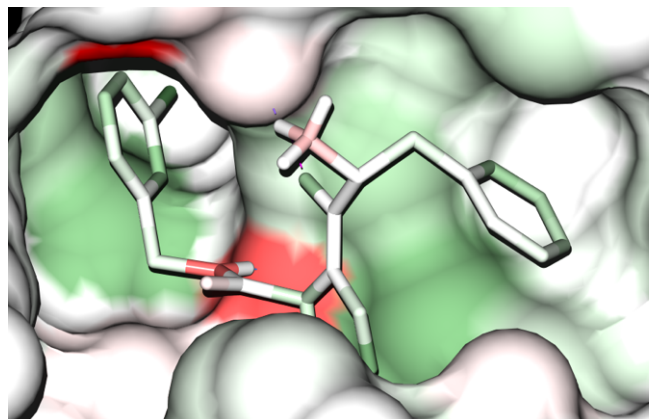


HYDE: Konsistente Bewertung von Protein-Ligand-Komplexen auf der Basis von Wasserstoffbrücken- und Dehydratationsenergie

Von der Theorie zur Anwendung



Dissertation
zur Erlangung des akademischen Grades

Dr. rer. nat.

an der Fakultät
für Mathematik, Informatik and Naturwissenschaften der
Universität Hamburg

eingereicht beim Fach-Promotionsausschuss Informatik von

Nadine Schneider

aus Merzig

Hamburg, September 2012

Erstgutachter: Prof. Dr. Matthias Rarey
Zweitgutachter: Prof. Dr. Bernd Meyer
Drittgutachter: Prof. Dr. Gerhard Klebe

Vom Fachbereich Informatik der Universität Hamburg als Dissertation angenommen
am 21.11.2012

Meinen lieben Eltern und André

Kurzfassung

Die Abschätzung der freien Bindungsenthalpie zwischen Biomolekülen ist ein Schlüsselproblem im computergestützten rationalen Wirkstoffentwurf. Die Suche nach neuen Leitstrukturen, in der mehrere hunderttausend Moleküle in einem virtuellen Screening getestet werden, erfordert dazu eine schnelle und zuverlässige Berechnung der Bindungsstärke der Moleküle zu einem Zielprotein. In dieser Dissertation wird die Weiterentwicklung von HYDE, einer Bewertungsfunktion zur Abschätzung der Bindungsenergie von Protein-Ligand-Komplexen, vorgestellt.

Die Bewertung eines Protein-Ligand-Komplexes in HYDE basiert auf einer konsistenten Beschreibung der Wasserstoffbrücken- und Dehydratationsenergie. Die Terme der HYDE-Funktion wurden von einem innovativen Konzept abgeleitet, das die intermolekularen Wechselwirkungen von Molekülen in wässriger Lösung als auch mit den Wassermolekülen selbst theoretisch beschreibt. Dadurch ist es möglich sowohl die energetisch günstigen Beiträge durch Wasserstoffbrücken-Interaktionen und den hydrophoben Effekt als auch destabilisierende Beiträge, die durch die Desolvatisierung von polaren Atomen entstehen, in der Abschätzung der Bindungsenergie eines Protein-Ligand-Komplexes zu berücksichtigen. Gerade diese ungünstigen Beiträge zur Bindungsenergie sind wichtig, um aktive von inaktiven Molekülen in der Leitstruktur-Suche zu unterscheiden. Im Gegensatz zu vielen anderen Bewertungsfunktionen wurde HYDE nicht anhand von Protein-Ligand-Komplexen und experimentell bestimmten Bindungsaffinitäten kalibriert. Stattdessen wurden gemessene Oktanol-Wasser-Verteilungskoeffizienten ($\log P$) von kleinen Molekülen zur Parametrisierung verwendet, wodurch eine konsistente Quantifizierung der ungünstigen polaren Dehydratationsenergie ermöglicht wurde.

Im Rahmen dieser Dissertation wurden verschiedene Aspekte der HYDE-Bewertungsfunktion überarbeitet, wie die Berechnung der Oberfläche, die zur Abschätzung der Dehydratationsenergie benötigt wird, die Modellierung der polaren Dehydratation als auch des Wasserstoffbrücken-Terms sowie die Bewertung von Metallionen. Weiterhin wurde das $\log P$ -Modell überarbeitet und die HYDE-Funktion neu parametrisiert. Die HYDE-Funktion wurde zudem in ein Optimierungsverfahren eingebettet, um eine genauere Abschätzung der Bindungsenergie eines Protein-Ligand-Komplexes zu ermöglichen. In einer intensiven Validierungsstudie zeigt die überarbeitete HYDE-Bewertungsfunktion verbesserte Ergebnisse und ist dadurch vergleichbar mit den bereits gut etablierten Protein-Ligand-Bewertungsfunktionen.

Abstract

The estimation of free binding energy of biomolecules is a key problem in computer-aided rational drug design. The search for new lead compounds, where several hundred thousand compounds are tested in a virtual screening, requires a fast and reliable calculation of the binding affinity of a compound to a protein target. In this thesis the further development of HYDE, which is a scoring function to estimate the binding energy of protein-ligand complex, is presented.

The assessment of a protein-ligand complex in HYDE is based on a consistent description of HYdrogen bond and DEhydration energies. The terms of the HYDE function were derived from an innovative concept, which theoretically describes the intermolecular interactions of molecules in aqueous solution as well as interactions of the molecules with water itself. This allows the consideration of energetically favorable contributions of hydrogen bond interactions and the hydrophobic effect as well as the destabilizing contributions arising from the desolvation of polar atoms with regard to the estimation of the binding energy of protein-ligand complexes. These unfavorable contributions are especially important in the discrimination of active and inactive molecules in a virtual screening. In contrast to many other scoring functions, HYDE is not calibrated on protein-ligand complexes and experimental binding affinities. Instead, it is parameterized using octanol/water partition coefficients ($\log P$) of small molecules which enables a consistent quantification of unfavorable polar desolvation.

Within the scope of this thesis several aspects of the HYDE scoring function were revised, such as the surface calculation which is required for the estimation of dehydration energies. Furthermore, the modeling of the polar dehydration as well as the hydrogen bond terms were changed and the scoring of metal ions was integrated. The $\log P$ model was revised as well and HYDE was re-calibrated. In addition, HYDE is integrated into an optimization procedure to allow a more accurate estimation of the binding energy of a protein-ligand complex. The good results of the revised HYDE scoring function in large-scale validation studies show the improvements and that HYDE is comparable to well established protein-ligand scoring functions in the field.

Danksagung

Besonders möchte ich mich bei Prof. Dr. Matthias Rarey bedanken, der mir die Möglichkeit gab ein sehr spannendes Thema in meiner Dissertation zu bearbeiten und es national sowie international auf Konferenzen zu präsentieren. Außerdem möchte ich mich bei ihm für die intensive fachliche Betreuung, die ständige Diskussionsbereitschaft und ein überaus motivierendes Arbeitsklima bedanken.

Ich bedanke mich auch bei Prof. Dr. Bernd Meyer und Prof. Dr. Gerhard Klebe für die Begutachtung meiner Dissertationsschrift.

Besonderer Dank gilt auch meinen Kooperationspartnern für die gute Zusammenarbeit und die vielen interessanten Projekttreffen, die zu immer neuen Anregungen führten. Ganz besonders möchte ich mich bei Dr. Gudrun Lange und Dr. Robert Klein bedanken, für unzählige wissenschaftliche Diskussionen und die intensive Betreuung, sowie den ambitionierten Einsatz in unserem Projekt.

Der BioSolveIT GmbH danke ich für die Bereitstellung der Flex*-Software-Bibliothek, für die Integration des HYDE-Moduls in die LeadIT-Software und die intensive Nutzung ihres Clusters zur Validierung der HYDE-Funktion. Besonders möchte ich mich bei Dr. Sally Hindle, Dr. Holger Claußen, Dr. Christian Lemmen, Dr. Marcus Gastreich und Dr. Carsten Detering bedanken, die mich bei der Softwareentwicklung, der Validierung und der Präsentation von HYDE so hilfreich unterstützt haben.

Allen meinen Kollegen und ehemaligen Kollegen am ZBH danke ich für die ausgezeichnete und abwechslungsreiche Arbeitsatmosphäre und für die interessanten und motivierenden Diskussionen. Ihr seid die besten Kollegen, die man sich vorstellen kann.

Ich möchte mich auch bei der Bayer CropScience AG und der Bayer HealthCare AG für die Finanzierung des HYDE-Projektes bedanken.

Für das Korrekturlesen meiner Dissertation bedanke ich mich bei André und besonders bei Lisa, die sich tapfer durch alle Seiten gekämpft hat.

Meinen Eltern danke ich für ihre immerwährende Unterstützung und den Glauben an mich, um auch die letzten sehr arbeitsintensiven Wochen zu überstehen.

André, danke für deine unendliche Geduld, Unterstützung und Motivation. Danke, dass du immer für mich da bist.

Inhaltsverzeichnis

Abkürzungsverzeichnis	XVII
1 Einleitung	1
1.1 Motivation	1
1.2 Projektbeschreibung	2
1.3 Übersicht Dissertation	3
2 Strukturbasierter Wirkstoffentwurf	7
2.1 Stufen des rationalen Wirkstoffentwurfs	8
2.1.1 Identifizierung des Zielproteins	8
2.1.2 Protein-Struktur-Bestimmung	8
2.1.3 Leitstruktur-Suche	10
2.1.4 Leitstruktur-Optimierung	11
2.1.5 Präklinische und klinische Phasen	12
2.2 Kristallstrukturen molekularer Komplexe	13
2.2.1 Einfache Qualitätsmaße von Kristallstrukturen	14
2.2.2 Mögliche Probleme im Umgang mit Kristallstrukturen	18
2.3 Molekulare Interaktionen	22
2.3.1 Bindungsenergie von Protein-Ligand-Komplexen	22
2.3.2 Interaktionstypen in Protein-Ligand-Komplexen	23
2.3.3 Entropische Beiträge in Protein-Ligand-Komplexen	27
2.4 Molekulares Docking	28
2.4.1 Verschiedene Ansätze im molekularen Docking	29
2.4.2 Herausforderungen im molekularen Docking	32
2.5 Protein-Ligand-Bewertungsfunktionen	33
2.5.1 Klassische Bewertungsfunktionen	34
2.5.2 Bestehende Herausforderungen	41
2.5.3 Alternative Ansätze	44

2.6	Strukturbasiertes virtuelles Screening	47
3	Konzept hinter der HYDE-Funktion	51
3.1	Das Wasserstoffbrücken-Netzwerk in Wasser	51
3.1.1	Sättigung des Wasserstoffbrücken-Netzwerkes	52
3.1.2	Mittlere Wasserstoffbrücken-Energie im Wassernetzwerk	55
3.1.3	Enthalpie und Entropie des Wasserstoffbrücken-Netzwerkes	55
3.2	Energie intermolekularer Wechselwirkungen in Wasser	57
3.2.1	Wechselwirkungen von polaren Gruppen mit Wasser	57
3.2.2	Wechselwirkungen von apolaren Gruppen mit Wasser	58
3.2.3	Wechselwirkungen zwischen funktionellen Gruppen in Wasser	59
3.3	Wechselwirkungen in Wasser als Basis von HYDE	61
4	Bestehende Methoden und Module	63
4.1	Der HYDE-Prototyp	63
4.1.1	Terme der HYDE-Bewertungsfunktion	64
4.1.2	Parametrisierung und Implementierung	66
4.1.3	Erste Resultate des HYDE-Prototyps	70
4.1.4	Erweiterungsmöglichkeiten	71
4.2	Optimierungsmethoden der Flex*-Bibliothek	72
4.2.1	ProToss: Optimierung des Wasserstoffbrücken-Netzwerkes	72
4.2.2	Optimierung der Protein-Ligand-Komplex-Geometrie	73
5	Parametrisierung des logP-Modells	75
5.1	Molekulare Oberfläche und logP-Wert	75
5.1.1	Verschiedene Oberflächen-Modelle eines Moleküls	76
5.1.2	logP - Logarithmus des Oktanol-Wasser-Verteilungskoeffizienten	77
5.1.3	Analyse der Beziehung von logP-Wert und Molekül-Oberfläche	78
5.2	Atombasierte logP-Parametrisierung in HYDE	83
5.2.1	Einteilung der neuen plogP-Atomtypen und Parametrisierung	84
5.2.2	Resultierende plogP-Inkrementen	85
6	Berechnung der HYDE-Funktion	89
6.1	HYDE - allgemeine Form	90
6.1.1	Berechnung der Dehydratationsenergie	90
6.1.2	Berechnung der Wasserstoffbrücken-Energie	96
6.2	Änderung der Zugänglichkeit von Atomen	97
6.2.1	Berechnung der molekularen Oberfläche	97

6.2.2	Abschätzung der Zugänglichkeitsänderung	101
6.3	Bewertung von Metallionen in HYDE	103
6.4	HYDE im Optimierungskontext	104
6.4.1	Die GeoHYDE-Funktion	105
6.4.2	Kombinierte Zielfunktion des Optimierungsverfahrens	107
6.5	Bewertung eines Protein-Ligand-Komplexes mit HYDE	108
6.5.1	Einzelbeiträge von Atomen	109
6.5.2	Einfluss einzelner Atome auf die Bindungsaffinität	111
7	Resultate und Diskussion	115
7.1	Qualitätsmaße und Bewertungskriterien	116
7.1.1	Bindungsaffinitätsvorhersage	116
7.1.2	Vorhersage von Protein-Ligand-Komplexen	117
7.1.3	Virtuelles Screening	118
7.2	Vorhersage von Bindungsaffinitäten	121
7.2.1	Affinitätsvorhersage von Molekülserien	121
7.2.2	Affinitätsvorhersage von Benchmark-Daten	128
7.3	Bestimmung der bioaktiven Ligandkonformation	133
7.3.1	Vergleich der Resultate von HYDE mit anderen Methoden	133
7.3.2	Detaillierte Analyse der HYDE-Redocking-Ergebnisse	135
7.4	Trennung von aktiven und inaktiven Molekülen	143
7.4.1	Detaillierte Analyse der Resultate des virtuellen Screenings	147
7.4.2	Vergleich der Resultate mit anderen Methoden	154
7.5	Mittlere Laufzeit der HYDE-Bewertungsfunktion	156
8	Weitere Anwendungen von HYDE	159
8.1	HYDE im biotechnologischen Kontext	160
8.2	Bewertung von Protein-Protein-Interaktionen	162
9	Fazit und Ausblick	167
	Literaturverzeichnis	171
	Anhang	196
A	Benutzung der Software	197
A.1	HYDE in der LeadIT GUI	197
A.2	HYDE als Screening-Tool	199

B Implementierung	203
B.1 Softwarestruktur	204
C Zusätzliche Resultate und Tabellen	205
C.1 DUD Enrichment: HYDE und HYDE-Prototyp	205
C.2 Strukturelle Klassifikation des PDBbind 2007 coreset	208
C.3 Strukturelle Klassifikation des DUD Datensatzes	214
D Veröffentlichungen	217
D.1 Veröffentlichungen in wissenschaftlichen Zeitschriften	217
D.2 Buchkapitel	218
D.3 Vorträge	218
D.4 Poster	218

Abkürzungsverzeichnis

ADME	Absorption Distribution Metabolisierung Exkretion
AMDE-Tox	Absorption Distribution Metabolisierung Exkretion Toxizität
AUC	Area Under the Curve
BFGS	Broyden-Fletcher-Goldfarb-Shanno
CSD	Cambridge Structural Database
DUD	Directory of Useful Decoys
FEP	Free Energy Perturbation
GB	Generalized Born
HTS	High Throughput Screening
HYDE	Hydrogen-bonds Dehydration
ITC	Isothermal Titration Calorimetry
IUPAC	International Union of Pure and Applied Chemistry
LIE	Linear Interaction Energy
MC	Monte Carlo
MD	Molecular Dynamics
MM	Molecular Mechanics
NMR	Nuclear Magnetic Resonance
PB	Poisson Boltzmann
PDB	Protein Data Bank
QSAR	Quantitative Structure Activity Relationship
RMSD	Root Mean Square Deviation
ROC	Receiver Operator Characteristic

Abkürzungsverzeichnis

SAS	Solvent Accessible Surface
TI	Thermodynamic Integration
VDW	Van-der-Waals
VHTS	Virtual High Throuhput Screening
WSAS	Weighted Solvent Accessible Surface
ZINC	ZINC Is Not Commercial
m	Milli (10^{-3})
μ	Mikro (10^{-6})
n	Nano (10^{-9})
p	Piko (10^{-12})
Å	Ångström
J	Joule
K	Kelvin
M	Molar
R	Allgemeine Gaskonstante
T	Temperatur
ΔG	Freie Enthalpie
ΔH	Änderung der Enthalpie
ΔS	Änderung der Entropie
k_b	Boltzmann-Konstante
K_d	Dissoziationskonstante
K_i	Inhibitionskonstante

1

Einleitung

1.1 Motivation

Die Entwicklung eines Medikaments ist ein sehr zeit- und kostenintensiver Prozess, der bis zu zwölf Jahre andauert und dabei meist über eine Milliarde US-Dollar beansprucht [1]. Computergestützte Methoden, die mittlerweile zusätzlich in den frühen Phasen der Entwicklung eingesetzt werden, dienen dazu, Kosten und Zeit einzusparen und die Entwicklung noch gezielter durchführen zu können. Viele dieser Methoden sind derzeit noch nicht so ausgereift wie die etablierten experimentellen Methoden, dennoch haben sie sich zu hilfreichen Werkzeugen im rationalen Wirkstoffentwurf entwickelt. Sie eröffnen neue Möglichkeiten, da sie nicht von Randbedingungen wie der Verfügbarkeit oder Löslichkeit von Substanzen abhängen, wie es z. B. in einem experimentellen Hochdurchsatz-Screening der Fall ist.

Eine computergestützte Methode, die zur Suche von neuen Leitstrukturen eingesetzt wird, ist das strukturbasierte virtuelle Hochdurchsatz-Screening. In dieser Methode werden mehrere hunderttausend Moleküle mit einem Docking-Programm in die Bindetasche eines Proteins eingepasst, und mit einer Bewertungsfunktion wird die Bindungsaffinität des Komplexes abgeschätzt. Dieser Prozess ist deutlich kostengünstiger und weniger zeitintensiv als ein experimentelles Hochdurchsatz-Screening. Die Erfolgsquote, d. h. die Anzahl der Moleküle, die dadurch gefunden werden, liegt zudem signifikant über

der des experimentellen Hochdurchsatz-Screenings [2]. Viele prospektive Studien der letzten Jahre dokumentieren die diversen Moleküle mit teilweise schon sehr hoher Bindungsaffinität, die durch verschiedene virtuelle Screening-Experimente, basierend auf unterschiedlichen computergestützten Methoden, gefunden wurden [3, 4].

Trotz dieser vielversprechenden Ergebnisse werden auch häufig falsche Vorhersagen durch computergestützte Methoden gemacht. Nicht jede Methode kann auf jede Proteinklasse angewendet werden [5–8]. Dies hat verschiedene Ursachen, wie z. B. die Abhängigkeit der Vorhersagen von den zugrunde liegenden Daten, vor allem aber eine unvollständige Modellierung der unterliegenden molekularen Prozesse und Wechselwirkungen. Diese sind zum Teil theoretisch nicht hinreichend erklärt, was ihre Modellierung in den computergestützten Methoden erschwert. Dennoch besteht ein sehr großes Interesse daran, eine Abschätzung z. B. der freien Bindungsenergie zu ermöglichen. In einem virtuellen Screening muss diese Abschätzung im Bereich von Millisekunden bis Sekunden erfolgen, um eine effiziente Anwendung zu garantieren. Die Bewertungsfunktionen, die dazu entwickelt wurden, sind derzeit nicht soweit ausgereift, dass eine hinreichend zuverlässige Vorhersage der Bindungsenergie erwartet werden kann. Ein möglicher Grund ist, dass in den meisten Bewertungsfunktionen nur günstige Energien bei der Komplexbildung erwogen werden, wodurch eine Überschätzung der Bindungsenergie eines Moleküles verursacht werden kann. Weiterhin werden häufig die Wassermoleküle vernachlässigt, die die Moleküle in Lösung umgeben und zugleich mit ihnen wechselwirken. Im Rahmen dieser Dissertation wurde die HYDE-Bewertungsfunktion weiterentwickelt, die sich gerade um die konsistente Beschreibung zuvor genannter Effekte und Energien bemüht. Die HYDE-Funktion basiert auf nur zwei Termen zur Abschätzung der Bindungsenergie, der Wasserstoffbrücken- (**HY**drogen bonds) und der **DE**hydratationsenergie. Dadurch können die wesentlichen Beiträge zur Bindungsenergie beschrieben werden.

1.2 Projektbeschreibung

Die Weiterentwicklung der HYDE-Bewertungsfunktion wurde in einem Kooperationsprojekt zwischen der Universität Hamburg, der Bayer CropScience AG, der Bayer HealthCare AG und der BioSolveIT GmbH durchgeführt. Ziel dabei war die Verbesserung des im Vorgängerprojekt entwickelten HYDE-Prototypen bis hin zu einer generellen Anwendbarkeit von HYDE und schließlich zur Marktreife. An der Universität Hamburg sollten dazu neue effizientere Berechnungsverfahren und Parameter entwickelt werden als auch die Modellierung einzelner Terme der HYDE-Funktion überarbeitet werden.

Weiterhin sollte eine modifizierte Form der HYDE-Funktion entstehen, die GeoHYDE-Funktion, die zur geometrischen Optimierung von Protein-Ligand-Komplexen verwendet werden kann. Die Kooperationspartner, insbesondere von Bayer CropScience, unterstützten das Projekt durch die Konzeption einer neuen Theorie, die die Basis von HYDE bildet als auch durch ihre wissenschaftliche Expertise in der Wirkstoffentwicklung. Durch das Mitwirken aller Kooperationspartner wurde eine umfangreiche kritische Untersuchung und Validierung der HYDE-Funktion möglich. Die BioSolveIT stellte zudem die Flex*-Software-Bibliothek bereit, in die die HYDE-Funktion als Modul an der Universität Hamburg implementiert wurde. Weiterhin sorgte die BioSolveIT für die Integration der HYDE-Funktion als Bewertungsmodul in die LeadIT-Software-Suite, so dass eine benutzerfreundliche Handhabung von HYDE möglich wurde. Zur Nutzung von HYDE in einem virtuellen Screening wurde zusätzlich ein kommandozeilenbasiertes Tool von der BioSolveIT zur Verfügung gestellt.

1.3 Übersicht Dissertation

Die vorliegende Arbeit wurde in der Abteilung für Algorithmisches Molekulares Design am Zentrum für Bioinformatik der Universität Hamburg in der Zeit von März 2008 bis September 2012 durchgeführt. Die Arbeit wurde bereits in zwei Publikationen, zwei Postern und fünf Vorträgen auf wissenschaftlichen Konferenzen veröffentlicht. Eine weitere Publikation wurde eingereicht und zwei weitere befinden sich in der Vorbereitung. In Anhang D sind diese Publikationen zusammenfasst aufgeführt.

In Kapitel 2 wird ein grundlegender Überblick zu Methoden des strukturbasierten Wirkstoffentwurfs gegeben. Nach einer Zusammenfassung der einzelnen Stufen des rationalen Wirkstoffentwurfs wird auf Kristallstrukturen molekularer Komplexe, die in der Regel den Ausgangspunkt des strukturbasierten Wirkstoffentwurfs bilden, eingegangen. Es werden verschiedene Kriterien erläutert, die zur Beurteilung der Qualität einer Kristallstruktur herangezogen werden sollten. Des Weiteren werden die molekularen Interaktionen, die zur Bildung eines Protein-Ligand-Komplexes führen, aufgezeigt und ihre Modellierung und der Beitrag, den sie zur Bindungsenergie leisten, wird kurz diskutiert. Sie bilden auch die Grundlage der Modelle, die im molekularen Docking und in der Bewertung von Protein-Ligand-Komplexen verwendet werden. Diese beiden computergestützten Methoden werden im Anschluss vorgestellt, wobei die verschiedenen Ansätze dargelegt und kritisch betrachtet werden. Das Kapitel schließt mit dem typischen Ablauf eines virtuellen Screenings ab.

Kapitel 3 stellt das Konzept, auf dem die HYDE-Bewertungsfunktion basiert, vor. Dieses wurde von Gudrun Lange und Robert Klein bei Bayer CropScience in Frankfurt entwickelt. Es liefert eine neuartige Beschreibung der intermolekularen Wechselwirkungen von Molekülen in wässriger Lösung als auch mit den Wassermolekülen selbst. Dadurch können Dehydratationsterme abgeleitet werden, so dass die Energie von Wasserstoffbrücken, dem hydrophoben Effekt und der polaren Dehydratation in konsistenter Weise abgeschätzt werden kann.

Das folgende Kapitel 4 beschreibt den HYDE-Prototyp, der im Rahmen der Dissertation von Ingo Reulecke am Zentrum für Bioinformatik der Universität Hamburg entwickelt und implementiert wurde. Des Weiteren werden Optimierungsmodule der Flex*-Bibliothek vorgestellt, die in der Überarbeitung der HYDE-Funktion verwendet wurden. In Kapitel 5 wird das neue Modell zur logP-Parametrisierung eingeführt, das zusammen mit Robert Klein von Bayer CropScience entwickelt wurde. Dazu wurden in einer Studie die einfache Korrelation des experimentell bestimmten logP-Wertes mit der molekularen Oberfläche generiert und analysiert. Die Erkenntnisse daraus führten zu einer Umstellung der bisher im HYDE-Prototyp verwendeten logP-Parametrisierung.

Die Berechnung der HYDE-Funktion wird in Kapitel 6 ausführlich erläutert. Die einzelnen Terme der überarbeiteten HYDE-Funktion werden aufgezeigt und die neue Modellierung der polaren Dehydratation und Wasserstoffbrücken-Terme werden erklärt. Weiterhin wird die neue Berechnungsmethode der Oberfläche und Zugänglichkeit zur Abschätzung der Dehydratationsenergie im Detail vorgestellt. Das Modell zur Bewertung von Metallionen durch HYDE wird ebenfalls erläutert. Eine modifizierte Form der HYDE-Funktion, die GeoHYDE-Funktion, die in ein Optimierungsverfahren integriert wurde, wird eingeführt. Zum Abschluss des Kapitels wird an einem Beispiel die Bewertung eines Protein-Ligand-Komplexes mit HYDE gezeigt.

In Kapitel 7 erfolgt eine intensive Validierung der überarbeiteten HYDE-Funktion. Die Vorhersagequalität von HYDE im Bezug auf Bindungsaffinitäten von Protein-Ligand-Komplexen wird untersucht als auch die Identifizierung der bioaktiven Konformation eines Liganden und die Trennung von aktiven und inaktiven Molekülen überprüft. Die Resultate werden diskutiert und wenn möglich mit dem HYDE-Prototypen und anderen etablierten Protein-Ligand-Bewertungsfunktionen verglichen.

Kapitel 8 zeigt weitere interessante Anwendungsmöglichkeiten von HYDE im biotechnologischen Kontext und in der Bewertung von Protein-Protein-Interaktionen auf.

In Kapitel 9 wird die gesamte Arbeit noch einmal zusammengefasst, bestehende Herausforderungen werden diskutiert und ein Ausblick auf weitere Verbesserungsmöglichkeiten der HYDE-Funktion wird gegeben.

Im Anhang sind weitere Details über die Verwendung der HYDE-Software, die Implementierung, zusätzliche Resultate und Tabellen sowie eine Liste der Publikationen, die aus dieser Arbeit hervorgegangen sind, zu finden.

2

Strukturbasierter Wirkstoffentwurf

Die Entwicklung oder vielmehr Entdeckung von Wirkstoffen war noch bis Anfang der achtziger Jahre geprägt von glücklichem Zufall und von Irrtum. Die Entdeckung von Penicillin durch Alexander Fleming im Jahr 1928 ist eines der bekanntesten Beispiele für einen glücklichen Zufall, doch auch viele andere bekannte Arzneimittel wurden eher zufällig gefunden, oder eine bessere Wirksamkeit wurde nachträglich in anderen Anwendungsgebieten festgestellt [9]. Die Erkenntnisse von Emil Fisher, der beschrieb, dass ein Substrat in seiner räumlichen Anordnung genau zu der Bindestelle eines Enzyms passen muss, um eine Reaktion auszulösen (Schlüssel-Schloss-Prinzip), und Paul Ehrlich, der erkannte, dass Arzneistoffe gebunden werden müssen, um ihre Wirkung zu entfalten, führten zu einem besseren Verständnis des Effektes, den ein Arzneistoff im Organismus auslöst. Durch die Röntgenstrukturanalyse und die Kristallographie, die die Aufklärung der dreidimensionalen Struktur eines Proteinkomplexes liefern, begann die gezieltere Entwicklung von Wirkstoffen. Der erste erfolgreiche strukturbasierte Wirkstoffentwurf wurde 1976 von Goodford und Kollegen veröffentlicht [10]. Anhand einer Struktur von Hämoglobin mit einem gebundenen 2,3-Bisphosphoglycerat-Molekül entwickelten sie verschiedene Verbindungen, die den gewünschten Effekt bewirkten und sich auch von dem natürlichen Liganden des Hämoglobins unterschieden. Eine weitere Verbesserung wurde mit dem Einsatz von computergestützten Methoden erzielt, die heutzutage in vielen Aspekten der Wirkstoffentwicklung ergänzend verwendet werden

[11]. Sie ermöglichen nicht nur komplexere Berechnungen, wie z. B. quantenchemischen Modelle, sondern stellen auch eine Visualisierung von Protein-Ligand-Komplexen bereit, die zur Generierung von neuen Ideen beiträgt und so eine wichtige Unterstützung bei der Entwicklung neuer Wirkstoffe liefert. Somit wird durch die Verwendung von strukturbasierten und computergestützten Methoden ein rationaler Entwurf von Wirkstoffen möglich.

2.1 Stufen des rationalen Wirkstoffentwurfs

Der rationale Wirkstoffentwurf kann in mehrere Stufen eingeteilt werden und beinhaltet viele verschiedene experimentelle und ergänzende computergestützte Methoden. In Abbildung 2.1 ist der Prozess des rationalen Wirkstoffentwurfs zusammengefasst dargestellt. Die einzelnen Phasen dieses Prozesses werden im Folgenden kurz beschrieben.

2.1.1 Identifizierung des Zielproteins

Am Anfang der rationalen Entwicklung eines neuen Wirkstoffs steht die Identifizierung des Krankheitsauslösers bzw. ein Verständnis für den Stoffwechselweg, der durch den Wirkstoff beeinflusst werden soll. Ein Zielprotein, gewöhnlich ein Enzym oder Rezeptor, muss bestimmt werden, dessen Aktivität oder Konformation mit einem niedermolekularen Molekül modifiziert werden kann. Dabei sollte eine gewisse Spezifität der Bindestelle gegeben sein, um unerwünschten Seiteneffekten durch die Interaktion des entwickelten Moleküls mit anderen Proteinen vorzubeugen. Weiterhin sollte das Zielprotein eine möglichst niedrige Mutationsrate aufweisen, um zu verhindern, dass ein Wirkstoff nach jahrelanger Entwicklung gar nicht mehr mit dem, vielleicht bis dahin mutierten, Zielprotein interagieren kann. Die Identifizierung des Zielproteins erfolgt experimentell, mit molekularbiologischen Methoden und stellt die Voraussetzung für einen rationalen Wirkstoffentwurf dar. Bei der Bestimmung des Zielproteins sollte auch berücksichtigt werden, dass dessen Gewinnung, Isolierung und Aufreinigung in größerem Maßstab möglich ist, um in späteren Phasen der Wirkstoffentwicklung beispielsweise Proteinkristalle zu erzeugen, ein Hochdurchsatz-Screening durchzuführen oder die Affinitätsänderung in der Leitstruktur-Optimierung zu messen.

2.1.2 Protein-Struktur-Bestimmung

Das Vorhandensein der dreidimensionalen Struktur des Zielproteins ist die Grundlage des strukturbasierten Wirkstoffentwurfs. Teilweise stehen zu Beginn der Entwicklung eines neuen Wirkstoffs bereits Strukturen des Proteins öffentlich in der Protein Data

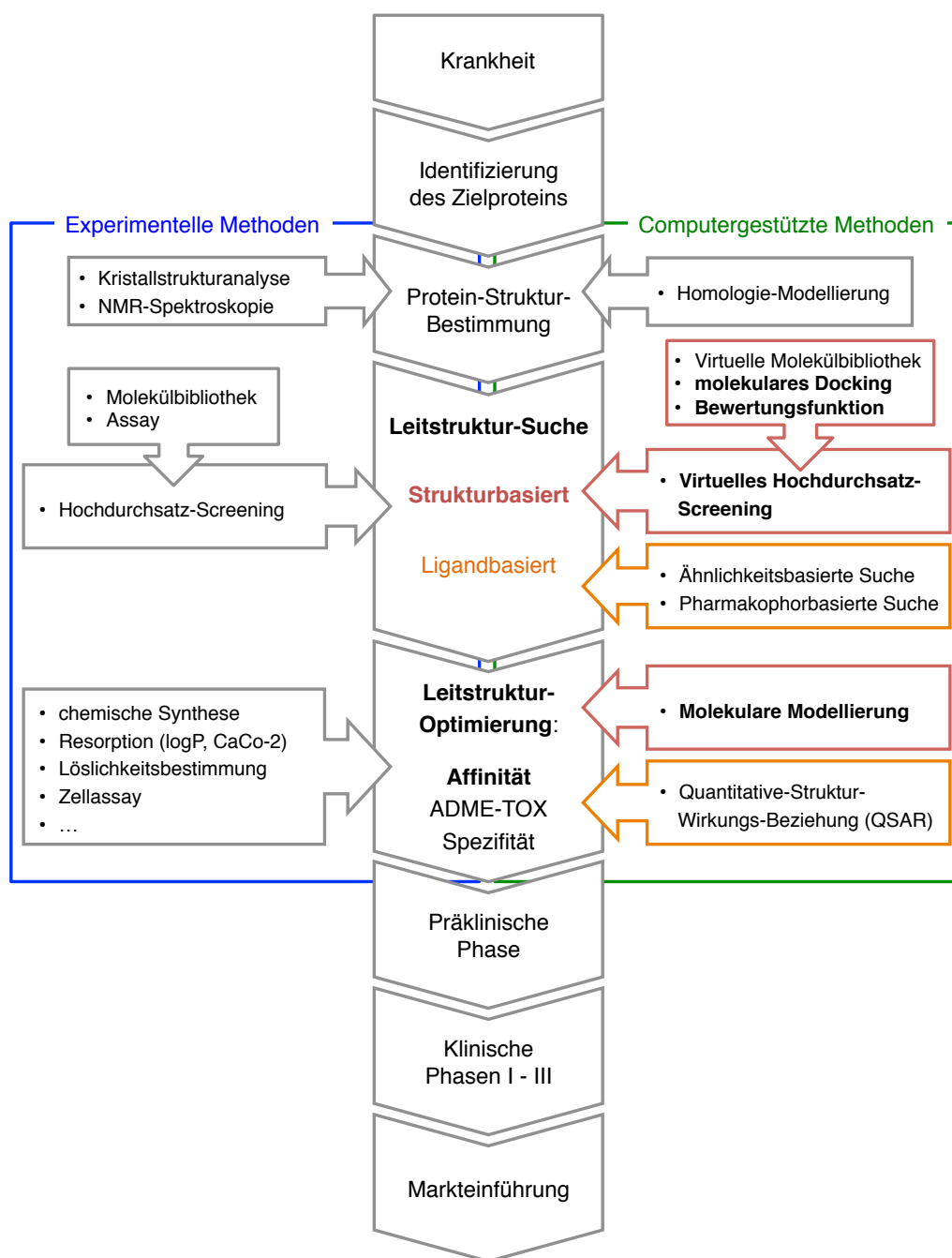


Abbildung 2.1: Wirkstoff-Entwicklungsprozess. In den Phasen, in denen ergänzend computergestützte Methoden der Chemieinformatik/Bioinformatik zur Verfügung stehen, sind auch einige der experimentellen, chemischen/biochemischen Methoden aufgezeigt. Die Schritte des Prozesses, die Gegenstand dieser Arbeit sind, wurden hervorgehoben.

Bank (PDB) [12] zur Verfügung oder sind in einer internen Datenbank eines Pharmaunternehmens gespeichert. Ist dies nicht der Fall, so kann die Struktur des Zielproteins durch experimentelle Methoden, wie Kristallstrukturanalyse oder NMR-Spektroskopie, aufgeklärt werden, oder am Computer kann mittels Homologie-Modellierung ein dreidimensionales Modell des Proteins erstellt werden. Am häufigsten werden im strukturbasierten Entwurf **Kristallstrukturen von Proteinen** (Kapitel 2.2) und Röntgenstrukturanalyse verwendet, da dieses Verfahren im Gegensatz zur NMR-Spektroskopie weniger aufwendig ist und auch größere Proteine strukturell aufgeklärt werden können. Nicht immer gelingt jedoch die Kristallisation eines Proteins, so dass die Kristallstrukturanalyse in diesen Fällen als Methode nicht in Frage kommt. Auch NMR-Spektroskopie kann nur bei Proteinen bis zu einer bestimmten Größe durchgeführt werden. Die Voraussetzung der Homologie-Modellierung ist die Existenz von dreidimensionalen Strukturen homologer Proteine. Hierbei ist die Qualität des Modells deutlich vom Grad der Homologie abhängig. Ist keine Struktur des Zielproteins verfügbar, dies ist z. B. oft bei membrangebundenen Rezeptoren der Fall, so kommen vorerst im weiteren Verlauf des rationalen Wirkstoffentwurfs neben den experimentellen Methoden nur ligandbasierte computergestützte Methoden bei der Leitstruktur-Suche in Frage.

2.1.3 Leitstruktur-Suche

Die Leitstruktur-Suche beginnt in der Regel mit einem experimentellen (HTS) oder virtuellen Hochdurchsatz-Screening (VHTS). Dabei werden Millionen von Verbindungen, die in chemischen Substanzbibliotheken oder virtuellen Molekülbibliotheken enthalten sind, auf eine Aktivität bezüglich des Zielproteins getestet oder es wird nach ähnlichen Molekülen zu bereits bekannten aktiven Verbindungen gesucht. Während im experimentellen Ansatz ein *in vitro*-Testsystem (Assay) zur automatischen Messung der Affinität der verschiedenen Moleküle der Substanzbibliothek vor dem HTS entwickelt werden muss [13], wird im virtuellen Experiment außer einer Molekülbibliothek eine **Docking-Methode** (Kapitel 2.4) benötigt, die die Moleküle in die vorgegebene Bindetasche einer Proteinstruktur einpasst als auch eine **Bewertungsfunktion** (Kapitel 2.5), die die Bindungsstärke der Moleküle in der Bindetasche abschätzt. Steht keine Struktur des Zielproteins im VHTS zur Verfügung, so werden ligandbasierte Verfahren eingesetzt. Hierzu müssen bereits aktive Verbindungen für das Zielprotein oder zumindest der natürliche Ligand des Zielproteins bekannt sein, so dass über eine Ähnlichkeitssuche, die auf verschiedenen Eigenschaften der Moleküle, z. B. pharmakophore Eigenschaften [14], beruhen kann, andere potentiell aktive Moleküle in der Bibliothek gefunden werden können [15]. Die Diversität der Moleküle, die im **strukturbasierten VHTS** (Kapitel 2.6)

entdeckt werden, ist somit höher. Grundlegend im experimentellen und im virtuellen Ansatz ist auch der Entwurf der entsprechenden Substanzbibliothek. Diese sollte möglichst divers sein und im virtuellen Fall auch nur synthetisierbare Moleküle enthalten. Im HTS werden meist nur sehr schwach bindende Substanzen im mikromolaren Bereich identifiziert, in manchen Fällen konnte kein einziges aktives Molekül gefunden werden [16, 17]. Durch die niedrige Erfolgsrate und die hohen Kosten eines HTS Experiments wird häufig auch ein VHTS als erster Schritt in der Leitstruktur-Suche durchgeführt. Dieses virtuelle Experiment ist deutlich schneller und kostengünstiger und liefert zudem höhere Erfolgsraten [2]. Dennoch sind auch in diesem Ansatz noch viele Probleme vorhanden, wie z. B. die Berücksichtigung von, für die Bindung wichtigen, Wassermolekülen, die Modellierung von Proteinflexibilität oder auch die Abschätzung der Bindungsstärke [3]. Nur durch das medizinisch-chemische Wissen und den Erfahrungsschatz von Experten können letztendlich geeignete Leitstrukturen, d. h. Moleküle, die sich auch weiter modifizieren und optimieren lassen, ausgewählt werden; HTS und VHTS Experimente können als Filter oder auch zur Generierung neuer Ideen dienen.

2.1.4 Leitstruktur-Optimierung

In der Leitstruktur-Optimierung wird in vielen Iterationen die Affinität und Spezifität der Leitstrukturen zum Zielprotein als auch die Absorption dieser Moleküle durch den Organismus und die Wirkdauer verbessert, während Nebenwirkungen und Toxizität der Moleküle reduziert werden muss (ADME-TOX Parameter). Um diese Ziele zu erreichen, d. h. aus den Leitstrukturen ein Molekül zu entwickeln, das in die klinische Phase übergeben werden kann, muss die chemische Struktur der gefundenen Substanzen verändert werden. Dabei ist eine Vielzahl an strukturellen Variationen der Moleküle möglich, wichtig ist nur, dass die Teile des Moleküls, die für die spezifische Bindung verantwortlich sind (der Pharmakophor) erhalten bleiben oder nur eingeschränkt verändert werden. Eine übliche Methode ist der bioisostere Austausch, d. h. bestimmte Gruppen/Atome eines Moleküls werden durch andere ersetzt, die ähnliche sterische und elektronische Eigenschaften besitzen und die die biologische Aktivität im Wesentlichen erhalten [18]. Eine weitere wichtige strukturelle Veränderung ist die Rigidisierung der Moleküle durch den Einbau planarer Ringsysteme, oder umgekehrt die Flexibilisierung durch die Ersetzung von Ringen durch aliphatische Ketten. Weiterhin wird meist die Lipophilie der Moleküle erhöht, so dass einerseits die Absorption und gleichzeitig meist auch die Affinität verbessert wird.

In dieser Phase des rationalen Entwurfs werden besonders häufig computergestützte

Methoden eingesetzt, da sie eine einfachere Analyse der Molekülkonformation ermöglichen als auch im strukturbasierten Entwurf durch die Visualisierung der **molekularen Interaktionen** (Kapitel 2.3) zwischen Protein und Ligand einen gezielteren Austausch von Gruppen zulassen. Dabei kommen verschiedene Methoden der molekularen Modellierung zum Einsatz, wie z. B. Konformationsanalyse durch Molekül-Kraftfelder, quantenchemische Verfahren, Berechnung physikochemischer Eigenschaften und ADME-TOX-Vorhersage, Überlagerung von Molekülen, 3D-QSAR Methoden [19], Moleküldynamik-Simulationen, molekulares Docking oder auch die Abschätzung der biologischen Aktivität durch Bewertungsfunktionen. Steht keine Information über die dreidimensionale Struktur des Proteins zur Verfügung, so kann mithilfe von Struktur-Wirkungs-Beziehungen (QSAR) [19, 20] das Ausmaß der Änderung der biologischen Aktivität, hervorgerufen durch die Variation von Atomen/Gruppen an einem Molekül, vorhergesagt werden. Die vorgeschlagenen Veränderungen an den Molekülen müssen anschließend im Labor synthetisiert werden und die vorhergesagten, verbesserten Eigenschaften der neuen Moleküle müssen experimentell validiert werden. Nur die Moleküle, die optimale Eigenschaften aufweisen, werden in den folgenden Phasen, der präklinischen Phasen und vier klinischen Phasen, weiter berücksichtigt. Dennoch scheitern die meisten Wirkstoffkandidaten an fehlender Effizienz (biologische Aktivität *in vivo*) oder schlechter Bioverfügbarkeit und Toxizität [21, 22].

2.1.5 Präklinische und klinische Phasen

In der präklinischen Phase werden die optimierten Moleküle auf ihre Wirksamkeit, Verträglichkeit und Toxizität für den Organismus getestet. Dazu werden weitere *in vitro* Versuche an Zellkulturen, isolierten Zellen und Organen durchgeführt als auch zum ersten Mal in *in vivo* Versuchen der Wirkstoff an Säugetiere, wie Mäuse oder Hunde, verabreicht. Meist resultiert aus dieser Phase nur noch ein Molekül, das weiter in den klinischen Phasen getestet werden kann. Dieses muss vor der klinischen Prüfung erst in eine geeignete Darreichungsform gebracht werden. Die klinische Prüfung ist in vier Phasen eingeteilt, wobei die ersten drei vor der Zulassung eines neuen Medikaments durchlaufen werden müssen, während sich die letzte, vierte Phase nach der Zulassung anschließt. Diese vier Phasen dienen dazu die pharmakokinetischen Eigenschaften im menschlichen Organismus zu überprüfen, ob das Medikament die gewünschte Wirkung zeigt und welche Nebenwirkungen zu erwarten sind. In der ersten Phase wird der neue Wirkstoff an einer kleinen Gruppe freiwilliger, gesunder Probanden getestet, um die Verträglichkeit des Medikaments zu kontrollieren und um Dosierungsrichtlinien zu finden, d. h. bei welcher Dosierung das Medikament zu Nebenwirkungen führt. In der

zweiten klinischen Phase wird das Arzneimittel einer etwas größeren Gruppe von Patienten verabreicht, um eine Wirksamkeit des neuen Arzneistoffs festzustellen und um eine Einschätzung tolerierbarer Nebenwirkungen zu ermöglichen. Weiterhin wird in dieser Phase die Dosis und Applikationsweise des neuen Wirkstoffs optimiert. In der dritten und letzten klinischen Phase vor der Zulassung des neuen Arzneistoffs werden Studien zur Unbedenklichkeit und Wirksamkeit als auch zur Feststellung seltener Nebenwirkungen oder Wechselwirkungen mit anderen Medikamenten durchgeführt. Dazu wird das Medikament einer großen Anzahl an Patienten, die in weitere Untergruppen einteilt werden, wie z. B. Patienten, die an anderen chronischen Erkrankungen leiden, verabreicht. Diese Phase kann mehrere Jahre andauern, bevor es zur Zulassung des Medikaments kommt. Nach der Zulassung schließt sich die vierte Phase der klinischen Prüfung an. In Langzeitstudien werden seltene Nebenwirkungen erfasst, die Darreichungsform wird gegebenenfalls angepasst und die Effizienz des Medikaments wird bestimmt. In dieser Phase kann eine erteilte Zulassung auch wieder zurückgezogen werden.

2.2 Kristallstrukturen molekularer Komplexe

Im strukturbasierten Wirkstoffentwurf ist die dreidimensionale Struktur des Zielproteins die Grundlage für die Entwicklung neuer Moleküle, die mit dem Zielprotein interagieren sollen. In den meisten Fällen wird dabei eine Kristallstruktur des Zielproteins genutzt, daher ist es wichtig bei der Auswahl einer passenden Kristallstruktur auf bestimmte Kriterien zu achten. Die Koordinaten eines Proteins, die durch Kristallstrukturanalyse generiert wurden, stellen lediglich ein atomares Modell der Proteinstruktur dar, das durch einen Kristallographen aus den Reflexen des Röntgenbeugungsexperiments abgeleitet wurde. Dies wird häufig im Umgang mit Kristallstrukturen vernachlässigt, d. h. Ungenauigkeiten, die in diesen Strukturen enthalten sind, werden nicht genügend hinterfragt, was zu schlechten oder auch falschen Resultaten z. B. im strukturbasierten virtuellen Screening führen kann. Im Folgenden wird auf einige Kriterien eingegangen, die einen Hinweis auf die Qualität einer Kristallstruktur liefern. Weiterhin werden mögliche Probleme aufgezeigt, die im Umgang mit Kristallstrukturen auftreten können und nicht durch die zuvor genannten Kriterien erfasst werden können.

Die Kristallstruktur eines Moleküls wird durch ein Röntgenbeugungsexperiment erzeugt. Dazu muss das Molekül zuvor kristallisiert werden. Dies kann bei größeren Molekülen, wie Proteinen, sehr aufwendig sein. Optimale Bedingungen (pH-Wert, Temperatur, Auswahl geeigneter Salze oder Detergenzien) müssen gefunden werden, in denen ein

regelmäßiges Kristallgitter des Moleküls entsteht. Nur solche Kristalle können später auch strukturell aufgeklärt werden. Die Röntgenstrahlung, die im Diffraktometer auf den Kristall trifft, wird an den Elektronen der Atome gebeugt und erzeugt ein, für den Kristall spezifisches Beugungsmuster, das je nach Element verschieden starke Intensitäten der Reflexe aufweist. Für einen Kristall werden im Diffraktometer sehr viele dieser Beugungsmuster aufgenommen, die jeweils einer Ebene des Kristalls (=Beugungsebene) entsprechen. Neben diesen experimentellen Daten, den Reflexen der Beugungsebenen, benötigt der Kristallograph Informationen über die sogenannte *relative Phasenlage* der Reflexe, um über eine Fourier-Transformation die Strukturfaktoren des Moleküls zu generieren, aus denen dann die Elektronendichtekarte konstruiert wird. Die relative Phasenlage kann jedoch nicht durch das Diffraktometer aufgezeichnet werden und ist als *Phasenproblem* der Kristallstrukturbestimmung bekannt [23]. Die relative Phasenlage kann durch verschiedene Verfahren initial bestimmt werden, so dass ein erstes Modell aus diesen Daten und der zugrunde liegenden Proteinsequenz erzeugt werden kann. Dieses ist jedoch sehr ungenau und enthält meist nicht alle Atome. In einem iterativen Verfeinerungsprozess wird das Modell immer wieder an die experimentellen Daten angepasst, so dass sich am Ende eine gute Übereinstimmung der beobachteten Elektronendichte F_o mit der, vom finalen Modell berechneten Elektronendichte F_c ergeben sollte. Die Differenzelektronendichte $F_o - F_c$ zeigt Abweichungen von Strukturmodell und Beugungsexperiment auf. Dabei führen fehlende Atome zu einer positiven Differenz, während falsch modellierte Atome eine negative Differenz ergeben. Meist wird die Elektronendichtekarte auch als kombinierte Differenzelektronendichte $2F_o - F_c$ [24] dargestellt, so dass bei falsch modellierten Atome eine geringere Dichte und bei fehlenden Atomen eine höhere Dichte zu finden ist (vgl. Abbildung 2.2).

2.2.1 Einfache Qualitätsmaße von Kristallstrukturen

Um die Qualität einer Kristallstruktur zu beurteilen, können viele verschiedene Kriterien berücksichtigt werden. Im Folgenden werden drei einfache Qualitätsmaße aufgezeigt, die in der Regel für eine Kristallstruktur angegeben werden. Diese beziehen sich auf die ganze Struktur und sind daher nicht optimal, um einzuschätzen wie gut z. B. die Binde-tasche eines Proteins modelliert werden konnte. Jedoch können durch diese einfachen Qualitätsmaße schnell Strukturen ausgeschlossen werden, die nicht im strukturbasierten Wirkstoffentwurf verwendet werden sollten.

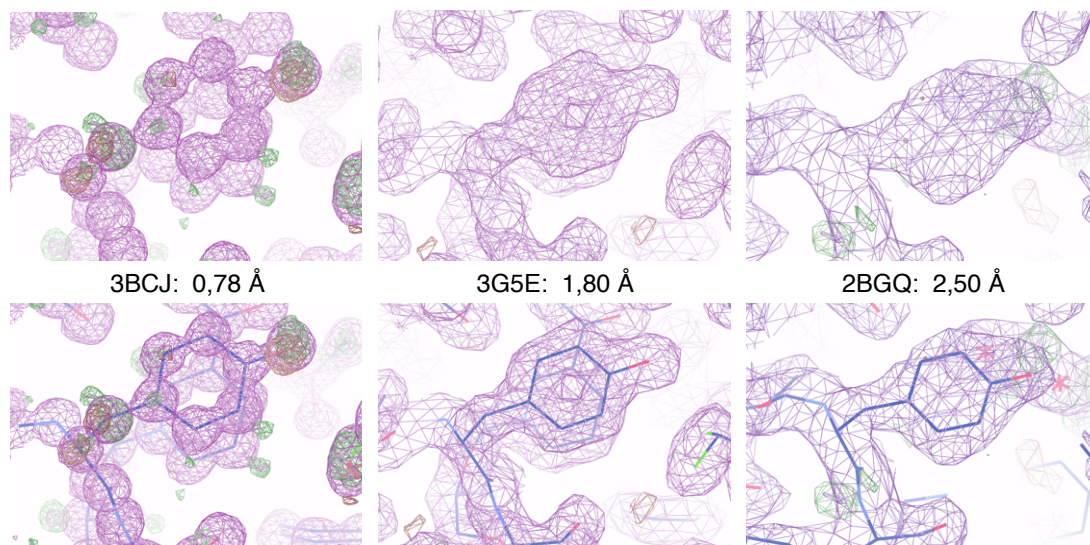


Abbildung 2.2: Verschiedene Auflösungen von Kristallstrukturen. Elektronendichte ($2F_o - F_c$ -Dichte bei 1σ -Level in Magenta) und Differenzelektronendichte ($F_o - F_c$ -Dichte bei $+3\sigma$ -Level in grün und -3σ -Level in rot) eines Tyrosins der Aldose-Reduktase bei drei unterschiedlichen Auflösungen: $0,78 \text{ \AA}$ (PDB: 3BCJ), $1,8 \text{ \AA}$ (PDB: 3G5E) und $2,5 \text{ \AA}$ (PDB: 2BGQ). Oben ist jeweils nur die Elektronendichte dargestellt, die über eine Fourier-Transformation aus den Reflexen des Röntgenbeugungsexperiments berechnet wurde. Unten ist jeweils die Interpretation der Elektronendichte, die Kristallstruktur des Proteins, aufgezeigt. [25]

Auflösung

Die Auflösung einer Kristallstruktur gibt den Grad an Detail an, der in den Reflexen enthalten war und anschließend in der daraus berechneten Elektronendichte sichtbar wird. Sie berechnet sich unter Verwendung der Bragg'schen Gleichung als minimaler Abstand d [\AA] der Beugungsebenen und hängt von der verwendeten Wellenlänge λ der Röntgenstrahlung und dem maximalen Austrittswinkel θ der berücksichtigten Reflexe ab. Je kleiner dieser Abstand d ist, desto höher/besser ist die Auflösung einer Kristallstruktur, d. h. desto höher ist der Grad an Detail, der in der Elektronendichte sichtbar wird. Die Auflösung einer Kristallstruktur wird vor allem durch die Qualität des Kristalls bestimmt, d. h. liegt ein besonders regelmäßiges Kristallgitter vor, so wird eine gute Auflösung erreicht. Bei Makromolekülen wie Proteinen, die eine hohe Dynamik aufweisen, ist das Kristallgitter meist nicht perfekt, da das Protein oftmals in leicht unterschiedlichen Konformationen darin vorliegt. Ab einer Auflösung von etwa

3 Å können die Seitenketten der Aminosäuren eines Proteins wahrgenommen werden, jedoch ist es erst ab einer Auflösung von etwa 2,5 Å möglich, die Seitenketten genauer zu modellieren. Ist die Auflösung besser als 1 Å, so wird auch die Lage von Wasserstoffatomen sichtbar. In Abbildung 2.2 sind beispielhaft drei verschiedene Auflösungsstufen einer Tyrosin-Seitenkette von Aldose-Reduktase-Strukturen dargestellt. In der ersten Struktur, die eine sehr hohe Auflösung von 0,78 Å besitzt, können die Positionen der einzelnen Atome wahrgenommen werden. Die meisten Protein-Kristallstrukturen haben eine Auflösung zwischen 1,5 und 3 Å, wobei Strukturen mit einer Auflösung von mehr als 2,5 Å schon nicht mehr besonders zuverlässig sind, was die genaue Lage der Atome betrifft.

R-Wert und freier R-Wert

Der R-Wert ist ein Maß, um zu überprüfen, wie gut das berechnete Modell mit den experimentellen Daten übereinstimmt. Dazu werden aus dem erstellten Modell die Reflexe berechnet und mit den experimentellen Werten verglichen. Der R-Wert, der sich für ein zufälliges Modell ergibt, liegt bei 0,63, während er bei einem perfekten Modell $R = 0$ beträgt. Im Verfeinerungsprozess wird der R-Wert optimiert und kann auch durch eine größere Anzahl an verwendeten Parametern beeinflusst und künstlich minimiert werden [26]. Im Gegensatz dazu wird der freie R-Wert [27] berechnet, indem das erstellte Modell mit experimentellen Daten verglichen wird, die vor der Erstellung/Verfeinerung des Modells entfernt wurden (üblich sind 10 % der Daten). Er ist daher ein besseres Qualitätsmaß für eine Kristallstruktur und stellt eine Art Kreuzvalidierung des Modells dar. Kristallstrukturen, die einen freien R-Wert von über 0,4 aufweisen, können als nicht vertrauenswürdig beurteilt werden [27]. Der konventionelle R-Wert und der freie R-Wert sollten bei einem guten Modell sehr ähnliche Werte liefern, ansonsten wurde das Modell im Verfeinerungsprozess zu sehr an die experimentellen Daten angepasst und damit auch an den experimentellen Fehler.

Temperaturfaktor

Der Temperaturfaktor oder B-Faktor, der für alle Atome des Modells angegeben wird, repräsentiert die mittlere Abweichung eines Atoms von seiner Position im Modell, er spiegelt die thermale Vibration eines Atoms im Modell wider. Während dem Verfeinerungsprozess werden jedoch viele der Fehler des Modells, wie z. B. Unregelmäßigkeiten im Kristall, durch den Temperaturfaktor kompensiert, so dass er nicht nur die thermale Vibration eines Atoms repräsentiert. Als Qualitätsmaß kann nicht der Wert eines

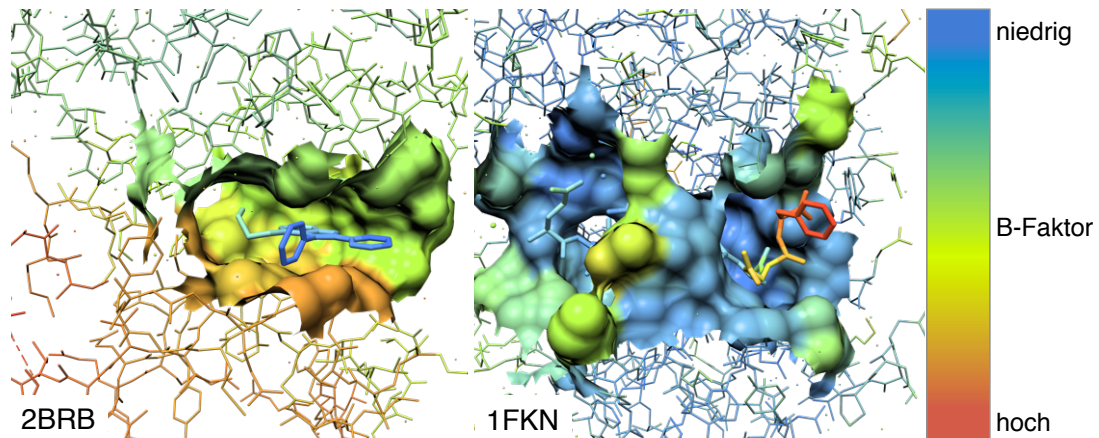


Abbildung 2.3: Temperaturfaktor/B-Faktor von Kristallstrukturen. Dargestellt sind zwei Protein-Ligand-Komplexe (PDB: 2BRB und 1FKN), die entsprechend ihrer Temperaturfaktoren (atombasiert) eingefärbt wurden. Da die Verteilung der Temperaturfaktoren von Kristallstruktur zu Kristallstruktur variiert, wurden die Atome nicht nach einem festen Wert farbcodiert, sondern die Farbcodierung wurde entsprechend des höchsten und niedrigsten Temperaturfaktors innerhalb jeweiligen Struktur bestimmt. [28]

einzelnen Temperaturfaktors eines Atoms dienen, da dieser auch durch den Verfeinerungsprozess beeinflusst wird. Stattdessen ist die Verteilung der Temperaturfaktoren im ganzen Modell interessant. Dabei sollten die Werte dieser relativ ähnlich für alle Atome sein. Teile des Modells, die überdurchschnittlich hohe Temperaturfaktoren im Vergleich zu der Umgebung aufweisen, deuten auf Ungenauigkeiten im Modell hin. Meist findet sich in diesen Teilen des Modells auch eine geringere oder fehlende Elektronendichte, so dass eine genaue Position der Atome im Modell eigentlich nicht festgelegt werden kann. Abbildung 2.3 zeigt die Verteilung der Temperaturfaktoren für zwei Protein-Ligand-Komplexe. In dem Chk1-Kinase Komplex (2BRB) sind die Temperaturfaktoren für einen Teil der Bindetasche erhöht, hier liegt auch eine schlechtere Elektronendichte vor. Die Koordinaten der Atome in diesem Teil der Bindetasche sind daher nicht so zuverlässig, wodurch sich diese Struktur für den strukturbasierten Wirkstoffentwurf nicht gut eignet. Der zweite Komplex, eine Beta-Sekretase mit einem Inhibitor (1FKN), zeigt eine relativ gleichmäßige Verteilung der Temperaturfaktoren auf Seiten des Proteins. Die Temperaturfaktoren des endständigen Phenylrings des Inhibitors hingegen sind jedoch stark erhöht, was auf eine Dynamik dieses Rings schließen lässt.

Neben diesen einfachen Qualitätsmaßen zur Beurteilung einer Kristallstruktur gibt es andere Möglichkeiten, wie das Erstellen eines Ramachandran-Plots, um die Torsions-

winkel des Proteinrückgrates des erzeugten Modells zu kontrollieren. Aber auch das Überprüfen der molekularen Wechselwirkungen eignet sich sehr gut, um die Zuverlässigkeit einer Kristallstruktur zu bestimmen, d. h. polare und geladene Atome, sollten möglichst nicht vergraben sein, ein gutes Wasserstoffbrücken-Netzwerk sollte im Protein ausgebildet werden können und hydrophobe Bereiche sollten gegen andere hydrophobe Bereiche gelagert sein. Ein gutes Modell ist immer chemisch (Bindungslängen und -winkel, Chiralität), physikalisch (keine Überlappung von nicht kovalent gebundenen Atomen) und kristallographisch (experimentelle Daten sollten erklärt sein) valide.

2.2.2 Mögliche Probleme im Umgang mit Kristallstrukturen

Die einfachen Qualitätsmaße, die im letzten Abschnitt vorgestellt wurden, lassen überwiegend eine globale Beurteilung einer Kristallstruktur zu. Durch Anwendung dieser Maße können unzuverlässige Strukturen ausgeschlossen werden, dennoch ist es möglich, dass auch Strukturen mit einer hohen Auflösung und einem niedrigen freien R-Wert Ungenauigkeiten oder Mehrdeutigkeiten in interessanten Bereichen, wie der Bindetasche des Proteins, aufweisen. Daher ist es für den strukturbasierten Wirkstoffentwurf wichtig, diese Mehrdeutigkeiten auszuschließen und sich zumindest, falls keine bessere Kristallstruktur zu Verfügung steht, über Ungenauigkeiten bewusst zu sein. Dieser Abschnitt zeigt einige dieser möglichen Probleme auf. Eine umfassendere Zusammenstellung inklusive der Konsequenzen, die dadurch im strukturbasierten Wirkstoffentwurf entstanden sind, ist in [29, 30] zu finden.

Schon die Kristallisationsbedingungen, wie z. B. der pH-Wert, können Einfluss auf die Kristallstruktur eines Protein-Ligand-Komplexes haben und zu unterschiedlichen Bindungsmodi des Liganden führen [31, 32]. Die verschiedenen Techniken, die bei der Kristallisation angewendet werden, wie z. B. *Kokristallisation* oder *Soaking* des Liganden, können ebenfalls leicht unterschiedliche Kristallstrukturen verursachen [33]. Durch eine fehlerhafte Proteinsequenz, hervorgerufen z. B. durch Mutationen, posttranslationale Modifikationen oder Sequenzierungsfehler, die zur Interpretation der Elektronendichte und damit zur Generierung des Modells genutzt wurde, wird die Elektronendichte während der Verfeinerung an den falschen Aminosäurerest angepasst. Diese Probleme und Einwirkungen auf eine Kristallstruktur können nur durch den Kristallographen selbst oder durch intensive Studie und gute Expertise mit Kristallstrukturen erkannt werden. Weiterhin können Mehrdeutigkeiten bei der Zuordnung von Stickstoff- und Sauerstoffatomen in Glutaminen und Asparaginen als auch bei der Festlegung von Sauerstoff-

und Kohlenstoffatomen in Histidinen entstehen. Diese Elemente weisen ähnliche Intensitäten im Beugungsexperiment auf und ihre Zuordnung in den betroffenen Aminosäureresten liegt in der Interpretation des Kristallographen, der dies anhand des Wasserstoffbrücken-Netzwerkes festlegt, das diese Reste mit umliegenden Aminosäuren ausbilden können. Zur Unterstützung wurden auch mehrere Computerprogramme entwickelt, die das Wasserstoffbrücken-Netzwerk optimieren und so die Zuordnung der Atomtypen überprüfen [34–37].

Unregelmäßigkeiten im Proteinkristall führen zu einer schlechteren Auflösung und zu geringerer oder fehlender Elektronendichte. Dadurch wird die Modellierung des Proteinrückgrates, aber vor allem von Aminosäure-Seitenketten erschwert, wodurch die Zuweisung falscher Konformationen resultieren kann. In manchen Fällen kann keine eindeutige Konformation zugewiesen werden, daher werden vom Kristallographen *alternative Konformationen* in der PDB-Datei hinterlegt. Abbildung 2.4 zeigt beispielhaft alternative Konformationen für zwei Protein-Ligand-Komplexe, die einmal die Bindetasche und im anderen Fall den Liganden betreffen. Flexible Seitenketten an der Oberfläche der Bindetasche, wie z. B. von Lysinen, Glutamaten oder Argininen, weisen oftmals eine schlechtere Elektronendichte auf und werden dann in einer Konformation modelliert, die statisch häufig für diese Reste in hochaufgelösten Strukturen zu finden ist [29].

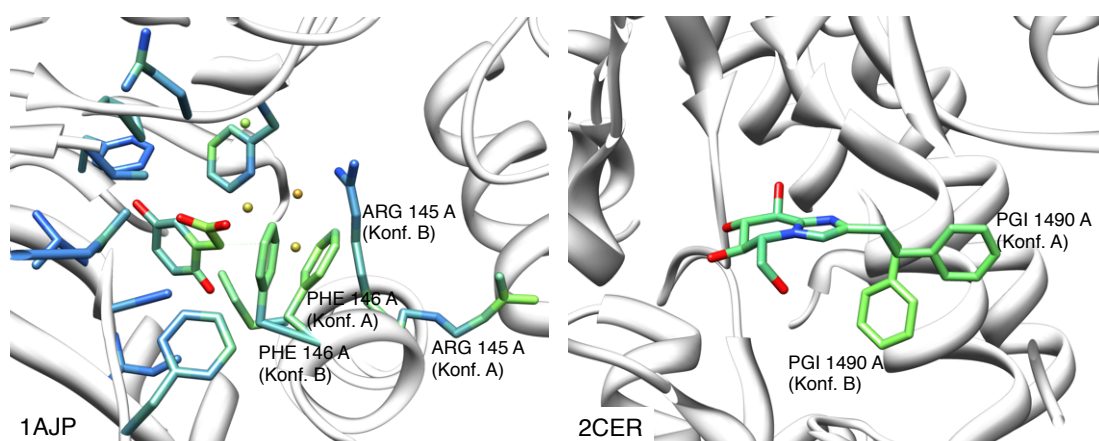


Abbildung 2.4: Alternative Konformationen in Kristallstrukturen. Zwei Strukturen von Protein-Ligand-Komplexen sind gezeigt, in denen alternative Konformationen A und B von Aminosäure-Seitenketten (PDB: 1AJP) oder des Liganden (PDB: 2CER) vorkommen. Die Atome wurden entsprechend ihres Temperaturfaktors eingefärbt, die Heteroatome des Liganden wurden zur besseren Übersicht in der Elementfärbung dargestellt.

Daher sollte im strukturbasierten Entwurf die Protein-Bindetasche anhand der Elektrodichtekarte untersucht werden, um festzustellen, ob solche flexiblen Reste vorhanden sind, deren Positionen nicht eindeutig festgelegt werden konnten, da sie maßgeblich das Resultat, z. B. im virtuellen Screening, beeinflussen können.

Ein weiteres Problem entsteht durch schlecht modellierte Liganden, d. h. falsche Konformationen oder falsche chemische Struktur. In älteren Kristallstrukturen wurden Liganden teilweise nicht modelliert, da die Struktur des Makromoleküls damals von größerem Interesse war [29, 38, 39]. Im strukturbasierten Wirkstoffentwurf ist natürlich auch ein korrekt modellierter Ligand von Bedeutung, um dessen molekulare Wechselwirkungen mit der Bindetasche untersuchen zu können, oder aber auch um die Bindetasche zu lokalisieren.

Die Koordinaten, die durch den Kristallographen für eine Proteinstruktur gewöhnlich angegeben werden, beziehen sich auf die *asymmetrische Einheit* im Proteinkristall, also der kleinste Baustein einer Kristallstruktur, der sich im Kristall immer wiederholt. Diese Einheit entspricht nicht zwangsläufig der *biologischen Einheit*, also dem funktionsfähigen Proteinkomplex, oder sie repräsentiert gleich mehrere biologische Einheiten des Proteins. Daher ist es im strukturbasierten Entwurf wichtig, auf die biologische Einheit zurückzugreifen, da teilweise nicht die gesamte Bindetasche in der asymmetrischen Einheit vorliegt, z. B. wenn sie sich innerhalb der Kontaktfläche zweier Proteinketten befindet. Ein Beispiel dazu ist in Abbildung 2.5 dargestellt. In dem Streptavidin-Komplex (1DF8) wird die biologische Einheit, ein Homotetramer, durch die asymmetrische Einheit (ein Dimer) und ein symmetrie-verwandtes Dimer gebildet. Hierbei ist in der asymmetrischen Einheit die Bindetasche nicht vollständig enthalten. In einem Proteinkristall liegen die einzelnen Proteinketten dicht gepackt vor, so dass es auch zu unnatürlichen Protein-Protein-Interaktionen, den sogenannten *Kristallpackungseffekten*, kommen kann. Dadurch kann die Konformation der interagierenden Aminosäurereste beeinflusst werden. Dies ist vor allem ungünstig, wenn im strukturbasierten Entwurf die molekularen Wechselwirkungen in der Bindetasche untersucht werden sollen und diese künstlich durch Kristallpackungseffekte verändert werden. Abbildung 2.5 stellt einen Kristallpackungseffekt dar, der die Geometrie der Bindetasche beeinflusst (PDB: 1O3P). Betrachtet man nur die asymmetrische Einheit, so gibt es keinen Hinweis auf die Beeinflussung der Geometrie der Bindetasche. Daher ist auch eine Untersuchung der symmetrie-verwandten Moleküle wichtig, um die Qualität einer Struktur für den strukturbasierten Entwurf zu beurteilen. Ein anderes Problem, das ebenfalls aus der Kristallpackung resultiert, ist die Einteilung in biologisch relevante und artifizielle Proteinkontakte [40] (vgl. Kapitel 8.2). Nicht immer ist bekannt, ob ein Protein als Mono- oder Multimer vorliegt, so dass anhand der Kristallstruktur entschieden werden muss, welche Proteinketten die

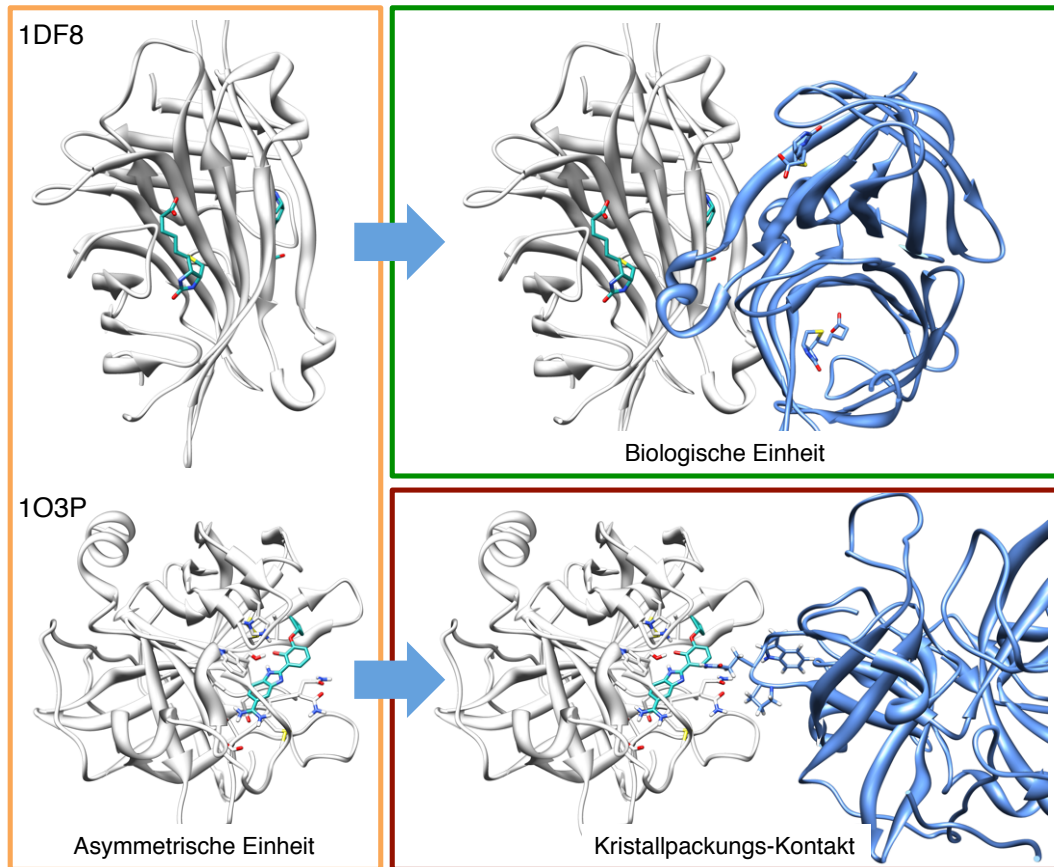


Abbildung 2.5: Asymmetrische Einheit und Kristallkontakte. Die asymmetrische Einheit von Streptavidin (1DF8) und Urokinase (1O3P) ist dargestellt. Bei Streptavidin bilden die benachbarten symmetrieverwandten Proteinketten (in blau) die biologische Einheit, ein Homotetramer. Bei der Urokinase ist die asymmetrische Einheit zugleich die biologische, die benachbarte symmetrieverwandte Proteinkette (in blau) führt hier zu einem Kristallpackungseffekt, der auch die Bindetasche der Urokinase beeinflusst.

biologische Einheit bilden. Eine falsche Annotation kann daher auch Probleme im strukturbasierten Entwurf verursachen.

Da gerade die Computerprogramme, die im strukturbasierten Entwurf verwendet werden, z. B. die Docking-Programme, sehr sensitiv auf kleine Veränderungen in den Koordinaten der Strukturen reagieren, ist es wichtig die möglichen Probleme, die im Umgang mit Kristallstrukturen auftreten können, bei der Auswertung von Ergebnissen zu berücksichtigen.

2.3 Molekulare Interaktionen

Bei der Entwicklung eines Wirkstoffs soll ein niedermolekulares Molekül gefunden werden, dass die Funktion eines Zielproteins beeinflussen kann. Ist das Zielprotein ein Enzym, so konkurriert der Wirkstoff meist mit dem natürlichen Substrat des Enzyms um die Besetzung der Protein-Bindetasche. Eine kompetitive Hemmung des Enzyms wird nur erreicht, wenn die molekularen Wechselwirkungen (auch Interaktionen) des Wirkstoffs in der Bindetasche energetisch günstiger sind als die des Substrates.

2.3.1 Bindungsenergie von Protein-Ligand-Komplexen

Die Ausbildung eines Komplexes hängt letztendlich davon ab, ob die Energie beider Moleküle durch ihre Assoziation günstiger ist als im ungebundenen Zustand. Im chemischen Gleichgewicht der Reaktion kann dies über ein Massenwirkungsgesetz als Verhältnis aus der Konzentration von ungebundenem Protein $[P]$ und Ligand $[L]$ zu der Konzentration des Protein-Ligand-Komplexes $[PL]$ bestimmt werden:

$$K_d = \frac{[P] \cdot [L]}{[PL]} \quad (2.1)$$

Dabei ist K_d die *Dissoziationskonstante* und steht für die Bindungsaffinität eines Liganden an das Protein. K_d wird in der Einheit einer Konzentration mol/l (M) angegeben. Je kleiner der Wert von K_d ist, desto stabiler ist der Protein-Ligand-Komplex, d. h. desto stärker bindet der Ligand. Bei der Entwicklung von Enzyminhibitoren wird der K_d -Wert des Inhibitors auch häufig als *Inhibitionskonstante* K_i bezeichnet. Die experimentelle Bestimmung der Inhibitionskonstante K_i ist aufwendig, daher wird auch häufig der IC_{50} -Wertes eines Inhibitors gemessen. Dieser Wert entspricht der Konzentration eines Inhibitors, die benötigt wird, um die Aktivität eines Enzyms um 50 % zu vermindern. Beide Werte, IC_{50} und K_i , hängen von experimentellen Bedingungen wie z. B. Temperatur oder pH-Wert ab und sind deshalb für unterschiedliche Protein-Ligand-Komplexe nur schwer vergleichbar. Bei konstanten Bedingungen jedoch können diese Werte genutzt werden, um zu entscheiden, welches Molekül stärker an das Protein bindet und sich daher besser als Inhibitor eignet.

Die Dissoziations- bzw. Inhibitionskonstante steht in direkter Beziehung zu der freien Bindungsenthalpie ΔG und lässt sich durch folgende Thermodynamische Gleichung in die freie Bindungsenthalpie umwandeln:

$$\Delta G = RT \cdot \ln(K_i) \quad (2.2)$$

Dabei ist R die allgemeine Gaskonstante und T die absolute Temperatur in Kelvin. Eine spontane Reaktion, bzw. die Bindung eines Liganden an das Protein, läuft ab, wenn $\Delta G < 0$ ist, d. h. Energie durch den Prozess frei wird. Die freie Enthalpie ΔG (auch Gibbs-Energie) kann auch durch folgende Gleichung, mithilfe der Enthalpieänderung ΔH und der Änderung der Entropie ΔS bei einer bestimmten Temperatur T , ausgedrückt werden:

$$\Delta G = \Delta H - T\Delta S \quad (2.3)$$

Hierbei wird deutlich, dass die freie Enthalpie aus einer enthalpischen und entropischen Komponente besteht, die beide berücksichtigt werden müssen. Die Bindungsenergie von nanomolaren Inhibitoren liegt im Bereich von $\Delta G = -35 - -55 \text{ kJ/mol}$, dabei können die einzelnen enthalpischen und entropischen Beiträge jedoch deutlich größer sein; bei der Bildung eines Protein-Ligand-Komplexes kommt es häufig zu einer sogenannten *Enthalpie-Entropie-Kompensation* [41, 42]. Dies ist ein bekannter Effekt in der Leitstruktur-Optimierung. Die freie Enthalpie eines Protein-Ligand-Komplexes kann experimentell durch Isotherme Titrationskalorimetrie (ITC) bestimmt werden. Dabei ist es auch möglich, den enthaltenen enthalpischen und entropischen Anteil zu ermitteln. Eine genaue Berechnung der freien Enthalpie hingegen stellt eine große Herausforderung dar. Daher werden zur Abschätzung der freien Bindungsenthalpie im strukturbasierten Wirkstoffentwurf einige Vereinfachungen herangezogen, wie die Ableitung der Bindungsenergie auf verschiedene Typen von molekularen Interaktionen. Weiterhin werden Annahmen bezüglich der Additivität einzelner Wechselwirkungsbeiträge gemacht. Eine besondere Schwierigkeit besteht auch in der Berücksichtigung von Desolvatisierungseffekten, die eine große Rolle bei der Bildung eines Protein-Ligand-Komplexes spielen. Die Bindungsenergie wird im strukturbasierten Wirkstoffentwurf meist mithilfe einer Bewertungsfunktion (Kapitel 2.5) abgeschätzt, die die verschiedenen Interaktionstypen in einem Protein-Ligand-Komplex ermittelt und über deren Beiträge die Gesamtenergie bestimmt.

2.3.2 Interaktionstypen in Protein-Ligand-Komplexen

Zur Leitstruktur-Optimierung ist eine Identifizierung von möglichen Interaktionen in der Protein-Bindetasche wichtig, so dass das potentielle Wirkstoffmolekül gezielt modifiziert werden kann, um die geeigneten Wechselwirkungsgruppen aufzuweisen. Dabei sind die wichtigsten Interaktionstypen, die in Protein-Ligand-Komplexen auftreten, wie folgt:

Wasserstoffbrücken

Eine Wasserstoffbrücke wird durch eine Dipol-Dipol-Interaktion bewirkt und stellt damit eine elektrostatische Wechselwirkung dar [43]. Sie bildet sich üblicherweise zwischen zwei elektronegativen Atomen (Stickstoff oder Sauerstoff), wobei eines der Atome eine Bindung zu einem Wasserstoffatom aufweist, der *Wasserstoffbrückendonator*, und das andere Atom ein freies Elektronenpaar besitzt, der *Wasserstoffbrückenakzeptor*. Wird eine Wasserstoffbrücke ausgebildet, so ist der Abstand der Schweratome meist kleiner als die Summe ihrer Van-der-Waals-Radien. Eine Wasserstoffbrücke ist stark gerichtet, d. h. sie ist vom Winkel der Interaktionspartner zueinander abhängig. Dabei sollte das Proton des Donors möglichst genau auf das freie Elektronenpaar des Akzeptors zeigen (vgl. Tabelle 2.1) [44]. Damit hat die Wasserstoffbrücke auch einen kovalenten Charakter, jedoch liegt ihre Bindungsstärke meist weit unter dem einer kovalenten Bindung. Die Bindungsenergie einer Wasserstoffbrücke beträgt *in vacuo* in der Regel weniger als -25 kJ/mol [45]. Ist jedoch einer der Bindungspartner geladen, so entsteht eine *ladungsunterstützte Wasserstoffbrücke*, die eine höhere Bindungsenergie *in vacuo* aufweisen kann. Die Interaktion von Protein und Ligand findet im Organismus hingegen in wässriger Lösung statt, deshalb ist der Beitrag einer Wasserstoffbrücke zur Bindungsenergie eines Protein-Ligand-Komplexes durch die Desolvatisierungskosten der Bindungspartner deutlich reduziert. Meist liegt die Energie von intermolekularen Wasserstoffbrücken im Bereich zwischen -2 und -6 kJ/mol [46], jedoch kann eine Wasserstoffbrücke, die stark von ihrer idealen Geometrie abweicht, durch Desolvatisierungseffekte auch ungünstig beitragen [47]. Ladungsunterstützte Wasserstoffbrücken leisten, bedingt durch höhere Desolvatisierungskosten, im Durchschnitt auch keinen größeren Beitrag zur Bindungsenergie eines Protein-Ligand-Komplexes [44]. Neben den klassischen Wasserstoffbrücken werden auch sogenannte *schwache Wasserstoffbrücken* untersucht, die nicht zwischen zwei elektronegativen Atomen ausgebildet werden. Alternativ sind CH-Gruppen, π -Elektronen oder auch Fluoratome daran beteiligt [48, 49].

Ionische Wechselwirkung

Ein weiterer Interaktionstyp elektrostatischen Ursprungs in Protein-Ligand-Komplexen bilden die ionischen Wechselwirkungen oder auch *Salzbrücken*. Diese entstehen, wenn zwei gegensätzlich geladene Gruppen, z. B. eine Carboxylat-Gruppe und ein Benzamidin, miteinander interagieren. Dabei kann sich auch die ionische Wechselwirkung mit einer Wasserstoffbrücke überlagern. Ionische Wechselwirkungen weisen eine hohe Bindungsstärke durch ihre elektrostatische Anziehung auf, die *in vacuo* bis zu -400 kJ/mol betragen kann [50]. Der Beitrag einer Salzbrücke zur Bindungsenergie in einem Protein-Ligand-Komplex ist durch die erhöhten Desolvatisierungskosten ähnlich dem einer unge-

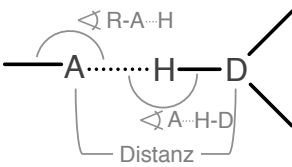
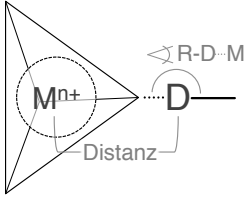
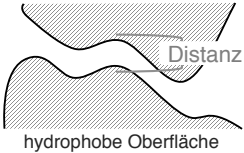
Typ	Allgemeine Beschreibung	Geometrie Parameter	Art des Beitrags	Energie [kJ/mol]
Wasserstoffbrücke		$D = 2,7 - 3,2 \text{ \AA}$ $\sphericalangle(R-A \cdots H - D)$ $= 160 - 200^\circ$ $\sphericalangle(R - A \cdots H)$ $= 100 - 260^\circ$	enthalpisch	<i>in vacuo</i> < -25 in Wasser $-2 - -6$
Metallkoordinierung		$D = 2 - 2,5 \text{ \AA}$ $\sphericalangle(R - D \cdots M)$ $= 160 - 200^\circ$	enthalpisch	
Hydrophobe Wechselwirkung		$D = 3,4 - 4,0 \text{ \AA}$	entropisch	<i>in vacuo</i> $-2 - -10$ in Wasser $-70 - -125$ J/mol \AA^{-1}

Tabelle 2.1: Typische molekulare Wechselwirkungen in Protein-Ligand-Komplexen.

ladenen Wasserstoffbrücke [44]. Ob sich anstatt einer Wasserstoffbrücke eine Salzbrücke in einem Protein-Ligand-Komplex bildet, hängt von dem pH-Wert der wässrigen Lösung und von den pK_a -Werten der Wechselwirkungspartner ab. Im strukturbasierten Entwurf muss daher auch der pH-Wert, bei dem die Komplexbildung stattfinden soll, berücksichtigt werden.

Metallkoordination

Ein besonderer Interaktionstyp, der in Metalloenzymen auftritt, ist die Metallkoordination. Das Metallion, das sich in der Protein-Bindetasche befindet, geht dabei eine koordinative Bindung mit den umgebenden Aminosäureresten, meist Histidine, Aspartate oder Cysteine, in einer ganz bestimmten, auch vom Metallion abhängigen, Koordinationsgeometrie ein. Das Metallion bildet den Akzeptor, der ein freies Elektronenpaar des Bindungspartners (Donor) annimmt. Wird das Metallion für die katalytische Funktion

benötigt, z. B. in Metalloproteasen, so bindet in der Regel auch der Ligand über eine koordinative Bindung an eine freie Koordinationsstelle des Metallions. In Protein-Ligand-Komplexen finden sich verschiedene Koordinationsgeometrien der Metallionen. Typisch sind Polyeder, wie Tetraeder, planares Rechteck, trigonale Bipyramide oder Oktaeder. Manche Metallionen können auch je nach Ligand verschiedene Koordinationsgeometrien annehmen [51]. Die Metallkoordination ist durch ihre Koordinationsgeometrie ebenso wie die Wasserstoffbrücken-Bindung eine stark gerichtete Wechselwirkung (vgl. Tabelle 2.1). Der Energiebeitrag einer koordinativen Bindung ist schwer zu bestimmen. Er liegt wahrscheinlich deutlich über dem anderer elektrostatischer Wechselwirkungen in Protein-Ligand-Komplexen, unterscheidet sich jedoch je nach Metallion, Komplex und Koordinationsgeometrie. Mutationsexperimente, wie sie zur Abschätzung des Beitrags von Wasserstoff- oder Salzbrücken durchgeführt werden, sind bei Metallionen nicht möglich. Die Relevanz vollständig besetzter Koordinationsgeometrien, wird in der Untersuchung von Kristallstrukturen bestätigt, in denen freie Koordinationsstellen immer durch ein Wassermolekül abgesättigt werden.

Hydrophobe Wechselwirkungen

Hydrophobe Wechselwirkungen entstehen, wenn sich aliphatische und aromatische Gruppen des Liganden an apolaren Aminosäuren des Proteins anlagern. Dieser Interaktionstyp ist ungerichtet und hängt nur von der Distanz beider Gruppen ab (vgl. Tabelle 2.1). Der energetische Beitrag einer solchen Wechselwirkung wird meist über die Änderung der hydrophoben Oberfläche durch die Komplexbildung bestimmt. Dabei wurden in verschiedenen Studien über die Verteilung von Molekülen im Oktanol-Wasser-System und die Stabilität von Proteinen Energiewerte ermittelt, die im Bereich von -70 bis $-125 \text{ kJ/mol } \text{Å}^{-2}$ liegen [52–54]. Der Hauptanteil der Energie entsteht jedoch nicht direkt durch die schwachen Anziehungskräfte der Gruppen zueinander (Van-der-Waals-Kraft) (vgl. Tabelle 2.1, *in vacuo* -2 - -10 kJ/mol [50]), sondern durch einen entropischen Desolvatisierungsbeitrag, den *hydrophoben Effekt* (Kapitel 2.3.3). Eine andere besondere Form der hydrophoben Wechselwirkungen stellen die π - π -Interaktionen dar, die von aromatischen Ringen gebildet werden können. Hierbei existieren zwei bevorzugte Geometrien der Ringe zueinander, die parallele, leicht verschobene Ausrichtung in einem Abstand von $3,4$ - $3,6 \text{ Å}$ und die vertikale T-förmige Ausrichtung in einem Abstand von $3,6$ - $3,8 \text{ Å}$ [44]. Diese entstehen durch eine günstige Überlappung der π -Orbitale der aromatischen Ringe und die damit verbundene Induzierung eines Quadrupol-Moments [55].

Neben diesen klassischen Interaktionstypen treten in Protein-Ligand-Komplexen auch *Kation- π -Interaktionen* und *Halogen-Brücken* auf, die durch die statistische Analy-

se von Kristallstrukturen gefunden wurden [56, 57]. Kation- π -Interaktionen entstehen durch die Wechselwirkung eines Kations, meist ein Ammonium, mit einem aromatischen Ring, in einem Abstand von 3,4 - 4,0 Å [44]. Diese Interaktion beruht auf einem Polarisierungseffekt zwischen der positiven Ladung des Kations und den π -Elektronen des aromatischen Rings [50]. Besonders stabil ist die Wechselwirkung, wenn es sich um quartäre Ammoniumverbindungen handelt, hierbei wird der Energiebeitrag nicht durch die hohen Desolvatisierungskosten der NH-Gruppen des Ammoniums kompensiert [58, 59]. Halogen-Brücken bilden sich zwischen Halogenen (Chlor, Brom oder Iod) und klassischen Wasserstoffakzeptoren, meist jedoch einer Carbonyl-Gruppe. Es handelt sich dabei um eine gerichtete, elektrostatische Wechselwirkung, wobei der Abstand der Wechselwirkungspartner wie bei Wasserstoffbrücken auch unterhalb der Summe der Van-der-Waals-Radien liegen kann (2,9 - 3,5 Å) [44, 60]. Der Energiebeitrag einer Halogen-Brücke ist deutlich schwächer als der einer Wasserstoffbrücke, er steigt mit der Größe des Halogenatoms [44].

Zur statistischen Analyse der verschiedenen Typen von molekularen Interaktionen in Protein-Ligand-Komplexen wurden Kristallstrukturen aus der PDB nach signifikant häufig auftretenden Wechselwirkungen untersucht [61–65]. Weiterhin wurden für verschiedene funktionellen Gruppen die Verteilung der Wechselwirkungspartner bestimmt, so dass bevorzugte Interaktionsdistanzen und -winkel abgeleitet werden können [44, 66–69]. Zusätzlich existieren Computerprogramme, wie z. B. GRID [70], die die Oberfläche der Bindetasche nach günstigen Interaktionen, sogenannten *Hot Spots*, absuchen, um das Ableiten strukturbasierter Pharmakophor-Hypothesen zu ermöglichen [71]. Ein Problem bei der Einteilung der Interaktionen in Typen und einzelne Energiebeiträge ist die Annahme der Additivität dieser Beiträge. Diese ist nicht immer gegeben, in manchen Protein-Ligand-Komplexen wurde eine Kooperativität einzelner Wechselwirkungen beobachtet [72, 73]. Letztere wird meist durch eine gegenläufige Entwicklung von enthalpischen und entropischen Beiträgen verursacht.

2.3.3 Entropische Beiträge in Protein-Ligand-Komplexen

Eine große Herausforderung im strukturbasierten Wirkstoffentwurf ist die Berücksichtigung von entropischen Beiträgen zur freien Bindungsenthalpie. Diese können im Grunde nicht in einzelne Interaktionen heruntergebrochen werden, sondern entstehen überwiegend durch die Wechselwirkungen der Moleküle mit den umgebenden Wassermolekülen. Im ungebundenen Zustand wird das tetraedrische Wasserstoffbrücken-Netzwerk, das zwischen den einzelnen Wassermolekülen besteht [74], durch Ligand und Protein,

besonders aber durch die apolaren Gruppen, gestört. Durch die Komplexbildung werden Ligand und Protein desolvatisiert, das Wassernetzwerk muss sich ebenfalls reorganisieren. Dies hat entropische als auch enthalpische Beiträge zur Bindungsenergie zur Folge. Durch das Aufbrechen von Wasserstoffbrücken zwischen den Molekülen erhöht sich die Entropie, während durch die Bildung neuer Wasserstoffbrücken im Protein-Ligand-Komplex als auch zwischen den Wassermolekülen ein enthalpischer Energiegewinn hervorgeht. Während die Desolvatisierung von apolaren Gruppen und ihre anschließende Zusammenlagerung im Komplex in einem energetisch günstigen hydrophoben Effekt resultiert, kann der Energieverlust durch die Desolvatisierung von polaren Atomen nur durch die Ausbildung neuer Wasserstoffbrücken im Protein-Ligand-Komplex kompensiert werden.

Eine weitere Entropieänderung wird durch die Einschränkung der Freiheitsgrade von Ligand und Rezeptor durch die Komplexbildung verursacht. Die Aminosäure-Seitenketten der Protein-Bindetasche und die Konformation des Liganden werden durch die Bindung der Moleküle teilweise fixiert, wodurch sich ein Entropieverlust ergibt. Auch hier ist die genaue Bestimmung der Größe dieses Entropieverlusts schwierig, da zum einen die Moleküle auch im Komplex eine gewisse Mobilität behalten [42] und zum anderen die Konformation der Moleküle auch schon im ungebundenen Zustand durch das Wassernetzwerk eingeschränkt wird.

2.4 Molekulares Docking

Molekulares Docking ist eine computerbasierte Methode, die in verschiedenen Phasen des strukturbasierten Wirkstoffentwurfs eingesetzt wird. Die Hauptanwendung liegt im strukturbasierten virtuellen Screening, jedoch dient es auch in der Leitstruktur-Optimierung zur Vorhersage des Bindungsmodus der modifizierten Leitstrukturen. Weiterhin wird die Docking-Methodik im *de novo*-Entwurf [75] verwendet, um kleine Fragmente in der Bindetasche des Proteins zu platzieren und anschließend miteinander zu einem vollständigen Molekül zu verbinden [76].

Im molekularen Docking wird versucht, den bioaktiven Bindungsmodus zweier Moleküle vorherzusagen, wobei es sich bei den Molekülen im strukturbasierten Wirkstoffentwurf meist um ein Protein, Enzym oder Rezeptor und einen niedermolekularen Liganden handelt. Darüber hinaus existieren auch Ansätze, die sich auf die Bindung von Makromolekülen, wie Proteinen oder DNS, konzentrieren [77, 78], diese werden hier jedoch nicht weiter beschrieben. Um den bioaktiven Bindungsmodus eines Liganden an ein Protein zu bestimmen, werden im Docking verschiedene Konformationen des Liganden in der Bindetasche des Proteins erzeugt (auch als *Posen* bezeichnet), bevor anschließend

mit einer Bewertungsfunktion (Kapitel 2.5) die beste Konformation bezüglich verschiedener Kriterien, z. B. Interaktionen mit dem Protein oder geometrische Kompatibilität, gesucht wird. Diese entspricht im idealen Fall der bioaktiven Konformation, die auch in einer Kristallstruktur des Protein-Ligand-Komplexes eingenommen werden würde. Alternativ zum molekularen Docking werden auch Moleküldynamik-Simulationen zur Lösung dieser Fragestellung eingesetzt. Diese sind jedoch zeit- und ressourcenintensiver, so dass sie nicht im virtuellen Screening eingesetzt werden können.

2.4.1 Verschiedene Ansätze im molekularen Docking

Die erste Docking-Methode DOCK, die 1982 von Kuntz *et al.* veröffentlicht wurde [79], behandelte beide Moleküle, Protein und Ligand, als starre Strukturen und nutzte ausschließlich geometrische Kriterien, um den Liganden in die Bindetasche einzupassen. In der weiteren Entwicklung von Docking-Methoden wurde der Ligand flexibel modelliert, d. h. verschiedene Konformationen des Liganden werden während einer Docking-Berechnung in der Bindetasche des Proteins konstruiert. Die Flexibilität der Bindetasche hingegen ist durch die Einführung zusätzlicher Freiheitsgrade und der damit verbundenen höheren Berechnungskomplexität ein großes Problem im molekularen Docking. Dies wird anschließend als eine der Herausforderungen diskutiert. Die Ligandflexibilität wird unterschiedlich in den verschiedenen Docking-Programmen realisiert. Eine grobe Einteilung in drei verschiedene Ansätze kann dabei vorgenommen werden: Fragmentbasierte Methoden, stochastische Methoden und Multikonformer-Methoden. Selbst nur die Berücksichtigung der Flexibilität des Liganden resultiert in einem hochdimensionalen Suchraum, so dass Vereinfachungen (z. B. Diskretisierung von Torsionswinkeln oder Wechselwirkungsgeometrien) und Heuristiken bei der Suche angewendet werden, um eine effiziente Lösungsstrategie des Problems zu ermöglichen.

Fragmentbasierte Methoden

In fragmentbasierten Methoden wird der Ligand vor dem Docking in kleinere, meist rigide Fragmente unterteilt. Anschließend wird er entweder durch einen *inkrementellen Aufbau* (z. B. in FLEXX [80] oder DOCK4.0 [81]) oder durch einen sogenannten *Place-and-Join-Algorithmus* (z. B. in SURFLEX [82] oder EHITS [83]) fragmentweise in der Bindetasche des Proteins rekonstruiert. Bei jedem Anbauschnitt wird der Konformationsraum des Liganden durch Anfügen der einzelnen Fragmente in unterschiedlichen Torsionen exploriert. Der Konformationsraum kann dabei nicht vollständig abgesucht

werden, sondern meist werden aus statistischen Analysen abgeleitete Torsionswinkelpotenziale (z. B. MIMUMBA [84, 85]) verwendet, die die bevorzugten Torsionswinkel zweier Fragmente enthalten. Während im inkrementellen Aufbau ausgehend von einem Basis-Fragment, das zuvor ausgewählt und in der Bindetasche platziert wurde, der Ligand vollständig rekonstruiert wird, werden in der Place-and-Join-Methode mehrere starre Fragmente in der Bindetasche platziert, bevor anschließend der Ligand mithilfe der flexiblen Fragmente zusammengesetzt wird. Jede Platzierung eines Fragments, auch der Anbau von Fragmenten, wird mit einer Bewertungsfunktion beurteilt. Dabei spielen geometrische als auch Kriterien, die die molekularen Interaktionen des Fragmentes mit der Bindetasche betreffen, eine Rolle. Um die Anzahl der generierten Posen weiter zu reduzieren, wird z. B. in FLEXX ein heuristischer *Greedy-Algorithmus* beim Aufbau des Liganden verwendet, so dass nur die besten Zwischenlösungen als Ausgangspunkt für den nächsten Anbauschnitt verwendet werden. Dadurch wird jedoch nicht immer das globale Optimum gefunden. Bei Place-and-Join-Methoden kann eine gespannte Konformation der Liganden erzeugt werden. Daher wird die finale Konformation des Liganden anschließend in der Bindetasche lokal minimiert [83].

Stochastische Methoden

In stochastischen Methoden wird im ersten Schritt eine beliebige Konformation des Liganden in der Bindetasche des Proteins platziert. Der Konformationsraum des Liganden wird anschließend in der Bindetasche abgesucht, indem die Torsionswinkel zufällig variiert werden. Der Ligand wird zusätzlich translatiert und rotiert, um den bioaktiven Bindungsmodus finden zu können. Dabei werden verschiedene Algorithmen verwendet, um die zufälligen Veränderungen zu bestimmen und um zu entscheiden, ob eine Konformation beibehalten wird und als Startpunkt im nächsten Veränderungsschritt dient. Zu den etabliertesten Methoden gehört dabei die *Monte-Carlo-Methode* (z. B. in ICM [86]) oder auch sogenannte *genetische Algorithmen* (z. B. in GOLD [87, 88] oder in AUTODOCK [89]).

In der Monte-Carlo-Methode wird in jedem Schritt eine zufällige Veränderung bezüglich der Torsionswinkel, der Rotation oder der Translation des Liganden durchgeführt. Anschließend wird die neue Konformation des Liganden minimiert, bevor sie mit einer Bewertungsfunktion beurteilt wird. Ist diese besser als die Ausgangskonformation, so wird sie zur Startkonformation des nächsten Schritts. Anderenfalls wird nach dem *Metropolis-Kriterium*, das auf einer Boltzmann-Verteilung beruht, entschieden, ob die neue Konformation den Startpunkt des nächsten Schritts darstellt, oder ob sie verworfen wird. Die Monte-Carlo-Methode kann auch mit einem *simulierten Abkühlungs-*

prozess gekoppelt werden, um in einem Energieminimum zu konvergieren. Zu Beginn der Suche wird dabei ein hoher Temperaturparameter gewählt, so dass durch das Metropolis-Kriterium fast keine Konformation des Liganden abgelehnt wird und damit auch Energiebarrieren überwunden werden können.

Genetische Algorithmen imitieren den evolutionären Prozess bei der Suche nach der bioaktiven Konformation des Liganden. Dazu wird die Konformation eines Liganden (hier als *Individuum* bezeichnet) in einem *Chromosom* kodiert, wobei die Torsionswinkel, die Rotation und die Translation jeweils die einzelnen *Gene* darstellen. Diese werden dann in jedem Schritt zufällig mutiert oder eine zufällige Rekombination der Chromosomen (*Crossover*) wird durchgeführt. Dadurch entsteht eine neue Generation von Individuen, die durch eine *Fitnessfunktion* (=Bewertungsfunktion) beurteilt werden. Nur die besten Individuen werden durch Selektion für die Erzeugung der nächsten Generation ausgewählt.

Durch die zufälligen Veränderungen und das probabilistische Akzeptanzkriterium dieser Methoden können die Resultate verschiedener Docking-Berechnungen variieren. Daher werden häufig mehrere Docking-Berechnungen durchgeführt, um eine bessere Konvergenz zu erhalten.

Multikonformer-Methoden

In Multikonformer-Methoden wird die Flexibilität des Liganden mithilfe eines Konformationsensembles modelliert. Dazu werden vor dem eigentlichen Docking-Prozess viele verschiedene Konformationen eines Liganden generiert, die anschließend starr in der Bindetasche des Proteins platziert und bewertet werden. Die Platzierung erfolgt entweder auf Basis eines Vergleichs der Form der Bindetasche mit der Form des Liganden (z. B. in FRED [90]) oder durch die Superpositionierung von spezifischen Charakteristika (z. B. Distanz von Wechselwirkungsgruppen) der beiden Moleküle (z. B. in TRIXX [91, 92]). Der Vorteil dieser Methoden liegt vor allem in ihrer Effizienz, jedoch ist die Chance die bioaktive Konformation des Liganden zu bestimmen an die zuvor generierten Konformationen gebunden. In GLIDE [93] werden daher die erzeugten Konformere nur grob in der Bindetasche platziert, bevor die besten Posen anschließend minimiert werden und mithilfe einer Monte-Carlo-Suche zusätzlich die Torsionen des Liganden optimiert werden.

2.4.2 Herausforderungen im molekularen Docking

Molekulares Docking ist aufgrund der vielen Freiheitsgrade ein multidimensionales Optimierungsproblem. Die Flexibilität des Liganden ist dabei nur ein Freiheitsgrad, der neben der Rotation und Translation des Liganden in den meisten Docking-Programmen berücksichtigt wird. Dies ist im Grunde jedoch nicht ausreichend, um die Bindung eines niedermolekularen Liganden an ein Protein zu beschreiben. Eine große Rolle spielt auch die Flexibilität von Proteinen, diese können durch eine *induzierte Anpassung* verschiedene Moleküle binden [94]. Die Flexibilität reicht von unterschiedlichen Konformationen der Seitenketten bis hin zur Bewegungen ganzer Protein-Domänen bei der Bindung. Bisher wird die Proteinflexibilität im molekularen Docking durch zwei unterschiedliche Ansätze, dem *Ensemble Docking* (z. B. FLEXE [95] oder FITTED [96]) oder dem *Induced Fit Docking* (z. B. SLIDE [97]), modelliert [98]. Während im Ensemble Docking verschiedene Konformationen des Proteins vor dem Docking erzeugt werden, in die der Ligand anschließend platziert wird, wird im Induced Fit Docking eine Überlappung zwischen Protein und Ligand bei der Platzierung zugelassen, die durch eine Minimierung des Komplexes kompensiert werden soll. Neben der zusätzlichen Komplexität in der Docking-Berechnung durch die Modellierung von Proteinflexibilität und der Schwierigkeit der Erzeugung einer validen Proteinkonformation ist vor allem die Bewertung und die Identifizierung der bioaktiven Konformation eine große Herausforderung. In den meisten Ansätzen werden die enthalpischen und entropischen Beiträge, die durch die Konformationsänderung innerhalb des Proteins entstehen, nicht berücksichtigt [99]. Einen weiteren Freiheitsgrad stellen auch die verschiedenen Protonierungszustände, Tautomere und freidrehbare funktionellen Gruppen beider Moleküle dar. Diese haben unmittelbar Einfluss auf die Wechselwirkungen, die zwischen den Molekülen ausgebildet werden können und somit auch auf die Resultate im molekularen Docking [100]. Die verschiedenen Zustände des Liganden können vor dem Docking erzeugt werden, jedoch sind auch in der Bindetasche verschiedene Protonierungszustände für Aminosäuren, wie Aspartat, Glutamat oder Histidin möglich. Die Realisierung aller möglichen Zustände während eines Dockings erhöht nicht nur zusätzlich die Komplexität der Berechnung, sondern ist auch nur dann sinnvoll, wenn die Bewertungsfunktion zwischen den verschiedenen Zuständen unterscheiden und den bioaktiven Zustand herausfiltern kann. Letzteres ist bisher die Hauptursache für schlechtere Docking-Resultate, trotz der Erzeugung verschiedener Zustände [100].

Protein und Ligand binden in wässriger Lösung. Daher sind in manchen Komplexen auch Wassermoleküle an der Bindung beteiligt. Diese sind teilweise sogar essentiell dafür. Die Einbeziehung dieser Wassermoleküle im molekularen Docking stellt eine weitere

Herausforderung dar. Es existieren Docking-Ansätze, in denen vorhandene Kristallwasser während der Docking-Berechnung an- und abgeschaltet werden können [96, 101–103]. In FLEXX können darüber hinaus auch günstige Positionen für potentielle Wassermoleküle im Docking vorhergesagt werden [101]. Die Schwierigkeit besteht primär jedoch in der Entscheidung, ob ein Wassermolekül bei einer Bindung benötigt wird. Das wiederum hängt von der Bindungsenergie jedes einzelnen Wassermoleküls ab [42] und müsste daher von der Bewertungsfunktion richtig abgeschätzt werden.

Im Grunde ist neben der erhöhten Komplexität bei der Berücksichtigung aller dieser Freiheitsgrade und der damit verbunden Herausforderung einer effizienten Berechnung insbesondere die richtige Beurteilung eines Protein-Ligand-Komplexes durch eine Bewertungsfunktion gefordert.

2.5 Protein-Ligand-Bewertungsfunktionen

Die energetische Bewertung eines Protein-Ligand-Komplexes ist eines der grundlegenden Probleme im strukturbasierten Wirkstoffentwurf. Erst eine realistische Abschätzung der Bindungsenergie ermöglicht die Entscheidung, ob ein Molekül als potentieller Wirkstoff in Frage kommt. Während es verschiedene experimentelle Messverfahren gibt, die die Affinität eines Moleküls zu einem Zielprotein bestimmen, so ist eine zuverlässige, computergestützte Berechnung der Bindungsenergie immer noch nicht möglich. Gute Ergebnisse liefern bisher Simulationsverfahren, die in Abschnitt 2.5.3 kurz vorgestellt werden. In diesen Verfahren werden alle bei der Bindung beteiligten Moleküle berücksichtigt, das bedeutet vor allem auch die umgebenden Wassermoleküle. Diese Verfahren sind aufgrund der komplexen Berechnungen sehr zeit- und speicherintensiv. Sie eignen sich daher eher zur Bewertung einzelner Komplexe. Gerade im molekularen Docking oder im virtuellen Screening wird aber eine Bewertung benötigt, die im Millisekunden- bis Sekundenbereich anhand einer einzigen Protein-Ligand-Komplexgeometrie die Entscheidung zulässt, ob sich ein Molekül für weitere Untersuchungen eignet. Dies führte zur Entwicklung sogenannter *Protein-Ligand-Bewertungsfunktionen* oder *Scoring-Funktionen*, die einen Kompromiss aus Genauigkeit und Effizienz darstellen.

Eine Bewertungsfunktion hat verschiedene Anwendungsbereiche im computergestützten Wirkstoffentwurf. Zum einen wird sie im molekularen Docking eingesetzt, um die bioaktive Konformation des Liganden unter allen vom Docking erzeugten Posen zu bestimmen. Im virtuellen Screening wird sie genutzt, um zwischen aktiven und inaktiven Molekülen eines Zielproteins zu unterscheiden. Weiterhin wird sie in der Leitstruktur-Optimierung zur Vorhersage der Bindungsaffinität benötigt. In verschiedenen Vergleichsstudien hat sich gezeigt, dass die Vorhersage der Bindungsaffinität mit Abstand die

größte Herausforderung für eine Bewertungsfunktion darstellt [5–8].

Wie bereits im letzten Abschnitt 2.4 beschrieben, wird schon während der Docking-Berechnung eine Bewertungsfunktion eingesetzt, um die Platzierung des Liganden in der Bindetasche in die richtige Richtung zu lenken. Hierbei ist vor allem wichtig, dass eine Überlappung zwischen den Atomen der Moleküle erkannt und verhindert wird sowie die Erzeugung einer energetisch günstigen Konformation des Liganden und die Maximierung der möglichen intermolekularen Wechselwirkungen angestrebt wird. Dies sind jedoch nicht unbedingt die Beiträge, die zu einer guten Abschätzung der Bindungsenergie führen, daher werden teilweise unterschiedliche Bewertungsfunktionen für die verschiedenen Phasen eingesetzt (= *Nachbewertung*) oder kombiniert (= *Konsensus-Bewertung*), um die Stärken jeder Funktion optimal zu nutzen. Durch die zentrale Bedeutung der Bewertungsfunktionen im computergestützten Wirkstoffentwurf, wurden in den letzten Jahren zahlreiche, verschiedene Bewertungsfunktionen entwickelt. Eine Übersicht darüber findet sich z. B. in [4, 42, 99, 104]. Im Folgenden werden die grundlegenden Ansätze der klassischen Bewertungsfunktionen vorgestellt und die bestehenden Herausforderungen diskutiert, des Weiteren werden alternative Ansätze aufgezeigt.

2.5.1 Klassische Bewertungsfunktionen

Aufgrund der unterschiedlichen Anforderungen an eine Bewertungsfunktion haben sich verschiedene Typen von Bewertungsfunktionen entwickelt, die sich generell in drei Kategorien einteilen lassen: *Kraftfeldbasierte* Bewertungsfunktionen, *empirische* Bewertungsfunktionen und *wissensbasierte* Bewertungsfunktionen. Bewertungsfunktionen stellen im Allgemeinen einen mathematischen Ausdruck dar, um die Energie von nicht-kovalenten Wechselwirkungen zwischen Protein und Ligand abzuschätzen. Dabei wird angenommen, dass die Energie in einzelne Beiträge aufgeteilt werden kann und eine Additivität der paarweisen Protein-Ligand-Interaktionen vorliegt.

Kraftfeldbasierte Bewertungsfunktionen

In kraftfeldbasierten Bewertungsfunktionen (z. B. in DOCK [105], Autodock [89, 106] oder GOLD [88]) sind einzelne Terme der Funktion von Kraftfeldmethoden abgeleitet. Kraftfelder wurden in der Chemie entwickelt, um die Energie eines Moleküls oder auch zwischen verschiedenen Molekülen zu bestimmen und damit die Vorhersage einer energetisch günstigen Konformation des Moleküls bzw. der Komplexgeometrie zu ermöglichen. Dazu werden einerseits kovalente Energiebeiträge berechnet, d. h. die Abweichung

von idealen Bindungslängen und Bindungswinkeln als auch die Potenziale der Torsionswinkel. Des Weiteren werden Van-der-Waals und elektrostatische Wechselwirkungen nicht unmittelbar benachbarter Atome (in einem Abstand von mindestens vier Bindungen) bzw. zwischen Atomen unterschiedlicher Moleküle berücksichtigt. Diese werden in der Regel durch einen Lennard-Jones-Term und ein Coulomb-Potenzial beschrieben. Es existieren verschiedene Kraftfelder, die speziell für Makromoleküle wie Proteine oder DNA angepasst oder entwickelt wurden, z. B. AMBER [107, 108], OPLS-AA [109, 110], CHARMM [111, 112] und GROMOS [113]. Diese Kraftfelder wurden anhand von experimentellen Daten und quantenchemischen Berechnung kalibriert. Sie enthalten viele verschiedene Parameter für die unterschiedlichen Atomtypen. Die klassischen Kraftfelder enthalten keine Terme, die die Entropie ermitteln, sondern sie werden meist in Kombination mit Simulationsmethoden, wie Monte-Carlo (MC) oder Moleküldynamik (MD) Simulation, verwendet. In diesen wird auch das Lösungsmittel entweder durch explizite Modellierung der Wassermoleküle oder impliziert, z. B. durch die Verwendung einer Dielektrizitätskonstante, berücksichtigt. Somit werden entropischen Beiträge, die durch die Dynamik der Moleküle selbst und durch die Interaktion der Moleküle mit dem Lösungsmittel entstehen, in die Berechnung miteinbezogen.

Eine der ersten kraftfeldbasierten Bewertungsfunktionen wurde im Zusammenhang mit dem Docking-Programm GOLD entwickelt [87, 88]. In der GoldScore-Funktion wurden Terme und Parameter aus dem Tripos Kraftfeld [114] zur Berechnung der Van-der-Waals-Energie und des Torsionswinkelpotenzials des Liganden übernommen. Zusätzlich enthält die GoldScore-Funktion einen Term zur Ermittlung der Wasserstoffbrücken-Energie und eine Abwandlung des Lennard-Jones-Potenzials (hier: (8,4)-Potenzial) zur Abschätzung der Van-der-Waals-Energie zwischen Ligand und Protein:

$$\begin{aligned} \text{GoldScore} &= -E_{\text{H-bond}} - (E_{\text{Internal}} + E_{\text{Complex}}) \\ &= -E_{\text{pair}} - (E_{\text{vdW, internal}} + E_{\text{torsional}} + E_{\text{vdW, external}}) \end{aligned} \quad (2.4)$$

Dabei werden die einzelnen Terme wie folgt berechnet:

$$E_{\text{pair}} = (E_{\text{da}} + E_{\text{WW}}) - (E_{\text{dW}} + E_{\text{aW}})$$

$$E_{\text{vdW, internal}} = \frac{C}{d_{ij}^{12}} - \frac{D}{d_{ij}^6}$$

$$E_{\text{torsional}} = \frac{1}{2} V_{ijkl} \left[1 + \frac{n_{ijkl}}{|n_{ijkl}|} \cos(|n_{ijkl}| \cdot \omega_{ijkl}) \right]$$

$$E_{\text{vdW, external}} = \frac{A}{d_{ij}^8} - \frac{B}{d_{ij}^4}$$

Der Wasserstoffbrücken-Term E_{pair} berücksichtigt auch die Desolvatisierung (w) von Donor (d) und Akzeptor (a) bei der Bildung einer Wasserstoffbrücke. Die Parameter wurden für verschiedene Donor- und Akzeptor-Typen mithilfe quantenchemischer Methoden bestimmt [87, 88]. Das (8,4)-Potenzial zur Abschätzung der Van-der-Waals-Energie zwischen Ligand und Protein wurde verwendet, da es weicher ist als das übliche (12,6)-Potenzial, d. h. es steigt nicht so schnell an und lässt damit eine leichte Überlappung zwischen Protein- und Ligandatomen zu.

Im Gegensatz zu den Kraftfeldmethoden werden die kraftfeldbasierten Bewertungsfunktionen in der Regel nur auf eine einzige Konformation eines Protein-Ligand-Komplexes angewendet. Die Energie wird daher nicht als Mittelwert eines Ensembles berechnet. Die Dynamik der Moleküle und der Einfluss des Lösungsmittels geht somit verloren. Verwendet werden meist nur die Terme, die sich auf die intermolekularen Wechselwirkungen beziehen, d. h. der Lennard-Jones-Term zur Abschätzung der Van-der-Waals-Energie und das Columb-Potenzial, um den Beitrag elektrostatischer Interaktionen zu ermitteln. Damit werden nur enthalpische Beiträge integriert, die zudem eine Vakuumenergie darstellen. Daher neigen kraftfeldbasierte Bewertungsfunktionen dazu, die Bindungsaffinität von großen, polaren Molekülen zu überschätzen. Kraftfeldbasierte Bewertungsfunktionen sind eher zur Bestimmung der Ligandkonformation im Docking geeignet als zur Vorhersage von Bindungsaffinitäten oder zur Unterscheidung von aktiven und inaktiven Molekülen eines Zielproteins [7, 115]. Zur Verbesserung der Vorhersagequalität von kraftfeldbasierten Bewertungsfunktionen wurden zusätzliche Terme für entropische Beiträge eingeführt oder die Abschätzung der Wasserstoffbrücken-Energie verfeinert (z. B. in Medusa [116]) als auch eine Kalibrierung der Funktion an experimentell gemessenen Bindungsaffinitäten durchgeführt (z. B. in Autodock [106] oder bei der RankScore-Funktion [117]).

Empirische Bewertungsfunktionen

Empirische oder *regressionsbasierte* Bewertungsfunktionen sind begründet auf der Annahme, dass die freie Bindungsenthalpie über die nichtkovalenten, intermolekularen Wechselwirkungen zwischen Protein und Ligand abgeleitet werden kann. Dazu werden Interaktionen/Deskriptoren bestimmt, von denen angenommen wird, dass sie den

größten Beitrag zur Bindungsaffinität leisten. Anhand von experimentell bestimmten Bindungsaffinitäten und Kristallstrukturen von Protein-Ligand-Komplexen wird über statistische Verfahren eine Gewichtung der einzelnen Interaktions-Terme bestimmt. Je nach Funktion unterscheiden sich diese Terme und auch die Regressionsverfahren, die zur Bestimmung der Koeffizienten eingesetzt wurden. Damit sind empirische Bewertungsfunktionen sehr ähnlich zu QSAR-Methoden [19, 20], in denen die Struktur oder die Eigenschaften eines Moleküls in Beziehung zu seiner Aktivität gesetzt wird. Als Interaktionen werden in den meisten Funktionen Wasserstoffbrücken, ionische Wechselwirkungen, hydrophobe Kontakte, Van-der-Waals-Energie oder Metallinteraktionen berücksichtigt. Aber auch entropische Beiträge, wie das Einfrieren von Bindungen durch die Komplexbildung oder Desolvatisierungsbeiträge, meist über die Änderung der apolaren Oberfläche modelliert, werden in empirische Bewertungsfunktionen miteinbezogen. Die Interaktionen oder Deskriptoren werden über geometrische oder strukturelle Eigenschaften der Protein-Ligand-Komplexe bestimmt (vgl. Abschnitt 2.3).

Die erste empirische Bewertungsfunktion SCORE1 [118] wurde von Böhm im Zusammenhang mit dem *de novo*-Design Programm LUDI [119, 120] entwickelt. Die Funktion berücksichtigt die energetischen Beiträge von Wasserstoffbrücken ΔG_{hb} , ionische Wechselwirkungen ΔG_{ionic} , die Änderung der apolaren Oberfläche ΔG_{lipo} und das Einfrieren von frei drehbaren Bindungen durch die Komplexbildung ΔG_{rot} :

$$\begin{aligned} \Delta G_{\text{binding}} = \Delta G_0 + \Delta G_{\text{hb}} \sum_{\text{h-bonds}} f(\Delta R, \Delta \alpha) + \Delta G_{\text{ionic}} \sum_{\text{ionic int.}} f(\Delta R, \Delta \alpha) \\ + \Delta G_{\text{lipo}} |A_{\text{lipo}}| + \Delta G_{\text{rot}} NROT \end{aligned} \quad (2.5)$$

Dabei ist $f(\Delta R, \Delta \alpha)$ eine Skalierungsfunktion, die die Abweichung von der idealen Wechselwirkungsgeometrie bestraft, und $NROT$ die Anzahl der frei drehbaren Bindungen, d. h. alle azyklischen sp^3 - sp^3 - und sp^3 - sp^2 -Bindungen. Die Gewichte ΔG_i der einzelnen Terme wurden mittels einer Regressionsanalyse auf einem Datensatz von 45 Protein-Ligand-Komplexen mit gemessenen K_i -Werten bestimmt. Dabei wurde die Regressionskonstante ΔG_0 als entropischer Beitrag zur Bindungsenergie interpretiert, der durch den Wegfall von Freiheitsgraden der Translation und Rotation des Liganden entsteht.

Von dieser Funktionen wurden viele weitere empirische Bewertungsfunktionen abgeleitet, z. B. die FlexScore-Funktion in dem Docking-Programm FlexX [80], die im Gegensatz zur SCORE1-Funktion hydrophobe und aromatische Kontakte über spezielle Wechselwirkungsgeometrien modelliert, oder die ChemScore-Funktion von Eldridge *et*

al. [121], die keinen extra Term für ionische Wechselwirkungen hat, dafür aber Metallinteraktionen bewertet. Die ChemScore-Funktion wiederum wird in vielen Docking-Programmen verwendet, z. B. in GOLD [122]), PRO_LEADS [123] oder FRED [124], und vielfach auch mit anderen oder zusätzlichen Termen variiert z. B. in der GlideSP- und GlideXP-Bewertungsfunktion in Glide [93, 125] oder in der X-Score-Bewertungsfunktion [126].

Empirische Bewertungsfunktionen unterscheiden sich nicht nur in der genauen Berechnung der Terme oder den Beiträgen, die modelliert werden, sondern auch teilweise stark in der Anzahl der Terme oder Deskriptoren. Die PLP (*Pairwise Linear Potenzial*) Bewertungsfunktion [127] besteht lediglich aus zwei Termen zur Berechnung der Wasserstoffbrücken- und der sterischen Energie zwischen zwei Atomen. Sie basiert auf vier Atomtypen und wurde auf nur drei Protein-Ligand-Komplexen parametrisiert. Im Gegensatz dazu besteht die kürzlich publizierte PHOENIX-Bewertungsfunktion aus 42 Deskriptoren und wurde auf einem Datensatz aus 112 hochaufgelösten Kristallstrukturen mit kalorimetrisch gemessenen Bindungsenergien kalibriert [128]. Dennoch war die Vorsagequalität bezüglich experimentell bestimmter Bindungsaffinitäten beider Funktionen annähernd gleich gut [7, 128].

Im Vergleich zu den kraftfeldbasierten Methoden erzielen empirische Bewertungsfunktionen meist bessere Ergebnisse, was die Vorhersage von Bindungsaffinitäten und die Trennung von aktiven und inaktiven Molekülen betrifft [6, 7, 129]. Jedoch ist auch bekannt, dass empirische Bewertungsfunktionen deutlich zuverlässigere Vorhersagen für Protein-Ligand-Komplexe erzielen, die ähnlich zu den Komplexen im Kalibrierungsdatensatz sind [3, 4, 42, 99, 130]. Die Vorsagequalität von empirischen Bewertungsfunktionen ist stark von der Qualität, der Größe und der Diversität des Kalibrierungsdatensatzes abhängig. In einer Zusammenarbeit, dem *Scoring Function Consortium* (SFC), zwischen verschiedenen akademischen Arbeitskreisen, pharmazeutischen Unternehmen und dem Cambridge Crystallographic Data Centre wurde dementsprechend ein Datensatz, bestehend aus 855 verschiedenen, hochaufgelösten Protein-Ligand-Strukturen zusammengestellt [131]. Aus über 60 verschiedenen Deskriptoren wurden acht neue Bewertungsfunktionen, die SFCscore-Funktionen, auf Teilmengen des Datensatzes entwickelt. Diese wurden intensiv validiert und erzielten im Vergleich zu anderen Bewertungsfunktionen bessere Resultate in der Vorhersage von Bindungsaffinitäten [131].

Eine weitere Schwierigkeit bei der Entwicklung von empirischen Bewertungsfunktionen oder Bewertungsfunktionen im Allgemeinen ist die Berücksichtigung von ungünstigen Beiträgen zur Bindungsenergie [44, 132]. Dies ist vor allem wichtig zur Trennung von aktiven und inaktiven Molekülen und wird als eine der bestehenden Herausforderungen

weiter unten in Abschnitt 2.5.2 diskutiert.

Wissensbasierte Bewertungsfunktionen

Die stetig steigende Anzahl aufgeklärter Kristallstrukturen von Protein-Ligand-Komplexen ermöglicht die statistische Analyse bevorzugter Distanzen/Kontakte verschiedener Atomtypen. Dabei ist die Annahme, dass statistisch häufig auftretende Atomtyp-Paare mit geringer Distanz energetisch günstig sein müssen. Diese Idee, ursprünglich im Kontext der Proteinstrukturvorhersage verwendet [133], wurde schnell auf die Vorhersage von Protein-Ligand-Komplexen übertragen [134]. In wissensbasierten Bewertungsfunktionen wird mittels der Boltzmann-Statistik aus der Häufigkeitsverteilung der verschiedenen Atomtyp-Paare ein Potenzial, das sogenannte *potential of mean force*, abgeleitet. Im Gegensatz zu empirischen Bewertungsfunktionen werden zur Kalibrierung nur Strukturdaten benötigt, so dass die Anzahl der Fehlerquellen reduziert wird und in der Regel eine größere Menge an Daten verwendet werden kann.

Einer der bekanntesten Repräsentanten von wissensbasierten Bewertungsfunktionen ist die PMF-Funktion, die von Muegge und Martin entwickelt wurde [63]. In dieser Funktion wurden distanzabhängige Potenziale $A_{ij}(r)$ aus der Häufigkeitsverteilung von Atomtyp-Paaren ij abgeleitet:

$$A_{ij}(r) = -k_B T \ln \left[f_{\text{Vol_corr}}^j(r) \frac{p_{\text{seg}}^{ij}(r)}{p_{\text{bulk}}^{ij}} \right]$$

$$\text{PMF_score} = \sum_{\substack{kl \\ r < r_{\text{cut-off}}^{ij}}} A_{ij}(r) \quad (2.6)$$

Das Potenzial A_{ij} ist eine Funktion über die Distanz r von zwei Atomtypen i und j . k_B ist die Boltzmann-Konstante und T die absolute Temperatur in Kelvin. $p_{\text{seg}}^{ij}(r)$ ist die Anzahl der Atomtyp-Paare ij des Parametrisierungsdatensatzes, in einem bestimmten Raumsegment seg , das über den Abstand r der Atomtypen i und j definiert wird. Diese wird durch einen Referenzzustand p_{bulk}^{ij} normalisiert, der die Anzahl des entsprechenden Atomtyp-Paares im übrigen Raum darstellt. Der Korrekturfaktor $f_{\text{Vol_corr}}^j(r)$ wurde eingeführt, um den Raum zu berücksichtigen, der durch die Ligandatome selbst besetzt ist, d. h. dort können keine Interaktionen stattfinden. Die PMF-Energie eines Protein-Ligand-Komplexes wird als Summe der Potenziale $A_{ij}(r)$ über alle Protein-Ligand-Atompaare kl vom Atomtyp ij mit einem Abstand r berechnet. Zur Kalibrierung

der PMF-Funktion wurden 16 unterschiedliche Proteinatomtypen und 34 Ligandatotypen definiert. Als Parametrisierungsdatensatz wurden 697 Protein-Ligand-Komplexe aus der PDB zusammengestellt. Daraus resultierten 294 statistisch signifikante Potenziale für die verschiedenen Atomtyp-Paare. Die PMF-Funktion wurde 2005 von Muegge auf einem ungefähr zehnfach größeren Datensatz, bestehend aus 7152 Protein-Ligand-Komplexen der PDB, rekaliert, so dass auch Metall- und Halogen-Wechselwirkungen statistisch signifikante Potenziale lieferten [135].

Wissensbasierte Bewertungsfunktionen unterscheiden sich vor allem in der Definition des Referenzzustandes und der verwendeten Atomtypen, aber auch in der Größe des Parametrisierungsdatensatzes und in der genauen Form der Funktion. Neben der PMF-Funktion, gehören die DrugScore- [64, 136, 137] und die ASP-Funktion [138] zu den etabliertesten, wissensbasierten Bewertungsfunktionen. Die DrugScore-Funktion enthält ein zusätzliches Potenzial, das von der lösungsmittelzugänglichen Oberfläche abhängt, und der Referenzzustand bezieht sich nicht auf ein entsprechendes Atomtyp-Paar, sondern auf die mittlere Anzahl irgendeines Kontaktpaares. Ursprünglich wurde die Funktion auf 1376 Protein-Ligand-Komplexen, die aus der ReLiBase [68, 139] entnommen wurden, kalibriert [64]. Eine weitere Version wurde auf Kristallstrukturen von kleinen Molekülen aus der *Cambridge Structural Database* (CSD) parametrisiert, wodurch deutlich mehr statistisch signifikante Potenziale für verschiedene, auch seltenere Atomtypen, wie z. B. Halogene, abgeleitet werden konnten. Diese Reparametrisierung führte zu besseren Resultaten in der Vorhersage der bioaktiven Konformation eines Liganden und der Bindungsaffinitäten [136]. Kürzlich wurde eine überarbeitete Version der DrugScore-Funktion, die DSX-Funktion, veröffentlicht, die zusätzlich auch noch Torsionswinkelpotenziale enthält und in der Referenzzustand neu definiert wurde [137]. Wissensbasierte Bewertungsfunktionen zeigen vor allem in der Vorhersage der bioaktiven Konformation sehr gute Resultate, jedoch liefern sie in der Abschätzung von Bindungsaffinitäten keine Verbesserung im Vergleich zu den empirischen Bewertungsfunktionen [7, 137]. In der Trennung von aktiven und inaktiven Molekülen erzielen eher empirische Bewertungsfunktionen ein besseres Ergebnis [7, 115]. Eine Schwierigkeit in der Ableitung wissensbasierter Bewertungsfunktionen ist die Unterrepräsentierung mancher Kontaktpaare, die zu keinem statistisch signifikanten Potenzial führt. Dies trifft auch auf ungünstige Kontaktpaare zu, die seltener in Kristallstrukturen von Protein-Ligand-Komplexen auftreten, jedoch wichtig zur Unterscheidung von aktiven und inaktiven Molekülen sind.

2.5.2 Bestehende Herausforderungen

Die Abschätzung der Bindungsaffinität eines Protein-Ligand-Komplexes stellt, wie bereits beschrieben, immer noch eine der größten Herausforderung im strukturbasierten Wirkstoffentwurf dar. Viele verschiedene Effekte, wie z. B. die Desolvatisierung der Moleküle bei der Bindung oder die Flexibilität beider Moleküle, beeinflussen die Bindung, insbesondere die Bindungsstärke, eines Liganden an ein Protein. Nicht alle Einflüsse sind derzeit vollständig verstanden und daher korrekt in den Protein-Ligand-Bewertungsfunktionen abgebildet, woraus unzuverlässige Vorhersagen der Bindungsaffinität resultieren.

Vor allem die Berücksichtigung der umgebenden Wassermoleküle ist ein grundlegendes Problem in der Abschätzung der Bindungsaffinität. Bei der Komplexbildung werden die Wassermoleküle aus der Bindetasche des Proteins verdrängt, die Wassermoleküle, die den Liganden umgeben haben, werden abgestreift. Dies verursacht eine Umstrukturierung des Wassernetzwerkes und damit eine Änderung der Enthalpie als auch der Entropie. Die Desolvatisierung hydrophober Atome der Moleküle resultiert in einem günstigen energetischen Beitrag, während das Aufbrechen von Wasserstoffbrücken zwischen polaren Atomen und Wassermolekülen ungünstig beiträgt. Die Modellierung, aber vor allem auch die Quantifizierung dieser Beiträge ist immer noch eine Herausforderung in der Entwicklung einer Bewertungsfunktion. Während die ersten Bewertungsfunktionen diese Beiträge ganz vernachlässigt haben oder nur den Gewinn des hydrophoben Effektes meist über die Änderung der apolaren Oberflächen modelliert haben, wird zum Teil in neueren Bewertungsfunktionen auch die ungünstige polare Desolvatisierung berücksichtigt. Werden Desolvatisierungsbeiträge polarer Atome vernachlässigt, so wird die Bindungsaffinität von polaren und geladenen Molekülen in der Regel überschätzt.

Die erste Bewertungsfunktion, die einen Term für die polare Desolvatisierung integrierte, war die Hammerhead-Bewertungsfunktion [140]. Dieser Term wurde über die maximal mögliche Anzahl an Wasserstoffbrücken (Anzahl Donoren und Akzeptoren in der Protein-Ligand-Kontaktfläche) und der tatsächlich eingegangenen Anzahl an Wasserstoffbrücken modelliert. Hierbei stellte die Quantifizierung das größte Problem dar, in der Kalibrierung auf Protein-Ligand-Komplexen und experimentellen Affinitäten wurde dieser Term kaum gewichtet. Aufgrund dessen wurden in einer Überarbeitung dieser Funktion "negative Daten", d. h. durch Docking von inaktiven Molekülen generierte artifizielle Komplexe, mit in die Parametrisierung einbezogen, um eine bessere Gewichtung von Termen, die ungünstige Beiträge repräsentieren, zu erhalten [141]. In der Bewertungsfunktion des Docking-Programms Autodock4 [89] wurde ebenfalls ein Term zur

Beschreibung der Desolvatisierung einbezogen, der zuerst eine Kombination aus apolarem und polarem Beitrag darstellte und über einen volumenbasierten Ansatz berechnet wurde. Dieser Term konnte wiederum nicht parametrisiert werden, wodurch der polare Anteil separat behandelt wurde, indem anhand des Kalibrierungsdatensatzes eine Abschätzung über die Energie einer Wasserstoffbrücke in Wasser und die Energie einer intermolekularen Wasserstoffbrücke gemacht wurde. Der Beitrag der Desolvatisierung eines polaren Atoms resultiert aus der Differenz dieser Abschätzungen, wodurch sich ein relativ geringer Wert von $0,2 \text{ kcal/mol}$ ergab [89]. Diese Modellierung wurde in einer nachfolgenden Version überarbeitet und wieder durch einen volumenbasierten Ansatz ersetzt, der mithilfe von experimentellen Solvatisierungsenergien und Partialladungen der Atome kalibriert wurde, wodurch ein Beitrag für polare Atome von $0,6 \text{ kcal/mol}$ resultierte [106]. Im HINT-Kraftfeld [142] werden die Wechselwirkungsbeiträge von Atomen über atombasierte logP-Werte abgeschätzt. Je nach Art des Kontaktes wird über eine logische Funktion das Vorzeichen verändert, so dass hydrophobe Kontakte und Wasserstoffbrücken belohnt werden, während nicht komplementäre Kontakte (polar-polar oder polar-apolar) ungünstig zur Energie beitragen [142]. In der LigScore-Funktion [143] werden polare und apolare Atome auf Basis der Partialladung unterschieden. Die Desolvatisierung von polaren Atomen durch die Komplexbildung wird berücksichtigt und durch einen oberflächenabhängigen Term bestraft, der auf Protein-Ligand-Komplexen und experimentellen Bindungsaffinitäten kalibriert wurde. Dieser Term leistete in der Validierung der Bewertungsfunktion je nach Komplex einen deutlichen Beitrag zur Bindungsenergie [143]. Die GlideXP-Bewertungsfunktion modelliert die polare Desolvatisierung durch einen gitterbasierten Ansatz, wobei polare Atome, die im Komplex nicht hinreichend solvatisiert sind, bestraft werden [125]. In die Bewertungsfunktion von DOCK 3.5.54 wurde ein Term zur Berechnung der kontextabhängigen Ligand-Desolvatisierungsenergie eingefügt [144], während vorher die vollständige Desolvatisierung des Liganden angenommen und über die Born-Gleichung abgeschätzt wurde [145]. Diese kontextabhängige Berechnung führt zu einer Verbesserung der Resultate im virtuellen Screening [144].

Neben den ungünstigen energetischen Beiträgen, die durch die Desolvatisierung von polaren Atomen hervorgerufen werden, entsteht durch die Komplexbildung beider Moleküle auch ein Entropieverlust, der zu einer Verringerung der Bindungsenergie führt. Dieser wird meist über die Anzahl der frei drehbaren Bindungen des Liganden abgeschätzt. Der Verlust an Entropie, durch Einschränkung von Translations- und Rotationsfreiheit des Liganden, wird in der Regel vernachlässigt [99]. Ebenso wie bei der Modellierung der polaren Desolvatisierung, ist die Quantifizierung dieses Beitrags schwierig [42].

Ein weiteres Problem stellt die Annahme der Additivität der einzelnen Beiträge dar.

Während sie zum Teil für enthalpische Beiträge als erste Näherung benutzt werden kann, trifft dies nicht auf die Entropie und die freie Energie zu [42]. Selbst in der Betrachtung von intermolekularen Wechselwirkungen ist nicht unbedingt eine einfache Additivität gegeben. In verschiedenen experimentellen Studien wurde eine Kooperativität von verschiedenen Wechselwirkungen gezeigt [42, 44]. Die Modellierung einer solchen Abhängigkeit von unterschiedlichen Interaktionen ist ein schwieriges und bisher kaum untersuchtes Problem. In der GlideXP-Funktion wird versucht über verschiedene spezielle Terme (besondere Anordnungen von Wasserstoffbrücken, hydrophobe Komplementarität) solche Effekte zu berücksichtigen [125]. Die neue empirische Bewertungsfunktion ScorpionScore konzentriert sich auf die Vorhersage und energetische Abschätzung solcher kooperativen Interaktionen, indem sie nicht einzelne Wechselwirkungen modelliert, sondern ein Interaktionsnetzwerk berechnet [146].

Neben diesen schwer zu erfassenden Effekten ist selbst die Abschätzung der Energie und die Modellierung "einfacher" intermolekularer Wechselwirkungen immer noch nicht vollständig gelöst. Gerade der energetische Beitrag einer Wasserstoffbrücke zur Bindungsenergie, der durch Desolvatisierungseffekte beeinflusst wird, variiert teilweise stark [44, 147]. In den meisten Bewertungsfunktionen wird dieser Betrag jedoch konstant und additiv modelliert. Dadurch werden insbesondere polare Moleküle, die multifurkate Wasserstoffbrücken ausbilden, überschätzt. In der AIScore-Bewertungsfunktion [148] wird der Beitrag von Wasserstoffbrücken zwischen verschiedenen Donoren und Akzeptoren unterschieden, zudem wurde eine besondere Behandlung von multifurkaten Wasserstoffbrücken eingeführt. Die energetischen Beiträge der Donoren und Akzeptoren wurden dabei über quantenchemische Berechnungen kalibriert. Andere seltenere Wechselwirkungen, wie Kation- π -Interaktionen oder Halogen-Brücken, werden in den meisten Bewertungsfunktionen nicht abgebildet, da bislang zu wenige Daten darüber vorliegen [99]. Eine besondere Herausforderung bei der Modellierung von Interaktionen stellt die Metallkoordinierung dar, die durch die große Anzahl an Metalloenzymen häufig vorkommt, jedoch in den meisten Bewertungsfunktionen gar nicht oder nur unzureichend abgebildet wird. Im Gegensatz zu Kation- π oder Halogen-Brücken, leisten Metallinteraktionen jedoch einen großen Beitrag zur Bindungsenergie, so dass die Vernachlässigung dieser Interaktionen meist zu einer unzuverlässigen Vorhersage führt [99]. Während Modelle zur Beschreibung der Koordinierungsgeometrie von Metallionen existieren [51, 82, 83, 149], so ist die Quantifizierung des Beitrags einer Metallinteraktion in einem Protein-Ligand-Komplex kaum möglich (vgl. Abschnitt 2.3.2).

2.5.3 Alternative Ansätze

Aufgrund der Schwächen der verschiedenen Bewertungsfunktionen, die in den vorherigen Abschnitten 2.5.1 und 2.5.2 diskutiert wurden, und den damit verbundenen unzuverlässigen Vorhersagen, wurden alternative Ansätze entwickelt, die zur Verbesserung mancher Aspekte führen, wiederum jedoch auch Defizite in anderer Hinsicht aufweisen.

Konsensus-Bewertung

In verschiedenen Vergleichsstudien (z. B. [5–8, 150]), die in den letzten Jahren immer wieder in Bezug auf die unterschiedlichen Bewertungsfunktionen und ihre Resultate in den verschiedenen Anwendungen gemacht wurden, hat sich keine der Bewertungsfunktionen herauskristallisiert, die immer in einer oder mehreren Disziplinen bessere Ergebnisse als andere geliefert hat. Demzufolge kann eine Kombination verschiedener Bewertungsfunktionen zu einem zuverlässigeren Ergebnis führen. Dieser Ansatz, die *Konsensus-Bewertung*, wurde von Charifson *et al.* [151] vorgeschlagen, die in ihrer Studie zwei Docking-Methoden und dreizehn Bewertungsfunktionen kombinierten und auf drei verschiedenen Zielproteinen in einem virtuellen Screening testeten. Die Kombination aus ChemScore, der AMBER-basierten Funktion von DOCK und der PLP-Funktion erzielte eine signifikante Verbesserung in dem virtuellen Screening.

In der Regel wird in einem Konsensus-Ansatz pro Ligand eine Anzahl an Docking-Posen generiert und mit verschiedenen Bewertungsfunktionen nachbewertet. Anschließend werden die Resultate der Funktionen kombiniert, wobei auch die Kombinationsmethoden großen Einfluss auf das Endergebnis haben [152]. Um die beste Kombination zu finden, werden teilweise auch aufwendige maschinelle Lernverfahren verwendet [153, 154]. Eine andere Möglichkeit ist eine direkte Verknüpfung von einzelnen Termen unterschiedlicher Bewertungsfunktionen miteinander [155].

In einer Konsensus-Bewertung gehen jedoch nicht nur die Stärken, sondern auch die Schwächen der jeweiligen Bewertungsfunktionen mit in das Ergebnis ein. In einer Studie von Yang *et al.* [156] wurde gezeigt, dass eine Konsensus-Bewertung nur zu einem besseren Resultat führt, wenn die einzelnen Funktionen selbst schon gute Ergebnisse liefern. Weiterhin muss der deutlich höhere Zeit- und Rechenaufwand berücksichtigt werden, der durch die Verwendung einer Konsensus-Bewertung entsteht.

Zielproteinspezifische Bewertung

Eine weitere Alternative ist die spezifische Ausrichtung einer Bewertungsfunktion auf ein bestimmtes Zielprotein hin. Dies kann durch eine Rekalibrierung oder Erweiterung einer bereits bestehenden Bewertungsfunktion, empirisch oder wissensbasiert, erreicht

werden. Eine Übersicht verschiedener Methoden und Funktionen findet sich in [157]. Eine Voraussetzung für diesen Ansatz ist die Kenntnis über bereits identifizierte, aktive Moleküle eines Zielproteins. Im besten Fall sind die Kristallstrukturen mehrerer Komplexe und die Affinitätsdaten gegeben, so dass die Bewertungsfunktion darauf parametrisiert werden kann. Daher kommt diese Methode meist nur für bereits gut untersuchte Zielproteine in Frage.

Die wissensbasierte Bewertungsfunktion ASP [138] beispielsweise wurde auf einem diversen Datensatz von Protein-Ligand-Komplexen parametrisiert, der zusätzlich mit einer großen Anzahl verschiedener CDK2-Komplexe ergänzt wurde. Die resultierenden Funktionen zeigen eine deutliche Verbesserung im virtuellen Screening und in der Vorhersage der bioaktiven Konformation von CDK2-Inhibitoren [138]. Basierend auf der PMF-Funktion [135] wurde eine neue, Kinase-spezifische Funktion (kinase-PMF) von 872 Kinase-Strukturen abgeleitet, die im Vergleich zu allgemeinen Bewertungsfunktionen bessere Vorhersagen im Docking und im virtuellen Screening für verschiedene Kinasen liefert [158]. In der ISAC-Methode [159] werden neben den unterschiedlichen Bindungsmodi verschiedener Liganden eines Zielproteins auch die Aktivitätsschwankungen, die durch kleine strukturelle Veränderungen des Liganden entstehen, die sogenannten *activity cliffs*, bei der Kalibrierung der Funktion ausgenutzt. Dadurch ist es möglich, in der Bindetasche des Proteins die Reste zu identifizieren, die an diesen Schwankungen beteiligt sind und so zusätzlich zu einer zielproteinspezifischen Bewertungsfunktion auch Pharmakophor-Hypothesen für das Protein abzuleiten. Um unabhängiger von bereits aufgeklärten Komplexstrukturen zu sein, werden in manchen Ansätzen auch Docking-Posen bekannter aktiver Moleküle zur Ableitung einer zielproteinspezifischen Bewertung herangezogen [160, 161].

Dieser alternative Ansatz hat den Nachteil, dass bereits viele Informationen, zumindest einige aktive Moleküle, über das Zielprotein bekannt sein müssen, um eine deutliche Verbesserung der Vorhersage zu erzielen.

Simulationsmethoden

Eine aus thermodynamischen Gesichtspunkten präzisere Vorhersage von relativen freien Bindungsenergien verschiedener Liganden eines Zielproteins wird mit sogenannten *free-energy-perturbation* (FEP) Berechnungen oder *thermodynamischer Integration* (TI) erzielt [11, 42]. Dabei wird der Unterschied der freien Energie berechnet, der durch die Transformation eines Liganden in einen ähnlichen Liganden entsteht. In dieser Methode werden MC- oder MD-Simulationen benötigt, um ein Konfigurations-Ensemble des Protein-Ligand-Komplexes und des ungebundenen Liganden zu erhalten, auf dem die FEP- oder TI-Berechnung durchgeführt werden kann. In diesen Methoden wird die

Flexibilität beider Moleküle, Protein und Ligand, und der Einfluss des Lösungsmittels durch die explizite Modellierung von Wassermolekülen in der Simulation berücksichtigt. Dies liefert in der Regel eine bessere Vorhersage der Bindungsenergie eines Protein-Ligand-Komplexes als mit einer Bewertungsfunktion [11, 104, 162], jedoch ist diese Berechnung sehr zeit- und rechenintensiv. Weiterhin ist es nur möglich für relativ ähnliche Liganden eine zuverlässige Vorhersage der relativen Bindungsenergie zu machen, so dass die Methode nicht auf die diversen Liganden, die in einem virtuellen Screening betrachtet werden, angewendet werden kann [104]. Die Genauigkeit dieser Methoden ist zudem auch an das unterliegende Kraftfeld geknüpft [162].

Eine Alternative zu diesen sehr aufwendigen Berechnung wurde von Åqvist *et al* publiziert [163]. Die *linear interaction energy* (LIE) Methode erfordert nur eine Simulation des Ausgangs- und Endzustandes, d. h. ungebunden und gebundener Zustand, und berücksichtigt auch nur die Änderung der elektrostatischen und der Van-der-Waals-Energie [163]. Diese Methode ist im Vergleich zu einer Bewertungsfunktion immer noch sehr aufwendig, erlaubt aber die Betrachtung von diversen Liganden und erzielte schon vielversprechende Resultate [4, 11].

Eine andere Methode, die ebenfalls enthalpische als auch entropische Beiträge mit in die Abschätzung der freien Bindungsenergie eines Protein-Ligand-Komplexes einbezieht, ist die MM/PBSA- oder MM/GBSA-Berechnung. Dazu werden Kraftfeldterme (MM) mit einem Term für die polare Solvatisierung, *Poisson-Boltzmann* (PB) oder *Generalized Born* (GB), und einem Term für die lösungsmittelzugängliche Oberfläche (SA) kombiniert [164, 165]. Mittels einer Simulation werden Trajektorien der ungebundenen Moleküle und des Protein-Ligand-Komplexes erzeugt. Die freie Bindungsenergie des Protein-Ligand-Komplexes wird über die Energiedifferenz der ungebundenen und gebundenen Moleküle abgeschätzt, wobei in der Berechnung eine implizite Repräsentation des Lösungsmittels verwendet wird (PB oder GB). In verschiedenen Studien wurde eine gute Abschätzungen relativer Bindungsenergien ähnlicher Liganden eines Proteins durch Anwendung dieser Methodik gezeigt, jedoch existieren ebenso Studien, in denen diese aufwendige Berechnung keine besseren Resultate als eine einfache Bewertungsfunktion lieferte [4, 162].

Eine alternative Methode stellt auch die WaterMap-Berechnung dar [166, 167], die die Bindetasche eines Protein-Ligand-Komplexes mit Wassermolekülen füllt (der Ligand wird zuvor entfernt) und, mithilfe einer kurzen MD-Simulation und der sogenannten *inhomogeneous fluid solvation theory* [168], die entropischen und enthalpischen Beiträge jedes einzelnen Wassermoleküls in der Bindetasche berechnet. Die Bindungsenergie des Liganden wird über die Summe der Energien der verdrängten Wassermoleküle abgeschätzt. [167, 169].

Diese Simulationsmethoden liefern in der Regel gute Abschätzungen relativer Bindungsenergien von Protein-Ligand-Komplexen, jedoch sind sie zu aufwendig, um in großem Maßstab, z. B. im virtuellen Screening, genutzt zu werden. Zusätzlich ist ihre Anwendbarkeit meist auf sehr ähnliche Liganden beschränkt.

Die vorgestellten alternativen Ansätze eignen sich meist nicht in den frühen Phasen des strukturbasierten Wirkstoffentwurfs, in denen eine Vielzahl an Molekülen untersucht wird und häufig noch nicht viele Informationen über den Bindungsmodus oder bekannte aktive Moleküle vorliegen. Zudem weisen sie auch eigene Schwächen auf und liefern nicht in jedem Fall bessere Resultate als weniger aufwendige Protein-Ligand-Bewertungsfunktionen.

2.6 Strukturbasiertes virtuelles Screening

Die in den vorherigen Abschnitten vorgestellten Docking-Methoden 2.4 und Bewertungsfunktionen 2.5, bilden die Basis des strukturbasierten virtuellen Screenings. Viele verschiedene Schwächen, Herausforderungen und Erweiterungsmöglichkeiten dieser Ansätze wurden aufgezeigt und diskutiert, dennoch ist das virtuelle Screening ein etabliertes Verfahren im strukturbasierten Wirkstoffentwurf. Es dient als fortgeschrittener Filter und vielleicht auch als Ideengenerator zur Unterstützung eines erfahrenen Medizinalchemikers. Mit dem Wissen dieser noch bestehenden Herausforderungen sollte ein virtuelles Screening nicht als Blackbox angesehen werden, sondern je mehr Expertise und Wissen in die einzelnen Schritte eines virtuellen Screenings fließt, desto vielversprechender werden die Resultate. Wird es so verwendet, können sehr erfolgreiche Resultate erzielt werden, wie viele prospektive Studien zeigen. Eine Zusammenfassung dieser findet sich in [3, 4]. Die Erfolgsraten im virtuellen Screening liegen im Mittel deutlich über denen eines HTS-Experiments [2], zudem ist es kostengünstiger und weniger zeitintensiv als letzteres.

Ein strukturbasiertes virtuelles Screening kann grob in vier Schritte eingeteilt werden: Aufbereitung der Moleküle, Aufbereitung der Proteinstruktur, Docking und Bewertung der Moleküle und als letzten Schritt, die Nachbearbeitung und Analyse der Resultate durch einen Experten (vgl. Abbildung 2.6) [170]. Jeder Schritt ist dabei wichtig und beeinflusst direkt die Qualität der Resultate.

Aufbereitung der Moleküle

Die Moleküle, die in einem virtuellen Screening durchgemustert werden sollen, können

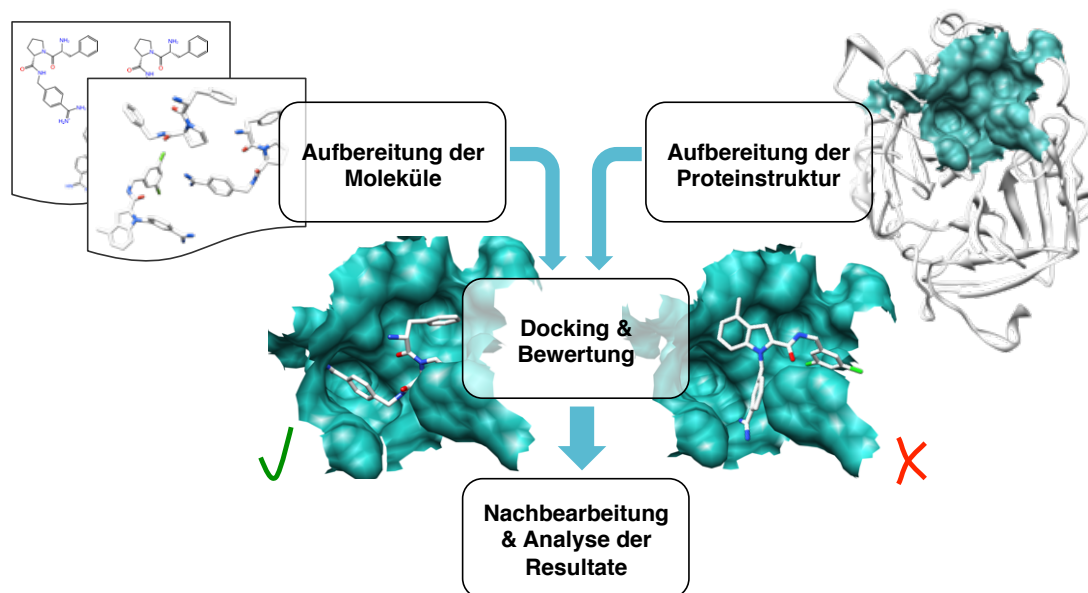


Abbildung 2.6: Typischer Ablauf eines virtuellen Screenings.

aus verschiedenen Quellen, wie hauseigene Moleküldatenbanken des Pharmaunternehmens oder aus Datenbanken von kommerziell erwerblichen Verbindungen (z. B. ZINC [171]), stammen. Sind schon viele Informationen über ein Zielprotein vorhanden (z. B. bekannte aktive Moleküle), können auch sogenannte *fokussierte Bibliotheken* erstellt werden, die nur interessante Moleküle für eine spezielle Proteinklasse (z. B. Kinasen oder GPCRs) enthalten. Eine weitere Möglichkeit, Moleküle für ein virtuelles Screening zu generieren, bieten auch *Fragmenträume*, die virtuelle kombinatorische Bibliotheken repräsentieren. Aus verschiedenen Molekülfragmenten werden dabei durch vorher definierte Regeln (Syntheseregeln) neue Moleküle aufgebaut. Eine ausführliche Zusammenfassung dieser Methoden ist in [172] zu finden. Bevor die Moleküle in einem virtuellen Screening durch eine Docking-Methode getestet werden, ist es sinnvoll, die Menge der Moleküle zu reduzieren [3]. Dazu können verschiedene Filter verwendet werden, um Moleküle, die sich nicht als Leitstruktur eignen, vorher auszusortieren, z. B. reaktive und toxische Moleküle. Weiterhin können die Moleküle nach pharmakophoren Eigenschaften oder nach sogenannter *druglikeness* [173] oder *leadlikeness* [174] gefiltert werden. Vor dem Docking müssen dreidimensionale Strukturen der Moleküle generiert werden. Je nach Docking-Programm sind dabei mehrere Konformationen der Moleküle erforderlich (vgl. Abschnitt 2.4). Weiterhin sollten verschiedene Protonierungszustände, Tautomere und Stereoisomere der Moleküle mit in das Docking einbezogen werden, wobei erste-

re vom pH-Wert abhängen und daher teilweise schwierig zu bestimmen sind [175] und auch nicht immer zu besseren Resultaten im Docking führen [100]. Außerdem sollte die chemische Korrektheit aller Moleküle überprüft werden, da sich in vielen Datenbanken auch fehlerhafte Strukturen befinden können [176] und Docking-Programme in der Regel sehr sensibel darauf reagieren [170].

Aufbereitung der Proteinstruktur

Ebenso wichtig für den Erfolg eines strukturbasierten virtuellen Screenings ist die Auswahl und Aufbereitung einer geeigneten Proteinstruktur. In Regel werden dabei Kristallstrukturen von Proteinen genutzt, aber auch Strukturen, die durch NMR-Spektroskopie bestimmt oder durch Homologie-Modellierung konstruiert wurden, können als Grundlage dienen. Dabei ist es wichtig die Strukturen vorher auf Korrektheit und Qualität zu überprüfen (vgl. Abschnitt 2.2). Ist die Bindetasche des Proteins noch nicht durch einen bereits gebundenen Liganden identifiziert, so ist es möglich diese mit verschiedenen Methoden vorherzusagen [177]. Weiterhin muss die Flexibilität eines Proteins in Betracht gezogen werden, dazu sollten, sofern vorhanden, verschiedene Protein-Ligand-Strukturen untersucht werden, um die relevanteste Konformation der Bindetasche zu bestimmen. Alternativ kann ein flexibler Docking-Ansatz verwendet werden. In der Bindetasche sollten auch alle wichtigen Kofaktoren, Metallionen und Wasser identifiziert und berücksichtigt werden. Die korrekte Bestimmung der Protonierungszustände und Ausrichtung der frei drehbaren Wasserstoffbrücken-Donoren hat ebenfalls maßgeblichen Einfluss auf die Resultate des Dockings. Ein Qualitätsmerkmal einer guten Struktur, ist auch ein vollständig ausgebildetes internes Wasserstoffbrücken-Netzwerk.

Docking & Bewertung

In verschiedenen Vergleichsstudien [5–7] wurde gezeigt, dass die unterschiedlichen Docking-Programme je nach Zielprotein in ihrer Vorhersagequalität stark variieren. Anhand dieser Studien kann eine Vorauswahl der Docking-Methode getroffen werden, jedoch sollte vor dem virtuellen Screening eine Validierung der Docking-Methode als auch der ausgewählten Proteinstruktur anhand eines *Redockings* durchgeführt werden. Dabei sollte der gebundene Ligand wieder in seiner bioaktiven Konformation in die Bindetasche platziert werden können und durch die verwendete Bewertungsfunktion korrekt identifiziert werden. Ist dies nicht der Fall, sollte die Proteinstruktur verworfen oder eine andere Docking-Methode in Betracht gezogen werden. Ein weiterer Test, der vor einem virtuellen Screening durchgeführt werden kann, ist die Identifizierung von bereits bekannten aktiven Molekülen durch das Docking und die Bewertung. Schlägt dieser Test fehl, so ist auch die Chance neue aktive Moleküle, die sich auch später im Experiment

nicht als Falsch-Positive herausstellen, eher gering. Auch die Behandlung von Proteinflexibilität, Wassermolekülen oder unterschiedlichen Protonierungszuständen durch die Docking-Methode und die Bewertungsfunktion sollten bei der Auswahl berücksichtigt werden. Ist die Zielprotein schon ausreichend untersucht, so können vor dem Docking auch Pharmakophor-Filter definiert werden, die festlegen welche Interaktionen im Protein durch den Liganden erfüllt werden müssen [178]. Weiterhin bietet die Generierung von mehreren Posen pro Ligand die Möglichkeit, im Anschluss an das Docking eine Nachbewertung mit einer oder mehreren anderen Bewertungsfunktionen durchzuführen.

Nachbearbeitung & Analyse der Resultate

Aufgrund der Probleme, die noch in Docking-Methoden und Bewertungsfunktionen existieren, ist es erforderlich die Docking-Ergebnisse durch zusätzliche Verfahren zu untersuchen und zu verifizieren. Häufig wird eine Nachbewertung durch eine aufwendiger zu berechnende Bewertungsfunktion durchgeführt oder auch eine Konsensus-Bewertung mit verschiedenen Bewertungsfunktionen (vgl. Abschnitt 2.5.3). Sind bereits interessante Docking-Resultate bestimmt, können rechenintensive Verfahren wie MM/PBSA (vgl. Abschnitt 2.5.3) angewendet werden [175]. Weiterhin können die Resultate durch sogenannte *Protein-Ligand Interaction Fingerprints* gefiltert werden [170, 175], diese beschreiben die Interaktionen zwischen Protein und Ligand durch einen einfachen binären Code, um die Liganden mit den gewünschten Interaktionen schnell zu finden. Eine weitere Hilfestellung kann durch eine intuitive Visualisierung der Ergebnisse erfolgen, da letztendlich zur finalen Auswahl der Moleküle, die synthetisiert und im Labor getestet werden sollen, die Expertise eines Medizinalchemikers immer noch unerlässlich ist.

3

Konzept hinter der HYDE-Funktion

Fast alle biochemischen Prozesse finden in wässriger Lösung statt. Dabei dient Wasser nicht nur als Lösungsmittel, sondern auch als Wechselwirkungspartner in vielen molekularen Komplexen. Die ungewöhnlichen Eigenschaften von Wasser faszinieren Wissenschaftler seit Jahrzehnten. Sie sind noch nicht vollständig erklärt, werden aber dem starken Wasserstoffbrücken-Netzwerk zwischen den Wassermolekülen zugeschrieben [179]. Die HYDE-Bewertungsfunktion basiert auf einem Konzept, das das Wasserstoffbrücken-Netzwerk in Wasser vereinfacht beschreibt, so dass die Energie von Wechselwirkungen funktioneller Gruppen mit Wasser als auch zwischen funktionellen Gruppen in Wasser abgeleitet werden kann [180, 181]. In HYDE wird das Konzept genutzt, um die Wasserstoffbrücken-Energie als auch den hydrophoben Effekt abzuschätzen. Des Weiteren wird eine konsistente Quantifizierung der ungünstigen Beiträge durch die Desolvatisierung von polaren Atomen möglich, die als eine Herausforderung im vorherigen Kapitel 2.5.2 diskutiert wurde. Dieses Konzept wurde von Gudrun Lange und Robert Klein entwickelt [180, 181] und wird in folgendem Kapitel vorgestellt.

3.1 Das Wasserstoffbrücken-Netzwerk in Wasser

Die Struktur des Wasserstoffbrücken-Netzwerkes von flüssigem Wasser wurde schon in vielen Studien theoretisch [182–185] und experimentell (Übersicht der Methoden in

[185]) untersucht. Für die theoretischen Methoden wurden verschiedene Modelle zur Beschreibung von Wasser entwickelt [186–188]. In vielen der Untersuchungen wurde dabei über die Anzahl der Wasserstoffbrücken, an denen ein Wassermolekül beteiligt ist, diskutiert. In der dreidimensionalen Struktur von gewöhnlichen hexagonalen Eiskristallen I_h formen die Wassermoleküle ein regelmäßiges Gitter und bilden in idealer tetraedrischer Geometrie vier Wasserstoffbrücken mit den vier benachbarten Wassermolekülen [189–191]. In flüssigem Wasser ist dagegen eine ungeordnetere Struktur vorzufinden, die ein breites, temperaturabhängiges Spektrum an Distanzen und Winkeln zwischen den Wassermolekülen aufweist, so dass nicht alle Wasserstoffbrücken zwischen den Wassermolekülen mit idealer Geometrie ausgebildet werden können [183, 184, 192, 193]. Die Wassermoleküle bilden ein sich ständig änderndes Netzwerk von transienten Wasserstoffbrücken, wobei einzelne Wasserstoffbrücken kontinuierlich gebrochen und gebildet werden. Bei einer bestimmten Temperatur bleibt der Anteil der gebrochenen und gebildeten Wasserstoffbrücken konstant [184]. Die Idee, dass nicht alle vier Wasserstoffbrücken in flüssigem Wasser gemacht werden, wurde schon 1960 von Pauling veröffentlicht [43]. Basierend auf dem Anteil der Schmelz- und Verdampfungsenthalpie, schätzte er, dass nur 85 % der Wasserstoffbrücken bei 273 K zwischen den Wassermolekülen ausgebildet werden.

In dem folgenden theoretischen Konzept, auf dem HYDE basiert, werden Wassermoleküle nicht individuell beschrieben, sondern das Wasserstoffbrücken-Netzwerk als Ganzes untersucht. Dabei besitzt jedes Wassermolekül vier sogenannte *Wasserstoffbrücken-Funktionen*, zwei Donor- und zwei Akzeptor-Funktionen, die in tetraedrischer Geometrie angeordnet sind. Der Zustand eines individuellen Wassermoleküls variiert zwischen vollständig abgesättigt ($f_{sat} = 1$), wenn alle seine Wasserstoffbrücken-Funktionen an idealen Wasserstoffbrücken beteiligt ist, und unabgesättigt ($f_{unsat} = 1$), wenn es keine Wasserstoffbrücke ausbilden kann.

3.1.1 Sättigung des Wasserstoffbrücken-Netzwerkes

Um die Sättigung des Wasserstoffbrücken-Netzwerkes im statistischen Ensemble von flüssigem Wasser zu beschreiben, wurde die Idee von Pauling aufgegriffen [43]. So kann mithilfe eines thermodynamischen Zyklus von Wasser und den folgenden Annahmen die temperaturabhängige Sättigung des Wasserstoffbrücken-Netzwerkes bestimmt werden:

- In hexagonalen Eiskristallen sind alle vier Wasserstoffbrücken-Funktionen aller Wassermoleküle vollständig abgesättigt [189–191]
- Im gasförmigen Zustand sind alle Wasserstoffbrücken zwischen den Wassermolekülen gebrochen

- Die Summe der Schmelzenthalpie (ΔH_{fusion}), der Wärmeenergie ($\Delta H_{273-373K}$), die benötigt wird, um Wasser von 273 K auf 373 K zu erhitzen, und die Verdampfungsenthalpie ($\Delta H_{evaporation}$) bildet die Energie, die benötigt wird, um alle vier Wasserstoffbrücken zwischen den Wassermolekülen aufzubrechen (Abbildung 3.1)
- Der Anteil der abgesättigten Wasserstoffbrücken-Funktionen in flüssigem Wasser ist proportional zur Enthalpie des Systems

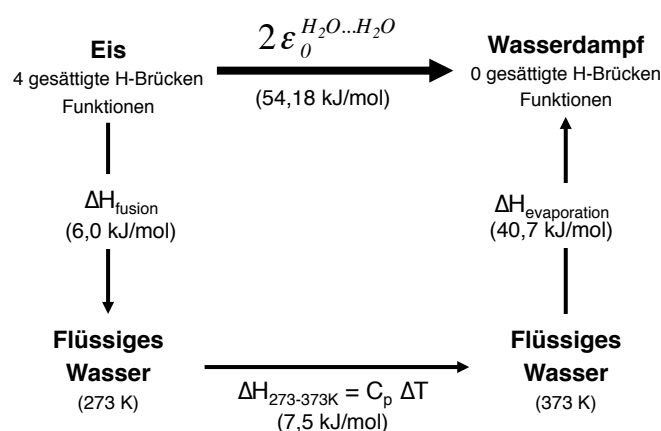


Abbildung 3.1: Thermodynamischer Zyklus von Wasser. Bild in Anlehnung an [181]

Die spezifische Wärmekapazität C_p von Wasser ist zwischen 273 K und 373 K konstant. Damit ergibt sich für die Wärmeenergie, um Wasser von 273 K auf 373 K bei konstantem Druck von 1000 hPa zu erhitzen: $\Delta H_{273-373K} = C_p \cdot \Delta T = 7,5 \text{ kJ/mol}$. Die Gesamtenergie, die benötigt wird, um alle vier Wasserstoffbrücken zwischen den Wassermolekülen aufzubrechen, berechnet sich wie folgt:

$$\Delta H_{fusion} + \Delta H_{273-373K} + \Delta H_{evaporation} = 2\epsilon_0^{H_2O...H_2O} = 54,18 \text{ kJ/mol} \quad (3.1)$$

Unter der Annahme, dass der Anteil der ungesättigten Wasserstoffbrücken-Funktionen $F_{unsat}(T)$ in flüssigem Wasser proportional zur zugeführten Wärme ist, kann dieser Anteil mit folgender Gleichung berechnet werden:

$$F_{unsat}(T) = \frac{\Delta H_{fusion} + C_p(T - 273K)}{\Delta H_{fusion} + \Delta H_{273-373K} + \Delta H_{evaporation}} \quad (3.2)$$

3. Konzept hinter der HYDE-Funktion

Den restlichen Anteil bilden folglich die *abgesättigten Wasserstoffbrücken-Funktionen* $F_{sat}(T)$ in flüssigem Wasser bei einer bestimmten Temperatur:

$$F_{sat}(T) = 1 - F_{unsat}(T) \quad (3.3)$$

Damit ergibt sich ein Anteil von 11 % unabgesättigten Wasserstoffbrücken-Funktionen im Wasserstoffbrücken-Netzwerk von flüssigem Wasser bei 273 K. Der prozentuale Anteil wächst auf ca. 17 % bei 313 K und erhöht sich auf 25 % bei 373 K (Abbildung 3.2). Der Übergang von flüssigem zu gasförmigem Wasser ist daher gekennzeichnet durch Wassermoleküle, die im Mittel drei abgesättigte Wasserstoffbrücken-Funktionen und eine unabgesättigte besitzen. Diese Werte liefern eine gute Übereinstimmung mit anderen theoretischen und experimentellen Studien zum Anteil der abgesättigten Wasserstoffbrücken-Funktionen im Wasserstoffbrücken-Netzwerk von flüssigem Wasser [184, 193–195].

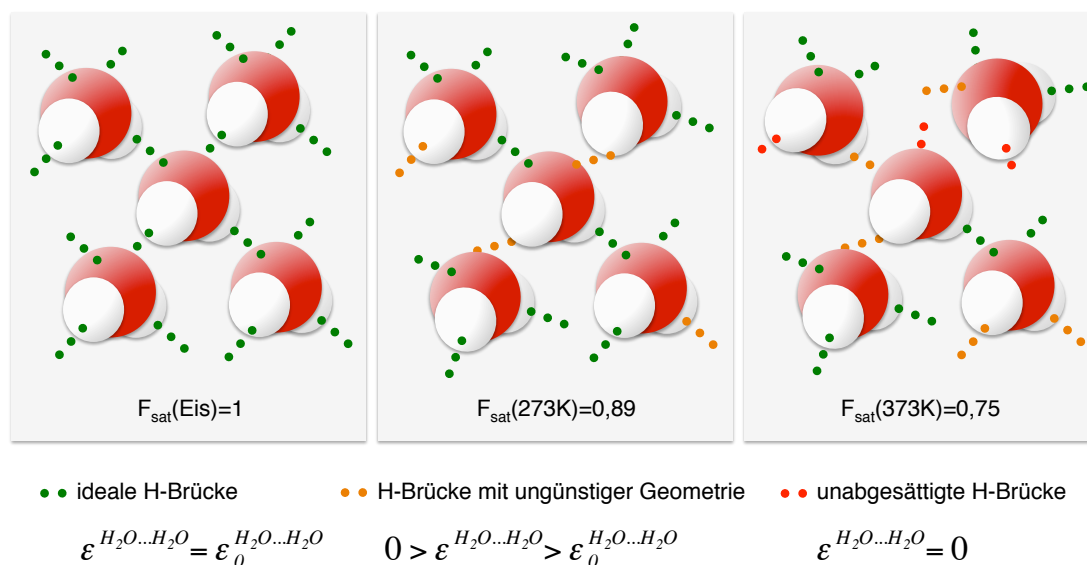


Abbildung 3.2: Schematische Darstellung der Sättigung des Wasserstoffbrücken-Netzwerkes von Wasser bei verschiedenen Temperaturen. Die Sättigung individueller Wasserstoffbrücken variiert zwischen vollständig abgesättigten Wasserstoffbrücken mit idealer Geometrie (grün dargestellt), nicht vollständig abgesättigten Wasserstoffbrücken mit ungünstiger Geometrie (orange dargestellt) und unabgesättigten Wasserstoffbrücken-Funktionen (rot dargestellt). Bild in Anlehnung an [181]

3.1.2 Mittlere Wasserstoffbrücken-Energie im Wassernetzwerk

In Eis bildet ein Wassermolekül vier Wasserstoffbrücken mit den vier benachbarten Wassermolekülen. Jeder Wasserstoffbrücken-Funktion eines Wassermoleküls kann somit ein Viertel der Gesamtenergie ($-54,18 \text{ kJ/mol}$), die in diesen vier Wasserstoffbrücken steckt (Gleichung 3.1), zugeteilt werden. Somit ergibt sich die Energie einer idealen Wasserstoffbrücke zwischen zwei Wassermolekülen aus der Summe der Energie der beiden beteiligten Wasserstoffbrücken-Funktionen: $\epsilon_0^{H_2O\dots H_2O} = -27,1 \text{ kJ/mol}$.

Es ist bekannt, dass das Potential einer Wasserstoffbrücke an bestimmte Optima für die Distanz als auch die Winkel zwischen Wasserstoffbrücken-Donor und -Akzeptor, gebunden ist. Die *volle Wasserstoffbrücken-Energie* ϵ_0 wird nur für eine geometrisch ideale Wasserstoffbrücke erreicht. Schon kleinere Abweichungen der idealen Geometrie führen zu einer niedrigeren Wasserstoffbrücken-Energie $\epsilon < \epsilon_0$. Die temperaturabhängige Sättigung des Wasserstoffbrücken-Netzwerkes von Wasser $F_{sat}(T)$ reflektiert die mittlere Qualität der Wasserstoffbrücken in diesem Netzwerk (Abbildung 3.2). Sie kann daher genutzt werden, um die mittlere Energie der Wasserstoffbrücken zu bestimmen, die je nach Temperatur mehr oder weniger von ihrer idealen Geometrie abweichen:

$$\epsilon^{H_2O\dots H_2O} = F_{sat}(T) \cdot \epsilon_0^{H_2O\dots H_2O} \quad (3.4)$$

Der Energie einer geometrisch idealen Wasserstoffbrücke zwischen zwei Wassermolekülen ($-27,1 \text{ kJ/mol}$) wird damit für die mittlere Wasserstoffbrücke um den Faktor $F_{sat}(T)$ reduziert. Bei einer Temperatur von 273 K ergibt sich somit eine Energie von -24 kJ/mol , bei 298 K eine Energie von -23 kJ/mol und bei 373 K ist die Energie auf -20 kJ/mol gesunken. In der kürzlich von der IUPAC veröffentlichten Definition einer Wasserstoffbrücke, wurde eine Energie von -20 bis -25 kJ/mol vorgeschlagen [45].

3.1.3 Enthalpie und Entropie des Wasserstoffbrücken-Netzwerkes

Die Enthalpie des Wasserstoffbrücken-Netzwerkes $H_{H-bond \text{ network}}$ entspricht der mittleren Wasserstoffbrücken-Energie. Mit steigender Temperatur werden weniger Wasserstoffbrücken gebildet und mehr unabsättigte Wasserstoffbrücken-Funktionen entstehen. Die Enthalpie des Wasserstoffbrücken-Netzwerkes ist somit temperaturabhängig und korreliert mit dem Anteil der abgesättigten Wasserstoffbrücken-Funktionen:

$$H_{H-bond \text{ network}}(T) = F_{sat}(T) \cdot 2\epsilon_0^{H_2O\dots H_2O} \quad (3.5)$$

3. Konzept hinter der HYDE-Funktion

Wird eine Wasserstoffbrücke gebrochen, so ändert sich die Enthalpie des Wasserstoffbrücken-Netzwerkes. Diese Änderung wird berechnet, indem ein beliebiges Wassermolekül des statistischen Ensembles aus dem Zustand $F_{sat}(T)$ in einen Zustand mit drei abgesättigten und einer unabgesättigten Wasserstoffbrücken-Funktion überführt wird:

$$\Delta H_{H-bond\ network} = (0.75 - F_{sat}(T)) \cdot 2\epsilon_0^{H_2O\dots H_2O} \quad (3.6)$$

Im Gegensatz zu nicht wechselwirkenden Teilchen, hängt die Entropie von Netzwerken nicht linear von der Anzahl der Teilchen ab, die das Netzwerk bilden, sondern die Anzahl der Zustände des Netzwerks entspricht der Netzwerk-Entropie, die mit den Fehlstellen des Netzwerks korreliert [196]. Die Fehlstellen des Wasserstoffbrücken-Netzwerkes von Wasser werden durch unvollständig abgesättigte Wasserstoffbrücken-Funktionen verursacht. Das bedeutet auch, dass jede strukturelle Veränderung des Wasserstoffbrücken-Netzwerkes eine Änderung der Enthalpie und Entropie zur Folge hat. Das Wasserstoffbrücken-Netzwerk von flüssigem Wasser ist ein dynamisches System, indem kontinuierlich Wasserstoffbrücken gebrochen und neu gebildet werden. Um dieses System bei einer bestimmten Temperatur T im Gleichgewicht ($\Delta G_{H-bond\ network} = 0$) zu halten, muss die Änderung der Enthalpie $\Delta H_{H-bond\ network}$ beim Brechen einer Wasserstoffbrücke durch eine Änderung der Entropie des Wasserstoffbrücken-Netzwerkes $\Delta S_{H-bond\ network}$ kompensiert werden. Somit kann die Änderung der Entropie für das Brechen einer Wasserstoffbrücke wie folgt berechnet werden:

$$\begin{aligned} T\Delta S_{H-bond\ network} &= \Delta H_{H-bond\ network} \\ &= (0.75 - F_{sat}(T)) \cdot 2\epsilon_0^{H_2O\dots H_2O} \\ &= -(0.25 - F_{unsat}(T)) \cdot 2\epsilon_0^{H_2O\dots H_2O} \\ \Rightarrow \Delta S_{H-bond\ network} &= -\left(\frac{0.25 - F_{unsat}(T)}{T}\right) \cdot 2\epsilon_0^{H_2O\dots H_2O} \end{aligned} \quad (3.7)$$

Die Änderung der Entropie des Wasserstoffbrücken-Netzwerkes $\Delta S_{H-bond\ network}$ bei einer bestimmten Temperatur T ist proportional zum Anteil der unvollständig abgesättigten Wasserstoffbrücken-Funktionen F_{unsat} und somit zum Anteil der Fehlstellen des Wasserstoffbrücken-Netzwerkes von flüssigem Wasser.

3.2 Energie intermolekularer Wechselwirkungen in Wasser

Verschiedene funktionelle Gruppen interagieren unterschiedlich in Wasser, während polare Gruppen sich gut in das Wasserstoffbrücken-Netzwerk von Wasser eingliedern können, führen apolare Gruppen zu einer Störung des Wasserstoffbrücken-Netzwerkes. Basierend auf denen im vorherigen Abschnitt vorgestellten Gleichungen können die energetischen Wechselwirkungsbeiträge von verschiedenen idealisierten Gruppen mit Wasser als auch zwischen diesen Gruppen in Wasser abgeschätzt werden.

3.2.1 Wechselwirkungen von polaren Gruppen mit Wasser

In Experimenten konnte gezeigt werden, dass die tetraedrische Struktur des Wasserstoffbrücken-Netzwerkes von Wasser in Anwesenheit von gelösten Substanzen, wie DMSO oder Aceton, erhalten bleibt und dass die polaren Sauerstoffe dieser Substanzen sich geometrisch gut in das Wasserstoffbrücken-Netzwerk eingliedern können [192, 197]. Daraus kann abgeleitet werden, dass die Energie dieser Wasserstoffbrücke zwischen der polaren Gruppe und einem Wassermolekül vergleichbar mit der Energie einer Wasserstoffbrücke zwischen zwei Wassermolekülen ist, d.h. $\epsilon_0^{polar...H_2O} = \epsilon_0^{H_2O...H_2O} = -27,1 \text{ kJ/mol}$. Die Anwesenheit dieser polaren Gruppen beeinflusst also nicht den Anteil an abgesättigten und unabgesättigten Wasserstoffbrücken-Funktionen F_{sat} und F_{unsat} . Die Entropie des Systems ändert sich nicht ($T\Delta S = 0$). Im Folgenden werden diese Gruppen als *polare Funktionen* bezeichnet.

Durch die Hydratation einer idealisierten polaren Funktion wird eine Wasserstoffbrücke zwischen zwei Wassermolekülen gebrochen. Eine der beiden beteiligten Wasserstoffbrücken-Funktionen kann sich dabei in das Wasserstoffbrücken-Netzwerk von Wasser reintegrieren, während die andere eine neue Wasserstoffbrücke mit der polaren Funktion ausbilden kann. Der Enthalpie-Verlust durch das Aufbrechen der Wasserstoffbrücke zwischen den zwei Wassermolekülen wird durch die Reintegration der einen Wasserstoffbrücken-Funktion ins Wassernetzwerk zur Hälfte kompensiert (Gleichung 3.8). Der umgekehrte Prozess findet sich bei der Dehydratation einer idealisierten polaren Funktion (Gleichung 3.9). Ähnlich wie die Wasserstoffbrücken zwischen Wassermolekülen haben Wasserstoffbrücken mit polaren Funktionen nicht zwangsweise eine ideale Geometrie, daher ist die mittlere Wasserstoffbrücken-Energie $\epsilon^{polar...H_2O}$ um den Sättigungsfaktor F_{sat} reduziert.

Für die freie Hydratations- und Dehydratationsenergie einer idealisierten polaren Funktion ergeben sich damit folgende Gleichungen:

$$\begin{aligned}\Delta G_{hydration}^{polar} &= \Delta H_{hydration}^{polar} - \underbrace{T \Delta S_{hydration}^{polar}}_{=0} \\ &= F_{sat}(T) \cdot \epsilon_0^{polar \dots H_2O} - F_{sat}(T) \cdot \frac{1}{2} \cdot \epsilon_0^{H_2O \dots H_2O} < 0\end{aligned}\quad (3.8)$$

$$\Delta G_{dehydration}^{polar} = -\Delta G_{hydration}^{polar} > 0 \quad (3.9)$$

Die Dehydratationsenergie einer idealisierten polaren Funktion ist immer positiv und somit energetisch ungünstig. Mit steigender Temperatur und/oder mit der Zunahme von unabgesättigten Wasserstoffbrücken-Funktionen wird dieser Beitrag kleiner. Bei 298 K führt die Dehydratation einer polaren Funktion zu einem ungünstigen Beitrag von ca. 10 bis 12 kJ/mol (unter der Annahme: $\epsilon_0^{polar \dots H_2O} = \epsilon_0^{H_2O \dots H_2O}$). Ein ähnlich hoher destabilisierender Beitrag wurde auch für nicht abgesättigte polare Funktionen in Protein-Ligand-Komplexen gemessen [47, 198, 199].

3.2.2 Wechselwirkungen von apolaren Gruppen mit Wasser

Im Gegensatz zu polaren Gruppen findet man bei apolaren Gruppen keine eindeutig bevorzugte Anordnung der Wassermoleküle um die apolare Gruppe [192, 197]. Dies deutet darauf hin, dass die Wechselwirkungen von Wassermolekülen und apolaren Gruppen nur sehr schwach sein können, für eine $CH \cdots O$ -Wasserstoffbrücke wird eine Energie von -1 kJ/mol angenommen [200]. Daher werden diese Gruppen im Folgenden als *apolare Funktionen* bezeichnet und die Wasserstoffbrücken-Funktion des Wassermoleküls, das in Kontakt mit der apolaren Funktion ist, wird als unabgesättigt angesehen. Somit haben Wassermoleküle, die in Kontakt mit einer idealisierten apolaren Funktion sind, im Mittel drei abgesättigte und eine unabgesättigte Wasserstoffbrücken-Funktion ($F_{sat} = 0,75$ und $F_{unsat} = 0,25$). Dies entspricht den Sättigungsfaktoren von Wasser bei einer Temperatur von 373 K. Die Integration von apolaren Funktionen in Wasser simuliert somit eine höhere Temperatur. Die Änderung der Enthalpie und Entropie durch die Einführung einer apolaren Funktion in Wasser wird daher zwischen dem Gleichgewichtszustand $F_{sat}(T)$ und $F_{unsat}(T)$ und dem Zustand, der durch $F_{sat} = 0,75$ und $F_{unsat} = 0,25$ charakterisiert ist, berechnet:

$$\begin{aligned}\Delta H_{hydration}^{apolar} &= (0.75 - F_{sat}(T)) \cdot 2\epsilon_0^{H_2O\dots H_2O} \\ &= -(0.25 - F_{unsat}(T)) \cdot 2\epsilon_0^{H_2O\dots H_2O}\end{aligned}\quad (3.10)$$

$$\Delta S_{hydration}^{apolar} = -\left(\frac{0.25}{373K} - \frac{F_{unsat}(T)}{T}\right) \cdot 2\epsilon_0^{H_2O\dots H_2O}\quad (3.11)$$

Durch die Einführung einer apolaren Funktion in Wasser kann die Änderung der Enthalpie bei einer bestimmten Temperatur nicht mehr durch die Änderung Entropie kompensiert werden. Der Entropie-Term ist um den Faktor $T/373$ kleiner als der Enthalpie-Term, dadurch ergibt sich ein energetisch ungünstiger Beitrag zur freien Energie:

$$\Delta G_{hydration}^{apolar} = -\frac{1}{2} \cdot \epsilon_0^{H_2O\dots H_2O} \cdot \left(1 - \frac{T}{373K}\right) > 0\quad (3.12)$$

$$\Delta G_{dehydration}^{apolar} = -\Delta G_{hydration}^{apolar} < 0\quad (3.13)$$

Die Hydratation einer apolaren Funktion ist bei jeder Temperatur energetisch ungünstig, der Beitrag wird mit steigender Temperatur jedoch schwächer. Im Gegensatz dazu ist die Dehydratation einer apolaren Funktion bei jeder Temperatur energetisch günstig, bei 273 k liegt der Beitrag bei $-3,6 \text{ kJ/mol}$, bei 313 K beläuft sich der Beitrag auf $-2,2 \text{ kJ/mol}$ und hat bei 373 K schließlich keinen Beitrag mehr. Um die apolare Hydratation und damit eine Erhöhung der unangesättigten Wasserstoffbrücken-Funktionen zu vermeiden, kommt es zur Zusammenlagerung von apolaren Funktionen in Wasser, dem hydrophoben Effekt.

3.2.3 Wechselwirkungen zwischen funktionellen Gruppen in Wasser

Mithilfe der hergeleiteten Gleichungen für die Dehydratationsenergie von idealisierten polaren und apolaren Funktionen (Gleichung 3.9 und Gleichung 3.13) kann die Energie von verschiedenen intermolekularen Wechselwirkungen in Wasser abgeschätzt werden. Manche dieser intermolekularen Wechselwirkungen tragen dabei günstig zur freien Bindungsenergie bei, während andere ungünstige Kontakte nur auftreten, wenn die Summe der günstigen Beiträge in einem molekularen Komplex überwiegt.

Kontakt zwischen zwei polaren Gruppen

Der Kontakt zwischen zwei polaren Gruppen kann energetisch günstig als auch ungünstig zur freien Bindungsenergie beitragen. In wässriger Lösung müssen beide polaren Funktionen zuerst dehydratisiert werden, bevor sich eine Wasserstoffbrücke zwischen ihnen ausbilden kann. Der energetische Beitrag vom Kontakt zweier polarer Funktionen zur freien Bindungsenergie setzt sich daher aus der Wasserstoffbrücken- und der Dehydratationsenergie beider Funktionen zusammen:

$$\begin{aligned}\Delta G_{binding}^{polar-polar} &= \epsilon_0^{polar...polar} + 2 \cdot \Delta G_{dehydration}^{polar} \\ &= \epsilon_0^{polar...polar} - 2 \cdot F_{sat}(T) \cdot \epsilon_0^{polar...H_2O} + F_{sat}(T) \cdot \epsilon_0^{H_2O...H_2O}\end{aligned}\quad (3.14)$$

Die mittlere Wasserstoffbrücken-Energie ist im Wassernetzwerk um den Faktor F_{sat} reduziert (Gleichung 3.4), dadurch sind die Dehydratationskosten ebenfalls um diesen Faktor reduziert und es ergibt sich ein schwach stabilisierender Beitrag für eine intermolekulare Wasserstoffbrücke. Bei 298 K beläuft sich der Beitrag auf $-3,9 \text{ kJ/mol}$ für eine geometrisch ideale Wasserstoffbrücke. Dieser Beitrag nimmt bei steigender Temperatur leicht zu (Abbildung 3.3), da die Dehydratationskosten weiter sinken. Sobald die Abweichung von der idealen Geometrie zu groß wird, wirkt sich die intermolekulare Wasserstoffbrücke destabilisierend auf die freie Bindungsenergie aus. Die Wasserstoffbrücken-Energie $\epsilon_0^{polar...polar}$ wird vernachlässigbar klein und die Dehydratationskosten der polaren Funktionen können nicht kompensiert werden. Eine ähnliche destabilisierende Konstellation tritt durch den Kontakt zweier Wasserstoffbrücken-Donoren oder -Akzeptoren auf, wo sich bei 298 K ein destabilisierender Beitrag von bis zu 22 kJ/mol ergibt.

Kontakt zwischen zwei apolaren Gruppen

Der Kontakt zweier apolarer Gruppen in einem molekularen Komplex in wässriger Lösung führt durch die Dehydratation der apolaren Funktionen immer zu einem stabilisierenden Beitrag zur freien Bindungsenergie:

$$\Delta G_{binding}^{apolar-apolar} = 2 \cdot \Delta G_{dehydration}^{apolar} = \epsilon_0^{H_2O...H_2O} \cdot \left(1 - \frac{T}{373K}\right) < 0 \quad (3.15)$$

Bei 298 K ergibt sich für ein apolares Kontaktpaar ein energetisch günstiger Beitrag von $-5,4 \text{ kJ/mol}$. Mit steigender Temperatur nimmt dieser Beitrag jedoch stark ab (Abbildung 3.3).

Kontakt zwischen polarer und apolarer Gruppe

Ein weiterer möglicher Kontakt, der in einem molekularen Komplex zustande kommen kann, ist der zwischen einer polaren und einer apolaren Gruppe. Dieser Kontakt ist in wässriger Lösung immer energetisch ungünstig, da die Dehydratationskosten der polaren Funktion nicht vollständig durch die energetisch günstige Dehydratation der apolaren Gruppe kompensiert werden können:

$$\begin{aligned} \Delta G_{binding}^{polar-apolar} &= \Delta G_{dehydration}^{polar} + \Delta G_{dehydration}^{apolar} \\ &= -F_{sat}(T) \cdot \epsilon_0^{polar...H_2O} + F_{sat}(T) \cdot \frac{1}{2} \cdot \epsilon_0^{H_2O...H_2O} \\ &\quad + \frac{1}{2} \cdot \epsilon_0^{H_2O...H_2O} \cdot \left(1 - \frac{T}{373K}\right) > 0 \end{aligned} \quad (3.16)$$

Der energetisch günstige Beitrag der apolaren Dehydratation nimmt mit steigender Temperatur schneller ab als der ungünstige Beitrag der polaren Dehydratation, daher wirkt sich dieses Kontaktpaar bei höheren Temperaturen noch destabilisierender auf die freie Bindungsenergie aus (Abbildung 3.3). Bei 298 K liegt der Beitrag bei $8,8 \text{ kJ/mol}$, während er bei 313 K auf $9,1 \text{ kJ/mol}$ ansteigt.

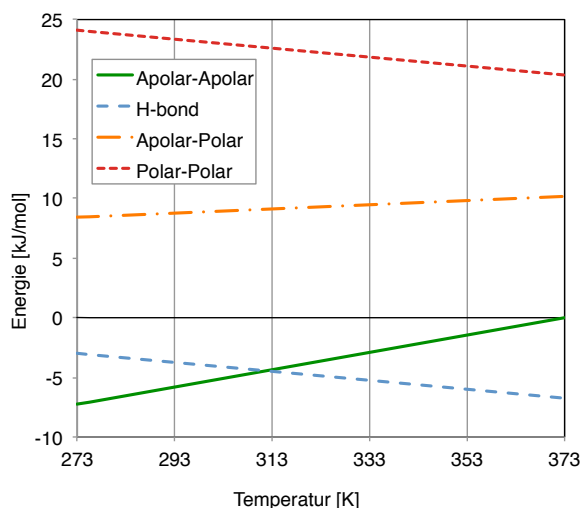


Abbildung 3.3: Temperaturabhängige Wechselwirkungsenergien verschiedener funktioneller Gruppen in Wasser. Dargestellt sind: Wasserstoffbrücken (blaue gestrichelte Linie), der hydrophobe Effekt (grüne durchgezogene Linie), der ungünstige Kontakt zwischen einer polaren und einer apolaren Gruppe (orange durchbrochene Linie) und der ungünstige Kontakt zwischen zwei Wasserstoffbrücken-Donoren oder -Akzeptoren (rote gestrichelte Linie). Bild in Anlehnung an [181]

3.3 Wechselwirkungen in Wasser als Basis von HYDE

Das im vorherigen Abschnitt vorgestellte Konzept zur Beschreibung der Energie von intermolekularen Wechselwirkung in wässriger Lösung dient als Grundlage der HYDE

3. Konzept hinter der HYDE-Funktion

Bewertungsfunktion. Dabei wird davon ausgegangen, dass der Hauptanteil der freien Bindungsenergie in molekularen Komplexen über grundlegende Wechselwirkungen, d.h. Wasserstoffbrücken, den hydrophoben Effekt und die polare Desolvatisierung abgeschätzt werden kann. Damit lässt sich die freie Bindungsenergie zweier Moleküle folgendermaßen berechnen [201]:

$$\Delta G_{binding} = \Delta G_{dehydration} + \sum_{i=1}^n \sum_{j=1}^m e^{polar1 \dots polar2} \cdot \delta_{H-bond}(i, j) \quad (3.17)$$

$$\delta_{H-bond}(i, j) = \begin{cases} 1 & \text{Wasserstoffbrücke zwischen Atom } i \text{ und Atom } j \\ 0 & \text{sonst} \end{cases}$$

Die Dehydratationsenergie $\Delta G_{dehydration}$ beinhaltet sowohl die energetisch günstigen Beiträge durch den hydrophoben Effekt als auch die destabilisierenden Beiträge, die bei der polaren Dehydratation entstehen. Die Wasserstoffbrücken-Energie zwischen Atom i und Atom j liefert immer einen stabilisierenden Beitrag zur freien Bindungsenergie, sofern die Abweichung von der idealen Wasserstoffbrücken-Geometrie nicht zu hoch ist. Um eine atom-basierte Bewertung in HYDE zu ermöglichen, wird die Wasserstoffbrücken-Energie auf die einzelnen Wasserstoffbrücken-Partner aufgeteilt. Die freie Bindungsenergie kann damit als Summe über die Beiträge aller Atome, die an der Komplex-Bindung beteiligt sind, abgeschätzt werden [201]:

$$\Delta G_{binding} = \sum_{i=1}^{n+m} \Delta G_{dehydration}^i + \Delta G_{H-bond}^i \quad (3.18)$$

In HYDE soll zwischen verschiedenen Atomtypen unterschieden werden, z. B. zwischen schwächeren und stärkeren Wasserstoffbrücken-Donoren oder -Akzeptoren. Dies ist mit dem in diesem Kapitel vorgestellten Konzept nicht direkt möglich, da hier nur idealisierte Wasserstoffbrücken-Funktion beschrieben wurden. Daher wurde im Kontext von HYDE der Oktanol-Wasser-Verteilungskoeffizient $\log P$ zur Parametrisierung der verschiedenen Atomtypen genutzt, der implizit auch die Dehydratationsenergie beschreibt. Hierauf wird in den folgenden zwei Kapiteln näher eingegangen.

4

Bestehende Methoden und Module

In diesem Kapitel werden bereits bestehende Module der Flex*-Bibliothek [80] vorgestellt, in die die HYDE-Bewertungsfunktion integriert wurde, sowie der erste HYDE-Prototyp [201, 202], der die Ausgangsposition der aktuellen HYDE-Version bildete. Zuerst erfolgt die genaue Beschreibung des HYDE-Prototyps, bezüglich des verwendeten Modells und der Implementierung, zudem werden Erweiterungsmöglichkeiten aufgezeigt. Im zweiten Teil des Kapitels werden verschiedene Optimierungsmethoden eingeführt, die als bereits bestehende Module der Flex*-Bibliothek zur Verfügung standen.

4.1 Der HYDE-Prototyp

Die erste Version von HYDE wurde 2008 von Reulecke *et al.* [201, 202] veröffentlicht. Die Grundlage der HYDE-Bewertungsfunktion bildet das theoretische Konzept zur Beschreibung von intermolekularen Wechselwirkungen in Wasser, das im vorhergehenden Kapitel 3 beschrieben wurde. Im Folgenden werden die einzelnen Terme des HYDE Prototyps, die Parametrisierung, die Implementierung und erste Resultate zusammengefasst [201, 202]. Anschließend werden Erweiterungsmöglichkeiten diskutiert.

4.1.1 Terme der HYDE-Bewertungsfunktion

Die HYDE-Bewertungsfunktion basiert auf zwei Termen zur Abschätzung der freien Bindungsenthalpie von Protein-Ligand-Komplexen: der Wasserstoffbrücken- und der Dehydratationsenergie. Für alle Atome i des Liganden und der Bindetasche des Proteins, wird die Änderung der Dehydratation und der Wasserstoffbrücken-Bindungen zwischen freiem und komplexiertem Zustand bestimmt:

$$\Delta G_{\text{HYDE}} = \sum_{\text{atoms } i} \Delta G_{\text{dehydration}}^i + \Delta G_{\text{H-bond}}^i \quad (4.1)$$

Während durch den Wasserstoffbrücken-Term nur energetisch günstige Beträge berücksichtigt werden, erfasst der Dehydratationsterm sowohl den energetisch günstigen Betrag durch den hydrophoben Effekt als auch die energetisch ungünstige polare Desolvatisierung.

Beide Terme der HYDE-Bewertungsfunktion sind von folgender thermodynamischer Gleichung abgeleitet, wobei man von einem chemischen Gleichgewicht ($\Delta G = 0$) und konstanter Temperatur ausgeht:

$$\Delta G = -RT \cdot \ln(K) \quad (4.2)$$

Dabei ist R die allgemeine Gaskonstante, T die Temperatur und K die Gleichgewichtskonstante einer Reaktion.

Berechnung der Dehydratationsenergie

Zur Quantifizierung der Wasserstoffbrücken- als auch Dehydratationsenergie werden in HYDE partielle logP-Atom-Inkrementen (plogP) verwendet. Der $\log P_{O/W}$ -Wert eines Moleküls ist der negative dekadische Logarithmus seines Oktanol-Wasser-Verteilungskoeffizienten. Ein positiver logP-Wert bedeutet, dass sich anteilig mehr Moleküle in der Oktanol-Phase befinden und das Molekül somit als *hydrophob* klassifiziert werden kann. Die freie Dehydratationsenergie kann daher mithilfe des logP-Wertes und Gleichung 4.2 berechnet werden:

$$\Delta G_{\text{dehydration}} = -RT \cdot \ln K_{O/W} = -2.3RT \cdot \log P_{O/W} \quad (4.3)$$

Zur Berechnung der Dehydratationsenergie eines einzelnen Atoms i werden in HYDE atombasierte logP-Inkrementen ($plogP^i$) verwendet:

$$\Delta G_{dehydration}^i = -2.3RT \cdot \frac{acc_{free}^i - acc_{complex}^i}{acc_{mean}^i} \cdot p \log P_i \quad (4.4)$$

Neben dem $plogP^i$ -Wert eines Atoms wird in HYDE zur Berechnung der Dehydratationsenergie auch die Änderung seiner Lösungsmittel-Zugänglichkeit acc^i durch die Komplexbildung berücksichtigt. Hierzu wird die Zugänglichkeit des Atoms im ungebundenen acc_{free}^i und im gebundenen Zustand $acc_{complex}^i$ bestimmt und es erfolgt eine Normalisierung über die mittlere statistische Zugänglichkeit acc_{mean}^i des entsprechenden Atomtyps. Für apolare Atome ($plogP$ -Wert > 0), deren Lösungsmittel-Zugänglichkeit durch die Komplexbildung reduziert wird, ergibt sich demnach und entsprechend des hydrophoben Effektes ein stabilisierender Beitrag ($\Delta G_{dehydration}^i < 0$) zur Gesamtenergie. Im Gegensatz dazu führt die Dehydratation eines polaren Atoms ($plogP$ -Wert < 0) zu einer ungünstigen Dehydratationsenergie ($\Delta G_{dehydration}^i > 0$) und trägt damit destabilisierend zur Gesamtenergie bei.

Berechnung der Wasserstoffbrücken-Energie

Der Energiegewinn einer Wasserstoffbrücke zwischen zwei polaren Wasserstoffbrücken-Funktionen wurde bereits in Kapitel 3 hergeleitet. Dieser ergibt sich aus der Tatsache, dass in Wasser im Mittel nicht alle Wasserstoffbrücken mit idealer Geometrie ausgebildet werden und so die mittlere Wasserstoffbrücken-Energie um den Sättigungsfaktor F_{sat} reduziert ist. Hieraus folgt, dass die Kosten der polaren Dehydratation ebenfalls um den Sättigungsfaktor F_{sat} geringer sind als die Energie einer geometrisch ideal ausgebildeten Wasserstoffbrücke zwischen zwei polaren Atomen im Protein-Ligand-Komplex. Für die Abschätzung der Wasserstoffbrücken-Energie in HYDE ergibt sich damit folgende Gleichung:

$$\Delta G_{H-bond}^i = -\frac{1}{F_{sat}} \cdot \Delta G_{dehydration}^i = \frac{2.3RT}{F_{sat}} \cdot p \log P_i \quad (4.5)$$

In HYDE wird auch die Wasserstoffbrücken-Energie atombasiert berechnet, so dass beiden Atomen, die an einer Wasserstoffbrücke beteiligt sind, ihr jeweiliger Betrag entsprechend ihres $plogP$ -Wertes zugeteilt wird. Als weiterer Faktor wurde die Änderung der Sättigung Δsat^i eines Atoms i zwischen gebundenem und ungebundenem Zustand in Gleichung 4.5 integriert:

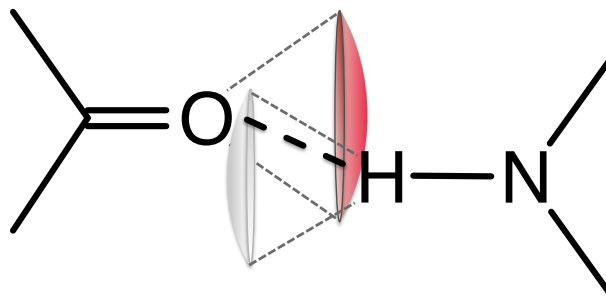
$$\Delta G_{H-bond}^i = \frac{2.3RT}{F_{sat}} \cdot (sat_{complex}^i - sat_{free}^i) \cdot p \log P_i \quad (4.6)$$

Die Sättigung sat^i eines Atoms i wird aus der Anzahl und der Güte der von ihm ausgebildeten Wasserstoffbrücken mit umgebenden Atomen j berechnet:

$$sat^i = \sqrt{\sum_j f_{dev}(i, j)} \quad (4.7)$$

Um einer Überbewertung aufgrund der Ausbildung von multiplen Wasserstoffbrücken durch ein polares Atom entgegenzuwirken, wurde eine Skalierung durch den Wurzelterm eingeführt. Die geometrische Güte der Wasserstoffbrücken wird über den Bestrafungsfaktor $f_{dev}(i, j)$, entsprechend des FlexX Wechselwirkungsmodells [80], berücksichtigt. Für eine geometrische ideale Wasserstoffbrücke ist $f_{dev}(i, j) = 1$ und nimmt linear gegen 0 ab, bis die Abweichung von der idealen Wechselwirkungs-Geometrie bezüglich Distanz und/oder Winkel einen Toleranzwert überschreitet. In Abbildung 4.1 ist beispielhaft die Wasserstoffbrücken-Wechselwirkung zweier polarer Atome dargestellt, ideale Distanz und Winkel werden jeweils durch die Wechselwirkungsflächen beschrieben. Eine Wechselwirkung wird ausgebildet, wenn die beteiligten Atome jeweils auf der Wechselwirkungsfläche des anderen Partners liegen. Diese Wechselwirkungsflächen wurden für HYDE, auf Basis der Kristallstrukturen kleiner Moleküle teilweise überarbeitet und angepasst [202].

Abbildung 4.1: Schematische Darstellung der Wechselwirkungsflächen und der Wasserstoffbrücke von zwei polaren Atomen. Das Wechselwirkungs-Modell wurde von FlexX abgeleitet [80].



4.1.2 Parametrisierung und Implementierung

plogP-Parametrisierung

Im Gegensatz zu den vielen empirischen Bewertungsfunktionen, die anhand von experimentellen Bindungsaffinitäten und Kristallstrukturen von Protein-Ligand-Komplexen kalibriert wurden (s. Kapitel 2.5), sind in HYDE lediglich die partiellen logP-Inkrementen (plogP) durch eine Regressions-Analyse bestimmt worden. Dadurch ist HYDE genereller

einsetzbar als andere empirische Bewertungsfunktionen und liefert nicht wie diese meist bessere Resultate für Komplexe, die im Parametrisierungsdatensatz enthalten waren. In HYDE werden die plogP-Werte genutzt, um die Beiträge einzelner Atome zur Wasserstoffbrücken- und Dehydratationsenergie abzuschätzen. Zur Kalibrierung der plogP-Werte, wurden 68 unterschiedliche plogP-Atomtypen definiert. Diese beziehen sich auf das Element, die Hybridisierung, die Anzahl und Art der ausgehenden Bindungen und die Anzahl gebundener Wasserstoffe des jeweiligen Atoms. Weiterhin wurde die Lösungsmittel-Zugänglichkeit acc^i eines Atoms i bei der Regressions-Analyse berücksichtigt. In der Parametrisierung und somit zur Bestimmung des logP-Wertes eines Moleküls mit n Atomen i wurde folgende Inkrementformel verwendet:

$$\log P = \sum_{i=1}^n \frac{acc^i}{acc_{mean}^k} \cdot p \log P_k(i) \quad (4.8)$$

acc_{mean}^k ist die mittlere Zugänglichkeit des Atomtyps k im Parametrisierungsdatensatz, $p \log P_k$ ist das entsprechende logP-Inkrement.

Die Parametrisierung erfolgte mit experimentell gemessenen logP-Werten von 696 kleinen Molekülen, die aus der Zusammenstellung von Hansch und Leo [203] stammen und der PHYSPROP Datenbank [204] entnommen wurden. Die Moleküle wurden so ausgewählt, dass sie mindestens zwei verschiedene funktionelle Gruppen enthalten. Mittels multipler linearer Regression wurden die plogP-Parameter der verschiedenen Atomtypen bestimmt. Eine genaue Auflistung der Atomtypen mit entsprechenden plogP-Parametern findet sich in [201].

Berechnung der Lösungsmittel-Zugänglichkeit

Die Abschätzung der lösungsmittelzugänglichen Oberfläche von Atomen wird in HYDE zur Berechnung der Dehydratationsenergie als auch in der logP-Parametrisierung benötigt. Die lösungsmittelzugängliche Oberfläche (SAS) ist definiert als Erweiterung der Van-der-Waals-Oberfläche eines Moleküls, um den Radius eines idealisierten Lösungsmittel-Moleküls [205–207]. Meist wird ein Radius von $r_{SAS} = 1,4 \text{ \AA}$ für das Lösungsmittel-Molekül gewählt, im HYDE-Prototyp wurde der Radius auf $r_{SAS} = 1,2 \text{ \AA}$ reduziert, um Ungenauigkeiten in den zugrunde liegenden Strukturen gerecht zu werden.

In HYDE wurde eine Abwandlung der SAS implementiert, die gewichtete SAS (WSAS). Hierbei werden für polare Atome nur die Bereiche der SAS berücksichtigt, die in der Vorzugsrichtung einer Wasserstoffbrücke liegen. Die Berechnung der WSAS beruht auf der FlexX-Implementierung der SAS. Diese wird über eine rekursive Ikosaeder-Tesselierung

abgeschätzt. Dabei wird ein Ikosaeder für jedes Schweratom des Moleküls erzeugt. Der Radius des Ikosaeders ergibt sich aus dem Van-der-Waals-Radius r_{VDW} des jeweiligen Atoms und der Erweiterung um den Radius des Lösungsmittel-Moleküls r_{SAS} . In dem rekursiven Algorithmus wird für die drei Eckpunkte eines Ikosaeder-Dreiecks jeweils die Zugänglichkeit unter Berücksichtigung der umliegenden Atome bestimmt. Sind nicht alle Eckpunkte zugänglich oder unzugänglich, so wird dieses Dreieck durch Halbierung der Dreiecksseiten in vier neue Dreiecke unterteilt und jeweils wieder auf Zugänglichkeit getestet. Abbildung 4.2 stellt schematisch die Ikosaeder-Tesselierung zur Abschätzung der SAS eines Atoms dar. Aus allen zugänglichen Eckpunkten eines Ikosaeders wird schließlich die SAS eines Atoms berechnet. Analog dazu erfolgt die Berechnung der WSAS in HYDE, mit dem Unterschied, dass bei polaren Atomen nur Eckpunkte des Ikosaeders als zugänglich markiert werden, die sich in Richtung der jeweiligen Wechselwirkungsfläche befinden. Eckpunkte mit einer geringen Abweichungen zur Wechselwirkungsfläche erhalten eine niedrigere Gewichtung bis zu einem Gewicht von 0, wenn die Winkelabweichung zu groß wird. Abbildung 4.2 zeigt schematisch die Einführung der Gewichtung der SAS. Zwei weitere Methoden zur besseren Beschreibung von schwer zugänglichen Stellen, die besonders auf Proteinseite zum Tragen kommen, wurden ebenfalls implementiert. Eine genaue Beschreibung findet sich in [202]. Die Komplexität der Algorithmen liegt bei einer Größenordnung von $O(n^2)$ bezüglich der Anzahl der Atome [202].

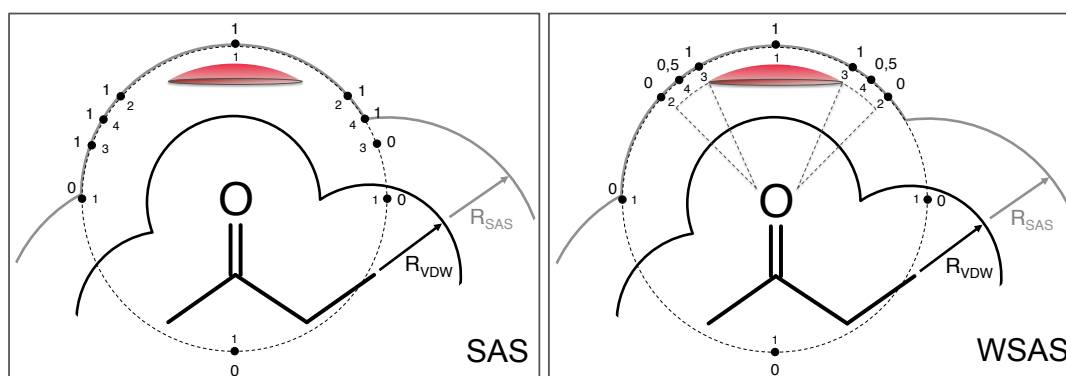


Abbildung 4.2: Vergleich SAS und gewichtete SAS (WSAS). Dargestellt ist die Oberflächendiskretisierung mittels Ikosaeder-Tesselierung (Punkte auf der SAS) mit Verfeinerungsgrad (Zahlen innerhalb der SAS), und die Zugänglichkeit/Gewichtung der Ikosaeder-Eckpunkte (Zahlen außerhalb der SAS). Bei der WSAS wird für polare Atome nur die Oberfläche in Richtung der Wechselwirkungsflächen berücksichtigt.

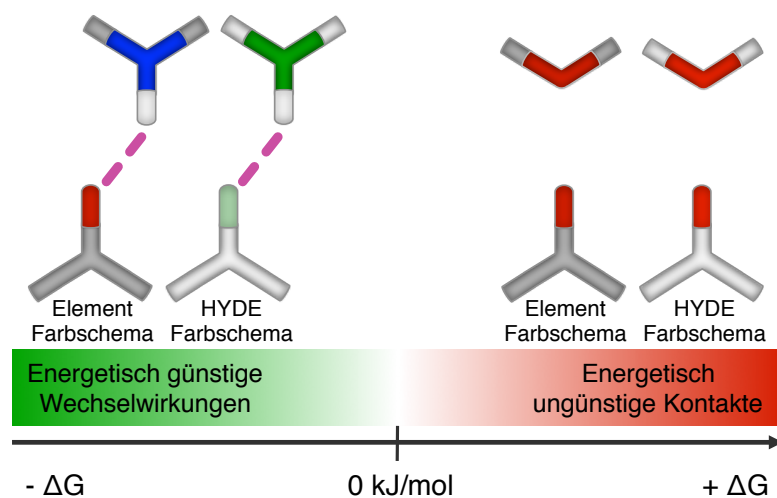


Abbildung 4.3: HYDE-Farbschema. Grün eingefärbte Atome tragen stabilisierend zur Gesamtenergie ΔG_{HYDE} bei; rot eingefärbte Atome wirken destabilisierend; weiße Atome haben keinen Beitrag. Links: Geometrisch gute Wasserstoffbrücke in Element- und HYDE-Färbung, Wasserstoffbrücken-Donor und Akzeptor, beide in grün. Rechts: Kontakt von zwei Wasserstoffbrücken-Akzeptoren, beide werden rot eingefärbt. Bild in Anlehnung an [208]

Atombasierte HYDE-Färbung

Zur Verbesserung der Analyse von Protein-Ligand-Komplexen wurde eine intuitive, atombasierte Einfärbung entwickelt [201]. Diese ermöglicht das direkte Erkennen von energetisch günstigen und ungünstigen Konstellationen im Protein-Ligand-Komplex. Hierzu wurde eine Farbskala von dunkel grün, für energetisch günstige Atombeiträge, über weiß, für Atome die keinen Beitrag zur Gesamtenergie leisten, bis hin zu rot, für energetisch ungünstige Kontakte, eingeführt. Dadurch erhalten Atome, die z. B. an einer geometrisch guten Wasserstoffbrücke, an einer Metallkoordinierung oder am hydrophoben Effekt beteiligt sind, eine grüne Einfärbung. Im Gegensatz dazu werden Atome in ungünstigen Konstellationen, wie z. B. der Kontakt von zwei Wasserstoffbrücken-Donoren oder -Akzeptoren, aber auch Atome, die an einer geometrisch schlechten Wasserstoffbrücke beteiligt sind, rot eingefärbt. Abbildung 4.3 zeigt beispielhaft die HYDE-Färbung einer energetisch günstigen Wechselwirkung als auch die eines energetisch ungünstigen Kontaktes.

Diese Einfärbung der Atome ermöglicht eine direkte Visualisierung des Einflusses individueller Atome auf die Gesamtenergie. Weiterhin wurde eine Zuordnung des Ener-

giebeitrags der Proteinatome zu den jeweils räumlich nächstgelegenen Ligandatomen eingeführt, so dass die einzelnen Ligandatome dann auf Basis dieses akkumulierten Energiebeitrags eingefärbt werden können. In der Leitstruktur-Optimierung können so potenzielle Optimierungsmöglichkeiten an der Leitstruktur leicht sichtbar gemacht werden.

4.1.3 Erste Resultate des HYDE-Prototyps

Der HYDE-Prototyp wurde auf gängigen Benchmark-Datensätzen, wie dem Directory of Useful Decoys (DUD) [209] oder dem PDBbind Datensatz [210, 211] als auch in einer prospektiven Studie experimentell validiert.

Der DUD-Datensatz wurde zur Simulation eines virtuellen Screenings entwickelt und dient damit zur Validierung der Fähigkeit einer Bewertungsfunktion zur Unterscheidung von Bindern und Nicht-Bindern. Dieser Datensatz besteht aus 40 für den Wirkstoffentwurf relevanten Zielproteinen und jeweils experimentell validierten Bindern. Anhand der Binder wurden aus der ZINC-Datenbank [171] topologisch unähnliche Moleküle, jedoch mit ähnlichen physikochemischen Eigenschaften, entsprechend für jedes Zielprotein zusammengestellt. Von diesen Molekülen wird angenommen, dass sie nicht an das Zielprotein binden, sie wurden dennoch nicht als inaktiv experimentell bestätigt. HYDE wurde auf allen Zielproteinen getestet und zeigte im Vergleich zu FlexX teilweise eine bessere Vorhersage der Binder [202]. Zudem wurden die Ergebnisse für drei der Zielproteine genauer untersucht. Dabei zeigte sich, dass schlechte Resultate bei Thrombin und der p38 MAP Kinase auf Fehler im Datensatz zurückzuführen waren und gut mit HYDE identifiziert werden konnten [201, 202].

Die Qualität von HYDE zur Vorhersage von Bindungs-Affinitäten wurde anhand des PDBbind 2004 Datensatzes (refined set) getestet. Dieser besteht aus 800 Protein-Ligand-Komplexen mit experimentell gemessenen Affinitäten. Hierbei erreichte HYDE eine durchschnittliche Vorhersagequalität im Vergleich zu anderen Bewertungsfunktionen die ebenfalls auf diesem Datensatz getestet wurden [150, 202]. Die Analyse des Resultats ergab, dass sich HYDE auch in diesem Fall sehr sensitiv gegenüber Ungenauigkeiten in den zugrunde liegenden Kristallstrukturen der Komplexe zeigte.

Der HYDE-Prototyp wurde 2007 von Bayer CropScience für ein virtuelles Screening eingesetzt. Dabei wurden 4,5 Millionen Substanzen aus einer virtuellen Bibliothek mithilfe von FlexX gedockt. Die erzeugten Platzierungen dieser Substanzen wurden anschließend mit HYDE bewertet. Dabei erhielten 2200 der Substanzen eine Bewertung, die auf eine aktive Substanz hindeutet. Aus diesen wurden dann nach verschiedenen Kriterien, wie Verfügbarkeit oder ADME-Parametern, 172 Substanzen ausgewählt und experimentell

in Assays auf Aktivität getestet. Für zwei Drittel der Verbindungen wurde ein IC_{50} -Wert im mikromolaren Bereich festgestellt, 20 % der Verbindungen erwiesen sich sogar als nanomolare Binder. Diese überaus hohe Trefferquote wird nur selten in einem virtuellen Screening erreicht [212] und zeigt damit die Fähigkeit von HYDE, Binder von Nicht-Bindern zu unterscheiden.

4.1.4 Erweiterungsmöglichkeiten

Die Verwendbarkeit von HYDE wurde eindrucksvoll in einem virtuellen Screening bei Bayer CropScience gezeigt, während die Resultate für Benchmark-Daten eher durchschnittlich waren. Eine Erklärung dafür ist die Sensitivität von HYDE auf Ungenauigkeiten in den zu bewertenden Kristallstrukturen. Diese öffentlich zugänglichen Strukturen weisen häufig weniger ideale Wasserstoffbrücken-Geometrien auf und werden somit in HYDE bestraft, dessen Wasserstoffbrücken-Modell von Kristallstrukturen kleiner Moleküle abgeleitet wurde [202] und damit eine höhere Genauigkeit aufweist. Weiterhin ist die korrekte Protonierung der potentiellen Wasserstoffbrücken-Partner von großer Bedeutung, sowie die richtige Ausrichtung frei drehbarer Gruppen, wie z. B. von Hydroxyl-Gruppen. Eine weitere Ursache ist die unzureichende Beschreibung der Wechselwirkungen von Metallionen im HYDE-Prototyp.

Zur adäquaten Bewertung von Protein-Ligand-Komplexen mit weniger gut aufgelösten Kristallstrukturen durch HYDE ist eine Vorooptimierung der Protein-Ligand-Komplex-Geometrie in Richtung des HYDE-Wasserstoffbrücken-Modells unabdingbar. In diesem Zusammenhang ist natürlich auch eine sinnvolle Protonierung des Liganden und des Proteins sowie die richtige Ausrichtung frei drehbarer Wechselwirkungspartner von großer Wichtigkeit. Die Vorooptimierung der Protein-Ligand-Komplex-Geometrie ist insofern auch notwendig, da in der HYDE-Bewertungsfunktion keine Terme zur Detektion einer Überlappung von Atomen vorgesehen sind. Diese finden sich zwar selten in Kristallstrukturen, treten jedoch häufiger in den, durch molekulares Docking vorhergesagten, Protein-Ligand-Komplexen auf. Ein weiterer Aspekt zur Verbesserung des HYDE-Prototyps stellt der Geschwindigkeitsaspekt dar. Durch die aufwendige Berechnung der molekularen Oberfläche in jeder Komplex-Bewertung ist diese der geschwindigkeitsbegrenzende Faktor. Um eine Vorooptimierung des Protein-Ligand-Komplexes in Richtung des HYDE-Modells durchführen zu können, sollte daher eine alternative Berechnungsmethode der Oberfläche implementiert werden.

4.2 Optimierungsmethoden der Flex*-Bibliothek

Die Analyse der Resultate des HYDE-Prototyps hat gezeigt, dass eine geometrische Voroptimierung des Protein-Ligand-Komplexes als auch eine Optimierung des Wasserstoffbrücken-Netzwerkes, grundlegend sind, um eine zuverlässige Bewertung des Protein-Ligand-Komplexes durch HYDE zu ermöglichen. Im Folgenden werden Optimierungsmethoden und -module vorgestellt, die bereits in der Flex*-Bibliothek zur Verfügung standen und in diesem Projekt mit HYDE kombiniert wurden.

4.2.1 ProToss: Optimierung des Wasserstoffbrücken-Netzwerkes

Die Auflösung von Kristallstrukturen von Protein-Ligand-Komplexen ist in der Regel zu niedrig, um die eindeutige Lokalisierung von Wasserstoffatomen festzustellen. Im Struktur-Verfeinerungsprozess kann die Position von Wasserstoffatomen durch entsprechende Programme berechnet werden, jedoch sind die meisten Kristallstrukturen von Protein-Ligand-Komplexen in der PDB [12] ohne Koordinaten von Wasserstoffatomen hinterlegt. Dadurch ist die Protonierung von polaren Atomen (z. B. bei Aminen, Carbonsäuren oder Phenolen) als auch ein bestimmter tautomerer Zustand einer funktionellen Gruppe (z. B. der Imidazolring in Histidin) nicht eindeutig gegeben. Die genaue Ausrichtung der Wasserstoffatome bei frei drehbaren Gruppen, wie Aminen oder Hydroxyl-Gruppen, ist daher auch nicht eindeutig bestimmt. Eine weitere Mehrdeutigkeit, die in Protein-Kristallstrukturen auftritt, ist eine mögliche 180° Drehung von, in der Elektronendichte symmetrischen, Aminosäure-Seitenketten. Dies betrifft den Imidazolring von Histidin sowie die Amid-Seitenketten von Asparagin und Glutamin (vgl. Kapitel 2.2). Alle diese Faktoren beeinflussen letztendlich die Ausbildung eines idealen Wasserstoffbrücken-Netzwerkes innerhalb des Proteins als auch zwischen Ligand und Protein. Die Ausbildung dieses idealen Wasserstoffbrücken-Netzwerkes ist wichtig für die Proteinstabilität [213], aber auch um die Bindungsaffinität eines Liganden zum Protein zuverlässig abschätzen zu können.

ProToss [37] ist ein Modul der Flex*-Bibliothek, das anhand einer Protein-Kristallstruktur das optimale Wasserstoffbrücken-Netzwerk innerhalb des Proteins sowie zwischen Protein und Ligand generiert. Dazu werden die verschiedenen, oben erwähnten Zustände der betreffenden Gruppen erzeugt. Mittels einer empirischen Bewertungsfunktion zur Bestimmung der Güte einer potentiellen Wasserstoffbrücke und dynamischer Programmierung wird das optimale Wasserstoffbrücken-Netzwerk berechnet. Die Bewertungsfunktion für die Güte einer Wasserstoffbrücke basiert auf dem FlexX-Wechselwirkungsmodell [80]. Im Kontext von HYDE werden in ProToss die für HYDE optimierten Wechselwirkungsflächen (vgl. Abschnitt 4.1.1) genutzt.

4.2.2 Optimierung der Protein-Ligand-Komplex-Geometrie

Die durch molekulares Docking erzeugten Protein-Ligand-Komplexe enthalten zum Teil Überlappungen von Atomen. Das passiert vor allem, wenn im Docking-Prozess die Proteinstruktur als starr behandelt wird. Der gedockte Liganden kann ebenfalls in einer ungünstigen Konformation vorliegen, da die im Docking eingesetzten Bewertungsfunktionen vorwiegend die Anzahl der Wechselwirkungen optimieren. Eine Optimierung der Protein-Ligand-Komplex-Geometrie ist auch nötig, um zwischen den verschiedenen Optima verschiedener Funktionen oder Programme zu wechseln. Das bedeutet, z. B., dass die Protein-Ligand-Kristallstruktur, die von einem Kristallographen mit einem Struktur-Verfeinerungsprogramm modelliert wurde, sich in dem Optimum des Kraftfeldes befindet, das diesem Programm zugrunde liegt. Die Platzierungen, die mit verschiedenen Docking-Programmen erzeugt werden, befinden sich auch jeweils im Optimum der dort eingesetzten Bewertungsfunktion. Um jetzt diese Protein-Ligand-Komplexe adäquat mit HYDE bewerten zu können, muss eine Optimierung entwickelt werden, die den Protein-Ligand-Komplex in Richtung des Optimums der HYDE-Funktion bewegt. Dabei sollten neben den Termen der HYDE-Funktion auch Terme zur Bewertung der Ligandkonformation und zur Ermittlung von Atom-Überlappungen integriert werden. Im Folgenden werden zwei Optimierungsmethoden, die in der Flex*-Bibliothek zur Verfügung stehen kurz beschrieben. Beide benötigen eine Zielfunktion zur Optimierung. Diese wurde im Hinblick auf HYDE neu entwickelt und wird in Kapitel 6.4.1 beschrieben.

Stochastische Optimierung

Als stochastisches Optimierungsverfahren wurde in der Flex*-Bibliothek eine Monte-Carlo-Optimierung mit einem simulierten Abkühlungsprozess implementiert. Dabei wird in jedem Schritt eine zufällige Veränderung, entweder eines Torsionswinkels des Liganden, der Translation oder der Rotation des Liganden, eingeführt. Anschließend wird bewertet, ob sich diese Veränderungen bezüglich der Zielfunktion positiv oder negativ auswirkt. Bei einer Verbesserung der Energie des Systems wird diese Veränderungen als Ausgangsposition für den nächsten Schritt gewählt. Im Falle einer Verschlechterung wird abhängig von der Temperatur entschieden, ob der Zustand trotzdem angenommen wird. Je höher die Temperatur ist, desto höher ist die Wahrscheinlichkeit, dass dieser Zustand angenommen wird. In jedem Optimierungsschritt wird iterativ die Temperatur abgesenkt, wodurch die Wahrscheinlichkeit sinkt, ungünstige Zustände anzunehmen, so dass das Verfahren in einem Optimum konvergiert. Dieses Verfahren ermöglicht das

Auffinden verschiedener Optima und dient der Auffindung des globalen Optimums hinsichtlich der gegebenen Zielfunktion. Mit diesem Optimierungsverfahren sind auch größere Bewegungen des Liganden in der Bindetasche möglich.

Numerische Optimierung

Die numerische Optimierung, die in der Flex*-Bibliothek implementiert wurde, verwendet ein Quasi-Newton-Verfahren, dem der BFGS-Algorithmus [214] zur Berechnung der Hesse-Matrix zugrunde liegt. Mit diesem Verfahren können nichtlineare Minimierungsprobleme numerisch gelöst werden. Die Hesse-Matrix wird dabei aus Effizienzgründen nur näherungsweise berechnet, um das Minimum der Zielfunktion zu ermitteln. In jedem Optimierungsschritt werden minimale Bewegungen am Liganden (Translation, Rotation und Torsionswinkel) durchgeführt. Der Unterschied der Hesse-Matrix zum vorherigen Zustand wird bestimmt. Wird ein niedrigeres Minimum erreicht, wird dieser Zustand die Startposition des nächsten Optimierungsschritts. Das Verfahren konvergiert, wenn kein niedrigeres Minimum gefunden wird oder der Unterschied der Minima unterhalb eines Schwellwertes liegt. Diese Optimierungsmethode liefert das nächstgelegene Minimum. Dieses muss nicht zwangsweise das globale Minimum hinsichtlich der Zielfunktion sein. Durch diese Methode wird die Lage und die Konformation des Liganden in der Bindetasche nur lokal optimiert, so dass z. B. eine Anpassung an das jeweilige Wechselwirkungsmodell erreicht wird.

Beide vorgestellten Optimierungsverfahren beinhalten einen Term, um eine Überlappung von Atomen innerhalb des Liganden zu verhindern, alle weiteren Terme der Zielfunktion werden dem Optimierungsverfahren über die GeoHYDE-Funktion (Kapitel 6.4.1) geliefert.

5

Parametrisierung des logP-Modells

In diesem Kapitel wird die Re-Parametrisierung der logP-Parameter der HYDE-Bewertungsfunktion beschrieben. Die atombasierten logP-Parameter wurden im HYDE-Prototyp unter Berücksichtigung der lösungsmittelzugänglichen Oberfläche von Atomen bestimmt (vgl. Kapitel 4.1.2). Bei der Re-Parametrisierung der neuen HYDE-Version wurde daher eine Studie zur Untersuchung des Zusammenhangs von logP-Wert und der Oberfläche eines Moleküls durchgeführt. Diese wird im ersten Teil des Kapitel vorgestellt. Im zweiten Teil wird die Re-Parametrisierung von HYDE bezüglich des Parametrisierungsdatensatzes, der verwendeten Typen und der Methode beschrieben.

5.1 Molekulare Oberfläche und logP-Wert

Die molekulare Oberfläche und der logP-Wert sind zwei wichtige Eigenschaften eines Moleküls. Sie werden z. B. als Deskriptoren in einigen Protein-Ligand-Bewertungsfunktionen [64, 118, 126, 131, 142, 143] genutzt, wurden aber vor allem in quantitativen Struktur-Wirkungs-Beziehungen (QSAR) (z. B. [20, 215, 216]) integriert. Bei der Oberfläche des Moleküls wird dabei teilweise zwischen polarer und apolarer Oberfläche unterschieden, wobei die polare Oberfläche häufig zur Vorhersage der Absorption von Molekülen herangezogen wird (z. B. [215, 217]). Im Folgenden werden kurz die verschiedenen Oberflächen, die für ein Molekül berechnet werden können, beschrieben. Anschließend

wird auf den logP-Wert eines Moleküls eingegangen und eine Beziehung zwischen diesen beiden Moleküleigenschaften hergestellt.

5.1.1 Verschiedene Oberflächen-Modelle eines Moleküls

Obwohl die Elektronen, die ein Molekül umgeben, keine wohldefinierte Oberfläche bilden, wurden dennoch verschiedene Modelle zur Beschreibung der Oberfläche eines Moleküls entwickelt. Diese erwiesen sich als vielversprechende Deskriptoren in verschiedenen Anwendungen [64, 118, 131, 143, 215, 216]. Die drei gebräuchlichsten Oberflächen-Modelle sind: die Van-der-Waals-Oberfläche (VDW), die lösungsmittelzugängliche Oberfläche (SAS) [205–207] und die molekulare Oberfläche (MS) (auch als Connolly-Oberfläche bezeichnet) [207, 218–220].

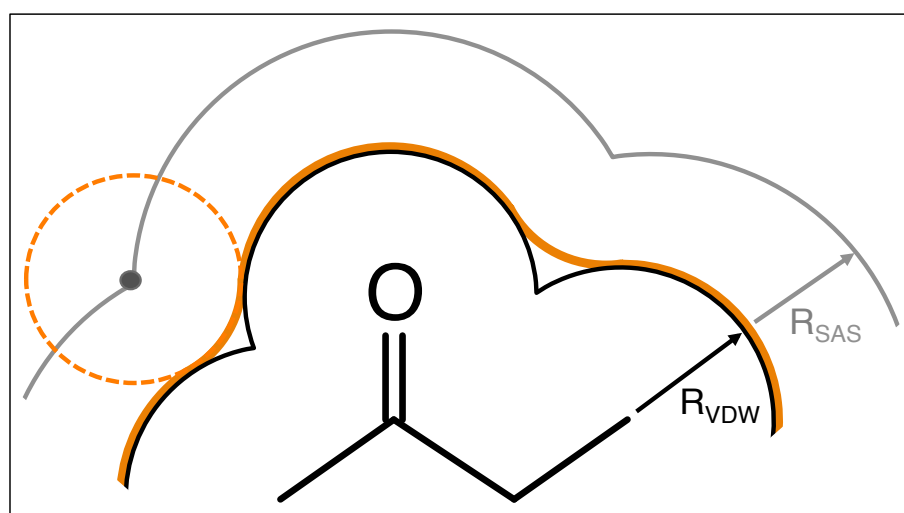


Abbildung 5.1: Schematische Darstellung der verschiedenen Oberflächen eines Moleküls. Dargestellt ist die Van-der-Waals-Oberfläche eines Moleküls (schwarze Linie, R_{VDW}), die molekulare oder Connolly-Oberfläche (orange Linie) und die lösungsmittelzugängliche Oberfläche (SAS) (graue Linie, R_{SAS}). Der orange-gestrichelte Kreis stellt die Lösungsmittel-Abrollkugel dar.

Um die Van-der-Waals-Oberfläche zu generieren, wird das Molekül als Vereinigung von ideellen harten Kugeln modelliert, wobei die Mittelpunkte der Kugeln den Atomkoordinaten und die Radien der Kugeln den Van-der-Waals-Radien der Atome [221] entsprechen (vgl. Abbildung 5.1). Die SAS repräsentiert die Oberfläche des Moleküls, die für den Mittelpunkt eines ideellen Lösungsmittel-Moleküls, das als Kugel modelliert

wird, zugänglich ist, ohne dass eine Überlappung mit den Van-der-Waals-Kugeln des Moleküls auftritt (vgl. Abbildung 5.1, schwarze Linie). Sie wird durch die Erweiterung des Atom-Radius um den Radius der Lösungsmittel-Kugel gebildet oder durch Abrollen der Lösungsmittel-Kugel über die Van-der-Waals-Oberfläche des Moleküls. Dabei entspricht die SAS dem Mittelpunkt der Abrollkugel (vgl. Abbildung 5.1, graue Linie) [205–207]. Die molekulare oder Connolly-Oberfläche besteht aus den konvexen Kontaktflächen der Abroll-Kugel mit der Van-der-Waals-Oberfläche des Moleküls und den sattelförmigen Flächen, die durch den gleichzeitigen Kontakt der Abroll-Kugel mit zwei oder drei Atomen entstehen (vgl. Abbildung 5.1, orange Linie) [207, 218–220]. Die Größe dieser Oberfläche entspricht der Größe der Kavität, die das gelöste Molekül im Lösungsmittel erzeugt.

5.1.2 logP - Logarithmus des Oktanol-Wasser-Verteilungskoeffizienten

Der Oktanol-Wasser-Verteilungskoeffizient ist der Quotient aus der Konzentration eines ungeladenen Moleküls in dem Zweiphasensystem, das aus einer Oktanol- und einer Wasser-Phase gebildet wird. Gewöhnlich wird nicht der Verteilungskoeffizient selbst angegeben, sondern der dekadische Logarithmus des Verteilungskoeffizienten $\log P$.

$$\log P = \log \left(\frac{[\text{gelöste Substanz}]_{\text{Oktanol}}}{[\text{gelöste Substanz}]_{\text{Wasser}}} \right) \quad (5.1)$$

Der logP-Wert ist eine stoffspezifische Konstante und dient als Maß für die Hydrophobizität eines Moleküls. Das bedeutet, Moleküle, die sich vorwiegend in der Oktanol-Phase aufhalten, werden als *hydrophob* bezeichnet und besitzen einen positiven logP-Wert. Im Gegensatz dazu werden Moleküle mit einem negativen logP-Wert als *hydrophil* bezeichnet, da diese vermehrt in der Wasser-Phase zu finden sind. Daher wird der logP-Wert auch häufig herangezogen, um zu entscheiden, ob ein Molekül theoretisch durch eine biologische Membran diffundieren kann und spielt somit in der Pharmakologie eine wichtige Rolle, z. B. als Parameter in der *Rule of Five* [173].

Experimentell kann der logP-Wert eines Moleküls durch die aufwendige Schüttelmethode sehr genau ermittelt werden [222–224]. Eine schnellere, jedoch weniger genaue, Bestimmung des logP-Wertes einer Substanz bietet die Hochdruck-Flüssigkeitschromatographie (HPLC) [225]. Weitere experimentelle Methoden sind in [226] zu finden. Neben den experimentellen Methoden wurden einige computerbasierte Methoden zur Vorhersage des logP-Wertes von Molekülen entwickelt [227]. Dazu wird von einer Additivität des logP-Wertes ausgegangen und anhand experimentell gemessener logP-Werte vieler verschiedener Moleküle werden atombasierte [228, 229] oder fragmentbasierte [230, 231]

logP-Inkrementen abgeleitet. Die verschiedenen Methoden unterscheiden sich in der Definition und Anzahl der Atom- bzw. Fragmenttypen als auch in der, bei der Parametrisierung eingesetzten, Regressionsmethode [227].

Bei der experimentellen Ermittlung des logP-Wertes stellen geladene Moleküle eine Schwierigkeit dar, da sie in verschiedenen protomeren Formen vorliegen. In diesen Fällen werden für die unterschiedlichen Spezies logD-Werte bestimmt. Diese werden bei einem pH-Wert gemessen, in dem das jeweilige Molekül in neutralem Zustand vorliegt. In den computerbasierten Methoden kann zwischen verschiedenen Atomtypen eine lineare Abhängigkeit auftreten, wodurch eine genaue Zuordnung der Einzelbeiträge dieser Atomtypen durch die Regression nicht möglich ist. Dieses Problem wurde bereits bei der Entwicklung der ersten Methoden festgestellt [228]. Weiterhin ist die Additivität nicht immer gegeben, da sich räumlich benachbarte funktionelle Gruppen gegenseitig durch intramolekulare Wechselwirkungen beeinflussen können [230]. In einer Vergleichsstudie zu verschiedenen logP-Vorhersagemethoden wurde gezeigt, dass die Vorhersage-Qualität dieser Methoden mit der Komplexität der Moleküle abnimmt [227].

5.1.3 Analyse der Beziehung von logP-Wert und Molekül-Oberfläche

Die logP-Parameter der HYDE-Bewertungsfunktion wurden im HYDE-Prototyp unter Berücksichtigung der lösungsmittelzugänglichen Oberfläche von Atomen kalibriert. Im Zuge der Re-Parametrisierung der logP-Parameter der HYDE-Bewertungsfunktion, wurde die Beziehung zwischen dem logP-Wert eines Moleküls und seiner Oberfläche näher untersucht [232]. Diese Beziehung als auch vergleichbare Zusammenhänge, wie z. B. zwischen der Wasserlöslichkeit eines Moleküls und seiner lösungsmittelzugänglichen Oberfläche, wurden bereits in verschiedenen Studien behandelt [52, 233–240]. Neben der Oberfläche des Moleküls wurden auch schon andere Moleküleigenschaften, wie z. B. Molekulargewicht [241] oder Volumen [238–240, 242, 243], in Beziehung zu dem logP-Wert bzw. der Wasserlöslichkeit eines Moleküls gesetzt. Durch die Analyse der Beziehung von logP-Wert und Oberfläche eines Moleküls sollen weitere Erkenntnisse über die Wechselwirkungen von polaren und apolaren Molekülen mit der wässrigen Phase gewonnen werden. Diese können zur Verbesserung des Wasserstoffbrücken- sowie des Dehydratations-Terms der HYDE-Bewertungsfunktion herangezogen werden.

LogP-Datensatz

Für diese Analyse wurde ein Datensatz aus 454 kleinen Molekülen mit experimentell gemessenen logP-Werten zusammengestellt, der auch zur Re-Parametrisierung der

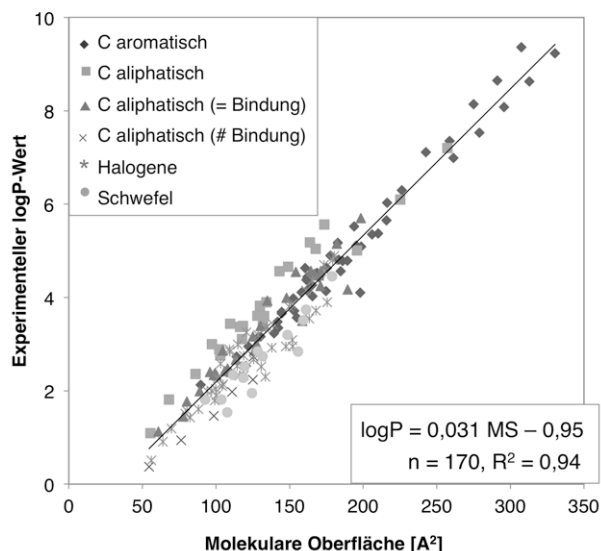
logP-Parameter von HYDE eingesetzt wurde. Die Moleküle wurden der PHYSPROP-Datenbank [204] entnommen und stammen aus der Sammlung von Hansch und Leo [203, 244, 245]. Es wurden nur Moleküle ausgewählt, die maximal ein Heteroatom - Sauerstoff, Stickstoff, Schwefel oder Halogen - enthalten. Dadurch sollen Nachbarschaftseffekte, wie intramolekulare Wechselwirkungen, vermieden werden, da diese Einfluss auf den logP-Wert haben [230]. Zur Analyse der Wechselwirkungen von polaren Gruppen mit der wässrigen Phase wurden zwei weitere Datensätze zusammengestellt: einer bestehend aus 49 Molekülen mit einer Amid-Gruppe und ein weiterer aus 91 Molekülen mit zwei identischen funktionellen Gruppen. Mithilfe von CORINA (Version 3.48) [246, 247] wurden 3D-Koordinaten der Moleküle, die im SMILES-Format vorlagen, generiert. Fehlende Wasserstoffatome wurden von CORINA hinzugefügt, so dass die Moleküle in neutraler Form vorlagen.

Korrelation von logP-Wert und Molekül-Oberfläche

In der Analyse wurden zwei verschiedene Molekül-Oberflächen getestet: die SAS und die molekulare Oberfläche (vgl. Abschnitt 5.1.1). Wasserstoffatome wurden bei beiden Oberflächen implizit berücksichtigt, indem die Van-der-Waals-Radien der Schweratome um 0,1 Å für jedes gebundene Wasserstoffatom erhöht wurden. Die genaue Berechnungsweise der Oberfläche findet sich in Kapitel 6.2.1. In Abbildung 5.2 ist die Korrelation des logP-Wert mit der molekularen Oberfläche von 170 hydrophoben Molekülen des Datensatzes dargestellt. Unabhängig vom Oberflächen-Modell - SAS oder molekulare Oberfläche - wird eine sehr hohe Korrelation zwischen dem experimentellen logP-Wert der Moleküle und ihrer Oberfläche erreicht (SAS: $R = 0,96$, molekulare Oberfläche: $R = 0,97$). Wird im Vergleich dazu anstatt der Molekül-Oberfläche das Molekulargewicht mit dem logP-Wert korreliert, sinkt der Korrelationskoeffizient auf $R = 0,8$.

In den 170 hydrophoben Molekülen sind aliphatische und aromatische Kohlenwasserstoffe enthalten, sowie Moleküle, die ein Schwefelatom oder ein Halogen enthalten. In Abbildung 5.2 wird deutlich, dass sowohl die schwefelhaltigen als auch die halogenhaltigen Moleküle sich gut in die Regression einfügen und keine signifikante Abweichung von der Regressionsgeraden aufweisen. Werden diese Moleküle von der Korrelationsanalyse ausgeschlossen, so ändert sich der Korrelationskoeffizient nicht ($R = 0,97$). Von diesem Resultat kann abgeleitet werden, dass sich Schwefel und Halogene in organischen Verbindungen ähnlich zu Kohlenstoff verhalten und keine oder nur sehr schwache Wasserstoffbrücken mit den umgebenden Wassermolekülen eingehen. Innerhalb der Genauigkeit, die in einem logP-Modell erreicht werden kann, können diese Moleküle alle

Abbildung 5.2: Korrelation des experimentellen logP-Wertes von hydrophoben Molekülen und ihrer molekularen Oberfläche. Bild in Anlehnung an [232]



gleich beschrieben werden, auch unabhängig vom Hybridisierungszustand (vgl. Abbildung 5.2 Beispiel Kohlenstoffe).

Einfluss von polaren Atomen auf den logP-Wert

Um den Einfluss von polaren Atomen auf den logP-Wert zu bestimmen, wurde die Korrelation des experimentellen logP-Wertes mit der molekularen Oberfläche von hydrophilen und hydrophoben Molekülen verglichen. Betrachtet man hydrophile Moleküle mit nur einer funktionellen Gruppe, so zeigt sich wiederum eine hohe Korrelation zwischen der molekularen Oberfläche und dem experimentellen logP-Wert dieser Moleküle ($R = 0,95 - 0,96$). Unabhängig von der funktionellen Gruppe ergibt sich eine parallele Verschiebung der Korrelationsgeraden der hydrophilen Moleküle im Vergleich zu den hydrophoben Molekülen in Richtung eines niedrigeren logP-Wertes (vgl. Abbildung 5.3).

Das bedeutet, dass polare Atome zu einer besseren Löslichkeit in der wässrigen Phase führen, die durch einen konstanten negativen Beitrag zum logP-Wert wiedergespiegelt wird. Abhängig von der funktionellen Gruppe variiert diese parallele Verschiebung (vgl. Abbildung 5.3). Polare Atome, die an einen aromatischen Rest gebunden sind, weisen einen niedrigeren Beitrag zum logP-Wert auf, was durch die Elektronegativität der aromatischen Gruppe ausgelöst werden kann.

Dieser additive Beitrag von polaren Atomen zum logP-Wert zeigt sich auch, wenn Moleküle mit zwei gleichen funktionellen Gruppen in den Vergleich einbezogen werden. Sind die funktionellen Gruppen räumlich voneinander getrennt, ergibt sich eine zweite

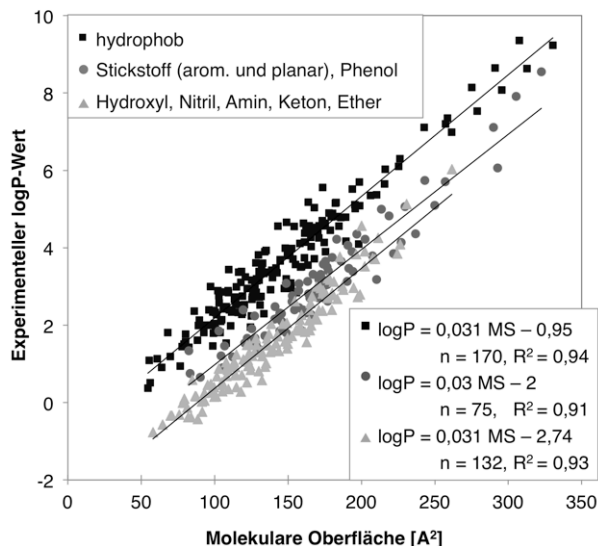


Abbildung 5.3: Einfluss von polaren Atomen auf den experimentellen logP-Wert. Gezeigt ist die Korrelationen von experimentellen logP-Werten mit der molekularen Oberfläche von hydrophoben Molekülen im Vergleich zu hydrophilen Molekülen mit verschiedenen funktionellen Gruppen. Bild in Anlehnung an [232]

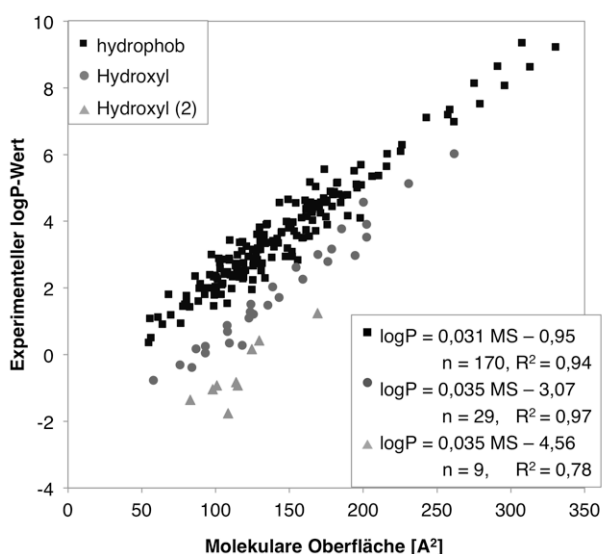


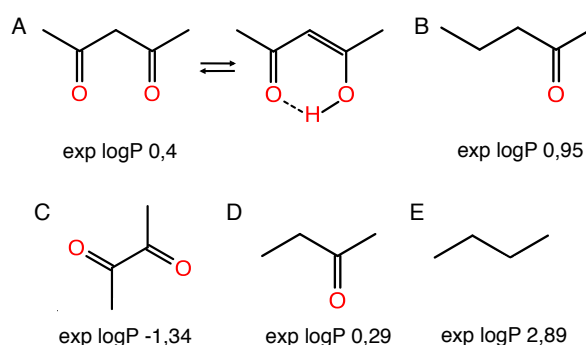
Abbildung 5.4: Beispiel für die Additivität die logP-Wert anhand von Molekülen mit einer und zwei Hydroxyl-Gruppen. Gezeigt ist die Korrelation des experimentellen logP von hydrophoben Molekülen und ihrer molekularen Oberfläche im Vergleich zu Molekülen mit einer oder zwei Hydroxyl-Gruppen. Bild in Anlehnung an [232]

parallel-verschobene Regressionsgerade. Diese befindet sich, verglichen mit der Regressionsgeraden der Moleküle mit nur einer funktionellen Gruppe, in fast doppelter Distanz zu der Regressionsgeraden der hydrophoben Moleküle. In Abbildung 5.4 ist dieses Resultat exemplarisch für Moleküle mit einer und zwei Hydroxyl-Gruppen dargestellt. Wie bereits oben erwähnt, können Nachbarschaftseffekte auftreten, sobald die funktionellen Gruppen räumlich-benachbart sind. Abbildung 5.5 zeigt die Auswirkungen eines solchen Nachbarschaftseffekts auf den logP-Wert anhand von zwei Molekülen mit jeweils zwei Carbonyl-Gruppen (Acetylaceton und Diaacetyl). In beiden Fällen sind die zwei

5. Parametrisierung des logP-Modells

Carbonyl-Gruppen eng benachbart, jedoch kann sich bei Acetylaceton (Abbildung 5.5 A) eine intramolekulare Wasserstoffbrücke ausbilden, da Keto- und Enolform in Lösung koexistieren. Vergleicht man die logP-Werte mit denen der entsprechenden Moleküle mit nur einer Carbonyl-Gruppe, so wird deutlich, dass im Fall von Acetylaceton keine additive Beziehung der zwei Carbonyl-Gruppe im Hinblick auf den logP-Wert gegeben ist. Bei Diaacetyl hingegen liegt diese Additivität annähernd vor (vgl. Abbildung 5.5).

Abbildung 5.5: Beispiel für einen Nachbarschaftseffekte von polaren Atomen. Intramolekulare Wechselwirkungen beeinflussen den logP-Wert. A) Acetylaceton, Keto- und Enolform koexistieren in Lösung. B) 2-Pentanon C) Diaacetyl D) 2-Butanon E) Butan



Viele der funktionellen Gruppen im Datensatz können theoretisch multiple Wasserstoffbrücken mit der wässrigen Phase ausbilden. Es stellt sich die Frage, wie sich die Anzahl der Wasserstoffbrücken, die ein polares Atom eingehen kann, auf seinen logP-Wert auswirkt. Normalerweise wird angenommen, dass jede zusätzliche Wasserstoffbrücke einen gleichen zusätzlichen Energiebeitrag liefert. Diese Annahme wird auch in den meisten Bewertungsfunktionen dementsprechend realisiert, wodurch es im Fall von multifurkaten Wasserstoffbrücken auch zu einer Überbewertung dieser Moleküle kommen kann (vgl. Kapitel 2.5.2). Um dies näher zu analysieren, wurde die molekulare Oberfläche von Aminen, Amiden, Hydroxyl-Gruppen und Ethern mit ihren experimentellen logP-Werten korreliert. Diese funktionellen Gruppen wurden ausgesucht, da sie sich jeweils nur in der Anzahl ihrer Wasserstoffbrücken-Funktionen unterscheiden. Der Stickstoff eines primären Amines hat drei Wasserstoffbrücken-Funktionen: zwei Donor- und eine Akzeptorfunktion oder, im protonierten Zustand, drei Donorfunktionen. Dementsprechend haben sekundäre Amine eine Donor- und eine Akzeptorfunktion oder zwei Donorfunktionen (im protonierten Zustand). Tertiäre Amine besitzen nur eine Donor- oder eine Akzeptorfunktion. Durch die Planarität von Amid-Gruppen (sp^2 -Hybridisierung) stehen dem primären Amid-Stickstoff nur zwei Donorfunktionen und keine Akzeptorfunktion zur Verfügung. Der sekundäre Amid-Stickstoff hat eine Donorfunktion und der tertiäre Amid-Stickstoff kann keine Wasserstoffbrücke ausbilden. Im dritten Fall wird die Hydroxyl-Gruppe mit der Ether-Gruppe verglichen. Während die Hydroxyl-Gruppe

eine Donor- und zwei Akzeptorfunktionen besitzt, hat der Ether-Sauerstoff nur zwei Akzeptorfunktionen. Abbildung 5.6 zeigt das Resultat dieses Experiments exemplarisch für Amine, Hydroxyl-Gruppe und Ether (Amid Resultat s. [232]). In allen drei Fällen hat die Anzahl der Wasserstoffbrücken, die ein polares Atom theoretisch ausbilden kann, keinen Einfluss auf seinen Beitrag zum logP-Wert. Dieses Resultat steht im Widerspruch zu dem Konzept von Wasserstoffbrücken oder Wasserstoffbrücken-Energie, wie es normalerweise in der molekularen Modellierung angewendet wird. In den meisten Anwendungen wird der Beitrag von Wasserstoffbrücken aufsummiert, auch wenn zwei oder mehr Wasserstoffbrücken vom gleichen Atom ausgebildet werden. Eine Ausnahme dazu bildet die AIScore Protein-Ligand-Bewertungsfunktion, die eine auf quanten-chemischen Berechnungen basierende Skalierung von bifurkaten Wasserstoffbrücken berücksichtigt [148] (vgl. Kapitel 2.5.2).

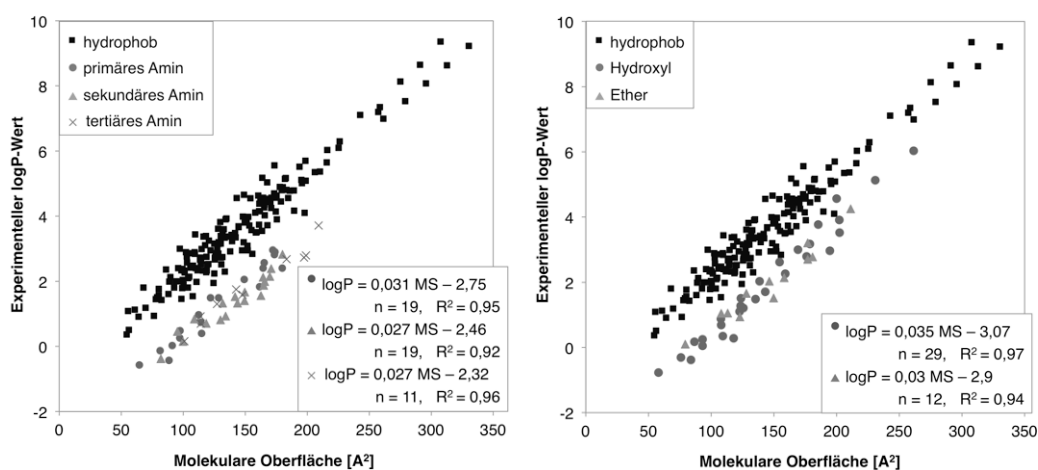


Abbildung 5.6: Einfluss von multiplen Wasserstoffbrücken-Donoren/Akzeptoren auf den experimentellen logP-Wert. Gezeigt ist die Korrelation von hydrophoben Molekülen und deren Oberfläche im Vergleich zu Molekülen mit einer Amid-Gruppe (links) und Molekülen mit einer Hydroxyl- oder Ether-Gruppe (rechts). Bild in Anlehnung an [232]

5.2 Atombasierte logP-Parametrisierung in HYDE

Zur Verbesserung der HYDE-Bewertungsfunktion und durch die Veränderung der Zugänglichkeitsberechnung in der neuen HYDE-Version (vgl. Kapitel 6.2.1) wurden die atombasierten logP-Inkrementen des HYDE-Prototyps [201] vollständig überarbeitet. Dabei verfolgte die erste Strategie eine weitere Verfeinerung der plogP-Typen, z. B. eine zusätzliche Unterscheidung zwischen Sulfonamid-Stickstoff und Amid-Stickstoff

oder auch die Trennung von Carbonyl-Sauerstoff und Sulfonamid/Sulfato/Phosphato-Sauerstoff. Dies führte zu zehn zusätzlichen plogP-Typen, die auf einem größeren und weitaus diverseren Parametrisierungsdatensatz als dem ursprünglichen kalibriert wurden. Die resultierenden logP-Inkrementen lagen deutlich unter den Beiträgen der ursprünglichen logP-Inkrementen. Da die plogP-Werte in HYDE auch zur Berechnung der Wasserstoffbrücken-Energie verwendet werden, reduzierte sich der energetische Beitrag von Wasserstoffbrücken stark unterhalb dem Wert, der für Wasserstoffbrücken in wässriger Phase angenommen wird (-2 bis -6 kJ/mol [46]). Außerdem wurde für einige der plogP-Typen eine lineare Abhängigkeit festgestellt, wodurch die in der Parametrisierung verwendete multiple lineare Regression diesen Typen keine vertrauenswürdigen Inkrementen zuweisen konnte. Die Vergrößerung des Parametrisierungsdatensatzes führte zudem zu mehr statistischem Rauschen in den resultierenden Inkrementen, da die enthaltenen Moleküle eine höhere Diversität und Komplexität aufwiesen. Aufgrund der Schwächen dieser Strategie wurde als nächstes eher eine Reduzierung der bestehenden plogP-Typen angestrebt. Ziel dabei war es, plogP-Inkrementen abzuleiten, die nicht durch Rauschen, wie z. B. Nachbarschaftseffekte von polaren Atomen, beeinflusst werden, um damit Basis-Beiträge von Atomen zu Wasserstoffbrücken als auch zum hydrophoben Effekt in HYDE abschätzen zu können. Daher wurde ein neuer Parametrisierungsdatensatz, wie bereits im vorherigen Abschnitt vorgestellt, aus 454 kleinen, einfachen Molekülen mit maximal einem Hetero-Atom zusammengestellt, so dass Nachbarschaftseffekte nicht auftreten können. Mit diesem Datensatz wurde die oben beschriebene Studie (Abschnitt 5.1.3) zur Beziehung von logP-Wert und Molekül-Oberfläche durchgeführt. In Bezug auf die HYDE Bewertungsfunktion konnten dabei für die Re-Parametrisierung wichtige Vorbedingungen, wie die Einteilung der plogP-Typen, abgeleitet werden als auch ein neues Konzept für die Bewertung von multifurkaten Wasserstoffbrücken erstellt werden.

5.2.1 Einteilung der neuen plogP-Atomtypen und Parametrisierung

Durch die Korrelation von logP-Wert mit molekularer Oberfläche wurde festgestellt, dass sich schwefel- und halogenhaltige Moleküle dabei sehr ähnlich den Kohlenwasserstoffen verhalten (Abschnitt 5.1.3). Daher wurden Schwefel, Halogene und Kohlenstoff in der Re-Parametrisierung der logP-Parameter der HYDE-Bewertungsfunktion in einem plogP-Atomtyp der *hydrophoben Oberfläche* zusammengefasst. Phosphor, der nicht im Parametrisierungsdatensatz enthalten war, da Phosphor nicht als einzelnes Hetero-Atom in organischen Verbindungen auftaucht, wurde ebenfalls diesem plogP-Typ zugeteilt. Im Vergleich zum HYDE-Prototypen konnten die plogP-Parameter dadurch um 22 Typen

reduziert werden. Damit genügt ein plogP-Typ, der nur von der zugänglichen molekularen Oberfläche des Atoms abhängt, zur Beschreibung von hydrophoben Atomen.

Bei polaren Atomen wurde festgestellt, dass sie einen konstanten additiven Beitrag zum logP-Wert eines Moleküls liefern, sofern sie sich nicht gegenseitig durch Nachbarschaftseffekte beeinflussen. Um dies zu verhindern, wurden im Parametrisierungsdatensatz nur Moleküle aufgenommen, die maximal ein Hetero-Atom besitzen. Außerdem zeigte die Untersuchung, dass die Anzahl der Wasserstoffbrücken-Donoren/Akzeptoren keinen Einfluss auf den logP-Wert hat, daher wird bei Sauerstoff- und Stickstoffatomen nur noch nach Hybridisierungszustand und Delokalisierung der Elektronen unterschieden. Für Stickstoff ergeben sich folglich drei plogP-Typen: *sp³-hybridisierter*, *sp²-hybridisierter* und *sp-hybridisierter Stickstoff*. Sauerstoffatome werden in vier plogP-Typen eingeteilt: *sp³-hybridisierter* und *sp³-hybridisierter delokalisierte Sauerstoff*, sowie *sp²-hybridisierter* und *sp²-hybridisierter delokalisierte Sauerstoff*. Im Vergleich zum HYDE-Prototyp konnten die plogP-Typen für polare Atome damit von 31 auf 7 Typen reduziert werden.

Mittels dieser 8 Typen wurde auf dem Parametrisierungsdatensatz eine multiple lineare Regression mit folgender Inkrementformel durchgeführt:

$$\log P = \sum_{\text{polar atoms } i}^n p \log P_k(i) + \sum_{\text{apolar atoms } j}^m acc(j) \cdot p \log P_{\text{hydrophobic_surface}(j)} \quad (5.2)$$

Dabei wird bei polaren Atomen i das logP-Inkrement $p \log P$ abhängig vom plogP-Typ k nach der Anzahl kalibriert, während bei apolaren Atomen j , die nur durch den plogP-Typ *hydrophobic_surface* beschrieben werden, das logP-Inkrement $p \log P_{\text{hydrophobic_surface}}$ abhängig von der zugänglichen molekularen Oberfläche acc des Atoms j bestimmt wird. Zur Durchführung der multiplen linearen Regression wurde das plogP-Parametrisierungs-Modul der Flex*-Bibliothek genutzt, das bereits für den HYDE-Prototyp implementiert wurde [202].

5.2.2 Resultierende plogP-Inkmente

Bei der Regression konnte eine gute Korrelation zwischen den experimentellen und den berechneten logP-Werten ($R = 0,93$) innerhalb des Parametrisierungsdatensatzes erzielt werden (vgl. Abbildung 5.7). Dies bestätigt die Anwendbarkeit der gewählten plogP-Typen und der Inkrementformel zur Bestimmung des logP-Wertes. Die plogP-Typen und die resultierenden plogP-Inkmente sind in Tabelle 5.1 zusammengefasst. Polare Atome erhalten hierbei einen konstanten negativen Beitrag zum logP-Wert, während

apolaren Atomen ein positives plogP-Inkrement zugewiesen wird. Damit werden polare Atome korrekt als hydrophil und apolare Atome korrekt als hydrophob klassifiziert. Weiterhin liegt die Größenordnung der plogP-Inkmente in einem realistischen Bereich, so dass für eine geometrisch ideale Wasserstoffbrücke, z. B. zwischen einem Carbonyl-Sauerstoff und einem Amid-Stickstoff, mittels der HYDE-Funktion ein energetischer Beitrag von $-3,1 \text{ kJ/mol}$ berechnet werden kann. Der hydrophobe Effekt, der aus dem plogP-Inkrement apolarer Atome mit HYDE abgeschätzt wird, liegt mit einem Wert von $143 \text{ J/\AA}^2 \text{ mol}$ etwas über dem Wert, den Eisenberg und McLachlan dafür ermittelt hatten ($110 \text{ J/\AA}^2 \text{ mol}$) [53].

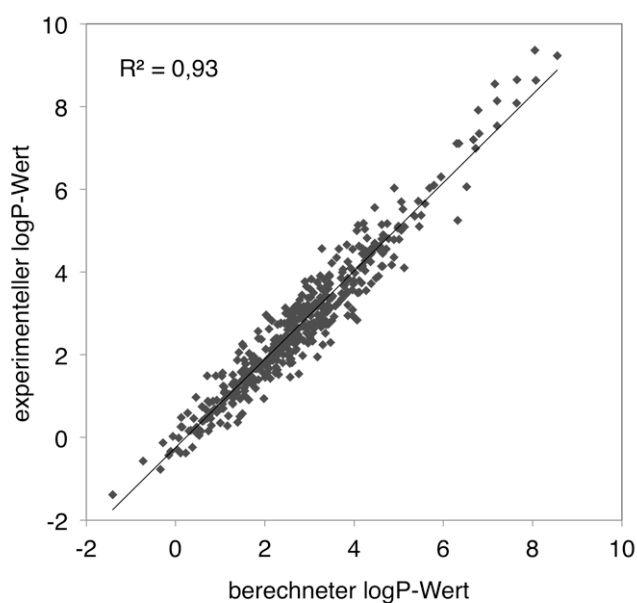


Abbildung 5.7: Korrelation der experimentellen und berechneten logP-Werte innerhalb des Parametrisierungsdatensatzes.

Besondere Gruppen bei der Parametrisierung

Geladenen Gruppen, wie z. B. Carboxylaten, wird der plogP-Wert ihrer ungeladenen Form zugewiesen, da es für diese Gruppen keine experimentellen logP-Werte gibt, sondern nur die pH-abhängigen logD-Werte. Der pH-Wert in der Bindetasche von Proteinen unterscheidet sich jedoch von Bindetasche zu Bindetasche und wahrscheinlich auch von dem des umgebenden wässrigen Mediums, daher ist die Berücksichtigung des pH-Wertes bei der Bewertung von Protein-Ligand-Komplexen nur schwer realisierbar. Polare Atome, die keine Wasserstoffbrücken ausbilden können, wie sp^2 -hybridisierte Stickstoffe, die keine Bindung zu einem Wasserstoffatom haben, wurden von der Parametrisierung ausgeschlossen, ihnen wird ein plogP-Wert von Null zugewiesen. Dies gilt auch für Gruppen, von denen angenommen wird, dass sie durch Delokalisierung der Elektronen nicht an

Wasserstoffbrücken beteiligt sein können, z. B. Ether-Gruppen, die zwei aromatische Systeme überbrücken.

plogP-Typ	plogP-Wert
hydrophobe Oberfläche (C, S, P, F, Cl, Br, I)	0,025 [1/Å]
Stickstoff	
sp ³ -hybridisiert	-1,77
sp ² -hybridisiert	-1,24
sp-hybridisiert	-1,44
Sauerstoff	
sp ³ -hybridisiert	-1,41
sp ³ -hybridisiert (delokalisiert)	-0,73
sp ² -hybridisiert	-1,51
sp ² -hybridisiert (delokalisiert)	-0,58

Tabelle 5.1: plogP-Parameter

6

Berechnung der HYDE-Funktion

In Kapitel 4.1 wurde bereits der HYDE-Prototyp vorgestellt. Dieser wurde in der hier zugrunde liegenden Arbeit vollständig überarbeitet, um bestehende Mängel zu verbessern (vgl. Kapitel 4.1.4) und neue Funktionalität zu integrieren [208]. Dabei wurde das Grundkonzept der HYDE-Bewertungsfunktion (Kapitel 3) beibehalten, jedoch wurde die Berechnung der energetischen Beiträge von polaren Atomen substantiell im Vergleich zum Prototypen verändert. Der Algorithmus zur Berechnung der Zugänglichkeit von Atomen wurde durch eine effizientere Berechnungsmethode ersetzt. Die bestehenden logP-Parameter wurden insgesamt auf einem neuen Parametrisierungsdatensatz mit einer stark reduzierten Anzahl von Atomtypen neu kalibriert (Kapitel 5). Weiterhin wurden neue Terme hinzugefügt, die die Anordnung von Wassermolekülen um beide Moleküle - Protein und Ligand - im ungebundenen Zustand berücksichtigen. Außerdem wurde die HYDE-Funktion um die Bewertung von Metallionen im Protein-Ligand-Komplex erweitert. Die überarbeitete HYDE-Funktion wurde in ein Optimierungsverfahren integriert, so dass eine verbesserte Vorhersage von Protein-Ligand-Komplexen ermöglicht wird. In diesem Kapitel werden die überarbeiteten Terme und die neuen Methoden im Detail beschrieben. Abschließend wird anhand eines Beispiels die Bewertung eines Protein-Ligand-Komplexes mit HYDE gezeigt und der Einfluss einzelner Atome auf die HYDE-Energie diskutiert.

6.1 HYDE - allgemeine Form

Die HYDE-Bewertungsfunktion unterliegt einem intuitiven Konzept: Beide Moleküle - Protein und Ligand - sind im ungebundenen Zustand von Wassermolekülen umgeben. Während des Bindungsprozesses werden die Wassermoleküle, die den Liganden umgeben, abgestreift und die, die sich in der Bindetasche des Proteins befinden, werden durch den Liganden verdrängt. Dabei werden die Wasserstoffbrücken zwischen den Wassermolekülen und den polaren Atomen der beiden Moleküle aufgebrochen, was zu einem ungünstigen enthalpischen Energiebeitrag führt, der nur durch neue, geometrisch gute Wasserstoffbrücken im Protein-Ligand-Komplex ausgeglichen werden kann. Apolare Teile des Liganden und des Proteins, die im ungebundenen Zustand in Kontakt mit Wasser sind, verursachen dagegen eine Störung des Wasserstoffbrücken-Netzwerkes von Wasser, was eine ungünstige Energie zur Folge hat. Werden die apolaren Teile beider Moleküle in der Komplexbildung zusammen gelagert und somit die Störung des Wasserstoffbrücken-Netzwerkes von Wasser behoben, resultiert daraus ein erheblicher Energiegewinn zur freien Bindungsenergie, der hydrophobe Effekt. In der HYDE-Bewertungsfunktion wird angenommen, dass diese Prozesse, die bei der Bindung eines Liganden an ein Protein stattfinden, den Hauptanteil zur freien Bindungsenergie liefern. Daher werden in der HYDE-Bewertungsfunktion ausschließlich Energiebeiträge von Wasserstoffbrücken und dem hydrophoben Effekt sowie energetisch ungünstige Beiträge durch die polare Dehydratation berücksichtigt.

Die allgemeine Form der HYDE-Funktion wurde bereits in Kapitel 4.1 eingeführt (Gleichung 4.1), diese wurde in der Überarbeitung der HYDE-Bewertungsfunktion nicht verändert:

$$\Delta G_{\text{HYDE}} = \sum_{\text{atoms } i} \Delta G_{\text{dehydration}}^i + \Delta G_{\text{H-bond}}^i$$

In HYDE wird für alle Atome i des Liganden und der Bindetasche des Proteins die Änderung der Dehydratations- und der Wasserstoffbrücken-Energie berechnet.

Im Folgenden werden die Einzeltermine $\Delta G_{\text{dehydration}}^i$ und $\Delta G_{\text{H-bond}}^i$ der HYDE-Funktion und ihre Überarbeitung vorgestellt.

6.1.1 Berechnung der Dehydratationsenergie

Während die Dehydratation von hydrophoben Atomen stabilisierend zur freien Bindungsenergie beiträgt, ist die Dehydratation von hydrophilen Atomen energetisch ungünstig. In der überarbeiteten Fassung der HYDE-Funktion, wurden zwei separate Ter-

me zur Abschätzung der Dehydratationsenergie von hydrophoben und hydrophilen Atomen entwickelt:

Hydrophobe/apolare Atome

$$\Delta G_{dehydration}^{i,hydrophobic} = -2.3RT \cdot p \log P^i \cdot (acc_{unbound}^i - acc_{bound}^i) \quad (6.1)$$

Für alle hydrophoben Atome i des Liganden und der Protein-Bindetasche wird die Änderung der Zugänglichkeit acc^i zwischen *ungebundenem* und *gebundenem* Zustand berechnet und mit dem jeweiligen logP-Inkrement $p \log P^i$ multipliziert, um die Dehydratationsenergie des Atoms abzuschätzen. Die Berechnung der Dehydratationsenergie von hydrophoben Atomen (Gleichung 6.1) ist damit fast identisch zur Berechnungsweise im HYDE-Prototyp (vgl. Gleichung 4.4). Eine Normierung der Zugänglichkeit durch die mittlere Zugänglichkeit eines Atomtyps ist durch die Re-Parametrisierung der logP-Inkmente nicht mehr erforderlich. In der überarbeiteten Fassung der HYDE-Funktion wird zudem die Änderung der Zugänglichkeit nicht über die lösungsmittelzugängliche Oberfläche, sondern über die molekulare Oberfläche mit einem neuen effizienteren Algorithmus berechnet. Der genaue Algorithmus zur Berechnung der Oberfläche wird in Abschnitt 6.2 beschrieben.

Hydrophile/polare Atome

$$\Delta G_{dehydration}^{i,hydrophilic} = -2.3RT \cdot p \log P^i \cdot f_{bur}^i \cdot f_{water}^i \cdot \sum_{H\text{-bond functions } j} w^j \cdot p_{dehyd}^j \quad (6.2)$$

Bei hydrophilen Atomen wurde eine neue Berechnungsweise der Dehydratationsenergie eingeführt. Zuvor wurde die Dehydratationsenergie von hydrophilen Atomen ebenso wie die von hydrophoben Atomen durch Gleichung 4.4 berechnet, indem eine gewichtete lösungsmittelzugängliche Oberfläche für hydrophile Atome verwendet wurde. Dabei wurden für hydrophile Atome nur die Teile der lösungsmittelzugänglichen Oberfläche hinzugenommen, die in der bevorzugten Wasserstoffbrücken-Richtung, d. h. in Richtung eines freien Elektronenpaares bzw. eines polaren Wasserstoffatoms, lagen (vgl. Kapitel 4.1). In der überarbeiteten Fassung der HYDE-Funktion wird die Dehydratationsenergie eines hydrophilen Atoms anhand der einzelnen Wasserstoffbrücken-Funktionen des Atoms abgeschätzt. Dazu wird für jede Wasserstoffbrücken-Funktion j eine Dehydratationswahrscheinlichkeit p_{dehyd}^j und eine Gewichtung w^j ermittelt. Außerdem wird die Vergrabenheit f_{bur}^i des hydrophilen Atoms i und die Anordnung der Wassermoleküle im ungebundenen Zustand f_{water}^i um das Atom i zur Berechnung der Dehydratationsenergie berücksichtigt.

Dehydrationswahrscheinlichkeit p_{dehyd}^j

Die Zugänglichkeit von hydrophilen Atomen wird überprüft, indem abgeschätzt wird, ob in der bevorzugten Wasserstoffbrücken-Richtung im Protein-Ligand-Komplex ausreichend Raum für ein potentielles Wassermolekül vorhanden ist. Dazu wird die Dehydrationswahrscheinlichkeit p_{dehyd}^j jeder einzelnen Wasserstoffbrücken-Funktion j eines hydrophilen Atoms berechnet. Wasserstoffbrücken-Funktionen, die durch die Komplexbildung vollständig abgedeckt werden oder an Wasserstoffbrücken beteiligt sind, erhalten eine Dehydrationswahrscheinlichkeit von $p_{dehyd}^j = 1$. In den anderen Fällen, nimmt die Dehydrationswahrscheinlichkeit linear ab, sobald in der bevorzugten Wasserstoffbrücken-Richtung mindestens Raum für das halbe Volumen eines Wassermoleküls vorzufinden ist. Die Berechnung der Zugänglichkeit wird in Abschnitt 6.2 genauer vorgestellt.

Gewichtung w^j

Zusätzlich wurde eine Gewichtung w^j jeder einzelnen Wasserstoffbrücken-Funktion j eines hydrophilen Atoms eingeführt. Diese Gewichtung geht aus der LogP-Studie des vorherigen Kapitels 5.1.3 hervor. In dieser Analyse zeigte sich, dass die Anzahl der Wasserstoffbrücken-Funktionen eines hydrophilen Atoms keinen Einfluss auf seinen logP-Wert hat. Man kann annehmen, dass dies durch geometrische Restriktionen, bedingt durch die tetraedrische Geometrie des Wasserstoffbrücken-Netzwerkes von Wasser (vgl. Kapitel 3) hervorgerufen wird. Das bedeutet, polaren Atomen mit mehr als einer Wasserstoffbrücken-Funktion ist es nur eingeschränkt möglich, sich an das Wasserstoffbrücken-Netzwerkes von Wasser anzupassen, um so geometrisch gute Wasserstoffbrücken mit den Wassermolekülen auszubilden. Da dies Einfluss auf die Dehydrations- und Wasserstoffbrücken-Energie eines hydrophilen Atoms hat, wurde folgendes Gewichtungsschema eingeführt: Bei multiplen Wasserstoffbrücken-Funktionen eines hydrophilen Atoms bekommt die Wasserstoffbrücken-Funktion, die an der geometrisch besten Wasserstoffbrücke beteiligt ist, eine Gewichtung von 100 %. Bei der zweitbesten ist die Gewichtung auf 20 % reduziert und die drittbeste trägt nur noch mit einem Gewicht von 10 % bei. Weitere Wasserstoffbrücken-Funktionen bekommen eine Gewichtung von 0 % und liefern somit keinen Beitrag mehr zur freien Bindungsenergie. Wasserstoffbrücken-Funktionen, die keine Wasserstoffbrücke ausbilden, werden entsprechend ihrer Dehydrationswahrscheinlichkeit gewichtet, so dass die Wasserstoffbrücken-Funktion mit der kleinsten Wahrscheinlichkeit die höchste Gewichtung bekommt.

Vergrabenheit f_{bur}^i

Für jedes hydrophile Atom i der Bindetasche des Proteins wird die Vergrabenheit f_{bur}^i im ungebundenen Zustand bestimmt. Dieser Faktor dient dazu, stark vergrabene Atome der Bindetasche, die keine Wasserstoffbrücke zum Solvens ausbilden können, höher zu gewichten als die exponierten Atome der Bindetasche, die mit dem Solvens oder dem Liganden Wasserstoffbrücken eingehen können. Mit diesem Faktor wird auch die höhere Flexibilität von exponierten Seitenketten berücksichtigt, die die genaue Abschätzung der Zugänglichkeit von diesen erschwert. Dieser Wert wird nach dem Strahlen-Ansatz von Stahl berechnet [248]. Vergrabenen Proteinatomen wird ein Faktor von 1 zugewiesen, exponierte Atome der Bindetasche erhalten je nach Zugänglichkeit einen Wert zwischen 1 und 0. Im Fall von Ligandatomen wird dieser Faktor auf 1 gesetzt.

Anordnung der Wassermoleküle f_{water}^i

In der HYDE-Bewertungsfunktion werden Wassermoleküle implizit behandelt. Daher wurde ein Korrekturfaktor f_{water}^i eingeführt, der die lokale Anordnung der Wassermoleküle um ein hydrophiles Atom i im ungebundenen Zustand berücksichtigt. Dieser Korrekturfaktor f_{water}^i wird folgendermaßen berechnet:

$$f_{water}^i = \sum_{\text{H-bond functions } j \text{ of atom } i} water_{overlap}^j \cdot water_{interaction}^j \quad (6.3)$$

Über die zwei Faktoren $water_{overlap}^j$ und $water_{interaction}^j$ wird für jede Wasserstoffbrücken-Funktion j eines hydrophilen Atoms eine Abschätzung über die Qualität der Wechselwirkungen von dieser mit dem Solvens im ungebundenen Zustand gemacht, da dies einen direkten Einfluss auf die Dehydratationsenergie hat. Das Konzept auf dem die HYDE-Funktion beruht (vgl. Kapitel 3) geht von einem idealen Zustand aus: Jede Wasserstoffbrücken-Funktion wird dabei durch ein Wassermolekül ideal im ungebundenen Zustand abgesättigt. Dies trifft bei isolierten Wasserstoffbrücken-Funktionen zu, jedoch ist die lokale Anordnung der Wassermoleküle in der Bindetasche des Proteins und bei Liganden mit vielen benachbarten polaren Gruppen geometrischen Restriktionen unterworfen. Dadurch können die Dehydratationskosten von Wasserstoffbrücken-Funktionen, die im solvatisierten Zustand nicht ideal abgesättigt werden, sinken. Diese Annahme wird durch die Beobachtung gestärkt, dass der logP-Wert von kleinen Molekülen nicht linear mit der Anzahl der gebundenen polaren Gruppen abfällt.

Überlappende Wasser Der $water_{overlap}^j$ -Term liefert eine Abschätzung der Anzahl der Wassermoleküle, die um ein hydrophiles Atom angeordnet werden können, so dass die Dehydratationskosten zwischen polaren Gruppen aufgeteilt werden können (vgl. Abbildung 6.1 A und B). Dazu werden hypothetische Wassermoleküle (Radius = 1,4 Å) auf die ideale Position eines Wasserstoffbrücken-Partners vor jeder Wasserstoffbrücken-Funktion der Protein-Bindestasche und des Liganden platziert (vgl. Abbildung 6.1 B). Für jedes dieser hypothetischen Wassermoleküle i wird die Überlappung mit den anderen, umgebenden, hypothetischen Wassermolekülen j berechnet:

$$water_{overlap}^i = 1 - \frac{\sum_{\text{surrounding waters } j} \frac{1}{2} \cdot \text{overlap volume}(water^i, water^j)}{\text{volume}(water)} \quad (6.4)$$

Abbildung 6.1 A zeigt schematisch eine kleine hydrophile Subtasche der Bindestelle des Proteins. Drei polare Gruppen interagieren hier mit dem gleichen Wassermolekül. Im Gegensatz dazu stellt Abbildung 6.1 B die Überlappung von drei ideal platzierten Wassermolekülen in dieser kleinen hydrophilen Tasche dar. Das Überlappungsvolumen von Wassermolekül i mit den beiden anderen Wassermolekülen j und k wird jeweils berechnet (vgl. Abbildung 6.1 C und D). Die Summe der Überlappungsvolumina wird mit dem Volumen eines hypothetischen Wassermoleküls normalisiert. In diesem Fall würde der $water_{overlap}^j$ -Term von Wasser i in einem Drittel resultieren, wodurch die Dehydratationskosten der Wasserstoffbrücken-Funktion, von der Wassermolekül i abgeleitet wurde, auf ein Drittel gesenkt werden.

Konservierte Wasser Der $water_{interaction}^j$ -Faktor ist komplementär zu dem $water_{overlap}^j$ -Term. Dieser Faktor liefert eine Abschätzung der Sättigung eines Wassermoleküls, das mit einer bestimmten polaren Gruppe im ungebundenen Zustand interagieren würde. Abhängig von der Anzahl der Wasserstoffbrücken, an denen ein Wassermolekül beteiligt ist, können die Dehydratationskosten stark variieren. Wassermoleküle, die in der Bindestasche des Proteins mehrere Wasserstoffbrücken ausbilden können (Abbildung 6.1 A), sind meist stark konserviert, wodurch ein Verdrängen dieser durch den Liganden enthalpisch sehr ungünstig ist. Im Gegensatz dazu sind Wassermoleküle in kleinen hydrophoben Subtaschen des Proteins, die nur mit einer polaren Gruppe interagieren können, entropisch und enthalpisch sehr ungünstig und die Dehydratationskosten dieser könnten überschätzt werden. Zur Berechnung des $water_{interaction}^j$ -Faktor wird wieder ein hypothetisches Wassermolekül auf die ideale Position eines Wasserstoffbrücken-Partners, vor jeder Wasserstoffbrücken-Funktion der

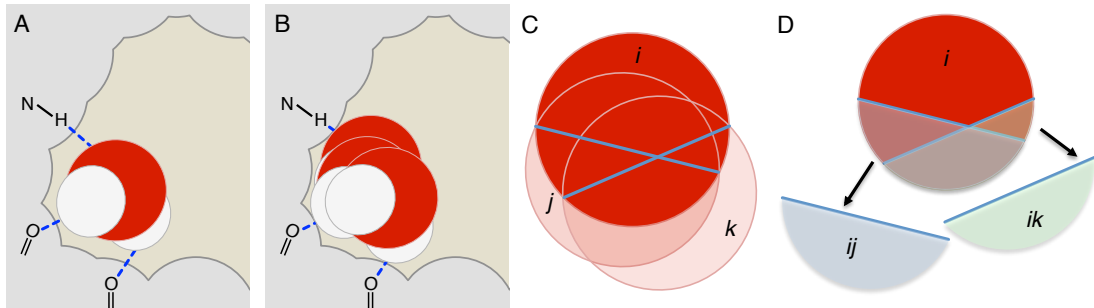


Abbildung 6.1: Lokale Anordnung von Wassermolekülen und Überlappung von ideal platzierten Wassermolekülen. A) Drei polare Gruppen interagieren mit dem gleichen Wassermolekül. B) Überlappung von drei ideal platzierten Wassermolekülen. C) Überlapp von Wassermolekül i mit den Wassermolekülen j und k . D) Überlappungsvolumen von Wassermolekül i mit den Wassermolekülen j und k . Bild aus [208]

Protein-Bindetasche und des Liganden platziert. Dieses Wassermolekül wird dann um 60 mögliche Richtungen rotiert, um das beste Wasserstoffbrücken-Netzwerk zu finden und die Anzahl der unangesättigten Wasserstoffbrücken-Funktion des Wassermoleküls zu reduzieren. Der $water_{interaction}^j$ Faktor ergibt sich aus der Kombination der Anzahl der gemachten Wasserstoffbrücken und der Anzahl der unangesättigten Wasserstoffbrücken-Funktion (vgl. Tabelle 6.1).

H-Brücken	unangesättigte H-Brücken-Funktionen	$water_{interaction}$ Faktor
3	0 ; 1	1,4
2	0	1,2
2	1 ; 2	1
1	0 ; 1	1
0	0 ; 1	0,8
1	2	0,7
0	2	0,4
0	3	0

Tabelle 6.1: Kombinationen und resultierende Werte des $water_{interaction}$ -Faktors

6.1.2 Berechnung der Wasserstoffbrücken-Energie

Um die Komplementarität der Wasserstoffbrücken- und der Dehydratationsenergie deutlich zu machen, wurde in der HYDE-Funktion ein ähnlicher Term zur Berechnung entwickelt. Die Wasserstoffbrücken-Energie wird in der HYDE-Funktion atombasiert ermittelt, d. h. die beiden beteiligten polaren Atome erhalten jeweils den, ihrem $\log P$ -Inkrement $p \log P^i$ entsprechenden, energetischen Beitrag. Daher kann für jedes polare Atom i des Liganden und der Protein-Bindetasche der Betrag zu einer Wasserstoffbrücke folgendermaßen abgeschätzt werden:

$$\Delta G_{H\text{-bond}}^i = \frac{2.3RT}{F_{sat}} \cdot p \log P^i \cdot f_{bur}^i \cdot \sum_{H\text{-bond } j} w^j \cdot f_{dev}^j \quad (6.5)$$

Basierend auf dem in Kapitel 3 beschriebenen Konzept, entsteht in HYDE ein stabilisierender Beitrag für eine intermolekulare Wasserstoffbrücke aus der Tatsache, dass nicht alle Wasserstoffbrücken im Wasserstoffbrücken-Netzwerk des die Moleküle umgebenden Wassers geometrisch ideal ausgebildet werden. Die Energie, die benötigt wird, um die Wasserstoffbrücken-Funktionen der Moleküle - Protein und Ligand - zu dehydratisieren ist damit niedriger, als der Energiegewinn einer geometrisch idealen, intermolekularen Wasserstoffbrücke. In Kapitel 3 wurde gezeigt, dass diese Dehydratationsenergie einer polaren Funktion genau um den Sättigungsfaktor $F_{sat}(T)$ niedriger ist. Daher wurde der reziproke Sättigungsfaktor $1/F_{sat}$ in die Berechnung der Wasserstoffbrücken-Energie integriert, um so den energetischen Gewinn einer Wasserstoffbrücke gegenüber den Dehydratationskosten der Wasserstoffbrücken-Funktionen, zu bestimmen (vgl. Kapitel 4.1). Als Temperatur wurde $T = 298 \text{ K}$ gewählt, da die experimentellen Bindungsaffinitäten von Protein-Ligand-Komplexen meist bei dieser Temperatur gemessen werden. Damit ergibt sich ein Sättigungsfaktor $F_{sat}(298\text{K}) = 0,85$. Der Energiegewinn einer idealen intermolekularen Wasserstoffbrücke ist somit um 17 % ($= 1/F_{sat}(298\text{K})$) höher als die Dehydratationsenergie der Wasserstoffbrücken-Funktion.

Der energetische Beitrag einer Wasserstoffbrücke hängt stark von ihrer geometrischen Qualität ab, d. h. Distanz und Winkel sollten innerhalb eines idealen Bereichs liegen. Die geometrische Qualität einer Wasserstoffbrücke j wird über den Faktor f_{dev}^j in Gleichung 6.5 berücksichtigt. Die idealen geometrischen Parameter der verschiedenen Atomtypen werden in HYDE über spherische Wechselwirkungsflächen modelliert. Diese Wechselwirkungsflächen repräsentieren die ideale Position eines Wechselwirkungspartners. Sie basieren auf dem Wechselwirkungsmodell von FlexX [80], wurden für HYDE aber nochmal verfeinert (vgl. Kapitel 4.4). Geometrisch ideale Wasserstoffbrücken j erhalten einen Faktor $f_{dev}^j = 1$. Bei einer Abweichung von der idealen Geometrie ($f_{dev}^j < 1$) werden die Wasserstoffbrücken linear gegen Null skaliert.

Die anderen beiden Faktoren - die Gewichtung w^j einer Wasserstoffbrücken-Funktion und die Vergrabenheit f_{bur}^i eines polaren Atoms - wurden bereits zur Berechnung der Dehydratationsenergie von polaren Atomen integriert und im vorherigen Abschnitt vorgestellt. Im Vergleich zum HYDE-Prototyp, indem die Überbewertung von multiplen Wasserstoffbrücken, die von einem Atom ausgebildet werden, mithilfe eines Wurzelterms verhindert werden sollte, wurde in der überarbeiteten Fassung der HYDE-Funktion stattdessen die Gewichtung w^j von Wasserstoffbrücken-Funktionen eingeführt.

6.2 Änderung der Zugänglichkeit von Atomen

Die Dehydratationsenergie wird in HYDE über die, durch die Komplexbildung induzierte, Änderung der Zugänglichkeit der Atome berechnet. Um diese Änderung zu bestimmen, wird nicht wie in vielen anderen Methoden/Funktionen oder auch dem HYDE-Prototyp dazu die Änderung der lösungsmittelzugänglichen Oberfläche (SAS) [205–207] berechnet, sondern die molekulare Oberfläche [207, 218–220] eines Atoms im ungebundenen und im gebundenen Zustand abgeschätzt. Die verschiedenen Oberflächenmodelle, die für Moleküle existieren, wurden in Kapitel 5.1.1 behandelt. Im Folgenden wird der Algorithmus zur Berechnung der molekularen Oberfläche in HYDE vorgestellt, bevor auf die Berechnung der Zugänglichkeitsänderung eingegangen wird.

6.2.1 Berechnung der molekularen Oberfläche

Zur atombasierten Berechnung der molekularen Oberfläche wird ein Oberflächennetz um jedes Molekül - Ligand und Bindetasche des Proteins - erzeugt. Jedem, der entstehenden Netzknoten, wird ein entsprechender Oberflächenanteil zugeordnet, so dass über die, dem Atom zugehörigen, Netzknoten, die molekulare Oberfläche jedes einzelnen Atoms bestimmt werden kann.

Generierung des Oberflächennetzes

Das Oberflächennetz wird auf der SAS des jeweiligen Moleküls generiert, wobei die Lösungsmittel-Abrollkugel einen Radius von $R_{SAS} = 1,4 \text{ \AA}$ hat (vgl. Kapitel 5.1.1). Für die Schweratome wird der Standard-Van-der-Waals-Radius [221] genommen. Wasserstoffatome werden implizit behandelt, indem der Radius des zugehörigen Schweratoms jeweils um $0,1 \text{ \AA}$ pro Wasserstoffatom erweitert wird. Um eine gleichmäßige Verteilung der Netzknoten des Oberflächennetzes zu erhalten, wird für jedes Atom ein einbeschriebener Ikosaeder generiert, d. h. der Mittelpunkt des Ikosaeders liegt im Zentrum des

Atoms und der Radius des Ikosaeders ergibt sich aus der Summe des Van-der-Waals-Radius (R_{VDW}) des Atoms und des Radius der Abrollkugel (R_{SAS}). Der Ikosaeder wird in einer zweistufigen rekursiven Verfeinerung weiter unterteilt. Dabei werden die Kanten des Ikosaeders in jeder Stufe halbiert. Dadurch entstehen pro Ikosaeder-Dreieck drei neue Eckpunkte, aus denen vier neue Dreiecke generiert werden können (Abbildung 6.2 A). Die neu generierten Punkte werden anschließend noch auf den Ikosaeder-Radius rausgeschoben (Abbildung 6.2 B). Nach der zweiten Verfeinerungsstufe sind 162 Ikosaeder-Eckpunkte entstanden, die eine sehr gleichmäßige Verteilung der Netzknoten garantieren und eine gute Beschreibung der Kugeloberfläche des zugrunde liegenden Atoms liefern (Abbildung 6.2 C).

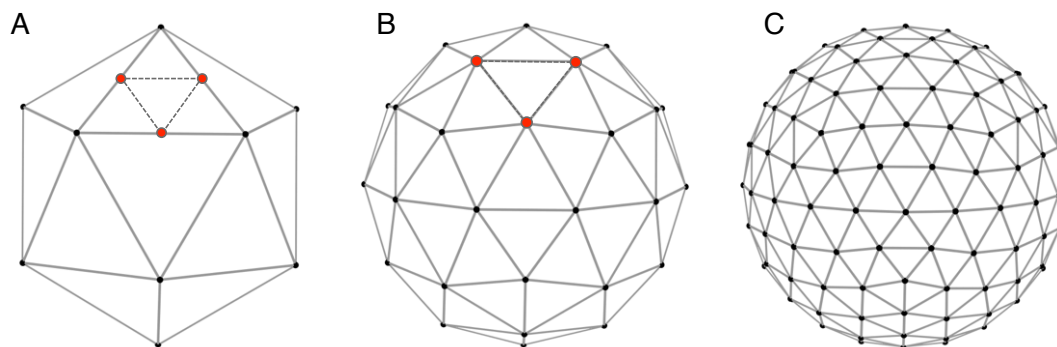


Abbildung 6.2: Rekursive Ikosaeder Verfeinerung. A) Ikosaeder mit 12 Eckpunkten, 20 Dreiecken und 30 Kanten. Die Verfeinerung eines Dreiecks ist in rot dargestellt. B) Erste Verfeinerungsstufe des Ikosaeders (42 Eckpunkte). C) Zweite Verfeinerungsstufe des Ikosaeders (162 Eckpunkte).

Innenliegende, d. h. durch benachbarte Atome abgedeckte Ikosaeder-Punkte müssen während der Generierung des Oberflächennetzes erkannt und ausgeschlossen werden. Dazu werden die Schnittebenen eines Atoms mit seinen Nachbaratomen gebildet. Alle Ikosaeder-Punkte, die hinter einer Schnittebene liegen, sind durch Nachbaratome verdeckt und werden vom Oberflächennetz ausgeschlossen. Die Schnittebenen entsprechen zugleich jeweils der Ebene eines Torus zwischen zwei Nachbaratomen. Ein Torus entsteht, wenn die Abrollkugel gleichzeitig zwei benachbarte Atome berührt. Die einzelnen Torussegmente einer Torusbahn werden durch die stationären Punkte bestimmt, die sich bilden, sobald die Abrollkugel Kontakt zu einem weiteren, dritten Nachbaratom hat. Torussegmente, die von anderen nahegelegenen Atomen abgedeckt werden sowie stationäre Punkte, die mit Nachbaratomen kollidieren, werden nicht in das Oberflächennetz

aufgenommen. Die verbleibenden Torussegmente werden entsprechend der euklidischen Distanz von zwei Ikosaeder-Punkten diskretisiert und die entstehenden Punkte werden als Netzknoten dem Oberflächennetz zugeordnet. Die zugänglichen stationären Punkte werden ebenfalls als Netzknoten in das Oberflächennetz integriert (vgl. Abbildung 6.3). Zur besseren Veranschaulichung des Algorithmus zur Generierung des Oberflächennetzes ist der Pseudocode dargestellt:

Algorithmus 6.1: Generierung des Oberflächennetzes

Data : Molekül A

Result : OberflächenNetzVektor

foreach Atom *a* in Molekül A **do**

generiere $P_I \leftarrow$ IkosaederPunkte(*a*, Verfeinerungsstufe *n*);

foreach Atom *b* in Nachbaratome(*a*) **do**

berechne $E_T \leftarrow$ SchnittEbenen(*a*, *b*);

if Index(*b*) > Index(*a*) **then**

foreach Atom *c* in (Nachbaratome(*a*) \cap Nachbaratome(*b*)) **do**

berechne $(P_K, S_T) \leftarrow$ KonkavePunkteUndTorusSegmente(*a*, *b*, *c*);

OberflächenNetzVektor $\leftarrow P_K$;

berechne $P_T \leftarrow$ TorusSegmentDiskretisierung(*a*, *b*, S_T);

OberflächenNetzVektor $\leftarrow P_T$;

berechne $P_{IO} \leftarrow$ InnenliegendeIkosaederPunkte(P_I , E_T);

OberflächenNetzVektor $\leftarrow P_{IO}$;

Berechnung der Oberflächen-Inkmente

Jedem der Netzknoten wird eine Oberflächen-Inkrement zugeordnet. Entsprechend ihres Ursprungs - Ikosaeder-Punkt, Torus-Punkt oder stationärer Punkt - entstehen unterschiedliche Oberflächen-Inkmente:

Ikosaeder-Punkt Das Oberflächen-Inkrement eines Ikosaeder-Punktes $A_{\text{Ikosaeder}}$ wird als Anteil an der Kugeloberfläche des jeweiligen Atoms ermittelt. Dazu wird die

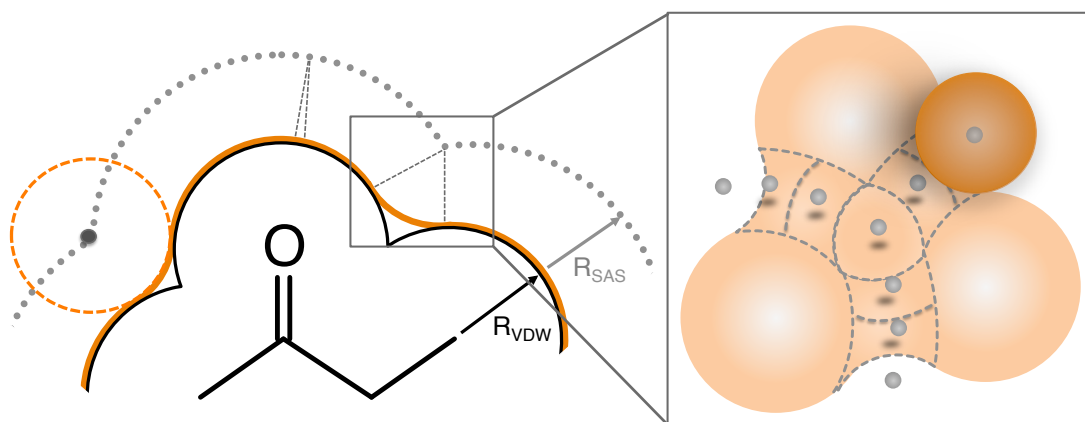


Abbildung 6.3: Generierung des Oberflächennetzes auf der SAS eines Moleküls. Links: Dargestellt ist das Oberflächennetz auf der SAS (graue Punkte), die Abrollkugel (orange gestrichelter Kreis), die molekulare Oberfläche (orange Linie) und die VDW-Oberfläche des Moleküls (schwarze Linie). Rechts: drei diskretisierte Torusbahnen und der stationärer Punkt (konkave Fläche) zwischen den Torusbahnen. Bild in Anlehnung an [208]

Kugeloberfläche durch die Anzahl der Ikosaeder-Punkte, die von der Verfeinerungsstufe n des Ikosaeders abhängen, dividiert:

$$A_{\text{Ikosaeder}} = \frac{\text{Kugeloberfläche des Atom}}{\text{Anzahl der Ikosaeder-Punkte}} = \frac{4\pi r_{VDW}^2}{2 + 10 \cdot 4^n} \quad (6.6)$$

Torus-Punkt Bei der Diskretisierung eines Torussegments entstehen rechteckige Sattelflächen (vgl. Abbildung 6.3). Die Kantenlänge der Rechtecke kann jeweils als Kreissegment über den Radius der Abrollkugel und den Winkel zwischen den Berührungspunkten bzw. über den Radius der Torusbahn und den Winkel zwischen den Diskretisierungsschritten berechnet werden. Die Fläche eines Torus-Punktes entspricht dann der Fläche des zugrunde liegenden rechteckigen Sattelfläche.

stationärer Punkt Das Oberflächen-Inkrement eines stationären Punktes wird als Oberfläche des unterliegenden sphärischen Dreiecks berechnet (vgl. Abbildung 6.3):

$$A_{\text{konkav}} = (\alpha + \beta + \gamma - \pi) \cdot R_{SAS}^2 \quad (6.7)$$

α , β und γ sind dabei die Innenwinkel des sphärischen Dreiecks in Radiant. R_{SAS} ist der Radius der Abrollkugel, die das sphärische Dreieck aufspannt.

Die Oberflächen-Inkmente werden jeweils an den Netzknoten gespeichert, so dass für jedes Atom mithilfe der zugehörigen Netzknoten seine molekulare Oberfläche abgeschätzt werden kann. Die Summe der Oberflächen-Inkmente aller Netzknoten entspricht der molekularen Oberfläche des jeweiligen Moleküls.

6.2.2 Abschätzung der Zugänglichkeitsänderung

Im HYDE-Prototyp wurde die Abschätzung der Zugänglichkeitsänderung über die WSAS berechnet (vgl. Kapitel 4.1.2). In der ersten Überarbeitung des Oberflächenmodells wurde daher eine gewichtete molekulare Oberfläche entsprechend der WSAS berechnet. Dadurch wurden weiterhin tief vergrabene polare Atome unterschätzt, wodurch das Modell dahingehend variiert wurde, dass die polaren Netzknoten der Oberfläche weiter rausgeschoben wurden. Dies führte ebenfalls nicht zur optimalen Beschreibung der polaren Dehydratation, da durch die Verwendung der Wechselwirkungsflächen von polaren Atome (vgl. Kapitel 4.1.2) nicht der Raum abgebildet wird, den ein Wassermolekül bei einer Interaktion im solvatisierten Zustand einnehmen würde. Aus diesem Grund wurde ein Oberflächenmodell entwickelt, in dem die polare Dehydratation durch die Verdrängung von hypothetischen Wassermolekülen modelliert wird.

Für beide Moleküle - Ligand und Bindetasche des Proteins - wird ein Oberflächennetz generiert (Abbildung 6.4). Dabei wird für den Liganden die gleiche Konformation im ungebundenen und im gebundenen Zustand angenommen und das Protein wird als starr betrachtet. Die Änderung der Zugänglichkeit von Atomen wird somit nur durch die Komplexbildung verursacht.

Bei hydrophoben Atomen wird die Zugänglichkeitsänderung ermittelt, indem die Oberflächen-Inkmente der durch die Schweratome des anderen Moleküls abgedeckten Netzknoten aufsummiert werden. Diese Änderung der Zugänglichkeit zwischen ungebundenem und gebundenem Zustand wird in Gleichung 6.1 genutzt, um die Dehydratationsenergie eines hydrophoben Atoms abzuschätzen. In Abbildung 6.4 B sind die durch die Atome der Bindetasche des Proteins abgedeckten Netzknoten des Liganden dargestellt und Abbildung 6.4 C zeigt die durch den gebundenen Liganden, abgedeckten Netzknoten der Bindetasche.

Für polare Atome wird die Zugänglichkeitsänderung nicht über die Abdeckung der molekularen Oberfläche, sondern über die Verdrängung von hypothetischen Wassermolekülen bestimmt. Dazu wird vor jeder Wasserstoffbrücken-Funktion eines polaren Atoms an der bevorzugten Position des Wasserstoffbrücken-Partners, ein hypothetisches Wassermolekül platziert (Abbildung 6.4 A). Im Komplex wird die Überlappung

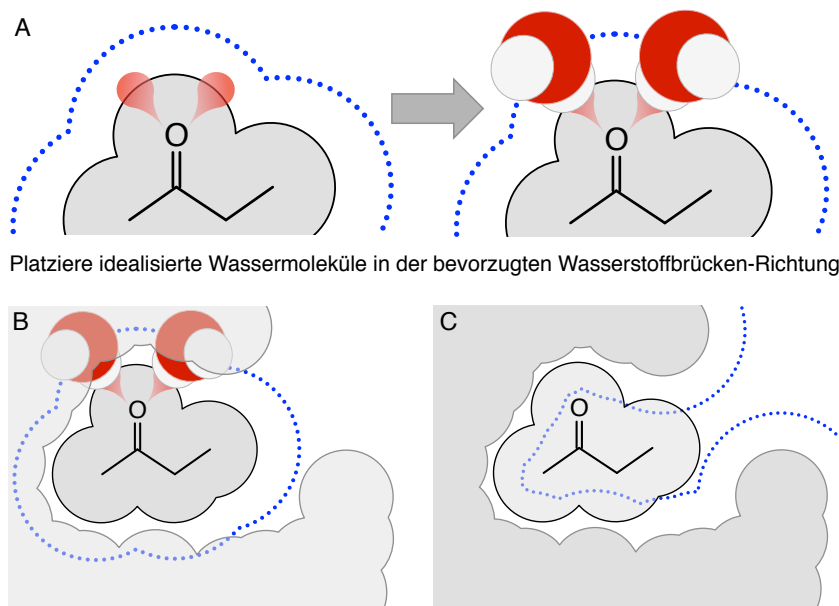


Abbildung 6.4: Bestimmung der Zugänglichkeit von Atomen. A) Generierung des Oberflächennetzes für den Liganden (blaue Punkte), für hydrophile Atome werden hypothetische Wassermoleküle an der bevorzugten Position der Wasserstoffbrücken-Partners platziert. B) Durch die Bindetasche induzierte Zugänglichkeitsänderung des Liganden. C) Durch den Liganden induzierte Zugänglichkeitsänderung der Bindetasche. Bild in Anlehnung an [208]

dieses Wassermoleküls mit den umgebenden Schweratomen des anderen Moleküls berechnet. Abbildung 6.4 B stellt die Überlappung von zwei hypothetischen Wassermolekülen einer Carbonyl-Gruppe des Liganden mit der Bindetasche des Proteins dar. Beträgt die Summe der Überlappungsvolumina mit den umgebenden Schweratomen mehr als das halbe Volumen eines hypothetischen Wassermoleküls (Radius = 1,4 Å), so wird die Wasserstoffbrücken-Funktion j , von der das hypothetische Wassermolekül abgeleitet wurde, als dehydratisiert angesehen und für die Dehydratationswahrscheinlichkeit ergibt sich ein Wert von $p_{dehyd}^j = 1$. Anderenfalls, wird die Dehydratationswahrscheinlichkeit linear in Abhängigkeit des Überlappungsvolumens herrunterskaliert:

$$p_{dehyd}^j = \begin{cases} 1 & ,\text{if } overlap > \frac{volume(water)}{2} \\ \frac{2 \cdot overlap}{volume(water)} & ,\text{else} \end{cases} \quad (6.8)$$

$$overlap = \sum_{\text{surrounding atoms } k} overlap\ volume(water^j, atom^k)$$

6.3 Bewertung von Metallionen in HYDE

Fast die Hälfte aller Enzyme benötigen ein Metallion in ihrer Bindetasche, um ihre Funktion ausüben zu können [249]. Dabei übernehmen diese Metalloenzyme auch Schlüsselfunktionen in den, z. B. für die Krebstherapie relevanten, Stoffwechselwegen [250]. Die Bewertung von Protein-Ligand-Komplexen, die ein Metallion als Kofaktor in der Bindetasche des Proteins beinhalten, führte im HYDE-Prototyp zu einer ungenauen Vorhersage, da der Beitrag des Metallions zur freien Bindungsenergie ebenfalls über den regulären Wasserstoffbrücken- und Dehydratationsterm abgeschätzt wurde. In der überarbeiteten Fassung der HYDE-Funktion wurde ein neues Modell für die Bewertung von Metallionen, die in der Bindetasche des Proteins eingebettet sind, integriert.

Die Metallionen von Metalloenzymen sind in der Bindetasche des Proteins in einer bestimmten Koordinierungsgeometrie gebunden, die auch von den jeweiligen Liganden abhängen kann. Das bedeutet, dass bei Metallionen, die an der katalytischen Reaktion beteiligt sind, mindestens eine freie Koordinierungsstelle existiert, mit der der Ligand interagiert. Bleibt eine Koordinierungsstelle unbesetzt, d. h. wird dehydratisiert oder nicht von Gegenionen abgesättigt, so ist dies energetisch sehr ungünstig. Daher wurde in HYDE ein Modell integriert, dass vor der energetischen Bewertung die optimale Koordinierungsgeometrie eines Metallions in der Bindetasche, anhand der umgebenden Proteinatome, bestimmt [51]. Dazu werden die Distanzen und Winkel vom Metallion zu den umliegenden Proteinatomen berechnet, die das Metallion koordinativ binden können, d. h. Stickstoff, Sauerstoff und Schwefel. Anhand dieser geometrischen Restriktionen wird dem Metallion ein kompatibler Polyeder - Tetraeder, planares Rechteck, trigonale Bipyramide oder Oktaeder - als Koordinierungsgeometrie zugeordnet.

Auf Basis dieser Koordinierungsgeometrie wird der energetische Beitrag eines Metallions zur freien Bindungsenergie in der HYDE-Funktion durch folgende Terme abgeschätzt:

$$\Delta G_{\text{HYDE}}^{\text{metal}} = \sum_{\text{metal ions } i} \Delta G_{\text{interaction}}^i + \Delta G_{\text{dehydration}}^i \quad (6.9)$$

$$\Delta G_{\text{interaction}}^i = \epsilon_{\text{interact}}^{\text{metal}} \cdot \sum_{\text{interactions } j} f_{\text{dev}}^j \quad (6.10)$$

$$\Delta G_{\text{dehydration}}^i = \epsilon_{\text{dehyd}}^{\text{metal}} \cdot \sum_{\text{coordination sites } j} p_{\text{dehyd}}^j \quad (6.11)$$

Die energetische Bewertung von Metallionen in HYDE setzt sich entsprechend der Bewertung von polaren Atomen auch wieder aus zwei Energiebeiträgen zusammen, dem Interaktionsterm $\Delta G_{interaction}^i$ und dem Dehydratationsterm $\Delta G_{dehydration}^i$. Diese Aufteilung der Energiebeiträge ist wichtig, um auch die energetisch ungünstigen, unbesetzten Koordinierungsstellen zu berücksichtigen.

Aufgrund fehlender bzw. unzuverlässiger logP-Werte für Metallionen, wurden durch Untersuchung von verschiedenen Metalloenzym-Komplexen empirisch Energie-Inkrementen für die Metallinteraktionen abgeleitet. Damit ergibt sich für, durch den Liganden besetzte, Koordinierungsstellen ein energetisch günstiges Inkrement von $\epsilon_{interact}^{metal} = -20kJ/mol$ und für unbesetzte Koordinierungsstellen ein ungünstiges Energie-Inkrement von $\epsilon_{dehyd}^{metal} = 10kJ/mol$.

Nur der energetische Beitrag von Interaktionen zwischen dem Liganden und dem Metallion werden in HYDE zur Berechnung der freien Bindungsenergie berücksichtigt, die restlichen durch Proteinatome besetzten Koordinierungsstellen tragen nicht zur freien Bindungsenergie bei. In HYDE dienen Stickstoff-, Sauerstoff- und Schwefelatome auf Ligandenseite als Akzeptoren für die Interaktion mit den Metallionen. Der energetische Beitrag von diesen wird weiterhin durch den Wasserstoffbrücken- und den Dehydratationsterm von HYDE ermittelt.

Metallinteraktionen werden in HYDE, wie Wasserstoffbrücken auch, durch überlappende Wechselwirkungsflächen des Metallions und des Ligand-Akzeptor-Atoms modelliert. Ähnlich wie bei Wasserstoffbrücken, ist auch bei Metallinteraktionen die geometrische Qualität der Interaktion zwischen dem Ligandatom und dem Metallion wichtig, um den energetischen Beitrag der Interaktion zu beurteilen. Daher wurde der Faktor f_{dev}^j zur Abschätzung der geometrischen Qualität einer Interaktion j in die Berechnung der Energie einer Metall-Interaktion $\Delta G_{interaction}^i$ integriert. Geometrisch ideale Interaktionen j erhalten einen Faktor von $f_{dev}^j = 1$, anderenfalls wird f_{dev}^j linear gegen Null herunterskaliert. Für alle nicht durch Proteinatome besetzten Koordinierungsstellen wird zudem die Dehydratationsenergie $G_{dehydration}^i$ berechnet. Dazu wird für jede Koordinierungsstelle j die Dehydratationswahrscheinlichkeit p_{dehyd}^j bestimmt, wie in Abschnitt 6.2.2 beschrieben.

6.4 HYDE im Optimierungskontext

In der HYDE-Funktion sind keine Terme zur Beurteilung der Protein-Ligand-Komplex-Geometrie oder zur Abschätzung der Konformationsenergie des Liganden enthalten. Dadurch wird die Überlappung von Protein- und Ligand-Atomen als auch eine energetisch

ungünstige Ligandkonformation, was in den durch Docking vorhergesagten Protein-Ligand-Komplexen durchaus vorkommen kann, nicht erkannt. Weiterhin hat jede Bewertungsfunktion ihr eigenes Optimum, so dass Protein-Ligand-Komplexe, die mit anderen Programmen generiert oder auch durch ein Kristallographie-Verfeinerungsprogramm bestimmt wurden, nicht zwingend mit dem Optimum der HYDE-Bewertungsfunktion übereinstimmen. Eine weitere wichtige Voraussetzung, um eine gute Vorhersage der freien Bindungsenergie eines Protein-Ligand-Komplexes mit HYDE zu erlangen, ist das Vorhandensein eines optimal abgesättigten Wasserstoffbrücken-Netzwerks zwischen Ligand und Protein als auch innerhalb des Proteins.

Die überarbeitete Fassung der HYDE-Funktion wurde daher in einen zweistufigen Optimierungsverfahren eingebettet, um so eine zuverlässige Bewertung eines Protein-Ligand-Komplexes mit HYDE zu ermöglichen. In der ersten Optimierungsstufe wird ein optimales Wasserstoffbrücken-Netzwerk des Protein-Ligand-Komplexes mithilfe von ProToss generiert (vgl. Kapitel 4.2.1). Im zweiten Optimierungsschritt wird die gegebene Protein-Ligand-Komplex-Geometrie in Richtung des Optimums der HYDE-Funktion adaptiert. Dabei wird der Ligand in der Bindetasche des starren Proteins translatiert und rotiert, falls nötig werden auch Torsionswinkel des Liganden verändert. Zwei verschiedene Optimierungsverfahren stehen in diesem Optimierungsschritt zur Verfügung, eine numerische Optimierung, um in das nächste lokale Optimum der HYDE-Funktion zu gelangen, und eine stochastische Optimierung, um das globale Optimum der HYDE-Funktion zu finden. Beide Strategien wurden bereits in Kapitel 4.2.2 genauer beschrieben. Als Zielfunktion beider Optimierungsverfahren dient eine Abwandlung der HYDE-Funktion, die GeoHYDE-Funktion, so dass die aufwendige molekulare Oberfläche nicht in jedem Optimierungsschritt berechnet werden muss.

6.4.1 Die GeoHYDE-Funktion

Die Zielfunktion eines Optimierungsverfahren soll den Optimierungsprozess in die gewünschte Richtung lenken, damit das Optimum dieser Funktion gefunden werden kann. Im Optimierungskontext von HYDE soll das Minimum der HYDE-Funktion gefunden werden. Dazu müsste in jedem Optimierungsschritt die HYDE-Funktion für den Protein-Ligand-Komplex neu berechnet werden, um zu entscheiden, ob die durch den Optimierungsprozess durchgeführte Änderung am Liganden eine Minimierung der HYDE-Energie bewirkt. Während die Wasserstoffbrücken- als auch die Dehydratationsenergie für polare Atome schnell berechnet werden können, ist die Abschätzung der Dehydratationsenergie apolarer Atome durch die Generierung der molekularen Oberfläche beider Moleküle ein aufwendiger Prozess. Obwohl durch den neuen Algorithmus zur Be-

6. Berechnung der HYDE-Funktion

rechnung der Zugänglichkeitsänderung von Atomen ein Oberflächennetz für die starre Protein-Bindetasche vorberechnet werden könnte, müsste für den Liganden, an dem Konformationsänderungen vorgenommen werden, in jedem Schritt ein neues Oberflächennetz generiert werden. Daher wurde die GeoHYDE-Funktion entwickelt, in der der Term für die apolare Dehydratationsenergie durch einen Lennard-Jones-Term ersetzt wurde:

$$\text{HYDE}_{opt} = w_{HB} \cdot \Delta G_{H-bond} + w_{DP} \cdot \Delta G_{dehydration}^{hydrophilic} + w_{LP} \cdot E_{\text{Lennard-Jones}}^{\text{inter}} \quad (6.12)$$

Die GeoHYDE-Funktion besteht somit aus drei Termen, dem Wasserstoffbrücken-Term ΔG_{H-bond} und dem polaren Dehydratationsterm $\Delta G_{dehydration}^{hydrophilic}$ der HYDE-Funktion, sowie einem Lennard-Jones Term $E_{\text{Lennard-Jones}}^{\text{inter}}$. Mit dem Lennard-Jones-(12,6)-Potential ist es möglich, einerseits überlappende Atome zwischen Protein und Ligand durch den repulsiven Term zu erkennen und zu bestrafen, und andererseits die Dehydratation apolarer Atome durch den anziehenden Term zu belohnen. Um die Berechnung des Lennard-Jones-Terms zu beschleunigen, wird das Potential nicht immer über alle Atome gerechnet, sondern nur zwischen Atomen bestimmt, deren Abstand nicht größer als die Summe ihrer Van-der-Waals-Radien r_{VDW}^i und r_{VDW}^j und einer Konstanten von 2,8 Å ist:

$$E_{\text{Lennard-Jones}}^{\text{inter}} = \sum_{atoms\ i} \sum_{\substack{atoms\ j \\ distance(i,j) \leq \\ r_{VDW}^i + r_{VDW}^j + 2,8\text{Å}}} \left(\left(\frac{r_{VDW}^i + r_{VDW}^j}{distance(i,j)} \right)^{12} - \left(\frac{r_{VDW}^i + r_{VDW}^j}{distance(i,j)} \right)^6 \right) \quad (6.13)$$

Jeder Term der GeoHYDE-Funktion ist zusätzlich gewichtet (w_{HB} , w_{DP} , und w_{LP}). Dadurch ist es möglich, den Optimierungsprozess besser zu steuern, z. B. die Optimierung der Wasserstoffbrücken-Geometrie in Richtung des HYDE-Modells, durch eine höhere Gewichtung w_{HB} des Wasserstoffbrücken-Terms, zu lenken. Die unterschiedliche Gewichtung der Terme führt aber auch dazu, dass das Optimum der GeoHYDE-Funktion und der HYDE-Bewertungsfunktion möglicherweise weiter auseinander liegen. Die Resultate, die im nächsten Kapitel 7 vorgestellt werden, basieren auf folgender Gewichtung: $w_{HB} = 2$, $w_{DP} = 1$, und $w_{LP} = 1$. Somit ist die Optimierung hauptsächlich auf die Verbesserung der Wasserstoffbrücken-Geometrie eines Protein-Ligand-Komplexes ausgerichtet.

6.4.2 Kombinierte Zielfunktion des Optimierungsverfahrens

Neben der GeoHYDE-Funktion wurden noch weitere Terme in die Zielfunktion des Optimierungsverfahrens integriert, die sich auf die Abschätzung der Ligand-Konformationsenergie beschränken.

numerische Optimierung In der Zielfunktion der numerischen Optimierung wird zusätzlich ein Torsionswinkelpotential und ein intramolekularer Lennard-Jones-Term für den Liganden berechnet:

$$F_{numopt} = \text{HYDE}_{opt} + w_{TP} \cdot E_{\text{Torsion}} + w_{LPL} \cdot E_{\text{Lennard-Jones}}^{\text{intra}} \quad (6.14)$$

Dabei wird das Torsionswinkelpotential E_{Torsion} und der intramolekulare Lennard-Jones-Term $E_{\text{Lennard-Jones}}^{\text{intra}}$ in der Zielfunktion folgendermaßen gewichtet: $w_{TP} = 3$ und $w_{LPL} = 1$. In einer numerischen Optimierung wurden maximal 1000 Optimierungsschritte durchgeführt, bzw. wurde der Prozess abgebrochen, sobald die Energieänderung der Zielfunktion zwischen zwei Optimierungsschritten unterhalb eines Schwellwertes sinkt.

stochastische Optimierung Die Zielfunktion der stochastischen Optimierung berücksichtigt außer der GeoHYDE-Funktion einen Bestrafungsterm für die Überlappung von Atomen innerhalb des Liganden:

$$F_{stochopt} = \text{HYDE}_{opt} + \sum_{atoms\ i} \sum_{atoms\ j} E_{\text{AtomOverlap}}(i, j) \quad (6.15)$$

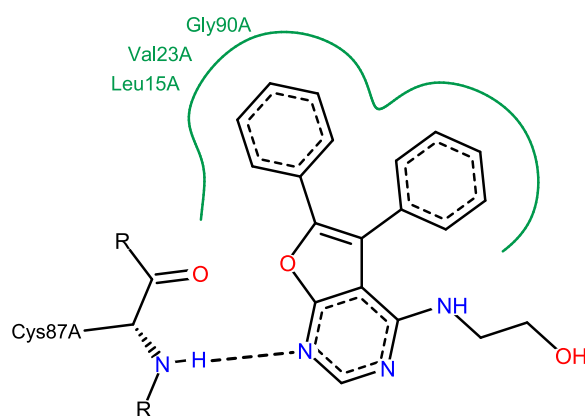
$$E_{\text{AtomOverlap}}(i, j) = \begin{cases} 0 & ,\text{if } distance(i, j) \geq (r_{VDW}^i + r_{VDW}^j) \\ \frac{r_{VDW}^i + r_{VDW}^j \cdot 10}{distance(i, j)^2} & ,\text{else} \end{cases}$$

Der Bestrafungsterm $E_{\text{AtomOverlap}}(i, j)$ wird nur für Atome i und j berechnet, die mindestens drei Bindungen voneinander entfernt und nicht im gleichen Ringsystem sind. Um eine bessere Konvergenz des stochastischen Verfahrens zu erlangen, wurden zehn Optimierungsläufe mit jeweils 1000 Optimierungsschritten durchgeführt. Das beste Resultat aus den zehn Läufen, entspricht dem Endresultat der stochastischen Optimierung.

6.5 Bewertung eines Protein-Ligand-Komplexes mit HYDE

Um die Bewertung eines Protein-Ligand-Komplexes mit HYDE zu verdeutlichen, wird im Folgenden die HYDE-Energie anhand eines Beispiels, dem Komplex der Chk1-Kinase mit einem Inhibitor (PDB: 2BRB), genau in die Einzelbeträge aufgeschlüsselt und erklärt. Weiterhin wird an diesem Beispiel auch gezeigt, wie sich der Austausch eines einzelnen Atoms auf die HYDE-Energie auswirkt.

Abbildung 6.5: Komplex der Chk1-Kinase mit einem Inhibitor (PDB: 2BRB). 2D-Darstellung der Wechselwirkungen mit Poseview [251]: Wasserstoffbrücke zwischen einem Pyrimidin-Stickstoff des Inhibitors und einem Stickstoff des Proteinrückgrates (CYS 87 A) (gestrichelte Linie) und hydrophober Kontakt zwischen den Phenyl-Ringen des Inhibitors und drei Aminosäuren (LEU 15 A, VAL 23 A, GLY 90 A) (grüne Linie).



Die Chk1-Kinase ist eine Serin/Threonin-Protein-Kinase, die eine wichtige Rolle im Zellzyklus während der Mitose übernimmt. Damit ist sie ein potentielles Zielprotein in der Krebstherapie. Abbildung 6.5 zeigt einen Komplex von Chk1 mit dem Inhibitor 2-[(5,6-Diphenylfuro[2,3-D]pyrimidin-4-yl)amino]ethanol (PDB: 2BRB). Dieser Inhibitor bindet kompetitiv mit einer mikromolaren Inhibitionskonstanten von $K_i = 13,7 \mu M$ in die ATP-Bindetasche der Kinase [199]. Dabei bildet der Pyrimidin-Stickstoff des Inhibitors eine Wasserstoffbrücke mit dem Amid-Stickstoff des Proteinrückgrates (CYS 87 A) in der Hinge-Region. Die beiden Phenyl-Ringe des Liganden werden durch hydrophobe Seitenketten des Proteins in der Bindetasche vergraben und tragen somit durch den hydrophoben Effekt zur Bindungsaffinität bei (vgl. Abbildung 6.5). Der Carbonyl-Sauerstoff von Cystein 87 A, der im Komplex mit ATP eine Wasserstoffbrücke ausbilden kann, wird durch diesen Inhibitor nicht abgesättigt.

Um diesen Komplex mit HYDE zu bewerten wurde die Proteinstruktur zuvor mit MOE (v 2009.10) [252] minimiert, da das Proteinrückgrat leicht verdreht vorlag und eine Optimierung der Inhibitor-Position in der Bindetasche durch den HYDE-Optimierungskontext nicht ausreichte, um die Wasserstoffbrücke zwischen dem Inhibitor und dem

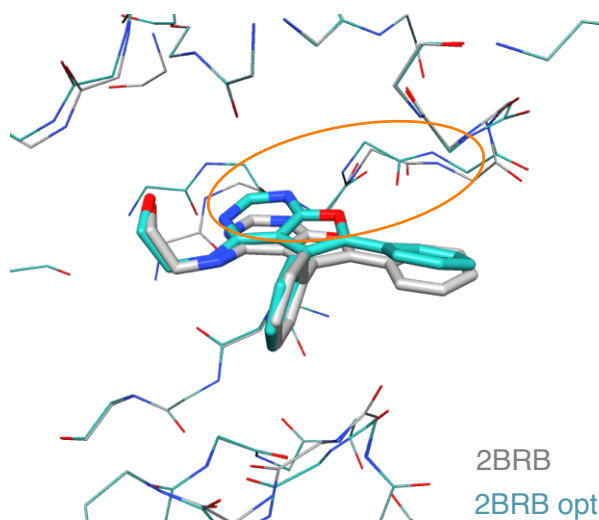


Abbildung 6.6: Optimierung des Chk1-Kinase-Komplexes (2BRB). Grau: Originalstruktur, Türkis: Optimierte Struktur. Oranger Kreis: CYS 87 A.

Proteinrückgrat in HYDE zu erkennen. Abbildung 6.6 zeigt den Komplex vor und nach der Optimierung. Die Änderung der Koordinaten auf Proteinseite ist mit 0,4 - 0,6 Å sehr klein, jedoch ändert sich dadurch der Bindungswinkel der Wasserstoffbrücke D-H...A von 150° auf 165° deutlich. Im Idealfall sollte dieser einen Wert von 180° haben. Diese voroptimierte Struktur wurde anschließend mit HYDE bewertet.

Die Bewertung mit HYDE ergibt eine Bindungsenergie von $\Delta G_{\text{HYDE}} = -18,5 \text{ kJ/mol}$, das entspricht einer mikromolaren Bindungsaffinität ($572 \mu\text{M}$). Im Vergleich zu der experimentell gemessenen Inhibitionskonstanten wird dieser Komplex von HYDE mit einer schwächeren Bindungsaffinität abgeschätzt.

6.5.1 Einzelbeiträge von Atomen

In Tabelle 6.2 sind die Einzelbeiträge der HYDE-Energie für jedes Atom des Inhibitors zusammengefasst. Die Beiträge der Proteinatome wurden auf das nächstgelegene Atom des Inhibitors übertragen. Abbildung 6.7 zeigt den Komplex im HYDE-Farbschema (vgl. Kapitel 4.1): Anhand der energetischen Beiträge aus Tabelle 6.2 werden die Inhibitoratome von grün für einen energetisch günstigen Beitrag ($\Delta G < 0$), über weiß für keinen energetisch Beitrag, bis hin zu rot, wenn ein Kontakt destabilisierend zur Gesamtenergie beiträgt ($\Delta G > 0$), eingefärbt. Die Aufschlüsselung der Benennung der Inhibitoratome nach Tabelle 6.2 ist in Abbildung 6.7 rechts dargestellt.

Der Pyrimidin-Stickstoff N1 bildet eine ideale Wasserstoffbrücke zum Amid-Stickstoff N CYS 87 A des Proteinrückgrates. Beide Wasserstoffbrücken-Partner werden durch die Komplexbildung dehydratisiert, wodurch eine ungünstige Dehydratationsenergie von $\Delta G_{\text{dehydration}} = 13,1 \text{ kJ/mol}$ entsteht. Diese wird durch die Ausbildung einer idealen

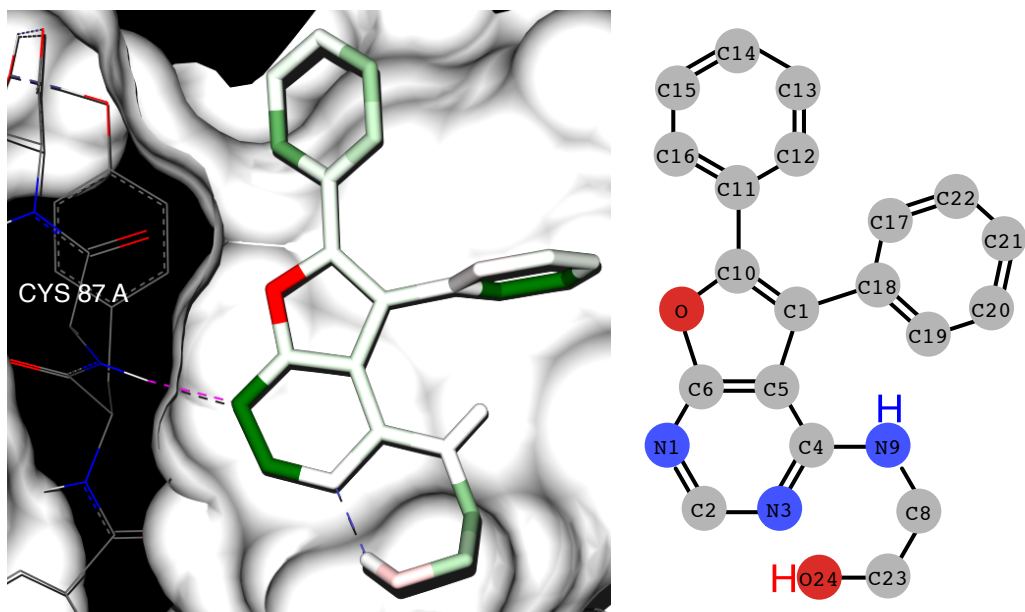


Abbildung 6.7: Bewertung eines Protein-Ligand-Komplexes (2BRB) mit HYDE. Links: HYDE-Farbcodierung der Atombeiträge: Grün: stabilisierender Beitrag zur freien Bindungsenergie ($\Delta G < 0$); Rot: destabilisierender Beitrag zur freien Bindungsenergie ($\Delta G > 0$); Weiß: kein Beitrag zur freien Bindungsenergie. Der Beitrag von Proteinatomen wurde auf das nächstgelegene Ligandatom übertragen. Rechts: Aufschlüsselung der Benennung der Ligandatome nach Tabelle 6.2.

Wasserstoffbrücke mit einer Wasserstoffbrücken-Energie von $\Delta G_{H-bond} = -15,8 \text{ kJ/mol}$ kompensiert. Zusätzlich zu diesen Energiebeiträgen wird dem Inhibitoratom N1 noch die Dehydratationsenergie umliegender hydrophober Proteinatome (CB CYS 87 A, SG CYS 87 A, CA TYR 86 A) zugeteilt, so dass sich ein Gesamtwert von $\Delta G_{HYDE} = -3,4 \text{ kJ/mol}$ für dieses Atom in Tabelle 6.2 ergibt.

Die Inhibitoratome C2, C19 und C20 hingegen leisten durch den hydrophoben Effekt einen großen Beitrag zur HYDE-Energie. Besonders Atom C19, das tief in einer hydrophoben Tasche des Proteins vergraben wird, liefert zusammen mit den umgebenden hydrophoben Resten des Protein, die ebenfalls durch die Bindung des Inhibitors dehydratisiert werden, einen energetisch günstigen Beitrag von $\Delta G_{HYDE} = -6,6 \text{ kJ/mol}$ (vgl. Tabelle 6.2).

Im Gegensatz zu diesen stabilisierenden Beiträgen wird in Abbildung 6.7 deutlich, dass das Sauerstoffatom des Inhibitors zu einer Destabilisierung des Chk1-Inhibitor-Komplex führt. Dies wird ausgelöst durch den Kontakt zweier Wasserstoffbrücken-Akzeptoren,

Ligandatome i	Ligand		Protein		ΔG_{HYDE}^i
	ΔG_{H-bond}^i	$\Delta G_{dehydration}^i$	ΔG_{H-bond}^i	$\Delta G_{dehydration}^i$	
C2	0,00	-1,53	0,00	-2,80	-4,32
C10	0,00	-0,43	0,00	0,00	-0,43
C11	0,00	-0,38	0,00	0,00	-0,38
C12	0,00	-0,99	0,00	0,00	-0,99
C13	0,00	-0,59	0,00	-0,90	-1,49
C14	0,00	-0,20	0,00	0,00	-0,20
C15	0,00	-0,32	0,00	-0,06	-0,37
C16	0,00	-1,43	0,00	-0,63	-2,05
C19	0,00	-1,80	0,00	-4,75	-6,55
C20	0,00	-1,53	0,00	-2,82	-4,35
C21	0,00	-0,18	0,00	0,00	-0,18
C22	0,00	-0,21	0,00	0,00	-0,21
N1	-8,29	7,09	-7,49	5,31	-3,38
C6	0,00	-0,46	0,00	0,00	-0,46
O	0,00	3,18	0,00	6,75	9,93
C5	0,00	-0,48	0,00	0,00	-0,48
N9	0,00	0,00	0,00	0,00	0,00
C4	0,00	-0,44	0,00	0,00	-0,44
N3	0,00	0,00	0,00	0,00	0,00
C17	0,00	-0,74	0,00	0,94	0,20
C1	0,00	-0,31	0,00	0,00	-0,31
C18	0,00	0,00	0,00	0,00	0,00
C23	0,00	-0,86	0,00	-0,30	-1,17
C8	0,00	-1,33	0,00	-0,20	-1,53
O24	0,00	2,40	0,00	-1,72	0,69
Gesamt	-8,29	-1,52	-7,49	-1,17	-18,47

Tabelle 6.2: Bewertung eines Protein-Ligand-Komplexes (2BRB) mit HYDE: Einzelbeiträge von Ligandatomen und zugeleiteter Beitrag von umliegenden Proteinatomen, aufgeschlüsselt in Wasserstoffbrücken- und Dehydratationsenergie der HYDE-Energie.

der Furan-Sauerstoff O des Inhibitors und der Carbonyl-Sauerstoff O CYS 87 A des Proteinrückgrates. Beide werden im Bindungsprozess dehydratisiert, jedoch kann keine neue, ausgleichende Wasserstoffbrücke zwischen diesen polaren Atomen gebildet werden. Dadurch entsteht ein ungünstiger Beitrag von $\Delta G_{HYDE} = 9,9 \text{ kJ/mol}$ zur Gesamtenergie. Diese destabilisierende Konstellation wird in HYDE erkannt und führt zu einer Reduzierung der Bindungsaffinität um zwei Größenordnungen.

6.5.2 Einfluss einzelner Atome auf die Bindungsaffinität

Wie im vorherigen Abschnitt beschrieben führt der Kontakt von zwei Wasserstoffbrücken-Akzeptoren zu einer reduzierten Bindungsaffinität. Um den Einfluss einzelner Atome auf die Bindungsaffinität zu testen, wurde dieser Wasserstoffbrücken-Akzeptor des Inhibi-

tors ausgetauscht.

Im ersten Fall wurde der Furan-Ring durch einen Pyrrol-Ring ersetzt, so dass ein Wasserstoffbrücken-Donor zur Verfügung steht (Abbildung 6.8 A und B). Dieser bildet eine ideale Wasserstoffbrücke mit dem Carbonyl-Sauerstoff 0 CYS 87 A des Proteins. Die beiden Wasserstoffbrücken-Partner werden zuvor durch die Bindung des Inhibitors an das Protein dehydratisiert. Die dabei entstehende, ungünstige Energie wird jedoch durch die Ausbildung einer idealen Wasserstoffbrücke kompensiert und der Gesamtbeitrag beider Atome zur freien Bindungsenergie beläuft sich auf $\Delta G_{HYDE} = -2,5 \text{ kJ/mol}$. Durch den Austausch des einen Atoms (Sauerstoff durch Stickstoff) verbessert sich die HYDE-Energie um $\Delta\Delta G_{HYDE} = -12,5 \text{ kJ/mol}$ im Vergleich zu dem vorherigen Komplex (2BRB). Die Bindungsenergie dieses Inhibitors beträgt somit $\Delta G_{HYDE} = -31 \text{ kJ/mol}$, was einer Bindungsaffinität von $3,7 \mu\text{M}$ entspricht. Dieser Wert stimmt gut mit der experimentell gemessenen Inhibitionskonstanten $K_i = 1,3 \mu\text{M}$ [199] überein.

Hypothetisch könnte der Furan-Ring auch durch einen Cyclopentadien-Ring ersetzt werden. Damit würde ein Kontakt zwischen einem polaren Atom, dem Wasserstoffbrücken-Akzeptor 0 CYS 87 A des Proteins und einem apolaren Atom, der CH_2 -Gruppe des Inhibitors entstehen (Abbildung 6.8 A und C). Dadurch ergeben sich für das polare Atom Dehydratationskosten von $\Delta G_{dehydration} = 6,7 \text{ kJ/mol}$, wohingegen eine günstige Dehydratationsenergie $\Delta G_{dehydration} = -1,7 \text{ kJ/mol}$ für das apolare Atom erzielt wird. Die Dehydratationskosten des polaren Atoms werden durch den Dehydratationsgewinn des apolaren Atoms nicht kompensiert, dennoch ist diese Konstellation nicht so ungünstig wie die ursprüngliche, der Kontakt zweier polarer Atome, die keine neue, ausgleichende Wasserstoffbrücke bilden können. Die HYDE-Energie verbessert sich in diesem Fall um $\Delta\Delta G_{HYDE} = -5 \text{ kJ/mol}$, was einer Größenordnung bei der Bindungsaffinität entspricht.

Dieses Beispiel zeigt, dass die HYDE-Bewertungsfunktion sehr sensitiv gegenüber kleinen Veränderungen im Protein-Ligand-Komplex ist und somit geeignet ist, günstige von ungünstigen Konstellationen zu unterscheiden. In vielen anderen Bewertungsfunktionen wird der Kontakt von apolaren und polaren Atomen als destabilisierend erkannt, jedoch werden die Kontakte von nicht kompatiblen polaren Atomen oft nicht behandelt (vgl. Kapitel 2.5.2), obwohl diese, wie gerade gezeigt, weitaus ungünstiger sind.

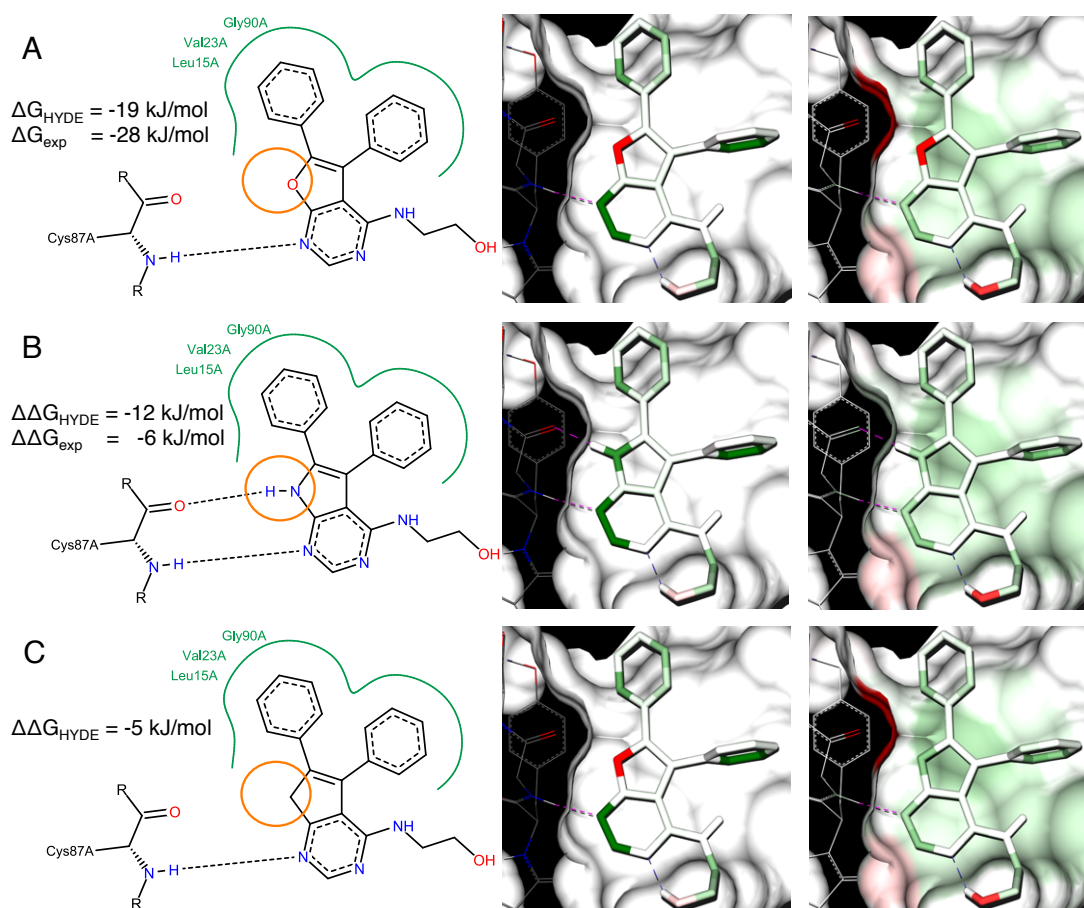


Abbildung 6.8: Einfluss einzelner Atome auf die Bindungsaffinität. Links: Schematische Darstellung von Wechselwirkungen von Protein-Ligand-Komplexen [251]. Mitte: HYDE-Farbcodierung der Komplexe, Beitrag der Proteinatome wird auf das nächstgelegene Ligandatom übertragen. Rechts: HYDE-Farbcodierung der Komplexe jedes Atom wird nach seinem eigenen Beitrag gefärbt. A) Protein-Ligand-Komplex 2BRB. HYDE-Energie und experimentell gemessene Bindungsenergie sind angegeben. B) Austausch des Sauerstoffatoms durch ein Stickstoffatom. Änderung der HYDE-Energie und der experimentell gemessene Bindungsenergie im Vergleich zu A) sind angegeben. C) Austausch des Sauerstoff-Atoms durch ein Kohlenstoff-Atom. Änderung der HYDE-Energie im Vergleich zu A) ist angegeben.

7

Resultate und Diskussion

Eine Protein-Ligand-Bewertungsfunktion wird im strukturbasierten Wirkstoffentwurf für verschiedene Zwecke eingesetzt. Zum einen sollte sie die Bindungsaffinität von Liganden an das Zielprotein vorhersagen können. Des Weiteren wird sie im molekularen Docking herangezogen, um aus der Menge der in der Protein-Bindetasche generierten Konformationen des Liganden die bioaktive auszuwählen. Im virtuellen Screening hingegen wird sie genutzt, um für Millionen von neuen Molekülen in Sekundenschnelle zu entscheiden, ob sie aktiv oder inaktiv gegenüber einem Zielprotein sind. Dies sind anspruchsvolle Aufgaben, wobei die richtige Vorhersage der Bindungsaffinität die kritischste darstellt, da sie die Voraussetzung der erfolgreichen Durchführung der beiden anderen Aufgaben bildet.

In diesem Kapitel wird die überarbeitete Fassung der HYDE-Bewertungsfunktion umfassend in allen diesen drei Aufgaben beurteilt. Weiterhin wird die Entwicklung gegenüber dem HYDE-Prototypen geprüft als auch Vergleiche zu anderen, in diesem Feld etablierten Methoden durchgeführt. Abschließend wird auf die mittlere Laufzeit der Berechnung der HYDE-Energie und der HYDE-Optimierung eingegangen. Als Einstieg werden zuerst die herangezogenen Qualitätsmaße und Bewertungskriterien kurz diskutiert.

7.1 Qualitätsmaße und Bewertungskriterien

Um eine Entscheidung über die Vorhersage-Qualität der Bewertungsfunktion treffen zu können, werden je nach Anwendung verschiedene Qualitätsmaße herangezogen. Im Folgenden werden die Bewertungskriterien für die drei, oben genannten Anwendungsgebiete einer Bewertungsfunktion zusammengefasst, bevor im zweiten Teil des Kapitels auf sie zurückgegriffen wird.

7.1.1 Bindungsaffinitätsvorhersage

Die Qualität der Vorhersage von Bindungsaffinitäten durch eine Bewertungsfunktion wird anhand verschiedener Protein-Ligand-Komplexe mit experimentell bestimmten Werten ermittelt. Dabei wird ein Vergleich von vorhergesagten Werten mit den experimentell bestimmten Werten durchgeführt, indem diese Werte miteinander korreliert werden. Dazu wird der Pearson-Korrelationskoeffizient R_p über alle Komplexe i berechnet:

$$R_p = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 (y_i - \bar{y})^2}} \quad (7.1)$$

Dabei sind x_i die experimentell bestimmten Werten und y_i die durch die Bewertungsfunktion vorhergesagten Werte. Weiterhin ist \bar{x} der Mittelwert der experimentell bestimmten Werte und \bar{y} der Mittelwert der vorhergesagten Werte. Der Wertebereich des Pearson-Korrelationskoeffizienten liegt zwischen -1 und 1 , wobei man von einem linearen Zusammenhang der Werte ausgehen kann, sobald $|R_p| > 0,5$ ist.

Oftmals werden die experimentell bestimmten und die vorhergesagten Werte auch durch eine einfache Regressionsanalyse in einen linearen Zusammenhang gesetzt:

$$y_i = ax_i + b \quad (7.2)$$

Hierbei wird häufig auch das Bestimmtheitsmaß R^2 berechnet. Bei einer einfachen Regression entspricht das Bestimmtheitsmaß dem Quadrat des Pearson-Korrelationskoeffizienten. Daher liegt hier ab einem Wert von $R^2 > 0,25$ ein linearer Zusammenhang zwischen den Werten vor.

Zusätzlich wurde auch die Standardabweichung SD der linearen Korrelation berechnet:

$$SD = \sqrt{\sum_{i=1}^n [y_i - ax_i + b]^2 / (n - 2)} \quad (7.3)$$

Als weiteres Maß wurde in der Analyse der Ergebnisse der Affinitätsvorhersage der Spearman-Korrelationskoeffizient genutzt:

$$R_s = 1 - \frac{6 \cdot \sum_{i=1}^n (rg(x_i) - rg(y_i))^2}{n \cdot (n^2 - 1)} \quad (7.4)$$

Dabei ist $rg(x_i)$ der Rang von x_i in einer nach experimenteller Bindungsaffinität sortierten Liste und entsprechend $rg(y_i)$ der Rang von y_i in einer nach vorhergesagter Bindungsaffinität sortierten Liste. Der Wertebereich des Spearman-Korrelationskoeffizienten liegt ebenfalls zwischen -1 und 1 . Dieser Korrelationskoeffizient ist jedoch im Vergleich zum Pearson-Korrelationskoeffizient robuster gegenüber Ausreißern. Weiterhin steht nicht die Vorhersage des genauen Wertes, sondern die richtige Klassifikation der Komplexe - in energetisch stabile und weniger stabile - im Vordergrund dieses Qualitätsmaßes.

Diese Maße zur Beurteilung der Affinitätsvorhersage-Qualität einer Bewertungsfunktion ermöglichen vor allem den Vergleich verschiedener Methoden. Ein Problem bei der Nutzung dieser Maße stellt jedoch der nicht berücksichtigte, experimentelle Fehler der zugrunde liegenden Daten dar. Die experimentelle Bestimmung von Affinitätsdaten hängt stark von den experimentellen Bedingungen ab, wie z. B. pH-Wert, Temperatur oder Menge und Art des organischen Lösungsmittels. Schon kleinere Veränderungen dieser Bedingungen sorgen für eine Abweichung der Bindungsaffinität eines Liganden an ein Protein. Beim strukturbasierten Wirkstoffentwurf ist die Kristallstruktur eines Protein-Ligand-Komplexes die Grundlage der Bewertung. Diese wird allerdings meist unter anderen Bedingungen erzeugt als die, die bei der Affinitätsmessung vorlagen. Diese kritischen Punkte erschweren somit die Beurteilung der Vorhersage-Qualität einer Bewertungsfunktion. Daher sollte bei der Auswahl der experimentell bestimmten Daten, die zur Beurteilung nötig sind, vor allem darauf geachtet werden, konsistente Daten, wenn möglich aus einer Messreihe, zu verwenden. Die Klassifikation der Liganden in schwache, mittlere und starke Binder wird dadurch zuverlässiger und der Einfluss des experimentellen Fehlers wird verringert.

7.1.2 Vorhersage von Protein-Ligand-Komplexen

Im strukturbasierten Wirkstoffentwurf werden mithilfe von molekularem Docking Bindungsmodi eines Liganden in der Bindetasche des Zielproteins generiert. In dieser Anwendung dient die Bewertungsfunktion dazu, die Konformation des Liganden zu bestimmen, die unter allen durch das Docking generierten Konformationen der bioaktiven am ähnlichsten ist. Um beurteilen zu können, wie gut die Bewertungsfunktion diese

Aufgabe erfüllt, werden die Koordinaten der Ligandatome von Kristallstrukturen (bioaktive Konformation) mit den durch das Docking erzeugten Koordinaten verglichen. Dazu wird die Wurzel der mittleren quadratischen Abweichung (RMSD) zwischen den Kristall-Koordinaten c_K und den, durch das Docking vorhergesagten, Koordinaten c_D über alle Schweratome i des Liganden berechnet:

$$RMSD = \sqrt{\frac{\sum_{i=1}^n (c_K^i - c_D^i)^2}{n}} \quad (7.5)$$

Üblicherweise wird dabei eine Konformation mit einem $RMSD \leq 2\text{\AA}$ noch als ähnlich zur Kristall-Konformation angesehen [253]. Eine Bewertungsfunktion, die die Ligandkonformation mit dem kleinsten RMSD zur Kristallstruktur auf den ersten Rang aller durch das Docking erzeugten Konformationen einordnet, würde ein perfektes Resultat liefern. Dieses Bewertungskriterium wird in den meisten Validierungsstudien abgeschwächt, indem nicht nur die am besten bewertete Konformation des Liganden betrachtet wird, sondern die x besten Konformationen, wobei oftmals $x = 20$ gewählt wird.

Die Verwendung des RMSD als Qualitätsmaß weist ebenfalls Schwächen auf [254], wie z. B. die Abhängigkeit von der Molekülgröße. Eine weitere Schwierigkeit, die durch den Vergleich von generierten Komplexen zu Kristall-Komplexen auftritt, ist der Einfluss der Qualität der zugrunde liegenden Kristallstrukturen. Diese enthalten in der Regel eine mittlere Unsicherheit von bis zu $0,5\text{\AA}$ in den Atomkoordinaten. Damit ist eine Ligandkonformation, deren RMSD kleiner als $0,5\text{\AA}$ ist, genauer als es der experimentelle Fehler eigentlich zulässt [255]. Im Extremfall existiert für manche Teile des Kristall-Komplexes keine Elektronendichte, wodurch für diese Teilbereiche keine Aussage über die Richtigkeit der Vorhersage getroffen werden kann, daher sollte bei der Auswahl der Strukturen, die zur dieser Art der Validierung herangezogen werden, immer die experimentelle Elektronendichte überprüft werden.

7.1.3 Virtuelles Screening

Im virtuellen Screening ist die Bewertungsfunktion für die Klassifizierung von aktiven und inaktiven Molekülen eines bestimmten Zielproteins zuständig. Um die Fähigkeit der Bewertungsfunktion zur Trennung von aktiven und inaktiven Molekülen ermitteln zu können, werden bekannte aktive Moleküle eines Zielproteins und bekannte inaktive Moleküle oder, wenn diese Information nicht vorliegt, unwahrscheinliche Binder in das Zielprotein gedockt. Diese Komplexe werden anschließend mit der Bewertungsfunktion klassifiziert. Dabei sollten die inaktiven Moleküle eine gewisse Ähnlichkeit zu den bekannten aktiven Molekülen haben, um ein reales Szenario zu schaffen. Alle klassifizierten

Moleküle werden entsprechend ihrer Bewertung sortiert, so dass Moleküle, für die eine hohe Bindungsaffinität abgeschätzt wurde, zuerst in die Liste eingeordnet werden. Anhand dieses Resultats, kann die Sensitivität SE und die Spezifität SP der Vorhersage berechnet werden [256]:

$$SE = \frac{N_{\text{gefundene Aktive}}}{N_{\text{Aktive gesamt}}} = \frac{TP}{TP + FN} \quad (7.6)$$

$$SP = \frac{N_{\text{verworfen Inaktive}}}{N_{\text{Inaktive gesamt}}} = \frac{TN}{TN + FP} \quad (7.7)$$

Dabei ist TP die Anzahl der richtig vorhergesagten aktiven Moleküle (*true positives*) und FN die Anzahl der falsch als inaktiv klassifizierten Moleküle (*false negatives*). Dagegen ist TN die Anzahl der korrekt bestimmten inaktiven Moleküle (*true negatives*) und FP die Anzahl der als aktiv klassifizierten inaktiven Moleküle (*false positives*). Die Sensitivität beschreibt den Anteil von korrekt klassifizierten Aktiven zu den aktiven Molekülen, die insgesamt im Datensatz vorhanden waren. Die Spezifität hingegen repräsentiert den Anteil an korrekt klassifizierten inaktiven Molekülen im Verhältnis zu allen betrachteten Inaktiven. Wird die Sensitivität gegen die inverse Spezifität ($1-SP$) für jedes klassifizierte Molekül des Datensatzes aufgetragen, so erhält man eine ROC-Kurve (*Receiver Operator Characteristic*) [256]. Abbildung 7.1 stellt Beispiele für verschiedene ROC-Kurven dar. Als Qualitätsmaß wird für jede ROC-Kurve der Flächeninhalt der Fläche unter Kurve AUC (*Area Under the Curve*) berechnet [257].

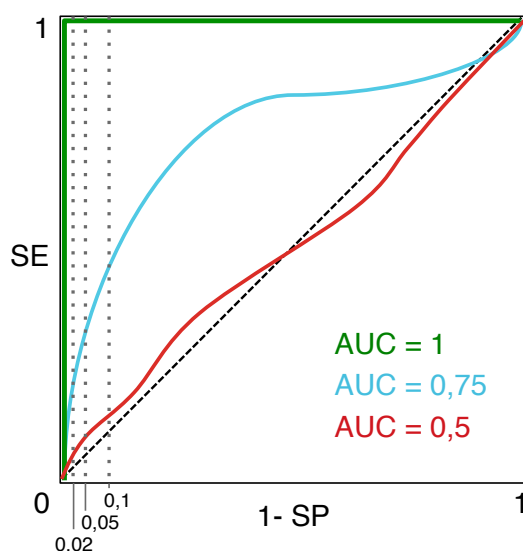


Abbildung 7.1: Verschiedene ROC-Kurven. Grün: ideales Resultat. ROC-Kurve mit einem AUC von 1. Blau: Mittelmäßiges Resultat. ROC-Kurve mit einem AUC von 0,75. Rot: schlechtes Resultat. ROC-Kurve mit einem AUC von 0,5. Diagonale: Entspricht einer zufälligen Auswahl. Grau gestrichelte Linie: Falsch-Positiv-Rate bei 2 %, 5 % und 10 % gefundenen Falsch-Positiven, entspricht dem y-Wert der ROC-Kurve bei diesem Schwellwert.

Eine Bewertungsfunktion liefert eine ideale Trennung von aktiven und inaktiven Molekülen eines bestimmten Zielproteins, wenn die ROC-Kurve einen AUC-Wert von $AUC = 1$ besitzt. Das bedeutet, alle aktiven werden besser gegenüber den inaktiven Molekülen bewertet, werden also in einer sortierten Liste der klassifizierten Moleküle höher eingeordnet als die inaktiven Moleküle (vgl. Abbildung 7.1 grüne Kurve). Erzielt die Klassifizierung durch eine Bewertungsfunktion einen AUC-Wert von $AUC = 0,5$ oder weniger, ist das Resultat nicht besser als das, das durch eine zufällige Klassifizierung der Moleküle in aktiv und inaktiv erreicht werden könnte (vgl. Abbildung 7.1 rote Kurve). Ein gutes Resultat liegt daher deutlich über einem AUC-Wert von 0,5, eher in einem Bereich zwischen 0,8 und 1.

Ein weiteres Qualitätsmaß zur Einschätzung der frühen Anreicherung von aktiven Verbindungen ist der ROC@(x %)-Wert. Da in einem virtuellen Screening teilweise Millionen von Molekülen bewertet werden, sind vor allem die Moleküle von Interesse, die als starke Binder durch die Bewertungsfunktion eingestuft werden, d. h. am Anfang der sortierten Liste stehen, da nur diese später experimentell validiert werden sollen. Daher wird für bestimmte Schwellwerte x der Falsch-Positiv-Rate ($1 - SP$) die Trefferquote (SE) ermittelt (vgl. Abbildung 7.1). Das bedeutet, ein ROC@(1 %)-Wert gibt den Anteil an korrekt klassifizierten Aktiven zu den tatsächlich aktiven Molekülen an, wenn bis dahin 1 % der inaktiven Molekülen falsch als aktiv eingestuft wurde. Je höher dieser Anteil ist, desto besser ist die Fähigkeit der Bewertungsfunktion aktive von inaktiven Molekülen zu trennen. Interessante Schwellwerte, um die frühe Anreicherung von aktiven Molekülen zu beurteilen, liegen bei einer Falsch-Positiv-Rate von 0,5 %, 1 %, 2% und vielleicht noch 5 % (bei einer Million Moleküle wären das die ersten 50 000 Moleküle) [255].

Die Validierung einer Bewertungsfunktion im virtuellen Screening kann durch die Zusammenstellung der eingesetzten Datensätze beeinflusst werden, d. h. werden die Inaktiven so ausgewählt, dass sie mit sehr hoher Wahrscheinlichkeit nicht an ein bestimmtes Zielprotein binden, so ist der molekulare Unterschied zu den Aktiven so groß, dass die Klassifizierung von aktiven und inaktiven Molekülen keine Schwierigkeit darstellt. Im anderen Fall sind sich aktive und inaktive Moleküle zu ähnlich und die Inaktiven wurden nicht experimentell validiert, so kann über die richtige Klassifizierung keine Aussage getroffen werden. Ein weiteres Problem stellt die Flexibilität der Protein-Struktur dar. Die Bindetasche eines Proteins kann sich den verschiedenen Molekülen, die mit dem Protein wechselwirken, anpassen. Im virtuellen Screening, wird jedoch in den meisten Fällen, je nach Docking Programm, von einer starren Bindetasche ausgegangen, so dass bei der Validierung nur aktive Moleküle korrekt klassifiziert werden können, die in diese Konformation des Proteins binden.

7.2 Vorhersage von Bindungsaffinitäten

Die Validierung der HYDE-Bewertungsfunktion bezüglich der Vorhersage von Bindungsaffinitäten ist in zwei Teile unterteilt. Wie bereits im vorherigen Abschnitt 7.1.1 angemerkt, ist die Auswahl valider und vor allem konsistenter experimenteller Daten als Grundlage der Validierung von großer Wichtigkeit. Daher wurden im ersten Teil zwei kleinere Molekülserien ausgesucht, die jeweils aus einer Messreihe und dem gleichen Labor stammen. Im zweiten Teil der Validierung hingegen wurde ein Benchmark-Datensatz genutzt, um einerseits die überarbeitete Fassung der HYDE-Bewertungsfunktion mit dem HYDE-Prototypen und anderen bereits etablierten Methoden zu vergleichen und andererseits, um HYDE in der Affinitätsvorhersage in einem größerem Maßstab zu testen.

7.2.1 Affinitätsvorhersage von Molekülserien

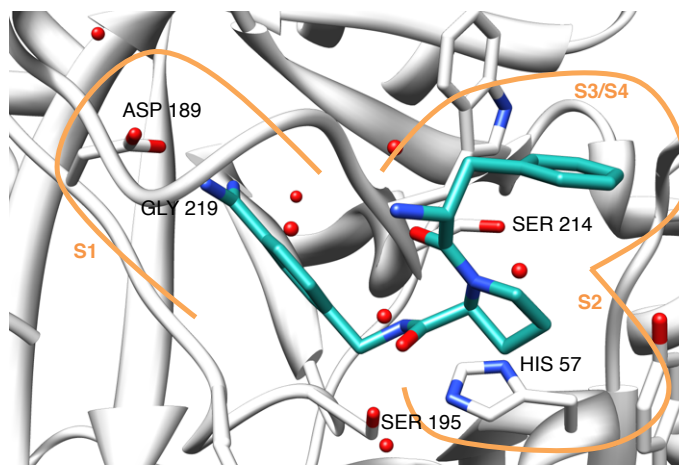
Um eine optimale Beurteilung in Bezug auf die Qualität der Affinitätsvorhersage der HYDE-Bewertungsfunktion zu erlauben, wurden zwei Serien von Molekülen, die aus einer Messreihe stammen, ausgewählt. Der Vorteil dabei ist, dass der Einfluss des experimentellen Fehlers minimiert wird. Weiterhin wurde auch die Qualität der Kristallstrukturen mithilfe der gemessenen Elektronendichte überprüft, um weitere Fehlerquellen, die nicht auf HYDE zurückzuführen sind, auszuschließen. Die Kristallisation der Komplexe und die Affinitätsmessungen wurden jeweils im gleichen Labor durchgeführt. Im Folgenden wird zuerst eine Serien von Thrombin-Inhibitoren untersucht, wobei der energetische Beitrag einzelner Atome diskutiert wird. Die zweite Serie beinhaltet verschiedene p38-MAP-Kinase-Inhibitoren. Hierbei wird die Entwicklung, ausgehend von einer Leitstruktur bis hin zu einem klinisch geprüften Wirkstoff, mit HYDE bewertet.

Thrombin

Thrombin ist eine Serinprotease, d. h. ein Enzym das Proteine mithilfe von einem Serin-Rest in seiner katalytischen Bindetasche an der Peptidbindung aufspaltet. Es befindet sich im Blut als inaktive Vorstufe Prothrombin und spielt eine entscheidende Rolle bei der Blutgerinnung. Nach seiner Aktivierung katalysiert es die Spaltung von Fibrinogen zu Fibrin, das polymerisiert und so zur Blutgerinnung führt. Damit ist Thrombin spezifischer als andere Serinproteasen, die im Verdauungstrakt viele verschiedene Arten von Polypeptiden aufspalten. Bei Thrombin wird in der Katalyse Serin 195 durch das benachbarte Histidin 57 polarisiert, um eine nucleophilen Angriff auf die Carbonyl-Gruppe der zu spaltenden Peptidbindung des Substrats zu ermöglichen. Dieser führt schließlich

zur Aufspaltung dieser Bindung. Um die Blutgerinnung zu unterbinden, wird Thrombin bevorzugt als Zielprotein gewählt. Abbildung 7.2 zeigt die Bindetasche von Thrombin mit einem Inhibitor (PDB: 2ZDA). Die katalytisch wichtigen Aminosäuren sind markiert und die verschiedenen Spezifitätstaschen von Thrombin sind dargestellt.

Abbildung 7.2: Bindetasche von Thrombin mit einem Inhibitor (PDB: 2ZDA). Wichtige Aminosäuren der Bindetasche sind beschriftet. Die S1-S4 Spezifitätstaschen von Thrombin sind in orange markiert.



Die Kristallstrukturen von fünf Thrombin-Komplexen mit verschiedenen Inhibitoren (PDB: 2ZFF, 2ZDV, 2ZC9, 2ZDA) [258] wurden mit HYDE bewertet. Diese Inhibitoren besitzen alle das gleiche Grundgerüst (vgl. Abbildung 7.3), nur der Teil, der die S1-Tasche von Thrombin besetzt, unterscheidet sich. Vier der Inhibitoren binden mit einem hydrophoben Phenylring, der mit Wasserstoff, Methyl, Fluor oder Chlor meta-substituiert ist, in die S1-Tasche. Im Gegensatz dazu bildet der fünfte Inhibitor über ein Benzamidin eine Salzbrücke zu Aspartat 189 in der S1-Tasche. Alle fünf Inhibitoren sind in Abbildung 7.3 im HYDE-Farbschema zusammen mit der berechneten Bindungsenergie ΔG_{HYDE} dargestellt. Die Bindungsenergie ΔG_{exp} dieser Komplexe wurde durch Isotherme Titrationskalorimetrie (ITC) auch experimentell bestimmt [258] und ist ebenfalls in Abbildung 7.3 zu finden. Die HYDE-Energie der Komplexe zeigt eine gute Übereinstimmung mit der experimentell gemessenen Bindungsaffinität. Der Korrelationskoeffizient zwischen den beiden Werten beträgt $R_p = 0,97$.

Betrachtet man die Komplexe im HYDE-Farbschema (Abbildung 7.3), so fällt auf, dass zwei Atome, ein Kohlenstoff des Phenylrings, der in die S1-Tasche bindet, und ein Amid-Stickstoff in den Komplexen der ersten vier Inhibitoren immer destabilisierend zur Gesamtenergie beitragen. Anhand eines Komplexes (2ZC9) wird im Folgenden genauer untersucht, welche Ursache diese destabilisierenden Beiträge haben und wie genau sich die HYDE-Energie bei diesen Atomen zusammensetzt. Außerdem wird der Einfluss des Substituenten, in diesem Fall Chlor, auf die Gesamtenergie bestimmt.

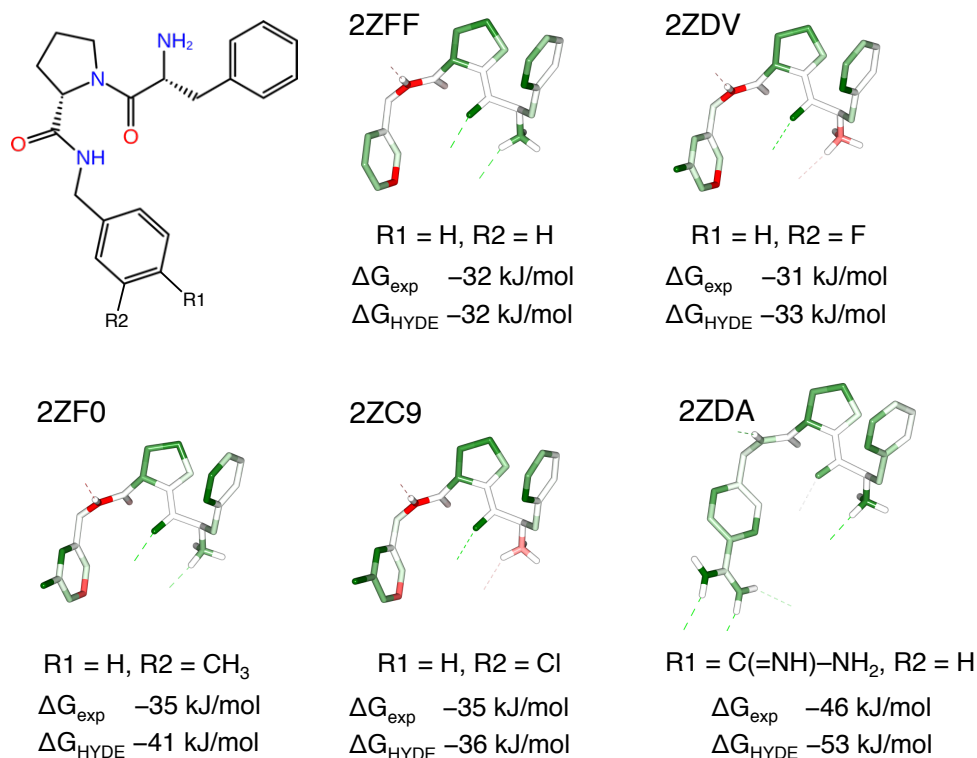


Abbildung 7.3: Serien von Thrombin Inhibitoren. Oben links: Grundgerüst der Inhibitoren mit zwei Substitutionsstellen R1 und R2. Alle Moleküle sind im HDYE-Farbschema dargestellt. Für jedes Molekül sind die PDB ID (oben links) und die Substituenten der beiden Substitutionsstellen R1 und R2 (unter dem Molekül), sowie die experimentell bestimmte (ΔG_{exp}) und die von HYDE vorhergesagte (ΔG_{HYDE}) Bindungsenergie angegeben. Bild in Anlehnung an [208]

In dem Thrombin-Komplex mit dem Inhibitor D-phenylalanyl-N-(3-chlorobenzyl)-L-prolinamid (2ZC9) bindet der D-phenylalanyl-N-Rest, wie auch bei den anderen vier Inhibitoren, in die S3/S4 Tasche von Thrombin, während sich der L-prolinamid-Rest in der S2-Spezifitätstasche befindet. In der S1-Tasche wird der 3-Chlorobenzyl-Ring über den hydrophoben Effekt stabilisiert (vgl. Abbildung 7.4 links).

Die Ursache des destabilisierenden Beitrags des Amid-Stickstoffs des Inhibitors ist in Abbildung 7.4 A dargestellt. Dieser Stickstoff bildet eine Wasserstoffbrücke mit dem Carbonyl-Sauerstoff von SER214 aus. Diese Wasserstoffbrücke weist eine größere Abweichung von der idealen Wasserstoffbrücken-Geometrie auf. Der Bindungswinkel bezüglich der Ebene der freien Elektronenpaare des Carbonyls beträgt 49° . In HYDE wird für diesen Winkel eine Abweichung bis zu 20° toleriert, somit liegt die aktuelle

Abweichung deutlich darüber. Dadurch ergibt sich für den Faktor f_{dev} , der die Qualität der Wasserstoffbrücken-Geometrie in der Berechnung der HYDE-Wasserstoffbrücken-Energie berücksichtigt, ein Wert von 0,5. Die Energie dieser Wasserstoffbrücke wird demzufolge auch um die Hälfte reduziert ($\Delta G_{H-bond} = -8,2 \text{ kJ/mol}$). Dies hat jedoch keinen Einfluss auf die Dehydratationskosten der beiden Wasserstoffbrücken-Partner ($\Delta G_{dehydration} = 14,6 \text{ kJ/mol}$), die durch diese geometrisch schlecht ausgebildete Wasserstoffbrücke nicht kompensiert werden können. Diese Konstellation trägt somit mit $\Delta G_{HYDE} = 6,4 \text{ kJ/mol}$ ungünstig zur Bindungsenergie bei.

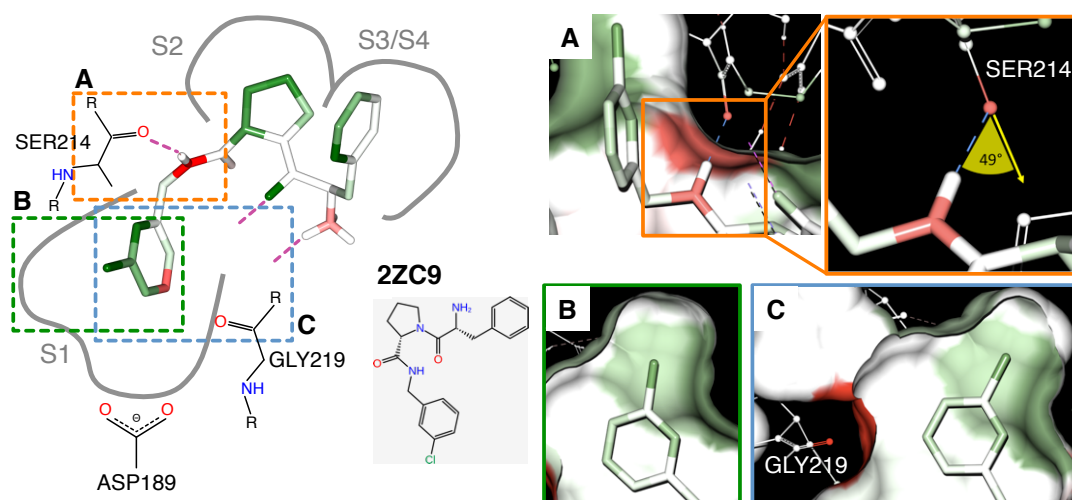


Abbildung 7.4: Thrombin-Komplex mit dem Inhibitor D-phenylalanyl-N-(3-chlorobenzyl)-L-prolinamid (2ZC9). Links: Schematische Darstellung des Inhibitors (HYDE-Farbschema) in der Bindetasche von Thrombin. Rechts: Drei Detailaufnahmen von Wechselwirkungen des Inhibitors in der Bindetasche. In dieser Darstellung sind die HYDE-Atombeiträge der Proteinatome nicht auf die Inhibitoratome übertragen. Jedes Atom wurde entsprechend seines eigenen Beitrags eingefärbt. A) Wasserstoffbrücke zwischen dem Amid-Stickstoff des Inhibitors und dem Carbonyl-Sauerstoff von SER214. Die Abweichung von der idealen Geometrie ist dargestellt. B) Hydrophober Effekt des Chloratoms in der kleinen hydrophoben Sub-Tasche der S1-Spezifitätstasche. C) Desolvatisierung von Carbonyl GLY219 durch den Phenyl-Ring. Bild aus [208]

Ein Kohlenstoffatom des 3-Chlorobenzyl-Rings liefert nach Abbildung 7.4 (links), ebenfalls einen ungünstigen Beitrag zur Gesamtenergie. In diesem Fall wird die destabilisierende Energie durch den Kontakt eines apolaren Atoms, dem Kohlenstoff, und eines polaren Atoms, dem Carbonyl-Sauerstoff von GLY219, ausgelöst (vgl. Abbildung 7.4 C). Der 3-Chlorobenzyl-Ring in der S1-Tasche von Thrombin desolvatisiert den

Carbonyl-Sauerstoff von GLY219. Dies führt zu Dehydratationskosten in Höhe von $\Delta G_{dehydration} = 6,4 \text{ kJ/mol}$ in HYDE. Diese können nicht durch die günstige Dehydratationsenergie $\Delta G_{dehydration} = -1,7 \text{ kJ/mol}$ des apolaren Kohlenstoff kompensiert werden.

Im Gegensatz zu diesen destabilisierenden Beiträgen liefert das Chloratom des 3-Chlorobenzyl-Rings durch den hydrophoben Effekt einen günstigen Beitrag zur Bindungsenergie. Das Chloratom füllt die hydrophobe Sub-Tasche der S1-Spezifitätstasche perfekt aus und verdrängt so ein ungünstiges Wassermolekül, das diese Tasche besetzt. Durch diese Konstellation wird die Sub-Tasche als auch das Chloratom selbst vollständig desolvatisiert, wodurch eine Dehydratationsenergie von $\Delta G_{dehydration} = -3 \text{ kJ/mol}$ für das Chloratom und von $\Delta G_{dehydration} = -3,2 \text{ kJ/mol}$ für die Atome der hydrophoben Tasche entsteht (vgl. Abbildung 7.4 B).

Dies zeigt, dass hydrophobe Reste in der S1-Spezifitätstasche einen guten Beitrag zur Bindungsenergie liefern können und somit eine Alternative zu geladenen Benzamidin-Resten darstellen, wobei diese durch die Wasserstoffbrücken-Bindungen zur Carboxylat-Gruppe von ASP189 und zum Carbonyl von GLY219 nicht zu unabgesättigten polaren Atomen in der S1-Tasche von Thrombin führen und damit deutlich stärker zur Bindungsenergie beitragen können.

p38-MAP-Kinase

Die p38-mitogenaktivierte-Proteinkinase ist eine Serin/Threonin Kinase, d. h. sie phosphoryliert andere Proteine spezifisch an Serinen oder Threoninen. Diese Kinasen sind ein wichtiger Bestandteil in Signaltransduktionswegen, die an der Regulation der Zelldifferenzierung, des Zellwachstums und der Apoptose beteiligt sind. Dadurch sind sie ein potentiell Zielprotein bei der Behandlung von chronisch entzündlichen Erkrankungen als auch in der Krebstherapie. Die p38-MAP-Kinase benötigt wie andere Kinasen auch für die Phosphorylierung die Bindung zu einem Adenosintriphosphat (ATP). Dieses wird während der Phosphorylierung durch die Kinase in ein Adenosindiphosphat (ADP) umgesetzt und die abgespaltene Phosphat-Gruppe wird auf das zu phosphorylierende Protein übertragen. Die Protein-Bindetasche für ATP, auch Hinge-Region genannt, ist in allen Kinasen stark konserviert und eignet sich daher gut als Bindestelle für Inhibitoren. Eine gute Selektivität für eine bestimmte Kinase durch diese Inhibitoren (Typ-I-Kinase-Inhibitoren) zu erreichen, ist nur eingeschränkt über die nahegelegenen hydrophoben Bereiche der Protein-Bindetasche möglich. Eine weitere Möglichkeit Kinasen zu inhibieren bietet sich über eine andere, überwiegend hydrophobe Tasche, die spezifischer für bestimmte Kinasen ist. Diese hydrophobe Tasche entsteht durch eine

Konformationsänderung des DFG-Motifs (ASP-PHE-GLY), d. h. sie ist vor der Aktivierung bestimmter Kinasen vorhanden (DFG-out-Konformation). Durch die Aktivierung einer Kinase ändert sich die Konformation des DFG-Motifs in die DFG-in-Konformation und diese hydrophobe Tasche wird verschlossen, jedoch wird die Bindetasche für ATP geöffnet. Typ-II-Kinase-Inhibitoren binden spezifisch, vor der Aktivierung einer Kinase in die hydrophobe Tasche und verhindern damit das Öffnen der ATP-Tasche.

Im Folgenden wird der Entwicklungsprozess einer Leitstruktur von Typ-II-Kinase-Inhibitoren zu einem klinisch getesteten Wirkstoff BIRB796 mit HYDE untersucht. Dazu wurden die zwei in der PDB vorhandenen Kristallkomplexe, einer mit gebundener Leitstruktur (PDB: 1KV1) und einer komplexiert mit dem Wirkstoff (PDB: 1KV2), genutzt. Beide Komplexe sind in Abbildung 7.5 dargestellt. Fünf verschiedene Zwischenstufen (Molekül 25, 46, 58, 65 und 75 aus [259]) in diesem Entwicklungsprozess, die größere Unterschiede in ihrer Bindungsaffinität aufwiesen, wurden ausgewählt und anhand der Kristallstruktur des Wirkstoffs modelliert. Die experimentell bestimmten Dissoziationskonstanten (K_d -Werte) zu diesen Strukturen wurden ebenfalls aus [259] entnommen und mithilfe von Gleichung 2.2 in die freie Bindungsenergie (ΔG_{exp} -Werte) umgerechnet.

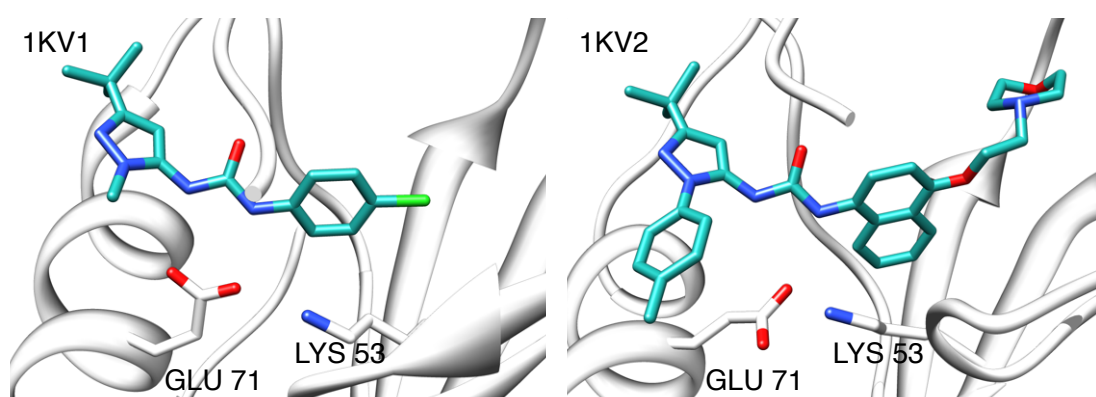


Abbildung 7.5: Kristallkomplexe der p38-MAP-Kinase mit zwei verschiedenen Inhibitoren. Links: Kristallstruktur von p38-MAP-Kinase mit der Leitstruktur in der hydrophoben Bindetasche der DFG-out-Konformation (1KV1). Die Carboxylat-Gruppe von GLU71 bildet eine bidentate Wasserstoffbrücke zur Harnstoff-Gruppe des Inhibitors. Rechts: Kristallstruktur von p38-MAP-Kinase mit dem Wirkstoff BIRB796 in der hydrophoben Bindetasche der DFG-out-Konformation (1KV2). Konformationsänderung der Carboxylat-Gruppe von GLU71, sie bildet hier eine bidentate Wasserstoffbrücke zum Amin von LYS53.

Alle sieben Protein-Inhibitor-Komplexe wurden mit HYDE bewertet (vgl. Abbildung 7.6). Die von HYDE berechneten Bindungsenergien zeigen eine gute Übereinstimmung mit den experimentell bestimmten Bindungsaffinitäten. Der Korrelationskoeffizient liegt

bei $R_p = 0,88$. In Abbildung 7.6 ist der Entwicklungsprozess ausgehend von der Leitstruktur über die fünf Zwischenstufen bis hin zu BIRB796 dargestellt. Die entsprechenden Modifikationen der Moleküle in der Leitstrukturoptimierung sind jeweils hervorgehoben und alle Moleküle sind zusätzlich im HYDE-Farbschema abgebildet. Die Änderung der experimentellen $\Delta\Delta G_{exp}$ und der von HYDE berechneten $\Delta\Delta G_{HYDE}$ Bindungsenergie ist ebenfalls angegeben. Für alle Moleküle, außer bei Verbindung 46, stimmt die Änderung der HYDE-Energie, hervorgerufen durch die jeweilige strukturelle Modifikation, gut mit der Änderung der experimentellen Bindungsaffinität überein.

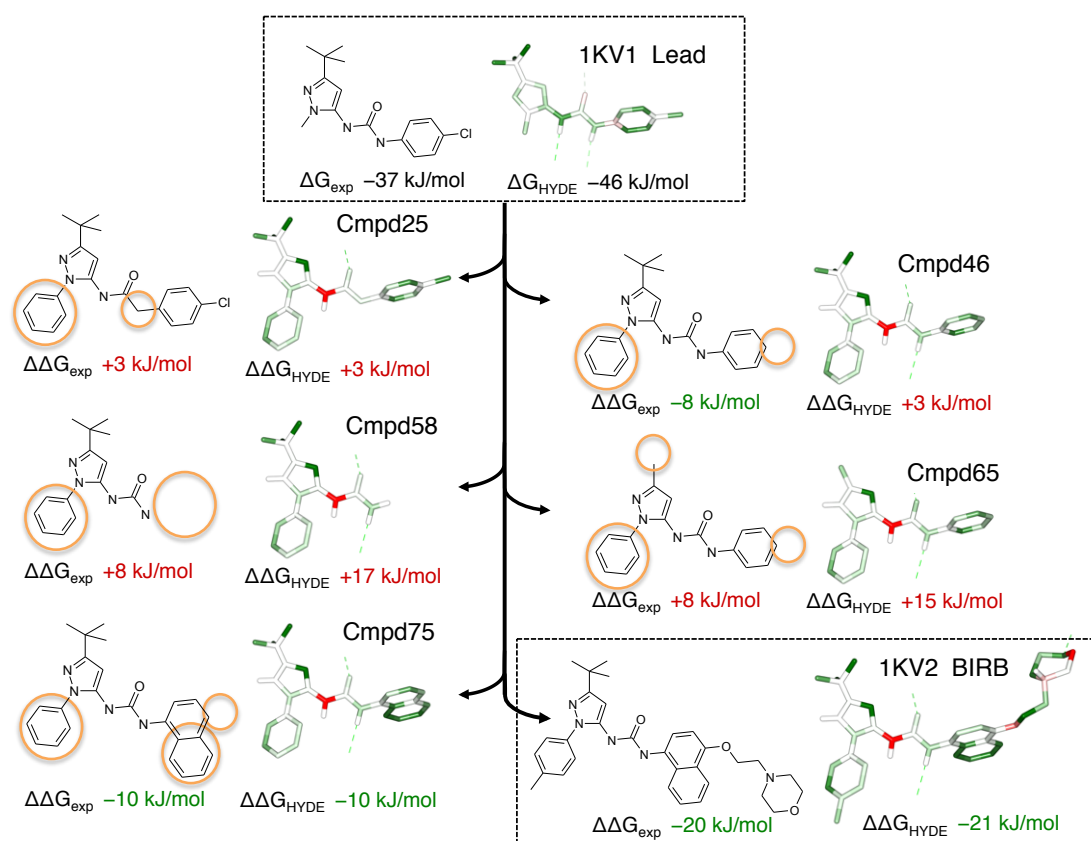


Abbildung 7.6: Leitstrukturoptimierung eines Typ-II-Kinase-Inhibitors der p38-MAP-Kinase. Ausgangspunkt der Entwicklung ist die Leitstruktur 1KV1 (oben Mitte). Die finale Struktur BIRB796 (1KV2) ist unten links gezeigt. Strukturelle Veränderungen der Zwischenstufen im Optimierungsprozess bzgl. der Leitstruktur sind orange markiert. Die Änderung der experimentellen $\Delta\Delta G_{exp}$ und der von HYDE berechneten $\Delta\Delta G_{HYDE}$ Bindungsenergie ist jeweils unter den Molekülen dargestellt. Bild aus [208]

Die Modifikationen, die an Molekül 46 im Vergleich zur Leitstruktur durchgeführt wurden, resultieren in einer Verbesserung der experimentellen Bindungsaffinität um $\Delta\Delta G_{exp} = -8 \text{ kJ/mol}$. Im Gegensatz dazu wird von HYDE eine Verschlechterung der Bindungsenergie um $\Delta\Delta G_{HYDE} = 3 \text{ kJ/mol}$ zur Leitstruktur vorhergesagt. Einer der Harnstoff-Stickstoffe wird in der HYDE-Färbung in rot dargestellt, trägt also ungünstig zur Bindungsenergie bei (vgl. Abbildung 7.6). Dieser Stickstoff wird im Komplex dehydratisiert, kann aber keine ausgleichende Wasserstoffbrücke zum Protein ausbilden. In der Leitstruktur hingegen sind beide Harnstoff-Stickstoffe über eine bidentate Wasserstoffbrücke mit der Carboxylat-Gruppe von GLU71 abgesättigt (vgl. Abbildung 7.5 links). Die Einführung des Phenyl-Rings an dem Pyrazol-Stickstoff N2 bewirkt eine Konformationsänderung der Seitenkette von GLU71, so dass der hydrophobe Phenyl-Ring einen engen Kontakt zum apolaren Alkyl-Rest der GLU71 Seitenkette aufbauen kann. Dies hat jedoch die Auflösung der bidentaten Wasserstoffbrücke zwischen den Harnstoff-Stickstoffen des Inhibitors und der Carboxylat-Gruppe von GLU71 zur Folge. Letztere bildet nur noch eine monodentate Wasserstoffbrücke zu einem der Harnstoff-Stickstoffe, kann jedoch jetzt eine zusätzliche bidentate Wasserstoffbrücke zu einem benachbarten Lysin (LYS53) bilden [259] (vgl. Abbildung 7.5 rechts). Die Dehydratation dieses Harnstoff-Stickstoffs wird in HYDE mit einem ungünstigen Beitrag von $\Delta G_{dehydration} = 6,7 \text{ kJ/mol}$ zur Bindungsenergie bewertet. Im Fall von Molekül 46 kann der durch den hydrophoben Effekt hervorgerufene stabilisierende Beitrag des eingeführten Phenyl-Rings diese Dehydratationskosten als auch den Verlust an Bindungsenergie durch die Substitution des Chloratoms durch ein Wasserstoffatom am anderen Phenyl-Ring nicht ausgleichen. Ein Grund dafür könnte sein, dass die Dehydratationskosten einer Harnstoff-Gruppe zur Zeit in HYDE überbewertet werden, da die plogP-Inkrementen von HYDE nur auf sehr einfachen Molekülen parametrisiert wurden.

7.2.2 Affinitätsvorhersage von Benchmark-Daten

Um die HYDE-Bewertungsfunktion in größerem Maßstab zu testen und mit anderen bereits etablierten Bewertungsfunktionen zu vergleichen, wurde der PDBbind 2007 core-set Benchmark-Datensatz [210, 211] ausgewählt. Auf diesem Datensatz wurde die Qualität der Vorhersage von Bindungsaffinitäten von 16 verschiedenen Bewertungsfunktionen durch Cheng *et al.* [7] getestet. Der Datensatz besteht aus 195 Protein-Ligand-Komplexen, für die experimentell bestimmte Inhibitionskonstanten (K_i -Wert) oder Dissoziationskonstanten (K_d -Wert) vorliegen. Weiterhin wurden nur Komplexe in den Datensatz aufgenommen, deren Kristallstruktur eine Auflösung von wenigstens $2,5 \text{ \AA}$ besitzt.

Vor der Bewertung der Komplexe mit HYDE wurden alle Komplexe mit der LeadIT-Software (v2.1.1) [260] vorbereitet: Die Bindetasche wurde mithilfe des Liganden bestimmt, indem alle Aminosäuren, Kofaktoren und Ionen ausgewählt wurden, die innerhalb eines Radius von 6,5 Å um die Schweratome des Liganden lagen. Kristallwasser wurden in der Bewertung aus Gründen der Vergleichbarkeit nicht hinzugenommen. Anschließend wurde das Wasserstoffbrücken-Netzwerk der Bindetasche und des Liganden grob mit ProToss voroptimiert (vgl. Kapitel 4.2.1). Die Metallkoordinierung wurde automatisch zugewiesen, jedoch mussten manche Koordinationsgeometrien nach der visuellen Untersuchung der Komplexe manuell korrigiert werden. Alle Liganden des Datensatzes wurden zusätzlich mit NAOMI [176] prozessiert, um fehlerhafte Moleküle zu vermeiden. Alle 195 Protein-Ligand-Komplexe wurden mit der überarbeiteten Fassung der HYDE-Funktion (HYDE2.0) und dem HYDE-Prototypen (HYDE1.0) bewertet. Zusätzlich wurde der Ligand für alle Komplexe im HYDE-Optimierungsverfahren, einmal durch die numerische Optimierung und einmal durch die stochastische Optimierung, geometrisch in der Bindetasche optimiert (vgl. Kapitel 4.2.2 und 6.4). Weiterhin wurden noch einzelne Terme und Term-Kombinationen der HYDE-Bewertungsfunktion getestet. Alle Ergebnisse sowie die Resultate anderer Bewertungsfunktionen sind in Tabelle 7.1 aufgeführt.

Im Vergleich zum HYDE-Prototypen (Tabelle 7.1 HYDE1.0) verbessert sich die Korrelation der experimentell bestimmten Affinität zur vorhergesagten Bindungsenergie durch die überarbeitete Fassung der HYDE-Funktion (Tabelle 7.1 HYDE2.0) und es ergibt sich ein Korrelationskoeffizient von $R_p = 0,323$. Die numerische Optimierung der Komplexe ändert die Korrelation nicht, durch die stochastische Optimierung wird der Korrelationskoeffizient leicht auf $R_p = 0,345$ verbessert (Tabelle 7.1 HYDE2.0, Spalte 7). Vergleicht man HYDE jedoch zu den anderen Bewertungsfunktionen, so erzielt HYDE nur einen Platz im unteren Drittel, in der nach dem Pearson-Korrelationskoeffizienten sortierten Tabelle 7.1. Einige dieser Bewertungsfunktionen, wie PHOENIX oder X-Score, wurden auf, zum PDBbind 2007 coreset ähnlichen, Datensätzen kalibriert, was ihre bessere Vorhersagequalität auf diesem Datensatz gegenüber HYDE erklären könnte. Andererseits kann ein niedrigerer Korrelationskoeffizient als der, der durch die Korrelation der Schweratome des Liganden mit der experimentell bestimmten Affinität erzielt werden kann (Tabelle 7.1 *by NHA*), nicht allein durch die Kalibrierung auf ähnlichen Daten verursacht werden.

Wie bereits erwähnt, ist HYDE sehr sensitiv bezüglich der Geometrie der Komplexe und nimmt so schon kleine Ungenauigkeiten in den Kristallstrukturen deutlich wahr. Infolgedessen wurde der PDBbind 2007 coreset Datensatz genauer untersucht und mithilfe

7. Resultate und Diskussion

Bewertungsfunktion	Original-Komplexstruktur				optimierte Komplexstruktur			
	N	R_p	SD	R_s	N	R_p	SD	R_s
X-Score::HMScore	195	0,644	1,83	0,705	195	0,649	1,82	0,701
HYDE2.0::HbondsHydrophobic	195	0,620	1,89	0,669	-	-	-	-
PHOENIX	194	0,616	2,16	0,644	-	-	-	-
HYDE2.0::Hydrophobic	195	0,602	1,91	0,642	-	-	-	-
DrugScoreCSD	195	0,569	1,96	0,627	195	0,589	1,93	0,649
SYBYL::ChemScore	195	0,555	1,98	0,585	194	0,622	1,87	0,668
DS::PLP1	195	0,545	2,00	0,588	194	0,529	2,03	0,569
GOLD::ASP	193	0,534	2,02	0,577	194	0,518	2,04	0,558
SYBYL::G-Score	195	0,492	2,08	0,536	195	0,522	2,03	0,579
DS::LUDI3	195	0,487	2,09	0,478	194	0,477	2,10	0,478
DS::LigScore2	193	0,464	2,12	0,507	194	0,479	2,10	0,505
GlideScore-XP	178	0,457	2,14	0,435	187	0,555	2,01	0,556
DS::PMF	193	0,445	2,14	0,448	194	0,471	2,11	0,482
GOLD::ChemScore	178	0,441	2,15	0,452	186	0,528	2,05	0,553
by NHA	195	0,431	2,15	0,517	195	0,431	2,15	0,517
SYBYL::D-Score	195	0,392	2,19	0,447	195	0,388	2,20	0,443
HYDE2.0	161	0,323	2,26	0,34	189	0,345	2,24	0,378
DS::Jain	189	0,316	2,24	0,346	190	0,339	2,26	0,362
GOLD::GoldScore	169	0,295	2,29	0,322	188	0,329	2,26	0,386
HYDE1.0	150	0,271	2,30	0,321	-	-	-	-
SYBYL::PMF-Score	190	0,268	2,29	0,273	180	0,235	2,31	0,235
SYBYL::F-Score	185	0,216	2,35	0,243	181	0,238	2,31	0,208

Tabelle 7.1: Korrelation zwischen experimentell bestimmten Bindungskonstanten und vorhergesagter Affinität für den PDBbind 2007 coreset Datensatz. Die Bewertungsfunktionen sind entsprechend ihres Pearson-Korrelationskoeffizienten R_p geordnet, der für die Original-Komplexstruktur berechnet wurde. Zusätzlich sind die Resultate für die optimierte Komplexstruktur angegeben (falls vorhanden). Die Resultate sind für alle Bewertungsfunktionen, außer HYDE und PHOENIX [128], aus [7] entnommen. N = Anzahl der Komplexe, denen durch die Bewertungsfunktion eine günstige Bindungsenergie zugewiesen wurde. R_p = Pearson-Korrelationskoeffizient. SD = Standardabweichung der linearen Korrelation (in $\log K_d$ Einheiten). R_s = Spearman-Korrelationskoeffizient. NHA = Anzahl der Schwertome des Liganden. Tabelle aus [208]

der verfügbaren Elektronendichtekarten wurden die Protein-Ligand-Komplexe nach bestimmten Kriterien bewertet. Die Bewertungskriterien und die finale Klassifikation der untersuchten Kristallstrukturen sind aus [208] entnommen und werden in Anhang C.2 aufgeführt. Viele der untersuchten Komplexe weisen Mängel bezüglich der gemessenen Elektronendichte und der dann zugewiesenen Temperaturfaktoren auf. Ebenfalls liegen

in einigen der Komplexe Kristallpackungseffekte vor. Abbildung 7.7 zeigt vier Komplexe des PDBbind 2007 coresets und beispielhaft die strukturellen Mängel der Komplexe, die in diesem Datensatz gefunden wurden. In allen vier Komplexen ist eine fehlende Elektronendichte für größere Teile des Liganden zu erkennen. Damit sind die für den Liganden abgeleiteten Koordinaten höchst fragwürdig und eigentlich nicht zur Validierung anderer Methoden geeignet. Ebenso wie fehlende Elektronendichte sind auch Kristallstrukturen, die alternative Konformationen der Bindetasche und des Liganden aufweisen, was auf eine hohe Flexibilität dieser Regionen schließen lässt, eher ungeeignet, um sie als Bewertungsgrundlage von Methoden zu nehmen, die den Protein-Ligand-Komplex meist als starre Momentaufnahme behandeln (vgl. Abbildung 7.7 C und D).

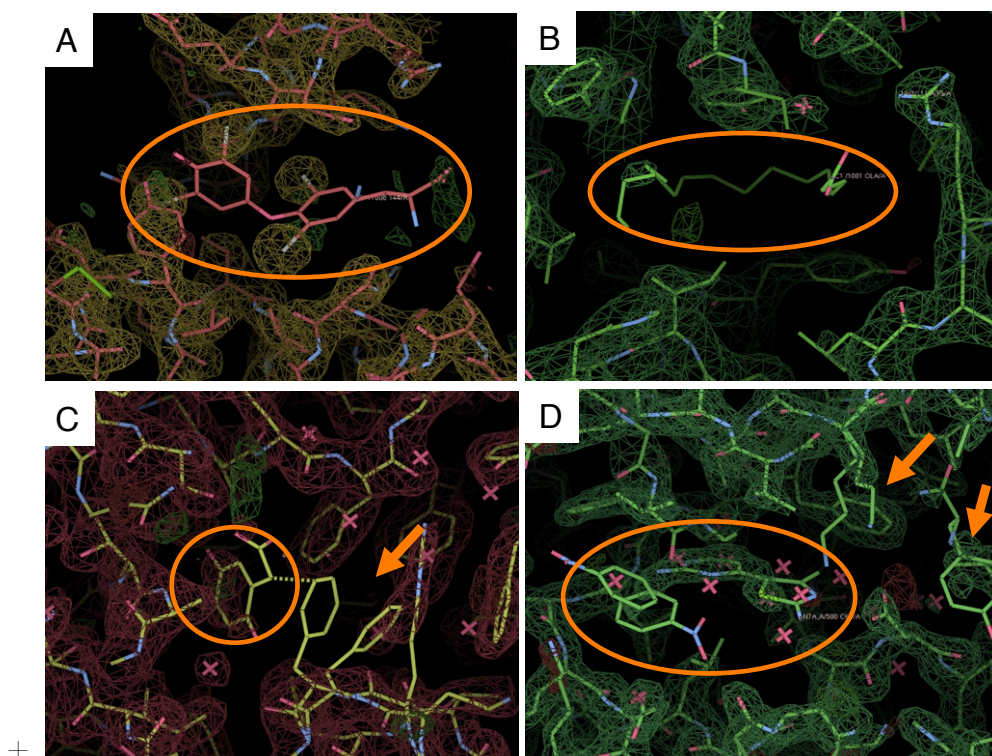


Abbildung 7.7: Beispiele der Elektronendichte von Komplexen des PDBbind 2007 coreset. Liganden sind durch orange Kreise markiert. A) Fast keine Elektronendichte für den Liganden ist vorhanden (1HK4). B) Keine Elektronendichte des Liganden (1GNI). C) Fehlende Elektronendichte für den Liganden und die Bindetasche, alternative Konformationen von PHE146 und ARG145 sind vorhanden (1AJP). D) Fehlende Elektronendichte für den Liganden und die Bindetasche, der Ligand liegt fragmentiert vor und hat eine alternative Konformation für den Nitrobenzol-Ring, Wasser wurden innerhalb des Liganden modelliert, eine alternative Konformation für LYS33 ist zu finden (1PXO). Bild aus [208]

Ein ähnliches Resultat hatte eine Studie von Sondergaard *et al.* [261], die den PDBbind 2007 refined Datensatz (aus diesem wurde der PDBbind 2007 coreset abgeleitet) nach Struktur-Artefakten untersuchten. Sie fanden, dass 36 % der Protein-Ligand-Komplexe des Datensatzes durch Kristallpackungseffekte beeinflusst werden und diese wiederum die Vorhersagequalität einer Bewertungsfunktion senken. Eine Ursache für die niedrige Korrelation zwischen der von HYDE vorhergesagten und der experimentell bestimmten Affinität könnte daher die Bestrafung von Wasserstoffbrücken, die zu stark von der idealen Wasserstoffbrücken-Geometrien abweichen, sein. Die sehr spezifischen Wasserstoffbrücken-Geometrie Parameter in HYDE, die von Kristallstrukturen kleiner Moleküle abgeleitet wurden (vgl. Kapitel 4.1), können im Fall von sehr ungenauen Kristallstrukturen eher ein Rauschen verursachen.

Um diese Annahme zu überprüfen und um besser zu verstehen, welche Terme der HYDE-Bewertungsfunktion am meisten durch die strukturelle Qualität von Kristallkomplexen beeinflusst werden, wurden verschiedenen Komponenten der HYDE-Funktion separat getestet: der hydrophobe Effekt, die Wasserstoffbrücken-Energie und die polare Dehydratation. Dazu wurde zuerst der Betrag des hydrophoben Effektes zur Bindungsenergie in HYDE berechnet und mit der experimentell bestimmten Affinität verglichen. Dies führte zu einer enormen Verbesserung des Korrelationskoeffizienten auf $R_p = 0,602$ (vgl. Tabelle 7.1 HYDE::Hydrophobic). Wird die Wasserstoffbrücken-Energie hinzugenommen, d. h. 17 % der Energie des HYDE-Wasserstoffbrücken-Terms, da dies den Gewinn einer intermolekularen Wasserstoffbrücke im Vergleich zu den Dehydratationskosten der polaren Wechselwirkungspartner darstellt (vgl. Kapitel 6.1.2), so erhöht sich der Korrelationskoeffizient wiederum auf $R_p = 0,620$ (vgl. Tabelle 7.1 HYDE::HbondsHydrophobic). Dies bestätigt, dass die Bestrafung von geometrisch schlechten Wasserstoffbrücken durch die polaren Dehydratationskosten die höchste Sensitivität gegenüber strukturellen Ungenauigkeiten in den Protein-Ligand-Komplexen aufweist.

Im Vergleich zu den anderen Bewertungsfunktionen, die in Tabelle 7.1 aufgeführt sind, ist die Korrelation von HYDE verhältnismäßig gut, wenn diese beiden, eher einfach zu berechnenden Terme genutzt werden. Im Gegensatz zur PHOENIX-Bewertungsfunktion, die eine gute Korrelation durch die Verwendung von 42 Deskriptoren erreichte, oder zur GlideScore-XP-Funktion, die aus 10 Termen besteht und damit eine Korrelation von $R_p = 0,457$ lieferte, kann durch den hydrophoben Effekt und die Wasserstoffbrücken-Energie von HYDE eine bessere Vorhersagequalität erzielt werden. Zudem wurde HYDE nicht mit experimentell gemessenen Affinitäten oder auf Kristallstrukturen von Protein-Ligand-Komplexen kalibriert, wie fast alle der Bewertungsfunktionen in Tabelle 7.1. Trotz dieses Resultats auf dem PDBbind 2007 coreset Datensatzes, ist die Beachtung

der polaren Dehydratationsenergie grundlegend bei der Unterscheidung von aktiven und inaktiven Molekülen und führt zur Reduzierung der Falsch-Positiven, was in Abschnitt 7.4 gezeigt wird.

7.3 Bestimmung der bioaktiven Ligandkonformation

Im Kontext von molekularem Docking wurde die HYDE-Bewertungsfunktion auf ihre Fähigkeit hin überprüft, die bioaktive Konformation unter allen durch das Docking generierten Ligandkonformationen zu bestimmen, wobei davon ausgegangen wird, dass diese in der Kristallstruktur eines Protein-Ligand-Komplexes vorliegt (= Redocking). Zu dieser Validierung wurde der Astex Diverse Set Datensatz [262] ausgewählt, der aus 85 hochaufgelösten Kristallstrukturen medizinisch-chemisch relevanter Protein-Ligand-Komplexe besteht. Im ersten Teil der Validierung wurde der original Astex Diverse Set verwendet, auf dem bereits zwei andere Docking-Methoden, GOLD [87, 88, 122] und PLANTS [149, 263, 264], getestet wurden, um so das von HYDE erzielte Resultat mit den Ergebnissen dieser beiden etablierten Methoden zu vergleichen. Im zweiten Teil wurde eine überarbeitete Version des Astex Diverse Set genommen, die als Benchmark Datensatz für ein Docking und Scoring Symposium veröffentlicht wurde, das auf dem 241. ACS Spring Meeting 2011 in Anaheim stattfand¹. Für diese Version, wurden die Koordinaten aller Komplexe des Datensatzes anhand ihrer Elektronendichte neu verfeinert. Mithilfe dieses Datensatzes wurde eine detaillierte Analyse der Redocking-Ergebnisse hinsichtlich der zugrunde liegenden Strukturen durchgeführt.

7.3.1 Vergleich der Resultate von HYDE mit anderen Methoden

Für alle 85 Komplexe des original Astex Diverse Set wurden Bindungsmodi des Liganden mithilfe der FlexX-Docking-Methode [80] erzeugt. Diese wurden im Anschluss mit HYDE nachbewertet, um die bioaktive Konformation des Liganden zu identifizieren. Dazu wurden die PDB-Strukturen aller Komplexe mit der LeadIT-Software (v2.1.1) [260] automatisch aufbereitet: Für die Bindetasche des Proteins wurden anhand des gegebenen Referenzliganden (Kristallkoordinaten) alle Aminosäuren, Kofaktoren und Ionen in einem Radius von 6,5 Å um die Schweratome dieses Liganden ausgewählt. Die automatisch zugewiesene Metallkoordination wurde durch eine visuelle Inspektion der Komplexe auf Korrektheit überprüft und gegebenenfalls angepasst. Das Wasserstoffbrücken-

¹<http://abstracts.acs.org/chem/241nm/program/divisionindex.php>

Division: COMP, Session: Docking and Scoring: A Review of Docking Programs

Netzwerk der Komplexe wurde mit ProToss (vgl. Kapitel 4.2.1) voroptimiert. Die Referenzliganden wurden mit NAOMI [176] aus dem *mol*-Format in das *mol2*-Format konvertiert. Konformationen der Referenzliganden wurden aus dem SMILES-Format [265] mit CORINA (v3.48) [246, 247] generiert und als Startkonformation für das molekulare Docking genutzt. Für jeden der 85 Liganden wurden 200 verschiedene Konformation (= Docking-Posen) in der Bindetasche des dazugehörigen Proteins mithilfe der neusten Version des FlexX-Docking-Algorithmus, der ebenfalls in der LeadIT-Software (v2.1.1.) integriert ist, erzeugt. Alle diese Konformationen wurden anschließend mit dem HYDE-Optimierverfahren vorbereitet und danach durch HYDE bewertet. Tabelle 7.2 zeigt die Resultate von HYDE im Vergleich zu FlexX (mit nativer Bewertungsfunktion FlexScore), GOLD und PLANTS.

		RMSD [Å] ≤		
		1,0	1,5	2,0
HYDE	Top 1	53 (2)	67 (2)	76 (1)
	Top 20	80 (1)	90 (1)	94 (1)
FlexX	Top 1	35	56	69
	Top 20	73	85	87
GOLD::GoldScore	Ø 25 beste	-	-	74 - 81
GOLD::ChemScore	Ø 25 beste	-	-	68 - 79
PLANTS::CHEMPLP	Ø 25 beste	-	-	76 - 89
PLANTS::PLP	Ø 25 beste	-	-	75 - 86

Tabelle 7.2: Redocking Resultate auf dem original Astex Diverse Set. Die Resultate sind in Prozent zur Anzahl der Komplexe angegeben. Die Ergebnisse von HYDE wurden aufgrund der stochastischen Optimierung über drei Iterationen gemittelt, die Standardabweichung ist in Klammern angegeben. Die Resultate von GOLD und PLANTS stammen aus Korb *et al.* [149]. Tabelle aus [208]

Im Redocking werden durch HYDE, mithilfe der stochastischen Optimierung, gute Ergebnisse erzielt, die vergleichbar mit den Ergebnissen der bereits etablierten Methoden FlexX, GOLD und PLANTS sind (vgl. Tabelle 7.2). Hierbei sei jedoch angemerkt, dass HYDE im Vergleich zu den Bewertungsfunktionen der anderen beiden Docking-Methoden, GOLD und PLANTS, nicht auf diesem Datensatz kalibriert wurde. In 76 % der Fälle wird von HYDE eine Ligandkonformation, die einen RMSD von unter 2 Å zu den Kristallkoordinaten des Liganden aufweist, als beste ausgewählt (Tabelle 7.2 Top 1). Diese Erfolgsrate erhöht sich auf 94 %, sobald die 20, von HYDE am besten bewerteten, Docking-Posen berücksichtigt werden (Tabelle 7.2 Top 20). Vergleicht man das

mit dem Resultat, das mit der nativen Bewertungsfunktion von FlexX erreicht wird, so liefert die Nachbewertung der Docking-Posen durch HYDE eine Steigerung der Erfolgsrate um jeweils sieben Prozentpunkte, wenn nur die beste bzw. die besten 20 Posen betrachtet werden. Noch deutlicher wird die Verbesserung der Ergebnisse durch HYDE, wenn nur die Ligandkonformationen einbezogen werden, die einen RMSD von unter 1 Å zu den Kristallkoordinaten besitzen. Hier wird eine Steigerung der Erfolgsrate um 18 Prozentpunkte erreicht bei Betrachtung der am besten bewerteten Pose.

7.3.2 Detaillierte Analyse der HYDE-Redocking-Ergebnisse

Anhand einer überarbeiteten Version des Astex Diverse Set wurde ein weiteres Redocking und eine anschließende Nachbewertung der Posen mit HYDE durchgeführt. In Hinblick auf den Einfluss der zugrunde liegenden Strukturen auf den Erfolg eines Redockings wurden die Ergebnisse im Detail untersucht. Diese Analyse wurde aufgrund des Docking und Scoring Symposiums, das auf dem 241. ACS Spring Meeting 2011 in Anaheim stattfand, durchgeführt [266]. Im Unterschied zum vorherigen Experiment wurde bei multimeren Proteinkomplexen nicht nur eine Bindetasche zum Redocking definiert, sondern alle Taschen wurden ausgewählt, in denen der Referenzligand in der Kristallstruktur gebunden vorlag. Dadurch ergaben sich 151 Bindetaschen für den Astex Diverse Set. Vier dieser Bindetaschen (Komplex PDB: 1TZ8) wurden in der späteren Analyse ausgeschlossen, da die Bindetaschen nicht vollständig in der asymmetrischen Einheit vorlagen. Mithilfe der LeadIT-Software (v2.0.2) wurde für alle gegebenen Referenzliganden jeweils die Bindetasche (Aminosäuren, Kofaktoren und Ionen) mit einem Radius von 6,5 Å um die Schweratome des Liganden bestimmt. Die Koordinationsgeometrie von Metallionen wurde in manchen Fällen manuell angepasst (vgl. [266]). In allen Protein-Ligand-Komplexen wurde das Wasserstoffbrücken-Netzwerk mit ProToss optimiert. Für jede Proteinstruktur wurde eine vorgegebene Startkonformation des Liganden mithilfe der in LeadIT integrierten Version von FlexX in alle vordefinierten Bindetaschen der Proteinstruktur gedockt. Dabei wurden für jede Bindetasche bis zu 400 Posen des Liganden mit zwei verschiedenen Platzierungsalgorithmen von FlexX erzeugt. Anschließend wurden alle Posen durch den HYDE-Optimierungskontext (numerische Optimierung, maximal 1000 Optimierungsschritte) vorbereitet, um danach mit der HYDE-Bewertungsfunktion neu eingeordnet (= Reranking) zu werden. Für alle optimierten Posen wurde der RMSD zu den Kristallkoordinaten des Liganden mit der MOE Software (v2010.10) [252] berechnet, die Ergebnisse werden in Tabelle 7.3 (optimierte Strukturen) zusammengefasst. Tabelle 7.3 zeigt auch die Resultate für die Original-Strukturen, d. h. in diesen Komplexen wurde vorher nicht das Wasserstoffbrücken-

Netzwerk mit ProToss optimiert, auch die Metallkoordinationsgeometrien wurden nicht korrigiert.

Rang	RMSD [\AA] \leq							
	Original-Strukturen				optimierte Strukturen			
	0,5	1	1,5	2	0,5	1	1,5	2
1	10	35	52	66	16	46	67	75
≤ 2	13	42	62	72	21	56	74	80
≤ 3	14	46	65	77	26	64	79	84
≤ 4	15	48	66	78	27	65	80	85
≤ 5	18	51	69	79	28	68	80	86
≤ 20	33	64	81	87	42	76	86	91
≤ 32	39	69	85	89	46	77	88	93

Tabelle 7.3: Redocking Ergebnisse auf allen Protein-Bindetaschen des Astex Diverse Set. Alle Resultate sind in Prozent zur Anzahl der Bindetaschen angeben: Originale Strukturen basierend auf 151 Bindetaschen, optimierte Strukturen basierend auf 147 Bindetaschen. Tabelle in Anlehnung an [266]

Die Redocking-Resultate in Tabelle 7.3, die auf der überarbeiteten Version des Astex Diverse Set basieren und alle Bindetaschen einer Proteinstruktur berücksichtigen, sind sehr ähnlich zu den in Tabelle 7.2 gezeigten Resultaten, die auf dem original Astex Diverse Set erzeugt wurden. In 75 % der Bindetaschen hat die von HYDE am besten bewertete Ligandkonformation einen RMSD von unter 2 \AA zu den Kristallkoordinaten des Liganden (Tabelle 7.3 optimierte Strukturen). Betrachtet man dagegen das Resultat basierend auf den Original-Strukturen in Tabelle 7.3, so wird durch eine fehlende Optimierung des Wasserstoffbrücken-Netzwerkes ein Verlust um neun Prozentpunkte hervorgerufen. Damit wird deutlich, wie wichtig ein vollständig ausgebildetes Wasserstoffbrücken-Netzwerk in einem Protein-Ligand-Komplex für eine zuverlässige Bewertung durch HYDE ist. Vor allem, das identifizieren von Posen, die nur wenig von der bioaktiven Konformation abweichen (RMSD $\leq 1 \text{\AA}$), wird im Fall eines nicht optimalen Wasserstoffbrücken-Netzwerkes erschwert.

In Tabelle 7.4 sind die Redocking-Resultate für die Original-Strukturen und die optimierten Strukturen statistisch zusammengefasst. Unter Berücksichtigung der jeweils am besten bewerteten Pose ergibt sich für die optimierten Strukturen ein mittlerer RMSD von 1,89 \AA und ein Median RMSD von 1,06 \AA über alle Bindetaschen des Astex Diverse Set.

Rang	Original-Strukturen		optimierte Strukturen	
	1	≤ 32	1	≤ 32
Mittelwert	2,15	0,91	1,89	0,78
Std. Abw.	1,97	0,85	2,06	0,74
Median	1,45	0,59	1,06	0,53
Minimum	0,27	0,19	0,22	0,16
Maximum	9,57	6,49	9,65	5,80

Tabelle 7.4: RMSD Statistik für Original-Strukturen und optimierte Strukturen. Alle Resultate sind in Å angegeben. Tabelle aus [266]

Im Folgenden werden die Resultate, basierend auf den optimierten Strukturen, genauer untersucht. In Abbildung 7.8 A ist die Verteilung der RMSD Werte der jeweils am besten bewerteten Pose (Top 1) für alle Bindetaschen der 85 Strukturen des Astex Diverse Set dargestellt. Zwischen den verschiedenen Taschen einer Proteinstruktur treten zum Teil erhebliche Schwankungen des RMSD Wertes der am besten bewerteten Pose auf. Diese werden z. B. durch leicht unterschiedliche Bindungsmodi in den verschiedenen Taschen (1SQ5, 1HVY), durch unterschiedliche Seitenkettenkonformationen (1IA1) oder die Hinzunahme von nicht-biologisch relevanten Taschen (112S), verursacht. Weiterhin haben nicht alle Taschen die gleiche Qualität in Bezug auf ihre gemessene Elektronendichte, z. B. in 1J3J ist eine der Bindetaschen hochaufgelöst (Proteinkette A), während man in der zweiten fehlende Elektronendichte und erhöhte Temperaturfaktoren vorfindet (Proteinkette B). Wird nur die Tasche ausgewählt, in der die am besten bewertete Pose den kleinsten RMSD zur bioaktiven Konformation aufweist (vgl. Abbildung 7.8 B), so ist dieser in 82 % der Komplexe kleiner als 2 Å. Dies zeigt, dass HYDE für fast alle Komplexe eine durch Docking generierte Ligandkonformation als beste auswählt, die ähnlich der bioaktiven Konformation ist. In Abbildung 7.8 B sind die wenigen Ausreißer dargestellt, zwei davon werden im nächsten Abschnitt genauer analysiert. Berücksichtigt man bei der Analyse nicht nur die von HYDE als beste bewertete Docking-Pose, sondern die besten 32, so vermindern sich die Schwankungen der RMSD Werte zwischen den Taschen einer Proteinstruktur deutlich (vgl. Abbildung 7.8 C). In 93 % der Komplexe liegt der RMSD von mindestens einer Docking-Pose der besten 32 Posen unter 2 Å. Wird wieder nur jeweils eine Tasche pro Proteinstruktur betrachtet, in der der niedrigste RMSD der besten 32 Posen zu finden ist, ist dieser in 96 % der Fälle unter 2 Å und für 82 % der Bindetaschen ist er besser als 1 Å, so dass mindestens eine gewählte Pose sehr ähnlich zur bioaktiven Konformation ist.

7. Resultate und Diskussion

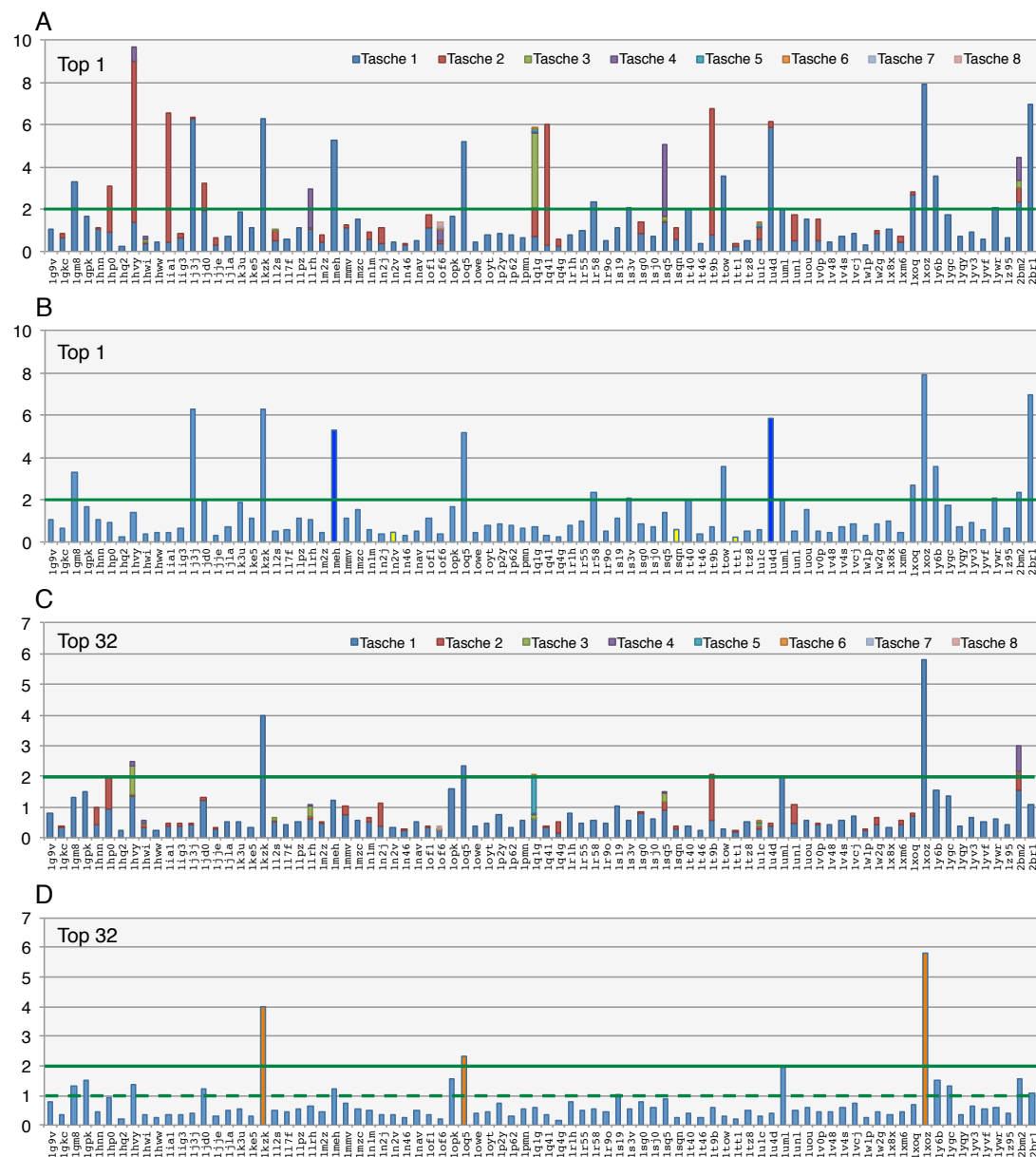
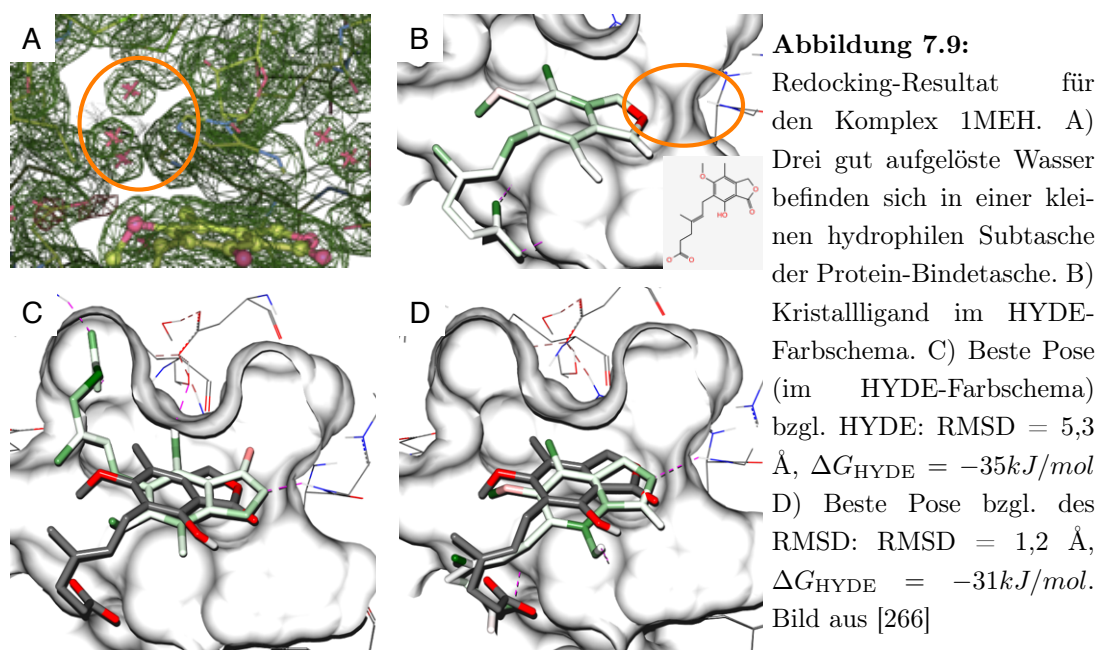


Abbildung 7.8: RMSD Verteilung der 85 Komplexe des Astex Diverse Set. A) Top 1 Pose aller 147 Bindetaschen. B) Top 1 Pose der jeweils besten Bindetasche (auf den RMSD bezogen) einer Kristallstruktur. Dunkelblau: Ausreißer, die im Text diskutiert werden. Gelb: Beispiele, auf die im Text eingegangen wird. C) Top 32 Pose aller 147 Bindetaschen. D) Top 32 Pose der jeweils besten Bindetasche (auf den RMSD bezogen) einer Kristallstruktur. Orange: Hier wurde keine bessere Pose durch das Docking-Programm geliefert. Bild in Anlehnung an [266]

In Abbildung 7.8 B sind in dunkelblau zwei Ausreißer bezüglich des RMSD markiert - 1MEH und 1U4D. Diese werden im Folgenden analysiert. Im Anschluss werden drei Beispiele diskutiert (Abbildung 7.8 B gelb markiert) in denen HYDE eine Pose, die sehr ähnlich der bioaktiven Konformation ist, als beste bewertet hat.



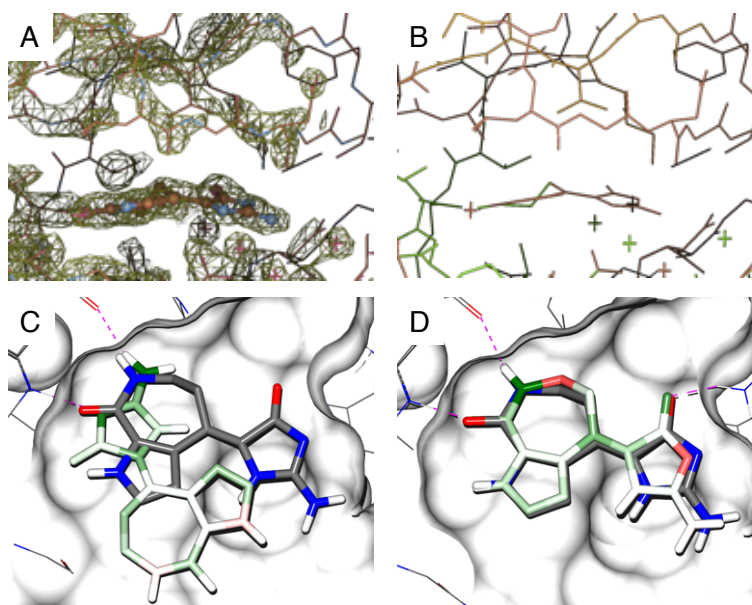
1MEH: Inosine Monophosphat Dehydrogenase mit Mycophenolsäure

In der Kristallstruktur von 1MEH werden nur zwei Wasserstoffbrücken zwischen der Carboxylat-Gruppe der flexiblen Seitenkette des Liganden und der Protein-Bindetasche ausgebildet (vgl. Abbildung 7.9 B). Durch zu große Abweichung von den idealen Wasserstoffbrücken-Parametern kann keine weitere Wasserstoffbrücke zwischen dem aromatischen Sauerstoff des Liganden (rotes Atom Abbildung 7.9 B) und einem Amid-Stickstoff des Proteins (N GLY 314 A) entstehen. Dies könnte ein Artefakt sein, das durch die Soaking-Technik bei der Kristallisation entstanden ist. Beide polaren Atome werden dadurch desolvatisiert und tragen damit ungünstig zur HYDE-Energie bei. Damit ist die HYDE-Energie von $\Delta G_{\text{HYDE}} = -14 \text{ kJ/mol}$ der Kristallstruktur deutlich niedriger als die experimentell bestimmte Affinität von $K_i = 9 \text{ nM}$ ($\Delta G_{\text{exp}} = -46 \text{ kJ/mol}$) [267]. Die Docking-Pose, die von HYDE am besten bewertet wird, hat eine HYDE-Energie von $\Delta G_{\text{HYDE}} = -35 \text{ kJ/mol}$ und einen RMSD zur Kristallstruktur von 5,3 Å (Abbildung 7.9 C). Diese Pose liegt genau umgedreht zur bioaktiven Konformation des Liganden in der Bindetasche vor und die Carboxylat-Gruppe der flexiblen Seitenketten bindet

über zwei Wasserstoffbrücken in einer kleinen hydrophilen Subtasche. In dieser Tasche befinden sich in der Kristallstruktur des Komplexes drei gut aufgelöste Wassermoleküle (vgl. Abbildung 7.9 A). Da alle Kristallwasser für das Redocking-Experiment aus der Bindetasche entfernt wurden, bildet diese auch eine attraktive Bindestelle für die flexible Seitenkette des Liganden. Die Pose mit dem niedrigsten RMSD unter den 32 am besten bewerteten Docking-Posen hat eine mittlere Abweichung von 1,2 Å zur bioaktiven Konformation des Liganden. Die HYDE-Energie dieser Pose ist nur um 4 kJ/mol niedriger ($\Delta G_{HYDE} = -31 kJ/mol$) als die der am besten bewerteten Pose. In diesem Fall ist der Ligand nur um etwa 1 Å im Vergleich zur Kristallstruktur in der Bindetasche verschoben, so dass sich die fehlende Wasserstoffbrücke zwischen dem aromatischen Sauerstoff des Liganden und dem Amid-Stickstoff des Protein-Rückgrates ausbilden kann. Zusätzlich zu dieser entsteht eine weitere Wasserstoffbrücke zwischen der Hydroxyl-Gruppe des Liganden und der Bindetasche.

Abbildung 7.10:

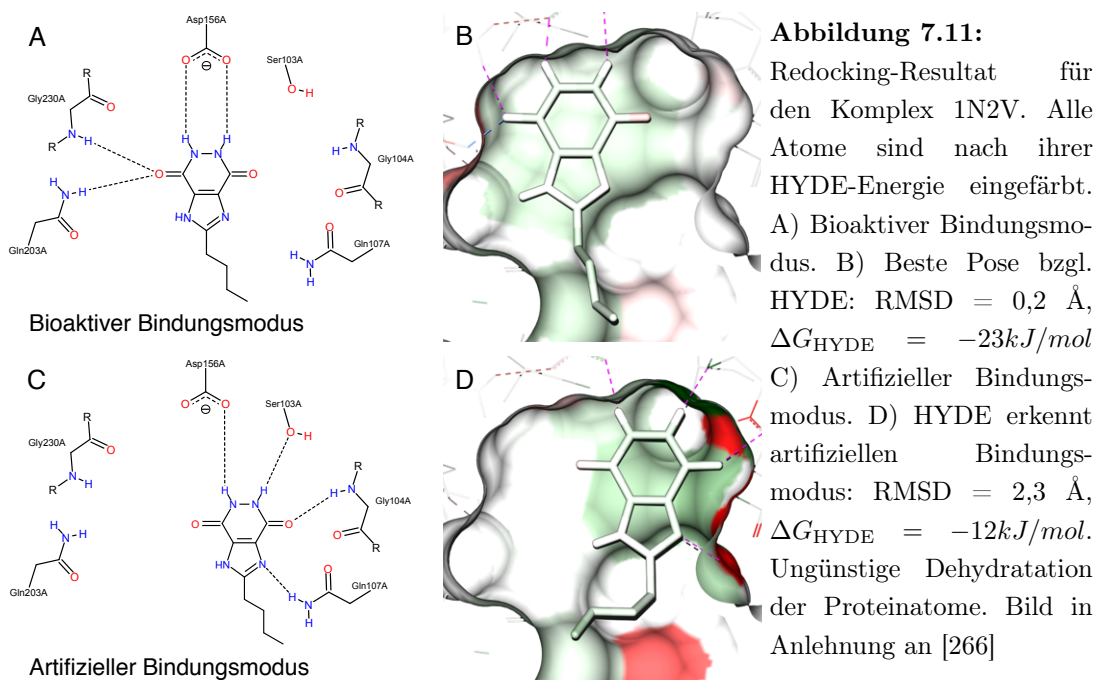
Redocking-Resultat für den Komplex 1U4D. A) Fehlende Elektronendichte der Bindetasche. B) Teilweise hohe Temperaturfaktoren der Bindetasche (rot/braun: hoch, grün/blau: niedrig) C) Beste Pose (im HYDE-Farbschema) bzgl. HYDE: RMSD = 5,9 Å, $\Delta G_{HYDE} = -27kJ/mol$ D) Beste Pose bzgl. des RMSD: RMSD = 0,4 Å, $\Delta G_{HYDE} = -21kJ/mol$. Bild in Anlehnung an [266]



1U4D: ACK1 Kinase mit Debromohymenialdisin

In der Kristallstruktur von 1U4D liegen für große Teile der Bindetasche keine Elektronendichte und hohe Temperaturfaktoren vor (vgl. Abbildung 7.10 A und B). Die Wasserstoffbrücken, die der Ligand mit den gut aufgelösten Teilen der Bindetasche in der Kristallstruktur ausbildet, weisen eine deutlich bessere geometrische Qualität auf und leisten damit einen stabilisierenden Beitrag zur HYDE-Energie, während die polaren Atome im übrigen Teil der Tasche durch ihre Dehydratationskosten destabilisierend

wirken. Daher ergibt sich für die Pose mit dem besten RMSD unter den besten 32 nur eine geringe HYDE-Energie von $\Delta G_{\text{HYDE}} = -21 \text{ kJ/mol}$ (Abbildung 7.10 D). Die Pose hingegen, die von HYDE am besten bewertet wird, hat einen RMSD von $5,9 \text{ \AA}$ und liegt im Vergleich zur bioaktiven Konformation umgekehrt und verschoben in der Bindetasche (Abbildung 7.10 C). Die HYDE-Energie ist in diesem Fall mit einem Wert von $\Delta G_{\text{HYDE}} = -27 \text{ kJ/mol}$ nicht viel höher. Diese Pose wird von HDYE als besser eingestuft, da die Dehydratationskosten der polaren Atome des schlecht aufgelösten Teils der Bindetasche wegfallen.



1N2V: tRNA-Guanin Transglycosylase mit Inhibitor

In der Bindetasche von 1N2V scheinen zwei verschiedene Bindungsmodi des Inhibitors möglich zu sein. In beiden Fällen kann der Inhibitor vier gute Wasserstoffbrücken mit dem Protein ausbilden (vgl. Abbildung 7.11 A und C). In der Kristallstruktur liegt ein konserviertes Wassermolekül (HOH 1404 A) auf der rechten Seite der Bindetasche vor, das mit drei Aminosäuren interagiert (GLY 104 A, TYR 106 A, GLN 107 A). Obwohl Kristallwasser nicht im Redocking vorhanden waren, konnte HYDE die bioaktive Konformation des Liganden gegenüber dem artifiziiellen Bindungsmodus als beste Pose mit einem RMSD von $0,5 \text{ \AA}$ eindeutig identifizieren (Abbildung 7.11 B). Die HYDE-Energie dieser Pose von $\Delta G_{\text{HYDE}} = -23 \text{ kJ/mol}$ entspricht der experimentell gemessenen Bindungsenergie von $K_i = 83 \text{ \mu M}$ ($\Delta G_{\text{exp}} = -23 \text{ kJ/mol}$). Im Vergleich dazu liefert der

artifizielle Bindungsmodus des Inhibitors (Abbildung 7.11 C) mit einem RMSD von $1,2 \text{ \AA}$ eine HYDE-Energie von nur $\Delta G_{\text{HYDE}} = -12 \text{ kJ/mol}$. Die Carbonyl-Gruppe des Liganden, die mit GLN 104 A eine Wasserstoffbrücke eingeht, kann jedoch die anderen beiden polaren Atome von TYR 106 A und GLN 107 A, die in der Kristallstruktur mit dem konservierten Wassermolekül wechselwirken, nicht absättigen. Dies wird von HYDE erkannt und mit den Dehydratationskosten dieser polaren Atome bestraft (vgl. Abbildung 7.11 D).

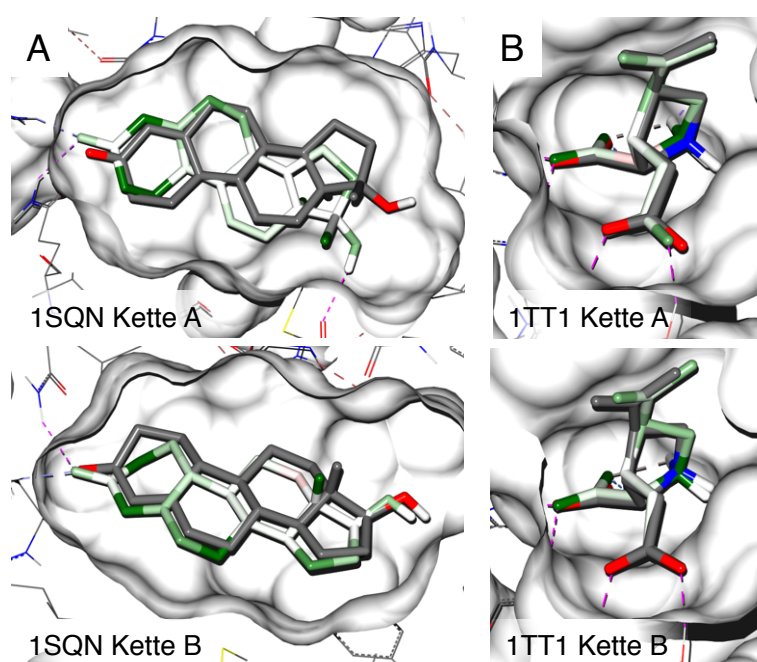


Abbildung 7.12: Redocking-Resultat für einen hydrophoben (1SQN) und einen hydrophilen Komplex (1TT1). A) 1SQN: Beste Pose (HYDE-Farbschema) bzgl. HYDE in Proteinkette A: $\text{RMSD} = 1,1 \text{ \AA}$, $\Delta G_{\text{HYDE}} = -50 \text{ kJ/mol}$. Beste Pose bzgl. HYDE in Proteinkette B: $\text{RMSD} = 0,6 \text{ \AA}$, $\Delta G_{\text{HYDE}} = -55 \text{ kJ/mol}$. B) 1TT1: Beste Pose bzgl. HYDE in Proteinkette A: $\text{RMSD} = 0,2 \text{ \AA}$, $\Delta G_{\text{HYDE}} = -44 \text{ kJ/mol}$. Beste Pose bzgl. HYDE in Proteinkette B: $\text{RMSD} = 0,4 \text{ \AA}$, $\Delta G_{\text{HYDE}} = -45 \text{ kJ/mol}$. Bild in Anlehnung an [266]

1SQN: Progesteronrezeptor mit Norethisteron

In dieser sehr hydrophoben und starren Bindetasche, in der nur wenige gerichtete Wechselwirkungen vorgefunden werden, weist die durch HYDE jeweils am besten bewertete Pose in beiden Bindetaschen der Proteinstruktur eine sehr hohe Ähnlichkeit zur bioaktiven Konformation des Liganden auf (Abbildung 7.12 A). In der Bindetasche von Proteinkette A ergibt sich für den RMSD der am besten bewerteten Pose ein Wert von $1,1 \text{ \AA}$

und eine HYDE-Energie von $\Delta G_{\text{HYDE}} = -50 \text{ kJ/mol}$. Diese Pose ist im Vergleich zur bioaktiven Konformation nur wenig verschoben, so dass die Hydroxyl-Gruppe des Liganden eine zusätzliche Wasserstoffbrücke mit dem Protein ausbilden kann. Die am besten bewertete Pose in der Bindetasche von Proteinkette B liefert einen RMSD von $0,6 \text{ \AA}$ zur Kristallstruktur des Liganden und eine HYDE-Energie von $\Delta G_{\text{HYDE}} = -55 \text{ kJ/mol}$. In beiden Fällen liegt eine gute Übereinstimmung der HYDE-Energie mit der experimentell bestimmten Dissoziationskonstante $K_d = 0,4 \text{ nM}$ ($\Delta G_{\text{exp}} = -54 \text{ kJ/mol}$) vor.

1TT1: Glutamatrezeptor mit Kainsäure

Der Glutamatrezeptor besitzt eine sehr hydrophile Bindetasche und geht acht geometrisch ideale Wasserstoffbrücken mit Kainsäure ein, die ein Strukturanalogon der Glutaminsäure ist. In beiden Bindetaschen identifiziert HYDE eine Pose mit einem RMSD von $0,2 \text{ \AA}$ bzw. $0,4 \text{ \AA}$ zur bioaktiven Konformation als beste (Abbildung 7.12 B). Die HYDE-Energie beider Posen von $\Delta G_{\text{HYDE}} = -44 \text{ kJ/mol}$ bzw. $\Delta G_{\text{HYDE}} = -45 \text{ kJ/mol}$ stimmt gut mit der experimentell bestimmten Inhibitionskonstante $K_i = 64,4 \text{ nM}$ ($\Delta G_{\text{exp}} = -43 \text{ kJ/mol}$) überein.

In diesem Abschnitt wurde gezeigt, dass die HYDE-Bewertungsfunktion für die meisten Protein-Ligand-Komplexe des Astex Diverse Set die bioaktive Konformation des Liganden unter vielen Docking-Posen als beste erkennt. Dies ist vor allem im folgenden Experiment, dem virtuellen Screening wichtig, da hierbei anhand des vorgeschlagenen Bindungsmodus und der davon abhängigen Bewertung eines Moleküls zwischen biologisch aktiven und inaktiven Molekülen unterschieden werden soll.

7.4 Trennung von aktiven und inaktiven Molekülen

Zur Validierung der Fähigkeit der HYDE-Bewertungsfunktion biologisch aktive (Binder) von biologisch inaktiven Molekülen (Nicht-Binder) zu trennen, wurde der DUD Datensatz [209] herangezogen. Für diesen Datensatz wurden 40 medizinal-chemisch relevante Zielproteine zusammengestellt zu welchen bereits mehrere biologisch aktive Moleküle bekannt sind. Anhand dieser bekannten Binder wurden physikochemisch ähnliche, jedoch topologisch unterschiedliche Moleküle aus der ZINC Datenbank [171] gesucht. Dabei wurden pro aktivem Moleküle 36 Moleküle aus der ZINC Datenbank entnommen, von denen vermutet wird, dass sie für das jeweils betrachtete Zielprotein keine messbare Aktivität zeigen. Dies wurde jedoch nicht experimentell bestätigt. Damit wurde für jedes Zielprotein jeweils ein Datensatz aus wenigen aktiven und sehr vielen inaktiven Molekülen zusammengestellt, um ein künstlich erschwertes Szenario für die Validierung

von Docking-Programmen zu liefern.

Die Kristallstrukturen der Zielproteine wurden vor dem Experiment mithilfe der experimentell bestimmten Elektronendichte durch eine visuelle Untersuchung auf ihre Qualität hin überprüft, da zum Teil auch ältere Strukturen im Datensatz enthalten waren. Dabei wurden vier Qualitätsstufen festgelegt: Q1 bedeutet keine oder nur kleinere Mängel; Q2 erhielten Strukturen, die nur Mängel am Kristallliganden aufwiesen oder die Hinzunahme von, für die Bindung wichtigen Wassern erfordern würden; Q3 und Q4 Strukturen zeigten fehlende Elektronendichte, hohe Temperaturfaktoren oder Packungseffekte auf. Eine genaue Charakterisierung der Strukturen findet sich in Tabelle C.3 in Anhang C.3. In diesem Experiment wurden alle Bindetaschen, jeweils eine Bindetasche pro Zielprotein, wiederum mit der LeadIT-Software (v2.0.2) vorbereitet: Anhand des Kristallliganden des jeweiligen Zielproteins wurden die Aminosäuren, die Kofaktoren und die Ionen der Bindetasche, die sich in einem Abstand von 6,5 Å um die Schweratome des Liganden befanden, bestimmt. Die automatisch bestimmte Koordinationsgeometrie des Metallions wurde bei dem Zielprotein Adenosin Deaminase (ADA) manuell korrigiert. Anschließend wurde mit ProToss das Wasserstoffbrücken-Netzwerk der Bindetasche optimiert. Alle aktiven und inaktiven Moleküle des Datensatzes wurden mit FlexX, wobei zwei verschiedene Platzierungsmethoden verwendet wurden, in die Bindetasche des jeweiligen Zielproteins gedockt. Pro Molekül wurden bis zu 40 Posen erzeugt und durch den HYDE-Optimierungskontext vorbereitet, bevor mit HYDE daraus die beste Pose bestimmt wurde. Diese wurde entsprechend ihrer HYDE-Energie in eine Liste aller erfolgreich gedockten Moleküle eines Zielproteins eingeordnet, so dass die Moleküle, die die beste HYDE-Energie aufweisen am Anfang der Liste auftauchen. Im optimalen Fall werden die Binder vor den Nicht-Bindern in der Liste eingeordnet. Abbildung 7.13 stellt das Resultat dieses virtuellen Screening-Experiments für alle 40 Zielproteine des DUD Datensatzes dar. Die AUC-Werte der ROC-Kurven sind für jedes der Zielproteine angegeben, um ein Qualitätsmaß für den gesamten virtuellen Screening-Lauf bereitzustellen. Dabei würde ein AUC-Wert von 0,5 bei einer zufälligen Auswahl der Moleküle erreicht werden. Um die frühe Anreicherung von aktiven Molekülen, d. h. eine Sortierung vorne in der Liste, zu überprüfen, wurde die Trefferquote (Sensitivität) bei einer Falsch-Positiv-Rate von 0,1 %, 1 % und 2 % bestimmt. Hierbei entspricht eine Trefferquote von 0,1 %, 1 % und 2 % einer zufälligen Auswahl von Molekülen. Weiterhin sind die Zielproteine in Abbildung 7.13 entsprechend der Qualitätseinstufung (Q1 - Q4) ihrer Kristallstruktur farblich unterlegt.

In Abbildung 7.13 ist zu erkennen, dass für die meisten der Zielproteine ein deutlich besseres Resultat als durch eine zufällige Auswahl der Moleküle erreicht wird (bei 82 %

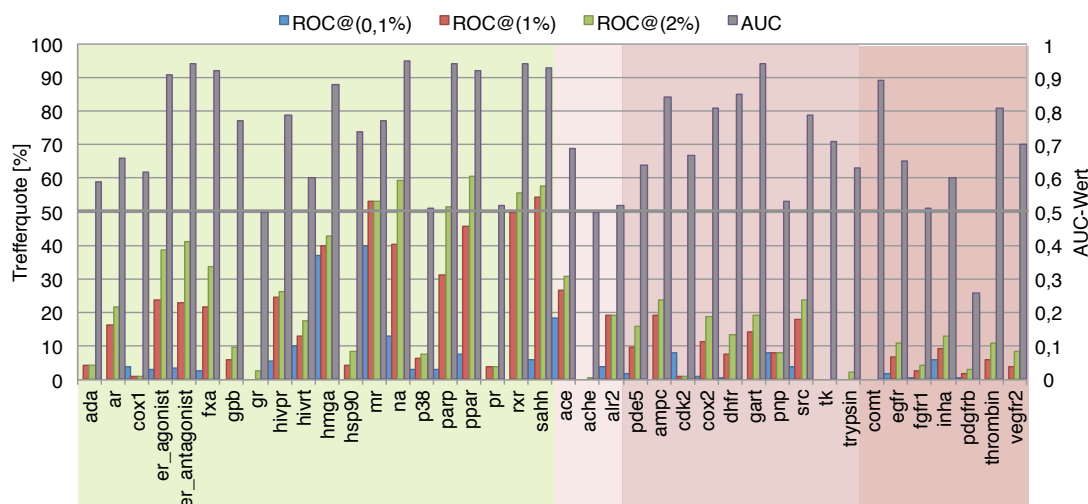


Abbildung 7.13: Resultat des virtuellen Screenings für alle Zielproteine des DUD Datensatzes. Die AUC-Werte (graue Balken) sowie die Trefferquote in Prozent bei einer Falsch-Positiv-Rate von 0,1 % (blaue Balken), 1 % (rote Balken) und 2 % (grüne Balken), sind für alle Zielproteine dargestellt. Die vier Farbabstufungen, mit welchen das Diagramm unterlegt ist, repräsentieren die vier Qualitätsstufen Q1 (gute Qualität, grüner Bereich) bis Q4 (schlechte Qualität, dunkel roter Bereich) der Strukturen der Zielproteine.

ist der AUC-Wert $> 0,5$). Im Mittel ergibt sich ein AUC-Wert von 0,72. Dieser verbessert sich auf 0,77, wenn nur die Zielproteine mit einer Q1-Kristallstruktur berücksichtigt werden. Die frühe Anreicherung von aktiven Molekülen ist im Vergleich zum AUC-Wert eindeutiger durch die Qualität der Kristallstruktur des Zielproteins (vgl. Abbildung 7.13) beeinträchtigt. Der Median der Trefferquote bei einer Falsch-Positiv-Rate von 1 % liegt bei 10 %, wenn alle Strukturen betrachtet werden, und steigt um zwölf Prozentpunkte, wenn die Statistik auf die Q1-Strukturen beschränkt wird, d. h. 22 % der aktiven Moleküle wurden schon identifiziert, wobei nur 1 % der inaktiven Moleküle als bindend vorhergesagt wurde. Weitere Resultate sind in Tabelle 7.5 für alle Zielproteine und für die Q1-Strukturen zusammengefasst.

Im DUD Datensatz sind fünf verschiedene Proteinfamilien und eine Anzahl verschiedener anderer Proteine enthalten. Um einen besseren Eindruck über die Resultate zu bekommen, wurden die Ergebnisse zusätzlich entsprechend der verschiedenen Proteinfamilien zusammengefasst (vgl. Tabelle 7.6). Damit soll überprüft werden, ob HYDE für spezielle Proteinfamilien bessere bzw. schlechte Resultate liefert. Dabei wird insgesamt deutlich, dass HYDE bei allen Proteinfamilien im Durchschnitt besser aktive Moleküle identifizieren kann, als mit einer zufälligen Auswahl der Moleküle erreicht würde.

7. Resultate und Diskussion

	Trefferquote in %, bei einer Falsch-Positiv-Rate von			
	AUC-Wert	ROC@(0,1 %)	ROC@(1 %)	ROC@(2 %)
Alle Strukturen				
Mittelwert	0,72	4,82	15,72	20,6
Std. Abw.	0,17	8,84	15,99	19,07
Median	0,73	1,88	9,61	14,56
Minimum	0,26	0,00	0,00	0,00
Maximum	0,95	40,00	54,55	60,49
Nur Q1 Strukturen				
Mittelwert	0,77	6,94	23,13	29,87
Std. Abw.	0,17	11,41	18,68	21,95
Median	0,78	3,08	22,46	30,11
Minimum	0,50	0,00	0,00	1,00
Maximum	0,95	40,00	54,55	60,49

Tabelle 7.5: Resultat virtuelles Screening auf dem DUD Datensatz. Tabelle in Anlehnung an [266]

Proteinfamilie	N	Trefferquote in %, bei einer Falsch-Positiv-Rate von							
		AUC-Wert		ROC@(0,1 %)		ROC@(1 %)		ROC@(2 %)	
		MW	ME	MW	ME	MW	ME	MW	ME
Kernrezeptoren	8	0,77	0,84	6,75	1,50	27,00	23,48	34,67	39,92
Kinasen	9	0,62	0,67	2,00	0,64	5,00	4,05	7,54	7,81
Serin-Proteasen	3	0,79	0,81	0,94	0,00	9,33	6,15	15,61	10,77
Metalloenzyme	4	0,70	0,67	5,08	0,98	10,17	7,08	12,66	10,02
Folat-Enzyme	2	0,90	0,90	0,25	0,25	10,88	10,88	16,24	16,24
Andere	14	0,73	0,78	6,97	4,83	19,80	15,97	24,92	19,03

Tabelle 7.6: Resultat virtuelles Screening auf dem DUD Datensatz zusammengefasst nach Proteinfamilie. N = Anzahl. MW = Mittelwert. ME = Median.

Beurteilt man nur den mittleren AUC-Wert, so liefert HYDE bei den Folat-Enzymen und den Serin-Proteasen die besten Ergebnisse, während die Kinasen den niedrigsten mittleren AUC-Wert aufweisen. Bei Kinasen ist bekannt, dass Proteinflexibilität eine große Rolle zur Ausübung ihrer Funktion spielt und damit auch zur richtigen Erkennung der Liganden notwendig ist. Ein Beispiel dafür ist die p38-MAP-Kinase, die im DUD Datensatz in der DFG-out-Konformation vorliegt. Die enthaltenen aktiven Moleküle, binden jedoch zu 83 % an die DFG-in-Konformation des Proteins und können daher mit HYDE nicht richtig identifiziert werden, so dass sich bei der p38-MAP-Kinase ein AUC-Wert von 0,51 ergibt. Werden hingegen nur die 17 % der aktiven Moleküle berücksichtigt, die auch tatsächlich in die DFG-out-Konformation binden, so erzielt HYDE einen AUC-Wert von 0,92 [201, 266]. Dies würde zu einer Verbesserung des mittleren AUC-Wertes bei Kinasen auf 0,66 führen. Wichtiger jedoch als der mittlere AUC-Wert ist wieder die frühe Anreicherung von aktiven Molekülen. Tabelle 7.6 zeigt eindeutig, dass HYDE hier vor allem bei den Kernrezeptoren und Proteinen, die keiner Familie zugeordnet wurden, gute Resultate im Bereich von 20 - 27 % identifizierte aktive Moleküle bei einer Falsch-Positiv-Rate von 1 % liefert. Für Kinasen, Serin-Proteasen und Metalloenzyme werden hierbei deutlich niedrigere Trefferquoten erzielt.

7.4.1 Detaillierte Analyse der Resultate des virtuellen Screenings

Im Folgenden wird die Ursache dieser niedrigen Trefferquote anhand einer Kinase, der Thymidin-Kinase TK, und einem Metalloenzym, der Catechol-O-Methyltransferase COMT, genauer analysiert. Anschließend wird auf das Resultat von drei Zielproteinen, einem Kernrezeptor (Estrogenrezeptor), der Neuraminidase und der HMG-CoA-Reduktase, bei denen durch HYDE eine sehr gute Anreicherung von aktiven Molekülen erreicht wurde, eingegangen.

Thymidin-Kinase

Die Thymidin-Kinase katalysiert die Phosphorylierung von Desoxythymidin (vgl. Kristallstruktur Abbildung 7.14) zu Desoxythymidin-5-Phosphat. Sie spielt damit eine wichtige Rolle bei der DNA-Synthese und dient z. B. als Zielprotein für antivirale Wirkstoffe. Die meisten Inhibitoren der Thymidin Kinase sind Substratanaloga, d. h. sie besitzen eine Thymin-ähnliche Gruppe und binden über diese spezifisch in die Desoxythymidin Bindetasche der Thymidin Kinase. Es ist bekannt, dass die Thymidin-Kinase durch Proteinflexibilität, wasserüberbrückte Wechselwirkungen und die niedrige Affinität vieler Inhibitoren eine Herausforderung im virtuellen Screening darstellt [268].

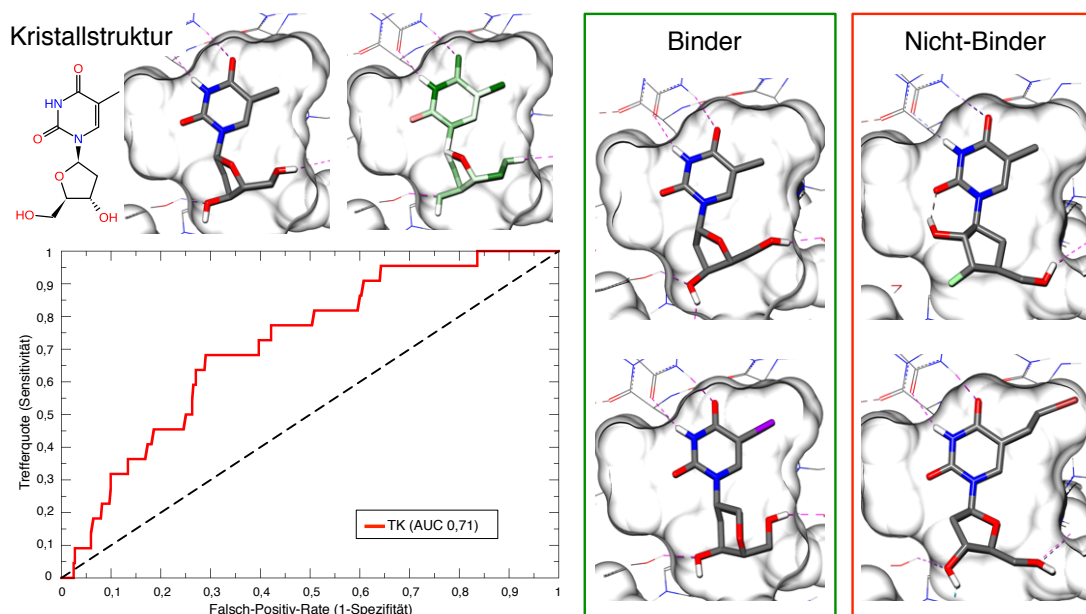


Abbildung 7.14: Resultat des virtuellen Screenings für die Thymidin Kinase (TK) (PDB: 1KIM). Die Kristallstruktur mit Thymidin als Liganden ist im Element-Farbschema und im HYDE-Farbschema ist dargestellt ($\Delta G_{exp} = -30 kJ/mol$, $\Delta G_{HYDE} = -31 kJ/mol$). Die ROC-Kurve mit einem AUC-Wert von 0,71 ist abgebildet. Rechts sind zwei Binder und zwei Nicht-Binder aus dem TK Datensatz gezeigt. Viele der Nicht-Binder besitzen eine Thymin-ähnliche Gruppe, die wichtig für die Erkennung durch das Enzym ist. Der Nicht-Binder rechts unten ist ein Agonist der Thymidin-Kinase-2. Bild in Anlehnung an [266]

Die Anreicherung an aktiven Molekülen, die sich für die Thymidin-Kinase im virtuellen Screening mit HYDE ergibt, ist mit einem AUC-Wert von 0,71 eher mäßig. Vor allem eine frühe Anreicherung von aktiven Molekülen ist nicht vorhanden (vgl. ROC-Kurve Abbildung 7.14). Die Bewertung, der im DUD Datensatz vorhandenen Thymidin-Kinase-Kristallstruktur (PDB: 1KIM) mit HYDE, liefert eine gute Übereinstimmung der HYDE-Energie ($\Delta G_{HYDE} = -31 kJ/mol$) mit der experimentell bestimmten Bindungsenergie ($\Delta G_{exp} = -30 kJ/mol$) [269] (vgl. Abbildung 7.14). Die Analyse des Bindungsmodus, der am besten bewerteten Moleküle zeigte, dass dieser ebenfalls eine hohe Ähnlichkeit zu dem Bindungsmodus der Kristallstruktur aufwies. Sowohl die gut bewerteten aktiven Moleküle des Datensatzes als auch die inaktiven Moleküle die von HYDE als aktiv eingestuft wurden, interagieren über die Thymin-ähnliche Gruppe mit der Bindetasche des Proteins (vgl. Abbildung 7.14) rechts). Bei der Untersuchung des Thymidin-Kinase-Datensatzes enthielten 72 der inaktiven Moleküle, die für die Erkennung

nung durch das Enzym wichtige Thymin-ähnliche Gruppe, das entspricht einem Anteil von 18 %. Wohingegen 18 der aktiven Moleküle (81 %) ebenfalls eine solche Gruppe enthalten. Eines der inaktiven Moleküle, das durch HYDE als Binder eingeordnet wurde, konnte als Agonist der Thymidin-Kinase-2 identifiziert werden [270] (vgl. Abbildung 7.14) rechts unten). Da die übrigen inaktiven Moleküle des Datensatzes nicht experimentell als nicht bindend validiert wurden, kann über diese durch die hohe Ähnlichkeit zu den aktiven Molekülen eigentlich keine Aussage getroffen werden. Somit kann, bei dieser Zusammensetzung des Datensatzes keine frühe Anreicherung der aktiven Moleküle erwartet werden.

Catechol-O-Methyltransferase

Die Catechol-O-Methyltransferase gehört zur Familie der Metalloenzyme und ist für die Inaktivierung verschiedener Botenstoffe im Körper zuständig. Dies geschieht durch die Methylierung einer phenolischen Hydroxyl-Gruppe der Catecholamine dieser Botenstoffe. Jedoch werden durch die Catechol-O-Methyltransferase auch Wirkstoffe, die gegen neurodegenerative Krankheiten wie *Morbus Parkinson* eingesetzt werden, methyliert und somit unwirksam. Daher wurden Catechol-O-Methyltransferase-Inhibitoren entwickelt, um so die Bioverfügbarkeit dieser Wirkstoffe zu erhöhen. Die Bindetasche der Catechol-O-Methyltransferase ist eine sehr kleine, flache Tasche an der Oberfläche des Proteins. Nur die Catechol-Gruppe des Liganden, die das Magnesiumion, das in einer oktaedrischen Geometrie vorliegt, koordiniert, wird in der kleinen Tasche vergraben (vgl. Kristallstruktur Abbildung 7.15). Im virtuellen Screening erreichte HYDE einen guten AUC-Wert von 0,89 bei der Catechol-O-Methyltransferase, jedoch konnte keine frühe Anreicherung von aktiven Molekülen gefunden werden (vgl. ROC-Kurve Abbildung 7.15). Die Analyse der Elektronendichtekarte der Kristallstruktur (PDB: 1H1D) zeigt, dass der Bindungsmodus des Kristallliganden durch einen Kristallpackungseffekt mit einem symmetrieverwandtem Molekül beeinflusst wird. Betrachtet man die HYDE-Energie des Komplexes ($\Delta G_{\text{HYDE}} = -54 \text{ kJ/mol}$), so wird fast die gesamte Bindungsenergie durch die Interaktion der Catechol-Gruppe mit dem Magnesiumion bewirkt ($\Delta G_{\text{HYDE}} = -42 \text{ kJ/mol}$) (vgl. Abbildung 7.15). Durch die exponierte Lage des Liganden kann man davon ausgehen, dass dies bei der experimentell gemessenen Inhibitionskonstante $K_i = 6 \text{ nM}$ ($\Delta G_{\text{exp}} = -47 \text{ kJ/mol}$) auch der Fall ist. Im Datensatz der inaktiven Moleküle sind ebenfalls 18 Moleküle (5 %) mit einer Catechol-Gruppe enthalten, die auch sterisch in der Lage wären an das Magnesiumion der Catechol-O-Methyltransferase zu binden. Dies sind mehr Moleküle als die 11 aktiven, die im Datensatz enthalten sind und sich nur marginal von diesen inaktiven Molekülen unterscheiden.

7. Resultate und Diskussion

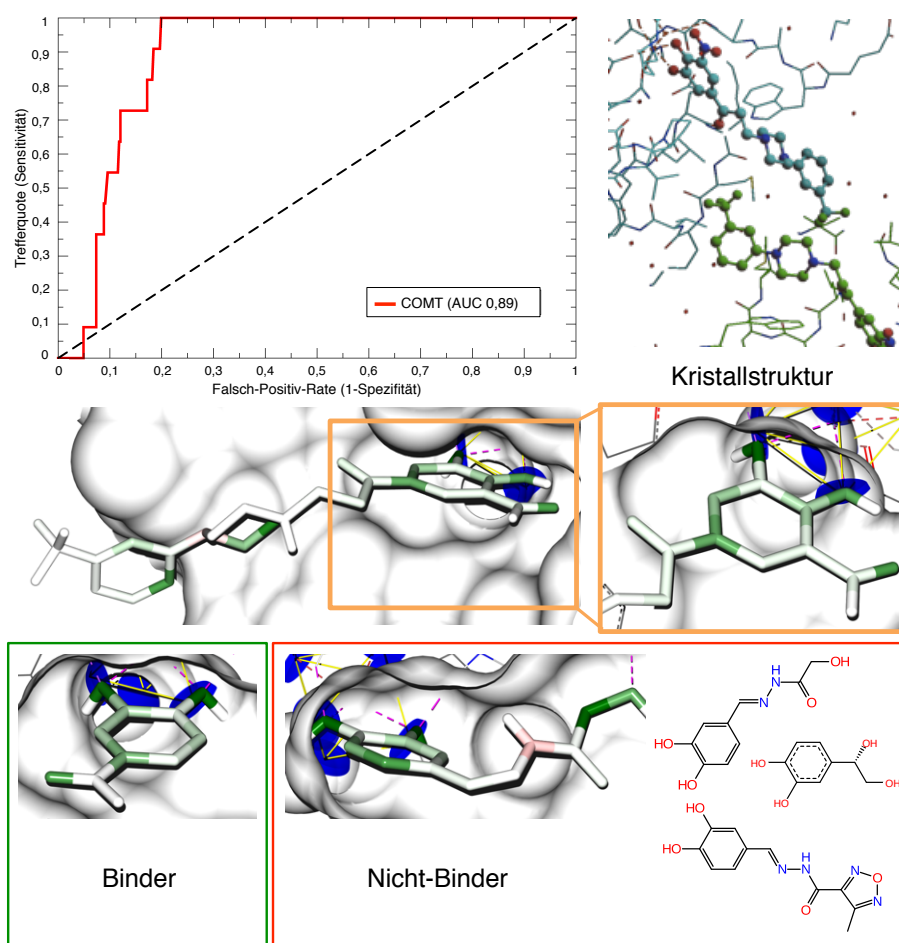


Abbildung 7.15: Resultat des virtuellen Screenings für die Catechol-O-Methyltransferase (COMT) (PDB: 1H1D). Die ROC-Kurve mit einem AUC-Wert von 0,89 ist dargestellt. Die Kristallstruktur zeigt den Kontakt zu einem symmetrieverwandten Molekül. Der Bindungsmodus des Liganden wird dadurch beeinflusst. Die Kristallstruktur im HYDE-Farbschema wird gezeigt ($\Delta G_{exp} = -47kJ/mol$, $\Delta G_{HYDE} = -54kJ/mol$), der Hauptanteil der Energie kommt von der Wechselwirkung der Catechol-Gruppe mit dem Magnesiumion ($\Delta G_{HYDE} = -42kJ/mol$). Die Docking-Pose eines Binders (zinc03814484) ist im HYDE-Farbschema dargestellt ($\Delta G_{HYDE} = -46kJ/mol$) sowie die eines Nicht-Binders (zinc00366295) ($\Delta G_{HYDE} = -54kJ/mol$). Bild in Anlehnung an [266]

Wie schon im Fall der Thymidin-Kinase sollten diese experimentell nicht als inaktiv validierten Moleküle nicht in einem solchen Datensatz enthalten sein. HYDE bewertet alle inaktiven Moleküle, die eine Catechol-Gruppe enthalten und somit eine gute Bindungsenergie erzielen können als aktiv, so dass keine frühe Anreicherung der als aktiv markierten Moleküle erreicht werden kann. Zur Verdeutlichung ist in Abbildung 7.15 ein durch

HYDE gut bewertetes sehr kleines, aktives Molekül (Binder) ($\Delta G_{\text{HYDE}} = -46 \text{ kJ/mol}$) und ein Nicht-Binder mit einer Catechol-Gruppe ($\Delta G_{\text{HYDE}} = -54 \text{ kJ/mol}$) dargestellt sowie die Strukturen weiterer Nicht-Binder mit Catechol-Gruppe, die im Datensatz enthalten sind.

Estrogenrezeptor

Der Estrogenrezeptor ist ein Kernrezeptor, der durch das Hormon Estrogen (Agonist) aktiviert wird und anschließend als DNA-Transkriptionsfaktor agiert. Zur Bindung des Steroidhormons wird auch die Bindetasche des Estrogenrezeptors hauptsächlich durch hydrophobe Proteinreste gebildet. Um Einfluss auf die Funktion des Estrogenrezeptors zu nehmen, wurden verschiedene Wirkstoffe entwickelt, die je nach Anwendungsgebiet, eine agonistische oder eine antagonistische Wirkung auf den Estrogenrezeptor ausüben. Agonisten werden z. B. zur Empfängnisverhütung oder zur Vorbeugung von Osteoporose eingesetzt, während Antagonisten vor allem in der Krebstherapie Verwendung finden. Die Struktur des Estrogenrezeptor verändert sich, je nachdem, ob ein Agonist oder Antagonist gebunden wird, dazu ist eine Konformationsänderung der Helix 12 notwendig. Diese bildet den Eingang zur Bindetasche und verschließt diese, im Fall der Bindung eines Agonisten wieder, nachdem sich der Agonist in der Bindetasche befindet. Im DUD Datensatz ist sowohl eine Struktur der Estrogenrezeptors in der Agonisten-Konformation als auch in der Antagonisten-Konformation vorhanden. Die aktiven Moleküle des Estrogenrezeptors sind dementsprechend auch nach Agonist und Antagonist aufgeteilt. In beiden virtuellen Screening-Läufen werden mit HYDE sehr gute Resultate erzielt. Auf dem Antagonisten Datensatz wird ein AUC-Wert von 0,94 erreicht, bei dem Agonisten Datensatz liegt der AUC-Wert bei 0,91. Die frühe Anreicherung von aktiven Molekülen ist ebenfalls in beiden Datensätzen mit Trefferquote von 23 % (Antagonist) bzw. 24 % (Agonist), gemessen bei einer Falsch-Positiv-Rate von 1 %, sehr hoch. Abbildung 7.16 A zeigt die ROC-Kurve und den Bindungsmodus des am besten bewerteten Moleküls - Diethylstilbestrol ein bekannter Agonist - im HYDE-Farbschema für den Agonisten Datensatz. Der Bindungsmodus von Diethylstilbestrol stimmt gut mit der bioaktiven Konformation des Kristallliganden überein (PDB: 1L2I). Die HYDE-Energie von $\Delta G_{\text{HYDE}} = -60 \text{ kJ/mol}$ von Diethylstilbestrol ist etwas höher als experimentell bestimmte Inhibitionskonstanten $K_i = 0,13 - 0,5 \text{ nM}$ ($\Delta G_{\text{exp}} = -53 - -56 \text{ kJ/mol}$) [271, 272].

HMG-CoA-Reduktase

Die HMG-CoA-Reduktase katalysiert eine Schlüsselreaktion des Cholesterolsyntheseweges, die Reduzierung von β -Hydroxy- β -Methylglutaryl-Coenzym-A zu Mevalonat.

7. Resultate und Diskussion

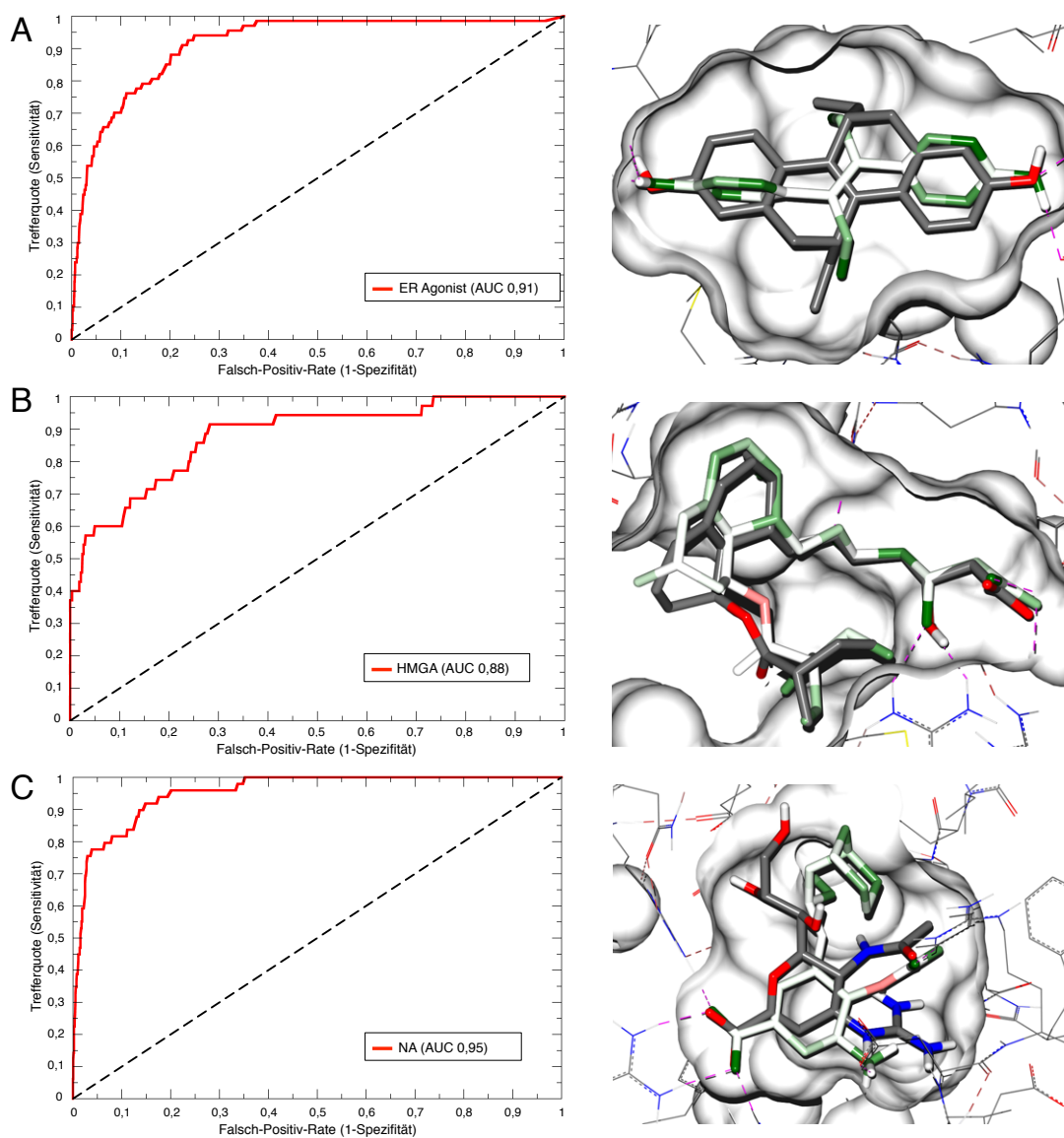


Abbildung 7.16: Resultate im virtuellen Screening für den Estrogenrezeptor (ER Agonist) (PDB: 1L2I), die HMG-CoA-Reduktase (HMGA) (PDB: 1HW8) und die Neuraminidase (NA) (PDB: 1A4G). Links sind die ROC-Kurven der drei Zielproteine dargestellt. Rechts ist jeweils die am besten bewertete Pose (im HYDE-Farbschema) und der Kristallligand (im Element-Farbschema) gezeigt. A) ROC-Kurve mit AUC-Wert = 0,91 von ER-Agonist. Am besten bewertetes Molekül im virtuellen Screening der ER-Agonisten Struktur: zinc03861549 ($\Delta G_{\text{HYDE}} = -60 \text{ kJ/mol}$). B) ROC-Kurve mit AUC-Wert = 0,88 von HMGA. Am besten bewertetes Molekül im virtuellen Screening von HMGA: zinc04134476 ($\Delta G_{\text{HYDE}} = -69 \text{ kJ/mol}$). C) ROC-Kurve mit AUC-Wert = 0,95 von NA. Am besten bewertetes Molekül im virtuellen Screening von NA: zinc04134491 ($\Delta G_{\text{HYDE}} = -51 \text{ kJ/mol}$). Bild in Anlehnung an [266]

Damit ist die HMG-CoA-Reduktase ein Zielprotein, das zur Senkung des Cholesterinspiegels inhibiert wird. Die Bindetasche der HMG-CoA-Reduktase ist relativ groß und eher hydrophil, um die flexible β -Hydroxy- β -Methylglutaryl-Kette des Substrats aufzunehmen. Das virtuelle Screening mithilfe von HYDE liefert für dieses Zielprotein ein gutes Resultat mit einem AUC-Wert von 0,88. Besonders die frühe Anreicherung von aktiven Molekülen ist mit einer Trefferquote von 40 %, während nur 1 % der inaktiven Moleküle von HYDE als aktiv klassifiziert wurden, sehr hoch. Die zwölf durch HYDE am besten bewerteten Moleküle sind ausschließlich bekannte aktive Verbindungen, wie z. B. Fluvastatin oder Simvastatin, die nanomolare Inhibitionskonstanten für die HMG-CoA-Reduktase aufweisen. Die HYDE-Energien von diesen liegen im Bereich von -69 bis -49 kJ/mol , was einer nanomalen Affinität entspricht. In Abbildung 7.16 B ist die ROC-Kurve und der Bindungsmodus des am besten bewerteten Moleküls im HYDE-Farbschema dargestellt. Der Bindungsmodus dieses Moleküls entspricht sehr gut der Konformation des Kristallliganden, vor allem der flexible HMG-Rest des Moleküls liegt in einer identischen Konformation vor. Die ROC-Kurve weist einen Einbruch auf, nachdem ca. 60 % der aktiven Moleküle identifiziert wurden. Die Analyse der aktiven Moleküle zeigt, dass 35 der Moleküle einen HMG-Rest besitzen, jedoch liegen 18 davon in der inaktiven Lactonform vor. Diese inaktive Form kann nicht in die Bindetasche der HMG-CoA-Reduktase binden, sondern wird *in vivo* enzymatisch hydrolysiert und in die aktive Hydroxycarbonsäure überführt [273]. Die betroffenen Moleküle weisen somit keine ähnliche Konformation zur bioaktiven Konformation auf und werden dadurch von HYDE nicht als aktive Verbindung klassifiziert.

Neuraminidase

Die Neuraminidase ist eine Glycosidase, d. h. sie spaltet mithilfe von Wasser die glycosidische Bindung der Sialinsäure. Am bekanntesten sind die viralen Neuraminidasen, da sie ein Zielprotein für die Bekämpfung von z. B. *Influenza*-Viren darstellen. Die Bindetasche der Neuraminidase wird durch viele polare und auch geladene Aminosäuren gebildet, so dass die Bindung des Substrats überwiegend durch Wasserstoff- oder Salzbrücken erfolgt. Im virtuellen Screening der Neuraminidase wird mit HYDE ein sehr hoher AUC-Wert von 0,95 erzielt als auch eine sehr gute Trefferquote von 40 %, bei einer Falsch-Positiv-Rate von 1 % (vgl. ROC-Kurve Abbildung 7.16 C). Zusätzlich wurde wieder der Bindungsmodus der gut bewerteten Moleküle überprüft, wobei dieser auch eine gute Übereinstimmung mit der bioaktiven Konformation des Kristallliganden (Zanamivir) zeigte. Die HYDE-Energien der am besten bewerteten Moleküle liegen im Bereich von $\Delta G_{\text{HYDE}} = -51$ bis -44 kJ/mol und entsprechen somit gut den experimentell bestimmten Werten bekannter Neuraminidase-Inhibitoren wie z. B. Zanamivir

($K_i = 0,1 \text{ nM}$, $\Delta G_{exp} = -55 \text{ kJ/mol}$) [274] oder Oseltamivir ($IC_{50} = 1 - 7 \text{ nm}$, $\Delta G_{exp} \approx -49 \text{ kJ/mol}$) [275–278].

Die Analyse der Resultate des virtuellen Screening-Experiments zeigt, dass die HYDE-Bewertungsfunktion bei vielen der Zielproteine des DUD Datensatzes eine gute Trennung von aktiven und inaktiven Molekülen erzielt. Die Vorhersagequalität von HYDE wird durch die Qualität der zugrunde liegenden Kristallstrukturen beeinflusst. Dies zeigte sich auch schon in den beiden anderen Experimenten, die Vorhersage der Bindungsaffinität und das Auffinden der bioaktiven Konformation eines Liganden. Dabei wurde im Zusammenhang der Vorhersage von Bindungsaffinität gezeigt, dass die Abschätzung der polaren Dehydratation am stärksten durch die Qualität von Strukturen beeinträchtigt wird. Berücksichtigt man nur die belohnenden Terme der HYDE-Funktion, so ergibt sich für die Strukturen des PDBBind 2007 coreset eine gute Korrelation mit den experimentell gemessenen Affinitäten (vgl. Kapitel 7.2.2). Wird der Term der polaren Dehydratation im Gegensatz dazu im virtuellen Screening-Experiment vernachlässigt, so sinkt der Median AUC-Wert deutlich von 0,73 auf 0,65. Somit erhöht sich die Rate der falsch-positiven Moleküle, wenn die vollständige Absättigung des Wasserstoffbrücken-Netzwerkes im Protein-Ligand-Komplex nicht beachtet wird.

7.4.2 Vergleich der Resultate mit anderen Methoden

Abschließend werden die Resultate von HYDE im virtuellen Screening mit dem HYDE-Prototyp und anderen, bereits etablierten Methoden verglichen, die auch auf dem DUD Datensatz evaluiert wurden (GlideSP [279], GOLD::ChemPLP [115], ICM [280], FRED::CG, Surflex::Ringflex [124] und DOCK 6 [281]). Dazu befindet sich im Anhang C.1 ein detaillierter Vergleich der Screening-Resultate auf dem DUD zwischen der überarbeiteten HYDE-Version und dem HYDE-Prototypen [202]. Anhand von Enrichment-Plots kann in vielen Fällen eine deutliche Verbesserung der Ergebnisse durch Verwendung der überarbeiteten HYDE-Version gezeigt werden. Abbildung 7.17 zeigt einen Vergleich mit anderen gut etablierten Methoden. Alle Resultate beruhen dabei auf einer starren Proteinstruktur, da im HYDE-Optimierungskontext bisher keine Minimierung der Proteinstruktur beinhaltet ist und so eine bessere Vergleichbarkeit der Methoden vorliegt. Mit einem Median AUC-Wert von 0,73 gehört die Kombination von einem FlexX-Docking in LeadIT und die Nachbewertung dieser Posen mit HYDE zu den besten der betrachteten Methoden. Die Resultate wurden zusätzlich nach den im DUD enthaltenen Proteinfamilien aufgeschlüsselt, um festzustellen, ob gerade bei Kinasen oder Metalloenzymen, bei denen HYDE einen niedrigeren mittleren AUC-Wert erzielt

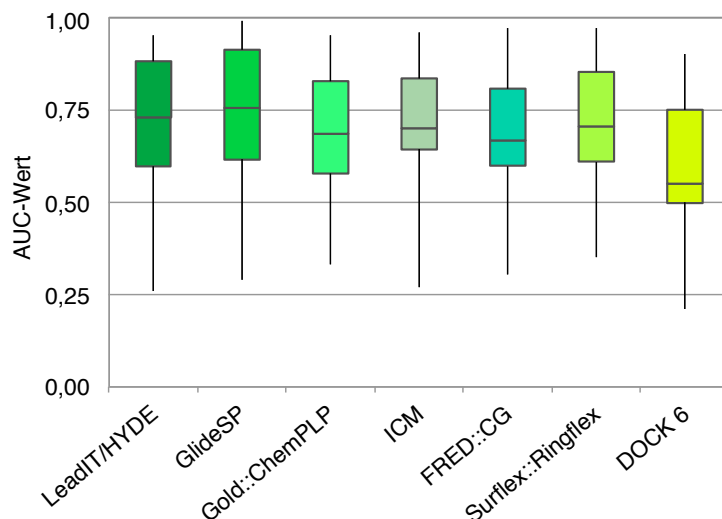


Abbildung 7.17: Vergleich der Resultate verschiedener Methoden im virtuellen Screening auf dem DUD Datensatz. Im Boxplot dargestellt ist der Median der AUC-Werte (graue Linie in der Box) das untere und obere Quartil (Box) und der Minimum- und Maximum-AUC-Wert (schwarze vertikale Linien). Bild in Anlehnung an [208]

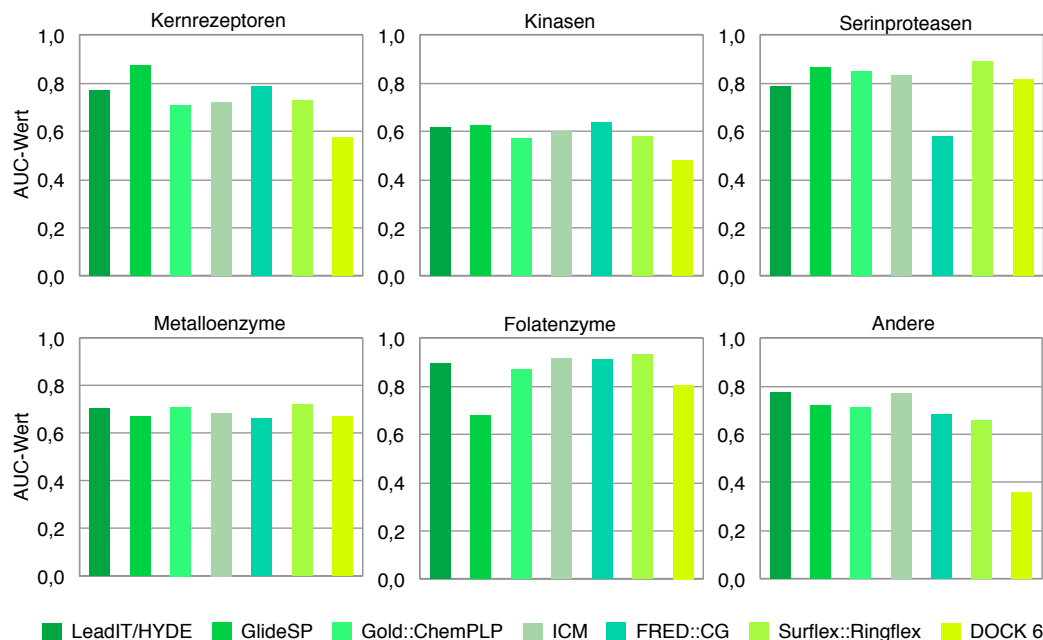


Abbildung 7.18: Vergleich der Resultate verschiedener Methoden im virtuellen Screening auf dem DUD Datensatz zusammengefasst nach Proteinfamilie. Dargestellt ist der mittlere AUC-Wert pro Proteinfamilie.

hat, ein besseres Resultat für die übrigen Methoden vorliegt. In Abbildung 7.18 sind die mittleren AUC-Werte pro Proteinfamilien der verschiedenen Methoden aufgetragen. Es wird deutlich, dass im Fall von Kinasen und Metalloenzymen alle Methoden ein ähnliches Resultat erzielen. Bei den Serinproteasen erreicht HYDE einen geringfügig niedrigeren mittleren AUC-Wert als andere Methoden, während es bei den Proteinen, die keiner Familie zugeordnet wurden (Andere), zusammen mit ICM das beste Resultat liefert. Die Folatenzyme führen bei fast allen Methoden, außer GlideSP und DOCK 6, zu einem sehr guten Ergebnis. In der letzte Gruppe, der Kernrezeptoren, kann GlideSP das beste Resultat aufweisen, wobei HYDE und FRED::CG in dieser Proteinfamilie ebenfalls gute Ergebnisse erzielen.

Im Kontext des virtuellen Screenings, liefert die HYDE-Bewertungsfunktion zusammen mit einem FlexX-Docking ähnlich gute Resultate wie die besten, bereits etablierten Methoden. Im Gegensatz zu diesen Methoden, ist die HYDE-Funktion zuvor nicht auf Protein-Ligand-Komplexen und experimentellen Bindungsaffinitäten kalibriert worden, sondern benötigt lediglich die logP-Werte kleiner Moleküle als Parameter, um die Bindungsenergie eines Protein-Ligand-Komplexes abzuschätzen.

7.5 Mittlere Laufzeit der HYDE-Bewertungsfunktion

Durch die komplexe Berechnung der Oberfläche in der HYDE-Bewertungsfunktion ist zur Abschätzung der Energie eines Protein-Ligand-Komplexes eine höhere mittlere Laufzeit zu erwarten, als z. B. bei der FlexX-Bewertungsfunktion, die aus einfacheren Termen besteht [80] und im Mittel eine Laufzeit von ca. 2 msec benötigt. Ein Ziel der Weiterentwicklung der HYDE-Bewertungsfunktion war eine Beschleunigung der Oberflächenberechnung, um so eine schnellere, geometrische Optimierung eines Protein-Ligand-Komplexes in Richtung der HYDE-Funktion zu ermöglichen.

Im Gegensatz zum HYDE-Prototypen (Kapitel 4.1) wird in der überarbeiteten Version der HYDE-Bewertungsfunktion die molekulare Oberfläche anstatt der gewichteten SAS berechnet. Dies wurde eingeführt, um besser enge Stellen im Protein zu repräsentieren, was als Problem des HYDE-Prototypen diskutiert wurde und durch eine sehr aufwendig zu berechnende Abwandlung der gewichteten Oberfläche (WRAY) in diesem realisiert wurde [202]. Durch eine neue Datenstruktur, dem Oberflächennetz, ist die Berechnung der Oberfläche auf Proteinseite, z. B. bei einem virtuellen Screening, sofern das Protein als starr angesehen wird, nur einmal notwendig und daher effizienter. Die mittlere Laufzeit der Generierung des Oberflächennetzes für den Liganden als auch für die Bindetasche des Proteins wurde anhand der 85 Komplexe des Astex Diverse Set Datensatzes [262] bestimmt (vgl. Tabelle 7.7). Die Generierung des Oberflächennetzes

für beide Moleküle ist mit einer mittleren Laufzeit von 956 msec um einen Faktor 478 langsamer als die Berechnung der gesamten FlexX-Bewertungsfunktion. Ein direkter Vergleich zum HYDE-Prototypen war leider nicht mehr möglich.

Bei der Berechnung der gesamten HYDE-Funktion nimmt die Generierung des Oberflächennetzes den Großteil der mittleren Laufzeit ein (89 %). Die Berechnung der Wasserstoffbrücken-Energie ist bezüglich der Laufzeit nicht relevant. Durch die Abschätzung der Dehydratationsenergie, die die Berechnung der Oberflächenänderung bei der Komplexbildung beinhaltet, entsteht hingegen eine mittlere Laufzeit von 2 msec. Die Zeit, die bei Initialisierung der Datenstrukturen zusätzlich zur Generierung des Oberflächennetzes benötigt wird, ist durch die Berechnung der internen Wasserstoffbrücken und der Wasserparameter $water_{overlap}$ und $water_{interaction}$ (vgl. Kapitel 6.1.1) auf Proteinseite bedingt. Somit kann in einem virtuellen Screening durch Vorberechnung der Oberfläche und der Wasserparameter der Protein-Bindetasche die HYDE-Energie eines Protein-Ligand-Komplexes im Mittel in 15 msec (Generierung des Oberflächennetzes für den Liganden und Berechnung der Dehydratationsenergie) abgeschätzt werden und ist damit nur noch um einen Faktor 7,5 langsamer als die FlexX-Bewertungsfunktion.

Zur geometrischen Optimierung eines Protein-Ligand-Komplexes in Richtung der HYDE-Funktion, wird die GeoHYDE-Funktion berechnet (vgl. Kapitel 6.4.1), die keine erneute Generierung des Oberflächennetzes des Liganden während der Optimierung erfordert. Daher kann ein lokales Minimum durch die numerische Optimierung (max. 50 Optimierungsschritte) im Mittel in 920 msec, inklusive anschließender Neuberechnung der HYDE-Energie, bestimmt werden. Die Suche nach dem globalen Minimum mithilfe der stochastischen Optimierung dauert, durch die erhöhte Anzahl der Optimierungsschritte (10.000), ungefähr sieben mal länger; die mittlere Laufzeit dabei beträgt 6,5 sec.

7. Resultate und Diskussion

Oberflächengenerierung		
Bindetasche	Generierung der Nachbarlisten	131
	Generierung des Oberflächennetzes	813
Ligand	Generierung der Nachbarlisten	0
	Generierung des Oberflächennetzes	13
Gesamt		956

Hyde-Energie Berechnung	
Initialisieren der Datenstrukturen	1069
Wasserstoffbrücken-Energie	0
Dehydratationsenergie	2
Gesamt	1071

Hyde-Optimierung		
Numerisch	max. 50 Schritte	920
Stochastisch	10 Läufe a 1000 Schritte	6507

Tabelle 7.7: Mittlere Laufzeit (in Millisekunden) der HYDE-Bewertungsfunktion. Gemittelt über alle 85 Komplexe des Astex Diverse Set. Die Analyse wurde auf einer Intel Core 2 Quad 2.66GHz Q9450 CPU unter einem Linux SUSE 11.3 Betriebssystem durchgeführt.

8

Weitere Anwendungen von HYDE

Die HYDE-Bewertungsfunktion wurde ursprünglich zur Abschätzung der Energie von Protein-Ligand-Komplexen im strukturbasierten Wirkstoffentwurf entwickelt. Im Gegensatz zu vielen anderen Bewertungsfunktionen in diesem Feld wurde HYDE nicht auf den Kristallstrukturen und experimentell bestimmten Affinitäten von Protein-Ligand-Komplexen kalibriert. Die plogP-Parameter, die HYDE zur Abschätzung der Wasserstoffbrücken- und der Dehydratationsenergie benötigt, wurden anhand von experimentellen logP-Werten kleiner Moleküle abgeleitet und sind atom-basiert, so dass HYDE sehr generell zur Abschätzung der Energie zwischen Molekülen angewendet werden kann. Die einzige Einschränkung besteht darin, dass angenommen wird, die Wechselwirkungen finden in einem wässrigen Medium statt.

In diesem Kapitel werden weitere Anwendungsgebiete der HYDE-Bewertungsfunktion aufgezeigt. Ähnlich zur Abschätzung der Bindungsenergie von potentiellen Inhibitoren im strukturbasierten Wirkstoffentwurf wurde HYDE in einem biotechnologischen Kontext zur Bewertung von Molekülen, die während einer enzymatischen Reaktion als Puffer dienen, eingesetzt. Dazu wurde die Affinität eines bestimmten Puffers zu einem Enzym in Beziehung zu der Affinität des Substrats gesetzt, um eine Beeinflussung der Aktivität des Enzyms durch den Puffer vorherzusagen. Des Weiteren wurde HYDE zur Berechnung von Protein-Protein-Interaktionen genutzt. Dabei wurde die Stabilität von Protein-Protein-Kontaktflächen mit HYDE abgeschätzt, um z.B. eine Aussage über

biologische oder nicht-biologische Kontakte in Kristallstrukturen treffen zu können. Im Folgenden werden diese Ansätze zusammengefasst und exemplarisch wiedergegeben.

8.1 HYDE im biotechnologischen Kontext

In einer kürzlich veröffentlichten Publikation von Schomburg *et al.* [282] wurde die HYDE-Bewertungsfunktion zur Vorhersage von kompetitiver Substrat-Hemmung durch Puffersubstanzen bei Enzymen eingesetzt. Dies ist eine wichtige biotechnologische Fragestellung, da in der Entwicklung von synthetischen Multi-Enzym-Stoffwechselwegen, in der die Aktivität jedes Enzyms maximiert werden soll, sich ein solcher Faktor als Engpass in der biotechnologischen Produktion auswirken würde. Die Aktivität eines Enzyms ist durch viele verschiedene Faktoren, wie z.B. Druck, Temperatur oder pH-Wert, bestimmt. Durch eine enzymatische Reaktion kann sich der pH-Wert des Mediums verändern, so dass Puffersubstanzen zugegeben werden, um über den gesamten Zeitraum des Prozesses einen optimalen pH-Wert für die Aktivität eines Enzyms zu garantieren. Bei der Auswahl des entsprechenden Puffers muss jedoch nicht nur darauf geachtet werden, dass dieser den optimalen pH-Bereich für das Enzym bereitstellt, sondern auch, dass die Puffersubstanz nicht mit dem Substrat um die Besetzung der Bindetasche des Enzyms konkurriert.

In einem neuen Ansatz haben Schomburg *et al.* mittels molekularem Docking und anschließender Nachbewertung durch HYDE ein computergestütztes Vorhersage-Protokoll entwickelt, das verschiedene Puffer bezüglich ihrer Affinität zu einem bestimmten Enzym testet. Dadurch können bestimmte Puffersubstanzen für die Optimierung von Stoffwechselreaktionen ausgeschlossen werden, ohne eine zeit- und kostenintensive Untersuchung im Labor durchzuführen. Das Verfahren wurde anhand verschiedener Fälle, die bereits in der Literatur beschrieben wurden, als auch mithilfe einer experimentellen Validierung, bestätigt [282]. Im Folgenden wird dieses Verfahren anhand eines Beispiels aus der Publikation [282] als eine weitere Anwendung von HYDE beschrieben.

Phosphoglucose-Isomerase

Die Phosphoglucose-Isomerase katalysiert sowohl in der Glycolyse die Umwandlung von Glucose-6-Phosphat in Fructose-6-Phosphat als auch in der Gluconeogenese die umgekehrte Reaktion von Fructose-6-Phosphat in Glucose-6-Phosphat (Abbildung 8.1 A). Eine Kristallstruktur der Phosphoglucose-Isomerase kokristallisiert mit Fructose-6-Phosphat (1HOX) wurde als Ausgangspunkt für das molekulare Docking verwendet. In der Bindetasche dieser Proteinstruktur wurden mithilfe eines FlexX-Dockings in der

LeadIT-Software jeweils bis zu 100 Konformationen von 14 verschiedenen Puffersubstanzen als auch des Substrats und des Produktes erzeugt. Anschließend wurden die Posen durch den HYDE-Optimierungskontext vorbereitet, bevor die Bindungsenergie wie im virtuellen Screening mit HYDE abgeschätzt wurde. Die HYDE-Energie, der jeweils am besten bewerteten Pose der Puffer, wurde durch die des Substrates normalisiert, so dass sich eine relative HYDE-Energie der Puffersubstanzen zum Substrat ergibt. Beträgt die relative HYDE-Energie eines Puffers mindestens 90 %, so wird der Puffer als kritisch angesehen und führt mit hoher Wahrscheinlichkeit zu einer kompetitiven Hemmung des Enzyms. Liegt die relative HYDE-Energie im Bereich von 75 - 90 %, so besteht immer noch eine höhere Wahrscheinlichkeit, dass der Puffer mit dem Substrat um die Besetzung der Bindetasche konkurriert. Bei einer relativen HYDE-Energie von weniger als 75 % wird davon ausgegangen, dass der Puffer keinen Einfluss auf die Aktivität eines Enzyms hat. Für vier der Puffersubstanzen (PIPES, TRIS, Diglycin und Karbonat)(Abbildung 8.1 A) wurde die von HYDE vorhergesagte Hypothese im Labor getestet.

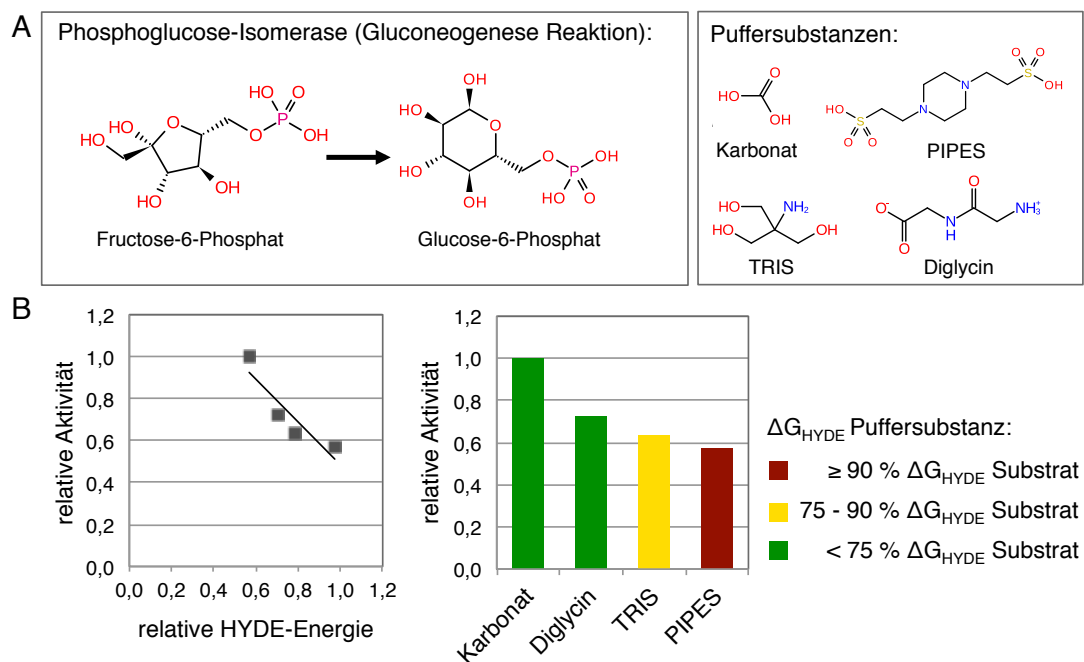


Abbildung 8.1: Kompetitive Hemmung der Phosphoglucose-Isomerase durch Puffersubstanzen. A) Links: Gluconeogenese Reaktion katalysiert durch Phosphoglucose-Isomerase. Rechts: Verschiedene Puffersubstanzen. B) Links: Korrelation der relativen Aktivität und der relativen HYDE-Energie. Rechts: Relative Aktivität beeinflusst durch die vier Puffer.

Abbildung 8.1 B zeigt das Resultat für die verschiedenen Puffersubstanzen. Die relative Aktivität, d.h. die Aktivität des Enzyms in Anwesenheit eines bestimmten Puffers normalisiert mit der höchsten Aktivität des Enzyms, korreliert sehr gut mit der relativen HYDE-Energie. In HYDE wird für PIPES eine Bindungsenergie von $\Delta G_{\text{HYDE}} = -36 \text{ kJ/mol}$ abgeschätzt, während sich für das Substrat eine HYDE-Energie von $\Delta G_{\text{HYDE}} = -37 \text{ kJ/mol}$ ergibt. Somit kann PIPES als kritisch für die Aktivität des Enzyms eingeschätzt werden. Wird PIPES im Medium als Puffer verwendet, so sinkt die experimentell bestimmte Aktivität des Enzyms auf 57 %. Die Puffersubstanz TRIS wird ebenfalls mit einer relativen HYDE-Energie von 78 % als potentieller Inhibitor der Phosphoglucose-Isomerase beurteilt. Dies kann mit einer Abnahme der gemessenen Aktivität um 37 % bestätigt werden. Diglycin, das als Puffersubstanz für Phosphoglucose-Isomerase-Assays empfohlen wird, liegt mit einer relativen HYDE-Energie von 70 % gerade nicht mehr im kritischen Bereich. Im Vergleich zum Karbonat-Puffer wird die Aktivität der Phosphoglucose-Isomerase jedoch um 28 % gesenkt. Der Karbonat-Puffer wurde auch von HYDE mit der niedrigsten Bindungsenergie von $\Delta G_{\text{HYDE}} = -21 \text{ kJ/mol}$ aller vier im Labor getesteten Puffer als beste Puffersubstanz für die Phosphoglucose-Isomerase eingeordnet.

Diese Anwendung von HYDE zeigt, dass die Bewertungsfunktion auch zu anderen Fragestellungen außerhalb des strukturbasierten Wirkstoffentwurfs genutzt werden kann.

8.2 Bewertung von Protein-Protein-Interaktionen

Die Interaktion von Proteinen ist in biochemischen Stoffwechselwegen und Signaltransduktionsprozessen essentiell. Anhand von Protein-Interaktionsdaten wurde abgeschätzt, dass das Protein-Interaktions-Netzwerk (Interaktom) im menschlichen Organismus einer Größe von etwa 650.000 verschiedenen Protein-Protein-Interaktionen entspricht [283]. Dabei existieren verschiedene Klassifikationen der Protein-Protein-Interaktionen, z. B. die Einteilung nach ihrer Lebensdauer in permanente und transiente Komplexe [284, 285]. Erstere entstehen meist gleich nach der Transkription der Proteine und zeichnen sich neben ihrer hohen Lebensdauer auch durch ihre große Stabilität aus. Transiente Protein-Protein-Interaktionen, z. B. die transienten Interaktionen von G-Proteinen, spielen hingegen durch ihre reversible Bindung vor allem in Signaltransduktionsprozessen eine wichtige Rolle. Daher sind die Proteine, die an diesen transienten Protein-Protein-Interaktionen beteiligt sind, auch interessante neue Ansatzpunkte in der Wirkstoffentwicklung [286]. Experimentelle Methoden zur Bestimmung der verschiedenen Protein-Protein-Interaktionen sind meist sehr zeitintensiv und können deshalb nur

in einem kleineren Maßstab durchgeführt werden, außerdem weisen sie teilweise hohe Falsch-Positiv-Raten auf [287, 288]. Aus diesem Grund wurden einige computergestützte Methoden zur Unterscheidung der verschiedenen Protein-Protein-Interaktionen entwickelt [289–297]. Diese benötigen zur Vorhersage des Typs der Interaktion meist eine dreidimensionale Struktur des Komplexes. In Kristallkomplexen findet sich ein zusätzlicher Typ von Protein-Protein-Interaktionen: Kristallkontakte. Diese werden im Kristallisationsprozess durch die Kristallpackung erzwungen und haben keine biologische Bedeutung [40, 298]. Zur Beurteilung von Kristallstrukturen ist daher auch eine Trennung zwischen biologisch relevanten und artifiziellen Proteinkomplexen erforderlich [299, 300].

In diesem Zusammenhang wurde ein Klassifizierungsverfahren für permanente und transiente Protein-Protein-Interaktionen als auch zur Identifizierung von Kristallkontakten mittels einer Support-Vector-Maschine (SVM) von Eva Vennmann in ihrer Bachelorarbeit entwickelt [301]. Dieses ist über einen Webserver frei verfügbar: ppi.zbh.uni-hamburg.de. Dabei wurde die HYDE-Bewertungsfunktion als einer der Deskriptoren zur Abschätzung der Stabilität von Protein-Protein Kontakten eingesetzt. Im Folgenden wird die Verwendung von HYDE im Kontext der Bewertung von Protein-Interaktionen anhand dieses Klassifizierungsverfahrens aufgezeigt.

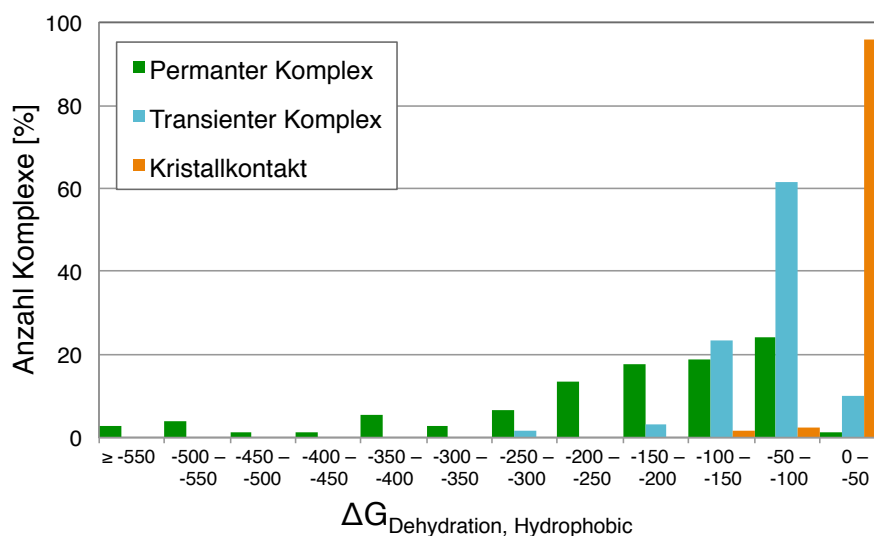


Abbildung 8.2: Bewertung von Protein-Protein-Kontaktflächen des Trainingsdatensatzes durch die hydrophobe Dehydratationsenergie der HYDE-Bewertungsfunktion. Diese ist für drei verschiedene Typen von Protein-Protein-Interaktionen aufgetragen: Permanente Komplexe, transiente Komplexe und Kristallkontakte.

8. Weitere Anwendungen von HYDE

Protein- und Ligand-Atome werden in HYDE gleich behandelt, daher kann die HYDE-Bewertungsfunktion direkt auf Proteine angewendet werden. Zur Klassifizierung von Protein-Protein-Interaktionen hat sich der hydrophobe Effekt der HYDE Funktion als vielversprechender Deskriptor erwiesen. Dazu wird die hydrophobe Dehydrationsenergie der Protein-Protein-Kontaktfläche eines Proteinkomplexes berechnet. In Abbildung 8.2 ist die Verteilung der hydrophoben Dehydrationsenergie der 254 Proteinkomplexe, die im Trainingsdatensatz [294] der SVM zur Trennung der verschiedenen Typen von Protein-Interaktionen enthalten waren, dargestellt. Daraus ist klar ersichtlich, dass sich biologisch relevante Komplexe in ihrer hydrophoben Dehydrationsenergie von artifiziellen Kristallkontakten eindeutig unterscheiden. Abbildung 8.2 zeigt auch, dass permanente Komplexe im Durchschnitt einen stärkeren hydrophoben Effekt ($\langle \Delta G_{dehydration}^{hydrophobic} = -204 \text{ kJ/mol} \rangle$) als transiente Komplexe ($\langle \Delta G_{dehydration}^{hydrophobic} = -88 \text{ kJ/mol} \rangle$) aufweisen, jedoch ist hier die Trennung der beiden Typen nicht so deutlich.

Die hydrophobe Dehydrationsenergie von HYDE wird in einer zweistufigen SVM als Deskriptor eingesetzt, wobei in der ersten Stufe biologisch relevante von artifiziellen Komplexen getrennt werden und in der zweiten Stufe zwischen permanenten und transienten Protein-Interaktionen unterschieden wird. Abbildung 8.3 zeigt schematisch den Aufbau der zweistufigen SVM, sowie die Resultate für die hydrophobe Dehydrationsenergie als einzigen Deskriptor in beiden Stufen.

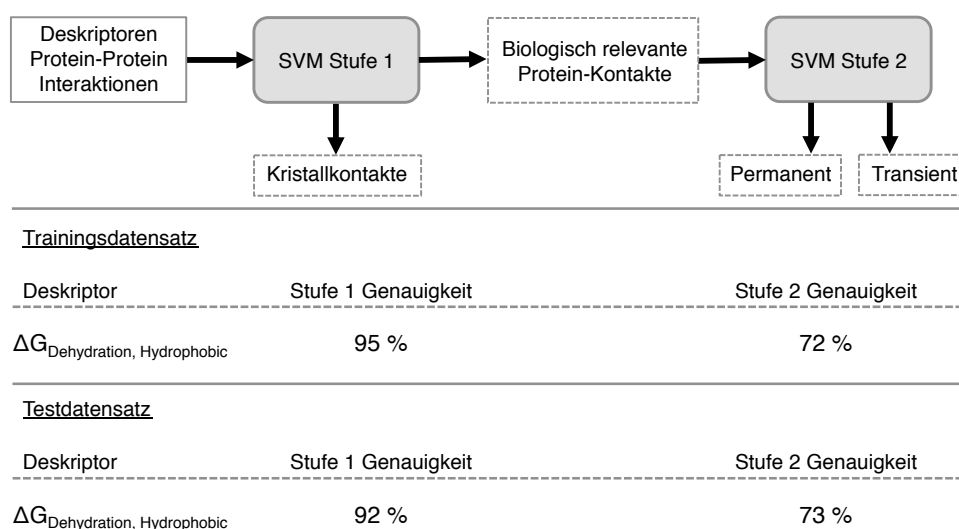


Abbildung 8.3: Schematische Darstellung der zweistufigen SVM und Resultate für die hydrophobe Dehydrationsenergie von HYDE als einziger Deskriptor in beiden Stufen. Die Resultate sind für den Trainings- und Testdatensatz angegeben.

In der ersten Stufe wurde mithilfe dieses einen Deskriptors innerhalb des Trainingsdatensatzes eine Genauigkeit von 95 % bei der Klassifizierung von biologisch relevanten und artifiziellen Protein-Kontakten erreicht. Bei der Trennung von permanenten und transienten Proteinkomplexen sinkt die Vorhersagegenauigkeit jedoch auf 72 %. Das Verfahren wurde auf einem unabhängigen Testdatensatz aus 152 Proteinkomplexen [292, 302–305] angewendet und erzielte auch hierbei in 92 % die richtige Einteilung in biologisch relevant und artifiziell. Die Klassifizierung von permanenten und transienten Protein-Interaktionen in der zweiten Stufe der SVM, ist auch bei diesem Datensatz nur in 73 % der Fälle richtig.

Durch die guten Ergebnisse in der Unterscheidung von biologisch relevanten und artifiziellen Protein-Kontakten, wurde die hydrophobe Dehydratationsenergie von HYDE als einziger Deskriptor in der ersten Stufe des Klassifizierungsverfahrens genutzt. Im Gegensatz zur anderen Klassifizierungsverfahren [294, 297], die ebenfalls diese zweistufige Form der Klassifizierung verwenden, benötigen diese jedoch deutlich mehr Deskriptoren, um eine ähnliche Vorhersagequalität zu erzielen.

Im Kontext der Bewertung von Protein-Protein-Interaktionen wurden mit HYDE bereits auch einzelne Aminosäuren innerhalb eines Proteins bewertet, um den Einfluss einer Mutation bestimmter Aminosäuren abzuschätzen oder um die Stabilität von Sekundärstrukturen in Proteinen zu beurteilen. Hierbei wurden ebenfalls vielversprechende Ergebnisse erzielt, so dass HYDE zukünftig auch in diesen Anwendungen eingesetzt werden kann.

9

Fazit und Ausblick

Im Rahmen dieser Dissertation wurde die Weiterentwicklung der HYDE-Bewertungsfunktion, die zur Abschätzung der Bindungsenergie von Protein-Ligand-Komplexen entwickelt wurde, beschrieben. Dazu wurden verschiedene Aspekte der HYDE-Funktion überarbeitet und neue Terme, Parameter und Berechnungsverfahren integriert und implementiert. Das theoretische Konzept von HYDE, demzufolge der Beitrag von intermolekularen Wechselwirkungen in wässriger Lösung direkt von den Eigenschaften des Wasserstoffbrücken-Netzwerkes von Wasser selbst abgeleitet werden kann, wurde dabei erhalten. Die HYDE-Funktion besteht nach wie vor aus zwei Termen zur Beschreibung der Wasserstoffbrücken- und der Dehydratationsenergie. Die Modellierung und Berechnung dieser Terme wurde hingegen weiterentwickelt.

Zur Berechnung der Zugänglichkeitsänderung wurde ein neuer effizienterer Algorithmus implementiert, der diese Änderung mithilfe eines zuvor generierten Oberflächennetzes ermittelt. Die Änderung der Oberfläche bei der Komplexbildung wird nicht mehr wie zuvor über die lösungsmittelzugängliche Oberfläche abgeschätzt, sondern durch die molekulare Oberfläche. Durch die Veränderung der Oberflächenberechnung wurde eine Re-Parametrisierung der atombasierten logP-Inkrementen benötigt, die in HYDE zur Bestimmung der Beiträge von Wasserstoffbrücken-Interaktionen als auch zur Quantifizierung der Dehydratationsenergien verwendet werden. Im Zuge dieser Re-Parametrisierung wurde eine Studie zur Untersuchung des Zusammenhangs von logP-Wert und

der Oberfläche eines Moleküls durchgeführt. Diese Studie resultierte nicht nur in einer deutlichen Reduzierung der verwendeten logP-Atomtypen und neuen Inkrementen, sondern führte auch zu einer neuen Modellierung der Wasserstoffbrücken- und Dehydratationsbeiträge von polaren Atomen mit multiplen Wasserstoffbrücken-Funktionen. Weiterhin wurden neue Parameter zur Beschreibung von Wassern im ungebundenen Zustand in die HYDE-Funktion integriert. In HYDE werden Wasser, wenn sie nicht explizit im Protein-Ligand-Komplex vorhanden sind (Kristallwasser), zurzeit nur implizit berücksichtigt, daher liefern diese Parameter auch eine Korrektur dieser impliziten Beschreibung, in die nicht alle Effekte miteinbezogen werden können. Des Weiteren wurde ein zusätzlicher Term zur Abschätzung des Beitrags von Metallionen zur Bindungsenergie entwickelt. Diese werden jetzt ähnlich zu Wasserstoffbrücken in HYDE behandelt, wobei ihr Beitrag deutlich über dem einer Wasserstoffbrücke liegt. Zudem wird die unterschiedliche Koordinierungsgeometrie von Metallionen in HYDE berücksichtigt. Die HYDE-Bewertungsfunktion wurde in ein Optimierungsverfahren eingebettet, so dass ein Protein-Ligand-Komplex, unabhängig davon wie er erzeugt wurde, optimal auf die Bewertung mit HYDE vorbereitet werden kann. Dabei wird sowohl das Wasserstoffbrücken-Netzwerk des Komplexes optimiert als auch die Komplex-Geometrie mithilfe der entwickelten GeoHYDE-Funktion in Richtung der HYDE-Funktion minimiert.

Die Verwendung von logP-Parametern zur Kalibrierung von HYDE ermöglicht eine Quantifizierung der ungünstigen Beiträge der polaren Desolvatisierung und führt zu einer generelleren Anwendbarkeit von HYDE. Die Vorhersagequalität von HYDE ist dadurch nicht wie bei vielen anderen Bewertungsfunktionen an bestimmte Protein-Ligand-Komplexe gebunden, die auch in der Parametrisierung verwendet wurden. Daher können neben der Abschätzung der Bindungsstärke von Inhibitoren im strukturbasierten Wirkstoffentwurf auch andere Fragestellungen mit HYDE behandelt werden, z. B. die in Kapitel 8 beschriebene Bewertung von Puffersubstanzen oder von Protein-Protein-Interaktionen. Weiterhin wird durch die Verwendung von einfachen logP-Daten anstatt von Protein-Ligand-Strukturen und Bindungsaffinitäten die Anzahl der Fehlerquellen, die in die Kalibrierung eingehen, reduziert. Ein weiterer Vorteil der HYDE-Funktion sind ihre atombasierten intuitiven Beiträge, die zur Analyse von Protein-Ligand-Komplexen hilfreich sind, was in zahlreichen Beispielen in dieser Arbeit demonstriert wurde. Die Berücksichtigung der ungünstigen Beiträge der polaren Desolvatisierung führt zu einer deutlichen Reduzierung von falsch-positiven Molekülen im virtuellen Screening. Dies wurde bereits eindrucksvoll in einer prospektiven Studie des HYDE-Prototypen gezeigt. Insgesamt ist HYDE sehr sensitiv bei der Bewertung eines Protein-Ligand-Komplexes.

Leichte Veränderungen in der Komplexstruktur werden durch eine genaue Modellierung der Wasserstoffbrücken-Geometrien registriert. Dies ermöglicht z. B. das Erkennen von Schwankungen in der Aktivität durch kleinere Strukturveränderungen (*activity cliffs*) als auch das Auffinden von Fehlern und Ungenauigkeiten in Kristallstrukturen.

Diese Sensitivität führt jedoch zu einer Problematik beim Umgang mit ungenaueren oder fehlerhaften Strukturen. Gerade die Verwendung von öffentlich verfügbaren Datensätzen, die zum Teil solche Strukturen enthalten, ist schwierig mit HYDE. Dies wurde bereits in der Validierung angesprochen. Im Bezug auf diese Strukturen führt die Bestrafung von geometrisch schlechteren Wasserstoffbrücken in HYDE eher zu einem Rauschen als zu einer zuverlässigen Vorhersage. Andere Bewertungsfunktionen, die besser mit diesen Strukturen umgehen können, wurden meist auf Protein-Ligand-Komplex-Daten trainiert und berücksichtigen daher intrinsisch diese Ungenauigkeiten. Um in HYDE auch solche Strukturen verwenden zu können, ist eine bessere Aufbereitung und Optimierung der Strukturen vor der Bewertung nötig. Dazu müsste das Optimierungsverfahren weiterentwickelt werden, wobei sich als erster Schritt die Überarbeitung der Parametrisierung und Gewichtung der einzelnen Terme lohnen würde. Die Integration von Proteinflexibilität während dem Optimierungsverfahren würde mit großer Wahrscheinlichkeit zu einer weiteren Verbesserung der Handhabung von Ungenauigkeiten in Kristallstrukturen führen. Die Energielandschaft der HYDE-Funktion ist eher zerklüftet und macht damit gerade eine numerische Optimierung schwierig, daher könnte eine Modellierung von weicheren Potentialen in der HYDE-Funktion, z. B. zur Berechnung der Abweichung von der optimalen Wechselwirkungsgeometrie, das Auffinden eines globalen Minimums in der Optimierung erleichtern.

In HYDE wird die gleiche Konformation des Liganden im ungebunden und im gebundenen Zustand verwendet. Dies kann einen Fehler bei der Berechnung der Dehydratationsenergie verursachen, da diese von der Oberfläche und damit auch von der Konformation des Liganden abhängt. Außerdem wird die Entropieänderung des Liganden, die durch unterschiedliche Konformationen im ungebunden und im gebundenen Zustand hervorgerufen wird, nicht beachtet. Zurzeit wird davon ausgegangen, dass diese Konformationen sich nicht so stark unterscheiden, da die Ligandkonformation auch im ungebundenen Zustand Restriktionen durch das Wassernetzwerk unterworfen ist. Jedoch sollte dies in einer Studie validiert werden, indem die Konformation des Liganden in wässriger Lösung erzeugt und mit der Konformation im Komplex verglichen wird. Nur so kann der wirkliche Fehler, der durch die Verwendung der gleichen Konformation entsteht abgeschätzt und gegebenenfalls korrigiert werden.

Weiterhin werden nicht alle Effekte und intermolekularen Wechselwirkungen in HYDE modelliert, wie z. B. die Kation- π -Interaktionen oder Halogen-Brücken. Diese Interaktionen treten jedoch nicht so häufig in Protein-Ligand-Komplexen auf, so dass eine spezielle Modellierung erst sinnvoll ist, wenn diese Wechselwirkungen auch theoretisch besser erforscht sind. Eine verbesserte Bewertung von Kristallwassern in HYDE ist dringlicher, da diese momentan als Teil der Proteinstruktur aufgefasst werden und so ungünstige Konstellationen nicht genügend berücksichtigt werden. Die explizite Platzierung von nicht vorhandenen Wassern im Protein-Ligand-Komplex, z. B. im virtuellen Screening, und ihre energetische Bewertung würde weiterhin zu einer genaueren Vorhersage der Bindungsenergie beitragen.

Die HYDE-Bewertungsfunktion zeigt sich als innovativer Ansatz zur Abschätzung der Energie molekularer Interaktionen zwischen Biomolekülen. Sie kann zur Trennung von aktiven und inaktiven Molekülen im virtuellen Screening eingesetzt werden als auch zur detaillierten Analyse einzelner Komplexe. Trotz der Erweiterungsmöglichkeiten die im vorherigen Abschnitt diskutiert wurden, zeigt HYDE auch jetzt schon sehr vielversprechende Resultate in den verschiedenen Anwendungsgebieten und ist dadurch vergleichbar mit den bereits gut etablierten Protein-Ligand-Bewertungsfunktionen.

Literaturverzeichnis

- [1] B. Munos. Lessons from 60 years of pharmaceutical innovation. *Nat. Rev. Drug Discovery*, 8(12):959–968, 2009. 1
- [2] B. K. Shoichet. Virtual screening of chemical libraries. *Nature*, 432(7019):862–865, 2004. 2, 11, 47
- [3] G. Klebe. Virtual ligand screening: strategies, perspectives and limitations. *Drug Discovery Today*, 11(13):580–594, 2006. 2, 11, 38, 47, 48
- [4] C. Sotriffer. Virtual Screening: Principles, Challenges, and Practical Guidelines. In R. Mannhold, H. Kubinyi, and G. Folkers, editors, *Methods and Principles in Medicinal Chemistry*, volume 130. Wiley-VCH, 2011. 2, 34, 38, 46, 47
- [5] P. Ferrara, H. Gohlke, D. J. Price, G. Klebe, and C. L. Brooks. Assessing scoring functions for protein-ligand interactions. *J. Med. Chem.*, 47(12):3032–3047, 2004. 2, 34, 44, 49
- [6] G. L. Warren, C. W. Andrews, A. M. Capelli, B. Clarke, J. LaLonde, M. H. Lambert, M. Lindvall, N. Nevins, S. F. Semus, S. Senger, G. Tedesco, I. D. Wall, J. M. Woolven, C. E. Peishoff, and M. S. Head. A critical assessment of docking programs and scoring functions. *J. Med. Chem.*, 49(20):5912–5931, 2006. 38
- [7] T. Cheng, X. Li, Y. Li, Z. Liu, and R. Wang. Comparative assessment of scoring functions on a diverse test set. *J. Chem. Inf. Model.*, 49(4):1079–1093, 2009. 36, 38, 40, 49, 128, 130
- [8] R. D. Smith, J. B. Dunbar, P. M. U. Ung, E. X. Esposito, C. Y. Yang, S. Wang, and H. A. Carlson. CSAR Benchmark Exercise of 2010: Combined evaluation across all submitted scoring functions. *J. Chem. Inf. Model.*, 51(9):2115–2131, 2011. 2, 34, 44
- [9] G. Klebe. Am Anfang stand der glückliche Zufall. In *Wirkstoffdesign*, pages 23–31. Spektrum Akademischer Verlag, 2009. 7
- [10] C. R. Beddell, P. J. Goodford, F. E. Norrington, S. Wilkinson, and R. Wootton. Compounds designed to fit a site of known structure in human haemoglobin. *Br. J. Pharmacol.*, 57(2):201, 1976. 7

- [11] W. L. Jorgensen. The many roles of computation in drug discovery. *Science*, 303(5665):1813–1818, 2004. 8, 45, 46
- [12] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne. The protein data bank. *Nucleic Acids Res.*, 28(1):235–242, 2000. 10, 72
- [13] R. P. Hertzberg and A. J. Pope. High-throughput screening: new technology for the 21st century. *Curr. Opin. Chem. Biol.*, 4(4):445–451, 2000. 10
- [14] L. B. Kier. Molecular orbital calculation of preferred conformations of acetylcholine, muscarine, and muscarone. *Mol. Pharmacol.*, 3(5):487–494, 1967. 10
- [15] H. Koeppen, J. Kriegl, U. Lessel, C. S. Tautermann, and B. Wellenzohn. Ligand-Based Virtual Screening. In C. A. Sotriffer, editor, *Virtual Screening: Principles, Challenges, and Practical Guidelines*, pages 61–85. Wiley-VCH Verlag GmbH & Co. KGaA, 2011. 10
- [16] R. Lahana. How many leads from HTS? *Drug Discovery Today*, 4(10):447 – 448, 1999. 11
- [17] H. J. Boehm, M. Boehringer, D. Bur, H. Gmuender, W. Huber, W. Klaus, D. Kostrewa, H. Kuehne, T. Luebbers, N. Meunier-Keller, and F. Mueller. Novel inhibitors of DNA gyrase: 3D structure based biased needle screening, hit validation by biophysical methods, and 3D guided optimization. A promising alternative to random screening. *J. Med. Chem.*, 43(14):2664–2674, 2000. 11
- [18] G. Klebe. Die Optimierung der Leitstruktur. In *Wirkstoffdesign*, pages 113–124. Spektrum Akademischer Verlag, 2009. 11
- [19] H. Kubinyi. QSAR and 3D QSAR in drug design Part 1: methodology. *Drug Discovery Today*, 2(11):457–467, 1997. 12, 37
- [20] C. Hansch. Quantitative approach to biochemical structure-activity relationships. *Acc. Chem. Res.*, 2(8):232–239, 1969. 12, 37, 75
- [21] S. Venkatesh and R. A. Lipper. Role of the development scientist in compound lead selection and optimization. *J. Pharm. Sci.*, 89(2):145–154, 2000. 12
- [22] C. A. Lipinski, F. Lombardo, B. W. Dominy, and P. J. Feeney. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Delivery Rev.*, 46(1-3):3–26, 2001. 12
- [23] G. Klebe. Experimentelle Methoden zur Strukturauflklärung. In *Wirkstoffdesign*, pages 189–205. Spektrum Akademischer Verlag, 2009. 14
- [24] R. J. Read. Improved Fourier coefficients for maps using phases from partial structures with errors. *Acta Crystallogr., Sect. A*, 42(3):140–149, 1986. 14

- [25] P. Emsley and K. Cowtan. Coot: model-building tools for molecular graphics. *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 60(12):2126–2132, 2004. 15
- [26] G. J. Kleywegt and T. Alwyn Jones. Model building and refinement practice. *Methods in enzymology*, 277:208–230, 1997. 16
- [27] A. T. Brünger. Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature*, 355(6359):472–475, 1992. 16
- [28] E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, and T. E. Ferrin. UCSF ChimeraÑa visualization system for exploratory research and analysis. *J. Comput. Chem.*, 25(13):1605–1612, 2004. 17
- [29] A. M. Davis, S. J. Teague, and G. J. Kleywegt. Application and Limitations of X-ray Crystallographic Data in Structure-Based Ligand and Drug Design. *Angew. Chem., Int. Ed.*, 42(24):2718–2736, 2003. 18, 19, 20
- [30] A. M. Davis, S. A. St-Gallay, and G. J. Kleywegt. Limitations and lessons in the use of X-ray structural information in drug design. *Drug Discovery Today*, 13(19-20):831–841, 2008. 18
- [31] M. T. Stubbs, S. Reyda, F. Dullweber, M. Möller, G. Klebe, D. Dorsch, W. W. K. R. Mederski, and H. Wurziger. pH-Dependent Binding Modes Observed in Trypsin Crystals: Lessons for Structure-Based Drug Design. *ChemBioChem*, 3(2-3):246–249, 2002. 18
- [32] H. Yang, M. Yang, Y. Ding, Y. Liu, Z. Lou, Z. Zhou, L. Sun, L. Mo, S. Ye, H. Pang, G. F. Gao, A. Kanchan, R. Bartlam, M. Hilgenfeld, and Z. Rao. The crystal structures of severe acute respiratory syndrome virus main protease and its complex with an inhibitor. *Proc. Natl. Acad. Sci. U. S. A.*, 100(23):13190, 2003. 18
- [33] H. Steuber, M. Zentgraf, C. Gerlach, C. A. Sotriffer, A. Heine, and G. Klebe. Expect the unexpected or caveat for drug designers: multiple structure determinations using aldose reductase crystals treated under varying soaking and co-crystallisation conditions. *J. Mol. Biol.*, 363(1):174–187, 2006. 18
- [34] R. A. Laskowski, M. W. MacArthur, D. S. Moss, and J. M. Thornton. Procheck: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.*, 26(2):283–291, 1993. 19
- [35] R. W. W. Hooft, G. Vriend, C. Sander, and E. E. Abola. Errors in protein structures. *Nature*, 381:272–272, 1996.
- [36] I. W. Davis, A. Leaver-Fay, V. B. Chen, J. N. Block, G. J. Kapral, X. Wang, L. W. Murray, W. B. Arendall, J. Snoeyink, J. S. Richardson, and D. C. Richardson. Molprobity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.*, 35(suppl 2):W375–W383, 2007.

- [37] T. Lippert and M. Rarey. Fast automated placement of polar hydrogen atoms in protein-ligand complexes. *J. Cheminformatics*, 1(1):1–12, 2009. 19, 72
- [38] J.W.M. Nissink, C. Murray, M. Hartshorn, M.L. Verdonk, J.C. Cole, and R. Taylor. A new test set for validating predictions of protein–ligand interaction. *Proteins: Struct., Funct., Bioinf.*, 49(4):457–471, 2002. 20
- [39] J. C. Cole, O. Korb, T. S. G. Olsson, and J. Liebeschuetz. *The Basis for Target-Based Virtual Screening: Protein Structures*, pages 87–113. Wiley-VCH Verlag GmbH & Co. KGaA, 2011. 20
- [40] O. Carugo and P. Argos. Protein-protein crystal-packing contacts. *Protein Sci.*, 6(10):2261–2263, 1997. 20, 163
- [41] J. D. Dunitz. Win some, lose some: enthalpy-entropy compensation in weak intermolecular interactions. *Chem. Biol.*, 2(11):709–712, 1995. 23
- [42] H. Gohlke and G. Klebe. Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptors. *Angew. Chem., Int. Ed.*, 41(15):2644–2676, 2002. 23, 28, 33, 34, 38, 42, 43, 45
- [43] L. C. Pauling. *The Nature of the Chemical Bond and the Structure of Molecules and Crystals: An Introduction to Modern Structural Chemistry*. Cornell Univ. Press, Ithaca, N. Y., 3rd edition, 1960. 24, 52
- [44] C. Bissantz, B. Kuhn, and M. Stahl. A Medicinal Chemist’s Guide to Molecular Interactions. *J. Med. Chem.*, 53(14):5061–5084, 2010. 24, 25, 26, 27, 38, 43
- [45] E. Arunan, G.R. Desiraju, R.A. Klein, J. Sadlej, S. Scheiner, I. Alkorta, D. C. Clary, R. H. Crabtree, J. J. Dannenberg, P. Hobza, H. G. Kjaergaard, Legon A. C., B. Mennucci, and D. J. Nesbitt. Definition of the hydrogen bond (IUPAC Recommendations 2011). *Pure Appl. Chem.*, 83(8):1637, 2011. 24, 55
- [46] A. R. Fersht, J. P. Shi, J. Knill-Jones, D. M. Lowe, A. J. Wilkinson, D. M. Blow, P. Brick, P. Carter, M. M. Y. Waye, and G. Winter. Hydrogen bonding and biological specificity analysed by protein engineering. *Nature*, 314:235–238, 1985. 24, 84
- [47] G. Lange, D. Lesuisse, P. Deprez, B. Schoot, P. Loenze, D. Bénard, J. P. Marquette, P. Broto, E. Sarubbi, and E. Mandine. Requirements for specific binding of low affinity inhibitor fragments to the SH2 domain of pp60Src are identical to those for high affinity binding of full length inhibitors. *J. Med. Chem.*, 46(24):5184–5195, 2003. 24, 58
- [48] S. Sarkhel and G. R. Desiraju. N-H...O, O-H...O, and C-H...O hydrogen bonds in protein–ligand complexes: Strong and weak interactions in molecular recognition. *Proteins: Struct., Funct., Bioinf.*, 54(2):247–259, 2004. 24

- [49] S. K. Panigrahi and G. R. Desiraju. Strong and weak hydrogen bonds in the protein-ligand interface. *Proteins: Struct., Funct., Bioinf.*, 67(1):128–141, 2007. 24
- [50] Gerhard Klebe. Protein-Ligand-Wechselwirkungen als Grundlage der Arzneimittelwirkung. In *Wirkstoffdesign*, pages 49–67. Spektrum Akademischer Verlag, 2009. 24, 26, 27
- [51] B. Seebeck, I. Reulecke, A. Kämper, and M. Rarey. Modeling of metal interaction geometries for protein-ligand docking. *Proteins Struct. Funct. Bioinform.*, 71:1237–1254, 2008. 26, 43, 103
- [52] R. B. Hermann. Theory of hydrophobic bonding. II. Correlation of hydrocarbon solubility in water with solvent cavity surface area. *J. Phys. Chem.*, 76(19):2754–2759, 1972. 26, 78
- [53] D. Eisenberg and A. D. McLachlan. Solvation energy in protein folding and binding. *Nature*, 319:199–203, 1986. 86
- [54] N. T. Southall, K. A. Dill, and A. D. J. Haymet. A view of the hydrophobic effect. *J. Phys. Chem. B*, 106(3):521–533, 2002. 26
- [55] S. Tsuzuki, K. Honda, T. Uchamaru, M. Mikami, and K. Tanabe. Origin of Attraction and Directionality of the π/π Interaction: Model Chemistry Calculations of Benzene Dimer Interaction. *J. Am. Chem. Soc.*, 124(1):104–112, 2002. 26
- [56] J. P. Gallivan and D. A. Dougherty. Cation- π interactions in structural biology. *Proc. Natl. Acad. Sci. U. S. A.*, 96(17):9459–9464, 1999. 27
- [57] N. Ramasubbu, R. Parthasarathy, and P. Murray-Rust. Angular preferences of intermolecular forces around halogen centers: Preferred directions of approach of electrophiles and nucleophiles around carbon-halogen bond. *J. Am. Chem. Soc.*, 108(15):4308–4314, 1986. 27
- [58] Y. Mo, G. Subramanian, J. Gao, and D. M. Ferguson. Cation- π Interactions: An Energy Decomposition Analysis and Its Implication in δ -Opioid Receptor-Ligand Binding. *J. Am. Chem. Soc.*, 124(17):4832–4837, 2002. 27
- [59] L. M. Salonen, C. Bucher, D. W. Banner, W. Haap, J.-L. Mary, J. Benz, O. Kuster, P. Seiler, W. B. Schweizer, and F. Diederich. Cation- π Interactions at the Active Site of Factor Xa: Dramatic Enhancement upon Stepwise N-Alkylation of Ammonium Ions. *Angew. Chem., Int. Ed.*, 48(4):811–814, 2009. 27
- [60] J. P. M. Lommerse, A. J. Stone, R. Taylor, and F. H. Allen. The nature and geometry of intermolecular interactions between halogens and oxygen or nitrogen. *J. Am. Chem. Soc.*, 118(13):3108–3116, 1996. 27

- [61] R. A. Laskowski, J. M. Thornton, C. Humblet, and J. Singh. X-SITE: Use of Empirically Derived Atomic Packing Preferences to Identify Favourable Interaction Regions in the Binding Sites of Proteins. *J. Mol. Biol.*, 259(1):175 – 201, 1996. 27
- [62] M. L. Verdonk, J. C. Cole, and R. Taylor. SuperStar: a knowledge-based approach for identifying interaction sites in proteins. *J. Mol. Biol.*, 289(4):1093–1108, 1999.
- [63] I. Muegge, Y.C. Martin, P.J. Hajduk, and S.W. Fesik. Evaluation of pmf scoring in docking weak ligands to the fk506 binding protein. *J. Med. Chem.*, 42(14):2498–2503, 1999. 39
- [64] H. Gohlke, M. Hendlich, and G. Klebe. Knowledge-based scoring function to predict protein-ligand interactions. *J. Mol. Biol.*, 295(2):337–356, 2000. 40, 75, 76
- [65] Z. Deng, C. Chuaqui, and J. Singh. Structural interaction fingerprint (SIFt): a novel method for analyzing three-dimensional protein-ligand binding interactions. *J. Med. Chem.*, 47(2):337–344, 2004. 27
- [66] I. J. Bruno, J. C. Cole, J. P. M. Lommerse, R. S. Rowland, R. Taylor, and M. L. Verdonk. IsoStar: a library of information about nonbonded interactions. *J. Comput. Aided Mol. Des.*, 11(6):525–537, 1997. 27
- [67] J. E. J. Mills and P. M. Dean. Three-dimensional hydrogen-bond geometry and probability information from a crystal survey. *J. Comput. Aided Mol. Des.*, 10(6):607–622, 1996.
- [68] M. Hendlich, A. Bergner, J. Günther, and G. Klebe. Relibase: design and development of a database for comprehensive analysis of protein–ligand interactions. *J. Mol. Biol.*, 326(2):607–620, 2003. 40
- [69] Z. Liu, G. Wang, Z. Li, and R. Wang. Geometrical Preferences of the Hydrogen Bonds on Protein-Ligand Binding Interface Derived from Statistical Surveys and Quantum Mechanics Calculations. *J. Chem. Theory Comput.*, 4(11):1959–1973, 2008. 27
- [70] P. J. Goodford. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J. Med. Chem.*, 28(7):849–857, 1985. 27
- [71] M. Baroni, G. Cruciani, S. Sciabola, F. Perruccio, and J. S. Mason. A common reference framework for analyzing/comparing proteins and ligands. Fingerprints for Ligands and Proteins (FLAP): theory and application. *J. Chem. Inf. Model.*, 47(2):279–294, 2007. 27
- [72] D. H. Williams, E. Stephens, D. P. O’Brien, and M. Zhou. Understanding Noncovalent Interactions: Ligand Binding Energy and Catalytic Efficiency from Ligand-Induced Reductions in Motion within Receptors and Enzymes. *Angew. Chem., Int. Ed.*, 43(48):6596–6616, 2004. 27

- [73] L. Muley, B. Baum, M. Smolinski, M. Freindorf, A. Heine, G. Klebe, and D. G. Hangauer. Enhancement of hydrophobic interactions and hydrogen bond strength by cooperativity: synthesis, modeling, and molecular dynamics simulations of a congeneric series of thrombin inhibitors. *J. Med. Chem.*, 53(5):2126–2135, 2010. 27
- [74] D. Eisenberg and W. Kauzmann. *The Structure and Properties of Water*. Oxford University Press, 1969. 27
- [75] G. Schneider and U. Fechner. Computer-based de novo design of drug-like molecules. *Nat. Rev. Drug Discovery*, 4(8):649–663, 2005. 28
- [76] J. Degen and M. Rarey. FlexNovo: Structure-Based Searching in Large Fragment Spaces. *ChemMedChem*, 1(8):854–868, 2006. 28
- [77] I. S. Moreira, P. A. Fernandes, and M. J. Ramos. Protein–protein docking dealing with the unknown. *J. Comput. Chem.*, 31(2):317–342, 2010. 28
- [78] M. van Dijk and A. M. J. J. Bonvin. Pushing the limits of what is achievable in protein–DNA docking: benchmarking HADDOCK’s performance. *Nucleic Acids Res.*, 38(17):5634–5647, 2010. 28
- [79] I. D. Kuntz, J. M. Blaney, S. J. Oatley, R. Langridge, and T. E. Ferrin. A geometric approach to macromolecule–ligand interactions. *J. Mol. Biol.*, 161(2):269–288, 1982. 29
- [80] M. Rarey, B. Kramer, T. Lengauer, and G. Klebe. A fast flexible docking method using an incremental construction algorithm. *J. Mol. Biol.*, 261:470–489, 1996. 29, 37, 63, 66, 72, 96, 133, 156, 204
- [81] T. J. A. Ewing, S. Makino, A. G. Skillman, and I. D. Kuntz. DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases. *J. Comput. Aided Mol. Des.*, 15(5):411–428, 2001. 29
- [82] A. N. Jain. Surflex: fully automatic flexible molecular docking using a molecular similarity-based search engine. *J. Med. Chem.*, 46(4):499–511, 2003. 29, 43
- [83] Z. Zsoldos, D. Reid, A. Simon, B. S. Sadjad, and A. P. Johnson. eHiTS: an innovative approach to the docking and scoring function problems. *Curr. Protein Pept. Sci.*, 7(5):421–436, 2006. 29, 30, 43
- [84] G. Klebe and T. Mietzner. A fast and efficient method to generate biologically relevant conformations. *J. Comput. Aided Mol. Des.*, 8(5):583–606, 1994. 30
- [85] J. Sadowski and J. Boström. MIMUMBA revisited: Torsion angle rules for conformer generation derived from X-ray structures. *J. Chem. Inf. Model.*, 46(6):2305–2309, 2006. 30

- [86] M. Totrov and R. Abagyan. Flexible protein-ligand docking by global energy optimization in internal coordinates. *Proteins: Struct., Funct., Genet.*, 29(1):215–220, 1997. 30
- [87] G. Jones, P. Willett, and R. C. Glen. Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation. *J. Mol. Biol.*, 245(1):43–53, 1995. 30, 35, 36, 133
- [88] G. Jones, P. Willett, R. C. Glen, A. R. Leach, and R. Taylor. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.*, 267(3):727–748, 1997. 30, 34, 35, 36, 133
- [89] G. M. Morris, D. S.Goodsell, R. S. Halliday, R. Huey, W. E. Hart, R. K. Belew, and A. J. Olson. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.*, 19(14):1639–1662, 1998. 30, 34, 41, 42
- [90] M. R. McGann, H. R. Almond, A. Nicholls, J. A. Grant, and F. K. Brown. Gaussian docking functions. *Biopolymers*, 68(1):76–90, 2003. 31
- [91] I. Schellhammer and M. Rarey. TrixX: structure-based molecule indexing for large-scale virtual screening in sublinear time. *J. Comput. Aided Mol. Des.*, 21(5):223–238, 2007. 31
- [92] J. Schlosser and M. Rarey. Beyond the virtual screening paradigm: Structure-based searching for new lead compounds. *J. Chem. Inf. Model.*, 49(4):800–809, 2009. 31
- [93] R. A. Friesner, J. L. Banks, R. B. Murphy, T. A. Halgren, J. J. Klicic, T. Daniel, M. P. Repasky, E. H. Knoll, M. Shelley, J. K. Perry, D. E. Shaw, P. Francis, and P. S. Shenkin. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.*, 47(7):1739–1749, 2004. 31, 38
- [94] D. E. Koshland. Application of a Theory of Enzyme Specificity to Protein Synthesis. *Proc. Natl. Acad. Sci. U. S. A.*, 44(2):98–104, 1958. 32
- [95] H. Claußen, C. Buning, M. Rarey, and T. Lengauer. FlexE: efficient molecular docking considering protein structure variations. *J. Mol. Biol.*, 308(2):377–395, 2001. 32
- [96] C. R. Corbeil, P. Englebienne, and N. Moitessier. Docking ligands into flexible and solvated macromolecules. 1. Development and validation of FITTED 1.0. *J. Chem. Inf. Model.*, 47(2):435–449, 2007. 32, 33
- [97] M. I. Zavodszky and L.A. Kuhn. Side-chain flexibility in protein–ligand binding: The minimal rotation hypothesis. *Protein Sci.*, 14(4):1104–1114, 2005. 32
- [98] A. M. Henzler and M. Rarey. In Pursuit of Fully Flexible Protein-Ligand Docking: Modeling the Bilateral Mechanism of Binding. *Mol. Inf.*, 29(3):164–173, 2010. 32

- [99] N. Moitessier, P. Englebienne, D. Lee, J. Lawandi, and C. R. Corbeil. Towards the development of universal, fast and highly accurate docking/scoring methods: a long way to go. *Br. J. Pharmacol.*, 153:7–26, 2008. 32, 34, 38, 42, 43
- [100] T. ten Brink and T. E. Exner. Influence of Protonation, Tautomeric, and Stereoisomeric States on Protein- Ligand Docking Results. *J. Chem. Inf. Model.*, 49(6):1535–1546, 2009. 32, 49
- [101] M. Rarey, B. Kramer, and T. Lengauer. The particle concept: placing discrete water molecules during protein-ligand docking predictions. *Proteins: Struct., Funct., Bioinf.*, 34(1):17–28, 1999. 33
- [102] M. L. Verdonk, G. Chessari, J. C. Cole, M. J. Hartshorn, C. W. Murray, J. W. M. Nissink, R. D. Taylor, and R. Taylor. Modeling water molecules in protein-ligand docking using GOLD. *J. Med. Chem.*, 48(20):6504–6515, 2005.
- [103] N. Huang and B. K. Shoichet. Exploiting ordered waters in molecular docking. *J. Med. Chem.*, 51(16):4862–4865, 2008. 33
- [104] A. R. Leach, B. K. Shoichet, and C. E. Peishoff. Docking and scoring. *J. Med. Chem.*, 49:5851–5855, 2006. 34, 46
- [105] E. C. Meng, B. K. Shoichet, and I. D. Kuntz. Automated docking with grid-based energy evaluation. *J. Comput. Chem.*, 13(4):505–524, 1992. 34
- [106] R. Huey, G. M. Morris, A. J. Olson, and D. S. Goodsell. A semiempirical free energy force field with charge-based desolvation. *J. Comput. Chem.*, 28(6):1145–1152, 2007. 34, 36, 42
- [107] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.*, 117(19):5179–5197, 1995. 35
- [108] D. A. Case, T. E. Cheatham, T. Darden, H. Gohlke, R. Luo, K. M. Merz Jr, A. Onufriev, C. Simmerling, B. Wang, and R. J. Woods. The amber biomolecular simulation programs. *J. Comput. Chem.*, 26(16):1668–1688, 2005. 35
- [109] W. L. Jorgensen and 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.*, 110(6):1657–1666, 1988. 35
- [110] G. A. Kaminski, R. A. Friesner, J. Tirado-Rives, and W. L. Jorgensen. Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *J. Phys. Chem. B*, 105(28):6474–6487, 2001. 35

- [111] A. D. MacKerell, D. Bashford, Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin, and M. Karplus. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B*, 102(18):3586–3616, 1998. 35
- [112] A. D. MacKerell Jr, N. Banavali, and N. Foloppe. Development and current status of the CHARMM force field for nucleic acids. *Biopolymers*, 56(4):257–265, 2000. 35
- [113] C. Oostenbrink, A. Villa, A. E. Mark, and W. F. Van Gunsteren. A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53A5 and 53A6. *J. Comput. Chem.*, 25(13):1656–1676, 2004. 35
- [114] M. Clark, R. D. Cramer, and N. Van Opdenbosch. Validation of the general purpose Tripos 5.2 force field. *J. Comput. Chem.*, 10(8):982–1012, 1989. 35
- [115] J. W. Liebeschuetz, J. C. Cole, and O. Korb. Pose prediction and virtual screening performance of GOLD scoring functions in a standardized test. *J. Comput. Aided Mol. Des.*, 26(6):737–748, 2012. 36, 40, 154
- [116] S. Yin, L. Biedermannova, J. Vondrasek, and N. V. Dokholyan. Medusascoring: an accurate force field-based scoring function for virtual drug screening. *J. Chem. Inf. Model.*, 48(8):1656–1662, 2008. 36
- [117] P. Englebienne and N. Moitessier. Docking ligands into flexible and solvated macromolecules. 5. Force-field-based prediction of binding affinities of ligands to proteins. *J. Chem. Inf. Model.*, 49(11):2564–2571, 2009. 36
- [118] 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.*, 8:243–256, 1994. 37, 75, 76
- [119] H. J. Böhm. LUDI: rule-based automatic design of new substituents for enzyme inhibitor leads. *J. Comput. Aided Mol. Des.*, 6:593–606, 1992. 37
- [120] H.J. Böhm. The computer program LUDI: a new method for the de novo design of enzyme inhibitors. *J. Comput. Aided Mol. Des.*, 6(1):61–78, 1992. 37
- [121] M. D. Eldridge, C. W. Murray, T. R. Auton, G. V. Paolini, and R. P. Mee. Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J. Comput. Aided Mol. Des.*, 11:425–445, 1997. 38

- [122] M. L. Verdonk, J. C. Cole, M. J. Hartshorn, C. W. Murray, and R. D. Taylor. Improved protein–ligand docking using GOLD. *Proteins: Struct., Funct., Bioinf.*, 52(4):609–623, 2003. 38, 133
- [123] C. A. Baxter, C. W. Murray, D. E. Clark, D. R. Westhead, and M. D. Eldridge. Flexible docking using tabu search and an empirical estimate of binding affinity. *Proteins: Struct., Funct., Bioinf.*, 33(3):367–382, 1998. 38
- [124] M. McGann. FRED pose prediction and virtual screening accuracy. *J. Chem. Inf. Model.*, 51(3):578–596, 2011. 38, 154
- [125] R. A. Friesner, R. B. Murphy, M. P. Repasky, L. L. Frye, J. R. Greenwood, and T. A. Halgren. Extra precision Glide: docking and scoring incorporating a model of hydrophobic enclosure for protein–ligand complexes. *J. Med. Chem.*, 49:6177–6196, 2006. 38, 42, 43
- [126] R. Wang, L. Lai, and S. Wang. Further development and validation of empirical scoring functions for structure-based binding affinity prediction. *J. Comput. Aided Mol. Des.*, 16(1):11–26, 2002. 38, 75
- [127] D. K. Gehlhaar, G. M. Verkhivker, P. A. Rejto, C. J. Sherman, D. R. Fogel, L. J. Fogel, and S. T. Freer. Molecular recognition of the inhibitor AG-1343 by HIV-1 protease: conformationally flexible docking by evolutionary programming. *Chem. Biol.*, 2(5):317–324, 1995. 38
- [128] Y. T. Tang and G. R. Marshall. PHOENIX: a scoring function for affinity prediction derived using high-resolution crystal structures and calorimetry measurements. *J. Chem. Inf. Model.*, 51(2):214–228, 2011. 38, 130
- [129] J. B. Cross, D. C. Thompson, B. K. Rai, J.C. Baber, K. Y. Fan, Y. Hu, and C. Humblet. Comparison of several molecular docking programs: pose prediction and virtual screening accuracy. *J. Chem. Inf. Model.*, 49(6):1455–1474, 2009. 38
- [130] M. A. Murcko. Computational methods to predict binding free energy in ligand–receptor complexes. *J. Med. Chem.*, 38(26):4953–4967, 1995. 38
- [131] C. A. Sotriffer, P. Sanschagrin, H. Matter, and G. Klebe. SFCscore: scoring functions for affinity prediction of protein–ligand complexes. *Proteins: Struct., Funct., Bioinf.*, 73(2):395–419, 2008. 38, 75, 76
- [132] H. J. Savage, C. J. Elliott, C. M. Freeman, and J. L. Finney. Lost hydrogen bonds and buried surface area: rationalising stability in globular proteins. *J. Chem. Soc. Faraday Trans.*, 89:2609–2617, 1993. 38
- [133] M. J. Sippl. Calculation of Conformational Ensembles from Potentials of Mean Force: An approach to the Knowledge-based Prediction of Local Structures in Globular Proteins. *J. Mol. Biol.*, 213(4):859–883, 1990. 39

- [134] R. S. DeWitte and E. I. Shakhnovich. SMOG: de novo design method based on simple, fast, and accurate free energy estimates. 1. Methodology and supporting evidence. *J. Am. Chem. Soc.*, 118(47):11733–11744, 1996. 39
- [135] I. Muegge. PMF scoring revisited. *J. Med. Chem.*, 49:5895–5902, 2006. 40, 45
- [136] H. F. G. Velec, H. Gohlke, and G. Klebe. DrugScoreCSD 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.*, 48(20):6296–6303, 2005. 40
- [137] G. Neudert and G. Klebe. DSX: A Knowledge-Based Scoring Function for the Assessment of Protein–Ligand Complexes. *J. Chem. Inf. Model.*, 51(10):2731–2745, 2011. 40
- [138] W. Mooij and M. L. Verdonk. General and targeted statistical potentials for protein–ligand interactions. *Proteins: Struct., Funct., Bioinf.*, 61(2):272–287, 2005. 40, 45
- [139] M. Hendlich. Databases for protein–ligand complexes. *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 54(6):1178–1182, 1998. 40
- [140] A. N. Jain. Scoring noncovalent protein–ligand interactions: a continuous differentiable function tuned to compute binding affinities. *J. Comput. Aided Mol. Des.*, 10(5):427–440, 1996. 41
- [141] T. A. Pham and A. N. Jain. Parameter estimation for scoring protein–ligand interactions using negative training data. *J. Med. Chem.*, 49(20):5856–5868, 2006. 41
- [142] G. E. Kellogg, J. C. Burnett, and D. J. Abraham. Very empirical treatment of solvation and entropy: a force field derived from Log Po/w. *J. Comput. Aided Mol. Des.*, 15(4):381–393, 2001. 42, 75
- [143] A. Krammer, P. D. Kirchhoff, X. Jiang, C. M. Venkatachalam, and M. Waldman. LigScore: a novel scoring function for predicting binding affinities. *J. Mol. Graphics Modell.*, 23(5):395–407, 2005. 42, 75, 76
- [144] M. M. Mysinger and B. K. Shoichet. Rapid context-dependent ligand desolvation in molecular docking. *J. Chem. Inf. Model.*, 50(9):1561–1573, 2010. 42
- [145] B. K. Shoichet, A. R. Leach, and I. D. Kuntz. Ligand solvation in molecular docking. *Proteins: Struct., Funct., Bioinf.*, 34(1):4–16, 1999. 42
- [146] B. Kuhn, J. E. Fuchs, M. Reutlinger, M. Stahl, and N. R. Taylor. Rationalizing Tight Ligand Binding through Cooperative Interaction Networks. *J. Chem. Inf. Model.*, 51(12):3180–3198, 2011. 43

- [147] H. Kubinyi. Hydrogen Bonding: The Last Mystery in Drug Design? In *Pharmacokinetic optimization in drug research: biological, physicochemical, and computational strategies*, pages 513–524. Wiley-VCH Verlag GmbH & Co. KGaA, 2001. 43
- [148] S. Raub, A. Steffen, A. Kämper, and C. M. Marian. AIScore: Chemically Diverse Empirical Scoring Function Employing Quantum Chemical Binding Energies of Hydrogen-Bonded Complexes. *J. Chem. Inf. Model.*, 48(7):1492–1510, 2008. 43, 83
- [149] O. Korb, T. Stüttzle, and T. E. Exner. Empirical scoring functions for advanced protein-ligand docking with plants. *J. Chem. Inf. Model.*, 49(1):84–96, 2009. 43, 133, 134
- [150] R. Wang, Y. Lu, X. Fang, and S. Wang. An extensive test of 14 scoring functions using the PDBbind refined set of 800 protein-ligand complexes. *J. Chem. Inf. Comput. Sci.*, 44(6):2114–2125, 2004. 44, 70
- [151] P. S. Charifson, J. J. Corkery, M. A. Murcko, and W. P. Walters. Consensus scoring: A method for obtaining improved hit rates from docking databases of three-dimensional structures into proteins. *J. Med. Chem.*, 42(25):5100–5109, 1999. 44
- [152] A. Oda, K. Tsuchida, T. Takakura, N. Yamaotsu, and S. Hirono. Comparison of consensus scoring strategies for evaluating computational models of protein-ligand complexes. *J. Chem. Inf. Model.*, 46(1):380–391, 2006. 44
- [153] S. Betzi, K. Suhre, B. Chétrit, F. Guerlesquin, and X. Morelli. Gfscore: a general non-linear consensus scoring function for high-throughput docking. *J. Chem. Inf. Model.*, 46(4):1704–1712, 2006. 44
- [154] R. Teramoto and H. Fukunishi. Consensus scoring with feature selection for structure-based virtual screening. *J. Chem. Inf. Model.*, 48(2):288–295, 2008. 44
- [155] M. Stahl and M. Rarey. Detailed analysis of scoring functions for virtual screening. *J. Med. Chem.*, 44(7):1035–1042, 2001. 44
- [156] J. M. Yang, Y. F. Chen, T. W. Shen, B. S. Kristal, and D. F. Hsu. Consensus scoring criteria for improving enrichment in virtual screening. *J. Chem. Inf. Model.*, 45(4):1134–1146, 2005. 44
- [157] M. H. J. Seifert. Targeted scoring functions for virtual screening. *Drug Discovery Today*, 14(11-12):562–569, 2009. 45
- [158] M. Xue, M. Zheng, B. Xiong, Y. Li, H. Jiang, and J. Shen. Knowledge-based scoring functions in drug design. 1. Developing a target-specific method for kinase-ligand interactions. *J. Chem. Inf. Model.*, 50(8):1378–1386, 2010. 45
- [159] B. Seebeck, M. Wagener, and M. Rarey. From Activity Cliffs to Target-Specific Scoring Models and Pharmacophore Hypotheses. *ChemMedChem*, 6(9):1630–1639, 2011. 45

- [160] H. Gohlke and G. Klebe. DrugScore meets CoMFA: adaptation of fields for molecular comparison (AFMoC) or how to tailor knowledge-based pair-potentials to a particular protein. *J. Med. Chem.*, 45(19):4153–4170, 2002. 45
- [161] T. A. Pham and A. N. Jain. Customizing scoring functions for docking. *J. Comput. Aided Mol. Des.*, 22(5):269–286, 2008. 45
- [162] D. L. Mobley and K. A. Dill. Binding of Small-Molecule Ligands to Proteins: „What You See“ Is Not Always „What You Get“. *Structure*, 17:489–498, 2009. 46
- [163] J. Åqvist, C. Medina, and J. E. Samuelsson. A new method for predicting binding affinity in computer-aided drug design. *Protein Eng.*, 7(3):385–391, 1994. 46
- [164] P. A. Kollman, I. Massova, C. Reyes, B. Kuhn, S. Huo, L. Chong, M. Lee, T. Lee, Y. Duan, W. Wang, O. Donini, P. Cieplak, S. Jaysharee, D. E. Case, and T. E. Cheatham. Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc. Chem. Res.*, 33(12):889–897, 2000. 46
- [165] N. Huang, C. Kalyanaraman, J. J. Irwin, and M. P. Jacobson. Physics-based scoring of protein-ligand complexes: enrichment of known inhibitors in large-scale virtual screening. *J. Chem. Inf. Model.*, 46(1):243–253, 2006. 46
- [166] T. Young, R. Abel, B. Kim, B. J. Berne, and R. A. Friesner. Motifs for molecular recognition exploiting hydrophobic enclosure in protein–ligand binding. *Proc. Natl. Acad. Sci. U. S. A.*, 104(3):808, 2007. 46
- [167] R. Abel, T. Young, R. Farid, B. J. Berne, and R. A. Friesner. Role of the active-site solvent in the thermodynamics of factor Xa ligand binding. *J. Am. Chem. Soc.*, 130(9):2817–2831, 2008. 46
- [168] T. Lazaridis. Inhomogeneous fluid approach to solvation thermodynamics. 1. Theory. *J. Phys. Chem. B*, 102(18):3531–3541, 1998. 46
- [169] T. Beuming, R. Farid, and W. Sherman. High-energy water sites determine peptide binding affinity and specificity of PDZ domains. *Protein Sci.*, 18(8):1609–1619, 2009. 46
- [170] D. Rognan. Docking Methods for Virtual Screening: Principles and Recent Advances. In C. A. Sotriffer, editor, *Virtual Screening: Principles, Challenges, and Practical Guidelines*, pages 153–176. Wiley-VCH Verlag GmbH & Co. KGaA, 2011. 47, 49, 50
- [171] J. J. Irwin and B. K. Shoichet. ZINC—a free database of commercially available compounds for virtual screening. *J. Chem. Inf. Model.*, 45(1):177–182, 2005. 48, 70, 143
- [172] M. Boehm. Virtual Screening of Chemical Space: From Generic Compound Collections to Tailored Screening Libraries. In C. A. Sotriffer, editor, *Virtual Screening: Principles, Challenges, and Practical Guidelines*, pages 1–33. Wiley-VCH Verlag GmbH & Co. KGaA, 2011. 48

- [173] C. A. Lipinski, F. Lombardo, B. W. Dominy, and P. J. Feeney. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Delivery Rev.*, 23(1-3):3–25, 1997. 48, 77
- [174] T. I. Oprea, A. M. Davis, S. J. Teague, and P. D. Leeson. Is there a difference between leads and drugs? a historical perspective. *J. Chem. Inf. Comput. Sci.*, 41(5):1308–1315, 2001. 48
- [175] B. Waszkowycz, D. E. Clark, and E. Gancia. Outstanding challenges in protein–ligand docking and structure-based virtual screening. *WIREs Comput. Mol. Sci.*, 1(2):229–259, 2011. 49, 50
- [176] S. Urbaczek, A. Kolodzik, J. R. Fischer, T. Lippert, S. Heuser, I. Groth, T. Schulz-Gasch, and M. Rarey. NAOMI: On the Almost Trivial Task of Reading Molecules from Different File formats. *J. Chem. Inf. Model.*, 51(12):3199–3207, 2011. 49, 129, 134
- [177] A. Volkamer, A. Griewel, T. Grombacher, and M. Rarey. Analyzing the topology of active sites: on the prediction of pockets and subpockets. *J. Chem. Inf. Model.*, 50(11):2041–2052, 2010. 49
- [178] S. Hindle, M. Rarey, C. Buning, and T. Lengauer. Flexible docking under pharmacophore type constraints. *J. Comput. Aided. Mol. Des.*, 16(2):129–149, 2002. 50
- [179] F. H. Stillinger. Water revisited. *Science*, 209(4455):451–457, 1980. 51
- [180] G. Lange, R. Klein, J. Albrecht, M. Rarey, and I. Reulecke. Method for the Determination of Intra- and Intermolecular Interactions in Aqueous Solution. European patent No. EP2084520, 2010. 51
- [181] G. Lange, N. Schneider, M. Rarey, and R. Klein. Impact of unsatisfied hydrogen bond functions on intermolecular interactions in water. *in Vorbereitung*, 2012. 51, 53, 54, 61
- [182] B. Chen, I. Ivanov, M. L. Klein, and M. Parrinello. Hydrogen bonding in water. *Physical review letters*, 91(21):215503, 2003. 51
- [183] P. Wernet, D. Nordlund, U. Bergmann, M. Cavalleri, M. Odelius, H. Ogasawara, L.Å. Näslund, TK Hirsch, L. Ojamäe, P. Glatzel, L. G. M. Pettersson, and A. Nilsson. The structure of the first coordination shell in liquid water. *Science*, 304(5673):995–999, 2004. 52
- [184] J. D. Eaves, J. J. Loparo, C. J. Fecko, S. T. Roberts, A. Tokmakoff, and P. L. Geissler. Hydrogen bonds in liquid water are broken only fleetingly. *Proc. Natl. Acad. Sci. U. S. A.*, 102(37):13019, 2005. 52, 54
- [185] R. Kumar, J. R. Schmidt, and J. L. Skinner. Hydrogen bonding definitions and dynamics in liquid water. *The J. Chem. Phys.*, 126:204107, 2007. 51, 52

- [186] F. H. Stillinger and A. Rahman. Improved simulation of liquid water by molecular dynamics. *J. Chem. Phys.*, 60:1545, 1974. 52
- [187] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, 79(2):926, 1983.
- [188] H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma. The missing term in effective pair potentials. *J. Phys. Chem.*, 91(24):6269–6271, 1987. 52
- [189] L. Pauling. The structure and entropy of ice and of other crystals with some randomness of atomic arrangement. *J. Am. Chem. Soc.*, 57(12):2680–2684, 1935. 52
- [190] P. G. Owston. The position of hydrogen bond atoms in ice. *J. Chim. Phys.*, 50:C13–C18, 1953.
- [191] H. Engelhardt and B. Kamb. X-ray determination of the structure of ice IV. *Journal of Glaciology*, 21(85), 1978. 52
- [192] A. K. Soper and A. Luzar. Orientation of water molecules around small polar and nonpolar groups in solution: a neutron diffraction and computer simulation study. *J. Phys. Chem.*, 100(4):1357–1367, 1996. 52, 57, 58
- [193] J. D. Smith, C. D. Cappa, K. R. Wilson, B. M. Messer, R. C. Cohen, and R. J. Saykally. Energetics of hydrogen bond network rearrangements in liquid water. *Science*, 306(5697):851–853, 2004. 52, 54
- [194] A. Luzar and J. Stefan. Dielectric behaviour of DMSO-water mixtures. A hydrogen-bonding model. *Journal of Molecular Liquids*, 46:221–238, 1990.
- [195] A. K. Soper. The radial distribution functions of water and ice from 220 to 673 K and at pressures up to 400 MPa. *Chemical Physics*, 258(2):121–137, 2000. 54
- [196] P. J. Flory. *Principles of Polymer Chemistry*. Cornell Univ. Press, Ithaca, N. Y., 1953. 56
- [197] S. E. McLain, A. K. Soper, and A. Luzar. Investigations on the structure of dimethyl sulfoxide and acetone in aqueous solution. *J. Chem. Phys.*, 127(17):174515–174700, 2007. 57, 58
- [198] C. C. McComas, B. M. Crowley, and D. L. Boger. Partitioning the loss in vancomycin binding affinity for D-Ala-D-Lac into lost H-bond and repulsive lone pair contributions. *J. Am. Chem. Soc.*, 125:9314–9315, 2003. 58
- [199] N. Foloppe, L. M. Fisher, R. Howes, P. Kierstan, A. Potter, A. G. S. Robertson, and A. E. Surgenor. Structure-Based Design of Novel Chk1 Inhibitors: Insights into Hydrogen

- Bonding and Protein-Ligand Affinity. *J. Med. Chem.*, 48(13):4332–4345, 2005. 58, 108, 112
- [200] Y. Gu, T. Kar, and S. Scheiner. Fundamental Properties of the CH-O Interaction: Is It a True Hydrogen Bond? *J. Am. Chem. Soc.*, 121(40):9411–9422, 1999. 58
- [201] I. Reulecke, G. Lange, J. Albrecht, R. Klein, and M. Rarey. Towards an integrated description of hydrogen bonding and dehydration: Reducing false positives in virtual screening using the HYDE scoring function. *ChemMedChem*, 3(6):885–897, 2008. 62, 63, 67, 69, 70, 83, 147
- [202] I. Reulecke. *Adaptives Scoring im strukturbasierten Wirkstoffentwurf. HYDE - Eine neue Scoring-Funktion für die konsistente Bewertung von Protein-Ligand Komplexen*. Dissertation, Zentrum für Bioinformatik, Universität Hamburg, 2008. 63, 66, 68, 70, 71, 85, 154, 156, 205, 206, 207
- [203] C. Hansch, A. Leo, and D. Hoekman. *Exploring QSAR: Hydrophobic, Electronic, and Steric Constants*. American Chemical Society, 1995. 67, 79
- [204] Syracuse. PHYSPROP Database. <http://www.syrres.com/esc/physprop.htm>. 67, 79
- [205] B. Lee and F. M. Richards. The interpretation of protein structures: Estimation of static accessibility. *J. Mol. Biol.*, 55(3):379 – IN4, 1971. 67, 76, 77, 97
- [206] A. Shrake and J. A. Rupley. Environment and exposure to solvent of protein atoms. Lysozyme and insulin. *J. Mol. Biol.*, 79(2):351 – 371, 1973.
- [207] F. M. Richards. Areas, volumes, packing, and protein structure. *Annu. Rev. Biophys. Bioeng.*, 6(1):151–176, 1977. 67, 76, 77, 97
- [208] N. Schneider, G. Lange, S. Hindle, R. Klein, and M. Rarey. A Consistent Description of HYdrogen Bond and DEhydration Energies in Protein-Ligand Complexes: Methods Behind the HYDE Scoring Function. *eingereicht*, 2012. 69, 89, 95, 100, 102, 123, 124, 127, 130, 131, 134, 155, 208
- [209] N. Huang, B. Shoichet, and J. J. Irwin. Benchmarking Set for Molecular Docking. *J. Med. Chem.*, 49(23):6789–6801, 2006. 70, 143
- [210] R. Wang, X. Fang, Y. Lu, and S. Wang. The PDBbind database: collection of binding affinities for protein-ligand complexes with known three-dimensional structures. *J. Med. Chem.*, 47(12):2977–2980, 2004. 70, 128
- [211] R. Wang, X. Fang, Y. Lu, C. Y. Yang, and S. Wang. The PDBbind database: methodologies and updates. *J. Med. Chem.*, 48(12):4111–4119, 2005. 70, 128

- [212] H. Kubinyi. Success stories of computer-aided design. In *Computer Applications in Pharmaceutical Research and Development*, pages 377–424. John Wiley & Sons, Inc., 2006. 71
- [213] I. K. McDonald and J. M. Thornton. Satisfying hydrogen bonding potential in proteins. *J. Mol. Biol.*, 238(5):777–793, 1994. 72
- [214] J. E. Dennis and R. B. Schnabel. *Numerical methods for unconstrained optimization and nonlinear equations*, volume 16. Society for Industrial Mathematics, 1996. 74
- [215] K. Palm, K. Luthman, A.L. Ungell, G. Strandlund, F. Beigi, P. Lundahl, and P. Artursson. Evaluation of dynamic polar molecular surface area as predictor of drug absorption: comparison with other computational and experimental predictors. *J. Med. Chem.*, 41(27):5382–5392, 1998. 75, 76
- [216] J. Polanski and B. Walczak. The comparative molecular surface analysis (COMSA): a novel tool for molecular design. *Comput. Chem. (Oxford)*, 24(5):615–625, 2000. 75, 76
- [217] P. Ertl, B. Rohde, and P. Selzer. Fast calculation of molecular polar surface area as a sum of fragment-based contributions and its application to the prediction of drug transport properties. *J. Med. Chem.*, 43(20):3714–3717, 2000. 75
- [218] M. L. Connolly. Analytical molecular surface calculation. *J. Appl. Crystallogr.*, 16(5):548–558, 1983. 76, 77, 97
- [219] M. L. Connolly. Solvent-accessible surfaces of proteins and nucleic acids. *Science*, 221(4612):709–713, 1983.
- [220] T. J. Richmond. Solvent accessible surface area and excluded volume in proteins: analytical equations for overlapping spheres and implications for the hydrophobic effect. *J. Mol. Biol.*, 178(1):63–89, 1984. 76, 77, 97
- [221] A. Bondi. van der Waals Volumes and Radii. *J. Phys. Chem.*, 68(3):441–451, 1964. 76, 97
- [222] A. Leo and D. Hansch, C. and Elkins. Partition coefficients and their uses. *Chem. Rev.*, 71(6):525–616, 1971. 77
- [223] J. C. Dearden and G. M. Bresnen. The measurement of partition coefficients. *Quant. Struct.-Act. Relat.*, 7(3):133–144, 1988.
- [224] A. Hersey, A. P. Hill, R. M. Hyde, and D. J. Livingstone. Principles of method selection in partition studies. *Quant. Struct.-Act. Relat.*, 8(4):288–296, 1989. 77
- [225] M. S. Mirrlees, S. J. Moulton, C. T. Murphy, and P. J. Taylor. Direct measurement of octanol-water partition coefficients by high-pressure liquid chromatography. *J. Med. Chem.*, 19(5):615–619, 1976. 77

- [226] J. Sangster. *Octanol-water partition coefficients: fundamentals and physical chemistry*. John Wiley & Son Ltd, 1997. 77
- [227] R. Mannhold, G. I. Poda, C. Ostermann, and I. V. Tetko. Calculation of molecular lipophilicity: State-of-the-art and comparison of log P methods on more than 96,000 compounds. *J. Pharm. Sci.*, 98(3):861–893, 2009. 77, 78
- [228] A. K. Ghose and G. M. Crippen. Atomic Physicochemical Parameters for Three-Dimensional Structure-Directed Quantitative Structure-Activity Relationships I. Partition Coefficients as a Measure of Hydrophobicity. *J. Comput. Chem.*, 7(4):565–577, 1986. 77, 78
- [229] R. Wang, Y. Gao, and L. Lai. Calculating partition coefficient by atom-additive method. *Perspect. Drug Discovery Des.*, 19(1):47–66, 2000. 77
- [230] A. J. Leo. Calculating log P from structures. *Chem. Rev.*, 93(4):1281–1306, 1993. 77, 78, 79
- [231] A. A. Petrauskas and E. A. Kolovanov. ACD/Log P method description. *Perspect. Drug Discovery Des.*, 19(1):99–116, 2000. 77
- [232] N. Schneider, R. Klein, G. Lange, and M. Rarey. Nearly no Scoring Function Without a Hansch-Analysis. *Mol. Inf.*, 31(6-7):503–507, 2012. 78, 80, 81, 83
- [233] S. M. J. Harris, T. Higuchi, and J. H. Rytting. Thermodynamic group contributions from ion pair extraction equilibria for use in the prediction of partition coefficients. Correlation of surface area with group contributions. *J. Phys. Chem.*, 77(22):2694–2703, 1973. 78
- [234] C. Chothia. Hydrophobic bonding and accessible surface area in proteins. *Nature*, 248(5446):338–339, 1974.
- [235] J. A. Reynolds, D. B. Gilbert, and C. Tanford. Empirical correlation between hydrophobic free energy and aqueous cavity surface area. *Proc. Natl. Acad. Sci. U. S. A.*, 71(8):2925, 1974.
- [236] G. L. Amidon, S. H. Yalkowsky, and S. Leung. Solubility of nonelectrolytes in polar solvents II: Solubility of aliphatic alcohols in water. *J. Pharm. Sci.*, 63(12):1858–1866, 1974.
- [237] G. L. Amidon, S. H. Yalkowsky, S. T. Anik, and S. C. Valvani. Solubility of nonelectrolytes in polar solvents. V. Estimation of the solubility of aliphatic monofunctional compounds in water using a molecular surface area approach. *J. Phys. Chem.*, 79(21):2239–2246, 1975.

- [238] R. Pearlman, S. H. Yalkowsky, and S. Banerjee. Water solubilities of polynuclear aromatic and heteroaromatic compounds. *J. Phys. Chem. Ref. Data*, 13(2), 1984. 78
- [239] A. Opperhuizen, F. A. P. C. Gobas, J. M. D. Van der Steen, and O. Hutzinger. Aqueous solubility of polychlorinated biphenyls related to molecular structure. *Environ. Sci. Technol.*, 22(6):638–646, 1988.
- [240] J. de Bruijn and J. Hermens. Relationships between octanol/water partition coefficients and total molecular surface area and total molecular volume of hydrophobic organic chemicals. *Quant. Struct.-Act. Relat.*, 9(1):11–21, 1990. 78
- [241] T. I. Oprea. Rapid estimation of hydrophobicity for virtual combinatorial library analysis. *SAR QSAR Environ. Res.*, 12(1-2):129–141, 2001. 78
- [242] A. Leo, C. Hansch, and P. Y. C. Jow. Dependence of hydrophobicity of apolar molecules on their molecular volume. *J. Med. Chem.*, 19(5):611–615, 1976. 78
- [243] N. El Tayar, B. Testa, and P. A. Carrupt. Polar intermolecular interactions encoded in partition coefficients: an indirect estimation of hydrogen-bond parameters of polyfunctional solutes. *J. Phys. Chem.*, 96(3):1455–1459, 1992. 78
- [244] C. Hansch and A. J. Leo. Medchem Project Issue. *Claremont, CA. Pomona College*, 26, 1985. 79
- [245] C. Hansch and A. Leo. The log P database. *Claremont, CA: Pomona College*, 398, 1987. 79
- [246] J. Gasteiger, C. Rudolph, and J. Sadowski. Automatic generation of 3D-atomic coordinates for organic molecules. *Tetrahedron Computer Methodology*, 3(6):537–547, 1990. 79, 134
- [247] Molecular Networks. CORINA. <http://www.molecular-networks.com/products/corina>. 79, 134
- [248] M. Stahl. Modifications of the scoring function in FlexX for virtual screening applications. *Perspect. Drug Discovery Des.*, 20(1):83–98, 2000. 93
- [249] J. Finkelstein. Metalloproteins. *Nature*, 460(7257):813–813, 2009. 103
- [250] F. E. Jacobsen, J. A. Lewis, and S. M. Cohen. The design of inhibitors for medically relevant metalloproteins. *ChemMedChem*, 2(2):152–171, 2007. 103
- [251] K. Stierand and M. Rarey. Drawing the PDB - Protein-Ligand Complexes in two Dimensions. *J. Med. Chem. Lett.*, 1:540–545, 2010. 108, 113
- [252] Chemical Computing Group Inc. Molecular Operating Environment (MOE). <http://www.chemcomp.com/software-moe2011.htm>. 108, 135

- [253] M. Vieth, J. D. Hirst, A. Kolinski, and C. L. Brooks. Assessing energy functions for flexible docking. *J. Comput. Chem.*, 19(14):1612–1622, 1998. 118
- [254] J. Kirchmair, P. Markt, S. Distinto, G. Wolber, and T. Langer. Evaluation of the performance of 3D virtual screening protocols: RMSD comparisons, enrichment assessments, and decoy selection—What can we learn from earlier mistakes? *J. Comput. Aided Mol. Des.*, 22(3):213–228, 2008. 118
- [255] A. N. Jain and A. Nicholls. Recommendations for evaluation of computational methods. *J. Comput. Aided Mol. Des.*, 22(3):133–139, 2008. 118, 120
- [256] N. Triballeau, F. Acher, I. Brabet, J. P. Pin, and H. O. Bertrand. Virtual Screening Workflow Development Guided by the „Receiver Operating Characteristic“ Curve Approach. Application to High-Throughput Docking on Metabotropic Glutamate Receptor Subtype 4. *J. Med. Chem.*, 48:2534–2547, 2005. 119
- [257] K. C. Yeh and K. C. Kwan. A Comparison of Numerical Integrating Algorithms by Trapezoidal, Lagrange, and Spline Approximation. *J. Pharmacokinet. Pharmacodyn.*, 6:79–89, 1978. 119
- [258] B. Baum, M. Mohamed, M. Zayed, C. Gerlach, A. Heine, D. Hangauer, and G. Klebe. More than a simple lipophilic contact: a detailed thermodynamic analysis of nonbasic residues in the S1 pocket of thrombin. *J. Mol. Biol.*, 390(1):56–69, 2009. 122
- [259] J. Regan, S. Breitfelder, P. Cirillo, T. Gilmore, A.G. Graham, E. Hickey, B. Klaus, J. Mardwed, M. Moriak, N. Moss, C. Pargellis, S. Pav, A. Proto, A. Swinamer, L. Tong, and C. Torcellini. Pyrazole urea-based inhibitors of p38 MAP kinase: from lead compound to clinical candidate. *J. Med. Chem.*, 45(14):2994–3008, 2002. 126, 128
- [260] BioSolveIT. LeadIT. <http://www.biosolveit.de/LeadIT>. 129, 133
- [261] C. R. Søndergaard, A. E. Garrett, T. Carstensen, G. Pollastri, and J. E. Nielsen. Structural Artifacts in Protein-Ligand X-ray Structures: Implications for the Development of Docking Scoring Functions. *J. Med. Chem.*, 52(18):5673–5684, 2009. 132
- [262] M. J. Hartshorn, M. L. Verdonk, G. Chessari, S. C. Brewerton, W. T. M. Mooij, P. N. Mortonson, and C. W. Murray. Diverse, High-Quality Test Set for the Validation of Protein-Ligand Docking Performance. *J. Med. Chem.*, 50(4):726–741, 2007. 133, 156
- [263] O. Korb, T. Stützle, and T. Exner. PLANTS: application of ant colony optimization to structure-based drug design. *Ant Colony Optimization and Swarm Intelligence*, pages 247–258, 2006. 133
- [264] O. Korb, T. Stützle, and T. E. Exner. An ant colony optimization approach to flexible protein-ligand docking. *Swarm Intelligence*, 1(2):115–134, 2007. 133

- [265] D. Weininger. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J. Chem. Inf. Comput. Sci.*, 28(1):31–36, 1988. 134
- [266] N. Schneider, S. Hindle, G. Lange, R. Klein, J. Albrecht, H. Briem, K. Beyer, H. Claußen, M. Gastreich, C. Lemmen, et al. Substantial improvements in large-scale redocking and screening using the novel HYDE scoring function. *J. Comput. Aided Mol. Des.*, 26(6):701–723, 2012. 135, 136, 137, 138, 139, 140, 141, 142, 146, 147, 148, 150, 152, 214
- [267] M. D. Sintchak, M. A. Fleming, O. Futer, S. A. Raybuck, S. P. Chambers, P. R. Caron, M. A. Murcko, and K. P. Wilson. Structure and Mechanism of Inosine Monophosphate Dehydrogenase in Complex with the Immunosuppressant Mycophenolic Acid. *Cell*, 85:921–930, 1996. 139
- [268] C. Bissantz, G. Folkers, and D. Rognan. Protein-Based Virtual Screening of Chemical Databases. 1. Evaluation of Different Docking/Scoring Combinations. *J. Med. Chem.*, 43(25):4759–4767, 2000. 147
- [269] R. Perozzo, I. Jelesarov, H. R. Bosshard, G. Folkers, and L. Scapozza. Compulsory order of substrate binding to herpes simplex virus type 1 thymidine kinase. *J. Biol. Chem.*, 275(21):16139–16145, 2000. 148
- [270] M. J. Pérez-Pérez, E. M. Priego, A. I. Hernández, O. Familiar, M. J. Camarasa, A. Negri, F. Gago, and J. Balzarini. Structure, physiological role, and specific inhibitors of human thymidine kinase 2 (TK2): Present and future. *Med. Res. Rev.*, 28:797–820, 2008. 149
- [271] S. Gauthier, B. Caron, J. Cloutier, Y. L. Dory, A. Favre, D. Larouche, J. Mailhot, C. Ouellet, A. Schwerdtfeger, G. Leblanc, C. Martel, J. Simard, Y. Mérand, A. Bélanger, C. Labrie, and F. Labrie. S-(+)-4-[7-(2, 2-dimethyl-1-oxopropoxy)-4-methyl-2-[4-[2-(1-piperidinyl)-ethoxy] phenyl]-2H-1-benzopyran-3-yl]-phenyl 2, 2-dimethylpropanoate (EM-800): a highly potent, specific, and orally active nonsteroidal antiestrogen. *J. Med. Chem.*, 40(14):2117, 1997. 151
- [272] J. M. Schmidt, J. Mercure, G. B. Tremblay, M. Pagé, A. Kalbakji, M. Feher, R. Dunn-Dufault, M. G. Peter, and P. R. Redden. De novo design, synthesis, and evaluation of novel nonsteroidal phenanthrene ligands for the estrogen receptor. *J. Med. Chem.*, 46(8):1408–1418, 2003. 151
- [273] A. Corsini, F. M. Maggi, and A. L. Catapano. Pharmacology of competitive inhibitors of HMG-CoA reductase. *Pharmacol. Res.*, 31(1):9–27, 1995. 153
- [274] A. C. Krueger, Y. Xu, W. M. Kati, D. J. Kempf, C. J. Maring, K. F. McDaniel, A. Molla, D. Montgomery, and W. E. Kohlbrenner. Synthesis of potent pyrrolidine influenza neuraminidase inhibitors. *Bioorg. Med. Chem. Lett.*, 18(5):1692–1695, 2008. 154

- [275] C. U. Kim, W. Lew, M. A. Williams, H. Wu, L. Zhang, X. Chen, P. A. Escarpe, D. B. Mendel, W. G. Laver, and R. C. Stevens. Structure-activity relationship studies of novel carbocyclic influenza neuraminidase inhibitors. *J. Med. Chem.*, 41(14):2451–2460, 1998. 154
- [276] Y. S. Babu, P. Chand, S. Bantia, P. Kotian, A. Dehghani, Y. El-Kattan, T.-H. Lin, T. L. Hutchison, A. J. Elliott, C. D. Parker, S. L. Ananth, L. L. Horn, G. W. Laver, and J. A. Montgomery. BCX-1812 (RWJ-270201): Discovery of a Novel, Highly Potent, Orally Active, and Selective Influenza Neuraminidase Inhibitor through Structure-Based Drug Design. *J. Med. Chem.*, 43(19):3482–3486, 2000.
- [277] D.A. Erlanson, R.S. McDowell, and T. O’Brien. Fragment-based drug discovery. *J. Med. Chem.*, 47(14):3463–3482, 2004.
- [278] K. Hata, K. Koseki, K. Yamaguchi, S. Moriya, Y. Suzuki, S. Yingsakmongkon, G. Hirai, M. Sodeoka, M. Von Itzstein, and T. Miyagi. Limited inhibitory effects of oseltamivir and zanamivir on human sialidases. *Antimicrob. Agents Chemother.*, 52(10):3484–3491, 2008. 154
- [279] M. P. Repasky, R. B. Murphy, J. L. Banks, J. R. Greenwood, I. Tubert-Brohman, S. Bhat, and R. A. Friesner. Docking performance of the glide program as evaluated on the Astex and DUD datasets: a complete set of glide SP results and selected results for a new scoring function integrating WaterMap and glide. *J. Comput. Aided Mol. Des.*, 26(6):787–799, 2012. 154
- [280] M. A. C. Neves, M. Totrov, and R. Abagyan. Docking and scoring with ICM: the benchmarking results and strategies for improvement. *J. Comput. Aided Mol. Des.*, 26(6):675–686, 2012. 154
- [281] S. R. Brozell, S. Mukherjee, T. E. Balius, D. R. Roe, D. A. Case, and R. C. Rizzo. Evaluation of DOCK 6 as a pose generation and database enrichment tool. *J. Comput. Aided Mol. Des.*, 26(6):749–773, 2012. 154
- [282] K. T. Schomburg, I. Ardao, K. Götz, F. Rieckenberg, A. Liese, A.-Z. Zeng, and M. Rarey. Computational Biotechnology: Prediction of competitive substrate inhibition of enzymes by buffer compounds with protein-ligand docking. *J. Biotechnol.*, 161(4):391–401, 2012. 160
- [283] M. P. H. Stumpf, T. Thorne, E. De Silva, R. Stewart, H. J. An, M. Lappe, and C. Wiuf. Estimating the size of the human interactome. *Proc. Natl. Acad. Sci. U. S. A.*, 105(19):6959, 2008. 162
- [284] S. Jones and J. M. Thornton. Principles of protein-protein interactions. *Proc. Natl. Acad. Sci. U. S. A.*, 93(1):13, 1996. 162

- [285] C.J. Tsai and R. Nussinov. Hydrophobic folding units at protein-protein interfaces: Implications to protein folding and to protein-protein association. *Protein Sci.*, 6(7):1426–1437, 1997. 162
- [286] J. A. Wells and C. L. McClendon. Reaching for high-hanging fruit in drug discovery at protein-protein interfaces. *Nature*, 450(7172):1001–1009, 2007. 162
- [287] C. M. Deane, Ł. Salwiński, I. Xenarios, and D. Eisenberg. Protein interactions: two methods for assessment of the reliability of high throughput observations. *Mol. Cell. Proteomics*, 1(5):349–356, 2002. 163
- [288] C. Von Mering, R. Krause, B. Snel, M. Cornell, S. G. Oliver, S. Fields, and P. Bork. Comparative assessment of large-scale data sets of protein-protein interactions. *Nature*, 317:399–404, 2002. 163
- [289] H. Ponstingl, K. Henrick, and J. M. Thornton. Discriminating between homodimeric and monomeric proteins in the crystalline state. *Proteins: Struct., Funct., Bioinf.*, 41(1):47–57, 2000. 163
- [290] Y. Ofran and B. Rost. Analysing six types of protein-protein interfaces. *J. Mol. Biol.*, 325(2):377–387, 2003.
- [291] J. Mintseris and Z. Weng. Atomic contact vectors in protein-protein recognition. *Proteins: Struct., Funct., Bioinf.*, 53(3):629–639, 2003.
- [292] P. R. Bahadur, P. Chakrabarti, F. Rodier, and J. Janin. A dissection of specific and non-specific protein-protein interfaces. *J. Mol. Biol.*, 336(4):943–955, 2004. 165
- [293] J. R. Bradford and D. R. Westhead. Improved prediction of protein-protein binding sites using a support vector machines approach. *Bioinformatics*, 21(8):1487–1494, 2005.
- [294] H. Zhu, F. Domingues, I. Sommer, and T. Lengauer. NOXclass: prediction of protein-protein interaction types. *BMC Bioinformatics*, 7(1):27, 2006. 164, 165
- [295] P. Block, J. Paern, E. Hüllermeier, P. Sanschagrin, C. A. Sotriffer, and G. Klebe. Physicochemical descriptors to discriminate protein-protein interactions in permanent and transient complexes selected by means of machine learning algorithms. *Proteins: Struct., Funct., Bioinf.*, 65(3):607–622, 2006.
- [296] S. H. Park, J. A. Reyes, D. R. Gilbert, J. W. Kim, and S. Kim. Prediction of protein-protein interaction types using association rule based classification. *BMC Bioinformatics*, 10(1):36, 2009.
- [297] Q. Liu and J. Li. Propensity vectors of low-ASA residue pairs in the distinction of protein interactions. *Proteins: Struct., Funct., Bioinf.*, 78(3):589–602, 2010. 163, 165

- [298] J. Janin and F. Rodier. Protein–protein interaction at crystal contacts. *Proteins: Struct., Funct., Bioinf.*, 23(4):580–587, 1995. 163
- [299] E. D. Levy. PiQSi: protein quaternary structure investigation. *Structure*, 15(11):1364–1367, 2007. 163
- [300] P. Mitra and D. Pal. Combining Bayes Classification and Point Group Symmetry under Boolean Framework for Enhanced Protein Quaternary Structure Inference. *Structure*, 19(3):304–312, 2011. 163
- [301] E. Vennmann, N. Schneider, G. Lange, and M. Rarey. Discrimination of permanent and transient heterodimers: A structurally unbiased classification of protein-protein interactions. *in Vorbereitung*, 2012. 163
- [302] S. De, O. Krishnadev, N. Srinivasan, and N. Rekha. Interaction preferences across protein-protein interfaces of obligatory and non-obligatory components are different. *BMC Structural Biology*, 5(1):15, 2005. 165
- [303] I. Nooren and J. M. Thornton. Structural characterisation and functional significance of transient protein-protein interactions. *J. Mol. Biol.*, 325(5):991–1018, 2003.
- [304] Y. C. Chen and C. Lim. Common physical basis of macromolecule-binding sites in proteins. *Nucleic Acids Res.*, 36(22):7078–7087, 2008.
- [305] H. Madaoui and R. Guerois. Coevolution at protein complex interfaces can be detected by the complementarity trace with important impact for predictive docking. *Proc. Natl. Acad. Sci. U. S. A.*, 105(22):7708, 2008. 165



Benutzung der Software

Die Bewertungsfunktion HYDE wurde durch die BioSolveIT als Modul in die Software Plattform LeadIT integriert. Dort ist sie zur Analyse und Bewertung von einzelnen Protein-Ligand-Komplexen direkt über die Benutzeroberfläche zugänglich. Zum Zweck der Nutzung in einem virtuellen Screening wurde sie von der BioSolveIT zusätzlich als Kommandozeilen-Tool zur Verfügung gestellt. In beiden Versionen sind auch die Optimierungsroutinen enthalten, so dass die Komplexe optimal für die HYDE-Funktion vorbereitet werden können, unabhängig mit welchem Programm die Docking-Posen erzeugt wurden.

A.1 HYDE in der LeadIT GUI

HYDE kann in der LeadIT-GUI über das *grüne Daumen* Symbol in der Hauptleiste gestartet werden (Abbildung A.1), um Kristallstrukturen, Docking-Lösungen oder Recore-Moleküle zu bewerten. Diese können direkt im HYDE-Menü geladen werden. Ein Editor wurde integriert, so dass auch der Austausch einzelner Atome bewertet werden kann (*Stift* Symbol) (Abbildung A.1). Der Komplex wird automatisch zuerst optimiert und dann bewertet, diese Einstellung kann in den **Preferences** geändert werden: `HYDE_OPTIMIZER_METHOD` und `HYDE_USE_HBONDNETWORK`. Die HYDE-Bewertung liefert

A. Benutzung der Software

Project - LeadIT 2.1.3 (13.06.13)

LeadIT Molecules Docking RefCore Scoring Display Window Help

Ligands: 32U-1000-H Load... Save...

Ligand After Assessment

Hyde Assessment

Hyde Score: mM μM nM

ΔG = -67 kJ/mol LE = 0.56

Atom	Hyde Score	Ligand		Receptor	
		Desolv.	Interact.	Desolv.	Interact.
_H13	-3.9	8.4	-11.4	6.4	-7.3
_C5	0.0	0.0	0.0	0.0	0.0
_C14	0.0	0.0	0.0	0.0	0.0
_O32	-8.4	7.5	-10.1	0.6	-6.3
_C33	-2.2	-1.4	0.0	-0.8	0.0
_C6	-2.1	-0.2	0.0	-1.9	0.0
_C8	-0.2	-0.2	0.0	0.0	0.0
_C12	-4.1	-1.9	0.0	-2.2	0.0
_C9	-0.5	-0.5	0.0	0.0	0.0
_C11	-3.4	-1.9	0.0	-1.5	0.0
_C10	-2.0	-1.2	0.0	-0.8	0.0
_H1	0.0	0.0	0.0	0.0	0.0
_C1	-4.0	-1.2	0.0	-2.8	0.0
_C7	0.0	0.0	0.0	0.0	0.0
_O22	-0.7	0.3	0.0	-1.0	0.0
_C2	-3.2	-1.5	0.0	-1.7	0.0
_C3	-3.5	-2.1	0.0	-1.4	0.0
_C4	-1.2	-0.7	0.0	-0.6	0.0
_H23	-2.8	7.1	-8.4	7.4	-8.0
_C24	-0.9	-0.8	0.0	-0.1	0.0
_C25	-0.2	-0.2	0.0	0.0	0.0
_C26	-0.4	-0.4	0.0	0.0	0.0
_C27	-4.1	-2.1	0.0	-2.1	0.0
_C28	-2.3	-0.7	0.0	-1.6	0.0

Hyde

Project Tree Hyde

Hyde Atom Contributions

Atom	Hyde Score	Ligand		Receptor	
		Desolv.	Interact.	Desolv.	Interact.
_H13	-3.9	8.4	-11.4	6.4	-7.3
_C5	0.0	0.0	0.0	0.0	0.0
_C14	0.0	0.0	0.0	0.0	0.0
_O32	-8.4	7.5	-10.1	0.6	-6.3
_C33	-2.2	-1.4	0.0	-0.8	0.0
_C6	-2.1	-0.2	0.0	-1.9	0.0
_C8	-0.2	-0.2	0.0	0.0	0.0
_C12	-4.1	-1.9	0.0	-2.2	0.0
_C9	-0.5	-0.5	0.0	0.0	0.0
_C11	-3.4	-1.9	0.0	-1.5	0.0
_C10	-2.0	-1.2	0.0	-0.8	0.0
_H1	0.0	0.0	0.0	0.0	0.0
_C1	-4.0	-1.2	0.0	-2.8	0.0
_C7	0.0	0.0	0.0	0.0	0.0
_O22	-0.7	0.3	0.0	-1.0	0.0
_C2	-3.2	-1.5	0.0	-1.7	0.0
_C3	-3.5	-2.1	0.0	-1.4	0.0
_C4	-1.2	-0.7	0.0	-0.6	0.0
_H23	-2.8	7.1	-8.4	7.4	-8.0
_C24	-0.9	-0.8	0.0	-0.1	0.0
_C25	-0.2	-0.2	0.0	0.0	0.0
_C26	-0.4	-0.4	0.0	0.0	0.0
_C27	-4.1	-2.1	0.0	-2.1	0.0
_C28	-2.3	-0.7	0.0	-1.6	0.0

Abbildung A.1: HYDE in der LeadIT-GUI

die HYDE-Energie ΔG , die Einordnung dieses Wertes in eine Bindungsaffinitätsskala (von μM bis nM) und die *Ligand Efficiency* LE , die die HYDE-Energie normalisiert über die Schweratome des Liganden angibt (je höher der Wert desto besser) (Abbildung A.1). Weiterhin werden die Atombeiträge des Liganden inklusive der zugeordneten Atombeiträge der nahegelegenen Proteinatome angegeben. Diese sind in ihre Anteile nach Dehydratationsenergie (**Desolv**) und Wasserstoffbrücken-Energie (**Interact.**) unterteilt. Durch Anklicken in der Tabelle oder auf Atome im Anzeigefenster wird das entsprechende Atom und sein Beitrag hervorgehoben (Abbildung A.1).

A.2 HYDE als Screening-Tool

Der `hydescorer` ist das kommandozeilenbasierte HYDE-Screening-Tool, das als Eingabe ein Protein im `mol2` oder `fxn` (LeadIT Projektdatei) Format und einen Liganden oder mehrere Docking-Posen im `mol2` oder `sdf` Format fordert. Der Aufruf des Tools sieht folgendermaßen aus:

```
./hydescorer --target 2zda.fxn --input 2zda_crysth.mol2
```

Als Ausgabe erhält man eine Tabelle mit HYDE-Energien in der Kommandozeile. In diesem Fall wird der Protein-Ligand-Komplex auch automatisch optimiert und dann bewertet. Verschiedene Einstellungen sind über optionale Parameter möglich, diese werden im Folgenden beschrieben.

Obligatorisch:

--target

Protein Eingabedatei in `mol2` oder `fxn` Format

--input

Ligand oder Docking-Posen Eingabedatei in `mol2` or `sdf` Format

Optionen:

--reference

Referenz-Ligand Eingabedatei

--Hnetwork

Angabe, ob eine Optimierung des Wasserstoffbrücken-Netzwerkes in der Protein-Bindetasche durchgeführt werden soll. (1=eingeschaltet, 0=ausgeschaltet)

Voreinstellung: 1

--exposed

Schwellwert für den HYDE-exposed-Term. Hydrophobe Ligandatome, die aus der Bindetasche reichen werden bestraft. Atome, deren Oberflächenänderung kleiner oder gleich dieses Wertes sind, tragen ungünstig zur Energie bei.

Voreinstellung: 0.0

--optimize

Optimierungs-Routine (1=stochastisch, 2=numerisch, 0=ausgeschaltet)

Voreinstellung: numerisch (2)

--optimize-steps

Maximale Anzahl der Optimierungsschritte für die numerische Optimierung

Voreinstellung: 100

--sto-opt-steps

Maximale Anzahl der Optimierungsschritte für die stochastische Optimierung

Voreinstellung: 1000

--sto-opt-runs

Anzahl der Läufe der stochastischen Optimierung

Voreinstellung: 10

--output

Ausgabedatei-Format: **all**, **csv**, **score**, **table**

Voreinstellung: **--output table** **--output all**

--output-file

Name der Ausgabedatei der HYDE-Energiewerte. Argument: Dateiname

Voreinstellung: <Proteinname>.log

--output-target

Name der Ausgabedatei des optimierten Proteins (Wasserstoffbrücken-Netzwerk) als mol2. Argument: Dateiname

Voreinstellung: <Proteinname>_target.mol2

--output-ligand

Name der Ausgabedatei des optimierten Liganden. Argument: Dateiname und Dateiformat (ext=mol2 oder ext= sdf)

Voreinstellung: <Ligandname>.<ext>

--binding-site

Angabe, ob nur die Bindetasche des Proteins in die Ausgabedatei geschrieben werden soll. (on oder off)

Voreinstellung: off

--version

Zeigt die Versions Nummer des Tools an

--license-info

Zeigt die Lizenz Information an und ob eine Lizenz vorliegt

--help

Zeigt die Hilfefunktion des Tools an

Benutzung: '--help long' oder --help --<option name>

B

Implementierung

B.1 Softwarestruktur

Die HYDE-Software wurde in ANSI C implementiert und als Modul in die Flex*-Bibliothek [80] integriert. Abbildung B.1 zeigt die Struktur der Software.

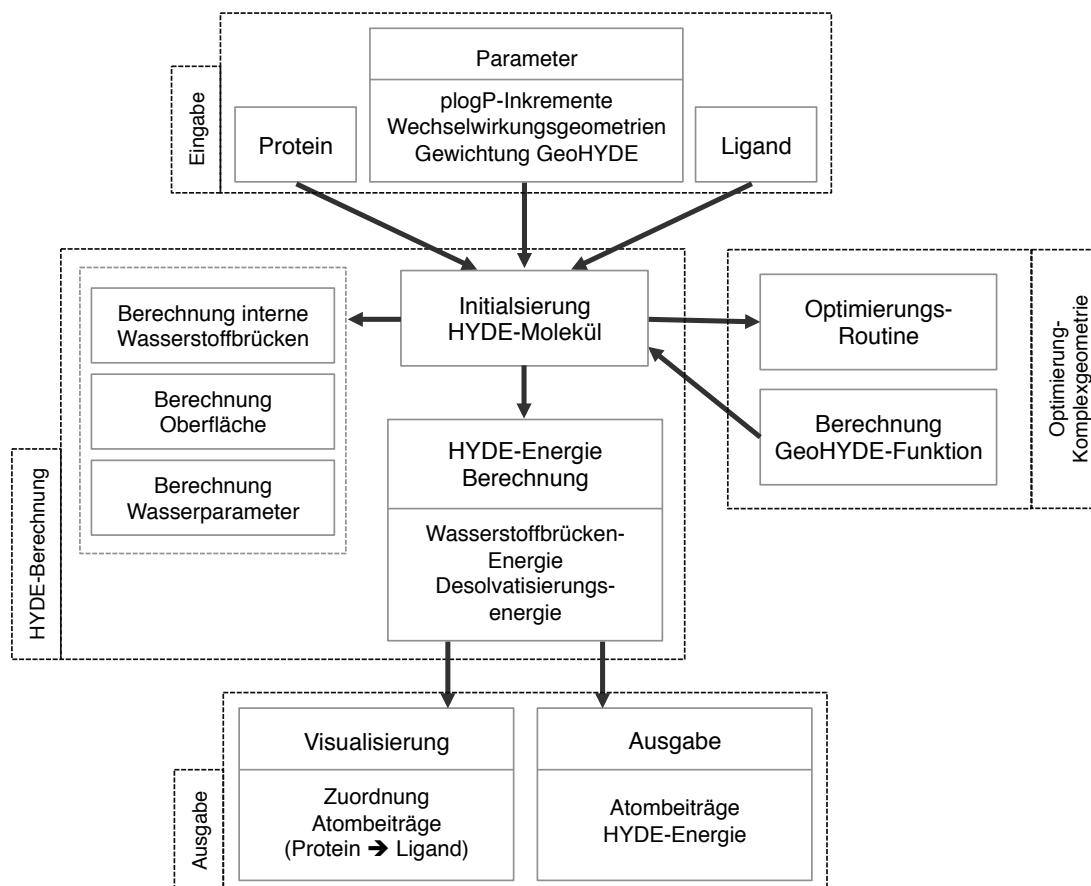


Abbildung B.1: Softwarestruktur von HYDE

C

Zusätzliche Resultate und Tabellen

C.1 DUD Enrichment: HYDE und HYDE-Prototyp

Zum Vergleich der überarbeiteten HYDE-Version und des HYDE-Prototyps sind die Enrichment-Ergebnisse beider Versionen auf dem DUD Datensatz dargestellt. Die Ergebnisse des HYDE-Prototyps wurden aus [202] entnommen.

Zur Erstellung des Enrichment-Plots werden die bereits gefundenen aktiven Moleküle (y-Achse) gegen die Anzahl der bereits durchsuchten Moleküle des Datensatzes (x-Achse) logarithmisch aufgetragen. Der untere graue Bereich stellt eine zufällige Auswahl der Moleküle dar, im optimalen Fall verläuft die Kurve entlang des oberen grauen Bereichs. Enrichment-Plots bieten eine alternative Darstellung von Screening-Ergebnissen.

C. Zusätzliche Resultate und Tabellen

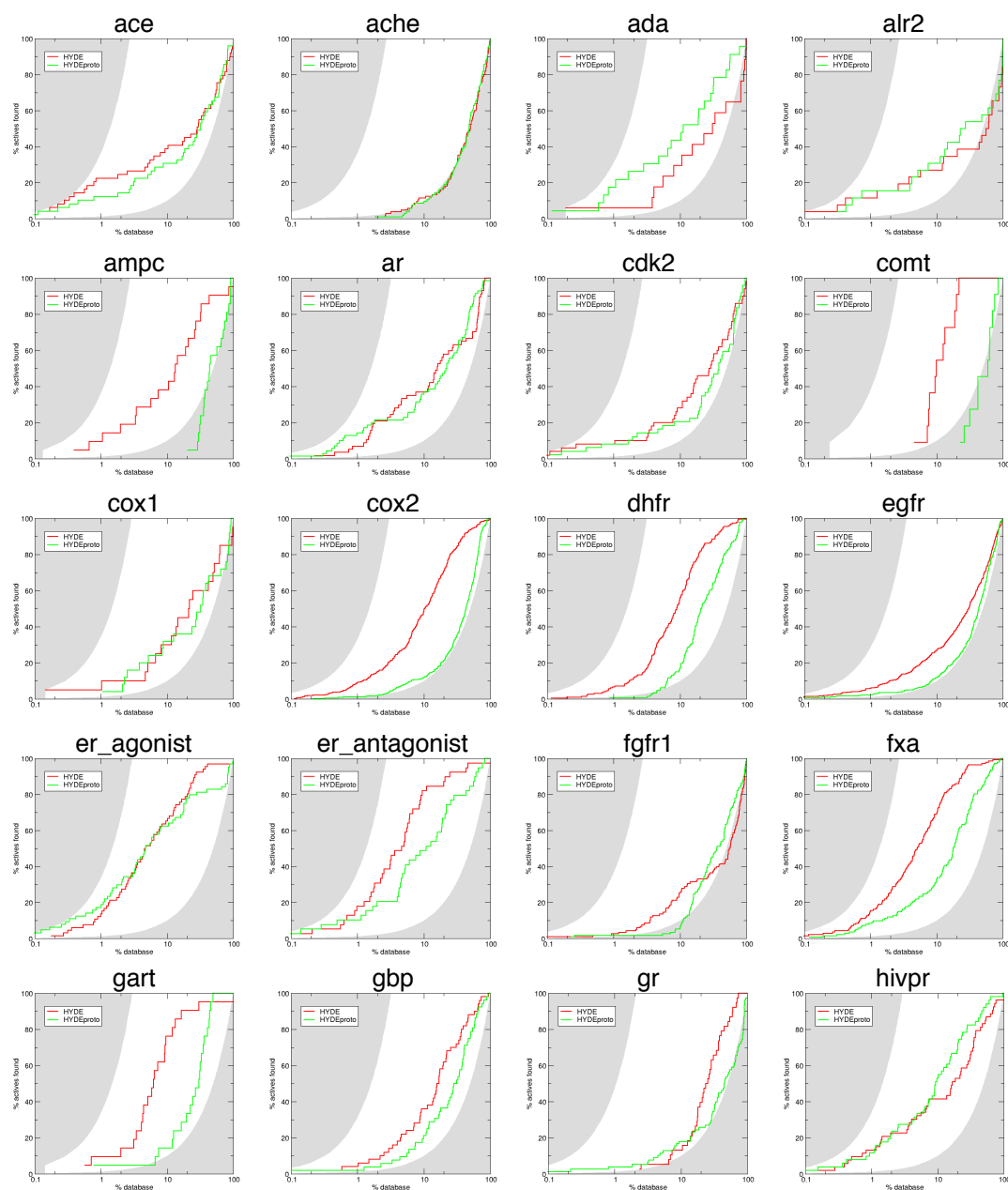


Abbildung C.1: Vergleich der Enrichment Ergebnisse DUD Datensatz. HYDE (rot) und HYDE-Prototyp (grün) (aus [202])

C.1 DUD Enrichment: HYDE und HYDE-Prototyp

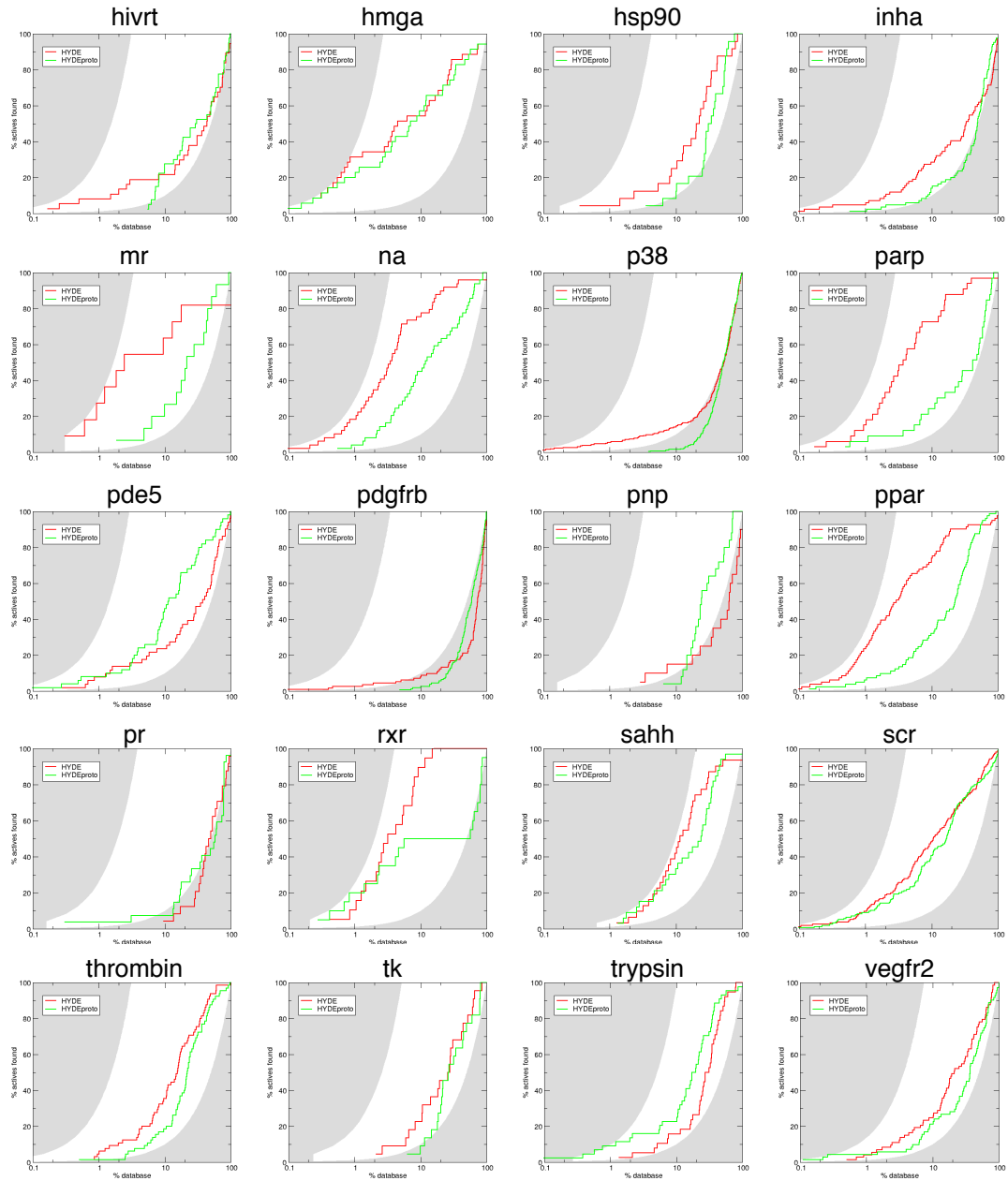


Abbildung C.2: Vergleich der Enrichment Ergebnisse DUD Datensatz (Fortsetzung). HYDE (rot) und HYDE-Prototyp (grün) (aus [202])

C.2 Strukturelle Klassifikation des PDBbind 2007 coreset

Klassifikations Kriterien

Die Klassifikation wurde aus [208] übernommen. Die meisten Kriterien beruhen auf der visuellen Untersuchung der Strukturen anhand der Elektronendichtekarten und sind daher subjektiv.

Tabelle C.1: Klassifikationskriterien

Elektronendichte Ligand (visuell) (Dichte Lig.)	1 = gute Dichte, keine Mängel 2 = kleinere Mängel, Dichte von einzelnen Atomen nicht vorhanden 3 = fehlenden Dichte an flexiblen Stellen des Liganden/ undefinierte Dichte im Liganden 4 = nur etwa 50 % der Dichte vorhanden 5 = weniger als 50 % der Dichte vorhanden
Elektronendichte Protein (visuell) (Dichte Prot.)	1 = gute Dichte, keine Mängel 2 = kleinere Mängel, Dichte von einzelnen Atomen/Aminosäuren nicht vorhanden 3 = fehlenden Dichte für größere Stellen Bindetasche (mehr als 3 Aminosäuren)
Alternative Konformationen Ligand/Bindetasche (Alt. Konf. Lig., Alt. Konf. Prot.)	1 = vorhanden 2 = nicht vorhanden
Temperaturfaktor Ligand/Bindetasche (Temp. Lig., Temp. Prot.)	1 = gleichmäßig 2 = höher für flexible/isolierte Teile 3 = höher für größere Teile (mehr als 3 Aminosäuren)
Kristallpackungseffekte (Kristallpackung)	1 = vorhanden 2 = nicht vorhanden
Viele Kristallwasser in der Tasche (Multiple Wasser)	1 = vorhanden 2 = nicht vorhanden

Klassifikation anhand der Kriterien:

$$\begin{aligned}
 QW = & (\text{Dichte Lig.} + \text{Dichte Prot.} \\
 & + \text{Alt. Konf. Lig.} + \text{Alt. Konf. Prot.} \\
 & + \text{Temp. Lig.} + \text{Temp. Prot.} \\
 & + (\text{Kristallpackung})^2 + \text{Multiple Wasser}) / \text{Anzahl der Kriterien}
 \end{aligned}$$

Die Anzahl der Kriterien hängt von der Verfügbarkeit der Elektronendichtekarten bzw. der Temperaturfaktoren ab.

Klassifikation:

$$QW < 1,25 \quad QK = 1$$

$$QW < 1,5 \quad QK = 2$$

$$QW < 1,75 \quad QK = 3$$

$$QW < 2 \quad QK = 4$$

$$QW \geq 2 \quad QK = 5$$

Tabelle C.2: Strukturelle Klassifikation des PDBbind 2007 coreset Datensatzes

PDB	RFree	Auflös.	Dichte		Alt. Konf.		Temp.		Kristall Packung	Multiple Wasser	QW	QK
			Lig.	Prot.	Lig.	Prot.	Lig.	Prot.				
1tok	0,227	1,85	1	1	1	1	1	1	1	1	1	1
2baj	0,303	2,25	1	1	1	1	1	1	1	1	1	1
2hdq	0,227	2,1	1	1	1	1	1	1	1	1	1	1
2h3e	0,25	2,3	1	1	1	1	1	1	1	1	1	1
1d09	0,234	2,1	1	1	1	1	1	1	1	1	1	1
2ctc	N.A.	1,4	1	1	1	1	1	1	1	1	1	1
1fcz	0,174	1,38	1	1	1	1	1	1	1	1	1	1
1hfs	0,217	1,7	1	1	1	1	1	1	1	1	1	1
2bz6	0,193	1,6	1	1	1	1	1	1	1	1	1	1
1nc1	0,223	2	1	1	1	1	1	1	1	1	1	1
1y6q	0,246	2,2	1	1	1	1	1	1	1	1	1	1
1df8	0,225	1,51	1	1	1	1	1	1	1	1	1	1
1olu	0,2	1,9	1	1	1	2	1	1	1	1	1,13	1
1ai5	0,222	2,36	2	1	1	1	1	1	1	1	1,13	1
1ajq	0,233	2,05	1	1	2	1	1	1	1	1	1,13	1
1e66	0,205	2,1	1	1	1	1	1	1	1	2	1,13	1

Fortsetzung auf der nächsten Seite

C. Zusätzliche Resultate und Tabellen

Tabelle C.2: (Fortsetzung der vorherigen Seite)

PDB	RFree	Auflös.	Dichte		Alt. Konf.		Temp.		Kristall Packung	Multiple Wasser	QW	QK
			Lig.	Prot.	Lig.	Prot.	Lig.	Prot.				
1b9j	0,219	1,8	1	1	1	1	1	1	1	2	1,13	1
2j77	0,258	2,1	1	1	1	1	1	1	1	2	1,13	1
2j78	0,201	1,65	1	1	1	1	1	1	1	2	1,13	1
2cet	0,244	1,97	2	1	1	1	1	1	1	1	1,13	1
1tmn	N.A.	1,9	2	1	1	1	1	1	1	1	1,13	1
1fh8	0,224	1,95	1	1	1	1	1	1	1	2	1,13	1
1vzq	0,209	1,54	1	2	1	1	1	1	1	1	1,13	1
1y1m	0,224	1,8	1	1	1	1	1	1	1	2	1,13	1
1p1q	0,27	2	1	1	1	1	1	1	1	2	1,13	1
1syh	0,199	1,8	1	1	1	1	1	1	1	2	1,13	1
1fcx	0,167	1,47	2	1	1	1	1	1	1	1	1,13	1
1j16	N.A.	1,6	1	2	1	1	1	1	1	1	1,13	1
1fh7	0,234	1,82	1	2	1	1	1	1	1	2	1,25	2
1fh9	0,219	1,72	1	2	1	1	1	1	1	2	1,25	2
1o3f	0,213	1,55	2	1	1	1	2	1	1	1	1,25	2
1gpk	0,2135	2,1	1	1	1	1	1	2	1	2	1,25	2
2ceq	0,215	2,14	1	2	1	1	1	1	1	2	1,25	2
1toi	0,197	1,9	1	2	1	1	1	2	1	1	1,25	2
1l2s	0,207	1,94	2	2	1	1	1	1	1	1	1,25	2
1fd0	0,158	1,38	2	1	2	1	1	1	1	1	1,25	2
2b1v	0,238	1,8	2	1	1	1	2	1	1	1	1,25	2
2b7d	0,289	2,24	1	2	1	1	1	2	1	1	1,25	2
1l0l	0,229	1,9	1	2	1	1	1	1	1	2	1,25	2
2f01	0,174	0,85	2	1	2	1	1	1	1	1	1,25	2
1d7j	0,254	1,85	2	1	2	1	1	1	1	1	1,25	2
2fdp	0,254	2,5	2	2	1	1	1	1	1	2	1,38	2
2flb	0,28	1,95	2	1	1	2	1	2	1	1	1,38	2
1zc9	0,237	2	1	1	1	1	1	1	2	1	1,38	2
1loq	0,198	1,5	1	1	1	1	1	1	2	1	1,38	2
1a1b	N.A.	2,2	1	1	1	1	1	1	2	1	1,38	2
1ols	0,199	1,85	1	2	1	2	1	2	1	1	1,38	2
1n2v	0,233	2,1	2	2	1	1	1	2	1	1	1,38	2
2brb	0,272	2,1	1	1	1	1	2	3	1	1	1,38	2
1nvq	0,234	2	1	3	1	1	1	2	1	1	1,38	2
1kv5	0,175	1,65	3	1	1	1	1	1	1	2	1,38	2
2gss	0,229	1,9	3	1	1	1	1	1	1	2	1,38	2
2bzz	N.A.	0,98	4	1	1	1	1	1	1	1	1,38	2
1bxo	0,125	0,95	1	1	1	1	1	1	2	2	1,5	3
2i0d	0,216	1,95	1	2	1	1	1	1	2	1	1,5	3
1u33	0,227	1,95	1	1	1	1	1	1	2	2	1,5	3
1fzj	0,222	1,6	2	1	1	1	1	1	2	1	1,5	3
1fzk	0,213	1,7	2	1	1	1	1	1	2	1	1,5	3
2d1o	0,206	2,02	2	1	1	1	1	1	2	1	1,5	3
1o0h	0,221	1,2	1	1	1	1	1	1	2	2	1,5	3
1is0	0,272	1,9	2	1	1	1	1	1	2	1	1,5	3
2d3z	0,232	1,8	2	2	1	1	2	2	1	1	1,5	3
2d3u	0,245	2	2	2	1	1	2	2	1	1	1,5	3

Fortsetzung auf der nächsten Seite

Tabelle C.2: (Fortsetzung der vorherigen Seite)

PDB	RFree	Auflös.	Dichte		Alt. Konf.		Temp.		Kristall Packung	Multiple Wasser	QW	QK
			Lig.	Prot.	Lig.	Prot.	Lig.	Prot.				
1h23	0,2148	2,15	2	1	1	1	3	1	1	2	1,5	3
1s39	0,242	1,95	1	3	1	1	1	3	1	1	1,5	3
2azr	0,25	2,4	3	2	1	1	2	1	1	1	1,5	3
1g7f	0,249	2,35	3	1	1	1	2	1	1	2	1,5	3
2brm	0,281	2,2	1	3	1	1	1	3	1	1	1,5	3
2ayr	0,263	1,9	1	3	1	1	1	3	1	1	1,5	3
2d0k	0,284	1,9	3	1	1	1	2	2	1	1	1,5	3
2c02	0,228	2	4	1	1	1	2	1	1	1	1,5	3
1rnt	N.A.	1,9	3	2	1	1	1	1	1	2	1,5	3
1zs0	0,235	1,56	2	2	1	1	1	1	2	1	1,63	3
1v48	0,242	2,2	1	2	1	1	1	2	2	1	1,63	3
1o3p	0,249	1,81	1	2	1	2	1	1	2	1	1,63	3
1a08	N.A.	2,2	1	2	1	1	1	2	2	1	1,63	3
1det	N.A.	1,8	2	1	1	1	1	1	2	2	1,63	3
1ha2	0,247	2,5	2	2	1	1	2	3	1	1	1,63	3
1toj	0,201	1,9	2	2	1	1	2	3	1	1	1,63	3
1xgj	0,215	1,97	3	1	1	1	3	1	1	2	1,63	3
1q8t	0,229	2	1	2	1	1	3	3	1	1	1,63	3
1bcu	0,212	2	2	2	1	1	3	2	1	1	1,63	3
2fai	0,24	2,1	2	1	1	1	3	3	1	1	1,63	3
1ttm	0,21	1,95	4	2	1	1	1	2	1	1	1,63	3
1a30	0,227	2	3	2	1	1	3	1	1	1	1,63	3
2f80	0,218	1,45	3	2	2	2	1	1	1	1	1,63	3
1pr5	0,259	2,5	2	3	1	1	3	2	1	1	1,75	4
1g7q	0,221	2,5	3	2	1	1	1	1	2	1	1,75	4
1zvx	0,244	1,87	1	3	1	1	1	2	2	1	1,75	4
1nhu	0,256	2	4	2	1	1	2	2	1	1	1,75	4
4tmn	N.A.	1,7	3	1	1	1	3	2	1	2	1,75	4
1mq6	0,268	2,1	2	2	1	1	3	3	1	1	1,75	4
1v2o	0,196	1,62	4	1	1	1	2	1	2	1	1,88	4
1hk4	0,267	2,4	5	2	1	1	2	2	1	1	1,88	4
1fkn	0,224	1,9	4	1	1	1	3	2	1	2	1,88	4
1k4g	0,241	1,7	4	3	1	1	2	2	1	1	1,88	4
1b39	0,27	2,1	3	3	1	1	2	3	1	1	1,88	4
1gni	0,242	2,4	5	2	1	1	3	2	1	1	2	5
2cer	0,246	2,29	3	2	2	1	2	3	1	2	2	5
2bak	0,25	2,2	3	3	1	1	3	3	1	1	2	5
1utp	0,183	1,3	5	1	1	2	3	1	1	2	2	5
1ajp	0,225	2,31	5	3	1	2	2	2	1	1	2,13	5
1u2y	0,198	1,95	5	2	1	1	3	2	1	2	2,13	5
1xd1	0,203	2,2	4	3	1	1	3	3	1	1	2,13	5
2fzc	0,25	2,1	5	3	1	1	3	2	1	1	2,13	5
1pxo	0,235	1,96	5	3	2	2	1	2	1	2	2,25	5
1j17	N.A.	2	4	3	1	1	3	3	2	1	2,5	5
2qwb	N.A.	2	N.A.	N.A.	1	1	1	1	1	1	1	1
2qwd	N.A.	2	N.A.	N.A.	1	1	1	1	1	1	1	1
2qwe	N.A.	2	N.A.	N.A.	1	1	1	1	1	1	1	1

Fortsetzung auf der nächsten Seite

C. Zusätzliche Resultate und Tabellen

Tabelle C.2: (Fortsetzung der vorherigen Seite)

PDB	RFree	Auflös.	Dichte		Alt. Konf.		Temp.		Kristall Packung	Multiple Wasser	QW	QK
			Lig.	Prot.	Lig.	Prot.	Lig.	Prot.				
1v16	0,17	1,9	N.A.	N.A.	1	1	1	1	1	1	1	1
1sl3	0,24	1,81	N.A.	N.A.	1	1	1	1	1	1	1	1
1pb9	0,227	1,6	N.A.	N.A.	1	1	1	1	1	1	1	1
1pbq	0,27	1,9	N.A.	N.A.	1	1	1	1	1	1	1	1
1b8o	0,232	1,5	N.A.	N.A.	1	1	1	1	1	1	1	1
1ftm	0,247	1,7	N.A.	N.A.	1	1	1	1	1	1	1	1
4tim	N.A.	2,4	N.A.	N.A.	1	1	1	1	1	1	1	1
1k9s	0,184	2	N.A.	N.A.	1	1	1	1	1	1	1	1
3pce	N.A.	2,06	N.A.	N.A.	1	1	1	1	1	1	1	1
3pch	N.A.	2,05	N.A.	N.A.	1	1	1	1	1	1	1	1
3pcj	N.A.	2,13	N.A.	N.A.	1	1	1	1	1	1	1	1
1pz5	0,254	1,8	N.A.	N.A.	1	1	1	1	1	1	1	1
2cgr	N.A.	2,2	N.A.	N.A.	1	1	1	1	1	1	1	1
10gs	0,214	2,2	N.A.	N.A.	1	1	1	1	1	1	1	1
3std	0,23	1,65	N.A.	N.A.	1	1	1	1	1	1	1	1
1uwt	0,239	1,95	N.A.	N.A.	1	1	1	1	1	2	1,17	1
2g94	0,227	1,86	N.A.	N.A.	1	1	1	1	1	2	1,17	1
2rkm	0,219	1,8	N.A.	N.A.	1	1	1	1	1	2	1,17	1
1b7h	0,26	2	N.A.	N.A.	1	1	1	1	1	2	1,17	1
1m0n	0,239	2,2	N.A.	N.A.	1	1	2	1	1	1	1,17	1
1kv1	0,284	2,5	N.A.	N.A.	1	1	1	2	1	1	1,17	1
1ndw	0,218	2	N.A.	N.A.	1	1	1	1	1	2	1,17	1
1ndy	0,21	2	N.A.	N.A.	1	1	1	1	1	2	1,17	1
1ndz	0,225	2	N.A.	N.A.	1	1	1	1	1	2	1,17	1
1ydt	N.A.	2,3	N.A.	N.A.	1	1	1	1	1	2	1,17	1
1re8	0,252	2,1	N.A.	N.A.	1	1	1	1	1	2	1,17	1
1zoe	0,277	1,77	N.A.	N.A.	1	1	1	1	1	2	1,17	1
1nny	0,238	2,4	N.A.	N.A.	1	1	1	1	1	2	1,17	1
1tsy	N.A.	2,2	N.A.	N.A.	1	1	1	1	1	2	1,17	1
1nja	N.A.	2,5	N.A.	N.A.	1	1	1	1	1	2	1,17	1
8cpa	N.A.	2	N.A.	N.A.	1	1	1	1	1	2	1,17	1
7cpa	N.A.	2	N.A.	N.A.	1	1	1	1	1	2	1,17	1
1jqd	0,257	2,28	N.A.	N.A.	1	1	1	1	1	2	1,17	1
1f4f	0,251	2	N.A.	N.A.	1	1	1	1	1	2	1,17	1
1f5k	0,257	1,8	N.A.	N.A.	1	1	1	1	1	2	1,17	1
2bok	0,237	1,64	N.A.	N.A.	1	1	1	1	1	2	1,17	1
1nfy	N.A.	2,1	N.A.	N.A.	1	1	1	1	1	2	1,17	1
1bra	N.A.	2,2	N.A.	N.A.	1	1	2	1	1	1	1,17	1
6std	0,224	1,8	N.A.	N.A.	1	1	1	1	1	2	1,17	1
1hi4	0,27	1,8	N.A.	N.A.	1	1	1	1	1	2	1,17	1
2g5u	0,22	1,8	N.A.	N.A.	1	1	2	1	1	1	1,17	1
5er1	N.A.	2	N.A.	N.A.	1	1	N.A.	N.A.	1	2	1,25	2
2er9	N.A.	2,2	N.A.	N.A.	1	1	N.A.	N.A.	1	2	1,25	2
4er2	N.A.	2	N.A.	N.A.	1	1	N.A.	N.A.	1	2	1,25	2
1x1z	0,18	1,45	N.A.	N.A.	1	2	1	1	1	2	1,33	2
4tln	N.A.	2,3	N.A.	N.A.	1	1	2	2	1	1	1,33	2
2aov	0,254	1,91	N.A.	N.A.	1	1	1	2	1	2	1,33	2

Fortsetzung auf der nächsten Seite

Tabelle C.2: (Fortsetzung der vorherigen Seite)

PDB	RFree	Auflös.	Dichte		Alt. Konf.		Temp.		Kristall Packung	Multiple Wasser	QW	QK
			Lig.	Prot.	Lig.	Prot.	Lig.	Prot.				
1vfn	N.A.	2,15	N.A.	N.A.	1	1	1	3	1	1	1,33	2
1f4g	0,214	1,75	N.A.	N.A.	2	1	1	1	1	2	1,33	2
1a69	N.A.	2,1	N.A.	N.A.	1	1	2	2	1	1	1,33	2
3gss	0,233	1,9	N.A.	N.A.	1	1	2	1	1	2	1,33	2
1e5a	0,262	1,8	N.A.	N.A.	2	1	2	1	1	1	1,33	2
1jq9	0,225	1,8	N.A.	N.A.	1	1	2	2	1	1	1,33	2
1jaq	N.A.	2,25	N.A.	N.A.	1	1	1	1	2	1	1,5	3
1dhi	N.A.	1,9	N.A.	N.A.	1	1	2	2	1	2	1,5	3
1sqa	0,266	2	N.A.	N.A.	1	1	1	1	2	1	1,5	3
1fki	N.A.	2,2	N.A.	N.A.	1	1	1	1	2	1	1,5	3
1fkb	N.A.	1,7	N.A.	N.A.	1	1	1	1	2	1	1,5	3
1m2q	0,232	1,79	N.A.	N.A.	1	1	1	3	1	2	1,5	3
2std	0,259	2,1	N.A.	N.A.	1	1	1	3	1	2	1,5	3
2drc	N.A.	1,9	N.A.	N.A.	1	1	2	1	2	1	1,67	3
1ppm	N.A.	1,7	N.A.	N.A.	1	1	1	1	2	2	1,67	3
1apw	N.A.	1,8	N.A.	N.A.	1	1	1	1	2	2	1,67	3
2usn	N.A.	2,2	N.A.	N.A.	1	1	2	1	2	1	1,67	3
1bma	N.A.	1,8	N.A.	N.A.	1	1	2	1	2	1	1,67	3
1elb	N.A.	2,1	N.A.	N.A.	1	1	2	1	2	1	1,67	3
1tyr	N.A.	1,8	N.A.	N.A.	2	1	1	1	2	1	1,67	3
6rnt	N.A.	1,8	N.A.	N.A.	1	1	1	1	2	2	1,67	3
2aou	0,258	2,3	N.A.	N.A.	1	1	3	3	1	1	1,67	3
1f4e	0,238	1,9	N.A.	N.A.	1	1	3	2	1	2	1,67	3
1trd	N.A.	2,5	N.A.	N.A.	1	1	3	3	1	1	1,67	3
1sv3	0,243	1,35	N.A.	N.A.	1	1	3	2	1	2	1,67	3
1m0q	0,24	2	N.A.	N.A.	1	1	2	1	2	1	1,67	3
1flr	N.A.	1,85	N.A.	N.A.	1	1	1	3	2	1	1,83	4
1if7	0,294	1,98	N.A.	N.A.	1	1	3	1	2	1	1,83	4
1avn	N.A.	2	N.A.	N.A.	1	1	2	1	2	2	1,83	4
2g8r	0,23	1,7	N.A.	N.A.	1	1	2	1	2	2	1,83	4
1ulb	0,258	2	N.A.	N.A.	1	1	3	1	2	1	1,83	4
1om1	0,227	1,68	N.A.	N.A.	2	1	3	3	1	1	1,83	4
1jys	0,239	1,9	N.A.	N.A.	1	1	3	3	1	2	1,83	4
1e1v	0,269	1,95	N.A.	N.A.	1	1	1	3	2	2	2	5
1ela	N.A.	1,8	N.A.	N.A.	1	1	3	1	2	2	2	5
1nje	N.A.	2,3	N.A.	N.A.	1	1	3	2	2	2	2,17	5
1slg	0,222	1,76	N.A.	N.A.	2	1	3	2	2	1	2,17	5
1q7a	0,213	1,6	N.A.	N.A.	1	1	3	3	2	1	2,17	5

N.A. = nicht verfügbar

C.3 Strukturelle Klassifikation des DUD Datensatzes

Tabelle C.3: Strukturelle Klassifikation des DUD Datensatzes. Tabelle aus [266]

Zielprotein	Kommentar	Qualität	PDB ID	Auflösung
ada	Elektronendichtekarte nicht vorhanden, wichtiges strukturelles Wasser (HOH 1149)	Q 1	1ndw	2
ar		Q 1	1xq2	1,9
cox1	ARG 120 sollte leicht rotiert werden um einen Überlapp von Liganden mit der Bindetasche zu vermeiden	Q 1	1q4g	2,1
er-agonist	Elektronendichtekarte nicht vorhanden	Q 1	1l2i	1,9
er-antagonist	Elektronendichtekarte nicht vorhanden	Q 1	3ert	1,9
fxa		Q 1	1f0r	2,7
gpb		Q 1	1a8i	1,8
gr		Q 1	1m2z	2,5
hivpr	Elektronendichtekarte nicht vorhanden	Q 1	1hpx	2
hivrt	Elektronendichtekarte nicht vorhanden, ausschließlich hydrophobe Bindetasche	Q 1	1rt1	2,6
hmga	wichtiges strukturelles Wasser (HOH 1003)	Q 1	1hw8	2,1
hsp90	wasserüberbrückte Wechselwirkungen zwischen Protein und Ligand (HOH 2059, HOH 2137)	Q 1	1uy6	1,9
mr		Q 1	2aa2	1,9
na	Elektronendichtekarte nicht vorhanden, wichtiges strukturelles Wasser (HOH 713)	Q 1	1a4g	2,2
p38		Q 1	1kv2	2,8
parp		Q 1	1efy	2,2
ppar		Q 1	1fm9	2,1
pr		Q 1	1sr7	1,9
rxr	Elektronendichtekarte nicht vorhanden	Q 1	1mvc	1,9
sahh		Q 1	1a7a	2,8
ace	Elektronendichtekarte nicht vorhanden, viele Wasser in der Bindetasche	Q 2	1o86	2
ache	symmetrieverwandtes Molekül nahe der Bindetasche, nur wasserüberbrückte Wechselwirkungen zwischen Protein und Ligand	Q 2	1eve	2,5
alr2	fehlende Elektronendichte und hohe Temperaturfaktoren für Teile der Bindetasche	Q 2	1ah3	2,3
ampc	Hohe Temperaturfaktoren des Liganden	Q 2	1xgj	2
pde5	Hohe Temperaturfaktoren des Liganden, wasserüberbrückte Wechselwirkung zwischen Protein und Ligand (HOH 1007)	Q 3	1xp0	1,8
cdk2	Elektronendichtekarte nicht vorhanden, Konformation von HIS 84 A in der Hinge-Region ist durch einen Kristallpackungseffekt beeinflusst	Q 3	1ckp	2,1

Fortsetzung auf der nächsten Seite

Tabelle C.3: (Fortsetzung der vorherigen Seite)

Zielprotein	Kommentar	Qualität	PDB ID	Auflösung
cox2	Ungeordnete Arginin-Reste in der Bindetasche	Q 3	1cx2	3
dhfr	sehr alte Struktur (1982), Elektronendichtekarte nicht vorhanden	Q 3	3dfr	1,7
gart	fehlende Elektronendichte	Q 3	1c2t	2,1
pnp	Elektronendichtekarte nicht vorhanden, symmetrieverwandtes Molekül nahe der Bindetasche, Phosphat-Molekül beeinflusst die Anordnung verschiedener Seitenketten	Q 3	1b8o	1,5
src	Elektronendichtekarte nicht vorhanden, sehr hohe Temperaturfaktoren der Bindetasche	Q 3	2src	1,5
tk	Phosphat-Molekül beeinflusst die Anordnung verschiedener Seitenketten unterhalb des Liganden	Q 3	1kim	2,1
trypsin	Phosphat-Molekül beeinflusst die Anordnung verschiedener Seitenketten	Q 3	1bjv	1,8
comt	symmetrieverwandtes Molekül nahe der Bindetasche, Kristallpackungseffekt	Q 4	1h1d	2
egfr	Kristallpackungseffekt, hohe Temperaturfaktoren und multiple Konformationen des Liganden, Soaking	Q 4	1m17	2,1
fgfr1	Hohe Temperaturfaktoren des Liganden, Soaking	Q 4	1agw	2,4
inha	fehlende Elektronendichte, Hohe Temperaturfaktoren des Liganden, Soaking	Q 4	1p44	2,7
pdfgrb	Homologie-Model	Q 4	-	n.a
thrombin	kovalent gebundener Inhibitor verursacht verzerrte Struktur der Bindetasche	Q 4	1ba8	1,8
vegfr2	Apo-Struktur mit modelliertem Liganden	Q 4	1vr2	2,4

D

Veröffentlichungen

D.1 Veröffentlichungen in wissenschaftlichen Zeitschriften

1. N. Schneider, S. Hindle, G. Lange, R. Klein, J. Albrecht, H. Briem, K. Beyer, H. Claußen, M. Gastreich, C. Lemmen, M. Rarey. Substantial improvements in large-scale redocking and screening using the novel HYDE scoring function. *Journal of Computer-Aided Molecular Design* 2012, *26(6)*: 701-723
2. N. Schneider, R. Klein, G. Lange, M. Rarey. Nearly no Scoring Function without a Hansch-Analysis. *Molecular Informatics* 2012, *31(6-7)*: 503-507
3. N. Schneider, G. Lange, S. Hindle, R. Klein, M. Rarey. A Consistent Description of Hydrogen Bond and Dehydration Energies in Protein-Ligand Complexes: Methods Behind the HYDE Scoring Function. *Journal of Computer-Aided Molecular Design*. *eingereicht*
4. G. Lange, N. Schneider, M. Rarey, R. Klein. Impact of unsatisfied hydrogen bond functions on intermolecular interactions. *in Vorbereitung*
5. E. Vennmann, N. Schneider, G. Lange, M. Rarey. Discrimination of permanent and transient heterodimers: A structurally unbiased classification of protein-protein interactions. *in Vorbereitung*

D.2 Buchkapitel

1. A. Kolodzik, N. Schneider, and M. Rarey, Structure Based Virtual Screening, Handbook of Cheminformatics

D.3 Vorträge

1. N. Schneider, G. Lange, R. Klein, C. Lemmen, M. Rarey, HYDE - Scoring for Lead Optimization, 5th Joint Sheffield Conference on Chemoinformatics, 2010, Sheffield, GB
2. N. Schneider, G. Lange, R. Klein, S. Hindle, C. Lemmen, M. Rarey, HYDEing the false positives: Scoring for lead optimization, 241st ACS National Meeting, 2011, Anaheim, CA
3. N. Schneider, G. Lange, R. Klein, C. Lemmen, M. Rarey, HYDE: A consistent Description of HYdration and DESolvation in Scoring of Protein-Ligand Complexes, Gordon Research Conference on Computer Aided Drug Design, 2011, West Dover, VT (invited Poster talk)
4. N. Schneider, G. Lange, R. Klein, C. Lemmen, M. Rarey, Consistent Estimation of Hydrogen Bond and Dehydration Energy using the HYDE Scoring Function, International Workshop, New Approaches in Drug Design and Discovery, 2012, Schloss Rauischholzhausen, Marburg (invited talk)
5. N. Schneider, S. Hindle, G. Lange, R. Klein, C. Lemmen, M. Rarey, HYDE in practice: Understanding protein-ligand interactions, 243rd ACS National Meeting, 2012, San Diego, CA
6. E. Vennmann, N. Schneider, G. Lange, M. Rarey, Discrimination of permanent and transient heterodimers: A real challenge in the classification of protein-protein interactions, 243rd ACS National Meeting, 2012, San Diego, CA

D.4 Poster

1. N. Schneider, G. Lange, R. Klein, C. Lemmen, M. Rarey, HYDEing the false positives - Scoring for lead optimization, 6th German Conference on Chemoinformatics, 2010, Goslar

2. N. Schneider, G. Lange, R. Klein, C. Lemmen, M. Rarey, HYDE: A consistent Description of HYdration and DEsolvation in Scoring of Protein-Ligand Complexes, Gordon Research Conference on Computer Aided Drug Design, 2011, West Dover, VT

