.5. Ergebnisse 2: Entwicklung eines optimierten Workflows und Anwendung auf Thienopyridine in der PFGSK-3

3.5.1. Evaluierung des Verfahrens DoReCCoM

Der in Abbildung 3.5 auf Seite 90 dargestellte Workflow sieht als ersten Schritt ein Docking mit den drei Programmen FlexX, AutoDock und FRED vor. Die Ergebnisse dieses Dockings und die Auswirkungen des Rescorings mit DrugScoreX werden im folgenden Abschnitt für jedes Dockingprogramm beschrieben. Dann folgt ein Clustering der gesamten Lösungsmenge, wodurch mehrfach von verschiedenen Programmen für ein Protein erzeugte Posen herausgefiltert werden, auch dieses Zwischenergebnis wird für sich betrachtet und ab Seite 128 beschrieben. Der letzte Schritt des Workflows ist die Komplexminimierung, deren Ergebnisse dann ab Seite 129 dargelegt werden.

3.5.1.1. Erster und zweiter Schritt: Ergebnisse der einzelnen Dockingprogramme und deren Rescoring

Ausgangspunkt der Suche nach dem Bindungsmodus sind die Protein-Ligand-Komplexe, die im Rahmen des Crossdockings von den unterschiedlichen Programmen erzeugt werden. Es liegt eine Matrix von 12 (Proteine) mal 9 (Liganden, Ausschluss dreier Liganden vgl. S. 92 ff.) mit insgesamt 108 Kombinationen von Ligand und Protein vor. Für jede Kombination wurde mit jedem der drei Programme eine Anzahl von Lösungsmöglichkeiten erzeugt.¹¹ Für jedes einzelne Docking gibt es Fragen, deren Beantwortung Aussagen über die Qualität des Ergebnisses macht.

- Wie weit weicht die beste Lösung, die das Programm erzeugt, von der Position des Liganden in der Kristallstruktur ab? Ist das Programm also überhaupt in der Lage eine Geometrie zu erzeugen, die der experimentell festgestellten nahe kommt – das Docking-Problem.
- Welche Abweichung zum kokristallisierten Liganden weist die Lösung auf, die vom Programm als beste eingeordnet wird? Diese Frage zielt vor allem auf die Qualität der Bewertung unterschiedlicher Lösungen ab – das Scoring-Problem.

Um diese Fragen vergleichend zwischen den Programmen zu beantworten, wurden vier Kennzahlen berechnet, die sich jeweils auf die Menge der 108 Crossdocking-Experimente (Kombinationen aus Protein und Ligand) beziehen:

¹¹Diese Anzahl wird bei AutoDock und FRED festgelegt, bei FlexX ergibt sie sich als Teil des Ergebnisses aus dem Algorithmus (vgl. exp. Teil S. 139 ff.).

Tabelle 3.8. | Vergleich FlexX, AutoDock und FRED.

Kennzahl	FlexX		AutoDock		FRED	
	gen.	DSX	gen.	DSX	CG	DSX
1. MW (kleinster rmsd)	2.29		1.47		1.59	
2. MW (Platzierung von 1.)	64	56	106	77	40	35
3. Anzahl (beste Lsg. $\leq 2 \text{ \AA}$ & Top10)	12	20	5	7	19	27
4. Anzahl (erste Lsg. $\leq 2 \text{ \AA}$)	13	27	27	26	19	27

gen.: generische Scoring-Funktion des jeweiligen Programmes; DSX: DrugScoreX; CG: ChemGauss; MW: Mittelwert; Lsg.: Lösung; Top10: Lösungen innerhalb der ersten 10.

1. Wie groß ist im Durchschnitt der rmsd der besten Lösung, also derjenigen mit dem kleinsten rmsd zum kokristallisierten Liganden?
2. An welcher Stelle wird diese beste Lösung platziert?
3. Wie oft kommt die folgende Kombination vor: Die beste Lösung nach Punkt 1. ist unter den ersten 10 platziert und die betreffende Pose ist gleichzeitig als gute Lösung (rmsd kleiner 2 \AA) einzuordnen?
4. In wie vielen der 108 Fällen weist die am besten *bewertete* Lösung einen rmsd von weniger als 2 \AA zum kokristallisierten Liganden auf?

Kennzahlen, die von der Bewertung abhängen, ändern sich durch das Rescoring. Das betrifft außer Punkt 1 alle genannten Parameter. Die Übersicht in Tabelle 3.8 gibt einen Überblick über das Abschneiden der einzelnen Programme, in den folgenden Absätzen sind Besonderheiten aufgeführt, die aus den im Anhang aufgeführten detaillierten Ergebnisübersichten hervorgehen.

3.5.1.1.1. FlexX Im Anhang in Tabelle B.4 auf Seite 239 sind die Ergebnisse des Dockings mit FlexX zusammengefasst. FlexX schafft es nur in 13 der 108 Fälle, dass eine auf Platz 1 gesetzte Lösung einen rmsd unter 2 \AA aufweist. Dabei ist besonders beachtlich, dass nur in zwei der Redockings ein solches Ergebnis erzielt wird. Besonders problematisch scheint die Verbindung UN3 (**47**), für die bestenfalls eine Lösung mit einem rmsd von 3.32 \AA berechnet wird. Es gelingt mit FlexX also mit keinem Protein, die in der Kristallstruktur vorhandene Geometrie zu reproduzieren. Auch die Verbindungen UN1 (**42**) und DRG (**31**) bereiten Probleme, hier gibt es jeweils nur zwei Lösungen mit einem rmsd von weniger als 2 \AA . Dafür gelingt es mit FlexX, in jedem Protein eine kristallstrukturnahe Pose für DFN (**43**) zu erzeugen. Für HBM (**44**) gibt es vier Proteine, in denen Pose mit rmsd unter 2 \AA berechnet wird – dies gelingt mit den anderen Programmen nur im Sonderfall des Redockings.

3.5.1.1.2. AutoDock Insgesamt werden von AutoDock die besseren Komplex-Geometrien gefunden, der durchschnittliche kleinste rmsd liegt mit 1.47 deutlich unter 2 Å und damit unter dem von FlexX erreichten Wert. Bei AutoDock ist es vor allem die Verbindung HBM (**44**), die schlecht gedockt wird (genaue Auflistung im Anhang S. 240). Immerhin wird hier beim Redocking eine der Kristallstruktur nahe Lösung berechnet (rmsd = 1.30 Å, Platz 40).

AutoDock profitiert weniger stark vom Rescoring mit DrugScoreX. Hinsichtlich der durchschnittlichen Platzierung der besten Lösung gibt es eine Verbesserung durch DrugScoreX. So rückt die beste Lösung nun auf Platz 26 vor. Auch die Anzahl der besten Lösungen in den jeweiligen 10 besten Lösungen ist zwar gegenüber dem mit FlexX erreichten Wert kleiner, verbessert sich aber durch die Bewertung mit DrugScoreX von 5 auf 7. Die Kennzahl 4. verschlechtert sich leicht (26 statt 27).

3.5.1.1.3. FRED Die Ergebnisse von FRED sind hinsichtlich der besten errechneten Lösung insgesamt etwas schlechter als die von AutoDock. Allerdings wird von FRED für alle Liganden wenigstens eine Lösung mit einem rmsd unter 2 Å gefunden – auch wenn es bei DFN (**43**) und HBM (**44**) jeweils nur eine einzige ist (interessanterweise im ersten Fall nicht im Redocking erzielt, Übersicht im Anhang S. 241).

3.5.1.1.4. Gegenüberstellung der Programme Im Vergleich zwischen den Programmen findet AutoDock im Durchschnitt die Lösungen mit dem kleinsten rmsd. FRED profitiert besonders vom Rescoring mit DrugScoreX, das zu günstigen Platzierungen guter Posen führt. Alle Programme profitieren in den Kennzahlen 2. und 3. von Seite 125 (vgl. Tab. 3.8). Die Anzahl der am besten bewerteten Lösungen mit einem rmsd von weniger als 2 Å zum kokristallisierten Liganden (Kriterium 4.) sinkt nur bei AutoDock durch das Rescoring von 27 auf 26, für die anderen beiden Programme ist auch in diesem Punkt eine Verbesserung durch die Anwendung von DrugScoreX erreicht.

Für alle Programme gilt, dass die Verbindungen ATU (**28**) und IXM (**36**) deutlich besser abschneiden als die übrigen. Der Grund dafür dürfte darin liegen, dass beide Verbindungen starre planare Moleküle sind, die aufgrund ihres Aufbaus nur eine geringe Anzahl möglichen Lösungen zulassen. Damit entfällt eines der Probleme beim Docking: die Abbildung des konformationellen Raumes. Zudem gehören die Komplexe dieser Liganden zu den besonders gut aufgelösten Strukturen (vgl. Tabelle 3.2 auf S. 92), während der Ligand HBM (**44**), für den insgesamt Schwierigkeiten bei unterschiedlichen Programmen auftreten, mit 3.2 Å am schlechtesten aufgelöst ist. Die Position des Liganden, die der Berechnung des rmsd zugrunde liegt, wird also durch die Kristallstruktur im letztgenannten Fall nur unscharf beschrieben, was eine Abweichung erklären könnte.

Bei HBM (**44**) tritt zudem ein im einleitenden Kapitel 3.1.3 geschildertes Problem auf. Es werden teilweise wichtige Interaktionen in der erhaltenen Pose ausgebildet (Kontakte zur Hinge

Region), der vergleichsweise flexible sulfonamidische Rest ragt jedoch weit aus der Bindetasche hinaus und wird deshalb weniger genau bestimmt. Es ergibt sich ein insgesamt höherer rmsd, so dass auch Lösungen mit einem rmsd von mehr als 2 Å nicht automatisch als schlecht gelten müssen.

3.5.1.2. Dritter Schritt: Clustering der zusammengeführten Ergebnisse

Werden Ergebnisse aller drei Dockingprogramme vereint, liegen für jede Kombination von Protein und Ligand zwischen 382 (DRG (**31**) in kubra) und 1427 (DFN (**43**) in 1Q3W) Posen vor. In Tabelle B.7 im Anhang auf Seite 242 sind die Ergebnisse für alle Protein-Ligand-Paare nach Rescoring mit DrugScoreX und die sich daraus ergebenden Kennzahlen in Entsprechung zum vorangegangenen Abschnitt aufgeführt. Das im Anschluss durchgeführte Clustering führt zu einer deutlichen Verringerung der Lösungsmenge, es liegen nun nur noch zwischen 27 (HT3 (**48**) in 1Q4L) und 291 (DFN (**43**) in 3DU8) Cluster – und damit potentielle Bindungsmodi – vor. Der prozentuale Anteil guter Lösungen sinkt zwar im Durchschnitt: Es sind nur noch 1.5 bis 3.0% der Lösungen gut, während es vor dem Clustering 4.2 bis 14.0% waren – allerdings waren diese Lösungen durch die hohe absolute Anzahl schwieriger ausfindig zu machen.

Dieses Problem wird durch das Clustering relativiert. Zwar ändert sich das Scoring nicht, die Lösungsmenge wird jedoch durch das Aussieben mehrfacher Lösungen auf eine so überschaubare Anzahl reduziert, dass die Identifikation guter Lösungen erleichtert wird. Dies äußert sich in der Kennzahl 4 von Seite 125: Nach dem Clustering sind bei 58 der 108 Dockings die besten Lösungen unter 2 Å *und* werden in den besten 10 platziert – diese Kombination gab es bei einzelnen Docking-Programmen trotz der Kombination mit DrugScoreX maximal in 27 Fällen (Kombination aus FRED und DrugScoreX, vgl. Tab. B.6 im Anhang auf S. 241).

Dass die beste Lösung im Durchschnitt mit 1.39 anstelle von 1.09 Å etwas weiter von der in der Kristallstruktur abweicht, ist insofern logisch, als dass die Auswahl des Repräsentanten aus einem Cluster nach DrugScoreX getroffen wird, wodurch ein etwas weiter von der experimentellen Konformation abweichende Lösung unter Umständen bevorzugt wird. Diese Verschlechterung wird jedoch durch die verbesserte Platzierung aufgewogen.

Werden im folgenden Schritt die geclusterten Lösungen von allen Proteinen für jeden Liganden zusammengeführt und nach DrugScoreX übergreifend sortiert, ergibt sich das in Tabelle 3.9 dargestellte Bild. Es ist wiederum festzuhalten, dass für jeden Liganden eine gute Lösung existiert. Allerdings weist die am höchsten bewertete nur für ATU (**28**) und IXM (**36**) einen rmsd unter 2 Å auf. Dafür findet sich unter den ersten 10 für 6 der Liganden eine gute Pose. Probleme zeichnen sich bei UN2 (**46**) und UN3 (**47**) ab, hier wird eine gute Lösung erst hinter Platz 50 aufgefunden.

Tabelle 3.9. | Ergebnisse von DoReCCoM nach Clustering. Werden die Cluster aller Proteine für jeden Liganden zusammengefasst, ergeben sich aus Tabelle B.7 (S. 242, rechter Teil) die aufgeführten Platzierungen, rmsd-Werte und Anzahlen.

Ligand	rmsd #1	Top10	erster rmsd < 2	bester rmsd	# bester rmsd	Cluster (Lösungen)
ATU	0.70	7	1	0.70	1	492 (6564)
IXM	0.48	8	1	0.40	39	732 (8360)
UN1	5.87	0	15	0.50	50	1411 (9592)
DFN	2.78	2	2	0.82	47	2741 (13914)
HBM	6.53	1	4	1.68	4	1657 (11544)
UN2	5.95	0	57	0.48	64	474 (8862)
UN3	5.19	0	67	0.91	207	1256 (10313)
HT3	6.53	2	9	0.54	39	437 (8655)
DRG	3.20	1	7	0.46	19	599 (5971)

Top10: Anzahl der Lösungen mit rmsd < 2 Å unter den ersten 10; # hier: Platzierung im Rahmen des DrugScoreX Scorings im Vergleich der Lösungen für alle Proteine.

3.5.1.3. Vierter und fünfter Schritt: Komplexminimierung und Rescoring

Die Komplexminimierung wirkt sich unterschiedlich auf die Liganden aus, einen Überblick gibt Tabelle 3.10. Mit der Ausnahme von HT3 (**48**) und IXM (**36**) wird der kleinste rmsd geringfügig verbessert bzw. bleibt im Fall von UN2 (**46**) gleich. Es gibt nach wie vor 3 Verbindungen, für die innerhalb der ersten 10 Lösungen keine mit einem rmsd unter 2 Å gesetzt wird. Allerdings sind es nach der Komplexminimierung andere Verbindungen als zuvor: Während UN1 (**42**) von dem Verfahren profitiert und sich nach der Minimierung 2 gute Lösungen unter den ersten 10 finden (die erste bereits auf Platz 2), fällt die erste Lösung mit einem rmsd unter 2 Å für HBM (**44**) von Platz 4 auf Platz 13 ab. Die Platzierung der entsprechenden Lösung für UN2 (**46**) wird leicht verbessert und ist nun innerhalb der ersten 50 Lösungen aufzufinden.

Tabelle 3.10. | Ergebnisse von DoReCCoM nach Komplexminimierung und Rescoring.

Ligand	rmsd #1	Top10	erster rmsd ≤ 2	bester rmsd	# bester rmsd
ATU	1.71	8	1	0.53	100
IXM	0.69	5	1	0.54	2
UN1	7.00	2	2	0.31	2
DFN	5.54	3	2	0.79	35
HBM	3.67	0	13	1.56	49
UN2	5.82	0	45	0.48	87
UN3	5.44	0	138	0.89	554
HT3	3.14	1	3	0.64	63
DRG	4.14	3	7	0.40	50

Top10: Anzahl der Lösungen mit rmsd < 2 Å; # hier: Platzierung im Rahmen des DrugScoreX Scorings.

Für die vorliegende Arbeit von besonderem Interesse ist die Veränderung für DRG (**31**). Die Geometrie des Komplexes wird marginal verbessert, nach wie vor wird die erste Lösung mit

einem rmsd unter 2 Å auf Platz 7 bewertet. Im Unterschied zum Scoring vor der Komplexminimierung befinden sich jedoch 2 weitere ähnlich gute Lösungen unter den ersten 10 (auf Platz 8 und 10) – DRG (**31**) profitiert also von dem Verfahren.

Insgesamt scheint das Auffinden des kristallstrukturnahen Bindungsmodus für die Verbindung UN3 (**47**) problematisch. Es werden von AutoDock und FRED gute Lösungen gefunden (vgl. Tab. B.5 und B.6 im Anhang auf S. 240 und 241), diese werden jedoch schlecht bewertet. Um den Effekt der Komplexminimierung für alle Verbindungen übergreifend abzuschätzen, wurden in Analogie zu den Kennzahlen zur Bewertung der Dockings mit einzelnen Programmen (vgl. S. 125) die entsprechenden Kennzahlen für die zusammengeführten Ergebnisse der Proteine ermittelt (vgl. Tab. 3.11).

Tabelle 3.11. | Vergleich der Kennzahlen vor bzw. nach der Komplexminimierung.

	vorher	(ohne 47)	⇒	nachher	(ohne 47)
1. Mittelwert (bester rmsd):	0.72	(0.70)	⇒	0.68	(0.66)
2. Mittelwert (Platzierung von 1.):	52	(33)	⇒	105	(49)
3. Anzahl (rmsd < 2 & Top10):	21	(21)	⇒	22	(22)
4. Mittelwert (erster rmsd < 2):	18	(12)	⇒	24	(9)

Der mittlere beste rmsd zur Kristallstruktur und die Kennzahl 3. verbessern sich in jedem der betrachteten Fälle leicht, während sich die Platzierung der Pose mit dem kleinsten rmsd (Kennzahl 2.) im Mittel in jedem Fall verschlechtert.

Das vierte Kriterium wurde gegenüber der Betrachtung der einzelnen Dockingprogramme angepasst. Da hier eine einzige Ergebnisliste für jeden der Liganden ermittelt wird, ist die durchschnittliche Platzierung der ersten Lösung mit einem rmsd unter 2 Å von Interesse. Für die 9 Liganden wird eine solche Lösung im Mittel auf Platz 18 vor und auf Platz 24 nach der Komplexminimierung gefunden.

Berechnet man die Kennzahlen ohne UN3 (**47**), erhält man eine Verbesserung des vierten Kriteriums, die erste Lösung mit einem rmsd unter 2 Å ist im Mittel auf Platz 9. Erklärungsansätze für das schlechte Abschneiden von UN3 (**47**) werden in der abschließenden Diskussion aufgegriffen (vgl. S. 190).

Da die Veränderungen insgesamt klein sind und es auch Beispiele für Verbindungen gibt, die nicht profitieren, sollte im Einzelfall über die Anwendung der Komplexminimierung entschieden werden.

3.5.1.4. Bewertung von DoReCCoM

Durch die Kombination verschiedener Dockingprogramme konnte im Crossdocking für alle eingesetzten Liganden in mehreren Proteinen eine kristallstrukturnahe Lösung gefunden werden,

die Kombination erbringt also einen echten Vorteil gegenüber dem Einsatz eines einzigen Programmes. Das Docking-Problem kann gelöst werden.

Durch das Rescoring konnte eine Verbesserung hinsichtlich der Platzierung guter Lösungen erreicht werden, zudem erbrachte das Clustering eine Verkleinerung der Lösungsmenge auf eine übersichtlichere Größe, so dass das Auffinden des Bindungsmodus im Falle der Inhibitoren der *HsGSK-3 β* möglich ist. Das Scoring-Problem kann zwar noch nicht als gelöst betrachtet werden, DoReCCoM bietet aber eine klare Verbesserung gegenüber dem Standardvorgehen mit einem einzigen der eingesetzten Dockingprogramme und dessen entsprechender generischer Scoringfunktion.

Hinsichtlich der Komplexminimierung muss die allgemeine Übertragbarkeit infrage gestellt werde. Betrachtet man die Gesamtheit der untersuchten Inhibitoren, so kann kein genereller Vorteil durch diesen Schritt erreicht werden. Hier muss daher im Einzelfall untersucht werden, ob der zu behandelnde Verbindungstypus profitiert oder nicht. Da dies für DRG (**31**) der Fall war, wird geschlossen, dass das Verfahren Vorteile für die Verbindungsklasse der Thienopyridine bieten könnte. Es wurde folglich auf den in Abschnitt 3.2.1 vorgestellten Datensatz angewendet.

3.5.2. Anwendung von DoReCCoM auf Thienopyridine in der *PfGSK-3*

Die Anwendung des Arbeitsablaufes auf die Daten des Projektes wurde mit den in Abschnitt 3.2.2.4 aufgeführten Liganden durchgeführt. Als Proteine wurden 10 der im Kapitel 2 entwickelten Homologie-Modelle eingesetzt.¹² In Entsprechung zur Vorgehensweise in der Evaluierung wurden die Ergebnisse für alle Proteine zusammengeführt und von diesen die ersten 50 Lösungen untersucht. Dabei wurde eine Platzierung der entsprechenden Pose unter den ersten 10 als besonders vorteilhaft eingeschätzt. Es werden im Folgenden für die Klassen 1a, 1b und 2 beobachtete Bindungsmodi diskutiert, die neben ihrer Platzierung im Arbeitsablauf dadurch bedeutsam erscheinen, dass sie

- die bekannten Struktur-Wirkungsbeziehungen unterstützen,
- bereits eine Rolle im ersten Docking spielten und
- im Bezug auf die Lage der MIFs interpretierbar sind.

3.5.2.1. Bindungsmodi der Klasse 1a

Für alle drei Verbindungen der Klasse 1a, die untersucht wurden, konnte in Entsprechung zum ersten Docking und zur Kristallstruktur Modus A unter den ersten 10 Lösungen wiedergefunden werden (vgl. Abb. 3.8 auf S. 102). Die Verhältnisse sind in Abbildung 3.23 für die Verbindung

¹²Es wurde jeweils das beste Modell pro Vorlage gewählt. Dabei wurde auf *pfgsk_1uv5* verzichtet, da die Vorlage 1UV5 im Rahmen der Evaluierung DoReCCoM nicht eingesetzt wurde.

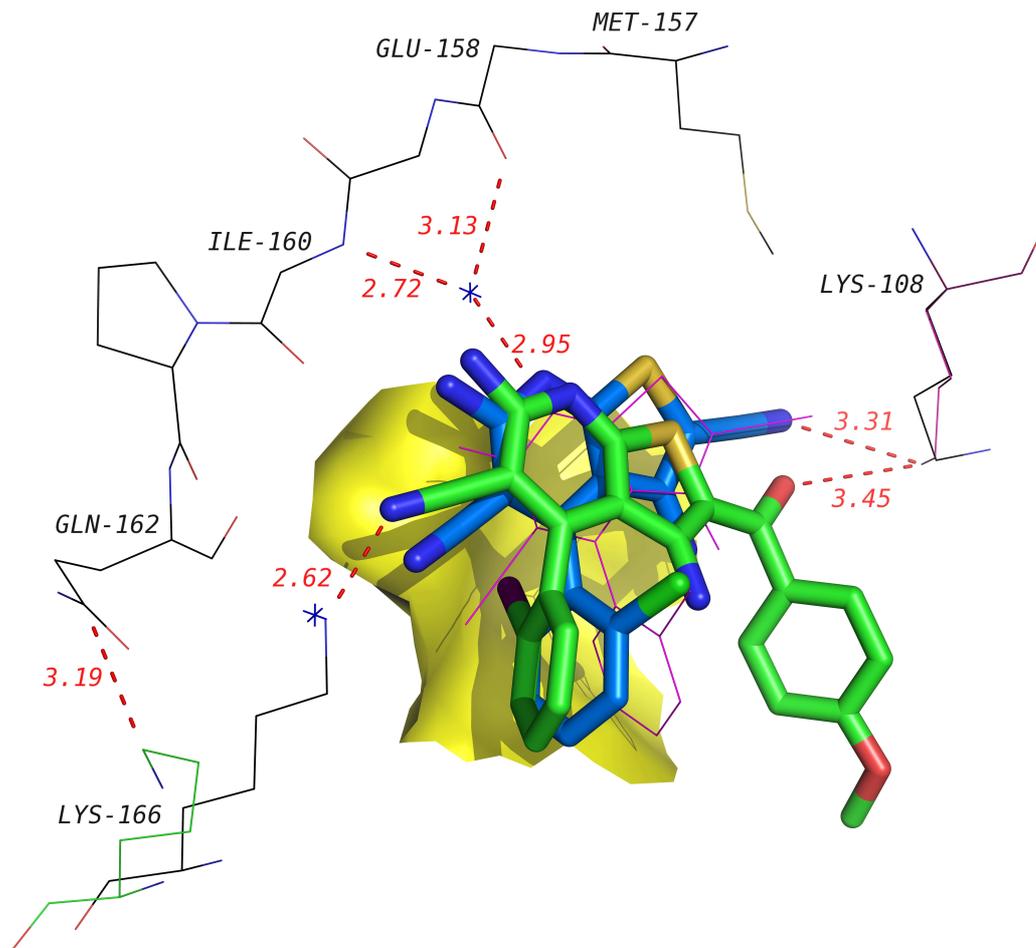


Abbildung 3.23. | Modus A der Verbindungen **31** und **69** als Vertreter der Klasse 1a bzw. 1b in der PfGSK-3. Das nicht minimierte Protein pfgsk_1q5k, in das das zugrunde liegende Docking durchgeführt wurde, ist in grau dargestellt. Am Beispiel des Lys108 ist zu erkennen, wie wenig sich das Protein durch die Minimierung verändert (Position nach Minimierung: magenta). Das Lys166 kann eine aus der Tasche hinaus gewandte Position in Entsprechung zur Situation in der HsGSK-3 β einnehmen. Als Beispiel ist in Grün das Lys166 von pfgsk_kubra im Vordergrund gezeigt.

31 (und als Vertreter der Klasse 1b für **69**) in der PfGSK-3 dargestellt. Dabei wurden zur Orientierung Wassermoleküle aus der humanen Kristallstruktur in die Szene projiziert, so dass die Analogie zu den Verhältnissen im humanen Enzym klar zu erkennen sind.

Auch die im ersten Docking als Modus C eingeführte Geometrie befindet sich für die drei Vertreter der Klasse 1a unter den ersten 10 Lösungen. Sie wird sogar vor Modus A platziert. Für alle drei Verbindungen an erster Stelle wird der auf Seite 103 vorgestellte Modus Xa/b gesetzt – es sprechen allerdings die dort genannten Gründe gegen diese beiden Hypothesen, so dass sie trotz ihrer guten Platzierung nicht weiter in Betracht gezogen werden.

Es wird auch der weiter unten beschriebene Modus H beobachtet (vgl. Abb. 3.25 auf S. 134), jedoch wegen der darin nicht vorhandenen Kontakte des Nitrils an C2 nicht weiter berücksichtigt.

3.5.2.2. Bindungsmodi der Klasse 1b

Es befinden sich auch für alle 9 untersuchten Substanzen dieser Klasse Lösungen im Modus A unter den ersten 50 Posen (vgl. Abb. 3.23) – allerdings trifft dies ebenso auf Modus B zu (vgl. Abb. 3.10 aus dem ersten Docking auf S. 104). Hier sind sogar in 6 Fällen die Lösungen innerhalb der ersten 10 gelistet, im Falle des Modus A sind es nur drei Moleküle, die unter den besten 10 Lösungen aufgefunden werden. Im Durchschnitt unterscheidet sich die Platzierung der ersten Lösung in dem betreffenden Modus für alle 9 Verbindungen kaum (13.1 für Modus B, 13.9 für Modus A).

Von der Platzierung her ist der Modus C sogar noch vor den beiden genannten unter den ersten 10 der Verbindungen der Klasse 1b aufzufinden. Allerdings gibt es hier einen zu engen Kontakt mit der Oberfläche des Taschengrundes mit dem 2-Halogensubstituenten des 4-Arylkernes am Thienopyridingrundkörper (1.7 \AA zum nächsten Wasserstoff des Leu213). Nicht zuletzt spricht gegen diese Hypothese, dass bei diesem Bindungsmodus ausgerechnet der 2-Benzoylrest, der offenbar für Selektivität mit verantwortlich ist, in Richtung einer hochkonservierten Region zeigen würde – zum DFG-Motiv. Hier sind zwischen *Hs*GSK-3 β und *Pf*GSK-3 keine Unterschiede beobachtet worden und es ist davon auszugehen, dass eine Selektivität über Interaktionen mit diesem Bereich des Proteins nicht möglich ist. Der Modus ist beispielhaft in Abbildung 3.24 dargestellt, wird aber abgelehnt.

Auch die Modi Xa/b sind wieder bei den Lösungen vertreten. Während bei der Klasse 1a gar kein Interaktionspartner der Nitrilgruppe auf der Seite des Proteins auszumachen war, gibt es für die Verbindungen der Klasse 1b jedenfalls die Möglichkeit, mit dem Carbonyl einen Kontakt zur Hydroxylgruppe des Tyr159 auszubilden. Allerdings ist auch hier die Geometrie so unvorteilhaft, dass die Argumente gegen diese Hypothese nicht aufgewogen werden (vgl. S. 103).

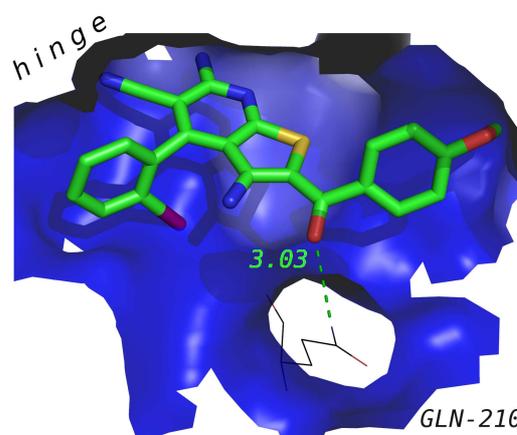


Abbildung 3.24. | Modus C am Beispiel von 69.

3.5.2.3. Bindungsmodi der Klasse 2

Für die Klasse 2 werden wie schon beim ersten Docking viele Posen beobachtet, die verworfen werden, da keine Interaktion mit der Hinge Region vorhanden ist. In den trotzdem oft gut bewerteten Modi C und D zeigt der anellierte hydrophobe Ring direkt auf die Hinge Region. Hier wird also nicht nur keine günstige Interaktion ausgebildet, es ist überhaupt unwahrscheinlich, dass an dieser Stelle ein so lipophiler Teil des Moleküls platziert sein kann.

Dafür wird bei allen eingesetzten Verbindungen der Klasse 2 ein Modus aufgefunden, der in Abbildung 3.25 dargestellt ist und als Modus H bezeichnet wird. Bei den Verbindungen **137**,

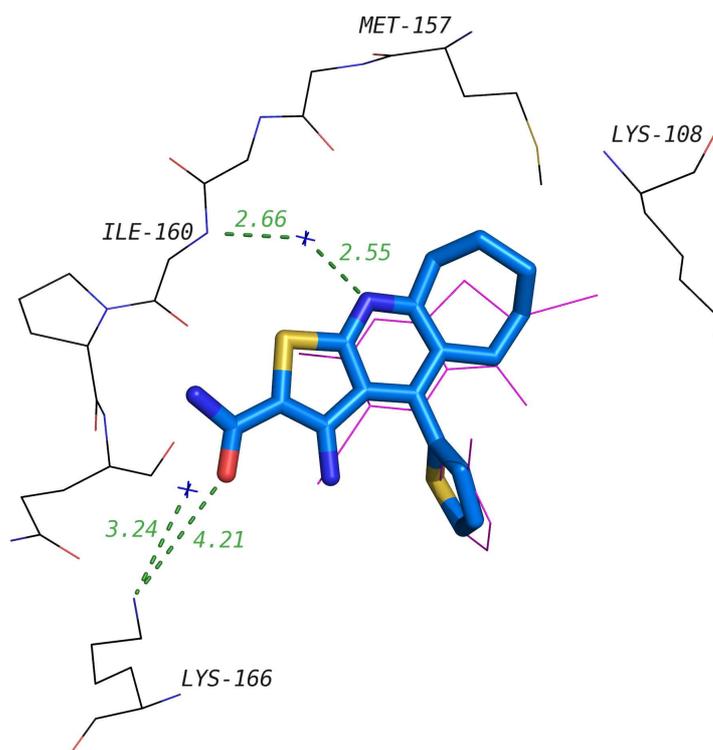


Abbildung 3.25. | Modus H am Beispiel von **140**. Das in Magenta abgebildete Molekül ist **31** aus der Kristallstruktur mit der *HsGSK-3 β* . In diesem Modus liegen die Pyridinringe übereinander, der C2-Substituent (hier das Amid) ragt aus der Tasche hinaus und könnte ein an dieser Position liegendes Wasser verdrängen (Wasser als blaue Sterne wieder projiziert aus der Struktur von **31** in der *HsGSK-3 β*).

138 und **140** wird er mit DoReCCoM auf Platz 1 gesetzt, bei der vierten gedockten Verbindung der Klasse 2 (**121**) erst auf Rang 15. Der Pyridinring des Scaffolds liegt dabei direkt über demjenigen der Verbindung **31** in der Kristallstruktur mit der *HsGSK-3 β* . Die Verbindung ist jedoch gespiegelt, so dass der Substituent an Position 2 aus der Tasche hinaus in die lösungsmittelzugewandte Region zeigt. In beiden Positionen könnte die indirekte Wasserstoffbrückenbindung zur Hinge Region ausgebildet werden, da die Stickstoffe der Pyridinringe in den beiden Modi direkt übereinanderliegen. In der Abbildung wurde das Wasser aus der Kristallstruktur in die Szene projiziert. Der Substituent an C2 könnte eine Interaktion mit Lys166 ausbilden und dabei das in der humanen Struktur beobachtete Wasser verdrängen.

Zusätzlich wird bei **121**, **138** und **140** der Modus A beobachtet – wenn auch im ersten und letzten Fall auf vergleichsweise schlechten Rängen (29, 2, 33).

3.6. Fazit DoReCCoM

Durch Anwendung von DoReCCoM auf einen erweiterten Anteil des Datensatzes konnten die im ersten Ergebnisteil beschriebenen Bindungsmodi durch den zweiten Ergebnisteil einerseits bestätigt werden. So finden sich bei der Anwendung des Arbeitsablaufes auf den erweiterten Datensatz für die Klasse 1 die möglichen Bindungsmodi A bis D wieder, sie werden dadurch

als Optionen gestärkt. Andererseits wird hinsichtlich der Bewertung dieser Modi das Bild verändert. Während im ersten Teil noch eine Abgrenzung der Klassen 1a und 1b durch die Modi A und B als Zwischenergebnis konstatiert wurde, treten die beiden Modi für Klasse 1b nun fast gleichwertig auf, für Klasse 1a wird Modus B gar nicht beobachtet.

Der experimentell festgestellte Bindungsmodus von **31** in der *HsGSK-3 β* kann losgelöst von den Dockingexperimenten als Hinweis auf ein ähnliches Verhalten der Verbindungen in der *PfGSK-3* gewertet werden. Grundsätzlich wäre ein unterschiedliches Verhalten der Klassen 1a und 1b möglich, aber ungewöhnlich. Da im zweiten Docking der Unterschied zwischen den Modi A und B für die Klasse 1b nivelliert wird, erscheint das Auftreten multipler Bindungsmodi unwahrscheinlich. Vor diesem Hintergrund wird als Schlussfolgerung aus den Ergebnissen des zweiten Dockings für die Klasse 1a und 1b gleichermaßen der Bindungsmodus A angenommen.

Für die Klasse 2 unterscheiden sich die Ergebnisse vom ersten und zweiten Docking deutlicher. Die Modi E bis G werden nicht als Ergebnisse von DoReCCoM erhalten. Dafür gibt es die im Zwischenfazit des ersten Teils erwähnte analoge Version des Modus A für die Klasse 2, die als Variante des Modus E aufgefasst werden kann. Bei dieser liegt der Grundkörper aus der Tasche hinausgedreht vor, so dass die überbrückte Wasserstoffbrückenbindung zur Hinge Region ausgehend vom Stickstoff des Thienopyridins möglich wird und die Geometrie insgesamt dem Modus A entspricht. Damit ist Modus A der einzige, der für den gesamten Datensatz mit guten Bewertungen gefunden wird. Allerdings sind die Lösungen der Hypothese H für Verbindungen der Klasse 2 hinsichtlich ihrer Platzierung in DoReCCoM klar überlegen, dieser Bindungsmodus wird daher als Ergebnis des zweiten Dockings für die Klasse 2 angenommen.

Als zusätzliche Bestätigung dieser Differenzierung sollte eine weitere Untersuchung angestellt werden, die die Möglichkeit unterschiedlicher Bindungsmodi der Klassen 1 und 2 von anderer Seite beleuchtet: Die Hypothesen wurden als Ausgangspunkt für Überlagerungen der Moleküle herangezogen, die zur Berechnung von im nächsten Kapitel beschriebenen 3D-QSAR-Modellen dienten. Dabei standen als Konsequenz der in diesem Kapitel erhaltenen Ergebnisse folgende Fragen im Mittelpunkt:

- Liegen für die Verbindungen der Klasse 1 und 2 die unterschiedlichen Bindungsmodi A und H vor?
- Binden alle Verbindungen auf die gleiche Weise, binden sie im Modus A?

3.7. Experimenteller Teil: Computational Details

Für Darstellungen von Bindungsmodi wurde in diesem Kapitel Pymol [232] verwendet. Berechnungen des rmsd wurden mit `fconv` ([233], in der Arbeit verwendete Version: v0.99q) durchgeführt, ebenso das Clustering im Rahmen von DoReCCoM. Für Arbeitsabläufe in Sybyl (Vorbereitung der Liganden) wurde die Version 8.0 verwendet [188].

Proteine, die als Ausgangspunkt für die verschiedenen Vorbereitungsroutinen der Dockingprogramme herangezogen wurden, waren bei der *HsGSK-3 β* die jeweiligen PDB-Dateien (ggf. zur Auswahl der Ketten vgl. Tabelle 2.2 auf S. 35) und bei der *PfGSK-3* die entsprechenden Homologie-Modelle. Die Überlagerung erfolgte in Entsprechung zum Vorgehen bei der Berechnung der MIFs (vgl. S. 75).

3.7.1. Erstes Docking mit FlexE und AutoDock

Vorbereitung der Proteine und Liganden für FlexX/E

Im ersten Abschnitt wurde die Version *FlexX 3.0.2* verwendet. Die einzelnen **Proteine** wurden mit der Protein Preparation Routine im GUI von FlexX 3.0.2 überprüft (Protonierung, N–O Positionen in Amiden, Torsionswinkel von Wasserstoffen). Daraufhin wurden im für das Docking maßgeblichen `ensemble description file` folgende Änderungen gegenüber den Voreinstellungen gewählt:

- Alle Cysteine werden als protoniert behandelt.
- Der Torsionswinkel der Hydroxylgruppe am Tyr159 wurde auf das Carbonyl von Pro161 ausgerichtet. Dazu wurde der Winkel auf 240° angepasst bei `pfgsk_1q3w`, 0° bei `pfgsk_2o5k` und 270° bei den übrigen Proteinen.
- Der Torsionswinkel der Hydroxylgruppe am Thr163 wurde abweichend vom voreingestellten Winkel von 180° verändert auf 90° bei `pfgsk_1j1b`, 210° bei `pfgsk_1q5k`, `pfgsk_1uv5`, `pfgsk_2ow3`, `pfgsk_kubra` und 240° bei `pfgsk_1r0e`.

Um das Zusammenführen zu einem Ensemble zu ermöglichen, mussten spezielle Einstellungen vorgenommen werden, da die verwendeten Strukturen sich insbesondere im Bereich des Glycine-rich Loops im Verlauf des Backbones unterscheiden, was die Konstruktion einer United Protein Structure erschwerte. Abweichend von den Standardeinstellungen wurden die folgenden Parameter gesetzt:

- `INST_EXT_ACT_RADIUS = 0.0`

Diese Angabe bewirkt, dass nur Aminosäuren innerhalb der Bindetasche mit in die Berechnungen einbezogen werden. Eine Ausdehnung auf den hier festzulegenden Radius führt zu noch unterschiedlicherer Beschaffenheit der am Ensemble beteiligten Proteine.

- `ENS_CLASH_HYDROGEN = 1`
Das Herabsetzen des minimalen Abstandes zwischen Wasserstoffen bei der Konstruktion des Ensembles von 1.5 Å auf 1 Å einerseits und ...
- `INST_OVERLAP_VOL = 10`
... andererseits das Heraufsetzen des erlaubten gemeinsamen Volumens der Instances auf 10 Å³ war notwendig, um die Strukturen zusammenführen zu können.
- `INST_CLUSTER_DELTA = 1`
Durch das Heraufsetzen (Vorgabe: 0.5 Å) dieses Wertes werden beim Zusammenfassen der Interaktionspunkte weniger, dafür größere Cluster gebildet. Dadurch werden Extreme in Cluster mit einbezogen, die dann besser kombiniert werden können.

Die Strukturen der **Liganden** wurden zunächst in Sybyl 8.0 aus vordefinierten Fragmenten und deren Erweiterung konstruiert, Gasteiger Ladungen zugewiesen und eine kurze Minimierung vorgenommen (Methode: Powell [234] mit kurzer initialer Simplex-Minimierung (20 Iterationen)[235], Tripos Kraftfeld, Abbruchkriterium: rmsd-Gradient = 0.05 kcal/(mol*Å) mit maximal 1000 Iterationen).

Vorbereitung der Proteine und Liganden für AutoDock

Für die ersten Dockings wurde AutoDock (und entsprechend AutoGrid) in der Version 4.0.0 verwendet. Es wurden die Homologie-Modelle `pfgsk_1j1b`, `pfgsk_1q3d`, `pfgsk_1q41`, `pfgsk_2ow3` und `pfgsk_kubra` verwendet. Die Auswahl repräsentiert möglichst unterschiedliche Größen der Bindetasche (vgl. Dimensionen in Tab. 2.6 auf S. 54). Es wurde die in ADT dokumentierte Routine zur Vorbereitung des Proteins mit Standardeinstellungen genutzt (Prüfung auf fehlerhafte Strukturelemente, Ergänzen polarer Wasserstoffe, Zuweisung von Kollmann United Atom Charges).

Für AutoDock wurde auf die für FlexE vorbereiteten **Liganden** eine in ADT verfügbare Routine angewendet (Entfernen nichtpolarer Wasserstoffe, da die Berechnungen nach dem United Atom Model vorgenommen werden, Zuweisung rotierbarer Bindungen). Zusätzlich musste manuell die Bindung zwischen dem Thienopyridin und dem in Position 4 befindlichen Aromaten als rotierbar definiert werden.

Docking

Beim Docking mit **FlexE** folgt nach der Konstruktion der United Protein Structure als Rezeptor das eigentliche Docking mit FlexX. Dieses wurde nach automatischer Auswahl des Basisfragmentes und Platzierung desselben mittels `triangle matching` durchgeführt. Es wurden jeweils 100 Lösungen begutachtet, indem rmsd-basiert Cluster gebildet wurden. Dazu wurde skriptgesteuert eine rmsd-Berechnung aller 100 Posen zueinander in Chimera [236] angestellt. Die

resultierende Matrix lag dem mit KNIME durchgeführten hierarchischen Clustering mit den folgenden Einstellungen zugrunde:

- Distance function: Euclidean
- Linkage type: Average
- number output clusters: 7

Wenn einzelne Cluster nach visueller Kontrolle inhomogen erschienen, wurden diese erneut nach dem gleichen Schema unterteilt. Resultierende Gruppierungen wurden als hypothetische Bindungsmodi betrachtet und mit AutoDock-Lösungen verglichen.

AutoDock benötigt zunächst ein Gitter, das dann zur Berechnung der Interaktionsenergien des Liganden in der entsprechenden Orientierung verwendet wird. Dieses wurde mit einer Auflösung von 0.375 Å mit allen in zu dockenden Verbindungen aufzufindenden Atomtypen mit AutoGrid (integriert in AutoDock) berechnet. Das Zentrum der zu berechnenden Box wurde als absolute Koordinate anhand des in 1J1B kokristallisierten Liganden ANP festgelegt, es werden 60 × 60 × 60 Gitterpunkte berechnet. Dann wurden 50 Durchläufe des eigentliche Dockings mit dem implementierten Lamarckian Genetic Algorithm (LGA) durchgeführt (default-Einstellungen bis auf `ga_num_evals = 250 000`), die zu den in der Analyse berücksichtigten 50 Posen führten.

Das Docking wurde wiederholt, allerdings unter Einbeziehung der Flexibilität der Seitenketten der Aminosäuren Lys108, Met157, Thr163, Lys166 und Gln210 durch Berücksichtigung derselben im genetischen Algorithmus. Auf diese Weise wurde in einem zweiten Durchlauf die Flexibilität des Rezeptors mit in die Berechnungen einbezogen.

Komplexminimierung

Es wurde die in Sybyl implementierte Routine **Staged Minimization** verwendet. Dabei werden zunächst automatisiert Anteile des Proteins fixiert, die Minimierung erfolgt dann sukzessive in den Schritten: (1) Wasserstoffe des Proteins, (2) Seitenketten, (3) Protein (nur C_α-Atome fixiert), (4) Ligand, (5) das ganze System. Dabei wurde das Tripos Kraftfeld in Kombination mit Gasteiger-Marsili-Ladungen verwendet (einzelne Parameter der Minimierung entsprechen den unter 3.7.1 auf S. 136 ff. aufgeführten).

3.7.2. Rescoring mit DrugScore(X)

Während in den ersten Dockings interaktiv die über das Internet verfügbare Version DrugScore^{ONLINE} und darin DrugScore^{CSD} gewählt wurde, sollte für die Validierung eine lokale Version von DrugScore verwendet werden, da diese die Verarbeitung größerer Datensätze erlaubt. Bei der Installation ergab sich die Möglichkeit, eine aktualisierte Fassung zu testen, deren Potentiale auf einem erweiterten Datensatz aus der PDB basieren. Bei dieser neueren Fassung handelt es sich um DrugScoreX, das seit 03.08.2010 die publizierte ursprüngliche Version von

DrugScore [224] ersetzt hat. Letztere wird nicht mehr unterstützt und es wird vom Arbeitskreis Klebe ausdrücklich darauf hingewiesen, dass Ergebnisse mit DrugScoreX [237] von den mit DrugScore erhaltenen abweichen (<http://pc1664.pharmazie.uni-marburg.de/drugscore/>, Stand 07.07.2011).

Es wurden im Verlauf des Projektes von G. Neudert verschiedene Versionsanpassungen vorgenommen. Alle Ergebnisse des Rescorings im Rahmen des zweiten Dockings ab Abschnitt 3.5 auf S. 125 wurden mit der Version DrugScoreX 0.48 (28.05.2010) berechnet.

3.7.3. Entwicklung eines optimierten Workflows

Die Vorbereitung von Protein und Ligand sowie das Docking erfolgte im Falle von FlexX und AutoDock prinzipiell wie im vorangegangenen Abschnitt erläutert. Während die Thienopyridine immer als neutral behandelt wurden, liegt die Carbonsäure in **42** deprotoniert vor, so dass hier eine Bruttoladung von -1 festgelegt wurde. Abweichend vom ersten Docking wurden zudem die folgenden Änderungen vorgenommen:

FlexX Im Rahmen des zweiten Dockings wurde FlexX als Teil von LeadIT [238] verwendet (veränderte Bibliotheken gegenüber Version 3.0.2). Wegen der im Ergebnisteil beschriebenen Probleme bei der Weiterverarbeitung der Protein-Ligand-Komplexe wurde nicht unter Einbeziehung des Moduls FlexE gearbeitet, sondern die einzelnen Proteine im Sinne eines klassischen Crossdockings eingesetzt. Vorbereitung von Proteinen und Liganden entspricht grundsätzlich Abschnitt 3.7.1. Die Dockings wurden immer doppelt durchgeführt, einmal mit Platzierung des Basisfragmentes mittels `single interaction scan` und einmal `triangle hashing techniques based`.

Für die Vorbereitung der Proteine im Rahmen von DoReCCoM wurde die Protein Preparation Routine in FlexX verwendet (Protonierung, N-O Positionen in Amiden, Torsionswinkel von Wasserstoffen). In der Vorbereitung der humanen Proteinstrukturen wurde zur Definition der Bindetasche der jeweilige kokristallisierte Ligand verwendet. Aminosäuren in einem Radius von 6.5 Å um diesen wurden als zur Bindetasche gehörig definiert (Ausnahme: 3F7Z: 8 Å, da sonst Arg141 nicht erfasst wird).

Bei den Homologie-Modellen wurde zur Definition der Bindetasche jeweils ANP aus 1J1B projiziert (standardmäßige Erweiterung auf einen Radius von 6.5 Å). Davon abweichend wurden folgende Kombinationen gewählt, um alle relevanten Aminosäuren der definierten Bindetasche (35 Aminosäuren, 6.5 Å um ANP in `pfgsk_1j1b`) zu erfassen:

Protein	–	Ligand	–	Radius [Å]
pfgsk_1q41	–	UN1 (42)	–	7.0
pfgsk_1q41	–	STU (25)	–	6.5
pfgsk_2o5k	–	HBM (44)	–	6.0
pfgsk_2ow3	–	ANP (44)	–	8.0
pfgsk_kubra	–	STU (25)	–	7.0

AutoDock Für die Entwicklung des Workflows wurde mit der Version 4.2.3 gearbeitet. Das Vorgehen hinsichtlich der Vorbereitung der Liganden und Proteine entspricht komplett den unter 3.7.1 geschilderten Abläufen. Um mögliche Konformationen noch besser abzudecken, wurden jeweils 200 Posen ermittelt (`ga_run` = 200).

Ein wesentlicher Unterschied ist die Behandlung der gesättigten anellierten Ringe. Während für die ersten Dockings noch eine starre Konformation als Ergebnis der in 3.7.1 beschriebenen Vorbereitung der Liganden auf das Docking mit FlexE eingesetzt wurde, können gesättigte Ringe seit der Version 4.2 flexibel behandelt werden, wovon Gebrauch gemacht wurde.

FRED: Vorbereitung der Strukturen und Docking Die Vorbereitung des Rezeptors wurde automatisiert mit dem dazu in FRED verfügbaren Protokoll durchgeführt. Dabei wurde die Bindetasche in der Evaluierung jeweils durch den kokristallisierten Liganden (bei der Evaluierung mit *HsGSK-3 β*) definiert. Für die Definition bei der Anwendung auf die Homologie-Modelle wurde hinsichtlich der Liganden und Radii in Entsprechung zu FlexX verfahren.

Zum Generieren der Konformerbibliotheken wurde OMEGA2 [239] verwendet. Dabei wurde unter Verwendung des MMFFs-Kraftfeldes die Anzahl an Konformeren auf maximal 400 begrenzt, Konformere wurden ab einem `rmsd` < 0.3 Å als Duplikate entfernt.

Es wurde das defaultmäßig aktivierte Consensus Scoring ausgeschaltet und ausschließlich mit ChemGauss3 bewertet, was in Vorversuchen zu den günstigsten Ergebnissen geführt hatte. Durch Erzeugen von 99 alternativen Lösungen (`num_alt_poses` = 99) wurde eine Anzahl von 100 möglichen Posen abgespeichert.

Komplexminimierung mit SZYBKI Bei der Komplexminimierung mit SZYBKI [227] wurde die Abbruchkriterien abweichend von den Voreinstellungen eingestellt: `max_iter` = 5000 (maximale Anzahl an Iterationen) und `grad_conv` = 0.02 (Konvergenzkriterium).

4. 3D-QSAR-Untersuchungen möglicher Bindungsmodi

4.1. Quantitative Struktur-Wirkungsbeziehungen

Während in den ersten Kapiteln dieser Arbeit der strukturbasierte Ansatz im Vordergrund stand, stellen die im Folgenden dargestellten Ergebnisse der Berechnung von dreidimensionalen quantitativen Struktur-Wirkungsbeziehungen (3D-QSAR, engl.: three-dimensional quantitative structure-activity relationships) eine klassische Form des ligandbasierten Ansatzes in der Arzneistoffentwicklung dar. Die Definition der IUPAC für QSAR lautet [240]:

Quantitative structure-activity relationships are mathematical relationships linking chemical structure and pharmacological activity in a quantitative manner for a series of compounds. (...)

Es wird also ein mathematischer Zusammenhang hergestellt zwischen der chemischen Konstitution C und der pharmakologischen Aktivität Φ (Begrifflichkeit und Gleichung nach CRUM BROWN *et al.* [241], 1868):

$$\Phi = f(C) \quad (4.1)$$

QSAR können dazu beitragen, die Bedeutung von Strukturelementen einer Substanz für deren Aktivität und die Auswirkungen von strukturellen Veränderungen zu verstehen.¹ Sie können außerdem für Vorhersagen der Aktivität bzw. der Eigenschaften unbekannter Strukturen verwendet werden. Durch moderne rechnergestützte Verfahren wird es so möglich, Substanzen virtuell am Computer zu untersuchen, ohne sie tatsächlich dargestellt zu haben. Bei allen Begrenzungen, die diese Technik auch heute nach wie vor mit sich bringt, ist sie mittlerweile ein etabliertes Mittel in der kostspieligen und langwierigen Entwicklung eines Arzneistoffes.

Im Kontext dieser Arbeit war das Ziel, verschiedene mögliche Bindungsmodi zu differenzieren. In diesen liegen die Moleküle in unterschiedlichen Konformationen und einer bestimmten Positionierung zueinander vor. Ein auf der biologisch aktiven Konformation basierendes Modell sollte gegenüber anderen überlegen sein – dieser Unterschied wurde als Argumentationshilfe auf der Suche nach dem Bindungsmodus von Thienopyridinen in der PfGSK-3 genutzt.

¹Im weiter gefassten Begriff der QSPR kann dies auch auf eine Eigenschaft P (engl.: quantitative structure-property relationship) zutreffen.

4.1.1. Grundlagen

Die grundlegende Idee, die biologische Aktivität einer Substanz als Funktion ihrer Struktur aufzufassen, ist bereits sehr alt. Gleichung 4.1 stammt aus dem Jahr 1868, in dem CRUM BROWN und FRAZER ihre Untersuchungen über den Einfluss der Methylierung eines basischen tertiären Stickstoffs auf die „physiologische Aktivität“ der untersuchten Verbindungen veröffentlichten [241].

Die konsequente Erweiterung dieser 1868 noch vagen Vorstellung auf die konkrete Korrelation physikochemischer und struktureller Eigenschaften mit der biologischen Aktivität einer Substanz wurde knapp 100 Jahre später durch HANSCH und FUJITA [242] bzw. FREE und WILSON [243] umgesetzt.

Eine wichtige Neuerung durch HANSCH war die Einführung eines additiven Parameters π_X für die relative Lipophilie eines Substituenten X an einem Grundsystem, der sich nach Gleichung 4.2 ergibt.

$$\log P_{R-X} - \log P_{R-H} = \pi_X \quad (4.2)$$

Es stand damit eine Entsprechung zu der bereits in den 1930ern von HAMMETT [244, 245] postulierten Konstante σ zur Verfügung. Letztere beschreibt für verschiedene Verbindungen die relativen elektronischen Eigenschaften eines Substituenten an einem gemeinsamen Grundgerüst. Dadurch wurde der zweite bahnbrechende Beitrag von HANSCH zur Weiterentwicklung von QSAR-Berechnungen möglich: Verschiedene physikochemische Parameter konnten mithilfe multipler Regression in einer einzigen Gleichung mit einer Konzentration C korreliert werden, die einen bestimmten biologischen Effekt ausübt – auf der resultierenden Gleichung 4.3 beruhende Berechnungen werden bis heute durchgeführt und im Allgemeinen als Hansch-Analyse bezeichnet [246, 247].

$$\log^{1/C} = a\pi + b\sigma + \dots + const \quad (4.3)$$

Unabhängig davon veröffentlichten FREE und WILSON einen anderen Ansatz. Darin gehen sie davon aus, dass alle Substituenten X_i eines Grundkörpers einen Beitrag a_i zur Aktivität leisten [243], so dass unter Annahme einer dem Grundkörper eigenen Aktivität μ sich die Aktivität des substituierten Derivates nach Gleichung 4.4 ergibt.

$$\log^{1/C} = \sum a_i + \mu \quad (4.4)$$

Durch das einfache Zuweisen von 0 und 1 für das (Nicht-) Vorhandensein eines Substituenten entsteht eine Matrix, die wiederum in einer Regressionsanalyse mit $\log^{1/C}$ korreliert werden kann. Auf diese Weise wird eine Free-Wilson Analyse durchgeführt.

Deskriptoren

Die in allen beschriebenen Ansätzen notwendige Voraussetzung für eine Berechnung ist, dass Eigenschaften der Struktur in eine mathematisch fassbare Größe überführt werden, die dann zur rechnerischen Lösung in die jeweilige Gleichung eingesetzt werden kann – diese Größe wird als Deskriptor bezeichnet. Die im vorangegangenen Abschnitt vorgestellten Parameter π_X und σ sind Beispiele für solche Deskriptoren.

Im einfachsten Fall können Deskriptoren aus abzählbaren Eigenschaften wie der Anzahl bestimmter Elemente oder funktioneller Gruppen abgeleitet werden. Des Weiteren können physikochemische Parameter wie π_X bzw. σ herangezogen werden oder es liegt beispielsweise die in der Free-Wilson Analyse erstellte Matrix des Substitutionsmusters zugrunde. Das Spektrum reicht bis hin zu komplexen Deskriptoren, die quantenmechanisch berechnet werden, auch die Überführung experimenteller Daten (z. B. Spektren) in Deskriptoren ist eine Option, einen Überblick über die große Vielfalt gibt TODSCHINI [248]. Auf zahlreiche Deskriptoren (derzeit mehr als 1600, Stand 09.05.2011) ist ein freier Zugriff über die Plattform E-Dragon möglich [249, 250].

Ein wichtiges Merkmal der Deskriptoren ist die Dimensionalität der zugrunde liegenden Struktur des Moleküls. Während eindimensionale Deskriptoren wie das Molekulargewicht sich aus der Summenformel ableiten lassen, haben beispielsweise Fingerprints, die Aussagen über die Konnektivität innerhalb des Moleküls machen, Information über die zweidimensionale Gestalt des Moleküls zur Voraussetzung.

Andere Eigenschaften, wie zum Beispiel die Oberfläche des Moleküls betreffende (Größe, Polarität bzw. Lipophilie), können besser unter Berücksichtigung der dreidimensionalen Gestalt eines Moleküls berechnet werden. Sie beeinflusst diese Eigenschaften, verschiedene Konformationen können sich unter Umständen deutlich unterscheiden, so dass die Berechnungen auf diese Weise an Genauigkeit gewinnen. Eine besondere Gruppe unter diesen 3D-Deskriptoren bilden diejenigen, die von der Ausrichtung im Raum abhängig sind. Hier sind insbesondere alle gitterbasierten Ansätze zu nennen, da hier an einem betrachteten Punkt des Gitters durch Rotation oder Translation des Moleküls der Wert des Deskriptor verändert wird. Für diese Ansätze ist demnach eine räumliche Überlagerung (engl.: Alignment) wichtige Voraussetzung.

Im Kapitel *Homologie-Modelle der PfgSK-3 und deren Untersuchung mit MIFs* wurde bereits GRID als Beispiel für ein gitterbasiertes Verfahren vorgestellt. Die Möglichkeiten der Kombination von GRID/GOLPE im Rahmen von 3D-QSAR-Untersuchungen beschreibt SIPPL [251]. Erstmals beschrieben wurde die Idee eines „Lattice Models“ 1979 von MILNE und CRAMER [252]: Moleküle werden in einer dreidimensionalen Gitterbox fixiert, deren Punkte mit einer Wechselwirkungssonde abgetastet werden.

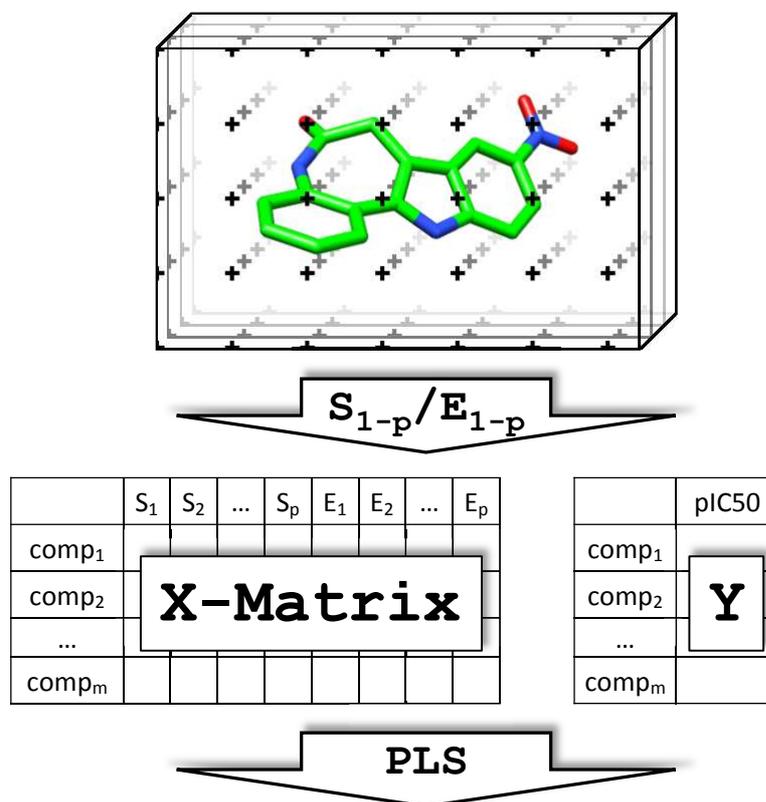
Das Modell wurde in der Folge zum allgemeinen Verfahren der vergleichenden molekularen Feldanalyse CoMFA (engl.: comparative molecular field analysis) weiterentwickelt [175], das heute häufig und in unterschiedlichen Varianten eingesetzt wird. Eine wichtige Variante ist die

von KLEBE *et al.* entwickelte CoMSIA-Methode (engl.: comparative molecular similarity indices analysis [176]).

4.1.2. 3D-QSAR: CoMFA und CoMSIA

4.1.2.1. Prinzip vergleichender Feldanalysen

Das Prinzip einer CoMFA oder CoMSIA ist grafisch in Abbildung 4.1 dargestellt. Der Analyse geht die Auswahl der Strukturen mit ihren biologischen Daten, das Generieren von Konformeren und deren sinnvolle dreidimensionale Überlagerung voraus. Die auf diese Weise präparierten Verbindungen werden in eine dreidimensionale Gitterbox eingebettet. An jedem der Gitterpunkte wird das abstandsabhängige Interaktionspotential zwischen dem Molekül und einer Sonde S berechnet. Dabei stehen bei der CoMFA-Methode die sterische und die elektrostatische Sonde zur Verfügung, bei einer CoMSIA-Berechnung zusätzlich eine hydrophobe Sonde sowie eine Wasserstoffbrückendonator- und -akzeptorsonde. Bei $x * y * z = p$ Gitterpunkten und s Sonden



$$\log 1/C = a_1S_1 + a_2S_2 + \dots + a_pS_p + b_1E_1 + b_2E_2 + \dots + b_pE_p + k$$

Abbildung 4.1. | Prinzip vergleichender Feldanalysen. Die Moleküle eines Datensatzes werden in einem Gitter positioniert. An allen Punkten des Gitters wird ein Interaktionspotential (oder bei CoMSIA ein Ähnlichkeitsindex) zwischen einem Molekül und einer Sonde berechnet. Die Werte werden zu der abgebildeten X-Matrix zusammengefasst, die gemeinsam mit der Y-Matrix, in der die Aktivitäten enthalten sind, in die PLS-Analyse eingeht (Abbildung in Anlehnung an [175, 253]).

werden für jedes Molekül des Datensatzes demnach $p * s = n$ Interaktionspotentiale berechnet. Diese Daten werden für jedes Molekül in einen Zeilenvektor geschrieben und bei einer Anzahl von m Molekülen in einer $m \times n$ –Deskriptormatrix X zusammengefasst. Die biologischen Daten werden in einer entsprechende Y -Matrix abgelegt. Diese Matrices werden nun korreliert, das Ergebnis ist prinzipiell aufgebaut wie Gleichung 4.3. Um die Qualität des Modells einschätzen zu können, muss es in geeigneter Weise validiert werden.

Die Ergebnisse dieser Analyse können einerseits grafisch dargestellt werden in Form von Konturdiagrammen, die anzeigen, wo Veränderungen im Datensatz parallel zur biologischen Aktivität variieren und wo dieser Zusammenhang in umgekehrter Weise besteht. Daraus können Schlüsse über wichtige Interaktionen des Liganden in der Bindetasche abgeleitet werden. Die Konturdiagramme können außerdem zu Ideen für Strukturabwandlungen an den Liganden inspirieren. Andererseits ist es möglich, die Affinität nicht vorliegender Verbindungen vorherzusagen, indem die betreffenden Moleküle virtuell konstruiert und ihre Aktivität (oder allgemeiner die beschriebene Eigenschaft) dem entsprechenden Modell folgend berechnet wird.

Der zentrale Unterschied zwischen CoMFA und CoMSIA ist der Verlauf der Potentiale. Während im ersten Fall ein Lennard-Jones- bzw. Coulomb-Potential zugrunde liegt, wird im zweiten auf eine Gaußfunktion zurückgegriffen, mit der der namensgebende Ähnlichkeitsindex (engl.: similarity) berechnet wird. Der Unterschied im Verlauf macht eine Begrenzung der berechneten Energielevel (engl.: cutoff) bei der CoMSIA überflüssig und führt so zu besser interpretierbaren Konturdiagrammen und dazu, dass der Bereich innerhalb des Moleküls mit in die Berechnung einbezogen werden kann – was ebenfalls zur Interpretierbarkeit beiträgt. Ein weiterer Vorteil von CoMSIA gegenüber CoMFA liegt im Vorhandensein der zusätzlichen Sonden, durch die wesentliche Interaktionsmöglichkeiten von Ligand und Protein beschrieben werden können.

4.1.2.2. Voraussetzungen für 3D-QSAR-Untersuchungen

Bei der Betrachtung von biologischen Aktivitäten als Funktion der Struktur verschiedener Verbindungen gemäß der eingangs zitierten Definition der IUPAC sind Voraussetzungen zu erfüllen, damit die Berechnungen zu einem sinnvollem Ergebnis führen können. Die Anforderungen und Probleme insbesondere bei 3D-QSAR-Analysen kleiner Datensätze werden beispielsweise von SCIOR *et al.* [254] zusammenfassend dargestellt:

Die Moleküle sollen am gleichen Target angreifen und den gleichen Wirkungsmechanismus haben, da ansonsten die gleiche Veränderung in verschiedenen Molekülen des Datensatzes bei unterschiedlichem Target oder Mechanismus zu gegensätzlichen Effekten führen könnte, was die Regressionsanalyse unmöglich macht.

Die Substanzen sollen aus einer kongenerischen Serie stammen, gemeinsame oder wenigstens vergleichbare Strukturelemente aufweisen. Der Datensatz muss dennoch eine

gewisse Variabilität aufweisen, da gerade Unterschiede zu Differenzen zwischen den Deskriptoren führen und eine Regressionsanalyse erst möglich machen. Sind Verbindungen zu ähnlich, ist das Modell nicht auf andere Moleküle übertragbar und somit nicht für valide Vorhersagen geeignet.²

Der Datensatz muss umfangreich genug sein. SCIOR *et al.* geben als Erfahrungswert minimal 20 Verbindungen an, der von CRAMER *et al.* zur Veröffentlichung der CoMFA verwendete Datensatz enthält 21 Steroide [252]. Häufig werden jedoch größere Datensätze verwendet. So stellen GARCÍA *et al.* in einem Übersichtsartikel (3D-)QSAR-Untersuchungen von Inhibitoren der *HsGSK-3 β* gegenüber [255]. Darin werden für 3D-QSAR-Untersuchungen mindestens 36 [256] und bis zu 80 Verbindungen [257] verwendet.

Es soll eine ausreichende Spannweite der Aktivität gegeben sein. Darunter werden in der Regel 3 dezimale Größenordnungen verstanden, über die die Aktivitäten möglichst gleichmäßig verteilt sein sollen.

Es werden einfache Testsysteme gefordert, da ansonsten der Zusammenhang zwischen Struktur und Aktivität verschleiert werden kann. Dazu können Effekte führen wie beispielsweise die Wirkung über unterschiedliche Mechanismen, Instabilität (Ist die Substanz selbst wirksam oder ein Zerfalls bzw. Abbauprodukt?) und jede Form von ADME-Vorgängen bei *in vivo*-Testsystemen. Je komplexer das Testsystem ist, desto größer ist auch die Fehlerhaftigkeit der biologischen Daten. Zudem sollten die Testbedingungen, unter denen die Daten erhoben wurden, möglichst identisch sein.

Der Datensatz sollte den oben genannten Anforderungen genügen und zusätzlich vor der Weiterverarbeitung einer sorgfältigen Überprüfung auf Fehlerhaftigkeit bzw. Inkonsistenz unterzogen werden – ein Schritt, der laut TROPSHA [258] oft vernachlässigt wird. Diese Feststellung bezieht sich allerdings insbesondere auf Untersuchungen mit sehr großen Datensätzen. In diesem Zusammenhang ist auch die Forderung nach aussagekräftigen Deskriptoren zu verstehen (z. B. [259, 260]), die bei einer CoMSIA/CoMFA durch den direkten Bezug zur Struktur gegeben sind.

4.1.2.3. Überlagerung

Von zentraler Bedeutung für eine 3D-QSAR Analyse ist die räumliche Überlagerung der Moleküle. Je mehr die Positionierung der Moleküle zueinander im Gitter und deren jeweilige Konformation den Gegebenheiten in der Bindetasche entspricht, je besser also die biologisch aktiven

²Ein verwandtes Problem ist die sogenannte QSAR blindness. Sie liegt vor, wenn wichtige Interaktionen oder Eigenschaften in der Analyse nicht auffallen, da sie allen Verbindungen innewohnen und deshalb in der mathematischen Analyse nicht erfasst werden, die gerade auf die Unterschiede zwischen den Verbindungen abzielt.

Formen repräsentiert werden, desto besser sollte die Korrelation mit den biologischen Daten gelingen – eine Grundidee, die zur Anwendung der Methode im Rahmen dieser Arbeit führte.

Grundsätzlich kann die Überlagerung mit oder ohne Einbeziehen von Informationen über das Protein erfolgen. Sind solche zugänglich, kann das Alignment beispielsweise durch ein Docking erfolgen. Ist dies nicht der Fall, müssen Überlagerungen ohne Informationen über die Bindeta-sche erzeugt werden.

4.1.3. Multivariate Analyse mittels PLS

Das Problem bei der Korrelation der biologischen Daten Y (abhängige Variable, jeweils ein IC_{50} , eine Spalte, für m Moleküle bzw. Zeilen) mit der in der CoMSIA erhaltenen Matrix X (unabhängige Variable, n Ähnlichkeitsindizes bzw. Spalten für die m Moleküle; vgl. Abb. 4.1 auf S. 144) besteht darin, dass das System in der Regel

- (a) massiv unterbestimmt ist (viele unabhängige Variablen stehen wenigen abhängigen gegenüber, $n \gg m$). Es existieren deshalb bei Anwendung einer multiplen linearen Regression (MLR) mehrere gleichwertige Lösungen der Gleichung.
- (b) Multikollinearität aufweist. Voneinander abhängige Spalten führen in der MLR dazu, dass die Varianzen der Regressionskoeffizienten unannehmbar groß werden.

4.1.3.1. Partial Least Square Regression

Die PLS (engl.: partial least square regression; oder aber: projection to latent structures, nach S. WOLD [261]) wurde bereits 1966 von H. WOLD beschrieben [262]. Die Anwendung auf chemometrische Verfahren wurde insbesondere von dessen Sohn S. WOLD verfolgt.

Das Verfahren zeigt einige Parallelen zur Hauptkomponentenanalyse (vgl. Abschnitt 2.3.6.1 auf S. 39 ff.): Wieder liegt eine Projektion von Daten in ein neues Koordinatensystem zugrunde, wodurch eine Reduktion der großen Anzahl an Variablen auf eine kleine Anzahl aussagekräftiger latenter Variablen gelingt. Im Unterschied zur PCA werden jedoch die Daten einer zweiten Matrix Y in die Berechnung mit einbezogen, so dass die Komponenten des Modells nicht nur die maximale Varianz der Matrix X beschreiben, sondern zugleich die optimale Korrelation mit Y gesucht wird.

Wie bereits bei der PCA ist eine wichtige Größe zur Beschreibung des Modells die Anzahl der Komponenten. Prinzipiell kann durch Erhöhung der Anzahl an Komponenten eine Beschreibung des Datensatzes immer besser gelingen, ohne dass jedoch die Vorhersagekraft des Modells sich verbessert (Phänomen des Overfittings). Die Anzahl sollte deshalb dann nicht mehr erhöht werden, wenn durch weitere Komponenten keine signifikante Verbesserung des Modells erzielt werden kann [261].

Das Ergebnis der PLS liefert eine Gleichung, die im Falle einer CoMSIA den einzelnen Gitterpunkten Regressionskoeffizienten zuordnet:

$$\log^1/c = a_1S_1 + a_2S_2 + \dots + a_pS_p + b_1E_1 + b_2E_2 + \dots + b_pE_p + k \quad (4.5)$$

Darin sind $a_1\dots a_p$ bzw. $b_1\dots b_p$ die Regressionskoeffizienten an den Gitterpunkten 1 bis p , $S_1\dots S_p$ bzw. $E_1\dots E_p$ sind die Wechselwirkungspotentiale (als Beispiel: S für sterisch und E für elektrostatisch) der jeweiligen Sonden an den Gitterpunkten 1 bis p und k ist ein konstanter Summand. Mit dieser Gleichung kann die Aktivität einer bisher nicht im Modell vorhandenen Verbindung vorhergesagt werden. Außerdem können zur visuellen Analyse Gitterpunkte angezeigt werden, deren Koeffizienten einen bestimmten Wert (Konturlevel) über- oder unterschreiten und dadurch besonders stark zur Vorhersage beitragen. Das Ergebnis sind die bei der Analyse des Modells hilfreichen Konturdiagramme.

Statistische Größen

Ein Maß für die Güte der Korrelation ist der quadrierte Korrelationskoeffizient r^2 , der sich für m Moleküle aus der Fehlerquadratsumme oder Summe der Abweichungsquadrate RSS (engl.: residual sum of squares) und der Summe der quadratischen Abweichungen vom Mittelwert SS (engl.: sum of squares) der Zielgröße y nach Gleichung 4.8 ergibt ($y_{pred,k}$: vorhergesagter Wert für das Molekül k , $y_{act,k}$: tatsächlicher Wert für das Molekül k , \bar{y} Mittelwert der Zielgröße im Datensatz). Der Wert liegt zwischen 0 und 1, bei einer optimalen Vorhersage geht RSS gegen 0 und somit r^2 gegen 1.³

$$RSS = \sum_{k=1}^m (y_{pred,k} - y_{act,k})^2 \quad (4.6)$$

$$SS = \sum_{k=1}^m (y_k - \bar{y})^2 \quad (4.7)$$

$$r^2 = 1 - \frac{RSS}{SS} \quad (4.8)$$

$$RMS = \sqrt{\frac{\sum_{k=1}^m (y_{pred,k} - y_{act,k})^2}{m - 1}} \quad (4.9)$$

$$F = \frac{r^2}{1 - r^2} \frac{m - a - 1}{a} \quad (4.10)$$

Ein zweites Gütemaß ist die Standardabweichung RMS (engl.: root mean square error) oder auch s . Sie wird nach Gleichung 4.9 berechnet, sollte möglichst klein sein und ist ebenfalls als Maß dafür anzusehen, wie gut das Modell den Datensatz beschreibt. Der Fischer-Wert F gibt

³Die Gleichungen für die grundlegenden statistischen Kennzahlen wurden aus der Dokumentation der Software entnommen [263].

unter Berücksichtigung der Anzahl a an Komponenten das Verhältnis von erklärter Varianz zu unerklärter an (vgl. Gleichung 4.10), ein hoher Wert zeigt daher ein günstiges Verhältnis an.

4.1.3.2. Validierung von PLS-Modellen

Die Validierung von statistischen Regressionsmodellen ist ein zentraler Punkt, um deren Verlässlichkeit und Vorhersagekraft zu gewährleisten [259]. Dabei kann grundsätzlich zwischen der Güte der Korrelation (intern) und der Anwendbarkeit auf andere Moleküle (extern) unterschieden werden. Die Qualität eines Modells ist also nicht automatisch durch einen hohen Korrelationskoeffizienten gegeben, die Robustheit muss vielmehr geprüft werden, wobei die im Folgende beschriebenen Ansätze allgemein anerkannt sind [264].

Kreuzvalidierung

Bei der Kreuzvalidierung wird dem Datensatz eine bestimmte Anzahl an Elementen entnommen. Für die verbleibenden Verbindungen werden Modelle erstellt, mit denen Zielgrößen der zuvor entfernten Elemente berechnet werden. Dieser Vorgang wird wiederholt, bis alle Elemente des Datensatzes auf der Basis des vorläufigen Modells vorhergesagt wurden.

Im einfachsten Fall der Leave-one-out-Kreuzvalidierung (LOO) wird jeweils eine der m Verbindungen aus dem Datensatz entfernt. Werden mehrere Elemente ausgespart, spricht man von Leave-some-out-Validierung (LSO). Es konnte theoretisch gezeigt werden, dass LSO gegenüber LOO überlegen ist [264, 265]. LOO stellt dennoch insbesondere bei kleineren Datensätzen im zweistelligen Bereich, wie sie bei 3D-QSAR-Analysen typischerweise vorliegen, ein Standardverfahren dar. Es wird deshalb hier näher beschrieben und zur Validierung der erstellten Modelle eingesetzt.

Der Vorgang der LOO kann m mal wiederholt werden, dabei wird analog zur SS die Fehlerquadratsumme der Vorhersage $PRESS$ (engl.: predictive sum of squares) berechnet. Der zu r^2 analoge Korrelationskoeffizient der LOO wird als q^2 (oder auch r_{cv}^2) bezeichnet und ergibt sich nach Gleichung 4.12. Er macht eine Aussage über die interne Vorhersagekraft des Modells – nicht über die externe [266].

$$PRESS = \sum_{k=1}^m (y_{pred,k} - y_{act,k})^2 \quad (4.11)$$

$$q^2 = 1 - \frac{PRESS}{\sum_{k=1}^m (y_{act} - \bar{y})^2} \quad (4.12)$$

$$s_{PRESS} = \sqrt{\frac{PRESS}{m - a - 1}} \quad (4.13)$$

Wird für y_{pred} immer der Mittelwert berechnet, ergibt sich somit ein q^2 von 0. Auch ein q^2 von weniger als 0 ist möglich, wenn $PRESS$ hohe Werte annimmt, die perfekte Vorhersage ergibt

in Analogie zu r^2 einen q^2 von 1. Das der Standardabweichung analoge Fehlermaß in der LOO ist der s_{PRESS} , der sich nach Gleichung 4.13 ergibt. Hier geht die Anzahl a der Komponenten ein, ein gutes Modell liefert einen kleinen s_{PRESS} .

Die LOO spielt nicht nur eine Rolle bei der Validierung der Modelle. Sie wird in der Regel genutzt, um optimale Parameter für das endgültige Modell zu ermitteln. Dies sind im Rahmen einer 3D-QSAR-Analyse die Anzahl an Komponenten und unter Umständen die Auswahl der Felder. Dazu kann das Verfahren SAMPLS (SAMple-distance Partial Least Squares, [267]) eingesetzt werden, mit dem die genannten statistischen Parameter für verschiedene Anzahlen an Komponenten sehr schnell berechnet wird. Nimmt der q^2 bei Steigerung der Anzahl an Komponenten nicht mehr signifikant zu (Schwellenwert konventionsgemäß 5 %) ist die optimale Anzahl erreicht.

Für CoMFA-Studien geben KIM *et al.* [268] einen q^2 von 0.3 als akzeptabel an. Auch wenn der q^2 keine Aussagekraft hinsichtlich der externen Prädiktivität hat (vgl. [259]) und von daher nicht überschätzt werden darf, ist dieser Wert allerdings niedrig angesetzt und Literaturwerte liegen in der Regel deutlich darüber. So werden in dem bereits zitierten Artikel von GARCÍA *et al.* [255] QSAR-Studien vorgestellt, in denen für den q^2 Werte zwischen 0.480 [269] und 0.805 [270] erreicht werden.

Externe Validierung durch einen Testsatz

Zur Ermittlung der externen Vorhersagekraft eines Modells müssen Aktivitäten von Verbindungen vorhergesagt werden, die nicht an der Entstehung des Modells beteiligt waren. Dazu wird vor der Analyse ein Teil des Datensatzes von den Berechnungen ausgenommen, der Testsatz (auch Validation Set). Mit den verbleibenden Verbindungen wird ein Modell erstellt, sie stellen in diesem Szenario den sogenannten Trainingssatz dar. Für den Testsatz werden dann mit dem zuvor mit dem Trainingssatz erstellten Modell Aktivitäten berechnet. Die Güte dieser Vorhersage gibt verlässliche Auskunft über die Möglichkeiten der externen Vorhersagekraft des Modells [264].

Randomisierung der abhängigen Variablen

Ein weiterer Ansatz zur Validierung des QSAR-Modells ist die Randomisierung der abhängigen Variablen, also der biologischen Daten (auch Y-Scrambling). Prinzipiell sollte bei zufälliger Anordnung der Zielvariablen kein sinnvolles Modell mehr zustande kommen. Zeigen sich bei der statistischen Beurteilung dennoch annehmbare Gütemaße, ist dies zunächst nur als das Anpassen von Zahlen an ein Rauschen zu interpretieren und das Modell muss als nicht valide angesehen werden.

Enthält der Datensatz viele Redundanzen, können auch bei randomisierten Zielvariablen sehr gute Modelle berechnet werden, da möglicherweise zufällig ähnliche abhängige Variablen ver-

tauscht worden sind. Kommt noch die Unschärfe biologischer Daten dazu, ist die Interpretation eines Y-Scramblings nicht trivial. Auch die Ergebnisse einer Validierung durch LOO können in diesem Fall zu optimistisch ausfallen. CLARK *et al.* [271, 272] entwickelten aus diesem Gedanken heraus das Progressive Scrambling.

Dabei wird der Datensatz ausgehend von vielen kleinen Gruppen sukzessive in immer weniger größere eingeteilt, innerhalb derer die Zielgröße zufällig angeordnet wird. Es entsteht eine kontrolliert ansteigende Durchmischung des Datensatzes. Für jede Anordnung wird mehrfach eine LOO durchgeführt und die statistischen Kennzahlen berechnet. Diese werden in Zusammenhang gesetzt mit der Korrelation des vermischten Datensatzes mit dem Ausgangsdatsatz $r_{yy'}^2$. Werden zu Beginn in vielen kleinen Gruppen die Werte vertauscht, ist die Veränderung von Y gegenüber der Ausgangssituation klein (werden Redundanzen vertauscht: nicht vorhanden), so dass bei hohem $r_{yy'}^2$ der Wert für q^2 des manipulierten Datensatzes hoch sein darf. Mit abnehmendem Zusammenhang zu den realen Daten sollten die Werte für q^2 drastisch sinken, ansonsten kann das Modell nicht akzeptiert werden.

4.2. Material und Methoden

4.2.1. Rationale

Im Rahmen dieser Arbeit sollten QSAR-Modelle genutzt werden, um die unterschiedlichen Hypothesen zu Bindungsmodi aus Kapitel 3 zu bewerten. Eine Überlagerung, die den räumlichen und konformativen Gegebenheiten im Komplex von Ligand und Protein nahekommt, sollte ein überzeugendes Modell ergeben. Dieses muss sowohl durch statistische Kennzahlen die Korrelation von biologischer Aktivität und strukturellen Merkmalen stützen, als auch bei visueller Analyse im Hinblick auf den Rezeptor sinnvolle Hinweise auf Struktur-Wirkungsbeziehungen geben. Ist dies der Fall, spricht das für die Sinnhaftigkeit der Überlagerung und damit für die zugrunde liegende Hypothese zum Bindungsmodus.

4.2.2. Der Datensatz

Die Strukturen mit ihren biologischen Daten sind in Tabelle B.3 auf Seite 234 ff. aufgeführt. Es liegen für 46 aktive Verbindungen diskrete IC_{50} -Werte an der PfGSK-3 vor. Davon zählen 31 zur Klasse 1 und 13 zur Klasse 2 (zur Einteilung der Klassen und bekannten Struktur-Wirkungsbeziehungen siehe S. 83 ff.). Die beiden Verbindungen **115** und **117** können keiner der beiden Klassen eindeutig zugeordnet werden, da sie weder über eine 6-Amino-5-cyano-Substitution noch über einen anellierten gesättigten Ring an den entsprechenden Positionen verfügen. Da die Berechnungen auf die Bewertung von Bindungsmodi der Klassen abzielen, werden die nicht zuzuordnenden Verbindungen nicht mit in die Analyse einbezogen. Es bleiben demnach 44 Verbindungen beider Klassen (Vorbereitung der Moleküle vgl. exp. Teil S. 178).

4.2.2.1. Aufteilung in Trainings- und Testsatz

Die Aktivitäten der 44 Verbindungen wurden in den negativen dekadischen Logarithmus (pIC_{50}) überführt. Alle in die Analyse mit einfließenden Verbindungen sind nach fallender Aktivität geordnet in Tabelle 4.1 aufgeführt.

Tabelle 4.1. | Datensatz für QSAR-Analyse.

Verbindung	pIC_{50}	Verbindung	pIC_{50}	Verbindung	pIC_{50}	Verbindung	pIC_{50}
108	6.886	66	6.041	72	5.620	107	5.000
31	6.886	139	6.000	78	5.569	84	5.000
63	6.745	76	5.959	143	5.523	101	4.939
56	6.398	140	5.959	127	5.523	142	4.824
77	6.319	64	5.796	132	5.523	128	4.796
83	6.301	40	5.796	71	5.495	125	4.523
62	6.292	69	5.770	75	5.420	121	4.367
74	6.215	80	5.770	141	5.420	104	4.301
138	6.097	137	5.745	135	5.398	103	4.222
58	6.097	105	5.658	85	5.222	102	4.168
79	6.041	81	5.658	87	5.071	68	4.155

pIC_{50} : negativer dekadischer Logarithmus des IC_{50} an der PfGSK-3.

Auf die Bedeutung eines Testsatzes zur Validierung von QSAR-Analysen wurde in Abschnitt 4.1.3.2 eingegangen. Aus den dort genannten Gründen wurden die Daten in Trainings- und Testsatz unterteilt. Dabei sollte aus jeder Klasse ein besonders aktiver und ein besonders wenig aktiver Vertreter ausgewählt werden [254]. Für Klasse 1 wurden zufällig fünf weitere Verbindungen ausgewählt, für Klasse 2 eine weitere. Dadurch wurde mit 10 Substanzen im Test gegenüber 33 im Trainingssatz ein vergleichsweise großer Testsatz bereitgestellt. Er wird als *Testsatz A* bezeichnet und enthält die folgenden 10 Verbindungen:

31, 40, 81, 83, 102, 104, 121, 135, 139.

Da die Prädiktivität des QSAR-Modells nicht im Fokus dieser Untersuchungen steht, wurde zudem ein verkleinerter *Testsatz B* erstellt, damit nicht wichtige Information für das Erstellen des Modells verloren geht, eine Überprüfung des Modells durch den unabhängigen Testsatz aber dennoch möglich bleibt. Prinzipiell wurde ebenso verfahren wie im ersten Fall, allerdings wurden nur fünf Verbindungen zugeordnet. Diese repräsentieren sowohl die strukturelle Vielfalt des Datensatzes (beide Klassen sind vertreten) als auch die Variabilität der Aktivitäten; neben diesen beiden Punkten fordern GOLBRAIKH *et al.* [266] minimal fünf Verbindungen für ein valides Testset, so dass diese gemeinhin akzeptierten Anforderungen mit den folgenden Verbindungen erfüllt sind:

56, 84, 102, 132, 140

4.2.2.2. Eigenschaften des Datensatzes

Da die biologischen Daten der Substanzen alle im gleichen Assay an isoliertem Enzym und im gleichen Labor vermessen wurden, kann davon ausgegangen werden, dass experimentelles Vorgehen und Protokolle der Datenverarbeitung gleich waren und der Datensatz als homogen anzusehen ist. Es ist allerdings anzumerken, dass zu unterschiedlichen Zeitpunkten und bis zu unterschiedlichen Endkonzentrationen (10 bzw. 100 μM) gemessen wurde (Details zu den biologischen Daten in Abschnitt 3.2.1.2 auf S. 85).

Eine wichtig Forderung an die einer QSAR-Berechnung zugrunde liegenden biologischen Daten ist deren ausreichende Spannweite. Die hier vorliegenden Aktivitäten reichen von 0.13 bis 70 μM , das entspricht pIC_{50} -Werten von 6.8861 bis 4.1549. Damit wird ein Bereich von 2.7 logarithmischen Einheiten abgedeckt. Als Daumenregel wird ein abzudeckender Bereich von mindestens drei dekadischen Größenordnungen gefordert [254]. Der Datensatz erfüllt diese Anforderung also fast und kann in dieser Hinsicht als akzeptabel gelten.

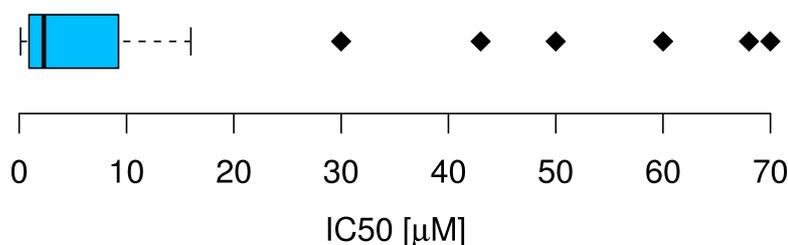


Abbildung 4.2. | Verteilung der biologischen Daten als Boxplot. Der Median liegt bei sehr kleinen Werten (dicker senkrechter Strich), der farbig markierte Bereich der beiden angrenzenden Quantile erstreckt sich bis 10 μM , mehr als die Hälfte des Datensatzes ist also auf diesen Bereich beschränkt.

Eine weitere interessante Betrachtung ist die Verteilung der biologischen Daten. Diese sollten nach Möglichkeit über den Bereich gleichmäßig sein. Diese Forderung wird leider nicht sehr gut vom vorliegenden Satz erfüllt (vgl. Abb. 4.2). Während 38 der 44 Verbindungen zwei Größenordnungen im Bereich zwischen 0.13 und 16 μM abdecken, stehen nur sechs Verbindungen als Repräsentanten des darüber liegenden Bereiches bis 70 μM zur Verfügung.

4.2.3. Methodik: Überlagerungen und 3D-CoMSIA

In dieser Arbeit sollte durch den Vergleich verschiedener Modelle auf die Bedeutung der zugrunde liegenden Überlagerung geschlossen werden. Dazu wurde zunächst die zentrale Frage bezüglich des Bindungsmodus, die sich aus den Ergebnissen des Kapitels 3 herauskristallisierte (vgl. S. 134), in unterschiedliche Überlagerungen überführt. Es wurden flexible Überlagerungen mit den Programmen FlexS [273–275] und ROCS [276, 277] auf Repräsentanten der unterschiedlichen hypothetischen Bindungsmodi des letzten Kapitels durchgeführt. Alignments, die

auf Dockinglösungen beruhen, hatten in Vorversuchen nicht zu interpretierbaren Modellen geführt.

FlexS überlagert paarweise Moleküle auf ein vorgegebenes Zielmolekül. Flexibilität wird dabei durch den fragmentbasierten Ansatz gewährleistet, die Vorgehensweise entspricht derjenigen von FlexX (vgl. S. 95): Nach Auswahl des Basisfragmentes werden die zu verknüpfenden Fragmente nach möglicher Interaktion mit einem Rezeptor platziert – der Rezeptor wird bei FlexS durch Projektion von Interaktionsmöglichkeiten des Referenzliganden simuliert. Weiterhin wird durch Translation das überlappende Volumen mit dem Referenzliganden maximiert.

ROCS (Rapid Overlay of Chemical Structures) verfolgt einen gestaltbasierten Ansatz (engl.: shape-based). Für zuvor generierte Konformerbibliotheken wird nach dem Element mit maximaler Überlappung mit dem Zielmolekül gesucht.

Basierend auf den unterschiedlichen Alignments wurden die fünf möglichen CoMSIA-Felder und mit SAMPLS die Kenndaten für alle 31 Feldkombinationen berechnet (vgl. Tab. B.8-B.10). Aus den Ergebnissen konnte das optimale Alignment und günstige Felderkombinationen abgeleitet werden – sofern Modelle vorhanden waren, die als signifikant anzusehen waren. Für diese wurden dann in der Folge endgültige Modelle berechnet, die eine visuelle Analyse erlauben.

Die endgültigen Modelle wurden jeweils für drei Szenarien erstellt. Neben der oben beschriebenen Aufteilung in Trainingssatz A bzw. B und den jeweils korrespondierenden Testsatz wurde zusätzlich eine Analyse des gesamten Datensatzes – also ohne Aufteilung in Trainings- und Testsatz – durchgeführt, da die externe Prädiktivität nicht im Fokus der Untersuchungen stand.

4.3. Ergebnisse

4.3.1. Überlagerungen und deren Analyse mittels SAMPLS

Für die im Folgenden beschriebenen Überlagerungen wurden alle 44 in Tabelle 4.1 aufgeführten Verbindungen herangezogen. Die orientierenden SAMPLS-Berechnungen wurden mit dem Trainingssatz A durchgeführt, für die im Detail beschriebenen Modelle (Abschnitt 4.3.2 auf S. 157 ff.) wurden die in Abschnitt 4.2.2.1 genannten unterschiedlichen Aufteilungen des Datensatzes betrachtet.

4.3.1.1. Überlagerung nach Modus A und Modus H

Die unterschiedlichen strukturellen Merkmale der Klassen 1 und 2 sowie die im letzten Kapitel geschilderten Dockingstudien legen unterschiedliche Bindungsmodi für die Verbindungen der Klassen nahe. Es werden daher Überlagerungen betrachtet, in denen die Klasse 1 gemäß

Modus A und die Klasse 2 gemäß Modus H bindet. Die überlagerten Verbindungen sind in Abbildung 4.3 dargestellt. Für die Verbindungen der Klasse 2 wurde beim Erstellen der Überlagerung (vgl. exp. Teil S. 178 ff.) eine Verbindung mit Arylsubstitution an Position 2 als Repräsentant des Modus H und damit als Zielmolekül der Überlagerung gewählt, damit auch dieser Substituent in einer der Bindetasche angepassten Geometrie vorliegt. Dadurch liegen die Pyridinringe der Verbindungen und folglich die 4-Arylsubstitution des Thienopyridins nicht mehr ganz übereinander. Die Stickstoffe des Pyridins hingegen, die die für beide Bindungsmodi charakteristische überbrückte Wasserstoffbrückenbindung zur Hinge Region ausbilden, werden in der Überlagerung gut aufeinander abgebildet.

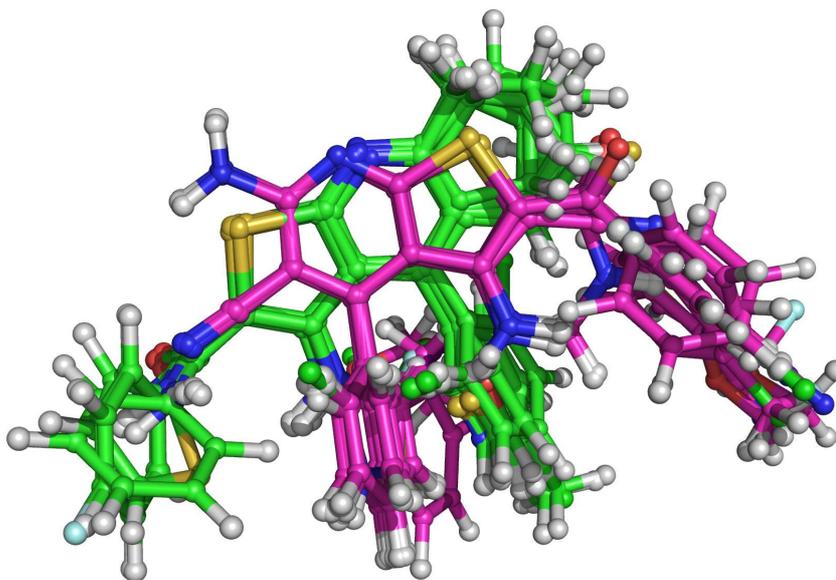


Abbildung 4.3. | Überlagerung der beiden Klassen in den Modi A (magenta) und H (grün).

Nach Berechnung der CoMSIA-Felder ergab die orientierende LOO aller Felderkombinationen jedoch für keine Felderkombination einen q^2 von mehr als 0.25. Selbst mit einem Schwellenwert von 0.3 für den q^2 eines aussagekräftigen QSAR-Modells, der im Vergleich zu in jüngerer Zeit veröffentlichten Untersuchungen als niedrig anzusehen ist (vgl. S. 149), kann die Qualität des Modells nicht als ausreichend angesehen werden, um weitere Schlüsse zu begründen. q^2 , s_{PRESS} und Anzahl der Komponenten sind für alle Felderkombinationen im Anhang auf Seite 243 dargestellt.

4.3.1.2. Überlagerung basierend auf Modus A

Das für die weiteren Berechnungen zugrunde liegende Alignment, das den Bindungsmodus A repräsentiert, ist in Abbildung 4.4 dargestellt. Die leichte Verschiebung der Atome des Thienopyridins gegeneinander ist in der Berechnung des Alignments durch FlexS begründet (vgl. exp. Teil S. 178 ff.). Der zugrunde liegende Algorithmus zielt auf potentielle Interaktionspunkte im Rezeptor ab und ist nicht auf den Vergleich einzelner Atome ausgerichtet. Dieses Ver-

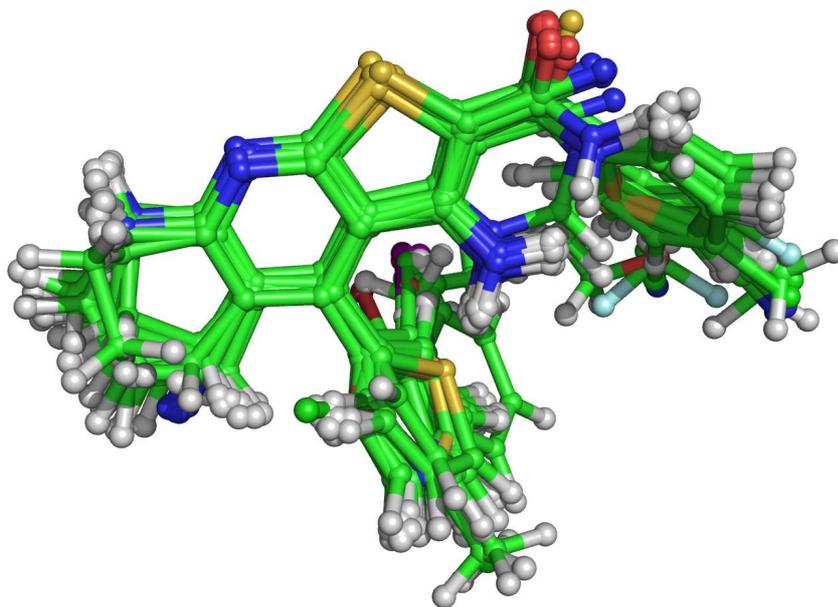


Abbildung 4.4. | Überlagerung der Verbindungen gemäß Modus A.

fahren hatte in vorausgehenden Untersuchungen gegenüber einer einfachen Überlagerung der Heteroaromaten im direkten Vergleich zu den besseren Ergebnissen geführt.

Die Ergebnisse der SAMPLS-Berechnungen für dieses Alignment sind vollständig im Anhang auf Seite 244 aufgeführt. Sie sind insgesamt deutlich besser als die im vorangegangenen Abschnitt vorgestellten und lassen eine nähere Untersuchung des Modells gerechtfertigt erscheinen. Die drei besten Felderkombinationen sind:

1. Wasserstoffbrückenakzeptor- und elektrostatisches Feld (im Folgenden AcEl) mit einem q^2 von 0.402
2. sterisches, Wasserstoffbrückenakzeptor- und -donorfeld (im Folgenden AcDoSt) mit einem q^2 von 0.396
3. elektrostatisches, Wasserstoffbrückenakzeptor- und -donorfeld (im Folgenden AcDoEl) mit einem q^2 von 0.392

4.3.1.3. Überlagerung basierend auf Modus B

Das auf Modus B gestützte Alignment ist in Abbildung 4.6 dargestellt. Bei vergleichender Betrachtung mit der zuletzt gezeigten Überlagerung fällt ins Auge, dass der Unterschied losgelöst von den absoluten Koordinaten im Wesentlichen im Torsionswinkel zwischen dem Heteroaromaten und demjenigen des Benzoylrestes liegt. Abbildung 4.5 verdeutlicht schematisch, dass in beiden Hypothesen der Benzoylrest auf der gleichen Seite wie der 4-Arylrest liegt. Der grün gekennzeichnete Torsionswinkel ist in allen Molekülen beider Überlagerungen in

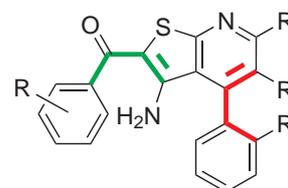


Abbildung 4.5. | Wichtige Torsionswinkel zur Unterscheidung der Konformationen, die Modus A und B einnehmen können.

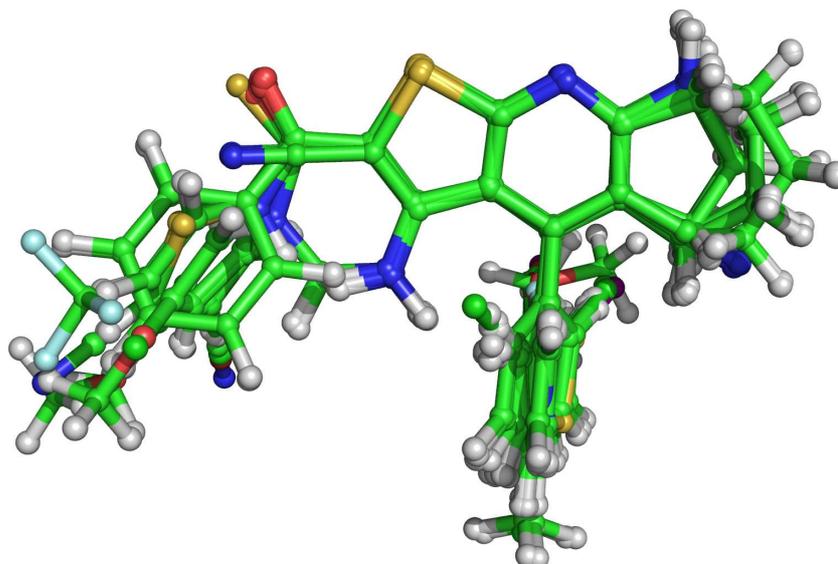


Abbildung 4.6. | Überlagerung der Verbindungen gemäß Modus B.

der Regel gleich 0° , höchstens 30° . Aufeinander abgebildet unterscheiden sich die Konformationen der Modi A und B folglich hauptsächlich durch die Orientierung des Substituenten an Position 2 des 4-Arylrestes (roter Torsionswinkel in Abbildung 4.5). Dadurch ist eine große Ähnlichkeit der Modelle an sich zu erwarten, die Interpretation der Konturdiagramme durch Projektion in den Rezeptor hingegen sollte sich unterscheiden.

Die Ergebnisse der SAMPLS-Berechnungen (Tab. B.10 auf S. 244) heben sich klar von den in Abschnitt 4.3.1.1 dargestellten ab, die auf der Überlagerung der Verbindungen nach den Modi A und H beruhen. Sie sind vergleichbar mit den im vorangegangenen Abschnitt aufgeführten. Die bereits dort als optimal ermittelte Kombinationen von Feldern schneiden wieder am besten ab, die Werte für den q^2 sind in absteigender Reihenfolge:

1. AcDoSt: $q^2 = 0.392$
2. AcEl: $q^2 = 0.319$
3. AcElSt: $q^2 = 0.288$

Während der erste Wert der entsprechenden Kombination aus Abschnitt 4.3.1.2 sehr nahe kommt, ergibt sich für AcEl und AcElSt ein niedrigerer q^2 .

4.3.2. Analyse resultierender 3D-QSAR-Modelle

4.3.2.1. Modelle basierend auf Modus A

In den SAMPLS-Berechnungen zur Ermittlung der optimalen Anzahl an Komponenten und der besten Feldkombinationen kristallisierten sich die oben genannten Kombinationen AcDoSt, AcEl und AcElSt als besonders günstig heraus. Es wurden deshalb vollständige Modelle für

Tabelle 4.2. | Statistische Parameter der auf Modus A beruhenden Modelle. Im unteren Teil der Tabelle sind die prozentualen Anteile der am jeweiligen Modell beteiligten, namensgebenden physikochemischen Eigenschaften aufgeführt.

	AcDoSt	AcEl	AcElSt	AcDoSt	AcEl	AcElSt	AcDoSt	AcEl	AcElSt
	<i>Trainingsatz A</i>			<i>Trainingsatz B</i>			<i>kompletter Datensatz</i>		
q^2	0.396	0.401	0.392	0.406	0.417	0.378	0.491	0.464	0.437
s_{PRESS}	0.583	0.522	0.535	0.631	0.551	0.569	0.581	0.535	0.548
r^2	0.698	0.635	0.653	0.786	0.629	0.607	0.785	0.632	0.612
s	0.412	0.407	0.404	0.378	0.440	0.452	0.378	0.443	0.455
F	7.237	26.955	18.850	10.312	30.552	27.832	12.042	35.181	32.321
# comp.	8	2	3	10	2	2	10	2	2
Ac.	35.8	41.3	36.1	32.9	41.7	41.4	32.5	44.5	43.8
Do.	41.2	–	–	37.9	–	–	40.1	–	–
St.	23.0	–	9.3	29.2	–	6.9	27.5	–	7.7
El.	–	58.7	54.6	–	58.3	51.6	–	55.5	48.5

q^2 : Korrelationskoeffizient im Rahmen der Kreuzvalidierung; s_{PRESS} : root mean **P**redictive **E**rroR **S**um of **S**quares; r^2 : Korrelationskoeffizient; s : Standardabweichung des Modells (*RMS*); F : Prüfgröße F des Fischertests; # comp.: Anzahl a der Komponenten (Gleichungen und Definitionen der statischen Kennzahlen vgl. S. 148 ff.); Ähnlichkeitsindizes der Modelle werden abgekürzt mit Ac.: Wasserstoffbrückenakzeptor, Do.: Wasserstoffbrückendonor, St.: sterische Beschaffenheit und El.: elektrostatische Beschaffenheit.

die entsprechenden Kombinationen von Feldeigenschaften berechnet, die Ergebnisse sind in Tabelle 4.2 in Form der statistische Kennzahlen dargestellt. Die Residualplots der Modelle sind überblicksartig in Abbildung 4.7 abgebildet.

Die erste Kombination von CoMSIA-Feldern schneidet hinsichtlich der Korrelationskoeffizienten im Vergleich am besten ab, die Kombination ist die einzige, bei der ein r^2 von 0.7 und mehr erreicht wird. Die Abgrenzung von zweiter und dritter Kombination ist schwieriger. Insgesamt zeigt die dritte mit kleinen Unterschieden gegenüber der zweiten die schlechtesten statistischen Kennzahlen. Der optische Eindruck bestätigt diesen Sachverhalt (vgl. Abb. 4.7). Wenn die erste Variante hier als beste genannt wird, muss gleichzeitig jedoch berücksichtigt werden, dass die hohe Anzahl an Komponenten zu einem Overfitting führen kann. Die im Vergleich zu den Modellen AcEl und AcElSt niedrige Prüfgröße F zeigt diese Problematik an. Nimmt man die Prüfgröße F als maßgeblichen Parameter, ist die Kombination AcEl sogar in jeder der drei verschiedenen Szenarien hinsichtlich der Auswahl des Testsatzes das beste Modell. Insgesamt müssen die statistischen Kennzahlen im Vergleich zu Literaturwerten allerdings als schlecht angesehen werden (vgl. S. 149).

Eine auf der Hand liegende Tendenz ist, dass die Modelle umso besser werden, je mehr Verbindungen an deren Entstehung beteiligt sind: 34 im Trainingsatz A, 39 im Trainingsatz B und 44 bei Verwendung des gesamten Datensatzes. Die Feldbeiträge des sterischen Feldes sind in allen Kombinationen die kleinsten, während das elektrostatische Feld einen großen Beitrag leistet. Ist es im Modell enthalten, trägt es etwa 50% und mehr zum Modell bei. Dass das

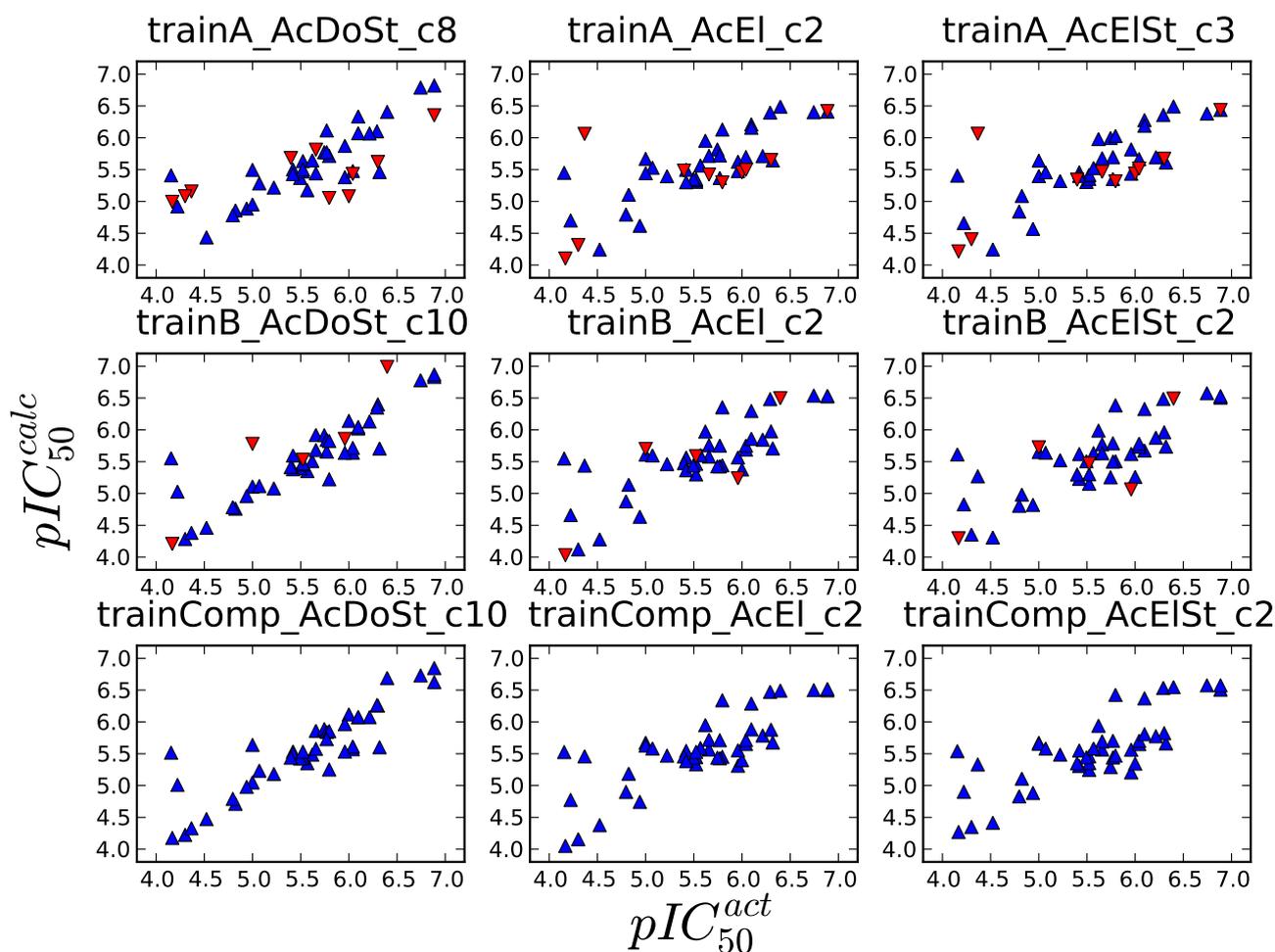


Abbildung 4.7. | Residualplots der verschiedenen Kombinationen von Feldern und Datensätzen (blaue Dreiecke: Verbindungen der Trainingsätze, rote Dreiecke: Testsätze; Bezeichnung der Modelle gemäß Tabelle 4.2 mit $c < n > =$ Anzahl der Komponenten). Aufgetragen sind die jeweils vom Modell berechneten pIC_{50} -Werte (pIC_{50}^{calc}) gegen die gemessenen (pIC_{50}^{act}).

Akzeptorfeld in allen Modellen beteiligt ist und der Beitrag in keinem Fall nennenswert unter $1/3$ ist (kleinstenfalls 32.9% bei AcDoSt mit Trainingsatz B), spricht für dessen Bedeutung für die Berechnung quantitativer Struktur-Wirkungsbeziehungen.

Für die detaillierte Analyse werden die beiden Kombinationen AcDoSt und AcEI ausgewählt. Die entsprechenden Kombinationen werden auf Seite 168 ff. vergleichend für die Überlagerung gemäß 4.3.1.3 basierend auf Modus B ausgewertet. Es wurde jeweils das mit dem Trainingsatz B erhaltene Modell untersucht. Dadurch bleibt der Vorteil der Möglichkeit einer *externen* Validierung durch den komplementären Testsatz B erhalten. Gleichzeitig wird möglichst viel der im Datensatz enthaltenen Information zur Erstellung des Modells eingesetzt.

4.3.2.2. Das Modell AcDoSt

In Tabelle B.11 im Anhang auf Seite 245 sind vorhergesagte und tatsächliche pIC_{50} -Werte sowie die entsprechenden Abweichungen (Residuen) für den Trainings- und den korrespondierenden

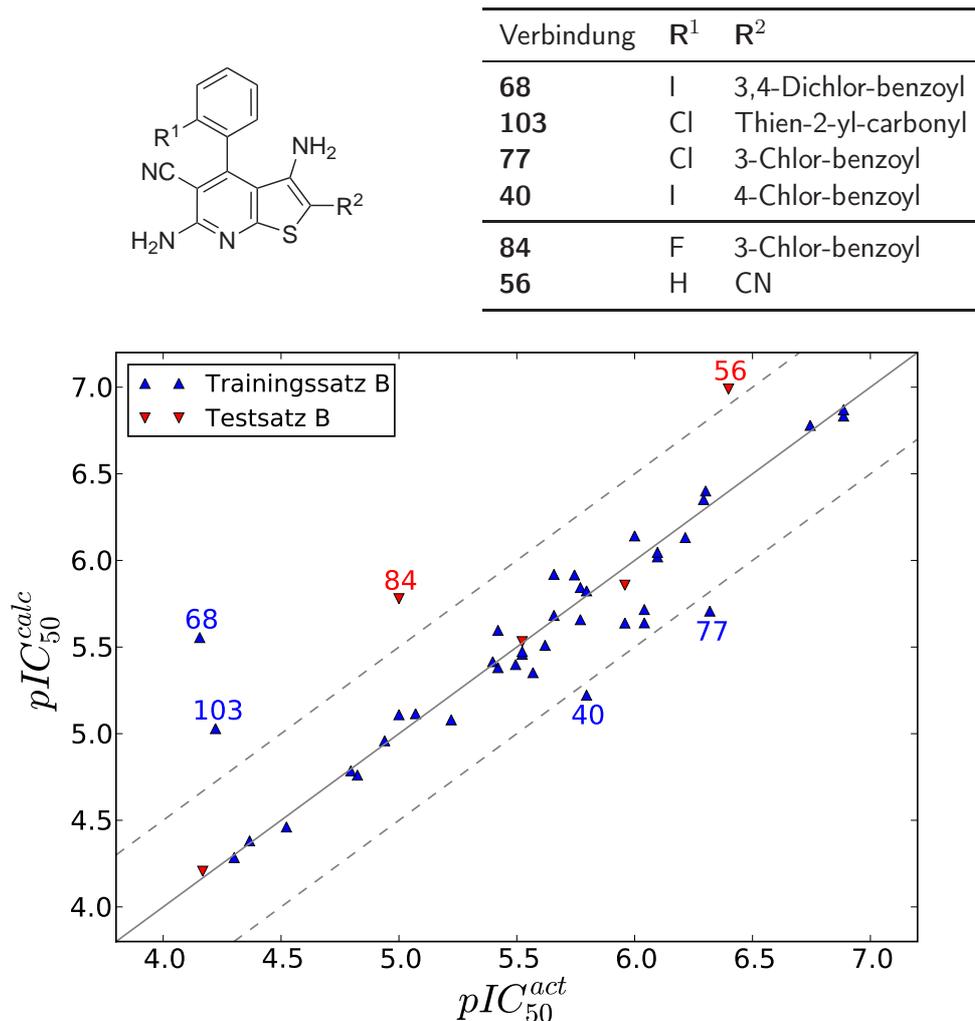


Abbildung 4.8. | Grafische Auftragung der Tabelle B.11 (Modell AcDoSt, Überlagerung nach Modus A). Bei den bezeichneten Verbindungen weichen berechneter und gemessener pIC_{50} -Wert mehr als 0.5 Einheiten voneinander ab.

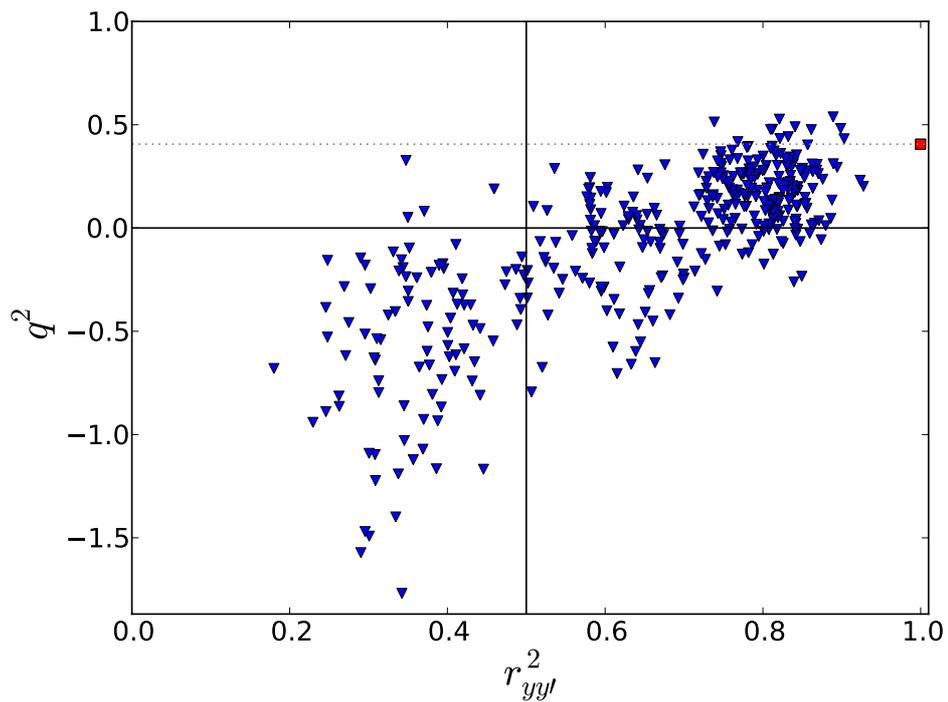
Testssatz aufgeführt. Den daraus folgenden Residualplot zeigt Abbildung 4.8. Darin bezeichnet sind Verbindungen, deren berechnete Aktivität mehr als 0.5 pIC_{50} -Einheiten von der gemessenen abweicht. Als erstes fällt auf, dass es sich bei den Verbindungen mit großen Residuen ausschließlich um solche der Klasse 1 handelt, alle Verbindungen der Klasse 2 werden vom Modell gut beschrieben. Bei diesen 13 Verbindungen beträgt die größte Abweichung von berechnetem und tatsächlichem pIC_{50} im Falle von **137** gerade einmal -0.172.

Besonders stark weicht die Verbindung **68** ab. Sie ist die einzige, deren berechnete Aktivität auch in allen anderen berechneten Modellvarianten um mehr als eine logarithmische Einheit vom bestimmten Wert abweicht. Ihre strukturelle Beschaffenheit (Iodsubstitution am 4-Arylrest, Dichlorsubstitution am 2-Benzoylrest) wird vom Trainingssatz gut beschrieben, weshalb die starke Abweichung überrascht. Bedenkt man die Abweichungen bei Wiederholmessungen bei den nahe verwandten Substanzen **40** und **77**, ist eine Abweichung durch experimentell bedingte Schwankungen im Testsystem denkbar (vgl. Tabelle B.3 im Anhang S. 234). Unter dieser

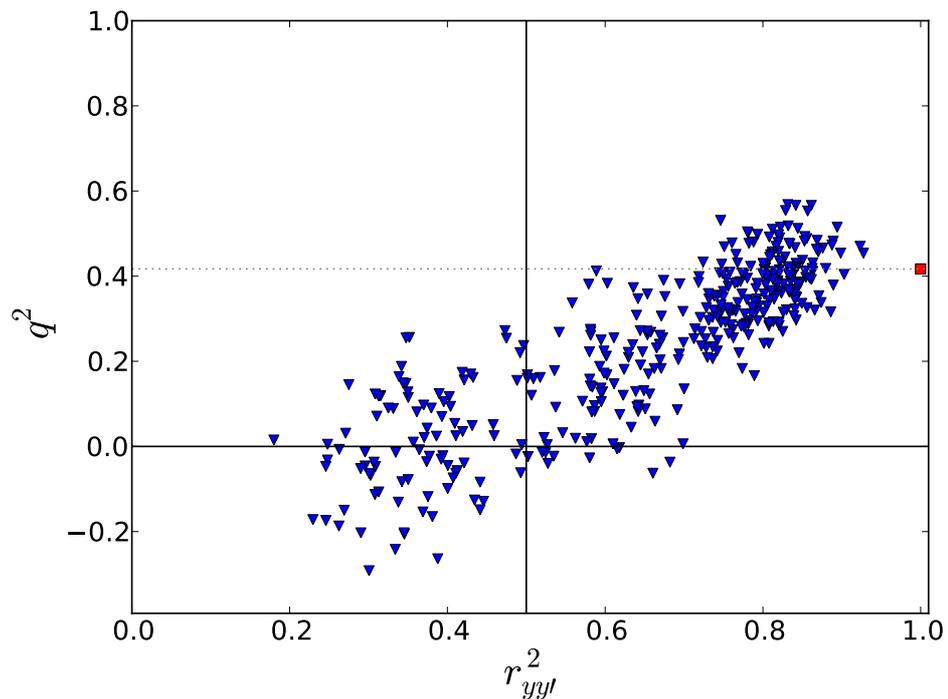
Annahme ist auch die schlechte Vorhersage von **40** und **77** verständlich. Mit einer Abweichung von -0.808 Einheiten wird die Aktivität der Verbindung **103** vom Modell schlecht beschrieben. Die tatsächliche Aktivität entspricht nur knapp $\frac{1}{6}$ der berechneten. Dieses Ergebnis ist insofern interessant, als dass die nahe verwandte Verbindung **102** im Testsatz wiederum sehr genau vorhergesagt wird. Im Rahmen der Darstellung der Ergebnisse der biologischen Testung von **101-104** wurde die Widersprüchlichkeit der gemessenen biologischen Daten diskutiert (vgl. S. 118). Dabei wurde darauf hingewiesen, dass die Aktivität des 2-chlorsubstituierten Derivates **103** den bekannten Struktur-Wirkungsbeziehungen nach im Vergleich zum 2-iodsubstituierten **101** die höhere Aktivität zeigen sollte. Der vom Modell errechnete Wert von 5.029 für **103**, entsprechend einem IC_{50} von knapp $10 \mu M$, wäre hinsichtlich der übrigen Schlüsse zu Struktur-Aktivitätsbeziehungen kongruent. Dies lässt einen Messfehler bei der biologischen Prüfung als mögliche Erklärung für die schlechte Vorhersage im Rahmen des Modells zu.

Die ungenaue Vorhersage der Verbindungen **56** und **84** im Testsatz lässt sich auf die nicht vorhandene bzw. kleine Substitution an Position 2 des 4-Arylrestes zurückführen. Während den im übrigen Datensatz beobachteten Struktur-Wirkungsbeziehungen zufolge die Aktivität mit abnehmender Größe des Substituenten steigen sollte, bilden die genannten Verbindungen hinsichtlich ihrer Aktivität Ausnahmen von dieser Regel. Sie stellen gewissermaßen das Bindeglied zu den im Modell wegen der fehlenden diskreten pIC_{50} -Werte nicht eingesetzten Verbindungen **97** und **106** dar, deren 4-Phenylsubstituent nicht weiter substituiert ist. Die Verbindungen sind an *Pf*GSK-3 und *Hs*GSK-3 β inaktiv (vgl. Abschnitt 3.2.1.1 auf S. 84).

Neben der LOO wird als zweiter Ansatz zur Validierung die in Abschnitt 4.1.3.2 erläuterte Randomisierung der abhängigen Variablen mittels Progressive Scrambling verwendet (siehe Abb. 4.9(a) auf S. 162). Aufgetragen werden die nach Randomisierung und durchgeführter Kreuzvalidierung erhaltenen Werte für q^2 gegen $r_{yy'}^2$. Mit abnehmender Korrelation von durchmischem Datensatz mit den tatsächlichen Ergebnissen nimmt der q^2 ab, allerdings gibt es 12 Durchläufe der LOO, in denen der Wert nach Randomisierung besser ist als im untersuchten Modell. Im Diagramm erscheinen diese Fälle als Punkte über der gestrichelten grauen Linie, die mit 0.406 den q^2 des ursprünglichen Modells anzeigt. Wie im allgemeinen Abschnitt dargelegt, ist dies jedoch einsichtig, bedenkt man die Möglichkeit, dass bei hohem $r_{yy'}^2$ nur benachbarte Werte getauscht wurden. Insbesondere bei einer so hohen Anzahl an nahe beieinanderliegenden Aktivitäten, wie sie im zugrunde liegenden Datensatz vorliegen (vgl. Abb. 4.2 auf S. 153), ist ein hoher q^2 also bei gleichzeitig hohem $r_{yy'}^2$ nicht verwunderlich. Zudem ist die Möglichkeit gegeben, dass zufällig Datenpunkte getauscht werden, die mit Messfehlern behaftet waren. Das Ergebnis des Y-Scramblings spricht also nicht gegen das Modell.



(a)



(b)

Abbildung 4.9. | Randomisierung der biologischen Daten mittels Progressive Scrambling für Modelle, die basierend auf Modus A berechnet wurden. Der rot markierte Wert stellt den q^2 des ursprünglichen Modells mit nicht durchmischten abhängigen Variablen dar. (a) zeigt die Veränderungen in Abhängigkeit von $r^2_{yy'}$ für die Felderkombination AcDoSt, (b) diejenigen für die Felderkombination AcEl.

Tabelle 4.3. | Farbkodierung für CoMSIA-Konturdiagramme.

CoMSIA-Feld	+		-	
	Farbe	Bedeutung	Farbe	Bedeutung
Akzeptor	violett	Donor im Rezeptor \uparrow	rot	Donor im Rezeptor \downarrow
Donor	cyan	Akzeptor im Rezeptor \uparrow	blau	Akzeptor im Rezeptor \downarrow
elektrostatisch	blau	pos. des Liganden \uparrow	gelb	neg. des Liganden \uparrow
sterisch	grün	Substitution in diesem Bereich \uparrow	gelb	Substitution in diesem Bereich \downarrow

\uparrow / \downarrow : günstig / ungünstig für die Aktivität des Liganden; pos.: positive Partialladung; neg.: negative Partialladung.

Analyse der CoMSIA-Konturdiagramme

Die visuelle Analyse der Felder erfolgt unter zwei Gesichtspunkten. Einerseits soll die Stimmigkeit des Modells in sich beurteilt werden, andererseits die Interpretierbarkeit im Bezug auf die Bindetasche. Dieser Aspekt ist besonders bei den Wasserstoffbrückendonor- und -akzeptorfeldern von Bedeutung, die jeweils das begünstigende Vorhandensein eines Wasserstoffbrückenantipoden *im Rezeptor* markieren. Auch bei der Analyse des sterischen Feldes ist der Bezug auf den Rezeptor besonders naheliegend. Die Farbkodierung für die Darstellung aller Konturdiagramme in dieser Arbeit ist in Tabelle 4.3 aufgeführt.

Die Wasserstoffbrückenakzeptor- und -donorfelder haben in diesem Modell einen gemeinsamen Anteil von 70,8 % an der Vorhersage der Aktivität (vgl. Tab. 4.2). Die beiden Felder sind in Abbildung 4.10 gemeinsam mit den beiden Inhibitoren **62** und **103** dargestellt. Die wichtige Rolle des Substituenten an der Position 2 des Thienopyridins wird deutlich. Für einen Liganden

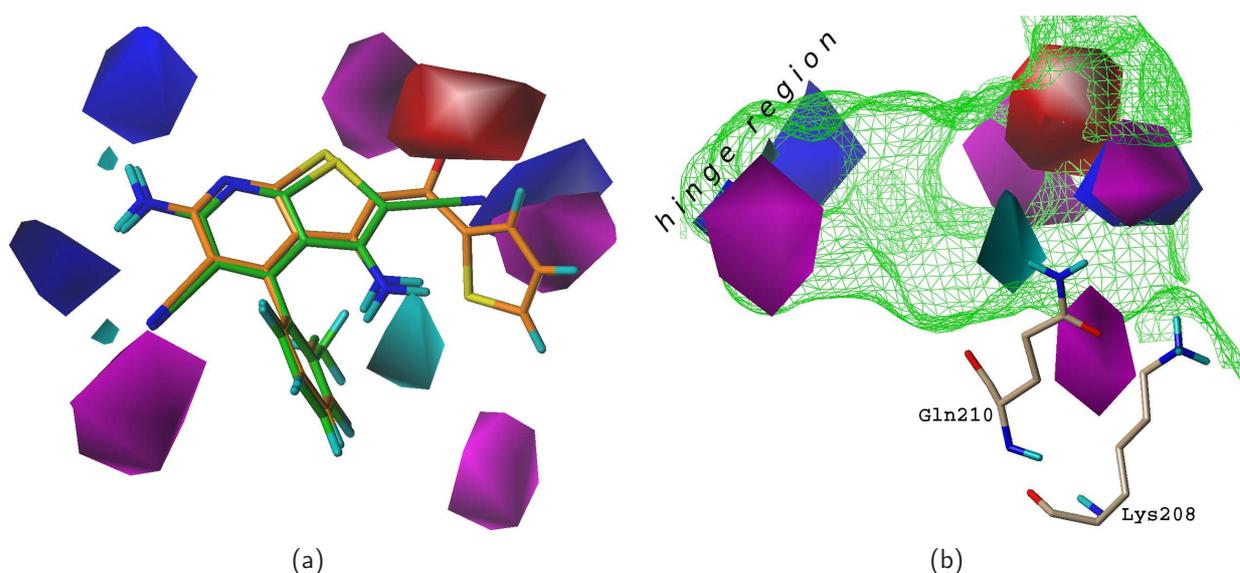


Abbildung 4.10. | **62** (grün) als aktive Verbindung und **103** (orange) als weniger aktive mit Wasserstoffbrückenakzeptor- und -donorfeldern gemäß AcDoSt / Modus A (a); Oberfläche der Bindetasche (b).

mit hoher Affinität wie **62** (grün in Abb. 4.10(a)) ist das blaue Donorfeld in Verlängerung der Dreifachbindung des 2-Nitrils folgerichtig. Es besagt, dass hier ein Akzeptor im Rezeptor negative Auswirkung auf die Aktivität hätte. Das direkt angrenzende violette Feld hingegen ist ein positives Akzeptorfeld – hier wird das begünstigende Vorhandensein eines Donors berechnet. Diese Position könnte von dem Lys108 eingenommen werden, das allerdings in den Modellen der PfGSK-3 im roten Bereich liegt, der an das Carbonyl von **103** angrenzt. Dass diese Interaktion des Carbonyls in einen negativen Feldbeitrag ragt, erscheint zunächst widersprüchlich, lässt sich aber mit der Zusammensetzung des Datensatzes erklären (vgl. S. 83 ff., unterschiedliche Repräsentation Klasse 1a/b): Aktivere Verbindungen an der PfGSK-3 sind insgesamt diejenigen der Klasse 1a. Nur **64** hat mit 1.6 μM einen IC_{50} über 1, für alle anderen Verbindungen dieser Gruppe liegt er darunter. Für die benzoylsubstituierte Verbindungen der Klasse 1b hingegen gibt es zahlreiche Beispiele mit IC_{50} -Werten über 1. Dieser Sachverhalt ist weniger darauf zurückzuführen, dass Verbindungen der Klasse 1b nicht hohe Aktivität aufweisen *können*, es wurden vielmehr wegen der Selektivität von Verbindungen dieses Typs deutlich mehr Derivate synthetisiert. Da dabei mehr variiert wurde, als es bei den offenbar nicht selektiven Verbindungen der Klasse 1a der Fall war, enthält der Datensatz eine höhere Anzahl von weniger aktiven Verbindungen der Klasse 1b. In der Folge werden die charakteristische Eigenschaften der Klasse 1b vom statistischen Modell als negativ für die Aktivität eingestuft. Dies könnte das roten Donorfeld in Abbildung 4.10(a) in Verlängerung des Carbonyls von Verbindungen der Klasse 1b erklären.

Im Inneren des Proteins liegt neben der oben diskutierten Region um das Lys108 ein positiver Bereich des Akzeptorfeldes zwischen Lys208 und Gln210 (Abb. 4.10(b)). Hier ist ein Donor dann vorhanden, wenn die Seitenkette des Lysins sich in diesen Bereich bewegt oder das Amid des Gln210 die umgekehrte Anordnung von Sauerstoff und Stickstoff einnimmt. Damit wäre auch das entsprechende Carbonyl im Bereich eines positiven Donorfeldes (cyan). Da in solchen Bereichen ein Akzeptor im Rezeptor günstige Auswirkungen auf die Aktivität der Verbindungen hat, ist dieses Feld mit den Gegebenheiten im Rezeptor konform, wenn die umgekehrte Anordnung von Sauerstoff und Stickstoff im Amid von Gln210 angenommen wird.

Die blauen Donorfelder in der Nähe der 6-Aminosubstitution des Grundkörpers sind schwer interpretierbar. Sie zeigen an, dass dem Modell zufolge hier ein Akzeptor im Rezeptor ungünstige Auswirkungen auf die Aktivität des Liganden hat – allerdings wird im zugrunde liegenden Modus A über das entsprechende Amin eine Wasserstoffbrücke zur Hinge Region ausgebildet (vgl. Abb. 3.8 auf S. 102).

Die sterischen Felder sind in Abbildung 4.11 dargestellt. Hier ergibt sich ein auf den ersten Blick überraschender Befund: Gerade am Grund der Tasche in der Region um Thr163, Leu213 und Cys224, in die die 2-Substitution des 4-Arylsubstituenten zeigt, befindet sich ein negatives Feld, das die besagte Aushöhlung ausfüllt. Ein Erklärungsansatz ist der in Verbindung mit der schlechten Vorhersage von **56** und **84** im Testsatz genannte: Im Datensatz gibt es mehrere

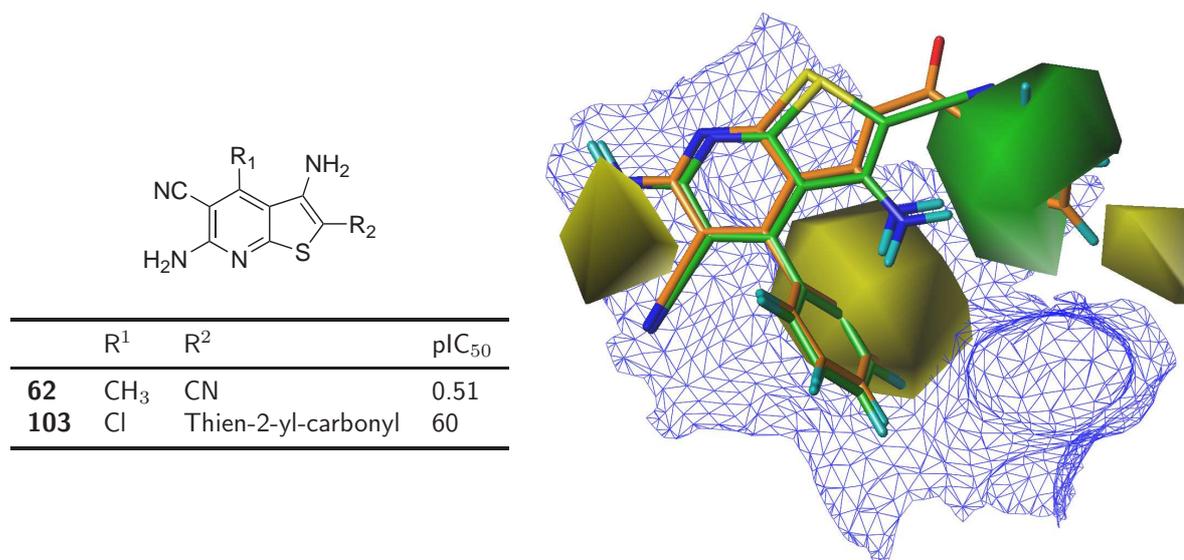


Abbildung 4.11. | Sterisches Feld mit Liganden **62** in Grün und **103** in Orange (pIC₅₀ an der PfgSK-3). Das Feld, das die Substitution des 4-Arylrestes in Position 2 verdeckt, passt genau auf das Profil des als blaues Gitter dargestellten Taschengrundes. Die Form des grünen Feldes verdeutlicht eine bevorzugte Substitution des Benzoylrestes in meta-Stellung.

Beispiele dafür, dass die Aktivität in der Folge I < Br < Cl zunimmt (Beispiele vgl. S. 84). Der kleinere Substituent an dieser Stelle bewirkt also die höhere Aktivität, was sich im Modell durch das dargestellten Feld zeigt. Dass unsubstituierte Derivate der Klasse 1b inaktiv sind, wird vom Modell nicht berücksichtigt, da diese Verbindungen (**97**, **106**) aufgrund der fehlenden diskreten Angabe des IC₅₀ nicht in das Modell eingingen.

Die Aktivität verstärkend wird dem Modell nach der durch den Benzoylsubstituenten besetzte Bereich, im Bild grün dargestellt. Diese Beobachtung steht im Widerspruch zu der Tatsache, dass Verbindungen der Klasse 1a häufig aktiver sind als diejenigen der Klasse 1b. Insbesondere die am 2-Benzoylrest unsubstituierten Derivate wie das in der Abbildung dargestellte **103** zählen zu den weniger aktiven Verbindungen des Datensatzes. Wird der Benzoylrest jedoch in meta-Position zum Carbonyl substituiert, sind die betreffenden Verbindungen mit Aktivitäten kleiner 1 μM aktiv (z. B. **66**, **74**, **77**). Dem Modell zufolge verstärkt also eine Benzoylsubstitution dann die Aktivität, wenn der betreffende Aromat in meta-Stellung substituiert vorliegt, während eine Substitution in para vergleichsweise eine geringere Aktivität von Liganden mit sich bringt, die diesen Raum (gelbes Feld in Abb. 4.11) besetzen. Auch diese Beobachtung deckt sich mit den bekannten Struktur-Wirkungsbeziehungen nach Abschnitt 3.2.1.1.

4.3.2.3. Das Modell AcEI

Die von diesem Modell berechneten pIC₅₀-Werte und die sich ergebenden Residuen sind im Anhang in Tabelle B.12 auf Seite 245 dargestellt, den entsprechenden Residualplot zeigt Abbildung 4.12. Substanzen, bei denen die Abweichung des berechneten vom gemessenen Wert mehr

Verbindung	R ¹	R ²	R ^{3a/3b}
68	2-Iod-phenyl	3,4-Dichlorbenzoyl	6-Amino-5-cyano
121	Thien-2-yl	Benzoyl	anel. Cyclopentan
139	Furan-2-yl	Carboxamid	anel. Cylohexan
107		<i>siehe links</i>	
77	2-Chlorphenyl	3-Chlor-benzoyl	6-Amino-5-cyano
64	Indol-3-yl	CN	6-Amino-5-cyano
87	2-Methoxy-phenyl	3-Chlor-benzoyl	6-Amino-5-cyano
140	Thien-2-yl	Carboxamid	anel. Cycloheptan
84	2-Fluor-phenyl	3-Chlor-benzoyl	6-Amino-5-cyano

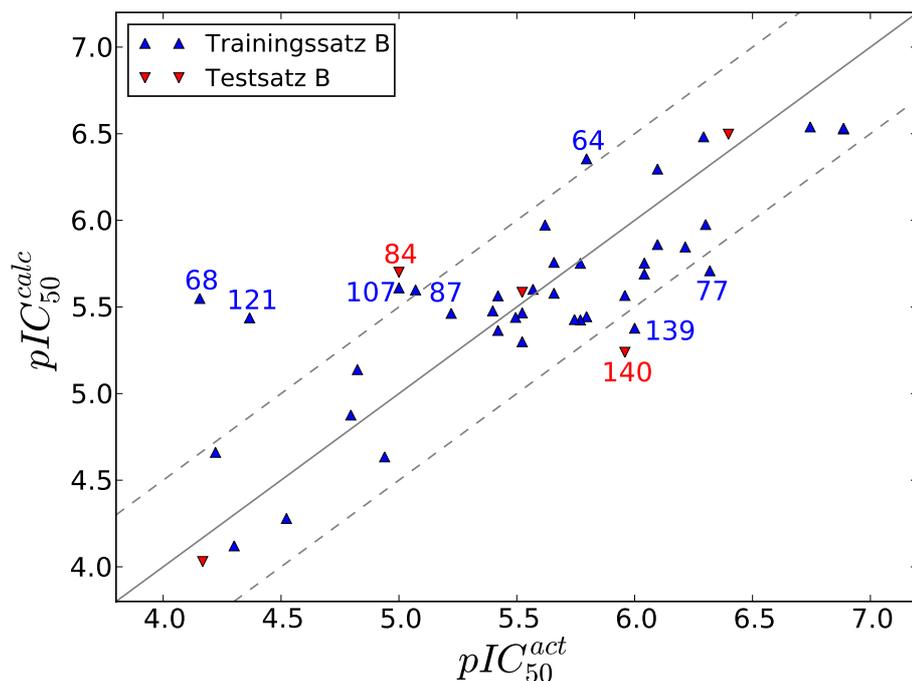
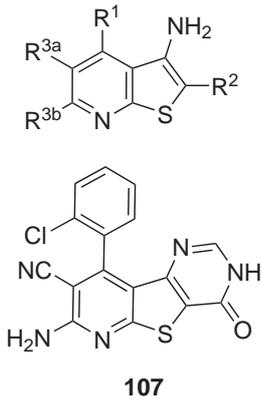


Abbildung 4.12. | Grafische Auftragung der Tabelle B.12. Bei den bezeichneten Verbindungen weichen berechneter und gemessener pIC_{50} -Wert mehr als 0.5 logarithmische Einheiten voneinander ab.

als 0.5 logarithmische Einheiten beträgt, sind wieder bezeichnet und als Struktur dargestellt. Im Vergleich zum vorangegangenen Modell AcDoSt trifft dies auf mehr Verbindungen zu, was aufgrund des schlechteren Korrelationskoeffizienten (0.629 *vs.* 0.786) zu erwarten war. Bezogen auf den Testsatz zeigen sich Parallelen: In beiden Modellen werden zwei Verbindungen schlecht vorhergesagt, in beiden ist eine davon **84**.

Bei den in Abbildung 4.12 dargestellten Verbindungen, deren pIC_{50} vom Modell schlecht berechnet werden, fällt im Unterschied zum Modell AcDoSt (vgl. vorangehender Abschnitt) auf, dass für Substanzen *beider* Klassen Abweichungen über 0.5 logarithmische Einheiten erhalten werden. Offenbar ist die Kombination von Akzeptor- und elektrostatischem Feld nicht so gut in der Lage, die Verbindungen der Klasse 2 zu beschreiben, wie dies mit den Ähnlichkeitsindizes des Akzeptor-, Donor- und des sterischen Feldes möglich ist.

Die größte Abweichung zeigt wiederum **68** (Erklärungsansatz siehe oben). Die schlechtere Vorhersage von Verbindungen der Klasse 2 ist insofern erklärbar, als dass gerade durch das Vorhandensein des anellierten gesättigten Ringes anstelle der polaren potentiellen Interaktionspunkte Amin und Nitril im Molekül das sterische Feld an Bedeutung gewinnt. Da dieses in AcEl nicht berücksichtigt wird, ermöglicht die Kombination AcDoSt für diese Verbindungen das zutreffendere Modell.

Wird zur Validierung durch Progressive Scrambling wieder schrittweise Unschärfe in den Datensatz eingeführt, so fallen auch hier die mit dem stufenweise durchmischten Datensatz berechneten Korrelationsmaße der Kreuzvalidierung q^2 ab (vgl. Abb. 4.9(b) auf S. 162). Allerdings sinken die Werte hier im Vergleich zu denen mit zufällig vertauschtem Datensatz im Modell AcDoSt weniger stark, der kleinste Wert liegt bei -0.292. Außerdem liegen 66 von 400 der permutierten q^2 über demjenigen, der für das Modell ohne randomisierte biologische Daten erhalten wurde – bei der Felderkombination AcDoSt waren es nur 12. Auch wenn dies erst ab einem Korrelationsgrad $r_{yy'}^2$ des Originaldatensatz mit dem durchmischten Datensatz von mehr als 0.746 der Fall ist, sollte die Interpretation des Modells unter Vorbehalt geschehen.

Analyse der CoMSIA-Konturdiagramme

Das Modell AcEl gleicht hinsichtlich des Akzeptorfeldes stark dem zuvor vorgestellten AcDoSt. In Abbildung 4.13 sind die Übereinstimmungen des Akzeptorfeldes mit Abbildung 4.10 klar erkennbar. Der dargestellte vergleichsweise aktive Ligand **83** zeigt gut die mögliche Interaktion des 3-Cyano-benzoyl-Restes mit dem Akzeptorfeld zwischen Lys208 und Gln210 auf.

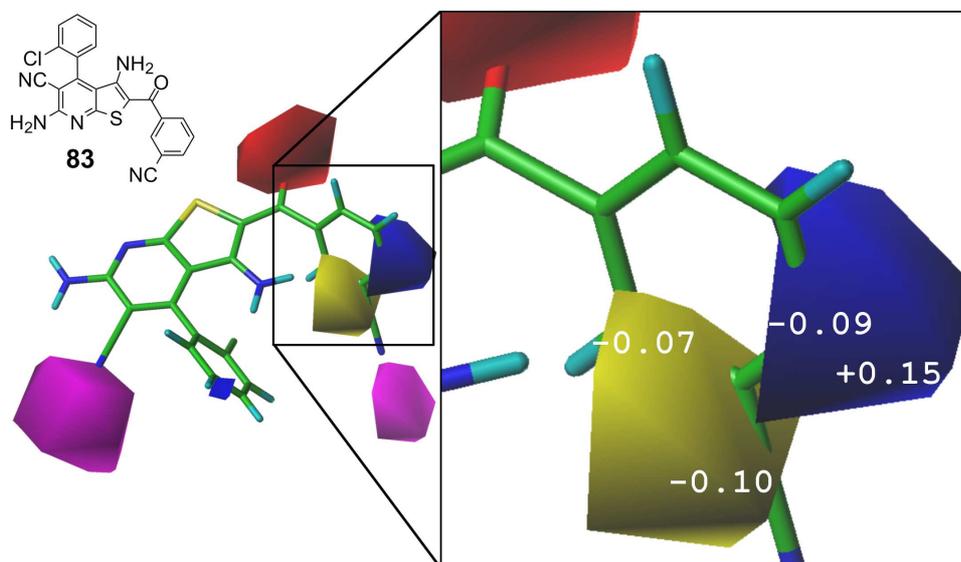


Abbildung 4.13. | Elektrostatisches und Akzeptorfeld mit dem aktiven Liganden **83**. Im gelb markierten Feld wird die Anforderung einer negativen Partialladung sowohl vom bezeichneten Kohlenstoff des Ringes (-0.07) als auch von dem in der Cyanogruppe liegenden (-0.10) erfüllt. Im blau markierten Raum erfüllt vor allem der Wasserstoff (0.15) die Annahme einer positiven Partialladung aktiver Substanzen. Der im aromatischen Ring liegende Kohlenstoff hingegen (-0.09) ist dem zuvor erwähnten Kohlenstoff recht ähnlich.

Dem Cyanosubstituenten an Position 5 steht im Modell ein positives Akzeptorfeld gegenüber. Tatsächlich könnte bei Projektion der Felder in die Bindetasche hier ein Donor im Rezeptor vorliegen: Es handelt sich um die Position, an der in der Röntgenstruktur der *HsGSK-3 β* mit **31** ein Wassermolekül liegt und die im plasmodialen Enzym von Lys166 am Tascheneingang eingenommen werden könnte. Das Modell zeigt nicht die schwer interpretierbaren negativen Akzeptorfelder, die das Modell AcDoSt bei der 6-Aminosubstitution des Liganden im Bereich der Hinge Region aufweist (vgl. Abb. 4.10 auf S. 163).

Zusätzlich kann das elektrostatische Feld zur Interpretation herangezogen werden. Hier unterscheiden sich die im Feld liegenden aromatischen Kohlenstoffe mit Partialladungen von -0.07 und -0.09 wenig. Die am Ring anschließenden Atome erfüllen jedoch gut die Annahme, dass bei einer aktiven Substanz hier eine negative Ladung (-0.10 des Nitrilkohlenstoffes im gelben Bereich) bzw. eine positive (0.15 des Wasserstoffes am para-Kohlenstoff im blauen Bereich) für die Aktivität von Vorteil ist.

4.3.2.4. Modelle basierend auf Modus B

Die verschiedenen Datensätze (Trainingssatz A und B bzw. kompletter Datensatz) ergeben mit der in Abschnitt 4.3.1.3 dargestellten Überlagerung 3D-QSAR-Modelle, deren statistische Kennzahlen in Tabelle 4.4 dargestellt sind. Bei dem Wechsel von Trainingssatz A zu Trainingssatz B verschlechtern sich die Werte für AcDoSt etwas. Diese Tendenz überrascht, da im Trainingssatz B mehr Verbindungen zum Erstellen des Modells zur Verfügung stehen und durch die umfangreichere Information das bessere Modell erwartet würde. Das Modell, das auf dem gesamten Datensatz beruht, erfüllt diese Erwartung. Es zeigt für alle Felderkombinationen den

Tabelle 4.4. | Statistische Parameter der auf Modus B beruhenden Modelle. Im unteren Teil der Tabelle sind in Entsprechung zu Tabelle 4.2 die prozentualen Anteile der am jeweiligen Modell beteiligten physikochemischen Eigenschaften aufgelistet.

	AcDoSt	AcEl	AcElSt	AcDoSt	AcEl	AcElSt	AcDoSt	AcEl	AcElSt
	<i>Trainingssatz A</i>			<i>Trainingssatz B</i>			<i>kompletter Datensatz</i>		
q^2	0.393	0.319	0.288	0.279	0.335	0.320	0.429	0.415	0.407
s_{PRESS}	0.585	0.548	0.569	0.650	0.588	0.595	0.581	0.559	0.562
r^2	0.666	0.481	0.562	0.640	0.607	0.579	0.679	0.618	0.595
s	0.434	0.478	0.446	0.459	0.452	0.468	0.436	0.451	0.465
F	6.221	29.699	19.905	9.477	27.828	24.726	13.040	33.233	30.088
# comp.	8	1	2	6	2	2	6	2	2
Ac.	23.3	55.7	41.4	43.5	40.9	41.4	42.4	42.2	42.2
Do.	41.7	–	–	31.7	–	–	33.0	–	–
St.	34.9	–	8.2	24.8	–	7.4	24.6	–	7.4
El.	–	44.3	50.4	–	59.1	51.3	–	57.8	50.4

Abkürzungen siehe Tab. 4.2 auf S. 158.

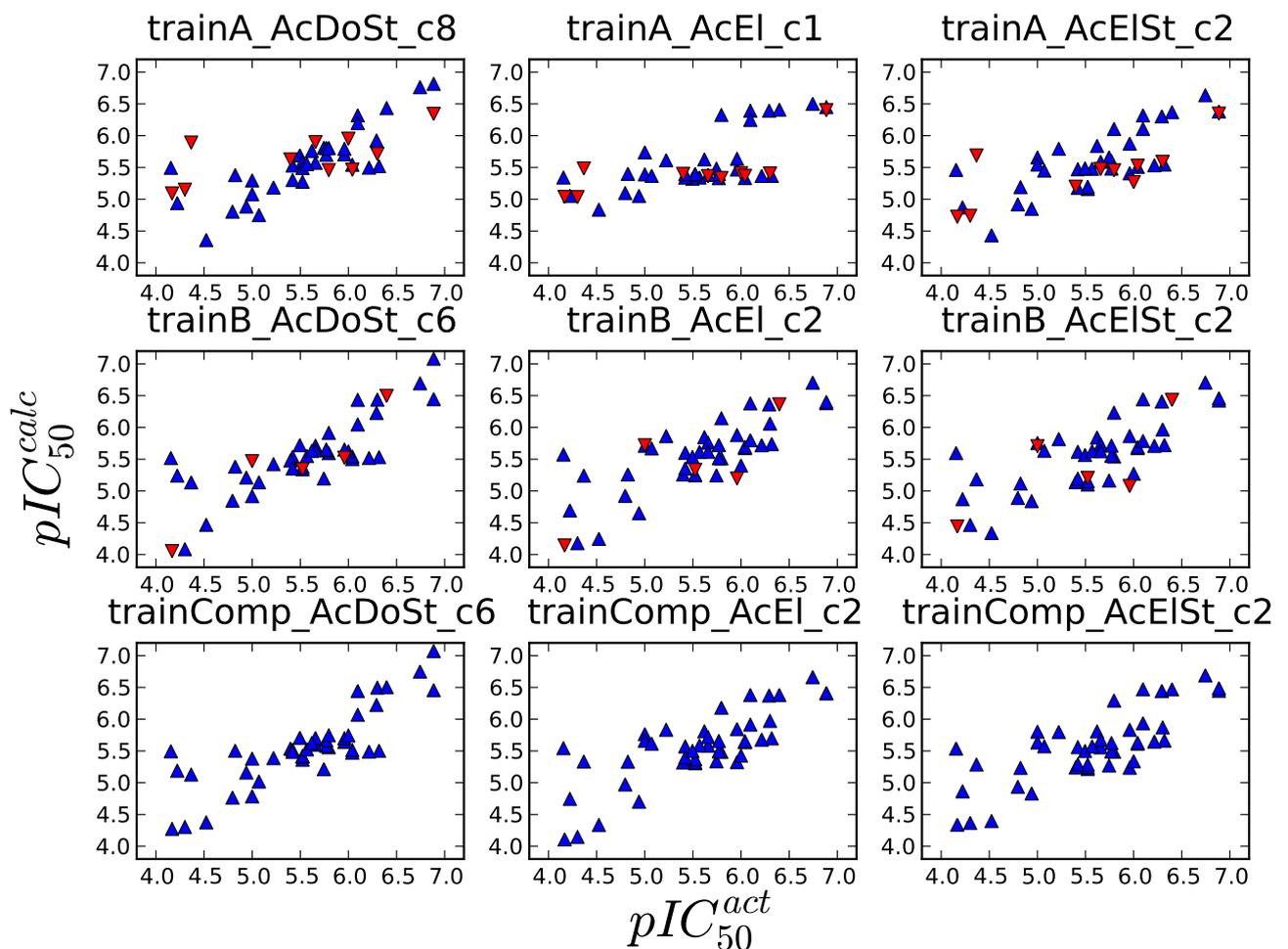
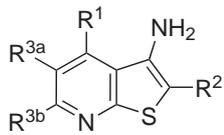


Abbildung 4.14. | Übersicht der Residualplots verschiedener Kombinationen von Feldern und Datensätzen (Bezeichnung analog zu Abb. 4.7).

jeweiligen besten q^2 und F -Wert, die anderen Parameter liegen ebenfalls entweder günstiger oder wenig schlechter als die Ergebnisse der alternativen Datensätze.

Im Vergleich zu den in Tabelle 4.2 dargestellten Ergebnissen, die mit der Überlagerung gemäß Modus A erzielt wurden, sind die Korrelationskoeffizienten der auf Modus B basierenden Modelle schlechter. Der q^2 bzw. r^2 der entsprechenden Felderkombinationen ist kleiner, die Standardabweichungen größer. So liegt bis auf die Modelle, die auf dem gesamten Datensatz beruhen, der q^2 nicht über 0.4 und es gibt kein Modell mit einem $r^2 > 0.7$. Hinsichtlich des F -Wertes ist das Bild im Vergleich zu den auf Modus A beruhenden Modellen uneinheitlich. Bei ausschließlicher Betrachtung des Trainingsatzes B sind die Modelle nach Modus B etwas schlechter als diejenigen nach Modus A. Eine Übersicht der Residualplots für die verschiedenen Kombinationen aus Feldern und Zusammenstellung der Trainings- und Testsätze ist in Abbildung 4.14 dargestellt. Es werden in Entsprechung zu Abschnitt 4.3.2.1 die Modelle näher untersucht, die auf der Basis der Kombination AcDoSt und AcEI mit dem Trainingsatz B berechnet wurden.



Verbindung	R ¹	R ²	R ^{3a/3b}
68	2-Iod-phenyl	3,4-Dichlor-benzoyl	6-Amino-5-cyano
103	2-Chlor-phenyl	Thien-2-yl-carbonyl	6-Amino-5-cyano
77	2-Chlor-phenyl	3-Chlor-benzoyl	6-Amino-5-cyano
121	Thien-2-yl	Benzoyl	anel. Cyclopentan
74	2-Brom-phenyl	3-Chlor-benzoyl	6-Amino-5-cyano
142	4-Chlor-phenyl	Carboxamid	anel. Cyloheptan
137	Thien-2-yl	Benzoyl	anel. Cyclohexan
66	2-Iod-phenyl	3-Chlor-benzoyl	6-Amino-5-cyano

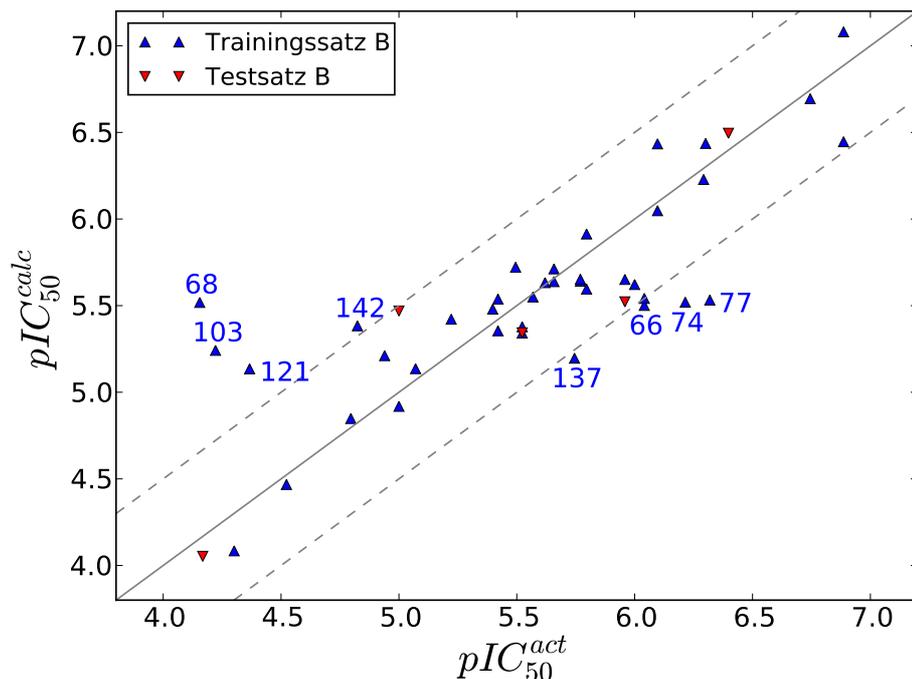


Abbildung 4.15. | Grafische Auftragung der Tabelle B.13. Bei den bezeichneten Verbindungen weichen berechneter und gemessener pIC_{50} -Wert mehr als 0.5 Einheiten voneinander ab.

4.3.2.5. Das Modell AcDoSt

Tabelle B.13 im Anhang auf Seite 246 gibt alle einzelnen Verbindungen und deren berechnete pIC_{50} -Werte durch das Modell AcDoSt wieder. Es fällt zunächst auf, dass im Trainingssatz bei 8 Verbindungen die Differenz zwischen berechnetem und tatsächlichem Wert größer als 0.5 ist (entsprechendes auf Modus A basierendes Modell: 3 Verbindungen), andererseits wird der Testsatz vergleichsweise gut vorhergesagt: Hier beschreibt das Modell alle fünf Verbindungen im Bereich von 0.5 logarithmische Einheiten vom gemessenen Wert. Die Tabelle wird in Entsprechung zu den vorangegangenen Abschnitten in Abbildung 4.15 als Residualplot dargestellt.

Einige der Verbindungen, bei denen sich große Abweichungen von berechnetem und tatsächlichem pIC_{50} zeigen, wurden schon in den Modellen schlecht vorhergesagt, die auf Modus A beruhen (68, 77, 103, 121). Die anderen in Abbildung 4.15 dargestellten Verbindungen haben keine strukturellen Besonderheiten, die die Abweichung erklären könnten. Sie liegen zudem

nur knapp über der Grenze einer Abweichung um 0.5 Einheiten, so dass die Ursache hier eher in der höheren Standardabweichung des Modells (0.459 *vs.* 0.378 bei AcDoSt mit Modus A) zu suchen ist und weniger auf individuelle Abweichungen einzelner Substanzen zurückgeführt werden kann.

In der Validierung mittels Progressive Scrambling ergibt sich für das Modell AcDoSt die in Abbildung 4.16(a) auf Seite 172 dargestellte Verteilung der q^2 -Werte in Abhängigkeit von $r_{yy'}^2$. Es liegen 60 von 400 Werten über dem Niveau des eigentlichen Modells (0.279, rote Markierung bzw. grau gestrichelte Linie). Bei der Kombination von Modus A mit der Felderkombination AcDoSt waren es nur 12. Bereits bei $r_{yy'}^2$ von knapp 0.6 werden vergleichbare Werte für q^2 berechnet, was Zweifel an der Validität des Modells rechtfertigt.

Analyse der CoMSIA-Konturdiagramme

Betrachtet man zunächst die Wasserstoffbrückendonor- und -akzeptorfelder (Abb. 4.17), so gibt es einige Parallelen zum auf Modus A beruhenden Modell. Auffällig ist als erstes das große in der Abbildung rot dargestellte Akzeptorfeld, das eine unvorteilhafte Position eines Donors im Rezeptor bedeutet. Die Carbonylfunktion der Liganden der Klasse 1b zeigt genau in dieses Feld, in dessen Verlängerung folgerichtig der Stickstoff des Backbones von Ile160 liegt. Diese zentrale Interaktion des Modus B entspricht derjenigen der Liganden mit dem Lys108 in Modus A. Auf die im ersten Moment widersprüchliche Tatsache, dass es sich um ein negatives Feld handelt, wurde oben eingegangen (vgl. S. 164).

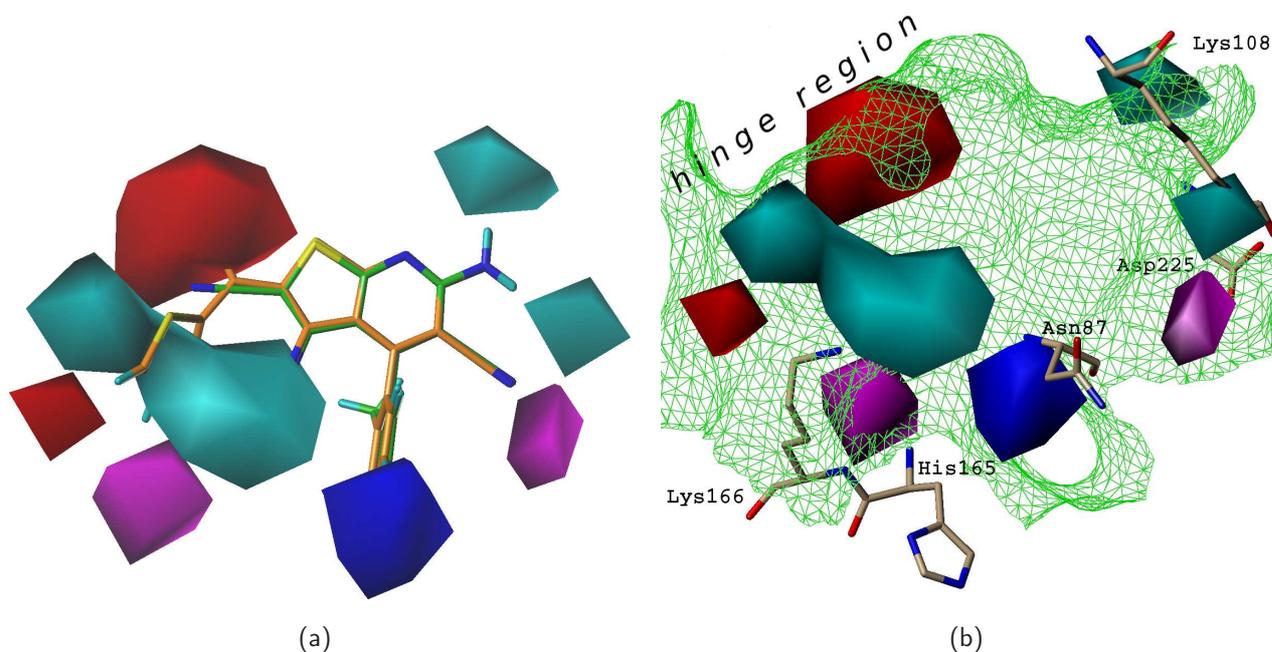
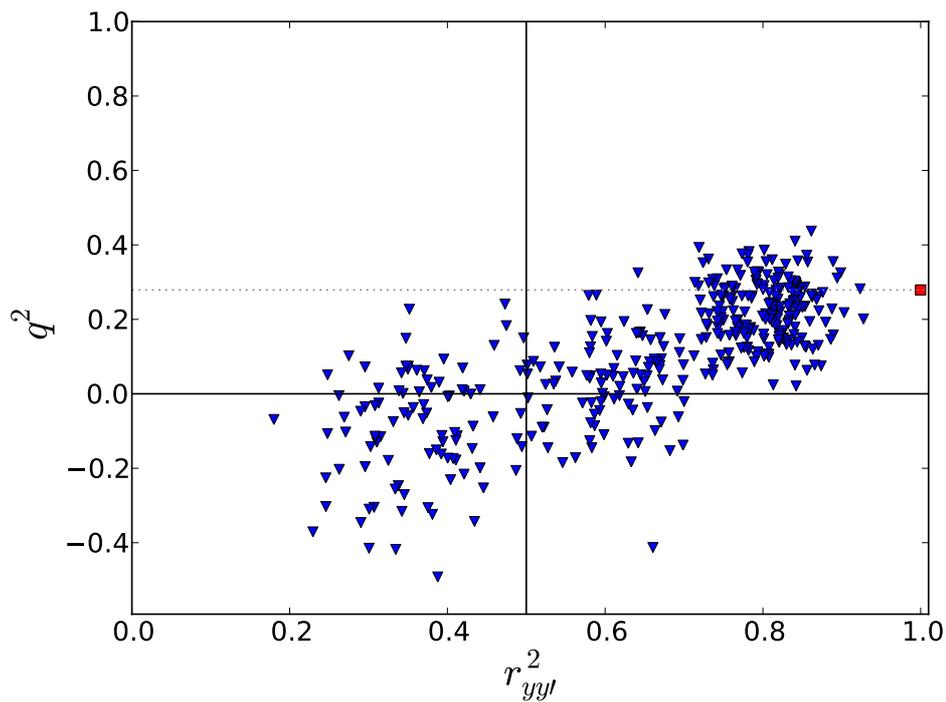
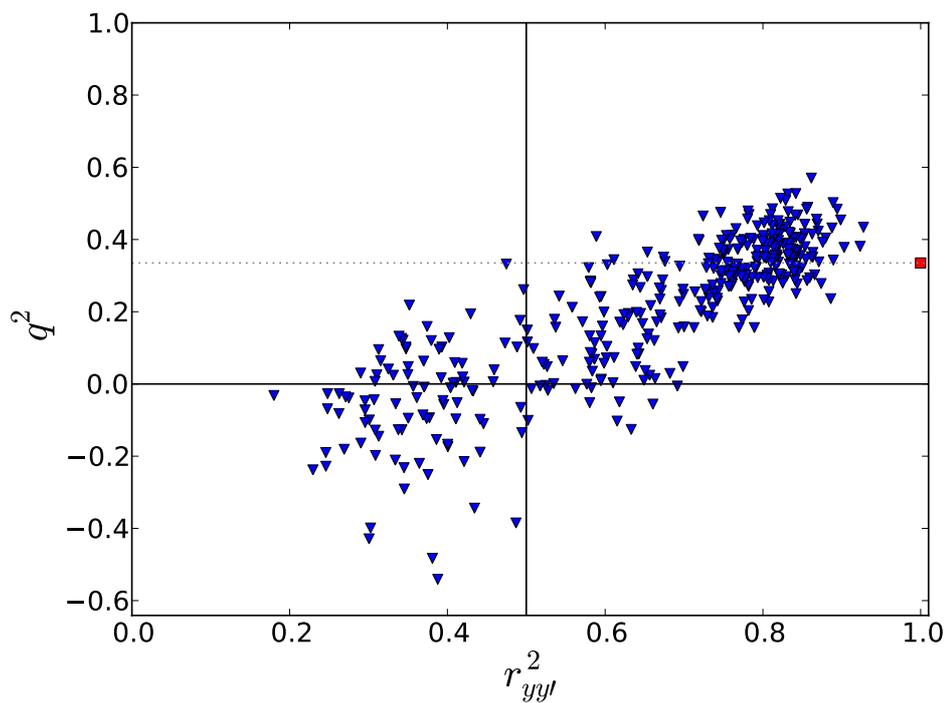


Abbildung 4.17. | **62** (grün) als aktive Verbindung und **103** (orange) als weniger aktive mit Wasserstoffbrückenakzeptor- und -donorfeldern gemäß AcDoSt / Modus B (a); Oberfläche der Bindetasche (b).



(a)



(b)

Abbildung 4.16. | Veränderungen des q^2 beim Progressive Scrambling für Modelle, die auf Modus B beruhen. Im oberen Teil (a) sind die Veränderungen bei Durchmischung des Datensatzes für die Felderkombination AcDoSt abgebildet, im unteren Teil (b) diejenigen für die Felderkombination AcEl. In beiden Fällen liegen im Vergleich zu den auf Modus A beruhenden Modellen mehr Werte über demjenigen, der durch für das eigentliche Modell berechnet wird.

Der einen Donor im Rezeptor begünstigende violett dargestellten Feldbereich in Verlängerung der 5-Cyanosubstitution der Liganden findet kein Gegenstück im Rezeptor. Im Gegenteil: Hier ist als Seitenkette diejenige des Asp225 in der Nähe, die in ihrer deprotonierten Form als Akzeptor einzuordnen wäre. Das zweite violettfarbene positive Akzeptorfeld im Bereich der lösungsmittelzugewandten Region nahe Asn87, His165 und Lys166 (links in Abb. 4.17) liegt zwar nicht direkt über dem Stickstoff von Lys166, wegen der hohen Beweglichkeit der Seitenkette ist jedoch eine Interaktion denkbar.

Die cyanfarbenen positiven Donorfelder nahe der 6-Aminosubstitution des Thienopyridins sind von der Seite des Liganden her betrachtet einsichtig. Bei Projektion in die Bindetasche hingegen sind sie nicht mit den Gegebenheiten im Rezeptor vereinbar. In den betreffenden Bereichen wird ein Akzeptor seitens des Rezeptors erwartet. Der eine Teil dieses Feldes liegt genau beim Stickstoff des Lys108, der bei Vorliegen von Modus B als Donor mit der 5-Cyanosubstitution des Liganden einen Kontakt ausbilden sollte. Wie bereits bei der Beschreibung des Modus B diskutiert, befindet sich das Lys108 jedoch praktisch in der Mitte zwischen der 6-Amino- und der 5-Cyanosubstitution, die drei Stickstoffe bilden ein Dreieck (vgl. Abb. 3.10 auf S. 104). Das Modell spiegelt also die ungewöhnliche Geometrie in diesem Bereich des Protein-Ligand-Komplexes bei Vorliegen des Modus B wider. Der hintere Bereich des cyanfarbenen Donorfeldes wird nahe der Seitenkette des Gatekeepers Met157 berechnet – hier ist ebenfalls kein Wasserstoffbrückenakzeptor seitens des Rezeptors vorhanden. Auch das cyanfarbene begünstigende Donorfeld im Bereich der lösungsmittelzugewandten Tascheneingangs ergibt seitens des Rezeptors keinen Sinn, da keine Seitenkette des Rezeptors mit Eigenschaften eines Wasserstoffbrückenakzeptors in der Lage ist, sich hierhin zu bewegen.

Das sterische Feld ist in Abbildung 4.18 dargestellt. Der gelbe Bereich am Taschengrund in Position 2 der 4-Arylsubstitution ähnelt stark dem in Abbildung 4.11 auf Seite 165 erläuterten des Modells, das auf Modus A beruht. Der grüne positive Feldbeitrag liegt bei Projektion in das Protein im Bereich der lösungsmittelzugewandten Region, so dass hier tatsächlich Platz für einen Arylsubstituenten der Verbindungen der Klasse 1b ist. Allerdings ergeben sich bei Betrachtung der Aktivitäten die in der Bildunterschrift beschriebenen Unstimmigkeiten, so dass die Bedeutung dieses Feldbeitrages zur Erklärung der biologischen Daten fraglich erscheint.

4.3.2.6. Das Modell AcEI

Die Gegenüberstellung der mit diesem Modell berechneten pIC_{50} -Werte mit den gemessenen ist im Anhang in Tabelle B.14 auf Seite 246 dargestellt. Es sind 8 Verbindungen des Trainings- und 2 des Testsatzes, die schlecht vom Modell beschrieben werden (vgl. Abb. 4.19), es handelt sich um Verbindungen beider Klassen.

Bis auf **108** und **85** handelt es sich ausschließlich um Verbindungen, die bereits von anderen Modellen schlecht beschrieben wurde. Im Falle von **108** ist dabei bemerkenswert, dass das Modell zwischen den beiden Analoga **107** und **108** offenbar schlecht differenziert. Mit 5.712

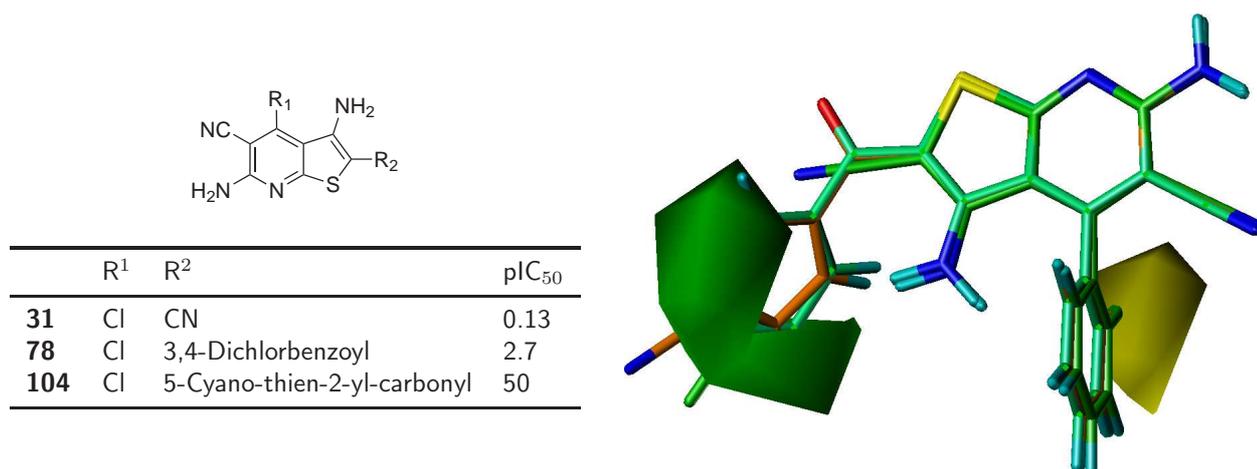
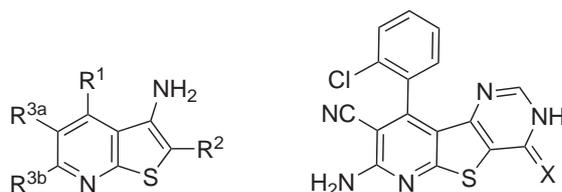


Abbildung 4.18. | Sterisches Feld mit Liganden **31**, **78** (beide als aktive Verbindungen grün) und **104** (als weniger aktive Verbindung orange), Aktivitäten an der PfGSK-3 in μM . Obgleich **78** und **104** beide gleichermaßen den grünen bevorzugten sterischen Bereich mit ihrer 2-Aryl-carbonyl-Substitution erreichen, zeigen sie deutliche Unterschiede in ihrer Aktivität, wohingegen **31** als Verbindung mit der vergleichsweise höchsten Aktivität dieses Feld in keiner Weise berührt. Auch eine Unterscheidung der verschiedenen Positionen einer Substitution am Benzoylsubstituenten, wie sie aus den Struktur-Wirkungsbeziehungen abzuleiten ist, wird durch das Modell nicht ermöglicht. Insgesamt stellen die sterischen Feldbeiträge die Sinnhaftigkeit des Modells demnach infrage.

und 6.384 werden ähnliche Werte für den pIC₅₀ berechnet, obwohl die tatsächliche Differenz bei knapp 2 Größenordnungen liegt (5.000 vs. 6.886). Der Grund dafür ist in dem abweichenden Grundkörper zu suchen, der sich von den anderen Verbindungen durch das anellierte Pyrimidin unterscheidet und deshalb vom Modell nicht gut beschrieben wird. Dieser Unterschied führte bereits im auf Modus A basierenden Modell mit der Felderkombination AcEl dazu, dass **107** schlecht vorhergesagt wurde. Die Modelle mit der Felderkombination AcDoSt hingegen in der Lage, diese strukturelle Variante zu erfassen.

Für das Modell spricht andererseits, dass die Verbindungen **101-104**, die von anderen Modellen schlecht beschrieben werden, alle mit Abweichungen von weniger als 0.5 Einheiten vorhergesagt werden können.

Bei der Validierung mittels Progressive Scrambling schneidet das Modell im Vergleich zu den zuvor betrachteten schlechter ab. So wird in Abbildung 4.16(b) auf Seite 172 deutlich, dass in 131 von 400 Fällen der q^2 von Modellen basierend auf dem zufällig durchmischten Datensatz größer als 0.335 ist und damit höher als der im eigentlichen Modell berechnete liegt. Selbst bei einem $r_{yy'}^2$ von 0.475 wird mit $q^2 = 0.332$ fast der vom Modell erzielte Wert in der Kreuzvalidierung erhalten. Dieser Befund zeigt, dass die Validität des Modells infrage zu stellen ist.



Verbindung	R ¹	R ²	R ^{3a/3b}	X
68	2-Iod-phenyl	3,4-Dichlor-benzoyl	6-Amino-5-cyano	
121	Thien-2-yl	Benzoyl	anel. Cyclopentan	
107				O
85	2-Methoxy-phenyl	Benzoyl	6-Amino-5-cyano	
139	Furan-2-yl	Carboxamid	anel. Cylohexan	
87	2-Methoxy-phenyl	3-Chlor-benzoyl	6-Amino-5-cyano	
77	2-Chlorphenyl	3-Chlor-benzoyl	6-Amino-5-cyano	
108				S
140	Thien-2-yl	Carboxamid	anel. Cycloheptan	
84	2-Fluor-phenyl	3-Chlor-benzoyl	6-Amino-5-cyano	

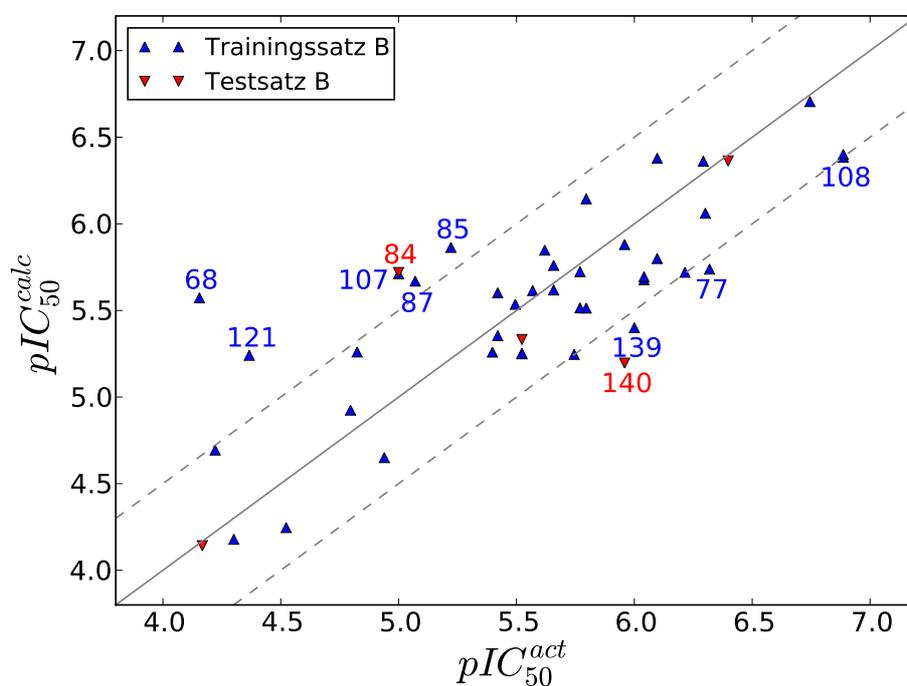


Abbildung 4.19. | Grafische Auftragung der Tabelle B.14. Bei den bezeichneten Verbindungen weichen berechneter und gemessener pIC_{50} -Wert mehr als 0.5 Einheiten voneinander ab.

Analyse der CoMSIA-Konturdiagramme

Aus Gründen der Konsistenz wird die Analyse der Konturdiagramme dennoch beschrieben. Die berechneten Felder sind in Abbildung 4.20 dargestellt. Hinsichtlich des Akzeptorfeldes sind wieder große Ähnlichkeiten zu den anderen Modellen zu erkennen (roter Bereich in Verlängerung der Carbonylgruppe an C2, violetter Bereich auf der rechten Seite der Abbildung). Interessant ist hier das elektrostatische Feld. Innerhalb des Modells ist die Anordnung der Felder klar verständlich: **125** als vergleichsweise wenig aktive Substanz mit einem IC_{50} von $30 \mu M$ hat im blauen Bereich Atome mit negativer Partialladung und im gelben den Schwefel mit der bezeichneten positiven. Die Felder besagen allerdings, dass für eine erhöhte Aktivität die umgekehrte Anordnung günstig wäre.

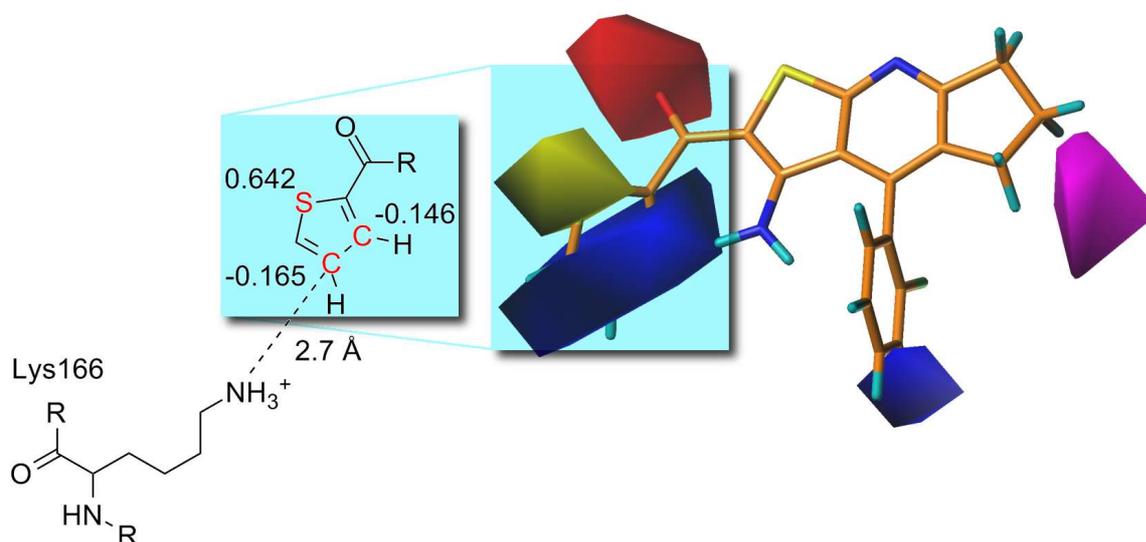


Abbildung 4.20. | Akzeptor- und elektrostatisches Feld von AcEI. In Orange eingezeichnet ist **125**.

Interessant im Bezug auf den Vergleich der Modelle ist die Projektion in die Bindetasche. Das genannte blaue Feld befindet sich im Abstand von etwa 2.7 \AA vom Lys166 im lösungsmittelzugewandten Bereich der Tasche. Dem Modell nach soll gerade hier eine positive Partialladung der Liganden günstig sein. Eine direkte Wechselwirkung mit dem Lys166, wie sie für den Modus B anzunehmen wäre, kommt also dem Modell nach nicht infrage. Das Modell widerspricht damit der Hypothese zum Bindungsmodus, die der Überlagerung zugrunde liegt und stützt den Modus B nicht.

4.4. Fazit 3D-QSAR

Die erste der drei als Ausgangspunkt dieses Kapitels gewählten Überlagerungen führte nicht zu interpretierbaren 3D-QSAR-Modellen. In ihr lagen die Verbindungen der Klassen 1 und 2 in den unterschiedlichen Modi A und H vor, es war nicht möglich, mit dieser Überlagerung akzeptable Modelle zu berechnen.

Ausgehend von den beiden Überlagerungen, bei denen alle Moleküle auf die gleiche Zielstruktur überlagert wurden, konnten hingegen unterschiedliche Modelle entwickelt werden, deren Qualität aufgrund der statistischen Kennzahlen und anhand der Interpretation der Konturdiagramme untersucht wurde. Hinsichtlich der Korrelationskoeffizienten und Standardabweichungen sind die Modelle insgesamt im Vergleich zu anderen publizierten Modellen weniger aussagekräftig, können jedoch mit Einschränkungen beim zuletzt beschriebenen Modell als interpretierbar angesehen werden.

Die Interpretation der Konturdiagramme spiegelte innerhalb der Modelle nachvollziehbar bekannte Struktur-Wirkungsbeziehungen wider. Bei Projektion in die Bindetasche der PfGSK-3 ergaben sich insbesondere bei den auf Modus A beruhenden Modellen sinnvolle Interpretationsmöglichkeiten.

Zusammenfassend können aus den Ergebnissen der 3D-QSAR-Analysen folgende Schlüsse gezogen werden, die zur Beantwortung der aus Kapitel 3 entwickelten Fragen herangezogen werden können:

1. Eine Unterscheidung der Klassen 1 und 2 hinsichtlich ihres Bindungsmodus gemäß A und H wird nicht bekräftigt, da die entsprechende Überlagerung keine interpretierbaren Modelle ergibt.
2. Modelle, die auf einer Überlagerung auf der Basis von Modus A entstehen, ergeben akzeptable Modelle, mit denen sich die Aktivitäten der untersuchten Datensatzes erklären lassen. Auch im Bezug auf den Rezeptor sind sinnvolle Entsprechungen von Konturdiagrammen und PfGSK-3 möglich.
3. Werden Modelle basierend auf Modus B als Beispiel eines alternativen Bindungsmodus entwickelt, sind die statistischen Kennzahlen der Modelle schlechter, die Validierung ergibt hinsichtlich Progressive Scrambling und Vorhersage eines externen Testsatzes unterlegene Ergebnisse und auch die Projektion in die Bindetasche ist weniger gut möglich.

Die 3D-QSAR-Untersuchungen weisen auf das gemeinsame Vorliegen des Bindungsmodus A bei allen untersuchten Thienopyridinen als Inhibitoren der PfGSK-3 hin.

4.5. Experimenteller Teil: Computational Details

Überlagerungen und deren Analyse mittels SAMPLS, Abschnitt 4.3.1

Für die orientierenden SAMPLS-Berechnungen zur Auffindung des optimalen Alignments wurde ausschließlich der Trainingsatz A verwendet. Die verschiedenen Überlagerungen wurden mit ROCS oder FlexS erstellt. Die Ausgangsdateien wurden für alle Verbindungen analog zur Vorbereitung von Liganden für das Docking mit FlexX/E vorbereitet (vgl. S. 137). Für die Alignments mit ROCS wurden zugrunde liegende Konformerbibliotheken jeweils mit OMEGA [239] generiert (`maxconfs=400`, `ewindow=100`, `rms=0.1`). Im Anschluss wurden Partiaalladungen mit QUACPAC [278] zugewiesen. Dabei wurden AM1-Ladungen [279] für die QSAR-Berechnungen gewählt. Im Falle von FlexS-Überlagerungen wurde der Zielstruktur AM1-BCC-Ladungen [280, 281] zugewiesen, das Alignment erfolgte mit Standardeinstellungen. Für die eigentliche QSAR-Berechnung wurden die Moleküle des erhaltenen Alignments wiederum AM1-Ladungen versehen, da die entsprechenden Ergebnisse den mit AM1-BCC erhaltenen überlegen waren.⁴

Für die Überlagerung 4.3.1.1 nach den Modi A und H wurde für Verbindungen der Klasse 1 eine zuvor als vorteilhaft ermittelte Überlagerung mit ROCS übernommen. Diese war mit **137** in einer Dockinglösung nach Modus A aus DoReCCoM als Zielmolekül konstruiert worden. Dabei mussten die Verbindungen **56**, **58**, **31** und **63** manuell angepasst werden, da ROCS für diese 2-cyano-substituierten Derivate in nicht nachvollziehbarer Weise von dem Grundkörper entfernte Lösungen berechnete. Dies geschah, indem der Thienopyridinring auf denjenigen eines der erwartungsgemäß überlagerten Derivate der Klasse 1a angepasst wurde. Die Verbindungen der Klasse 2 wurden auf den Repräsentanten **121** im Modus H überlagert.

Die Überlagerung 4.3.1.2 aller Moleküle gemäß Modus A wurde mit FlexS auf die Vorlage von **78** erstellt. In Entsprechung dazu wurde bei Modus B vorgegangen (Überlagerung in Abschnitt 4.3.1.3), hier erzielte die mit ROCS erstellte Überlagerung bei den orientierenden SAMPLS-Berechnungen die besseren statistischen Kennzahlen und wurde daher übernommen.

Die SAMPLS Berechnungen wurden mit bis zu 10 Komponenten durchgeführt. Es wurde die Anzahl an Komponenten für das endgültige Modell übernommen, für die der geringste *s_{PRESS}* berechnet wurde.

Gegenüberstellung der Modelle, Abschnitt 4.3.2

Für alle 3D-QSAR-Berechnungen und die Darstellungen der Moleküle mit korrespondierenden Konturdiagrammen wurde in diesem Kapitel Sybly-X 1.2 [183] verwendet. Für die Berechnung der CoMSIA-Felder wurden Standardparameter gewählt: Gitterabstand: 2 Å, automatische Auswahl der Region (minimaler Abstand von 4 Å von jedem Molekül), Sonden: Ø 1 Å mit

⁴Auf diese Weise wird das Vorgehen dem in der Literatur beschriebenen Vorteil semiempirischer Ladungen gegenüber Gasteiger-Hückel [282] gerecht. Allerdings überrascht die Beobachtung insofern, als dass es in der Literatur Hinweise auf die Überlegenheit der AM1-BCC-Ladungen gegenüber AM1 gibt [281, 283].

einfacher Ladung bzw. einfachen Hydrophobizitäts- und Wasserstoffbrückeneigenschaften, kein column filtering).

Für das Progressive Scrambling wurde schrittweise von 5 auf 2 Gruppen reduziert. Es wurden für jede Einteilung des Datensatzes 100 SAMPLS-Analysen berechnet (\Rightarrow 400 Datenpunkte für die Abbildungen 4.9 bzw. 4.16 mit q^2 vs. $r_{yy'}^2$).

5. Diskussion und Ausblick

Der Versuch einer Erklärung der Selektivität von Liganden für ein bestimmtes Protein kann vonseiten der Zielstruktur oder des Liganden ausgehend erfolgen. Wird vom Protein ausgegangen, so ist der erste Schritt der Vergleich der Sequenzen der abzugrenzender Proteine, die Suche nach Unterschieden in deren Primärstruktur. Davon ausgehend kann der Blick auf die Sekundär- und Tertiärstruktur geweitet werden. Schließlich kann auch die Quartärstruktur für einen Protein-Ligand-Komplex von Bedeutung sein, so dass die genannten Blickwinkel Stationen in der Diskussion dieser Arbeit darstellen.

Für das Design neuer Inhibitoren ist der Bindungsmodus der Liganden von entscheidender Bedeutung. Zudem kann die Analyse der Struktur-Wirkungsbeziehungen Impulse zur Weiterentwicklung geben, so dass diese beiden Themenkomplexe den zweiten Teil der Diskussion ausmachen.

5.1. Homologie-Modelle und MIFs

5.1.1. Die Sequenzen von *PfGSK-3* und *HsGSK-3 β* als Ansatzpunkt der Entwicklung selektiver Inhibitoren

Im Abschnitt 2.4.1.1 auf Seite 47 ist das Sequenzalignmet von *PfGSK-3* und *HsGSK-3 β* dargestellt. Zu den hinsichtlich der Entwicklung selektiver Inhibitoren vielversprechenden Abweichungen zählt der Gatekeeper (human: Leu132, plasmodial: Met157). Unterschiedliche Aminosäuren an dieser Schlüsselposition der ATP-Bindetasche spielten in anderen Fällen bei der Entwicklung selektiver Kinaseinhibitoren eine entscheidende Rolle (vgl. Abschnitt 1.2.4 auf S. 13). Allerdings zählen Methionin und Leucin mit 40.0% (häufigste Aminosäure als Gatekeeper) und 17.1% (dritthäufigste, vgl. [59]) zu den häufig vertretenen Aminosäuren bei humanen Proteinkinasen. Ein großer Teil dieser Kinasen hat somit den gleichen Gatekeeper. Gegenüber diesen Kinasen kann eine auf dieser Unterscheidung begründete Selektivität demnach nicht erzielt werden. Auch wenn in frühen Stadien der Entwicklung zunächst die Abgrenzung von *HsGSK-3 β* und *PfGSK-3* im Vordergrund der Untersuchungen steht, zeichnet sich hier ein Problem ab, das bei der Weiterentwicklung der Inhibitoren eine Rolle spielen könnte.

Ausgehend von der Sequenz entwickelten HUANG *et al.* 2010 eine Klassifizierung humaner Proteinkinasen in der Art eines Fingerprints, in dem für Selektivität relevante Aminosäuren in

der Bindetasche kodiert werden – der Gatekeeper zählt dabei zu den wichtigen Aminosäuren [284]. Die Autoren leiten aus dem Fingerprint die Wahrscheinlichkeit ab, mit der die Entwicklung selektiver Inhibitoren vom Typ I für eine Proteinkinase möglich ist. Die *HsGSK-3 β* wird dabei als relativ wenig geeignet eingestuft, da sie nur in 2 von 9 Kriterien von der größten Gruppe von Proteinkinasen abweicht, die den Autoren zufolge als promisk gilt. Erstellt man den analogen Fingerprint für die *PfGSK-3*, so ist dieser identisch mit dem der *HsGSK-3 β* . Auf den Gatekeeper bezogen liegt dies darin begründet, dass Leucin und Methionin in die gleiche Gruppe von Aminosäuren eingeordnet werden. Die Identität innerhalb der Bindetasche ist außerdem so groß, dass sich keines der anderen Bits des Fingerprints unterscheidet. Auch wenn die Sequenzen einen ersten Hinweis auf Unterschiede zwischen *HsGSK-3 β* und *PfGSK-3* geben, reicht diese zweidimensionale Betrachtungsweise offenbar nicht zur hinreichenden Erklärung einer Selektivität von Verbindungen an *HsGSK-3 β* und *PfGSK-3* aus.

Überträgt man die Unterschiede in den Sequenzen ins Dreidimensionale, so entsteht ein Eindruck von möglichen Ansatzpunkten im Design von selektiven Inhibitoren. Die Position sich unterscheidender Aminosäuren ist in Abbildung 5.1 im humanen Enzym dargestellt. Um die Veränderungen in der *PfGSK-3* konkret fassen zu können, war die Konstruktion von Homologie-Modellen als Vehikel zur Darstellung der Unterschiede zwischen *HsGSK-3 β* und *PfGSK-3* nötig (vgl. Kapitel 2).

5.1.2. Homologie-Modelle der *PfGSK-3*

Dass eine unerlässliche Voraussetzung zur Arbeit mit Homologie-Modellen die ausführliche Evaluierung der Strukturen ist, führte im Kontext der Entwicklung antiplasmodialer Wirkstoffe zur Arbeit von GOWTHAMAN *et al.* [285]. Die Autoren stellten 2005 eine Datenbank zusammen, die Homologie-Modelle für 476 plasmodiale Enzyme beinhaltet. Die Modelle wurden mit MODELLER „semiautomatisch“ erstellt (manuelle Anpassung des Alignments im Einzelfall, Energie-Minimierung mit AMBER, Kontrolle stereochemischer Parameter mit PROCHECK). Die Autoren betonen die Fehleranfälligkeit einer vollautomatischen Erstellung der Modelle ohne Evaluierung im Einzelfall. Sie sehen aber durch die erstellten Modelle eine verbesserte Möglichkeit, Einblicke in die Mechanismen einer plasmodialen Infektion zu erhalten und außerdem mögliche Targets zu identifizieren. Die *PfGSK-3* ist allerdings nicht in der Datenbank enthalten.

Vor dem Hintergrund der Notwendigkeit einer eingehenden Evaluierung der eigenen Modelle war das in Kapitel 2 beschriebene Bewertungsverfahren der Modelle mit *RCS* eine wertvolle Hilfe bei der Beurteilung der Modelle. So konnte zunächst eine Vielzahl an Homologie-Modellen erstellt werden, damit der konformative Raum möglichst gut abgedeckt werden kann. Ein Problem bei der Beurteilung der dadurch erhaltenen großen Anzahl an Modellen war, dass die eingesetzten Evaluierungsprogramme teils widersprüchliche Reihungen der Modelle nahelegten. Mit der Berechnung des *RCS* und einer entsprechenden Priorisierung konnten diejenigen Homologie-Modelle ausgewählt werden, die im Vergleich zu den auf der gleichen Vorlage ba-

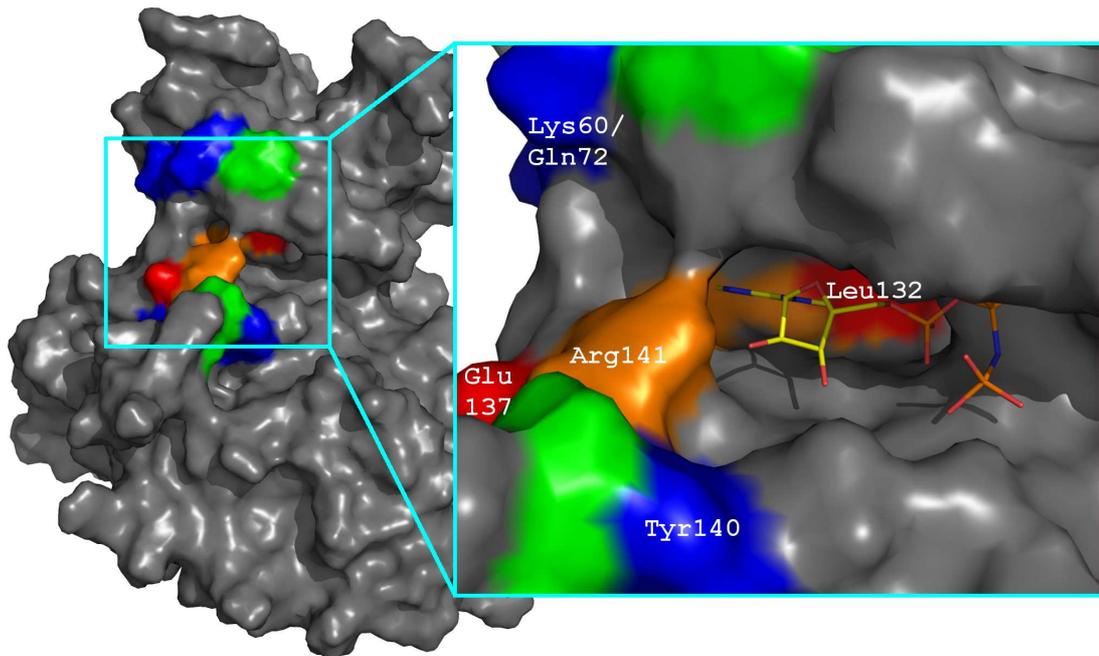


Abbildung 5.1. | Positionen von Aminosäuren, die sich in *HsGSK-3β* und *PfGSK-3* unterscheiden. Abgebildet ist der PDB-Eintrag 1J1B der *HsGSK-3β*. Rot und orange markierte Aminosäuren liegen innerhalb der Bindetasche, wobei die Abweichungen zwischen *HsGSK-3β* und *PfGSK-3* bei Leu132 und Glu137 eher für das Design selektiver Inhibitoren geeignet sind als bei den orange gekennzeichneten Aminosäuren, die sich zwar unterscheiden, hinsichtlich der physikochemischen Eigenschaften allerdings hohe Ähnlichkeit aufweisen. Blau und grün markierte Aminosäuren sind ebenfalls unterschiedlich in *HsGSK-3β* und *PfGSK-3*, allerdings weiter von der eigentlichen Bindetasche entfernt und im Falle der grünen Aminosäuren wiederum so nahe verwandt, dass sie als Ansatzpunkt für Selektivität weniger geeignet erscheinen (vgl. Tab. 2.4 auf S. 49).

sierenden Modellen Vorteile in mehreren der berechneten Parametern aufweisen. Dabei wurde durch die parallele Bewertung von Vorlagen und Modellen (vgl. Abb. 2.8 und Tab. 2.5) klar, dass die Abstufung des *RCS* in den Vorlagen angelegt war und sich durch diese auf die Modelle überträgt.

Die beiden am besten bewerteten Vorlagen waren 1J1B und 1Q41. Auch bei den resultierenden Homologie-Modelle belegen auf diesen Vorlagen basierende Strukturen die vordersten Plätze. Das am besten bewertete Homologie-Modell *pfgsk_1j1b_128* beruht auf 1J1B. Es zeichnet sich auf den ersten Blick dadurch aus, dass es im gesamten Protein nur einen einzigen zu engen Kontakt aufweist, dieser liegt nicht in der Bindetasche. Zudem sind Anteile an Aminosäuren in günstigen Bereichen des Ramachandran Plots hoch. Das unter allen 4400 Modellen als zweitbeste eingeordnete *pfgsk_1q41_72* basiert auf der Vorlage 1Q41, es zeichnet sich vor allem durch gute Werte in den anderen ermittelten Parametern aus (vgl. Tab. B.1 auf S. 231).

Schlecht schneidet die Kristallstruktur 205K und die aus ihr resultierenden Modelle im *RCS* ab. Dies ist offenbar auf den schon in der Kristallstruktur vergleichsweise geringen Anteil an Aminosäuren in der Kernregion des Ramchandran Plots (71.1%) sowie die hohe Anzahl an zu dichten Kontakten der Struktur insgesamt und insbesondere in der Bindetasche (16 bzw.

3) zurückzuführen. Die zu dichten Kontakte in der Bindetasche sind verteilt auf die Bereiche Glycine-rich Loop (Gly65<O> \Rightarrow Ser66<C>), Hinge Region (Pro136<O> \Rightarrow Glu137<O>) und den Grund der Bindetasche (Ile109<O> \Rightarrow Val110<C>). Die beiden Carbonylsauerstoffe in der Hinge Region sind nur 2.3 Å voneinander entfernt – hier zeigt sich bereits, dass die Struktur auch direkt in einer für die Bindung von Liganden entscheidenden Region in der Bindetasche nicht optimal aufgebaut ist. Die Ramachandran Plots bestätigen diesen Eindruck, in Abbildung 5.2 sind die Ramachandran Plots der Vorlage und der beiden Modelle `pfgsk_2o5k_52` und `pfgsk_2o5k_HBM_88` im Vergleich zu denen der Vorlage 1Q41 und den daraus resultierenden Modellen abgebildet. In jeder der drei Reihen (Vorlagen: (a) und (b), Homologie-Modelle ohne Berücksichtigung der Liganden: (c) und (d), Homologie-Modelle unter Berücksichtigung der Liganden: (e) und (f)) sind auf der rechten Seite, also bei 205K und den resultierenden Modellen, mehr Aminosäuren in unvorteilhafteren Bereichen als bei den entsprechenden Darstellungen für 1Q41. Wie bereits bei 1J1B und dem korrespondierenden besten Homologie-Modell gemäß *RCS* (vgl. Abb. 2.7 auf S. 51) ist zu erkennen, dass problematische Aminosäuren in den Modellen teilweise auf die Vorlagen zurückzuführen sind. In Abbildung 5.2(a) bzw. (c) und (e) trifft dies z. B. auf die Aminosäuren Cys218 und Glu290 zu, deren korrespondierende Aminosäuren Cys243 und Asp315 in den entsprechenden Homologie-Modellen ähnlich ungünstige Kombinationen aus Φ - und Ψ -Winkeln aufweisen. Andererseits ist es durchaus möglich, dass im Rahmen des Homology Buildings Probleme behoben werden können, die in der Vorlage aufzufinden waren (z. B. Lys91, Lys122 und Asn285 in Abb. 5.2(b)).

Aus Tabelle B.2 auf Seite 232 ist ersichtlich, dass die mit Ligandinformation erstellten Modelle sich insbesondere durch schlechtere Gesamtanzahl an engen Kontakten sowie höheren Wert des MODELLER *pdf* und schlechteren *G-factor* auszeichnen. Diese Tendenz ist in Bezug auf die Verteilung der Φ - und Ψ -Winkel auch in Abbildung 5.2 zu erkennen. Visuell unterscheiden sich die dreidimensionalen Strukturen der jeweiligen Modelle gegenüber den ohne Ligandinformation erstellten zwar wenig, da die genannten Unterschiede jedoch konsistent für die ohne Ligandinformation erstellten Homologie-Modelle sprechen, wurden diese für weitere Untersuchungen herangezogen.

5.1.3. Möglichkeiten und Grenzen der Analyse mit molekularen Interaktionsfeldern und deren statistischer Auswertung

5.1.3.1. Die visuelle Analyse molekularer Interaktionsfelder

Besonders vielversprechende Unterschiede zwischen *Hs*GSK-3 β und *Pf*GSK-3 ergibt die visuelle Auswertung des N1+-Feldes. Das ausgeprägtere Feld der *Pf*GSK-3 im Bereich der Back Pocket (vgl. Abb. 2.19 auf S. 64) ist von besonderem Interesse, da hier zusätzlich Unterschiede hinsichtlich des Oberflächenprofils festzustellen sind und der Bereich sich in der Nähe des Gatekeepers befindet. Hier ergibt sich bei Weitung des Blickfeldes eine Beziehung zu Unterschieden

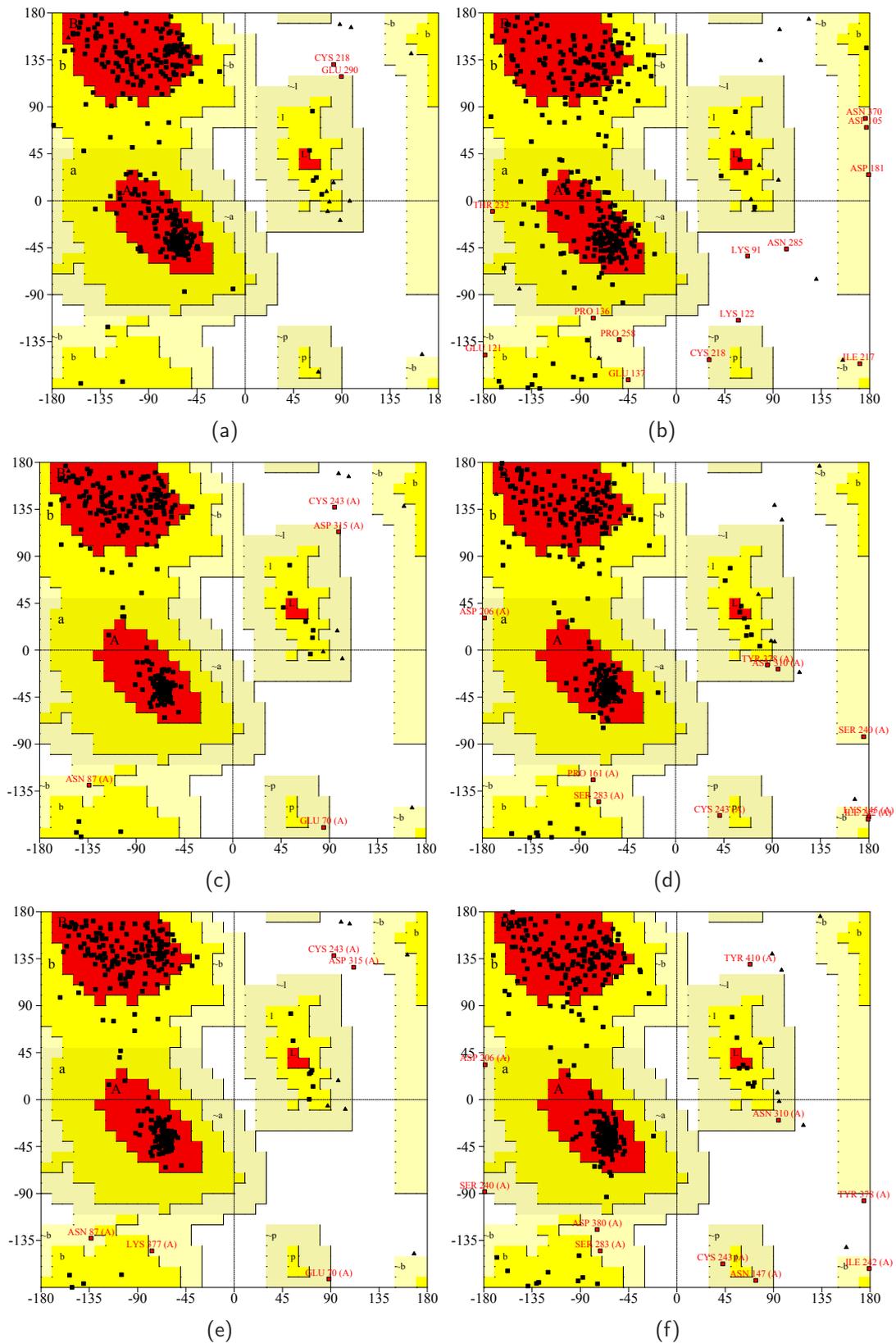


Abbildung 5.2. | Ramachandran Plots der Vorlagen 205K (a) bzw. 1Q41 (b), der Modelle pfgsk_1q41_72 (c) bzw. pfgsk_2o5k_52 (d) und der Modelle pfgsk_1q41_IXM_120 (e) bzw. pfgsk_2o5k_HBM_88 (f).

in der Sequenz: In Tabelle 2.4 auf Seite 49 sind auch weiter vom Zentrum der Bindetasche entfernte Aminosäuren aufgeführt, die sich in *Hs*GSK-3 β und *Pf*GSK-3 unterscheiden. Hier fällt in Zusammenhang mit dem N1+-Feld der Unterschied Lys60/Gln72 (*Hs*GSK-3 β) zu Asn83/Glu95 (*Pf*GSK-3) auf. Im plasmodialen Enzym entfällt damit gegenüber dem humanen eine positive Ladung, es kommt mit dem Glu95 sogar eine negative hinzu, was sich prinzipiell für die Interaktion eines potentiellen Liganden günstig auswirken kann, wenn dieser die Eigenschaften der N1+-Sonde in Form eines entsprechenden Substituenten hat.

Allerdings liegt das Glu95 der *Pf*GSK-3 gut 10 Å vom oben beschriebenen Feld nahe dem Gatekeeper entfernt (vgl. Abb. 5.1 auf S. 183). Somit ergibt sich zwar eine Erklärung für die Unterschiede, da elektrostatische Wechselwirkungen durchaus über weitere Distanzen Einfluss auf die Wechselwirkungsenergien haben (vgl. Berechnung der Interaktionsenergien S. 37). Eine direkte Interaktion von Liganden mit dem Glu95 ist jedoch an dieser Position ausgeschlossen, wenn es sich um einen klassischen Typ-I-Inhibitor handeln soll.

Die zweite Differenz, die sich in mehreren Homologie-Modell/Vorlage-Paaren bestätigt und in der paarweise PCA lokalisiert wird, betrifft die Felder der I- und C3-Sonde. Sie wird in Abschnitt 5.2.2 aufgegriffen werden, ebenso die Verschiebung der Felder verschiedener Wasserstoffbrückenakzeptorsonden entlang der Hinge Region ins Innere der Tasche hinein.

Ein weiteres zentrales Ergebnis der visuellen Analyse der MIFs bestand in Veränderungen der Felder am Gatekeeper, die verstärkte Interaktionsmöglichkeiten mit der DRY-Sonde bei der *Pf*GSK-3 anzeigen. Da der Bereich der Back Pocket hohes Potential hinsichtlich der Entwicklung von selektiven Proteinkinaseinhibitoren bietet (vgl. Abschnitt 1.2.4 auf S. 13) und sich der Unterschied zudem in der paarweisen PCA bestätigt, wäre eine lipophile Interaktion in dieser Position ein erstrebenswertes Ergebnis des Designs. Vor diesem Hintergrund waren einige der Modifikationsvorschläge (1) von Seite 114 zu sehen. Beim Vorliegen des Modus B wäre bei den Verbindungen bis auf 40_mod02a bzw. 69_mod02a ein lipophiler Kontakt der Methylgruppe denkbar gewesen, während bei Vorliegen anderer Modi mit einem Aktivitätsverlust zu rechnen wäre.

5.1.3.2. (C)PCA zur Differenzierung plasmodialer und humaner Enzyme

Die statistische Auswertung der MIFs ergab nur im Fall der paarweisen Betrachtung mittels PCA auswertbare Ergebnisse. Dieses Resümee überrascht insofern, als dass das Verfahren in der Literatur verschiedentlich erfolgreich angewendet wurde (vgl. Abschnitt 2.2) – wie einleitend ausgeführt, war dies der Anlass, das Verfahren einzusetzen.

Die Schwierigkeiten, die Proteine der beiden Spezies mit der CPCA voneinander zu differenzieren, sollen an dieser Stelle in Bezug gesetzt werden zu den Untersuchungen von NAUMANN *et al.* [120], denen mittels CPCA eine Charakterisierung von Kinasen gelang. Die Autoren konnten Familien von ähnlichen Proteinkinasen identifizieren, so wurde eine Aufteilung von MAPK und CDK über die zweite Hauptkomponente erreicht. Dabei wurden Bereiche in der Ribose

Pocket sowie der phosphatbindenden Region mit der DRY-Sonde als entscheidend für die unterschiedlichen Interaktionsmöglichkeiten von Liganden lokalisiert (vgl. Abbildung 5 in [120]). Diese Bereiche ähneln den in Abbildung 2.22 auf Seite 69 dargestellten Regionen, die durch die Projektion der Loadings in der paarweisen PCA als wichtig für die Differenzierung von *HsGSK-3 β* und *PfGSK-3* ermittelt wurden. Offenbar ist die Diskriminierung von *HsGSK-3 β* und *PfGSK-3* über ähnliche Regionen möglich wie diejenige von MAPK und CDK. MAPK und CDK sind nahe verwandt (beide sind nach [50] in die Gruppe der CMGC einzuordnen), so dass die Diskriminierung der ebenfalls nahe verwandten *HsGSK-3 β* und *PfGSK-3* über eine ähnliche Region mindestens als interessante Parallele gelten kann. Diese Region wird im Allgemeinen eher als uninteressant hinsichtlich ihrer Bedeutung für Selektivität von Kinaseinhibitoren eingestuft [58, 59]. Offenbar scheint ihr in besonderen Fällen, wie in der oben zitierten Studie sowie in der vorliegenden Arbeit, eine Bedeutung zuzukommen, die sie als Fokus in weiteren Untersuchungen qualifiziert. Bezogen auf die in dieser Arbeit durchgeführten PCA-/CPCA-Berechnungen bestätigt sich also, dass losgelöst von den Schwierigkeiten bei der Berechnung eines statistischen Modelles für *alle* Homologie-Modell/Vorlage-Paare eine sinnvolle Analyse der einzelnen Paare möglich ist.

Bei dem Vergleich mit anderen Studien muss berücksichtigt werden, dass diesen in der Regel (so auch bei [120]) Röntgenkristallstrukturen zugrunde lagen. Bei der Verwendung von Homologie-Modellen hingegen wurde die hier beobachtete Problematik der Nähe von Modell und Vorlage bereits beschrieben [126]. Dass diese Problematik eine Anwendung auf Homologie-Modelle nicht ausschließt, konnten MYSHKIN *et al.* [125] zeigen. So gelang den Autoren bei ihrer Analyse von Eph Kinasen eine vergleichsweise klaren Auftrennung der Strukturen, die zum Teil Homologie-Modelle sind. Andererseits konstatieren sie, dass der angenommene Induced-Fit-Bindungsmechanismus ein Problem darstellt, dem die Methode nicht gerecht wird. Es ist möglich, dass ein solcher Mechanismus die Unterscheidung per CPCA im vorliegenden Fall verhindert.

5.2. Bindungsmodus von Thieno[2,3-*b*]pyridinen als Inhibitoren der *PfGSK-3*

5.2.1. Docking mit DoReCCoM

Durch die Anwendung des Arbeitsablaufes DoReCCoM konnten die Ergebnisse des Cross-dockings von Inhibitoren der *HsGSK-3 β* gegenüber denjenigen verbessert werden, die durch Anwendung eines einzelnen Dockingprogramms erzielt worden waren. Dies zeigt zum einen den Nutzen des angewandten Verfahrens, zum anderen unterstreicht es die Bedeutung der Ergebnisse der damit im zweiten Docking berechneten Bindungsmodi im Vergleich zu den heterogenen Ergebnissen des ersten Durchlaufes.

Die verschiedenen Stufen des Arbeitsablaufes sind in Abbildung 3.5 auf Seite 90 abgebildet. Ihre Bedeutung für das Endergebnis soll hier im Einzelnen betrachtet werden. Der erste Schritt des Dockings mit verschiedenen Programmen führt zu den in Tabelle 3.8 auf Seite 126 dargestellten Ergebnissen. Diese sind in Abbildung 5.3 als Radialplot der Kennzahlen aus Tabelle 3.8 dargestellt. Am weitesten außen sind jeweils die besten Ergebnisse aufgetragen (niedrige Platzierung bzw. kleiner rmsd, hohe Anzahl an Lösungen für Kennzahl 3 und 4). Dabei schneidet FRED in der Mehrzahl der Kriterien als bestes Programm ab. AutoDock bietet Lösungen, die am nächsten an der Kristallstruktur liegen. Außerdem weichen beim in AutoDock implementierten Scoring bereits 27 bestbewertete Lösungen weniger als 2 Å von der Kristallstruktur ab. FlexX liegt bis auf Kennzahl 4, bei der gemeinsam mit FRED ein kleiner Vorteil gegenüber AutoDock besteht, in der Mitte.

Bei der Beurteilung der Kennzahlen muss bedacht werden, dass die Gesamtheit der 108 Crossdockings betrachtet wird. Im Einzelfall bietet jedes Dockingprogramm Vorteile. So wird für die Verbindung DFN (**43**) von FRED nur in einem der Proteinkomplexe eine Lösung mit einem rmsd von weniger als 2 Å gefunden – in FlexX gelingt dies mit Ausnahme eines einzigen Komplexes bei allen. Auch bei HBM (**44**) gelingt es bei FRED und AutoDock nur mit einem der Proteine, eine kristallstrukturnahe Pose zu generieren. Mit FlexX hingegen ist dies bei 4 Komplexen der Fall. Es ist demnach sinnvoll, die Programme zu kombinieren, um ihre jeweiligen Stärken auszunutzen. Als Nachteil dieses Vorgehens ist die Vergrößerung der Lösungsmenge (bis zu 1427 Posen für eine Kombination von Ligand und Protein, vgl. S. 128) anzusehen – allerdings ein Nachteil, der durch das im dritten Schritt durchgeführte Clustering aufgewogen wird.

Das Rescoring mit DrugScoreX erbringt über alle Lösungen betrachtet Vorteile für alle Programme, wie aus Tabelle 3.8 auf Seite 126 und aus Abbildung 5.3 ersichtlich ist. Die einzige Kennzahl, die sich im Fall von AutoDock marginal verschlechtert, ist die Anzahl an Lösungen auf Platz 1 mit einem rmsd von weniger als 2 Å. Allerdings bleibt sie mit 26 Lösungen knapp unter den maximal erzielten 27, die bei den berechneten Kombinationen von Dockingprogramm und Scoringfunktion möglich sind.

Das Clustering der vereinigten Lösungen der Dockingprogramme in Kombination mit dem Rescoring ist ein wesentlicher Schritt innerhalb von DoReCCoM. Er führt zu einer Reduktion der Datenmenge auf eine überschaubare Lösungsmenge (zwischen 27 und 291 Cluster für eine Kombination von Ligand und Protein, vgl. S. 128) und trägt dadurch dazu bei, dass es machbar wird, die Ergebnisse für mehrere Proteine auch visuell auszuwerten und auf diese Weise die Flexibilität des Proteins mit in das Verfahren einzubeziehen. Bei der Auswertung bestünde auch die Möglichkeit, die Lösungen nicht nur aller Dockingprogramme, sondern auch diejenigen aller Proteine zu vereinigen und das Clustering für diese Gesamtmenge durchzuführen. Dann entfällt allerdings die Möglichkeit, im weiteren Verlauf des Arbeitsablaufes zu beobachten, ob eine Pose konsistent in unterschiedlichen Proteinen mit günstigen Scores berechnet wird. Eine solche Beobachtung kann als Argument für die betreffende Pose verstanden werden und sollte

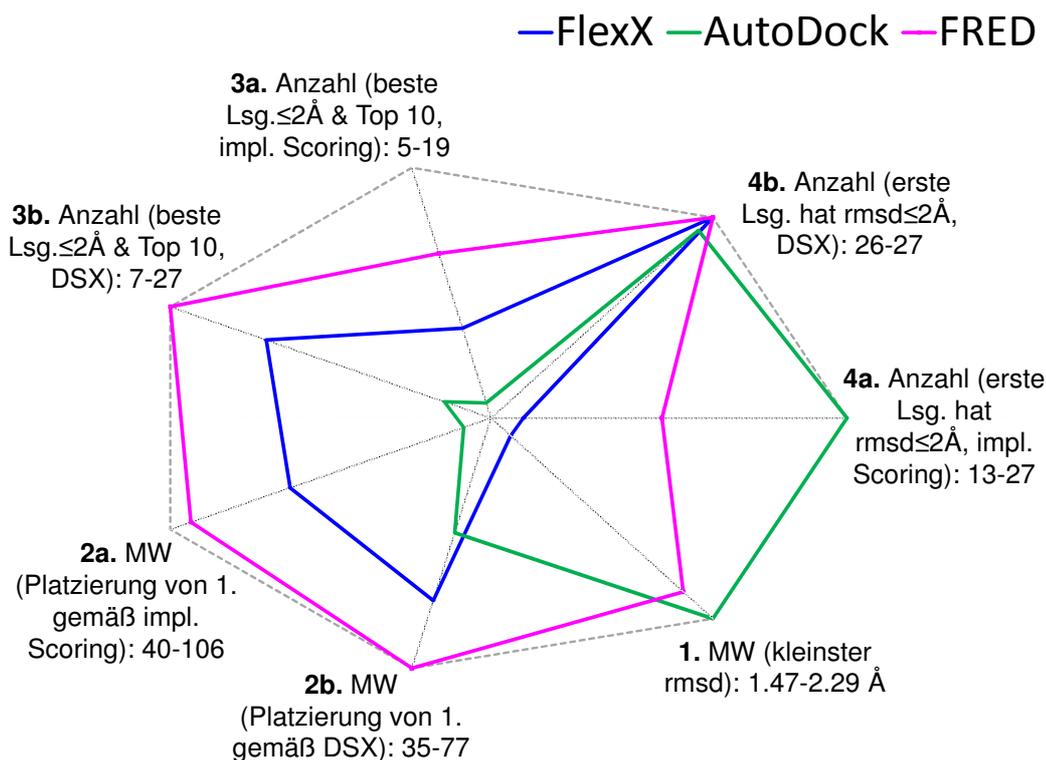


Abbildung 5.3. | Vergleich der eingesetzten Dockingprogramme (MW: Mittelwert; Lsg.: Lösung; impl.: implementiert; DSX: DrugScoreX). Die Kennzahlen 2a bis 4a beziehen sich auf Ergebnisse mit der implementierten Scoringfunktion, 2b bis 4b auf solche, die mit DrugScoreX erhalten wurden.

möglich bleiben, so dass das Verfahren auf die beschriebene Art konzipiert wurde.

Bei der Betrachtung der Ergebnisse der Komplexminimierung innerhalb DoReCCoM stellt sich die Frage, unter welchen Umständen dieser letzte Schritt im Allgemeinen angewendet werden sollte. Er ist der zeitaufwändigste und führt für die zur Evaluierung eingesetzten Verbindungen zu sehr unterschiedlichen Ergebnissen (vgl. Tab. 3.10 auf S. 129). Auffällig sind die Verschlechterungen für die Verbindung UN3 (**47**). Die Verbindung bietet bei näherer Betrachtung allerdings zwei Besonderheiten, die dieses schlechte Abschneiden erklären können. Einerseits bindet UN3 (**47**) zur Hinge Region nur über eine einzige Wasserstoffbrückenbindung. Diese wird zudem von einem Phenolether ausgebildet, der als schwacher Wasserstoffbrückenakzeptor eingeordnet werden kann [286].

Vor allem aber ist die in der Kristallstruktur gefundene Geometrie des Komplexes insofern besonders, als dass das Molekül in der Kristallstruktur in einer extrem kompakten Konformation vorliegt, in der der 3-Fluor-4-methoxyphenylsubstituent in die Bindetasche hinein gedreht ist (vgl. Abb. 5.4). Auf diese Weise steht er weniger gut zur Interaktion mit dem Protein bereit, wenn die Bindetasche in einer durchschnittlich weit geöffneten Konformation vorliegt. Allerdings ist in der Kristallstruktur 3F7Z (blau in Abb. 5.4) das Phe67 an der Spitze des Glycine-rich Loops weit in die Tasche hinein gedreht (vgl. Torsionswinkel in Tab. 2.3.4.2

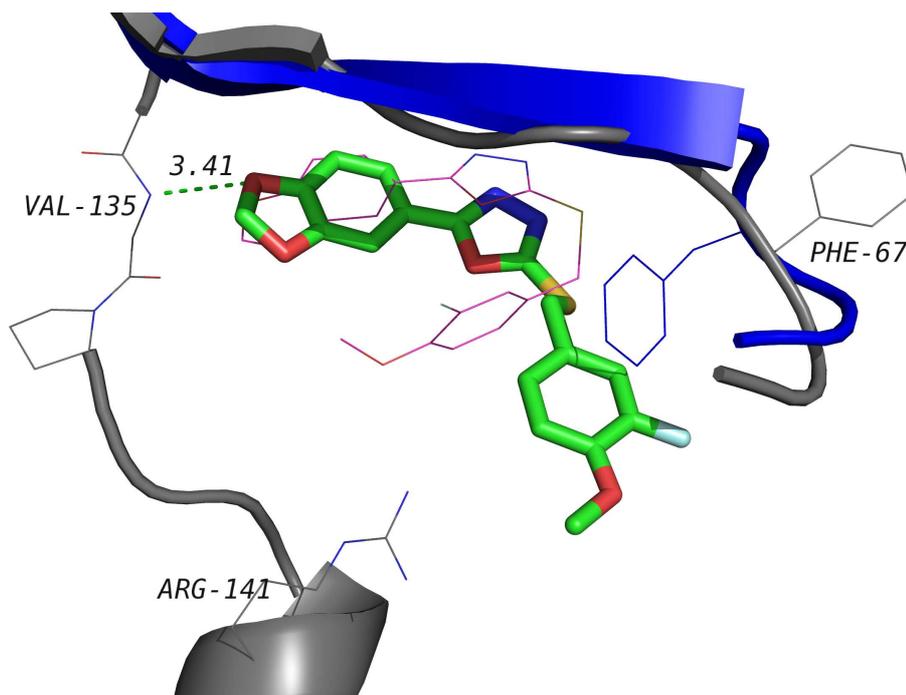


Abbildung 5.4. | UN3 (47) in der *HsGSK-3β*. Die in Grün abgebildete Dockinglösung wurde von FlexX in der Kristallstruktur 205K (grau) berechnet. Bei vergleichender Betrachtung der Lösungen in allen Proteinen liegt sie nach Rescoring mit DrugScoreX an Platz 29 und ist damit die einzige unter den ersten 50 Lösungen, in der der Benzodioxol-Grundkörper in der richtigen Ausrichtung vorliegt, so dass die charakteristische Wasserstoffbrückenbindung zum Val135 der Hinge Region ausgebildet werden kann. In Magenta ist die Lage von UN3 (47) in der Kristallstruktur dargestellt, in Blau ein Ausschnitt des Glycine-rich Loops von 3F7Z.

auf S. 34). Dadurch wird eine hydrophobe Wechselwirkung der Phenylkerne vom 3-Fluor-4-methoxyphenylsubstituenten und dem Phe67 möglich, die nur bei einer entsprechenden Konformation des Phe67 infrage kommt.

Lässt man die Ergebnisse von UN3 (47) aus den genannten Gründen aus der vergleichenden Betrachtung der Kennzahlen heraus, können die Veränderungen durch Komplexminimierung im Rahmen von DoReCCoM mit den auf Seite 130 dargestellten Ergebnissen als leichte Verbesserung gewertet werden. Insbesondere die durchschnittliche Platzierung der ersten guten Lösung auf Platz 9 stellt eine Verbesserung gegenüber der Bewertung nicht minimierter Komplexe dar, in der die entsprechende Lösung im Mittel auf Platz 12 gesetzt wurde. Bei den anderen Kennzahlen ist keine deutliche Veränderung zu verzeichnen.

DRG (31) profitiert von der Komplexminimierung. Obwohl das in der Kristallstruktur wichtige kokristallisierte Wasser nicht explizit in das Docking einbezogen wurde, wird nach der Minimierung eine verbesserte Platzierung erreicht. Betrachtet man die 3 Lösungen unter den ersten 10, deren rmsd zur Kristallstruktur unter 2 Å liegt (vgl. Tab. 3.10 auf S. 129), so ist deren Veränderung durch die Komplexminimierung gering. Der rmsd der Liganden vor der Minimierung zu der Struktur danach beträgt nur 0.47 bis 0.79 Å. Das Scoring mit DrugScoreX profitiert dennoch, so dass die betreffenden Posen unter den ersten 10 Lösungen bewertet werden. Ein

weiteres interessantes Detail ist, dass jede der Lösungen in einem anderen Protein gefunden wird (1Q41, 1Q4L und kubra). Durch die Minimierung wird also tatsächlich die Flexibilität des Proteins so gut in das Verfahren einbezogen, dass auch in den nicht auf den Liganden präformierten Bindetaschen von 1Q41 und 1Q4L eine kristallstrukturnahe Lösung auf einer Platzierung unter den ersten 10 möglich wird.

Eine in diesem Zusammenhang zu erwähnende Beobachtung ist, dass aromatische Systeme nach der Minimierung teilweise leichte Abweichungen von der (Ko-)Planarität zeigten. Als Beispiel sei die in Abbildung 5.6(a) dargestellte Pose nach Komplexminimierung von **102** genannt. Offensichtlich geht bei der Minimierung die Energie des Liganden nur zu einem kleineren Teil in die Gesamtenergie des Komplexes ein, so dass das lokale Minimum des Komplexes insgesamt zu einer nicht optimalen Geometrie des Liganden führt. Die Abweichungen waren jedoch so gering, dass sie nicht als Argument gegen die Interpretation der berechneten Posen gewertet wurde.

Berücksichtigt man den vergleichsweise großen Zeitaufwand der Minimierung im Verhältnis zum dargestellten Nutzen, ist die Anwendung von DoReCCoM inklusive des letzten Schrittes kleinen Datensätzen vorbehalten, bei denen für verhältnismäßig wenige Verbindungen verhältnismäßig viele unterschiedliche Posen betrachtet werden. In einem solchen Szenario kann DoReCCoM die im Ergebnisteil dargestellten Vorteile gegenüber der Anwendung einzelner Dockingprogramme bieten.

5.2.2. Argumente für mögliche Bindungsmodi

Ein zentraler Punkt der Zielsetzung dieser Arbeit war es, den Bindungsmodus von Thieno[2,3-*b*]pyridinen als Inhibitoren der *Pf*GSK-3 zu untersuchen. Es soll an dieser Stelle daher die Entwicklung der Argumente für verschiedene Modi im Verlauf der Arbeit aufgezeigt werden. Als Ausgangspunkt können dabei die in Abschnitt 3.2.1.1 geschilderten bekannten Struktur-Wirkungsbeziehungen und der aus dem humanen Enzym bekannte Bindungsmodus A gelten. Zudem legen die unterschiedlichen Charakteristika der Klassen 1 und 2 verschiedene Modi nahe. In Abbildung 5.5 sind **40** und **129** als Repräsentanten der Klassen dargestellt. Die klare Abgrenzung der Aktivität ist kein allgemein auf die Klasse 2 übertragbarer Unterschied, es finden sich aktive Substanzen sowohl mit 2-Cyanosubstitution (z. B. **138**, analog zu Klasse 1a) als auch 2-Aroylsubstitution (z. B. **137**, analog zu Klasse 1b).

Die nahe Verwandtschaft der Strukturen schlägt sich in ähnlichen Oberflächeneigenschaften im Bereich der 4-Aryl- und der 2-Aroylsubstitution nieder. Unterschiede zeigen sich vor allem im Bereich der Positionen 5 und 6 (negative Partialladung der Cyanosubstitution, lipophileres Potential des anellierten gesättigten Ringes in Verbindungen der Klasse 2 gegenüber der 6-Amino-5-cyano-Substitution von Verbindungen der Klasse 1). Interessant ist zudem, dass sich auch das elektrostatische Potential im Bereich des 3-Aminosubstituenten verändert. Hier gibt es bei der Klasse 1 einen positiven Bereich (in der Abbildung rot dargestellt), der bei **129** fehlt.

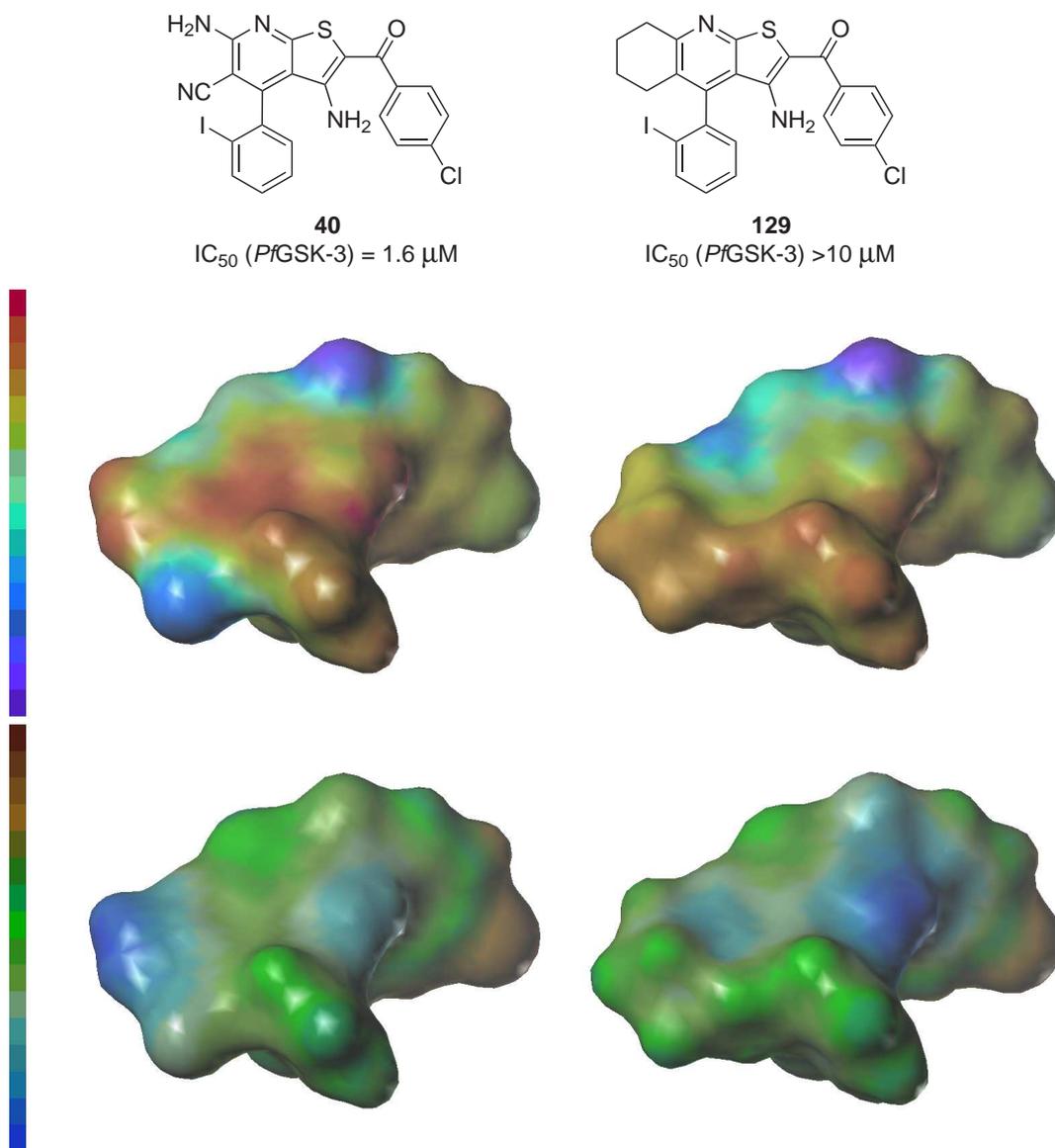


Abbildung 5.5. | Elektrostatisches Potential und Lipophile von Verbindungen der Klassen 1 und 2. Die Verbindungen **40** und **129** als Repräsentanten der Klassen 1 und 2 verdeutlichen exemplarisch die Unterschiede hinsichtlich des elektrostatischen und lipophilen Potentials. In der oberen Reihe ist das elektrostatische Potential dargestellt, rote Bereiche markieren positive Partiaalladungen der Atome, blaue Bereiche negative. Die untere Reihe zeigt das lipophile Potential, die Farbskala reicht hier von braun (lipophil) bis blau (hydrophil). Die dargestellten Oberflächen sowie das lipophile und elektrostatische Potential wurden mit dem MOPAC-Modul in Syblyl [183] berechnet (Fast Connolly-Oberfläche mit dot density = 6 und probe radius = 1.4 Å; Gasteiger-Hückel-Ladungen; Lipophilie nach [287]).

Auch die Ladungsverteilung im Aromaten verändert sich: So wird durch den Zug der Cyano-substitution der Aromat weniger elektronenreich. Die Folge ist das weniger negative Potential im Bereich des Stickstoffes des Thienopyridins in der Klasse 1. Beim lipophilen Potential unterscheidet sich erwartungsgemäß die Region der 6-Amino-5-cyanosubstitution der Klasse 1 vom korrespondierenden Bereich der Klasse 2. Dabei ist durch die Farbkodierung zu erkennen, dass bereits im Bereich des 5-Cyanosubstituenten kein so extremer Unterschied mehr zwischen den

Liganden besteht wie im benachbarten Bereich des 6-Aminosubstituenten. Es unterscheiden sich hingegen die lipophilen Potentiale in der Region der 3-Aminosubstitution.

Die Verschiedenheit der Verbindungen beider Klassen führte zu den unterschiedlichen Ergebnissen des orientierenden Dockings und zu deren Interpretation. Es wurden für die beiden Klassen voneinander abweichende Ergebnisse gefunden, die einen gemeinsamen Bindungsmodus für die Klassen 1 und 2 nicht möglich erscheinen ließen. Zudem legten die Ergebnisse unterschiedliche Modi für die Klassen 1a und 1b nahe (vgl. S. 112).

Ein Argument für diese Unterscheidung waren die Ergebnisse des Kapitels 2. Hier wurden die Unterschiede hinsichtlich der Sequenz von *Hs*GSK-3 β und *Pf*GSK-3 geschildert, die insbesondere im Bereich der lösungsmittelzugewandten Region Ansätze zur Erklärung einer Selektivität bieten. Im Modus B ragt der 2-Aroylrest der Moleküle der Klasse 1b in diese Region. Dass die selektiven Moleküle des Datensatzes damit eine Region besetzen, die sich in *Hs*GSK-3 β und *Pf*GSK-3 unterscheidet, machte den Modus B für diese Verbindungen plausibel, während andererseits für die nicht selektiven Verbindungen der Klasse 1a ein Bindungsmodus analog zur Situation im humanen Enzym naheliegt und vom Docking bestätigt wurde. Gegen den Modus B spricht allerdings die ungewöhnliche Geometrie der Wasserstoffbrückenbindungen der Substituenten an Position 5 und 6 (vgl. Abb. 3.10 auf S. 104 bzw. Abb. 5.6(b) auf S. 195).

Unterschiede von *Hs*GSK-3 β und *Pf*GSK-3 hinsichtlich ihrer MIFs waren für die Argumentation an diesem Punkt weniger hilfreich. So betreffen die Unterschiede des N1+-Feldes einen Bereich, der in keinem der gefundenen potentiellen Bindungsmodi von den Verbindungen besetzt wird, das Gleiche gilt für die durch den unterschiedlichen Gatekeeper verursachten Differenzen. Andersherum können in Modus A und B das I bzw. C3-Feld am Grund der Tasche von einem Substituenten an Position 2 des 4-Arylringes eingenommen werden.

Ein Zusammenhang zwischen den MIFs und den Modi A und B ist hinsichtlich der Verschiebung der Felder verschiedener Wasserstoffbrückenakzeptorsonden an der Hinge Region entlang in Richtung des Gatekeepers möglich. Im Modus A wird von keinem Akzeptor des Moleküls eine direkte Wasserstoffbrückenbindung ausgebildet, der entsprechende Kontakt wird über Wasser vermittelt (vgl. Abb. 1.17 auf S. 20). Da das Feld bei einigen Kristallstrukturen weiter in die Tasche hinein verschoben liegt (vgl. Tab. 2.8 auf S. 65), wäre hier eine direkte Wasserstoffbrücke des Liganden in der *Pf*GSK-3 in selektiven Molekülen schlüssig. Einschränkend muss allerdings bedacht werden, dass diese Unterschiede weniger konsistent in den verschiedenen Kristallstrukturen zu beobachten waren und sich nicht in der paarweisen PCA bestätigten.

Die biologischen Daten der im Rahmen dieser Arbeit dargestellten Verbindungen **101-104** erbrachten keine eindeutigen Hinweise auf den Bindungsmodus. Eine Interpretation wurde zudem dadurch erschwert, dass die Daten in sich widersprüchliche Tendenzen aufweisen (vgl. Abschnitt 3.4.2 auf S. 118). Allerdings können sie als selektive (bzw. präferentielle im Falle von **103**) Inhibitoren der *Pf*GSK-3 angesehen werden und bestätigen so den grundsätzlichen Befund, dass die 2-Aroylsubstitution von Verbindungen der Klasse 1b zu Selektivität führt.

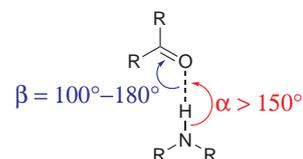
Bei dem in Abschnitt 3.5.2 beschriebenen zweiten Docking lag der wesentliche Unterschied zum ersten Durchlauf im angewandten Verfahren (DoReCCoM) sowie im erweiterten Datensatz. Die Ergebnisse stellten die Unterscheidung der Modi für die Klassen 1a und 1b infrage. Es findet sich zwar wiederum der Modus B für Verbindungen der Klasse 1b, allerdings stehen die Modi A und B hinsichtlich der Bewertung innerhalb DoReCCoM für die Klasse 1a und 1b nahezu gleichwertig nebeneinander. Die in den folgenden Absätzen dargelegten Punkte sprechen für die Annahme von Modus A für die Klassen 1a und 1b.

Treten eng verwandte Verbindungen in unterschiedlichen Bindungsmodi auf, so werden diese als alternative oder multiple Bindungsmodi bezeichnet [288]. Dieser Sachverhalt wurde anhand einer Reihe von Beispielen als Ursache für Ausreißer innerhalb von QSAR-Studien beschrieben [289]. Er widerspricht der grundlegenden Annahme, dass ähnliche Strukturen ähnliche biologisch aktive Konformationen und Positionen in der Bindetasche einnehmen – eine Voraussetzung für jede (3D-) QSAR-Untersuchung (vgl. Abschnitt 4.1.2.2). Auch losgelöst von der Perspektive auf QSAR-Untersuchungen nennen z. B. TEAGUE *et al.* einige Beispiele, in denen nahe verwandte Verbindungen auf unterschiedliche Art und Weise binden [290]. Allerdings können diese als Ausnahme betrachtet werden, bedenkt man die große Mehrzahl an Untersuchungen, bei denen in kongenerischen Serien von Verbindungen äquivalente Bindungsmodi beobachtet wurden. Da durch die Anwendung von DoReCCoM praktisch kein Unterschied zwischen den Modi festgestellt werden konnte, wurde nicht von multiplen Bindungsmodi als Ausnahme von der Regel ausgegangen. Es wurde vielmehr der Bindungsmodus A für alle Verbindungen der Klasse 1 als gemeinsamer Bindungsmodus angenommen, da er sich mit ähnlicher Bewertung als Ergebnis des Dockings herauskristallisierte. Er ist in Abbildung 5.6(a) am Beispiel der Verbindung **102** dargestellt, deren Synthese in Kapitel 3.4 beschrieben ist.

Die Winkel bzw. Entfernungen der Wasserstoffbrückenbindungen sind mit 126° bzw. 3.4 \AA (Carbonyl des Ile160 zum 6-Aminosubstituenten von **102**) und 114° bzw. 3.1 \AA (Stickstoff des Lys108 zum Carbonylsauerstoff von **102**) günstig.¹ Das Oberflächenprofil von Ligand und Protein ist in der Abbildung 5.6(a) als komplementär zu erkennen. Ein weiteres Argument für diesen Bindungsmodus ist seine Nähe zur Geometrie der Kristallstruktur von **31** in der *Hs*GSK-3 β . Allerdings gibt es einen interessanten Unterschied zwischen den Posen im Modus A und der genannten Kristallstruktur: Der 2-Chlorsubstituent in **31** in der *Hs*GSK-3 β zeigt nach oben unter den Glycine-rich Loop. In der *Pf*GSK-3 findet sich der Modus A jedoch in den Dockings mehrheitlich mit nach unten ausgerichteten 2-Halogensubstituenten. Da zudem die Berechnungen der MIFs für diese Anordnung sprechen, wird sie wie abgebildet angenommen.

Diese Region der Bindetasche kann auch in dem vom orientierenden Docking favorisierten

¹Es ist der Winkel β gemeint, wenn man die optimalen Winkel für Wasserstoffbrückenbindungen wie allgemein üblich für α und β wie abgebildet annimmt [292]. Sauerstoff, Wasserstoff und Stickstoff liegen in den betrachteten Fällen nahezu in einer Linie, $\alpha \approx 180$.



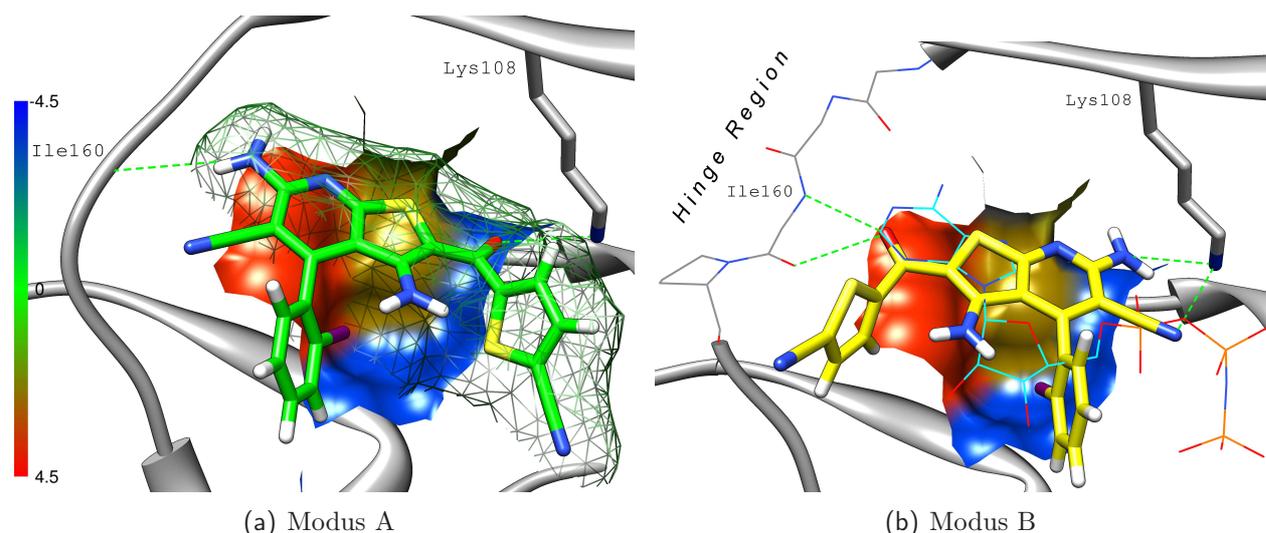


Abbildung 5.6. | **102** in den Bindungsmodi A und B mit pfgsk_1j1b_128. Ein gemeinsames Merkmal der beiden abgebildeten Modi ist die Ausrichtung des 2-Iodsubstituenten am 4-Arylkern. Er zeigt zum Grund der Tasche in eine Auswölbung hinein, in der das für die *PfGSK-3* stark ausgeprägte I- bzw C3-Feld liegt. Diese Orientierung kann demnach zur Selektivität der Verbindungen beitragen – im humanen Enzym ist der Substituent zum Glycine-rich Loop hin nach oben ausgerichtet. Als weitere Gemeinsamkeit ist in beiden Modi hinter dem Liganden ausreichend Platz für ein Wassermolekül, über das ein indirekter Kontakt zum Protein ausgebildet werden kann (vgl. z. B. Abb. 3.8 auf S. 102). Die in (a) als Gitter in Grün dargestellte Oberfläche des Liganden im Bindungsmodus A ist sterisch kompatibel zu derjenigen des Proteins. In Teil (b) der Abbildung ist **102** in Modus B dargestellt, zudem ist in Cyan die Position des ATP aus 1J1B in die Bindetasche projiziert. Die Oberfläche des Proteins am Grund der Tasche wurde jeweils farbkodiert nach Lipophilie (Darstellung in UCSF Chimera [236], Lipophilie farbkodiert nach kdHydrophobicity [291]). Es ist zu erkennen, dass die Auswölbung hinsichtlich ihrer Lipophilie den Übergang von der eher hydrophilen (blauen) Region im weiter außen gelegenen Bereich der Bindetasche zu hydrophoben (roten) Bereichen weiter innerhalb in Richtung der Hinge Region darstellt. Die Verhältnisse entsprechen sich diesbezüglich in *HsGSK-3 β* und *PfGSK-3*, da die Aminosäuren sich nicht unterscheiden, die am Grund der Tasche die in der Abbildung zu erkennende Ausbuchtung bilden.

Modus B entsprechend besetzt werden. In Abbildung 5.6(b) sind allerdings die Nachteile dieses Modus verdeutlicht, die letztlich mit zur Bevorzugung des Modus A führten. So ist die Anordnung der Wasserstoffbrücke zum Lys108 ungünstig: Es wird ein Dreieck zwischen den Stickstoffen des Lys108 und den beiden der Amino- bzw. Cyanosubstitution des Liganden ausgebildet, so dass sich Donor und Donor bzw. Akzeptor und Donor im gleichen Abstand von etwa 3.7 Å befinden. Eine weitere Ungereimtheit tritt auf, wenn man die Position des ATP in die Bindetasche projiziert. Das C2 des Puringrundgerüsts, das in der Bindung des ATP die Rolle eines Wasserstoffbrückendonors einnimmt (vgl. S. 14), liegt genau über dem Carbonyl des Liganden, der in Modus B als Wasserstoffbrückenakzeptor einen entscheidenden Kontakt zur Hinge Region ausbildet. Diese Anordnung ist von den Winkeln her zwar möglich, die Entfernungen zum Amin bzw. Carbonyl des Ile160 zeigen jedoch mit 3.2 bzw. 3.7 Å den Vorteil der erstgenannten Wasserstoffbrückenbindung an – ein weiteres Argument gegen Modus B.

Für die Klasse 2 ergab das zweite Docking eine deutliche Überlegenheit des Bindungsmodus

H gegenüber dem Modus A (vgl. Abb. 3.25 auf S. 134). Es wurden allerdings ebenso Posen im Modus A identifiziert, so dass im Gegensatz zur Situation nach dem ersten, orientierenden Docking für alle Verbindungen ein einziger Bindungsmodus grundsätzlich möglich schien. Bei Vorliegen des Modus A für Verbindungen der Klasse 2 ist zu bedenken, dass kein direkter Kontakt zur Hinge Region mehr vorliegen würde. Von den charakteristischen Wasserstoffbrücken der Typ-I-Inhibitoren bliebe nur noch die über Wasser vermittelte Wasserstoffbrücke. In diesem Zusammenhang geben die auf die Oberfläche der Moleküle projizierten elektrostatischen Potentiale von Abbildung 5.5 eine Zusatzinformation. Der Stickstoff im Thienopyridin, von dem aus die Wasserstoffbrückenbindung ausgebildet wird, ist in Verbindungen der Klasse 2 stärker negativ geladen, da dem aromatischen System weniger Elektronen von der Cyanogruppe entzogen werden. Möglicherweise führt diese stärkere Partialladung zu einer verstärkten Wasserstoffbrückenbindung zu einem Wassermolekül, so dass dieser indirekte Kontakt in der Lage ist, den ansonsten ausgebildeten direkten zu ersetzen. Auch wenn der Modus H DoReCCoM folgend höher bewertet wird als der Modus A, sollte die Plausibilität alternativer Bindungsmodi für Verbindungen der beiden Klassen durch 3D-QSAR-Untersuchungen überprüft werden, um den Bindungsmodus auf diese Weise durch ein zum Docking alternatives Verfahren zu überprüfen.

5.2.3. 3D-QSAR-Untersuchungen zur Interpretation möglicher Bindungsmodi

Ausgehend von Überlagerungen, in denen die Bindungsmodi A (Verbindungen der Klasse 1) und H (Verbindungen der Klasse 2) gleichzeitig repräsentiert wurden, konnten keine interpretierbaren Modelle erstellt werden (vgl. Abschnitt 4.3.1.1 auf S. 154). Andererseits ergaben Überlagerungen aller Substanzen auf eine einzige Zielstruktur CoMSIA-Modelle mit akzeptablen Kennzahlen. Dieser Befund stützt die Annahme eines einzigen gemeinsamen Bindungsmodus für alle Verbindungen. Er steht im Gegensatz zur Priorisierung der Modi gemäß DoReCCoM, die den Modus H für Verbindungen der Klasse 2 als überlegen einordnet. Dieser Widerspruch wurde allerdings als weniger gravierend eingestuft, da die alternative Hypothese (alle Klassen binden in einem Modus) mit den Dockinglösungen vereinbar war. Zudem entspricht diese Annahme dem grundlegenden Konzept, dass ähnliche Analoga den gleichen Bindungsmodus aufweisen.

Diejenigen Überlagerungen, aus denen interpretierbare 3D-QSAR-Modelle abgeleitet werden konnten, basieren auf einem der beiden Modi A oder B. Obwohl Modus B in den Dockingergebnissen nicht für alle Verbindungen beobachtet wurde, sollte er dennoch als mögliche Alternative untersucht werden. Daraus konnten folgende Hinweise hinsichtlich des Bindungsmodus gewonnen werden.

- Für unterschiedliche Felderkombinationen zeigten Modelle basierend auf Modus A gegenüber denen basierend auf Modus B bessere statistische Kennzahlen. Zudem musste

die Validität des Modelles AcEl basierend auf Modus B infrage gestellt werden (vgl. Progressive Scrambling S. 162 bzw. 172). Eine mögliche Erklärung hierfür ist, dass die erstgenannte Überlagerung eher die Konformationen biologisch aktiver Liganden wiedergibt als die zweite. Losgelöst von der Tatsache, dass die zweite Überlagerung auf Modus B beruht, kann generell konstatiert werden, dass mit Modus A QSAR-Modelle berechnet werden konnten, die anderen Überlagerungen überlegen waren – ein Argument für den Bindungsmodus A.

- Die Interpretation der Konturdiagramme unter Bezugnahme auf den Rezeptor gelang insbesondere bei den Modellen basierend auf Modus A (vgl. S. 163 ff.), während es hinsichtlich der Modelle basierend auf Modus B (vgl. S. 171 ff.) im ohnehin anzuzweifelnden Modell AcEl im Bezug auf den Rezeptor widersprüchliche Feldbeiträge gab. Das Modell AcElSt ließ zwar Erklärungsansätze zu, die allerdings nicht zuletzt auf Parallelen zu Modus A hinsichtlich der Konformation (bzw. Torsionswinkel, vgl. Abb. 4.5 auf S. 156) zurückzuführen sind, die zu vergleichbaren Bezügen hinsichtlich der SAR führen.

Zusätzlich wurden durch die QSAR-Berechnungen Hinweise auf Ausreißer bei den biologischen Daten erhalten. Im Ergebnisteil wurden grundsätzlich keine Verbindungen aus den Berechnungen ausgeschlossen, auch wenn sie vom entsprechenden Modell schlecht beschrieben wurden. Es stand nicht die Entwicklung von prädiktiven Modellen mit möglichst guten Kennzahlen im Mittelpunkt, sondern die Interpretation der Modelle zum Vergleich der zugrunde liegenden Bindungsmodi. Deshalb wurden stattdessen Abweichungen von mehr als 0.5 pIC₅₀-Einheiten markiert, um schlecht vorhergesagte Moleküle zu identifizieren und nach Erklärungen für diese Abweichung zu suchen. CUNHA *et al.* betrachten Verbindungen als Ausreißer, die mehr als das doppelte der Standardabweichung s [293] vom tatsächlichen Wert abweichen. Wendet man diese Grenze auf die Verbindungen an, so stellen sich die drei Verbindungen **68**, **103** und **121** als Ausreißer dar. Die Einstufung von **103** als Ausreißer würde Widersprüchlichkeiten hinsichtlich der Struktur-Wirkungsbeziehungen erklären (vgl. S. 118). Die Einstufung von **68** als Ausreißer würde erklären, warum die Verbindung trotz ihrer 3,4-Dichlorbenzoylsubstitution als schwacher Inhibitor bestimmt wird – **78** bzw. **75** als Verbindungen mit dem gleichen Substitutionsmuster sind um den Faktor 26 bzw. 18 aktiver. Die Abweichung von **121** überschreitet die Grenze von $2s$ nicht im auf Modus A beruhenden Modell AcDoSt. Das Modell beschreibt den Datensatz also besser als die Alternativen. Da dieses Modell zudem hinsichtlich der Validierung mittels Progressive Scrambling den anderen gegenüber als klar überlegen gelten kann, wird die Kombination von AcDoSt mit der Überlagerung aller Verbindungen basierend auf Modus A als bestes untersuchtes Modell eingeschätzt.

5.2.4. Die Quartärstruktur als Erklärungsansatz für Selektivität

In Kristallstrukturen liegt die *HsGSK-3 β* oft als Dimer vor, so sind 15 der 19 in Tabelle 2.2 auf Seite 35 aufgeführten Strukturen Dimere. Die genaue Funktion der Dimere im humanen Organismus ist nicht vollständig verstanden, es wird jedoch davon ausgegangen, dass es sich nicht um ein experimentell verursachtes Artefakt handelt [109].

Es ist nicht bekannt, ob die *PfGSK-3* in Analogie zum humanen Enzym Dimere ausbilden kann oder ob das Enzym als Monomer vorliegt. Binden nun Verbindungen der Klasse 1b im Modus A, so zeigt der für Selektivität wichtige 2-Aroylsubstituent in Richtung der zweiten Kette

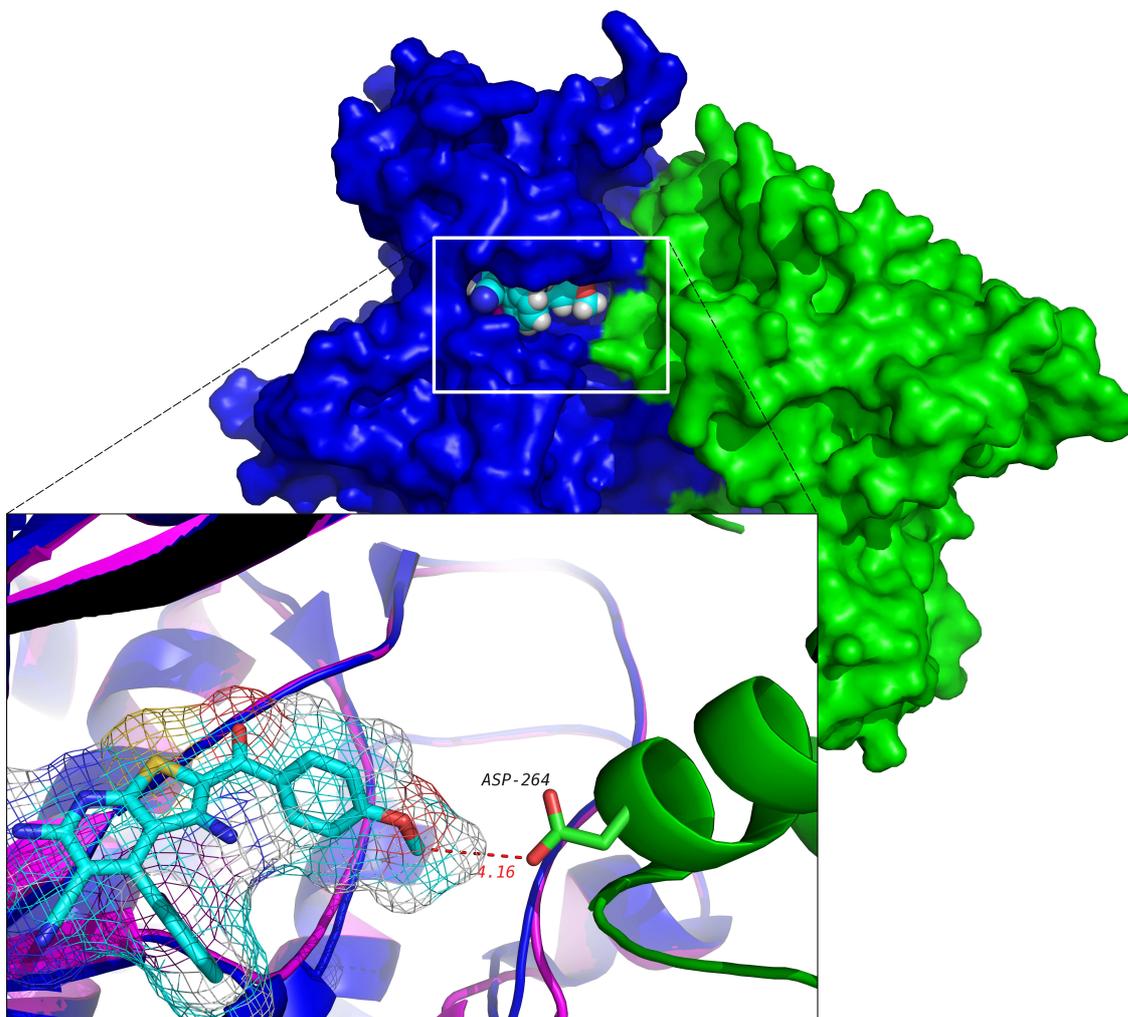


Abbildung 5.7. | Mögliche Erklärung für die Selektivität der Verbindungen der Klasse 1b durch die Quartärstruktur von *HsGSK-3 β* und *PfGSK-3*. Die Abbildung zeigt das Dimer der *HsGSK-3 β* (PDB-Eintrag 1Q5K, Kette A blau, B grün). Bindet ein Inhibitor der Klasse 1b wie z. B. die abgebildete Verbindung **69** (in Cyan: Docking-Lösung gemäß Bindungsmodus A; Protein: pfgsk_2o5k_160, magenta in vergrößertem Ausschnitt), so würde der Benzoylrest in Richtung der zweiten Kette des Proteins ragen. Dies hätte einen unvorteilhaften, zu engen Kontakt des Asp264 der B-Kette zur Folge, was die Bindung im Falle eines Dimers möglicherweise erschwert. Liegt das Enzym hingegen als Monomer vor, kann der Ligand binden.

des Dimers im humanen Enzym (vgl. Abb. 5.7), was eine Bindung behindern oder erschweren könnte. Möglicherweise können die Verbindungen nicht an das humane Enzym binden, wohl aber an das plasmodiale, wenn letzteres im Gegensatz zum erstgenannten als Monomer vorliegen würde. Dieser Erklärungsansatz ermöglicht die sinnvolle Kombination der beobachteten Struktur-Wirkungsbeziehungen mit Bindungsmodus A und vorliegenden strukturellen Daten und kann daher als weiterer Hinweis auf das Vorliegen von Modus A verstanden werden.

ALS zentrales Ergebnis dieser Arbeit wird für Thieno[2,3-*b*]pyridine der Klassen 1 und 2 in der *Pf*GSK-3 der gemeinsame Bindungsmodus A postuliert, wie er in Abbildung 5.6(a) dargestellt ist. Ein entscheidender Fortschritt in der Entwicklung potenter und selektiver Inhibitoren der *Pf*GSK-3 wäre die Aufklärung der dreidimensionalen Struktur des Enzyms im Komplex mit einem Inhibitor – diesem ausstehenden experimentellen Ergebnis bleibt die endgültige Beantwortung der Frage nach dem Bindungsmodus vorbehalten.

5.3. Ausblick

Ausgehend von den Ergebnissen dieser Arbeit ergeben sich einerseits Ansätze zur Optimierung der eingesetzten Verfahren, andererseits entwickelten sich aus den Ergebnissen der Untersuchungen neue Fragen, die sich als Gegenstand von Folgearbeiten anbieten.

5.3.1. Optimierung eingesetzter Verfahren

Die offenbar problematische Nähe der Homologie-Modell/Vorlage-Paare legt Veränderungen im Prozess des Homology Modellings nahe. Hier wären mögliche Alternativen zu dem gewählten Verfahren beispielsweise, durch die Kombination mehrerer verschiedener Vorlagen zu einem einzigen Modell die enge Verwandtschaft von Homologie-Modell und Vorlage zu lösen. Ein weiterer Ansatz, um die Besonderheiten der Homologie-Modelle gegenüber den Vorlagen herauszuarbeiten, bestünde in der Durchführung umfangreicherer MD-Simulationen. Diese könnten Aufschluss geben über Unterschiede zwischen Vorlage und Modell, die bei Betrachtung statischer Strukturen nicht zu erkennen sind.

Im Rahmen des Arbeitsablaufes DoReCCoM besteht prinzipiell die Möglichkeit, jeden der Schritte durch alternative Programme auszutauschen. Da das Dockingproblem mit den eingesetzten Programmen als gelöst bezeichnet werden kann, betrifft dies insbesondere den Schritt des Rescorings. Da dieser Bereich Gegenstand aktueller Forschung ist, eröffnen sich durch verbesserte Scoringfunktionen auch für DoReCCoM neue Möglichkeiten.

Hinsichtlich des Clusterings wäre die Variante denkbar, alle Lösungen verschiedener Proteine gemeinsam als Ausgangspunkt zu verwenden. Dadurch würde eine weitere Reduktion der Datenmenge gelingen, allerdings wäre die Beobachtung, ob eine Pose für mehrere Proteine als günstig bewertet wird, nicht mehr möglich.

Der größte Spielraum innerhalb des Verfahrens ist hinsichtlich einer Verbesserung der Komplexminimierung zu suchen. Hier wären insbesondere Untersuchungen interessant, in denen die Funktion zur Ermittlung der potentiellen Energie im Rahmen der Minimierung auf die im Rescoring eingesetzte Scoringfunktion angepasst wird. Außerdem könnten durch Anpassungen des Kraftfeldes verbesserte Ergebnisse erzielt werden. Eine weitere denkbare Variante wäre, den Liganden entweder partiell (z. B. aromatische Systeme) oder sogar ganz zu fixieren – es konnte schließlich gezeigt werden, dass eine kristallstrukturnahe Lösung existiert und nur gefunden werden muss.

5.3.2. Weiterführende Untersuchungen

Während in der vorliegenden Arbeit der Fokus der Untersuchungen auf dem Datensatz der Thieno[2,3-*b*]pyridine und auf direkt von diesen abgeleiteten Verbindungen bzw. deren Modifikationen lag, bieten sich die im ersten Teil erstellten Homologie-Modelle als Ausgangspunkt

eines virtuellen Screenings an, um neuartige Inhibitoren der *Pf*GSK-3 aufzufinden. Diese könnten sich durch einen abweichenden Grundkörper auszeichnen oder sogar als Kinaseinhibitoren vom Typ 1^{1/2} nach [73] bzw. Typ III in Bereichen binden, die an die ATP-Bindetasche angrenzen (vgl. S. 15 ff.).

Die in Abschnitt 3.3.3 erarbeiteten Synthesevorschläge wurden im Rahmen dieser Arbeit nur zu einem kleinen Teil umgesetzt, da der Schwerpunkt nicht in der präparativen Arbeit lag. Da die vorgestellten Modifikationen weitere Aufschlüsse über die Struktur-Wirkungsbeziehungen eröffnen würden und zudem basierend auf den Dockingergebnissen aktive Verbindungen zu erwarten sind, wäre die Umsetzung eine aussichtsreiche Arbeit.

In Analogie zum genannten Abschnitt 3.3.3 könnten zudem basierend auf der Annahme des Bindungsmodus A für Verbindungen beider Klassen Synthesen entworfen werden, die in Kapitel 2 dargestellte Unterschiede zwischen *Hs*GSK-3 β und *Pf*GSK-3 nutzen. Insbesondere die Unterschiede im Bereich des Gatekeepers geben Anlass, ein Design von Modifikationen zu betreiben, die beispielsweise mit protoniertem Amin den Bereich des sich unterscheidenden N1+-Feldes adressieren.

Weitere Untersuchungen zum Bindungsmodus durch MM-PBSA bzw. MM-GBSA-Berechnungen wären reizvoll, einerseits um ein genaueres Bild hinsichtlich der freien Bindungsenergie zu erhalten, andererseits weil die dafür durchzuführenden MD-Simulationen von Protein-Ligand-Komplexen zum Verständnis der Interaktionen beitragen könnten.

In diesem Zusammenhang ist insbesondere die Rolle von konstitutionell gebundenem Wasser in der ATP-Bindetasche zu sehen. Die ungewöhnliche Interaktion der Thienopyridine als Typ-I-Kinaseinhibitoren mit nur einem einzigen Kontakt zur Hinge Region könnte durch die MD-Simulation näher untersucht werden. Bei vergleichender Betrachtung von Komplexen mit *Hs*GSK-3 β und *Pf*GSK-3 sind Einsichten zur molekularen Grundlage von Selektivität möglich.

Besonders hilfreich im weiteren Design wäre die Aufklärung der dreidimensionalen Struktur der *Pf*GSK-3 – ein solches experimentelles Ergebnis würde nicht nur die Frage nach dem Bindungsmodus abschließend klären, sondern darüber hinaus einen wichtigen Anstoß für alle folgenden strukturbasierten Untersuchungen geben.

6. Zusammenfassung

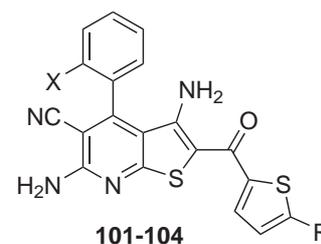
In der vorliegenden Dissertation wurde mit Methoden des Molecular Modellings ein verbessertes Verständnis biologischer und chemischer Daten zu Thieno[2,3-*b*]pyridin-Derivaten in ihrer Eigenschaft als selektive Inhibitoren der Glykogen Synthase Kinase-3 von *Plasmodium falciparum* (*Pf*GSK-3) entwickelt. Ausgangspunkt waren einerseits Untersuchungen von DROUCHEAU *et al.*, in denen die *Pf*GSK-3 als mögliche Zielstruktur in der Bekämpfung der durch *P. falciparum* verursachten Malaria vorgeschlagen wurde [80]. Andererseits konnte BRANDT im Rahmen präparativer Arbeiten eine Serie von Thieno[2,3-*b*]pyridinen als selektive Inhibitoren der *Pf*GSK-3 darstellen [92].

Da die dreidimensionale Struktur der *Pf*GSK-3 bislang nicht experimentell ermittelt werden konnte, beschäftigt sich der erste Teil der Arbeit mit der Erstellung von Homologie-Modellen als Voraussetzung für die Anwendung strukturbasierter Methoden der Leitstrukturoptimierung. Dazu wurden basierend auf dreidimensionalen Strukturen des homologen humanen Enzyms, der humanen Glykogen Synthase Kinase-3 β (*Hs*GSK-3 β), Homologie-Modelle entworfen. Durch den Einsatz 11 unterschiedlicher Vorlagen und das Erstellen von insgesamt 4400 Homologie-Modellen wurde der konformative Raum der *Pf*GSK-3 erschöpfend repräsentiert. Die Modelle wurden einer umfassenden Evaluierung mit literaturbekannten Methoden unterzogen. Dazu wurde ein eigenes Bewertungsverfahren eingeführt, das Ergebnisse unterschiedlicher Evaluierungswerkzeuge zu einer übergreifenden Bewertung zusammenfasst (Rank-by-rank based consensus scoring, *RCS*). Dieses Vorgehen führte zur Auswahl qualitativ hochwertiger Homologie-Modelle, deren Stabilität zusätzlich in Moleküldynamiksimulationen bestätigt wurde [163].

Diese Modelle und die Kristallstrukturen der *Hs*GSK-3 β , die jeweils als Vorlagen eingesetzt worden waren, wurden vergleichend mit molekularen Interaktionsfeldern (MIFs) untersucht. Die visuelle Analyse ergab Unterschiede im Bereich der ATP-Bindetasche, die für die Aktivität und Selektivität von Inhibitoren der *Pf*GSK-3 von Bedeutung sein können [294]. Die paarweise Gegenüberstellung der Daten von Homologie-Modellen und deren Vorlagen mittels Hauptkomponentenanalyse (PCA) führte zu auswertbaren Ergebnissen, die die visuellen Befunde bekräftigten. Eine hierarchische PCA (CPCA) mit der Gesamtheit der untersuchten Proteine führte hingegen nicht zu interpretierbaren Modellen.

In der Folge wurden in orientierenden Docking-Studien unter Verwendung der entwickelten Homologie-Modelle verschiedene Hypothesen zum Bindungsmodus der untersuchten

Thieno[2,3-*b*]pyridine entwickelt, aus denen Synthesevorschläge abgeleitet wurden. Einige davon konnten umgesetzt werden und führten zur Darstellung der bislang nicht beschriebenen Substanzen **101-104** (X: Cl, I und R: H, CN). Bei biologischen Testungen wurde ermittelt, dass **101-104** im mikromolaren Bereich (IC_{50} von 11.5 bis 68 μM) präferentiell bis selektiv die *Pf*GSK-3 gegenüber der im Testsystem eingesetzten Säuger-GSK-3 hemmen.



Zur genaueren Analyse des Bindungsmodus der untersuchten Thieno[2,3-*b*]pyridine wurde der Arbeitsablauf DoReCCoM entwickelt (Docking, Rescoring, Clustering, Complex Minimisation). Darin werden die Ergebnisse verschiedener Dockingprogramme kombiniert, es folgt ein Clustering der Verbindungen zur Reduktion der Datenmenge und ein Rescoring mit einer externen Scoringfunktion. Zuletzt werden die erhaltenen Protein-Ligand-Komplexe einer Energieminimierung unterzogen, ein abschließendes Rescoring in Entsprechung zum ersten ergibt die durch das Verfahren vorhergesagte Priorisierung von Posen. Die Leistungsfähigkeit von DoReCCoM wurde durch ein Crossdocking literaturbekannter Inhibitoren der *Hs*GSK-3 β , für die Protein-Ligand-Komplexe vorliegen, überprüft. Es konnte eine klare Überlegenheit gegenüber der isolierten Anwendung der einzelnen in DoReCCoM integrierten Dockingprogramme gezeigt werden. Durch Anwendung von DoReCCoM auf den durch **101-104** erweiterten Datensatz der Thieno[2,3-*b*]pyridine wurden unterschiedliche Bindungsmodi der ersten Docking-Studie reproduziert, es ergab sich jedoch eine andere Priorisierung der Modi.

Basierend auf Bindungsmodi, die dieser Priorisierung zufolge als wahrscheinlich gelten konnten, wurden Überlagerungen generiert und zur Erstellung von 3D-QSAR-Modellen herangezogen. Die bei der Analyse der CoMSIA-Modelle erhaltenen Konturdiagramme ermöglichten nach Projektion in die ATP-Bindetasche der Homologie-Modelle eine Kombination von struktur- und ligandbasierten Informationen. Als Resultat wird für den gesamten Datensatz ein gemeinsamer Bindungsmodus postuliert, der in Abbildung 6.1 beispielhaft dargestellt ist und der Ähnlichkeit zu einer nicht veröffentlichten Kristallstruktur eines Thieno[2,3-*b*]pyridins im Komplex mit der *Hs*GSK-3 β aufweist. Eine experimentelle Bestätigung des postulierten Bindungsmodus bleibt nachfolgenden Arbeiten vorbehalten.

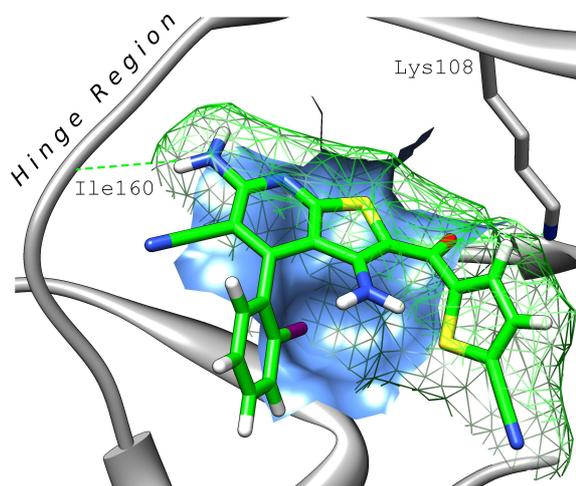


Abbildung 6.1. | Verbindung **102** als Beispiel für den postulierten Bindungsmodus der untersuchten Thieno[2,3-*b*]pyridine in der *Pf*GSK-3.

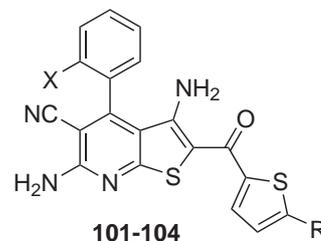
7. Summary

The aim of this thesis was to improve the understanding of biological and chemical data of thieno[2,3-*b*]pyridines as selective inhibitors of glycogen synthase kinase-3 of *Plasmodium falciparum* (*Pf*GSK-3) by using molecular modelling methods. A starting point was the work of DROUCHEAU *et al.*, who had proposed that the *Pf*GSK-3 might be target for the treatment of malaria tropica [80]. Additionally, the thesis of BRANDT describes the synthesis of a series of thieno[2,3-*b*]pyridines as selective inhibitors of *Pf*GSK-3 [92].

An experimental three-dimensional structure of *Pf*GSK-3 has not been published yet. Therefore the first step in the project was to build homology models for structure-based drug design. To this end, homology models were constructed based on 11 different crystal structures of *Hs*GSK-3 β , the human analogue of *Pf*GSK-3. This process led to 4400 3D models representing the conformational space of *Pf*GSK-3. The structures were evaluated according to established evaluation tools for homology models. The evaluation data were combined in a novel assessment system by calculating a rank-by-rank based consensus score (*RCS*). This results in the selection of high quality homology models the stability of which was confirmed in molecular dynamic simulations [163].

The homology models and corresponding templates were analysed and compared by means of molecular interaction fields (MIFs). The visual analysis elucidated differences between the ATP-binding site of the plasmodial and human enzymes that may contribute to selectivity of the ATP competitive inhibitors [294]. A pairwise statistical analysis of models and templates via principal component analysis (PCA) confirmed the results of the visual analysis, whereas a consensus PCA (CPCA) did not lead to interpretable models for the entirety of investigated protein structures.

Subsequent docking studies were performed to provide a presumed binding mode for the compounds, in order to generate proposals for the synthesis of new derivatives. Some of these proposals were carried out leading to the new compounds **101-104** (X: Cl, I and R: H, CN). **101-104** were tested for their inhibitory activity against plasmodial and mammal GSK-3 revealing values against *Pf*GSK-3 in the micromolar range (IC_{50} of 11.5 to 68 μ M) with selectivity towards the plasmodial enzyme.



In order to analyse the binding mode of the investigated thieno[2,3-*b*]pyridines in-depth, the workflow DoReCCoM was developed (Docking, Rescoring, Clustering, Complex Minimisation), where the results of different docking programs are combined, followed by a rescoring scheme with an external scoring function. Afterwards the poses are clustered to reduce the amount of data. At last, a complex minimisation with subsequent rescoring is carried out to prioritise the predicted docking poses. The capabilities of DoReCCoM were tested by a crossdocking experiment using published protein-inhibitor complexes of *Hs*GSK-3 β . The results obtained using DoReCCoM were superior compared to those of each docking program individually and, therefore, the procedure was applied to the dataset of thieno[2,3-*b*]pyridines in the *Pf*GSK-3. Binding modes of the first docking were reproduced, but the prioritisation was different.

Finally, 3D-QSAR-CoMSIA models were developed based on alignments reflecting the prioritisation of the proposed binding modes. Analysis of the respective contour maps enabled the combination of ligand and structure-based information. As a consequence, one common binding mode for all investigated thieno[2,3-*b*]pyridines is postulated. It is shown in figure 7.1 and can be characterised as similar to an unpublished X-ray crystal structure of a member of the dataset in complex with the *Hs*GSK-3 β . Experimental confirmation of the proposed binding mode of thieno[2,3-*b*]pyridines as selective inhibitors of *Pf*GSK-3 has to be performed in future studies.

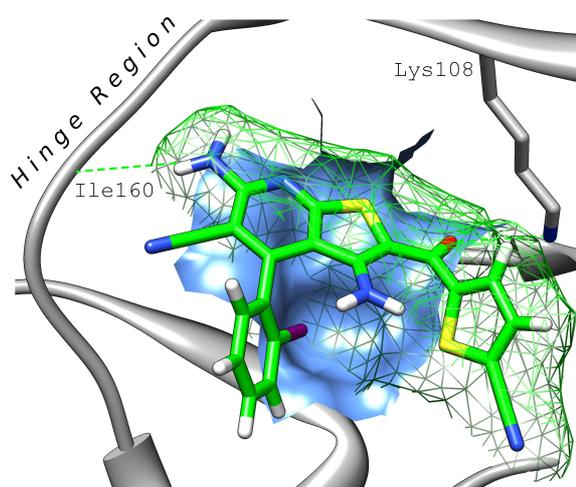


Figure 7.1. | Compound 102 as an example for the postulated binding mode of thieno[2,3-*b*]pyridines bound to *Pf*GSK-3.

Literatur

- 1 *World Malaria Report 2010*, World Health Organization, **2010**, Geneva, Switzerland.
- 2 *Global Malaria Action Plan*, Roll Back Malaria Partnership, **2008**, hosted by the WHO, Geneva, Switzerland.
- 3 S. I. HAY, C. A. GUERRA, P. W. GETHING, A. P. PATIL, A. J. TATEM, A. M. NOOR, C. W. KABARIA, B. H. MANH, I. R. F. ELYAZAR, S. BROOKER, D. L. SMITH, R. A. MOYEED, R. W. SNOW: A world malaria map: *Plasmodium falciparum* endemicity in 2007. *PLoS Med.* **2009**, *6*, e1000048.
- 4 C. A. GUERRA, P. W. GIKANDI, A. J. TATEM, A. M. NOOR, D. L. SMITH, S. I. HAY, R. W. SNOW: The limits and intensity of *Plasmodium falciparum* transmission: implications for malaria control and elimination worldwide. *PLoS Med.* **2008**, *5*, e38.
- 5 F. H. KAYSER: *Medizinische Mikrobiologie: Immunologie, Hygiene, Infektiologie, Bakteriologie, Mykologie, Virologie, Parasitologie*, Georg Thieme Verlag, Stuttgart, **2010**, 593–607.
- 6 T. N. C. WELLS, P. L. ALONSO, W. E. GUTTERIDGE: New medicines to improve control and contribute to the eradication of malaria. *Nat. Rev. Drug Discov.* **2009**, *8*, 879–891.
- 7 R. IGNATIUS: Plasmodien, in *Medizinische Mikrobiologie und Infektiologie*, H. Hahn, S. H. Kaufmann, T. F. Schulz, S. Suerbaum (Herausgeber), Springer Berlin, Berlin, **2008**, 660–664.
- 8 W. A. KROTOSKI, D. M. KROTOSKI, P. C. GARNHAM, R. S. BRAY, R. KILLICK-KENDRICK, C. C. DRAPER, G. A. TARGETT, M. W. GUY: Relapses in primate malaria: discovery of two populations of exoerythrocytic stages. Preliminary note. *Br. Med. J.* **1980**, *280*, 153–154.
- 9 S. CASARES, T.-D. BRUMEANU, T. L. RICHIE: The RTS,S malaria vaccine. *Vaccine* **2010**, *28*, 4880–4894.
- 10 P. BEJON *et al.*: Efficacy of RTS,S/AS01E vaccine against malaria in children 5 to 17 months of age. *N. Engl. J. Med.* **2008**, *359*, 2521–2532.
- 11 S. ABDULLA *et al.*: Safety and immunogenicity of RTS,S/AS02D malaria vaccine in infants. *N. Engl. J. Med.* **2008**, *359*, 2533–2544.
- 12 M. SCHLITZER: Antimalarial drugs - what is in use and what is in the pipeline. *Arch. Pharm. Chem. Life Sci.* **2008**, *341*, 149–163.
- 13 *Guidelines for the treatment of malaria – 2nd edition*, World Health Organization, **2010**, Geneva, Switzerland.
- 14 X. D. LUO, C. C. SHEN: The chemistry, pharmacology, and clinical applications of qinghaosu (artemisinin) and its derivatives. *Med. Res. Rev.* **1987**, *7*, 29–52.
- 15 S. R. MESHNICK, T. E. TAYLOR, S. KAMCHONWONGPAISAN: Artemisinin and the antimalarial endoperoxides: from herbal remedy to targeted chemotherapy. *Microbiol. Rev.* **1996**, *60*, 301–315.
- 16 R. JAMBOU, E. LEGRAND, M. NIANG, N. KHIM, P. LIM, B. VOLNEY, M. T. EKALA, C. BOUCHIER, P. ESTERRE, T. FANDEUR, O. MERCEREAU-PUJALON: Resistance of *Plasmodium falciparum* field isolates to in-vitro artemether and point mutations of the SERCA-type PfATPase6. *Lancet* **2005**, *366*, 1960–1963.

- 17 H. NOEDL, Y. SE, K. SCHAECHER, B. L. SMITH, D. SOCHEAT, M. M. FUKUDA, ARTEMISININ RESISTANCE IN CAMBODIA 1 (ARC1) STUDY CONSORTIUM: Evidence of artemisinin-resistant malaria in western Cambodia. *N. Engl. J. Med.* **2008**, *359*, 2619–2620.
- 18 A. M. DONDORP *et al.*: Artemisinin resistance in *Plasmodium falciparum* malaria. *N. Engl. J. Med.* **2009**, *361*, 455–467.
- 19 H. C. HOPPE, D. A. van SCHALKWYK, U. I. M. WIEHART, S. A. MEREDITH, J. EGAN, B. W. WEBER: Antimalarial quinolines and artemisinin inhibit endocytosis in *Plasmodium falciparum*. *Antimicrob. Agents Chemother.* **2004**, *48*, 2370–2378.
- 20 M. SCHLITZER: Wirkstoffe gegen Malaria: Was ist in der Pipeline? Wenig Neues im Kampf gegen die Malaria. *Pharm. unserer Zeit* **2009**, *38*, 522–526.
- 21 W. SNEADER: *Drug discovery – a history*, John Wiley & Sons Ltd, West Sussex, England, **2005**, 34–36.
- 22 T. J. EGAN: Haemozoin (malaria pigment): a unique crystalline drug target. *Targets* **2003**, *2*, 115–124.
- 23 C. P. SANCHEZ, J. E. MCLEAN, W. STEIN, M. LANZER: Evidence for a substrate specific and inhibitable drug efflux system in chloroquine resistant *Plasmodium falciparum* strains. *Biochemistry* **2004**, *43*, 16365–16373.
- 24 C. P. SANCHEZ, J. E. MCLEAN, P. ROHRBACH, D. A. FIDOCK, W. D. STEIN, M. LANZER: Evidence for a *pfcr1*-associated chloroquine efflux system in the human malarial parasite *Plasmodium falciparum*. *Biochemistry* **2005**, *44*, 9862–9870.
- 25 W. SNEADER: *Drug discovery – a history*, John Wiley & Sons Ltd, West Sussex, England, **2005**, 384–386.
- 26 D. S. PETERSON, D. WALLIKER, T. E. WELLEMS: Evidence that a point mutation in dihydrofolate reductase-thymidylate synthase confers resistance to pyrimethamine in falciparum malaria. *Proc. Natl. Acad. Sci. U. S. A.* **1988**, *85*, 9114–9118.
- 27 C. H. SIBLEY, J. E. HYDE, P. F. SIMS, C. V. PLOWE, J. G. KUBLIN, E. K. MBERU, A. F. COWMAN, P. A. WINSTANLEY, W. M. WATKINS, A. M. NZILA: Pyrimethamine-sulfadoxine resistance in *Plasmodium falciparum*: what next? *Trends Parasitol.* **2001**, *17*, 582–588.
- 28 M. W. MATHER, E. DARROUZET, M. VALKOVA-VALCHANOVA, J. W. COOLEY, M. T. MCINTOSH, F. DALDAL, A. B. VAIDYA: Uncovering the molecular mode of action of the antimalarial drug atovaquone using a bacterial system. *J. Biol. Chem.* **2005**, *280*, 27458–27465.
- 29 H. NAKATO, R. VIVANCOS, P. R. HUNTER: A systematic review and meta-analysis of the effectiveness and safety of atovaquone proguanil (Malarone) for chemoprophylaxis against malaria. *J. Antimicrob. Chemother.* **2007**, *60*, 929–936.
- 30 Deutsche Gesellschaft für Tropenmedizin und Internationale Gesundheit (DTG): Empfehlungen zur Malariavorbeugung, Stand April **2011**; Download von <http://www.dtg.org/>, 28.05.2011.
- 31 *Medicines for Malaria Venture Annual Report 2009*, Medicines for Malaria Venture, **2009**, Geneva, Switzerland.
- 32 C. L. YEATES *et al.*: Synthesis and structure-activity relationships of 4-pyridones as potential antimalarials. *J. Med. Chem.* **2008**, *51*, 2845–2852.
- 33 S. A. CHARMAN *et al.*: Synthetic ozonide drug candidate OZ439 offers new hope for a single-dose cure of uncomplicated malaria. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 4400–4405.

- 34 F.-J. GAMO, L. M. SANZ, J. VIDAL, C. de COZAR, E. ALVAREZ, J.-L. LAVANDERA, D. E. VANDERWALL, D. V. S. GREEN, V. KUMAR, S. HASAN, J. R. BROWN, C. E. PEISHOFF, L. R. CARDON, J. F. GARCIA-BUSTOS: Thousands of chemical starting points for antimalarial lead identification. *Nature* **2010**, *465*, 305–310.
- 35 C. A. LIPINSKI, F. LOMBARDO, B. W. DOMINY, P. J. FEENEY: Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* **2001**, *46*, 3–26.
- 36 S. EKINS, A. J. WILLIAMS: When pharmaceutical companies publish large datasets: an abundance of riches or fool's gold? *Drug Discov. Today* **2010**, *15*, 812–815.
- 37 H. JOMAA, J. WIESNER, S. SANDERBRAND, B. ALTINCICEK, C. WEIDEMEYER, M. HINTZ, I. TÜRBACHOVA, M. EBERL, J. ZEIDLER, H. K. LICHTENTHALER, D. SOLDATI, E. BECK: Inhibitors of the nonmevalonate pathway of isoprenoid biosynthesis as antimalarial drugs. *Science* **1999**, *285*, 1573–1576.
- 38 S. BORRMANN, S. ISSIFOU, G. ESSER, A. A. ADEGNIKA, M. RAMHARTER, P.-B. MATSIEGUI, S. OYAKHIROME, D. P. MAWILI-MBOUMBA, M. A. MISSINOU, J. F. J. KUN, H. JOMAA, P. G. KREMSNER: Fosmidomycin-clindamycin for the treatment of *Plasmodium falciparum* malaria. *J. Infect. Dis.* **2004**, *190*, 1534–1540.
- 39 S. BORRMANN, A. A. ADEGNIKA, P.-B. MATSIEGUI, S. ISSIFOU, A. SCHINDLER, D. P. MAWILI-MBOUMBA, T. BARANEK, J. WIESNER, H. JOMAA, P. G. KREMSNER: Fosmidomycin-clindamycin for *Plasmodium falciparum* Infections in African children. *J. Infect. Dis.* **2004**, *189*, 901–908.
- 40 S. BORRMANN, A. A. ADEGNIKA, F. MOUSSAVOU, S. OYAKHIROME, G. ESSER, P.-B. MATSIEGUI, M. RAMHARTER, I. LUNDGREN, M. KOMBILA, S. ISSIFOU, D. HUTCHINSON, J. WIESNER, H. JOMAA, P. G. KREMSNER: Short-course regimens of artesunate-fosmidomycin in treatment of uncomplicated *Plasmodium falciparum* malaria. *Antimicrob. Agents Chemother.* **2005**, *49*, 3749–3754.
- 41 M. J. GARDNER *et al.*: Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature* **2002**, *419*, 498–511.
- 42 V. PATEL, R. MAZITSCHKEK, B. COLEMAN, C. NGUYEN, S. URGAONKAR, J. CORTESE, R. H. BARKER, E. GREENBERG, W. TANG, J. E. BRADNER, S. L. SCHREIBER, M. T. DURASINGH, D. F. WIRTH, J. CLARDY: Identification and characterization of small molecule inhibitors of a class I histone deacetylase from *Plasmodium falciparum*. *J. Med. Chem.* **2009**, *52*, 2185–2187.
- 43 C. DOERIG, O. BILLKER, D. PRATT, J. ENDICOTT: Protein kinases as targets for antimalarial intervention: Kinomics, structure-based design, transmission-blockade, and targeting host cell enzymes. *Biochim. Biophys. Acta* **2005**, *1754*, 132–150.
- 44 E. H. FISCHER, D. J. GRAVES, E. R. S. CRITTENDEN, E. G. KREBS: Structure of the site phosphorylated in the phosphorylase *b* to *a* reaction. *J. Biol. Chem.* **1959**, *234*, 1698–1704.
- 45 P. COHEN: The role of protein phosphorylation in human health and disease. The Sir Hans Krebs Medal Lecture. *Eur. J. Biochem.* **2001**, *268*, 5001–5010.
- 46 J. M. BERG, J. L. TYMOCZKO, L. STRYER: *Biochemie*, Spektrum Akademischer Verlag, Heidelberg, **2009**, 318–319.
- 47 S. KLUMPP, J. KRIEGLSTEIN: Reversible phosphorylation of histidine residues in vertebrate proteins. *Biochim. Biophys. Acta* **2005**, *1754*, 291–295.
- 48 P. COHEN: Protein kinases—the major drug targets of the twenty-first century? *Nat. Rev. Drug Discov.* **2002**, *1*, 309–315.

- 49 R. EGLÉN, T. REISINE: Drug discovery and the human kinome: recent trends. *Pharmacol. Ther.* **2011**, *130*, 144–156.
- 50 G. MANNING, D. B. WHYTE, R. MARTINEZ, T. HUNTER, S. SUDARSANAM: The protein kinase complement of the human genome. *Science* **2002**, *298*, 1912–1934.
- 51 S. K. HANKS, T. HUNTER: Protein kinases 6. The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification. *FASEB J.* **1995**, *9*, 576–596.
- 52 J. J.-L. LIAO: Molecular recognition of protein kinase binding pockets for design of potent and selective kinase inhibitors. *J. Med. Chem.* **2007**, *50*, 409–424.
- 53 M. RABILLER, M. GETLIK, S. KLÜTER, A. RICHTERS, S. TÜCKMANTEL, J. R. SIMARD, D. RAUH: Proteus in the world of proteins: conformational changes in protein kinases. *Arch. Pharm. Chem. Life Sci.* **2010**, *343*, 193–206.
- 54 D. J. MATTHEWS, M. E. GERRITSEN: *Targeting protein kinases for cancer therapy*, John Wiley & Sons, Inc., Hoboken, New Jersey, **2010**, 89–91.
- 55 M. VALIEV, R. KAWAI, J. A. ADAMS, J. H. WEARE: The role of the putative catalytic base in the phosphoryl transfer reaction in a protein kinase: first-principles calculations. *J. Am. Chem. Soc.* **2003**, *125*, 9926–9927.
- 56 M. VALIEV, J. YANG, J. A. ADAMS, S. S. TAYLOR, J. H. WEARE: Phosphorylation reaction in cAPK protein kinase-free energy quantum mechanical/molecular mechanics simulations. *J. Phys. Chem. B* **2007**, *111*, 13455–13464.
- 57 J. A. ADAMS: Kinetic and catalytic mechanisms of protein kinases. *Chem. Rev.* **2001**, *101*, 2271–2290.
- 58 P. TRAXLER: Tyrosine kinase inhibitors in cancer treatment (Part II). *Expert Opin. Ther. Patents* **1998**, *8*, 1599–1625.
- 59 A. VULPETTI, R. BOSOTTI: Sequence and structural analysis of kinase ATP pocket residues. *II Farmaco* **2004**, *59*, 759–765.
- 60 A. C. PIERCE, K. L. SANDRETTO, G. W. BEMIS: Kinase inhibitors and the case for CH...O hydrogen bonds in protein-ligand binding. *Proteins* **2002**, *49*, 567–576.
- 61 A. C. PIERCE, E. ter HAAR, H. M. BINCH, D. P. KAY, S. R. PATEL, P. LI: CH...O and CH...N hydrogen bonds in ligand design: a novel quinazolin-4-ylthiazol-2-ylamine protein kinase inhibitor. *J. Med. Chem.* **2005**, *48*, 1278–1281.
- 62 O. PRIEN: The gatekeeper: friend or foe in identifying the next generation of kinase inhibitors. *ChemMedChem* **2006**, *1*, 1195–1196.
- 63 N. ZHANG, Y. JIANG, J. ZOU, B. ZHANG, H. JIN, Y. WANG, Q. YU: 3D QSAR for GSK-3 β inhibition by indirubin analogues. *Eur. J. Med. Chem.* **2006**, *41*, 373–378.
- 64 D. M. GOLDSTEIN *et al.*: Discovery of S-[5-amino-1-(4-fluorophenyl)-1H-pyrazol-4-yl]-[3-(2,3-dihydroxypropoxy)phenyl]methanone (RO3201195), an orally bioavailable and highly selective inhibitor of p38 MAP kinase. *J. Med. Chem.* **2006**, *49*, 1562–1575.
- 65 Y. LIU, N. S. GRAY: Rational design of inhibitors that bind to inactive kinase conformations. *Nat. Chem. Biol.* **2006**, *2*, 358–364.
- 66 Research Collaboratory for Structural Bioinformatics PDB, www.pdb.org.
- 67 H. M. BERMAN, J. WESTBROOK, Z. FENG, G. GILLILAND, T. N. BHAT, H. WEISSIG, I. N. SHINDYALOV, P. E. BOURNE: The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242.

- 68 C.-H. YUN, T. J. BOGGON, Y. LI, M. S. WOO, H. GREULICH, M. MEYERSON, M. J. ECK: Structures of lung cancer-derived EGFR mutants and inhibitor complexes: mechanism of activation and insights into differential inhibitor sensitivity. *Cancer Cell* **2007**, *11*, 217–227.
- 69 A. K. KUTACH, A. G. VILLASEÑOR, D. LAM, C. BELUNIS, C. JANSON, S. LOK, L.-N. HONG, C.-M. LIU, J. DEVAL, T. J. NOVAK, J. W. BARNETT, W. CHU, D. SHAW, A. KUGLSTATTER: Crystal structures of IL-2-inducible T cell kinase complexed with inhibitors: insights into rational drug design and activity regulation. *Chem. Biol. Drug Des.* **2010**, *76*, 154–163.
- 70 J. A. BERTRAND, S. THIEFFINE, A. VULPETTI, C. CRISTIANI, B. VALSASINA, S. KNAPP, H. M. KALISZ, M. FLOCCO: Structural characterization of the GSK-3 β active site using selective and non-selective ATP-mimetic inhibitors. *J. Mol. Biol.* **2003**, *333*, 393–407.
- 71 B. NAGAR, W. G. BORNMANN, P. PELLICENA, T. SCHINDLER, D. R. VEACH, W. T. MILLER, B. CLARKSON, J. KURIYAN: Crystal structures of the kinase domain of c-Abl in complex with the small molecule inhibitors PD173955 and imatinib (STI-571). *Cancer Res.* **2002**, *62*, 4236–4243.
- 72 Z. A. KNIGHT, K. M. SHOKAT: Features of selective kinase inhibitors. *Chem Biol* **2005**, *12*, 621–637.
- 73 F. ZUCCOTTO, E. ARDINI, E. CASALE, M. ANGIOLINI: Through the "gatekeeper door": exploiting the active kinase conformation. *J. Med. Chem.* **2010**, *53*, 2681–2694.
- 74 J. F. OHREN *et al.*: Structures of human MAP kinase kinase 1 (MEK1) and MEK2 describe novel noncompetitive kinase inhibition. *Nat. Struct. Mol. Biol.* **2004**, *11*, 1192–1197.
- 75 M. A. BOGOYEVITCH, D. P. FAIRLIE: A new paradigm for protein kinase inhibition: blocking phosphorylation without directly targeting ATP binding. *Drug Discov. Today* **2007**, *12*, 622–633.
- 76 N. EMBI, D. B. RYLATT, P. COHEN: Glycogen synthase kinase-3 from rabbit skeletal muscle. Separation from cyclic-AMP-dependent protein kinase and phosphorylase kinase. *Eur. J. Biochem.* **1980**, *107*, 519–527.
- 77 B. W. DOBLE, J. R. WOODGETT: GSK-3: tricks of the trade for a multi-tasking kinase. *J. Cell Sci.* **2003**, *116*, 1175–1186.
- 78 J. R. WOODGETT: Glycogen synthase kinase 3: an introductory synopsis, in *Glycogen synthase kinase 3 (GSK-3) and its inhibitors*, A. Martinez, A. Castro, M. Medina (Herausgeber), aus: Wiley series in drug discovery and development (B. Wang), Hoboken, New Jersey, **2006**, 3–23.
- 79 S. FRAME, P. COHEN: GSK3 takes centre stage more than 20 years after its discovery. *Biochem. J.* **2001**, *359*, 1–16.
- 80 E. DROUCHEAU, A. PRIMOT, V. THOMAS, D. MATTEI, M. KNOECKAERT, C. RICHARDSON, P. SALLICANDRO, P. ALANO, A. JAFARSHAD, B. BARATTE, C. KUNICK, D. PARZY, L. PEARL, C. DOERIG, L. MEIJER: *Plasmodium falciparum* glycogen synthase kinase-3: molecular model, expression, intramolecular localisation and selective inhibitors. *Biochim. Biophys. Acta* **2004**, *1697*, 181–196.
- 81 P. COHEN, M. GOEDERT: GSK3 inhibitors: development and therapeutic potential. *Nat. Rev. Drug Discov.* **2004**, *3*, 479–487.
- 82 M. P. MAZANETZ, P. M. FISCHER: Untangling tau hyperphosphorylation in drug design for neurodegenerative diseases. *Nat. Rev. Drug Discov.* **2007**, *6*, 464–479.
- 83 S. PHUKAN, V. S. BABU, A. KANNOJI, R. HARIHARAN, V. N. BALAJI: GSK3 β : role in therapeutic landscape and development of modulators. *Br. J. Pharmacol.* **2010**, *160*, 1–19.
- 84 R. DAJANI, E. FRASER, S. M. ROE, M. YEO, V. M. GOOD, V. THOMPSON, T. C. DALE, L. H. PEARL: Structural basis for recruitment of glycogen synthase kinase 3 β to the axin-APC scaffold complex. *EMBO J.* **2003**, *22*, 494–501.

- 85 C. J. FIOL, A. M. MAHRENHOLZ, Y WANG, R. W. ROESKE, P. J. ROACH: Formation of protein kinase recognition sites by covalent modification of the substrate. Molecular mechanism for the synergistic action of casein kinase II and glycogen synthase kinase 3. *J. Biol. Chem.* **1987**, *262*, 14042–14048.
- 86 C. SCHULTZ, A. LINK, M. LEOST, D. W. ZAHAREVITZ, R. GUSSIO, E. A. SAUSVILLE, L. MEIJER, C. KUNICK: Paullones, a series of cyclin-dependent kinase inhibitors: synthesis, evaluation of CDK1/cyclin B inhibition, and in vitro antitumor activity. *J. Med. Chem.* **1999**, *42*, 2909–2919.
- 87 C. KUNICK, K. LAUENROTH, M. LEOST, L. MEIJER, T. LEMCKE: 1-Azakenpaullone is a selective inhibitor of glycogen synthase kinase-3 β . *Bioorg. Med. Chem. Lett.* **2004**, *14*, 413–416.
- 88 A. MARTINEZ, M. ALONSO, A. CASTRO, C. PÉREZ, F. J. MORENO: First non-ATP competitive glycogen synthase kinase 3 β (GSK-3 β) inhibitors: thiadiazolidinones (TDZD) as potential drugs for the treatment of Alzheimer's disease. *J. Med. Chem.* **2002**, *45*, 1292–1299.
- 89 A. MARTINEZ, M. ALONSO, A. CASTRO, I. DORRONSORO, J. L. GELPÍ, F. J. LUQUE, C. PÉREZ, F. J. MORENO: SAR and 3D-QSAR studies on thiadiazolidinone derivatives: exploration of structural requirements for glycogen synthase kinase 3 inhibitors. *J. Med. Chem.* **2005**, *48*, 7103–7112.
- 90 L. SERENÓ, M. COMA, M. RODRÍGUEZ, P. SÁNCHEZ-FERRER, M. B. SÁNCHEZ, I. GICH, J. M. AGULLÓ, M. PÉREZ, J. AVILA, C. GUARDIA-LAGUARTA, J. CLARIMÓN, A. LLEÓ, T. GÓMEZ-ISLA: A novel GSK-3 β inhibitor reduces Alzheimer's pathology and rescues neuronal loss in vivo. *Neurobiol. Dis.* **2009**, *35*, 359–367.
- 91 Pressemitteilung der Firma Noscira zur Gewährung des Fast-Track-Status für Tideglusib (NP12=TDZD-8) durch die FDA, Zugriff 11.06.2011, http://www.noscira.com/media/docs/NP%20FAST%20TRACK_INGLES%20para%20env%C3%ADo.pdf.
- 92 W. BRANDT: Neue Inhibitoren der Proteinkinasen *Pf*GSK-3 und RET, Dissertation, Technische Universität Braunschweig, **2009**.
- 93 A. BAHL *et al.*: PlasmoDB: the Plasmodium genome resource. A database integrating experimental and computational data. *Nucleic Acids Res.* **2003**, *31*, 212–215.
- 94 P. WARD, L. EQUINET, J. PACKER, C. DOERIG: Protein kinases of the human malaria parasite *Plasmodium falciparum*: the kinome of a divergent eukaryote. *BMC Genomics* **2004**, *5*, 1–19.
- 95 ANAMIKA, N. SRINIVASAN, A. KRUPA: A genomic perspective of protein kinases in *Plasmodium falciparum*. *Proteins* **2005**, *58*, 180–189.
- 96 S. K. HANKS, A. M. QUINN: Protein kinase catalytic domain sequence database: identification of conserved features of primary structure and classification of family members. *Methods Enzymol.* **1991**, *200*, 38–62.
- 97 S. K. HANKS: Genomic analysis of the eukaryotic protein kinase superfamily: a perspective. *Genome Biol.* **2003**, *4*, 111.
- 98 C. DOERIG, O. BILLKER, T. HAYSTEAD, P. SHARMA, A. B. TOBIN, N. C. WATERS: Protein kinases of malaria parasites: an update. *Trends Parasitol.* **2008**, *24*, 570–577.
- 99 C. DOERIG, A. ABDI, N. BLAND, S. ESCHENLAUER, D. DORIN-SEMBLAT, C. FENNELL, J. HALBERT, Z. HOLLAND, M.-P. NIVEZ, J.-P. SEMBLAT, A. SICARD, L. REININGER: Malaria: Targeting parasite and host cell kinomes. *Biochim. Biophys. Acta* **2010**, *1804*, 604–612.
- 100 D. DORIN-SEMBLAT, A. SICARD, C. DOERIG, L. RANFORD-CARTWRIGHT, C. DOERIG: Disruption of the *Pf*PK7 gene impairs schizogony and sporogony in the human malaria parasite *Plasmodium falciparum*. *Eukaryot. Cell* **2008**, *7*, 279–285.

- 101 A. MERCKX, A. ECHALIER, K. LANGFORD, A. SICARD, G. LANGSLEY, J. JOORE, C. DOERIG, M. NOBLE, J. ENDICOTT: Structures of *P. falciparum* protein kinase 7 identify an activation motif and leads for inhibitor design. *Structure* **2008**, *16*, 228–238.
- 102 E. PIZZI, C. FRONTALI: Low-complexity regions in *Plasmodium falciparum* proteins. *Genome Res.* **2001**, *11*, 218–229.
- 103 N. BOULOC, J. M. LARGE, E. SMILJANIC, D. WHALLEY, K. H. ANSELL, C. D. EDLIN, J. S. BRYANS: Synthesis and in vitro evaluation of imidazopyridazines as novel inhibitors of the malarial kinase PfPK7. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 5294–5298.
- 104 S. HOLTON, A. MERCKX, D. BURGESS, C. DOERIG, M. NOBLE, J. ENDICOTT: Structures of *P. falciparum* PfPK5 test the CDK regulation paradigm and suggest mechanisms of small molecule inhibition. *Structure* **2003**, *11*, 1329–1337.
- 105 K. L. ROCH, C. SESTIER, D. DORIN, N. WATERS, B. KAPPES, D. CHAKRABARTI, L. MEIJER, C. DOERIG: Activation of a *Plasmodium falciparum* cdc2-related kinase by heterologous p25 and cyclin H. Functional characterization of a *P. falciparum* cyclin homologue. *J. Biol. Chem.* **2000**, *275*, 8952–8958.
- 106 S. MARTINEK, S. INONOG, A. S. MANOUKIAN, M. W. YOUNG: A role for the segment polarity gene shaggy/GSK-3 in the Drosophila circadian clock. *Cell* **2001**, *105*, 769–779.
- 107 S. BARIK, R. E. TAYLOR, D. CHAKRABARTI: Identification, cloning, and mutational analysis of the casein kinase 1 cDNA of the malaria parasite, *Plasmodium falciparum*. Stage-specific expression of the gene. *J. Biol. Chem.* **1997**, *272*, 26132–26138.
- 108 G. KLEBE: *Wirkstoffdesign*, Spektrum Akademischer Verlag, Heidelberg, **2009**, 17–18.
- 109 R. DAJANI, E. FRASER, S. M. ROE, N. YOUNG, V. GOOD, T. C. DALE, L. H. PEARL: Crystal structure of glycogen synthase kinase 3 β : structural basis for phosphate-primed substrate specificity and autoinhibition. *Cell* **2001**, *105*, 721–732.
- 110 S. L. MCGOVERN, B. K. SHOICHET: Information decay in molecular docking screens against holo, apo, and modeled conformations of enzymes. *J. Med. Chem.* **2003**, *46*, 2895–2907.
- 111 A. BORDOGNA, A. PANDINI, L. BONATI: Predicting the accuracy of protein-ligand docking on homology models. *J. Comput. Chem.* **2011**, *32*, 81–98.
- 112 R. C. WADE: Molecular Interaction Fields, in *3D QSAR in Drug Design. Theory, Methods and Applications*, H. Kubinyi (Herausgeber), ESCOM, Leiden, **1993**, 486–505.
- 113 P. J. GOODFORD: A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J. Med. Chem.* **1985**, *28*, 849–857.
- 114 *GRID 22*, Molecular Discovery Limited, **2004**, 4 Chandos street, London W1A 3BQ, England.
- 115 G. CRUCIANI: *Molecular interaction fields: Applications in drug discovery and ADME prediction*, John Wiley & Sons, Chichester, **2006**.
- 116 H. MATTER, W. SCHWAB: Affinity and selectivity of matrix metalloproteinase inhibitors: a chemometrical study from the perspective of ligands and proteins. *J. Med. Chem.* **1999**, *42*, 4506–4523.
- 117 E. FILIPPONI, V. CECCHETTI, O. TABARRINI, D. BONELLI, A. FRAVOLINI: Chemometric rationalization of the structural and physicochemical basis for selective cyclooxygenase-2 inhibition: toward more specific ligands. *J. Comput.-Aided Mol. Des.* **2000**, *14*, 277–291.
- 118 M. A. KASTENHOLZ, M. PASTOR, G. CRUCIANI, E. E. HAAKSMA, T. FOX: GRID/CPCA: a new computational tool to design selective ligands. *J. Med. Chem.* **2000**, *43*, 3033–3044.

- 119 M. RIDDERSTRÖM, I. ZAMORA, O. FJELLSTRÖM, T. B. ANDERSSON: Analysis of selective regions in the active sites of human cytochromes P450, 2C8, 2C9, 2C18, and 2C19 homology models using GRID/CPCA. *J. Med. Chem.* **2001**, *44*, 4072–4081.
- 120 T. NAUMANN, H. MATTER: Structural classification of protein kinases using 3D molecular interaction field analysis of their ligand binding sites: target family landscapes. *J. Med. Chem.* **2002**, *45*, 2366–2378.
- 121 G. E. TERP, G. CRUCIANI, I. T. CHRISTENSEN, F. S. JØRGENSEN: Structural differences of matrix metalloproteinases with potential implications for inhibitor selectivity examined by the GRID/CPCA approach. *J. Med. Chem.* **2002**, *45*, 2675–2684.
- 122 P. BRAIUCA, C. EBERT, L. FISCHER, L. GARDOSSI, P. LINDA: A homology model of penicillin acylase from *Alcaligenes faecalis* and in silico evaluation of its selectivity. *ChemBioChem* **2003**, *4*, 615–622.
- 123 H. JI, H. LI, M. FLINSPACH, T. L. POULOS, R. B. SILVERMAN: Computer modeling of selective regions in the active site of nitric oxide synthases: implication for the design of isoform-selective inhibitors. *J. Med. Chem.* **2003**, *46*, 5700–5711.
- 124 M. KURZ, V. BRACHVOGEL, H. MATTER, S. STENGELIN, H. THÜRING, W. KRAMER: Insights into the bile acid transportation system: the human ileal lipid-binding protein-cholelytaurine complex and its comparison with homologous structures. *Proteins* **2003**, *50*, 312–328.
- 125 E. MYSHKIN, B. WANG: Chemometrical classification of ephrin ligands and Eph kinases using GRID/CPCA approach. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1004–1010.
- 126 L. AFZELIUS, F. RAUBACHER, A. KARLÉN, F. S. JØRGENSEN, T. B. ANDERSSON, C. M. MASIMIREMBWA, I. ZAMORA: Structural analysis of CYP2C9 and CYP2C5 and an evaluation of commonly used molecular modeling techniques. *Drug Metab. Dispos.* **2004**, *32*, 1218–1229.
- 127 P. BRAIUCA, G. CRUCIANI, C. EBERT, L. GARDOSSI, P. LINDA: An innovative application of the "flexible" GRID/PCA computational method: study of differences in selectivity between PGAs from *Escherichia coli* and a *Providentia rettgeri* mutant. *Biotechnol. Prog.* **2004**, *20*, 1025–1031.
- 128 H. MATTER, P. KOTSONIS: Biology and chemistry of the inhibition of nitric oxide synthases by pteridine-derivatives as therapeutic agents. *Med. Res. Rev.* **2004**, *24*, 662–684.
- 129 A. VULPETTI, P. CRIVORI, A. CAMERON, J. BERTRAND, M. G. BRASCA, R. D'ALESSIO, P. PEVARELLO: Structure-based approaches to improve selectivity: CDK2-GSK3 β binding site analysis. *J. Chem. Inf. Model.* **2005**, *45*, 1282–1290.
- 130 A. BAIROCH, B. BOECKMANN, S. FERRO, E. GASTEIGER: Swiss-Prot: juggling between evolution and stability. *Brief. Bioinform.* **2004**, *5*, 39–55.
- 131 C. N. CAVASOTTO, S. S. PHATAK: Homology modeling in drug discovery: current trends and applications. *Drug Discov. Today* **2009**, *14*, 676–683.
- 132 T. ATTWOOD, R. L. DUNBRACK: Dictionary of Bioinformatics and Computational Biology, Comparative Modeling (Homology Modeling), Article Online Posting Date: October 15, **2004**.
- 133 M. A. MARTÍ-RENOM, A. C. STUART, A. FISER, R. SÁNCHEZ, F. MELO, A. ŠALI: Comparative protein structure modeling of genes and genomes. *Annu. Rev. Biophys. Biomol. Struct.* **2000**, *29*, 291–325.
- 134 S. F. ALTSCHUL, W. GISH, W. MILLER, E. W. MYERS, D. J. LIPMAN: Basic local alignment search tool. *J. Mol. Biol.* **1990**, *215*, 403–410.
- 135 S. F. ALTSCHUL, T. L. MADDEN, A. A. SCHAFFER, J. ZHANG, Z. ZHANG, W. MILLER, D. J. LIPMAN: Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **1997**, *25*, 3389–3402.

- 136 A. TRAMONTANO: Homology modeling with low sequence identity. *Methods* **1998**, *14*, 293–300.
- 137 S. B. NEEDLEMAN, C. D. WUNSCH: A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol.* **1970**, *48*, 443–453.
- 138 A. SALI, T. L. BLUNDEL: Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **1993**, *234*, 779–815.
- 139 *MODELLER 9v2*, <http://salilab.org/modeller/>.
- 140 R. A. LASKOWSKI, M. W. MACARTHUR, D. S. MOSS, J. M. THORNTON: PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **1993**, *26*, 283–291.
- 141 G. N. RAMACHANDRAN, V. SASISEKHARAN: Conformation of polypeptides and proteins. *Adv. Protein Chem.* **1968**, *23*, 283–438.
- 142 M. J. SIPPL: Recognition of errors in three-dimensional structures of proteins. *Proteins* **1993**, *17*, 355–362.
- 143 M. WIEDERSTEIN, M. J. SIPPL: ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res.* **2007**, *35*, W407–410.
- 144 H. J. C. BERENDSEN, D. van der SPOEL, R. van DRUNEN: GROMACS: A message-passing parallel molecular dynamics implementation. *Comput. Phys. Commun.* **1995**, *91*, 43–56.
- 145 E. LINDAHL, B. HESS, D. V. D. SPOEL: GROMACS 3.0: A package for molecular simulation and trajectory analysis. *J. Mol. Mod.* **2001**, *7*, 306–317.
- 146 D. V. D. SPOEL, E. LINDAHL, B. HESS, G. GROENHOF, A. E. MARK, H. J. C. BERENDSEN: GROMACS: Fast, flexible, and free. *J. Comput. Chem.* **2005**, *26*, 1701–1718.
- 147 M. R. BERTHOLD, N. CEBRON, F. DILL, G. DI FATTA, T. R. GABRIEL, F. GEORG, T. MEINL, P. OHL, C. SIEB, B. WISWEDEL: In *Proc. of the 4th Annual Industrial Simulation Conference (ISC), Workshop on Multi-Agent Systems and Simulation, Palermo, Italy*, 2006.
- 148 PlasmoDB, The Plasmodium Genome Resource, <http://plasmodb.org/plasmo/> Release 5.4, Februar **2008** (Feb. 2011: Release 7.10).
- 149 S. BOWMAN *et al.*: The complete nucleotide sequence of chromosome 3 of *Plasmodium falciparum*. *Nature* **1999**, *400*, 532–538.
- 150 M. AOKI, T. YOKOTA, I. SUGIURA, C. SASAKI, T. HASEGAWA, C. OKUMURA, K. ISHIGURO, T. KOHNO, S. SUGIO, T. MATSUZAKI: Structural insight into nucleotide recognition in tau-protein kinase I/glycogen synthase kinase 3 β . *Acta Crystallogr. D Biol. Crystallogr.* **2004**, *60*, 439–446.
- 151 R. BHAT *et al.*: Structural insights and biological effects of glycogen synthase kinase 3-specific inhibitor AR-A014418. *J. Biol. Chem.* **2003**, *278*, 45937–45945.
- 152 L. MEIJER *et al.*: GSK-3-selective inhibitors derived from Tyrian purple indirubins. *Chem. Biol.* **2003**, *10*, 1255–1266.
- 153 D. SHIN, S.-C. LEE, Y.-S. HEO, W.-Y. LEE, Y.-S. CHO, Y. E. KIM, Y.-L. HYUN, J. M. CHO, Y. S. LEE, S. RO: Design and synthesis of 7-hydroxy-1*H*-benzoimidazole derivatives as novel inhibitors of glycogen synthase kinase-3 β . *Bioorg. Med. Chem. Lett.* **2007**, *17*, 5686–5689.
- 154 H.-C. ZHANG, L. V. R. BOÑAGA, H. YE, C. K. DERIAN, B. P. DAMIANO, B. E. MARYANOFF: Novel bis(indolyl)maleimide pyridinophanes that are potent, selective inhibitors of glycogen synthase kinase-3. *Bioorg. Med. Chem. Lett.* **2007**, *17*, 2863–2868.
- 155 M. MENICHINCHERI *et al.*: First Cdc7 kinase inhibitors: pyrrolopyridinones as potent and orally active antitumor agents. 2. Lead discovery. *J. Med. Chem.* **2009**, *52*, 293–307.

- 156 M. SAITOH, J. KUNITOMO, E. KIMURA, Y. HAYASE, H. KOBAYASHI, N. UCHIYAMA, T. KAWAMOTO, T. TANAKA, C. D. MOL, D. R. DOUGAN, G. S. TEXTOR, G. P. SNELL, F. ITOH: Design, synthesis and structure-activity relationships of 1,3,4-oxadiazole derivatives as novel inhibitors of glycogen synthase kinase-3 β . *Bioorg. Med. Chem.* **2009**, *17*, 2017–2029.
- 157 M. SAITOH *et al.*: 2-3-[4-(Alkylsulfinyl)phenyl]-1-benzofuran-5-yl-5-methyl-1,3,4-oxadiazole derivatives as novel inhibitors of glycogen synthase kinase-3 β with good brain permeability. *J. Med. Chem.* **2009**, *52*, 6270–6286.
- 158 A. M. ARONOV *et al.*: Structure-guided design of potent and selective pyrimidylpyrrole inhibitors of extracellular signal-regulated kinase (ERK) using conformational control. *J. Med. Chem.* **2009**, *52*, 6362–6368.
- 159 M. ARNOST, A. PIERCE, E. t. HAAR, D. LAUFFER, J. MADDEN, K. TANNER, J. GREEN: 3-Aryl-4-(arylhyaazono)-1*H*-pyrazol-5-ones: Highly ligand efficient and potent inhibitors of GSK3 β . *Bioorg. Med. Chem. Lett.* **2010**, *20*, 1661–1664.
- 160 A. SCHAFFERHANS, G. KLEBE: Docking ligands onto binding site representations derived from proteins built by homology modelling. *J. Mol. Biol.* **2001**, *307*, 407–427.
- 161 S. KUMAR, R. NUSSINOV: Close-Range Electrostatic Interactions in Proteins. *ChemBioChem* **2002**, *3*, 604–617.
- 162 H. GOHLKE, G. KLEBE: Ansätze zur Beschreibung und Vorhersage der Bindungsaffinität niedermolekularer Liganden an makromolekulare Rezeptoren. *Angew. Chem.* **2002**, *114*, 2764–2798.
- 163 S. KRUGGEL, T. LEMCKE: Generation and evaluation of a homology model of *Pf*GSK-3. *Arch. Pharm. Chem. Life Sci.* **2009**, *342*, 327–332.
- 164 L. A. ROWLEY, D. NICHOLSON, N. G. PARSONAGE: Monte Carlo grand canonical ensemble calculation in a gas-liquid transition region for 12-6 Argon. *J. Comp. Phys.* **1975**, *17*, 401–414.
- 165 J. E. LENNARD-LONES: Cohesion. *Proc. Phys. Soc.* **1931**, *43*, 461–482.
- 166 B. R. BROOKS, R. E. BRUCCOLERI, B. D. OLAFSON, D. J. STATES, S. SWAMINATHAN, M. KARPLUS: CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **1983**, *4*, 187–217.
- 167 R. WADE: Calculation and application of molecular interaction fields, in *Molecular Interaction Fields. Applications in Drug Discovery and ADME Prediction*, G. Cruciani (Herausgeber), Wiley-VCH, Weinheim, **2006**, 27–42.
- 168 K. PEARSON: On lines and planes of closest fit to systems of points in space. *Philos. Mag.* **1901**, *2*, 559–572.
- 169 S. WOLD: Principal component analysis. *Chemometr. Intell. Lab. Syst.* **1987**, *2*, 37–52.
- 170 L. ERIKSSON, E. JOHANNSON, N. KETTANEH-WOLD, J. TYRGG, C. WIKSTRÖM, S. WOLD: *Multi- and Megavariate Data Analysis - Part I: Basic Principles and Applications*; Umetrics Academy, Umeå, Schweden, **2006**.
- 171 G. CRUCIANI, P. J. GOODFORD: A search for specificity in DNA-drug interactions. *J. Mol. Graphics Modell.* **1994**, *12*, 116–129.
- 172 M. PASTOR, G. CRUCIANI: A novel strategy for improving ligand selectivity in receptor-based drug design. *J. Med. Chem.* **1995**, *38*, 4637–4647.
- 173 M. PASTOR, M. BARRONI: GOLPE: Generating Optimal Linear PLS Estimations, Multivariate Infometric Analysis S.r.l., Viale dei Castagni, 16.
- 174 M. BARONI, G. COSTANTINO, G. CRUCIANI, D. RIGANELLI, R. VALIGI, S. CLEMENTI: Generating Optimal Linear PLS Estimations (GOLPE): an advanced chemometric tool for handling 3D-QSAR problems. *Quant. Struct.-Act. Relat.* **1993**, *12*, 9–20.

- 175 R. D. CRAMER, D. E. PATTERSON, J. D. BUNCE: Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. *J. Am. Chem. Soc.* **1988**, *110*, 5959–5967.
- 176 G. KLEBE, U. ABRAHAM, T. MIETZNER: 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.
- 177 T. FOX: Protein selectivity studies using GRID-MIFs, in *Molecular Interaction Fields. Applications in Drug Discovery and ADME Prediction*, G. Cruciani (Herausgeber), Wiley-VCH, Weinheim, **2006**, 45–82.
- 178 S. WOLD, S. HELLBERG, T. LUNDSTEDT, M. SJOSTROM UND H. WOLD, *Proc. Symp. on PLS Model Building: Theory and Application*, Frankfurt am Main, **1987**; auch: *Tech. rep., Department of Organic Chemistry*, Umeå University **1987**.
- 179 G. ERMONDI, C. ANGHILANTE, G. CARON: A combined in silico strategy to describe the variation of some 3D molecular properties of β -cyclodextrin due to the formation of inclusion complexes. *J. Mol. Graphics and Model.* **2006**, *25*, 296–303.
- 180 M. A. LARKIN, G. BLACKSHIELDS, N. P. BROWN, R. CHENNA, P. A. MCGETTIGAN, H. MCWILLIAM, F. VALENTIN, I. M. WALLACE, A. WILM, R. LOPEZ, J. D. THOMPSON, T. J. GIBSON, D. G. HIGGINS: Clustal W and Clustal X version 2.0. *Bioinformatics* **2007**, *23*, 2947–2948.
- 181 J. THOMPSON, D. HIGGINS, T. GIBSON: CLUSTAL W: improving the selectivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acid Research* **1994**, *22*, 4673–4680.
- 182 P. RICE, I. LONGDEN, A. BLEASBY: EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet.* **2000**, *16*, 276–277.
- 183 *SYBYL-X 1.2*, Tripos International, 1699 South Hanley Rd., St. Louis, Missouri, 63144, USA, www.tripos.com.
- 184 W. L. JORGENSEN, D. S. MAXWELL, J. TIRADO-RIVES: 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.
- 185 W. L. JORGENSEN, J. TIRADO-RIVES: The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *J. Am. Chem. Soc.* **1988**, *110*, 1657–1666.
- 186 H. J. C. BERENDSEN, J. P. M. POSTMA, W. F. van GUNSTEREN, A. DI NOLA, J. R. HAAS: Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81*, 3684–3690.
- 187 B. HESS, H. BEKKER, H. J. C. BERENDSEN, J. G. E. M. FRAAIJE: LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* **1997**, *18*, 1463–1472.
- 188 *Sybyl 8.1*, Tripos International, 1699 South Hanley Rd., St. Louis, Missouri, 63144, USA, www.tripos.com.
- 189 H. van de WATERBEEMD, R. E. CARTER, G. GRASSY, H. KUBINYI, Y. C. MARTIN, M. S. TUTE, P. WILLET: Glossary of terms used in computational drug design (IUPAC Recommendations 1997). *Pure Appl. Chem.* **1997**, *69*, 1142.
- 190 G. SCHNEIDER, K.-H. BARINGHAUS: *Molecular Design: Concepts and Applications*; Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, **2008**, 74 ff.
- 191 I. D. KUNTZ, J. M. BLANEY, S. J. OATLEY, R. LANGRIDGE, T. E. FERRIN: A geometric approach to macromolecule-ligand interactions. *J. Mol. Biol.* **1982**, *161*, 269–288.

- 192 U. RESTER: Dock around the clock – current status of small molecule docking and scoring. *QSAR Comb. Sci.* **2006**, *25*, 605–615.
- 193 M. RAREY, B. KRAMER, T. LENGAUER, G. KLEBE: A fast flexible docking method using an incremental construction algorithm. *J. Mol. Biol.* **1996**, *261*, 470–489.
- 194 G. M. MORRIS, D. S. GOODSSELL, R. S. HALLIDAY, R. HUEY, W. E. HART, R. K. BELEW, A. J. OLSON: Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.* **1998**, *19*, 1639–1662.
- 195 M. R. MCGANN, H. R. ALMOND, A. NICHOLLS, J. A. GRANT, F. K. BROWN: Gaussian docking functions. *Biopolymers* **2003**, *68*, 76–90.
- 196 R. A. FRIESNER, J. L. BANKS, R. B. MURPHY, T. A. HALGREN, J. J. KLICIC, D. T. MAINZ, M. P. REPASKY, E. H. KNOLL, M. SHELLEY, J. K. PERRY, D. E. SHAW, P. FRANCIS, P. S. SHENKIN: Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* **2004**, *47*, 1739–1749.
- 197 T. A. HALGREN, R. B. MURPHY, R. A. FRIESNER, H. S. BEARD, L. L. FRYE, W. T. POLLARD, J. L. BANKS: Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *J. Med. Chem.* **2004**, *47*, 1750–1759.
- 198 S. K. KEARSLEY, D. J. UNDERWOOD, R. P. SHERIDAN, M. D. MILLER: Flexibases: a way to enhance the use of molecular docking methods. *J. Comput.-Aided Mol. Des.* **1994**, *8*, 565–582.
- 199 P. COZZINI, G. E. KELLOGG, F. SPYRAKIS, D. J. ABRAHAM, G. COSTANTINO, A. EMERSON, F. FANELLI, H. GOHLKE, L. A. KUHN, G. M. MORRIS, M. OROZCO, T. A. PERTINHEZ, M. RIZZI, C. A. SOTRIFFER: Target flexibility: an emerging consideration in drug discovery and design. *J. Med. Chem.* **2008**, *51*, 6237–6255.
- 200 C. B-RAO, J. SUBRAMANIAN, S. D. SHARMA: Managing protein flexibility in docking and its applications. *Drug Discov. Today* **2009**, *14*, 394–400.
- 201 A. R. LEACH, B. K. SHOICHET, C. E. PEISHOFF: Prediction of protein-ligand interactions. Docking and scoring: successes and gaps. *J. Med. Chem.* **2006**, *49*, 5851–5855.
- 202 T. TUCCINARDI, M. BOTTA, A. GIORDANO, A. MARTINELLI: Protein kinases: docking and homology modeling reliability. *J. Chem. Inf. Model.* **2010**, *50*, 1432–1441.
- 203 G. KLEBE: *Wirkstoffdesign*, Spektrum Akademischer Verlag, Heidelberg, **2009**, 312.
- 204 R. WANG, Y. LU, X. FANG, S. WANG: An extensive test of 14 scoring functions using the PDBbind refined set of 800 protein-ligand complexes. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 2114–2125.
- 205 T. CHENG, X. LI, Y. LI, Z. LIU, R. WANG: Comparative assessment of scoring functions on a diverse test set. *J. Chem. Inf. Model.* **2009**, *49*, 1079–1093.
- 206 D. PLEWCZYNSKI, M. ŁAŻNIEWSKI, R. AUGUSTYNIAK, K. GINALSKI: Can we trust docking results? Evaluation of seven commonly used programs on PDBbind database. *J. Comput. Chem.* **2011**, *32*, 742–755.
- 207 D. YUSUF, A. M. DAVIS, G. J. KLEYWEGT, S. SCHMITT: An alternative method for the evaluation of docking performance: RSR vs RMSD. *J. Chem. Inf. Model.* **2008**, *48*, 1411–1422.
- 208 R. T. KROEMER, A. VULPETTI, J. J. McDONALD, D. C. ROHRER, J.-Y. TROSSET, F. GIORDANETTO, S. COTESTA, C. McMARTIN, M. KIHLEN, P. F. W. STOUTEN: Assessment of docking poses: interactions-based accuracy classification (IBAC) versus crystal structure deviations. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 871–881.
- 209 <http://blast.ncbi.nlm.nih.gov/Blast.cgi>.

- 210 H. FAN, J. J. IRWIN, B. M. WEBB, G. KLEBE, B. K. SHOICHET, A. SALI: Molecular docking screens using comparative models of proteins. *J. Chem. Inf. Model.* **2009**, *49*, 2512–2527.
- 211 E. M. NOVOA, L. R. d. POUPLANA, X. BARRIL, M. OROZCO: Ensemble docking from homology models. *J. Chem. Theory Comput.* **2010**, *6*, 2547–2557.
- 212 N. MISHRA, A. BASU, V. JAYAPRAKASH, A. SHARON, M. BASU, K. K. PATNAIK: Structure based virtual screening of GSK-3 β : importance of protein flexibility and induced fit. *Bioorg. Med. Chem. Lett.* **2009**, *19*, 5582–5585.
- 213 H. ALONSO, A. A. BLIZNYUK, J. E. GREASY: Combining docking and molecular dynamic simulations in drug design. *Med. Res. Rev.* **2006**, *26*, 531–568.
- 214 S. B. NABUURS, M. WAGENER, J. de Vlieg: A flexible approach to induced fit docking. *J. Med. Chem.* **2007**, *50*, 6507–6518.
- 215 R. S. ARMEN, J. CHEN, C. L. BROOKS: An evaluation of explicit receptor flexibility in molecular docking using molecular dynamics and torsion angle molecular dynamics. *J. Chem. Theory Comput.* **2009**, *5*, 2909–2923.
- 216 M. RAREY, B. KRAMER, T. LENGAUER: Multiple automatic base selection: protein-ligand docking based on incremental construction without manual intervention. *J. Comput.-Aided Mol. Des.* **1997**, *11*, 369–384.
- 217 *FlexX 3.0.2*, BioSolveIT, 53757 St. Augustin, Germany, **2008**.
- 218 H. CLAUSSEN, C. BUNING, M. RAREY, T. LENGAUER: FlexE: efficient molecular docking considering protein structure variations. *J. Mol. Biol.* **2001**, *308*, 377–395.
- 219 H. J. BÖHM: 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.
- 220 R. HUEY, G. M. MORRIS, A. J. OLSON, D. S. GOODSSELL: A semiempirical free energy force field with charge-based desolvation. *J. Comput. Chem.* **2007**, *28*, 1145–1152.
- 221 M. F. SANNER: Python: a programming language for software integration and development. *J. Mol. Graphics Modell.* **1999**, *17*, 57–61.
- 222 G. M. MORRIS, R. HUEY, W. LINDSTROM, M. F. SANNER, R. K. BELEW, D. S. GOODSSELL, A. J. OLSON: AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J. Comput. Chem.* **2009**, *30*, 2785–2791.
- 223 *FRED version 2.2.5*, OpenEye Scientific Software, Inc., Santa Fe, NM, USA, www.eyesopen.com, **2009**.
- 224 H. GOHLKE, M. HENDLICH, G. KLEBE: Knowledge-based scoring function to predict protein-ligand interactions. *J. Mol. Biol.* **2000**, *295*, 337–356.
- 225 H. F. G. VELEC, H. GOHLKE, G. KLEBE: 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.
- 226 F. H. ALLAN: The Cambridge Structural Database: a quarter of a million crystal structures and rising. *Acta Crystallogr., Sect. B: Struct. Sci.* **2002**, *58*, 380–388.
- 227 *SZYBKI version 1.3.4*, OpenEye Scientific Software Inc., Santa Fe, NM, USA, www.eyesopen.com, **2009**.
- 228 T. A. HALGREN: MMFF VII. Characterization of MMFF94, MMFF94s, and other widely available force fields for conformational energies and for intermolecular-interaction energies and geometries. *J. Comput. Chem.* **1999**, *20*, 730–748.

- 229 M. CHIOUA, A. SAMADI, E. SORIANO, O. LOZACH, L. MEIJER, J. MARCO-CONTELLES: Synthesis and biological evaluation of 3,6-diamino-1*H*-pyrazolo[3,4-*b*]pyridine derivatives as protein kinase inhibitors. *Bioorg. Med. Chem. Lett.* **2009**, *19*, 4566–4569.
- 230 Q.-Y. ZHUANG, J.-N. XU, S.-J. TU, R.-H. JIA, J.-Y. ZHANG, C.-M. LI, D.-X. ZHOU: A simple one-pot synthesis of 3,5-dicyano-1,2-dihydropyrid-2-ones in aqueous media. *Chin. J. Chem.* **2007**, *25*, 1568–1572.
- 231 Y. A. SHARANIN, S. G. KRIVOKOLYSKO, V. D. DYACHENKO: Nitrile cyclization reactions. LIV. Synthesis and properties of 6-amino-4-aryl-3,5-dicyanopyridin-2(1*H*)-ones, -thiones, -ylidene-malononitriles and their hydrogenated analogs. *Russ. J. Org. Chem.* **1994**, *30*, 581–587.
- 232 W. L. DELANO, The PyMOL Molecular Graphics System v0.99, **2006**, DeLano Scientific LLC, Palo Alto, CA, USA. <http://www.pymol.org>.
- 233 G. NEUDERT, G. KLEBE: fconv: format conversion, manipulation and feature computation of molecular data. *Bioinformatics* **2011**, *27*, 1021–1022.
- 234 M. J. D. POWELL: Restart procedures for the conjugate gradient method. *Math. Program.* **1977**, *12*, 241–254.
- 235 W. H. PRESS, B. P. FLANNERY, S. A. TEUKOLSKY, W. T. VETTERLING: *Numerical Recipes in C - The Art of Scientific Computing*; Cambridge University Press, Cambridge, **1988**.
- 236 E. F. PETERSEN, T. D. GODDARD, C. C. HUANG, G. S. COUCH, D. M. GREENBLATT, E. C. MENG, T. E. FERRIN: UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004**, *25*, 1605–1612.
- 237 *drugscoreX 0.48*, Gerd Neudert, AK Prof. Dr. G. Klebe, University of Marburg, unpublished test version, pers. communication, **2010**.
- 238 *LeadIT 1.0.1*, BioSolveIT, 53757 St. Augustin, Germany, **2010**.
- 239 *OMEGA version 2.3.2*, OpenEye Scientific Software, Inc., Santa Fe, NM, USA, www.eyesopen.com, **2008**.
- 240 H. van de WATERBEEMD, R. E. CARTER, G. GRASSY, H. KUBINYI, Y. C. MARTIN, M. S. TUTE, P. WILLETT: Glossary of terms used in computational drug design (IUPAC Recommendations 1997). *Pure Appl. Chem.* **1997**, *69*, 1149.
- 241 A. CRUM BROWN, T. FRAZER: On the connection between chemical constitution and physiological action. Part 1. On the physiological action of the ammonium bases, derived from strychnia, brucia, thebaia, codeia, morphia and nicotia. *Trans. Roy. Soc. Edinburgh* **1868**, *25*, 151–203.
- 242 C. HANSCH, T. FUJITA: $\rho - \sigma - \pi$ analysis. A method for the correlation of biological activity and chemical structure. *J. Am. Chem. Soc.* **1964**, *86*, 1616–1626.
- 243 S. M. FREE, J. W. WILSON: A mathematical contribution to structure-activity studies. *J. Med. Chem.* **1964**, *7*, 395–399.
- 244 L. P. HAMMETT: Some relations between reaction rates and equilibrium constants. *Chem. Rev.* **1935**, *17*, 125–136.
- 245 L. P. HAMMETT: The effect of structure upon the reactions of organic compounds. Benzene derivatives. *J. Am. Chem. Soc.* **1937**, *59*, 96–103.
- 246 C. HANSCH: Quantitative approach to biochemical structure-activity relationships. *Acc. Chem. Res.* **1969**, *2*, 232–239.
- 247 H. KUBINYI: From narcosis to hyperspace: The history of QSAR. *Quant. Struct.-Act. Relat.* **2002**, *21*, 348–356.

- 248 R. TODESCHINI, V. CONSONNI: *Molecular descriptors for chemoinformatics: Volume I: Alphabetical listing / Volume II: Appendices, references (methods and principles in medicinal chemistry)*, H. Kubinyi, G. Folkers (Herausgeber), Wiley-VCH, Weinheim, **2009**.
- 249 VCCLAB, Virtual Computational Chemistry Laboratory, <http://www.vcclab.org>, **2005**.
- 250 I. V. TETKO, J. GASTEIGER, R. TODESCHINI, A. MAURI, D. LIVINGSTONE, P. ERTL, V. A. PALYULIN, E. V. RADCHENKO, N. S. ZEFIROV, A. S. MAKARENKO, V. Y. TANCHUK, V. V. PROKOPENKO: Virtual computational chemistry laboratory—design and description. *J. Comput.-Aided Mol. Des.* **2005**, *19*, 453–463.
- 251 W. SIPPL: 3D-QSAR using the GRID/GOLPE approach, in *Molecular Interaction Fields. Applications in Drug Discovery and ADME Prediction*, G. Cruciani (Herausgeber), Wiley-VCH, Weinheim, **2006**, 145–170.
- 252 R. D. I. CRAMER, M. MILNE: The lattice model: a general paradigm for shape-related structure/activity correlation. *Abstracts of papers of the Am. Chem. Soc. Meeting* **1979**, COMP 44.
- 253 G. KLEBE: *Wirkstoffdesign*, Spektrum Akademischer Verlag, Heidelberg, **2009**, 273.
- 254 T. SCIOR, J. L. MEDINA-FRANCO, Q.-T. DO, K. MARTÍNEZ-MAYORGA, J. A. Y. ROJAS, P. BERNARD: How to recognize and work around pitfalls in QSAR studies: a critical review. *Curr. Med. Chem.* **2009**, *16*, 4297–4313.
- 255 I. GARCÍA, Y. FALL, G. GÓMEZ: QSAR, docking, and CoMFA studies of GSK3 inhibitors. *Curr. Pharm. Des.* **2010**, *16*, 2666–2675.
- 256 N. DESSALEW, P. V. BHARATAM: 3D-QSAR and molecular docking study on bisarylmaleimide series as glycogen synthase kinase 3, cyclin dependent kinase 2 and cyclin dependent kinase 4 inhibitors: an insight into the criteria for selectivity. *Eur. J. Med. Chem.* **2007**, *42*, 1014–1027.
- 257 J. XIAO, Z. GUO, Y. GUO, F. CHU, P. SUN: Inhibitory mode of N-phenyl-4-pyrazolo[1,5-*b*]pyridazin-3-ylpyrimidin-2-amine series derivatives against GSK-3: molecular docking and 3D-QSAR analyses. *Protein Eng., Des. Sel.* **2006**, *19*, 47–54.
- 258 A. TROPSHA: Best practices for QSAR model development, validation, and exploitation. *Mol. Inf.* **2010**, *29*, 476–488.
- 259 A. TROPSHA, P. GRAMATICA, V. K. GOMBAR: The importance of being earnest: Validation is the absolute essential for successful application and interpretation of QSPR models. *QSAR Comb. Sci.* **2003**, *22*, 69–77.
- 260 H. KUBINYI: Validation and predictivity of QSAR models, in *QSAR & Molecular Modelling in Rational Design of Bioactive Molecules*, Proceedings of the 15th European Symposium on QSAR & Molecular Modelling, Istanbul, Turkey, 2004, CADDD Society, Ankara, Turkey, E. A. Sener, I. Yalcin (Herausgeber), **2006**, 30–33.
- 261 S. WOLD, M. SJÖSTRÖM, L. ERIKSSON: PLS-regression: a basic tool of chemometrics. *Chemometr. Intel. Lab. Syst.* **2001**, *58*, 109–130.
- 262 H. WOLD: Estimation of principal components and related models by iterative least squares, aus *Multivariate Analysis*, P. R. Krishnaiah (Herausgeber), Academic Press, **1966**.
- 263 Tripos Bookshelf, *SYBYL-X 1.2*, Tripos International, 1699 South Hanley Rd., St. Louis, Missouri, 63144, USA.
- 264 S. WOLD, L. ERIKSSON: Statistical validation of QSAR results, in *Chemometric methods in molecular design*, H. van de Waterbeemd (Herausgeber), VCH Publishers, Inc., New York, NY (USA), **1995**, 309–318.
- 265 J. SHAO: Linear model selection by cross-validation. *J. Am. Stat. Assoc.* **1993**, *88*, 486–494.

- 266 A. GOLBRAIKH, A. TROPSHA: Beware of q^2 ! *J. Mol. Graph. Model.* **2002**, *20*, 269–276.
- 267 B. L. BUSH, R. B. NACHBAR: Sample-distance partial least squares: PLS optimized for many variables, with application to CoMFA. *J. Comput.-Aided Mol. Des.* **1993**, *7*, 587–619.
- 268 K. H. KIM, G. GRECO, E. NOVELLINO: A critical review of recent CoMFA applications. *Perspect. Drug Discov.* **1998**, *12–14*, 257–315.
- 269 N. DESSALEW, D. S. PATEL, P. V. BHARATAM: 3D-QSAR and molecular docking studies on pyrazolopyrimidine derivatives as glycogen synthase kinase-3 β inhibitors. *J. Mol. Graphics Modell.* **2007**, *25*, 885–895.
- 270 E. LESCOT, R. BUREAU, J. S. de OLIVEIRA SANTOS, C. ROCHAIS, V. LISOWSKI, J.-C. LANCELOT, S. RAULT: 3D-QSAR and docking studies of selective GSK-3 β inhibitors. Comparison with a thieno[2,3-*b*]pyrrolizinone derivative, a new potential lead for GSK-3 β ligands. *J. Chem. Inf. Model.* **2005**, *45*, 708–715.
- 271 R. D. CLARK, D. G. SPROUS, J. M. LEONARD: Validation models based on large data sets, in *Rational Approaches to Drug Design*, H.-D. Höltje und W. Sippl (Herausgeber), Prous Science SA, **2001**, 475–485.
- 272 R. CLARK, P. FOX: Statistical variation in progressive scrambling. *J. Comput.-Aided Mol. Des.* **2004**, *18*, 563–576.
- 273 *FlexS 2.0.0*, BiosolveIT GmbH, An der Ziegelei 79, 53757 St. Augustin, Germany, www.biosolveit.de.
- 274 C. LEMMEN, T. LENGAUER: Time-efficient flexible superposition of medium-sized molecules. *J. Comput.-Aided Mol. Des.* **1997**, *11*, 357–368.
- 275 C. LEMMEN, T. LENGAUER, G. KLEBE: FLEXS: a method for fast flexible ligand superposition. *J. Med. Chem.* **1998**, *41*, 4502–4520.
- 276 *ROCS, version 2.4.1*, OpenEye Scientific Software, Inc., Santa Fe, NM, USA, www.eyesopen.com.
- 277 J. A. GRANT, M. A. GALLARDO, B. T. PICKUP: A fast method of molecular shape comparison: A simple application of a Gaussian description of molecular shape. *J. Comput. Chem.* **1996**, *17*, 1653–1666.
- 278 *QUACPAC, version 1.7.4*, OpenEye Scientific Software, Inc., Santa Fe, NM, USA, www.eyesopen.com.
- 279 M. J. S. DEWAR, E. G. ZOEBISCH, E. F. HEALY, J. J. P. STEWART: Development and use of quantum mechanical molecular models. 76. AM1: a new general purpose quantum mechanical molecular model. *J. Am. Chem. Soc.* **1985**, *107*, 3902–3909.
- 280 A. JAKALIAN, B. L. BUSH, D. B. JACK, C. I. BAYLY: Fast, efficient generation of high-quality atomic charges. AM1-BCC model: I. Method. *J. Comput. Chem.* **2000**, *21*, 132–146.
- 281 A. JAKALIAN, D. B. JACK, C. I. BAYLY: Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *J. Comput. Chem.* **2002**, *23*, 1623–1641.
- 282 R. R. MITTAL, L. HARRIS, R. A. MCKINNON, M. J. SORICH: Partial charge calculation method affects CoMFA QSAR prediction accuracy. *J. Chem. Inf. Model.* **2009**, *49*, 704–709.
- 283 K.-C. TSAI, Y.-C. CHEN, N.-W. HSIAO, C.-L. WANG, C.-L. LIN, Y.-C. LEE, M. LI, B. WANG: A comparison of different electrostatic potentials on prediction accuracy in CoMFA and CoMSIA studies. *Eur. J. Med. Chem.* **2010**, *45*, 1544–1551.
- 284 D. HUANG, T. ZHOU, K. LAFLEUR, C. NEVADO, A. CAFLISCH: Kinase selectivity potential for inhibitors targeting the ATP binding site: a network analysis. *Bioinformatics* **2010**, *26*, 198–204.

-
- 285 R. GOWTHAMAN, D. SEKHAR, M. K. KALITA, D. GUPTA: A database for *Plasmodium falciparum* protein models. *Bioinformatics* **2005**, *1*, 50–51.
- 286 J. SCHWÖBEL, R.-U. EBERT, R. KÜHNE, G. SCHÜÜRMAN: Prediction of the intrinsic hydrogen bond acceptor strength of organic compounds by local molecular parameters. *J. Chem. Inf. Model.* **2009**, *49*, 956–962.
- 287 A. K. GHOSE, V. N. VISWANADHAN, J. J. WENDOLOSKI: 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.
- 288 G. SCHNEIDER, K.H. BARINGHAUS: *Molecular design: concepts and applications*, Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, **2008**, 113–114.
- 289 K. H. KIM: Outliers in SAR and QSAR: is unusual binding mode a possible source of outliers? *J. Comput.-Aided Mol. Des.* **2007**, *21*, 63–86.
- 290 S. J. TEAGUE: Implications of protein flexibility for drug discovery. *Nat. Rev. Drug Discov.* **2003**, *2*, 527–541.
- 291 J. KYTE, R. F. DOOLITTLE: A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **1982**, *157*, 105–132.
- 292 G. SCHNEIDER, K.H. BARINGHAUS: *Molecular design: concepts and applications*, Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, **2008**, 56.
- 293 E. F. F. da CUNHA, W. SIPPL, T. de CASTRO RAMALHO, O. A. CEVA ANTUNES, R. B. de ALENCASTRO, M. G. ALBUQUERQUE: 3D-QSAR CoMFA/CoMSIA models based on theoretical active conformers of HOE/BAY-793 analogs derived from HIV-1 protease inhibitor complexes. *European Journal of Medicinal Chemistry* **2009**, *44*, 4344–4352.
- 294 S. KRUGGEL, T. LEMCKE: Comparative investigation of the ATP-binding site of human and plasmodial glycogen synthase kinase-3. *QSAR Comb. Sci.* **2009**, *28*, 885–890.

Anhang

A. Ergänzende Daten

A.1. Alignments

Aus dem dargestellten Multialignment in Entsprechung zu Abbildung 2.5 sind die Lücken in den unterschiedlichen Vorlagen ersichtlich.

```
nr(Hs) 35...40.....60.....80.....90....
1J1B(B) SKVTTVVATPGQGPDRPQEVSYDTKVIENGSGVGVVYQAKLCDSGELVAIKKVLQDKRFK
1Q3D(B) SKVTTVVATPGQGPDRPQEVSYDTKVIENGSGVGVVYQAKLCDSGELVAIKKVLQDKRFK
1Q3W   SKVTTVVATPGQGPDRPQEVSYDTKVIENGSGVGVVYQAKLCDSGELVAIKKVLQDKRFK
1Q4L   SKVTTVVATPGQGPDRPQEVSYDTKVIENGSGVGVVYQAKLCDSGELVAIKKVLQDKRFK
1Q41(B) SKVTTVVATPGQGPDRPQEVSYDTKVIENGSGVGVVYQAKLCDSGELVAIKKVLQDKRFK
1Q5K(B) SKVTTVVATPGQGPDRPQEVSYDTKVIENGSGVGVVYQAKLCDSGELVAIKKVLQDKRFK
1R0E   SKVTTVVATPGQGPDRPQEVSYDTKVIENGSGVGVVYQAKLCDSGELVAIKKVLQDKRFK
1UV5   SKVTTVVATPGQGPDRPQEVSYDTKVIENGSGVGVVYQAKLCDSGELVAIKKVLQDKRFK
205K   SKVTTVVATPGQGPDRPQEVSYDTKVIENGSGVGVVYQAKLCDSGELVAIKKVLQDKRFK
20W3(B) SKVTTVVATPGQGPDRPQEVSYDTKVIENGSGVGVVYQAKLCDSGELVAIKKVLQDKRFK
kubra  -KVTTVVATPGQGPDRPQEVSYDTKVIENGSGVGVVYQAKLCDSGELVAIKKVLQDKRFK
nr(Pf) 58.....70.....90.....110.....
pfc0525c SEDEDEERTIDNEINRSPNKSYKLGNIIGNGSFGVVVEAICIDTSEQAIAIKKVLQDPQYK

95.....110.....130.....150..
1J1B(B) NRELQIMRKL DHCNIVRLRYFFYSSGEKKDE--VYLNVLVDYVPETVYRVARHYSRAKQT
1Q3D(B) NRELQIMRKL DHCNIVRLRYFFYSS---DE--VYLNVLVDYVPETVYRVARHYSRAKQT
1Q3W   NRELQIMRKL DHCNIVRLRYFFYSSG---E--VYLNVLVDYVPETVYRVARHYSRAKQT
1Q4L   NRELQIMRKL DHCNIVRLRYFFYSSG---DE--VYLNVLVDYVPETVYRVARHYSRAKQT
1Q41(B) NRELQIMRKL DHCNIVRLRYFFYSS-----VYLNVLVDYVPETVYRVARHYSRAKQT
1Q5K(B) NRELQIMRKL DHCNIVRLRYFFYSSGEKKDE--VYLNVLVDYVPETVYRVARHYSRAKQT
1R0E   NRELQIMRKL DHCNIVRLRYFFYSS-----E--VYLNVLVDYVPETVYRVARHYSRAKQT
1UV5   NRELQIMRKL DHCNIVRLRYFFYSSGEKKDE--VYLNVLVDYVPETVYRVARHYSRAKQT
205K   NRELQIMRKL DHCNIVRLRYFFYSSGEKKDE--VYLNVLVDYVPETVYRVARHYSRAKQT
20W3(B) NRELQIMRKL DHCNIVRLRYFFYSSGEKKDE--VYLNVLVDYVPETVYRVARHYSRAKQT
kubra  NRELQIMRKL DHCNIVRLRYFFYSSGAAKDA--VYLNVLVDYVPETVYRVARHYSRAKQT
118.....130.....150.....170.....
pfc0525c NRELMIMKLNHNHINIIYLKDYIYTESFKKNEKNIFLNVVMEYIPQTVHKYMKYYSRNNQA
```

	153.....170.....190.....210..
1J1B(B)	LPVIYVKLYMYQLFRSLAYIHSFGICHRDIKPQNLLDPDTAVLKLCDFGSAKQLVRGEP
1Q3D(B)	LPVIYVKLYMYQLFRSLAYIHSFGICHRDIKPQNLLDPDTAVLKLCDFGSAKQLVRGEP
1Q3W	LPVIYVKLYMYQLFRSLAYIHSFGICHRDIKPQNLLDPDTAVLKLCDFGSAKQLVRGEP
1Q4L	LPVIYVKLYMYQLFRSLAYIHSFGICHRDIKPQNLLDPDTAVLKLCDFGSAKQLVRGEP
1Q41(B)	LPVIYVKLYMYQLFRSLAYIHSFGICHRDIKPQNLLDPDTAVLKLCDFGSAKQLVRGEP
1Q5K(B)	LPVIYVKLYMYQLFRSLAYIHSFGICHRDIKPQNLLDPDTAVLKLCDFGSAKQLVRGEP
1R0E	LPVIYVKLYMYQLFRSLAYIHSFGICHRDIKPQNLLDPDTAVLKLCDFGSAKQLVRGEP
1UV5	LPVIYVKLYMYQLFRSLAYIHSFGICHRDIKPQNLLDPDTAVLKLCDFGSAKQLVRGEP
205K	LPVIYVKLYMYQLFRSLAYIHSFGICHRDIKPQNLLDPDTAVLKLCDFGSAKQLVRGEP
20W3(B)	LPVIYVKLYMYQLFRSLAYIHSFGICHRDIKPQNLLDPDTAVLKLCDFGSAKQLVRGEP
kubra	LPVIYVKLYMYQLFRSLAYIHSFGICHRDIKPQNLLDPDTAVLKLCDFGSAKQLVAGAA
	178.....200.....220.....237
pfc0525c	LPMFLVKLYSYQLCRALSYIHSKFIHRDLKPQNLLIDPRHTLKLCDFGSAKNLLAGQR
	213.....230.....250.....270..
1J1B(B)	NVSYICSRYYRAPELIFGATDYTSSIDVWSAGCVLAELLLGQPIFPGDSGVDQLVEIIKV
1Q3D(B)	NVSYICSRYYRAPELIFGATDYTSSIDVWSAGCVLAELLLGQPIFPGDSGVDQLVEIIKV
1Q3W	NVSYICSRYYRAPELIFGATDYTSSIDVWSAGCVLAELLLGQPIFPGDSGVDQLVEIIKV
1Q4L	NVSYICSRYYRAPELIFGATDYTSSIDVWSAGCVLAELLLGQPIFPGDSGVDQLVEIIKV
1Q41(B)	NVSYICSRYYRAPELIFGATDYTSSIDVWSAGCVLAELLLGQPIFPGDSGVDQLVEIIKV
1Q5K(B)	NVSYICSRYYRAPELIFGATDYTSSIDVWSAGCVLAELLLGQPIFPGDSGVDQLVEIIKV
1R0E	NVSYICSRYYRAPELIFGATDYTSSIDVWSAGCVLAELLLGQPIFPGDSGVDQLVEIIKV
1UV5	NVSYICSRYYRAPELIFGATDYTSSIDVWSAGCVLAELLLGQPIFPGDSGVDQLVEIIKV
205K	NVSYICSRYYRAPELIFGATDYTSSIDVWSAGCVLAELLLGQPIFPGDSGVDQLVEIIKV
20W3(B)	NVSYICSRYYRAPELIFGATDYTSSIDVWSAGCVLAELLLGQPIFPGDSGVDQLVEIIKV
kubra	NVSYICSRYYRAPELIFGATAYTSSIDVWSAGCVLAELLLGQPIFPGDSGVDQLVEIIKV
	238.....260.....280.....297
pfc0525c	SVSYICSRFYRAPELMLGSTNYTTHIDLWSLGCIIAEMILGYPIFSGQSSVDQLVRIIQV
	273.....290.....310.....330..
1J1B(B)	LGTPTREQUIREMNPNYTEFKFPQIKAHPWTKVFRPRTPPEAIALCSRLLEYTPTARLTPL
1Q3D(B)	LGTPTREQUIREMNP-----FPQIKAHPWTKVFRPRTPPEAIALCSRLLEYTPTARLTPL
1Q3W	LGTPTREQUIREMNP----KFPQIKAHPWTKVFRPRTPPEAIALCSRLLEYTPTARLTPL
1Q4L	LGTPTREQUIREMNP-----FPQIKAHPWTKVFRPRTPPEAIALCSRLLEYTPTARLTPL
1Q41(B)	LGTPTREQUIREMNPNYTEFKFPQIKAHPWTKVFRPRTPPEAIALCSRLLEYTPTARLTPL
1Q5K(B)	LGTPTREQUIREMNP---EFKFP---AHPWTKVFRPRTPPEAIALCSRLLEYTPTARLTPL
1R0E	LGTPTREQUIREMNPNYTEFKFPQIKAHPWTKVFRPRTPPEAIALCSRLLEYTPTARLTPL
1UV5	LGTPTREQUIREMNPNYTEFKFPQIKAHPWTKVFRPRTPPEAIALCSRLLEYTPTARLTPL
205K	LGTPTREQUIREMNPNYTEFKFPQIKAHPWTKVFRPRTPPEAIALCSRLLEYTPTARLTPL
20W3(B)	LGTPTREQUIREMNPNYTEFKFPQIKAHPWTKVFRPRTPPEAIALCSRLLEYTPTARLTPL
kubra	LGTPTAEQIAEMNPNTAAAFPAIAAAPWTKVFRPATPPEAIALCSRLLEYTPTARLTPL
	298.....320.....340.....
pfc0525c	LGTPTEDQLKEMNPYADIKFPDVKSKDLRKFVFKGTPDEAINLITQFLKYEPLKRLNPI

```

330.....350.....370.....
1J1B(B) EACAHSFFDELSDPNVKLPNGRD-TPALFNFTTQELSSNPPLAT-ILIPPHARIQA
1Q3D(B) EACAHSFFDELSDPNVKLPNGRD-TPALFNFTTQELSSNPPLAT-ILIPPHAR---
1Q3W EACAHSFFDELSDPNVKLPNGRD-TPALFNFTTQELSSNPPLAT-ILIPPHARIQA
1Q4L EACAHSFFDELSDPNVKLPNGRD-TPALFNFTTQELSSNPPLAT-ILIPPHARIQA
1Q41(B) EACAHSFFDELSDPNVKLPNGRD-TPALFNFTTQELSSNPPLAT-ILIPPHAR---
1Q5K(B) EACAHSFFDELSDPNVKLPNGRD-TPALFNFTTQELSSNPPLAT-ILIPPHARI--
1R0E EACAHSFFDELSDPNVKLPNGRD-TPALFNFTTQELSSNPPLAT-ILIPPHAR---
1UV5 EACAHSFFDELSDPNVKLPNGRD-TPALFNFTTQELSSNPPLAT-ILIPPHAR---
205K EACAHSFFDELSDPNVKLPNGRD-TPALFNFTTQELSSNPPLAT-ILIPPHARI--
20W3(B) EACAHSFFDELSDPNVKLPNGRD-TPALFNFTTQELSSNPPLAT-ILIPPHARIQA
kubra EACAHSFFDELSDPNVKLPNGRD-TPALFNFTTQELSSNPPLAT-ILIPPHAAI--
358.....380.....400.....
pfc0525c EALADPFFDELSDPCIKLPKYIDKLPFLNFCKEEIQEMSMCCRKIIPKNVYEEF

```

An den Positionen 121, 122, 125, 209, 211-212, 233, 278, 282, 290-292, 295, 297-299, 308 und 383 der Sequenz von kubra sind in den Rohdaten jeweils Alanine angegeben. Es ist davon auszugehen, dass hier die Seitenketten nicht vollständig aufgelöst waren.

Wie in Abschnitt 2.5.1 dargestellt wurde das Alignment für die Homologie-Modelle paarweise erstellt. Dabei wurden für die letzten Aminosäuren abweichende Sequenzalignments bei unterschiedlichen Vorlagen berechnet und als Grundlage bei der Erstellung der Modelle verwendet.

```

330.....350.....370.....
1J1B(B) EACAHSFFDELSDPNVKLPNG-RDTPALFNFTTQELSSNPPLATILIPPHARIQ-A
1Q3D(B) EACAHSFFDELSDPNVKLPN-GRDTPALFNFTTQELSSNPPLATILIPPHAR----
1Q3W EACAHSFFDELSDPNVKLPNG-RDTPALFNFTTQELSSNPPLATILIPPHARIQ-A
1Q4L EACAHSFFDELSDPNVKLPNG-RDTPALFNFTTQELSSNPPLATILIPPHARIQ-A
1Q41(B) EACAHSFFDELSDPNVKLPN-GRDTPALFNFTTQELSSNPPLATILIPPHAR----
1Q5K(B) EACAHSFFDELSDPNVKLPNG-RDTPALFNFT---TQELSSNPPLATILIPPHARI
1R0E EACAHSFFDELSDPNVKLPN-GRDTPALFNFTTQELSSNPPLATILIPP----HAR
1UV5 EACAHSFFDELSDPNVKLPN-GRDTPALFNFTTQELSSNPPLATILIPP----HAR
205K EACAHSFFDELSDPNVKLPNGR-DTPALFNFTTQELSSNPPLATILIPPHARI---
20W3(B) EACAHSFFDELSDPNVKLPN-GRDTPALFNFTTQELSSNPPLATILIPPHARIQA-
kubra EACAHSFFDELSDPNVKLPN-GRDTPALFNFTTQELSSNPPLATILIPPHAAI---
358.....380.....400.....
pfc0525c EALADPFFDELSDPCIKLPKYIDKLPFLNFCKEEIQEMSMCCRKIIPKNVYEEF

```

Die betreffenden Aminosäuren liegen mehr als 30 Å von der Bindetasche entfernt C-terminal am untersten Ende des Proteins. Die Unterschiede führten nicht zu problematischen Strukturen, was einerseits durch visuelle Analyse des C-terminalen Bereiches der Modelle und andererseits durch die Ergebnissen der Validierung mittels *RCS* sichergestellt wurde.

A.2. Abbildungen

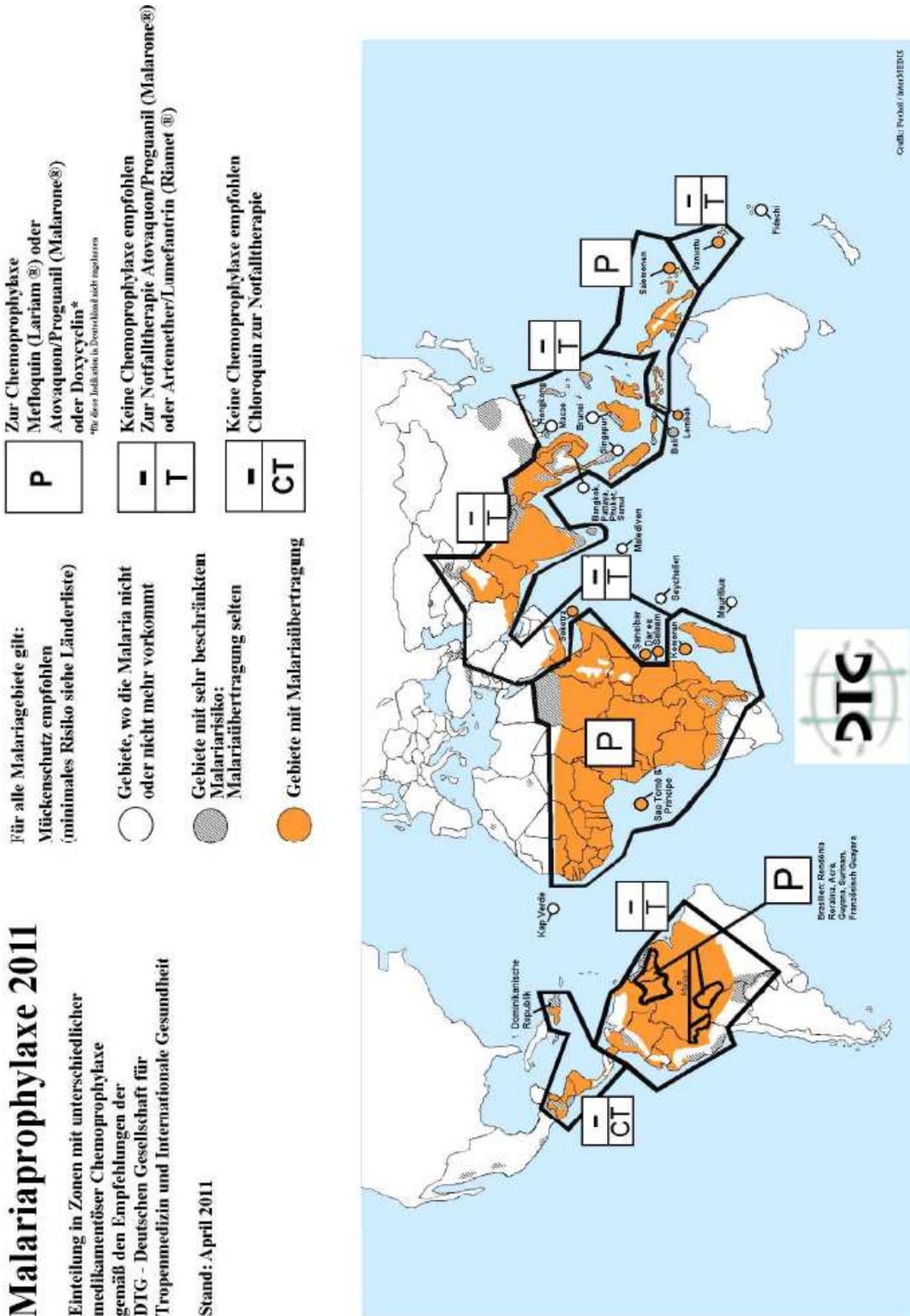


Abbildung A.1. | Aktuelle Empfehlungen der Malariaphylaxe nach DTG [30].

B. Tabellen

B.1. Validierung der Vorlagen und Homologie-Modelle

In den Tabellen B.1 und B.2 sind die Werte der einzelnen Kriterien der Evaluierung aufgelistet, die zur Berechnung des *RCS* gemäß Gleichung 2.1 herangezogen wurden.

Tabelle B.1. | Evaluierungsergebnisse der Vorlagen.

PDB- Eintrag	Ramachandran				bad contacts		G-factor			ProSa2003
	core	allowed	gen. all.	disall.	Σ	BT	GF_{ov}	GF_{dih}	GF_{cov}	z-score
1J1B	85.1	13.9	0.3	0.6	6	0	0.29	0.11	0.56	-10.33
1Q3D	87.3	11.0	0.3	1.4	2	0	0.29	0.17	0.46	-9.35
1Q3W	86.8	12.2	0.0	1.0	5	1	0.28	0.15	0.47	-9.62
1Q41	88.9	10.5	0.3	0.3	4	0	0.28	0.17	0.43	-9.93
1Q4L	87.8	10.5	1.0	0.7	4	0	0.28	0.14	0.48	-9.58
1Q5K	87.8	9.8	1.7	0.7	4	1	0.34	0.17	0.57	-9.42
1ROE	89.6	9.4	0.3	0.7	1	0	0.07	-0.19	0.46	-9.92
1UV5	83.8	12.6	2.3	1.3	8	0	0.03	-0.04	0.14	-8.70
205K	71.1	25.3	2.6	1.0	16	3	0.07	-0.20	0.48	-9.39
20W3	84.2	14.1	1.3	0.3	3	0	0.05	-0.28	0.56	-9.57
kubra	84.5	14.1	0.3	1.0	4	0	0.22	-0.08	0.67	-9.18

gen. all.: generously allowed, disall.: disallowed; BT: ATP-Bindetasche; GF_{ov} : G-factor (overall); GF_{dih} : G-factor (dihedral); GF_{cov} : G-factor (covalent).

Tabelle B.2. | Evaluierungsergebnisse der Homologie-Modelle. Es ist jeweils das beste Modell im Rahmen der vergleichenden Bewertung innerhalb der 200 generierten Strukturen pro Vorlage beschrieben. Einen Eindruck von dessen Qualität im Vergleich zu den anderen Modellen gibt die Platzierung gemäß *RCS* im Vergleich aller 4400 Proteine (*RCS* overall).

Homologie-Modell	Ramachandran				bad contacts		G-factor			ProSa2003	MODELLER	<i>RCS</i>
	core	allowed	gen. all.	disall.	Σ	BT	GF_{ov}	GF_{dih}	GF_{cov}	z-score	pdf	overall
pfgsk_1j1b_128	91.0	7.8	0.9	0.3	1	0	0.00	0.06	-0.10	-8.76	1745.97	1
pfgsk_1q3d_41	88.5	10.2	0.6	0.6	6	0	-0.04	0.03	-0.18	-8.49	2038.89	385
pfgsk_1q3w_196	89.8	8.7	1.2	0.3	6	1	-0.02	0.05	-0.15	-8.43	1915.15	387
pfgsk_1q41_72	90.1	8.7	0.9	0.3	7	0	0.02	0.09	-0.12	-8.99	1733.91	2
pfgsk_1q41_188	89.1	9.6	0.9	0.3	4	1	-0.04	0.03	-0.16	-8.36	2187.32	700
pfgsk_1q5k_160	88.5	9.3	1.9	0.3	9	0	-0.08	0.02	-0.26	-8.87	2449.73	447
pfgsk_1r0e_71	91.0	7.8	0.9	0.3	6	1	-0.03	0.05	-0.17	-8.58	1858.69	254
pfgsk_1uv5_41	88.5	9.0	1.2	1.2	4	0	-0.05	0.01	-0.17	-8.78	1987.52	426
pfgsk_2o5k_52	83.9	13.7	2.5	0.0	6	1	-0.11	-0.07	-0.18	-7.83	2217.37	1664
pfgsk_2ow3_78	89.4	9.9	0.3	0.3	9	0	-0.04	0.00	-0.10	-8.89	1954.49	379
pfgsk_kubra_197	90.1	8.4	0.6	0.9	7	1	-0.03	0.03	-0.13	-9.02	2056.96	439
pfgsk_1j1b_ANP_27	91.0	8.4	0.3	0.3	8	0	-0.08	-0.04	-0.16	-8.83	2319.71	334
pfgsk_1q3d_STU_53	87.3	10.9	1.2	0.6	11	1	-0.12	-0.07	-0.22	-8.68	2479.44	1879
pfgsk_1q3w_ATU_155	89.8	9.0	0.9	0.3	15	0	-0.10	-0.05	-0.19	-8.34	2379.59	1370
pfgsk_1q41_IXM_120	90.7	7.8	0.9	0.6	5	0	-0.09	-0.04	-0.19	-8.78	2150.23	325
pfgsk_1q41_679_68	89.4	8.7	1.2	0.6	9	1	-0.10	-0.05	-0.19	-8.30	2660.65	2393
pfgsk_1q5k_TMU_24	87.0	10.6	2.2	0.3	9	0	-0.18	-0.09	-0.36	-8.49	2909.62	3740
pfgsk_1r0e_DFN_83	90.4	8.4	0.9	0.3	9	1	-0.13	-0.08	-0.23	-8.82	2296.02	956
pfgsk_1uv5_BRW_31	89.4	9.0	0.3	1.2	9	0	-0.11	-0.06	-0.21	-8.63	2428.47	1669
pfgsk_2o5k_HBM_88	84.5	12.4	2.5	0.6	11	0	-0.20	-0.18	-0.24	-7.94	2677.90	3334
pfgsk_2ow3_BIM_187	89.4	9.0	0.9	0.6	11	0	-0.13	-0.11	-0.18	-8.61	2420.11	1419
pfgsk_kubra_DRG_143	88.8	9.6	0.6	0.9	10	1	0.28	0.55	-0.17	-8.62	2407.44	1297

B.2. Übersicht der Strukturen und biologischen Daten

Tabelle B.3 gibt einen vollständigen Überblick über Verbindungen und deren biologische Daten. Es liegen für gleiche Verbindungen teilweise mehrere Werte vor. Der chronologisch erste stammt aus dem Screening, bei dem die Verbindung 80083874 (**40**) als Hit identifiziert wurde. Gemeinsam mit den Werten aus dem Screening lagen biologische Daten für den größten Teil der durch W. BRANDT synthetisierten Verbindungen [92] zu Beginn der Promotion Anfang 2008 vor.

Ein zweiter Datensatz wurde im Frühjahr 2009 zugänglich. Die Werte entsprechen der Messreihe, deren Ergebnisse in der zitierten Dissertation veröffentlicht sind. Verbindungen, die ergänzend zum ersten Datensatz hinzukamen, sind in der Tabelle durch <Verbindung>^x gekennzeichnet. Zu den später in den Datensatz aufgenommenen Verbindungen zählen auch die in Abschnitt 3.4 beschriebenen Verbindungen **101** bis **104**, die im Rahmen der vorliegenden Arbeit in Braunschweig dargestellt werden konnten.

Unterscheiden sich die beiden Werten, ist der chronologisch zunächst bekanntgewordene zur Information in Klammern mit angegeben. *Der jeweils neueste Wert wurde für die QSAR-Berechnungen herangezogen* (eine Ausnahme bildet diesbezüglich **77**, da hier eine isolierte Wiederholmessung durchgeführt wurde, so dass für die QSAR-Berechnungen der im Zusammenhang mit den anderen ermittelte Wert angenommen wurde).

In diesem Zusammenhang sind auch die Daten der 2-brom-4-arylsubstituierten Verbindung **73** zu diskutieren. In den zunächst vorliegenden Daten wurde die Verbindung im Vergleich zur analogen 2-iodsubstituierten Hitstruktur 80083874 (**40**) als weniger aktiv gemessen, obgleich die in Abschnitt 3.2.1.1 geschilderten Struktur-Wirkungsbeziehungen die Substitution als vorteilhafter erscheinen lassen. In der zweiten Messreihe im Rahmen der Bestimmung von IC₅₀-Werten der übrigen Thienopyridine wurde überhaupt keine Aktivität der Verbindung gemessen. Die biologischen Daten von **73** wurden wegen ihrer widersprüchlichen experimentellen Befunde weder in Berechnungen noch in die Diskussion der Struktur-Wirkungsbeziehungen einbezogen.

Tabelle B.3. | Biologische Daten, IC₅₀ an PFGSK und SsGSK in μ M.

Verbindung	Struktur	R ¹	R ²	PfGSK	SsGSK
56 ^a		Phenyl	CN	0.4	2.5
57		2-Iod-phenyl	CN	>10	>10
58		2-Brom-phenyl	CN	0.8	1.3
31		2-Chlor-phenyl	CN	0.13	0.4
59		2-Methoxy-phenyl	CN	>10	>10
60		2-Ethoxy-phenyl	CN	>10	>10
61		2-Ethoxy-phenyl	COOCH ₃	>100	>100
62 ^a		2-Methyl-phenyl	CN	0.51	2
63		Thien-2-yl	CN	0.18	2.3
64		Indol-3-yl	CN	1.6	3.5
65		H	CN	>10	>10
40		2-Iod-phenyl	4-Chlor-benzoyl	1.6 (0.45)	>10 (>100)
66		2-Iod-phenyl	3-Chlor-benzoyl	0.91	>10
67		2-Iod-phenyl	2-Chlor-benzoyl	>10	>10
68		2-Iod-phenyl	3,4-Dichlor-benzoyl	70	>100
69		2-Iod-phenyl	4-Methoxy-benzoyl	1.7 (3.2)	>10
70		2-Iod-phenyl	4-Trifluormethyl-benzoyl	>10 (0.91)	>10
71		2-Iod-phenyl	4-Methyl-benzoyl	3.2	>10
72		2-Iod-phenyl	Benzoyl	2.4	>10
73		2-Brom-phenyl	4-Chlor-benzoyl	>100 (1.7)	>100
74		2-Brom-phenyl	3-Chlor-benzoyl	0.61	>10
75		2-Brom-phenyl	3,4-Dichlor-benzoyl	3.8	>10
76		2-Brom-phenyl	Benzoyl	1.1 (2.7)	>10
77 ^b		2-Chlor-phenyl	3-Chlor-benzoyl	0.48 [0.06]	>10
78		2-Chlor-phenyl	3,4-Dichlor-benzoyl	2.7	>10
79		2-Chlor-phenyl	3-Brom-benzoyl	0.91 (0.48)	>10
80		2-Chlor-phenyl	3-Methoxy-benzoyl	1.7	>10
81 ^x		2-Chlor-phenyl	3-Trifluormethyl-benzoyl	2.2	>10
82 ^x		2-Chlor-phenyl	3-Fluor-benzoyl	>100	>100
83 ^x		2-Chlor-phenyl	3-Cyano-benzoyl	0.5	>10

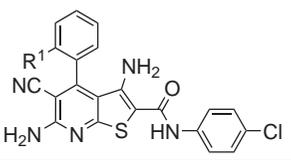
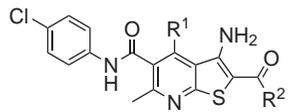
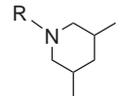
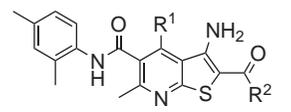
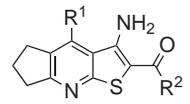
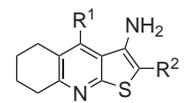
Fortsetzung auf der nächsten Seite. . .

Tabelle B.3 – Fortsetzung

Verbindung	Struktur	R ¹	R ²	PfGSK	SsGSK
84 ^x		2-Fluor-phenyl	3-Chlor-benzoyl	10	>100
85 ^a		2-Methoxy-phenyl	Benzoyl	6	>100
86		2-Methoxy-phenyl	2-Chlor-benzoyl	>10	>10
87		2-Methoxy-phenyl	3-Chlor-benzoyl	8.5	>100
88		2-Methoxy-phenyl	4-Chlor-benzoyl	>10	>10
89		2-Methoxy-phenyl	4-Methoxy-benzoyl	>10	>10
90		2-Methoxy-phenyl	4-Methyl-benzoyl	>10	>10
91		2-Methoxy-phenyl	4-Trifluormethyl-benzoyl	>10	>10
92		2-Methoxy-phenyl	4-Cyano-benzoyl	>10	>10
93		4-Ethoxy-phenyl	Benzoyl	>100	>100
94		4-Ethoxy-phenyl	4-Chlor-benzoyl	>100	>100
95		2-Ethoxy-phenyl	4-Chlor-benzoyl	>10	>10
96		4-Methyl-phenyl	4-Chlor-benzoyl	>100	>100
97		Phenyl	Benzoyl	>100	>100
98		Thien-2-yl	4-Brom-benzoyl	>100	>100
99 ^x		Thien-2-yl	3-Chlor-benzoyl	>10	>10
100		Furan-2-yl	4-Chlor-benzoyl	100	>100
101		2-Iod-phenyl	Thien-2-yl-carbonyl	11.5	>100
102		2-Iod-phenyl	(5-Cyano-thien-2-yl)carbonyl	68	>100
103		2-Chlor-phenyl	Thien-2-yl-carbonyl	60	>100
104		2-Chlor-phenyl	(5-Cyano-thien-2-yl)carbonyl	50	>100
105 ^x		2,6-Dichlor-phenyl	3-Chlor-benzoyl	2.2	>100
106		Phenyl	4-Chlor-benzoyl	>100	>100
107 ^{c/x}				≈10	>10
108 ^x				0.13	0.06
109		siehe Abbildung B.1		>10	>10
110 ^x				>100	>100
111 ^x				>10	>10

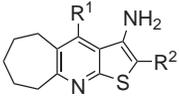
Fortsetzung auf der nächsten Seite. . .

Tabelle B.3 – Fortsetzung

Verbindung	Struktur	R ¹	R ²	PfGSK	SsGSK
112		Cl		>10	>10
113		Br		>10	>10
114		I		>10	>10
115		4-Bromphenyl	-NH ₂	4	9
116		4-Bromphenyl		>100	>100
117		Furan-2-yl	-NH ₂	0.8	5
118		Pyridin-3-yl	2-Methoxy-phenylamin	>100	>100
119		4-Chlor-phenyl	4-Ethyl-phenylamin	>100	>100
120		4-Brom-phenyl	4-Fluor-phenylamin	>100	>100
121		Thien-2-yl	Phenyl	43	>100
122		Thien-2-yl	4-Chlor-phenyl	>100	>100
123		Thien-2-yl	4-Methoxy-phenyl	>100	>100
124		Thien-2-yl	-OC ₃ H ₇	>100	>100
125		2-Chlor-phenyl	Thien-2-yl	30	>100
126		4-Chlor-phenyl	4-Chlor-phenylamin	>100	>100
127		4-Methyl-phenyl	-NH ₂	3	4.3
128		Pyridin-3-yl	4-Fluor-benzoyl	16 (6)	>100
129		2-Iod-phenyl	4-Chlor-benzoyl	>10	>10
130		2-Iod-phenyl	Benzoyl	>10	>10
131		2-Brom-phenyl	Benzoyl	>10	>10
132		2-Chlor-phenyl	Carboxamid	3 (0.43)	3 (0.43)
133		2-Chlor-phenyl	Thien-2-yl-carbonyl	>100	>100
134		2-Chlor-phenyl	COOCH ₃	>100	>100
135		4-Chlor-phenyl	Carboxamid	4	4
136		4-Methyl-phenyl	Carboxamid	>100	>100

Fortsetzung auf der nächsten Seite...

Tabelle B.3 – Fortsetzung

Verbindung	Struktur	R ¹	R ²	PfGSK	SsGSK
137		Thien-2-yl	Benzoyl	1.8	>100
138^a	wie 128-136	Thien-2-yl	CN	0.8	2
139		Furan-2-yl	Carboxamid	1	3.3
140		Thien-2-yl	Carboxamid	1.1	>100
141		2-Chlor-phenyl	Carboxamid	3.8 (0.3)	9 (1.4)
142		4-Chlor-phenyl	Carboxamid	15 (2.3)	80 (8.5)
143		4-Methyl-phenyl	Carboxamid	3	>100

^aBestimmung in 10 % DMSO Endkonzentration; ^bEckig eingeklammerter Wert aus Wiederholungsmessung im zeitlichen Abstand von 10 Monaten zur ersten Messung; ^cKein IC₅₀ Wert, Bestimmung der Restaktivität bei einer Konzentration von 10 μM: ≈50 %.

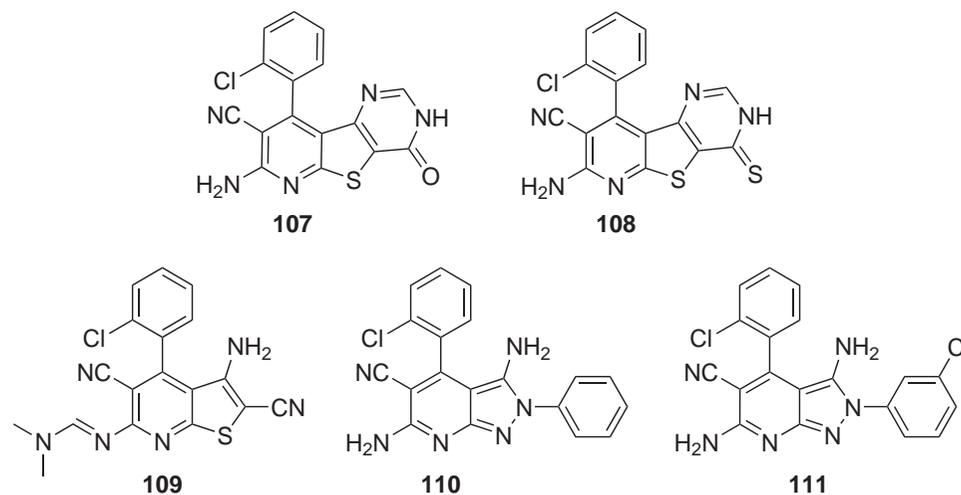


Abbildung B.1. | Strukturen zu Tabelle B.3

B.3. Docking

Ergebnisse der Dockings im Rahmen der Evaluierung

Auf den Seiten 239 bis 241 sind die Übersichten der einzelnen Dockings im Rahmen der Evaluierung des Verfahrens DoReCCoM dargestellt. Es ist in der oberen Tabelle für jede Paarung Protein-Ligand im Crossdocking der rmsd der am besten bewerteten Lösung zur Kristallstruktur angegeben. In der unteren Tabelle werden die Werte der Lösungen dargestellt, die den kleinsten rmsd haben, außerdem ist ihre Platzierung durch die jeweilige Scoringfunktion angegeben (linke Seite: generische Funktion des Programmes, rechte Seite: DrugScoreX).

Wichtige Kennzahlen bei der Auswertung waren die auf Seite 125 angegebenen. Die Farbkodierung der unteren Tabelle gibt einen ersten visuellen Eindruck hinsichtlich zweier Parameter. Die erste der zwei Spalten pro Ligand ist grün unterlegt, wenn der beste berechnete rmsd kleiner als 2 Å ist. Wenn eine solche gute Lösung zudem innerhalb der ersten 10 Lösungen bewertet wird, ist die daneben liegenden zweite Spalte ebenfalls grün unterlegt. In jedem anderen Fall sind die Felder rot gefärbt.

In Tabelle B.7 sind die zusammengefassten Ergebnisse nach Rescoring mit DrugScoreX dargestellt. Nach dem Clustering ist bereits auf den ersten Blick zu erkennen, dass mehr beste Lösungen mit gutem rmsd unter den ersten 10 zu finden sind (grüne Felder in der mittleren Tabelle). In der unteren Tabelle sind zusätzlich die Anzahl der Lösungen aller drei Programme angegeben (links) bzw. die Anzahl der Cluster (rechts).

In allen Tabellen sind die Redockings eingerahmt. Hier ist ein besonders gutes Ergebnis zu erwarten, da das Protein auf den entsprechenden Liganden präformiert ist. Durch den Ausschluss einiger Liganden (vgl. Abschnitt 3.2.2.3 auf S. 92) ergibt sich die nicht quadratische Matrix.

Tabelle B.4. | Ergebnisse des Dockings mit FlexX, Auswirkungen des Rescorings mit DrugScoreX.

rmsd der besten Lösung gemäß FlexX-Score zur Kristallstruktur

Prot \ Lig	ATU	IXM	UN1	DFN	HBM	UN2	UN3	HT3	DRG	MW	<2
hs1j1b	7.24	6.28	2.95	1.61	5.81	6.21	8.94	6.66	5.56	5.69	1
hs1q3d	6.61	6.44	5.85	5.82	2.97	9.42	6.42	6.93	3.66	6.01	0
hs1q3w	0.75	6.34	5.81	2.75	5.97	6.35	5.59	1.98	5.46	4.55	2
hs1q41	1.38	1.84	3.47	1.67	2.78	6.33	7.01	6.73	2.85	3.78	3
hs1q4l	2.10	2.61	2.86	2.73	9.06	5.98	6.56	2.21	5.83	4.44	0
hs1r0e	1.42	6.42	5.49	2.76	3.14	5.99	6.51	6.61	5.03	4.82	1
hs2o5k	1.26	2.61	3.85	2.91	4.61	1.88	5.27	6.55	6.19	3.90	2
hs2ow3	2.46	2.50	6.57	4.25	6.75	6.12	5.82	6.33	4.50	5.03	0
hs3du8	1.31	2.04	5.77	4.50	4.13	3.36	8.92	6.68	5.66	4.71	1
hs3f7z	1.26	6.29	5.37	4.26	2.38	6.08	8.26	6.71	2.88	4.83	1
hs3f88	0.82	2.61	5.98	3.76	7.44	6.06	7.61	2.18	5.35	4.65	1
kubra	6.11	6.24	4.69	1.42	4.20	3.44	8.55	2.24	5.01	4.66	1
MW	2.72	4.35	4.89	3.20	4.93	5.60	7.12	5.15	4.83	4.76	
<2	7	1	0	3	0	1	0	1	0	13	

kleinster rmsd, Platzierung gemäß FlexX-Score (FX)

Prot \ Lig	ATU	FX	IXM	FX	UN1	FX	DFN	FX	HBM	FX	UN2	FX	UN3	FX	HT3	FX	DRG	FX
hs1j1b	0.59	36	0.78	69	2.46	169	1.17	2	5.56	40	2.24	44	4.26	4	1.80	68	2.37	6
hs1q3d	1.22	18	1.34	165	1.77	68	1.17	24	2.30	25	2.30	129	3.32	10	1.56	90	2.44	15
hs1q3w	0.72	2	0.73	121	2.90	71	1.03	116	5.46	178	5.73	55	5.05	122	1.42	10	3.66	21
hs1q41	1.27	3	0.41	25	1.98	137	1.17	49	1.93	102	2.50	204	5.14	139	1.48	93	2.85	1
hs1q4l	0.68	48	0.89	63	1.48	121	1.57	85	1.69	21	3.92	39	5.57	26	1.49	107	2.22	30
hs1r0e	0.64	7	0.60	17	0.77	124	0.85	150	2.08	5	5.51	126	5.08	55	2.91	59	1.40	9
hs2o5k	0.81	19	0.47	67	2.26	6	1.64	47	1.83	252	1.31	67	5.26	3	3.96	27	2.13	8
hs2ow3	1.01	55	0.89	196	1.80	42	2.59	37	2.55	98	1.85	218	5.03	22	6.32	2	2.67	27
hs3du8	0.59	8	1.11	67	0.97	113	1.32	18	1.85	211	2.82	5	5.71	334	1.33	98	3.22	35
hs3f7z	0.87	8	0.64	135	1.81	44	1.44	76	2.15	33	2.75	28	5.53	52	1.59	69	2.54	24
hs3f88	0.66	8	0.63	51	1.12	77	1.29	4	3.27	324	3.09	176	7.45	8	2.01	11	3.44	4
kubra	0.56	17	0.79	109	1.56	40	1.08	7	3.15	6	3.15	8	4.45	2	2.21	5	1.41	4
0.80	19	0.77	90	1.74	84	1.36	51	2.82	108	3.10	92	5.15	65	2.34	53	2.53	15	
12	6	12	0	9	0	11	3	4	0	2	0	0	0	7	1	2	2	
MW (rmsd best solution):	2.29			MW (pl. best solution):			64			Σ (rmsd<2 & pl<10):			12					

rmsd der besten Lösung gemäß DrugScoreX zur Kristallstruktur

Prot \ Lig	ATU	IXM	UN1	DFN	HBM	UN2	UN3	HT3	DRG	MW	<2
hs1j1b	0.65	0.97	2.47	5.55	9.25	3.24	8.03	6.51	2.82	4.39	2
hs1q3d	1.52	1.42	6.52	1.20	7.18	6.23	7.89	6.79	3.06	4.64	3
hs1q3w	0.75	2.14	4.88	2.75	5.90	6.03	6.35	1.77	5.03	3.95	2
hs1q41	1.38	0.47	3.07	5.65	4.09	6.11	5.42	6.49	5.20	4.21	2
hs1q4l	1.07	0.89	5.87	2.78	3.33	5.93	8.68	2.17	4.87	3.95	2
hs1r0e	1.69	6.51	6.39	0.88	5.80	5.84	5.33	6.32	6.24	5.00	2
hs2o5k	1.18	1.32	6.15	6.11	2.33	6.06	6.79	5.48	6.19	4.62	2
hs2ow3	2.07	0.96	6.20	5.90	6.79	6.12	5.37	6.81	4.66	4.99	1
hs3du8	1.21	1.63	3.19	1.74	3.64	6.01	8.54	6.49	4.24	4.08	3
hs3f7z	1.22	1.13	5.88	5.43	3.42	6.03	7.41	1.94	3.00	3.94	3
hs3f88	0.82	1.25	2.86	5.57	7.27	6.03	7.61	6.62	4.14	4.68	2
kubra	0.56	1.90	6.17	1.49	3.82	5.88	8.60	2.26	2.11	3.64	3
MW	1.17	1.71	4.97	3.75	5.23	5.79	7.17	4.97	4.30	4.34	
n<2	11	10	0	4	0	0	0	2	0	27	

kleinster rmsd, Platzierung gemäß DrugScoreX (DSX)

Prot \ Lig	ATU	DSX	IXM	DSX	UN1	DSX	DFN	DSX	HBM	DSX	UN2	DSX	UN3	DSX	HT3	DSX	DRG	DSX
hs1j1b	0.59	5	0.78	4	2.46	82	1.17	77	5.56	17	2.24	119	4.26	7	1.80	35	2.37	8
hs1q3d	1.22	14	1.34	8	1.77	56	1.17	50	2.30	59	2.30	46	3.32	35	1.56	41	2.44	9
hs1q3w	0.72	4	0.73	6	2.90	46	1.03	73	5.46	37	5.73	141	5.05	58	1.42	67	3.66	6
hs1q41	1.27	5	0.41	2	1.98	29	1.17	113	1.93	256	2.50	190	5.14	44	1.48	68	2.85	4
hs1q4l	0.68	5	0.89	1	1.48	92	1.57	67	1.69	12	3.92	113	5.57	34	1.49	194	2.22	30
hs1r0e	0.64	8	0.60	2	0.77	46	0.85	54	2.08	187	5.51	124	5.08	100	2.91	3	1.40	6
hs2o5k	0.81	36	0.47	52	2.26	7	1.64	64	1.83	105	1.31	91	5.26	107	3.96	9	2.13	10
hs2ow3	1.01	46	0.89	29	1.80	65	2.59	79	2.55	62	1.85	199	5.03	41	6.32	44	2.67	4
hs3du8	0.59	5	1.11	2	0.97	38	1.32	56	1.85	19	2.82	7	5.71	357	1.33	71	3.22	14
hs3f7z	0.87	3	0.64	48	1.81	24	1.44	11	2.15	373	2.75	58	5.53	37	1.59	90	2.54	21
hs3f88	0.66	3	0.63	10	1.12	34	1.29	39	3.27	452	3.09	100	7.45	8	2.01	14	3.44	7
kubra	0.56	1	0.79	41	1.56	10	1.08	27	3.15	32	3.15	130	4.45	3	2.21	3	1.41	3
0.80	11	0.77	17	1.74	44	1.36	59	2.82	134	3.10	110	5.15	69	2.34	53	2.53	10	
12	9	12	8	9	1	11	0	4	0	2	0	0	0	7	0	2	2	
MW (rmsd best solution):	2.29			MW (pl. best solution):			56			Σ (rmsd<2 & pl<10):			20					

DrugScoreX

Tabelle B.5. | Ergebnisse des Dockings mit AutoDock4, Auswirkungen des Rescorings mit DrugScoreX.

rmsd der besten Lösung gemäß AutoDock4-Score zur Kristallstruktur												
Prot \ Lig	ATU	IXM	UN1	DFN	HBM	UN2	UN3	HT3	DRG	MW	<2	
hs1j1b	1.46	1.11	3.31	5.31	4.96	0.34	5.87	6.61	3.31	3.59	3	
hs1q3d	2.37	5.22	2.82	5.59	6.16	6.04	6.09	6.83	3.54	4.96	0	
hs1q3w	1.05	0.53	5.71	1.21	6.29	0.93	6.00	1.87	5.44	3.22	5	
hs1q41	1.31	0.97	4.87	5.48	3.06	6.10	6.02	6.73	1.69	4.03	3	
hs1q4l	6.25	5.00	2.16	1.74	7.63	5.93	5.45	2.47	2.78	4.38	1	
hs1r0e	1.59	4.97	6.95	5.01	6.25	6.03	6.90	2.52	1.53	4.64	2	
hs2o5k	1.51	1.03	2.95	5.41	5.83	1.57	5.98	2.50	6.14	3.66	3	
hs2ow3	7.68	6.68	5.60	5.77	7.15	3.53	2.89	2.64	5.87	5.31	0	
hs3du8	1.68	1.53	3.25	5.61	6.57	0.92	5.66	6.62	2.20	3.78	3	
hs3f7z	1.34	4.96	5.72	2.70	3.06	6.08	5.63	6.48	4.64	4.51	1	
hs3f88	1.81	0.92	2.96	1.54	2.92	1.02	5.77	6.70	4.57	3.14	4	
kubra	1.15	0.66	4.08	2.66	5.36	3.39	5.15	2.61	2.02	3.01	2	
MW	2.43	2.80	4.20	4.00	5.44	3.49	5.62	4.55	3.64	4.02		
n<2	9	7	0	3	0	5	0	1	2			27

kleinster rmsd, Platzierung gemäß AutoDock4-Score (AD4)																			
Prot \ Lig	ATU	AD4	IXM	AD4	UN1	AD4	DFN	AD4	HBM	AD4	UN2	AD4	UN3	AD4	HT3	AD4	DRG	AD4	
hs1j1b	0.56	123	0.84	133	2.34	152	2.79	2	2.97	33	0.32	2	0.91	35	1.12	103	1.56	113	
hs1q3d	0.94	199	0.56	176	2.15	22	0.96	38	2.21	45	1.05	119	1.56	139	1.57	150	1.92	137	
hs1q3w	0.34	191	0.37	122	1.74	84	0.71	20	2.80	17	0.47	22	1.18	93	1.38	77	1.15	182	
hs1q41	0.80	144	0.41	165	2.48	39	1.61	23	2.68	93	0.50	67	0.66	90	1.44	166	1.47	92	
hs1q4l	1.72	172	1.00	187	1.09	27	1.74	1	3.78	11	1.48	151	4.20	47	1.89	137	2.53	194	
hs1r0e	0.82	196	0.39	164	2.17	17	0.90	101	2.28	79	1.50	45	1.58	70	2.28	11	1.01	160	
hs2o5k	1.08	188	0.64	189	2.42	46	1.60	69	1.30	40	1.00	30	1.96	45	1.36	165	1.63	137	
hs2ow3	1.89	131	0.63	163	1.50	86	0.87	73	2.59	61	0.82	181	1.85	5	1.47	107	1.49	147	
hs3du8	0.49	166	0.98	153	2.43	41	1.02	38	2.35	150	0.42	63	1.08	24	1.73	196	0.78	199	
hs3f7z	0.44	195	0.35	174	1.98	187	1.65	35	2.83	75	0.52	78	0.56	10	1.13	200	1.48	149	
hs3f88	1.04	190	0.56	163	2.31	21	1.20	6	2.39	55	0.82	39	1.22	95	1.80	109	0.90	146	
kubra	0.69	166	0.38	184	2.21	137	2.63	131	2.86	126	2.78	127	2.83	173	2.22	195	0.35	162	
0.90	172	0.59	164	2.07	72	1.47	45	2.59	65	0.97	77	1.63	69	1.62	135	1.36	152		
12	0	12	0	4	0	10	2	1	0	11	1	10	2	10	0	11	0		
MW (rmsd best solution):																			1.47
MW (pl. best solution):																			106
Σ (n<2 und pl<10):																			5

rmsd der besten Lösung gemäß DrugScoreX zur Kristallstruktur												
Prot \ Lig	ATU	IXM	UN1	DFN	HBM	UN2	UN3	HT3	DRG	MW	<2	
hs1j1b	0.95	0.86	2.60	7.13	6.10	5.55	5.37	1.21	3.18	3.66	3	
hs1q3d	2.19	5.10	3.28	5.61	5.88	6.02	5.39	6.81	3.48	4.86	0	
hs1q3w	0.72	0.73	3.18	2.84	5.62	5.73	4.22	2.11	5.13	3.36	2	
hs1q41	1.66	0.48	3.15	5.73	6.53	5.79	5.96	2.25	2.55	3.79	2	
hs1q4l	1.93	1.18	5.94	1.74	8.64	5.95	4.72	6.31	2.68	4.34	3	
hs1r0e	1.73	0.61	2.43	5.52	5.31	5.99	6.24	6.76	2.76	4.15	2	
hs2o5k	1.74	1.09	2.78	5.93	1.68	5.74	5.02	3.16	3.43	3.40	3	
hs2ow3	2.16	1.58	6.27	5.73	7.20	3.72	5.91	6.90	3.67	4.79	1	
hs3du8	1.46	1.18	2.72	6.23	2.45	0.56	5.61	6.37	4.33	3.44	3	
hs3f7z	1.70	0.86	2.60	5.64	5.67	5.97	5.19	6.38	4.25	4.25	2	
hs3f88	1.46	0.78	2.46	5.82	2.40	3.58	5.45	6.30	3.20	3.50	2	
kubra	1.15	0.41	4.51	6.88	5.44	6.08	4.61	2.69	0.87	3.63	3	
MW	1.57	1.24	3.49	5.40	5.24	5.06	5.31	4.77	3.29	3.93		
n<2	10	11	0	1	1	1	0	1	1			26

kleinster rmsd, Platzierung gemäß DrugScoreX (DSX)																			
Prot \ Lig	ATU	pl.	IXM	pl.	UN1	pl.	DFN	pl.	HBM	pl.	UN2	pl.	UN3	pl.	HT3	pl.	DRG	pl.	
hs1j1b	0.56	73	0.84	5	2.34	152	2.79	76	2.97	20	0.32	15	0.91	28	1.12	22	1.56	45	
hs1q3d	0.94	160	0.56	145	2.15	35	0.96	19	2.21	25	1.05	123	1.56	105	1.57	161	1.92	142	
hs1q3w	0.34	49	0.37	30	1.74	175	0.71	14	2.80	38	0.47	103	1.18	72	1.38	130	1.15	31	
hs1q41	0.80	14	0.41	27	2.48	23	1.61	2	2.68	72	0.50	50	0.66	182	1.44	53	1.47	132	
hs1q4l	1.72	3	1.00	26	1.09	94	1.74	1	3.78	90	1.48	184	4.20	2	1.89	60	2.53	22	
hs1r0e	0.82	190	0.39	37	2.17	4	0.90	32	2.28	37	1.50	182	1.58	171	2.28	101	1.01	128	
hs2o5k	1.08	87	0.64	3	2.42	4	1.60	55	1.30	26	1.00	188	1.96	113	1.36	112	1.63	116	
hs2ow3	1.89	87	0.63	144	1.50	50	0.87	22	2.59	43	0.82	160	1.85	149	1.47	94	1.49	171	
hs3du8	0.49	153	0.98	45	2.43	14	1.02	10	2.35	18	0.42	4	1.08	37	1.73	176	0.78	196	
hs3f7z	0.44	160	0.35	33	1.98	180	1.65	65	2.83	48	0.52	34	0.56	98	1.13	138	1.48	136	
hs3f88	1.04	63	0.56	112	2.31	6	1.20	28	2.39	18	0.82	36	1.22	101	1.80	122	0.90	106	
kubra	0.69	25	0.38	33	2.21	65	2.63	16	2.86	91	2.78	38	2.83	164	2.22	194	0.35	66	
0.90	89	0.59	53	2.07	67	1.47	28	2.59	44	0.97	93	1.63	102	1.62	114	1.36	108		
12	1	12	2	4	0	10	3	1	0	11	1	10	0	10	0	11	0		
MW (rmsd best solution):																			1.47
MW (pl. best solution):																			77
Σ (n<2 und pl<10):																			7

DrugScoreX

Tabelle B.6. | Ergebnisse des Dockings mit FRED, Auswirkungen des Rescorings mit DrugScoreX.

rmsd der besten Lösung gemäß ChemGauss-Score zur Kristallstruktur																
Prot \ Lig	ATU	IXM	UN1	DFN	HBM	UN2	UN3	HT3	DRG	MW	<2					
hs1j1b	6.65	5.87	6.85	6.22	4.98	5.92	5.95	6.70	1.66	5.64	1					
hs1q3d	7.24	1.73	2.56	2.40	5.66	5.61	6.22	2.22	3.57	4.13	1					
hs1q3w	0.42	6.00	2.92	3.05	6.77	5.91	5.74	6.69	5.42	4.77	1					
hs1q41	6.40	5.94	4.89	5.69	8.30	5.75	4.35	6.75	4.28	5.82	0					
hs1q4l	7.30	6.01	1.55	2.50	6.39	5.85	7.32	1.91	2.54	4.60	2					
hs1r0e	1.21	6.03	1.70	2.35	2.46	5.67	5.36	2.24	1.25	3.14	3					
hs2o5k	5.76	1.29	3.87	5.48	2.72	6.29	5.67	1.72	4.04	4.09	2					
hs2ow3	7.07	6.53	5.78	2.71	7.09	1.96	5.89	6.97	5.45	5.49	1					
hs3du8	1.29	5.93	2.81	2.97	3.00	5.96	5.51	3.40	1.45	3.59	2					
hs3f7z	0.79	5.94	2.79	2.95	5.95	5.98	5.41	5.42	1.62	4.10	2					
hs3f88	0.46	5.83	2.78	2.47	2.46	5.97	1.47	5.12	1.36	3.10	3					
kubra	6.52	5.95	4.28	4.31	5.45	6.13	5.97	4.97	1.81	5.04	1					
MW	4.26	5.25	3.57	3.59	5.10	5.58	5.41	4.51	2.87	4.46						
<2	5	2	2	0	0	1	1	2	6		19					

rmsd der besten Lösung gemäß DrugScoreX zur Kristallstruktur																
Prot \ Lig	ATU	IXM	UN1	DFN	HBM	UN2	UN3	HT3	DRG	MW	<2					
hs1j1b	6.65	0.65	6.96	7.08	6.04	6.15	2.47	6.69	1.86	4.95	2					
hs1q3d	6.70	1.73	2.56	5.53	6.87	6.03	5.28	2.22	3.72	4.51	1					
hs1q3w	0.42	0.71	2.92	4.01	5.27	5.56	4.87	1.61	5.19	3.40	3					
hs1q41	0.40	0.45	4.89	6.54	5.46	6.14	4.38	6.12	4.85	4.36	2					
hs1q4l	1.20	1.36	0.62	5.98	6.90	5.88	5.26	6.83	4.03	4.23	3					
hs1r0e	1.26	1.10	0.50	4.75	2.51	5.92	5.49	6.79	1.48	3.31	4					
hs2o5k	3.47	0.96	4.11	5.32	2.75	6.30	6.06	6.50	3.17	4.29	1					
hs2ow3	3.51	6.54	5.69	5.26	6.93	2.37	5.72	7.02	5.45	5.39	0					
hs3du8	1.28	1.33	2.85	3.70	2.72	0.87	1.70	6.71	3.68	2.76	4					
hs3f7z	0.85	0.95	1.94	2.62	2.99	1.15	5.12	2.32	4.00	2.44	4					
hs3f88	0.46	0.87	2.02	5.59	2.81	6.14	5.11	6.33	2.94	3.59	2					
kubra	6.52	0.80	4.66	5.37	5.23	6.13	5.71	2.76	4.93	4.68	1					
MW	2.73	1.45	3.31	5.15	4.71	4.89	4.76	5.16	3.77	3.99						
<2	7	11	3	0	0	2	1	1	2		27					

DrugScoreX

kleinster rmsd, Platzierung gemäß ChemGauss-Score (CG)																		
Prot \ Lig	ATU	CG	IXM	CG	UN1	CG	DFN	CG	HBM	CG	UN2	CG	UN3	CG	HT3	CG	DRG	CG
hs1j1b	0.54	12	0.65	42	2.16	38	3.32	88	2.62	87	0.66	24	1.58	31	1.60	48	1.20	92
hs1q3d	0.87	36	0.94	26	0.87	63	2.01	42	2.34	90	1.24	66	1.72	11	1.64	18	1.08	94
hs1q3w	0.30	2	0.56	6	0.65	88	2.41	13	2.62	84	1.26	90	1.58	19	1.54	24	1.31	96
hs1q41	0.40	9	0.45	4	0.62	51	2.80	7	2.89	49	0.72	38	1.43	5	1.46	42	1.69	15
hs1q4l	1.20	32	1.17	7	0.62	3	2.40	3	3.07	98	2.87	73	4.73	99	1.55	7	1.51	27
hs1r0e	0.68	9	0.95	6	0.50	38	2.35	1	2.25	13	2.12	86	4.75	96	1.53	16	0.89	28
hs2o5k	1.88	93	0.96	10	2.08	86	2.73	52	1.78	74	1.19	34	3.38	54	1.72	1	1.01	48
hs2ow3	1.15	37	0.80	62	0.89	46	1.38	11	2.71	91	1.27	79	2.63	20	1.53	63	1.05	66
hs3du8	0.41	6	1.07	29	0.98	88	2.36	2	2.61	60	0.68	65	1.33	39	1.50	40	1.33	29
hs3f7z	0.25	3	0.69	15	0.67	20	2.37	25	2.67	38	0.73	28	1.29	14	1.59	17	0.63	67
hs3f88	0.41	3	0.64	15	0.94	8	2.38	38	2.30	15	1.09	59	1.41	2	1.47	82	1.24	3
kubra	0.33	13	0.80	14	2.07	70	2.97	76	2.49	92	4.02	72	4.34	9	2.06	81	0.61	7
MW	0.70	21	0.81	20	1.09	50	2.46	30	2.53	66	1.49	60	2.51	33	1.60	37	1.13	48
MW (rmsd best solution):	1.59																	
MW (pl. best solution):	40																	
Σ (n<2 und pl<10):	19																	

kleinster rmsd, Platzierung gemäß DrugScoreX (DSX)																		
Prot \ Lig	ATU	DSX	IXM	DSX	UN1	DSX	DFN	DSX	HBM	DSX	UN2	DSX	UN3	DSX	HT3	DSX	DRG	DSX
hs1j1b	0.54	5	0.65	1	2.16	45	3.32	86	2.62	66	0.66	75	1.58	47	1.60	15	1.20	65
hs1q3d	0.87	32	0.94	8	0.87	22	2.01	55	2.34	72	1.24	38	1.72	15	1.64	13	1.08	84
hs1q3w	0.30	3	0.56	3	0.65	43	2.41	8	2.62	23	1.26	63	1.58	43	1.54	4	1.31	83
hs1q41	0.40	1	0.45	1	0.62	25	2.80	77	2.89	97	0.72	70	1.43	16	1.46	34	1.69	38
hs1q4l	1.20	1	1.17	2	0.62	1	2.40	22	3.07	84	2.87	52	4.73	3	1.55	14	1.51	77
hs1r0e	0.68	7	0.95	3	0.50	1	2.35	44	2.25	14	2.12	74	4.75	83	1.53	43	0.89	25
hs2o5k	1.88	15	0.96	1	2.08	14	2.73	97	1.78	79	1.19	64	3.38	53	1.72	17	1.01	38
hs2ow3	1.15	38	0.80	14	0.89	6	1.38	55	2.71	77	1.27	68	2.63	7	1.53	43	1.05	77
hs3du8	0.41	16	1.07	2	0.98	60	2.36	38	2.61	4	0.68	18	1.33	61	1.50	33	1.33	5
hs3f7z	0.25	9	0.69	3	0.67	34	2.37	60	2.67	32	0.73	31	1.29	93	1.59	27	0.63	44
hs3f88	0.41	2	0.64	3	0.94	6	2.38	79	2.30	10	1.09	61	1.41	92	1.47	8	1.24	8
kubra	0.33	3	0.80	1	2.07	12	2.97	82	2.49	62	4.02	72	4.34	3	2.06	39	0.61	18
MW	0.70	11	0.81	4	1.09	22	2.46	59	2.53	52	1.49	57	2.51	43	1.60	24	1.13	47
MW (rmsd best solution):	1.59																	
MW (pl. best solution):	35																	
Σ (n<2 und pl<10):	27																	

Tabelle B.7. | Ergebnisse aller drei Programme nach Rescoring mit DrugScoreX und anschließendes Clustering.

rmsd2cryst of top ranked solution																				
Prot \ Lig	ATU	IXM	UN1	DFN	HBM	UN2	UN3	HT3t	DRG	MW	<2									
hs1j1b	6.65	0.86	2.60	5.55	6.10	6.15	5.37	1.28	4.04	4.29	2									
hs1q3d	2.19	5.10	3.28	5.61	5.88	6.02	5.39	6.83	3.48	4.86	0									
hs1q3w	0.72	0.73	3.18	2.80	5.62	6.12	4.87	6.70	4.99	3.97	2									
hs1q41	0.70	0.48	3.07	5.65	6.53	6.14	5.96	6.34	2.55	4.16	2									
hs1q4l	1.07	1.18	5.87	2.78	8.64	5.95	4.72	6.46	2.68	4.37	2									
hs1r0e	1.73	0.61	2.43	1.03	2.51	5.99	6.24	6.86	1.48	3.21	4									
hs2o5k	6.85	1.09	2.78	5.93	1.68	5.74	5.02	3.16	3.43	3.96	2									
hs2ow3	3.51	1.58	6.27	5.90	6.93	3.72	5.91	7.04	5.45	5.15	1									
hs3du8	1.46	1.60	3.19	6.23	3.64	5.98	5.61	6.59	4.33	4.29	2									
hs3f7z	1.70	0.86	2.60	5.64	5.67	6.20	5.19	4.39	4.69	4.10	2									
hs3f88	1.46	0.78	2.46	5.56	2.40	6.03	5.45	6.53	3.20	3.77	2									
kubra	1.15	1.86	4.51	1.49	5.44	6.00	4.61	6.39	4.93	4.04	3									
MW	2.43	1.39	3.52	4.52	5.09	5.84	5.36	5.71	3.77	4.18										
<2	8	11	0	2	1	0	0	1	1	24										
best rmsd, placement according dsx																				
Prot \ Lig	ATU	pl.	IXM	pl.	UN1	pl.	DFN	pl.	HBM	pl.	UN2	pl.	UN3	pl.	HT3t	pl.	DRG	pl.	MW	
hs1j1b	0.51	193	0.65	32	1.84	329	1.07	49	2.62	376	0.32	36	0.91	38	0.59	190	1.10	329	1.07	175
hs1q3d	0.87	209	0.56	182	0.87	148	0.96	69	2.21	63	1.05	214	1.56	195	1.18	118	1.08	401	1.15	178
hs1q3w	0.30	105	0.37	33	0.65	308	0.71	42	2.62	67	0.47	225	1.18	122	0.85	14	1.15	56	0.92	108
hs1q41	0.40	5	0.37	8	0.62	206	1.17	301	1.93	948	0.50	186	0.66	498	0.80	153	1.32	230	0.86	282
hs1q4l	0.68	44	0.90	7	0.62	9	1.51	125	1.69	119	1.48	501	4.20	2	1.00	212	1.51	388	1.51	156
hs1r0e	0.64	226	0.39	48	0.50	9	0.85	194	2.08	626	1.50	578	1.58	466	0.94	422	0.89	203	1.04	308
hs2o5k	0.81	328	0.56	369	1.68	99	1.60	130	1.30	109	1.00	349	1.31	390	0.62	344	1.01	244	1.10	262
hs2ow3	1.01	566	0.63	164	0.89	42	0.87	83	2.52	535	0.82	447	1.85	261	0.94	289	1.05	344	1.17	303
hs3du8	0.41	188	0.98	56	0.97	189	1.02	16	1.61	198	0.42	5	1.08	37	0.49	50	0.78	346	0.86	121
hs3f7z	0.25	119	0.35	40	0.67	317	0.98	43	2.13	994	0.52	70	0.56	223	0.74	76	0.63	239	0.76	236
hs3f88	0.41	170	0.56	120	0.94	83	0.97	75	2.30	28	0.82	131	1.22	174	0.54	43	0.90	130	0.96	106
kubra	0.33	13	0.38	35	1.28	38	1.08	43	2.49	94	2.04	735	2.83	302	1.43	420	0.35	77	1.36	195
MW (rmsd best solution):	0.55	181	0.56	91	0.96	148	1.07	98	2.12	346	0.91	290	1.58	226	0.84	194	0.98	249		
MW (pl. best solution):	12	1	12	2	12	2	12	0	4	0	11	1	10	0	10	0	12	0		
Σ (n<2 und pl<10):	1.06								202					6						
number of solutions																				
Prot \ Lig	ATU	<2A	IXM	<2A	UN1	<2A	DFN	<2A	HBM	<2A	UN2	<2A	UN3	<2A	HT3t	<2A	DRG	<2A	n<2	%
hs1j1b	509	180	503	184	830	1	1060	44	1126	0	773	167	723	26	733	273	417	57	932	14.0
hs1q3d	627	81	882	125	993	13	1273	115	856	0	744	30	730	12	732	66	560	23	465	6.3
hs1q3w	557	258	685	220	901	7	1427	101	896	0	699	76	982	5	854	153	547	47	865	11.5
hs1q41	549	193	856	233	924	4	850	93	1091	3	778	52	961	7	842	99	441	62	746	10.2
hs1q4l	535	74	626	95	724	53	1000	29	620	6	784	13	950	0	793	127	526	7	404	6.2
hs1r0e	519	272	555	151	774	45	1174	189	720	0	699	8	786	2	642	36	479	117	820	12.9
hs2o5k	572	161	546	233	456	2	1382	20	1149	18	662	113	846	13	633	214	408	18	792	11.9
hs2ow3	623	39	881	91	977	39	1391	40	1056	0	793	36	972	2	736	45	627	44	336	4.2
hs3du8	608	230	724	179	964	9	1385	27	1018	7	779	82	1102	29	703	193	554	43	799	10.2
hs3f7z	522	232	669	206	876	18	669	17	1077	0	669	59	902	27	703	89	545	48	696	10.6
hs3f88	462	226	834	214	514	23	1137	134	1197	0	775	75	482	10	714	138	485	82	902	13.7
kubra	481	39	599	122	659	7	1166	144	738	0	749	0	877	0	570	15	382	184	611	9.8
MW	547	165	697	179	799	18	1160	79	962	3	739	59	859	11	721	121	498	61	697	10.1

rmsd2cryst of top ranked solution																				
Prot \ Lig	ATU	IXM	UN1	DFN	HBM	UN2	UN3	HT3t	DRG	MW	<2									
hs1j1b	6.65	0.86	2.60	5.55	6.10	6.15	5.37	1.28	4.04	4.29	2									
hs1q3d	2.19	5.10	3.28	5.61	5.88	6.02	5.39	6.83	3.48	4.86	0									
hs1q3w	0.72	0.73	3.18	2.80	5.62	6.12	4.87	6.70	4.99	3.97	2									
hs1q41	0.70	0.48	3.07	5.65	6.53	6.14	5.96	6.34	2.55	4.16	2									
hs1q4l	1.07	1.18	5.87	2.78	8.64	5.95	4.72	6.46	2.68	4.37	2									
hs1r0e	1.73	0.61	2.43	1.03	2.51	5.99	6.24	6.86	1.48	3.21	4									
hs2o5k	6.85	1.09	2.78	5.93	1.68	5.74	5.02	3.16	3.43	3.96	2									
hs2ow3	3.51	1.58	6.27	5.90	6.93	3.72	5.91	7.04	5.45	5.15	1									
hs3du8	1.46	1.60	3.19	6.23	3.64	5.98	5.61	6.59	4.33	4.29	2									
hs3f7z	1.70	0.86	2.60	5.64	5.67	6.20	5.19	4.39	4.69	4.10	2									
hs3f88	1.46	0.78	2.46	5.56	2.40	6.03	5.45	6.53	3.20	3.77	2									
kubra	1.15	1.86	4.51	1.49	5.44	6.00	4.61	6.39	4.93	4.04	3									
MW	2.43	1.39	3.52	4.52	5.09	5.84	5.36	5.71	3.77	4.18										
<2	8	11	0	2	1	0	0	1	1	24										
best rmsd, placement according dsx																				
Prot \ Lig	ATU	pl.	IXM	pl.	UN1	pl.	DFN	pl.	HBM	pl.	UN2	pl.	UN3	pl.	HT3t	pl.	DRG	pl.	MW	
hs1j1b	0.95	2	0.86	1	1.84	80	1.86	2	2.96	15	0.48	9	0.91	22	1.28	1	1.64	32	1.42	18
hs1q3d	1.16	13	1.22	3	1.89	15	0.98	2	2.30	73	1.24	26	1.68	17	1.46	22	1.43	22	1.48	21
hs1q3w	0.72	1	0.40	2	0.65	61	0.82	2	2.84	4	0.85	8	1.18	33	1.36	9	1.53	12	1.15	15
hs1q41	0.70	1	0.48	1	0.62	45	1.49	4	2.31	100	0.53	11	1.43	43	1.13	8	1.61	3	1.14	24
hs1q4l	1.07	1	1.18	1	0.62	4	1.62	3	1.69	19	1.95	19	4.52	5	1.41	10	2.14	27	1.80	10
hs1r0e	0.92	15	0.61	1	0.50	7	1.03	1	2.18	95	1.82	31	2.71	16	1.17	5	1.30	8	1.36	20
hs2o5k	1.18	7	1.09	1	1.68	33	1.91	19	1.68	1	1.57	5	1.72	56	1.04	4	1.66	9	1.50	15
hs2ow3	1.89	8	0.79	8	0.89	21	0.87	11	2.62	28	1.91	8	2.00	34	1.76	7	1.09	26	1.53	17
hs3du8	1.17	8	1.18	2	1.61	24	1.02	5	1.84	9	0.56	2	1.08	9	0.76	6	1.59	6	1.20	8
hs3f7z	1.04	4	0.84	2	0.89	35	0.98	9	2.59	39	0.62	6	1.52	16	0.79	8	0.63	26	1.10	16
hs3f88	1.39	9	0.78	1	1.64	25	1.31	10	2.40	1	1.22	18	1.47	12	0.54	11	1.36	3	1.35	10
kubra	1.15	1	0.41	2	1.28	16	1.42	2	2.49	22	2.59	36	2.83	72	2.22	8	0.46	4	1.65	18
MW	1.11	6	0.82	2	1.18	31	1.28	6	2.32	34	1.28	15	1.92	28	1.24	8	1.37	15		
<2	10	12	12	2	12	10	3	2	11	6	9	1	11	9	11	6				
Σ (n<2 und pl<10):	1.39				16				16				58							
number of solutions																				
Prot \ Lig	ATU	<2A	IXM	<2A	UN1	<2A	DFN	<2A	HBM	<2A	UN2	<2A	UN3	<2A	HT3t	<2A	DRG	<2A	n<2	%
hs1j1b	39	4	48	3	140	1	286	1	136	0	36	3	113	3	30	2	47	4	21	2.4
hs1q3d	59	3	83	5	133	1	234	3	148	0	52	2	103	2	44	5	59	2	23	2.5
hs1q3w	43	5	57	5	125	3	270	3	142	0	33	2	114	1	39	4	52	2	25	2.9
hs1q41	44	2	71	5	135	1	205	2	147	0	44	2	124	1	29	2	50	3	18	2.1
hs1q4l	32	4	47	3	109	4														

B.4. QSAR

B.4.1. Ergebnisse der SAMPLS-Berechnungen

Tabelle B.8. | SAMPLS Berechnungen nach 4.3.1.1 auf Seite 154 (Überlagerung der Klassen 1 bzw. 2 basierend auf den Bindungsmodi A bzw. H).

Felder	q^2	# comp	s_{PRESS}
AcDoHyElSt	0.091	2	0.643
DoHyElSt	0.032	2	0.663
AcHyElSt	0.147	2	0.623
AcDoElSt	0.133	4	0.649
AcDoHySt	0.054	2	0.656
AcDoHyEl	0.093	2	0.642
AcDoHy	0.052	2	0.656
AcDoEl	0.144	4	0.645
AcDoSt	0.140	6	0.670
AcHyEl	0.155	2	0.620
AcHySt	0.111	2	0.636
AcElSt	0.221	3	0.605
DoHyEl	0.025	2	0.666
DoHySt	-0.042	1	0.677
DoElSt	0.081	2	0.647
HyElSt	0.073	2	0.649
AcDo	0.085	2	0.645
AcHy	0.115	2	0.634
AcEl	0.241	2	0.587
AcSt	0.148	2	0.622
DoHy	-0.045	1	0.678
DoEl	0.079	2	0.647
DoSt	-0.028	1	0.673
HyEl	0.071	2	0.650
HySt	-0.052	2	0.692
ElSt	0.214	2	0.598
Ac	0.153	2	0.683
Do	-0.033	1	0.675
Hy	-0.094	2	0.705
El	0.233	2	0.591
St	0.045	2	0.659

q^2 : Korrelationskoeffizient im Rahmen der Kreuzvalidierung; s_{PRESS} : root mean **P**redictive **E**rror **S**um of **S**quares; # comp.: Anzahl a der Komponenten (Gleichungen und Definitionen der statischen Kennzahlen vgl. S. 148 ff.); Ähnlichkeitsindizes werden abgekürzt mit Ac.: Wasserstoffbrückenakzeptor, Do.: Wasserstoffbrückendonator, St.: sterische Beschaffenheit und El.: elektrostatischen Beschaffenheit.

Tabelle B.9. | SAMPLS Berechnungen nach 4.3.1.2 auf Seite 155 (Überlagerung aller Verbindungen des Datensatzes basierend auf Modus A). Bei den grau unterlegten Kombinationen handelt es sich um die besten hinsichtlich des q^2 .

Felder	q^2	# comp	s_{PRESS}
AcDoHyElSt	0.268	2	0.577
DoHyElSt	0.195	2	0.605
AcHyElSt	0.309	3	0.570
AcDoElSt	0.359	8	0.601
AcDoHySt	0.163	2	0.617
AcDoHyEl	0.255	2	0.582
AcDoHy	0.149	1	0.612
AcDoEl	0.380	7	0.601
AcDoSt	0.396	8	0.583
AcHyEl	0.315	3	0.567
AcHySt	0.237	6	0.631
AcElSt	0.392	3	0.535
DoHyEl	0.209	2	0.599
DoHySt	0.119	2	0.648
DoElSt	0.202	2	0.602
HyElSt	0.228	3	0.602
AcDo	0.213	2	0.599
AcHy	0.229	6	0.605
AcEl	0.402	2	0.522
AcSt	0.336	4	0.568
DoHy	0.096	2	0.653
DoEl	0.212	2	0.598
DoSt	0.143	2	0.636
HyEl	0.246	3	0.595
HySt	0.012	2	0.666
ElSt	0.279	3	0.582
Ac	0.286	4	0.567
Do	0.123	2	0.642
Hy	0.020	9	0.674
El	0.294	3	0.576
St	0.147	3	0.645

Tabelle B.10. | SAMPLS Berechnungen nach 4.3.1.3 auf Seite 156 (Überlagerung aller Verbindungen des Datensatzes basierend auf Modus B).

Felder	q^2	# comp	s_{PRESS}
AcDoHyElSt	0.245	2	0.586
DoHyElSt	0.201	2	0.603
AcHyElSt	0.239	3	0.598
AcDoElSt	0.265	2	0.578
AcDoHySt	0.132	1	0.618
AcDoHyEl	0.226	2	0.593
AcDoHy	0.165	2	0.616
AcDoEl	0.251	2	0.583
AcDoSt	0.393	8	0.585
AcHyEl	0.242	3	0.597
AcHySt	0.204	4	0.622
AcElSt	0.288	2	0.569
DoHyEl	0.217	2	0.596
DoHySt	0.168	2	0.615
DoElSt	0.129	1	0.619
HyElSt	0.253	4	0.637
AcDo	0.201	2	0.603
AcHy	0.168	1	0.605
AcEl	0.319	1	0.548
AcSt	0.231	1	0.582
DoHy	-0.007	1	0.666
DoEl	0.218	2	0.596
DoSt	0.102	2	0.639
HyEl	0.280	3	0.582
HySt	0.116	4	0.655
ElSt	0.281	2	0.572
Ac	0.250	1	0.575
Do	0.081	2	0.646
Hy	0.156	6	0.664
El	0.271	2	0.576
St	0.145	3	0.634

B.4.2. Vorhersagen näher untersuchter Modelle

Tabelle B.11. | $\text{pIC}_{50}^{\text{act}}$ / $\text{pIC}_{50}^{\text{calc}}$ und Residuen gemäß Modell Modus A – AcDoSt (S. 159 ff.).

Verbindung	$\text{pIC}_{50}^{\text{act}}$	$\text{pIC}_{50}^{\text{calc}}$	Differenz
<i>Trainingsatz B</i>			
68	4.155	5.555	-1.400
103	4.222	5.029	-0.808
77	6.319	5.708	0.611
40	5.796	5.223	0.573
66	6.041	5.640	0.401
79	6.041	5.717	0.324
76	5.959	5.639	0.320
81	5.658	5.920	-0.262
78	5.569	5.352	0.216
75	5.420	5.596	-0.176
137	5.745	5.916	-0.172
85	5.222	5.080	0.142
139	6.000	6.142	-0.142
69	5.770	5.658	0.111
72	5.620	5.510	0.110
107	5.000	5.110	-0.110
83	6.301	6.401	-0.100
71	5.495	5.399	0.096
74	6.215	6.132	0.083
58	6.097	6.021	0.076
80	5.770	5.844	-0.074
143	5.523	5.458	0.065
142	4.824	4.761	0.063
125	4.523	4.462	0.061
62	6.292	6.352	-0.059
108	6.886	6.833	0.053
138	6.097	6.047	0.050
127	5.523	5.474	0.049
87	5.071	5.115	-0.045
141	5.420	5.381	0.039
63	6.745	6.780	-0.035
64	5.796	5.825	-0.029
105	5.658	5.683	-0.025
101	4.939	4.959	-0.020
135	5.398	5.416	-0.018
31	6.886	6.870	0.017
104	4.301	4.285	0.016
121	4.367	4.381	-0.015
128	4.796	4.785	0.011
<i>komplementärer Testsatz B</i>			
84	5.000	5.779	-0.779
56	6.398	6.989	-0.591
140	5.959	5.857	0.102
102	4.168	4.206	-0.039
132	5.523	5.531	-0.008

Tabelle B.12. | $\text{pIC}_{50}^{\text{act}}$ / $\text{pIC}_{50}^{\text{calc}}$ und Residuen gemäß Modell Modus A – AcEI (S. 165 ff.).

Verbindung	$\text{pIC}_{50}^{\text{act}}$	$\text{pIC}_{50}^{\text{calc}}$	Differenz
<i>Trainingsatz B</i>			
68	4.155	5.549	-1.394
121	4.367	5.438	-1.071
139	6.000	5.377	0.623
107	5.000	5.610	-0.610
77	6.319	5.709	0.610
64	5.796	6.355	-0.559
87	5.071	5.598	-0.528
103	4.222	4.661	-0.439
76	5.959	5.567	0.392
74	6.215	5.847	0.368
108	6.886	6.528	0.359
31	6.886	6.532	0.354
72	5.620	5.973	-0.353
40	5.796	5.444	0.352
66	6.041	5.690	0.351
69	5.770	5.425	0.344
83	6.301	5.976	0.325
137	5.745	5.428	0.317
142	4.824	5.139	-0.315
101	4.939	4.635	0.305
79	6.041	5.753	0.288
125	4.523	4.279	0.244
85	5.222	5.463	-0.241
138	6.097	5.860	0.237
143	5.523	5.300	0.223
63	6.745	6.539	0.205
58	6.097	6.296	-0.199
62	6.292	6.482	-0.190
104	4.301	4.121	0.180
75	5.420	5.565	-0.144
105	5.658	5.758	-0.100
128	4.796	4.876	-0.080
135	5.398	5.478	-0.080
81	5.658	5.579	0.078
127	5.523	5.466	0.057
141	5.420	5.365	0.056
71	5.495	5.440	0.055
78	5.569	5.602	-0.033
80	5.770	5.752	0.018
<i>komplementärer Testsatz B</i>			
140	5.959	5.238	0.720
84	5.000	5.700	-0.700
102	4.168	4.030	0.138
56	6.398	6.497	-0.099
132	5.523	5.584	-0.062

$\text{pIC}_{50}^{\text{act}}$ / $\text{pIC}_{50}^{\text{calc}}$: negativer dekadischer Logarithmus des gemessenen / berechneten IC_{50}

Tabelle B.13. | pIC_{50}^{act} / pIC_{50}^{calc} und Residuen gemäß Modell **Modus B – AcDoSt** (S. 170 ff.).

Verbindung	pIC_{50}^{act}	pIC_{50}^{calc}	Differenz
<i>Trainingsatz B</i>			
68	4.155	5.519	-1.364
103	4.222	5.242	-1.020
77	6.319	5.533	0.786
121	4.367	5.135	-0.769
74	6.215	5.520	0.695
142	4.824	5.384	-0.560
137	5.745	5.197	0.548
66	6.041	5.501	0.540
79	6.041	5.542	0.499
31	6.886	6.446	0.440
139	6.000	5.622	0.378
58	6.097	6.434	-0.337
76	5.959	5.651	0.307
101	4.939	5.211	-0.272
71	5.495	5.722	-0.227
104	4.301	4.085	0.216
40	5.796	5.596	0.200
85	5.222	5.422	-0.200
108	6.886	7.081	-0.195
143	5.523	5.342	0.181
127	5.523	5.378	0.145
83	6.301	6.437	-0.136
80	5.770	5.641	0.129
64	5.796	5.914	-0.118
75	5.420	5.538	-0.118
69	5.770	5.653	0.117
135	5.398	5.480	-0.082
107	5.000	4.919	0.081
87	5.071	5.136	-0.066
141	5.420	5.355	0.066
62	6.292	6.229	0.064
125	4.523	4.468	0.055
81	5.658	5.712	-0.055
128	4.796	4.848	-0.053
63	6.745	6.695	0.050
138	6.097	6.048	0.049
105	5.658	5.639	0.019
78	5.569	5.550	0.018
72	5.620	5.633	-0.013
<i>komplementärer Testsatz B</i>			
84	5.000	5.469	-0.469
140	5.959	5.522	0.436
132	5.523	5.343	0.180
102	4.168	4.053	0.115
56	6.398	6.495	-0.098

Tabelle B.14. | pIC_{50}^{act} / pIC_{50}^{calc} und Residuen gemäß Modell **Modus B – AcEI** (S. 173 ff.).

Verbindung	pIC_{50}^{act}	pIC_{50}^{calc}	Differenz
<i>Trainingsatz B</i>			
68	4.155	5.573	-1.418
121	4.367	5.241	-0.875
107	5.000	5.712	-0.712
85	5.222	5.864	-0.642
139	6.000	5.400	0.600
87	5.071	5.670	-0.599
77	6.319	5.739	0.580
108	6.886	6.384	0.502
137	5.745	5.246	0.498
74	6.215	5.720	0.494
31	6.886	6.400	0.486
103	4.222	4.694	-0.472
142	4.824	5.260	-0.436
79	6.041	5.677	0.364
64	5.796	6.144	-0.348
66	6.041	5.694	0.347
138	6.097	5.800	0.297
101	4.939	4.651	0.289
40	5.796	5.514	0.282
58	6.097	6.379	-0.282
125	4.523	4.247	0.276
143	5.523	5.250	0.273
127	5.523	5.253	0.270
69	5.770	5.515	0.255
83	6.301	6.061	0.240
72	5.620	5.848	-0.228
75	5.420	5.602	-0.182
135	5.398	5.260	0.138
128	4.796	4.925	-0.129
104	4.301	4.179	0.122
105	5.658	5.760	-0.103
76	5.959	5.881	0.078
62	6.292	6.362	-0.069
141	5.420	5.354	0.066
78	5.569	5.615	-0.046
80	5.770	5.724	0.046
71	5.495	5.536	-0.041
81	5.658	5.618	0.039
63	6.745	6.706	0.039
<i>komplementärer Testsatz B</i>			
140	5.959	5.196	0.763
84	5.000	5.719	-0.719
132	5.523	5.332	0.191
56	6.398	6.363	0.035
102	4.168	4.142	0.026

pIC_{50}^{act} / pIC_{50}^{calc} : negativer dekadischer Logarithmus des gemessenen / berechneten IC_{50}

C. Verwendete Gefahrstoffe und KMR-Substanzen

Mit den im Folgenden aufgeführten, besonders relevanten Gefahrstoffen wurde im Rahmen der Arbeit umgegangen. Dabei wurden die aufgeführten Gefahrenmerkmale (Hazard) und Sicherheitsratschläge (Precaution, auszugswise angegeben) beachtet. Die Angaben erfolgen gemäß Globally Harmonized System of Classification, Labelling and Packaging of Chemicals (GHS) bzw. Verordnung Nr. 1272/2008 (Annex VI) des Europäischen Parlaments und des Rates (GHS-Verordnung) oder den Sicherheitsdatenblättern folgend.

Stoff	Hazard	Precaution	Piktogramme
2-Chlorbenzaldehyd	H314	P303+P361+P353, P305+P351+P338, P310, P405, P501	
2-Cyanothioacetamid	H302, H312, H315, H319, H332, H335	P280, P302+P352, P261, P301+P312, P304+P340, P305+P351+P338	
Dichlormethan	H351	P260, P281, P308+P313	
N,N-Dimethylformamid	H226, H360D, H312, H332, H319	P280, P281, P305+P351+P338, P304+P340, P312, P405	

Stoff	Hazard	Precaution	Piktogramme
2-Iodbenzaldehyd	H315, H319, H335	P261, P305+351+P338	
Malonodinitril	H301, H311, H331, H400, H410	P261, P264, P273, P280, P301+P310, P311, P501	 
Piperidin	H225, H311, H331, H314	P303+P361+P353, P305+P351+P338, P310, P361, P405, P501	  

N,N-Dimethylformamid fällt nach der GHS-Verordnung 1272/2008 in die Kategorie R_E 1B fruchtschädigender (entwicklungsschädigender) Stoffe. Es wurde daher zusammen mit der Dissertation die geforderte Dokumentation nach der **Leitlinie für Tätigkeiten mit krebserzeugenden, erbgutverändernden und fortpflanzungsgefährdenden Stoffen (KMR-Stoffen) der Kategorie I und II im Fachbereich Chemie der Universität Hamburg** eingereicht.

D. Danksagung

An dieser Stelle möchte ich allen Personen meinen herzlichen Dank aussprechen, die mich während meiner Promotion unterstützt und begleitet haben. Neben Herrn Prof. Dr. H.-J. Duchstein möchte ich insbesondere Herrn Dr. T. Lemcke herzlich danken für die zahlreichen fachlichen und privaten, inspirierenden Gespräche und Diskussionen, seine ausdauernde, engagierte und freundliche Betreuung und Unterstützung, die hervorragende Zusammenarbeit und die schöne Zeit in der Arbeitsgruppe.

Mein besonderer Dank gilt des Weiteren Herrn Prof. Dr. C. Kunick, der eine Kooperation im Rahmen des Projektes zur Untersuchung von Thieno[2,3-*b*]pyridinen als selektive Inhibitoren der *Pf*GSK-3 stets unterstützt hat, entscheidende Ausgangsdaten zur Verfügung gestellt und die biologischen Testung in den Laboratorien von Herrn Prof. Dr. L. Meijer ermöglicht hat, dem ich ebenfalls danken möchte.

Insbesondere danke ich Herrn Kunick für die Ermöglichung des Forschungsaufenthaltes in Braunschweig. Aus seinem Arbeitskreis möchte ich mich insbesondere bei Frau Dr. W. Brandt für ihre großartige Unterstützung bei den präparativen Arbeiten bedanken. Ihr persönlicher Einsatz, ihre Hilfsbereitschaft und die vielen interessanten Gespräche führten dazu, dass die Zeit in Braunschweig gleichermaßen produktiv und eine persönliche Bereicherung war. Dazu trug auch die freundliche Aufnahme durch die anderen Braunschweiger Doktoranden bei, denen ich dafür danken möchte.

Für die Möglichkeit der Nutzung von DrugScoreX und den Austausch darüber möchte ich meinen Dank dem Arbeitskreis von Herrn Prof. Dr. G. Klebe und insbesondere Herrn G. Neudert aussprechen, der immer spontan hilfsbereit war, wenn die Funktionalität von DrugScoreX oder *fconv* optimiert werden konnte und Anregungen dieser Art freundlicherweise in kürzester Zeit umsetzte.

Auch Herrn Prof. Dr. M. Rarey und seinem Arbeitskreis möchte ich danken. Bei meiner Arbeit mit FlexE hatte ich die Gelegenheit, mich mit Frau A. Henzler auszutauschen und bei der zweidimensionalen Darstellung von Protein-Ligand-Komplexen mit Poseview war Frau K. Stierand so freundlich mich zu unterstützen.

Am Institut für Pharmazie der Universität Hamburg möchte ich in der pharmazeutischen Chemie den aktuellen Mitgliedern des Arbeitskreises von Herrn Prof. Dr. H.-J. Duchstein für die gute Zusammenarbeit danken. Insbesondere möchte ich den Herren J.-H. Löhr, B. Poll und M. Münsterberg für die Unterstützung beim Korrekturlesen, den alltäglichen wissenschaftlichen und privaten Austausch und, allein schon, für das hervorragende Arbeitsklima danken. Zudem bedanke ich mich bei den gegenwärtigen und ehemaligen Mitgliedern weiterer Arbeitskreise des Instituts für Pharmazie Frau K. Neumann, Frau Dr. A. Klack, Herrn Dr. F. Hansen, Herrn Dr. M. Khankischpur, Frau L. Dörre, Herrn S. Rehder, Frau Dr. A. Fenske, Frau E. Griewel und Herrn M. Münsterberg für die gute Zusammenarbeit in der Betreuung der Praktika

Arzneistoffanalytik unter besonderer Berücksichtigung der Arzneibücher und Arzneimittelanalytik, Drug-Monitoring, toxikologische und umweltrelevante Untersuchungen.

Für die erfolgreiche gemeinsame Durchführung und den damit verbundenen, weit über das Fachliche hinausgehenden Austausch von Gedanken im Rahmen von Weiterbildungs- und Kongressreisen möchte neben den Genannten zusätzlich den Mitgliedern des Arbeitskreises der pharmazeutischen Technologie danken – ganz besonders herzlich verbunden fühle ich mich Herrn M. Michaelis, Herrn S. Rehder und Frau I. Saniocki, mit denen ich nicht nur in diesem Zusammenhang viele sinnstiftende Gespräche führen durfte.

In den ersten Monaten meiner Arbeit wurde ich durch ein Promotionsstipendium der Universität Hamburg nach dem Hamburgischen Gesetz zur Förderung des wissenschaftlichen und künstlerischen Nachwuchses gefördert, wofür ich außerordentlich dankbar bin. In den Abschnitten der vorliegenden Arbeit zu experimentellen Details ist nachzulesen, dass Abbildungen teilweise mit dem Programmpaket UCSF Chimera der *Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco* erstellt wurden. Es wird gefördert durch die National Institutes of Health (NIH P41 RR001081).

Da die vorliegende Arbeit zum großen Teil computergestützte Methoden anwendet, war die reibungslose Zusammenarbeit mit dem immer hilfsbereiten IT-Service des Fachbereichs Chemie eine große Unterstützung, für die ich Herrn Prof. V. Vill, Herrn C. Schmidt und Herrn S. Ziehe danken möchte.

Vor allem danke ich meiner Familie, deren Vergrößerung den Beginn und das Ende meiner Promotion markiert. Die Begleitung, Unterstützung und Liebe, die ich von ihr erfahren habe, hat mir jeden Tag Energie gegeben, zur gleichen Zeit die Gefahr einer eindimensionalen Versteifung auf die Arbeit gebannt und die Entstehung der Dissertation so auf die beste denkbare Art und Weise gefördert.

E. Lebenslauf

Name	Sebastian Kruggel
Geboren am	23.02.1974 in Kiel
Kinder	Jonathan Friedrich Listmann (*04.02.2008) Carl David Listmann (*18.06.2011)
Sommer 1993	Abitur am Gymnasium Wyk auf Föhr
1994–2002	Diplom-Musiklehrer und Diplom Klassische Gitarre an der Hochschule für bildende Künste (Hamburg)
1999–2000	Studienjahr an der <i>Escuela Luthier de Artes musicales</i> (Barcelona)
2002–2006	Pharmaziestudium an der Universität Hamburg
2006–2007	Pharmaziepraktikant, Hafen-Apotheke (Hamburg)
2007	Pharmaziepraktikant, Bayer Schering AG (Berlin)
12/2007	Approbation als Apotheker
03/2008	Diplompharmazeut (Martin-Luther-Universität Halle-Wittenberge, Prof. Dr. W. Sippl, Prof. Dr. C. Sotriffer), Titel der Arbeit: <i>P-gp Substrate Differentiation by Pharmacophore Modelling</i>
01/2008–11/2008	Promotionsstipendium der Universität Hamburg, Anfertigung der vorliegenden Dissertation im Arbeitskreis von Herrn Prof. Dr. H.-J. Duchstein
seit 12/2008	Wechsel auf eine Stelle als wissenschaftlicher Mitarbeiter
seit 2008	Weiterbildung zum Fachapotheker für Pharmazeutische Analytik

F. Eidesstattliche Erklärung

Ich versichere hiermit an Eides statt, dass ich die vorliegende Dissertation selbständig und ohne fremde Hilfe angefertigt und mich nicht anderer als der im beigefügten Verzeichnis angegebenen Quellen und Hilfsmittel bedient habe.

Ferner versichere ich, dass dies mein erster Promotionsversuch ist und ich diese Dissertation noch an keiner anderen Universität eingereicht habe, um ein Promotionsverfahren eröffnen zu lassen.

Hamburg, den 12. Juli 2011

Sebastian Kruggel