

# **Computersimulationen zur Untersuchung von Wassermolekülen in Protein-Ligand Komplexen am Beispiel einer Modellbindetasche**

Dissertation zur Erlangung des naturwissenschaftlichen Doktorgrads  
der  
Julius-Maximilians-Universität Würzburg

vorgelegt von  
Daniel Cappel  
aus Bad Homburg v. d. Höhe

Mannheim, Januar 2011

Eingereicht bei der Fakultät für Chemie und Pharmazie am \_\_\_\_\_

Gutachter der schriftlichen Arbeit

1. Gutachter: \_\_\_\_\_

2. Gutachter: \_\_\_\_\_

Prüfer des öffentlichen Promotionskolloquiums

1. Prüfer: \_\_\_\_\_

2. Prüfer: \_\_\_\_\_

3. Prüfer: \_\_\_\_\_

Datum des öffentlichen Promotionskolloquiums \_\_\_\_\_

Doktorurkunde ausgehändigt am \_\_\_\_\_

Für meine liebe Frau



# Vorwort

Die Untersuchungen zur vorliegenden Arbeit wurden auf Anregung von Herrn Prof. Dr. C. A. Sotriffer am Institut für Pharmazie und Lebensmittelchemie der Julius-Maximilians-Universität Würzburg in der Zeit von Oktober 2006 bis März 2010 durchgeführt. Sie umfasst eine ausführliche Analyse struktureller, dynamischer sowie energetischer Eigenschaften von Wassermolekülen in unterschiedlichen Komplexen einer Modellbindetasche mittels verschiedener Computersimulationsmethoden.

Im Rahmen von Kooperationen wurden in diesem Zeitraum zusätzlich weitere Untersuchungen durchgeführt, die nicht Gegenstand der vorliegenden Arbeit sind. Dies war insbesondere die Simulation der Dynamik von Komplexen der Enoyl-ACP Reductase (InhA) mittels MD-Simulationen in Zusammenarbeit mit Benjamin Schaefer. Daraus ist ein Posterbeitrag hervorgegangen:

- B. Schaefer, D. Cappel, C. A. Sotriffer, *Docking Studies and MD Simulations on Mycobacterium Tuberculosis Enoyl-ACP Reductase (InhA)*. 4th Joint Ph.D. Students Meeting of the Collaborative Research Centers SFB 630, SFB 766, SFB 544, Bronnbach (2008)

Desweiteren wurden, ebenso mit Hilfe von MD-Simulationen, die Struktur und Dynamik der *Leishmania Major* Trypanothion Synthetase-Amidase untersucht. Dies geschah in Zusammenarbeit mit Dr. Oliver Koch und den Firmen Intervet Innovation GmbH (Schwabenheim) sowie MOLISA GmbH (Magdeburg). Im Rahmen dieses Projektes entstanden zwei gemeinsame Vorträge.

- O. Koch, D. Cappel, T. Jaeger, L. Flohé, C. Sotriffer, P. Selzer, *Having a closer look into trypanothione biosynthesis – A computational analysis using molecular dynamics simulations*. 2nd Joint Expert Meeting - COST Action CM0801 „New drugs for neglected Diseases“ Siena, Italien (28.–29. Mai 2010).
- O. Koch, D. Cappel, M. Nocker, T. Jaeger, L. Flohé, C. Sotriffer, P. Selzer, *Virtual Screening Using Structure-based Consensus Pharmacophore Models and Ensemble Docking Based on MD-generated Conformations*. 6. German Conference on Chemoinformatics, Goslar (2010).

Schließlich wurden zusammen mit Claudia Raffauf die Solvatationseigenschaften verschiedener Arzneistoffe mittels MD-Simulationen sowie TI-Rechnungen analysiert.

Energieangaben in der vorliegenden Arbeit wurden zum überwiegenden Teil in der Einheit Kalorie vorgenommen, Längenangaben häufig in der Einheit Å. Dem Autor ist bewusst, dass dies nicht dem internationalen Einheitensystem (Système international d'unités) entspricht. Dennoch wurden diese Größen gewählt, da dies der Größenordnung der Genauigkeit bei Energieberechnungen entspricht bzw. den atomaren Bindungsmaßstäben am nächsten kommt. Für die Umrechnung in SI-Einheiten gilt:  $1 \text{ cal} = 4.1868 \text{ J}$  bzw.  $1 \text{ Å} = 10^{-10} \text{ m}$ .

Teilergebnisse dieser Arbeit waren Gegenstand von Veröffentlichungen sowie von Postern und einem Kurzvortrag.

### Journalbeiträge:

- D. Cappel, R. Wahlström, R. Brenk, C. A. Sotriffer, *Probing the dynamic nature of water molecules and their influences on ligand binding in a model binding site*. Journal of Molecular Biology, submitted.
- D. Cappel, C. A. Sotriffer, *Probing dynamic water molecules with rigid-ligand complexes*. Chemistry Central Journal, 3(Suppl 1), O20, 2009.

### Vortrag:

- D. Cappel, C. A. Sotriffer, *Probing dynamic water molecules with rigid-ligand complexes*. 4. German Conference on Chemoinformatics, Goslar (2008).

### Posterbeiträge:

- D. Cappel, R. Wahlström, R. Brenk, C. A. Sotriffer, *Water Molecules in Computer-Aided Drug Design: Method Evaluation on a Model Binding Site*. 2nd International Symposium of the Collaborative Research Center 630, Würzburg (2009).
- D. Cappel, C. Sotriffer, *Effect of Water Molecules on Protein-Ligand Interactions in a Charged Model Binding Site*. 21st "Darmstadt" Molecular Modeling Workshop, Erlangen (2007).
- D. Cappel, C. Sotriffer, *Dynamics and Interactions of Water Molecules in Protein-Ligand Complexes*. Chem-SyStM 2007, Würzburg (2007).

Ohne die vielfältige Unterstützung, die ich von verschiedenen Seiten erhalten habe, wäre das Erstellen dieser Arbeit so nicht möglich gewesen.

Zu besonderem Dank bin ich Prof. Dr. Christoph A. Sotriffer verpflichtet, der mich in den vergangenen Jahren betreut hat. Er hatte stets ein offenes Ohr und war bereit, sich in komplexe Zusammenhänge einzuarbeiten, um dann mit Rat und Tat zur Seite zu stehen. Durch seine kritischen Rückfragen und Anmerkungen, ebenso wie durch seine Anstrengungen Zusammenhänge herauszuarbeiten, wurde die Arbeit an vielen Stellen erleichtert und Fehler entlarvt.

Für die Übernahme des Zweitgutachtens dieser Arbeit sowie für seine zahlreichen Hinweise zum Fortran-Code für die Entropieberechnungen bin ich Herrn PD Dr. Reinhold Fink ebenso sehr dankbar. Frau Prof. Dr. Caroline Kisker danke ich für Ihr Interesse an dieser Arbeit und die Übernahme der Drittprüferfunktion im mündlichen Kolloquium.

Im Laufe des Projektes war es immer wieder gut, dass Dr. Ruth Brenk beratend zur Seite stand und Vorschläge wie auch Kritik zum Vorgehen lieferte. Von ihr stammen auch sehr viele Hinweise über das untersuchte Proteinsystem, die ich dankenswerterweise in die Arbeit mit einfließen lassen konnte. Das Programm für die Entropieberechnungen wurde mir freundlicherweise von Prof. Themis Lazaridis (City College, New York) überlassen.

Mein Dank gilt weiterhin der Arbeitsgruppe (Monika Nocker, Ulrich Peinz, Benjamin Schaefer, Martin Sippel, Christine Topf, Constanze Waltenberger, Armin Welker und David Zilian), den (ehemaligen) Kollegen des 3. Semesterpraktikums (Christian Büchold, Monika Herb, Jessica Klöckner und Christian Markl) sowie Verena Buback die maßgeblich zur guten Arbeitsatmosphäre beigetragen haben und mit denen ich viele nette Stunden verbringen konnte.

Das besondere Engagement von Dr. Bernd Reyer bei allen Computer-Problemen, möchte ich an dieser Stellen ebenfalls dankend erwähnen. Ich danke außerdem dem Sekretariat (Frau Dreher, Frau Ebner und Frau Möhler) wie auch der Werkstatt für Hilfe und Entlastung.

Viele Fragen rund um das Programm AMBER konnte ich nur mit Hilfe der zugehörigen Mailing-Liste lösen, weshalb ich allen Nutzern für die investierte Zeit danke.

Desweiteren danke ich dem Leibniz Rechenzentrum in München für die zur Verfügung gestellte Rechenzeit, der deutschsprachigen Sektion der Molecular Graphics and Modelling Society (MGMS) und der Gesellschaft Deutsche Chemiker (GDCH) für die finanzielle Unterstützung durch Reisestipendien sowie der Deutschen Forschungsgemeinschaft (DFG) in Form des Sonderforschungsbereichs (SFB) 630 für die finanzielle Unterstützung im Allgemeinen.

Dr. Jianxin Duan danke ich für seine Hinweise bei der Korrektur dieser Arbeit.

Ein ganz besonderer Dank gilt zum Schluss meiner lieben Frau Ulrike. Ohne Sie, ihr Verständnis, Fürsorge und den Rückhalt, hätte ich mir die Zeit in Würzburg nicht vorstellen können. Ebenso danke ich für ihre hilfreichen Anmerkungen bei der Korrektur dieser Arbeit.





# Inhaltsverzeichnis

<b>1. Einführung und Problemstellung</b>	<b>1</b>
1.1. Bedeutung von Wasser für Struktur und Reaktivität von Biomakromolekülen	2
1.1.1. Wasser als Lösungsmittel und enzymatische Funktionen . . . . .	2
1.1.2. Wasser und die Protein-Ligand Bindung . . . . .	3
1.1.3. Strukturelle Charakterisierung von Wasser in Proteinstrukturen . .	5
1.2. Aktuelle Methoden zur Behandlung von Wasser im computergestützten Wirkstoffdesign . . . . .	8
1.2.1. Klassifizierung von Wasserpositionen . . . . .	8
1.2.2. Wasser im Docking Prozess . . . . .	9
1.2.3. Wasser in der Pharmakophorsuche . . . . .	10
1.2.4. Wasser in MD-Simulationen . . . . .	11
1.3. Cytochrom <i>c</i> Peroxidase . . . . .	12
1.3.1. Auswahl der untersuchten Komplexstrukturen . . . . .	14
1.4. Fragestellungen dieser Arbeit . . . . .	19
<b>I. Methoden</b>	<b>21</b>
<b>2. Modellierung der Dynamik von Biomakromolekülen</b>	<b>23</b>
2.1. Das Kraftfeld . . . . .	24
2.1.1. Solvensmodelle . . . . .	28
2.2. Durchsuchung des Konfigurationsraums . . . . .	30
2.2.1. Molekular-Dynamik Simulationen . . . . .	30
2.2.2. Monte-Carlo Simulationen . . . . .	34
<b>3. Bestimmung thermodynamischer Größen für Protein-Ligand Komplexe</b>	<b>37</b>
3.1. Isotherme Titrationskalorimetrie . . . . .	39
3.2. Scoring-Funktionen . . . . .	41
3.3. Thermodynamische Integration und Freie-Energie Störungsrechnung . . . .	43
3.3.1. Begrifflichkeiten . . . . .	43
3.3.2. Allgemeine Vorüberlegungen . . . . .	44
3.3.3. Freie-Energie Störungsrechnung . . . . .	45
3.3.4. Thermodynamische Integration . . . . .	46
3.3.5. Verbesserung des Samplings . . . . .	47
3.3.6. Anwendung am Beispiel des Thermodynamischen Kreisprozesses . .	48

3.3.7.	Besondere Aspekte beim Durchführen der Transformationen . . . . .	50
3.4.	Thermodynamische Berechnungen unter Verwendung der Theorie der inhomogenen Lösungen . . . . .	51
3.4.1.	Korrelationsfunktionen . . . . .	52
3.4.2.	Berechnung von Entropiewerten aus Simulationsdaten . . . . .	54
3.4.3.	WaterMap . . . . .	58
<b>4.</b>	<b>Verfahren der computerbasierten Wirkstoffentwicklung</b>	<b>61</b>
4.1.	Analyse der Bindetasche . . . . .	61
4.2.	Docking . . . . .	63
<b>II.</b>	<b>Untersuchungen, Ergebnisse und Auswertungen</b>	<b>67</b>
<b>5.</b>	<b>Strukturelle Betrachtungen an der Modellbindetasche der Cytochrom c Peroxidase W191G Mutante</b>	<b>69</b>
5.1.	Strukturelle Datengrundlage für die durchgeführten Analysen . . . . .	69
5.1.1.	Vorstellung der in dieser Arbeit untersuchten Komplexstrukturen . . . . .	69
5.1.2.	Berechnung molekularer Wechselwirkungsfelder . . . . .	75
5.2.	Docking-Studien . . . . .	78
5.2.1.	Zugrunde liegende Fragestellungen . . . . .	78
5.2.2.	Verwendete Einstellungen und Parameter der eingesetzten Dockingprogramme . . . . .	78
5.2.3.	Ergebnisse des Redockings in die ursprünglichen Kristallstrukturen . . . . .	79
5.2.4.	Ergebnisse des Dockings in die ligandfreie Bindetasche . . . . .	92
5.3.	Weitergehende strukturelle Untersuchungen der Komplexe . . . . .	102
5.3.1.	Deskriptoren der Datenbank Relibase+ . . . . .	102
5.3.2.	B-Faktoren . . . . .	103
5.3.3.	Zusammenfassung . . . . .	104
<b>6.</b>	<b>Molekulardynamik-Simulationen von Strukturen der Cytochrom c Peroxidase W191G Mutante</b>	<b>109</b>
6.1.	Einstellungen und Parameter der angewandten Methode . . . . .	109
6.2.	Auswertung der erhaltenen Trajektorien . . . . .	113
6.2.1.	Analyse der strukturellen Stabilität von Protein, Bindetasche und Ligand-Bindungsmodus . . . . .	113
6.2.2.	Betrachtung der Dynamik der Bindetaschen-Wassermoleküle . . . . .	115
6.3.	Weitergehende Untersuchungen am Komplex von CCP W191G mit 3-Aminopyridin . . . . .	129
6.4.	Vergleich mit Trajektorien mit artifiziellen Zusatzkräften . . . . .	132
6.5.	Kationenfreie Modellbindetasche der CCP W191G . . . . .	140

<b>7. Bestimmung thermodynamischer Größen einzelner Wassermoleküle in Protein-Ligand Komplexstrukturen</b>	<b>147</b>
7.1. Berechnungen mittels thermodynamischer Integration . . . . .	147
7.1.1. Verwendete Parameter und Einstellungen . . . . .	148
7.1.2. Entfernung eines Wassermoleküls aus Wasser . . . . .	151
7.1.3. Testsimulationen . . . . .	151
7.1.4. Ausblick . . . . .	160
7.2. Anwendung der Solvatationstheorie inhomogener Lösungen . . . . .	161
7.2.1. Implementierung . . . . .	161
7.2.2. Berechnung entropischer Beiträge einzelner Wassermoleküle zur Protein-Ligand Bindung . . . . .	164
7.3. Zusammenfassung, Konsequenzen, Ausblick . . . . .	175
7.3.1. Thermodynamische Berechnungen . . . . .	175
7.3.2. Wassermoleküle in der untersuchten Modellbindetasche . . . . .	175
<b>III. Zusammenfassung, Summary</b>	<b>179</b>
<b>8. Zusammenfassung</b>	<b>181</b>
<b>9. Summary</b>	<b>187</b>
<b>IV. Anhang</b>	<b>191</b>
<b>A. Verwendete Hard- und Software</b>	<b>193</b>
A.1. Verwendete Hardware . . . . .	193
A.2. Verwendete Software . . . . .	193
<b>B. Tabelle mit ergänzenden Informationen</b>	<b>195</b>
<b>Tabellenverzeichnis</b>	<b>199</b>
<b>Abbildungsverzeichnis</b>	<b>201</b>
<b>Abkürzungsverzeichnis</b>	<b>203</b>
<b>Literaturverzeichnis</b>	<b>205</b>



# 1. Einführung und Problemstellung

*Überall geht ein frühes Ahnen dem späten Wissen voraus.*

ALEXANDER VON HUMBOLDT (Naturforscher und Diplomat, 1769-1859)

„Wir trinken es, wir waschen uns damit, wir fischen darin oder kochen damit. Es macht gut die Hälfte von uns aus und ohne es sterben wir in ein paar Tagen“<sup>1</sup>. Es ist offensichtlich, hier geht es um ein relativ einfaches, aber umso wichtigeres Molekül für jede Form von Leben: Wasser. Das zentrale Thema der hier vorliegenden Arbeit ist dieser weltweit am meisten untersuchte aber bemerkenswerterweise so schlecht verstandene Stoff. Martin Chaplin hat auf seiner Webseite<sup>1</sup> zahlreiche Besonderheiten von Wasser zusammen getragen, die die Wissenschaft auch heute noch gewissermaßen in Erklärungsnot bringen. In dieser Einführung, wie auch in der gesamten Arbeit, soll es jedoch nicht um Wasser im Allgemeinen, sondern speziell um seine Einflüsse auf die Wechselwirkungen zwischen Proteinen und ihren Liganden gehen. Dieser Aspekt ist deshalb von großer Bedeutung, weil er bei zahlreichen Arzneistoffen und Arzneistoffkandidaten, die sich in Funktion eines Liganden in die Bindetasche eines Proteins anlagern, relevant sein kann und zugleich schwer vorherzusagen ist. Diese Herausforderung bei der Prognose hat unterschiedliche Ursachen. Zum Einen gibt es einen Mangel an experimentellen Details, da es z.B. nicht möglich ist, bestimmte physikalische Größen für einzelne Wassermoleküle zu messen. Zum Anderen liegen dem auch ein mangelndes Verständnis der beobachteten Phänomene und demnach häufig unzureichende Modelle für deren Beschreibung zu Grunde. Dies kommt natürlich ebenfalls im Bereich der computerbasierten Methoden des Wirkstoffdesigns zum Tragen, die in der vorliegenden Arbeit besonders im Vordergrund stehen.

Im Folgenden soll eine kurze Einführung in das Thema sowie ein Literaturüberblick zu den jeweiligen Bereichen gegeben werden. Dabei wird es um die strukturelle Charakterisierung der Wassermoleküle in Proteinstrukturen, allgemeine Einflüsse von Wasser auf Proteine und verschiedene Konzepte bezüglich ihres Einflusses auf die Ligandbindung gehen. Anschließend werden aktuelle Methoden des computerbasierten Wirkstoffdesigns vorgestellt, die die Einflüsse von Wasser berücksichtigen. Zum Abschluss der Einleitung wird das Protein und die Bindetasche vorgestellt, an dem die Untersuchungen dieser Arbeit durchgeführt wurden.

## 1.1. Bedeutung von Wasser für Struktur und Reaktivität von Biomakromolekülen

### 1.1.1. Wasser als Lösungsmittel und enzymatische Funktionen

Auf der Solvens zugänglichen Oberfläche von Proteinen finden sich vorwiegend Aminosäuren mit polaren oder geladenen Seitenketten, die beim Lösen durch das Solvens Wasser elektrisch abgeschirmt werden. Die Tertiär-Struktur, die für die Funktion eines Proteins essentiell ist, hängt entscheidend von eben dieser Lösemittelumgebung und ihren Eigenschaften ab.<sup>2</sup> Leichte Modulationen, z.B. durch Zugabe von organischen Lösemitteln oder auch pH-Wert Verschiebungen, verändern die *dreidimensionale Struktur des Proteins* mehr oder weniger stark. Unpolare Aminosäuren sind vorwiegend im Inneren des Proteins vergraben und die Faltung durch Minimierung der Zahl ihrer Kontakte zu Wasser bestimmt. Dispersionswechselwirkungen von unpolaren Gruppen mit Wassermolekülen sind energetisch nicht ungünstig, jedoch schwächer als gerichtete Wechselwirkungen. Diese sind aber zwischen Wassermolekülen und hydrophoben Resten nicht möglich. Gleichzeitig ist es energetisch nicht vorteilhaft, wenn beim Lösen des Stoffes Wasser-Wasser Wechselwirkungen aufgebrochen werden. Eben aus diesen Gründen wird die Zahl der hydrophoben Gruppen auf der Proteinoberfläche minimiert. Die Proteinfaltung in Lösung ergibt sich somit gewissermaßen „von selbst“. Allgemein bezeichnet man das als *hydrophoben Effekt*.

Das gängige molekulare Modell von der Lösung eines Proteins erklärt die günstigen Lösungsenthalpien sowie die Zunahme der Wärmekapazität durch stärkere Wasserstoffbrücken an der Oberfläche des gelösten Moleküls im Vergleich zur reinen Wasserphase und der damit verbundenen Ordnung des Wassers<sup>3-5</sup>. Diese Vorstellung wird in<sup>6</sup> einem anderen Konzept gegenübergestellt. Demnach beruht die für die Lösung auch von unpolaren Teilchen verantwortliche, kurzreichweitige und repulsive Kraft auf der Beschaffenheit der Oberfläche. Die beobachteten Effekte seien somit ein Oberflächenphänomen und Wasser nimmt keine Sonderrolle im Vergleich zu anderen Lösemitteln ein. Aktuelle theoretische Studien beschreiben den Lösungsvorgang als eine Art „Austrocknen“ an der Proteinoberfläche, bei denen sich die Wassermoleküle vom gelösten Stoff wegbewegen und ordnen<sup>7</sup>.

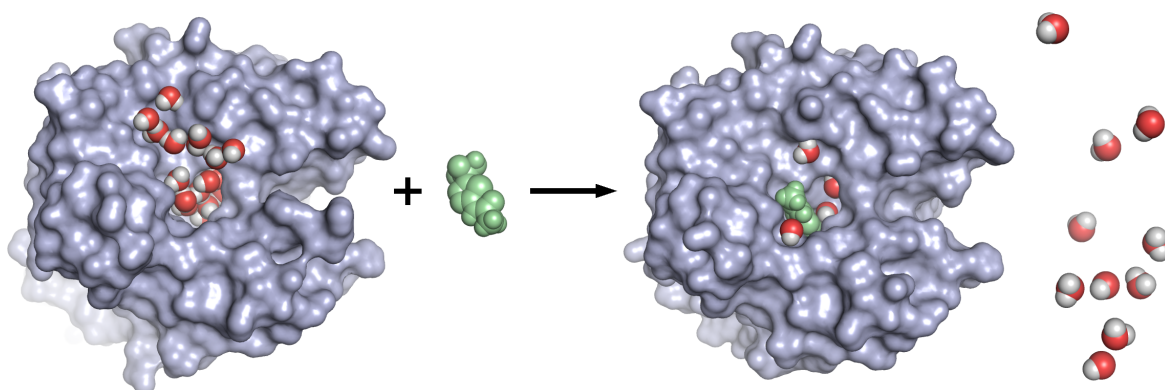
Eine ebenso wichtige Rolle kommt den Wassermolekülen zu, die in zahlreichen Enzymen an der *katalysierten Reaktion* beteiligt sind. Dies ist beispielsweise bei *Proteasen*, die nahezu in allen Zellen vorkommen, der Fall: Sowohl Cysteinproteasen wie Papain, Serinproteasen wie Trypsin oder Thrombin oder Aspartylproteasen wie die HIV-Protease benötigen ein Wassermolekül in der durch sie katalysierten Reaktion, um eine Peptidbindung zu spalten. Analog zur Spaltung eines Peptides katalysieren *Nucleasen* unter Verwendung eines Wassermoleküls den Abbau der Nukleinsäuren, *Glykosidasen* die Hydrolyse einer glykosidischen Bindung, *Esterasen* katalysieren die Verseifung von Estern wie sie z.B. in Fetten vorkommen während *Phosphatasen* die Reaktion der Kinasen umkehren. Überall sind dabei Wassermoleküle entscheidend beteiligt. Daraus ergibt sich ein mögliches therapeutisches Potential von Inhibitoren, die in der Lage sind, essentielle Wassermoleküle abzuschirmen oder zu ersetzen und so die Funktion des Enzyms zu unterbinden.

### 1.1.2. Wasser und die Protein-Ligand Bindung

Der bereits erwähnte hydrophobe Effekt spielt auch bei der Protein-Ligand Bindung eine entscheidende Rolle. Wassermoleküle in einer solvatisierten Bindetasche weisen eine höhere Ordnung auf als Wassermoleküle, die das Protein umgeben. Das Aufbrechen dieser Ordnung bei der Ligandbindung und der damit zusammenhängenden Verdrängung dieser Wassermoleküle liefert einen entropisch günstigen Beitrag zur Protein-Ligand Bindung.<sup>8</sup> Schematisch ist dieser Vorgang stark vereinfacht in Abb. 1.1 dargestellt. Hierbei wird aus Gründen der Übersichtlichkeit die Solvathülle, die sowohl das Protein als auch den unkomplexierten Liganden umgibt, nicht gezeigt.

Die Stärke der Ligandbindung wird neben der direkten Wechselwirkung zwischen Ligand und Aminosäuren durch weitere Beiträge bestimmt. Außer der Desolvatation der Tasche ist es nötig, die Solvenshülle des Liganden zu entfernen. Außerdem muss sehr wahrscheinlich Arbeit aufgewendet werden, um den Liganden aus seiner Minimumskonformation in Lösung in die Bindungskonformation zu bringen. Der Bindungsvorgang kann weiterhin mit konformationellen Änderungen des Proteins bzw. der Bindetasche einhergehen. Aber auch entropische Beiträge spielen eine Rolle. Sowohl die Beweglichkeit des Liganden als auch der Protein-Seitenketten wird eingeschränkt. Da aus zwei Teilchen eines wird, führt der Bindungsprozess zum Verlust von drei Translations- und drei Rotations-Freiheitsgraden. Energetisch entspricht das ungefähr einem Wert von  $6 \cdot \frac{1}{2}RT \approx 7.5 \text{ kJ/mol}$ . Kommt es bei der Ligandbindung zu konformationellen Änderungen des Rezeptors, im Speziellen der Bindetasche, so entspricht das nicht dem klassischen Schlüssel-Schloss-Prinzip, wonach ein Ligand wie ein Schlüssel ins Schloss zur Bindetasche passt. Man spricht in diesem Zusammenhang häufig vom so genannten *induced fit* Phänomen, wonach ein Ligand die Konformationsänderung des Rezeptors induziert.<sup>9</sup> Andererseits existiert auch die Vorstellung, dass verschiedene Konformationen einer Bindetasche bereits existieren und der Ligand eine davon durch die Bindung zum Rezeptor stabilisiert. Letzteres konnte am Beispiel der Aldose-Reduktase beobachtet werden<sup>10</sup>. Die Komplexität einer Bindungsvorhersage wird also durch die Kombination und Kompensation von enthalpischen und entropischen Effekten verursacht.

Durch die Ligandbindung werden nicht notwendigerweise alle Wassermoleküle einer Bindestelle verdrängt. In vielen Fällen bleiben Wassermoleküle bei der Ligandbindung in der Bindetasche des Rezeptors konserviert und sind dann verbrückend an der Wechselwirkung zwischen Protein und Ligand beteiligt. Ein Beispiel hierfür ist die oben erwähnte HIV-1 Protease. Die erste Generation ihrer Inhibitoren wurde als Peptidmimetika entworfen, die unter anderem durch Verbrückung über ein Wassermolekül an den Rezeptor gebunden sind. Dazu gehört z.B. eine hochaffine Verbindung, die letztlich alle klinischen Tests bestand und unter dem Namen Saquinavir im Handel ist. Wie erwähnt, kann aber die Verdrängung von Wassermolekülen aus der Bindetasche durch den Liganden einen entropisch günstigen Beitrag liefern. Eine neuere Klasse von Inhibitoren der HIV-Protease basiert deshalb auf dem Grundgerüst eines cyclischen Harnstoffes. Der Sauerstoff der Carbonyl-Gruppe übernimmt hierbei die Funktion als Wasserstoffbrückendonator, die zuvor das so genannte *Flap*-Wasser inne hatte<sup>11,12</sup>.



**Abb. 1.1.:** Die trichterförmige Bindetasche von Carboanhydrase II, einem Enzym, das die Hydratisierung von Kohlenstoffdioxid zu Kohlensäure (wie auch die Rückreaktion) katalysiert und deshalb z.B. an der Regulierung des Säurehaushaltes der Niere beteiligt ist, ist in der unkomplexierten Form mit verschiedenen Wassermolekülen gefüllt. Bei der Bindung des Arzneistoffes Acetazolamid werden einige dieser Wassermoleküle verdrängt und sind in der Komplexstruktur nicht mehr in der Bindetasche vorzufinden. Die hier gezeigte schematische Darstellung beruht auf den Kristallstrukturen mit den PDB-Indizes 2CBC und 1YDA.

Eine systematische Untersuchung über Wassermoleküle in Protein-Ligand Bindungen wurde von Lu *et al.*<sup>13</sup> 2007 angefertigt. Es zeigte sich, dass in einem ausgewählten Satz von hochaufgelösten Röntgenkristallstrukturen knapp 55 % der gefundenen Wassermoleküle, die an einen Liganden gebunden sind, auch mit dem Protein verknüpft sind und damit eine verbrückende Bindestelle darstellen. Die Studie zeigte ebenfalls, dass Wassermoleküle mit weniger als zwei polaren Bindungen zum Komplex eine höhere Mobilität zeigen als die Proteinatome. Im Schnitt werden außerdem mehr polare Wechselwirkungen mit dem Protein als mit dem Liganden eingegangen. In der Dissertation von J. Günther<sup>14</sup> wird ebenfalls auf diese Thematik eingegangen. Hierbei wird die Verschiebung oder Verdrängung von Wasserpositionen aus dem *statistischen Vergleich* von Komplexen und unkomplexierten Proteinen betrachtet. Gemäß der Auswertung dieses Vergleiches ist die Konservierung von Wasserpositionen im Bindetaschenbereich um mehr als 10 Prozentpunkte niedriger als auf der gesamten Proteinoberfläche. Der Vergleich zeigt außerdem, dass vor allem vergrabene Wassermoleküle eher an der Protein-Ligand Bindung beteiligt sind und somit im Design berücksichtigt werden müssen, weil sie den Bindemodus des Liganden mitbestimmen. Besonders für diese vergrabenen Wassermoleküle ist eine Konservierung um ein Vielfaches wahrscheinlicher als eine Verschiebung der Position. Etwa die Hälfte der vergrabenen, an das Protein gebundenen Wassermoleküle einer Tasche bleiben (bezogen auf den Vergleich unkomplexiertes Protein/Komplex) bei der Ligandbindung konserviert. Diese Zahl bezieht sich selbstverständlich nur auf die Zahl der kristallographisch aufgelösten Wasserpositionen. Vergrabene Wassermoleküle innerhalb eines Radius von 5 Å um einen Liganden sind im Vergleich mehrerer Komplexstrukturen sogar zu fast 80 % konserviert. Diese Ergebnisse werden aber maßgeblich auch durch die hohe Ähnlichkeit der Liganden eines Proteins



bestimmt.

Wassermoleküle, die als Brücke in einer Protein-Ligand Bindung beteiligt sind, können durch ihre Beweglichkeit sehr leicht die *geometrischen Anforderungen der Wasserstoffbrückenbindungen* erfüllen. Umgekehrt ist das Protein nicht notwendigerweise in der Lage eine Konformationsänderung zur Optimierung einer Wasserstoffbrückenbindung einzugehen. Da eine Abweichung von einem Bruchteil eines Å bereits die Stärke einer H-Brücke signifikant verringert, sollten beim Wirkstoffdesign nur dann Wasserstoffbrückenbindungs-donoren und -akzeptoren eingeplant werden, wenn diese Wechselwirkungen in einer optimalen Geometrie möglich sind. Ansonsten könnten apolare Gruppen an diesen Stellen mit dem hydrophoben Effekt zur Bindungsaffinität beitragen. Die Verbrückung durch ein Wassermolekül hingegen liefert den bereits erwähnten Vorteil konformationeller Flexibilität und ist von der Stärke her mit einer direkten Wechselwirkung vergleichbar<sup>15</sup>.

Im Falle einer hydrophoben Bindetasche kann die entsprechende Situation bei der Ligandbindung durchaus anders aussehen. Die hauptsächlich treibende Kraft für die Assoziation zweier hydrophober Teilchen und damit auch die für Bindung in eine hydrophobe Tasche ist die Erhöhung der Entropie der ursprünglich gebundenen und besonders geordneten Wassermoleküle, der hydrophobe Effekt. In dem im Jahr 2007 erschienenen Artikel von S. W. Homans<sup>16</sup> werden Beispiele diskutiert, dass hydrophobe Taschen nicht notwendigerweise irgend ein Wassermolekül enthalten. Da bei der Ligandbindung in diesen Fällen keine Desolvatisierung der Tasche notwendig ist, lässt sich die Affinität der Liganden dann vor allem dadurch vergrößern, dass Ligand und Tasche möglichst eine optimal aufeinander abgestimmte *Form* haben und die Dispersionswechselwirkungen maximiert werden. Dieses Konzept widerspricht dem bislang gängigen Bild der Protein-Ligand Bindung grundlegend. Von Matthews and Liu erschien vor kurzem ein Übersichtsartikel<sup>17</sup>, der sich mit der Frage auseinandersetzt, ob völlig apolare Bindetaschen in Proteinen wirklich *leer* sein können. Die Autoren kommen zu dem Ergebnis, dass es eher nicht der Fall ist, dass vorhandene Wassermoleküle in den Bindetaschen einer entsprechenden Röntgenstruktur nicht beobachtet werden. Das Ergebnis lässt sich ungefähr wie folgt zusammenfassen: Eine rein hydrophobe Bindetasche bleibt dann leer, wenn sie so klein ist, dass sie nur ein einzelnes Wassermolekül fassen kann. Ein Cluster aus drei bis vier Wassermolekülen ist erst ab einer Größe von ca. 500 Å<sup>3</sup> möglich. Experimentell sind diese Fälle im Prinzip ohne weiteres durch Röntgenstrukturanalyse zu bestimmen.

### 1.1.3. Strukturelle Charakterisierung von Wasser in Proteinstrukturen

Die Basis für zahlreiche Methoden des computergestützten Wirkstoffdesigns bilden strukturelle Informationen von Proteinen und Protein-Ligand Komplexen. Diese werden für gewöhnlich mit Hilfe der Röntgenstrukturanalyse (selten durch Neutronenbeugung), durch NMR-Methoden oder mit Hilfe von Homologie-Modellen gewonnen. Wassermolekülen muss hierbei besondere Aufmerksamkeit zukommen. Die PDB-Datenbank<sup>18</sup> enthält über 60 000 frei zugängliche Proteinstrukturen. Eine *Röntgenstruktur* wird mit Hilfe eines dreidimensionalen Strukturmodells in mehreren Schritten durch Optimierung an Hand der im Beugungsexperiment gemessenen Elektronendichte (ED) gewonnen, die für publizierte Struk-

turen über den Uppsala ED Server<sup>19</sup> öffentlich zugänglich gemacht wird. Der wichtigste Parameter, der die Qualität einer Kristallstruktur beschreibt, ist die sogenannte Auflösung. B-Faktoren (oder auch Temperaturfaktoren oder Debye-Waller-Faktoren) hingegen sind eine Funktion der Auslenkung eines Atoms von einer gemessenen Mittelwertsposition. Zwar können diese nicht direkt zwischen verschiedenen Strukturen verglichen werden (außer nach vorangegangener Normalisierung), dienen aber dennoch als Maß für die unterschiedliche Beweglichkeit z.B. von Wassermolekülen. Weitere Parameter wie der  $R$ -Wert

$$R = \frac{\sum |F_{\text{obs}}| - |F_{\text{calc}}|}{\sum |F_{\text{obs}}|} \quad (1.1)$$

(die gewichtete Differenz aus gemessenen Struktur Faktoren und denen, die aus dem Modell berechnet wurden),  $R_{\text{free}}$ -Wert oder der RSR-Wert („real space  $R$ -factor“) erlauben einen Vergleich zwischen der aufgrund des Modells berechneten und der gemessenen ED. Die beiden ersten der genannten Werte ermöglichen dies nur global für das ganze Protein, während der RSR-Wert für einzelne Reste bestimmt werden kann. Eine Übersicht über einzelne Schritte der Strukturbestimmung findet sich beispielsweise bei Klebe<sup>9</sup>.

In einem Übersichtsartikel zu den *Limitationen kristallographischer Daten* im Wirkstoff-Design von Kleywegt und Mitarbeitern<sup>20</sup> wird auch auf das Thema Wassermoleküle eingegangen. Bedingt durch den Auflösungsbereich der Röntgenstrukturallographie sind Wassermoleküle, bis auf wenige Ausnahmen, nicht vollständig aufgelöst. Nur das Sauerstoffatom kann mit der Elektronendichte lokalisiert werden; Wasserstoffatome beugen die Röntgenstrahlung nicht stark genug. Da Wasser isoelektronisch zu Natrium- oder Ammoniumionen ist, sind diese drei Moleküle/Ionen allein durch die Elektronendichte nicht unbedingt voneinander zu unterscheiden. Das Vorhandensein oder Fehlen eines Wassermoleküls in einer Struktur kann nicht mit Sicherheit bestimmt werden. Der Vergleich gleicher Strukturen, die von unterschiedlichen Kristallographen gelöst wurden, zeigt, dass die Positionierung der Wassermoleküle zum Teil recht subjektiv ist<sup>21,22</sup>. Das Hinzufügen von Wasser zu einer Struktur kann auch dazu dienen, Probleme dieser Struktur zu tarnen, indem die Unterschiede zwischen gemessenen und berechneten Struktur Faktoren künstlich verringert werden. Wassermoleküle, die dann keine Wasserstoffbrückenbindung mit irgendeinem anderen Atom bilden, sind sehr wahrscheinlich Artefakte. Bei schlecht aufgelösten Kristallstrukturen kann die Elektronendichte von Wasser möglicherweise gar nicht vom Rauschen unterschieden werden.

Im Gegensatz zu Elektronenstrahlen werden *Neutronenstrahlen* von Atomkernen gebeugt. Mit Hilfe der Neutronenkristallographie lassen sich so auch die Positionen von Wasserstoffatomen detektieren. Aufgrund der hierfür benötigten sehr großen Kristalle und langen Messzeiten gibt es keinen so großen Fundus an Strukturdaten wie durch Röntgenstrukturallographie. Da hier aber H-Brückenbindungsgeometrien und Wasserpositionen genau analysiert werden können, sind diese wenigen vorhandenen Datensätze von großem Wert.

Die *NMR-Strukturanalyse* hingegen bietet nun die Möglichkeit neben der Strukturinformation zusätzlich auch Dynamik-Informationen aus einem Experiment zu erhalten. Man

benötigt dafür jedoch größere Mengen an Protein und längere Messzeiten. Das führt aber speziell bei Interesse an Wassermolekülen zu einem entscheidenden Nachteil. Man erhält durch ein NMR-Experiment eher gemittelte Informationen über Solvens- als über einzelne strukturelle Wassermoleküle, da die Austauschzeiten gemessen an der Zeitskala des NMR-Experiments zu kurz sind.

Levitt und Park stellen fest<sup>23</sup>, dass im Schnitt etwa ein halbes bis zwei Wassermoleküle pro Aminosäure zu finden sind. Das summiert sich bei einem Protein typischer Größe dann auf ungefähr 200 Wassermoleküle. Der Übersichtsartikel befasst sich vor allem mit den Bedingungen, unter denen ein Wassermolekül im Rahmen der Röntgenstrukturanalyse überhaupt detektiert werden kann. Die Sichtbarkeit von Atomen durch Röntgenbeugungsexperimente hängt sehr stark von ihrer *Beweglichkeit* ab. In der Kristallstrukturanalyse werden ja gemittelte Daten über einen langen Zeitraum gesammelt. Die Beweglichkeit eines Wassermoleküls wird aber nicht nur dadurch bestimmt, ob eine Atomposition energetisch günstig ist, sondern auch durch andere Faktoren, die im Folgenden näher erläutert werden sollen.

Man führe sich folgende Beispiele vor Augen:

- Ein Wassermolekül ist sehr fest an eine Seitenkette gebunden, die wiederum aber sehr flexibel ist. Das Wassermolekül ist dann folglich ebenfalls schwierig oder nicht zu detektieren.
- Generell finden sich Wassermoleküle häufiger in Vertiefungen des Proteins als anderswo. Nur in einer sehr engen Vertiefung ist die Beweglichkeit des Solvensmoleküls sehr eingeschränkt. Dort ist es dann entsprechend wahrscheinlicher ein Wassermolekül zu beobachten verglichen mit einer weiten Vertiefung.
- Ein weiteres Indiz für das Auffinden von Wassermolekülen sind polare Wechselwirkungspartner. Unpolare Atome interagieren weniger spezifisch und geben mehr Spielraum für Bewegung. Ein Wassermolekül, das beispielsweise an einer rigiden Seitenkette wenig spezifisch gebunden ist, bewegt sich möglicherweise zwischen verschiedenen energetisch ähnlichen Positionen hin- und her.

Schließlich können auch *Austauschvorgänge* die Detektion von Wassermolekülen erschweren. Durch ein Röntgenstrukturerperiment können keine Aussagen über den Wasseraustausch des Proteins in Lösung gemacht werden, obwohl dieser selbstverständlich zumindest bei allen Oberflächenwassermolekülen (aber auch anderen) stattfindet. Alle Wassermoleküle haben identische physikalische und chemische Eigenschaften und sind ununterscheidbar. Die Wahrscheinlichkeit für einen Austausch ist durch den Übergangszustand, in dem kein Wasser an der betrachteten Position sitzt, charakterisiert. Die Höhe dieser Barriere hängt von der Bindungsstärke des Wassers ab, ist aber auch darin begründet, wie leicht diese Position zugänglich ist. Vergrabene Wasserpositionen zeigen demnach einen viel geringeren Wasseraustausch. Dieser wie auch die Energetik vergrabener Wassermoleküle wurde z.B. von Halle und seinen Mitarbeitern untersucht<sup>24,25</sup>.

Wassermoleküle können entscheidend an der korrekten Ausbildung der *Sekundär- und Tertiärstruktur* eines Proteins beteiligt sein, indem sie über Wasserstoffbrückenbindungen

ausgebildete Wechselwirkungen vermitteln. Für gewöhnlich bezeichnet man solche Wassermoleküle als *strukturelles Wasser*. Dies wurde beispielsweise von Yutani *et al.*<sup>26</sup> untersucht. Fischer und Verma hingegen<sup>27</sup> untersuchten die Einflüsse struktureller Wassermoleküle auf die Proteinflexibilität.

### 1.2. Aktuelle Methoden zur Behandlung von Wasser im computergestützten Wirkstoffdesign

Im Bereich des strukturbasierten und computergestützten Wirkstoffdesigns wird mit Hilfe des Wissens über die Struktur einer Bindetasche oder bekannter Inhibitoren eines Enzyms versucht, neue, potentere Liganden zu finden, die nach Abschluss eines Optimierungsprozesses als Wirkstoff Verwendung finden könnten. Zwei essentielle Schritte, die hierbei von Computerprogrammen übernommen werden, sind die Vorhersage der *Bindungsmodi* neuer Liganden und die Abschätzung ihrer *Affinität*. Gerade hierfür ist es, wie aus den obigen Ausführungen deutlich wird, wichtig, evtl. vorhandene Wassermoleküle im Bereich der Bindetasche in diese Überlegungen mit einzubeziehen. Dabei müssen einerseits individuelle Wassermoleküle, die an Protein und/oder Ligand gebunden sind, wie auch die des Solvens berücksichtigt werden. Im Folgenden sollen Methoden aus verschiedenen Bereichen vorgestellt werden, die versuchen, diese Einflüsse zu erfassen. Eine gute Übersicht hierzu findet sich ebenfalls bei Liedl *et al.*<sup>28</sup>. Allgemein kann man sagen, dass alle computerbasierten Ansätze die Tendenz haben, sich entweder auf ein individuelles Bindetaschenwassermolekül oder auf Solvatations- und Desolvatationseffekte zu fokussieren. Im so genannten virtuellen Screening geht der Trend dahingehend, entweder Wasser vollständig zu vernachlässigen oder nur wenige ausgewählte Wassermoleküle zu betrachten.

Die strukturellen Informationen über Wassermoleküle erhält man für gewöhnlich auf den gleichen Wegen wie die Informationen über Protein und Ligand. Liegen diese jedoch nicht vor, weil die Struktur des Proteins einem Homologie-Modell entspringt oder z.B. in der Röntgenstrukturanalyse wegen höherer Flexibilität nicht aufgelöst wurde, so können günstige Wasserpositionen in einer Bindetasche auf unterschiedliche Weise berechnet werden. Programmbeispiele hierfür sind AQUARIUS<sup>29</sup>, GRID<sup>30</sup>, MCSS<sup>31</sup>, SuperStar<sup>32</sup>, CS-MAP<sup>33</sup>, Drugscore<sup>34,35</sup> oder SiteMap<sup>36,37</sup>. Zum Teil wird hierauf in Kapitel 4 näher eingegangen.

#### 1.2.1. Klassifizierung von Wasserpositionen

Um die *energetische Bewertung* des verbrückenden Einflusses eines Wassermoleküls in der Bindetasche vorzunehmen, sind viele Scoring-Funktionen und Kraftfelder gebräuchlich. Ein spezielles Beispiel hierfür ist das HINT Kraftfeld<sup>38-40</sup>, das sich von den vielen auf Newtonscher Mechanik basierten Kraftfeldern konzeptionell unterscheidet. HINT steht für *hydrophobic interactions* und verwendet das bekannte Hydrophobizitäts-Maß  $\log P_{\text{Octanol/Wasser}}$  für ein Wechselwirkungsmodell. Das Kraftfeld wurde an experimentellen Daten für den Verteilungskoeffizienten Octanol/Wasser kalibriert. Für ein neues Molekül kann nun durch

das HINT-Kraftfeld ein Wert für  $\log P$  abgeschätzt werden. Zusätzlich können aber auch Wechselwirkungskarten für eine Bindetasche berechnet werden. Da die Hydrophilie eines Moleküls nativ berücksichtigt wird, eignet sich diese Funktion, um unter anderem aufgrund des HINT Scores über die Verdrängbarkeit von Wassermolekülen aus der Bindetasche zu urteilen<sup>41,42</sup>. Das empirische Fold-X Kraftfeld<sup>43</sup> wurde entwickelt, um die Positionen von Wassermolekülen in Proteinen zu bestimmen und wurde an hochaufgelösten Protein-Kristallstrukturen parametrisiert.

Ein weiteres Werkzeug entwickelt von Wissenschaftlern der Universität in Michigan (USA), Consolv<sup>44</sup>, nutzt eine „k-nearest-neighbor“ Klassifizierung an Hand von vier *Deskriptoren* (Atomdichte, Hydrophobizität, H-Brückenbindungen, B-Faktoren), um wasser-verbundene Protein-Ligand Bindungen vorherzusagen. WaterScore<sup>45</sup> nutzt ähnliche Parameter (B-Faktor, Größe der solvens-zugänglichen Oberfläche, H-Brückenbindungsenergie und die Anzahl der Protein-Wasser-Kontakte), um zwischen bei der Ligandbindung verdrängbaren und konservierten Wassermolekülen zu unterscheiden.

### 1.2.2. Wasser im Docking Prozess

Mit den heute verfügbaren Dockingprogrammen (wie Glide, Autodock, DOCK) können Wassermoleküle in einer Bindetasche als Wechselwirkungspartner mit berücksichtigt werden. Voraussetzung ist jedoch, dass die Positionen dieser Wassermoleküle bekannt sind. Sind Wasserpositionen aus einer apo-Struktur verfügbar, ist jedoch ungeklärt, ob diese Moleküle im Falle des gebundenen Liganden noch in der Bindetasche vorhanden sind. Nur bei wenigen Dockingprogrammen ist es möglich, diese Fragen automatisiert während des Docking Vorganges abzuwägen. Das explizite Berücksichtigen vergrößert den Suchraum des Optimierungsproblems „Docking“ erheblich. Dabei herrscht in Fachkreisen keine einhellige Meinung darüber, ob die Betrachtung von Wassermolekülen beim Docking die Ergebnisse verbessert oder nicht<sup>46</sup>.

Im Jahr 1999 wurde von M. Rarey und Kollegen<sup>47</sup> das so genannte *Particle Concept* vorgestellt. Der in FlexX<sup>48</sup> implementierte Algorithmus hat das Ziel, Wassermoleküle auf günstigen Positionen innerhalb des aktiven Zentrums zu finden. In FlexX wird der Ligand während des Docking-Vorgangs aus Inkrementen zusammengesetzt. Das Verfahren läuft demnach wie folgt ab: Zunächst werden für Wasser günstige Positionen in der Bindetasche berechnet und gespeichert. Anschließend wird das Basisfragment des Liganden in die Tasche positioniert. Während des inkrementellen Zusammensetzens des Liganden werden dann Wassermoleküle auf den zuvor berechneten Positionen platziert, sofern sie zusätzliche Wasserstoffbrückenbindungen mit dem Liganden ausbilden. Auf diese Weise wird der sterische Anspruch der Wassermoleküle, wie auch die Geometrie der Wechselwirkungen beim Aufbau der Liganden berücksichtigt. Seit Version 3.1.1 des Programms FlexX ist es auch möglich zu überprüfen, ob bekannte Wassermoleküle einer Bindetasche zusammen mit dem Liganden binden. Das Programm entscheidet dabei während des Dockings selbständig, ob Kristallwassermoleküle an der Ligandbindung beteiligt sind oder nicht.

Auch mit dem Programm GOLD<sup>49-51</sup> ist es möglich, Vorhersagen bezüglich der verbrückenden Wirkung von Wassermolekülen zu treffen. Aufgrund des genetischen Algorith-

musses für das Docking unterscheidet sich das Vorgehen von dem in FlexX. M. L. Verdonk und Mitarbeiter<sup>52</sup> implementierten hierfür einen *zusätzlichen Term* in die Goldscore- und die Chemscore- *Scoringfunktion*. Dieser entspricht der Verringerung der Entropie für jedes Wassermolekül, dessen Wechselwirkungen in der Tasche angeschaltet werden. Gleichzeitig werden die Wechselwirkungen des Wassermoleküls mit der Bindetaschen-Umgebung über die Scoringfunktion berechnet. Auf diese Weise ist das Programm in der Lage zwischen dem Verdrängen und Konservieren eines Wassermoleküls mit Hilfe des Scores zu unterscheiden.

An der Universität von Konstanz wurde ein Dockingprogramm mit Namen PLANTS<sup>53-56</sup> entwickelt, das die *ant colony optimization* verwendet, um die Ligand Konformation und Position in der Bindetasche zu ermitteln. Die Verdrängung von Wassermolekülen wird in diesem Algorithmus folgendermaßen gehandhabt: Um ein Wassermolekül im Bereich der Bindetasche wird ein Bereich definiert, in dem sich dieses frei bewegen darf. Muss ein Wassermolekül während des Dockings aufgrund einer Ligandkonformation den entsprechenden Bereich verlassen, um einem Zusammenstoß der Atome zu entgehen, so gilt es als durch den Liganden verdrängt und wird für die Wechselwirkungen dieser Docking Lösung nicht weiter berücksichtigt.<sup>57</sup>

Ein ähnlicher Algorithmus ist im Programm SLIDE<sup>58</sup> implementiert. Wenn beim Docking ein Ligand mit einem Wassermolekül kollidiert, so ist hier eine Translationsbewegung dieses Wassermoleküls erlaubt. Nur wenn eine Kollision dann immer noch nicht vermieden werden kann, wird das Wassermolekül aus der Bindetasche entfernt.

Mit einem genetischen Algorithmus versucht das Programm FITTED<sup>59</sup>, sowohl Flexibilität von Ligand als auch Rezeptor beim Docking zu berücksichtigen und gleichzeitig Vorhandensein oder Verdrängen von Wassermolekülen wie deren Orientierung vorherzusagen.

Das Programm GLIDE wiederum plaziert für alle energetisch gleichwertigen Posen eines Liganden Wassermoleküle in die Bindetasche.<sup>60</sup> Der Score wird danach neu berechnet und auf diese Weise statt eines Kontinuum basierten Solvensmodells ein explizites verwendet. Dadurch soll die Anzahl falsch positiver Lösungen im Ergebnis verringert werden.

### 1.2.3. Wasser in der Pharmakophorsuche

Die Möglichkeiten, wasser-verbrückte Wechselwirkungen oder die Verdrängung von Wassermolekülen vorherzusagen, sind bei einer Pharmakophorsuche naturgemäß sehr eingeschränkt. Es ist aber ohne weiteres möglich, bei der Entwicklung eines Pharmakophormodells strukturelle Informationen über wasser-verbrückte Wechselwirkungen mit einfließen zu lassen. Wassermoleküle werden dabei als Teil der Bindetasche betrachtet und entsprechend deren geometrischer Anordnung H-Brücken-Donor und -Akzeptor Eigenschaften im Pharmakophormodell platziert. Das wurde beispielsweise in einer Arbeit aus der Gruppe von G. Klebe aus dem Jahr 2003 gemacht<sup>61</sup>. Darüber hinausgehend versuchen Lloyd *et al.*<sup>62</sup> mit einem Hybridmodell Aussagen über die Verdrängung von Wassermolekülen zu machen. Ein ligandbasiertes Pharmakophormodell wird aus bekannten Inhibitoren erstellt und dann auf die Proteinstruktur überlagert. Gebundene Wassermoleküle werden über WaterScore identifiziert und so kann für andere Liganden eine Aussage über die Verdrängung

dieser Wassermoleküle aus sterischen Gründen getroffen werden.

### 1.2.4. Wasser in MD-Simulationen

Aufgrund der heute verfügbaren Rechenleistung ist es ohne weiteres möglich, eine MD-Simulation eines Proteins nicht nur in einem Kontinuum-Solvensmodell, sondern auch in einer explizit definierten Wasserumgebung durchzuführen. Eine sehr große Herausforderung im Kontext von MD-Simulationen ist aber die geeignete Auswertung und Erfassung der durch die Simulation erhaltenen Datenmenge. Dies kommt selbstverständlich auch bei der Analyse von Wassermolekülen im Zusammenhang mit der Protein-Ligand Bindung zum Tragen. Es gibt verschiedene Messgrößen, die hier üblicherweise Verwendung finden und relevante Anhaltspunkte zum Verständnis liefern.

Eine Möglichkeit, um die örtliche Konservierung eines Wassermoleküles zu beschreiben, ist die Berechnung einer radialen Verteilungsfunktion (RDF). Diese verdeutlicht den Aufenthaltsbereich eines Wassermoleküls und ebenso das Auftreten von weiteren in den verschiedenen Solvatationssphären um die Referenzposition. Aus den Daten der generierten Trajektorie kann ebenfalls ein B-Faktor für ein Wassermolekül berechnet werden, um so fester gebundene Wassermoleküle erkennen zu können. Gleichfalls kann basierend auf der Trajektorie eine Aufenthaltszeit ermittelt werden. Das ist die Zeit, in der ein bestimmtes Wassermolekül in einem vorher festgelegten Bereich anzutreffen ist. Hierfür wird jedoch ein Referenzpunkt seitens des Proteins benötigt. Oftmals ist das aber nicht unmittelbar möglich, da Proteinbewegungen mit berücksichtigt werden müssen. In der Literatur sind verschiedene Verfahren zu finden, die dieses Problem umgehen. Eine Koordinaten-Mittelung in Bezug auf jede einzelne Konfiguration der Trajektorie ist möglich, führt aber zu einem sehr hohen Rauschen der Werte<sup>63</sup>. Die Gruppe von McCammon beschreibt eine nur über eine bestimmte Zeit gemittelte Position (TAP, „time averaged position“) und kombiniert dies mit Koordinaten für ein Wassermolekül, die nur von der lokalen Proteinumgebung abhängen (ARC, „averaged residue coordinate“)<sup>64-66</sup>. Für eine zuvor festgelegt Raumregion kann außerdem auch ganz allgemein ein Besetzungsgrad ermittelt werden. Zwischen diesem und einer Aufenthaltszeit besteht dann aber nicht notwendigerweise eine direkte Korrelation, da es während der Simulation zu Austauschereignissen kommen kann. Als ein weiteres Maß bleibt noch die so genannte „inter-site jump time“<sup>64</sup> (also die Zeit, die vergeht, bis ein Wassermolekül von einer Bindestelle zu anderen wandert) zu nennen.

Thermodynamische Beiträge von Wassermolekülen zu Bindungsaffinitäten lassen sich üblicherweise mit Hilfe von freien Energie Rechnungen wie der „free energy perturbation“ (FEP) oder „thermodynamic integration“ (TI) Methode bestimmen. Diese Verfahren erfordern lange Rechenzeiten und liefern Genauigkeiten im Bereich des Kraftfeldfehlers. Eine ausführliche Darstellung findet sich in Kap. 3. Im Rahmen der Verfahren „Linear Interaction Energy“ (LIE)<sup>67,68</sup> und des MM-PBSA Ansatzes<sup>69-72</sup>, wie auch des MM-GBSA<sup>73,74</sup> Ansatzes zur Abschätzung der Ligandbindungsaffinität finden hingegen Kontinuums-Solvensmodelle ihren Einsatz. Ein relativ neuer Ansatz mit dem Namen „Watermap“<sup>75-78</sup> beruht auf Arbeiten von Lazaridis<sup>15,79-83</sup> im Zuge der inhomogenen Solvatationstheorie (vgl. Kap. 3.4). So können neben den enthalpischen Beiträgen auch entropische

einer wasser-verbrückten Protein-Ligand Bindung aus Simulationsdaten erhalten werden. Diese Methode wurde bereits sehr erfolgreich für verschiedene Systeme angewandt. Beispiele hierfür kommen von den Firmen Schrödinger, Pfizer und Novartis<sup>84-88</sup>.

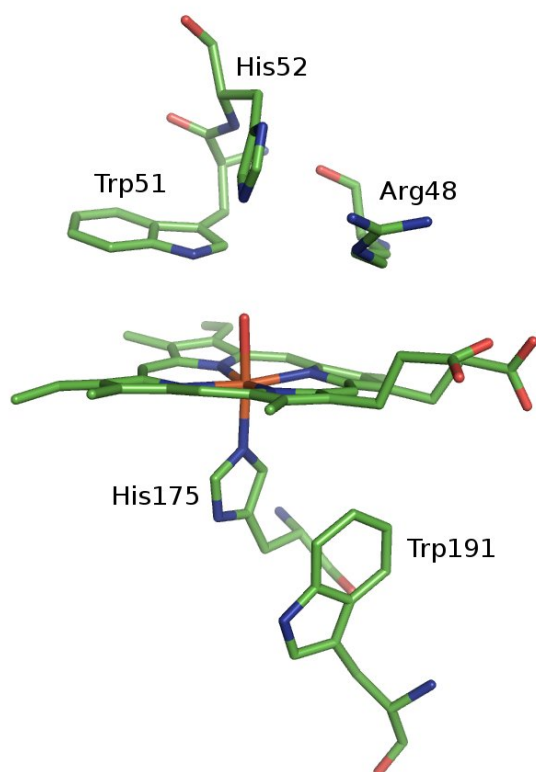
### 1.3. Cytochrom c Peroxidase

Der größte Teil der Untersuchungen dieser Arbeit wurde an einem Testsystem durchgeführt. Dieses bietet die Möglichkeit, die komplexen Zusammenhänge vereinfacht zu betrachten und neue Methoden und Verfahren der Auswertung zu etablieren bzw. validieren. Sind die Fragestellungen hier verstanden, lässt sich das Vorgehen leichter auf ein neues Proteinsystem übertragen. Hauptkriterium für die Auswahl des Testsystems ist hierbei selbstverständlich, ob die relevanten Fragestellungen dort überhaupt beobachtet werden können. Gesucht wurde also eine Bindetasche, in die Liganden zusammen mit Wasser binden. Vereinfachend ist eine solche Bindetasche zusätzlich relativ vergraben, sodass, anders als bei offenen und solvensexponierten Kavitäten, wenig Austausch der Wassermoleküle mit dem das Protein umgebenden Lösemittel zu beobachten ist. Idealerweise sind die Liganden klein und rigide und binden häufig dominiert von einer Wechselwirkung mit der Modellbindetasche. Das System sollte experimentell gut zugänglich und wenn möglich bereits gut charakterisiert sein. Die Wahl fiel in unserem Fall auf die Modellbindetasche der Mutante W191G der Cytochrom *c* Peroxidase, die alle Kriterien erfüllte. Andere Modellbindetaschen wie die von T4 Lysozym L99A oder die L99A/M102Q Tasche des gleichen Proteins sind ungeeignet, da sie teilweise völlig apolar sind und keine verbrückenden Wechselwirkungen mit Wasser beobachtet werden können.

Das Enzym Cytochrom *c* Peroxidase (CCP) katalysiert die Reduktion von Wasserstoffperoxid (oder anderen Hydroperoxiden) zu Wasser. Die Reduktionsäquivalente werden dafür von ferro-Cytochrom *c* zur Verfügung gestellt. Auf diese Weise ist CCP am Abbau toxischer Verbindungen des Intermembran-Raumes der Mitochondrien während des aeroben Metabolismus beteiligt. Das Enzym CCP liegt als Monomer mit 293 Aminosäuren vor und enthält ein Häm im aktiven Zentrum. Bei der katalytischen Reaktion bindet zunächst ein Peroxid an das Eisen des Häm. Daran ist His52 beteiligt. Das Peroxid wird dann heterolytisch gespalten. Dafür wird ein Proton zum anderen Sauerstoffatom übertragen und anschließend die O-O Bindung gebrochen. Das semi-stabile Intermediat wird aus historischen Gründen „Compound I“ genannt. Wenn das Wasser wegdiffundiert, verbleibt ein O am Häm, das von Arg48 und Trp51 koordiniert wird. Die Bindung wird dann zur Fe=O Doppelbindung oxidiert. Das zweite Elektron wird aus dem System durch Oxidation des Trp191 entfernt. In einem zweiten, zweistufigen Schritt wird Compound I über „Compound II“ zu CCP reduziert. Im zweiten Teil dieses Schrittes wird das am Eisen gebundene O freigesetzt<sup>89</sup>. Einen Überblick über den Bereich des aktiven Zentrums und die Konformation der beteiligten Aminosäuren findet man in Abb. 1.2.

Durch Mutation des Trp191 in ein Gly wurde 1994 von der Arbeitsgruppe um D. B. Goodin eine neue Tasche geschaffen<sup>90</sup>. Diese befindet sich, bezogen auf das katalytische Zentrum, auf der gegenüberliegenden Seite des Cofaktors. Die Tasche ist relativ stark ver-





**Abb. 1.2.:** Darstellung des aktiven Zentrums der Cytochrom *c* Peroxidase an Hand der Kristallstruktur (PDB-Code: 4CCX).

graben. Liganden finden Zugang zu dieser Tasche durch einen flexiblen Loop zwischen Pro190 und Asn195<sup>91</sup>. Durch Entfernen des kationischen Trp191 Radikals verbleibt eine negativ geladene Tasche, die bevorzugt kationische Liganden, im wesentlichen kleine, aromatische Heterozyklen, bindet. Entscheidend für die Bindung der Liganden (wie auch für die Stabilisierung des Trp191 im Wild-Typ) ist hierbei das Asp235<sup>92</sup>. Die unkomplezierte Form von CCP W191G enthält fünf Wassermoleküle und ein Kation (z.B. Kalium) in der Bindetasche. Desweiteren verbleiben aber eins oder mehrere der Wassermoleküle in der Kavität, wenn ein Ligand bindet. In der Vergangenheit konnten zahlreiche Kristallstrukturen von Komplexen der CCP W191G bestimmt werden<sup>93-97</sup>. Zahlreiche Liganden wurden dabei auch über computerbasierte Verfahren entdeckt<sup>96,97</sup>. Aktuell sind 52 Kristallstrukturen der CCP W191G bekannt. Neben dieser hervorragenden strukturellen Charakterisierung sind weiterhin von einigen Komplexen ITC Daten<sup>95</sup> wie auch UV Assay Daten<sup>96</sup> bekannt.

Dieses System wurde bereits in der Vergangenheit häufig unter verschiedensten Aspekten als Modellbindetasche für computerbasierte Studien herangezogen. 2003 führten Olson und Mitarbeiter<sup>98</sup> eine Untersuchung durch, bei der die Fähigkeiten, einen Satz bekannter Liganden von nicht-bindenden Molekülen mittels Docking zu separieren und experimentell bekannte Bindemoden zu reproduzieren, geprüft wurde. Später führten Brenk und Kollegen<sup>96</sup> einen Test durch, inwieweit die Balance zwischen geladenen Wechselwirkungen und Desolvatisierungstermen im Docking gegeben ist. Dafür wurden unter Anderem über 5000 Verbindungen aus einer Datenbank in die CCP W191G Tasche platziert. Eine große Her-

ausforderung beim virtuellen Screening ist die Bestimmung der korrekten Reihenfolge der Liganden in Hinblick auf die Bindungsaffinität. Gerade das ist entscheidend, um falsch positive oder falsch negative Ergebnisse zu vermeiden. Deng und Verlinde haben zu diesem Zweck drei verschiedene Docking-Programme und verschiedene Scoring-Funktionen an dieser Modellbindetasche verglichen<sup>99</sup>. Aus ähnlichem Grund wurde in der Gruppe von B. Shoichet ein Rescoring von Docking-Ergebnissen für verschiedene Modellbindetaschen durchgeführt<sup>97</sup>. Die Dynamik dieses Modellsystems wurde 2007 von Baron und McCammon mittels Molekular-Dynamik Simulationen in Hinblick auf die PGG2AAN Loop Region und ihre Einflüsse auf Dynamik des Liganden wie auch der Wassermoleküle im Bereich der Bindetasche untersucht<sup>100</sup>. Später wurden dann die Einflüsse der Loop-Flexibilität auf die Bindungsentropie eines Liganden betrachtet<sup>101</sup>. Schließlich wurde der Versuch unternommen, nicht nur die Flexibilität des Liganden sondern auch des Rezeptors am Beispiel der CCP W191G im computerbasierten Wirkstoffdesign zu berücksichtigen<sup>102</sup>.

### 1.3.1. Auswahl der untersuchten Komplexstrukturen

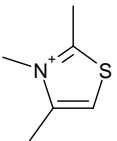
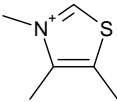
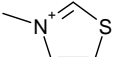
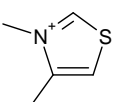
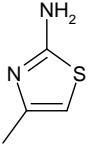
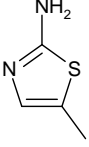
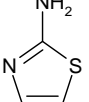
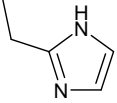
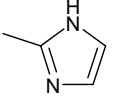
Eine Übersicht über alle aus Röntgenstrukturen bekannten Liganden der CCP W191G Tasche findet sich in Tab. 1.1. Für die Untersuchungen in dieser Arbeit wurde zunächst eine Auswahl aus den zum Startpunkt dieser Arbeit zu Verfügung stehenden Komplexstrukturen getroffen. Als Auswahlkriterium wurde hierbei deren *chemische Ähnlichkeit* herangezogen. Ähnliche oder sogar gleiche funktionelle Gruppen der Liganden ermöglichen einen weitaus einfacheren Vergleich der Ergebnisse bezüglich der zu untersuchenden Bindetaschenwassermoleküle und der Wechselwirkungen mit diesen. Außerdem ist die alchemische Transformation der Liganden untereinander in den freien Energierechnungen praktisch leichter durchführbar.

Die für diese Arbeit getroffene Auswahl für eine nähere Untersuchung fiel auf die Liganden der Komplexe (PDB-Codes) 1AEO, 1AEF, 1AEG, 2AQD und 2EUP: 2-Aminopyridin, 3-Aminopyridin, 4-Aminopyridin, 2,5-Diaminopyridin und 2-Amino-5-picolin. Bei diesen fünf untersuchten Systemen stellten sich zwei Aspekte als besonders interessant heraus:

- Obwohl alle fünf Liganden nur einfache Derivate voneinander sind, wechselwirken sie bzw. binden sogar mit einer unterschiedlichen *Anzahl* von Kristallwassermolekülen.
- Auch der *Bindungsmodus* der Moleküle ist bemerkenswert. Kim stellt in seiner Studie<sup>103</sup> über Ausreißer in QSAR Studien fest, dass strukturell ähnliche Liganden der CCP Tasche einen unerwartet unterschiedlichen Bindungsmodus einnehmen können. In dem von uns gewählten System mit 2,5-Diaminopyridin als Ligand ist es z.B. nicht offensichtlich, weshalb hier der Bindemodus von 2-Aminopyridin angenommen wird, obwohl dieses Molekül sowohl ein Derivat von 2- wie auch von 3-Aminopyridin ist.

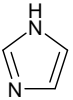
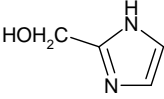
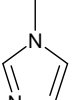
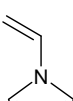


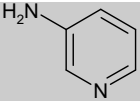


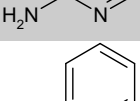
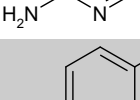
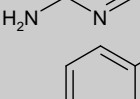
Die strukturellen Eigenschaften der Komplexe werden an späterer Stelle (Abschnitt 5.1.1) eingehender besprochen.

**Tab. 1.1.:** Liganden der CCP W191G Tasche, von denen Kristallstrukturen gelöst wurden (Stand: März 2010). Hinter dem PDB-Code ist die Auflösung in Å angegeben. Die Liganden wurden nach Derivaten in Gruppen eingeteilt und die in dieser Arbeit verwendeten Komplexe grau hinterlegt. Soweit in der Literatur bekannt sind außerdem Werte für die freie Dissoziationskonstante  $K_d$  angegeben. Für die Komplexe, bei denen kein Affinitätswert aufgeführt ist, wurde diese Größe nicht quantifiziert. Die Bestimmung von  $K_d$  erfolgte spektroskopisch. Details dazu finden sich in den angegebenen Referenzen. In den Fällen, in denen zwei Messwerte angegeben sind, wurde der erste von beiden durch ein ITC-Experiment erhalten

Struktur	PDB-Code/Auflösung (Å)	$K_d$ (mM)	Literatur
	1AC4 / 2.10	1.50	Fitzgerald <i>et al.</i> <sup>93</sup>
	1AC8 / 2.10	0.20	Fitzgerald <i>et al.</i> <sup>93</sup>
	1AEB / 2.10	0.30/0.11	Musah <i>et al.</i> <sup>95</sup>
	1AED / 2.10	0.05/0.11	Musah <i>et al.</i> <sup>95</sup>
	1AEH / 2.10	0.23/0.25	Musah <i>et al.</i> <sup>95</sup>
	1AEN / 2.10	0.006/0.008	Musah <i>et al.</i> <sup>95</sup>
	1AEV / 2.10	0.04/0.04	Musah <i>et al.</i> <sup>94</sup>
	1AEQ / 2.10	0.73/0.78	Musah <i>et al.</i> <sup>95</sup>
	1AEU / 2.10	0.07	Fitzgerald <i>et al.</i> <sup>91</sup>

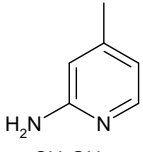
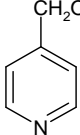
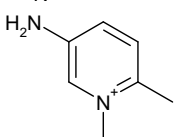
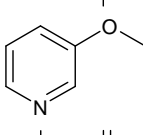
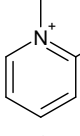
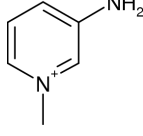
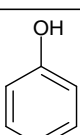
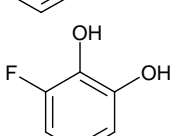
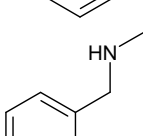
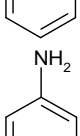
*Fortsetzung auf der nächsten Seite*

## 1. Einführung und Problemstellung

Struktur	PDB-Code/Auflösung (Å)	$K_d$ (mM)	Literatur
	1AES / 2.10	0.07	Fitzgerald <i>et al.</i> <sup>91</sup>
	2EUU / 1.45	-	Brenk <i>et al.</i> <sup>96</sup>
	1AET / 2.10	0.06/0.05	Fitzgerald <i>et al.</i> <sup>91</sup>
	1AEJ / 2.10	0.15/0.14	Musah <i>et al.</i> <sup>95</sup>
	1CMP / 1.24	0.027	Fitzgerald <i>et al.</i> <sup>90</sup>
	2RBY / 1.50	-	Graves <i>et al.</i> <sup>97</sup>
	1AEF / 2.10	0.04/0.07	Musah <i>et al.</i> <sup>95</sup>
	1AEG / 2.10	0.04/0.05	Musah <i>et al.</i> <sup>95</sup>
	1AEO / 2.10	0.05/0.07	Musah <i>et al.</i> <sup>95</sup>
	2ANZ / 1.75	0.06	Brenk <i>et al.</i> <sup>96</sup>
	2AQD / 1.35	-	Brenk <i>et al.</i> <sup>96</sup>
	2EUP / 1.40	0.04	Brenk <i>et al.</i> <sup>96</sup>

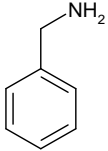
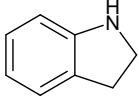
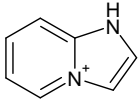
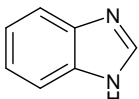
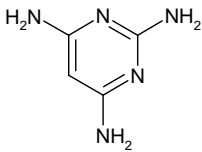
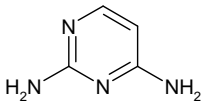
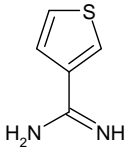
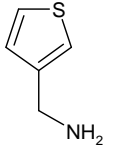
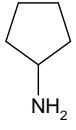
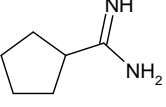
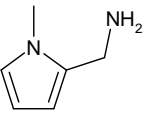
Fortsetzung auf der nächsten Seite

## 1.3. Cytochrom c Peroxidase

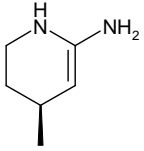
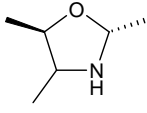
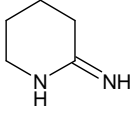
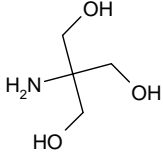
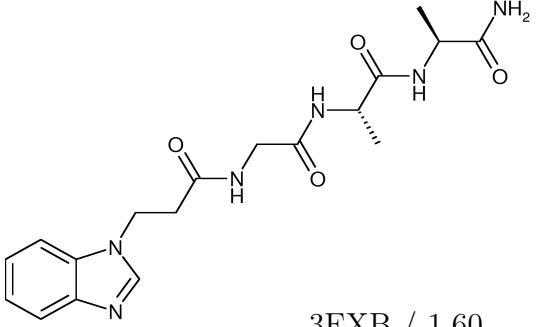
Struktur	PDB-Code/Auflösung (Å)	$K_d$ (mM)	Literatur
	2EUT / 1.12	-	Brenk <i>et al.</i> <sup>96</sup>
	2EUR / 1.39	-	Brenk <i>et al.</i> <sup>96</sup>
	2RBW / 1.50	-	Graves <i>et al.</i> <sup>97</sup>
	2RBZ / 1.80	-	Graves <i>et al.</i> <sup>97</sup>
	2RC2 / 1.50	-	Graves <i>et al.</i> <sup>97</sup>
	2EUO / 1.45	-	Brenk <i>et al.</i> <sup>96</sup>
	2AS3 / 2.40	4.10	Brenk <i>et al.</i> <sup>96</sup>
	2AS4 / 1.30	-	Brenk <i>et al.</i> <sup>96</sup>
	2RBT / 1.24	-	Graves <i>et al.</i> <sup>97</sup>
	1AEE / 2.10	0.03	Musah <i>et al.</i> <sup>95</sup>

Fortsetzung auf der nächsten Seite

## 1. Einführung und Problemstellung

Struktur	PDB-Code/Auflösung (Å)	$K_d$ (mM)	Literatur
	2EUS / 1.55	-	Brenk <i>et al.</i> <sup>96</sup>
	1AEK / 2.10	0.16	Musah <i>et al.</i> <sup>95</sup>
	1AEM / 2.10	0.09	Musah <i>et al.</i> <sup>95</sup>
	1RYC / 1.80	-	Fitzgerald <i>et al.</i> <sup>91</sup>
	2RBX / 1.50	-	Graves <i>et al.</i> <sup>97</sup>
	2EUN / 1.70	0.05	Brenk <i>et al.</i> <sup>96</sup>
	2AS1 / 1.55	0.02	Brenk <i>et al.</i> <sup>96</sup>
	2EUQ / 1.30	0.05	Brenk <i>et al.</i> <sup>96</sup>
	2AS6 / 1.45	-	Brenk <i>et al.</i> <sup>96</sup>
	2RBU / 1.80	-	Graves <i>et al.</i> <sup>97</sup>
	2RBV / 1.39	-	Graves <i>et al.</i> <sup>97</sup>

Fortsetzung auf der nächsten Seite

Struktur	PDB-Code/Auflösung (Å)	$K_d$ (mM)	Literatur
	2RC0 / 1.50	-	Graves <i>et al.</i> <sup>97</sup>
	2RC1 / 2.49	-	Graves <i>et al.</i> <sup>97</sup>
	2AS2 / 1.45	-	Brenk <i>et al.</i> <sup>96</sup>
	1CPF / 2.20	-	Miller <i>et al.</i> <sup>104</sup>
	3EXB / 1.60	-	Puntham <i>et al.</i> <sup>105</sup>

## 1.4. Fragestellungen dieser Arbeit

In Hinblick auf die bereits erwähnten Herausforderungen, die sich, verursacht durch die Beteiligung bzw. Nicht-Beteiligung von Wassermolekülen an der Protein-Ligand Bindung ergeben, ist es ein wichtiges Ziel der computerbasierten Wirkstoffentwicklung, das Vorliegen einer wasser-verbrückten Wechselwirkung bzw. den Grad der Desolvatisierung der Bindetasche durch den Liganden sicher vorherzusagen.

In der hier vorliegenden Arbeit war es zunächst notwendig, die Kristallstrukturen der fünf vorgestellten Komplexe von CCP W191G wie auch die unkomplexierte Form des Proteins detailliert strukturell auszuwerten und miteinander zu vergleichen. Dies umfasste unter anderem auch die Fragestellung, ob die in der Struktur beobachteten Wassermoleküle möglicherweise auf unzureichende Verfeinerung zurückzuführen sind, ob diese also wirklich plausibel erscheinen. Zusätzlich kann man die Strukturen auf Anhaltspunkte für

weitere Wassermoleküle im Bereich der Bindetasche untersuchen, die experimentell nicht beobachtet werden konnten.

Daraufhin galt es, die Fähigkeit zweier ausgewählter Docking-Programme bei der Vorhersage des Bindungsmodus der fünf Komplexliganden zu untersuchen. Insbesondere sollte hier der Frage Beachtung geschenkt werden, ob diese Programme ebenfalls das Wassernetzwerk in der Bindetasche korrekt vorhersagen können, wenn das Docking z.B. in die ligandfreie Bindetasche mit mehreren Wassermolekülen durchgeführt wird. Die Docking-Programme mussten also nach dem Kriterium ausgewählt werden, ob sie selbstständig zwischen einer Verdrängung von Wassermolekülen bei der Komplexierung und einer Beteiligung an der Ligandbindung unterscheiden können. Weiterhin konnte in dieser Docking-Analyse noch betrachtet werden, ob die Durchforstung des Konfigurationsraums die strukturell richtige Lösung für ein System generell findet und ob diese dann auch noch als energetisch günstigste bewertet wurde.

Durch MD Simulationen und der damit einhergehenden Kenntnis der Wasserdynamik soll weiterhin eine verlässlichere Bestimmung und Überprüfung von Kristallwasserpositionen in der Protein-Bindetasche ermöglicht werden. Dadurch können die der Röntgenstrukturanalyse inhärenten Schwierigkeiten bei der Bestimmung struktureller Wassermoleküle ausgeglichen werden. Neben der strukturellen Stabilität und Dynamik des ganzen Systems sollten die MD Trajektorien auf Rückschlüsse über die Stabilität bzw. Verdrängbarkeit von Wassermolekülen auf einzelnen Positionen in der Bindetasche untersucht werden. Die Ergebnisse der einzelnen Systeme sollten danach untereinander verglichen werden.

Um endgültig über die Desolvatisierung der Bindetasche bei der Ligandbindung zu entscheiden, ist eine genaue Kenntnis der Enthalpie- wie auch Entropie-Änderung eines Wassermoleküls bei der Überführung aus der Kavität ins Solvens unerlässlich. Diese thermodynamische Daten sollten deshalb ausgehend von MD Simulationen bestimmt werden. Dafür musste ein geeignetes Verfahren gefunden, unter Umständen implementiert, anschließend validiert und schließlich am gewählten Modellsystem erprobt werden.



# **Teil I.**

## **Methoden**



## 2. Modellierung der Dynamik von Biomakromolekülen

Sehr häufig werden inzwischen chemische Fragestellungen nicht nur mit experimentellen Methoden sondern auch rechnergestützt bearbeitet. Dabei konstruiert man ein Modell, um ausgehend davon messbare Größen vorherzusagen und die Ergebnisse mit denen der Messungen zu vergleichen. Erweist sich ein Modell als geeignet, ist es dann auch häufig in der Lage, nicht messbare Eigenschaften vorherzusagen. Die Computerchemie spielt vor allem dann ihre Vorteile aus, wenn es darum geht, makroskopische Eigenschaften mit mikroskopischen Vorgängen in einem System in Zusammenhang zu bringen. Im Falle von bio- und medizinisch-chemischen Fragestellungen bestehen diese Systeme häufig aus Proteinen. Eine Fragestellung hierbei kann sein, die thermodynamischen Eigenschaften einer Ligandbindung an das Protein mit den strukturellen in Zusammenhang zu bringen. Strukturinformationen von Proteinen können durch röntgenkristallographische Untersuchungen oder durch NMR-spektroskopische Experimente gewonnen werden. Die Dynamik eines Proteins ist mit Beugungsexperimenten jedoch weitgehend nicht zu erfassen und nur im Falle von NMR-Messungen zugänglich. Gerade bei der Frage nach dynamischem Verhalten von Biomakromolekülen leisten Computersimulationen einen sehr entscheidenden Beitrag.

Die quantenchemische Behandlung von einfachen Systemen in der Gasphase ist inzwischen ohne weiteres möglich. Die korrekte Beschreibung von z.B. Proteinen in der flüssigen Phase ist jedoch aufgrund der sehr großen Zahl an Freiheitsgraden auch heute nur sehr eingeschränkt oder gar nicht möglich. Als Alternative bietet sich deshalb eine Behandlung der Wechselwirkung zwischen den Atomen mit Methoden der klassischen Mechanik an. Man spricht hierbei von Molekularer Mechanik. Van Gunsteren<sup>106,107</sup> spricht von vier grundlegenden Problemen der biomolekularen Modellierung:

1. Das **Kraftfeldproblem** bezeichnet die Herausforderung, *sehr kleine Energiedifferenzen* und die *entropischen Einflüsse* in den *unterschiedlichen Atom- und Molekültypen* mit einer klassischen mechanischen Beschreibung der Wechselwirkungen hinreichend genau zu erfassen. Dazu wurden in der Vergangenheit viele unterschiedliche Kraftfelder entwickelt, die jeweils z.B. für kleine Moleküle oder Proteine gut geeignet sind. Dennoch können, trotz der vielen Fortschritte auf diesem Gebiet, bei weitem nicht alle Eigenschaften der kovalenten und nicht-kovalenten Bindungen gleich gut erfasst werden.
2. Mit dem Begriff **Suchproblem** meint man die Aufgabe, die Potentialenergiehyperfläche eines Systems mit sehr vielen Freiheitsgraden in geeigneter Weise nach den

Bereichen zu durchsuchen, die am stärksten zur freien Energie des Systems beitragen. Man sucht also die energetisch günstigsten Bereiche, die gemäß der Boltzmann-Gewichtung am stärksten besetzt sind.

3. Biomolekulare Systeme verhalten sich nach den Gesetzen der statistischen Mechanik. Deshalb reicht es nicht aus, das System allein durch das Minimum auf der Potentialenergiehyperfläche zu charakterisieren. Das System wird vielmehr durch ein Ensemble von Konfigurationen charakterisiert. Dieses muss in einer Simulation generiert werden, und man spricht in diesem Zusammenhang vom **Ensembleproblem**.
4. Die Entwicklung von Kraftfeldern wäre ohne experimentelle Daten nicht denkbar. Andererseits stehen die Ergebnisse einer jeden Simulation nicht für sich alleine, sondern sind im Vergleich zu experimentellen Daten zu betrachten. Daher rührt aber auch das vierte Problem in dieser Reihe, das so genannte **Experimentalproblem**. Experimentelle Daten sind teilweise von *ungenügender Genauigkeit*, von *ungenügender Anzahl*, oder die ermittelten Größen stellen *Mittelungen* dar.

Eine Methode zur Durchsuchung des Konfigurationsraums, *Molekular-Dynamik-Simulationen*, die häufig verwendet wird und auch in dieser Arbeit eingesetzt wurde, soll im Folgenden eingehend erläutert werden. Eine andere, ebenfalls verbreitete Suchmethode, die *Monte-Carlo-Simulation*, wird anschließend zum Vergleich ebenfalls kurz vorgestellt.

### 2.1. Das Kraftfeld

Wie bereits erwähnt, werden die kovalenten und nicht kovalenten Bindungen in biomolekularen Simulationen mithilfe eines Kraftfeldes beschrieben. Die Entwicklung der Kraftfelder für Proteine begann in den frühen 1980er Jahren und wird bis heute voran getrieben. An dieser Stelle seien nur einige der wichtigsten Kraftfelder erwähnt. Das OPLS Kraftfeld, das von Jorgensen und seinen Mitarbeitern entwickelt wurde, wird häufig bei MC-Simulationen eingesetzt<sup>108,109</sup>. Die Entwicklung des CHARMM-Kraftfelds wurde von einer Gruppe rund um Karplus begonnen<sup>110-112</sup> und ist zusammen mit dem gleichnamigen MD-Programmpaket weit verbreitet<sup>113</sup>. Ebenfalls zusammen mit den jeweiligen MD-Programmen GROMOS<sup>114</sup>/GROMACS und AMBER<sup>115</sup> (ein Akronym für „Assisted Model Building with Energy Refinement“) haben das GROMOS-<sup>116,117</sup> und AMBER-Kraftfeld weite Verbreitung gefunden. Das letztgenannte wurde in dieser Arbeit verwendet und entstand in der ersten Hälfte der 1980er Jahre in der Gruppe von Peter Kollman in San Francisco zunächst als so-genanntes *united atom* Kraftfeld. Die damalige Leistung der Computer war im Vergleich zu heutigen sehr begrenzt und man betrachtete deshalb nicht alle Wasserstoffatome explizit. Stattdessen wurden sie zusammen mit den gebundenen Kohlenstoffatomen als ein *united atom* modelliert. In den 90er Jahren wurde dann ein neues AMBER Kraftfeld entwickelt, in dem alle Atome explizit betrachtet werden. Die Entwicklung dieser *all atom* Kraftfelder hält bis heute an. Eine ausführlichere Darstellung über diese Entwicklung findet sich bei Ponder und Case<sup>118</sup>.

In den bis Ende des letzten Jahrzehnts entwickelten Kraftfeldern werden neben den bindenden nur Paarwechselwirkungen berücksichtigt. Mehrkörpereffekte wie Polarisation werden nicht berechnet. Eine noch genauere Annäherung an experimentelle Daten ist aber nur durch Berücksichtigung dieser Effekte im Modell zu erreichen. Speziell in unpolaren Lösungsmitteln sind diese Effekte von großer Bedeutung. In den letzten Jahren wurden Kraftfelder mit unterschiedlicher funktioneller Implementierung entwickelt, die die Berechnung dieser Effekte ermöglichen. Der erhöhte Rechenaufwand führt jedoch zu teureren Simulationen, und es wird diskutiert, ob diese zusätzliche Rechenzeit im Moment die noch wenig erhöhte Genauigkeit rechtfertigt.<sup>118</sup>

Für alle in dieser Arbeit durchgeführten MD Simulationen wurde das AMBER Kraftfeld in der Version von 1999 (Parm99)<sup>119</sup> verwendet. Die Parametrisierung erfolgte hier, wie auch bei den anderen erwähnten Kraftfeldern, an Hand von experimentellen Daten oder quantenchemischen Rechnungen. Bei der Entwicklung wurde hier bewusst ein kleinerer Basissatz (6-31G\*) verwendet, der die Polarität im Vakuum überschätzt und so ein gutes Abbild des Dipolmoments eines Referenzmoleküls im Dielektrikum Wasser ergibt. Die sehr weit verbreitete Nutzung dieses Kraftfeldes bei biomolekularen Simulationen ermöglicht zusätzlich auch eine gute Vergleichbarkeit der Ergebnisse. Das in AMBER verwendete Kraftfeld (ohne Polarisationsterme) hat folgende grundsätzliche Form:

$$V_{\text{total}} = \sum_{\text{bonds}} K_r (r - r_{\text{eq}})^2 + \sum_{\text{angles}} K_\theta (\theta - \theta_{\text{eq}})^2 + \sum_{\text{dihedrals}} \frac{V_n}{2} [1 + \cos(n\phi - \gamma)] + \sum_{i < j}^{\text{atoms}} \left[ \frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^6} \right] + \sum_{i < j}^{\text{atoms}} \frac{q_i q_j}{\epsilon R_{ij}}. \quad (2.1)$$

**Bindungsdehnung.** Der erste Term

$$V(r) = \sum_{\text{bonds}} K_r (r - r_{\text{eq}})^2 \quad (2.2)$$

auf der rechten Seite von Gl. (2.1) beschreibt, analog zum Hookeschen Gesetz, die Dehnung der Bindung zwischen zwei Atomen mit einem harmonischen Potential.  $K_r$  ist hierbei eine Kraftkonstante, die die Rückstellkraft beinhaltet. Der Gleichgewichtsabstand der Bindung wird mit  $r_{\text{eq}}$  bezeichnet, der Abstand mit  $r$ .

**Winkelbeugung.** Ganz analog dazu wird die Auslenkung der Winkel von ihrem Referenzwert durch ein harmonisches Potential beschrieben.

$$V(\theta) = \sum_{\text{angles}} K_\theta (\theta - \theta_{\text{eq}})^2 \quad (2.3)$$

Die Proportionalitätskonstante zwischen Auslenkung und Energie in Gl. (2.3) wird mit  $K_\theta$  bezeichnet, der Referenzwert für den Winkel mit  $\theta_{\text{eq}}$ , der Winkel selbst mit  $\theta$ .

**Torsionsterm.** Unter einer Torsion versteht man den Diederwinkel zwischen zwei Flächen durch jeweils drei Atome. Das zugehörige Potential ist von der Form:

$$V(\phi) = \sum_{\text{dihedrals}} \frac{V_n}{2} [1 + \cos(n\phi - \gamma)]. \quad (2.4)$$

$V_n$  ist ein qualitatives Abbild der Barrierenhöhe. Die absolute Höhe wird durch die nicht-kovalenten Wechselwirkungsterme des Kraftfeldes mitbestimmt.

Im sich anschließenden Cosinus-Term bezeichnet  $n$  die Multiplizität. Das ist die Anzahl der Minima durch die die Potentialfunktion geht, wenn der Torsionswinkel  $\theta$  einmal um  $360^\circ$  variiert wird. Die Drehung um eine Einfachbindung zwischen zwei  $\text{sp}^3$ -Kohlenstoffen kann z.B. mit  $n = 3$  und der Phasenverschiebung  $\gamma = 0^\circ$  beschrieben werden.

Das im Rahmen dieser Arbeit eingesetzte Programmpaket AMBER verwendet bis auf wenige Ausnahmen (z.B. die Torsion um O-C-C-O) nur diesen einen Cosinus-Term zur Beschreibung des Torsionspotentials.

Es kann weiterhin nötig sein, einen zusätzlichen, so genannten „improper torsion“ Term einzuführen. Betrachtet man z.B. Cyclobutanon, so würde ohne diesen Term der Sauerstoff nicht in der Ringebene liegen. Um diese „out-of-plane“ Biegung zu verhindern, werden Pseudo-Torsionspotentiale zwischen vier Atomen, die chemisch nicht seriell sondern verzweigt miteinander verbunden sind, eingeführt.

**Lennard-Jones Potential.** Ein Teil der nicht-kovalenten Kräfte in und zwischen Molekülen sind van-der-Waals Wechselwirkungen. Diese relativ schwache, kurzreichweitige Kraft beinhaltet Wechselwirkungen zwischen Dipolen (Keesom-Kräfte), zwischen Dipol und polarisierbarem Teilchen (Debye-Kräfte) und zwischen polarisierbaren Teilchen (London-Kräfte). Die Wechselwirkungsenergie fällt etwa mit der sechsten Potenz des Abstandes ab. Die gängige, empirische Beschreibung stammt von John Lennard-Jones und ist eine so genannte 12-6-Funktion:

$$V(R) = \sum_{i < j}^{\text{atoms}} \left[ \frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^6} \right]. \quad (2.5)$$

Diese beinhaltet zusätzlich noch einen Term für die Repulsion der Atome. Der Abstand zwischen zwei Teilchen  $i$  und  $j$  wird mit  $R_{ij}$  bezeichnet. Es wird hier ein großer Buchstabe verwendet, um den Unterschied von den Bindungslängen  $r$  im ersten Term von Gl. (2.1) deutlich zu machen. Die Konstanten  $A_{ij}$  und  $B_{ij}$  beinhalten die Tiefe der Potentialmulde und den Gleichgewichtsabstand der jeweiligen Teilchen.

**Elektrostatik-Beiträge.** In der Molekularen Mechanik werden Atome zunächst nur als Punktmassen behandelt. Die Ladungen für die einzelnen Atome müssen so berechnet werden, dass die Gesamtladungsverteilung die molekularen Eigenschaften möglichst gut widerspiegelt. Die Wechselwirkung wird dann mit Hilfe des Coulombschen Gesetzes beschrieben

(letzter Term in Gl. (2.1)).

$$V(R) = \sum_{i < j}^{\text{atoms}} \frac{q_i q_j}{\varepsilon R_{ij}} \quad (2.6)$$

Der Abstand zwischen den beiden Atomen  $i$  und  $j$  wird hierbei mit  $R_{ij}$  bezeichnet. Die zugehörigen Ladungen sind  $q_i$  und  $q_j$ . Die Dielektrizitätskonstante trägt die Bezeichnung  $\varepsilon$ .

Da die nicht-kovalenten Wechselwirkungen zwischen allen Atompaaaren des Systems ermittelt werden müssen, bestimmen sie maßgeblich die Rechenzeit. Während es für die van-der-Waals Wechselwirkungen durchaus üblich ist, einen Abbruchradius zu verwenden, ist dies für die Coulomb-Wechselwirkung nicht zulässig. Das Volumen einer Kugel und damit die Zahl der Teilchen, die in einem Abstand  $R$  von einem Punkt vorgefunden werden können, wächst  $\propto R^3$ . Die Wechselwirkung nimmt hingegen nur mit  $1/R$  ab. Coulomb-Kräfte werden also für zunehmende Abstände immer bedeutsamer.

**RESP-Ladungen.** Wie bereits erwähnt sind geeignete atomare Partiaalladungen eine notwendige Voraussetzung, um die nicht-kovalenten Wechselwirkungen adäquat zu beschreiben. Die Berechnung erfolgt in AMBER mit Hilfe der so genannten RESP-Methode<sup>120</sup>. Das Akronym steht hierbei für „restrained electrostatic potential“.

Zunächst wird aus der Wellenfunktion  $\Psi$  des Moleküls das elektrostatische Potential an ausgewählten Punkten berechnet. Anschließend werden mit der Methode der kleinsten Fehlerquadrate atomare Partiaalladungen berechnet, die dieses Potential am besten widerspiegeln und gleichzeitig in der Summe der Gesamtladung des Moleküls entsprechen. Die Punkte  $i$ , an denen das Potential optimiert wird, werden aus einer Region direkt außerhalb der van-der-Waals Radien gewählt, da hier die Wechselwirkungen am wichtigsten sind. Der RESP Algorithmus<sup>120</sup> verwendet zusätzliche Beschränkungen für Wasserstoffatome. Das aus den Partiaalladungen berechnete, elektrostatische Potential  $\hat{V}_i$ , das über

$$\hat{V}_i = \sum_j \frac{q_j}{r_{ij}} \quad (2.7)$$

von den Ladungen  $q_j$  der Atomzentren  $j$  abhängt, wird mit der Methode der kleinsten Fehlerquadrate bezüglich des aus den quantenchemischen Berechnungen bekannten Potentials  $V_i$  optimiert. Das heißt,  $\chi_{\text{esp}}^2$  muss minimiert werden:

$$\chi_{\text{esp}}^2 = \sum_i \left( V_i - \hat{V}_i \right)^2. \quad (2.8)$$

Im RESP-Verfahren wird nun eine zusätzliche Beschränkung ( $\chi_{\text{rstr}}^2$ ) eingeführt:

$$\chi^2 = \chi_{\text{esp}}^2 + \chi_{\text{rstr}}^2. \quad (2.9)$$

Es hat sich gezeigt, dass man diese Beschränkung am besten in Form einer hyperbolischen

Funktion ansetzt:

$$\chi_{\text{rstr}}^2 = a \sum_j \left( (q_j^2 + b^2)^{1/2} - b \right). \quad (2.10)$$

Hierbei bezeichnet  $a$  einen Skalierungsfaktor, der die Stärke der Beschränkung festlegt und  $b$  eine Konstante für die Breite der hyperbolischen Funktion um das Minimum.

Dieses Vorgehen führt dazu, dass an Kohlenstoffatomen keine artifiziell hohen Ladungen auftreten und die Ladungen sich mit der Konformation des Moleküls nicht mehr stark ändern. Die benötigten Wellenfunktionen werden alle in einer Hartree-Fock-Rechnung mit dem 6-31G(d) Basissatz ermittelt, der sich als bester Kompromiss zwischen Aufwand und Qualität des Ergebnisses herausgestellt hat. Dieser Basissatz überschätzt die Dipolmomente bei Berechnungen im Vakuum gerade so weit, dass ein gutes Abbild der Dipolmomente in einem Dielektrikum wie Wasser erreicht wird.

### 2.1.1. Solvensmodelle

Im Rahmen der Diskussion des Kraftfeldes ist es sinnvoll, sich über den Einsatz verschiedener Solvensmodelle Gedanken zu machen. Bei der Auswahl eines Solvensmodells ist man unabhängig von der Wahl des Kraftfeldes. Dennoch gehört dieser Bereich ebenso zur Parametrisierung in der molekularen Mechanik. Solvensmodelle können in zwei Bereiche unterteilt werden: explizite Berücksichtigung von Lösungsmittelmolekülen und implizite Berücksichtigung. Im Folgenden werden ausschließlich Wassermodelle diskutiert, da nur dieses Lösungsmittel in der vorliegenden Arbeit von Bedeutung ist.

**Implizite Berücksichtigung von Wasser.** In vielen Simulationen wird primär das Verhalten des gelösten Stoffes und nicht des Solvens betrachtet. In diesen Fällen ist es oft sinnvoll, die Einflüsse von Wasser zu berücksichtigen, ohne explizit Wassermoleküle in die Simulation zu integrieren und auf diese Weise wertvolle Rechenzeit einzusparen.

In AMBER ist dies mit Hilfe des ‚generalized Born/surface area‘ Modells möglich. Die interne Energie des Systems wird über das Kraftfeld bestimmt. Diese Funktion wird um zwei Terme erweitert.

$$\Delta G_{\text{solv}} = \Delta G_{\text{el}} + \Delta G_{\text{nonelect}} \quad (2.11)$$

Der zweite Term kann als die Freie Enthalpie betrachtet werden, um ein ladungsfreies Molekül aus dem Vakuum ins Solvens zu überführen.  $\Delta G_{\text{nonelect}}$  enthält somit den Effekt der günstigen van-der-Waals Wechselwirkungen zwischen Solvens und gelöstem Stoff sowie die ungünstigen Beiträge um gegen den Solvensdruck Platz für das Molekül zu schaffen sowie für die Reorganisation der Wasserstruktur. Dieser Beitrag wird über die solvenszugängliche Oberfläche berechnet.

Der erste Term  $\Delta G_{\text{el}}$  ist demnach die Freie Enthalpie, die benötigt wird, um alle Partialladungen im Vakuum zu entfernen und sie dann in Lösung wieder hinzuzufügen. Dieser wird über die Poisson-Boltzmann Gleichung bestimmt. Jedes Molekül wird dabei als eine Kugel mit Radius  $R_i$  und Ladung  $q_i$  betrachtet, in deren Innerem eine Dielektrizitätskonstante von 1 gilt. Das Molekül ist umgeben von einem Medium mit hoher Dielektrizitätskonstante



**Tab. 2.1.:** Ein Vergleich der Parameter verschiedener Wassermodelle.

	SPC <sup>122</sup>	SPC/E <sup>123</sup>	TIP3P <sup>124</sup>	TIP4P <sup>124</sup>
$r(\text{OH})/\text{\AA}$	1.0	1.0	0.9572	0.9572
$\angle \text{HOH}$	109.47	109.47	104.52	104.52
$A \cdot 10^{-3}/(\text{kcal} \cdot \text{\AA}^{12}/\text{mol})$	629.4	629.4	582.0	600.0
$C/(\text{kcal} \cdot \text{\AA}^6/\text{mol})$	625.5	625.5	595.0	610.0
$q(\text{O})$	-0.82	-0.8472	-0.834	0.0
$q(\text{H})$	0.41	0.4238	0.417	0.52
$q(\text{M})$	0.0	0.0	0.0	-1.04
$r(\text{OM})/\text{\AA}$	0.0	0.0	0.0	0.15

$\varepsilon$  (80 für Wasser bei 300 K). Es gilt dann die analytische Lösung:

$$\Delta G_{\text{el}} = -\frac{1}{2} \sum_{ij} \frac{q_i q_j}{f_{\text{GB}}(r_{ij}, R_i, R_j)} \left( 1 - \frac{\exp[-\kappa f_{\text{GB}}]}{\varepsilon} \right) \quad (2.12)$$

Hierbei ist  $r_{ij}$  der Abstand zwischen den Atomen  $i$  und  $j$ . Die elektrostatischen Effekte von Salzen sind über den Debye-Hückel Parameter  $\kappa$  berücksichtigt. Die effektiven Born-Radien  $R_i$  sind ein Maß für die Vergrabenheit eines Atoms. Eine verbreitete Wahl für die Funktion  $f_{\text{GB}}$  ist:<sup>121</sup>

$$f_{\text{GB}} = \left[ r_{ij}^2 + R_i R_j \exp\left(\frac{-r_{ij}^2}{4R_i R_j}\right) \right]^{1/2} \quad (2.13)$$

**Explizite Wassermodelle.** Häufig, wie auch in der hier vorliegenden Arbeit, liegt der Fokus der Betrachtungen speziell auf den Wassermolekülen. An dieser Stelle sollen nun einige bekannte explizite Wassermodelle vorgestellt werden, die heute in MD Simulationen von Biomolekülen häufig Verwendung finden. Wassermodelle können prinzipiell in drei Typen unterteilt werden. Zur ersten Gruppe gehören die Modelle mit starrer Geometrie und einfachen Wechselwirkungspunkten. Die Modelle der zweiten Gruppe berücksichtigen strukturelle Änderungen der Wassergeometrie. Zur dritten Gruppe der Wassermodelle zählen schließlich diejenigen, die Polarisation und Mehrteilchen-Effekte berücksichtigen können.

In Tab. 2.1 findet sich eine Übersicht der verschiedenen Parameter in häufig verwendeten Wassermodellen. Hierbei bezeichnet  $r(\text{OH})$  den Abstand und  $\angle \text{HOH}$  den Winkel zwischen Sauerstoff- und Wasserstoffatomen. Die Parameter  $A$  und  $C$  gehören zur Lennard-Jones Funktion, die die van-der-Waals Wechselwirkungen beschreibt. Mit  $q$  wird die Ladung bezeichnet, wobei  $q(\text{M})$  die Ladung an einem speziellen, für das jeweilige Modell charakteristischen Punkt darstellt. Alle aufgeführten Modelle sind einfach in dem Sinne, dass sie eine starre Geometrie des Wassermoleküls beschreiben. SPC/E stellt eine neuere Version des SPC Modells dar. Dieses unterscheidet sich nur geringfügig vom TIP3P Modell. Elek-

trostatische Wechselwirkungen berücksichtigt dieses Modell zu den Atomen des Wassers, während zur Beschreibung der van-der-Waals Wechselwirkungen ein auf das Sauerstoffatom zentrierter Punkt verwendet wird. Beim TIP4P Modell wird die negative Ladung vom Sauerstoffatom auf einen Punkt auf der Winkelhalbierenden von  $\angle \text{HOH}$  verschoben. Alle Modelle überschätzen das Dipolmoment von gasförmigem Wasser deutlich, erreichen aber recht gut den Wert in kondensierter Phase. Selbstverständlich können diese Modelle nicht alle Eigenschaften von Wasser adäquat beschreiben. Möchte man z.B. ein Schwingungsspektrum vorhersagen, ist es unabdingbar, dass Bindungslängen nicht rigide sind. Aufgrund des größeren Rechenaufwandes sollte das jeweilige Modell den Anforderungen entsprechend gewählt werden.<sup>125</sup>

Die hier vorliegende Arbeit stützt sich vollständig auf das TIP3P Modell<sup>124</sup>. Viele heute vorliegenden Arbeiten verwenden ebenfalls dieses Modell, so dass eine hohe Vergleichbarkeit gewährleistet ist. Im Gegensatz zum TIP4P Modell, das wie TIP3P ebenfalls auf das in AMBER vorhandene Kraftfeld abgestimmt ist, müssen bei einem 3-Punkt-Modell nur neun Abstände und nicht sechzehn für die Wechselwirkung zwischen zwei Wassermolekülen ermittelt werden. Da sich der Rechenaufwand bereits dadurch deutlich verringert, bietet sich die Verwendung dieses Modells bei größeren Systemen an.

## 2.2. Durchsuchung des Konfigurationsraums

### 2.2.1. Molekular-Dynamik Simulationen

Die Methoden, um den Konfigurationsraum eines Systems zu durchsuchen, können prinzipiell in zwei Gruppen unterteilt werden: systematische und heuristische Suchmethoden. Molekulare Dynamik (MD) zählt zu den *heuristischen* Methoden, da ausgehend von einer bekannten Konfiguration eine neue Konfiguration des Systems erzeugt wird. Zwei Faktoren bestimmen diesen Konfigurationswechsel. Zum einen der negative Wert des lokalen Gradienten (die Kraft)  $\partial V/\partial x$  und zum anderen die Trägheit der Freiheitsgrade (eine Art Kurzzeitgedächtnis über den bisher zurückgelegten Weg).<sup>107</sup> Bei der MD-Methode wird die Trajektorie eines Molekülsystems durch die gleichzeitige Integration der Newtonschen Bewegungsgleichung aller Atome des Systems erzeugt. Somit können Gleichgewichtseigenschaften des Systems betrachtet werden. Gleichzeitig können aber auch dynamische Informationen gewonnen werden, da MD-Simulationen eine zeitliche Korrelation aufweisen. Es ist außerdem möglich auch Nichtgleichgewichtszustände zu simulieren.<sup>106</sup>

**Konzept.** Die Newtonsche Bewegungsgleichung

$$\vec{F}_i = m\vec{a}_i = m_i \frac{\partial^2 \vec{r}_i}{\partial t^2} \quad (2.14)$$

beschreibt die Bewegung eines Teilchens mit der Masse  $m_i$  entlang einer Koordinate  $\vec{r}_i$  mit einer Kraft  $\vec{F}_i$ , die auf dieses Teilchen in diese Richtung wirkt. Umformung ergibt

$$\frac{\partial^2 \vec{r}_i}{\partial t^2} = \frac{1}{m_i} \vec{F}_i. \quad (2.15)$$

Die Kraft, die auf ein Teilchen wirkt, ist aus dem Potential  $V$  zugänglich, es gilt:

$$\vec{F}_i = - \frac{\partial V(\vec{r}_1, \vec{r}_2, \dots, \vec{r}_n)}{\partial \vec{r}_i}, \quad (2.16)$$

und damit

$$\frac{\partial^2 \vec{r}_i}{\partial t^2} = \frac{1}{m_i} \left[ - \frac{\partial V(\vec{r}_1, \vec{r}_2, \dots, \vec{r}_n)}{\partial \vec{r}_i} \right]. \quad (2.17)$$

Um die verschiedenen Zustände eines Systems bei einer Temperatur  $T$  zu beschreiben, wird eine Abhängigkeit  $\vec{r}_i(t)$  gesucht, die durch Integration der Bewegungsgleichung erhalten werden kann.

**Integration.** Diese Integration muss auf numerischem Wege erfolgen, wofür verschiedene Methoden zur Verfügung stehen. Die einfachste Variante ist die *Euler* Methode. Hier wird die Ableitung durch einen einfachen Differenzenquotienten angenähert. Die Fehler der Euler Methode sind jedoch nicht zu unterschätzen. Effizienter können Differentialgleichungen numerisch z.B. mit der *Runge-Kutta* Methode gelöst werden. Hierbei wird in jedem Integrationsintervall die Funktion durch verschiedene Zwischenpunkte besser angenähert. Damit stellt das Euler Verfahren nur den Spezialfall erster Ordnung der Runge-Kutta Methoden dar. Weit verbreitet bei der Lösung der Bewegungsgleichung ist der *Verlet*-Algorithmus. Der Fehler bei der Berechnung der Position zum Zeitpunkt  $t + 2\Delta t$  wird hierbei durch das Einbeziehen der Informationen vom Zeitpunkt  $t + \Delta t$  verringert. Ähnlich funktioniert auch der *Leapfrog* Algorithmus, der im Programm AMBER standardmäßig verwendet wird. Hierbei werden Geschwindigkeiten und Positionen versetzt zueinander berechnet. Ausgehend von der Geschwindigkeit zum Zeitpunkt  $t - \frac{1}{2}\Delta t$  wird die Geschwindigkeit für ein Zeitintervall in der Zukunft berechnet.

$$\begin{aligned} \vec{v}_i \left( t + \frac{1}{2} \Delta t \right) &= \vec{v}_i \left( t - \frac{1}{2} \Delta t \right) + \vec{a}_i(t) \Delta t \\ &= \vec{v}_i \left( t - \frac{1}{2} \Delta t \right) + \frac{1}{m_i} \left[ - \frac{\partial V(\vec{r}_1(t), \vec{r}_2(t), \dots, \vec{r}_n(t))}{\partial \vec{r}_i} \right] \Delta t \end{aligned} \quad (2.18)$$

Ausgehend davon kann die neue Position zum Zeitpunkt  $t + \Delta t$  ermittelt werden.

$$\vec{r}_i(t + \Delta t) = \vec{r}_i(t) + \vec{v}_i \left( t + \frac{1}{2} \Delta t \right) \Delta t. \quad (2.19)$$

Zu jedem beliebigen Zeitpunkt kann  $\vec{v}_i(t)$  mit

$$\vec{v}_i(t) = \frac{1}{2} \left[ \vec{v}_i \left( t + \frac{1}{2} \Delta t \right) + \vec{v}_i \left( t - \frac{1}{2} \Delta t \right) \right] \quad (2.20)$$

ermittelt werden. Dieses gegenseitige Überspringen, das der Methode ihren Namen gibt, wird immer weiter fortgesetzt und so die gesamte Trajektorie berechnet. Ein Nachteil dieser Methode ist, dass die Geschwindigkeiten und damit der Beitrag der kinetischen Energie zur Gesamtenergie nicht gleichzeitig mit der Position bestimmt werden können.<sup>125</sup>

**Weitere Aspekte.** Es wird klar, dass eine geeignete Wahl des Zeitschrittes  $\Delta t$  notwendig ist, um die interessanten Bewegungen des Systems hinreichend zu erfassen. Es gilt:

$$\Delta t \ll \nu_{\max}^{-1} \quad (2.21)$$

Die höchstfrequenten Schwingungen eines Systems sind die Valenzschwingungen der Wasserstoffatome. Im Allgemeinen ist man jedoch nicht an Valenzschwingungen interessiert. Fixiert man die Bindungslängen, so können größere Zeitschritte von 1-2 fs gewählt werden und die Kosten einer Simulation deutlich reduziert werden. Es hat sich herausgestellt, dass der Einfluss derartiger Nebenbedingungen unproblematisch ist. Meist wird der so genannte SHAKE-Algorithmus<sup>126</sup> für die praktische Durchführung verwendet.

Die Generierung der Trajektorie erfolgt bei einer MD Simulation üblicherweise im *NVE* Ensemble. Abhängig von der jeweiligen Fragestellung ist die Betrachtung des *NVT* oder *NPT* Ensembles jedoch ebenso relevant. Es gibt verschiedene Ansätze, die Temperatur eines Systems zu beeinflussen. Der einfachste Weg macht sich den direkten Zusammenhang von kinetischer Energie und Temperatur zu Nutze.

$$E_{\text{kin}} = \frac{3}{2} N k_B T = \frac{1}{2} \sum_{i=1}^N m_i \vec{v}_i^2 \quad (2.22)$$

Demnach reicht es aus, die Geschwindigkeiten zu skalieren, um die Temperatur des Systems zu verändern.

$$\lambda = \sqrt{T_{\text{neu}}/T_{\text{aktuell}}} \quad (2.23)$$

Eine Alternative stellt die Kopplung an ein *externes Wärmebad* dar. Das Bad dient dazu bei Bedarf Wärme an das System abzugeben oder von ihm aufzunehmen. In jedem Simulationsschritt werden die Geschwindigkeiten proportional zur Temperaturdifferenz zwischen System und Wärmebad skaliert.

$$\lambda^2 = 1 + \frac{\Delta t}{\tau} \left( \frac{T_{\text{Bad}}}{T(t)} - 1 \right) \quad (2.24)$$

Hierbei bezeichnet  $\tau$  den Kopplungsparameter, der aussagt wie stark System und Wärmebad miteinander verbunden sind. Ein kleines  $\tau$  bedeutet starke Kopplung, ein großes  $\tau$  eine

schwache Kopplung. Bei einem Zeitschritt von 1 fs hat sich  $\tau = 0.4$  ps als geeignet herausgestellt. Durch diesen Ansatz, der auch im Programm AMBER verwendet wird, kann auch weiterhin die Temperatur des Systems fluktuieren.<sup>125</sup>

Da Experimente häufig unter der Bedingung konstanten Drucks durchgeführt werden, ist es von ebenso großer Relevanz auch diese Fälle zu simulieren. Die Methoden hierfür unterscheiden sich nur geringfügig von dem Vorgehen im *NVT* Ensemble. Ganz analog kann durch einen Skalierungsfaktor das Volumen des Systems verändert werden. Volumen und Druck sind über die isotherme Kompressibilität  $\kappa$  miteinander verknüpft.

$$\kappa = -\frac{1}{V} \left( \frac{\partial V}{\partial P} \right)_T \quad (2.25)$$

Alternativ kann auch durch Einführung eines ‚*Druckbades*‘ der Druck direkt angepasst werden.

Führt man eine MD Simulation mit expliziten Solvensmolekülen durch, dann stellt sich die Frage wie man den Rand, beispielsweise bei Verwendung eines Quaders aus Wassermolekülen, behandeln soll. Es ist klar, dass am Rand der Wasserbox um das Protein nicht ein Vakuum beginnen kann. Man kann also den Rand der Box durch ein implizites Solvensmodell beschreiben, oder aber man wählt die so genannten *periodischen Randbedingungen*. Die verwendete Solvensbox wird dabei in alle drei Raumrichtungen unendlich oft aneinander gereiht. Jedes Molekül, das im Verlauf der Simulation die ursprüngliche Box in eine Richtung verlässt, tritt damit automatisch an der gegenüberliegenden Seite wieder in sie ein.

Bedingt durch dieses Konzept ergibt sich eine weitere Herausforderung. Durch die funktionelle Form der van-der-Waals Wechselwirkungen nimmt deren Bedeutung mit größer werdendem  $r$  ab. Für diese nicht-kovalenten Wechselwirkungen kann also gut ein oberer Abstand gewählt werden, bis zu dem diese zum Gesamtpotential beitragen. Der Abstand zwischen Oberfläche eines Proteins und Rand der Wasserbox sollte also deutlich größer sein als dieser Wert, damit das Molekül von Interesse nicht artifiziell mit sich selbst wechselwirkt. Bei den Coulombschen Wechselwirkungen kann jedoch nicht analog vorgegangen werden. Es ist nötig, dass diese Wechselwirkungen aufgrund ihrer Skalierung mit  $r^{-1}$  vollständig erfasst werden. Die Berechnung ist aber auch mit heutigen Computern bei größeren Systemen problematisch. Die nach Ewald benannte *particle mesh Ewald* Methode<sup>127</sup> löst dieses Problem geschickt und ist heute weit verbreitet.

Insgesamt bieten MD-Simulationen eine hervorragende Möglichkeit, um neben Gleichgewichtseigenschaften von Systemen von der Größe von Biomolekülen auch dynamische Prozesse zu beobachten. Aufgrund des zugrundeliegenden Konzeptes kann jedoch nicht der Vorgang einer Bindungsbildung oder des Bindungsbruches beschrieben werden. Dies ist nur durch eine Kombination von Quantenmechanik und MM in der so genannten QM/MM MD möglich. Alle Ergebnisse, die durch MD Simulationen erhalten werden, hängen von der richtigen Wahl der Parameter und des Kraftfeldes ab. Man ist vollständig auf adäquate Parameter angewiesen, da keine Informationen *ab initio* erhalten werden. Durch die ständig

steigende Computerleistung ist es möglich, immer längere Trajektorien zu generieren. Dennoch können komplexe Vorgänge wie die Faltung eines Proteins (noch) nicht beschrieben werden.

Die Herausforderung bei der Auswertung der generierten Trajektorien ist häufig das Erkennen der geeigneten Messgrößen, die aus der Fülle der erhaltenen Informationen Hilfe zum Beantworten der Fragestellung geben. Unterstützung kann man hierbei durch mathematische Verfahren erhalten, die aus den Bewegungen eines Systems vermeintlich wichtigere herausfiltern. Als Beispiele sind hier die Principle Component Analysis oder Essential Dynamics zu nennen. Mit Hilfe von Cluster-Algorithmen können wiederum auch verschiedene Strukturen eines Systems bezüglich ihrer Ähnlichkeit gruppiert werden.

### 2.2.2. Monte-Carlo Simulationen

Im Gegensatz zur Molekular-Dynamik Simulation werden in *Monte-Carlo* Simulationen verschiedene Konformationen und Konfigurationen eines Systems durch zufällige Änderungen erzeugt. Der Wechsel zwischen den verschiedenen Konformationen basiert dabei nur auf der Energie der jeweiligen Konformation. Diese Art der Durchforstung des Konformationsraumes stellt eine Alternative zu den MD Simulationen dar und soll hier kurz erläutert werden. Metropolis Monte-Carlo Algorithmen sind im Bereich der Computerchemie heute weit verbreitet und die erste Computersimulation eines molekularen Systems wurde auf diese Weise durchgeführt. Um zu verstehen, warum diese Methode einige Nachteile im Vergleich zu MD Simulationen aufweist, muss man zunächst verstehen, was diese Methode auszeichnet.

Ausgehend von einer Startkonfiguration werden die kartesischen Koordinaten des Systems eines zufällig ausgewählten Teilchens zufällig verändert. Statt nur ein Teilchen auszuwählen, können auch die Koordinaten von mehreren Teilchen gleichzeitig verändert werden. Anschließend wird die Energie der neuen Konfiguration berechnet. Hierfür wird wieder ein Kraftfeld und damit klassische Mechanik verwendet. Liegt die Energie der neuen Konfiguration unter der Energie der ursprünglichen, wird diese neue Konfiguration des Systems beibehalten. Liegt die Energie jedoch höher, so wird der Boltzmann Faktor mit einer Zufallszahl zwischen 0 und 1 verglichen. Die neue, energetisch höher liegende Konfiguration wird behalten, wenn gilt:

$$\text{rand}(0, 1) \leq \exp\left(\frac{-\Delta V(\vec{r}^N)}{k_B T}\right). \quad (2.26)$$

Hierbei bezeichnet  $\text{rand}(0, 1)$  die Zufallszahl zwischen 0 und 1. Der Term  $\Delta V(\vec{r}^N)$  steht für die Energie des Systems mit seinen  $N$  veränderbaren Koordinaten  $\vec{r}$ .

Ist die von einer Generation zur nächsten durchgeführte Koordinatenänderungen, klein, so führt das dazu, dass möglicherweise (zu) viele Konfigurationen eines Zustandes simuliert werden. Ist die Änderung jedoch zu groß, werden viele Konfigurationen nach dem obigen Ausschlusskriterium aussortiert. Die Wahl der erlaubten Größe der Koordinatenänderun-

gen kann deshalb im Verlauf der Simulation automatisch erfolgen.

Es ist deutlich einfacher, MC-Simulationen in bestimmten Ensembles durchzuführen, d.h. die exakte Kontrolle von Temperatur oder beispielsweise des Druckes ist möglich. Andererseits ist es nicht möglich, zeitabhängige Größen zu betrachten. Bei kleinen Systemen kann durch MC Methoden schnell ein Gleichgewichtswert einer Größe ermittelt werden. Bei großen Systemen ist dies jedoch nicht mehr der Fall. Durch häufige unphysikalische Konfigurationsänderungen ist ein längeres Durchforsten des Phasenraums notwendig. Andererseits wird durch eine MD-Simulation häufig der Konformationsraum eines Zustandes des Systems durchforstet, während es durch MC Algorithmen möglich ist, in einen komplett anderen Bereich des Phasenraums zu gelangen.





### 3. Bestimmung thermodynamischer Größen für Protein-Ligand Komplexe

Der Begriff der *biologischen Aktivität* eines Arzneistoffes bezeichnet, wenn man von den zahlreichen und komplexen Wechselwirkungs- und Verstoffwechslungsvorgängen im menschlichen Körper absieht, die Affinität dieser Substanz für ein bestimmtes Protein. Die Stärke der Bindungsbildung lässt sich mit thermodynamischen Zustandsgrößen beschreiben. Ein Ziel computergestützter Methoden in der Wirkstoffsuche ist eine möglichst exakte Vorhersage dieser Bindungsaffinität. Diese Vorhersage ist jedoch mit unterschiedlichen Herausforderungen verbunden. In diesem Kapitel soll zunächst ein kurzer Überblick über die wichtigsten Prinzipien und Probleme gegeben werden und danach sollen ausgewählte Methoden, die für die vorliegende Arbeit relevant waren, genauer erläutert werden.

**Theoretisches Konzept.** Betrachtet man eine reversible, nichtkovalente Bindung eines Liganden L an ein Protein P gemäß:



so gilt für die Gleichgewichtskonstante der Assoziationsreaktion  $K_a$  bzw. dem dazu reziproken Wert, die Dissoziationsreaktion  $K_d$

$$K_a = K_d^{-1} = \frac{[P'L'_{\text{aq}}]}{[P_{\text{aq}}][L_{\text{aq}}]}. \quad (3.2)$$

Im Gleichgewichtszustand sind die chemischen Potentiale von Hin- und Rückreaktion gleich und es gilt für die Freie Enthalpie (Gibbsche Fundamentalgleichung):

$$\Delta G^\circ = -RT \ln K_a = \Delta H^\circ - T\Delta S^\circ. \quad (3.3)$$

Ziel der Affinitätsbewertung mit Hilfe von Computermethoden ist die möglichst korrekte Vorhersage von  $\Delta G^\circ$ .

**Wechselwirkungsarten und deren Einflüsse auf Enthalpie und Entropie.** Durch den Einsatz einer quantenchemischen Methode würden die verschiedenen Arten der intermolekularen Wechselwirkungen zwischen Ligand und Rezeptor intrinsisch erfasst. Diese sehr genauen, theoretischen Methoden können aber wegen der benötigten Rechenzeit für Protein-Ligand Komplexe nur sehr eingeschränkt angewendet werden. Der Fehler eines Energiewer-

tes aus einer quantenchemischen Rechnung liegt für gewöhnlich im Bereich von 1 kcal/mol. Ist man nun z.B. an unterschiedlichen Bindungsaffinitäten einer Serie von Liganden interessiert, die sich nur durch wenige Funktionalitäten unterscheiden, so können die Differenzen durchaus nur wenige kcal betragen. Deshalb ist es, je nach Fragestellung, wichtig, eine möglichst hohe Genauigkeit bei der Berechnung zu erreichen. In den alternativ eingesetzten Methoden, die auf physikalischen und nicht rein statistischen Modellen basieren, werden die vorhandenen Wechselwirkungen in Einzeltermine separiert oder sie basieren auf Simulationen, die ihrerseits Kraftfelder zur Grundlage haben, die eine Aufteilung in *verschiedene Wechselwirkungsarten* vornehmen.

Ein wichtiger Anteil an der Bindungsenthalpie ist die Stärke der gebildeten *Wasserstoffbrückenbindungen*. Die Ermittlung dieser Größe ist mit einigen Hindernissen verbunden. So ist hierfür selbstverständlich die exakte Kenntnis aller Protonierungszustände nötig. Die Protonierung einiger Liganden sowie verschiedener Aminosäuren wie Histidin, Arginin, Lysin, Asparaginsäure oder Glutaminsäure hängt sehr stark von den elektrostatischen Verhältnissen in der unmittelbaren Umgebung ab, die sich selbstverständlich während der Ligandbindung auch ändern können. Durch die unterschiedlichen Dielektrizitätskonstanten im Inneren eines Proteins und an der Grenzfläche zum Wasser unterscheidet sich auch die Stärke der dort ausgebildeten Wasserstoffbrückenbindungen. Desweiteren können Wasserstoffbrückenbindungen zwischen Ligand und Rezeptor durch Solvensmoleküle vermittelt werden. Ob dies der Fall ist, ob Solvensmoleküle in der Bindetasche vorhanden sind und wie die Vorhersage solcher Wechselwirkungen vorzunehmen ist, ist unter anderem Gegenstand dieser Arbeit. Wünschenswert ist trotz allem eine möglichst exakte Beschreibung der Wasserstoffbrückenbindungen, da sie für gewöhnlich die spezifische Erkennung eines Liganden durch ein Protein ausmachen. Ein Beispiel für eine Scoring-Funktion, in der zwischen unterschiedlichsten Arten von Wasserstoffbrückenbindungen unterschieden wird, ist die Funktion Glide XP<sup>128</sup>.

In diesem Zusammenhang muss ein weiteres Phänomen betrachtet werden, das in Kapitel 1 unter dem Begriff *hydrophober Effekt* eingeführt wurde. Dieser ist für die Überführung eines hydrophoben Teilchens aus Wasser in die Protein-Bindetasche von Bedeutung. Die Größe dieses Effekts ist weitgehend proportional zur hydrophoben, vergrabenen Oberfläche und man hat damit ein Maß für die quantitative Abschätzung dieser daraus resultierenden Freien Enthalpieänderung.

Die Bestimmung des entropischen Beitrags jedoch stellt häufig die größte Herausforderung dar; die Ermittlung dieser Größe ist notwendig, da die Enthalpie alleine, wie oben gezeigt, keine Aussagekraft für die Affinität hat. Obwohl die Freie Enthalpie eine Zustandsfunktion ist, kann sie nicht in Einzelbeiträge aus Atompaarwechselwirkungen separiert werden. Das ist für die Energie möglich, für die Entropie jedoch nicht. Außerdem konnte wiederholt beobachtet werden, dass durch die Vergößerung der Bindungswechselwirkung die Flexibilität eines Liganden eingeschränkt und dadurch die Entropie erniedrigt wurde. Diese zwei Anteile der Freien Enthalpie kompensieren (man spricht von *Enthalpie-Entropie-Kompensation*) sich mitunter weitgehend, sodass letztlich für  $\Delta G^\circ$  keine Veränderung beobachtet werden kann.

**Verfügbare Methoden.** Verschiedenste experimentelle und theoretische Methoden stehen in der Wirkstoffforschung zur Verfügung, um die Bindungskonstanten eines Protein-Ligand Komplexes zu bestimmen. Eine ausführliche Übersicht findet man z.B. bei Gohlke und Klebe<sup>129</sup>. In der Gruppe der experimentellen Verfahren unterscheidet man zwischen *direkten* und *indirekten* Messmethoden. Von den direkten Verfahren hat die isotherme Titrationskalorimetrie die größte Bedeutung. Indirekte Verfahren messen, üblicherweise in einem *Enzymassay*, die Enzymkinetik, beispielsweise über eine spektroskopische Methode, und schließen dann aus den Veränderungen von Konzentrationen und Kinetik auf die Bindungskonstante.

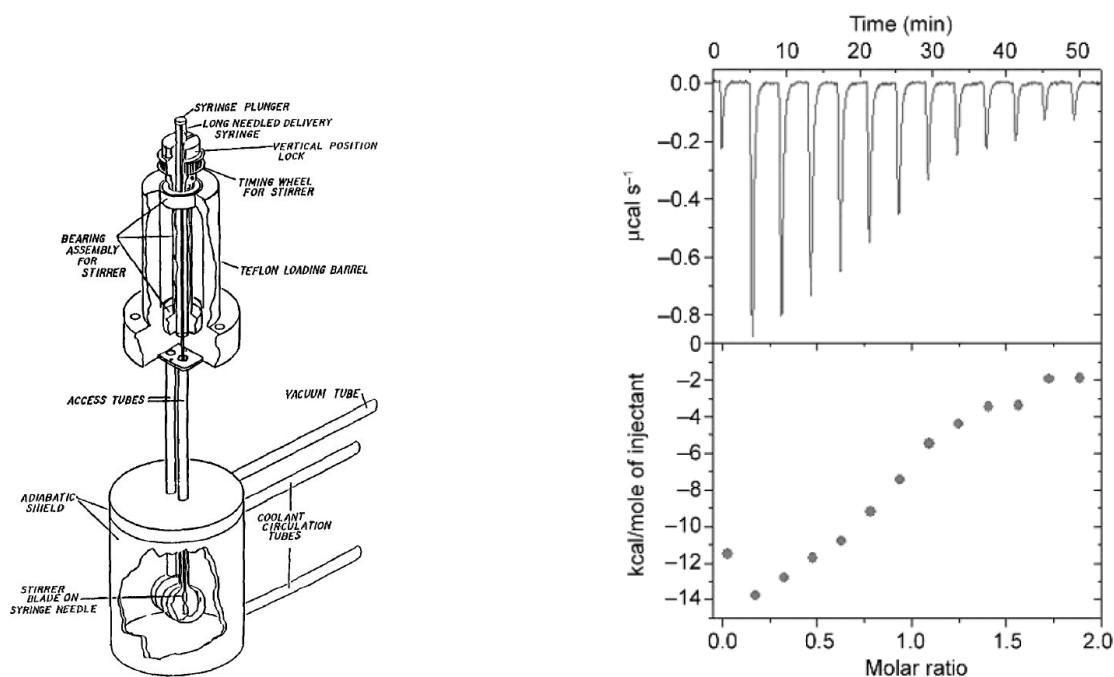
Bei den theoretischen Verfahren muss man zwischen denen, für die die *Kenntnis der Rezeptorstruktur* nicht benötigt wird oder denen für die diese erforderlich ist, unterscheiden. Die erste Gruppe umfasst eine große Zahl so genannter quantitativer Struktur-Wirkungs-Beziehungen (QSAR), die an einem Datensatz von Liganden mit bekannten Affinitätsdaten trainiert und dann zur Vorhersage eingesetzt werden. Zur zweiten Gruppe gehören Scoring-Funktionen oder Methoden, die auf vorangegangenen MD- oder MC-Simulationen beruhen, wie MM-PBSA, thermodynamische Integration (TI)<sup>130</sup>, Freie Energie Störungsrechnung (FEP)<sup>131</sup> oder LIE. Einige Verfahren ermöglichen die Berechnung von absoluten Werten, andere Methoden liefern nur relative Unterschiede in den freien Enthalpien. Im Falle einer schnellen Abschätzung mit Hilfe einer Scoring-Funktion sind die erhaltenen Ergebnisse möglicherweise auch dimensionslos und korrelieren nicht direkt mit Energieeinheiten.

Im Folgenden werden die für diese Arbeit relevanten Verfahren beschrieben.

### 3.1. Isotherme Titrationskalorimetrie

Unter den *mikrokalorimetrischen* Methoden hat die isotherme Titrationskalorimetrie (ITC) eine große Bedeutung zur direkten Messung von  $\Delta G^\circ$ . Hierbei wird schrittweise zu einer gepufferten Lösung des Rezeptors der Ligand hinzugefügt. Dabei wird die Temperaturdifferenz zwischen dieser Proben- und einer Referenzzelle kontinuierlich gemessen. Im Falle des Auftretens einer Reaktion und der damit verbundenen Reaktionswärme wird die Temperatur durch Hinzufügen von Wärme oder Kühlen konstant gehalten. Für jede Injektion wird die auftretende Wärmemengenänderung gemessen. Eine entsprechende Apparatur ist in Abb. 3.1 schematisch dargestellt. Dort findet sich ebenfalls exemplarisch eine ITC Auftragung aus Ladbury und Doyle<sup>132</sup>. Jede einzelne Zugabe von Ligand-Lösung zur Reaktionsmischung führt zu einer detektierbaren Energieänderung. Die einzelnen Signale nehmen jedoch im Zuge der immer weiter voranschreitenden Sättigung der Bindestelle am Protein ab. Aus dieser Primärkurve (Abb. 3.1 rechts oben) kann dann die zweite Auftragung (Abb. 3.1 rechts unten) erhalten werden. Aus dem Verlauf dieser Kurve kann dann die Bindungskonstante berechnet werden.<sup>133</sup> Der erhaltene Wert umfasst sämtliche ablaufende Haupt- und Nebenreaktionen. Nur bei Verwendung unterschiedlicher Pufferlösungen in mehreren Experimenten kann z.B. eine überlagerte Protonenverschiebung erkannt werden. Aus der Titrationskurve kann außerdem auf die Bindungsenthalpie geschlossen werden.

### 3. Bestimmung thermodynamischer Größen für Protein-Ligand Komplexe



**Abb. 3.1.:** Links: Schematische Darstellung einer Messapparatur, die bei einem ITC-Experiment Verwendung findet. Wiedergabe nach Brandts<sup>133</sup> mit freundlicher Genehmigung von Elsevier Inc., Amsterdam, Niederlande. Rechts: Beispiel eines Messergebnisses (oben) aus einem ITC-Experiment und des dazugehörigen Fits (unten) basierend auf dem entsprechenden Reaktionsmodell. Wiedergabe nach Doyle<sup>132</sup> mit freundlicher Genehmigung von John Wiley & Sons Ltd., Chichester, England.

Die Temperaturabhängigkeit der erhaltenen  $\Delta G$ -Werte ist oft gering. Über die Gibbsche Fundamentalgleichung schließt man dann aus den Werten für  $\Delta G$  und  $\Delta H$  auf  $\Delta S$ . Eine Übersicht über den Einsatz von ITC Experimenten im Wirkstoffdesign und über die Schritte beim Durchführen der Messungen findet sich beispielsweise bei Doyle<sup>132</sup>.

Die Möglichkeit, das vollständige thermodynamische Profil einer Reaktion mit einer Messung zu bestimmen und die *Verlässlichkeit* wie auch die *Genauigkeit* der Methode haben dazu geführt, dass ITC sich zum Standard bei der Validierung theoretischer Vorhersagen thermodynamischer Werte entwickelt hat. Nichtsdestoweniger sind Entropie und Enthalpie alleine oft nicht ausreichend für eine strukturelle Interpretation der Ergebnisse. Die Bindungsenthalpien wurden ja für die Gesamtreaktionen gemessen. Sämtliche Nebenreaktionen, wie Protonierungen, Deprotonierungen, Tautomerisierungen oder Konformationsänderungen sind in dem Wert enthalten. In einem systematischen Vergleich zwischen den gemessenen Bindungsstärken und Eigenschaften der Liganden kann man aber durchaus Zusammenhänge beobachten. In der SCORPIO („Structure/Calorimetry of Reported Protein Interactions Online“) Datenbank wurden von Olsson *et al.*<sup>134</sup> 254 Protein-Ligand Komplexe bezüglich der Korrelation von Bindungsaffinität und der Größe der apolaren Oberfläche untersucht.

## 3.2. Scoring-Funktionen

Bei der Diskussion der Scoring-Funktionen müssen drei Typen unterschieden werden: *empirische*, *kraftfeldbasierte* und *wissensbasierte* Scoring-Funktionen. Allen drei Gruppen ist gemein, dass sie versuchen, eine *schnelle* Abschätzung über die Bindungsenergie des Liganden zu treffen und an experimentellen Strukturdaten kalibriert wurden. Die erhaltenen Werte unterschiedlicher Funktionen sind dabei in ihrer absoluten Skala nicht miteinander vergleichbar, wohl aber die erhaltene Rangfolge der Liganden. Auch sind die Werte nicht notwendigerweise mit absoluten Werten der freien Bindungsenthalpie in Beziehung zu setzen, da z.B. Konformationsänderungen des Rezeptors unberücksichtigt bleiben. Die Korrelation mit der Bindungsstärke ist möglich, wenn man statt einer einfachen Funktion ein weitaus komplexeres Verfahren wie eine vollständige statistisch-thermodynamische Betrachtung durchführen würde. Dies ist jedoch aufgrund der benötigten Rechenzeit nicht praktikabel. Ausführlichere Übersichten<sup>129,135–137</sup> und Vergleiche<sup>138–141</sup> zu den verfügbaren Scoring-Funktionen finden sich in der Literatur.

**Empirische Scoring-Funktionen.** Mit dem Begriff „empirische Scoring-Funktion“ bezeichnet man die Gruppe von Funktionen, die die Energie der Ligand-Bindung durch eine Summe von Termen ermitteln, die z.B. die Anzahl der Wasserstoffbrückenbindungen, die hydrophoben Kontaktflächen oder den Verlust an Entropie durch das Einschränken von rotierbaren Bindungen enthalten. Die Gewichtung dieser Terme wurde an experimentellen Affinitäten kalibriert. Beispiele für derartige Funktionen sind ChemScore<sup>142,143</sup>, X-Score<sup>144</sup> und SFCscore<sup>145</sup>. Alle empirischen Scoring-Funktionen unterscheiden sich dabei in Anzahl und Art der verwendeten Terme.

**Kraftfeldbasierte Scoring-Funktionen.** Ebenso ist es möglich, die Bindungsenergie mit einer klassischen mechanischen Kraftfeldfunktion abzuschätzen. Kraftfelder, die entweder zur Beschreibung von kleinen Molekülen oder von Makromolekülen entwickelt wurden, tendieren dazu, die Protein-Ligand Bindungsenergie zu überschätzen. Deshalb werden häufig zusätzliche Terme, die entropische oder Solvatationseffekte erfassen, eingeführt. Ein Beispiel für eine kraftfeldbasierte Scoring-Funktion ist die AutoDock Scoring-Funktion. Hier wurde eine Kombination aus einem Kraftfeld-Term basierend auf dem AMBER Kraftfeld und zwei empirischen Termen für Solvatation und Ligand Entropie gewählt. Auch DockScore verwendet die AMBER Kraftfeldparameter, während GoldScore die von Tripos verwendet.

**Wissensbasierte Scoring-Funktionen.** Die dritte große Gruppe von Scoring-Funktionen wurde aus statistischen Analysen von Kristallstrukturen entwickelt, die die Verteilung von bestimmten Protein-Ligand Atom-Kontakten beschreiben. Daraus wiederum werden Paar-potentiale entwickelt. Der Score einer Docking-Lösung setzt sich aus der Summe aller Atompaar-Wechselwirkungen zusammen. Beispiele für solche Scoring-Funktionen sind die verschiedenen Varianten von Drugscore<sup>34,35,146</sup> oder PMF<sup>147,148</sup>.

**Fitness-Funktionen im Programm GOLD.** Von den im Docking-Programm GOLD implementierten Scoring-Funktionen wurden zwei in dieser Arbeit verwendet. GoldScore<sup>50</sup> bewertet die Docking Lösung mit der Summe dreier Energiekomponenten. Das sind die aufgrund von Abständen und Winkeln gewichtete Summe aller Wasserstoffbrückenbindungsenergien zwischen Ligand und Rezeptor, die Wechselwirkungsenergie des Komplexes, die mit Hilfe eines 4-8 Potentials berechnet wird, und die interne Energie des Liganden. Die sterische Energie wird dabei über ein 12-6 Potential, die Torsions-Energie wie im Tripos Kraftfeld ermittelt. Aufgrund dieser Kombination ist GoldScore den kraftfeldbasierten Scoring-Funktionen zuzuordnen.

Chemscore<sup>142,143</sup> bestimmt die Bindungswechselwirkung als Summe von vier verschiedenen Wechselwirkungstermen. Die Gewichtung jedes einzelnen Terms wurde an einem Datensatz aus 82 Protein-Ligand Komplexen mit experimentell gemessenen Affinitätsdaten kalibriert. Die Funktion enthält zusätzlich noch Terme, die zu dichte Kontakte zwischen Rezeptor und Ligand sowie ungünstige Konformationen des Liganden beim Docking verhindern sollen.

$$\text{ChemScore} = \underbrace{\Delta G_0 + \Delta G_{\text{hbond}} + \Delta G_{\text{metal}} + \Delta G_{\text{lipo}} + \Delta G_{\text{rot}}}_{\Delta G_{\text{binding}}} + P_{\text{clash}} + c_{\text{internal}}P_{\text{internal}} + (c_{\text{covalent}}P_{\text{covalent}} + P_{\text{constraint}}) \quad (3.4)$$

Mit  $P$  sind hier die sogenannten Strafterme (englisch: penalty) bezeichnet. Das sind einerseits Beiträge gegen zu dichte Kontakte („clash“), ungünstige Konformationen des Liganden („internal“) oder zusätzliche kovalente Beiträge („covalent“) und optionale vom Anwender gesetzte Beschränkungen („constraint“).

Für den Fall, das GOLD eingesetzt wird, um die Verdrängung oder Konservierung von Wassermolekülen aus der Bindetasche zu simulieren, werden die Wechselwirkungen dieser Wassermoleküle mit der gleichen Funktion wie die Protein-Ligand Interaktionen ermittelt. Im Fall von GoldScore ist das der Term für Wasserstoffbrückenbindungen und der van-der-Waals-Term. Bei Chemscore sind das die Terme für H-Brücken, Metall und ‚Clash‘. Da der Entropieverlust eines Wassermoleküls bei Bindung an das Protein in GOLD vereinfachend als konstant angenommen wird, wird bei beiden Scoring-Funktionen ein Term der Form

$$\sum o(w)(\sigma_p + \sigma_i(w)) \quad (3.5)$$

hinzugefügt. Hierbei bezeichnet  $o(w)$  den Besetzungsgrad des Wassermoleküls  $w$  (der Wert ist gleich 1 wenn das Wassermolekül vorhanden und gleich 0 wenn es nicht vorhanden ist),  $\sigma_p$  stellt die Änderung von  $\Delta G$  durch den Entropieverlust dar,  $\sigma_i(w)$  schließlich stellt die Wechselwirkung des Wassermoleküls mit der Umgebung dar und wurde eingeführt, um die Orientierung des Teilchens mit vorhersagen zu können.<sup>52</sup>

**Scoring in FlexX.** Die Abschätzung der freien Bindungsenthalpie in FlexX erfolgt über eine Funktion der Form<sup>48</sup>:

$$\begin{aligned} \Delta G = & \Delta G_0 + N_{\text{rot}} \Delta G_{\text{rot}} + \Delta G_{\text{hb}} \sum_{\text{neutral H-bonds}} f(\Delta R, \Delta \alpha) \\ & + \Delta G_{\text{io}} \sum_{\text{ionic int.}} f(\Delta R, \Delta \alpha) + \Delta G_{\text{aro}} \sum_{\text{aro int.}} f(\Delta R, \Delta \alpha) + \Delta G_{\text{lipo}} \sum_{\text{lipo cont.}} f^*(\Delta R). \end{aligned} \quad (3.6)$$

Hierbei ist  $N_{\text{rot}}$  die Zahl der rotierbaren Bindungen,  $f(\Delta R, \Delta \alpha)$  eine Skalierungsfunktion, die Abweichungen von der idealen Geometrie bestraft ebenso wie  $f^*(\Delta R)$ , welche nur von den Atom-Atom Kontakten abhängig ist und ideale Distanzen belohnt und zu dichte bestraft. Die Parameter für die einzelnen Bindungsbeiträge wurden bei der Entwicklung teilweise aus einer früheren Scoring-Funktion übernommen.

Wassermoleküle werden in FlexX nach dem sogenannten „Particle Concept“ gehandhabt.<sup>47</sup> Die energetische Bewertung der Wechselwirkungen eines Wassermoleküls erfolgt hierbei nicht aufgrund von experimentellen Daten oder quantenchemischen Berechnungen. Für die Wechselwirkungen wurden folgende Parameter gesetzt: Von den vier gerichteten Wechselwirkungen, die ein Wassermolekül eingehen kann, wird jede freie mit 1 kJ/mol bewertet. Die H-Brücke zu einem Liganden wird mit einem Betrag von 2.7 kJ/mol gewichtet (bei optimaler Geometrie) und der optimale Winkel für eine Wechselwirkung wird auf 110° gesetzt.

## 3.3. Thermodynamische Integration und Freie-Energie Störungsrechnung

### 3.3.1. Begrifflichkeiten

Die statistische Thermodynamik verwendet, im Gegensatz zur klassischen Thermodynamik, mikroskopische Theorien, um Aussagen über makroskopische thermodynamische Vorgänge zu machen. Größen wie z.B. Temperatur, Wärme oder Entropie werden mit Hilfe von klassischer Mechanik oder Quantenmechanik auf die Bewegung und Wechselwirkung einzelner Teilchen zurückgeführt. Jeder *Makrozustand* wird hierbei durch eine große Zahl verschiedener *Mikrozustände* realisiert. Die *Zustandssumme*  $Z$ , die aus einer Verteilung von Mikrozuständen ermittelt werden kann, ermöglicht dann die Berechnung von klassischen physikalisch-chemischen Größen. Nach der *Ergodenhypothese* kann man die statistischen Informationen über ein System auf zwei äquivalente Weisen bekommen. Man erhält das Scharmittel des Systems, wenn man die Zustände von hinreichend vielen Teilchen festhält und mittelt. Andererseits kann man auch die Zustände eines Teilchens über einen hinreichend langen Zeitraum festhalten und mitteln. Der Begriff des *Ensembles* bezeichnet die Gesamtheit aller Systeme, die zur Berechnung des Scharmittels herangezogen werden. Je nach gewählten Randbedingungen unterscheidet man unterschiedliche Ensemble.

- Das *kanonische Ensemble* besteht aus Systemen, die gleiche Teilchenzahl, gleiches Volumen und gleiche Temperatur, also kinetische Energie, haben.
- Das *mikrokanonische Ensemble* besteht aus Systemen, die neben gleicher Teilchenzahl und gleichem Volumen auch die gleiche Energie haben.
- Das *großkanonische Ensemble* besteht aus Systemen, die gleiches Volumen und gleiche Temperatur haben, zwischen denen aber ein Teilchenaustausch möglich ist.
- Das *isobar-isotherme Ensemble* besteht aus Systemen die gleiche Temperatur, gleichen Druck und gleiche Teilchenzahl aufweisen.

Die Methoden der statistischen Thermodynamik können nun dazu genutzt werden, um aus dem Ensemble, das z.B. mit Hilfe einer MD-Simulation generiert wurde, eine makroskopische thermodynamische Größe abzuleiten. Das Vorgehen dafür soll nun im Einzelnen näher erläutert werden.

#### 3.3.2. Allgemeine Vorüberlegungen

Für den Anteil der Zustandssumme  $Z$ , der von der potentiellen Energie des Realisierungszustandes des Systems abhängt, gilt

$$Z_{NVT} = \int \exp \left[ -\frac{V(\vec{r})}{k_B T} \right] d\vec{r}. \quad (3.7)$$

Hierbei bezieht  $V(\vec{r})$  das Potential an den Koordinaten  $\vec{r}$ ,  $k_B$  ist die Boltzmannkonstante. Da die Behandlung im kanonischen Ensemble zu weniger komplexen Gleichungen führt als die im isobar-isothermen Ensemble, wird der Einfachheit halber im Folgenden beispielhaft die Freie Energie  $A$ , die aus der Zustandssumme bei konstantem Volumen erhalten werden kann, diskutiert. Die Behandlung von  $G$  ist jedoch prinzipiell analog. Die Freie Energie ist definiert als  $A = U - TS$ . Hierbei bezeichnet  $U$  die innere Energie eines Systems. Zur Umrechnung zwischen  $\Delta G$  und  $\Delta A$  nutzt man den Zusammenhang

$$\Delta G = \Delta A + p\Delta V. \quad (3.8)$$

Für die kondensierte Phase ist  $\Delta V$  jedoch häufig vernachlässigbar. Die Freie Energie wiederum ist mit der Zustandssumme über

$$A = -k_B T \ln \left( \frac{Z_{NVT}}{V^N} \right) \quad (3.9)$$

( $N$  Teilchenzahl) verknüpft. Bei Kenntnis aller möglichen Zustände eines Systems könnte auf diese Weise die Freie Energie des Systems direkt ermittelt werden. Durch eine MD-Simulation ist jedoch niemals der gesamte Konfigurationsraum zugänglich. Es werden vorwiegend Konfigurationen in einem Bereich niedriger Energie durchmustert. Die



hochenergetischen Konfigurationen sind jedoch wegen der Energiefunktion im Exponenten der Zustandssumme von entscheidender Bedeutung.

Absolute Werte für die Freie Energie sind also aus den Trajektorien einer MD-Simulation nicht ohne Weiteres zugänglich. Anders ist dies aber bei den Differenzen der Freien Energie zweier wenig unterschiedlicher Systeme. Unterscheiden sich zwei Systeme chemisch nämlich nicht sehr stark voneinander, kann man annehmen, dass sich auch deren Verteilung der hochenergetischen Konfigurationen nur wenig unterscheidet. Da die niederenergetischen Konfigurationen durch Simulationen gut zugänglich sind, kann die Differenz der freien Energie zwischen diesen zwei Systemen berechnet werden. Bezeichnet man die beiden Systeme mit ,0‘ und ,1‘ und führt weiterhin einen Parameter  $\lambda$  ein, der eine lineare Variation von ,0‘ nach ,1‘ möglich macht, so lässt sich die Potentialfunktion folgendermaßen schreiben:

$$V(\vec{r}, \lambda) = (1 - \lambda)V_0(\vec{r}) + \lambda V_1(\vec{r}). \quad (3.10)$$

Die Freie Energie ist eine Zustandsfunktion und ihre Differenz ist somit nur von den beiden Endpunkten abhängig. Aus diesem Grund ist es gerechtfertigt, eine beliebige Variation von ,0‘ nach ,1‘ vorzunehmen.

### 3.3.3. Freie-Energie Störungsrechnung

Die *Freie-Energie Störungsrechnung* wurde in den 1950er Jahren von R. Zwanzig begründet<sup>131</sup>. Im Folgenden soll die Entwicklung der nötigen Gleichungen kurz dargelegt werden. Für  $\Delta A$  gilt

$$\Delta A = A_1 - A_0 = -k_B T \ln \left( \frac{Z_{NVT1}}{Z_{NVT0}} \right). \quad (3.11)$$

Verwendet man die Definition der Zustandssumme, so erhält man für den Quotienten  $Z_1/Z_0$

$$\frac{Z_{NVT1}}{Z_{NVT0}} = \int \exp \left[ -\frac{V_1(\vec{r})}{k_B T} \right] d\vec{r} / \int \exp \left[ -\frac{V_0(\vec{r})}{k_B T} \right] d\vec{r} \quad (3.12)$$

$$= \frac{\int \exp \left[ -\frac{V_1(\vec{r})}{k_B T} \right] \overbrace{\exp \left[ \frac{V_0(\vec{r})}{k_B T} \right] \exp \left[ -\frac{V_0(\vec{r})}{k_B T} \right]}^{=1} d\vec{r}}{\int \exp \left[ -\frac{V_0(\vec{r})}{k_B T} \right] d\vec{r}} \quad (3.13)$$

$$= \int \exp \left[ -\frac{V_1(\vec{r}) - V_0(\vec{r})}{k_B T} \right] \frac{\exp \left[ -\frac{V_0(\vec{r})}{k_B T} \right]}{Z_{NVT0}} d\vec{r} \quad (3.14)$$

$$= \int \exp \left[ -\frac{V_1(\vec{r}) - V_0(\vec{r})}{k_B T} \right] P_0(\vec{r}) d\vec{r} \quad (3.15)$$

$$= \left\langle \exp \left[ -\frac{\Delta V(\vec{r})}{k_B T} \right] \right\rangle_0. \quad (3.16)$$

Diese Umformung verwendet die Boltzmannsche Wahrscheinlichkeitsfunktion

$$P(\vec{r}) = \frac{\exp\left[-\frac{V(\vec{r})}{k_B T}\right]}{Z_{NVT}}. \quad (3.17)$$

$P(\vec{r})$  wird letztlich durch die Simulation erhalten und so ergibt sich der Quotient durch den Ensemblemittelwert, der durch die Bracketnotation gekennzeichnet wird. Das Subskript „0“ bedeutet hierbei, dass das Ensemble unter Verwendung des Potentials  $V_0(\vec{r})$  generiert wurde. Es folgt für die Differenz der Helmholtz-Energien

$$\Delta A = -k_B T \ln \left\langle \exp\left[-\frac{\Delta V(\vec{r})}{k_B T}\right] \right\rangle_0. \quad (3.18)$$

Analog erhält man

$$\Delta A = k_B T \ln \left\langle \exp\left[\frac{\Delta V(\vec{r})}{k_B T}\right] \right\rangle_1. \quad (3.19)$$

Man unterteilt nun das Intervall zwischen Potentialfunktion  $V_0(\vec{r})$  von Ausgangs- und Endzustand  $V_1(\vec{r})$  in mehrere Teilintervalle mit Hilfe der so genannten  $\lambda$ -Kopplung. Obige Ableitung setzte voraus, dass beide Potentialfunktionen  $V_0(\vec{r})$  und  $V_1(\vec{r})$  sich nur gering voneinander unterscheiden. Auf diese Weise kann man die Energiedifferenz für die gesamte Transformation mit der Freien-Energie Störungsrechnung ermitteln. Für die gesamte Energieänderung bildet man die Summe aus den Energieänderungen der einzelnen Teilschritte zwischen  $\lambda_i$  und  $\lambda_{i+1}$ . Es gilt:

$$\Delta A = \sum_{i=0}^{k-1} \Delta A_{\lambda_i \rightarrow \lambda_{i+1}}. \quad (3.20)$$

#### 3.3.4. Thermodynamische Integration

Ein weiteres, weit verbreitetes Verfahren, um die Energieänderung bei der Transformation von einem Zustand „0“ nach „1“ zu beschreiben, ist die so genannte *Thermodynamische Integration*. Auch hierbei wird der gesamte Prozess in mehrere Teilvorgänge mit Hilfe eines Parameters  $\lambda$  zerlegt. Die Helmholtz-Energie ist somit eine Funktion von  $\lambda$  und es gilt

$$\Delta A = \int_0^1 \frac{\partial A(\lambda)}{\partial \lambda} d\lambda. \quad (3.21)$$

Es gilt also die Ableitung der Energie nach dem Parameter  $\lambda$  zu bestimmen. Es ergibt sich<sup>125</sup>

$$\frac{\partial A(\lambda)}{\partial \lambda} = -k_B T \frac{\partial \ln Z_{NVT}(\lambda)}{\partial \lambda} \quad (3.22)$$

$$= \frac{-k_B T}{Z_{NVT}(\lambda)} \frac{\partial Z_{NVT}(\lambda)}{\partial \lambda} \quad (3.23)$$

Mit Hilfe der Kettenregel erhält man unter Verwendung von Gl. (3.7) für die Ableitung

$$\frac{\partial Z_{NVT}(\lambda)}{\partial \lambda} = \int \frac{\partial}{\partial \lambda} \exp \left[ -\frac{V(\vec{r}, \lambda)}{k_B T} \right] d\vec{r} \quad (3.24)$$

$$= -\frac{1}{k_B T} \int \frac{\partial V(\vec{r}, \lambda)}{\partial \lambda} \exp \left[ -\frac{V(\vec{r}, \lambda)}{k_B T} \right] d\vec{r}. \quad (3.25)$$

Fügt man nun beide Ausdrücke zusammen, so ergibt sich

$$\frac{\partial A(\lambda)}{\partial \lambda} = \frac{1}{Z_{NVT}(\lambda)} \int \frac{\partial V(\vec{r}, \lambda)}{\partial \lambda} \exp \left[ -\frac{V(\vec{r}, \lambda)}{k_B T} \right] d\vec{r} \quad (3.26)$$

$$= \int \frac{\partial V(\vec{r}, \lambda)}{\partial \lambda} \frac{\exp \left[ -\frac{V(\vec{r}, \lambda)}{k_B T} \right]}{Z_{NVT}(\lambda)} d\vec{r} \quad (3.27)$$

$$= \int \frac{\partial V(\vec{r}, \lambda)}{\partial \lambda} P(\vec{r}, \lambda) d\vec{r}. \quad (3.28)$$

Damit gilt für die gesamte Energieänderung im Intervall  $\lambda = 0$  bis  $\lambda = 1$  mit Hilfe des Ensemblemittelwertes für jedes  $\lambda$

$$\Delta A = \int_0^1 \left\langle \frac{\partial V(\vec{r}, \lambda)}{\partial \lambda} \right\rangle_\lambda d\lambda. \quad (3.29)$$

### 3.3.5. Verbesserung des Samplings

Voraussetzung für eine möglichst exakte Vorhersage einer Freien Energiedifferenz ist ein weitgehend vollständiges Abtasten des Konfigurationsraums. Dies gelingt mit der konventionellen Techniken der MD Simulation im Zusammenhang mit der heute verfügbaren Rechenleistung bereits sehr gut. Ist man aber an besonders exakten Ergebnissen für größere strukturelle Änderungen oder gar absoluten Änderung der Freien Energie interessiert, kommen die Standardverfahren an ihre Grenzen. Es existieren unterschiedliche Varianten von MD Simulationen, um das Durchmustern des Konfigurationsraums effektiver zu gestalten, wie *Replica Exchange Molecular Dynamics*<sup>149–151</sup> oder *Umbrella Sampling*<sup>152–154</sup>. Beim adaptiven Umbrella Sampling in Berechnungen der Freien Energie wird  $\lambda$  als eine weitere Koordinate betrachtet und der Konfigurationsraum auch hinsichtlich dieser Größe

durchsucht. Letztlich kann dann aus dem Potential in Abhängigkeit von  $\lambda$  die Änderung der Freien Energie bestimmt werden. Bei dieser Methode kann es jedoch passieren, dass der Wert für  $\lambda$  zu schnell verändert wird und das System nicht ausreichend schnell darauf reagieren kann. Die Vorteile der Replica Exchange Methoden hingegen wurden erst vor nicht allzu langer Zeit auch für Freie Energie Berechnung mit MD Simulationen zugänglich gemacht<sup>155,156</sup>. Beim so genannten *Parallel Tempering* werden Kopien (*Replicas*) des Systems bei unterschiedlichen Temperaturen simuliert. In regelmäßigen Abständen wird dann geprüft, ob die Koordinaten einer Kopie  $i$  mit einer anderen Kopie  $j$  getauscht werden können. Die Bedingung orientiert sich (hier am Beispiel des *NPT* Ensembles) am MC Algorithmus

$$\exp \left[ \left( \frac{1}{k_B T_j} - \frac{1}{k_B T_i} \right) (E(j) - E(i) + P(V(j) - V(i))) \right] \geq \text{rand}(0, 1) \quad (3.30)$$

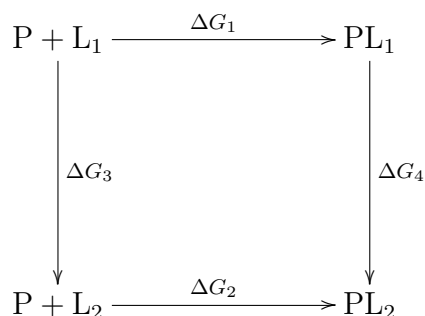
Hierbei bezeichnet  $\text{rand}(0, 1)$  eine Zufallszahl im Intervall  $[0, 1]$ . Jede Trajektorie ergibt dabei ein vollständiges Ensemble bei einer festgelegten Temperatur. Der Vorteil dieses Verfahrens ist, dass im Laufe der Simulation mögliche Minima in der Potentialenergiehyperfläche durch den Übergang zu Konformationen, die bei einer anderen Temperatur generiert wurden, verlassen werden können. Im Falle einer TI Simulation muss nun für jeden der  $n_\lambda$   $\lambda$ -Werte und jede der  $n_T$  Temperaturen eine Kopie des Systems simuliert, insgesamt also  $n_\lambda \times n_T$  Simulationen berechnet werden. Die Methode ist demnach sehr rechenzeitintensiv, das so genannte „Sampling“ für jeden Wert von  $\lambda$  wird aber deutlich verbessert. Alternativ zur Parallel Tempering Thermodynamischen Integration (PTTI) werden in der Replica Exchange Thermodynamic Integration (RETI) Kopien des Systems für verschiedene  $\lambda$ -Werte gleichzeitig simuliert. In festgelegten Zeitabständen wird getestet, ob die Koordinaten für verschiedene Replicas und damit  $\lambda$ -Werte getauscht werden können. Die Bedingung hierfür ist unabhängig vom *NVT* oder *NPT* Ensemble.

$$\exp \left[ \frac{1}{k_B T} (E_{\lambda_j}(j) - E_{\lambda_j}(i) - E_{\lambda_i}(j) + E_{\lambda_i}(i)) \right] \geq \text{rand}(0, 1) \quad (3.31)$$

Diese Methode ist hinsichtlich ihrer Rechenzeit weitaus günstiger als die Parallel Tempering Variante und umgeht die Probleme des Umbrella Samplings. Die Akzeptanz für einen Wechsel des  $\lambda$ -Wertes zwischen zwei Kopien des Systems beruht auf einer ausreichend großen Überlappung zwischen den Replicas. Damit ist die Konvergenz weitgehend sicher gestellt.

#### 3.3.6. Anwendung am Beispiel des Thermodynamischen Kreisprozesses

Zur konkreten Berechnung der Differenz der Freien Bindungsenthalpie zweier Liganden zu einem Protein greift man auf das Konzept des *Thermodynamischen Kreisprozesses* zurück, das erstmals von Tembe und McCammon 1982 für die Protein-Ligand Bindung genutzt



**Abb. 3.2.:** Darstellung des thermodynamischen Kreisprozesses zur Berechnung relativer freier Bindungsenthalpien. Mit „P“ wird das Protein, mit „L“ der Ligand bezeichnet. Die Indizes „1“ und „2“ bezeichnen die beiden unterschiedlichen Ligandmoleküle.

wurde<sup>157</sup>. Die direkte Berechnung der Freien Bindungsenergie zwischen unkomplexiertem Protein und Ligand ist nicht möglich, da bei obigen Ausführungen vorausgesetzt wurde, dass sich die Systeme nur gering voneinander unterscheiden. Eine schematische Darstellung des Kreisprozesses findet sich in Abb. 3.2. Zum einen wird die Transformation von Ligand 1 in Ligand 2 in wässriger Lösung und zum zweiten die äquivalente Transformation in der Bindetasche simuliert. Dabei erhält man die Freien Enthalpiendifferenzen  $\Delta G_3$  und  $\Delta G_4$ . In Abb. 3.2 werden reale Prozesse, also die Bindungsbildung zwischen Ligand und Protein, horizontal dargestellt. Die vertikalen Transformationen, bei denen Ligand 1 zu Ligand 2 wird, spiegeln hingegen keine chemische Reaktion wider. Man spricht deshalb auch häufig von einer *alchemischen Transformation*. Da  $G$  eine Zustandsfunktion ist, darf sich ihr Wert nicht ändern, wenn Ausgangs- und Endzustand identisch sind. Bewegt man sich also einmal entlang des Kreisprozesses, so muss für die freien Energien der Einzelschritte gelten:

$$0 = \Delta G_1 + \Delta G_4 - \Delta G_2 - \Delta G_3. \quad (3.32)$$

Umformen ergibt:

$$\Delta \Delta G = \Delta G_2 - \Delta G_1 = \Delta G_4 - \Delta G_3. \quad (3.33)$$

Die eigentlich interessante Größe  $\Delta G_2 - \Delta G_1$ , also der Unterschied in den Freien Bindungsenthalpien der beiden Liganden, wird rechnerisch durch  $\Delta G_4 - \Delta G_3$ , die Energien für die Mutation von Ligand 1 in Ligand 2, bestimmt.

Auf eine äquivalente Weise kann aber auch näherungsweise die absolute freie Bindungsenthalpie bestimmt werden. Hierfür muss der Ligand in ein Molekül frei von Wechselwirkungen mit umgebenden Atomen transformiert werden. Es wird also gewissermaßen von der Umgebung entkoppelt. Man spricht in diesem Zusammenhang, da diese Entkopplung zweimal (in wässriger Lösung und in der Bindetasche) simuliert werden muss, von einer *Double Decoupling* Methode<sup>158,159</sup>. Selbstverständlich sind hierfür weitaus längere Simulationszeiten nötig, um konvergente Ergebnisse zu erhalten. Außerdem ist es nur bei relativ kleinen Liganden möglich, eine derart große Störung (gewissermaßen das Verschwinden eines ganzen Moleküls) zu simulieren.

### 3.3.7. Besondere Aspekte beim Durchführen der Transformationen

Bei der Verwendung des thermodynamischen Kreisprozesses hat es sich als günstig herausgestellt, die TI Simulationen zur Berechnung der einzelnen  $\Delta G$  Werte in weitere Schritte zu unterteilen. Dabei wird die Transformation in drei Schritte zerlegt. Im ersten Transformationsschritt werden alle Ladungen von Atomen, für die auch Lennard-Jones-Parameter geändert werden, zum Wert 0 mutiert. Im zweiten Schritt werden dann die jeweiligen Lennard-Jones-Parameter der Atome verändert und im dritten Schritt Ladungen für die neuen Atome hinzugefügt. Dieses Vorgehen ist notwendig, um zu vermeiden, dass zwei Atome mit entgegengesetzter Ladung die gleiche Position einnehmen<sup>160</sup>. Sobald die Lennard-Jones Wechselwirkungen eines Atomes verringert werden, ist es ab einem gewissen Wert von  $\lambda$  möglich, dass zwei Atome überlappen. Haben diese zwei Atome Ladungen mit unterschiedlichem Vorzeichen und damit eine attraktive Coulomb Wechselwirkung und damit wiederum ein Energieminimum bei  $r_{AB} = 0$ , so werden die Atome aufeinander liegen. Die Simulation für die Entfernung und das Hinzufügen von Partialladungen konvergiert für gewöhnlich relativ schnell. Auch deshalb ist es sinnvoll, diese Berechnung von der Änderung des Lennard-Jones Beitrages zu trennen.

Neben diesem allgemeinen Vorgehen muss im zweiten Schritt ein weiterer Punkt Beachtung finden. Für den Fall, dass das Lennard-Jones-Potential eines Atoms ausgeschaltet wird, dass also ganze Atome in der Transformation gelöscht werden, gibt es zwei unerwünschte Konsequenzen<sup>160</sup>. Da

$$V(\lambda) = (1 - \lambda)V_0 + \lambda V_1, \quad (3.34)$$

skaliert die Lennard-Jones Wechselwirkung für die Atome, die gelöscht werden, bei kleinen Abständen  $r$  mit

$$\propto (1 - \lambda)/r^{12}. \quad (3.35)$$

Das führt dazu, dass eine Unstetigkeit der Funktion bei  $\lambda$  nahe dem Wert 1 im Übergang zum Wert 1 auftritt. Dadurch ist die Funktion an dieser Stelle nicht mehr mit numerischen Methoden zu integrieren. Für Werte  $\lambda = 1 - \varepsilon$  mit  $\varepsilon$  nahe 0 ist die Potentialfunktion von großer Bedeutung, während sie bei  $\lambda = 1$  den Wert 0 hat.

Um dieses Problem teilweise zu umgehen, kann die Lennard-Jones Energiefunktion mit einem Term der Form

$$\propto (1 - \lambda)^k/r^{12} \quad (3.36)$$

mit  $k \geq 4$  beschrieben werden. Dadurch ist es möglich, die entstehende Singularität numerisch zu integrieren. Dennoch bleibt aber das Problem, dass durch die Singularität in der Potentialfunktion Probleme bei der Simulation selbst, z.B. durch enorm große Kräfte, auftreten.

Eine weitaus bessere Lösung bieten die sogenannten *Soft-Core Potentiale*. Hierbei wird die funktionelle Form des Lennard-Jones-Potentials dahingehend modifiziert, dass der  $1/r^{12}$ -Term langsam als Funktion von  $\lambda$  abgeschaltet wird. Im Programmpaket AMBER sind die

Soft-Core-Potentiale in der folgenden funktionellen Form implementiert<sup>121,161</sup>.

$$V_{0,\text{verschwinden}} = 4\varepsilon(1 - \lambda) \left[ \frac{1}{\left[\alpha\lambda + \left(\frac{r_{ij}}{\sigma}\right)^6\right]^2} - \frac{1}{\alpha\lambda + \left(\frac{r_{ij}}{\sigma}\right)^6} \right] \quad (3.37)$$

$$V_{0,\text{erscheinen}} = 4\varepsilon\lambda \left[ \frac{1}{\left[\alpha(1 - \lambda) + \left(\frac{r_{ij}}{\sigma}\right)^6\right]^2} - \frac{1}{\alpha(1 - \lambda) + \left(\frac{r_{ij}}{\sigma}\right)^6} \right] \quad (3.38)$$

Hierbei sind  $\varepsilon$  und  $\sigma$  die Parameter des gewöhnlichen Lennard-Jones Potentials und  $\alpha$  eine veränderliche Konstante, die in der Implementierung in AMBER den Wert 0.5 hat.

### 3.4. Thermodynamische Berechnungen unter Verwendung der Theorie der inhomogenen Lösungen

Ein entscheidender Nachteil der zuvor vorgestellten Freien Energie Störungsrechnung und der Thermodynamischen Integration ist, dass sie einen sehr hohen Rechenaufwand erfordern. So wird der Hauptteil der Rechenzeit für die Simulation von chemisch unrelevanten Zuständen genutzt, die den Übergang von Ausgangs- in Endkonfiguration beschreiben. Effektiver ist es, eine Trajektorie für die Berechnung von Bindungsenergien zu verwenden, die für die Simulation der Dynamik eines Protein-Ligand Komplexes berechnet wurde und somit auch selbst von Interesse ist. Verbreitete Methoden wie MM-PBSA<sup>69-72</sup> oder LIE<sup>67,68,162</sup>, die zur Auswertung vorhandener Simulationsdaten geeignet sind, weisen jedoch Defizite bei der Genauigkeit auf<sup>163</sup> oder benötigen experimentelle Daten zur Kalibrierung (LIE).

Im Folgenden soll der von Lazaridis und Li<sup>81-83</sup> entwickelte Ansatz, die so genannte „inhomogeneous fluid solvation theory“ (IFST) vorgestellt werden, die 2003 auf Protein-Ligand Komplexe übertragen werden konnte und auf Arbeiten von Lazaridis, Paulaitis und Karplus<sup>79,80,164-166,166</sup> Ende der 90er Jahre sowie der Arbeit von Morita und Hiroike von 1961<sup>167</sup> zurückgeht. Dieses Vorgehen kann bereits vorhandene Trajektorien nutzen und erfordert keine zusätzliche Simulation. Ermittelt werden hierbei - ausgehend von Mehrteilchen-Korrelationsfunktionen - nur die thermodynamischen Beiträge von *Wassermolekülen* an der gesamten Bindungswechselwirkung. Speziell lässt sich der entropische Beitrag bestimmen, der aus einer klassischen freien Energierechnung nur indirekt durch Bestimmung von  $\Delta G$  für unterschiedliche Temperaturen  $T$  und anschließende Differenzierung ( $S = -\left(\frac{\partial G}{\partial T}\right)_{N,p}$ ) erhalten werden kann<sup>168</sup>. Durch eine Normalmoden- oder quasiharmonische Analyse<sup>101</sup> ist es nämlich nicht möglich, entropische Beiträge einzelner Wassermoleküle zu erhalten.

### 3.4.1. Korrelationsfunktionen

Radiale Dichteverteilungsfunktionen (RDF)<sup>169,170</sup> sind in der IFST und allgemein in der theoretischen Behandlung von Flüssigkeiten von großer Bedeutung. Sie beschreiben den Verlauf der Teilchenzahldichte als Funktion des Abstandes zwischen Paaren von Sauerstoffatomen in Wassermolekülen relativ zur Gesamtdichte. Diese Definition kann auf mehrtomige Teilchen erweitert werden. RDFs für Moleküle können auf die gleiche Weise wie für Atome gewonnen werden, wenn Positions- und Orientierungskordinaten voneinander separiert werden oder die Atome des Moleküls einzeln zur Ermittlung der RDF herangezogen werden. Ähnlich ist nun auch das Vorgehen im Rahmen der Solvatationstheorie inhomogener Lösungen. Zur Berechnung thermodynamischer Größen (wie der Entropie) von Wassermolekülen in Protein-Strukturen müssen Mehrteilchen-Korrelationsfunktionen der beteiligten Wassermoleküle ermittelt werden. Das geschieht jeweils getrennt für Translations- und Orientierungskordinaten. Die Position eines Wassermoleküls wird dabei über Kugelkoordinaten, die Orientierung hingegen über Euler-Winkel ausgedrückt.

#### Euler-Winkel

Es existieren mehrere leicht unterschiedliche Definitionen der Euler Winkel, ein Beispiel dafür findet sich z.B. bei Bronstein<sup>171</sup>. Eulerwinkel beschreiben vollständig die Orientierung eines neuen Koordinatensystems relativ zu einem alten. In der Arbeit von Lazaridis<sup>82</sup> wird das Vorgehen zur Berechnung dieser Winkel für unterschiedlich orientierte Wassermoleküle erläutert und soll hier kurz skizziert werden. Demnach lässt sich eine Rotation durch die drei Winkel  $\theta$ ,  $\psi$ ,  $\phi$  durch eine Matrix  $A$

$$\mathbf{A} = \mathbf{BCD} = \begin{pmatrix} \cos \phi & \sin \phi & 0 \\ -\sin \phi & \cos \phi & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \theta & \sin \theta \\ 0 & -\sin \theta & \cos \theta \end{pmatrix} \begin{pmatrix} \cos \psi & \sin \psi & 0 \\ -\sin \psi & \cos \psi & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (3.39)$$

angeben. Diese beschreibt die Transformation des Ausgangskordinatensystems  $X$  in das entsprechende nach der Rotation des Moleküls  $X'$ . Es gilt:

$$X = \mathbf{A}^{-1} X'. \quad (3.40)$$

Zur Ermittlung des Ausgangskordinatensystems wird die Mittelwertsposition des jeweiligen Wassermoleküles in einer Raumregion herangezogen. Für die transformierten Koordinaten wird zunächst das Wassermolekül einer Struktur der Trajektorie auf seine Mittelwertsposition translatiert. Ausgehend von der obigen Matrix erhält man sechs Gleichungen. Durch Lösen dieses Gleichungssystems bestimmt man die Euler-Winkel  $\theta$ ,  $\psi$  und  $\phi$ . Dabei kann das Koordinatensystem an einem Wassermolekül, z.B. wie folgt definiert werden:  $y$ -Achse ist die Winkelhalbierende zwischen den Atomen HOH. Die  $z$ -Achse liegt senkrecht zur  $y$ -Achse in der Ebene, die durch die drei Atome des Wassermoleküls aufgespannt wird. Die  $x$ -Achse schließlich steht senkrecht auf dieser Ebene. Diese Definition kann selbstverständlich variieren, was wiederum die Verteilung der Winkel beeinflusst. Für



eine später durchzuführende Integration ist es zweckmäßig, das Referenzkoordinatensystem so zu wählen, dass die Verteilung der Winkel stetig und möglichst gleichmäßig ist.

### Definition der Mehrteilchen-Korrelationsfunktionen

Ausgehend von der Trajektorie lassen sich die benötigten Korrelationsfunktionen (per Definition der Quotient einer Häufigkeit  $n$  und der erwarteten Häufigkeit  $n_{\text{id}}$  einer idealen Lösung<sup>169</sup>) durch die Berechnung einer einfachen Verteilungsfunktion (Histogramm) und anschließender Normalisierung für jede der Koordinaten bestimmen. Für  $r$  (Kugelkoordinate) ergibt sich für die Korrelationsfunktion  $g(r)$ :

$$g(r) = \frac{n_{\text{his}}/N}{4\pi r^2 dr}. \quad (3.41)$$

Hierbei bezeichnet  $n_{\text{his}}$  die absolute Anzahl der Wassermoleküle im gegebenen Abstandsintervall  $[r - dr, r + dr]$ .  $N$  ist der absolute Besetzungsgrad der betrachteten Bindetaschenregion mit Wassermolekülen. Für die weiteren Kugelkoordinaten erhält man ähnliche Funktionen. Es gilt:\*

$$g(\theta') = \frac{n_{\text{his}}/N \cdot 2}{\sin(\theta') \sum (n_{\text{his}}/N d\theta')} \quad (3.42)$$

und

$$g(\phi') = \frac{n_{\text{his}}/N \cdot 2\pi}{\sum (n_{\text{his}}/N d\phi')} \quad (3.43)$$

Die Berechnung der Orientierungs-Korrelationsfunktionen verläuft analog:

$$g(\theta) = \frac{n_{\text{his}}/N \cdot 2}{\sin(\theta) \cdot \sum (n_{\text{his}}/N d\theta)} \quad (3.44)$$

und

$$g(\psi) = \frac{n_{\text{his}}/N \cdot 2\pi}{\sum (n_{\text{his}}/N d\psi)} \quad (3.45)$$

und

$$g(\phi) = \frac{n_{\text{his}}/N \cdot 2\pi}{\sum (n_{\text{his}}/N d\phi)} \quad (3.46)$$

Näherungsweise lässt sich die Translationskorrelationsfunktion  $g_{\text{SW}}^{\text{tr}}(\mathbf{r}) = g_{\text{SW}}^{\text{tr}}(r, \theta', \phi')$  als Produkt aus drei einzelnen Funktionen ausdrücken.

$$g_{\text{SW}}^{\text{tr}}(r, \theta', \phi') = g_{\text{SW}}^{\text{tr}}(r) g_{\text{SW}}^{\text{tr}}(\theta') g_{\text{SW}}^{\text{tr}}(\phi') \quad (3.47)$$

Dabei ist jeweils  $g_{\text{SW}}^{\text{tr}}(r)$ ,  $g_{\text{SW}}^{\text{tr}}(\theta')$  und  $g_{\text{SW}}^{\text{tr}}(\phi')$  die über die beiden anderen Koordinaten gemittelte Gesamtfunktion. Diese Näherung wird eingeführt, um die Berechnung der Integrale

---

\*Im Folgenden sind die Winkel der Kugelkoordinaten zur besseren Unterscheidbarkeit von den Euler-Winkeln mit einem „'“ markiert.

über  $g_{\text{SW}}^{\text{tr}}(\mathbf{r})$  zu vereinfachen. In<sup>81</sup> wurde beschrieben, dass diese Annahme gerechtfertigt ist und zu den gleichen Ergebnissen wie die Integration der dreidimensionalen Korrelationsfunktion führt.

Die Orientierung  $\omega$  des Wassermoleküls nach unterschiedlichen Simulationszeiten lässt sich mit Hilfe der oben eingeführten Euler-Winkel  $\theta, \phi$  und  $\psi$  beschreiben. Somit gilt nach Faktorisierung:

$$g_{\text{SW}}^{\text{or}}(\omega) = g_{\text{SW}}^{\text{or}}(\theta, \phi, \psi) = g_{\text{SW}}^{\text{or}}(\theta)g_{\text{SW}}^{\text{or}}(\phi)g_{\text{SW}}^{\text{or}}(\psi). \quad (3.48)$$

Auch diese Näherung ist nach<sup>81</sup> gerechtfertigt. Damit ist außerdem (für das Volumenelement vgl.<sup>164</sup>)

$$\Omega = \int \sin \theta d\theta d\psi d\phi = \int_0^\pi \sin \theta d\theta \int_0^{2\pi} d\psi \int_0^{2\pi} d\phi = 8\pi^2. \quad (3.49)$$

### 3.4.2. Berechnung von Entropiewerten aus Simulationsdaten

Mit Hilfe der Paar-Korrelationsfunktion  $g^{(2)\dagger}$  kann der Ensemble-Mittelwert jeder Funktion  $a$ , die von einem Atompaar mit den Koordinaten  $\mathbf{r}_i$  und  $\mathbf{r}_j$  abhängt, als<sup>169</sup>:

$$\langle a(\mathbf{r}_i, \mathbf{r}_j) \rangle = \frac{1}{V^2} \int g(\mathbf{r}_i, \mathbf{r}_j) a(\mathbf{r}_i, \mathbf{r}_j) d\mathbf{r}_i d\mathbf{r}_j \quad (3.50)$$

geschrieben werden. Dabei steht  $V$  für das Volumen des Systems. Im Falle der molaren Energie einer einatomigen Flüssigkeit heiße das<sup>79</sup>:

$$e = \frac{3}{2}kT - \frac{1}{2}\rho \int g(\mathbf{r})^{(2)}u(\mathbf{r})d\mathbf{r} \quad (3.51)$$

Dabei bezeichnet  $u$  das Wechselwirkungspotential zwischen den Atompaaren,  $k$  die Boltzmann-Konstante,  $T$  die Temperatur und  $\mathbf{r}$  den relativen Abstand. Auch weitere thermodynamische Größen wie der Druck  $p$  oder das chemische Potential  $\mu$  lassen sich auf diese Weise berechnen.

Für die molare Entropie  $s$  einer reinen, einatomigen Flüssigkeit gilt<sup>79</sup>:

$$s = s^{\text{id}} + s^{(2)} + s^{(3)} + \dots \quad (3.52)$$

Hierbei wird mit  $s^{\text{id}}$  die Entropie des idealen Gases, das die gleiche Temperatur und Dichte wie die betrachtete Flüssigkeit hat, bezeichnet. Die Terme  $s^{(i)}$  gehören zu dem Anteil der Entropie, der aus der Flüssigkeit stammt. Für  $s^{(2)}$  kann man wiederum

$$s^{(2)} = -\frac{1}{2}k\rho \int [g^{(2)} \ln g^{(2)} - g^{(2)} + 1] d\mathbf{r} \quad (3.53)$$

---

<sup>†</sup>An dieser Stelle wird die Paar-Korrelationsfunktion mit dem Superskript (2) versehen, um sie deutlich von Triplet-Korrelationsfunktion oder höheren Termen zu unterscheiden.

schreiben. Üblicherweise werden die Drei-Teilchen-Korrelation und alle höheren Terme vernachlässigt. Diese Näherung ist nur gerechtfertigt, wenn diese Korrelationen keinen entscheidenden Beitrag liefern, was sich für einfache Systeme gezeigt hat<sup>79,80</sup>.

Dieses Vorgehen lässt sich auch auf eine Lösung (z.B. eines Proteins in Wasser) übertragen und die entsprechenden Ausdrücke für Energie und Entropie können abgeleitet werden. Das Verfahren hierbei soll im Folgenden nur skizziert werden. Eine ausführlichere Darstellung findet man bei<sup>79,80,83</sup>. Energie und Entropie werden hierbei in Terme für Beiträge aus Wechselwirkungen zwischen Solvens und gelöstem Stoff ( $E_{SW}$ ,  $S_{SW}$ ) und zwischen Solvens und Solvens ( $\Delta E_{WW}$ ,  $\Delta S_{WW}$ ) separiert.<sup>‡</sup> Hiermit wird der Transfer eines Wassermoleküls aus dem Solvens in die Bindetasche beschrieben. Die Größen können durch die Korrelationsfunktionen  $g_{SW}(\mathbf{r}, \omega)$  sowie  $g_{WW}(\mathbf{r}, \mathbf{r}', \omega, \omega')$  ausgedrückt werden und sind deshalb gut durch Computersimulationen zu ermitteln. Mit  $\mathbf{r}$  werden die Positionen des Sauerstoffatoms eines geordneten Wassermoleküls, das untersucht wird, während der Simulation relativ zu seiner Mittelwertsposition bezeichnet. Mit  $\omega$  wird die Orientierung des Wassermoleküls während der Simulation beschrieben. Die mit ' gekennzeichneten Orts- und Orientierungsvariablen markieren das jeweils zweite Wassermolekül, da in den ‚WW‘ Termen die Beiträge verschiedener Wassermoleküle untereinander enthalten sind. Man spricht in diesem Zusammenhang von einem ‚inhomogenen‘ Ansatz, da das System der Simulation des Protein-Komplexes keine größeren Mengen Solvens enthält und somit die Korrelationsfunktionen keine Mittelung über die Gesamtheit der Flüssigkeit darstellen, also in den verschiedenen Raumrichtungen unterschiedlich verlaufen.

Die Werte für  $E_{SW}$  und  $E_{WW}$  lassen sich zwar mit Hilfe von Paar-Korrelationsfunktionen ermitteln, viel einfacher gelingt das laut Lazaridis jedoch direkt aus der Simulation. Dafür werden die während der Simulation durch das Kraftfeld ermittelten Energiebeiträge des Wassermoleküls aufsummiert und gemittelt. Von besonderem Interesse ist die Differenz der Werte im Komplex und in der reinen Lösung. Wegen der Ununterscheidbarkeit der einzelnen Wassermoleküle erhält man<sup>83</sup> (für eine Region  $i$ ):

$$\Delta E_{\text{solv}} = E_{SW(i)} + \frac{1}{2}E_{W(i)W} - (-10.1 \text{ kcal/mol}). \quad (3.54)$$

Der Wert von -10.1 kcal/mol stellt den Betrag in reiner wässriger Phase dar. Im Rahmen dieser Arbeit wurde dieser für das verwendete Kraftfeld und Wassermodell erneut berechnet und ein Wert von -9.5 kcal/mol erhalten (vgl. Kap 7.2).

Für  $S$  kann man, wenn nur die Zwei-Teilchen-Beiträge zur Korrelation berücksichtigt werden, schreiben:

$$S_{SW} = -k \frac{\rho}{\Omega} \int g_{SW}(\mathbf{r}, \omega) \ln [g_{SW}(\mathbf{r}, \omega)] d\mathbf{r} d\omega. \quad (3.55)$$

<sup>‡</sup>Die Schreibweise orientiert sich in dieser Arbeit so weit wie möglich an der Schreibweise in den Publikationen von Lazaridis auch wenn man statt  $E$  korrekterweise von  $H$  sprechen müsste.

$$\begin{aligned} \Delta S_{\text{WW}} = & -\frac{1}{2}k\frac{\rho^2}{\Omega^2} \int g_{\text{SW}}(\mathbf{r}, \omega) \left[ g_{\text{SW}}(\mathbf{r}', \omega') \left( g_{\text{WW}}^{\text{inh}}(\mathbf{r}, \mathbf{r}', \omega, \omega') \ln g_{\text{WW}}^{\text{inh}}(\mathbf{r}, \mathbf{r}', \omega, \omega') \right. \right. \\ & \left. \left. - g_{\text{WW}}^{\text{inh}}(\mathbf{r}, \mathbf{r}', \omega, \omega') + 1 \right) - \left( g_{\text{WW}}^{\circ}(\mathbf{r}, \mathbf{r}', \omega, \omega') \ln g_{\text{WW}}^{\circ}(\mathbf{r}, \mathbf{r}', \omega, \omega') - g_{\text{WW}}^{\circ}(\mathbf{r}, \mathbf{r}', \omega, \omega') + 1 \right) \right] d\mathbf{r}d\mathbf{r}'d\omega d\omega' \end{aligned} \quad (3.56)$$

Hierbei bezeichnet  $\omega^{\text{rel}}$  die fünf Winkel, die die relative Orientierung der Wassermoleküle untereinander ausdrücken (mit  $\Omega^{\text{rel}} = \int d\omega^{\text{rel}} = 32\pi^3$ ), und  $R$  den Abstand der beiden Moleküle. Außerdem steht das Superskript ‚inh‘ für die Beiträge aus dem Komplex und ‚ $\circ$ ‘ aus der reinen Flüssigkeit. Das  $\Delta$  weist somit auf die Differenz zwischen Beiträgen aus der Bindetasche und der reinen Lösung hin.

Für den Fall, dass isolierte Wassermoleküle in einem biomolekularen Komplex betrachtet werden, sind die Beiträge zur Solvens–Solvens Entropie und Energie vernachlässigbar und  $\Delta S_{\text{WW}}$  entspricht dem Beitrag beim Entfernen eines Wassermoleküls aus der Flüssigkeit. Von Null verschiedene Werte für  $g_{\text{SW}}$  sind nur für die Raumregionen zu erwarten, die durch einen gelösten Stoff besetzt sind. Da man das Integral über den gesamten Raum  $V$  in Teilintegrale über einzelne Unterregionen aufteilen kann, ist es möglich, Beiträge einzelner Wassermoleküle zu bestimmen. Teilt man die Korrelationsfunktion zusätzlich in zwei Terme für Translation und Orientierung erhält man:

$$g_{\text{SW}}(\mathbf{r}, \omega) = g_{\text{SW}}^{\text{tr}}(\mathbf{r})g_{\text{SW}}^{\text{or}}(\omega|\mathbf{r}). \quad (3.57)$$

$$g_{\text{WW}}(\mathbf{r}, \mathbf{r}', \omega, \omega') = g_{\text{WW}}^{\text{tr}}(\mathbf{r}, \mathbf{r}')g_{\text{WW}}^{\text{or}}(\omega, \omega'|\mathbf{r}, \mathbf{r}'). \quad (3.58)$$

Die Schreibweise  $\omega|\mathbf{r}'$  deutet dabei an, dass die Orientierung des Wassermoleküls prinzipiell nicht unabhängig von der Position ist. Nimmt man aber an, dass in einem begrenzten Bereich des Raums  $v_i$   $g^{\text{or}}$  unabhängig von  $\mathbf{r}$  ist, also:

$$g_{\text{SW}}^{\text{or}}(\omega|\mathbf{r}) \approx g_{\text{SW}}^{\text{or}}(\omega), \quad (3.59)$$

kann man schreiben:

$$S_{\text{SW}(i)} = -k\rho \int_{v_i} g_{\text{SW}}^{\text{tr}}(\mathbf{r}) \ln [g_{\text{SW}}^{\text{tr}}(\mathbf{r})] d\mathbf{r} - k\frac{N_{\text{wat}(i)}}{\Omega} \int_{v_i} g_{\text{SW}}^{\text{or}}(\omega) \ln [g_{\text{SW}}^{\text{or}}(\omega)] d\omega. \quad (3.60)$$

Das Subskript am Integral bezeichnet die Region des Raums in der die oben getroffene Näherung gilt. Aus Gründen der Übersichtlichkeit kann die Abkürzung  $N_{\text{wat}(i)}$  eingeführt werden. Sie bezeichnet die Anzahl der Wassermoleküle in der Region des Raums um Wassermolekül  $i$ :

$$N_{\text{wat}(i)} = \rho \int_{v_i} g_{\text{SW}}^{\text{tr}}(\mathbf{r}) d\mathbf{r}. \quad (3.61)$$

Die Entropiedifferenz für den Übergang eines Wassermoleküls aus der wässrigen Phase in den gebundenen Zustand berechnet man am leichtesten als Differenz zwischen den Ter-

men für ein gebundenes Wassermolekül und denen, die für die Lösung vorberechnet wurden. Damit gilt bei Vernachlässigung der reinen Flüssigkeitsterme und mit Aufteilung in Translations- und Orientierungsanteil:

$$S_{W(i)W(j)}^{\text{trans}} = -\frac{1}{2}k\rho^2 \int_{v_i v_j} g_{SW(i)}^r(\mathbf{r}) g_{SW(j)}^r(\mathbf{r}') \left[ g_{W(i)W(j)}^{r,\text{inh}}(\mathbf{r}, \mathbf{r}') \ln g_{W(i)W(j)}^{r,\text{inh}}(\mathbf{r}, \mathbf{r}') - g_{W(i)W(j)}^{r,\text{inh}}(\mathbf{r}, \mathbf{r}') + 1 \right] d\mathbf{r} d\mathbf{r}' \quad (3.62)$$

$$S_{W(i)W(j)}^{\text{or}} = -\frac{1}{2}k\rho^2 \int_{v_i v_j} g_{SW(i)}^r(\mathbf{r}) g_{SW(j)}^r(\mathbf{r}') g_{W(i)W(j)}^{r,\text{inh}}(\mathbf{r}, \mathbf{r}') S_{W(i)W(j)}^m(\mathbf{r}, \mathbf{r}') d\mathbf{r} d\mathbf{r}' \quad (3.63)$$

mit (bei Annahme der Unabhängigkeit der Orientierungskordinaten von der Translation)

$$S_{WW}^m(\mathbf{r}, \mathbf{r}') = \frac{1}{\Omega^2} \int g_{WW}^{\text{or}}(\omega) g_{WW}^{\text{or}}(\omega') g_{WW}^{\text{or},\text{inh}}(\omega, \omega' | \mathbf{r}, \mathbf{r}') \ln g_{WW}^{\text{or},\text{inh}}(\omega, \omega' | \mathbf{r}, \mathbf{r}') d\omega d\omega'. \quad (3.64)$$

Während die Translations- und Orientierungs-Korrelationsfunktionen zwischen Wasser und gelöstem Stoff als Produkt der eindimensionalen Funktionen gewonnen werden, ist die sechsdimensionale Funktion  $g_{WW}^{\text{or},\text{inh}}(\omega, \omega' | \mathbf{r}, \mathbf{r}')$  schwer zu ermitteln. Gemäß der „Kirkwood superposition approximation“ (KSA)<sup>130</sup> ist diese inhomogene Paar-Korrelationsfunktion gleich der der reinen Flüssigkeit und hängt nur vom Abstand und Orientierung der Solvens-Moleküle ab.

$$g_{WW}^{\text{or},\text{inh}}(\omega, \omega' | \mathbf{r}, \mathbf{r}') = g_{WW}^{\text{or},\text{inh}}(\omega^{\text{rel}}, R). \quad (3.65)$$

Diese Näherung kann separat auf den Translations- und Orientierungsteil angewandt werden und somit ergibt sich

$$S_{WW}^m(\mathbf{r}, \mathbf{r}') = \frac{1}{\Omega^2} \int g_{WW}^{\text{or}}(\omega) g_{WW}^{\text{or}}(\omega') g_{WW}^{\text{or},\text{inh}}(\omega^{\text{rel}}, R) \ln g_{WW}^{\text{or},\text{inh}}(\omega^{\text{rel}}, R) d\omega d\omega'. \quad (3.66)$$

Man wählt nun für  $R$  den Abstand der beiden Mittelwertspositionen der untersuchten Wassermoleküle und berechnet für jeden Abstand und jede Orientierung die relative Orientierung der Wassermoleküle zueinander. Zur Bestimmung von  $g_{WW}^{\text{or},\text{inh}}(\omega^{\text{rel}}, R)$  wird auch diese Funktion in ein Produkt aus eindimensionalen Funktionen faktorisiert.

In den letzten Schritten wurden nur die Funktionen für die Komplexe betrachtet. Bildet man nun die Differenz zum vorberechneten Term der reinen Lösung, erhält man

$$\Delta S_{W(i)W} = \sum_j S_{W(i)W(j)}^{\text{trans}} + \sum_j S_{W(i)W(j)}^{\text{or}} - (-15.2 \text{ cal}/(\text{mol} \cdot \text{K})). \quad (3.67)$$

Die Entropie von  $-15.2 \text{ cal}/(\text{mol} \cdot \text{K})$ , um ein Wassermolekül aus reiner Phase zu entfernen, wurde hierbei von früheren Arbeiten von Lazaridis zusammen mit Karplus übernommen (vgl. hierzu<sup>82</sup> und<sup>172</sup>). Die Solvatationsentropie wiederum ist die Summe aus Entropieanteil

von gelöstem Stoff mit untersuchtem Wassermolekül und der Reorganisationsentropie aller gebundenen Wassermoleküle.

$$\Delta S_{\text{solv}} = \sum_i (\Delta S_{\text{W}(i)\text{W}} + S_{\text{SW}(i)}) \quad (3.68)$$

#### 3.4.3. WaterMap

Eine Implementierung des hier dargestellten Ansatzes um die Gruppe von Lazaridis findet sich im Programm WaterMap, das von den Gruppen um R. Friesner und B. Berne an der Columbia Universität entwickelt und von der Firma Schrödinger vertrieben wird<sup>75,78</sup>. WaterMap unterscheidet sich bezüglich des gewählten Referenzzustandes und der Implementierung jedoch geringfügig von dem bei Lazaridis beschriebenen Vorgehen<sup>173,174</sup>. Es wurde der solvationstheoretische Ansatz implementiert und zusätzlich ein Algorithmus hinzugefügt, der es ermöglicht, Wassermoleküle zu clustern. Speziell bei offenen Bindetaschen ist das von großer Bedeutung, da hier während der Simulation häufige Austauschvorgänge stattfinden. Aufgrund der Fülle der zu betrachtenden Wassermoleküle und deren fluktuierenden Positionen ist es notwendig, allen relevanten Wassermolekülen Raumregionen zuzuordnen, die anschließend einzeln analysiert werden. Aus den bisherigen Ableitungen wird ersichtlich, dass die Integration zum Ermitteln des Entropiebetrages voraussetzt, dass die Orientierung der Wassermoleküle unabhängig von der Position ist. Dies ist aber nur in kleinen Teilbereichen einer Tasche gewährleistet.

Der Cluster-Algorithmus in WaterMap findet zunächst die Wasserposition mit der größten Anzahl von benachbarten Wasserpositionen innerhalb eines Radius von 1 Å, die man bei Überlagerung aller Strukturen der Trajektorie vorfindet. Diese Wasserposition und die innerhalb der 1-Å-Kugel werden entfernt und die Suche nach einer neuen „principle hydration site“ wiederholt. Damit können dann Teilvolumina festgelegt werden, für die die gewünschten Energie- und Entropiewerte berechnet werden können. Die Beschreibung der Orientierung der Wassermoleküle einer Hydratisierungsregion erfolgt mittels Quaternionen und nicht mehr mit den nicht eindeutigen Euler-Winkeln. Auf dieser Basis wird auch das Clustering vorgenommen. Für jedes Wassermolekül wird ein Haupt-Quaternion berechnet, das dieses auf die Referenz-Orientierung dreht. Sodann wird der Abstand eines jeden Wassermoleküls im Winkel-Raum bestimmt. Dieser liegt in den Grenzen zwischen 0 und 1, wobei 0 eine identische Orientierung bedeutet. Analog zum Vorgehen beim Auffinden der Hydratisierungs-Regionen werden die zwei größten Orientierungscluster ermittelt. Anschließend werden alle Wassermoleküle einem der beiden Teilbereiche zugeordnet und dabei an Hand der Ebene orthogonal zum Verbindungsvektor der Quaternionen separiert.

Dieses Vorgehen hat auf viele Hydratisierungsregionen nur einen sehr geringen Effekt. Generell ist es aber dazu geeignet, die Präzision der Berechnung zu erhöhen, da so sichergestellt werden kann, dass die Orientierung der Wassermoleküle wirklich unabhängig von der Position ist.

Die so berechneten Entropiewerte wurden dann für die Implementierung eines Deskriptors genutzt<sup>76</sup>, der helfen kann, die Änderung der Bindungsaffinität in einer Serie von

### *3.4. Thermodynamische Berechnungen unter Verwendung der Theorie der inhomogenen Lösungen*

---

Protein-Ligand Komplexen vorherzusagen. Die implementierten WaterMap-Algorithmen haben sich bei den in <sup>75-77,84-88</sup> diskutierten Anwendungsbeispielen als sehr hilfreich herausgestellt. Das Programm war jedoch zu Beginn der hier vorgestellten Arbeit noch nicht veröffentlicht und wird derzeit nicht an nicht-kommerzielle Institutionen vertrieben.

### 3. Bestimmung thermodynamischer Größen für Protein-Ligand Komplexe

---



## 4. Verfahren der computerbasierten Wirkstoffentwicklung

Der Begriff des *virtuellen Screenings* umfasst den Prozess der Arzneimittelforschung, in dem durch *in silico* Methoden neue Wirkstoffmoleküle gefunden werden sollen. Dabei wird die Bindungsaffinität dieser potentiellen Inhibitoren nicht wie in einem klassischen Screening durch Experimente gemessen, sondern durch computerbasierte Methoden vorhergesagt. Ausführliche Beschreibungen zu den dafür verwendeten Verfahren sowie zu deren Vorzügen und Schwächen finden sich in der Literatur.<sup>9,175-177</sup> In diesem Kapitel soll nur auf Methoden eingegangen werden, die im Rahmen der vorliegenden Arbeit relevant waren. Außerdem sollen die Besonderheiten von Wassermolekülen im Rahmen eines virtuellen Screenings beleuchtet werden. Hierbei kamen ausschließlich strukturbasierte Verfahren der Wirkstoffsuche zum Einsatz. Methoden z.B. aus dem Bereich der Chemoinformatik wurden für diese Arbeit nicht verwendet.

### 4.1. Analyse der Bindetasche

Bei Vorliegen der 3D-Strukturinformation eines Proteins, z.B. aus einem NMR-Experiment oder durch röntgenkristallographische Bestimmung, ist es mitunter sehr hilfreich, mögliche Interaktionsstellen wie z.B. Wasserstoffbrücken oder hydrophobe Bereiche in der Bindetasche zu identifizieren und die Tasche zu charakterisieren. Für den Fall, dass kein Ligand mit dem Rezeptor co-kristallisiert werden konnte, ist es unter Umständen notwendig, die Bindetaschenregion grundsätzlich zu identifizieren. Verschiedene Computerprogramme bieten hierbei Hilfe.

Die Identifizierung einer möglichen Bindetasche wird z.B. vom Programm SiteMap<sup>36,37</sup> durchgeführt. Dieses bietet desweiteren eine Analysemöglichkeit von hydrophoben Regionen oder beispielsweise von Bereichen mit günstigen Wasserstoffbrückenbindungen. Auch mit Hilfe der Paarpotentiale aus der Scoringfunktion DrugScore<sup>34,35,146</sup> kann eine Charakterisierung der Bindetasche durch Visualisierung von Kontur-Diagrammen vorgenommen werden. Das Programmpaket MOE<sup>178</sup> oder die Programme SuperStar<sup>32</sup> oder MCSS<sup>31</sup> bieten ein ähnliches Werkzeug.

Vorreiter in diesem Bereich war das Programm GRID<sup>30</sup>, das die günstigen Regionen ebenfalls über ein Gitter ermittelt, das in der Bindetasche definiert wird und an jedem Gitterpunkt die Wechselwirkungsenergie für eine Sonde mit Hilfe eines Kraftfeldes berechnet. Dieses Kraftfeld enthält im Unterschied zum Kraftfeld für MD Simulationen keine intramolekularen Terme. Die nichtbindende Wechselwirkungsenergie  $E$  besteht im We-

sentlichen aus drei Teilen:

$$E = \sum E_{\text{LJ}} + \sum E_{\text{el}} + \sum E_{\text{HB}}. \quad (4.1)$$

Hierbei steht  $E_{\text{LJ}}$  für die van-der-Waals-Energie, die mit einer Lennard-Jones Funktion berechnet wird. Mit  $E_{\text{el}}$  werden die elektrostatischen und mit  $E_{\text{HB}}$  die Wasserstoffbrückenbindungsbeiträge zur Gesamtenergie ermittelt. Jeder individuelle Term ist dabei die Summe über alle paarweisen Wechselwirkungen. Das Lennard-Jones-Potential ist eine Funktion des Abstandes  $d$  mit zwei stoffspezifischen Konstanten  $A$  und  $B$ .

$$E_{\text{LJ}} = \frac{A}{d^{12}} - \frac{B}{d^6} \quad (4.2)$$

Für die Berechnung der elektrostatischen Wechselwirkungen wird ein etwas aufwändigerer Ansatz gewählt, als es das einfache Coulomb-Gesetz darstellt. Je nachdem, ob sich ein Sondenatom eher im Inneren des Proteins oder im Lösungsmittel befindet, wird eine andere Dielektrizitätskonstante gewählt. Dazu wird zunächst die Anzahl der benachbarten Proteinatome innerhalb von  $4 \text{ \AA}$  bestimmt. Liegt diese bei 12 oder höher, befindet sich die Sonde auf einem Gitterpunkt im Inneren des Proteins und man verwendet die Dielektrizitätskonstante  $\zeta$ . Liegt diese Anzahl unter 7, wird die Konstante  $\varepsilon$  genutzt. Man erhält eine Beziehung der Form

$$E_{\text{el}} = \frac{pq}{K\zeta} \left[ \frac{1}{d} + \frac{(\zeta - \varepsilon)/(\zeta + \varepsilon)}{\sqrt{d^2 + 4s_p s_q}} \right] \quad (4.3)$$

Hierbei bezeichnen  $p$  und  $q$  die Ladungen von Sonden- und Proteinatom,  $K$  ist eine Kombination aus Naturkonstanten ( $\pi$ ,  $\varepsilon$ , vgl.<sup>113</sup>) und geometrischen Faktoren,  $s_p$  ist die Vergrabenheit eines Protein-Atoms innerhalb des Proteins und  $s_q$  der analoge Wert für eine Sonde.  $4s_p s_q$  wird 0 gesetzt, falls die Anzahl der Kontakte unter vier liegt. Für die beiden Dielektrizitätskonstanten werden Werte von  $\varepsilon = 80$  und  $\zeta = 4$  verwendet.

Die Wasserstoffbrückenbindungen werden mit einer richtungsabhängigen Funktion ermittelt.

$$E_{\text{HB}} = \left[ \frac{C}{d^6} - \frac{D}{d^4} \right] \cos^m \theta \quad (4.4)$$

Dabei bestimmen  $C$  und  $D$  den Gleichgewichtsabstand zwischen zwei Atomen und  $\theta$  den Winkel der Wasserstoffbrückenbindung zwischen den Schweratomen. Der Parameter  $m$  wird genutzt, um eine winkelabhängige Gewichtung der Brückenbindungsstärke vorzunehmen. Weitere Details finden sich bei Goodford und Kollegen.<sup>30</sup>

Wie bereits erwähnt, werden die Wechselwirkungsenergien für verschiedene Sonden auf den Positionen der Gitterpunkte berechnet. Als Sonden kommen dabei verschiedene Atome und funktionelle Gruppen (wie z.B. Wasser, Amine oder Carboxylat) zum Einsatz, die relevant für kleine organische Moleküle sind. Eine Übersicht findet man im Anhang in Tab. B.1. Nach Ermittlung der Wechselwirkungsenergien können diese dann grafisch

zusammen mit der Bindetasche in Form von Kontur-Diagrammen dargestellt werden. Die Darstellung der Kontur-Linien kann dabei leicht für verschiedene Energiegrenzen variiert werden. Der Nachteil der Methode ist jedoch, dass verschiedene Sonden unabhängig voneinander berechnet werden und dadurch keine kooperativen Effekte berücksichtigt werden können.

Eine dieser Sonden im Programm GRID ist ein Wassermolekül. Auf diese Weise lassen sich für Wassermoleküle energetisch vorteilhafte Regionen der Bindetasche detektieren und visualisieren. Eine analoge Analysemöglichkeit der hydrophilen Regionen einer Kavität findet man ebenso in vergleichbaren Programmen wie SiteMap oder der entsprechenden Funktion in MOE. Nach Identifikation der Wasserpositionen kann eine Klassifizierung mit verschiedenen Methoden erfolgen. Einen Überblick hierüber findet man in Abschnitt 1.2.1.

In der vorliegenden Arbeit wurden weiterhin Informationen aus der Datenbank Relibase+<sup>179-181</sup> ausgewertet. Darin bietet das WaterBase Modul<sup>182-184</sup> zahlreiche, vorberechnete Deskriptoren für Wassermoleküle in Kristallstrukturen. Durch die Möglichkeit eigene Strukturdaten zu prozessieren, können diese auch für neue Kristallstrukturen, die (noch) nicht in der PDB abrufbar sind, genauso wie für Strukturen aus MD-Simulationen ermittelt werden. Die einzelnen Deskriptoren wurden von J. Günther<sup>14</sup> eingehend vorgestellt. Hier soll deshalb nur ein kurzer Überblick über diese Größen gegeben werden:

- *Mobilität*: Auf B-Faktoren aller Wassermoleküle und deren Besetzungsgrad normierter B-Faktor;
- *Kontakte*: Anzahl der polaren Kontakte eines Wassermoleküls zu Protein, Ligand oder Solvens innerhalb von 3.3 Å;
- *Polarität*: Maß für die polaren Eigenschaften der Wasserumgebung;
- *DrugScore*: Potentiale der wissensbasierten Scoring-Funktion DrugScore<sup>184</sup> für das jeweilige Wassermolekül;
- *Atomdichte in der lokalen Umgebung*: Zwei Deskriptoren (für erste und zweite Koordinationssphäre) als Maß für die Vergrabenheit eines Wassermoleküls;
- *Solvens-zugängliche-Oberfläche* als Beschreibung der Vergrabenheit eines Wassermoleküls.
- *Koordinationsgeometrie eines Wassermoleküls*: Abweichung (RMSD) von einer idealen tetraedrischen bzw. oktaedrischen Koordinationsgeometrie des Wassermoleküls.

Aufgrund dieser Werte ist somit ebenfalls eine Klassifizierung einzelner Wasserpositionen möglich.

## 4.2. Docking

Eine der am meisten genutzten Methoden im virtuellen Screening ist das Docking kleiner Moleküle. Dabei wird versucht, einen Liganden optimal in die Bindetasche einzupassen

und damit den Bindungsmodus in einer möglichen Kristallstruktur vorherzusagen. Die Bindungsenergie der einzelnen Lösungen dieses Optimierungsproblems wird mit Hilfe einer Scoring-Funktion (vgl. Abschnitt 3.2) abgeschätzt. Eine ausführlichere Darstellung findet sich z.B. in<sup>185</sup>. Da häufig Bindemodi einer großen Zahl von Liganden gefunden werden sollen, müssen der Konformationsraum eines Liganden also möglichst schnell und effizient durchmustert (*Suchproblem*) und die Lösungen gleichzeitig energetisch möglichst exakt und ebenfalls rasch *bewertet* werden. Docking-Programme kann man hierbei im Wesentlichen nach den verwendeten Such-Algorithmen aufteilen:

- *Multikonformationen Docking*: Konformationen werden vor dem Docking generiert. Beispiele sind die Programme FRED<sup>186</sup> oder Glide<sup>60</sup>;
- *Inkrementeller Aufbau*: Das Molekül wird in der Bindetasche aus Fragmenten zusammen gesetzt. Beispiele für Programme sind FlexX oder DOCK<sup>187-190</sup>;
- *Stochastische Suchmethoden*: Die Suche erfolgt z.B. durch genetische Algorithmen (GOLD oder Autodock<sup>191,192</sup>).

Zusätzlich ist im Prinzip ebenfalls die *Flexibilität des Rezeptors* zu berücksichtigen. Das wird aber wegen des erhöhten Rechenaufwands nicht standardmäßig durchgeführt. Es gibt jedoch ein standardisiertes Protokoll für *induced fit* Docking in Glide<sup>193</sup> oder die Möglichkeiten in Programmen wie z.B. FlexX<sup>48</sup> oder GOLD<sup>49-51</sup> unterschiedliche Konformationen von Seitenketten im Voraus zu definieren. Eine Übersicht über dieses Thema findet sich beispielsweise bei<sup>194</sup>.

In Hinblick auf die Berücksichtigung von Wasser beim Docking können die Programme weiter differenziert werden. Die Docking-Simulation kann einige wenige *strukturelle Wassermoleküle* mit einbeziehen oder diese vollständig ignorieren, was für einen Großteil der verfügbaren Programme zutrifft. Dies vernachlässigt aber ganz offensichtlich einen entscheidenden Aspekt, da einzelne Wassermoleküle wichtige gerichtete Wechselwirkungen vermitteln können und die Form der Bindetasche modifizieren. Zudem stammt ein entscheidender Beitrag zur Bindungsaffinität aus der Desolvatation der Tasche. Auch aus diesem Grund sind Rescoring-Methoden wie MM-PBSA oder MM-GBSA, die ein implizites Solvensmodell beinhalten, populärer geworden. Es gibt prinzipiell drei Wege, um Wassermoleküle explizit zu berücksichtigen. Diese können als Teil der Rezeptorstruktur angesehen und deren Konservierung in der Tasche vorhergesagt werden. Alternativ kann man Wassermoleküle gleichzeitig mit den Freiheitsgraden des Liganden behandeln. In manchen Programmen wiederum werden Wassermoleküle nach dem Finden einer potentiellen Ligandpose hinzugefügt, um das Scoring zu verbessern.

Die Berücksichtigung eines Wassermoleküls als Teil des Rezeptors ist durch verschiedene Grid-basierte Methoden wie AutoDock, Glide oder auch mit GOLD und FRED möglich. Das Particle Concept in FlexX hingegen platziert Wassermoleküle während der Konstruktion des Liganden. Glide schließlich dockt Wassermoleküle nach Abschluss des Ligand-Dockings in die Bindetasche, um die Abschätzung für die Bindungsaffinität zu verbessern. In der Literatur findet man zahlreiche Beispiele für Erfolge und Misserfolge dieser

Methoden, sowie zahlreicher modifizierter Varianten einiger ‚Standard‘-Programme. Eine Übersicht findet sich bei Liedl<sup>28</sup>.

Die erhaltene Dockingpose wird selbstverständlich nur dann von der Position der umgebenden Wassermoleküle beeinflusst, wenn diese nicht nachträglich hinzugefügt werden. In Hinblick auf ein automatisiertes Vorgehen, mit dem man über die Verdrängung und Konservierung von Wassermolekülen während des Dockings entscheiden kann und die dadurch neue Bindungsmodi zugänglich machen, können nur sehr wenige Programme angeführt werden. Das an der Universität Konstanz entwickelte Programm PLANTS<sup>53-57</sup> definiert kugelförmige Aufenthaltsbereiche um Wassermoleküle in der Bindetasche. In diesen Volumenelementen können sich die Wassermoleküle frei bewegen, damit ungünstige Wechselwirkungen mit dem Liganden und Protein vermieden werden. Sollte ein Wassermolekül diesen Bereich verlassen, so wird dieses als verdrängt angesehen und vollständig aus der Tasche entfernt. Dies entspricht ebenfalls dem prinzipiellen Vorgehen in SLIDE<sup>58</sup>. FITTED<sup>59</sup> geht einen anderen Weg. Hier wird das Vorhandensein der Wassermoleküle und deren Orientierung in das Chromosom eingefügt und über den genetischen Algorithmus zusammen mit der Konfigurationssuche des Liganden simuliert. In der vorliegenden Arbeit wurden die entsprechenden Methoden der Programme GOLD und FlexX eingesetzt. GOLD berechnet hierfür die Wechselwirkungsenergie des Wassermoleküls mit der Umgebung und vergleicht diesen Wert mit einem festgelegten Parameter, der den Freien Energiegewinn bei der Verdrängung darstellt.<sup>52</sup> Analog zur Vorhersage von Wasserpositionen in einer Bindetasche mit dem Particle Concept<sup>47</sup> wird in FlexX die Verdrängung gehandhabt. Für den Fall, dass ein Wassermolekül zusätzliche Wasserstoffbrückenbindungen mit dem Liganden ausbilden kann, wird es konserviert, falls nicht, werden dessen Wechselwirkungen mit der Umgebung abgeschaltet. Inwieweit die Ergebnisse der beiden letztgenannten Programme aussagekräftig sind, wurde unter anderem in dieser Arbeit untersucht.

Das konzeptionelle Vorgehen der vorgestellten Methoden beinhaltet jedoch einige grundsätzliche Schwierigkeiten bei der Vorhersage von Verdrängung und Konservierung von Wassermolekülen in der Bindetasche. Der Beitrag zur Freien Bindungsenthalpie beinhaltet einen Enthalpie- und einen Entropie-Term. Erst bei Kenntnis beider Größen kann über eine Änderung des Wassernetzwerkes bei der Ligandbindung geurteilt werden. Im Folgenden wird mit dem Begriff *Verdrängung* das Entfernen eines Wassermoleküls aus der Bindetasche ohne Substitution der H-Brücken dieses Wassermoleküls durch eine funktionelle Gruppe des Liganden bezeichnet. Mit *Ersetzen* hingegen wird der Prozess bezeichnet, in dem nicht nur das Wassermolekül verdrängt sondern dessen vorherige H-Brückenbindungen (zumindest teilweise) durch den Liganden ausgebildet werden. Folgende Möglichkeiten sind denkbar (alle thermodynamischen Größen werden für den Transfer des Wassermoleküls aus der Lösung in die Bindetasche angenommen):

- Für Wassermoleküle auf einer energetisch ungünstigen Position gilt  $\Delta G > 0$ .
  - Gilt für eine Wasserposition  $\Delta H > 0$  sowie  $-T\Delta S > 0$ , so ist die Verdrängung dieses Wassermoleküls durch einen Liganden aus thermodynamischer Sicht anzustreben.

- Gilt hingegen  $\Delta H < 0$  und  $-T\Delta S > 0$ , sodass jedoch  $\Delta G > 0$ , kann dieses Wassermolekül durch eine entsprechende funktionelle Gruppe des Liganden ersetzt werden. Bei dieser Ersetzung wird der Betrag des ungünstigen Entropie-terms an Energie frei und gleichzeitig können die an dieser Position günstigen H-Brückenwechselwirkungen durch den Liganden übernommen werden.
- Der letzte denkbare Fall mit  $\Delta H > 0$  und  $-T\Delta S < 0$ , wobei  $|\Delta H| > |-T\Delta S|$ , ist in der Praxis nicht relevant. Da die thermodynamischen Größen sich hier auf den Prozess des Wassertransfers vom umgebenden Solvens in die Bindetasche beziehen, wird die Entropie eines Wassermoleküls im Solvens stets günstiger sein, da hier die Beweglichkeit geringeren Restriktionen unterliegt.
- Es ist sinnvoll thermodynamisch stabile Wassermoleküle mit  $\Delta G < 0$  bei der Ligandbindung nicht zu verdrängen, da dies keinen energetisch vorteilhaften Beitrag zur Bindungsaffinität liefert. Auf Grund des oben bereits erwähnten sehr unwahrscheinlichen Falles mit  $-T\Delta S < 0$  ist hier nur die Variante mit  $\Delta H < 0$  und  $-T\Delta S > 0$  aber  $|\Delta H| > |-T\Delta S|$  relevant. Beim Design des Liganden ist die Verdrängung eines solchen Wassermoleküls gänzlich zu meiden. Auf Grund der günstigen enthalpischen Wechselwirkungen kann aber in Erwägung gezogen werden ein solches Molekül für eine Wasser-verbrückte Wechselwirkung zu verwenden.

Es wird deutlich, dass die exakte Kenntnis des thermodynamischen Profils einzelner Wassermoleküle für das Wirkstoffdesign von großem Wert sein kann. Dies wird durch die einfachen Ansätze, die in wenigen, aktuellen Docking-Programmen implementiert sind, nicht ermöglicht. Eine routinemäßige Klassifizierung von Wasserpositionen nach ihrem thermodynamischen Profil wäre also wünschenswert.

## **Teil II.**

# **Untersuchungen, Ergebnisse und Auswertungen**





# 5. Strukturelle Betrachtungen an der Modellbindetasche der Cytochrom c Peroxidase W191G Mutante

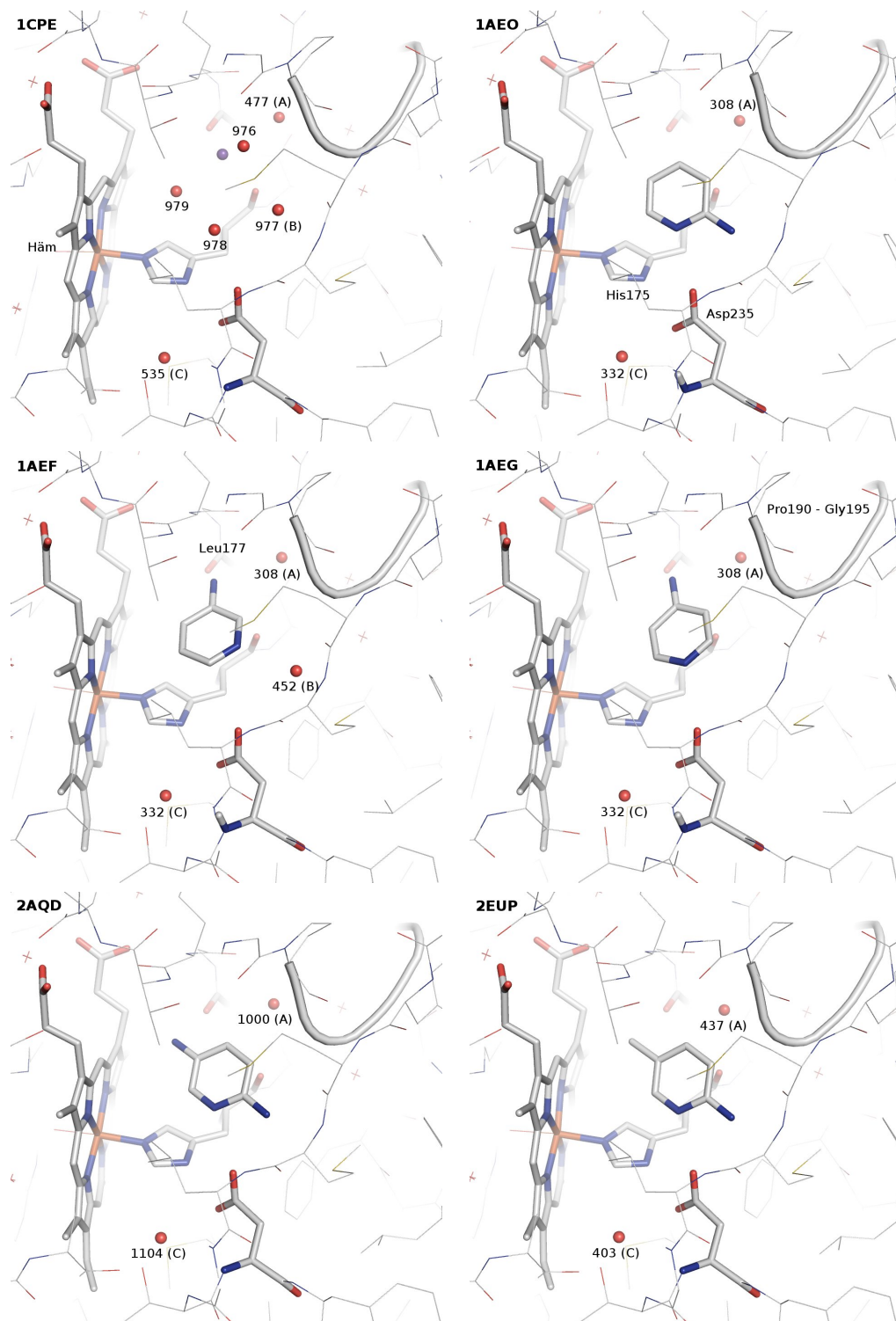
## 5.1. Strukturelle Datengrundlage für die durchgeführten Analysen

### 5.1.1. Vorstellung der in dieser Arbeit untersuchten Komplexstrukturen

Bereits in der Einführung zu dieser Arbeit wurden das Proteinsystem und die gewählten Komplexstrukturen, für die die Mehrheit der hier beschriebenen Untersuchungen und Analysen durchgeführt wurden, kurz vorgestellt. Es handelt sich hierbei um die Modellbindetasche der Cytochrom c Peroxidase Mutante W191G (CCP W191G). Diese stellt durch ihre tiefe Vergrabenheit, detaillierte strukturelle Charakterisierung und polare Umgebung einen idealen Kandidaten für die Untersuchung von Solvens-Effekten auf Stärke und Art der Protein-Ligand Bindung dar. Für die vorliegende Arbeit wurden aus der Fülle der bekannten und strukturell charakterisierten Komplexe der CCP W191G (siehe Tab. 1.1) fünf besonders interessante ausgewählt (vgl. Abb. 5.1). Diese sind strukturell sehr ähnlich aufgebaut, haben eine (soweit bekannt) ähnliche Bindungsaffinität und weisen dennoch interessante Unterschiede bei den Bindungsmodi und Wechselwirkungsschemata auf, gerade auch in Bezug auf die in der Tasche vorhandenen strukturellen Wassermoleküle. 2-, 3-, und 4-Aminopyridin sind allesamt einfache Konstitutionsisomere und binden dennoch auf sehr unterschiedliche Weise unter Beteiligung von keinem, zwei oder einem Wassermolekül. 2,5-Diaminopyridin wiederum lässt sich sowohl von 2- als auch 3-Aminopyridin ableiten und nimmt einen mit 2-Aminopyridin vergleichbaren Bindungsmodus ein. 2-Amino-5-picolin schließlich ist strukturell sehr ähnlich zu 2,5-Diaminopyridin und weist durch die fehlende Amino-Gruppe zwei polare H-Atome weniger auf, die für H-Brückenbindungen zur Verfügung stehen. Dennoch ist der Bindungsmodus dieser beiden Moleküle analog. Selbstverständlich wären auch andere Liganden komplexiert mit CCP W191G für die durchgeführten Untersuchungen von Interesse. Da diese jedoch keine unmittelbaren Konstitutionsisomere der ausgewählten fünf darstellen musste, ebenso aus zeitlichen Gründen, eine Beschränkung auf nur wenige zu untersuchende Systeme vorgenommen werden.

Neben den beschriebenen Unterschieden und Ähnlichkeiten der Ligand-Bindungsmodi weisen alle fünf Komplexe ein Wassermolekül nahe Leu177 (siehe Abb. 5.1) auf. Da die

## 5. Strukturelle Betrachtungen der Modellbindetasche



**Abb. 5.1.:** Übersicht über die dieser Arbeit zugrunde liegenden Kristallstrukturen der fünf untersuchten Komplexe von CCP W191G und der ligandfreien Struktur. Die angegebenen Nummern entstammen der Nummerierung der Wassermoleküle in den entsprechenden Kristallstrukturen.

**Tab. 5.1.:** RMSD Werte in Å für die untersuchten Kristallstrukturen berechnet für ein Alignment mit PyMOL basierend auf dem Proteinrückgrat.

	1AEO	1AEF	1AEG	2AQD	2EUP
1CPE	0.20	0.21	0.21	0.22	0.19
1AEO		0.19	0.19	0.30	0.26
1AEF			0.00	0.23	0.26
1AEG				0.23	0.26
2AQD					0.16

Nummerierung der Wassermoleküle in den Kristallstrukturen unterschiedlich vorgenommen wird, wird diese Wasserposition im Verlauf der vorliegenden Arbeit mit A gekennzeichnet. Das zweite Wassermolekül, das im Komplex 1AEF eine Bindung zum Liganden eingeht, wird in Analogie dazu mit B gekennzeichnet. Schließlich wurde in verschiedenen Auswertungen aus Gründen der Vergleichbarkeit das nicht an der Ligandbindung beteiligte Wassermolekül, das an Asp235 gebunden ist, mit untersucht. Es erhält im Folgenden die Bezeichnung C.

Alle fünf ausgewählten Strukturen waren bereits zu Beginn dieser Arbeit in der PDB verfügbar und ihre Auflösung ist in einem brauchbaren bis sehr guten Bereich. Die von Brenk et al.<sup>96</sup> gelösten Strukturen 2AQD und 2EUP haben eine Auflösung von 1.35 Å und 1.40 Å. Die bereits 2002 veröffentlichten<sup>95</sup> Strukturen 1AEO, 1AEF und 1AEG zeigen Auflösungen von 2.10 Å. Beim Lösen der Strukturen 1AEF und 1AEG wurden Protein- und Wasserkoordinaten aus der Struktur 1CMQ übernommen und eine Verfeinerung nur für den Liganden vorgenommen. Die daraus resultierenden Defizite bei den durchgeführten MD-Simulationen werden später an entsprechender Stelle erläutert (siehe Abschnitt 6.3). Für die Komplexe 2AQD und 2EUP ist auch ein  $R$ -Wert (0.134 und 0.153) bzw. der  $R_{\text{free}}$ -Wert (0.163 und 0.181) veröffentlicht. Letzterer misst, wie gut das Strukturmodell Messwerte vorhersagt, die nicht zur Anpassung des Modells dienen. Allgemein sollten diese Werte kleiner als 0.25 bzw. 0.4 sein, was bei den hier ausgesuchten Komplexen der Fall ist. Davis *et al.*<sup>20</sup> beschreiben, dass idealerweise die Differenz  $R_{\text{free}} - R < 0.05$  sein sollte, was im hier gezeigten Beispiel (0.029 und 0.028) ebenfalls erfüllt wird. In Tab. 5.1, 5.2 und 5.3 sind unterschiedliche RMSD Werte einander gegenübergestellt. Diese wurden nach dem Alignment der Strukturen bestimmt. Die Berechnung erfolgte entweder für alle Proteinrückgrat-Atome, nur für die Aminosäuren der Bindetasche ohne Seitenketten und H-Atome oder für den flexiblen PGGAAN Loop, der für Liganden ein Tor zur Bindetasche öffnen kann. Anhand dieser Werte erkennt man sehr gut die hohe strukturelle Ähnlichkeit nicht nur des gesamten Proteins, sondern insbesondere der untersuchten Modellbindetasche. In diese Auflistung wurde zusätzlich zu den Kristallstrukturen der fünf Komplexe noch die der ligandfreien Bindetasche (PDB-Code: 1CPE) aufgenommen.

Neben den Komplex-Strukturen der CCP W191G wurde für die in dieser Arbeit vor-

**Tab. 5.2.:** RMSD Werte in Å für die Aminosäurereste (ohne Seitenketten) in der Bindetasche basierend auf dem Alignment mit PyMOL für das Proteinrückgrat.

	1AEO	1AEF	1AEG	2AQD	2EUP
1CPE	0.17	0.17	0.17	0.29	0.27
1AEO		0.15	0.15	0.29	0.26
1AEF			0.00	0.31	0.29
1AEG				0.31	0.29
2AQD					0.11

**Tab. 5.3.:** RMSD Werte in Å für die Aminosäuren (ohne Seitenkette) aus der P190-G-G-A-A-N195 Schleife basierend auf dem Alignment mit PyMOL für das Proteinrückgrat.

	1AEO	1AEF	1AEG	2AQD	2EUP
1CPE	0.20	0.15	0.15	0.41	0.34
1AEO		0.19	0.19	0.40	0.31
1AEF			0.00	0.39	0.33
1AEG				0.39	0.33
2AQD					0.17

**Tab. 5.4.:** Vergleich der ligandfreien CCP W191G Kristallstrukturen. RMSD Werte wurden aus dem Alignment auf die Struktur 1CPE mit dem Programm PyMOL entnommen.

Struktur	RMSD	Auflösung	<i>R</i>
1CPE	-	2.20 Å	0.161
1CPD	0.17 Å	2.20 Å	0.161
1BEJ	0.11 Å	2.40 Å	0.164
1CMQ	0.25 Å	2.30 Å	0.190
1AA4	0.26 Å	2.10 Å	0.189
1CMT	0.26 Å	2.10 Å	0.168

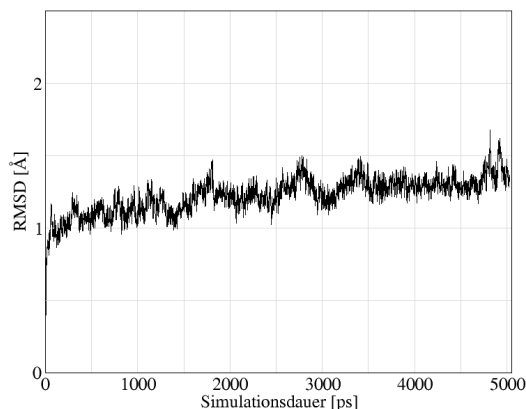
gestellten Untersuchungen noch eine weitere Kristallstruktur des Proteins verwendet, die keinen Liganden enthält. Sechs derartige Strukturen sind zur Zeit literaturbekannt (PDB-Codes: 1BEJ, 1CMQ, 1AA4, 1CMT, 1CPD, 1CPE). In allen liegt der PGGGAN-Loop in der geschlossenen Konformation vor. Sie sind damit prinzipiell geeignet, um im Vergleich zu den Komplexen analysiert zu werden. Diese Schleifen-Region ermöglicht in ihrer offenen Konformation den Zugang von Liganden zur Bindetasche.<sup>95</sup> Auflösung, *R*-Wert der Kristallstruktur und RMSD des Alignments auf die Struktur 1CPE als Referenz sind in Tab. 5.4 aufgelistet. Man sieht, dass alle sechs Strukturen eine sehr hohe strukturelle Ähnlichkeit aufweisen.

Ein wesentlicher Unterschied ist jedoch die Anzahl der Wassermoleküle in der Bindetasche von CCP W191G. Die Strukturen 1BEJ, 1CMQ, 1AA4 und 1CMT zeigen hier sechs Wassermoleküle, die im Vergleich der Strukturen annähernd die gleichen Positionen einnehmen. Die Strukturen 1CPE und 1CPD weisen hingegen nur fünf konservierte Wassermoleküle in der Bindetasche auf, deren Positionen mit denen der übrigen Strukturen übereinstimmen. Das sechste Wassermolekül fehlt zugunsten eines  $K^+$ -Ions im Falle von 1CPE und  $NH_4^+$  im Falle von 1CPD. Aus verschiedenen Gründen kann man annehmen, dass stets ein Kation in der ligandfreien Bindetasche vorliegt.

- Es ist bekannt<sup>20</sup>, dass die Elektronendichte von Wasser und z.B. einem  $Na^+$ -Ion kristallographisch praktisch nicht unterscheiden ist.
- Die Bindetasche trägt durch das Asp235 eine negative Ladung, weshalb die Mehrzahl der bekannten Liganden Kationen sind.
- Das vermeintliche sechste Wassermolekül ist ungefähr oktaedrisch koordiniert.
- Die Kristallationsbedingungen von z.B. 1CMQ (Soaking in  $Na_3PO_4$  Puffer) bedingen, dass  $Na^+$ -Ionen vorhanden sind.

Das vermeintliche sechste Bindetaschenwassermolekül ist somit sehr wahrscheinlich ein  $Na^+$ -Ion. Prinzipiell wäre aber auch einfach die Anwesenheit eines  $H_3O^+$ -Ions möglich. Dies wurde in einer Arbeit aus der Gruppe von G. Klebe<sup>195</sup> für die Bindetasche der tRNA-Guanin Transglycosylase vermutet. Für die hier dargestellten Untersuchungen wurde deshalb die Struktur 1CPE herangezogen. Hierfür kann in Analogie zu den Komplexstrukturen den Wassermolekülen 477, 977 und 535 ebenfalls die Bezeichnung A, B und C zugeordnet werden, da sie den entsprechenden Wasserpositionen der Komplexstrukturen am nächsten liegen.

Da ein Ion in der Bindetasche eigentlich auch ein Ligand ist, fehlt bei der bisher getroffenen Auswahl an Strukturen noch die Vergleichsmöglichkeit zwischen einer Komplexstruktur der CCP W191G und einer vollständig ligandfreien Bindetasche. Aus diesem Grund wurde mit einer Kombination verschiedener computergestützter Methoden ein Modell einer solchen ligandfreien Kavität erstellt. Als strukturelle Grundlage diente hierbei die Kristallstruktur 2AQD, da sie im hier verwendeten Datensatz die höchste Auflösung hat. Zunächst wurde mit Hilfe des Programms GRID<sup>30,196-198</sup> ein molekulares Wechselwirkungsfeld (weitere Details zu diesem Verfahren siehe unten) für eine Wasser-Sonde basierend auf einer

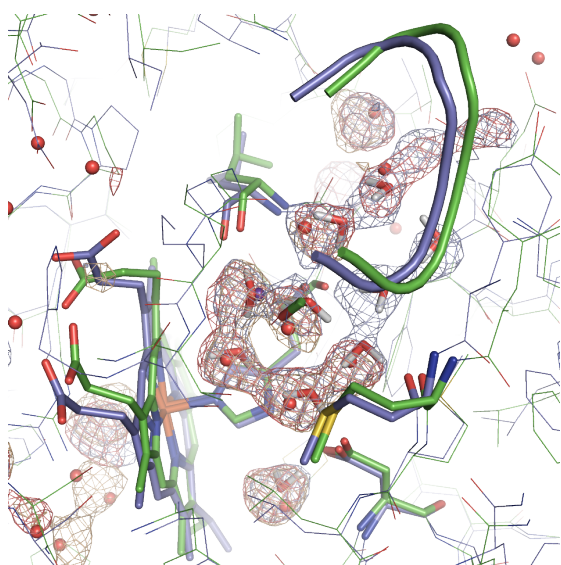


**Abb. 5.2.:** Auftragung des RMSD-Werte der  $C_{\alpha}$ -Atome der MD-Simulation des Modells der CCP W191G Tasche ohne Ion.

vollständig leeren Bindetasche berechnet. Mit Hilfe des Programms MOE wurde danach ein Wassermolekül an jene Stelle der Bindetasche platziert, die die niedrigste molekulare Wechselwirkungsenergie zeigte. Ausgehend von diesem Modell wurde erneut eine Berechnung des molekularen Wechselwirkungsfeldes durchgeführt. Bei dieser Berechnung wurde nun das bereits hinzugefügte Wasser mit einbezogen. Dieses wechselseitige Berechnen molekularer Wechselwirkungsfelder und Hinzufügen von Wassermolekülen wurde so oft wiederholt, bis aus dem berechneten Gitter keine separierte Wasserposition mehr ermittelt werden konnten. Auf diese Weise wurden fünf Wassermoleküle in der Bindetasche platziert, weil ab einem Energiewert von  $-10$  kcal/mol die Darstellung des molekularen Wechselwirkungsfeldes keine separierte Wasserposition mehr zeigte.

In dieses Modell wurden anschließend die Wassermoleküle 445, 494, 500, 535 und 611 der 1CPE Kristallstruktur integriert. Diese finden sich nicht direkt in der Bindetasche, sind jedoch für eine Ausgangsstruktur einer MD-Simulation relevant. Von diesem Modell wurde sodann eine MD-Simulation (Details zu den dafür verwendeten Parametern und Einstellungen in Abschnitt 6.1) über 5 ns durchgeführt. Bereits nach kurzer Zeit zeigte eine Auftragung von RMSD-Werten (Abb. 5.2) der Protein-Atome bezogen auf die Startstruktur, dass die Positionen dieser Atome weitgehend equilibriert sind. Der Grad der Equilibrierung der Wasserpositionen wurde mit Hilfe der Berechnung mittlerer Hydratationsdichtekarten (Details siehe Abschnitt 6.1) für 1 ns Abschnitte der Trajektorie analysiert. Beim Vergleich dieser Dichten (siehe Abb. 5.3) zeigte sich, dass nach 4 ns Simulation die Positionen der Wassermoleküle in der Bindetasche in einem hohen Maße konserviert sind. Aus dem Trajektorienabschnitt zwischen 4 und 5 ns wurde deshalb jene Konformation (nach 4569 ps) als Modellstruktur ausgewählt, die den geringsten RMSD der Proteinatome bezüglich einer gemittelten Struktur über diesen Trajektorienabschnitt zeigte. Diese Struktur wurde anschließend einer kurzen (200 Schritte) Minimierung mit SANDER und Generalized-Born implizitem Solvensmodell unterzogen.

Die aus der Trajektorie ausgewählte und minimierte Struktur zeigt nur sehr geringe strukturelle Unterschiede zur Basis dieses Modells, der Kristallstruktur 2AQD. Der RMSD-Wert der Proteinrückgrat-Atome beträgt  $0.94$  Å. Auch im Vergleich zu den Röntgenkristallstrukturen ohne Ligand in der Bindetasche zeigen sich nur sehr geringe strukturelle



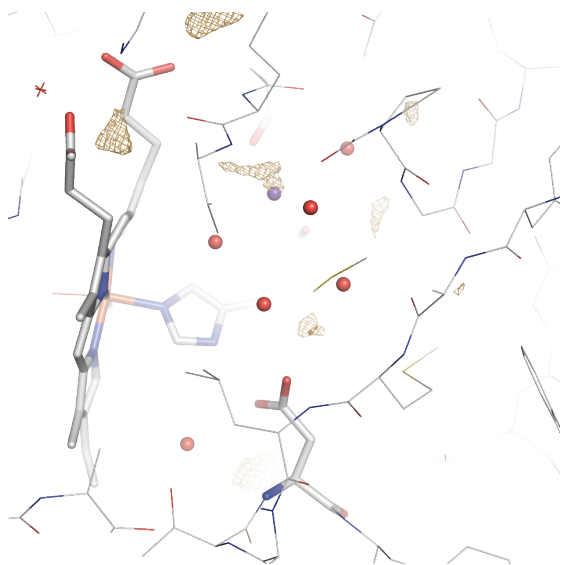
**Abb. 5.3.:** Die Abbildung zeigt die minimierte MD-Struktur nach 4569 ps als ionenfreies Modell der CCP W191G (grün) zusammen mit der Kristallstruktur von 1CPE (blau). Zusätzlich sind auch die für unterschiedliche Abschnitte der Trajektorie berechneten Hydratationsdichten gezeigt (in Braun für 2-3 ns, in Rot für 3-4 ns, in Blau für 4-5 ns). Beide Proteinstrukturen zeigen eine hohe Übereinstimmung, die Wasserpositionen des Modells (Stab-Darstellung) stimmen ebenfalls mit denen der Kristallstruktur (Kugeln) überein. Auch im Vergleich der Hydratationsdichte für unterschiedliche Zeitabschnitte erkennt man annähernd equilibrierte Wasserpositionen.

Unterschiede für die Proteinatome. Der RMSD bezogen auf das Protein-Rückgrat beträgt für 1CPE 0.87 Å, für 1CPD 0.86 Å, für 1BEJ 0.88 Å, für 1CMQ 0.86 Å, für 1CMT 0.86 Å. Die Abweichung der Wasserpositionen in der Tasche sind insgesamt gering. Die Position des  $K^+$ -Ions in 1CPE ist mit einem Wassermolekül besetzt, wie dies aber auch für die anderen ligandfreien Strukturen der Fall ist. Für sechs der Bindetaschenwassermoleküle liegen die Abstände der Sauerstoffatompositionen zu denen der ligandfreien Kristallstrukturen im Bereich von 1.5 bis 2.0 Å, für Wassermolekül C sogar unter 1 Å und somit im Bereich der Auflösung der Kristallstrukturen. Der einzige Unterschied ist im Bereich von Gly189 zu beobachten. In der aus der Simulation erhaltenen Modellstruktur ist dieser Schleifenbereich etwas geöffnet und an der Stelle des Carbonylsauerstoffatoms der Aminosäure findet sich ein zusätzliches Wassermolekül.

### 5.1.2. Berechnung molekularer Wechselwirkungsfelder

Bereits im vorangegangenen Abschnitt wurde die Berechnung molekularer Wechselwirkungsfelder mit dem Programm GRID<sup>30,196-198</sup> angesprochen. Zu diesem Zweck wurden die jeweils zu untersuchenden Strukturdaten im PDB-Dateiformat über die grafische Benutzeroberfläche GREATER eingelesen. Dabei wurde im Vorfeld die Benennung der Atome des Cofaktors so angepasst, dass das Programm eine Zuordnung der für Häm gespeicherten Parameter vornehmen konnte. Als Bereich, in dem die Berechnung des molekularen Wechselwirkungsfeldes vorgenommen werden soll, wurde ein Quader, dessen Kanten 10 Å vom Häm entfernt sind, definiert. Die Werte der Wechselwirkungsenergie wurden für Gitterpunkte im Abstand von 0.5 Å berechnet. Das Ausgabeformat wurde so gewählt, dass die Ergebnisse mit PyMOL<sup>199</sup> visualisiert werden können. Für alle anderen Einstellungen wurden die Standardwerte beibehalten.

Um die vorgenommene Interpretation der experimentellen Daten zum Vorliegen eines Kations in den Kristallstrukturen zu untermauern, wurde, wie bereits für die Wasser-Sonde



**Abb. 5.4.:** Darstellung des mit GRID für die Bindetasche der Kristallstruktur 1CPE berechneten molekularen Wechselwirkungsfeldes für die  $K^+$ -Sonde. Die Konturlinien sind für einen Energiewert von  $-25.0$  kcal/mol eingezeichnet. Wassermoleküle der Kristallstruktur sind durch rote Kugel, das  $K^+$ -Ion durch eine violette Kugel dargestellt.

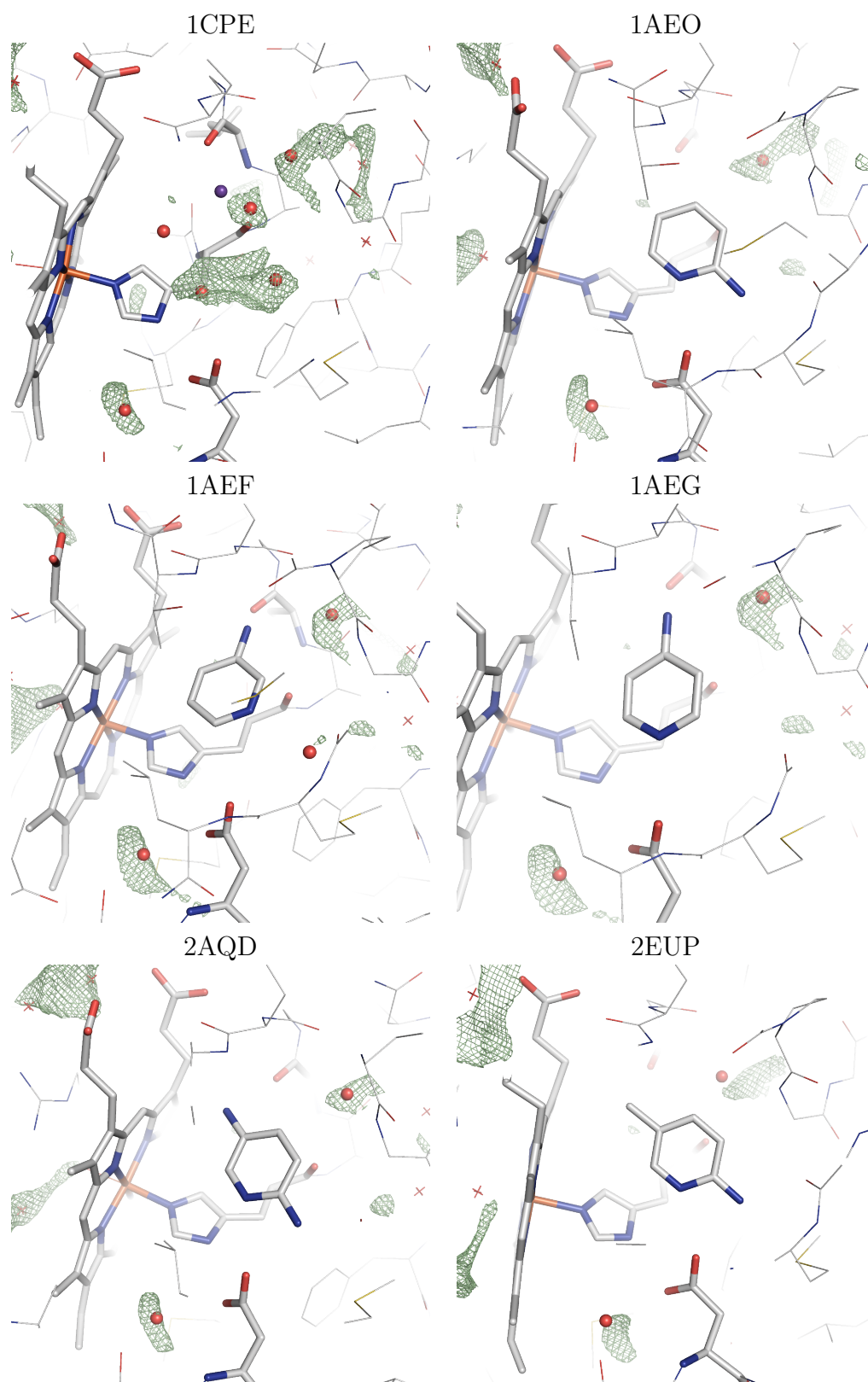
beschrieben, das molekulare Wechselwirkungsfeld für die  $K^+$ -Sonde in der 1CPE-Struktur berechnet. Aus Abbildung 5.4 ist ersichtlich, dass sich das  $K^+$ -Ion in einem mit  $-25$  kcal/mol energetisch sehr günstigen Bereich befindet.

Analog dazu wurde die Wechselwirkungsenergie für eine Wasser-Sonde für die 1CPE-Bindetasche mit  $K^+$ -Ion berechnet. Die Ergebnisse hiervon sind in Abb. 5.5 dargestellt. Eine Visualisierung der molekularen Wechselwirkungsfelder für unterschiedliche Energien zeigt, dass alle Wassermoleküle im Strukturmodell auf energetisch günstigen Positionen zu finden sind. Bei einem Energie-Wert von  $-8.0$  kcal/mol, wie in Abb. 5.5 dargestellt, ist dies für fünf der sechs gezeigten Moleküle der Fall. Die weniger negativen Energiewerte im Falle der Wasser-Sonde im Vergleich zum  $K^+$ -Kation kommen durch deren neutrale Ladung zustande.

Für die fünf untersuchten Komplexe wurde ebenfalls eine Berechnung der molekularen Wechselwirkungsfelder für die Wasser-Sonde durchgeführt. Die Ergebnisse sind ebenfalls in Abb. 5.5 dargestellt. Es zeigt sich, dass beide relevanten Wassermoleküle A und B, wie auch Wasser C in energetisch vorteilhaften Positionen zu finden sind.

Zusammenfassend kann man festhalten, dass die durchgeführten Energie-Berechnungen die experimentell gefundenen Positionen von Wassermolekülen und  $K^+$ -Ion untermauern, sodass davon ausgegangen werden kann, dass hier keine Ungenauigkeiten oder Fehler in der Kristallstruktur vorliegen. Ebenso konnten auch keine weiteren energetisch vorteilhaften Regionen in unmittelbarer Nähe zum Liganden gefunden werden, an denen Wassermoleküle zu erwarten sind, die experimentell nicht beobachtet wurden. Das wäre prinzipiell möglich, da sehr mobile Wassermoleküle auf der Zeitskala eines Röntgenstruktur-Experimentes nicht notwendigerweise detektiert werden. Nicht jedes mobile Wassermolekül wäre aber so aufzuspüren, da die erhöhte Beweglichkeit ja auch durch sehr schwache Wechselwirkungen verursacht werden kann.





**Abb. 5.5.:** Darstellung der mit GRID für die hier untersuchten Bindetaschen der CCP W191G berechneten molekularen Wechselwirkungsfelder für die Wasser-Sonde. Die Konturlinien sind für einen Energiewert von  $-8.0$  kcal/mol eingezeichnet.

## 5.2. Docking-Studien

### 5.2.1. Zugrunde liegende Fragestellungen

Eingangs wurde bereits erwähnt, dass einige der heute verfügbaren Dockingprogramme während des Docking-Prozesses zu bestimmen versuchen, ob es günstig ist, ein Wassermolekül aus der Bindetasche durch den Liganden zu verdrängen oder es dort zu belassen. Eine Intention dieser Arbeit war es zu überprüfen, inwieweit diese Funktion auch an der untersuchten Modellbindetasche der CCP W191G zuverlässig funktioniert und ob Bedarf an einer Verbesserung besteht, bzw. wie diese umgesetzt werden müsste. Für diesen Test wurden explizit zwei (kommerziell vertriebene) Programme ausgewählt, die im Bereich der computerbasierten Wirkstoffentwicklung weit verbreitet sind und eine Funktion zur Vorhersage der Wasserverdrängung beinhalten. Dabei handelt es sich um die Programme GOLD<sup>49-52</sup> und FlexX<sup>47,48</sup>. Diese wurden laut Literatur noch nicht im Zusammenhang mit der CCP W191G Tasche eingesetzt. Wie in Abschnitt 1.3 erwähnt, wurden in der Vergangenheit bereits Docking-Studien an CCP W191G mit Autodock, DOCK, FRED, QXP/FLO und GLIDE durchgeführt. Darin wurden aber keine Untersuchungen in Hinblick auf die Wasserverdrängung durchgeführt. Zunächst wurde in der vorliegenden Arbeit ein Redocking durchgeführt. Dabei werden die Liganden aus einer Komplexstruktur entfernt, um dann mit Hilfe des Dockingprogramms wieder in diese platziert zu werden. Aufgrund des für diese Proteinkonformation bekannten Bindungsmodus können die Ergebnisse des Dockingprogramms validiert werden. Danach wurde das Docking der Liganden in die ligandfreie Rezeptorstruktur durchgeführt.

### 5.2.2. Verwendete Einstellungen und Parameter der eingesetzten Dockingprogramme

Die Vorbereitung der Docking-Simulationen unterscheidet sich für die beiden verwendeten Programme nur geringfügig voneinander. Strukturelle Grundlage für das Redocking bildeten die PDB-Strukturen 1AEO, 1AEF (ebenso die neu kristallisierte Struktur dieses Komplexes, vgl. Abschnitt 6.3), 1AEG, 2AQD und 2EUP. Das Docking in die ligandfreie Tasche wurde auf Grundlage der 1CPE Struktur sowie des Strukturmodells für eine ionenfreie Tasche, das aus der MD Simulation erhalten wurde (siehe dazu Abschnitt 5.1.1), durchgeführt.

Für das Docking mit GOLD<sup>49-51</sup> wurden mit Hilfe des Programms Sybyl 8.0<sup>200</sup> die fehlenden Wasserstoffatome zu den Protein-Kristallstrukturen hinzugefügt. Alle Wassermoleküle außerhalb der Bindetasche wurden aus dem Strukturmodell entfernt. Im Fall des Systems 1CPE wurde außerdem das K<sup>+</sup>-Ion aus der Tasche gelöscht. Zum Modell der Bindetasche, das aus der MD-Simulation entnommen wurde, mussten hingegen keine Wasserstoffatome hinzugefügt werden. Für das Programm FlexX<sup>47,48</sup> wiederum wurden die PDB-Dateien ohne Wasserstoffatome verwendet. Die weiteren Vorbereitungsschritte wurden mit den grafischen Oberflächen der Docking-Programme durchgeführt. Dies umfasste die Definition der Bindetaschenregion, Auswahl von Proto- und Tautomeren in der Bindetaschenregion und

die Wahl der Parameter für das Docking.

Für das Docking wurde GOLD in der Version 4.0 mit Standardeinstellungen für den GA (Populationsgröße 100, Selektionsdruck: 1.1, Anzahl der Operationen 100 000, Anzahl der Inseln: 5, Nischengröße: 2, Crossoverfrequenz: 95, Mutationsfrequenz: 95, Migrationsfrequenz: 10) eingesetzt. Die Docking-Simulation für jede der fünf Verbindungen wurde 50-mal durchgeführt und die Option zum vorzeitigen Beenden abgeschaltet. Außerdem wurde der Grenzwert für das Clustering auf 1.0 Å gesetzt. Während des Redockings wurden die Wechselwirkungen aller Bindetaschenwassermoleküle jeweils einzeln an oder aus geschaltet bzw. in den so genannten „togglig“ Modus gesetzt. Für das Docking in die ligandfreie Bindetasche wurde für alle Wassermoleküle das An- bzw. Abschalten der Wechselwirkungen dem Programm überlassen und der „togglig“-Modus gewählt. Alle Docking-Experimente mit GOLD wurden sowohl mit der GoldScore wie auch der Chemscore Scoring-Funktion durchgeführt.

In Analogie dazu wurden in der Version 3.1.1 von FlexX ebenfalls Standardeinstellungen gewählt. Das Basisfragment wurde mit Hilfe der triangle hashing Technik platziert. Die maximale Zahl der Lösungen, die pro Iteration und Fragmentierung mitgenommen wurden, lag jeweils bei 200. Am Ende wurden dann alle Dockingposen ausgegeben. Mit Hilfe der grafischen Benutzeroberfläche wurde die Bindetasche 6.5 Å um den Liganden gewählt. Darin wurden Proto- und Tautomere der Aminosäuren sowie die Torsionswinkel einiger funktioneller Gruppen, an denen Wasserstoffatome hinzugefügt wurden, durch visuelle Betrachtung gewählt. Beim so durchgeführten Redocking wurden alle Wassermoleküle jeweils in den Modus „an“, „aus“ oder „verdrängbar“ versetzt. Beim Docking in die ligandfreie Bindetasche wurde für alle Wassermoleküle das Attribut „verdrängbar“ gewählt.

Die anschließende Auswertung erfolgte im Falle der FlexX Ergebnisse mit der grafischen Benutzeroberfläche des Programms. Die GOLD Ergebnisse wurden mit Hilfe von PyMOL<sup>199</sup> visualisiert. Die Berechnung der RMSD-Werte im Vergleich zur Position des Liganden in der entsprechenden Kristallstruktur erfolgte mit dem zu GOLD gehörenden Programm smart\_rms. Durch die Überlagerung der Röntgenkristallstruktur eines jeweiligen Komplexes mit der Rezeptor-Struktur des Dockings in die ligandfreie Tasche (1CPE oder aus der MD-Simulation) wurden neue Koordinaten für das Ligand-Molekül erhalten. Diese wurden dann als Referenz verwendet, um die RMSD-Werte für das Docking in die ligandfreie Bindetasche zu berechnen. Alle Lösungen wurden untereinander bezüglich ihres RMSD-Wertes mit einem „complete linkage“ hierarchischen Clustering gruppiert.

### 5.2.3. Ergebnisse des Redockings in die ursprünglichen Kristallstrukturen

Um zu überprüfen, inwieweit die verwendeten Docking-Algorithmen und Scoring-Funktionen die Orientierung der Liganden in den Kristallstrukturen reproduzieren können, wurde zunächst ein Redocking der fünf Liganden in ihre jeweilige Rezeptor-Proteinstruktur durchgeführt. Für die Wassermoleküle wurden dabei unterschiedliche Einstellungen gewählt. Die Wechselwirkungen der Bindetaschenwassermoleküle können in den verwendeten Docking-

Programmen an- und ausgeschaltet werden. Außerdem ist es im Programm FlexX möglich, das Wassermolekül beim Docking zu verdrängen bzw. im Programm GOLD während der GA-Suche die Wechselwirkungen des Wassermoleküls an- und abzuschalten. Bei den fünf untersuchten Komplexen wurden jeweils alle Kombinationen an Einstellungen für das konservierte Wassermolekül A, im Falle von 1AEO, 1AEG, 2AQD und 2EUP, bzw. die zwei Wassermoleküle A und B, im Falle von 1AEF, gewählt. Auf diese Weise sollte es möglich sein, die Leistungsfähigkeit dieser Optionen im Falle des hier untersuchten Rezeptors grob abzuschätzen. Die Ergebnisse dieser Docking-Simulationen sind in Tab. 5.5, 5.6 und 5.7 zusammengestellt. Hier sind bereits die Ergebnisse einer neueren Kristallstruktur des 1AEF-Komplexes aufgeführt. Einzelheiten zu dieser Röntgenkristallstruktur finden sich in Abschnitt 6.3. Im Folgenden werden die Ergebnisse diskutiert.

**GOLD/GoldScore.** Das Redocking mit der Software GOLD in Kombination mit der Scoring-Funktion GoldScore liefert sehr unterschiedliche Ergebnisse für die fünf untersuchten Komplexe. Für die Systeme 1AEG, 2AQD und 2EUP sind unabhängig vom Vorhandensein von Wassermolekülen in der Bindetasche die best-bewerteten Lösungen auch korrekt (RMSD im Vergleich zur Kristallstruktur  $< 1 \text{ \AA}$ ). Für den Komplex 1AEO sind die Lösungen mit wechselwirkendem Wassermolekül A ebenfalls korrekt. In Abwesenheit des konservierten Wassermoleküls ist die best-bewertete Lösung sehr unterschiedlich im Vergleich zur Referenz-Kristallstruktur (RMSD  $> 3 \text{ \AA}$ ). Eine korrekte Lösung wurde aber auch hier gefunden, jedoch deutlich schlechter bewertet (Rang 43). Allgemein kann festgehalten werden, dass der „toggling“-Modus dazu führt, dass die Wechselwirkungen der Wassermoleküle korrekterweise eingeschaltet werden. Die Docking-Lösungen des Komplexes 1AEF wiederum müssen differenzierter betrachtet werden. Keine der best-bewerteten Lösungen kommt hier zufriedenstellend nah an das experimentelle Ergebnis heran (RMSD  $> 1.5 \text{ \AA}$ ). Jedoch unterschreitet in allen Docking Simulationen dieses Systems mindestens eine Lösung (die mit kleinstem RMSD-Wert) diese Grenze. Prinzipiell kann also die korrekte Lösung auch für diesen Komplex vom Algorithmus gefunden werden, sie wird jedoch nicht ausreichend hoch bewertet. Im Vergleich dazu sind die Ergebnisse für die neue 1AEF-Struktur deutlich besser. Besonders gute Resultate können erzielt werden, wenn beide Wassermoleküle in der Bindetasche vorhanden sind. Entweder, weil dies im Vorfeld so festgelegt wurde, oder weil diese Lösung durch den „toggling“-Modus gefunden wurde. Interessanterweise zeigte aber auch die Lösung, in der die Wechselwirkungen beider Wassermoleküle von Beginn an ausgeschaltet wurden, einen guten RMSD.

**GOLD/Chemscore.** Die Kombination aus dem Docking-Algorithmus in GOLD und der Scoring Funktion Chemscore findet für die Komplexe 1AEO, 2AQD und 2EUP sehr gute Ergebnisse (RMSD  $< 1.0 \text{ \AA}$ ). Für die Komplexe 1AEG und 1AEF hingegen wird gar nicht oder nur bei einer schlecht bewerteten Lösung die richtige Orientierung des Liganden in der Bindetasche gefunden. Die RMSD-Werte der Ligandatome im Vergleich zur Referenzstruktur nehmen Werte  $> 3.0 \text{ \AA}$  an. Bemerkenswert ist aber vor allem, dass die Resultate für vier von fünf Komplexen annähernd vollständig unabhängig davon sind, ob in

der Bindetasche ein oder mehrere Wassermoleküle vorhanden sind. Nur im Falle der neuen 1AEF-Struktur kann für die Kombinationen Wasser A/Wasser B on/on, off/on, off/toggle und toggle/on eine richtige Lösung gefunden und somit ein Einfluss der Wasser-Präsenz in der Bindetasche beobachtet werden. Diese Lösungen sind aber allesamt nicht die best-bewertete und außerdem nicht im gleichen Cluster wie die best-bewertete Lösung zu finden. Bei der praktischen Anwendung des Dockings in einem virtuellen Screening wären diese Resultate unter Umständen nicht zur Auswertung herangezogen worden. Allgemein zeichnet sich für diese Kombination aus Docking-Programm und Scoring-Funktion in Bezug auf die „toggle“-Funktionalität in GOLD eine Tendenz zum Abschalten der Wasserwechselwirkung ab.

**FlexX.** Wie unterscheiden sich die Lösungen bei Verwendung eines anderen Docking-Programms, FlexX? Die Redocking Lösungen der Komplexe 1AEO, 2AQD und 2EUP liefern hervorragende Ergebnisse. Die RMSD-Werte liegen unter  $0.5 \text{ \AA}$  und in vielen Fällen ist die best-bewertete Lösung auch die mit dem kleinsten RMSD-Wert. Für keinen der drei Komplexe wurde das Wassermolekül A aus der Bindetasche verdrängt. Die Ergebnisse für den Komplex 1AEG sind hier überraschend schlecht. Die korrekte Lösung mit dem kleinsten RMSD-Wert wird durch die Scoring-Funktion sehr schlecht bewertet. Hier, wie für die anderen drei besprochenen Komplexe, sind die aufgelisteten Lösungen völlig unabhängig davon, ob ein Wassermolekül in der Bindetasche vorhanden ist oder nicht. Das System 1AEF sollte auch hier wieder eingehender betrachtet werden. Die best-bewerteten Ergebnisse sind hier relativ schlecht. Einzige Ausnahme bildet die Docking-Simulation, in der beide Wassermoleküle in den „displacable“-Modus gesetzt wurden. Hierbei konnte ein sehr gutes Resultat erzielt werden (RMSD  $0.88 \text{ \AA}$ ). Auch für alle anderen Wasser-Modi konnte FlexX die richtige Lösung finden, bewertete sie aber sehr schlecht. Vergleicht man dies mit der neuen 1AEF-Struktur so ändert sich das Ergebnis grundlegend. Das Docking mit der neu gelösten Kristallstruktur dieses Komplexes liefert für einen Großteil der eingestellten Wasseroptionskombinationen eine korrekte Lösung auf Rang 1. Die einzige Ausnahme findet man, wenn beide Bindetaschenwassermoleküle während des Dockings nicht vorhanden sind. Zusammenfassend bleibt außerdem festzuhalten, dass der „displacable“-Modus in fast allen Fällen das Wasser in der Bindetasche belassen hat.

Es hat sich gezeigt, dass die verwendeten Docking-Programme, Scoring-Funktionen und Einstellungen prinzipiell geeignet sind, um die Bindungsmodi der untersuchten Liganden zu reproduzieren. Insbesondere FlexX lieferte sehr gute Ergebnisse. Es zeigt sich an einigen Stellen Optimierungsbedarf im Bereich der Scoring-Funktionen. In einigen Fällen wurden die strukturell besten Lösungen schlecht bewertet und waren weit entfernt von Rang 1. Zudem wurde in einigen Systemen die Wasserverdrängung nicht korrekt vorher gesagt. Die Unterschiede im Score sind in diesen Fällen aber oftmals sehr gering, weshalb eine Identifizierung eines nicht optimalen Terms äußerst schwierig ist. Allgemein weisen die Ergebnisse der Wasserverdrängungsvorhersage darauf hin, dass die dafür relevanten Parameter noch einmal überprüft werden sollten. Möglicherweise sollten man weitere ex-

perimentelle Daten wie auch Simulationsergebnisse einfließen lassen. Überraschenderweise ist das Auffinden einer korrekten Lösung in vielen Fällen unabhängig vom Vorhandensein von Wassermolekülen in der Bindetasche. Gerade die Scoring-Funktion ChemScore zeigte sich hier als völlig ungeeignet, um eine Vorhersage über das Wassernetzwerk zu treffen. Umgekehrt wurde von FlexX die korrekte Lösung der alten 1AEF-Struktur nur dann auf Platz 1 bewertet, wenn für beide Wassermoleküle der „displaceable“ Modus eingestellt war. Auch wenn man im Falle eines falsch vorhergesagten Wassernetzwerkes mit einem korrekten Bindungsmodus des Liganden zufrieden sein könnte, ist diese Tatsache jedoch nur bedingt nützlich. In einem virtuellen Screening kann das Vorhandensein oder die Abwesenheit von Wassermolekülen einen signifikanten Einfluss auf das Wechselwirkungsschema und damit das Design eines Inhibitors haben. Selbstverständlich kann von den Untersuchungen an einem einzigen Proteinsystem kein allgemeiner Schluss auf die Leistungsfähigkeit der Docking-Algorithmen und Bewertungs-Funktionen gezogen werden. Aber auch in der Eingangs erwähnten Studie<sup>96</sup> wurde diese Modellbindetasche herangezogen, um die Leistungsfähigkeit der Docking-Methoden unter komplexeren Randbedingungen (geladene Tasche) zu erproben. Das System 1AEF stellte sich in beiden Programmen als herausfordernd dar. Das Auffinden einer korrekten Docking-Lösung gelang hier oftmals weniger gut, als bei den anderen Komplexen der CCP W191G. Auffallend ist aber auch die deutliche Verbesserung, die durch die neu kristallisierte 1AEF Struktur in diesen Fällen erreicht werden konnte. Die geringen strukturellen Änderungen waren hier offenbar ausreichend. Es bleibt zu überlegen, ob die Standard-Proteinpräparierung in GOLD und FlexX deshalb nicht eine Option zur kurzen Minimierung der Rezeptorstruktur enthalten sollte, damit mögliche Defizite der experimentellen Daten bei Bedarf ausgeglichen werden können. Die ähnliche Affinität der Liganden zum Protein, spiegelt sich im Übrigen in den erhaltenen Scores wider. Im Vergleich der Programme beobachtet man zudem einen analogen Trend zwischen den Liganden. 4-Aminopyridin (1AEG) wird etwas schlechter und der Ligand in 2EUP besser bewertet als der Rest.

**Tab. 5.5.:** Ergebnisse des Redockings mittels GOLD 4.0 und der Scoring-Funktion GoldScore. Angeführt sind nur jene Lösungen, die am besten bewertet wurden und jene, deren RMSD-Werte am kleinsten sind. Die Abkürzung „CR“ steht für den Cluster Rang, die Abkürzung „GS“ für GoldScore. Alle Lösungen wurden bezogen auf ihre RMSD Abweichungen untereinander geclustert. Die Cluster wiederum sind nach der enthaltenen best-bewerteten (Score) Lösung geordnet. Die hier gezeigten RMSD Werte wurden zwischen der Docking-Lösung und der Kristallstruktur als Referenz bestimmt. In der letzten Spalte ist angegeben, ob für Wassermoleküle im „toggling“-Modus in der entsprechenden Lösung die Wechselwirkung an- oder ausgeschaltet wurde.

System	Wasser A	Wasser B	Rang	CR	GS	RMSD (Å)	Wasser?	
1AEO	on	-	1	1	42.21	1.03	-	
			41	1	41.22	0.56	-	
	off	-	1	1	40.07	3.30	-	
			43	2	39.20	0.59	-	
	toggle	-	1	1	43.27	1.01	an	
			10	1	43.08	0.64	an	
	1AEF		on	1	1	41.19	2.19	-
				44	2	38.37	0.51	-
on		off	1	1	41.02	2.16	-	
			44	2	38.37	0.51	-	
		toggle	1	1	42.08	2.35	an	
			28	2	40.72	0.43	an	
		on		1	1	40.88	1.57	-
				40	1	38.97	1.38	-
off		off		1	1	40.14	1.51	-
				13	1	39.64	1.41	-
		toggle		1	1	43.13	1.50	an
				20	1	42.12	1.42	an
	on		1	1	43.27	2.17	an	
			35	3	41.31	1.36	an	
toggle	off		1	1	43.46	2.10	an	
			44	1	42.16	1.48	an	
	toggle		1	1	44.62	2.23	an/an	
			26	1	43.09	1.40	an/an	

*Fortsetzung auf der nächsten Seite*

## 5. Strukturelle Betrachtungen der Modellbindetasche

System	Wasser A	Wasser B	Rang	CR	GS	RMSD (Å)	Wasser?
1AEF (neu)		on	1	1	46.38	1.12	-
			2	1	46.33	0.45	-
	on	off	1	1	44.87	3.16	-
			40	2	42.75	0.90	-
		toggle	1	1	48.36	0.45	an
			41	1	47.28	0.41	an
		on	1	1	44.48	1.13	-
			33	1	43.08	0.94	-
	off	off	1	1	42.06	1.04	-
			45	1	39.12	0.80	-
		toggle	1	1	43.43	1.14	an
			39	1	41.48	0.93	an
		on	1	1	46.37	1.04	an
			29	1	45.34	0.42	an
toggle	off	1	1	44.32	1.61	an	
		42	1	42.45	0.88	an	
	toggle	1	1	48.61	0.47	an/an	
		13	1	48.06	0.43	an/an	
1AEG	on	-	1	1	40.99	0.78	-
			7	1	40.69	0.61	-
	off	-	1	1	39.73	0.86	-
			36	1	38.73	0.62	-
	toggle	-	1	1	41.00	0.87	an
			49	1	38.46	0.64	an
2AQD	on	-	1	1	40.07	0.35	-
			9	1	39.98	0.35	-
	off	-	1	1	42.23	0.40	-
			5	1	42.10	0.31	-
	toggle	-	1	1	47.17	0.42	an
			18	1	46.66	0.29	an

*Fortsetzung auf der nächsten Seite*



---

System	Wasser A	Wasser B	Rang	CR	GS	RMSD (Å)	Wasser?
2EUP	on	-	1	1	46.98	0.57	-
			6	1	46.78	0.40	-
	off	-	1	1	45.87	0.60	-
			26	1	45.52	0.42	-
	toggle	-	1	1	51.08	0.57	an
			39	1	50.51	0.43	an

---

## 5. Strukturelle Betrachtungen der Modellbindetasche

**Tab. 5.6.:** Ergebnisse des Redockings mittels GOLD 4.0 und der Scoring-Funktion Chemscore. Angeführt sind nur jene Lösungen, die am besten bewertet wurden und jene, deren RMSD-Werte am kleinsten sind. Die Abkürzung „CR“ steht für den Cluster Rang, die Abkürzung „CS“ für Chemscore. Alle Lösungen wurden bezogen auf ihre RMSD Abweichungen untereinander geclustert. Die Cluster wiederum sind nach der enthaltenen best-bewerteten (Score) Lösung geordnet. Die hier gezeigten RMSD Werte wurden zwischen der Docking-Lösung und der Kristallstruktur als Referenz bestimmt. In der letzten Spalte ist angegeben, ob für Wassermoleküle im „toggling“-Modus in der entsprechenden Lösung die Wechselwirkung an- oder ausgeschaltet wurde.

System	Wasser A	Wasser B	Rang	CR	CS	RMSD (Å)	Wasser?	
1AEO	on	-	1	1	18.18	0.44	-	
			38	1	17.56	0.29	-	
	off	-	1	1	23.43	0.46	-	
			27	1	23.15	0.30	-	
	toggle	-	1	1	23.48	0.44	aus	
			34	1	23.05	0.28	aus	
	1AEF		on	1	1	11.25	3.47	-
				43	2	10.24	3.40	-
on		off	1	1	18.13	3.68	-	
			50	1	16.46	3.64	-	
		toggle	1	1	18.08	3.68	aus	
			3	1	18.04	3.66	aus	
		on	1	1	17.37	3.48	-	
			45	1	15.91	3.37	-	
off		off	1	1	24.07	3.73	-	
			36	1	23.81	3.68	-	
		toggle	1	1	24.16	3.68	aus	
			48	1	23.75	3.66	aus	
		on	1	1	17.52	3.48	aus	
			28	1	16.81	3.40	aus	
toggle	off	1	1	24.25	3.68	aus		
		17	1	24.07	3.66	aus		
	toggle	1	1	24.25	3.68	aus/aus		
		47	1	23.89	3.64	aus/aus		

*Fortsetzung auf der nächsten Seite*

System	Wasser A	Wasser B	Rang	CR	GS	RMSD (Å)	Wasser?
1AEF (neu)		on	1	1	16.26	3.63	-
			49	3	14.44	0.64	-
	on	off	1	1	21.57	3.51	-
			30	1	20.63	3.48	-
		toggle	1	1	21.45	3.50	an
			47	3	20.15	3.36	an
		on	1	1	17.47	3.59	-
			9	2	16.93	1.13	-
	off	off	1	1	22.72	3.64	-
			32	1	22.11	3.46	-
		toggle	1	1	23.10	3.67	aus
			49	3	21.35	0.81	aus
		on	1	1	17.81	3.59	aus
			42	2	16.50	0.83	aus
toggle	off	1	1	22.96	3.58	aus	
		23	1	22.48	3.48	aus	
	toggle	1	1	23.05	3.61	aus/aus	
		6	1	22.91	3.48	aus/aus	
1AEG	on	-	1	1	16.68	3.57	-
			19	1	16.58	3.51	-
	off	-	1	1	22.79	3.57	-
			39	1	22.59	3.51	-
	toggle	-	1	1	16.68	3.57	aus
			40	1	16.54	3.52	aus
2AQD	on	-	1	1	18.77	0.45	-
			32	1	17.84	0.18	-
	off	-	1	1	23.16	0.30	-
			28	1	22.49	0.22	-
	toggle	-	1	1	23.20	0.34	an
			26	1	22.61	0.22	aus

*Fortsetzung auf der nächsten Seite*

## 5. Strukturelle Betrachtungen der Modellbindetasche

System	Wasser A	Wasser B	Rang	CR	GS	RMSD (Å)	Wasser?
2EUP	on	-	1	1	24.04	0.28	-
			10	1	23.89	0.24	-
	off	-	1	1	25.88	0.21	-
			15	1	25.73	0.16	-
	toggle	-	1	1	28.63	0.30	an
			2	1	28.60	0.18	aus

**Tab. 5.7.:** Ergebnisse des Redockings mit FlexX 3.1.4. Angeführt sind nur jene Lösungen, die am besten bewertet wurden und jene, deren RMSD-Werte am kleinsten sind. Die Abkürzung „FS“ steht für FlexX Score. Die hier gezeigten RMSD Werte wurden zwischen der Docking-Lösung und der Kristallstruktur als Referenz bestimmt. In der letzten Spalte ist angegeben, ob Wassermoleküle im „displacable“-Modus in der entsprechenden Lösung aus der Bindetasche verdrängt wurden oder in der Tasche verbleiben (ja).

System	Wasser A	Wasser B	Rang	FS	RMSD (Å)	Wasser?
1AEO	on	-	1	-16.99	0.37	-
	off	-	1	-16.88	0.37	-
	displacable	-	1	-16.88	0.37	ja
1AEF		on	1	-13.79	1.84	-
			246	-5.58	0.45	-
	on	off	1	-15.53	3.16	-
			363	-3.24	0.52	-
		displacable	1	-15.53	3.16	verdrängt
			354	-3.45	0.52	ja
		on	1	-13.56	1.84	-
			260	-5.46	0.45	-
	off	off	1	-15.33	3.16	-
			375	-3.15	0.52	-
		displacable	1	-15.33	3.16	verdrängt
			357	-3.36	0.52	ja
	on	1	-12.02	1.42	ja	
		2	-11.58	0.88	ja	
displacable	off	1	-15.33	3.16	ja	
		373	-3.16	0.52	ja	
	displacable	1	-11.58	0.88	ja/ja	

*Fortsetzung auf der nächsten Seite*

## 5. Strukturelle Betrachtungen der Modellbindetasche

System	Wasser A	Wasser B	Rang	FS	RMSD (Å)	Wasser?
1AEF (neu)		on	1	-17.71	1.41	-
				-16.29	0.52	-
	on	off	1	-14.78	0.60	-
			displaceable	1	-15.49	0.60
	off	on	1	-14.28	0.98	-
			2	-14.26	0.58	-
		off	1	-14.71	2.10	-
			3	-14.39	0.58	-
		displaceable	1	-13.36	0.58	ja
		on	1	-22.06	0.70	ja
	displaceable	off	1	-15.86	1.34	ja
			2	-15.36	0.58	ja
displaceable		1	-16.74	0.58	ja/ja	
1AEG	on	-	1	-11.81	2.78	-
			268	-4.35	0.75	-
	off	-	1	-12.79	2.78	-
			296	-4.26	0.75	-
	displaceable	-	1	-12.72	2.78	verdrängt
			301	-4.26	0.75	ja
2AQD	on	-	1	-18.20	0.49	-
	off	-	1	-17.90	0.49	-
	displaceable	-	1	-17.90	0.49	ja

*Fortsetzung auf der nächsten Seite*

---

System	Wasser A	Wasser B	Rang	FS	RMSD (Å)	Wasser?
2EUP	on	-	1	-19.57	0.48	-
	off	-	1	-19.58	0.48	-
			2	-15.84	0.46	-
	displacable	-	1	-19.58	0.48	ja
			2	-15.84	0.46	ja

---

## 5.2.4. Ergebnisse des Dockings in die ligandfreie Bindetasche

### Kristallstruktur 1CPE als Rezeptor

Ein weitaus realistischerer Testfall im Hinblick auf ein virtuelles Screening soll im Folgenden diskutiert werden. Dabei soll die Wasserverdrängung bei der Platzierung der fünf Liganden in die ligandfreie Bindetasche, die mit mehreren Wassermolekülen gefüllt ist, vorhergesagt werden. Während des virtuellen Screenings werden für gewöhnlich nur wenige, sehr gut bewertete Docking-Lösungen einer genaueren Auswertung unterzogen. Aus diesem Grund wurde im Folgenden der Fokus bei der Auswertung der Ergebnisse ebenfalls hauptsächlich auf die best-bewerteten Lösungen des Dockings in die 1CPE-Bindetasche gelegt. Hierbei befanden sich alle fünf Wassermoleküle im „toggling“ bzw. „displaceable“-Modus. Zusätzlich wurde das  $K^+$ -Ion aus der Tasche entfernt. Das jeweilige Docking-Programm hatte also die anspruchsvolle Aufgabe, nicht nur die optimale Positionierung des Liganden in der Bindetasche herauszufinden. Zudem galt es, das Optimierungsproblem zu lösen, welche der fünf Wassermoleküle in der Bindetasche verbleiben sollen, um Wechselwirkungen mit Ligand und Protein einzugehen. Zur Überprüfung der Güte einer Docking-Lösung wurde das Proteingerüst der Kristallstrukturen der Komplexe mit der Struktur 1CPE überlagert und die Ligandkoordinaten als Referenz für die RMSD-Berechnung verwendet.

Eine Übersicht der Resultate dieser Docking-Simulationen findet sich in Tab. 5.8. Die best-bewerteten Lösungen von FlexX und GOLD mit GoldScore sind zusammen mit der Kristallstruktur-Referenz in Abb. 5.6 dargestellt. In diese Abbildung wurden keine Lösungen aufgenommen, die mit der *Chemscore* Scoring-Funktion erhalten wurden, da hier in allen Fällen die Wasserwechselwirkungen ausgeschaltet wurden. Die richtige Vorhersage der Wasserverdrängung stellt im vorliegenden Beispiel ja ein entscheidendes Kriterium dar. Mit GOLD in Kombination mit Chemscore konnte für die Liganden 2-Aminopyridin, 2,5-Diaminopyridin und 2-Amino-5-picolin ein sehr gutes Ergebnis erzielt werden, während die Vorhersage der Bindungsmodi bei 3- und 4-Aminopyridin sich als völlig unzureichend erwies. Der Bindungsmodus von *2-Aminopyridin* wird durch GOLD/GoldScore weitgehend richtig vorhergesagt. Zusätzlich zum vorhandenen Wassermolekül A wird auch Wasser 976 als präsent prognostiziert. Auch das FlexX Ergebnis ist hier (wie in fast allen Fällen) korrekt. Der Bindungsmodus von *3-Aminopyridin* wird weder durch GOLD/GoldScore noch durch FlexX richtig vorhergesagt. Im Gegensatz dazu haben für *4-Aminopyridin* die Lösungen mit höchstem GoldScore und niedrigstem Flexscore die richtige Orientierung. Auch hier wird in beiden Fällen ein experimentell unbeobachtetes Wassermolekül in der Bindetasche belassen. Die Orientierung von *2,5-Diaminopyridin* wird sehr gut von FlexX und etwas schlechter von GOLD/GoldScore vorhergesagt. FlexX bestimmt außerdem das Wassernetzwerk völlig korrekt\*. Bei GOLD findet sich erneut zusätzlich zu Wasser A auch Wasser 976. Schließlich ist im Falle von *2-Amino-5-picolin* die Vorhersage beider Docking-Methoden erfolgreich, wenn auch nur FlexX die Wasserverdrängung korrekt bestimmt.

---

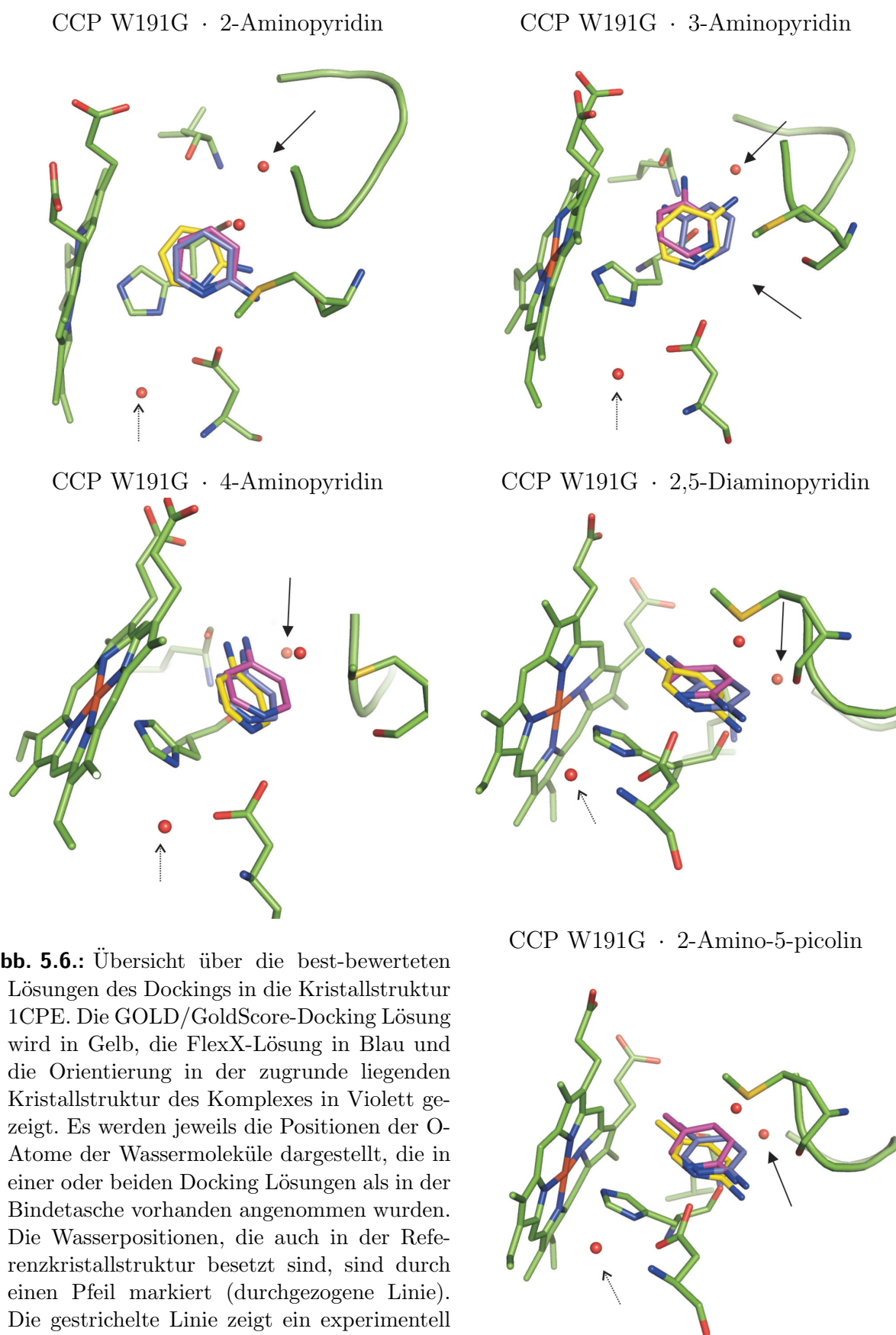
\*Hier wie in allen anderen Beispielen wird Wasser 979 von FlexX im eigentlichen Sinne nicht verdrängt. Das Programm nimmt dieses Wassermolekül bereits von Beginn an aus der Berechnung heraus, da keine Wechselwirkungen zum Protein vorliegen. Für die hier diskutierten Fragestellungen ist diese Behandlung jedoch auch angebracht.



Aus dem oben bereits diskutierten Grund ist eine falsche Vorhersage des Wassernetzwerkes während des virtuellen Screening-Prozesses unter Umständen irreführend und deshalb unerwünscht. Generell kann aufgrund dieser insgesamt (insbesondere aber von FlexX) sehr guten Ergebnisse festgehalten werden, dass für fast jedes System ein oder mehrere Docking-Methoden die richtige Lösung finden konnten und auch auf Rang 1 bei der Bewertung setzten. Einzige Ausnahme bildet der Ligand 3-Aminopyridin, und es stellt sich die Frage, was die Ursache für dieses Ergebnis ist. Im Re-Docking wurde eine deutliche Abhängigkeit für das Auffinden einer korrekten Lösung von der exakten Rezeptorstruktur beobachtet. Dies kann erst recht beim Docking in die nicht native Rezeptorstruktur relevant sein. Möglicherweise müssten hier Bewegungen der Seitenketten oder Schleifen-Regionen im Rahmen des „induced fit“ Konzeptes berücksichtigt werden. Da in den hier verwendeten Programmen kein automatisiertes Protokoll dafür zur Verfügung steht, müssten einzelne flexible Seitenketten definiert werden. Dies widerspräche aber dem Gedanken dieser Untersuchung, in der angenommen wurde, dass keine Informationen über die Komplexstrukturen sondern nur über die ligandfreie Form vorliegen. Auffällig ist außerdem die immer wieder beobachtete Vorhersage eines zusätzlichen Wassermoleküls (Nummer 976) in der Bindetasche. Dies wird also nach Vorhersage der Docking-Programme weder aus energetischen noch aus sterischen Gründen verdrängt, ist aber dennoch nicht in den Kristallstrukturen beobachtet worden. Ist dies ein Hinweis auf ein Defizit in den Kristallstrukturen oder in den Docking-Algorithmen? Vielleicht geben verschiedene, strukturelle Deskriptoren für dieses Wassermolekül eine Erklärung für dessen Besonderheit (vergleiche hierzu Abschnitt 5.3.3). Ebenso könnte auch die MD-Simulation helfen, Unzulänglichkeiten der Kristallstruktur aufzuklären (vergleiche hierzu Abschnitt 6.2.2). Abschließend ist zu bedenken, dass 1CPE möglicherweise keine ideale Referenz für eine ligandfreie Bindetasche darstellt, denn das in der Kristallstruktur vorhandene  $K^+$ -Ion beeinflusst selbstverständlich die Positionen der Wassermoleküle in der Bindetasche.

### **Ionenfreie Modellbindetasche der CCP W191G**

Da aus genau diesem Grund ein Modell einer ionenfreien Bindetasche der CCP W191G erstellt wurde, wurde auch ein Docking in diese Rezeptorstruktur durchgeführt. Hierbei wurden alle sieben Wassermoleküle innerhalb der Bindetasche in den „toggle“ bzw. „displaceable“ Modus gesetzt. Diese Rechnung war somit kombinatorisch noch anspruchsvoller als die vorherige und dementsprechend teurer. Um die Güte einer Docking-Lösung zu überprüfen, wurde auch hier das Proteingerüst der Kristallstrukturen der Komplexe mit der Modell-Struktur überlagert und die daraus resultierenden Ligandkoordinaten als Referenz für die RMSD-Berechnung verwendet. Die Ergebnisse dieser Simulation sind in Tab. 5.9 dargestellt. Zudem wurde überprüft, wie weit die Docking-Programme für den gegebenen Rezeptor das erwünschte Ergebnis überhaupt erreichen können. Dafür wurden alle Wassermoleküle bis auf Wasser A (bzw. zusätzlich Wasser B beim Liganden 3-Aminopyridin) aus dem Rezeptor entfernt. Mit diesen Wassermolekülen im eingeschalteten Zustand (kein „toggle“ bzw. „displaceable“ Modus !) wurde das Docking wie oben beschrieben erneut durchgeführt. Die Ergebnisse hiervon sind in Tab. 5.10 dargestellt.



**Abb. 5.6.:** Übersicht über die best-bewerteten Lösungen des Dockings in die Kristallstruktur 1CPE. Die GOLD/GoldScore-Docking Lösung wird in Gelb, die FlexX-Lösung in Blau und die Orientierung in der zugrunde liegenden Kristallstruktur des Komplexes in Violett gezeigt. Es werden jeweils die Positionen der O-Atome der Wassermoleküle dargestellt, die in einer oder beiden Docking Lösungen als in der Bindetasche vorhanden angenommen wurden. Die Wasserpositionen, die auch in der Referenzkristallstruktur besetzt sind, sind durch einen Pfeil markiert (durchgezogene Linie). Die gestrichelte Linie zeigt ein experimentell verifiziertes Wassermolekül, das im Docking nicht untersucht wurde.

Die Resultate des Docking bei „freier“ Wasserauswahl sind insgesamt nicht zufriedenstellend. Zum einen sind die strukturellen Abweichungen des Bindungsmodus vieler GOLD Lösungen sehr groß. Zum anderen zeigt jede der best-bewerteten wie auch der strukturell besten Lösungen ein oder mehrere Wassermoleküle zu viel in der Bindetasche. Hierbei handelt es sich auf jeden Fall immer um das Molekül Nr. 305, für das es in der 1CPE-Struktur keine Entsprechung gibt. Für manche Systeme bzw. durch manche Docking-Programme werden jedoch strukturell korrekte Lösungen gefunden. Die entsprechende mit der geringsten Abweichung davon ist ebenfalls in Tab. 5.9 dargestellt. Oftmals sind diese Lösungen aber deutlich schlechter bewertet, als die auf Rang 1, finden sich gar in einem ganz anderen Cluster, als die Lösung mit dem besten Score und zeigen ebenfalls zu viele Wassermoleküle in der Bindetasche. Für manche Kombinationen aus Dockingprogramm und Scoring-Funktion kann keine Lösung mit einem RMSD-Wert unter 1 Å gefunden werden. Dies ist bei 2-Aminopyridin und 2,5-Diaminopyridin mit GOLD/GoldScore sowie bei 3-Aminopyridin mit GOLD unabhängig von der Scoring-Funktion der Fall.

Setzt man nun die Wassermoleküle so, wie man sie erwarten würde (vgl. Ergebnisse in Tab. 5.10), werden die Ergebnisse zum Teil besser, zum Teil schlechter. Betrachtet man ausschließlich die best-bewerteten Lösungen, so

- verbessert sich das Ergebnis bei: 2AP, 4AP, DA1, LG4 mit GOLD/GoldScore sowie DA1 mit FlexX,
- verschlechtert sich das Ergebnis bei: 3AP mit GOLD/GoldScore sowie 2AP, 4AP mit GOLD/Chemscore,
- bleibt das Ergebnis weitgehend gleich bei: 3AP, DA1, LG4 mit GOLD/Chemscore sowie 2AP, 3AP, 4AP, LG4 mit FlexX.

Insgesamt zeigen alle Lösungen von FlexX (in Tab. 5.10) einen RMSD kleiner 1.25 Å. Die Vorhersagekraft des korrekten Bindungsmodus scheint bei FlexX in den hier aufgeführten Fällen relativ unabhängig davon, ob die Anwesenheit der Wassermoleküle im Voraus definiert wurde oder nicht. Die Kombination GOLD/GoldScore ist offenbar prinzipiell in der Lage, deutlich bessere Ergebnisse zu erzielen als bei voller Freiheit über die Verdrängung der Bindetaschenwassermoleküle. Beim Einsatz von GOLD zusammen mit Chemscore führt das Eingreifen durch den Nutzer bei der Wahl der Wassermoleküle umgekehrt zu einer leichten Verschlechterung des Ergebnisses. Zumindest bei der Kombination GOLD/GoldScore könnte es sich als sehr nützlich herausstellen, wenn das Optimierungsproblem bei der Wahl, aus sieben Möglichkeiten Wasserwechselwirkungen ein- oder auszuschalten, verringert werden könnte. Eine Möglichkeit dafür bietet eine Vorabauswahl von Wassermolekülen. An Hand unterschiedlicher Deskriptoren für Wassermoleküle kann versucht werden, bereits vor dem Docking zu entscheiden, welche Wassermoleküle sehr wahrscheinlich durch den Liganden verdrängt werden und im Docking nicht berücksichtigt werden müssen. Diese Deskriptoren können entweder aus Struktur- oder Dynamikinformatoren abgeleitet werden. Im folgenden Abschnitt wird darauf näher eingegangen.

## 5. Strukturelle Betrachtungen der Modellbindetasche

---

Um zu überprüfen, inwieweit Wassermoleküle überhaupt notwendig sind, um korrekte Ergebnisse zu erhalten, wurde auch ein Docking in eine vollständig wasserfreie Modellbindetasche durchgeführt. Bei Betrachtung der best-bewerteten Lösungen zeigt sich eine

- Verbesserung des Ergebnisses für 2AP und 4AP mit GOLD/Goldscore, 4AP mit FlexX, DA1 für GOLD/Chemscore und FlexX sowie für LG4,
- ein annähernd gleiches Ergebnis für 2AP, 3AP und 4AP mit GOLD/Chemscore sowie 2AP mit FlexX,
- eine Verschlechterung bei 3AP mit GOLD/Goldscore und FlexX sowie DA1 und GOLD/Goldscore.

Auch hier bestätigt sich der bereits beobachtete Trend, dass die Bewertung von Chemscore weitgehend unabhängig vom Vorhandensein von Wassermolekülen ist. Goldscore reagiert deutlich sensitiver auf diesen strukturellen Unterschied des Rezeptors. Es wird deutlich, dass Wassermoleküle zum Teil notwendig sind, um korrekte Bindungsmodi vorherzusagen. Gleichzeitig ist aber auch das Vorhandensein oder die Abwesenheit des Wassermoleküls an sich von Interesse, um beim Wirkstoffdesign das Wechselwirkungsmuster optimieren zu können. Die zum Teil beobachtete Verbesserung und Verschlechterung des Ergebnisses weist erneut darauf hin, dass die Parametrisierung der Wassereinflüsse noch weiter optimiert werden kann.

### Zusammenfassung

In Hinblick auf das Docking in die ligandfreie Tasche zeigten sich beide Dockingprogramme GOLD und FlexX als gleichermaßen geeignet/ungeeignet. Insgesamt wurden durchweg strukturell sehr gute Lösungen erhalten. Durchweg lässt aber die Vorhersage der Wasserverdrängung weiter Raum zur Verbesserung. Ist man speziell daran interessiert, kann der Einsatz von GOLD mit Chemscore an dieser Stelle nicht empfohlen werden. FlexX liefert strukturell sehr gute Lösungen und ist auch aufgrund seiner Geschwindigkeit für dieses System sehr zu empfehlen. Die Ergebnisse des Modells einer ionenfreien Bindetasche konnten letztlich nur unter Vorbehalt diskutiert werden. Es ist nicht bewiesen, dass dieses Strukturmodell so korrekt ist. Es gibt vielmehr den begründeten Verdacht, dass keine stabile Form einer ionenfreien CCP W191G-Tasche mit geschlossener PGGAAN-Schleifen-Konformation vorliegt und dementsprechend auch kein Modell erstellt werden kann. Eine geschlossene Loop-Konformation ist aber entscheidendes Kriterium, um den Zustand der Komplexe modellieren zu können. Mehr dazu findet man in Abschnitt 6.5.

**Tab. 5.8.:** Ergebnisse (best-bewertete Lösungen sowie solche mit kleinstem RMSD-Wert) des Dockings mit GOLD 4.0 und FlexX 3.1.4 in die K<sup>+</sup>-freie Bindetasche von 1CPE. Die Wechselwirkungen aller fünf Wassermoleküle können durch das Programm GOLD an- oder ausgeschaltet werden. FlexX ist in der Lage alle fünf Wassermoleküle zu verdrängen („v“). Die verwendete Abkürzung „CR“ bedeutet Cluster Rang, „GS“ steht für GoldScore, „CS“ für Chemscore und „FS“ für FlexX Score. Alle Lösungen wurden bezogen auf ihre RMSD Abweichungen untereinander einem Clustering unterzogen. Die Cluster wiederum sind nach der enthaltenen best-bewerteten (Score) Lösung geordnet. Die hier gezeigten RMSD Werte wurden zwischen der Docking-Lösung und der Kristallstruktur als Referenz bestimmt. In der jeweiligen Rezeptorstruktur findet sich das Pendant zu Wassermolekül 477 (Position A) und zusätzlich beim Liganden 3-Aminopyridin auch das von Wassermolekül 977 (Position B). Wassermoleküle, die vorhanden sein sollten, sind in der Tabelle fett gedruckt. Wassermolekül 979 wurde nicht durch das Program FlexX verdrängt, sondern bereits im Vorfeld von der Berechnung ausgenommen, weil keine Wechselwirkungen mit dem Protein vorliegen.

Ligand	Aus Komplex	Software	Rang	CR	GS/CS/FS	RMSD (Å)	Wassermoleküle: 477(A)/976/977(B)/978/979
2AP	1AEO	GOLD/GS	1	1	49.91	1.42	<b>an</b> /an/aus/aus/aus
			43	1	46.58	1.10	<b>an</b> /aus/aus/aus/aus
		GOLD/CS	1	1	23.53	0.37	<b>aus</b> /aus/aus/aus/aus
			39	1	23.28	0.23	<b>aus</b> /aus/aus/aus/aus
		FlexX	1	1	-16.45	0.36	<b>ja</b> /v/v/v/-
			2	2	-15.18	0.34	<b>ja</b> /v/v/v/-
3AP	1AEF	GOLD/GS	1	1	45.73	2.10	<b>an</b> /aus/ <b>aus</b> /aus/aus
			47	1	44.52	1.69	<b>an</b> /aus/ <b>aus</b> /aus/aus
		GOLD/CS	1	1	23.23	3.57	<b>aus</b> /aus/ <b>aus</b> /aus/aus
			46	1	22.92	3.52	<b>aus</b> /aus/ <b>aus</b> /aus/aus
		FlexX	1	1	-14.49	2.01	<b>ja</b> /v/ <b>v</b> /v/-
			35	25	-9.82	0.74	<b>ja</b> /v/ <b>v</b> /v/-

*Fortsetzung auf der nächsten Seite*

Ligand	Aus Komplex	Software	Rang	CR	GS/CS/FS	RMSD (Å)	Wassermoleküle: 477(A)/976/977(B)/978/979
4AP	1AEG	GOLD/GS	1	1	47.71	1.01	<b>an</b> /an/aus/aus/aus
			40	1	46.39	0.79	<b>an</b> /an/aus/aus/aus
		GOLD/CS	1	1	23.02	3.51	<b>aus</b> /aus/aus/aus/aus
			31	1	22.79	3.51	<b>aus</b> /aus/aus/aus/aus
		FlexX	1	1	-15.55	0.58	<b>ja</b> /ja/v/v/-
			22	14	-10.26	0.50	<b>ja</b> /v/v/v/-
DA1	2AQD	GOLD/GS	1	1	46.83	1.39	<b>an</b> /an/aus/aus/aus
			39	1	44.86	1.01	<b>an</b> /aus/aus/aus/aus
		GOLD/CS	1	1	23.03	0.36	<b>aus</b> /aus/aus/aus/aus
			46	1	22.68	0.26	<b>aus</b> /aus/aus/aus/aus
		FlexX	1	1	-17.27	0.95	<b>ja</b> /v/v/v/-
			12	12	-15.08	0.60	<b>ja</b> /v/v/v/-
LG4	2EUP	GOLD/GS	1	1	47.98	1.10	<b>an</b> /an/aus/aus/aus
			50	1	44.70	0.67	<b>an</b> /aus/aus/aus/aus
		GOLD/CS	1	1	25.65	0.37	<b>aus</b> /aus/aus/aus/aus
			37	1	25.04	0.29	<b>aus</b> /aus/aus/aus/aus
		FlexX	1	1	-16.91	0.71	<b>ja</b> /v/v/v/-
			7	7	-13.31	0.70	<b>ja</b> /v/v/v/-

**Tab. 5.9.:** Ergebnisse (best-bewertete Lösungen sowie solche mit kleinstem RMSD-Wert) des Dockings mit GOLD 4.0 und FlexX 3.1.4 in die ionenfreie Modellbindetasche von CCP W191G. Die Wechselwirkungen aller sieben Wassermoleküle können durch das Program GOLD an- oder ausgeschaltet werden. FlexX ist in der Lage alle sieben Wassermoleküle zu verdrängen („v“). Die verwendete Abkürzung „CR“ bedeutet Cluster Rang, „GS“ steht für GoldScore, „CS“ für Chemscore und „FS“ für FlexX Score. Alle Lösungen wurden bezogen auf ihre RMSD Abweichungen untereinander einem Clustering unterzogen. Die Cluster wiederum sind nach der enthaltenen best-bewerteten (Score) Lösung geordnet. Die hier gezeigten RMSD Werte wurden zwischen der Docking-Lösung und der Kristallstruktur als Referenz bestimmt. In der jeweiligen Rezeptorstruktur findet sich das Pendant zu Wassermolekül 1573 (Position A) und zusätzlich beim Liganden 3-Aminopyridin auch das von Wassermolekül 2238 (Position B). Wassermoleküle, die vorhanden sein sollten, sind in der Tabelle fett gedruckt. Wassermolekül 307 wurde nicht durch das Program FlexX verdrängt, sondern bereits im Vorfeld von der Berechnung ausgenommen, weil keine Wechselwirkungen mit dem Protein vorliegen.

Ligand	Aus Komplex	Software	Rang	CR	GS/CS/FS	RMSD (Å)	Wassermoleküle:
							305/308/1573(B)/2238(A)/5226/6667/307
2AP	1AEO	GOLD/GS	1	1	48.14	2.74	an/aus/aus/ <b>an</b> /an/aus/aus
			47	2	45.45	1.05	an/aus/aus/ <b>an</b> /aus/aus/aus
		GOLD/CS	1	1	27.19	2.54	an/aus/aus/ <b>an</b> /aus/aus/aus
			FlexX	1	1	-16.69	0.68
3AP	1AEF	GOLD/GS	1	1	45.58	1.88	an/aus/ <b>an</b> / <b>an</b> /an/aus/aus
			43	2	42.37	1.66	an/aus/ <b>aus</b> / <b>an</b> /aus/aus/aus
		GOLD/CS	1	1	29.76	1.57	an/aus/ <b>aus</b> / <b>an</b> /aus/aus/aus
			11	1	29.19	1.40	an/aus/ <b>aus</b> / <b>an</b> /aus/aus/aus
		FlexX	1	1	-16.66	1.24	ja/v/ <b>ja</b> / <b>ja</b> /ja/v/-
			3	3	-14.64	0.92	ja/v/ <b>v</b> / <b>ja</b> /ja/v/-

*Fortsetzung auf der nächsten Seite*

Ligand	Aus Komplex	Software	Rang	CR	GS/CS/FS	RMSD (Å)	Wassermoleküle: 305/308/1573(B)/2238(A)/5226/6667/307
4AP	1AEG	GOLD/GS	1	1	44.58	1.47	an/aus/an/ <b>an</b> /an/aus/aus
			43	1	42.88	0.40	an/aus/aus/ <b>an</b> /aus/aus/aus
		GOLD/CS	1	1	28.39	1.00	an/aus/aus/ <b>an</b> /aus/aus/aus
			49	1	27.21	0.78	an/aus/aus/ <b>an</b> /aus/aus/aus
		FlexX	1	1	-18.48	1.17	ja/v/v/ <b>ja</b> /v/v/-
			7	5	-10.69	0.73	ja/v/v/ <b>ja</b> /v/v/-
DA1	2AQD	GOLD/GS	1	1	46.59	2.50	an/aus/aus/ <b>an</b> /an/aus/aus
			21	1	45.00	1.61	an/aus/aus/ <b>an</b> /aus/aus/aus
		GOLD/CS	1	1	25.63	4.75	an/aus/aus/ <b>an</b> /aus/aus/aus
			7	2	25.43	0.96	an/aus/aus/ <b>an</b> /aus/aus/aus
		FlexX	1	1	-18.69	3.28	ja/v/v/ <b>ja</b> /v/v/-
			3	3	-15.37	0.86	ja/v/v/ <b>ja</b> /v/v/-
LG4	2EUP	GOLD/GS	1	1	48.40	7.37	aus/an/aus/ <b>aus</b> /an/an/an
			21	3	45.03	0.94	an/aus/aus/ <b>an</b> /aus/aus/aus
		GOLD/CS	1	1	28.53	2.35	an/aus/aus/ <b>an</b> /aus/aus/aus
			50	1	27.33	0.96	an/aus/aus/ <b>an</b> /aus/aus/aus
		FlexX	1	1	-19.33	1.12	ja/v/v/ <b>ja</b> /v/v/-
			27	23	-11.06	0.60	ja/v/v/ <b>ja</b> /ja/v/-



**Tab. 5.10.:** Ergebnisse (best-bewertete Lösungen sowie solche mit kleinstem RMSD-Wert) des Dockings mit GOLD 4.0 und FlexX 3.1.4 in die ionenfreie Modellbindetasche von CCP W191G. Wassermoleküle wurden im Vorfeld so gesetzt, wie sie in der Referenzkristallstruktur des Komplexes beobachtet wurden. Die verwendete Abkürzung „CR“ bedeutet Cluster Rang, „GS“ steht für GoldScore, „CS“ für Chemscore und „FS“ für FlexX Score. Alle Lösungen wurden bezogen auf ihre RMSD Abweichungen untereinander einem Clustering unterzogen. Die Cluster wiederum sind nach der enthaltenen best-bewerteten (Score) Lösung geordnet. Die hier gezeigten RMSD Werte wurden zwischen der Docking-Lösung und der Kristallstruktur als Referenz bestimmt.

Ligand	Aus Komplex	Software	Rang	CR	GS/CS/FS	RMSD (Å)
2AP	1AEO	GOLD/GS	1	1	40.58	1.23
			45	1	39.12	0.70
		GOLD/CS	1	1	20.44	3.37
			48	3	19.27	0.42
		FlexX	1	1	-16.86	0.68
		3AP	1AEF	GOLD/GS	1	1
12	1				41.73	2.19
GOLD/CS	1			1	19.14	1.51
	20			1	17.98	1.37
FlexX	1			1	-16.41	1.24
	203			60	-6.44	0.66
4AP	1AEG	GOLD/GS	1	1	39.56	0.69
			35	1	38.50	0.43
		GOLD/CS	1	1	22.50	1.39
			39	1	20.67	0.80
		FlexX	1	1	-15.82	1.17
			4	1	-10.69	0.73
DA1	2AQD	GOLD/GS	1	1	39.36	1.43
			23	2	37.82	0.88
		GOLD/CS	1	1	19.57	4.75
			48	4	17.64	0.83
		FlexX	1	1	-17.34	1.15
			3	3	-15.71	0.86

*Fortsetzung auf der nächsten Seite*

Ligand	Aus Komplex	Software	Rang	CR	GS/CS/FS	RMSD (Å)
LG4	2EUP	GOLD/GS	1	1	40.47	1.22
			49	1	37.43	0.80
		GOLD/CS	1	1	21.35	2.42
			30	2	20.19	0.66
		FlexX	1	1	-19.67	1.12
			23	17	-11.19	0.60

### 5.3. Weitergehende strukturelle Untersuchungen der Komplexe

#### 5.3.1. Deskriptoren der Datenbank Relibase+

Im Zusammenhang mit der Frage nach den Unterschieden zwischen den einzelnen Wassermolekülen in den vorausgegangenen Docking-Experimenten, ist es hilfreich, einen detaillierteren Blick auf strukturelle Deskriptoren zu werfen, die für Wassermoleküle berechnet werden können. Hierfür wurde auf die Datenbank Relibase+<sup>179-181</sup> und die im Modul WaterBase<sup>182-184</sup> implementierten Größen (vgl. Abschnitt 4.1) zurückgegriffen.

Die Werte wurden aus der Datenbank für die Komplexe 1AEO, 1AEF, 1AEG, 2AQD, 2EUP, die neu kristallisierte 1AEF-Struktur sowie für 1CPE und das Modell der ligandfreien CCP W191G Bindetasche entnommen und in Tab. 5.12 eingetragen. Vergleicht man zunächst nur die Werte der vier Komplexe miteinander, wird deutlich, dass die Mobilität von Wasser A im Schnitt deutlich größer (zumindest aber ähnlich) ist wie die von Wasser C. Beide Strukturen des Komplexes 1AEF unterscheiden sich deutlich voneinander bezüglich der Werte für die Wasserdiskriptoren. Aufgrund der höheren Auflösung und der vollständigen Verfeinerung der neuen Struktur soll im Folgenden diese zum Vergleich herangezogen werden. Wasser B zeigt hier eine ähnlich niedrige Mobilität in der gleichen Größenordnung wie Wasser A und C. Außer Wasser B in der Original 1AEF-Struktur hat gemäß des Distanz-Kriteriums keines der Wassermoleküle einen polaren Kontakt zum Liganden. Die höchste Zahl an Protein-Kontakten (4) zeigt Wasser A in Komplex 2AQD und 2EUP. Auch bei der Polarität kehren sich die Trends bei der neuen 1AEF-Struktur im Vergleich zu den anderen Komplexen um. Während dort Wasser A stets die höchste Polarität zeigt, ist dies für 1AEF (neu) bei Wasser B der Fall. Wasser A zeigt hier wiederum den geringsten Wert. Für die DrugScore Werte ist hingegen (auch im Hinblick auf die Aufteilung in Einzelbeiträge von Protein-, Ligand- und Solvens-Atomen) kein eindeutiger Trend erkennbar. Hierbei gilt es unter anderem zu beachten, dass die Einzelbeiträge des „DrugScore-Ligand“ Terms auch die Wechselwirkungen zum Cofaktor beinhalten. Das ist insbesondere beim Vergleich der ligandfreien Strukturen (siehe unten) offensichtlich. Die umgebende Atomdichte ist außer in 1AEF (neu) in der ersten Koordinationssphäre für

Wasser A am höchsten. In der zweiten Koordinationssphäre liegen die Werte für Wasser C höher. Der gleiche Trend wie für den AD1-Deskriptor spiegelt sich auch bei der solvenszugänglichen-Oberfläche wider. Hier hat in 1AEF (neu) der Deskriptor für Wasser A auch den höheren Wert. Eine oktaedrische Koordinationsgeometrie konnte für keines der Wassermoleküle beobachtet werden. Annähernd tetraedrisch sind hingegen Wasser A in 2AQD und 2EUP koordiniert.

Analog können nun die beiden ligandfreien Strukturen untereinander und mit den Komplexen verglichen werden. Für die Konformation, die aus der MD-Simulation erhalten wurde, wurden keine B-Faktoren und somit auch keine Mobilitäten ermittelt. Die Zahl der Kontakte ist hier für die Wassermoleküle, verursacht durch die häufigeren Wasser-Wasser-Kontakte, im Schnitt höher als in 1CPE. In beiden Strukturen hat (wie auch in den meisten Komplexen) Wasser A die Umgebung mit der höchsten Polarität. Bei den DrugScore Werten heben sich zwei Wassermoleküle besonders ab. Wasser 979 in 1CPE hat nur einen sehr wenig negativen Gesamtwert, was in dem positiven Score zum Protein begründet liegt. Wasser 307 aus der Modellstruktur hat sogar insgesamt einen positiven DrugScore. Auch hier ist der Wert für die Protein-Atome die Ursache dafür. In beiden Fällen handelt es sich hierbei um das Wassermolekül direkt neben dem Cofaktor über His175. In diesen zwei Strukturen können weiterhin sehr niedrige Werte für die umgebende Atomdichte ( $\approx 1$ ) beobachtet werden. Das Maß für die Vergrabenheit einzelner Wassermoleküle, die solvenszugängliche-Oberfläche, zeigt insbesondere bei der Modellstruktur recht hohe Werte (über 50%) und somit eine geringe Vergrabenheit dieses Moleküls.

### 5.3.2. B-Faktoren

Die thermische Bewegung eines jeden Atoms einer Kristallstruktur wird durch den so genannten B-Faktor  $B$  beschrieben. Dieser berechnet sich unmittelbar aus der mittleren Auslenkung  $U$  (in Å) eines schwingenden Atoms.

$$B = 8\pi^2 U^2. \quad (5.1)$$

Der B-Faktor beschreibt also die Größe der Oszillation um die wahrscheinlichste Position eines Atoms. Gleichzeitig kann aber auch die statische Unordnung des Kristalls den Wert beeinflussen. Der Werte hängen jedoch auch sehr stark von der Verfeinerungsmethode der Kristallstruktur ab<sup>201,202</sup>. Aus diesem Grund sind B-Faktoren unterschiedlicher Kristallstrukturen nicht direkt miteinander vergleichbar. Um dies zu ermöglichen, können die B-Faktoren normalisiert werden. Wang *et al.*<sup>13</sup> führen das mit dem Mittelwert aller B-Faktoren  $\langle B \rangle$  und dessen Standardabweichung  $\sigma(B)$  durch.

$$B_{\text{norm}} = \frac{B - \langle B \rangle}{\sigma(B)} \quad (5.2)$$

In dieser Arbeit wurden für die Normalisierung nur die Werte der B-Faktoren aller Protein-, Ligand- und Cofaktor-Atome (d.h. nicht der Wassermoleküle) herangezogen. Die Ergebnis-

se dieser Standardisierung finden sich in Tab. 5.11. Neben den Werten für die untersuchten Wassermoleküle A und B wurden, zwecks Vergleich, auch die für Wasser C, Ligand und Häm aufgelistet. Werte kleiner Null stehen dementsprechend für eine Beweglichkeit des Atoms, die niedriger ist als der Durchschnitt. Man sieht, dass im Durchschnitt der Ligand, im Speziellen aber der Cofaktor, eine sehr niedrige Beweglichkeit zeigt. Ausnahme bildet hierbei nur das  $K^+$ -Ion, das im Falle von 1CPE als Ligand deklariert wurde, sowie 3-Aminopyridin in der neuen 1AEF-Struktur. Bei den Wassermolekülen hingegen zeichnet sich kein so eindeutiger Trend ab. Die normalisierten B-Faktoren der Wassermoleküle A, B und C in Komplex 1CPE zeigen durchweg eine relativ niedrige Beweglichkeit. In den Komplexen weist nur Wasser A für 2EUP und Wasser C bei 1AEO, 1AEF und 1AEG eine sehr niedrige Beweglichkeit auf. Die normalisierten B-Faktor-Werte liegen hier zum Teil noch unter denen für den Cofaktor. Im Gegensatz dazu findet sich der höchste Wert (1.68) für Wasser A in 2AQD. Die hohe Beweglichkeit dieses Wassermoleküls findet sich auch im Zuge der durchgeführten MD-Simulationen (siehe Abschnitt 6.2.2). Auch bei dem hier dargestellten Deskriptor beobachtet man wieder unterschiedliche Relationen beim Vergleich von 1AEF und 1AEF (neu). Aufgrund der vollständigen Verfeinerung sind die Ergebnisse von 1AEF (neu) sehr wahrscheinlich aussagekräftiger. Das bedeutet, dass Wasser B eine höhere thermische Beweglichkeit als Wasser A oder C zeigt und damit nicht so rigide wie Wasser A an Protein und Ligand gebunden ist. Ist das möglicherweise ein Hinweis darauf, weshalb Wasser B in keinem anderen Komplex zu finden ist? Ist die hohe Beweglichkeit gleichzusetzen mit einer leichten Verdrängbarkeit? Dieses Ergebnis wird jedoch, wie bereits erwähnt, nicht vom Wert in 1CPE untermauert. Möglicherweise wird diese Position erst durch Anwesenheit eines Liganden thermodynamisch ungünstig.

### 5.3.3. Zusammenfassung

Aus der Fülle an Daten über einzelne Wassermoleküle, die allein aus der Analyse der experimentellen Strukturdaten vorliegen, konnte keine sichere Aussage zu deren Verdrängbarkeit bzw. Konservierung bei Ligandbindung in die CCP W191G Tasche getroffen werden. Zwar zeichnet sich allein durch die Konservierung von Wasser A in fast allen Kristallstrukturen der CCP W191G dessen Bedeutung ab. Ebenso deuten die schlechten DrugScore Werte sowie sehr geringe Kontaktzahl von Wasser 979 (1CPE) auf dessen Verdrängbarkeit. Es fehlt aber ein eindeutiger Deskriptor, der einem gewissermaßen eine Vorauswahl vor dem Docking absichert. Exakte Energiewerte könnten so eine Größe sein. Thermodynamische Informationen sind in den Deskriptoren gewissermaßen in indirekter Form enthalten, aber reichen nicht zu einer vollständigen Charakterisierung. Die Anzahl der Wasserstoffbrückenbindungen liefert wichtige Informationen darüber, wie ein Wassermolekül mit seiner Umgebung verknüpft ist. Darin wird aber naturgemäß nicht die unterschiedliche energetische Stärke von Wasserstoffbrückenbindungen abhängig von den jeweiligen Donoren, Akzeptoren und der Umgebung berücksichtigt. Die DrugScore Funktion wiederum versucht das thermodynamische Profil abzubilden. Bei ihrer Parametrisierung lagen jedoch keine experimentellen, energetischen Daten von Wassermolekülen vor. Aus diesem Grund könnte die Aussagekraft der Scores beeinflusst sein. Ohnehin lässt sich so zwar ein enthalpischer

**Tab. 5.11.:** Normalisierte B-Faktoren berechnet nach<sup>13</sup> für einige der untersuchten Wassermoleküle und zwecks Vergleich ebenso für Ligand und Cofaktor der sieben untersuchten Systeme. Die Normalisierung basiert auf einem gemittelten B-Faktor für alle Protein, Ligand und Cofaktor Atome. Die Wassermoleküle wurden mit den systemunabhängigen Beschriftungen „A“, „B“ und „C“ versehen.

System	Ligand	Häm	A	B	C
1CPE	0.85	-0.40	0.07	-0.03	-0.25
1AEO	0.07	-0.46	0.57	-	-0.25
1AEF	0.08	-0.42	0.50	0.08	-0.77
1AEF (neu)	0.37	0.03	0.17	0.51	0.27
1AEG	0.07	-0.42	0.50	-	-0.77
2AQD	0.16	-0.25	1.68	-	0.10
2EUP	-0.18	-0.12	0.07	-	0.14

Beitrag zur Bindung aber keinesfalls der entropische Teil abbilden. Letzterer ist aber entscheidend, um über die Verdrängbarkeit zu urteilen.

Bisher wurde aber ein Aspekt bei den Untersuchungen außer Acht gelassen. Kristallstrukturen bieten naturgemäß immer nur einen Einblick in die Proteinstruktur im festen Zustand. In Lösung wird ein Protein-Ligand-Komplex jedoch durch eine Vielzahl repräsentativer Strukturen beschrieben. Um insbesondere für diesen, in biologischer Umgebung relevanten Zustand ein Verständnis der Protein-Ligand-Wasser-Interaktionen zu erhalten, wurden MD-Simulationen durchgeführt, die im folgenden Kapitel diskutiert werden. Daraus lassen sich dann ebenfalls die oben angesprochenen entropischen Informationen berechnen. Der B-Faktor eines Atoms bzw. die Mobilität ersetzt nicht die Simulation. Wie bereits eingangs erwähnt, fließen in diese auch die Verfeinerungsmethode oder die strukturelle Unordnung im Kristall ein.

**Tab. 5.12.:** Deskriptoren aller Kristallwassermoleküle in und um die Bindetasche der untersuchten Systeme der CCP W191G. Die aufgeführten Werte sind dem Waterbase Modul der Datenbank Relibase+ 3.0.0 entnommen. Die Abkürzungen stehen im einzelnen für: MOB Mobilität, PK Protein Kontakte, LK Ligand Kontakte, WK Wasser Kontakte, GK Summe aller Kontakte, POL Polarität in der Umgebung des Wassermoleküls, DSP DrugScore<sup>PDB</sup>-Protein, DSL DrugScore<sup>PDB</sup>-Ligand, DSW DrugScore<sup>PDB</sup>-Wasser, DSG Summe der DrugScore<sup>PDB</sup> Terme, AD1 und AD2 Atom Dichte in der Koordinationssphäre von Wasser, SAS Solvenz zugängliche Oberfläche, RMSDtet RMSD in Bezug auf eine ideale tetraedrische Koordinationsgeometrie. Für diese Größe wurde dann kein Wert aufgeführt, wenn auch in der Datenbank keiner abgelegt ist.

System	Wasser	MOB	PK	LK	WK	GK	POL	DSP	DSL	DSW	DSG	AD1	AD2	SAS	RMSDtet
1CPE	477 (A)	0.55	3	0	0	3	4.806	-18198.92	0.0	5611.19	-12587.74	6.06	59.05	31.00 %	0.15
	976	0.41	2	0	1	3	4.320	-6084.78	0.0	-2216.69	-8301.46	5.23	65.61	44.00 %	0.24
	977 (B)	0.52	1	0	2	3	3.000	-2053.03	0.0	-3268.77	-5321.80	1.38	63.99	58.00 %	-
	978	0.41	1	0	2	3	3.244	-9564.14	277.65	-6505.93	-15792.42	1.25	69.02	58.00 %	-
	979	1.00	0	0	1	1	2.265	2557.31	-2281.23	-932.74	-656.66	4.54	70.83	46.00 %	-
	535 (C)	0.46	2	0	0	2	2.972	-12363.89	-6429.63	1688.15	-17105.38	4.89	65.98	22.00 %	-
1AEO	308 (A)	0.95	3	0	0	3	4.301	-16427.38	-1282.80	0.00	-17710.19	6.00	63.72	28.00 %	-
	332 (C)	0.56	2	0	0	2	2.821	-15255.25	-4532.53	0.00	-19787.78	4.56	66.94	23.00 %	-
1AEF	308 (A)	0.90	3	0	0	3	3.965	-16435.07	-2155.28	3326.88	-15263.47	5.61	63.80	26.00 %	-
	452 (B)	0.69	1	1	0	2	3.217	-6961.11	-3090.49	3252.70	-6798.90	6.11	67.50	30.00 %	-
	332 (C)	0.26	2	0	0	2	3.042	-13897.63	-3002.05	0.00	-16899.69	4.42	67.26	25.00 %	-
1AEF (neu)	3 (A)	0.38	2	0	0	2	2.677	-21930.96	0.00	0.00	-21930.96	4.32	64.15	23.00 %	-
	4 (B)	0.45	3	0	0	3	4.745	-19124.24	0.00	2992.71	-16133.53	6.93	74.09	12.00 %	-
	18 (C)	0.40	2	0	0	2	3.082	-14548.79	-7295.50	0.00	-21844.29	5.38	69.86	16.00 %	-
1AEG	308 (A)	0.90	3	0	0	3	3.796	-16435.07	-2777.95	1638.73	-17574.29	5.44	63.82	25.00 %	-
	332 (C)	0.26	2	0	0	2	3.042	-13897.63	-3002.05	0.00	-16899.69	4.42	67.17	25.00 %	-
2AQD	1000 (A)	0.86	4	0	0	4	4.852	-17128.02	-754.57	1816.84	-16065.75	7.62	66.10	21.00 %	0.22
	1104 (C)	0.48	2	0	0	2	2.856	-14407.04	-6716.67	0.00	-21123.71	5.28	69.66	17.00 %	-
2EUP	37 (A)	0.54	4	0	0	4	4.635	-20237.21	-1467.40	2484.30	-19220.30	7.22	67.47	23.00 %	0.53

*Fortsetzung auf der nächsten Seite*

System	Wasser	MOB	PK	LK	WK	GK	POL	DSP	DSL	DSW	DSG	AD1	AD2	SAS	RMSDtet
	3 (C)	0.55	2	0	0	2	2.891	-14124.09	-6031.32	0.00	-20155.42	5.34	69.65	17.00 %	-
Modell CCP W191G	305	-	1	0	2	3	3.702	-7393.12	0.00	-4037.91	-11431.03	1.25	55.90	58.00 %	-
	307	-	1	0	1	2	3.367	6199.58	-2468.18	-573.69	3157.710	7.13	69.15	50.00 %	-
	308	-	1	0	1	2	2.817	-4100.48	0.00	-1060.14	-5160.62	1.47	60.29	57.00 %	-
	1573 (B)	-	1	0	3	4	4.104	-1625.50	0.00	-6350.34	-7975.83	1.98	59.02	59.00 %	0.13
	2238 (A)	-	2	0	2	4	5.539	-6597.20	0.00	-2178.23	-8775.43	3.18	53.79	48.00 %	0.33
	5226	-	1	0	3	4	4.263	-2307.79	0.00	-6372.3	-8680.10	1.71	60.77	59.00 %	0.39
	6667	-	1	0	2	3	3.000	-2421.58	794.95	-4352.06	-5978.68	1.00	64.91	65.00 %	-
	309 (C)	-	3	0	0	3	4.036	-14478.58	-1078.12	1688.15	-13868.55	7.54	67.01	19.00 %	-





# 6. Molekulardynamik-Simulationen von Strukturen der Cytochrom c Peroxidase W191G Mutante

## 6.1. Einstellungen und Parameter der angewandten Methode

Neben den strukturellen Eigenschaften, die aus einer einzelnen Kristallstruktur für einen Protein-Ligand Komplex abgeleitet werden können, sind maßgeblich auch die Eigenschaften des jeweiligen Komplexes in Lösung von großem Interesse. Aus diesem Grund wurden Dynamik-Simulationen aller in dieser Arbeit untersuchten Kristallstrukturen durchgeführt. Dabei wurde den Molekular-Dynamik-Simulationen der Vorzug vor Metropolis-Monte-Carlo-Simulationen gegeben, da man hier nicht nur ein Ensemble von Einzelstrukturen sondern eine komplette Trajektorie mit einer zeitlich relevanten Bewegungsabfolge erhält. Diese Simulationen wurden alle mit dem Programm AMBER durchgeführt, da es sich in der Vergangenheit bereits bei vielen Protein-Simulationen als sehr geeignet erwiesen hat (die Fülle der bekannten Literatur-Referenzen ist zu umfangreich, um sie an dieser Stelle aufzuführen, eine Übersicht findet sich beispielsweise bei Kollman und Kollegen<sup>203</sup>). Außerdem verfügt das Programmpaket über einige Hilfs-/Zusatzprogramme, die dem Nutzer wesentliche Vorbereitungs- und Auswertungsschritte erleichtern.

Die grundlegenden Schritte sind bei allen durchgeführten MD-Simulationen identisch und sollen im Folgenden dargestellt werden. Spezielle Vorgangsweisen, die nur in Einzelfällen angewandt wurden, werden jeweils an der entsprechenden Stelle erläutert. Außer im Falle des ionenfreien Bindetaschen-Modells wurden die Kristallstrukturen 1AEO, 1AEF, 1AEG, 2AQD, 2EUP, 1CPE, 1AA4 und die neu kristallisierte Struktur des Systems 1AEF als Ausgangspunkt für die MD-Simulation verwendet. Aus diesen wurden, sofern vorhanden, die Wasserstoffatome und alle für die Fragestellung nicht relevanten Wassermoleküle entfernt. Eine Übersicht darüber, welche Wassermoleküle in die Startstruktur der MD-Simulation übernommen wurden, findet sich in Tab. 6.1. Nur die Wassermoleküle, die nahe am Liganden, nahe an diesen Wassermolekülen oder an Asp235 bzw. das Fe des Cofaktors koordiniert sind, wurden in die MD-Simulation übernommen. Es kann davon ausgegangen werden, dass diese Wassermoleküle entweder direkte Wechselwirkungen mit dem Liganden eingehen, in der Bindetasche an Austauschvorgängen beteiligt sein können oder wie im Falle des an das Fe koordinierten H<sub>2</sub>O die sechste Bindungsstelle des Metalls absättigen und somit entscheidend für die Dynamik des Systems sein werden.

Von allen Kristallstrukturen wurde das Protonierungs-Schema der Histidine visuell überprüft. Zusätzlich wurden auch Informationen, die mit Hilfe der *Protonate 3D*-Funktionalität von MOE<sup>204</sup> bzw. dem What If Webserver<sup>205</sup> erhalten wurden, herangezogen. Es zeigte sich, dass His96 in  $N_\epsilon$  und insbesondere das proximale His175 in  $\delta$ -Position protoniert sind. Für alle anderen His-Reste dieses Proteins wurde ebenfalls ein Wasserstoffatom am  $N_\delta$  angenommen, wobei dies für die Dynamik im Bereich der Bindetasche keinen signifikanten Einfluss haben sollte, da diese Aminosäuren relativ weit von der Kavität entfernt sind. Die MD-Simulationen wurden mit den Versionen 9 und 10 des AMBER Programmpakets<sup>206,207</sup> durchgeführt. Zur Bestimmung der RESP-Ladungen<sup>120</sup> des Liganden und Häm Cofaktors wurden mit Hilfe der HF/6-31G\* Methode in Gaussian 03<sup>208</sup> elektrostatische Potentiale berechnet und anschließend mit ANTECHAMBER<sup>209</sup> die RESP-Ladungen bestimmt. Für das Fe(III) wurde in der Berechnung des elektrostatischen Potentials ein Radius von 69 pm<sup>210</sup> angenommen. Die Kraftfeldparameter des Cofaktors wurden der AMBER-Parameter-Datenbank<sup>211,212</sup> entnommen. Fehlende Kraftfeldparameter des Liganden wurden mit dem Programm PARMCHK<sup>209</sup> passend zum jeweiligen GAFF-Atomtyp<sup>213</sup> bestimmt. Für die Protein-Atome wurde das Parm99SB Kraftfeld<sup>119</sup> aus AMBER verwendet. Alle fehlenden Protonen des Proteins wurden mit Hilfe des Programms LEaP hinzugefügt. Desweiteren wurden die für die Simulation benötigten Dateien ebenfalls mit diesem Programm erstellt. Dabei wurde explizit eine Bindung zwischen Häm-Fe und  $N_\epsilon$  von His175 definiert.

Alle Komplexe wurden sodann mit dem Programmmodul SANDER über 200 Schritte mit einem „steepest descent“ Algorithmus minimiert. Da noch keine Solvens-Atome vorhanden waren, wurde hier ein modifiziertes generalized Born implizites Solvens-Modell<sup>214</sup> verwendet. Zu der minimierten Struktur wurden anschließend bis zur Neutralität  $Na^+$ -Ionen hinzugefügt. Mit LEaP wurden die Komplexe danach mit dem TIP3P Solvensmodell<sup>124</sup> in einer Box, deren Kanten mindestens 8.0 Å Abstand vom Protein haben, solvatisiert. Abhängig vom jeweiligen Komplex enthält das System zwischen 8400–9300 Wassermoleküle. Die genaue Anzahl der hinzugefügten Wassermoleküle und weitere Details der simulierten Systeme finden sich in Tab. 6.2. Das System wurde dann von 100 K auf 300 K in 20 ps aufgeheizt und dann in 5 ps auf 100 K im *NVT*-Ensemble abgekühlt. Hierbei wurden die Positionen aller Atome außer die der Wassermoleküle und Ionen festgehalten. Im nachfolgenden Schritt wurde das jeweilige System im *NPT*-Ensemble über 25 ps auf 300 K erwärmt ohne positionsbeschränkende Kräfte auf den Atomen. Anschließend wurde jeweils eine 7 ns Trajektorie unter periodischen Randbedingungen, konstantem Druck (1 atm) und Temperatur (300 K)<sup>215</sup> sowie einem Zeitschritt von 2 fs simuliert. Der SHAKE<sup>126</sup> Algorithmus wurde auf alle Bindungen zu Wasserstoffatomen angewandt und die „particle mesh Ewald“ Methode<sup>127</sup> verwendet, um langreichweitige, elektrostatische Wechselwirkungen zu handhaben. Für andere nicht-kovalente Wechselwirkungen wurde ein Distanz-Ausschlusskriterium von 8.0 Å eingestellt. Die Koordinaten der Trajektorien wurden alle 1 ps und Energiewerte alle zehn Zeitschritte abgespeichert. Für fast alle der eingangs aufgeführten Systeme wurde eine Simulation ohne und mit einem zusätzlichen harmonischen Potential von 5.0 kcal/mol auf die PGGAAN 190-196 Loop  $C_\alpha$ -Atome bezüglich ihrer Ausgangsposition durchgeführt.

Die so erzeugten Trajektorien wurden alle mit dem Programm PTRAJ auf das Pro-

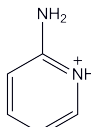
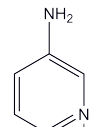
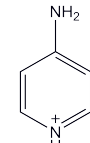
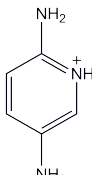
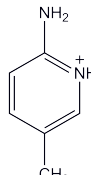
**Tab. 6.1.:** Liste der Kristallwassermoleküle, die in den Simulationen verwendet wurden.

PDB-Code	Kristallwassermoleküle in den MD-Simulationen
1CPE	77, 494, 535, 611, 976, 977, 978, 979
1AEO	308, 313, 332
1AEF	308, 313, 332, 406, 407, 452
1AEF (neu)	3, 4, 18, 141, 165, 206
1AEG	308, 313, 332, 406, 407
2AQD	1000, 1104, 1242, 1434, 1415
2EUP	3, 37, 62, 83, 312

tein zentriert, in ihre ursprüngliche Solvensbox zurück projiziert und einem RMSD-Fit bezüglich der  $C_\alpha$ -Atompositionen zu Beginn der Trajektorie unterzogen. Ein großer Teil der im Folgenden dargestellten strukturellen Deskriptoren wie Atomabstände, RMSD-Werte, Anteil von Sekundärstrukturelementen wurden mit PTRAJ berechnet. Die Visualisierung einzelner Konformationen erfolgte mit PyMOL<sup>199</sup>, die Visualisierung der Trajektorien mit VMD<sup>216</sup>. Mit dem Volmap-Plug-in dieses Programms wurden auch die Hydratationsdichtekarten bei einer Auflösung von 0.5 Å berechnet. Die lokalen Maxima wurden daraus mit Hilfe eines selbst-entwickelten Python-Skriptes extrahiert, das die APBS Python-Tools<sup>217</sup> mit verwendet. Hierbei wird aus der Wasserdichte das globale Dichtemaximum gefunden und abgespeichert. Anschließend wird die Wasserdichte an diesem Maximum und im umgebenden Bereich von 2.5 Å auf Null gesetzt. Diese Distanz wurde gewählt, weil sie im Rahmen der Auflösung des Dichtegitters dem Abstand einer Wasserstoffbrückenbindung entspricht und angenommen werden kann, dass sich zwei gleichzeitig in der Tasche vorhandene Wassermoleküle nicht weiter annähern werden. Aus der verbleibenden Dichteverteilung wird nun das nächste globale Maximum gefunden und dann wie zuvor verfahren. Diese Schritte werden wiederholt, bis keine weiteren Maxima mehr gefunden werden.

Die hier gewählte Art der Parametrisierung für die Simulationen hat sich erst nach mehreren fehlgeschlagenen Versuchen als optimal herausgestellt. Insbesondere die explizite Definition der Bindung zwischen Cofaktor und proximalen His stellte sich dabei als essentiell heraus. Falls man diesen Schritt weg lässt, sind beide Systemteile (Protein und Cofaktor) nur schwach miteinander verknüpft. Dies führt zu einer erhöhten Beweglichkeit und strukturellen Destabilisierung im Bereich der Bindetasche.

**Tab. 6.2.:** Details zu den simulierten Systemen. Die Bindungskonstanten der Komplexe 1AEO, 1AEF und 1AEG wurden mit zwei voneinander unabhängigen Methoden bestimmt. Der erste der beiden angegebenen Werte stammt aus ITC Messungen, die zweite Dissoziationskonstante wurde aus der Störung der Häm-Soret Absorption bestimmt<sup>95</sup>. Diese Methode wurde auch zur Bestimmung der Bindungskonstante in 2EUP angewandt. Für 2AQD liegen keine Werte vor. Die Anzahl der Kristallwassermoleküle ist diejenige Menge an Wassermolekülen, die aus der Kristallstruktur in die Simulation übernommen wurden. Zusätzlich ist die Gesamtzahl der Solvens-Moleküle im System angegeben. Die Werte in Klammern für den Komplex 1AEF beziehen sich auf die neu kristallisierte Struktur dieses Systems (siehe Abschnitt 6.3).

PDB-Code	1CPE <sup>218</sup>	1AEO <sup>95</sup>	1AEF <sup>95</sup>	1AEG <sup>95</sup>	2AQD <sup>96</sup>	2EUP <sup>96</sup>
Ligand	K <sup>+</sup>	 2-Aminopyridin	 3-Aminopyridin	 4-Aminopyridin	 2,5-Diaminopyridin	 2-Amino-5-picolin
$K_d$ [mM]	-	0.05, 0.07	0.04, 0.07	0.04, 0.05	-	0.04
Auflösung	2.20 Å	2.10 Å	2.10 (1.25) Å	2.10 Å	1.35 Å	1.40 Å
Anz. Kristallwasser	9	3	6	5	5	5
Anz. Atome gelöste Substanz	4649	4661	4661 (4656)	4661	4664	4679
Anz. Solvens Moleküle	8613	8433	8762 (9295)	8820	9254	9246
Anz. Atome in System	30488	29960	30947 (32541)	31121	32426	32426

## 6.2. Auswertung der erhaltenen Trajektorien

### 6.2.1. Analyse der strukturellen Stabilität von Protein, Bindetasche und Ligand-Bindungsmodus

Bei einer MD-Simulationen werden unterschiedliche Konformationen eines Systems unabhängig davon ob die gewählten Einstellungen sinnvoll sind oder nicht generiert. Deshalb ist es nach Abschluss der Berechnung angebracht, die erhaltene Trajektorie auf mögliche Auffälligkeiten zu überprüfen und so deren Aussagekraft zu verifizieren. Das können z.B. eine kontinuierlich größer werdende strukturelle Änderung des Proteins im Vergleich zur Ausgangsstruktur oder eine merkliche Veränderung in der Zusammensetzung der Sekundärstrukturelemente wie z.B. Helices sein. Aber auch häufige Temperaturschwankungen, die durch den Kopplungsalgorithmus immer wieder ausgeglichen werden müssen, können ein Indiz für eine nicht adäquate Parametrisierung sein. In Abb. 6.2 sind RMSD-Werte für verschiedene Atomgruppen jeweils für die fünf Komplexsimulationen und die MD-Simulationen der Struktur 1CPE und der ionenfreien Modellbindetasche aufgetragen. An den schwarzen und roten Auftragungen sieht man, dass die Proteinstrukturen und im speziellen die Bindetaschen im Verlauf der Simulation sich nicht wesentlich im Vergleich zur Ausgangsstruktur ändern. Der grüne Graph zeigt, dass die Änderung der Bindetaschenkonformation nicht nur absolut sondern auch relativ zum Rest des Proteins gering ist. Größere strukturelle Änderungen werden jedoch teilweise für die jeweiligen Liganden sichtbar. Die hohe Fluktuation der blauen Linie für das System 2EUP wird durch die Rotation der Methyl-Gruppe des Liganden verursacht, da die RMSD-Werte für alle Atome inkl. der Wasserstoffatome bestimmt wurden.

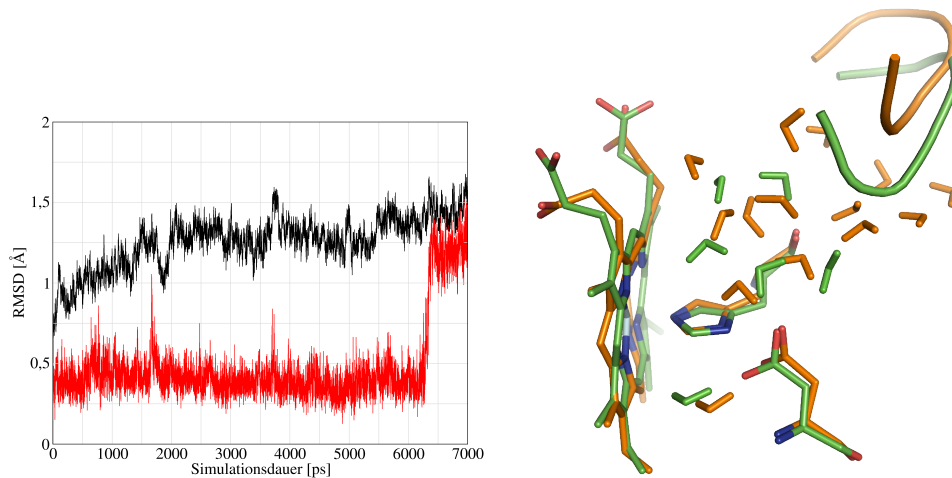
Der Bindungsmodus der Liganden ist in den fünf Komplex-Simulationen weitgehend konserviert im Vergleich zur Kristallstruktur. Einzige Ausnahme bildet das System 1AEF. Bereits nach sehr kurzer Simulationszeit dreht sich der Ligand in der Bindetasche so, dass eine direkte Wechselwirkung zwischen aromatischem N-Atom und Asp235 möglich ist. Der in Abb. 6.10 dargestellte Abstand zum rigiden Leu174 kann als charakteristisch für diese Drehung angesehen werden. Weitere Analyseschritte zu dieser Beobachtung und die daraus resultierenden Konsequenzen sind in Abschnitt 6.3 ausführlich dargestellt. In diesem Zusammenhang steht auch die Lösung einer neuen Struktur des Systems 1AEF. In der damit durchgeführten Simulation zeigt sich ebenfalls eine Umorientierung des Liganden und damit verbunden ein erhöhter RMSD-Wert (Abb. 6.2). Hier dreht sich der Ligand jedoch sehr rasch wieder in die Ausgangskonfiguration zurück.

Auch im Falle des holo-Systems 1CPE kann anhand des Verlaufs des RMSD-Wertes die Positionsänderung des  $K^+$ -Ions beobachtet werden. Dieses bewegt sich von seiner ursprünglichen Position auf Koordinaten in der Mitte der Bindetasche (siehe Abb. 6.4) und wird dort durch fünf bis sechs Wassermoleküle und den Carbonyl-O von His175 annähernd oktaedrisch koordiniert. In der Arbeit von Ross und Hardin<sup>219</sup> wurde gezeigt, dass die Standard  $K^+$  Lennard-Jones 6-12-Parameter angemessen für die Beschreibung der Wechselwirkungen mit TIP3P-Wassermolekülen und ihrem großen van-der-Waals Radius sind, jedoch nicht für die zu DNA oder Proteinen. Aus diesem Grund wurde eine zusätzliche

MD-Simulation des Systems 1CPE durchgeführt, um zu überprüfen, ob die Bewegung des  $K^+$ -Ions in die Mitte der Tasche in ungeeigneten Parametern begründet liegt. Diese 2 ns lange Simulation wurde mit einem veränderten  $K^+$ -Parameter  $r^* = 1.93 \text{ \AA}$  durchgeführt. Der Standardwert im AMBER Kraftfeld beträgt hierfür hingegen  $2.658 \text{ \AA}$ . Auch in dieser Trajektorie verlässt das Kation seine ursprüngliche Position und bewegt sich hier nahe an Asp235. Dort sind dann zwischen drei und vier Wassermoleküle an das  $K^+$  koordiniert. Die Parametrisierung ist somit sehr wahrscheinlich nicht die Ursache für die Bewegung des Ions im Verlauf der Simulation.

Die in Abb. 6.3 dargestellten Diagramme untermauern den bisherigen Eindruck einer weitgehend stabilen Protein-Konformation. Hier sind die Anteile unterschiedlicher Sekundärstrukturelemente an der Gesamtstruktur im Verlauf der Simulation dargestellt. Man sieht unmittelbar, dass keine Änderung der Proteinfaltung in den generierten Trajektorien zu beobachten ist. Eine Auftragung von Temperatur, kinetischer, potentieller oder Gesamtenergie im Verlauf der Simulationen (Diagramme hier nicht gezeigt) bestätigt außerdem das Bild, dass die generierten Strukturenssembles in einem equilibrierten Zustand vorliegen. Maximal die ersten 500 ps einer jeden Trajektorie können als notwendiger Zeitraum für die Equilibrierung des Proteins angesehen werden. In diesem Zusammenhang ist auch die Frage zu erörtern, ob ebenfalls eine Equilibrierung der Bindetaschen-Wassermoleküle innerhalb dieser Zeitspanne vorliegt. Das kann an den weiter unten diskutierten Hydratationsdichtekarten abgelesen werden. Gerade die Wasserdynamik während einer möglichen Equilibrierungsphase im Vergleich zur Röntgenstruktur ist außerdem sehr interessant. Zusätzlich wurden verschiedene Test-Simulationen durchgeführt, die überprüfen sollten, wie lange Wassermoleküle benötigen, um eine wasserfreie 1CPE Bindetasche zu befüllen und einen equilibrierten Zustand zu erreichen. Diese Trajektorien wurden ohne (2.5 ns), wie auch unter Zuhilfenahme von Zusatzkräften (unterschiedlicher Stärke oder graduell abnehmend) (3 ns), die eine strukturelle Destabilisierung einer leeren Tasche verhindern sollen, durchgeführt. Hierbei, wie auch in einer MD-Simulation des Systems 2EUP ohne Bindetaschenwassermoleküle, konnte keine Besetzung charakteristischer Wasserpositionen innerhalb des simulierten Zeitraums beobachtet werden. Die vermeintliche Wasserequilibrierungszeit scheint also länger als die hier generierten Trajektorienabschnitte zu sein.

Im Falle der MD-Simulation der Struktur 1AA4<sup>91</sup> erhält man ein anderes Bild. Die Trajektorie für diese Kristallstruktur wurde berechnet, um zu überprüfen, ob die Zweifel am sechsten Wassermolekül in der Bindetasche an Stelle eines  $Na^+$ -Ions berechtigt sind und deshalb die Konstruktion eines ionenfreien Bindetaschenmodells überhaupt notwendig ist. In Abb. 6.1 findet man eine Auftragung des  $C_\alpha$ -RMSDs in Abhängigkeit der Simulationsdauer. Während dieser Wert kontinuierlich größer wird, verzeichnet der RMSD der PGGAAN 190-195 Loop- $C_\alpha$ -Atome nach ca. 6200 ps einen plötzlichen Anstieg. Zu diesem Zeitpunkt öffnet sich der Loop-Bereich und zahlreiche zusätzliche Wassermoleküle strömen in die Bindetasche. Diese Destabilisierung (siehe ebenfalls Abb. 6.1) ist sehr wahrscheinlich auf den fehlenden Ladungsausgleich in der Bindetasche zurück zu führen.



**Abb. 6.1.:** MD-Simulation der CCP W191G Struktur 1AA4. Links: Diagramm der RMSD-Werte der  $C_{\alpha}$ -Atome des ganzen Proteins (schwarz) bzw. nur des PGGAAN 190-195 Loops (rot). Rechts: Blick auf Bindetasche und Loop zusammen mit Wassermolekülen nach 1 ps (grün) und 6365 ps (orange) Simulationsdauer.

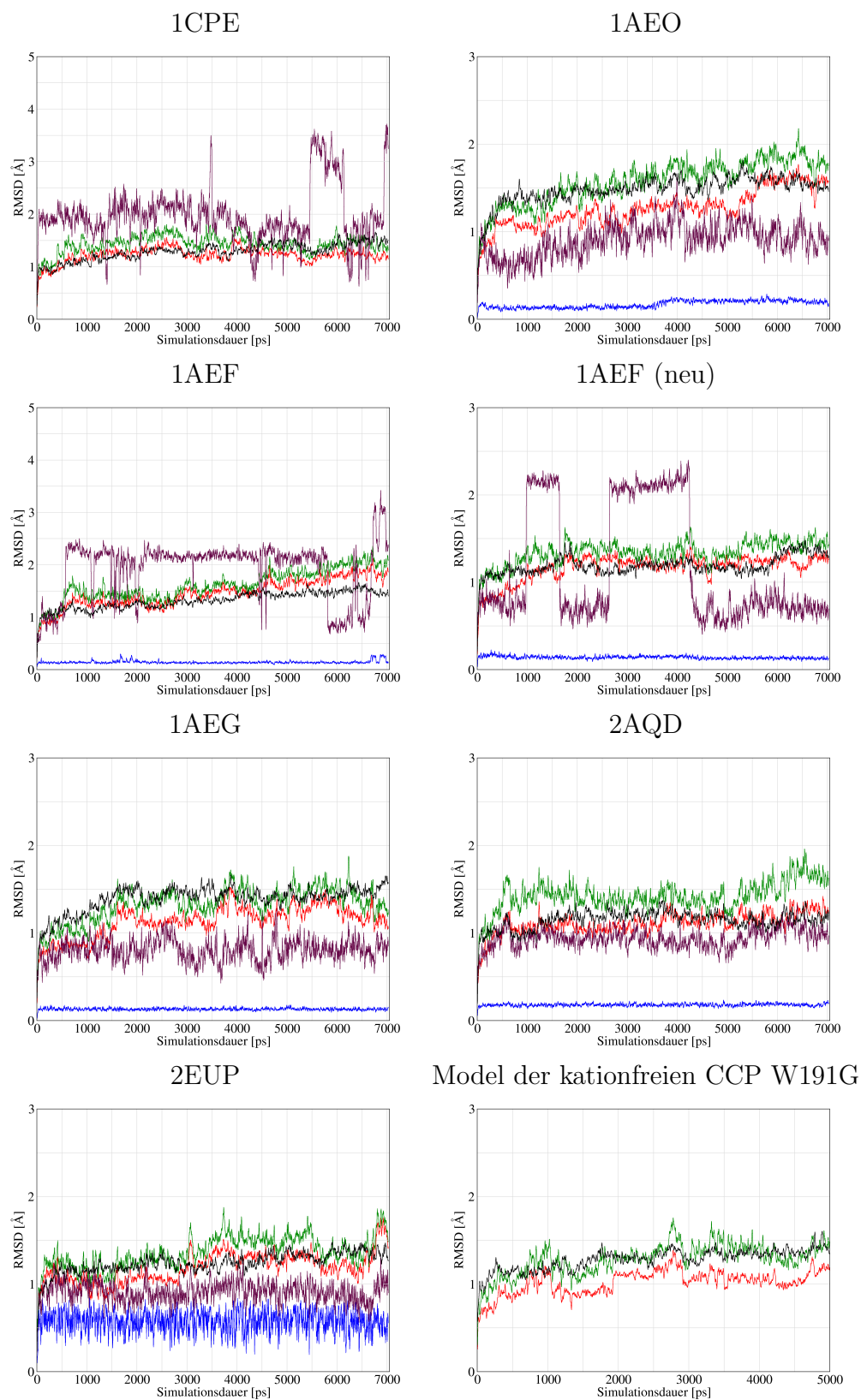
### 6.2.2. Betrachtung der Dynamik der Bindetaschen-Wassermoleküle

Neben verschiedenen weit verbreiteten Analysemethoden zur Erfassung von Dynamik von Ligand und Rezeptor sind bei der Bestimmung der Wasserdynamik einige Vorüberlegungen notwendig. Um im Rahmen der allgemeinen Protein-Bewegung während der Simulation eine Wasserposition sinnvoll zu definieren und aussagekräftige Werte z.B. für die Aufenthaltszeit zu erhalten, bedarf es - wie bereits in Abschnitt 1.2.4 angesprochen - einiger zusätzlicher Berechnungsschritte in Form von gemittelten Positionen und einem relativen Koordinatensystem. Eine andere übliche Herangehensweise bei der Analyse von Wechselwirkungen mit Wassermolekülen wäre die Bestimmung von Besetzungsgraden von Wasserstoffbrückenbindungen. Im Vorfeld der Erstellung dieser Arbeit wurde versucht, sowohl den Besetzungsgrad einzelner Wasserpositionen zu berechnen als auch die Absättigung von Wasserstoffbrückenbindungen zu analysieren. Im Folgenden soll erläutert werden, warum beide Ansätze dennoch nicht weiter verfolgt wurden.

#### Analyse der H-Brücken

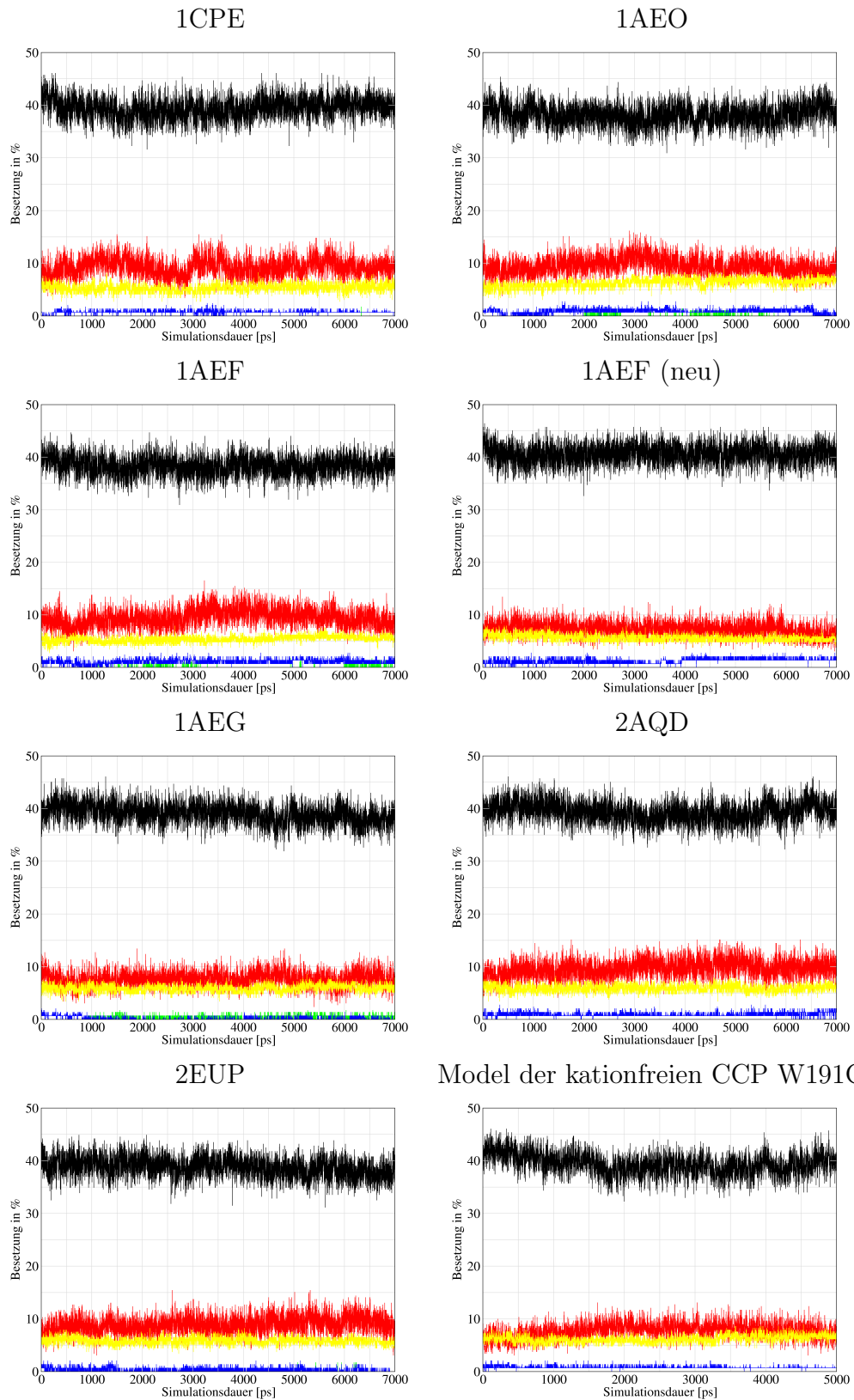
Bei einer Analyse von H-Brückenbindungen stellt sich zunächst die Frage, welche Grenzen (d.h. maximaler Abstand und größter Winkel) gewählt werden, in denen eine H-Brückenbindung als besetzt gilt. Das Analyse-Programm PTRAJ wählt hierfür standardmäßig Werte von 3.5 Å für den Abstand zwischen den beiden Heteroatomen. Dies scheint relativ hoch und passt nicht zu Angaben bei Jeffrey und Saenger<sup>220</sup> sowie Creighton<sup>221</sup>. Böhm *et al.*<sup>222</sup> geben einen durchschnittlichen Wert von 2.8 Å bis 3.2 Å für den N–O-Abstand und einen Wert von  $> 150^{\circ}$  für den N–H–O Winkel, sowie  $100^{\circ}$  bis  $180^{\circ}$  für den C=O–H Winkel

## 6. Auswertung der Dynamik



**Abb. 6.2.:** Übersicht über den Verlauf (Mittelung in 10 ps Intervallen) verschiedener RMSD-Werte der durchgeführten MD-Simulationen der CCP W191G. In Schwarz sind die Werte der C $\alpha$ -Atome, in Rot die der Bindetaschenatome und in Blau die der Ligandatome basierend auf einem Fit zur Referenzstruktur aufgetragen. In Grün dargestellt findet man die Werte der Bindetaschenatome basierend auf dem Fit aller C $\alpha$ -Atome des Proteins sowie in Violett der Ligandatome (für 1CPE von K $^+$ ) basierend auf dem Fit der C $\alpha$ -Atome.





**Abb. 6.3.:** Anteil verschiedener Sekundärstrukturelemente im Verlauf der generierten Trajektorien der CCP W191G. Die schwarze Linie zeigt den Anteil der  $\alpha$ -Helices und die rote den der 3-10-Helices. Der Anteil der antiparallelen  $\beta$ -Sheets wird durch die gelbe Linie, der Anteil der  $\pi$ -Helices durch die grüne und der der parallelen  $\beta$ -Sheets durch die blaue dargestellt.

an. Nach Verwertung dieser Informationen wurde für die H-Brückenbindungsanalyse ein maximaler Heteroatomabstand von 3.2 Å festgelegt. Der Winkel X–H–X sollte zwischen 180° und 100° liegen.

Die Besetzungsgrade der H-Brücken wurden für Trajektorienabschnitte von 100 ps jeweils separat für Protein–Protein, Wasser–Protein, Protein–Ligand, Wasser–Ligand sowie Wasser–Wasser H-Brückenbindungen bestimmt. Auf diese Weise konnte der Verlauf während der Simulation analysiert werden. Zum Vergleich wurde die entsprechende Analyse ebenfalls für die protonierten und mit SANDER minimierten Kristallstrukturen durchgeführt. Da hier naturgemäß nur eine Konformation vorliegt und keine Mittelung wie bei den Trajektorienabschnitten stattfindet, wurden zusätzlich die Werte für ein weniger spezifisches Distanzkriterium von 3.5 Å berechnet.

Als Nachteil stellte sich hierbei jedoch zunächst einmal der hohe Aufwand heraus, diese Analyse pro Trajektorie für 65 Abschnitte (500 ps Equilibrierung wurden angenommen) jeweils fünfmal (Aufteilung in Wasser-Wasser, Wasser-Ligand, Protein-Wasser, Protein-Protein und Protein-Ligand Anteile) durchzuführen. Desweiteren erhält man bei dieser Betrachtung Ergebnisse für individuelle Wassermoleküle. Diese sind aber nur in der Simulation, nicht aber in der experimentellen Wirklichkeit unterscheidbar. Relevant sind vielmehr die Eigenschaften einzelner Wasserpositionen in Ligandnähe unabhängig davon, welches Wassermolekül im Moment dort anzutreffen ist. Der etwas höher frequentierte Austausch auf Position A in manchen Komplexen brachte zudem eine weitere Komplikation mit sich. Selbstverständlich muss die Analyse für alle Wassermoleküle, die jemals in der Bindetasche anzutreffen sind durchgeführt werden. Das bedeutet aber auch, dass die Wechselwirkungen von Wassermolekülen, die gerade die Tasche verlassen haben, z.B. zum Protein oder anderen Wassermolekülen weiter erfasst werden. Ansonsten wäre es notwendig, die Gruppe der zu analysierenden Moleküle in jedem Trajektorienabschnitt neu festzulegen, was einen praktisch kaum zu handhabbaren Aufwand bedeuten würde. Diese zusätzliche Erfassung weiterer Wechselwirkungen außerhalb der Bindetasche verfälscht dann aber die Ergebnisse, wenn man an der Gesamtzahl der H-Brücken-Wechselwirkungen interessiert ist und sie unter den Komplexen vergleichen möchte.

Aufgrund dieser Schwierigkeiten wurde im Folgenden darauf verzichtet, diese systematische Analyse der Trajektorien weiter auszuwerten. Es ist zu erwarten, dass aufgrund der Nachteile des Auswertungsverfahrens keine aussagekräftigen Ergebnisse erhalten werden können.

### **Besetzungsgrad von Wasserpositionen**

Im vorangegangenen Abschnitt wurde bereits angesprochen, dass das primäre Interesse bei den Eigenschaften der für die Bindetasche charakteristischen Wasserpositionen nicht bei den Wassermolekülen, die diese im Verlauf der Trajektorie besetzen, liegt. Stattdessen ist man gewissermaßen an der Eigenschaft der Position an sich interessiert. Aus diesem Grund wurde versucht, die Besetzungsgrade dieser Positionen während des simulierten Zeitraumes zu bestimmen. Da diese Funktionalität nicht im Standard-MD-Auswertungsprogramm PTRAJ enthalten ist, wurde ein PYTHON-Skript entwickelt, das diese Aufgabe

lösen kann. Dieses verwendete wiederum eine fertige Klasse zum Lesen von PDB-Dateien<sup>223</sup> aus Biopython<sup>224</sup>. Mit diesem Programm wird dann erfasst, inwieweit jedes O-Atom irgend eines Wassermoleküls im Verlauf der Simulation innerhalb einer Kugel um eine gegebene Referenzposition anzutreffen ist. Für die praktische Anwendung ergeben sich damit zwei Fragen: (1) Wo genau wird die Referenzposition gewählt? (2) Wie groß wählt man den Kugel-Radius um die jeweilige Position?

Für die Kugelgröße wurde die RMS-Auslenkung der Bindetaschen-Aminosäuren (His175, Leu177, Gly178, Lys179, Thr180, Pro190, Gly191, Asn195, Phe202, Met230, Met231, Met232, Asp235) berechnet. Die Summe aus diesem Mittelwert und der Standardabweichung ergab den Radius. Damit ist es möglich, das unterschiedliche Maß an Bewegung in den verschiedenen Trajektorien zu erfassen. Als Referenzkoordinaten wurden die Koordinaten des Wassermolekül-O-Atoms in der Kristallstruktur, nach der MD-Aufheizphase auf 300 K sowie nach der angenommenen Equilibrierungsphase von 500 ps gewählt. Die Ergebnisse für alle drei Referenzkonfigurationen sind unterschiedlich. Man kann argumentieren, dass die beiden aus der Simulation gewählten Referenzkoordinaten nur eine Momentaufnahme des Systems darstellen, während die Kristallstruktur eine Mittelung über einen längeren Zeitraum wiedergibt und zusätzlich eine experimentell überprüfbare Grundlage hat; andererseits sind Abstände, Winkel und damit Atompositionen hier nicht unbedingt optimal für das in der Simulation verwendete Kraftfeld.

Die Ergebnisse, die auf diese Weise erhalten werden, basieren somit auf Randbedingungen, die im Vorfeld der Berechnung festgelegt werden müssen. Die entsprechenden Informationen über charakteristische Wasserpositionen in der CCP W191G Bindetasche lassen sich jedoch auch auf andere Art gewinnen. Dies soll im Folgenden eingehender erörtert werden. Die so gewonnenen Erkenntnisse können danach aber genutzt werden, um Besetzungsgrade zu berechnen (vgl. Abb. 6.5 und 6.6).

### **Gemittelte Hydratationsdichtekarten**

In Abschnitt 5.1.1 wurde bereits die Berechnung der gemittelten Hydratationsdichtekarten für die simulierten Trajektorien erwähnt und in Abschnitt 6.1 wurden die Details zur Berechnung vorgestellt. Diese Art der Analyse hat sich als äußerst hilfreich bei der Erfassung der Wasserbewegung im Verlauf der simulierten Zeitspanne und bei der Charakterisierung einzelner Wasserpositionen in der Bindetasche erwiesen. Die Hydratationsdichtekarten werden allein aus der Trajektorie selbst berechnet, man benötigt im Vorfeld keine Informationen über die Koordinaten und Größe der zu untersuchenden Wasserpositionen. Neben dieser nützlichen Darstellung der Wasserverteilung eines Zeitabschnittes wurden außerdem die Maxima aus der Verteilung extrahiert, um die Ergebnisse zu quantifizieren. Im Folgenden werden diese Ergebnisse zusammen mit den Beobachtungen zur Wasserdynamik bei visueller Inspizierung der Trajektorien vorgestellt und diskutiert.

Die berechneten gemittelten Hydratationsdichtekarten sind in Abb. 6.4 zusammen mit repräsentativen Proteinkonformationen für die jeweilige Simulation aufgetragen. „Repräsentativ“ bedeutet hierbei, dass eine Konformation gewählt wurde, die den geringsten RMSD der Proteinatome in Bezug auf eine gemittelte Struktur hat. Diese Mittelung wurde über

alle Proteinatornkoordinaten für die gesamte Trajektorie berechnet. Die dargestellten Hydratationsdichten bei einem Kontur-Wert von 70 % bedeuten, dass während 70 % der simulierten Zeitspanne das O-Atom eines Wassermoleküls innerhalb des dargestellten Bereichs zu finden ist. Die aus diesen Karten extrahierten lokalen Maxima wiederum sind in Tab. 6.4 aufgeführt. Hydratationsdichtekarten und Maxima wurden nicht nur für die vollständige Trajektorie sondern zusätzlich separat für deren erste und zweite Hälfte ermittelt. Auf diese Weise können Änderungen im Verlauf der Simulation verfolgt werden. Nur die durch den Algorithmus zum Finden der lokalen Maxima klar separierbaren Dichtemaxima sind in Tab. 6.4 zusammengestellt. Dabei wurde Hydratationsdichte im direkten Umfeld dieser Maxima vernachlässigt. Da drei individuelle Analysen für die vollständige Trajektorie und die beiden Hälften durchgeführt wurden, bedingt dies, dass die Summe für die Gesamttrajektorie nicht dem Mittelwert aus erster und zweiter Hälfte entspricht.

**1CPE.** Durch die Bewegung des  $K^+$ -Ions in die Mitte der Bindetasche ändert sich selbstverständlich auch die Wasserstruktur während der Simulation im Vergleich zur Kristallstruktur. Das Kation ist umgeben von fünf wohl definierten Wasserpositionen, die abwechselnd durch die in der Bindetasche vorhandenen Moleküle besetzt sind. Die Position A hat dabei ein lokales Hydratationsdichte-Maximum von 86 %. Für Wasserposition B liegt das Maximum bei 71 % am nächsten. Man sieht also deutlich, dass das Vorhandensein des  $K^+$ -Ions wie erwartet die Wasserstruktur signifikant beeinflusst. Dieses System stellt deshalb nicht unbedingt die optimale Referenz für einen ligandfreien Zustand dar. Das Wassermolekül auf Position C wiederum ist hochgradig konserviert auf seiner Position und an dieser Stelle findet sich ein Dichtemaximum von 95 %. Die Darstellung der Dichte zeigt hier, wie im Folgenden auch an vielen anderen Stellen, annähernd Kugelform und weist somit auf die sehr geringe Bewegung/Fluktuation des Wassermoleküls hin. Dies wird durch die Werte in Tab. 6.3 untermauert. Hier sind sowohl die RMSD-Werte wie auch die Fluktuationen einzelner Wassermoleküle aufgetragen. In allen Systemen sind diese für Wassermolekül C sehr niedrig.

**1AEO.** In diesem System liegen keine gerichteten Wechselwirkungen zwischen dem Liganden 2-Aminopyridin und dem Wassermolekül auf Position A vor. Höchstwahrscheinlich als Konsequenz hiervon ist ein häufigerer Austausch der Wassermoleküle auf dieser Position zu beobachten. Die Werte in Tab. 6.3 wurden deshalb auch nur für einen Teil der Trajektorie berechnet und zeigen dennoch einen RMSD-Wert  $> 2 \text{ \AA}$ . Die Wasserposition ist jedoch wohl definiert und zu einem hohen Grad besetzt (80 %). Bei Wassermolekül C wird kein Austausch aber ebenfalls eine hohe Konservierung und ein hoher Besetzungsgrad beobachtet (99 %). Wählt man als Darstellungsgrenze der Hydratationsdichte ein niedrigeres Level, so erkennt man, dass an einer Position vergleichbar mit denen der Wassermoleküle 976 bzw. 977 aus der Struktur 1CPE zusätzliche Dichte (Maximum bei 45 %) beobachtet werden kann. Interessanterweise ist das genau die Position, auf der in vielen Docking-Simulationen (siehe Abschnitt 5.2.4) ein Wassermolekül vorhergesagt wurde.

**Tab. 6.3.:** Gemittelte RMSD Werte und Fluktuationen (Abkürzung Flukt.) in Å bezogen auf den Anfang einer Trajektorie. Aufgeführt sind die Werte für die Wassermoleküle auf den Positionen A, B und C, die dort die längste Aufenthaltszeit im Verlauf der Simulation zeigen. Die Werte wurden für die Aufenthaltszeit (Angabe in Klammern in ns) ermittelt. Falls keine Zeitspanne aufgeführt wird, wurden die Werte für die komplette Trajektorie und für das Wassermolekül berechnet, das am Beginn der Simulation auf der Position zu finden war.

System	Position A		Position B		Position C	
	RMSD	Flukt.	RMSD	Flukt.	RMSD	Flukt.
1CPE	3.17	2.24	-	-	1.20	0.62
1AEO	2.38	1.30 (0-4.75)	-	-	1.09	0.60
1AEF	1.49	0.88 (3.6-6.25)	4.70	0.62	1.14	0.58
1AEF (neu)	1.87	0.96	0.95	0.62	1.00	0.57
1AEG	1.45	1.18 (0-6.75)	-	-	1.29	0.56
2AQD	4.25	1.14 (4.35-6.15)	-	-	1.12	0.58
2EUP	3.10	0.88 (1-7)	-	-	1.11	0.56

**1AEF.** Einzelheiten zur Wasserdynamik dieses Systems finden sich in Abschnitt 6.3.

**1AEG.** Bei der Simulation des Komplexes von CCP W191G und 4-Aminopyridin existiert eine direkte Wechselwirkung zwischen Ligand und Wassermolekül A. Auf dieser Position wird auch so gut wie gar kein Austausch während der simulierten Zeitspanne beobachtet. Daraus resultierend liegt der Besetzungsgrad dieser Position mit 86% in einem hohen Bereich sowie Fluktuation und RMSD bei geringen Werten. Analog ist auch Position C hoch konserviert (Besetzungsgrad 99%).

**2AQD.** Für diesen Komplex zeigt die entsprechende Simulation einen häufigen Austausch und eine hohe Beweglichkeit an Position A. Dadurch ergeben sich zwei Dichtemaxima mit einem Besetzungsgrad von 83% und 76%. Das Wassermolekül, das am längsten dort verbleibt hat einen RMSD von 4.25 Å und eine Fluktuation von 1.14 Å. Auch hier besteht keine Wechselwirkung dieses Wassermoleküls zum Liganden. Erneut findet man Wassermolekül C während der ganzen Simulation an der ursprünglichen Position (Besetzungsgrad 99%).

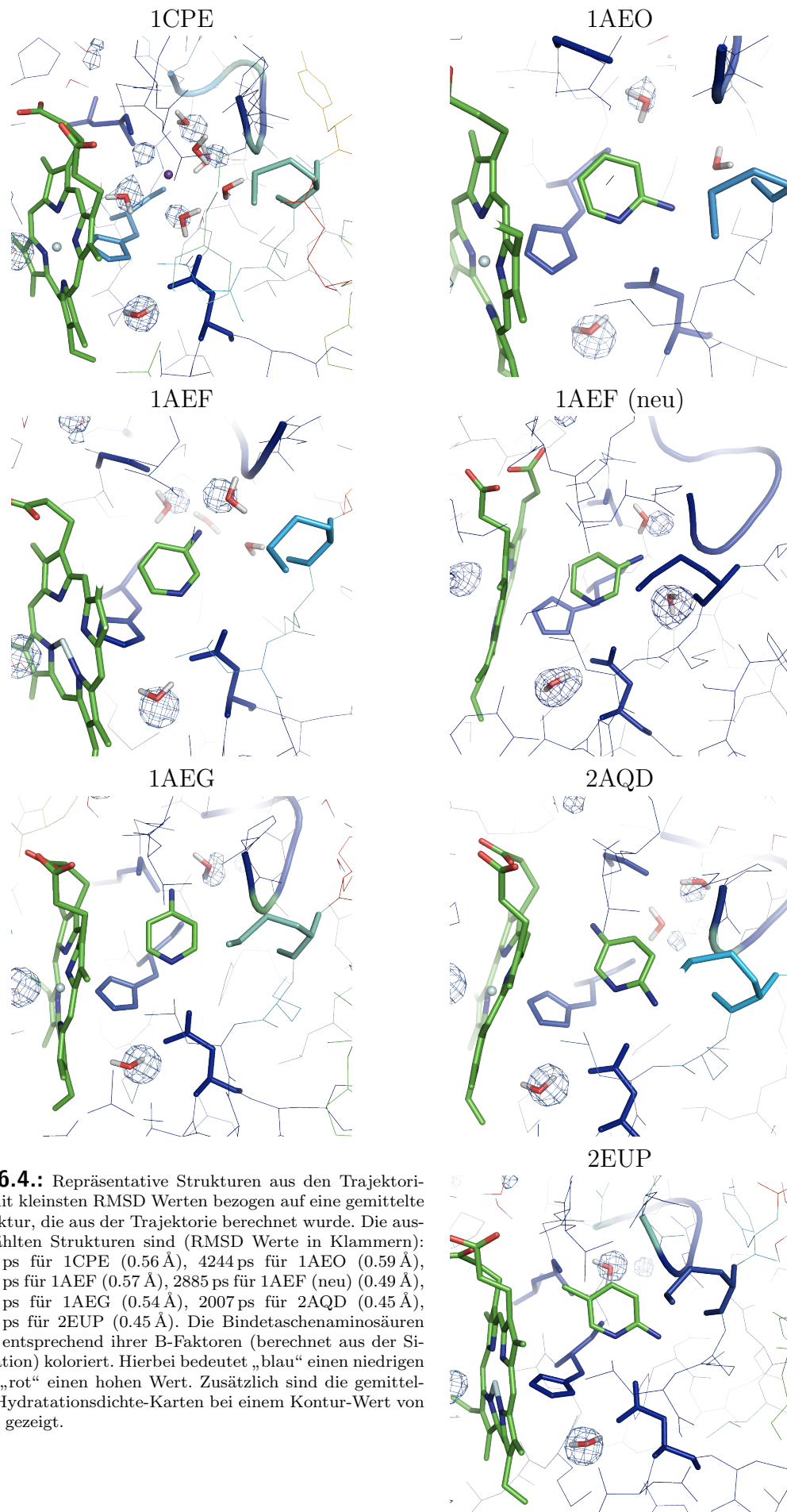
**2EUP.** Obwohl auch beim Liganden dieses Systems keine direkte Wechselwirkung mit dem Wassermolekül auf Position A ausgebildet wird, kann man keinen Austausch auf dieser Wasserposition beobachten. Das Maximum der Hydratationsdichte liegt bei 93%. Diese Beobachtung spiegelt sich auch im niedrigen Fluktuationwert ( $< 1$  Å) wider. Die hohe Konservierung von Position C schlägt sich ebenfalls in einem hohen Dichte-Wert (99%) nieder.

Beim allgemeinen Vergleich der Summe der Hydratationsdichtemaxima über die gesamte Trajektorie sowie der ersten und zweiten Hälfte mit der Anzahl der Wassermoleküle in der Kristallstruktur fällt auf, dass diese Werte für jedes der Systeme grob übereinstimmen. Tendenziell finden sich jedoch in den MD-Simulationen etwas mehr Wassermoleküle in der Bindetasche, als in der Röntgenstruktur, insbesondere bei den Systemen 1AEO, 1AEF (neu) und 2AQD. Für den Komplex 1AEO tritt die zusätzliche Wasserdichte (siehe oben) erst im späteren Verlauf der Simulation auf. Beim System 1AEF (neu) nimmt die Gesamt-Wasserdichte innerhalb der Bindetasche im Verlauf der Simulation ab. Der häufige Austausch auf Position A bringt es naturgemäß mit sich, dass zusätzliche Wassermoleküle die Bindetasche erreichen. Im System 2AQD hingegen beobachtet man teilweise zwei Wassermoleküle in unmittelbarer Nachbarschaft, die man beide Position A zuweisen könnte, und demnach auch einen erhöhten Wert der Hydratationsdichte im Vergleich zur experimentellen Struktur.

**Wasserdichte.** Basierend auf diesen Daten, die den Hydratationsgrad der Bindetasche beschreiben, entstand die Idee, lokale Wasserdichtewerte zu bestimmen. Dazu mussten jedoch zunächst die Volumina der Bindetasche in den unterschiedlichen Komplexen berechnet werden. Die Datenbank Relibase+ enthält für die darin abgelegten Bindetaschen auch eine Volumeninformation. Da die Tasche bei der Cytochrom *c* Peroxidase jedoch in unmittelbarer Nachbarschaft des Cofaktors zu finden ist, kann man der Datenbank nur ein Gesamtvolumen für beide Taschen entnehmen. Da eine einfache Abänderung des Algorithmus nicht möglich ist\*, musste eine Alternative für die Berechnung dieser Werte gefunden werden. Da es nicht gelang, mit Sybyl eine abgeschlossene Oberfläche der Bindetasche zu erzeugen, konnten hiermit nur die Volumina der Liganden bestimmt werden. Das Taschen-volumen wurde mit MOE auf Basis des „sitefinder“ Algorithmus mit Hilfe der so genannten „AlphaSpheres“ berechnet. Somit sollte das erhaltene Volumen dem Wert ausgehend von einer Connolly-Oberfläche entsprechen<sup>225</sup>. Die Ergebnisse sind in Tab. 6.5 zusammengefasst. Die Tabelle beinhaltet zusätzlich auch Daten, die auf Grundlage von Strukturen, die erst später in dieser Arbeit vorgestellt werden, berechnet wurden. Die Größe der ionenfreien Modellbindetasche liegt bei  $154 \text{ \AA}^3$  und ist somit etwas größer als die Taschen der übrigen Kristallstrukturen. Insgesamt ist die ermittelte Größe der Bindetasche jedoch sehr empfindlich gegenüber geringen strukturellen Änderungen. Beim System 1AEF liegt die Tasche in Richtung der Loop-Region in einer etwas anderen Konformation vor. Der Algorithmus erkennt so jedoch nicht, dass die Binderegion hier endet und zählt den außerhalb der Bindetasche liegenden Bereich bei der Volumenberechnung mit. Am Beispiel der Ligandvolumina sieht man, welche deutlichen Einflüsse die Wahl der das Volumen begrenzenden Oberfläche hat. Da für eine effektive Dichteberechnung die Differenz aus Taschen- und Ligandvolumen gebildet werden muss, stellt sich vor diesem Hintergrund die Frage, welche Aussagekraft die so berechneten Dichten haben können. Letztlich wurde deshalb auch keine Dichteberechnung vorgenommen.

---

\*Persönliche Mitteilung von Nils Weskamp



**Abb. 6.4.:** Repräsentative Strukturen aus den Trajektorien mit kleinsten RMSD Werten bezogen auf eine gemittelte Struktur, die aus der Trajektorie berechnet wurde. Die ausgewählten Strukturen sind (RMSD Werte in Klammern): 4286 ps für 1CPE (0.56 Å), 4244 ps für 1AEO (0.59 Å), 3412 ps für 1AEF (0.57 Å), 2885 ps für 1AEF (neu) (0.49 Å), 4582 ps für 1AEG (0.54 Å), 2007 ps für 2AQD (0.45 Å), 3092 ps für 2EUP (0.45 Å). Die Bindetaschenamino­säuren sind entsprechend ihrer B-Faktoren (berechnet aus der Simulation) koloriert. Hierbei bedeutet „blau“ einen niedrigen und „rot“ einen hohen Wert. Zusätzlich sind die gemittelten Hydratationsdichte-Karten bei einem Kontur-Wert von 70 % gezeigt.

**Tab. 6.4.:** Auflistung der lokalen Maxima sowie deren Summe der gemittelten Hydratationsdichtekarten für die betrachteten MD-Simulationen. Die Wasserdichte auf Position C (nahe Asp235 außerhalb der Bindetasche) ist als Vergleich angegeben und mit einem \* gekennzeichnet. Die Summe wurde nur für die Wasserdichte nahe am Liganden, also ohne Position C, bestimmt. Zum Vergleich hierfür ist wiederum in der letzten Zeile die Anzahl der Wassermoleküle nahe am Liganden in der Kristallstruktur aufgeführt.

System	1CPE	1AEO	1AEG	1AEF	1AEF (neu)	2AQD	2EUP
<b>Maxima</b>	0.95*, 0.86, 0.85, 0.81, 0.80, 0.79, 0.71, 0.51	0.99*, 0.80, 0.70, 0.45, 0.38	0.99*, 0.86	0.99*, 0.90, 0.83, 0.69	1.00*, 0.99, 0.90, 0.56, 0.33, 0.33	0.99*, 0.83, 0.76	0.99*, 0.93
<b>Summe</b>	5.33	2.33	0.86	2.42	3.11	1.59	0.93
<b>Maxima 1. Hälfte</b>	0.98*, 0.89, 0.88, 0.88, 0.86, 0.85, 0.71, 0.67	0.99*, 0.90, 0.88	0.99*, 0.96	0.99*, 0.87, 0.79, 0.74	1.00*, 0.99, 0.93, 0.43, 0.34, 0.20	1.00*, 0.78, 0.78	0.99*, 0.90
<b>Summe 1. Hälfte</b>	5.74	1.78	0.96	2.40	2.89	1.56	0.90
<b>Maxima 2. Hälfte</b>	0.92*, 0.86, 0.86, 0.85, 0.82, 0.75, 0.70	1.00*, 0.91, 0.81, 0.63, 0.62	0.99*, 0.86, 0.79	0.98*, 0.97, 0.87, 0.66, 0.59, 0.39	1.00*, 1.00, 0.93, 0.45, 0.32	0.99*, 0.89, 0.78	0.98*, 0.96
<b>Summe 2. Hälfte</b>	4.84	2.97	1.65	3.51	2.70	1.67	0.96
<b>Kristall</b>	5	1	1	2	2	1	1



**Tab. 6.5.:** Mit MOE berechnete Bindetaschenvolumina verschiedener CCP W191G Strukturen. Als strukturelle Grundlage diente bei den Werten für die MD-Simulationen diejenige Proteinkonformation, die die geringste RMSD-Abweichung zur Mittelwertsstruktur aufwies. Das Subskript „lr“ steht dabei für die Trajektorien, die unter Anwendung von Zusatzkräften auf die PGGAAN-Schleifenatome ermittelt wurden. Weiterhin sind Ligandvolumina aufgeführt, die mit dem Programm SYBYL berechnet wurden. Der erste Wert davon basiert auf der Connolly-, der zweite auf der van-der-Waals-Oberfläche. Alle Angaben sind in  $\text{\AA}^3$ . Die Werte sind mit keiner, einer oder zwei Nachkommastellen Genauigkeit angegeben, wie dies vom jeweiligen Programm berechnet wurde.

System	$V_{\text{Kristall}}$	$V_{\text{MD}}$	$V_{\text{MDlr}}$	$V_{\text{Ligand}}^1$	$V_{\text{Ligand}}^2$
1CPE	115	134	161	13.40	9.2
1AEO	124	133	114	97.90	85.4
1AEF	126	334	277	98.02	85.7
1AEF (neu)	110	92	125	98.02	85.7
1AEG	126	132	151	97.96	84.8
2AQD	114	137	140	113.79	100.2
2EUP	120	128	106	116.32	101.4

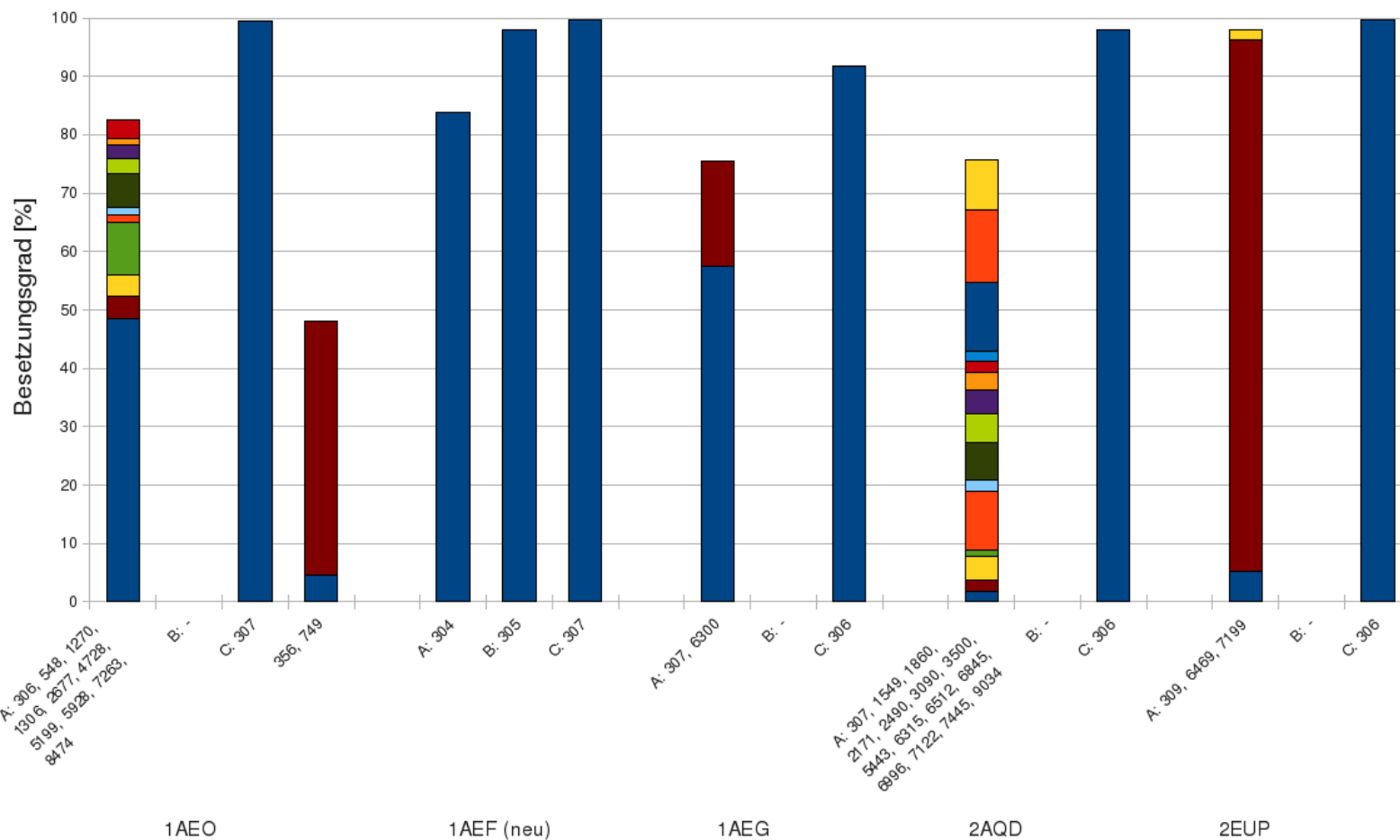
### Austauschvorgänge auf den einzelnen Wasserpositionen

An dieser Stelle wird nun wieder auf die eingangs diskutierte Berechnung von Besetzungsgraden einzelner Wasserpositionen mittels des selbst entwickelten Skriptes zurückgegriffen. Die damals diskutierte Fragestellung, welche Referenzposition für die Berechnungen gewählt werden soll, ist nun sehr einfach zu beantworten. Am besten eignen sich hierfür die Maxima der Hydratationsdichteverteilungen. Da diese ja direkt aus der Trajektorie berechnet wurden, ist nun die Auswertung frei von Vorinformationen möglich. Die auf diese Weise berechneten Besetzungsgrade für die untersuchten Komplexe sind in Abb. 6.5 dargestellt. Wie oben beschrieben, wurde hierbei der Mittelwert des RMSD der Bindetaschenamino-säuren als Kugelradius in der Berechnung verwendet. Die absolute Größe der Besetzungsgrade unterscheidet sich hierbei von den Maxima der Hydratationsdichteverteilung. Die beobachteten relativen Unterschiede werden aber hier auch sehr gut wiedergegeben. Die absoluten Größenunterschiede kommen durch die unterschiedliche Berechnungsweise zustande. Bei den Hydratationsdichteverteilungen wird ein Gitter über den zu untersuchenden Bereich gelegt. Liegt ein Gitterpunkt innerhalb des van-der-Waals Radius eines O-Atoms, so wird der Dichtewert am Gitterpunkt erhöht. Das daraus resultierende Maximum liegt nun zwangsläufig auf einem Gitterpunkt. Um diesen Gitterpunkt wird in der hier dargestellten Auswertung wiederum eine Kugel definiert, in der die Koordinaten eines O-Atoms liegen müssen. Wassermoleküle, deren Abstand  $d$  des O-Atoms zum Gitterpunkt des Maximums größer ist als der Radius zum Berechnen des Besetzungsgrades  $r_{\text{occ}}$ , werden also nicht erfasst. Durch das Programm VMD wurde ein spezielles Solvensmolekül aber

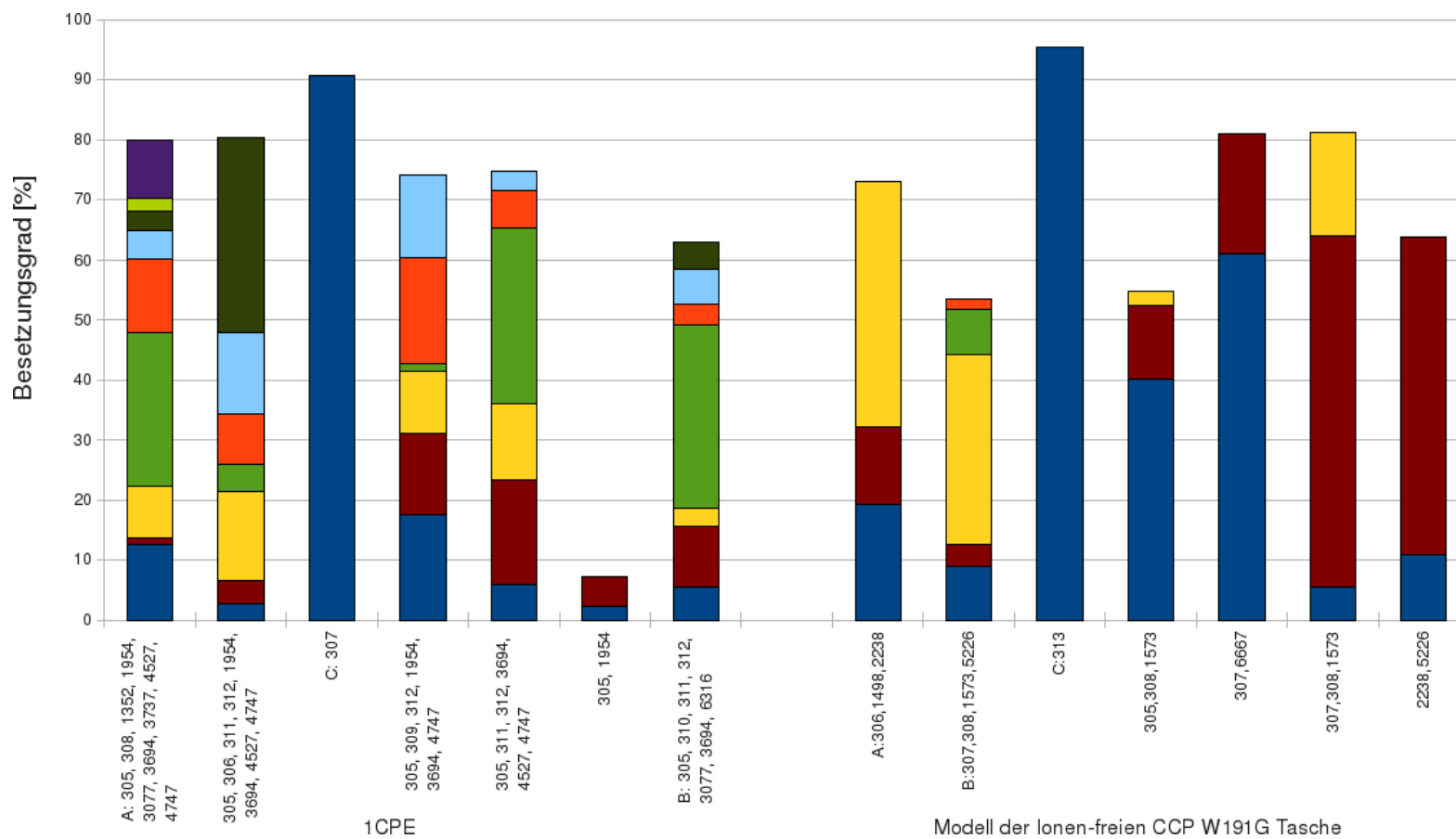
immer dann berücksichtigt, wenn  $d < r_{\text{vdW}}$  ist. Für den Fall, dass  $r_{\text{occ}}$  kleiner ist als der van-der-Waals Radius  $r_{\text{vdW}}$  des O-Atoms, werden durch das eigene Skript die Fälle nicht erfasst, die aber zuvor im Volmap-Skript berücksichtigt wurden. Grundsätzlich wäre es also eine Alternative, als Referenzradius an Stelle der Fluktuation der Bindetaschenamino-säuren den van-der-Waals Radius eines Sauerstoffatoms zu verwenden. Da dieser Wert mit  $1.77 \text{ \AA}$  relativ groß ist, lagen die gewählten Radien stets darunter. Eine damit neu durchgeführte Berechnung zeigte für fast alle Positionen Besetzungsgrade um 100 %. Durch den sehr großen Kugelradius nivellieren sich die feinen Unterschiede zwischen den einzelnen Regionen.

Die dargestellten Auftragungen zeigen sehr übersichtlich, welche unterschiedlichen Wassermoleküle im Verlauf der Simulation eine Position besetzen. Durch das Volmap Plugin geht diese Information verloren. Dieses Ergebnis ist dabei unabhängig von der Wahl des Kugelradius und der absoluten Größe der Besetzungszahl. Man erkennt, dass nahezu alle Wasserpositionen der Strukturen fast ausschließlich durch ein Wassermolekül besetzt sind. Bei der neu auftretenden Position in 1AEO und bei Position A in 1AEG finden sich zwei Wassermoleküle, wobei eines von beiden den überwiegenden Teil der Simulationszeit auf der Position verweilt. Dies ist auch bei Position A in 2EUP der Fall, die insgesamt von drei Wassermolekülen aufgesucht wird. Ganz anders ist dies bei Position A der Systeme 1AEO und 2AQD. Beide Bereiche werden während der simulierten Zeitspanne von vielen verschiedenen Wassermolekülen besetzt. Speziell im System 2AQD ist hierbei überhaupt kein Wassermolekül erkennbar, das diese Position mit einer absoluten Mehrheit bevorzugt einnimmt.

Analog wurde die Auswertung auch für CCP W191G ohne Liganden, das heißt für die MD-Simulation der Kristallstruktur 1CPE sowie der ionenfreien Modelltasche, durchgeführt. Die entsprechenden Ergebnisse sind in Abb. 6.6 aufgetragen. Im Fall des Bindetaschenmodells wurde nur der Trajektorienabschnitt zwischen 4 und 5 ns verwendet, in dem die Wasserpositionen bereits equilibriert sind. Die Ergebnisse sind hier, im Vergleich zum Komplex, zum Teil anders. Zwar sind die Wasserpositionen im System 1CPE, wie bereits beschrieben, sehr gut konserviert, jedoch tauschen die Wassermoleküle häufig ihre Positionen innerhalb der Bindetasche und besetzen so verschiedene Regionen. Genau diese Dynamik spiegelt sich im gezeigten Diagramm wider. Es handelt sich hierbei um diskrete Positionswechsel der einzelnen Wassermoleküle und nicht um z.B. eine Rotation des gesamten Wasserclusters. Der Ausschnitt aus der 1 ns langen Trajektorie nach Equilibrierung der Wasserpositionen in der Modelltasche zeigt nur sehr wenige Austauschvorgänge zwischen den Positionen. Weitere Auswertungen der Simulation der ionenfreien Modellstruktur werden in Abschnitt 6.5 diskutiert.



**Abb. 6.5.:** Besetzungsgrad von Wasserpositionen anteilig für die unterschiedlichen Solvensmoleküle auf diesen Positionen für verschiedene Komplexe der CCP W191G.



**Abb. 6.6.:** Besetzungsgrad von Wasserpositionen anteilig für die unterschiedlichen Solvensmoleküle auf diesen Positionen für die zwei ligandfreien Strukturen der CCP W191G.

### 6.3. Weitergehende Untersuchungen am Komplex von CCP W191G mit 3-Aminopyridin

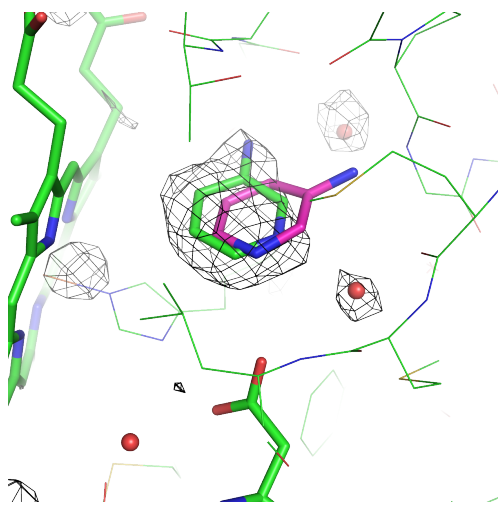
Die Sonderrolle, die dem System 1AEF aufgrund der Docking-Ergebnisse, nicht zuletzt aber vor allem aufgrund der Ergebnisse der MD-Simulation zukommt, wurde schon mehrfach angedeutet. In diesem Abschnitt soll deshalb ein detaillierterer Blick auf die Simulationsdaten dieses Komplexes geworfen werden.

In der Trajektorie, die ausgehend von der in der PDB deponierten 1AEF-Struktur berechnet wurde, beobachtet man bereits nach relativ kurzer Simulationsdauer (ca. 600 ps) eine Drehung des Liganden 3-Aminopyridin in der Bindetasche, sodass eine direkte Wechselwirkung zwischen Pyridin-N und Asp235 möglich ist. Der in Abb. 6.10 dargestellte Abstand zum rigiden Leu174 ist charakteristisch für diese Drehung. Während der simulierten Zeitspanne kehrt der Ligand nur sehr kurz in seine ursprüngliche Orientierung zurück. Mit dieser Drehung geht eine Bewegung der flexiblen Loop-Region im Bereich PGGAAN 190-195 einher (max. RMSD der Rückgrat-Atome in Bezug auf Startstruktur 2.10 Å, Durchschnitt 1.13 Å). Selbstverständlich ist diese signifikante Strukturänderung auch mit einer Änderung der Hydrathülle in der Bindetasche verbunden. Die Position B ist nicht länger günstig und das Wassermolekül bewegt sich in eine Region oberhalb (bezogen auf Abb. 6.4) des Liganden, wo eine direkte Wechselwirkung mit der Amino-Gruppe möglich wird. Andererseits wird durch die etwas geöffnete Konformation der Loop-Region ein häufigerer Austausch für Wassermoleküle auf Position A möglich. Diese Beobachtungen spiegeln sich selbstverständlich in der Hydratationsdichte-Verteilung wie auch z.B. in dem sehr hohen RMSD-Wert des Wassermoleküls B (4.70 Å) wider.

Bei diesem zu einem sehr hohen Anteil populierten Bindungsmodus stellt sich selbstverständlich die Frage, ob es für dieses Ergebniss auch experimentelle Hinweise gibt. In Abb. 6.7 ist die Elektronendichte zusammen mit der daraus abgeleiteten Orientierung des Liganden sowie der in der MD-Simulation bevorzugten Konfiguration dargestellt. Man kann deutlich erkennen, dass kein experimenteller Hinweis auf eine zweite Ligandorientierung vorliegt.

Die Vermutung, dass ein Fehler bei den Einstellungen für die MD-Simulation dazu geführt hat, dass dieses unerwartete Resultat erhalten wurde, lag nun sehr nahe. Aus diesem Grund wurden zahlreiche Analysen und Testsimulationen durchgeführt, um das ursächliche Problem zu erkennen. Diese Arbeiten umfassen:

- MD-Simulation mit einer Startkonfiguration, in der die Amino-Gruppe des Liganden in Richtung der Stelle von Wasser B orientiert ist. Das entsprechende Wassermolekül wurde deshalb aus der Startstruktur entfernt. Die Idee für diese Ligand-Orientierung liegt in der Vermutung begründet, dass möglicherweise kein zweites Wassermolekül in der Bindetasche vorhanden ist und die entsprechende Elektronendichte dem N-Atom der Amino-Gruppe zuzuordnen ist. Die Trajektorie zeigt zwar eine stabile Ligand-Orientierung aber einen steil ansteigenden Verlauf der RMSD-Werte der C<sub>α</sub>-Atome (Daten hier nicht gezeigt) und dementsprechend eine deutliche strukturelle Änderung des Proteins.

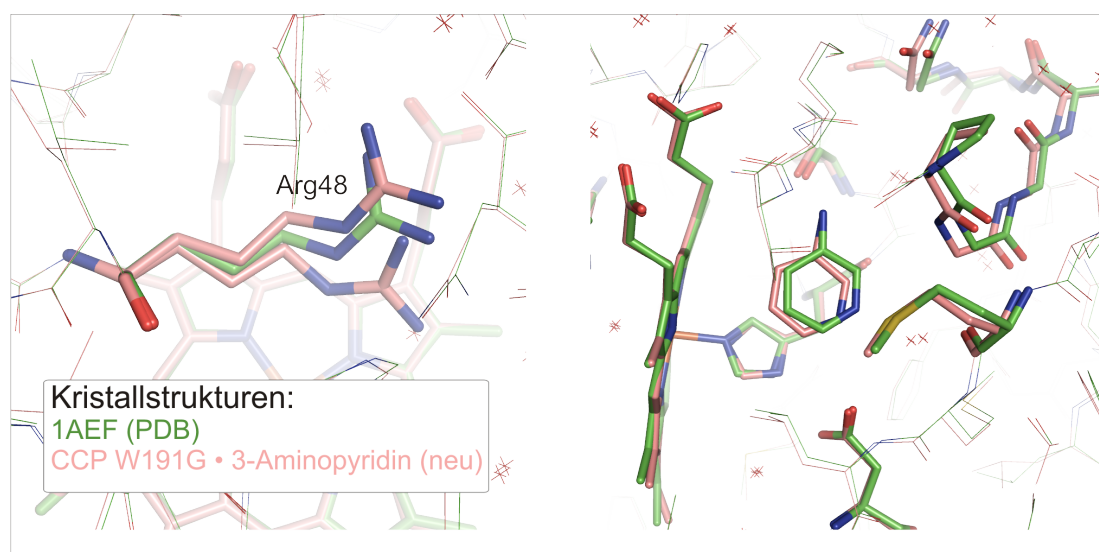


**Abb. 6.7.:** Kristallstruktur 1AEF (grün) zusammen mit dem Liganden in neuer Orientierung entsprechend der MD-Simulation (Konfiguration nach 600 ps in Violett) im Vergleich mit der  $3.0\sigma F_o - F_c$  Elektronendichte (Konturlinien).

- Die entsprechende Simulation wurde auch zusätzlich mit Wasser B (platziert direkt neben der Amino-Gruppe) durchgeführt. Auch hier wird ein stabiler Ligand-Bindungsmodus in der simulierten Zeitspanne beobachtet. Nach ca. 1.5 ns öffnet sich der PGGAAN Loop teilweise und zahlreiche Wassermoleküle strömen in die Binde-tasche ein.
- Um zu überprüfen, ob neben der Startorientierung des Liganden möglicherweise auch dessen Parametrisierung falsch gewählt wurde, wurde eine Simulation durchgeführt, bei der ein neutrales 3-Aminopyridin (deprotonierter Pyridinstickstoff) eingesetzt wurde. Hierbei dreht sich der Ligand jedoch bereits während der Equilibrierungsphase in eine veränderte Position.

Weiterhin denkbar sind natürlich auch ein nicht adäquates Ladungsmodell zur Parametrisierung des Liganden oder ein anderer Protonierungszustand von Asp235 in Gegenwart eines Liganden. In diesem Falle ist es aber fraglich, warum diese Parameter nur in einem Fall und nicht auch für die anderen vier Liganden von so großer Bedeutung sein sollten. Prinzipiell wäre es natürlich ebenso möglich, dass die Beschreibung durch das gewählte Kraftfeld grundsätzlich ungeeignet ist. Fraglich ist dann jedoch erneut, warum das nur bei einem der Komplexe offensichtliche Auswirkungen hat. Außerdem verwendeten die Docking-Programme andere Parametrisierungen, bevorzugen jedoch auch teilweise die veränderte 3-Aminopyridin Orientierung. Die wahrscheinlichste Ursache bleibt also weiterhin struktureller Natur.

Betrachtet man die eingesetzte Kristallstruktur näher, so fällt neben ihrer schlechten Auflösung (2.1 Å) auf, dass die Struktur nicht vollständig verfeinert wurde und alle Proteinkoordinaten aus der Struktur 1CMQ übernommen wurden (siehe Abschnitt 5.1.1). Es wäre also prinzipiell möglich, dass aufgrund dieser Defizite in der Kristallstruktur eine zweite Ligandorientierung unentdeckt geblieben ist. R. Wahlström aus der Gruppe von R. Brenk in Dundee (Schottland) gelang es, diesen Komplex erneut zu kristallisieren und eine hochaufgelöste Struktur (1.25 Å) zu bestimmen.<sup>226</sup> Diese zeigt (Abb. 6.8 rechts), dass



**Abb. 6.8.:** Vergleich zwischen der 1AEF-Struktur, wie sie in der PDB abgelegt ist, und der neu kristallisierten und gelösten Struktur. Besonders hervorgehoben sind die unterschiedlichen Konformationen der Seitenkette von Arg48 (links) und die hohe Übereinstimmung der Konformation im Bereich der Bindetasche (rechts).

auch hier keine Hinweise auf einen anderen als den zuvor bekannten experimentellen Bindungsmodus vorliegen. Desweiteren wird auch hier die Präsenz von zwei konservierten Wassermolekülen in der Bindetasche bestätigt. Die Unterschiede zwischen dieser und der Original-1AEF-Struktur sind gering. Der RMSD aller Protein-Atome beträgt  $0.27 \text{ \AA}$  zwischen alter und neuer Struktur. Berechnet man den entsprechenden Wert nur für das Protein-Rückgrat, so beträgt dieser  $0.24 \text{ \AA}$ . Der offensichtlichste Unterschied zwischen beiden Strukturen ist die alternative Seitenkettenkonformation des Arg48. Zu 40 % ist die Konformation aus der Original-Struktur besetzt. Zu 60 % liegt die Arg-Seitenkette in einer Konformation näher am Cofaktor vor, ähnlich zu jener in den Strukturen 2AQD oder 2EUP. Ein weiterer Unterschied besteht darin, dass in der neuen Struktur Aminosäure 272, die in 1AEF noch fälschlicherweise als Asp gekennzeichnet war, zu Asn korrigiert wurde. Da diese jedoch sehr weit von der Bindetasche entfernt zu finden ist, wird sie keinen Einfluss auf die Dynamik des Liganden ausüben.<sup>†</sup>

Mit der neuen Komplexstruktur von CCP W191G und 3-Aminopyridin wurde eine neue MD-Simulation mit den analogen Einstellungen wie zuvor durchgeführt. Als Seitenkettenkonformation wurde die gewählt, die nicht der in der ursprünglichen 1AEF PDB-Struktur entsprach. Der Verlauf der RMSD-Werte, der Anteil der Sekundärstrukturelemente sowie Deskriptoren für die Wasserdynamik finden sich oben. Die gewählte Seitenkettenkonformation zeigt sich als stabil während der Simulation. Der Torsionswinkel, der charakteristisch

<sup>†</sup>Um das zu überprüfen, wurde zusätzlich eine MD-Simulation mit Asp272 durchgeführt, die zu den gleichen Ergebnissen wie mit Asn272 und alternativer Arg48 Seitenkette kommt.

für die Orientierung der Seitenkette von Arg48 ist, ist in Abb. 6.9 aufgetragen. In der ursprünglichen Kristallstruktur liegt der Wert dafür bei  $164.5^\circ$  (grüne Linie zum Vergleich in Abb. 6.9). In der neu gelösten Kristallstruktur von CCP W191G im Komplex mit 3AP findet man die Werte  $166.1^\circ$  und  $-164.1^\circ$ . Die Konformation, der der letztgenannte Winkel zuzuordnen ist, wurde für die Generierung der neuen Trajektorie verwendet. Im Mittel zeigt sich im Verlauf der Simulation für diese Torsion ein Winkel von  $-176.0^\circ$ . Auch der mittlere RMSD-Wert dieser Aminosäure liegt mit  $0.49 \text{ \AA}$  in einem niedrigen Bereich. Der anhand der Fluktuation in der Trajektorie berechnete B-Faktor ist mit 15.2 unter dem Durchschnitt aller Aminosäuren (26.8).

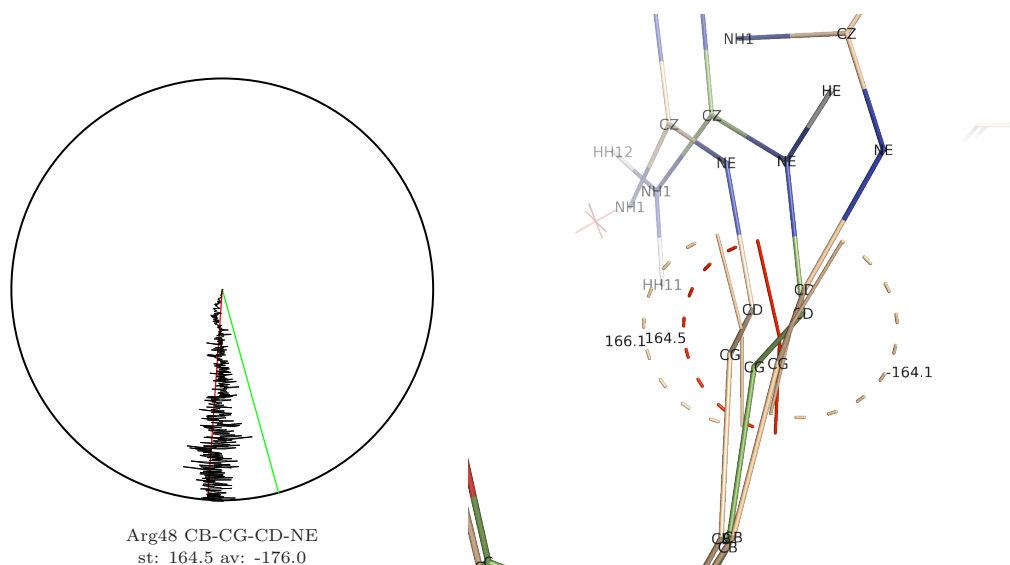
Die Bewegung des Liganden spiegelt sich sowohl im Verlauf der RMSD-Werte des Liganden (Abb. 6.2) als auch im für die Rotation charakteristischen Abstandsdiagramm (Abb. 6.10) wider. Man sieht, dass sich der Ligand auch in dieser Trajektorie nach ca. 1000 ps in den zweiten Bindungsmodus dreht. Bereits nach ca. 500 ps kehrt er jedoch zurück in die ursprüngliche Orientierung. Nach ca. 2600 ps Simulationsdauer wechselt er erneut zurück in die neue Lage, kehrt aber auch daraus in den ursprünglichen Bindungsmodus zurück. Insgesamt überwiegt deutlich der native Bindungsmodus. In dieser Simulation geht mit der Neuorientierung des Liganden auch keine Bewegung in der PGGAAN Loop-Region einher. Der RMSD der Loop-Rückgrat-Atome liegt im Mittel bei  $0.46 \text{ \AA}$  (Maximum  $1.06 \text{ \AA}$ ). Das veränderte Bewegungsmuster des Liganden hat selbstverständlich entscheidende Auswirkungen auf die Dynamik der Wassermoleküle in der Bindetasche. Alle drei Positionen A, B und C sind zu einem hohen Maße konserviert und zeigen Dichtemaxima von 80 %, 99 % und 100 %. Auf den Positionen B und C wird kein Austausch beobachtet. Position A hingegen wird im Verlauf der Simulation von unterschiedlichen Wassermolekülen besetzt. RMSD und Fluktuation (vgl. Tab 6.3) liegen hier auch etwas höher als für die ursprüngliche 1AEF-Trajektorie, jedoch im Falle von Wasser A deutlich niedriger als für Komplex 2AQD. Insgesamt konnte auch in dieser Simulation keine zusätzliche Wasserposition, die während der Simulation über einen längeren Zeitraum besetzt wird, beobachtet werden.

Insgesamt bleibt festzuhalten, dass die mangelnde Verfeinerung der Original-Kristallstruktur eine wesentliche strukturelle Information vorenthalten hat. Diese wiederum hat einen signifikanten Einfluss auf die Generierung der Trajektorie und damit auf die Wasserdynamik. In der aus der neuen Simulation erhaltenen Trajektorie konnten die experimentell bestimmten Wasserpositionen in der Bindetasche bestätigt werden.

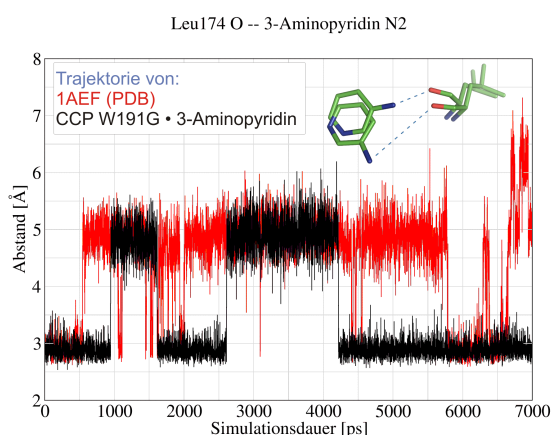
### 6.4. Vergleich mit Trajektorien mit artifiziellen Zusatzkräften

Eine überraschende Erkenntnis der oben diskutierten MD-Simulationen war, dass ein relativ häufiger Wasseraustausch für eine vergrabenen Bindetasche wie die der CCP W191G möglich ist. Da dies zu einigen technischen Schwierigkeiten bei der Durchführung der TI MD-Simulationen (siehe Abschnitt 7.1.3) führt, wurde ein Versuch unternommen, diese Austauschvorgänge innerhalb der betrachteten Zeitskala zu verhindern. Die Loop-Region





**Abb. 6.9.:** Links: Verlauf des Torsionswinkels in Arg48 ( $C_{\beta}$ - $C_{\gamma}$ - $C_{\delta}$ - $N_{\epsilon}$ ) in der Simulation der neu kristallisierten 3-Aminopyridin-Komplexstruktur. In Grün ist zum Vergleich der Winkel in der 1AEF Kristallstruktur und in Rot der Mittelwert während der Simulation aufgetragen. Rechts: Vergleich der Torsionswinkel in der 1AEF Kristallstruktur (grün) und dem neu kristallisierten Analogon (braun).



**Abb. 6.10.:** Charakteristischer Abstand (zwischen Leu174 und Liganden) für die Umorientierung von 3-Aminopyridin in der Bin-detasche während der MD-Simulation. Der schwarze Graph zeigt die Veränderung dieser Distanz für die Trajektorie, die aus der neuen Ausgangsstruktur erstellt wurde, während der rote die entsprechende Größe aus der Simulation der PDB-Struktur 1AEF darstellt.

der Aminosäuren Pro190-Asn195 ist bekannt für ihre Flexibilität und die Möglichkeit eine Konformationsänderung einzugehen, welche wiederum ein Tor zur Bindetasche öffnet.<sup>91</sup> Da für den Ein- und Austritt von Wassermolekülen eine weitaus kleinere Öffnung notwendig ist, als für einen Liganden, lag die Vermutung nahe, dass bereits geringe Konformationsänderungen in dieser Loop-Region die Möglichkeit des Wasseraustausches - vor allem an der benachbarten Position A - ermöglichen.

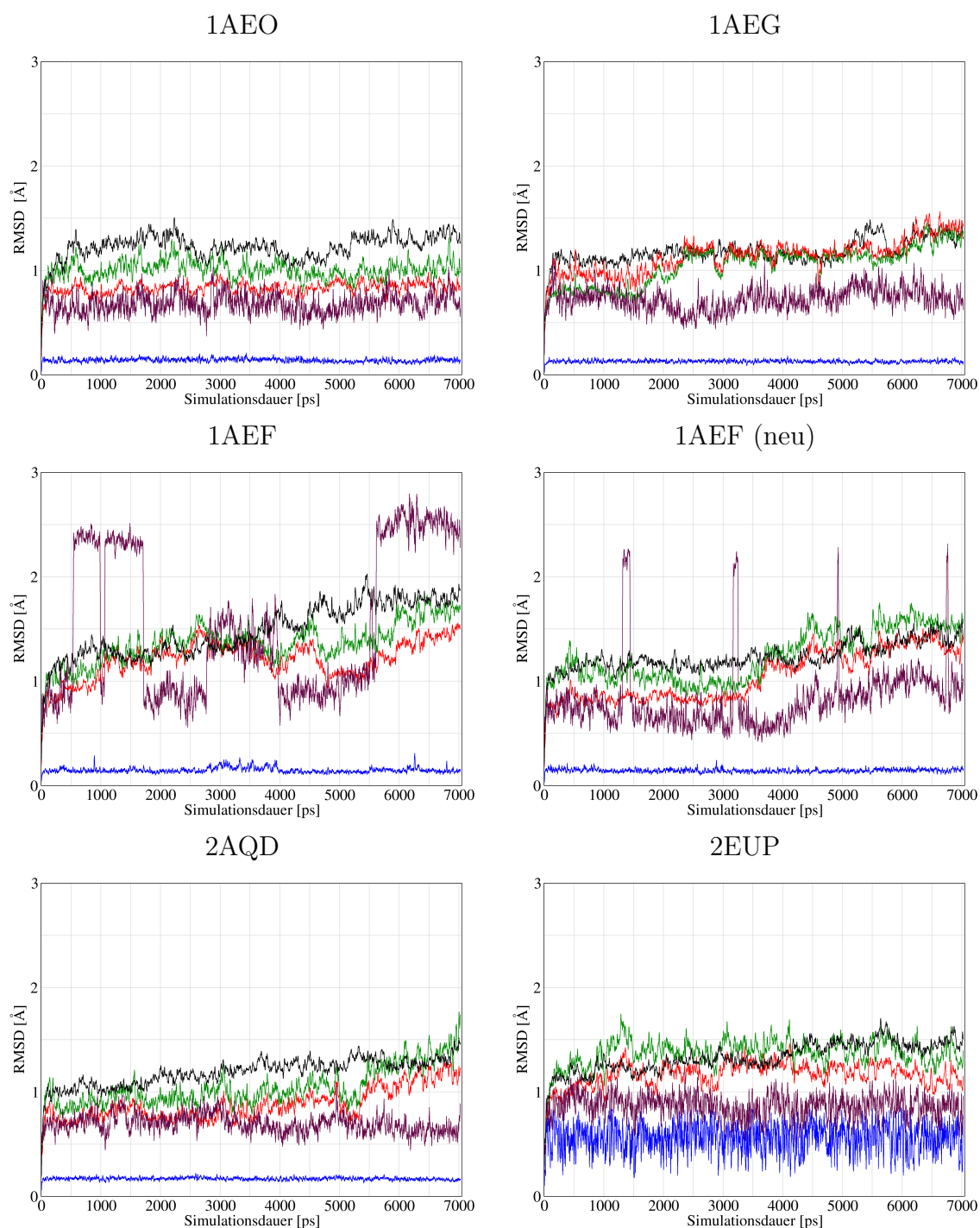
Aufgrund dieser Vorüberlegungen wurde von allen Komplexen eine weitere MD-Simulation über 7 ns durchgeführt, in der geringe Zusatzkräfte (Kraftkonstante 5.0 kcal/mol) auf die C<sub>α</sub>-Atome der Aminosäuren P190-G191-G192-A193-A194-N195 ausgeübt wurden, die diese mit einem harmonischen Potential zurück in ihre Ausgangsposition am Beginn der Simulation ziehen. Alle übrigen Einstellungen und Parameter wurden wie in den vorangegangenen Simulationen gewählt.

In den Abb. 6.11 und 6.12 finden sich Auftragungen der RMSD-Werte für C<sub>α</sub>-, Ligand- und Bindetaschenatome sowie der Besetzungsgrad unterschiedlicher Sekundärstrukturanteile im Verlauf der simulierten Zeitspanne. Es ist deutlich zu erkennen, dass die Tertiärstruktur der Komplexe im Verlauf der Simulationen erhalten bleibt. Insgesamt sind jeweils nur wenige strukturelle Änderungen der Proteinstruktur zu beobachten. Die Orientierung des Liganden ändert sich auch hier in der 1AEF-Simulation. Da aber nun die Loop-Region nicht auf diese Änderung reagieren kann, kommt es hier zu größeren strukturellen Änderungen der Bindetaschen-Aminosäuren, die nicht durch Zusatzkräfte in ihrer Position festgehalten werden. Die entsprechende Simulation ausgehend von der neuen Kristallstruktur des Komplexes zwischen CCP W191G und 3-Aminopyridin zeigt ebenfalls Umorientierungen des Liganden. Dies hat hier, wie man am Verlauf der RMSD-Werte sieht, jedoch keine Adaption der Protein-Umgebung zur Folge, sondern resultiert in einer sofortigen Rückkehr des Liganden in seinen ursprünglichen Bindungsmodus.

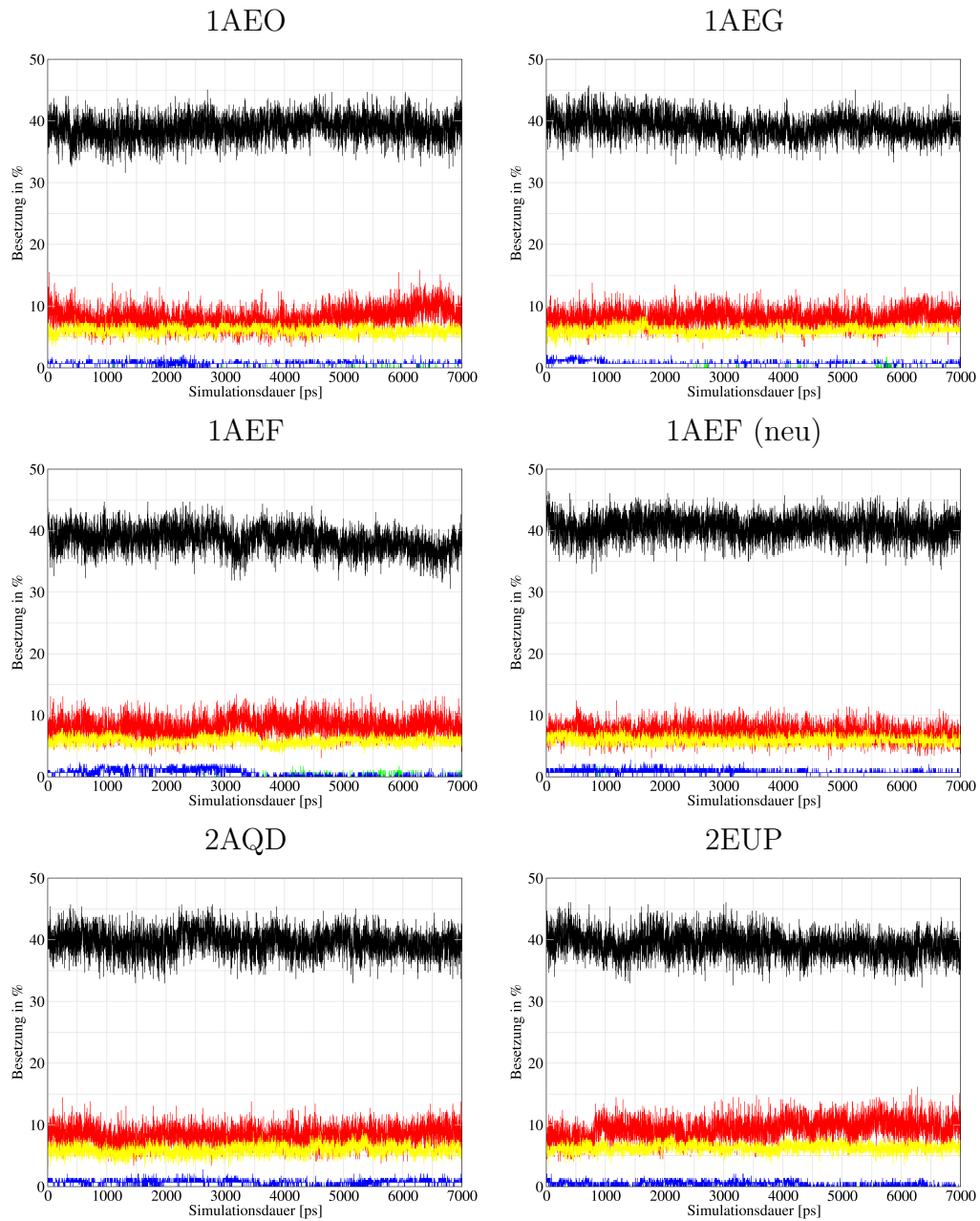
Die eingeführten Zusatzkräfte erfüllten in den erhaltenen Trajektorien im Übrigen genau den Zweck, für den sie eingeführt wurden. In keiner der durchgeführten Simulationen konnten Austauschvorgänge für die Wassermoleküle auf den Positionen A, B oder C mit Wassermolekülen aus der umgebenden Solvensphase beobachtet werden. In Abb. 6.13 sind die berechneten Hydratationsdichtekarten aufgetragen. In Tab. 6.6 finden sich die zugehörigen lokalen Maxima. Mit Ausnahme des 1AEF-Systems, in dem es deutliche Änderungen der Bindetaschenstruktur zu verzeichnen gibt, sind die charakteristischen Wasserpositionen aller Bindetaschen zu einem hohen Maße konserviert und besetzt.

Der Vergleich der Summe der Hydratationsdichtemaxima mit den entsprechenden Werten aus den Trajektorien ohne Zusatzkräfte zeigt, dass die Werte nun insgesamt niedriger liegen als zuvor. Durch den fehlenden Austausch mit dem umgebenden Solvens sind temporär keine zusätzlichen Wassermoleküle im Bereich der Bindetasche zu beobachten, die zur Wasserdichte beitragen könnten. Vergleicht man die Summen mit der Anzahl der Wassermoleküle in den Röntgenstrukturen, kann man deutliche Übereinstimmungen feststellen. Im zweiten Teil der 1AEO Trajektorie bewegt sich Wassermolekül A zwischen zwei benachbarten Positionen hin- und her. Bei der Berechnung der Hydratationsdichte wird jedoch ein Wassermolekül mehreren Gitterpunkten zugeordnet und kann somit gleichzeitig zu zwei sehr nahe beieinander liegenden Maxima beitragen. Aus diesem Grund ist der gegenüber

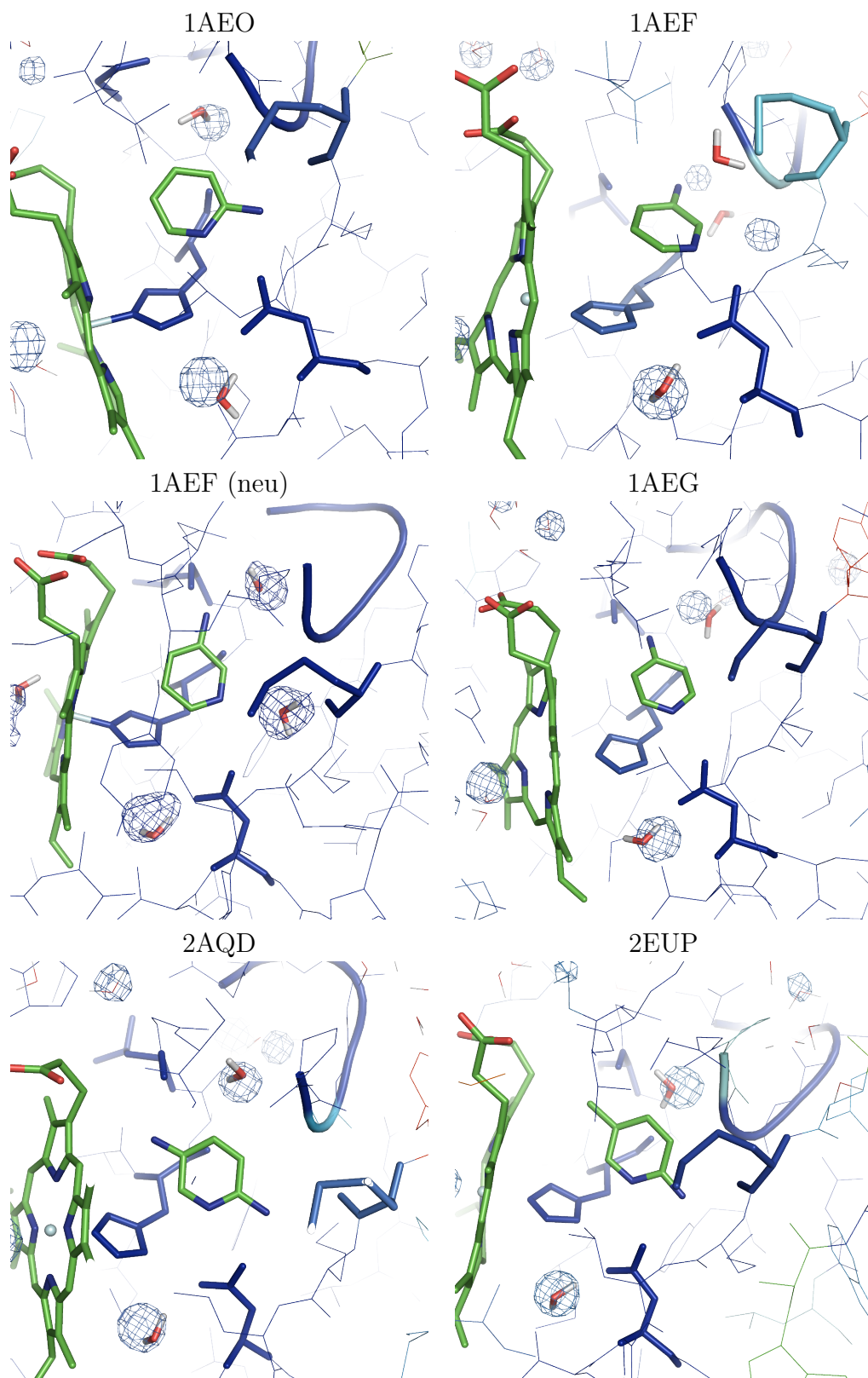
der Kristallstruktur von 1AEO geringfügig erhöhte Dichtewert leicht zu verstehen. Auch in der Simulation der neuen 1AEF Struktur bewegen sich beide Wassermoleküle (A und B) in einem Bereich zwischen zwei Maxima. In diesen Fällen tragen sie dann auch bei beiden Maxima zur Wasserdichte bei. Es wäre also ein komplexerer Algorithmus notwendig um diese Fälle durch das Skript korrekt zu erfassen. Man kann zusammenfassend festhalten, dass die Beobachtungen bezüglich der Wassermoleküle in den Simulationen mit Zusatzkräften gut mit denen der Experimente übereinstimmen.



**Abb. 6.11.:** Übersicht über den Verlauf (Mittelung in 10 ps Intervallen) verschiedener RMSD-Werte der durchgeführten MD-Simulationen der CCP W191G mit Zusatzkräften. In Schwarz sind die Werte der C<sub>α</sub>-Atome, in Rot die der Bindetaschenatome und in Blau die der Ligandatome basierend auf einem Fit zur Referenzstruktur aufgetragen. In Grün dargestellt findet man die Werte der Bindetaschenatome basierend auf dem Fit aller C<sub>α</sub>-Atome der Proteinstruktur, sowie in Violett die Werte der Ligandatome basierend auf dem gleichen Fit.



**Abb. 6.12.:** Anteil verschiedener Sekundärstrukturelemente im Verlauf der mit Zusatzkräften generierten Trajektorien der CCP W191G. Die schwarze Linie zeigt den Anteil der  $\alpha$ -Helices, die rote den der 3-10-Helices, die gelbe den der antiparallelen  $\beta$ -Sheets, die grüne Linie den der  $\pi$ -Helices und die blaue den der parallelen  $\beta$ -Sheets.



**Abb. 6.13.:** Repräsentative Strukturen aus den Trajektorien (mit Zusatzkräften) mit kleinsten RMSD Werten bezogen auf eine gemittelte Struktur, die aus der Trajektorie berechnet wurde. Die ausgewählten Strukturen sind (RMSD Werte in Klammern): 4216 ps für 1AEO (0.54 Å), 3412 ps für 1AEF (0.57 Å), 4026 ps für 1AEF (neu) (0.53 Å), 4760 ps für 1AEG (0.53 Å), 3378 ps für 2AQD (0.48 Å), 2863 ps für 2EUP (0.49 Å). Die Bindetaschen-Aminosäuren sind entsprechend ihrer B-Faktoren (berechnet aus der Simulation) koloriert. Hierbei bedeutet „Blau“ einen niedrigen und „Rot“ einen hohen Wert. Zusätzlich sind die gemittelten Hydratationsdichte-Karten bei einem Kontur-Wert von 70 % gezeigt.

**Tab. 6.6.:** Auflistung der lokalen Maxima sowie deren Summe der gemittelten Hydratationsdichtekarten für die betrachteten MD-Simulationen mit Zusatzkräften auf PGGAN C<sub>α</sub>-Loop-Atome. Die Wasserdichte auf Position C (nahe Asp235 außerhalb der Bindetasche) ist als Vergleich angegeben und mit einem \* gekennzeichnet. Die Summe wurde nur für die Wasserdichte nahe am Liganden, also ohne Position C, bestimmt. Zum Vergleich hierfür ist wiederum in der letzten Zeile die Anzahl der Wassermoleküle nahe am Liganden in der Kristallstruktur aufgeführt.

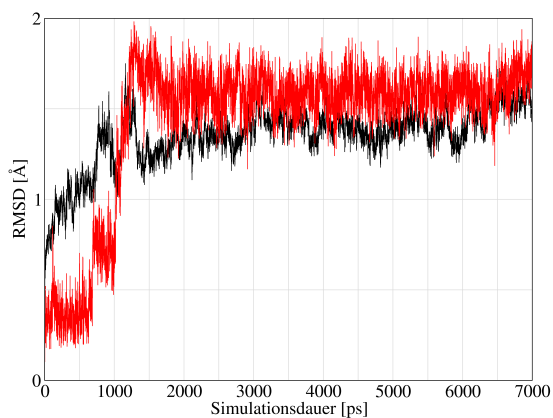
System	1AEO	1AEG	1AEF	1AEF (neu)	2AQD	2EUP
Maxima	0.98*, 0.82, 0.51	0.94*, 0.86	0.97*, 0.95, 0.94	1.0*, 0.99, 0.96, 0.42, 0.32, 0.27, 0.22	0.99*, 0.95	0.98*, 0.95
Summe	1.33	0.86	1.89	3.18	0.95	0.95
Maxima 1. Hälfte	0.99*, 0.89	0.96*, 0.89	0.96*, 0.93, 0.91	1.00*, 1.00, 0.96, 0.41, 0.34, 0.30	0.98*, 0.96	0.99*, 0.92
Summe 1. Hälfte	0.89	0.89	1.84	3.01	0.96	0.92
Maxima 2. Hälfte	0.98*, 0.88, 0.76	0.91*, 0.77	0.99, 0.99*, 0.94	0.99*, 0.99, 0.97, 0.64, 0.54	0.99*, 0.98	0.99*, 0.98
Summe 2. Hälfte	1.64	0.91	1.93	3.14	0.98	0.98
Kristall	1	1	2	2	1	1

## 6.5. Kationenfreie Modellbindetasche der Cytochrom c Peroxidase W191G

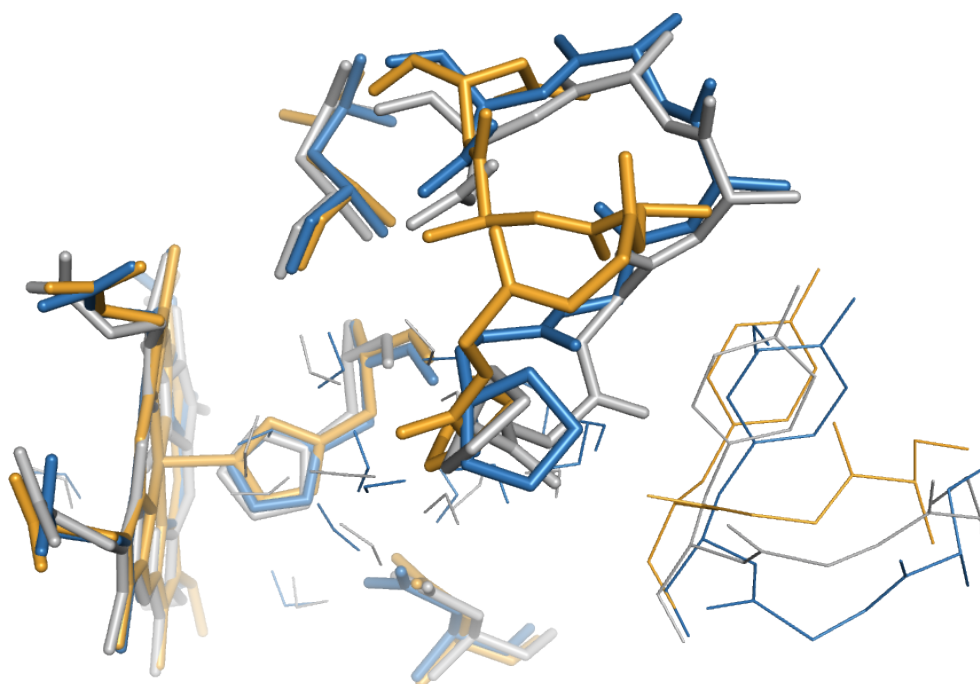
Ausgehend von der kationenfreien Modellbindetasche, die mit Hilfe von molekularen Wechselwirkungsfeldern und MD Simulationen konstruiert wurde (siehe oben), wurde eine 7 ns lange MD-Simulation durchgeführt, um die Dynamik der Wassermoleküle mit der aus den Simulationen der Komplexe vergleichen zu können. Insgesamt zeigt die Proteinstruktur hierbei nur geringe Konformationsänderungen. Die RMSD-Abweichung im Vergleich zum Startmodell bewegt sich dabei im Verlauf der simulierten Zeitspanne in einem akzeptablen Bereich (um 1.5 Å). Die wichtigste strukturelle Änderung tritt im Bereich der PGGAAN-Loop Region in unmittelbarer Nachbarschaft der Bindungsregion auf. Abb. 6.14 zeigt den Verlauf der RMSD-Werte für die Proteinrückgrat-Atome der Schleifen-Region im Vergleich zum Wert für die C<sub>α</sub>-Atome des gesamten Proteins. Die Konformationsänderung im Bereich der Loop-Region tritt zwischen 1000 und 1200 ps auf, also im ersten Fünftel der simulierten Zeitspanne. Im weiteren Verlauf der Trajektorie bleibt diese Tür zur Bindetasche dann im geöffneten Zustand. In Abb. 6.15 findet sich ein Vergleich zwischen einem Schnappschuss aus der Trajektorie vor und nach der Konformationsänderung. Zusätzlich ist hier die Kristallstruktur 1RYC abgebildet. In diesem Komplex der Cytochrom c Peroxidase liegt die flexible Loop-Region in der vollständig geöffneten Konformation vor. Der Vergleich zeigt, dass die Öffnung während der Simulation nur eine Vorstufe der vollständigen Öffnung ist. Durch diesen kleineren Kanal können zwar keine Liganden die Tasche verlassen oder in sie eindringen, der Austausch von Wassermolekülen ist jedoch sehr wohl möglich. Die dadurch häufigeren Bewegungen unter den Wassermolekülen ergeben, dass in dieser Trajektorie nur fünf Wasserpositionen als konserviert angenommen werden können. Davon war eine im Ausgangsmodell nicht durch ein Wassermolekül besetzt. Die gemittelte Hydratationsdichtekarte ist in Abb. 6.16 zusammen mit der MD-Konformation gezeigt, die die geringste strukturelle Abweichung zur Mittelwertsstruktur der Trajektorie zeigte (4672 ps, RMSD 1.00 Å). Es stellt sich die Frage, weshalb in der Simulation diese deutlichen strukturellen Änderungen im Bereich der Bindetasche beobachtet werden konnten. Die mögliche Antwort hierauf könnte lauten: (1) Es wurde eine Simulation einer experimentell bisher so nicht beobachteten Struktur durchgeführt. (2) Es ist bekannt, dass die Öffnung der Loop-Region den Eintritt eines Liganden in die Tasche ermöglicht. Es kann also sein, dass eine geschlossene Loop-Konformation der vollständig ligandfreien Tasche nicht existiert. Möglicherweise öffnet sich der Loop eben, um einem Liganden den Zugang zu ermöglichen. In Rahmen der, bezogen auf die Zeitskala von Proteinbewegungen, relativ kurzen Simulationsdauer wurde dann aber noch nicht die abgeschlossene Loop-Bewegung beobachtet. Nur ein Ligand kann demnach durch Ladungsausgleich die geschlossene Form der Tasche stabilisieren.

Der eigentliche Zweck dieser MD-Simulation bestand jedoch darin, die Dynamik der Wassermoleküle zu untersuchen, um anhand von geeigneten Deskriptoren das Docking zu vereinfachen, also die Anzahl der verdrängbaren Wassermoleküle durch eine Vorabauswahl zu verringern. Durch die gemittelte Hydratationsdichtekarte (vgl. Abb 6.16) zeigte sich,

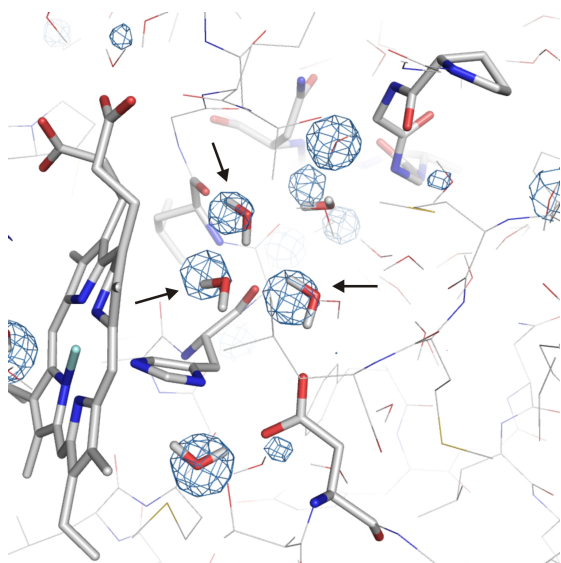




**Abb. 6.14.:** RMSD-Werte aus der MD-Simulation des kationenfreien Bindetaschenmodells. In Schwarz dargestellt sind die Werte der C $\alpha$ -Atome, in Rot die der PGGAAN 190-195 Loop-Region Protein-Rückgrat-Atome (jeweils bezogen auf die Startstruktur der Simulation).



**Abb. 6.15.:** Änderung der PGGAAN Loop-Konformation (dargestellt im Stick-Modus) während der MD-Simulation des kationenfreien Bindetaschenmodells der CCP W191G. Gezeigt sind die Konformation nach 500 ps (grau), nach 1200 ps (blau) sowie zum Vergleich die Kristallstruktur 1RYC (orange) mit geöffneter Loop-Region.



**Abb. 6.16.:** Gemittelte Hydratationsdichteverteilung bei einem Kontur-Wert von 70 % zusammen mit der Proteinkonformation nach 4672 ps, die die geringste RMSD-Abweichung (1.00 Å) im Vergleich zur Mittelwertsstruktur der Trajektorie zeigt.

dass nur wenige Wasserpositionen in der Bindetasche konserviert waren. Auch die Besetzung einzelner Wasserpositionen durch verschiedene Wassermoleküle wurde bezogen auf die Maxima der Hydratationsdichteverteilung bestimmt. Abb. 6.17 zeigt deutlich die hohe Zahl an stattfindenden Austauschvorgängen. Durch die Bewegung der Loop-Region ist der Mittelwert der RMSD-Änderung der Bindetaschenamino-säuren relativ hoch (1.89 Å). Die Sphären für die Analyse der Besetzungsgrade sind dementsprechend auch sehr groß und deshalb zu nahezu 100 % besetzt.

Daraufhin wurde eine Docking-Simulation mit den Programmen GOLD und FlexX wie in Abschnitt 5.2.2 beschrieben, durchgeführt. Als Rezeptorstruktur wurde hierbei der Schnappschuss nach 4672 ps verwendet, der als repräsentativste Proteinkonformation der Trajektorie angesehen werden kann. Alle vier (exklusive Wasser C) im Stab-Modus dargestellten Wassermoleküle sowie das Solvensteilchen innerhalb der in Abb. 6.16 dargestellten Dichte neben dem Loop wurden als verdrängbar definiert. Letzteres war nicht im ursprünglichen Strukturmodell enthalten, sondern diese Position wurde erst während der Simulation hydratisiert. Die Ergebnisse dieser Berechnungen zeigen nur für wenige Ausnahmen den richtigen Bindungsmodus. Durch die vergrößerte Bindetasche ist es für die verwendeten Algorithmen offenbar sehr schwierig, die Platzierung des Liganden exakt zu ermöglichen. Stimmt die oben aufgestellte These, so würde das Eindiffundieren des Liganden in die Tasche aber ohnehin wieder eine Veränderung der Loop-Konformation induzieren.

Realistischer für die Möglichkeiten eines Dockingprogramms sollte also wieder das Docking in die Ausgangsstruktur der Simulation mit geschlossener Schleifenkonformation sein. Für dieses Docking-Experiment wurde nun also die ursprünglich erstellte Modellstruktur verwendet. Dabei wurden aber nur die Wassermoleküle behalten, die sich in der Simulation als konserviert gezeigt hatten. Hierbei waren jetzt nur die drei in Abb. 6.16 mit Pfeilen markierten Wassermoleküle im Docking präsent. Wasser C wurde – wie bisher auch – nicht ins Docking aufgenommen. Das nahe der Schleifenregion neu hinzugekommene Wasser, dessen Position in der Simulation zwar konserviert blieb, war jedoch nicht in der Aus-

gangsstruktur enthalten. Die Ergebnisse des Dockings sind in Tab. 6.7 zusammengefasst. Im Vergleich mit den Ergebnissen in Tab. 5.9 ohne vorselektierte Wassermoleküle hat dieses Vorgehen für die Kombination aus GOLD mit GoldScore bei den Liganden 2AP und 4AP zu einer Verbesserung des RMSD-Wertes um 1.47 bzw. 0.77 Å geführt und auch die Wassermoleküle richtig gesetzt. Die bestbewerteten Lösungen der drei übrigen Liganden liegen hingegen weiter von der Kristallstruktureferenz entfernt. Der RMSD-Wert im Falle von 3AP verschlechtert sich um 0.48, der von DA1 um 3.63 Å. Der RMSD-Wert für LG4 bleibt im sehr schlechten Bereich von  $> 6$  Å. Zusätzlich gilt es zu beachten, dass Wasserposition B in diesem Docking Experiment stets unbesetzt ist, da sich in der MD Simulation dort keine konservierte Hydratationsdichte gezeigt hatte. Bei den Ergebnissen mit GOLD/Chemscore ist zwischen dem Docking mit sieben (vgl. Tab. 5.9) und dem mit drei Wassermolekülen kein großer Unterschied (Mittelwert der Beträge der Änderung der RMSD-Werte der best-bewerteten Lösungen: 0.11 Å) festzustellen. Die Wasservorhersage trifft im Docking-Experiment mit vorselektierten Bindetaschenwassermolekülen nun jedoch nur noch bei 4AP zu. Es hatte sich bereits früher gezeigt, dass die Ergebnisse von GOLD/Chemscore weitgehend unabhängig von evtl. vorhandenen Wassermolekülen sind. Die Ergebnisse bei den Berechnungen durch FlexX sind sehr ähnlich wie im Docking-Experiment ohne vorselektierte Wassermoleküle. Die außer für DA1 ohnehin bereits sehr guten Resultate wurden durch die Begrenzung der Wassermenge nicht weiter verbessert.

Zusammenfassend bleibt festzuhalten, dass nach Auswertung dieser MD-Simulation grundsätzliche Zweifel an der chemischen Relevanz des zu Grunde liegenden Rezeptormodells bestehen. Es wurde überprüft, ob eine Vorauswahl der Wassermoleküle die Docking-Ergebnisse verbessern könnte. Dies war jedoch nicht durchweg der Fall. Für GOLD/Goldscore wurde nur in zwei von fünf Fällen eine Verbesserung beobachtet, in den drei anderen wurden die Ergebnisse jedoch schlechter. Dabei wurden die erwarteten Wassermoleküle nicht aus der Tasche verdrängt, und der kleine Ligand an anderer Position platziert. Mit GOLD/Chemscore und FlexX wurden sehr ähnliche Resultate wie im Docking-Experiment mit sieben verdrängbaren Bindetaschenwassermolekülen erzielt. Dieses etwas enttäuschende Ergebnis stellt jedoch nicht grundsätzlich die Idee der Vorabauswahl von Wassermolekülen vor der Durchführung einer Docking-Simulation in Frage. Fraglich ist vielmehr, ob die hier untersuchte Proteinstruktur für einen solchen Test geeignet ist. Außerdem kann auch die Größe anhand derer diese Entscheidung getroffen wurde, nicht geeignet sein. Ein thermodynamisch instabiles Wassermolekül kann z.B. kinetisch gehemmt sein und deshalb seine Position nicht verlassen und einen hohen Besetzungsgrad zeigen. Ein Besetzungsgrad muss demnach nicht notwendigerweise die thermodynamischen Eigenschaften einer Wasserposition vollständig widerspiegeln, welche aber wie oben diskutiert für die Affinität des Liganden entscheidend sein können. Deshalb ist es unabdinglich, die thermodynamischer Daten direkt zu berechnen.



**Tab. 6.7.:** Ergebnisse (best-bewertete Lösungen sowie solche mit kleinstem RMSD-Wert) des Dockings mit GOLD 4.0 und FlexX 3.1.4 in die ionenfreie Modellbindetasche von CCP W191G. Die Wechselwirkungen aller drei vorselektierten Wassermoleküle können durch das Program GOLD an- oder ausgeschaltet werden. FlexX ist in der Lage alle drei Wassermoleküle zu verdrängen („v“). Die verwendete Abkürzung „CR“ bedeutet Cluster Rang, „GS“ steht für GoldScore, „CS“ für Chemscore und „FS“ für FlexX Score. Alle Lösungen wurden bezogen auf ihre RMSD Abweichungen untereinander einem Clustering unterzogen. Die Cluster wiederum sind nach der enthaltenen best-bewerteten (Score) Lösung geordnet. Die hier gezeigten RMSD Werte wurden zwischen der Docking-Lösung und der Kristallstruktur als Referenz bestimmt. In der jeweiligen Rezeptorstruktur findet sich das Pendant zu Wassermolekül 2238 (Position A). Wassermoleküle, die vorhanden sein sollten, sind in der Tabelle fett gedruckt.

Ligand	Aus Komplex	Software	Rang	CR	GS/CS/FS	RMSD (Å)	Wassermoleküle: 308/2238(A)/6667
2AP	1AEO	GOLD/GS	1	1	40.96	1.27	aus/ <b>an</b> /aus
			46	1	39.32	0.79	aus/ <b>an</b> /aus
		GOLD/CS	1	1	20.83	2.60	aus/ <b>aus</b> /aus
			36	1	20.33	0.45	aus/ <b>aus</b> /aus
FlexX	1	1	-16.70	0.69	v/ <b>ja</b> /v		
3AP	1AEF	GOLD/GS	1	1	38.87	2.36	aus/ <b>an</b> /aus
			43	1	37.14	2.18	aus/ <b>an</b> /aus
		GOLD/CS	1	1	23.70	1.51	aus/ <b>aus</b> /aus
			34	1	22.07	1.19	aus/ <b>aus</b> /aus
FlexX	1	1	-16.65	1.25	v/ <b>ja</b> /v		
			3	3	-14.64	0.92	v/ <b>ja</b> /v
4AP	1AEG	GOLD/GS	1	1	39.74	0.65	aus/ <b>an</b> /aus
			41	1	38.75	0.42	aus/ <b>an</b> /aus
		GOLD/CS	1	1	22.31	1.28	aus/ <b>an</b> /aus
			50	1	20.90	0.77	aus/ <b>aus</b> /aus
FlexX	1	1	-23.08	1.18	v/ <b>ja</b> /v		
			7	4	-10.69	0.76	v/ <b>ja</b> /v

*Fortsetzung auf der nächsten Seite*

## 6. Auswertung der Dynamik

---

Ligand	Aus Kom- plex	Software	Rang	CR	GS/CS/FS	RMSD (Å)	Wassermoleküle: 308/ <b>2238(A)</b> /6667
DA1	2AQD	GOLD/GS	1	1	42.98	6.13	an/ <b>aus</b> /an
			26	5	39.29	1.42	aus/ <b>an</b> /aus
		GOLD/CS	1	1	20.11	4.73	aus/ <b>aus</b> /aus
			47	2	18.93	0.97	aus/ <b>aus</b> /aus
		FlexX	1	1	-17.37	3.01	v/ <b>ja</b> /v
			4	4	-15.37	0.87	v/ <b>ja</b> /v
LG4	2EUP	GOLD/GS	1	1	42.63	6.57	an/ <b>aus</b> /an
			43	6	40.44	1.16	aus/ <b>an</b> /aus
		GOLD/CS	1	1	20.17	2.46	aus/ <b>aus</b> /aus
			34	1	21.02	0.64	aus/ <b>aus</b> /aus
		FlexX	1	1	-19.33	1.14	v/ <b>ja</b> /v
			29	24	-10.89	0.60	v/ <b>ja</b> /v

---

# 7. Bestimmung thermodynamischer Größen einzelner Wassermoleküle in Protein-Ligand Komplexstrukturen

## 7.1. Berechnungen mittels thermodynamischer Integration

Neben wenigen experimentellen Möglichkeiten, exakte thermodynamische Beiträge zur Bindung eines Liganden an einen biomakromolekularen Rezeptor zu erhalten, existieren verschiedene computergestützte Verfahren, die eben das ermöglichen sollen (siehe Abschnitt 1.2.4). Hiervon sind jedoch (fast) nur die Freie Energie Störungsrechnung (FEP) und die Thermodynamische Integration (TI) (vgl. Abschnitt 3.3) – basierend auf MD- oder MC-Simulationen – in der Lage, exakte Ergebnisse zu liefern. Noch genauere, quantenchemische Verfahren sind in der Praxis aufgrund ihres enormen Rechenaufwandes für Proteine nicht weit verbreitet. Zudem bieten computerbasierte Methoden wie TI oder FEP prinzipiell die Möglichkeit, thermodynamische Größen einzelner Wassermoleküle zu erfassen, was in dieser Arbeit von großer Relevanz ist.

In der Vergangenheit wurden TI und FEP an verschiedenen Systemen zur Untersuchung des Einflusses einzelner Wassermoleküle eingesetzt. Die im Folgenden dargestellte Auswahl an Veröffentlichungen erhebt keinerlei Anspruch auf Vollständigkeit. Bereits 1991 untersuchten Wade und Kollegen<sup>227</sup> zwei Wassermoleküle in einem sulfatbindenden Protein aus *Salmonella typhimurium* mittels TI-Berechnungen auf Basis von MD-Simulationen mit dem Programm GROMOS. Später berichteten Helms und Wade<sup>228</sup> über die Identifizierung des wahrscheinlichsten Hydratationsgrades der vergrabenen Bindetasche von Cytochrom P450cam. Die MD-Simulationen für die TI-Berechnungen wurden in diesem Fall mit dem Programm ARGOS durchgeführt. 1996 wurden von Roux *et al.*<sup>229</sup> die energetischen Eigenschaften von Wassermolekülen in unterschiedlichen Positionen innerhalb der Protonenpumpe von Bacteriorhodopsin untersucht. In einer neueren Arbeit berichteten Hammelberg und McCammon<sup>230</sup> über TI-Rechnungen mit der „Double Decoupling“-Methode für MD-Simulationen mit dem Simulationsprogramm AMBER, die energetisch günstige Wassermoleküle in Komplexen von Trypsin und HIV-1 Protease bestimmen. In der Arbeit von Lu, Yang und Wang<sup>231</sup> wurde ebenfalls das Programm AMBER mit dem dort implementierten TI-Modul für die Berechnungen an den Wassermolekülen in zwei Komplexen der HIV-1 Protease eingesetzt. Olano und Rick veröffentlichten 2004 eine Studie<sup>232</sup>, in der mit FEP-Rechnungen auf Basis von MD-Simulationen die thermodynamischen Beiträge einzel-

ner Wassermoleküle einer polaren Bindetasche in BPTI mit denen in einer hydrophoben Bindetasche der I76A Mutante der Barnase berechnet und dann gegenübergestellt wurden. Später beschrieb Rick zusammen mit Yu<sup>233</sup> TI-Rechnungen mit AMBER für Wassermoleküle in Komplexen der DNA Gyrase. Essex *et al.*<sup>234</sup> berichteten 2007 von thermodynamischen Analysen von Wassermolekülen in sechs Protein-Ligand Komplexen mittels „Replica Exchange“-TI-Rechnungen auf Basis von MC-Simulationen. Auch wenn Freie Energierechnungen für Wassermoleküle also schon mehrfach in der Literatur beschrieben sind, zeigen diese Quellen kein einheitliches Protokoll und insbesondere für das hier verwendete Programm AMBER auch nur wenige Beispiele. Zudem liefert der Literaturvergleich auf die später in der vorliegenden Arbeit auftretenden Probleme mit der Immobilisierung der Wassermoleküle keine nützliche Antwort. Bei TI und FEP-Rechnungen sollten also für jedes Protein-System die Einstellungen neu validiert werden. Die beschriebenen Beispiele zeigen aber, dass die Rechnungen für die jeweiligen Fragestellungen sehr nützlich waren und zu einem Erkenntnisgewinn geführt haben.

### 7.1.1. Verwendete Parameter und Einstellungen

Für die im Folgenden dargestellten TI-Rechnungen aus dieser Arbeit wurde das Modul SANDER aus dem AMBER Programmpaket in der Version 10<sup>207</sup> verwendet. Als strukturelle Basis für die Simulation der Komplexe wurden die jeweiligen Kristallstrukturen verwendet. Die vorbereitenden Schritte wurden analog für die unterschiedlichen Systeme durchgeführt. Individuelle Änderungen für einzelne Komplexe oder Liganden sind an der jeweiligen Stelle vermerkt. Die Aufbereitung der Strukturdaten, Protonierung und Parametrisierung erfolgte analog wie in Abschnitt 6.1 beschrieben. Die Definition der zu simulierenden Störung erfolgte auf zwei unterschiedliche Weisen. Beide Varianten unterscheiden sich bei der Programm-internen Handhabung und haben dementsprechend Vor- und Nachteile. Gemeinsam ist beiden Methoden, dass sie die Berechnung einer freien Energieänderung in drei Einzelschritte aufspalten. Es ist notwendig<sup>160</sup>, zunächst die Ladung der Atome, für die die van-der-Waals-Parameter geändert werden, zu entfernen. Anschließend erfolgt die Simulation, in der die Parameter der Atomtypen verändert werden, um das Zielmolekül zu erhalten. Schließlich werden die neuen Ladungen für die Atome, von denen diese zuvor entfernt wurden bzw. an denen sich die Ladung änderte, wieder hinzugefügt.

1. In der einen Variante wird die Datei mit dem minimierten aber noch nicht solvatisierten Komplex in LEaP eingelesen. Hier werden sowohl die Gegenionen wie auch die Wassermoleküle hinzugefügt, die benötigt werden, um eine Box zu erzeugen, deren Seitenwände mindesten 8 Å von der Proteinoberfläche entfernt sind. Anschließend werden – wie oben erläutert – die Ladungen einzelner oder aller Ligandatome entfernt und Topologie- sowie Koordinaten-Datei zum Start der Simulation abgespeichert. Danach werden neue van-der-Waals-Parameter für unterschiedliche Ligandatome festgesetzt und erneut die Informationen in Dateien geschrieben. Im letzten Vorbereitungsschritt wird dann der Komplex mit dem wieder geladenen Liganden in der Topologie- und Koordinaten-Datei gespeichert.



2. Als Alternative können mit LEaP nur die Ausgangs- und Endstruktur der alchemischen Transformation in zwei Sätzen aus Topologie- und Koordinaten-Datei geschrieben werden. Die beiden Zwischenzustände werden später über das Modul SANDER definiert. Diese Dateien dürfen sich nur bei den zu transformierenden Atomen unterscheiden. Deshalb wird hierfür die bereits bei 300 K equilibrierte, solvatisierte Komplexstruktur verwendet.\* Über den Umweg einer PDB-Datei wird hier die Störung definiert. Die PDB-Datei wird dann mit LEaP eingelesen und neu die umgebende Box-Information definiert. Anschließend werden die für die Simulation benötigten Dateien gespeichert.

Der Vorteil der zweiten Variante kommt vor allem dann zum Tragen, wenn Atome, Atomgruppen oder ganze Moleküle verschwinden sollen. Dies ist hier ohne zusätzliches Einführen so genannter „Dummy“-Atome möglich. Von Nachteil ist jedoch, dass zwei Sätze aus Dateien erstellt werden müssen, die sich nur bei den zu transformierenden Atomen unterscheiden dürfen. Dies verlangt insbesondere bei anderen strukturellen Veränderungen als dem Verschwinden von Atomen intensive Vorüberlegungen. Die Simulationsschritte in denen Atomladungen entfernt bzw. hinzugefügt werden, werden hierbei nicht mittels LEaP sondern direkt in SANDER definiert, was ebenfalls einen höheren Aufwand bei der Erstellung der hierfür benötigten Dateien bedeutet.

Die anschließende MD-Simulation wurde bei 300 K und einem Zeitschritt von 2 fs im *NPT*-Ensemble mit dem Programm SANDER durchgeführt. Der SHAKE-Algorithmus wurde nur auf solche C-H-Bindungen angewandt, bei denen keine Änderungen der Atomparameter erfolgten. Die „particle mesh Ewald“ Methode<sup>127</sup> wurde verwendet, um langreichweitige, elektrostatische Wechselwirkungen zu handhaben. Für andere nicht-kovalente Wechselwirkungen wurde ein Distanz-Ausschlusskriterium von 8.0 Å eingestellt. Die Koordinaten der Trajektorien wurden alle 1 ps und Energiewerte alle zehn Zeitschritte abgespeichert. Jeder der oben genannten Schritte wurde in neun Prozesse mit den  $\lambda$ -Werten 0.01592, 0.08198, 0.19331, 0.33787, 0.50000, 0.66213, 0.80669, 0.91802, 0.98408 unterteilt. Die Wahl fiel auf diese Werte, da sie eine spätere Integration über Gauß'sche Quadratur mit vorberechneten Koeffizienten erlauben. Jede dieser Simulationen wurde über 500 ps durchgeführt. Als Ausgangskonfiguration jedes Schrittes wurde die Endkonfiguration des Systems für den vorherigen  $\lambda$ -Wert verwendet. Insgesamt ergibt sich somit eine Simulationsdauer von 13.5 ns. Dies benötigte auf einem AMD Opteron 2.2 GHz Cluster mit Infiniband-Vernetzung ca. 290 h umgerechnet auf einen Prozessorkern. Der zweite Transformationsschritt, die Änderung der van-der-Waals-Parameter einzelner Atome, wurde unter Zuhilfenahme der Soft-Core Potentiale<sup>161</sup> (vgl. Abschnitt 3.3.7) durchgeführt. Die Werte für  $\langle \frac{\partial V}{\partial \lambda} \rangle$  dieser insgesamt sehr aufwendigen Berechnung wurden mittels eines PYTHON-Skriptes gemäß Gl. 3.29 integriert. Hierbei wurden die ersten 100 ps eines jeden 500 ps-Abschnittes als Equilibrierungsphase betrachtet und für die Energieberechnung außer Acht gelassen.

---

\*In Variante 1 schließt sich ja für jede der erzeugten Startkoordinaten noch eine Aufheiz- und Equilibrierungsphase an. Dies hätte in Variante 2 jedoch zur Folge, dass sich die Koordinaten von Start- und Zielkonfiguration unterscheiden würden.

Die Fehlerberechnung erfolgte gemäß dem Gauß'schen Fehlerfortpflanzungsgesetz über

$$\Delta = \sqrt{\Delta_{\lambda_1}^2 + \Delta_{\lambda_2}^2 + \dots + \Delta_{\lambda_9}^2}. \quad (7.1)$$

Hierbei stellen  $\Delta_\lambda$  die Unsicherheiten der einzelnen  $\lambda$ -Werte dar. Diese wurden aus der während der Simulation auftretenden statistischen Schwankung von  $\frac{\partial V}{\partial \lambda}$  um den Mittelwert erhalten.

In einzelnen Simulationen wurden zudem zusätzlich Zusatzkräfte auf Atomgruppen angewandt, vornehmlich um den Austausch bzw. die Bewegung von Wassermolekülen zu verhindern. Beim Vernichten und Erzeugen eines Wassermoleküls ist es z.B. notwendig, dass (1) dieses Wassermolekül trotz schwacher Wechselwirkungen mit der Umgebung die Position nicht verlässt und (2) keine anderen Wassermoleküle diese Position besetzen (vgl. hierzu Essex *et al.*<sup>234</sup> sowie Olano und Rick<sup>232</sup>). Zusätzlich zu den TI-Berechnungen der Komplexe wurde auch eine Reihe Transformationen der Liganden in Wasser durchgeführt. Die Parametrisierung und Durchführung der Simulation erfolgte hierbei völlig analog zu den Komplexsystemen. Durch die geringere Systemgröße wurde hierbei auch eine kürzere Rechenzeit benötigt.

Gemäß Gilson *et al.*<sup>159</sup> sind die über Simulationen bestimmten  $\Delta G$ -Werte bezüglich der Standardbedingungen mit verschiedenen Termen zu korrigieren (vgl. Gl. 7.2). McCammon und Gilson separieren den Komplex der Moleküle A und B in interne und externe kartesische Koordinaten. Damit wird dann die Position von Molekül A relativ zu der von Molekül B angegeben.  $\mathbf{r}_A$  sind die internen Koordinaten von Molekül A sowie  $\zeta_B$  die externen Koordinaten von B. Mit  $\mathbf{r}_S$  werden die kartesischen Koordinaten der Solvens-Moleküle, mit  $\sigma$  die Symmetriezahl von Molekül bzw. Komplex, mit  $C^\circ$  die Standard-Konzentration, mit  $\xi$  die Orientierungskordinaten und mit  $\bar{V}$  das Partialvolumen des Moleküls bzw. des Komplexes bezeichnet.  $I(\zeta)$  steht für die Schrittfunktion, die im Falle des komplexierten Zustandes 1 ist und im unkomplexierten Zustand den Wert 0 annimmt. Auf diese Weise werden nur die Konfigurationen betrachtet, die relevant sind, d.h. wo die Moleküle A und B nahe beieinander zu finden sind. Für die so korrigierte Freie Enthalpie gilt:

$$\begin{aligned} \Delta G_1^\circ = \int \left\langle \frac{\partial U(\lambda, \mathbf{r}_A, \mathbf{r}_B, \zeta_B, \mathbf{r}_S)}{\partial \lambda} \right\rangle_{\lambda, I(\zeta)=1} d\lambda - RT \ln \left( \frac{\sigma_{AB}}{\sigma_A \sigma_B} \right) \\ + RT \ln(C^\circ V_1) + RT \ln(\xi_1/8\pi^2) + P^\circ(\bar{V}_A - \bar{V}_{AB}) \end{aligned} \quad (7.2)$$

Der dritte Term der Gleichung beschreibt die Änderung der freien Enthalpie, wenn es dem zuvor festgehaltenen Liganden (in dieser Arbeit das Wassermolekül) erlaubt ist, sich im Volumen  $1/C^\circ$  zu bewegen. Der vierte Term beschreibt die Energieänderung, wenn der während der Simulation festgehaltene Ligand frei rotieren darf. Die mathematische Form beider Beiträge hängt von der Art der gewählten Zusatzkräfte ab (in Gl. 7.2 für ein „hard-wall“Potential). Ein harmonisches Potential entspräche einem dritten Term der Form  $-RT \ln[C^\circ(2\pi RT/k)^{3/2}]$ . Der letzte Term in Gl. 7.2 hingegen ist bei Simulationen mit normalem Druck vernachlässigbar. Selbstverständlich sind diese Korrekturen nicht bei

der Berechnung relativer freier Enthalpie-Unterschiede notwendig. Wo immer Korrekturen benötigt werden, werden diese im Folgenden aufgeführt.

### 7.1.2. Entfernung eines Wassermoleküls aus Wasser

Die berechnete Freie Enthalpieänderung für das Entfernen eines Wassermoleküls aus einem Protein-Ligand Komplex ist alleine noch nicht aussagekräftig. Es ist notwendig, diese mit dem entsprechenden Wert vom Entfernen eines TIP3P-Wassermoleküls aus der Wasserphase in Beziehung zu setzen. Erst in Summe ergeben diese beiden Werte die Information, ob der Transfer eines Wassermoleküls aus der Wasserphase in die Bindetasche energetisch vorteilhaft ist oder nicht.

Bei der dafür durchgeführten Simulation ergab sich ein Wert von  $\Delta G = -6.11 \pm 0.82$  kcal/mol für die Solvatisierung eines Wasser-Moleküls in der TIP3P-Wasserphase. Dieser Wert stimmt sehr gut mit Werten aus der Literatur überein. Experimentelle Ergebnisse liegen im Bereich von  $-5.7$  kcal/mol<sup>235</sup> und  $-6.3$  kcal/mol<sup>236</sup>. Für das TIP3P-Wassermodell wurde von Hermans *et al.*<sup>237</sup> ein Wert von  $-5.4$  kcal/mol berechnet. Die Freie Enthalpie für TIP4P ist ähnlich:  $-6.1 \pm 0.3$  kcal/mol<sup>238</sup>. Weiterhin bestimmten Wade *et al.*<sup>228</sup> einen Wert von  $-30.1 \pm 1.2$  kJ/mol. Das entspricht  $-7.19 \pm 0.29$  kcal/mol. Frühere Ergebnisse waren  $-6.4 \pm 0.4$  kcal/mol bei Quiocho und Kollegen<sup>227</sup> sowie vor kurzem bei Yu und Rick<sup>233</sup>  $-7.04 \pm 0.04$  kcal/mol.

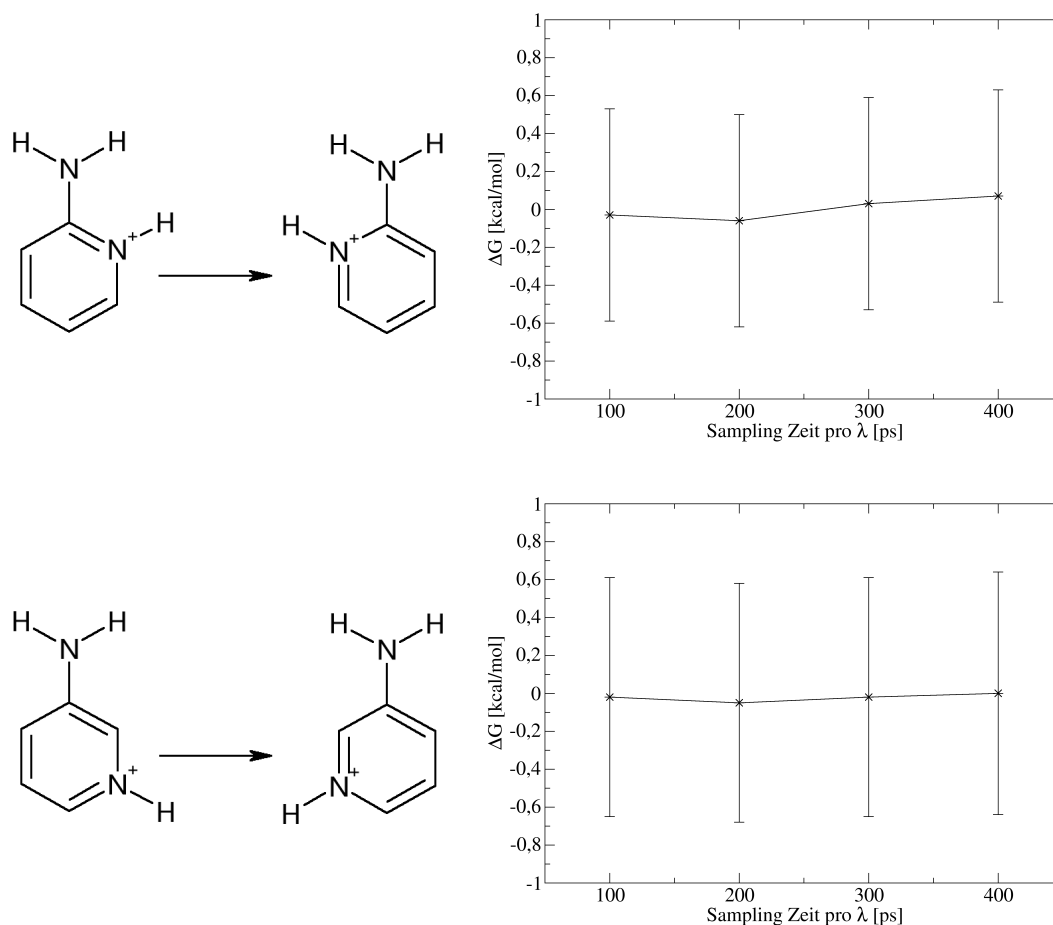
### 7.1.3. Testsimulationen

Um zu überprüfen, ob die gewählte Anzahl an  $\lambda$ -Werten sowie die veranschlagte Simulationsdauer ausreichend ist, um zu aussagekräftigen Ergebnissen zu gelangen, wurden eine Reihe von Testsimulationen durchgeführt, bei denen die Summe aller Freien Enthalpieänderungen Null ergeben sollte. Die sehr guten Ergebnisse hierfür sollen im Folgenden dargestellt werden.

#### Verschiebung des aromatischen N-Atoms in 2- und 3-Aminopyridin

Eine relativ einfach durchzuführende Testtransformation ist die Verschiebung der Position des aromatischen N-Atoms in den Liganden 2- und 3-Aminopyridin. Hierfür wird das ursprüngliche N-Atom in ein C-Atom transformiert, während ein anderes C-Atom zu einem N-Atom wird. Durch die Protonierung der Liganden sind zusätzlich die Kraftfeldparameter und Ladung des angrenzenden H-Atoms abzuändern. Auf diese Weise wurden die beiden unsymmetrischen Moleküle 2- und 3-Aminopyridin in Wasser in sich selbst transformiert. Da Ausgangs- und Endzustand der Transformation identisch sind, sollte für diesen Prozess gelten:  $\Delta G = 0$  kcal/mol.

Die Ergebnisse der Simulation sind in Abb. 7.1 dargestellt. Die freien Energiewerte wurden für unterschiedlich lange Simulationszeiten pro  $\lambda$ -Wert bestimmt. Man sieht, dass die Ergebnisse konvergiert sind und auch die Fehlerwerte liegen in einem sehr niedrigen Bereich. Außerdem kann nicht beobachtet werden, dass die Unsicherheiten bei längerer

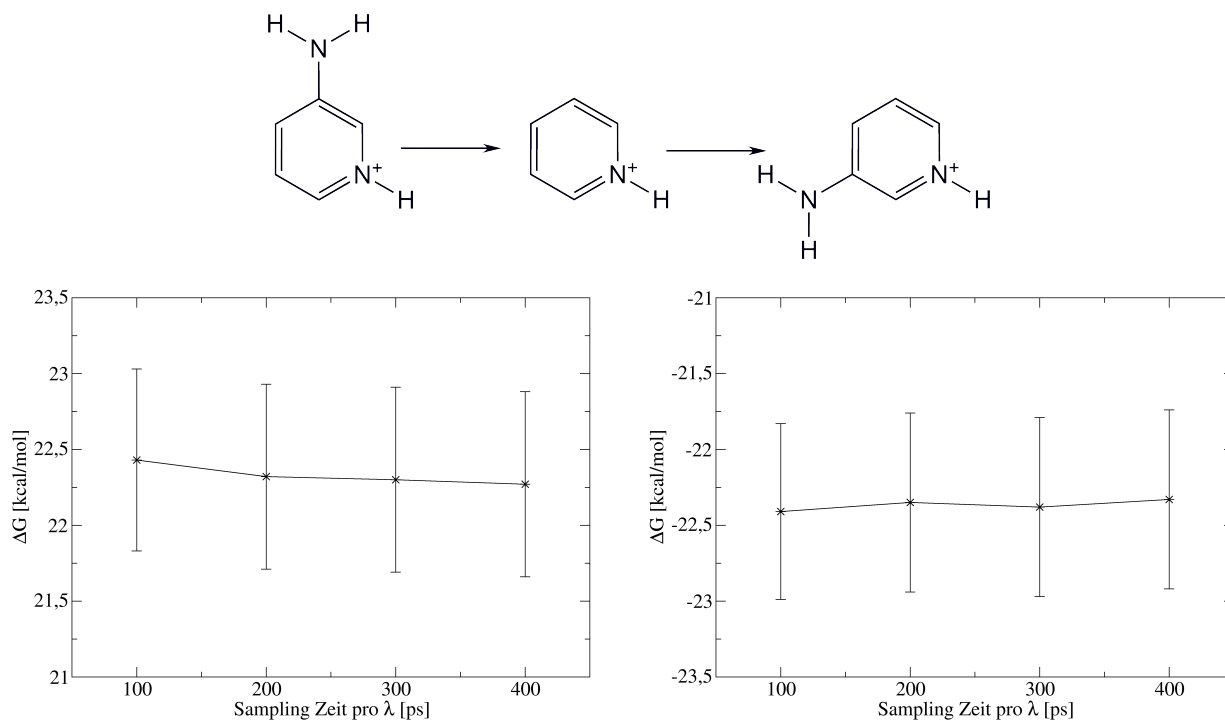


**Abb. 7.1.:** Ergebnisse der TI-Berechnungen für die Verschiebung des aromatischen N-Atoms in 2- und 3-Aminopyridin.

Simulationsdauer merklich kleiner werden. Für die Verschiebung des Pyridin-N-Atoms in 2-Aminopyridin bei der längsten durchgeführten Simulationsdauer erhält man  $\Delta G = 0.07 \pm 0.56$  kcal/mol. Der entsprechende Wert bei 3-Aminopyridin ist  $\Delta G = 0.00 \pm 0.64$  kcal/mol.

### Entfernen und Erzeugen der Amino-Gruppe an 3-Aminopyridin

Ebenso ist es möglich die Amino-Gruppe z.B. aus 3-Aminopyridin zu entfernen bzw. zu Pyridin hinzuzufügen. Die Summe von  $\Delta G$  für beide Prozesse sollte auch hier Null ergeben. In Abb. 7.2 findet sich die Reaktionsgleichung der alchemischen Transformationen sowie die Auftragung der Werte für  $\Delta G$  in Abhängigkeit von der Simulationszeit pro  $\lambda$ -Wert. Im Vergleich zum Verschieben der Position des N-Atoms im Ring benötigt das Entfernen und Erzeugen mehrerer Atome eine geringfügig längere Simulationsdauer. Dennoch sind die erhaltenen Unsicherheiten stets gering im Vergleich zum Absolutwert für  $\Delta G$ . Für die

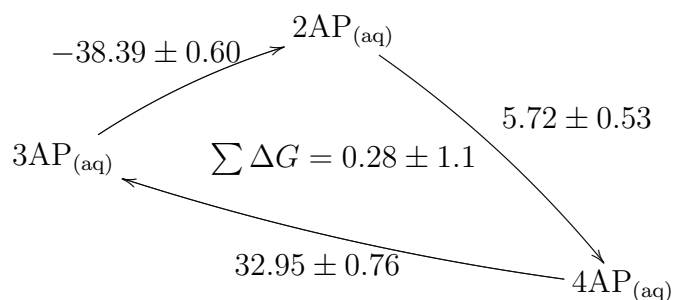


**Abb. 7.2.:** Ergebnisse der TI-Berechnungen für das Entfernen (links) und Erzeugen (rechts) einer Aminogruppe an 3-Aminopyridin bzw. Pyridin.

längste durchgeführte Simulationszeit erhält man für den Prozess der Entfernung einer  $\text{NH}_2$ -Gruppe  $\Delta G = 22.27 \pm 0.61$  kcal/mol. Das Erzeugen der gleichen Gruppe am einfach protonierten Pyridin ergibt  $\Delta G = -22.33 \pm 0.59$  kcal/mol. In Summe ergibt sich somit eine freie Energieänderung von  $-0.06 \pm 0.85$  kcal/mol.

### Transformation von 2-, 3- und 4-Aminopyridin untereinander

Durch die hohe strukturelle Ähnlichkeit der Liganden ist es möglich, einfache, formale Derivatisierungen zwischen den Aminopyridinen durchzuführen. Die Ergebnisse für diesen Prozess der Liganden in Wasser sind in Abb. 7.3 dargestellt. Die Summe aller freien Enthalpie-Änderungen beträgt  $\sum \Delta G = 0.28 \pm 1.10$  kcal/mol. Der zu erwartende Wert von 0 kcal/mol liegt somit innerhalb der Fehlergrenzen. In Tab. 7.1, 7.2 und 7.3 sind die Beiträge für die einzelnen  $\lambda$ -Werte und für Coulomb- sowie van-der-Waals-Term aufgelistet. Besonders fällt hier der sehr geringe Beitrag des zweiten Transformationsschrittes, Änderung der van-der-Waals Parameter, in Bezug auf die Gesamtenergieänderung auf. Die Änderung der Position des Pyridin N-Atoms im Ring resultiert nur in einer Änderung der H-Brückenbindungsgeometrie. Es bleiben ja die gleichen Atomtypen im Molekül. Dies hat in der isotropen Solvensumgebung keinen Einfluss auf den Lennard-Jones Anteil



**Abb. 7.3.:** Ergebnisse der TI-Berechnung in kcal/mol für den Kreisprozess der Transformation von 2- in 4-, 4- in 3- und 3- in 2-Aminopyridin in wässriger Lösung.

der freien Energie. Neben der Geometrie der Wechselwirkungen ändert sich aber auch das Dipolmoment durch die Partiaalladungen, was sich in den Werten der Coulomb-Beiträge widerspiegelt. Vergleicht man außerdem die Ergebnisse für die verschiedenen  $\lambda$ -Werte, so zeigen diese untereinander einen Fehler in ähnlicher Größenordnung. Es kann also davon ausgegangen werden, dass die Simulation für einzelne  $\lambda$ -Schritte nicht deutlich länger sein muss.

Insgesamt hat sich aus den bisher dargestellten Testsimulationen gezeigt, dass die gewählten Bedingungen für die Transformation kleiner Moleküle in Wasser vollkommen geeignet sind.

Die Transformation der drei Aminopyridine untereinander kann selbstverständlich nicht nur in Lösung sondern ebenso komplexiert mit dem Protein CCP W191G erfolgen. Die Simulation des Übergangs von 2- zu 4-Aminopyridin ergibt eine Energieänderung von  $\Delta G = 9.24 \pm 0.71$  kcal/mol. Die Transformation von 4-Aminopyridin zu 3-Aminopyridin ergibt  $\Delta G = 50.79 \pm 1.38$  kcal/mol. Schließlich erhält man für den Prozess der Umwandlung von 3-Aminopyridin in 2-Aminopyridin komplexiert mit CCP W191G einen Wert von  $\Delta G = -33.84 \pm 1.14$  kcal/mol. In Summe ergibt das für diesen Kreisprozess  $\sum \Delta G = 26.19 \pm 1.93$  kcal/mol. Dieser Wert ist sehr weit vom zu erwartenden Wert von 0 kcal/mol entfernt. Diese Diskrepanz ist nicht in größeren Unsicherheiten aufgrund einer für konvergierte Ergebnisse zu kurz gewählten Simulationsdauer begründet. Der Wert Null liegt weit außerhalb des angegebenen Fehlerbereiches. Auch eine nötige Korrektur bezüglich der Standardbedingungen (vgl. Gilson *et al.*<sup>159</sup>) kann nicht die Ursache sein, da diese nur bei der Berechnung absoluter Freier Enthalpieänderungen relevant ist. Alle drei Transformationen wurden unter den gleichen (Nicht-Standard-)Bedingungen durchgeführt. In diesem Fall kommt vielmehr der Einfluss des zweiten Wassermoleküls in der Bindetasche von CCP W191G im Komplex mit 3-Aminopyridin zum Tragen: Diesem konservierten Wassermolekül wurde in den vorangegangenen Simulationen keine Beachtung geschenkt. Es kann nicht davon ausgegangen werden, dass sich die Solvens-Umgebung in der Bindetasche so rasch dem veränderten Liganden anpassen kann, wie das in der Lösung der Fall ist. Aus diesem Grund müssen die Energie-Beiträge für das Erzeugen bzw. Vernichten von Wasser B bei diesem Kreisprozess noch hinzugefügt bzw. berücksichtigt werden.

**Tab. 7.1.:** Beiträge zu  $\Delta G$  für die Transformation von 2-Aminopyridin zu 4-Aminopyridin in Wasser über die Verschiebung des Pyridin N. Hierbei bezeichnet  $w$  den Gewichtungsfaktor für die Gauss-Integration.

$\lambda$	$w \cdot \langle \partial V / \partial \lambda \rangle_{\lambda} / \text{kcal} \cdot \text{mol}^{-1}$		
	Summe	Lennard-Jones	Coulomb
0.01592	$0.93 \pm 0.06$	$-0.10 \pm 0.03$	$1.03 \pm 0.06$
0.08198	$1.82 \pm 0.13$	$-0.23 \pm 0.06$	$2.05 \pm 0.12$
0.19331	$2.02 \pm 0.19$	$-0.31 \pm 0.09$	$2.33 \pm 0.17$
0.33787	$1.70 \pm 0.22$	$-0.25 \pm 0.08$	$1.95 \pm 0.20$
0.50000	$0.98 \pm 0.22$	$0.01 \pm 0.08$	$0.97 \pm 0.21$
0.66213	$0.12 \pm 0.23$	$0.26 \pm 0.10$	$-0.14 \pm 0.21$
0.80669	$-0.58 \pm 0.21$	$0.30 \pm 0.09$	$-0.88 \pm 0.19$
0.91802	$-0.81 \pm 0.14$	$0.23 \pm 0.06$	$-1.04 \pm 0.12$
0.98408	$-0.47 \pm 0.06$	$0.11 \pm 0.03$	$-0.58 \pm 0.05$
<b>Summe</b>	$5.72 \pm 0.53$	$0.01 \pm 0.22$	$5.71 \pm 0.48$

**Tab. 7.2.:** Beiträge zu  $\Delta G$  für die Transformation von 4-Aminopyridin zu 3-Aminopyridin in Wasser über die Verschiebung des aromatische N-Atoms. Hierbei bezeichnet  $w$  den Gewichtungsfaktor für die Gauss-Integration.

$\lambda$	$w \cdot \langle \partial V / \partial \lambda \rangle_{\lambda} / \text{kcal} \cdot \text{mol}^{-1}$		
	Summe	Lennard-Jones	Coulomb
0.01592	$3.44 \pm 0.09$	$-0.13 \pm 0.03$	$3.57 \pm 0.09$
0.08198	$6.96 \pm 0.20$	$-0.28 \pm 0.07$	$7.24 \pm 0.19$
0.19331	$8.37 \pm 0.28$	$-0.40 \pm 0.09$	$8.77 \pm 0.27$
0.33787	$7.63 \pm 0.33$	$-0.35 \pm 0.09$	$7.98 \pm 0.32$
0.50000	$5.41 \pm 0.33$	$-0.01 \pm 0.07$	$5.41 \pm 0.32$
0.66213	$2.72 \pm 0.32$	$0.34 \pm 0.09$	$2.37 \pm 0.31$
0.80669	$0.21 \pm 0.28$	$0.40 \pm 0.09$	$-0.18 \pm 0.27$
0.91802	$-1.00 \pm 0.19$	$0.29 \pm 0.06$	$-1.29 \pm 0.18$
0.98408	$-0.78 \pm 0.09$	$0.13 \pm 0.03$	$-0.91 \pm 0.09$
<b>Summe</b>	$32.95 \pm 0.76$	$-0.01 \pm 0.21$	$32.95 \pm 0.73$

**Tab. 7.3.:** Beiträge zu  $\Delta G$  für die Transformation von 3-Aminopyridin zu 2-Aminopyridin in Wasser über die Verschiebung des Pyridin N. Hierbei bezeichnet  $w$  den Gewichtungsfaktor für die Gauss-Integration.

$\lambda$	$w \cdot \langle \partial V / \partial \lambda \rangle_{\lambda} / \text{kcal} \cdot \text{mol}^{-1}$		
	Summe	Lennard-Jones	Coulomb
0.01592	$-0.54 \pm 0.07$	$-0.10 \pm 0.02$	$-0.44 \pm 0.07$
0.08198	$-1.53 \pm 0.15$	$-0.22 \pm 0.05$	$-1.31 \pm 0.14$
0.19331	$-3.08 \pm 0.22$	$-0.31 \pm 0.09$	$-2.77 \pm 0.20$
0.33787	$-4.89 \pm 0.27$	$-0.26 \pm 0.09$	$-4.62 \pm 0.25$
0.50000	$-6.32 \pm 0.27$	$-0.00 \pm 0.07$	$-6.32 \pm 0.26$
0.66213	$-7.17 \pm 0.27$	$0.27 \pm 0.09$	$-7.44 \pm 0.25$
0.80669	$-6.95 \pm 0.20$	$0.31 \pm 0.08$	$-7.26 \pm 0.19$
0.91802	$-5.34 \pm 0.14$	$0.21 \pm 0.05$	$-5.56 \pm 0.13$
0.98408	$-2.57 \pm 0.07$	$0.10 \pm 0.02$	$-2.67 \pm 0.06$
<b>Summe</b>	$-38.39 \pm 0.60$	$0.00 \pm 0.21$	$-38.39 \pm 0.56$

### Entfernen von Wassermolekül B in der Bindetasche von Cytochrom c Peroxidase W191G

Am Ende des vorherigen Abschnitts wurde klar, dass der Beitrag des zweiten Wassermoleküls in unmittelbarer Umgebung des Liganden in keinsten Weise zu vernachlässigen ist. Bei der Transformation des Systems 1AEG in 1AEF bewegt sich kein zusätzliches Wassermolekül in die Bindetasche. Bei der Umwandlung von 1AEF in 1AEO hingegen verlässt das zweite Wassermolekül die Tasche nicht und verhindert so, dass der neu entstandene Ligand den korrekten Bindungsmodus von 2-Aminopyridin einnehmen kann. Es ist also notwendig, eine weitere Simulation durchzuführen. Um den Kreisprozess abzuschließen, müssen die Wechselwirkungen von Wasser B ausgeschaltet werden bzw. dieses Wassermolekül in der Bindetasche erzeugt werden. Die Berechnung der dafür notwendigen Freien Enthalpieänderung ist nicht nur der Abschluss der durchgeführten Testrechnungen sondern ebenso an sich von Interesse. Im Folgenden sollen ja die Beiträge einzelner Wassermoleküle zur Protein-Ligand Bindung bestimmt werden. Dafür wird es stets notwendig sein, die Wechselwirkungen einzelner Wassermoleküle auszuschalten. Der für die Testrechnungen ermittelte Wert von Wasser B kann dann z.B. mit denen von Wassermolekül A der verschiedenen Komplexe verglichen werden.

Leider konnte die Suche nach den richtigen Einstellungen bis zum Ende dieser Arbeit nicht erfolgreich abgeschlossen werden. Es wurden zahlreiche Versuche durchgeführt, den gewünschten Energiebeitrag zu berechnen. Dabei traten jedoch unterschiedliche Probleme auf, die im Folgenden ausführlich diskutiert werden sollen.

Entfernt man Wasser B, nachdem die Transformation von 3-Aminopyridin zu 2-Amino-



pyridin im Komplex mit CCP W191G abgeschlossen ist, führt das dazu, dass sich dieses Wassermolekül von seiner ursprünglichen Position weg bewegt. Durch das Ausschalten der van-der-Waals Wechselwirkungen und das Entfernen der Ladungen wird das Wassermolekül an der entsprechenden Stelle nicht mehr ausreichend festgehalten. Sobald es diesen Bereich der Bindetasche, oder sogar die Kavität an sich verlässt, werden die Freien Enthalpiebeiträge an den neuen Positionen und nicht mehr in der Bindetasche bestimmt, was nicht erwünscht ist. Aus diesem Grund wurden im Folgenden zusätzliche Kräfte auf ein oder mehrere Wassermoleküle des Systems ausgeübt, um die fehlenden „natürlichen“ Interaktionen zu kompensieren und die Besetzung der entsprechenden Position zu garantieren. Diese Zusatzkräfte können auf unterschiedliche Weise definiert und auch in ihrer Stärke variiert werden.

Ein zusätzliches harmonisches Potential auf das verschwindende Wassermolekül führte zu unterschiedlichen Ergebnissen. Wählt man eine Kraftkonstante von 1.0 kcal/mol, so ist dies offenbar zu gering, und das Wassermolekül B verlässt trotzdem die Tasche. Erhöht man den Wert auf 5.0 kcal/mol, so ist dies offenbar zu rigide, und die Simulation bricht aufgrund eines Fehlers im SHAKE-Algorithmus ab. Olano und Rick<sup>232</sup>, Hammelberg und McCammon<sup>230</sup> sowie Roux und Kollegen<sup>229</sup> berechnen die Stärke der benötigten Zusatzkraft aus der Fluktuation des jeweiligen Wassermoleküls. Gemäß

$$k_{\text{harm}} = \frac{3RT}{\langle \delta r^2 \rangle} \quad (7.3)$$

ergibt sich bei einer Fluktuation von 0.62 Å (z.B. für das Wassermolekül B in der oben diskutierten MD-Simulation von 1AEF) ein Wert von 2.88 kcal/mol. Wählt man aber eine Stärke von 3.0 kcal/mol für die Zusatzkraft, ergibt sich trotz allem das gleiche Bild wie zuvor. Wassermolekül B bewegt sich während der Entkopplung seiner Wechselwirkungen aus der Bindetasche heraus.

Möglicherweise ist die Position durch die bereits erfolgte Ligandtransformation zu 2-Aminopyridin (wo ja in der Kristallstruktur nur ein Wassermolekül in unmittelbarer Nähe des Liganden beobachtet werden kann) thermodynamisch ungünstig, sodass eine geringfügige Verringerung der Wechselwirkung in einer beinahe freien Beweglichkeit resultiert. Deshalb war es wichtig, zu überprüfen, ob die Entfernung eines Wassermoleküls *vor* der Ligand-Transformation erfolgreicher verläuft. Eine Simulation mit nur geringen Zusatzkräften führte auch hier dazu, dass das entsprechende Wassermolekül Position B im Laufe der Transformation verlässt. Da eine sehr große Kraftkonstante aber wegen des zu erwartenden Fehlers mit dem SHAKE-Algorithmus nicht in Frage kam, wurde überprüft, ob eine graduelle Erhöhung des Wertes möglich ist. Auch hierfür musste erst die richtige Größe für die Kraftkonstanten gefunden werden. Eine Kombination war erfolgreich. Diese Simulation wurde beim ersten  $\lambda$ -Wert mit einer Kraft von 3.0 kcal/mol begonnen und endete für  $\lambda_9$  bei 15.0 kcal/mol. Dieses sehr starke harmonische Potential ermöglicht nun zwar das Festhalten von Wasser B auf seiner Ausgangsposition, führt aber durch diese absolute Fixierung im Raum zu starken Veränderungen der Protein-Struktur, Deformation des Cofaktors und Wanderung des Liganden.

Das SANDER-Programm bietet neben diesem harmonischen Zusatzpotential in Bezug auf einen Referenzkoordinatensatz die Möglichkeit, die positionsbeschränkende Kraft relativ zu einem anderen Atom des Systems zu definieren. Dies kann möglicherweise eine extreme Konformationsänderung der Umgebung (wie zuvor beobachtet) verhindern, da ein einzelnes Atom nicht absolut im Raum fixiert wird. Auch diese Simulation wurde durchgeführt. Dabei wurde der Abstand zwischen O von Wassermolekül B und Carbonyl-O von Asp235, der während der Simulation bei einem Gleichgewichtswert von ungefähr 3.5 Å liegt, für die Beschränkung ausgewählt. Die Umsetzung erfolgte mit dem NMR-Modul in SANDER. Der Bereich von 2.5 bis 3.5 Å wurde zur freien Beweglichkeit ohne eine Zusatzkraft belegt. Von 1.5 bis 2.5 Å sowie zwischen 3.5 und 4.5 Å wurde eine quadratische Rückstellkraft mit der Stärke 5.0 kcal/mol hinzugefügt. Außerhalb dieser Grenzen wächst die Rückstellkraft linear an. In dieser Simulation bewegte sich das Wassermolekül jedoch auf die gegenüberliegenden Seite von Asp235 und blieb somit innerhalb des definierten Abstandsbereichs. In einer zweiten Simulation wurde zusätzlich auch die Distanz zwischen O von Wasser B und N von Gly188 mit einem Potential belegt. Damit konnte der gewünschte Effekt, die Fixierung des entkoppelten Wassermoleküls auf Position B erreicht werden. Dabei traten jedoch erneut größere Deformationen der Proteinumgebung auf. Für einen chemisch gesehenen artifiziellen Zustand ist dies nicht unbedingt überraschend. Nun musste aber noch ein anderer Aspekt beachtet werden. Um den gewünschten Endzustand zum Schließen des thermodynamischen Kreisprozesses zu erhalten, darf sich natürlich nach Entkopplung von Wasser B kein neues Molekül auf dieser Position befinden. Genau das ist aber in der letztgenannten Simulationsvariante nicht gewährleistet. Hier bewegt sich ein anderes Wassermolekül auf die leere Position B.

Alle bisherigen Versuche, das zweite Wassermolekül im Komplex 1AEF zu vernichten, wurden nicht exakt mit den eingangs beschriebenen Einstellungen durchgeführt. Die Simulation für jeden  $\lambda$ -Wert begann hier nicht bei der Endstruktur des vorherigen Simulationsabschnittes sondern stets bei der Anfangsstruktur der gesamten Simulation. Das hatte den Vorteil, dass alle Rechnungen parallel gestartet werden konnten und somit Zeit eingespart werden sollte. Dieses Verfahren wurde jedoch im Laufe der Arbeit zu dem oben beschriebenen umgestellt, da die Ergebnisse so schneller und besser konvergieren. Aber auch hier zeigten sich die analogen Probleme für die Entkopplung von Wasser B. Harmonische Positionsbeschränkungen von 3.0 bzw. 5.0 kcal/mol zeigen, dass das Wassermolekül die (relativ zur Umgebung bestimmte) Position B verlässt. Die Variante mit Abstandsbeschränkungen liefert ein im Rahmen der simulierten Zeitspanne noch nicht konvergiertes Ergebnis für  $\Delta G$ , aber auch erneut ein neues Wassermolekül, das Position B einnimmt, während die Wechselwirkungen des ursprünglichen Wassermoleküls ausgeschaltet werden. Führt man hier zusätzlich eine Positionsbeschränkung auf die Atome der PGGAAN Schleifenregion ein, verhindert das zwar ein Eindringen eines Wassermoleküls aus der Lösung in die Binde-tasche, dann nimmt aber das Wassermolekül, das ursprünglich an das Häm-Fe koordiniert war, Position B ein. Das konnte auch bei einer Simulation mit Positionsbeschränkung für alle Protein-Rückgrat Atome beobachtet werden. In diesem Fall ist es aber ohnehin fraglich, inwieweit noch ein adäquates Sampling des Systems während der Transformation gewährleistet ist.

Bisher wurde die Transformation von System 1AEF in 1AEO stets in zwei separaten Simulationen durchgeführt. Unabhängig von der Reihenfolge war die Umwandlung der Liganden und Anpassung der Solvensumgebung stets getrennt. Obwohl so für die TI-Rechnung zwar die geringere Störung definiert wird, ist es in Hinblick auf die Adaption der Lösungsmittelumgebung möglicherweise hilfreicher, beide Transformationen in einem Schritt durchzuführen. Die dafür durchgeführte Simulation zeigte wiederum ein neues Wassermolekül, das Position B einnimmt, und neben ungewöhnlichen Konformationen der Proteinumgebung auch einen nicht vollständig konvergierten Energiewert. Nachdem nun bereits eine Vielzahl unterschiedlicher Varianten erprobt worden waren, stellte sich die Frage, wie die Entfernung eines Wassermoleküls aus einem Protein-Ligand Komplex in der Literatur parametrisiert wurde.

Bereits 1995 untersuchten V. Helms und R. C. Wade<sup>228</sup> die thermodynamischen Eigenschaften von Wassermolekülen in Cytochrom P450cam. Hierbei wurde eine harmonische Zusatzkraft auf das Wasser, an dem die Störung durchgeführt wurde, angewandt. Das Wassermolekül wurde so davon abgehalten, einen bestimmten kugelförmigen Bereich zu verlassen. Innerhalb dieses Bereichs kann es sich jedoch frei bewegen. Die Kraftkonstante von  $2000 \text{ kJ}/(\text{nm}^2 \cdot \text{mol})$  entspricht  $47.8 \text{ kcal}/(\text{\AA}^2 \cdot \text{mol})$ . In dieser Arbeit wurde jedoch keine Vorkehrung getroffen, Wassermoleküle davon abzuhalten, die frei werdende Position zu besetzen. Solche Vorkehrungen können wegen der deutlich kürzeren Simulationszeiten, in denen ein solches Phänomen nicht beobachtet werden konnte oder auch wegen der Geometrie der Bindetasche nicht notwendig gewesen sein. Auch bei Hammelberg und McCammon<sup>230</sup> wurde ein harmonisches Potential verwendet, um das untersuchte Wassermolekül in der Bindungsregion festzuhalten, während andere Wassermoleküle nicht davon abgehalten wurden, die entstehende Lücke zu füllen. Bei C. Barillari und Kollegen<sup>234</sup> wird ein sog. „hard wall“ Potential verwendet, um Wassermoleküle auf bestimmten Positionen festzuhalten. Dieses Potential entspricht dem Geschehen, wenn ein Teilchen auf eine Wand trifft. Bevor es die Wand erreicht, bestehen keinerlei Einschränkungen auf die Beweglichkeit. Sobald es jedoch die starre Wand erreicht, erfährt es eine (unendlich) große Rückstellkraft. Diese Art des Zusatzpotentials hindert andere Wassermoleküle in gleichem Maße am Eindringen in den Bereich innerhalb der Wände. Dieses Vorgehen ist so nicht auf die hier durchgeführten Simulationen übertragbar. Barillari *et al.* führten keine MD- sondern MC-Simulationen durch. Die beschriebene Potentialfunktion ist zudem im hier eingesetzten Simulationsprogrammpaket AMBER nicht verfügbar. Schließlich beschreiben Olano und Rick<sup>232</sup> das Entfernen von Wassermolekülen unter Zuhilfenahme eines harmonischen Potentials, um das betrachtete Wassermolekül am Verlassen einer Position zu hindern. Gleichzeitig wurde aber eine zusätzliche, kurzreichweitige Kraft auf *alle* Lösungsmittelmoleküle ausgeübt. Da diese relativ zur Position des verschwindenden Moleküls definiert wird, verhindert sie das Eindringen von Wassermolekülen in die leere Taschenregion. Das Potential hat die Form

$$U_r = \sum_k \varepsilon_r \left( \frac{r_{\text{Oh}}}{\sigma_r} \right)^{-12}. \quad (7.4)$$

Darin bezeichnet  $r_{\text{Oh}}$  den Abstand des O-Atoms des Referenzwassermoleküls in der Binde-

tasche und der O-Atome aller Solvensmoleküle. Die Simulation wurde mit dem Programm Gibbs durchgeführt, das zwar auch zum AMBER Paket gehört, jedoch in den aktuellen Versionen nicht mehr enthalten ist und zudem keine Parallelisierung auf mehrere Prozessoren ermöglicht.

In Analogie hierzu wurde nun auch versucht, mit den in SANDER zur Verfügung stehenden Methoden eine Beschränkung für andere Wassermoleküle im System zu definieren. Zusätzlich zu den oben bereits eingeführten Zusatzkräften im Bereich der flexiblen Loop-Region wurden mit Hilfe der über das NMR-Modul zur Verfügung stehenden Abstandsbeschränkungen alle Wassermoleküle in unmittelbarer Umgebung von Wasser B (Wassermolekül koordiniert an Fe, Wasser C und Wasser A) mit einem harmonischen Potential (Stärke 5.0 kcal/mol) daran gehindert, der Position B mehr als 1.5 Å näher als der Gleichgewichtsabstand (je 11.0 Å, 7.6 Å und 5.0 Å) zu kommen. Dieses Vorgehen war zwar erfolgreich, jedoch erreicht im Verlauf der Simulation nun ein Wasser aus dem umgebenden Lösungsmittel Position B. Als Konsequenz daraus wurde versucht, alle Wassermoleküle im System (mehr als 8000) relativ zu Position B mit einem Potential zu belegen. Das ist aber offenbar technisch nicht möglich. Setzt man diese neue Beschränkung wiederum über Positionsbeschränkungen um, kann zumindest ein Teil der Simulation durchgeführt werden. Schließlich bricht aber auch diese Rechnung mit einem Fehler ab. Bei der immer weiter ausgedehnten Menge an Beschränkungen kann immer weniger von einem wirklichen Durchforsten des Konfigurationsraums gesprochen werden und ebenso gut auf den Einsatz eines expliziten Solvensmodells verzichtet werden. Ein erster Versuch in diese Richtung wurde ebenfalls unternommen. Aufgrund der nicht vorhandenen Expertise mit dieser Art von Simulationen konnte die entsprechende Berechnung in der vorhandenen Zeit nicht durchgeführt werden und musste zurück gestellt werden.

### 7.1.4. Ausblick

In der vorliegenden Arbeit sollten TI-Rechnungen eingesetzt werden, um die thermodynamischen Beiträge einzelner Wassermoleküle in Protein-Ligand Bindungen zu ermitteln. Leider konnten diese Berechnungen im zur Verfügung stehenden Zeitraum aufgrund nicht absehbarer Probleme nicht abgeschlossen werden. In den vorangegangenen Abschnitten wurde dargestellt, wie ein robustes Vorgehen zur Durchführung einer solchen Berechnung verifiziert wurde, sodass aussagekräftige Ergebnisse erwartet werden können. Anhand von einfachen Transformationen, deren Freie Enthalpieänderung Null entsprechen muss, konnte dies für verschiedene Aminopyridine gelöst in Wasser überprüft werden. Die entsprechenden Simulationen mit den identischen Molekülen als Liganden in der Bindetasche von CCP W191G bereiteten bei der Anpassung der Solvensumgebung – wie oben dargestellt – verschiedene Probleme.

Die Wahl von Zusatzkräften brachte hierbei nur Teilerfolge. Es ist möglich, ein Wassermolekül, dessen Wechselwirkungen mit seiner Umgebung schrittweise reduziert werden, im Bereich seiner ursprünglichen Position festzuhalten. Andererseits ist es aber nicht gelungen, diese Position, die ja am Ende der Transformation leer sein soll, frei von anderen Wassermolekülen zu halten. Zusätzlich wurden bei den durchgeführten alchemischen Transforma-

tionen immer wieder sehr ungewöhnliche Konformationen des Protein-Systems beobachtet. Dies ist möglicherweise aber gar nicht bemerkenswert, da das Konzept der TI die Simulation einer Vielzahl chemisch nicht existierender, artifizieller Zustände benötigt. Nachdem die Ideen und Variationsmöglichkeiten bei der Parametrisierung dieser Rechnungen so gut wie ausgeschöpft wurden, verbleibt als mögliche Alternative nur eine aufwändige Modifizierung des Programmcodes oder der Einsatz einer anderen Simulationssoftware, die andere Definitionsmöglichkeiten für Zusatzpotentiale bietet. Dieser Schritt zieht wiederum eine gründliche Überprüfung aller eingestellten Optionen und Parameter der MD-Simulation nach sich und ist somit ebenfalls zeitintensiv.

Die Parametrisierung beim Entfernen eines Wassermoleküls aus dem Protein-Ligand Komplex ist sehr aufwändig. Möglicherweise muss dieses Vorgehen aber für jedes andere Wassermolekül neu angepasst werden. Dadurch und durch die hohen Kosten der Rechnungen ist dieses Vorgehen nur sehr eingeschränkt auf zahlreiche Wassermoleküle und ebenso in anderen Proteinsystemen einsetzbar. Andere Biomoleküle erfordern möglicherweise ein modifiziertes Vorgehen. Die Komplexität der TI-Rechnungen in Hinblick auf die Reversibilität der Simulationen erlaubt kein einheitliches Arbeitsprotokoll. Auf die Notwendigkeit von energetischen Informationen für eine große Zahl von Wassermolekülen, z.B. zur Verbesserung von Scoring-Funktionen, wurde schon mehrfach in dieser Arbeit hingewiesen. Gleichzeitig benötigt man, z.B. in einem virtuellem Screening, Informationen für ein einzelnes Wassermolekül. Der grundsätzlichen Problemstellung steht also der konkrete Bedarf in einem speziellen Fall gegenüber. Für den Wissenschaftler wäre es letztlich also erstrebenswert, bereits vorhandene Daten, wie die MD-Trajektorie des Systems, die eine chemisch relevante Dynamik beschreibt, mit den gesuchten Informationen zu verknüpfen. Im Prinzip ist das mit dem LIE-Ansatz<sup>67,68</sup> möglich, nur bisher noch nicht durchgeführt worden. Der Nachteil dabei ist, dass die Methode an experimentellen Werten für ein System kalibriert werden muss.  $\Delta G$ -Werte liegen aber höchstens für Liganden vor, nicht für einzelne Wassermoleküle. Eine lohnenswerte Alternative stellt der in den folgenden Abschnitten dargestellte Ansatz dar.

## 7.2. Anwendung der Solvatationstheorie inhomogener Lösungen

In Abschnitt 3.4 wurde die Methode von Lazaridis und Li vorgestellt, die entropische Beiträge einzelner Wassermoleküle in der Protein-Ligand Bindung mit Hilfe von Zusammenhängen der Solvatationstheorie berechnet. Im Folgenden sollen sowohl die Schwierigkeiten bei der praktischen Umsetzung dieses Ansatzes, wie auch die Ergebnisse für die in dieser Arbeit untersuchten Protein-Kristallstrukturen vorgestellt werden.

### 7.2.1. Implementierung

In Ermangelung einer verfügbaren Implementation dieser Methode wurde bereits in<sup>239</sup> und später dann für die hier vorliegende Arbeit versucht, eine Umsetzung aller relevanten

Schritte in der Programmiersprache PYTHON zu realisieren. Es stellte sich hierbei jedoch heraus, dass dies aufgrund der Komplexität der Methode (insbesondere Berechnung der Orientierungs-Koordinaten, Integration der Mehrteilchen-Korrelationsfunktionen) und der nicht sehr detaillierten Darstellung bei Lazaridis und Li (beispielsweise die Berechnung der Korrelationsfunktionen) nicht sehr erfolgversprechend war.

Dieses Vorgehen wurde eben deshalb aufgegeben. Freundlicherweise wurde der Arbeitsgruppe für diese Studie der Original-Programm-Code aus der Lazaridis-Gruppe zur Verfügung gestellt. Die praktische Anwendung dieser Methode verzögerte sich dennoch aus verschiedenen Gründen. Die für Lazaridis und Li zugrundeliegenden Simulationen wurden mit dem Programm CHARMM durchgeführt. Deshalb ergab sich zwangsläufig die Notwendigkeit, die vorliegenden Trajektorien-Dateien zu konvertieren. Um dies zu umgehen, wurde geprüft, ob aufgrund der nun vorliegenden Details zu den Berechnungen eine eigene Implementierung, die auf den mit AMBER generierten Dateien basiert, einfach möglich ist. Dies war jedoch nicht der Fall, größtenteils deshalb, weil der vorliegende Original-Programmcode schlecht bis gar nicht kommentiert vorlag. Mit einem modifizierten Fortran-Programm von T. Lazaridis<sup>240</sup> wurde deshalb die Trajektorie in einzelne Dateien im CHARMM-Koordinaten-Format umgewandelt. Die anschließende Kompilierung der Programme für die Berechnungen nach dem Ansatz über die Theorie inhomogener Lösungen stellte sich ebenfalls als nicht geradlinig heraus. Zwar führte die Kompilierung mit dem freien GCC-Fortran-Compiler<sup>241</sup> zu einem ausführbaren Programm. Dieses konnte aber die gewünschten Berechnungen nicht durchführen. Höchstwahrscheinlich lag das in einer zu weit verschachtelten Schleife im Abschnitt für die numerische Integration der multidimensionalen Korrelationsfunktionen begründet. Abhilfe schaffte hier der Einsatz des Portland Fortran-Compilers<sup>242</sup>. Mit der damit erzeugten Binär-Programmdatei erfolgte dann die Durchführung der Berechnungen wie von Lazaridis und Li beschrieben.

Als Datengrundlage wurde bei den Berechnungen die Trajektorie der Komplexe verwendet, in denen während der Simulation Zusatzkräfte auf Atome der Schleifen-Region der Aminosäuren 190-195 angewandt wurden. Dies war notwendig, um sicher zu stellen, dass kein Austausch der Wassermoleküle auf den charakteristischen Positionen in der Bindetasche stattfindet. Technisch ist es mit der zur Verfügung stehenden Software nur möglich, einzelne Wassermoleküle entropisch und enthalpisch zu charakterisieren. Im Vergleich zu den bei Lazaridis und Li<sup>81</sup> angewandten Beschränkungen der Systemdynamik, sind die hier eingesetzten Zusatzkräfte moderat und vertretbar. Für das System 1AEF wurde die Trajektorie verwendet, die auf der neuen Startstruktur beruht. Für die Auswertung des Systems 1CPE musste eine zusätzliche Trajektorie berechnet werden, da in der bisher vorhandenen zwar die Positionen aller Wassermoleküle in der Bindetasche konserviert sind, die Wassermoleküle auf diesen Positionen aber sehr häufig wechseln. Die neue Trajektorie wurde unter Zuhilfenahme von Zusatzkräften auf verschiedene Atomgruppen berechnet. Dies umfasst im Einzelnen eine Kraft mit der Stärke 5.0 kcal/mol auf die Protein-Rückgrat-Atome, von 5.0 kcal/mol auf das K<sup>+</sup>-Ion, sowie von 1.0 kcal/mol auf alle Bindetaschenwassermoleküle. Dieses zusätzliche harmonische Potential zieht die entsprechenden Atome zurück auf die Position eines vorherigen Trajektorienabschnitts. Zu diesem Zweck wurde die Trajektorie in Abschnitten von 50 ps berechnet und das Ende des vorangegangenen Abschnittes als Re-

ferenz für die Rückstellkraft des folgenden verwendet. Auf diese Weise ist immer noch eine leichte Bewegung der Atome möglich, wobei sie für die Bindetaschenwassermoleküle (unter Anderem durch die relativ rigide Proteinumgebung) weitgehend eingeschränkt ist. Um sicher zu stellen, dass die Simulation ausgehend von einer in Lösung sehr gut equilibrierten Struktur durchgeführt wird, wurde die Proteinkonformation der zuvor durchgeführten freien Simulation von 1CPE verwendet, die einer gemittelten Struktur am Nächsten kommt (nach 3783 ps). Darin stimmen die Wasserpositionen in der Bindetasche sehr gut mit denen überein, die über die gemittelten Hydratationsdichteverteilungen beobachtet werden können. Die Auswahl an Zusatzkräften führte dazu, dass zwar ein Durchforsten des Konfigurationsraums möglich war, aber die Wassermoleküle auf den Positionen konserviert blieben.

Nach der Konvertierung der Rohdaten wurden Regionen um ein geordnetes Wassermolekül, das untersucht werden soll, bestimmt, in denen andere Wassermoleküle zu finden sind. Diese müssen separat definiert werden, da die Methode (vgl. Gl. 3.62, 3.63) voraussetzt, dass Orientierungs- und Translationskoordinaten voneinander unabhängig sind. Durch Auffinden der Unterregionen ist gewährleistet, dass für die Wassermoleküle dieser Region keine wesentliche Änderung der Translationskoordinaten zu beobachten ist. Anschließend werden dann die Euler-Winkel, Korrelationsfunktionen und schließlich die Entropie berechnet. Für die numerische Evaluation der Integrale wurde die Intervallgröße folgendermaßen festgelegt:  $dr = 0.1 \text{ \AA}$ ,  $d\phi' = \pi/3$ ,  $d\theta' = \pi/6$ ,  $d\phi = \pi/5$ ,  $d\theta = \pi/10$ ,  $d\psi = \pi/5$ . Aus den einzelnen Entropietermen erhält man schließlich  $\Delta S_{\text{solv}}$ . Vergleiche hierzu Gl. 3.67 und 3.68.

Wie bei Lazaridis angegeben<sup>81-83</sup>, werden die energetischen Beiträge  $E_{\text{SW}}$  und  $E_{\text{WW}}$  direkt aus den vorliegenden Trajektorien gewonnen. Für diese Berechnungen wurde das MM-PBSA Modul aus AMBER 10 verwendet. Hierbei wurden jedoch einzig die Routinen verwendet, um an Hand der Trajektorien die Wechselwirkungsenergien für jede Konfiguration zu bestimmen und anschließend statistisch auszuwerten. Da das System ja bereits solvatisiert war, mussten nur die „vermeintlichen“ Gasphasen-Wechselwirkungen berechnet werden. Hierfür konnten keine periodischen Randbedingungen genutzt werden. Die Berechnung wurde separat für die Interaktion von untersuchtem Wassermolekül mit dem Rezeptor (inkl. Cofaktor) und dem Solvens bei einer Dielektrizitätskonstante von 1.0 vorgenommen. Außerhalb des Systems/der Wasserbox wurde eine Dielektrizitätskonstante von 78.5 verwendet. Nicht-kovalente Wechselwirkungen wurden bis zu einer Distanz von 99.0  $\text{\AA}$  berechnet. Entsprechend der Standardeinstellungen in SANDER wurden die van-der-Waals-Wechselwirkungen mit einem Wert von 1/2.0 (Option `scnb`) und die elektrostatischen Wechselwirkungen mit 1/1.4 (Option `scee`) skaliert. Das Coulomb-Potential wurde dabei unmittelbar ohne sog. „switch function“ bestimmt. Aus diesen Daten kann dann der Wert  $\Delta E_{\text{solv}}$  bestimmt werden.

$$\Delta E_{\text{solv}} = E_{\text{SW}} + \frac{1}{2}E_{\text{WW}} - (-9.5 \text{ kcal/mol}) \quad (7.5)$$

Der Wert von -9.5 kcal/mol ist hierbei die Hälfte der Solvens-Solvens Wechselwirkungsenergie im reinen Lösungsmittel, der aus einer reinen Wassersimulation bestimmt wurde.

Lazaridis<sup>83</sup> verwendet hierfür einen Wert von -10.1 kcal/mol für das von ihm verwendete Kraftfeld.

### 7.2.2. Berechnung entropischer Beiträge einzelner Wassermoleküle zur Protein-Ligand Bindung

#### Vorgehen und Fehlerabschätzung

Die Berechnungen entropischer Beiträge mit der solvationstheoretischen Methode nach Lazaridis wurden für alle geordneten Wassermoleküle A, B und C der untersuchten Komplexe der CCP W191G durchgeführt. Ebenso wurden alle Wassermoleküle in der Bindetasche von 1CPE untersucht. Zunächst mussten hierfür die Unterregionen identifiziert werden, in denen Translations- und Orientierungskordinaten weitgehend unabhängig voneinander sind. Die berechneten Verteilungen der Orientierungskordinaten wurden aufgetragen und die klar voneinander separierten Regionen identifiziert. Eine Auswahl für bestimmte Abstandsintervalle aus allen erstellten Diagrammen ist in Abb. 7.6 und 7.7 gezeigt. Zusätzlich sind die gewählten Unterregionen durch rote Linien gekennzeichnet. Teilweise wurden diese Bereiche jedoch nicht im kompletten Intervall  $2 \text{ \AA} < r < 5 \text{ \AA}$  definiert, um eine bessere Abtrennung der Unterregionen zu erreichen.

Anschließend wurden mit Hilfe des Programms aus der Gruppe von Lazaridis die Entropien und mit SANDER die Energiebeiträge bestimmt. Die Ergebnisse dieser Berechnungen sind in Tab. 7.6 und 7.7 zusammengefasst. Die Analyse der Trajektorien hinsichtlich der Energien liefert, da es sich um einen Mittelwert aus den Wechselwirkungsenergien für jede Konfiguration der Trajektorie handelt, auch eine Standardabweichung, die hier als Fehlerbereich mit angegeben wurde. Die Größenordnung dieses Fehlers ist folglich ebenso für die freie Gibbs-Energie anzusetzen. Die hier angegebenen Entropiewerte haben keine Fehlerangaben, da nur ein Wert aus der Korrelationsfunktion berechnet wurde. Prinzipiell wäre es möglich Entropien für unterschiedliche Trajektorienabschnitte zu ermitteln und zu mitteln. Dennoch ist es wichtig zu bedenken, welche Fehlerquellen bei der Berechnung grundsätzlich relevant sind. Neben den bei der Entwicklung der Methode getroffenen Näherungen ist dies zum einen ein möglicherweise unvollständiges Durchforsten des Konfigurationsraums durch die MD Simulation. Die zuvor durchgeführten Untersuchungen der Trajektorien legen diese Vermutung jedoch nicht nahe. Zum anderen kann eine weitere Unsicherheit durch die Berechnung der Orientierungskordinaten entstehen. Wie bereits erwähnt, gibt es leicht unterschiedliche Möglichkeiten, die Referenz bei der Berechnung der Euler-Winkel zu wählen. Somit können unterschiedliche Winkelsätze die gleiche Orientierung eines Moleküls beschreiben. Dadurch ergeben sich folglich auch unterschiedliche Verteilungen bzw. Korrelationsfunktionen in Abhängigkeit dieser Winkel. Eine Zustandsgröße wie die Entropie sollte wiederum unabhängig von der Form der Verteilung sein. Nach der hier durchgeführten numerischen Integration können aber geringe Unterschiede auftreten<sup>†</sup>. Exemplarisch wurde deshalb die Bestimmung des Orientierungsanteils der Entropie

---

<sup>†</sup>In einer persönlichen Mitteilung von T. Lazaridis wurde auf diese Schwierigkeit hingewiesen.



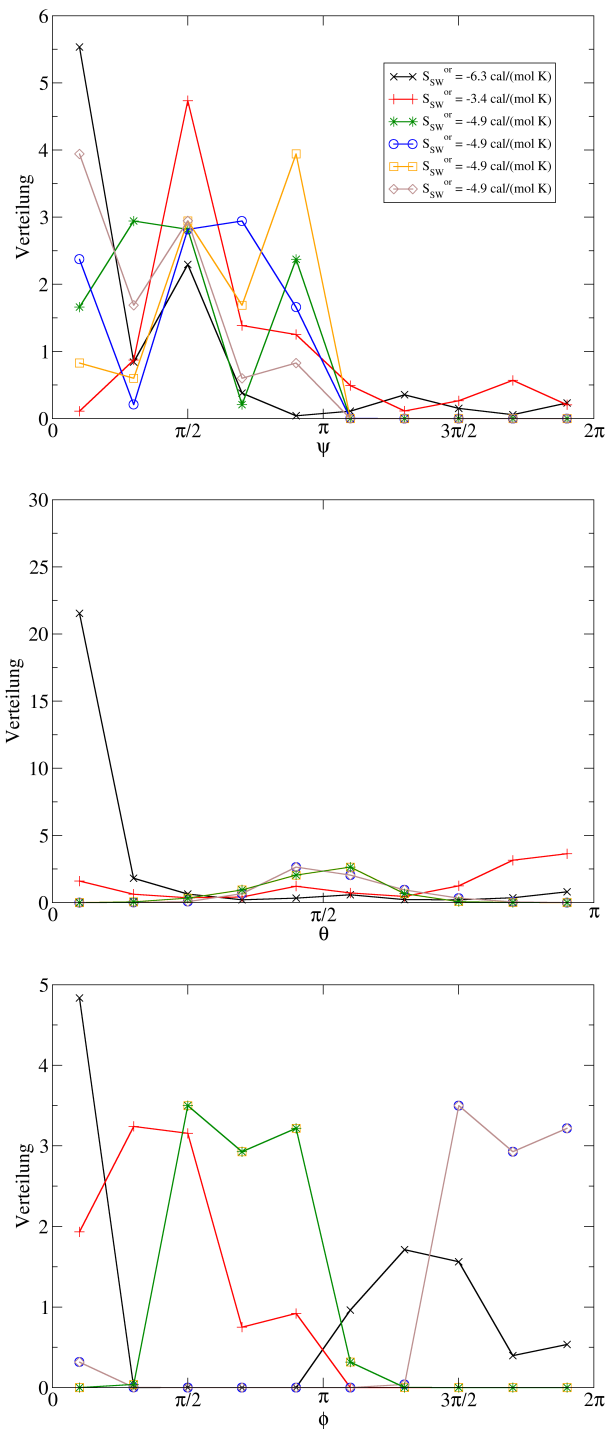
für Wassermolekül 307 aus Komplex 1AEG für die Differenz bezogen auf den Zustand in reiner Solvensphase und gelöstem Stoff für alle implementierten Methoden zur Bestimmung der Eulerwinkel durchgeführt. Die dadurch erhaltenen Verteilungen der Euler-Winkel sind in Abb. 7.4 zusammen mit den daraus resultierenden Entropiewerten gezeigt. Die Unterschiede für diese Größe sind dabei nur sehr gering. Einzig die Werte, die zu der in Schwarz und Rot aufgetragenen Kurve gehören, zeigen eine Differenz von 1 bis 2 cal/(mol · K) im Vergleich zu den übrigen Ergebnissen. Das jedoch ist eine akzeptable Größenordnung für die Unsicherheit. Für die endgültige Auswertung wurde in diesem Fall die in grün dargestellte Verteilung gewählt, da sie für alle drei Winkel relativ gleichmäßig ist.

### Vergleich der Systeme

**Komplexe.** Was bedeuten nun die ausgerechneten Werte? Für welchen Prozess wird die Änderung der thermodynamischen Größen beschrieben? Nach Li und Lazaridis<sup>83</sup> ist  $\Delta S_{\text{solv}}$  der Beitrag zur Solvatationsentropie des jeweils untersuchten, geordneten Wassermoleküls.  $\Delta G_{\text{solv}}$  beschreibt also die Freie Enthalpieänderung beim Transfer des untersuchten Wassermoleküls aus dem das Protein umgebenden Solvens in die Bindetasche. Eine intuitive Interpretation der Daten ist möglich, wenn man die Werte der Wasserpositionen untereinander vergleicht.

In beinahe allen Komplexen (Daten vgl. Tab 7.6) mit Ausnahme von 2AQD ist Wassermolekül C enthalpisch gesehen schwächer gebunden als Wassermolekül A, d.h. es muss eine geringere Energiemenge zur Verdrängung ins Solvens aufgewendet werden. Im Rahmen der Größenordnung der diskutierten Unsicherheiten sind diese Unterschiede jedoch gering. Wassermolekül A kann durch die größere Zahl der Bindungspartner das Wasserstoffbrückenbindungsnetzwerk offenbar besser absättigen. Überraschenderweise wird für alle Komplexe ein Entropiegewinn beim Transfer des Wassermoleküls aus dem Solvens in die Tasche vorhergesagt. Im Vergleich ist dieser für Wasser A stets höher als für Wasser C. Das ist generell überraschend, da durch die vermeintliche Einschränkung der Translations- und Rotationsfreiheitsgrade eine Zunahme der Ordnung zu erwarten gewesen war, obwohl  $\Delta S_{\text{solv}}$  ja der Wert relativ zum reinen Solvens ist. Dieser Trend spiegelt jedoch eigentlich nicht die Beobachtungen aus den Simulationen wider. Während Wasser C „rigide“ ist, wurde bei Wasser A eine erhöhte Beweglichkeit festgestellt. Es gilt jedoch zu beachten, dass für die hier vorliegende Auswertung die Trajektorien mit Zusatzkräften verwendet wurden. In diesen findet man aber gerade keine erhöhte Fluktuation an Position A. Die Entropieterme enthalten im Übrigen keinen Anteil an Wasser-Wasser Korrelation, da bei den Komplexen im Wesentlichen nur isolierte Lösungsmittelmoleküle betrachtet werden. Im Vergleich zu Position C sind die Wassermoleküle auf Position A entropisch betrachtet instabiler. Aufgrund der geringen numerischen Unterschiede der Einzeltermine  $S_{\text{SW}}^{\text{tr}}$  und  $S_{\text{SW}}^{\text{or}}$  variieren die Werte für  $\Delta S_{\text{solv}}$  selbstverständlich auch nur in einem kleinen Bereich. In Summe führen die Enthalpie- und Entropie-Werte dazu, dass die Freie Solvatisierungsenthalpie von Wassermolekül C in den jeweiligen Komplexen den im Vergleich ungünstigeren Wert annimmt. Desweiteren ist deutlich erkennbar, dass Position A in 1AEG zu den thermodynamisch vorteilhaftesten gehört.

## 7. Bestimmung thermodynamischer Größen



**Abb. 7.4.:** Exemplarische Darstellung der Abhängigkeit des Orientierungsanteils der Entropie von der Verteilung der Euler-Winkel, berechnet für verschiedene Referenzzustände. Die dargestellten Werte sind dem System 1AEG, Wassermolekül 307 entnommen.

Zwei Ausnahmen von diesen Trends sind detaillierter zu betrachten. Wassermolekül C in System 2AQD ist zwar entropisch gesehen weniger stabil als Position A, enthalpisch betrachtet kehrt sich dieser Trend jedoch um, sodass letztlich beide Positionen einen annähernd gleich großen Betrag der Freien Solvatisierungsenthalpie zeigen. In der Trajektorie frei von Zusatzkräften zeigten gerade die Wassermoleküle auf Position A einen häufigen Austausch. In der entsprechenden Trajektorie mit Zusatzkräften, die aus den oben genannten Gründen für diese Analysemethode verwendet werden musste, konnte diese hohe Fluktuation jedoch nicht mehr beobachtet werden. Diese Dynamik kann sich demnach, wie bereits angemerkt, auch nicht in den Daten der thermodynamischen Analyse widerspiegeln. Dies zeigt, wie weit der solvationsstheoretische Ansatz von der Datengrundlage abhängt und wo eine seiner Limitationen anzufinden ist. Stände wie im Programm WaterMap eine Clusteringmethode zur Verfügung und könnte man somit vom technischen Standpunkt aus unterschiedliche Wassermoleküle auf einer Position handhaben, würden die thermodynamischen Größen im Fall 2AQD wahrscheinlich anders aussehen.

Desweiteren ist Wasserposition B in System 1AEF von den drei darin analysierten Wassermolekülen am instabilsten. Das deckt sich mit der Beobachtung, dass das entsprechende Wassermolekül in allen anderen Komplexen verdrängt wurde. Bemerkenswert ist aber, dass der Unterschied zu Position C nur sehr gering ist. Aus dem Wissen um die Abwesenheit von Wassermolekül B in fast allen CCP W191G Komplexen kann man demnach schließen, dass Wassermolekül C höchstwahrscheinlich ebenfalls eine andere, thermodynamisch stabilere Position einnehmen würde, wenn es nicht aus z.B. sterischen oder kinetischen Gründen daran gehindert würde.

Aus dem Bindungsmodus der Liganden ist bekannt, in welchen Fällen eine Wechselwirkung zwischen Wassermolekül A und dem Liganden vorliegt. Diese spiegelt sich in den thermodynamischen Daten entsprechend wider. Für die Systeme 1AEF und 1AEG, in denen diese Wasserstoffbrückenbindung beobachtet wird, ist die Energie für Position A sehr niedrig. Gleichzeitig zeigt das System 2AQD einen im Vergleich dazu hohen Wert.  $\Delta G_{\text{solv}}$  für Position A in 1AEO und 2EUP ist jedoch negativer als das auf Grund der fehlenden Wechselwirkung zu erwarten war. Möglicherweise liegt das in stärkeren Wechselwirkungen mit dem umgebenden Protein begründet. Das kann jedoch auf Grund der hier vorliegenden Daten nur vermutet werden. Zum einen weil bei der durchgeführten Berechnung der Wechselwirkungsenergien aus den Trajektorien keine Aufteilung in Einzelbeiträge von Protein, Solvens, Cofaktor vorgenommen wurde. Zum anderen erlaubt die aktuelle Implementierung des solvationsstheoretischen Ansatzes ebenfalls keine solche Aufteilung.

**Ligandfreie Tasche.** Der Vergleich mit dem System 1CPE, also der Struktur ohne komplexiertes Ligandmolekül in der Bindetasche, ist von ebenso großem Interesse. Die berechneten Werte hierfür finden sich in Tab. 7.7. Hier lag während der Simulation ein  $K^+$ -Ion vor, das, wie weiter oben erwähnt, die Positionen der Wassermoleküle beeinflusst. Gleichzeitig finden die Wassermoleküle aber weiterhin eine geladene Wechselwirkung in der Bindetasche vor, weshalb die enthalpischen Größen von Komplexen und 1CPE miteinander vergleichbar sein sollten. Die in dieser Trajektorie angewandten Zusatzkräfte beeinflussen jedoch die

## 7. Bestimmung thermodynamischer Größen

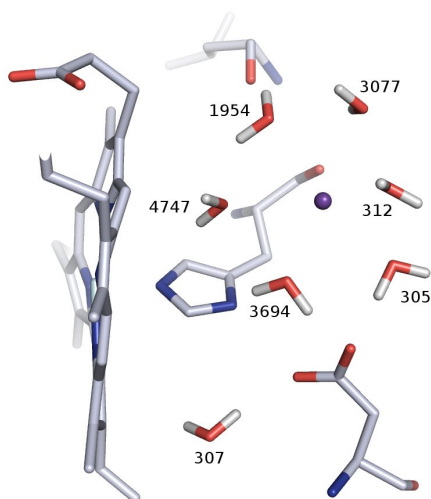
**Tab. 7.4.:** Beiträge der Unterregionen zur Solvens-Solvens Entropie einzelner Wasserpositionen im System 1CPE.

Wassermolekül	Nachbarregion	korrespondierendes Wassermolekül	$S_{\text{WW}} / (\text{cal}/(\text{mol} \cdot \text{K}))$
305	1	3694, 4747	-0.14
	1	4747	-0.18
1954	2	4747	-0.13
	3	3694	-0.004
	4	312	0.02
	5	3077	-0.36
	1	1954	-0.37
3077	2	312	0.04
	3	305	0.03
	1	4747	-0.12
3694	1	4747	-0.12
4747	1	3694	-0.30
	2	312	0.01

Dynamik der Wassermoleküle. Das ist höchstwahrscheinlich der Grund für die im Vergleich geringeren Entropiewerte. Der Wert für die Wasser-Wasser Korrelation ist nun von Null verschieden (vgl. Tab 7.4), da die Tasche mit mehreren Wassermolekülen gefüllt ist.

Beim direkten Vergleich der Positionen A, B und C fällt auf, dass Position A ähnlich stabil ist wie Position C. Wasserposition B hingegen ist nun deutlich stabiler als die Vergleichspositionen A oder C. Das wurde so nicht erwartet. Besonders die Wechselwirkungen mit dem gelösten Stoff, also in diesem Fall dem  $\text{K}^+$ -Ion, haben stark zugenommen. Offenbar tritt aber in Abwesenheit des  $\text{K}^+$ -Ions eine entsprechende Destabilisierung ein, sodass die Verdrängung auf Grund der Enthalpie-/Entropie-Bilanz bevorzugt ist.

Die Wassermoleküle 312, 1954, und 4747 (zur Übersicht über die Nummerierung vergleiche Abb. 7.5) werden in keiner der Komplexstrukturen beobachtet und man findet sie dementsprechend auch auf den im Vergleich weniger stabilen Positionen in 1CPE. Mit einem Wert von -13.8 kcal/mol für die Freie Solvatisierungsenthalpie von Wassermolekül 3694 ist diese Position im Vergleich gesehen relativ stabil. Das liegt vor allem an der sehr günstigen enthalpischen Wechselwirkung, die direkt mit der benachbarten Säurefunktionalität des Asp ausgebildet werden kann. Infolgedessen ist die Verdrängung dieses Wassermoleküls in den Komplexen nur möglich, weil die Wechselwirkung der Liganden mit dem Asp235 den Beitrag des Wassermoleküls kompensieren kann.



**Abb. 7.5.:** Wassernummerierung in 1CPE wie in den solvatationstheoretischen Berechnungen verwendet. Da für die thermodynamischen Berechnungen eine Trajektorie ausgehend von einer equilibrierten Struktur verwendet wurde, unterscheidet sich die Wassernummerierung darin von der Kristallstruktur. Für den Vergleich siehe Tab. 7.5.

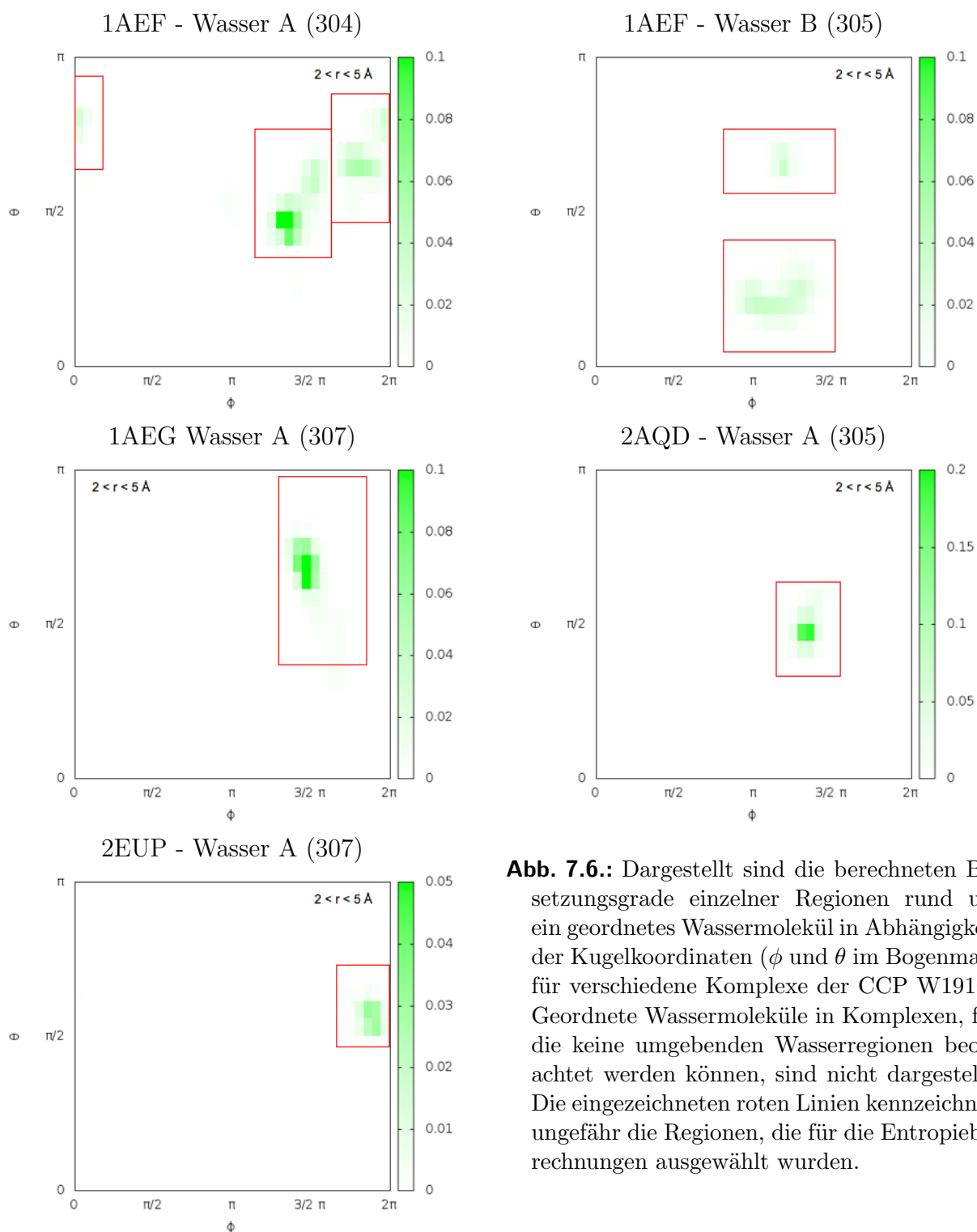
**Tab. 7.5.:** Vergleich der Wassernummerierung zwischen der MD-Trajektorie von 1CPE wie sie für die solvatationstheoretischen Berechnungen verwendet wurde mit der Kristallstruktur, soweit eine strukturelle Entsprechung vorhanden ist.

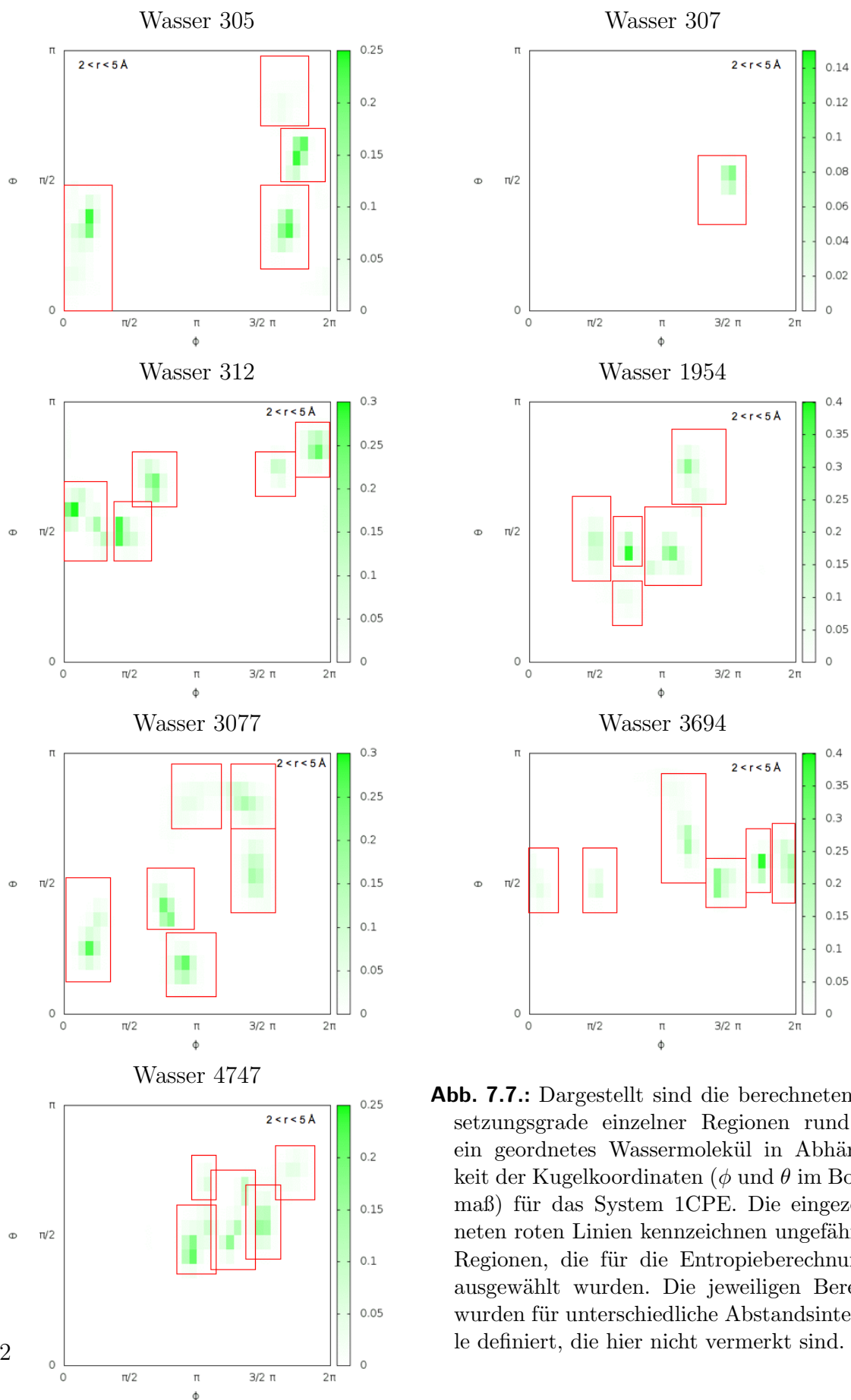
MD-Simulation	Kristallstruktur
305	977
307	535
3077	477
3694	978
4747	979

### Literaturvergleich

Bei Lazaridis wurde 2003<sup>81</sup> für das sogenannte Flap-Wasser der HIV-1 Protease ein Wert für  $S_{\text{SW}}$  von  $-25.0 \text{ cal}/(\text{mol K})$  und für  $E_{\text{SW}}$  von  $-28.2 \text{ kcal/mol}$  bestimmt. Die Werte für das untersuchte Wassermolekül von Concanavalin A<sup>82</sup> liegen mit  $S_{\text{SW}} = -22.1 \text{ cal}/(\text{mol} \cdot \text{K})$  und  $E_{\text{SW}} = -30.2 \text{ kcal/mol}$  in der gleichen Größenordnung. In beiden Fällen wurden keine Beiträge durch Wasser-Wasser Korrelation beobachtet. Bereits damit zeigt sich, dass die Wechselwirkungsenergien der Wassermoleküle mit ihrer Umgebung ähnlich stark sind, obwohl in diesen Arbeiten die Simulationen mit einem anderen Kraftfeld und stärkeren Beschränkungen/Zusatzkräften durchgeführt wurden als hier. Die jeweiligen Entropiewerte  $\Delta S_{\text{solv}}$  in den Arbeiten von Lazaridis liegen jedoch im negativen Bereich, während in der hier vorliegenden Arbeit positive Werte bestimmt wurden. In der Arbeit von Li und Lazaridis aus 2006<sup>83</sup> wurden verschiedene Wassermoleküle untersucht, die Wechselwirkungen untereinander zeigten. Die Werte für  $E_{\text{SW}}$  liegen in einem Bereich von  $-3.7$  bis  $-25.6 \text{ kcal/mol}$ , die für  $S_{\text{SW}}$  zwischen  $-10$  und  $-21.3 \text{ cal}/(\text{mol K})$ . Weiterhin ist  $E_{\text{WW}}$  zwischen  $3.4$  und  $-15.6 \text{ kcal/mol}$  sowie  $S_{\text{WW}}$  bei  $0.2$  bis  $-5.6 \text{ cal}/(\text{mol K})$ . Daraus resultiert eine Spanne für die Freie Enthalpie für die Solvatisierung von  $-1.9$  bis  $-14.8 \text{ kcal/mol}$ . Diese Ergebnisse zeigen, wie unterschiedlich die entropischen Beiträge allein für verschiedene Wassermoleküle in einer Bindetasche sein können. Gleichzeitig stellt sich die Frage, warum in der vorliegenden Arbeit positive Werte für  $\Delta S_{\text{solv}}$  bestimmt wurden.  $S_{\text{SW}}$  ist mit Werten von  $-8.4$  bis  $-13.9 \text{ cal}/(\text{mol K})$  in der vorliegenden Arbeit zwar in einer ähnlichen Größenordnung, wie die Werte in den Arbeiten von Li und Lazaridis. Nach Subtraktion des konstanten Terms von  $-15.2 \text{ cal}/(\text{mol K})$  für das reine Solvens erhält man jedoch einen Wert mit entgegengesetztem Vorzeichen bezogen auf die Angaben bei Lazaridis. Der verwendete Term für die reine wässrige Phase ist sehr wahrscheinlich von den Kraftfeldparametern abhängig. Der Wert von  $-15.2 \text{ cal}/(\text{mol K})$  wurde für das TIP4P Wassermolekülmodell im CHARMM Kraftfeld bestimmt. Während der experimentelle Wert bei  $-14.1 \text{ cal}/(\text{mol K})$ <sup>172</sup> liegt, könnte der Wert im AMBER Kraftfeld für das TIP3P Wassermolekülmodell noch geringfügig höher liegen und sich somit das Vorzeichen der Entropiewerte umkehren.

Auch das Programm WaterMap der Firma Schrödinger berechnet enthalpische und entropische Beiträge von Wassermolekülen nach dem gleichen solvationstheoretischen Modell jedoch unter Verwendung eines geringfügig anderen Referenzzustands. Bei Abel *et al.*<sup>76</sup> liegen die gesamten Wechselwirkungsenergien für verschiedene Positionen bei ca.  $-20 \text{ kcal/mol}$ , der entsprechende entropische Beitrag bei ca.  $-2.5$  bis  $-13 \text{ cal}/(\text{mol K})$ . Beuming und Kollegen<sup>77</sup> hingegen bestimmten Entropien im Bereich von  $-93 \text{ cal}/(\text{mol K})$ , während enthalpische Beiträge einzelner Wassermoleküle nur bei ungefähr  $-7 \text{ kcal/mol}$  liegen. Robinson schließlich klassifiziert in einer erst kürzlich erschienenen Veröffentlichung<sup>84</sup> nur zwischen „high energy hydration sites“ ( $\Delta G > 2.5 \text{ kcal/mol}$ ) und anderen Wasserpositionen. Da in WaterMap ein anderer Referenzzustand verwendet wurde, sind die Ergebnisse nicht unmittelbar miteinander vergleichbar. Zudem wurden die jeweiligen Trajektorien ohne Liganden generiert. Der Einfluss des  $\text{K}^+$ -Ions wird sehr wahrscheinlich die Ergebnisse verändern. Da jedoch keine Trajektorie einer ligandfreien Bindetasche vorliegt, die stabile Wasserpositionen zeigt, kann diese Frage nicht abschließend beantwortet werden.





**Abb. 7.7.:** Dargestellt sind die berechneten Besetzungsgrade einzelner Regionen rund um ein geordnetes Wassermolekül in Abhängigkeit der Kugelkoordinaten ( $\phi$  und  $\theta$  im Bogenmaß) für das System 1CPE. Die eingezeichneten roten Linien kennzeichnen ungefähr die Regionen, die für die Entropieberechnungen ausgewählt wurden. Die jeweiligen Bereiche wurden für unterschiedliche Abstandsintervalle definiert, die hier nicht vermerkt sind.



**Tab. 7.6.:** Thermodynamische Beiträge von einzelnen geordneten Wassermolekülen berechnet aus der MD-Simulation mittels des solvatationstheoretischen Ansatzes nach Lazaridis. Energiewertangaben sind in kcal/mol. Entropien haben die Einheit cal/(mol·K)

System	1AEO		1AEF			1AEG		2AQD		2EUP	
	A 306	C 307	A 304	B 305	C 307	A 307	C 306	A 305	C 306	A 307	C 306
$E_{SW}$	$-20.1 \pm 2.1$	$-20.3 \pm 1.9$	$-20.2 \pm 3.1$	$-18.2 \pm 2.2$	$-18.7 \pm 1.9$	$-21.6 \pm 2.4$	$-20.3 \pm 1.9$	$-18.6 \pm 2.7$	$-20.4 \pm 1.6$	$-20.1 \pm 2.1$	$-19.9 \pm 1.7$
$E_{WW}$	$-0.8 \pm 0.7$	$-0.2 \pm 0.6$	$-0.1 \pm 2.0$	$-0.8 \pm 2.8$	$-1.0 \pm 0.6$	$-1.7 \pm 2.1$	$-0.2 \pm 0.5$	$-1.1 \pm 2.0$	$-0.3 \pm 0.5$	$-0.8 \pm 1.0$	$-0.9 \pm 0.4$
$S_{SW}^{ctr}$	-3.3	-4.6	-3.7	-5.0	-5.3	-3.5	-5.0	-4.2	-5.3	-4.6	-4.9
$S_{SW}^{cor}$	-6.2	-6.0	-4.8	-6.4	-6.5	-4.9	-8.0	-5.2	-7.0	-6.2	-6.8
$S_{SW}$	-9.5	-10.6	-8.6	-11.4	-11.9	-8.4	-13.0	-9.4	-13.3	-10.9	-11.7
$S_{WW}^{ctr}$	0.0	0.0	-0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
$S_{WW}^{cor}$	0.0	0.0	-0.0	0.0	0.0	0.0	0.0	0.0	0.0	-0.1	0.0
$S_{WW}$	0.0	0.0	-0.1	0.0	0.0	0.0	0.0	0.0	0.0	-0.1	0.0
$\Delta S_{WW}$	15.2	15.2	15.1	15.2	15.2	15.2	15.2	15.2	15.2	15.1	15.2
$\Delta E_{solv}$	-11.0	-10.9	-10.8	-9.1	-9.7	-12.9	-11.0	-9.7	-11.0	-11.0	-10.9
$\Delta S_{solv}$	5.7	4.6	6.5	3.8	3.3	6.8	2.2	5.8	1.9	4.2	3.5
$\Delta G_{solv}$	-12.8	-12.2	-12.7	-10.2	-10.7	-15.0	-11.6	-11.4	-11.6	-12.3	-11.9

**Tab. 7.7.:** Thermodynamische Beiträge von einzelnen geordneten Wassermolekülen berechnet aus der MD-Simulation mittels des solvationstheoretischen Ansatzes nach Lazaridis für das System 1CPE. Energiewertangaben sind in kcal/mol. Entropien haben die Einheit cal/(mol · K)

Wasser	305 (B)	307 (C)	312	1954	3077 (A)	3694	4747
$E_{\text{SW}}$	$-28.9 \pm 3.0$	$-21.1 \pm 1.8$	$-22.0 \pm 3.0$	$-19.0 \pm 3.3$	$-22.9 \pm 3.4$	$-23.8 \pm 3.2$	$-16.0 \pm 3.3$
$E_{\text{WW}}$	$5.6 \pm 2.1$	$-0.2 \pm 0.8$	$4.4 \pm 3.0$	$-0.8 \pm 2.7$	$4.4 \pm 3.8$	$3.5 \pm 4.2$	$-4.0 \pm 2.8$
$S_{\text{SW}}^{\text{tr}}$	-5.0	-6.6	-6.9	-4.8	-6.4	-5.6	-4.6
$S_{\text{SW}}^{\text{or}}$	-8.3	-6.8	-7.0	-7.4	-6.1	-5.2	-5.5
$S_{\text{SW}}$	-13.3	-13.3	-13.9	-12.2	-12.5	-10.8	-10.1
$S_{\text{WW}}^{\text{tr}}$	-0.1	0.0	0.0	-0.4	-0.1	-0.1	-0.2
$S_{\text{WW}}^{\text{or}}$	0.0	0.0	0.0	-0.3	-0.2	0.0	-0.1
$S_{\text{WW}}$	-0.1	0.0	0.0	-0.7	-0.3	-0.1	-0.3
$\Delta S_{\text{WW}}$	15.1	15.2	15.2	14.5	14.9	15.1	14.9
$\Delta E_{\text{solv}}$	-16.6	-11.7	-10.3	-9.9	-11.2	-12.6	-8.5
$\Delta S_{\text{solv}}$	1.8	1.9	1.3	2.3	2.4	4.3	4.8
$\Delta G_{\text{solv}}$	-17.1	-12.3	-10.7	-10.6	-11.9	-13.8	-10.0

## 7.3. Zusammenfassung, Konsequenzen, Ausblick

### 7.3.1. Thermodynamische Berechnungen

Die Berechnung enthalpischer und entropischer Beiträge sowie freier Enthalpieunterschiede einzelner Wassermoleküle stellte sich im Rückblick als deutlich aufwändiger heraus als ursprünglich erwartet. Die Schwierigkeiten und Nachteile beim Durchführen von TI-Rechnungen für einzelne Bindetaschenwassermoleküle wurden oben bereits ausführlich dargestellt. Die nach anfänglichen Herausforderungen inzwischen etablierte, solvationstheoretische Methode nach Lazaridis hingegen zeigt gute Resultate. Neben der Tatsache, dass dieses Vorgehen einfach auf existierende Trajektorien zu übertragen ist, lassen sich so getrennt entropische wie auch enthalpische Beiträge analysieren. Wie bei den TI-Rechnungen können hier jedoch auch nur einzelne Wassermoleküle untersucht werden. Es ist aufgrund der jetzigen Implementierung der Methode nicht möglich, die Berechnung auf mehrere Wassermoleküle, die abwechselnd eine bestimmte Position besetzen, anzuwenden.

Aufgrund der ligandfreien Referenzbindetasche, die ein Kation enthält, konnten Rückschlüsse auf die bessere oder schlechtere Verdrängbarkeit einzelner Wassermoleküle nur unter Vorbehalt getroffen werden. Eine bis auf Wassermoleküle, vollständig leere Bindetasche scheint aber im geschlossenen Zustand nicht zu existieren (siehe Ergebnisse der MD-Simulationen). Auf diese Weise fehlt eine geeignete Referenz zum Vergleich. Vor diesem Hintergrund stellt sich die Frage nach der Eignung des Modellsystems. Auch wenn durch die Untersuchungen an dieser Tasche viele Methoden und Programme zur Behandlung von Wasser erprobt werden konnten und mögliche Schwierigkeiten und Nachteile aufgedeckt wurden, so ist diese Modellbindetasche unter Umständen nicht optimal, um die Vorhersage der Verdrängung von Wassermolekülen bei der Ligandbindung zu optimieren. Andere Modellbindetaschen, z.B. die zweier Mutanten des Proteins T4 Lysozym, sind jedoch zu hydrophob, um Einflüsse von Wassermolekülen untersuchen zu können.

Die Vielzahl an Beispielen aus der Literatur zeigt, dass die Anwendung der solvationstheoretischen Untersuchungen bereit ist für den Einsatz im computer-basierten Wirkstoffdesign. In den aufgeführten Literaturstudien zeigte der Ansatz von Lazaridis implementiert im Programm WaterMap stets sehr gute Resultate, die in mehreren Fällen die Erklärung für bis dahin beobachtete Affinitätsunterschiede von Inhibitoren liefern konnte. Die Anwendung dieser Methode auf ein noch breiteres Feld von Zielproteinen, in denen Fragestellungen bezüglich Wasser aufgeworfen worden waren, kann demnach erfolgen.

### 7.3.2. Wassermoleküle in der untersuchten Modellbindetasche

Im Anschluss an die Diskussion der jeweiligen Ergebnisse wurden an verschiedenen Stellen dieser Arbeit bereits die Vorteile und Nachteile der einzelnen, eingesetzten Methoden diskutiert. Die jeweiligen Schlussfolgerungen ermöglichen es, dass der Leser das für seine Fragestellung am besten geeignete Verfahren auswählen kann. Speziell die Analyse der Wasserdynamik, wie sie z.B. aus MD-Simulationen erhalten wird, erwies sich als nicht trivial. Die Berechnung von Hydratationsdichteverteilungen hat sich dabei aber als ein wertvolles

Hilfsmittel herausgestellt. Nun soll aber abschließend noch ein Blick auf die Ergebnisse der Untersuchungen der Dynamik und Thermodynamik der Wassermoleküle, die in der CCP W191G Bindetasche zu finden sind, geworfen werden.

Die zu Beginn dieser Arbeit über das Modellsystem CCP W191G verfügbare Literatur legte den Schluss nahe, dass die artifiziell erzeugte Bindetasche sich durch ihre Abgeschlossenheit und Isolierung vom das Protein umgebenden Lösungsmittel auszeichnet. Trotzdem konnte in den durchgeführten MD Simulationen erstaunlicherweise ein hoher Austausch von Wassermolekülen auf einzelnen Positionen beobachtet werden. Wie die Simulationen mit eingeschränkter Beweglichkeit der der Bindetasche nahen Schleifenregion gezeigt haben, ist die Flexibilität dieses Loops für den Austausch verantwortlich. Wäre dies im Vorfeld bekannt gewesen, hätte das unter Umständen die Auswahl des Modellsystems beeinflusst.

Die Auswertung der durchgeführten MD-Simulationen hat neben der Offenbarung von Ungenauigkeiten in der 1AEF Struktur gezeigt, dass es zumindest in einem System 1AEO Hinweise auf eine zusätzliche, zumindest temporär besetzte, Wasserposition in der Bindetasche gibt, die durch Röntgenstrukturanalyse nicht detektiert werden konnte. Dies befindet sich ebenfalls in Übereinstimmung mit einigen Ergebnisse der Docking-Simulationen, die die Anwesenheit von Wassermolekül 976 der 1CPE Struktur im Komplex prognostizierten. Desweiteren ergab die Analyse der Trajektorien eine hohe strukturelle Konservierung der Bindetasche, des Ligand-Bindungsmodus (mit Ausnahme von 1AEF) wie auch der Wasserpositionen in der Tasche. Die aus Experimenten bekannten Strukturinformationen wurden also durch die Simulationen untermauert.

Aufgrund des Umstandes, dass es vermutlich keine strukturellen Daten aus Experimenten mit einer vollständig ligandfreien Bindetasche der CCP W191G gibt, in der auch kein Ion vorhanden ist, fand sich zunächst keine geeignete Basis für einen direkten Vergleich mit der Wasserdynamik in den Komplexen. Auch die Konstruktion eines vollständig ligand- und ionenfreien Modells dieser Tasche erwies sich als schwierig. Man kann davon ausgehen, dass eine solche Tasche mit geschlossener Loop-Konformation nicht vorliegt. Die Dynamik der Wassermoleküle in der 1CPE Tasche zeigte erneut stabile Wasserpositionen aber einen häufigen Austausch der einzelnen Solvensmoleküle zwischen den konservierten Aufenthaltsbereichen. Außerdem wurde das Wassernetzwerk durch die Bewegung des Kations in der Simulation ebenfalls verändert.

Die durch den solvationstheoretischen Ansatz erhaltenen thermodynamischen Profile einzelner Wassermoleküle decken sich mit den Erwartungen und Erkenntnissen, die aus den vorangegangenen Untersuchungen vorhanden waren. Die Stabilität einzelner Wasserpositionen und die Wechselwirkungen einzelner Wassermoleküle mit dem Liganden spiegeln sich in den beobachteten Trends wider und ergeben ein insgesamt konsistentes Bild. So zeigen beispielsweise die Systeme 1AEF und 1AEG eine energetisch besonders niedrige Wasserposition A. In diesen wird, im Gegensatz zu System 2AQD, auch eine direkte H-Brückenbindung zum Liganden von dieser Wasserposition ausgebildet. Zudem zeigt sich das in allen Systemen beobachtete Wassermolekül A als thermodynamisch stabiler als Wasser B. Der Vergleich mit der ligandfreien 1CPE Tasche erwies sich auch in diesem Fall als herausfordernd, da die häufigen Austauschvorgänge zwischen den Wasserpositionen nicht vom Programm verfolgt werden können. Es ist nicht auszuschließen, dass die

deshalb eingesetzte Trajektorie kein vollständiges Sampling ermöglicht und so die Ergebnisse verändert hat. Dennoch ergab sich hier ebenfalls ein weitgehend konsistentes Bild. Drei von vier Wassermolekülen, die in keinem Komplex zu finden sind, fanden sich auch auf einer im Vergleich thermodynamisch instabileren Position ( $\approx -10$  kcal/mol). Eines befindet sich relativ gesehen in einer thermodynamisch stabileren Position (-13.8 kcal/mol), wird aber dennoch in den Komplexen verdrängt. Hierbei gilt es den Einfluss des  $K^+$ -Ions zu beachten, das eine Stabilisierung im Vergleich zu einer völlig ligandfreien Tasche bewirken kann. Die herausragende Lage von Wasser B zeigte sich auch durch diese Analyse. Die hohe energetische Stabilisierung von  $-17.1$  kcal/mol in der ligand-freien Struktur war so nicht erwartet und kann auch durch das direkt benachbarte Kation verursacht werden. Dieser große Unterschied zwischen ligand-freiem System und Komplex gibt einen Hinweis auf die Herausforderungen, denen ein Docking-Algorithmus, der auch die Wasserverdrängung vorhersagen soll, ausgesetzt ist. Ebenso unerwartet war die Tatsache, dass die durchweg konservierte Position C ebenfalls thermodynamisch nicht zu den stabilsten gehört.



## **Teil III.**

### **Zusammenfassung, Summary**





## 8. Zusammenfassung

Wasser ist als natürliches Solvens bei biologischen Prozessen allgegenwärtig. Aus diesem Grund ist es auch bei vielen Vorgängen auf atomarer Größenebene von entscheidender Bedeutung wie beispielsweise in seiner Rolle als Lösungsmittel für Biomakromoleküle oder bei der Beteiligung von Wassermolekülen an enzymatischen Reaktionen. Im Rahmen dieser Arbeit wurden deshalb die Einflüsse von Wassermolekülen bei der Bindung von Liganden an Proteine mittels computerbasierter Studien untersucht. Gerade in diesem Kontext wird der Einfluss von Wasser nämlich häufig unterschätzt oder vernachlässigt. Im Bereich des computerbasierten Wirkstoffdesigns sind verschiedene Aspekte von Wassermolekülen von Bedeutung. Beispielsweise ermöglicht es die Eigenschaft, Wasserstoffbrückenbindungsakzeptor und -donor zugleich zu sein, Strukturelemente eines Proteins zu stabilisieren, indem polare Wechselwirkungen verbrückt werden. Es ist außerdem lange bekannt, dass strukturelle Wassermoleküle eine Schlüsselrolle in Protein-Ligand Wechselwirkungen spielen können, unter anderem durch Verbrückung von Wechselwirkungen mit dem Rezeptor. Die *in silico* Behandlung dieses Aspektes ist aber mit Schwierigkeiten verbunden, insbesondere weil präzise energetische und entropische Informationen schwierig zu erhalten sind und strukturelle Informationen Unzulänglichkeiten enthalten können. Ein bedeutendes Beispiel dieser Schwierigkeiten stellt die exakte Vorhersage der Protein-Ligand Bindungsaffinität dar. Einer der größten Beiträge zur Freien Bindungsenthalpie stammt aus der Desolvatation der Bindetasche und demnach ist es eigentlich unerlässlich, vorhersagen zu können, welche Wassermoleküle aus einer Bindetasche bei der Ligandbindung verdrängt werden und welche darin verbleiben zusammen mit den kooperierenden oder sich kompensierenden enthalpischen und entropischen Beiträgen. In aktuellen computerbasierten Ansätzen wird die Verdrängung von Wassermolekülen bei der Ligandbindung in einer Proteinbindetasche jedoch kaum erfasst.

Gegenstand der vorliegenden Arbeit ist die Anwendung verschiedener Computermethoden, um die Effekte von Wassermolekülen in Protein-Ligand Komplexen zu untersuchen. Hierbei stehen nicht nur das untersuchte Proteinsystem selbst, sondern auch die verfügbaren Methoden mit den jeweiligen Vorteilen und Limitationen im Zentrum des Interesses. Langfristiges Ziel war es dabei die gewonnenen Erkenntnisse zu nutzen, um später, z.B. in einem virtuellem Screening, aufkommende Fragen über die Einflüsse der Wassermoleküle gezielt mit dem richtigen Verfahren untersuchen und beantworten zu können. Im Folgenden sollen die wichtigsten Vorgehensweisen und Ergebnisse noch einmal im Überblick dargestellt werden.

Die Untersuchungen der vorliegenden Arbeit wurden an einer Modellbindetasche durchgeführt. Dieses System besteht aus Komplexen der Cytochrom *c* Peroxidase (CCP) W191G Mutante und einem Satz strukturell ähnlicher, kleiner, rigider Liganden mit deutlich un-

terschiedlichem Wechselwirkungsmuster in Hinblick auf die Kristallwassermoleküle. Die speziell entworfene Tasche ist polar, solvatisiert, tief vergraben und experimentell sehr gut charakterisiert. Eine flexible Schleifenregion neben der Bindetasche kann eine Konformationsänderung eingehen und für Ligandmoleküle einen Eingang zur Tasche öffnen. Verschiedene andere Aspekte dieses Modellsystems wurden bereits in anderen Studien untersucht und in der Literatur beschrieben.

Für die vorliegende Arbeit wurden die PDB-Komplexe 1AEO, 1AEF, 1AEG, 2AQD, 2EUP mit den Liganden 2-Aminopyridin, 3-Aminopyridin, 4-Aminopyridin, 2,5-Diaminopyridin, und 2-Amino-5-picolin ausgewählt. Zusätzlich wurde die ligandfreie Form 1CPE mit einem  $K^+$ -Ion in der Bindetasche untersucht. Der umfassende strukturelle Vergleich der Röntgenkristallstrukturen zeigte nur sehr geringe Unterschiede. Der Vergleich von 1CPE mit anderen ligandfreien Strukturen der CCP W191G hingegen zeigte ein zusätzliches Wassermolekül auf der Position des Kations. Aufgrund der experimentellen Bedingungen, der Koordinationsgeometrie des vermeintlichen Wassermoleküls, der Ladung der Tasche, bekannter Probleme der Röntgenstrukturaufklärung und der Ergebnisse aus MD Simulationen wurde deutlich, dass das zusätzliche Wassermolekül, das in den anderen ligandfreien Strukturen zu finden ist, höchstwahrscheinlich ein Kation ist.

Die Bindetasche wurde dann mit Hilfe molekularer Wechselwirkungsfelder charakterisiert. Auch diese Ergebnisse untermauerten die Vermutung, dass in der Tasche ein  $K^+$ -Ion vorhanden ist. Ebenso zeigten sich alle Kristallwassermoleküle in energetisch vorteilhaften Bereichen. Es konnte weiterhin kein zusätzlicher Bereich, der für Wassermoleküle vorteilhaft wäre, ausgemacht werden. Diese Beobachtung ist von Relevanz, da Wassermoleküle mit einer erhöhten Beweglichkeit nicht notwendigerweise durch die Röntgenstrukturaufklärung detektiert werden können.

Die Fähigkeiten zweier verbreiteter Dockingprogramme, den Bindungsmodus der Heterozyklen wie auch die Wasserverdrängung während der Ligandbindung korrekt vorherzusagen, wurde sodann an dieser Modellbindetasche untersucht. In einem ersten Schritt wurden alle Verbindungen in einer Dockingsimulation zurück in ihre jeweilige Proteinbindetasche platziert. Dabei wurden die Wechselwirkungen der Wassermoleküle in der Bindetasche vor der Dockingsimulation an- oder ausgeschaltet. Zusätzlich wurden alle Wassermoleküle auch in den so genannten „displaceable“ bzw. „togglig“ Modus gesetzt, in dem die Programme während des Dockings selbstständig die Verdrängung einzelner Wassermoleküle oder das Aufrechterhalten der Wechselwirkungen zu bestimmen versuchen. Insgesamt war die Qualität der erhaltenen Bindemodi sehr gut, dies war jedoch von der gewählten Kombination aus Algorithmus und Scoringfunktion abhängig. Insbesondere der Komplex 1AEF stellte sich als anspruchsvoll für beide Dockingprogramme heraus. In Hinblick auf die Vorhersage der Wasserwechselwirkungen für die fünf Komplexe erreichte man mit FlexX und GOLD zusammen mit Goldscore gute Ergebnisse. In den meisten Fällen wurden die experimentell beobachteten Wassermoleküle als nicht verdrängt vorhergesagt. Im Gegensatz dazu fand GOLD zusammen mit Chemscore die korrekten Bindungsmodi ohne Anwesenheit von Wassermolekülen in der Bindetasche (mit Ausnahme von 2AQD).

Im Rahmen der Docking-Studie wurden die fünf Liganden in die Bindetasche von 1CPE platziert, nachdem daraus zuvor das  $K^+$ -Ion entfernt worden war und alle Wassermoleküle

---

in der Tasche in den „displaceable“ bzw. „togglig“ Modus gesetzt worden waren. Die Auswertung der best-bewerteten Lösungen bestätigte einige der vorherigen Schlussfolgerungen. Beide Dockingprogramme konnten richtige Bindungsmodi der Liganden vorhersagen und zeigten nur für das System 1AEF Ergebnisse, die nicht gut mit der experimentellen Komplexstruktur übereinstimmten. Die Vorhersage der Wasserverdrängung stellte sich in diesem Test als schwieriger dar. Von den fünf Wassermolekülen wurden zwar die experimentell beobachteten behalten, in mehreren Fällen fand man aber auch ein zusätzliches. Die Ergebnisse mit der Chemscore Scoringfunktion zeigten in drei von fünf Fällen den richtigen Bindungsmodus der Liganden aber gleichzeitig erneut keinerlei Wassermoleküle. Der Einsatz dieser Scoringfunktion kann im Zusammenhang mit dem Programm GOLD zumindest für dieses System nicht empfohlen werden, wenn man an der Vorhersage des Wassernetzwerkes in der Bindetasche interessiert ist.

Das ursprünglich vorhandene Ion beeinflusste die Positionen der Wassermoleküle und somit auch indirekt die Dockingergebnisse. Aus diesem Grund wurde mit Hilfe eines kombinierten Protokolls aus Berechnung von molekularen Wechselwirkungsfeldern und Relaxation durch MD-Simulationen versucht, ein ionenfreies Modell der CCP W191G Tasche zu konstruieren. Die erhaltene Struktur zeigte eine hohe strukturelle Ähnlichkeit mit der 1CPE Struktur. Ebenso stimmten die Wasserpositionen im Wesentlichen überein. Zusätzlich konnten aber weitere Wassermoleküle in der Nähe der Binderegion beobachtet werden. Weitere Docking-Simulationen mit diesem Strukturmodell ergaben keine überzeugenden Ergebnisse. Sowohl die Bindungsmodi wie auch die vorhandenen Wassermoleküle unterschieden sich deutlich von den Daten aus dem Experiment. Es wurde überprüft, ob der korrekte Bindungsmodus wenigstens bei vorherigem richtigen Festlegen des Wassernetzwerkes gefunden werden kann. Dieses Vorgehen führte jedoch nur in wenigen Fällen zu einer Verbesserung, in anderen sogar zu einer Verschlechterung.

Da das Optimierungsproblem beim Docking durch die mehrfache Vorhersage der Wasserverdrängung noch vergrößert wurde, wurde versucht, die Resultate durch eine Vorselektion der wahrscheinlich verdrängten Wassermoleküle zu verbessern. Deshalb wurden strukturelle Deskriptoren der Wassermoleküle aller Kristallstrukturen aus der Proteindatenbank Relibase+ entnommen und analysiert. Mit der konstruierten Modellstruktur wurde analog verfahren. Es zeigte sich, dass die Wasserpositionen, die in den fünf Komplexen konserviert sind, mehr Proteinkontakte, einen niedrigeren DrugScore-Wert und eine geringe solvenszugängliche Oberfläche haben als die Positionen, die nicht in allen Komplexen beobachtet werden können. Basierend auf diesen Deskriptoren alleine konnte jedoch kein Trend beobachtet werden, der helfen könnte, die übrigen Wasserpositionen in leicht verdrängbar oder nicht verdrängbar zu klassifizieren. In diese Analyse wurden außerdem die B-Faktoren der Wassermoleküle mit eingeschlossen, um einen begrenzten Einblick in die Dynamik des Systems über das Strukturmodell hinaus zu erhalten.

Um die Dynamik der untersuchten Systeme weiter zu analysieren, wurden MD-Trajektorien berechnet. Die Proteinstruktur und -faltung zeigte sich in allen Simulationen als konserviert. Die Analyse der Wasserdynamik ließ sich am besten mittels gemittelter Hydratationsdichteverteilungskarten durchführen, aus denen die Maxima extrahiert und die Anzahl der verschiedenen Wassermoleküle, die eine bestimmte Position besetzen, ermit-

telt wurden. Interessanterweise zeigte sich in verschiedenen Systemen ein erhöhter Austausch von Wassermolekülen in der Bindetasche mit der umgebenden Solvensphase. Dieser unerwartet häufige Wasseraustausch ist eng verbunden mit der Flexibilität der Schleifenregion in Bindetaschennähe. Zusätzliche MD-Simulationen mit geringen Zusatzkräften auf die  $C_\alpha$ -Atome dieses Loops zeigten keinerlei Wasseraustausch. Die Wasserpositionen in der Bindetasche, die in den Röntgenstrukturen beobachtet werden können, waren in den Simulationen ohne Zusatzkräfte auch trotz des Austausches hochgradig konserviert. In einem Komplex konnte außerdem ein zusätzliches Wassermolekül in der Bindetasche an einer Stelle beobachtet werden, an der in der Röntgenstruktur keines zu finden war. Diese Position ist interessanterweise nahe an der, wo in einigen Dockingsimulationen mit 1CPE ein zusätzliches Wassermolekül vorhergesagt wurde. Das deutet darauf hin, dass die Docking-Ergebnisse in dieser Hinsicht sehr wahrscheinlich richtig sind, wovon bis dahin nicht ausgegangen werden konnte. Die Hydratationsdichte an dieser Stelle ist auf einem niedrigen Niveau, was bedeutet, dass diese Position nur teilweise während der Simulation besetzt ist. Demnach ist es gut möglich, dass ein entsprechendes Wassermolekül in einem Röntgenstrukturexperiment wegen seiner höheren Mobilität nicht detektiert wurde. Der Bindungsmodus von 3-Aminopyridin aus dem Komplex 1AEF war während der simulierten Zeitspanne nicht konserviert. Stattdessen fand bereits früh eine Rotation um das Zentrum des aromatischen Rings statt. Diese Diskrepanz des Bindungsmodus regte die erneute Bestimmung dieser Kristallstruktur bei einer höheren Auflösung und vollständiger Verfeinerung an. Diese neue Struktur zeigte den gleichen Bindungsmodus des Liganden wie zuvor, jedoch eine andere Seitenkettenkonformation für Arg48 nahe am Cofaktor. In der daraus neu berechneten Trajektorie war die dominierende Ligandorientierung nun die der Kristallstruktur. Auch die Re-Dockingergebnisse verbesserten sich deutlich für diese Struktur. Geringe strukturelle Unterschiede der Bindetaschenregion wirkten sich hier also signifikant auf die Dockingergebnisse und Trajektorie aus. Die MD-Simulation des ionenfreien Bindetaschenmodells zeigte größere Veränderungen der Schleifenregion und ein Eindringen von Wasser in die Bindetasche. Eine Beschränkung auf die wenigen konservierten Wassermoleküle während eines Dockings analog zu dem oben beschriebenen zeigte keine Verbesserung der Ergebnisse. Insgesamt untermauert dies die ursprüngliche Vermutung, dass eine vollständige ionenfreie Tasche der CCP W191G nicht existiert.

Der letzte und entscheidende Schritt der Charakterisierung von Wassermolekülen in einer Bindetasche ist die Bestimmung von thermodynamischen Eigenschaften. Aus diesem Grund sollten Freie Energie Rechnungen mittels Computersimulationen durchgeführt werden. In der vorliegenden Arbeit wurden hierfür MD-Simulationen verwendet. Zunächst wurde das Protokoll für die Thermodynamische Integration an den reinen Liganden in Lösung etabliert und validiert. Beim Transfer in die Proteinumgebung traten jedoch Probleme auf. Der Vorgang des Entferns eines einzelnen Wassermoleküls aus der Bindetasche konnte nicht simuliert werden, da das Wassermolekül entweder nicht auf seiner Position festgehalten oder andere Wassermoleküle nicht davon abgehalten werden konnten diese Position zu besetzen. Dieser Ansatz war daher nicht hilfreich, um die Frage nach den thermodynamischen Größen zu beantworten.

Stattdessen wurde ein relativ neues, solvationstheoretisches Vorgehen gewählt, um

---

Enthalpie, Entropie und Freie Enthalpie einzelner Wassermoleküle zu bestimmen. Diese Methode erwies sich als äußerst nützlich, konnte problemlos etabliert werden, kann aber in der vorliegenden Implementierung nicht mehrere Wassermoleküle auf einer Position handhaben. Die zuvor beobachtete Dynamik auf den unterschiedlichen Wasserpositionen in der Bindetasche ebenso wie die Interaktion mit dem Liganden und dem Protein spiegelten sich in den erhaltenen thermodynamischen Daten wider. Die entsprechenden Ergebnisse der ligandfreien Tasche halfen nur in begrenztem Umfang, die Wasserverdrängung während der Ligandbindung zu erklären, da das vorhandene Kation die thermodynamischen Größen mit beeinflusst. Insgesamt wäre es jedoch sehr lohnenswert, diese Verfahren auf mehrere andere Proteinsysteme zu übertragen und die erhaltenen Ergebnisse dann für Entscheidungen in einem strukturbasierten virtuellen Screening zu nutzen.

Zusätzlich zu den Einblicken in das studierte Modellsystem gibt diese Arbeit dem Leser somit einen Überblick über Vorzüge und Nachteile unterschiedlicher Methoden, die bei der Analyse der Effekte von Wassermolekülen auf Protein-Ligand Wechselwirkungen von Interesse sind.



## 9. Summary

Water is the essential medium of any being, it is omnipresent and has a deep impact on life at the atomic level, which is often underestimated or disregarded. It plays a key role as solvent for biomacromolecules and participates in several enzymatic reactions. Particular in the field of computer-aided drug design water molecules are of importance. Their ability to act as an H-bond donor and acceptor simultaneously enables them, e.g., to stabilize structural motifs of a protein by bridging polar interactions. It is also long known that structural water molecules can play a key-role in protein-ligand interactions. In the present work the influences of water molecules on protein-ligand binding are investigated. The process of structure-based drug design still faces many difficulties in this context, in particular since the precise energetic and entropic contributions are difficult to estimate and the experimental structural information about water molecules in protein-ligand complexes are prone to ambiguities. One important example is the exact prediction of the protein-ligand binding affinity. One of the major contributions to the overall binding free energy arises from desolvation of the binding site. Hence, being able to predict which water molecules are displaced from a cavity upon ligand binding as well as their compensating or cooperating entropy and enthalpy terms is of fundamental importance. However, these effects are difficult to capture by current computational approaches.

The objective of this thesis was the application of various computer methods to study the effects of water molecules in a protein-ligand complex. Not only the protein system of this study itself but also the available methods and their advantages and limitations were of interest here. Expertise gained during this study could later on help to choose the right method to address questions, e.g. during a virtual screening campaign. Below the most important procedures and the obtained results are summarized.

In this work, a model system was used, consisting of complexes between the cytochrome *c* peroxidase (CCP) W191G mutant and a set of structurally similar, small and rigid ligands with significantly different interaction patterns to crystal water molecules. The engineered cavity is known to be polar, solvated, deeply buried and experimentally well characterized. A flexible loop next to the binding site can undergo conformational changes and open a gate for ligands to enter the binding site. Various aspects of this model system have been subject of earlier studies described in the literature.

The PDB complexes 1AEO, 1AEF, 1AEG, 2AQD, 2EUP with the ligands 2-aminopyridine, 3-aminopyridine, 4-aminopyridine, 2,5-diaminopyridine and 2-amino-5-picoline, respectively, have been chosen. Also the ligand-free form 1CPE with a potassium ion in the binding site was investigated. The overall comparison of the X-ray crystal structures revealed only minor differences. But the comparison among 1CPE and other ligand-free crystal structures of CCP W191G showed an additional water molecule at the position of the cation. Due

to the experimental conditions, the coordination geometry of the putative water molecule, the charged nature of the pocket, known issues in X-ray crystallography, and results from the MD simulations, it seemed most likely, that the additional water molecule, observed in the other ligand-free structures, is a cation.

The binding site was then analysed by calculating molecular interaction fields. The results supported the position of the potassium ion. Also all crystal water molecules were found in energetically favorable regions. Furthermore, unoccupied favorable regions were not found, which is an interesting aspect since highly mobile water molecules are not necessarily detected by X-ray crystallography.

The ability of two common docking tools to correctly predict water displacement upon ligand binding as well as the binding modes of the small heterocyclic compounds was then tested in this model environment. In the first step all compounds were re-docked into their respective protein crystal structure. In this part, interactions of water molecules in the binding pocket were switched on or off prior to the docking runs. Water molecules were also set in the displacable / toggling mode in which the decision about formation of interactions or displacement is made by the docking program. Overall, the quality of the predicted binding modes turned out to be fairly good. But, predicting the correct binding mode for complex 1AEF seemed challenging for the two docking engines. With respect to the prediction of the water interaction in the pocket, FlexX and GOLD/Goldscore performed very well. In most of the cases the experimentally observed water molecules were correctly predicted. In contrast, GOLD together with Chemscore found the correct binding modes without any water molecule in the pocket (with the exception of 2AQD).

In the second step, the study was continued by docking the five ligands in the 1CPE binding site after removing the potassium cation and setting all binding-site water molecules in the displacable / toggling mode. As far as the binding mode of the top scored poses is concerned, previously drawn conclusion could be confirmed. Both docking programs were able to find very good poses and only in the case of 1AEF deviating results were obtained. The prediction of the water displacement turned out to be more complex in this case. Out of the five initially present water molecules the experimentally observed ones were kept during docking, but in several cases also an additional one. Again the results obtained with GOLD and the Chemscore scoring function showed accurate binding modes in three out of five investigated cases but did not show any water molecule in the final results. Thus, using this scoring function together with GOLD cannot be recommended when one is interested in the prediction of the water network in the binding site for this system.

The originally present ion will most likely influence the positions of the water molecules in the binding site and, thus, also affect the docking results. Therefore, it was tried to construct an ion-free model of this pocket by using a combined protocol of calculating molecular interaction fields and relaxation by means of MD simulation. The obtained structural model showed a very high similarity with the crystal structure 1CPE. Also the water positions are comparable. Nevertheless, additional water molecules were observed close to the binding site. Further docking simulations with this structural model did not show convincing results. In several cases as well the obtained binding modes as the water prediction of the docking programs were significantly different from the experimental results. It was



---

tested whether at least correct binding modes could be obtained if the binding-site water molecules were set according to the experimental structures prior to the docking. But this experiment only resulted in improved accuracy in some cases while other results got worse.

Since the optimization problem increased significantly when trying to predict multiple water displacement or conservation during docking, it was tried to improve the results by preselecting water molecules which will probably be displaced by a ligand. For this purpose structural descriptors for the water molecules of all examined crystal structures were taken from the protein database Relibase+ and analyzed accordingly. The ion-free structural model was processed analogously. It turned out that the water positions conserved in all five complexes showed more protein contacts, lower DrugScore values as well as a lower solvens-accessible surface areas than the rest. But, based on these descriptors alone, a clear estimate whether displacement or retention is more likely to occur can hardly be given for the remaining water molecules. In this analysis also the temperature factors of the water molecules were considered in order to take into account also the dynamic information beyond the single structural model of the respective system.

In order to elucidate the dynamic behavior of these systems, MD trajectories were calculated for all of them. The protein structure and the fold was conserved in all of the investigated simulations. For analysing the water dynamics, calculations of average hydration density maps combined with extraction of the maxima as well as evaluating how many different water molecules occupy a single position turned out to be a good approach. Interestingly, in several systems a high number of exchange events between cavity and bulk was found. This unexpected frequent water exchange was closely connected to the flexible loop region close to the cavity. In an additional series of MD simulations small restraining forces on the  $C_{\alpha}$  atoms of this loop removed the water exchange. However, the positions in the binding site observed in the crystal structures were highly conserved also in the simulations without additional restraints despite the water exchange. In one complex also additional water density was found in the binding site, where no water molecule was observed in the X-ray structure. Interestingly, this position is very close to the water molecule which was retained in several of the docking simulations with 1CPE. This indicates that the observed docking results might be more correct than originally thought. The hydration density for this position was low, meaning that this position was not fully occupied during the simulation. Thus, an X-ray experiment might not detect that structural water molecule due to its high mobility. The ligand binding mode in 1AEF was not conserved during the simulation, instead an early in plane rotation around the centre of the aromatic ring was observed. The discrepancy in the binding mode triggered the re-determination of the same (fully refined) structure at higher resolution. This new structure showed the same ligand binding mode but a different side chain conformation for Arg48 close to the cofactor. In the newly calculated trajectory the dominating ligand orientation was then the native one. Also the re-docking results improved significantly by using this structure. Thus, small structural differences of the binding site were found to have a significant impact on docking results and dynamics. Finally, the MD simulation of the constructed ion-free pocket model showed larger movements of the loop region and a massive water influx into the binding site. Using only the few conserved water molecules of this simulation in a docking run similar to the

ones described above did not show an improvement of the results. Overall this supports the idea that a complete ion-free model of the CCP W191G site does not exist.

The last and essential step of characterizing water molecules in a binding site is the estimation of their thermodynamic properties. Thus, free energies should be calculated by means of computer simulation. In the present work MD simulations were used. In the first part a thermodynamic integration protocol was established and validated on the pure ligands in solvent environment. When transferring this in the second part to the protein environment unexpected problems arose. The process of removing a single water molecule from the binding site could not be simulated, since the water molecule could not be successfully restrained at its initial position or other water molecules could not be prevented from entering this area. Thus, this approach turned out to be not the right one for this kind of question.

Instead, a relatively new approach based on solvation theory was used to map enthalpy, entropy and free energy of single water molecules. This method turned out to be of high benefit but could not handle exchange events at a water position in the current implementation. The previously observed dynamics at the different water positions in the cavity as well as the interactions with the ligand and protein environment were reasonably well reflected in the thermodynamic data. The accordant data for the ligand-free pocket helped to explain the observed water displacement upon binding of the different ligands only to a smaller extent, because the cation present affected the results. However, it would be worthwhile applying this method to more protein systems and use the obtained results for decisions in a structure based virtual screening.

In addition to the insights on the studied model binding site, this work illustrates the benefits and shortcomings of various methods that are of interest, when analyzing water effects on protein-ligand interactions.

**Teil IV.**

**Anhang**



# Anhang A.

## Verwendete Hard- und Software

### A.1. Verwendete Hardware

Die Berechnungen dieser Arbeit wurden auf verschiedenen Computern durchgeführt. Dies sind im einzelnen:

- Ein Linux Arbeitsplatzrechner mit Intel Pentium 4 3.2 GHz Prozessor und 3 GB RAM
- Eine Linux Workstation mit zwei Intel Xeon QuadCore 3.0 GHz Prozessoren und 8 GB RAM
- Ein Linux Cluster bestehend aus 15 Rechnern mit AMD Opteron DualCore 2.2 GHz Prozessor, 4 GB RAM sowie einem Rechner mit AMD Opteron QuadCore 2.1 GHz Prozessor, 8 GB RAM und Infiniband Interprozessor Vernetzung
- Ein Linux Cluster aus Rechnern mit Itanium2 1.3 GHz Prozessoren und Myrinet Interprozessor Vernetzung am Leibniz Rechenzentrum (München).

### A.2. Verwendete Software

Sämtliche Molekulardynamik Simulationen wurden mit dem Programmpaket Amber in den Versionen 9 und 10 durchgeführt. Für die Rechner der Arbeitsgruppe wurde der Quellcode dieses Paketes wiederum mit Hilfe der GNU Compiler in ausführbare Programme übersetzt. Für die Parallelisierung auf mehrere Prozessoren wurden, je nach verwendeter Hardware-Architektur, die Open MPI und die MVAPICH2 Bibliothek verwendet.

Quantenchemische Rechnungen wurden mit dem Programm Gaussian in der Version 03, Revision D.01 durchgeführt.

Die Berechnung der molekularen Wechselwirkungsfelder erfolgte mit dem Programm GRID in der Version 22a.

Die Protonierung und Editierung der Liganden erfolgte entweder mit dem Programm MOE in den Versionen 2007.09, 2008.10 und 2009.10 oder dem Programm Sybyl in Version 8.0 bzw. als SYBYL-X.

Docking Experimente wurden mit den Programmen GOLD in der Version 4.0 und FlexX in der Version 3.1.4 durchgeführt.

Kristallstrukturdaten wurden der PDB entnommen. Strukturelle Analyse von Kristallstrukturen wurden zum Teil mit der Datenbank Relibase+ in der Version 3.0.0 durchgeführt.

Die Visualisierung der Molekülmodelle und Dynamik erfolgte mit den Programmen PyMOL (Version 0.99) und VMD (Version 1.8.6). Mit dieser Software wurden ebenso die entsprechenden Abbildungen erstellt.

Diagramme und Graphen wurden mit den Programmen Grace 5.1.21 und Gnuplot 4.2 erzeugt.

# Anhang B.

## Tabelle mit ergänzenden Informationen

**Tab. B.1.:** Übersicht der Sonden, die in der aktuellen Version des Programms GRID zur Verfügung stehen.<sup>30</sup>

Sonde	Beschreibung
OH2	Wasser
DRY	hydrophob
H	neutrales Wasserstoffatom
C3	Methylgruppe
C1=	sp <sup>2</sup> CH (aromatisch oder vinyl)
N3+	sp <sup>3</sup> Amin NH <sub>3</sub> (kationisch)
N2+	sp <sup>2</sup> Amin NH <sub>2</sub> (kationisch)
N2:	sp <sup>3</sup> NH <sub>2</sub> mit freiem Elektronenpaar
N2=	sp <sup>2</sup> Amin NH <sub>2</sub> (kationisch)
N2	neutrales, planares NH <sub>2</sub>
N1+	sp <sup>3</sup> Amin NH (kationisch)
N1:	sp <sup>3</sup> NH mit freiem Elektronenpaar
N1=	sp <sup>2</sup> Amin NH (kationisch)
N1	neutrales, planares NH
NH=	sp <sup>2</sup> NH mit freiem Elektronenpaar
N1#	sp NH mit einem H
N:	sp <sup>3</sup> N mit freiem Elektronenpaar
N:=	sp <sup>2</sup> N mit freiem Elektronenpaar
N:#	sp N mit freiem Elektronenpaar
N::	sp <sup>2</sup> N mit zwei freien Elektronenpaaren und einer Doppelbindung
N-	anionisches Tetrazol N
NM3	Trimethyl-ammonium Kation
N+	sp <sup>3</sup> cationischer N
O	sp <sup>2</sup> Carbonyl O
O::	sp <sup>2</sup> Carboxyl O
O-	sp <sup>2</sup> Phenolat O
O=	Phosphat O

Anhang B. Tabelle mit ergänzenden Informationen

Sonde	Beschreibung
OH	Phenol oder Carboxyl OH
O1	Alkyl OH Gruppe
OC2	Ether O
OC1	aromatischer/aliphatischer Ether O
OC=	aromatischer Ether O
OES	sp <sup>3</sup> Ester O
ON	O in Nitrogruppe
OS	O in Sulfoxid/Sulfinat
O=S	O in Sulfon
OFU	Furan O
PO4	Phosphatdianion
PO4H	Phosphatanion
S1	neutrale SH Gruppe
SO	pyramidaler S
STH	Thiophen oder Thioether S
F	organisches F
F3	CF <sub>2</sub> /CF <sub>3</sub> Gruppen F
FS	F gebunden an S
F-	Fluoridanion
CL	organisches Cl
CL-	Chloridanion
BR	organisches Br
BR-	Bromidanion
I	organisches I
I-	Iodidanion
LI+	Lithium Kation
NA+	Natrium Kation
K+	Kalium Kation
RB+	Rubidium Kation
CS+	Caesium Kation
MG+2	Magnesium Kation
CA+2	Calcium Kation
ZN+2	Zink Kation
CU+2	Kupfer Kation
FE+2	Eisen(II) Kation
FE+3	Eisen(III) Kation
BOTH	amphipatisch
COOH	aliphatisches Carboxylat
AR.COO-	aromatisches Carboxylat
CONH2	aliphatisches Amid



---

Sonde	Beschreibung
AR.CONH2	aromatisches Amid
CONHR	aliphatisches cis/trans Amid
AR.CONHR	aromatisches cis/trans Amid
AMIDINE	aliphatisches Amidin
AR.AMIDINE	aromatisches Amidin
M-DIAMINE	m-Diamino-Benzol
C:#	sp C mit 1 freien Elektronenpaar und 1 Dreifachbindung
N0:	sp <sup>3</sup> N mit 1 freien Elektronenpaar, ohne H
O-1	anionischer sp <sup>3</sup> O
O#	O in Cyanat

---



# Tabellenverzeichnis

1.1. Kristallstrukturen der CCP W191G Komplexe . . . . .	15
2.1. Wassermodelle . . . . .	29
5.1. RMSD der Kristallstrukturen . . . . .	71
5.2. RMSD der Bindetaschen der Kristallstrukturen . . . . .	72
5.3. RMSD des Loops der Kristallstrukturen . . . . .	72
5.4. Vergleich der ligandfreien CCP W191G Kristallstrukturen . . . . .	72
5.5. Ergebnisse Redocking mit GOLD/GoldScore . . . . .	83
5.6. Ergebnisse Redocking mit GOLD/Chemscore . . . . .	86
5.7. Ergebnisse Redocking mit FlexX . . . . .	89
5.8. Ergebnisse des Dockings in 1CPE Bindetasche . . . . .	97
5.9. Ergebnisse des Dockings in die ionenfreie Modellbindetasche der CCP W19G	99
5.10. Ergebnisse des Dockings in die ionenfreie Modellbindetasche der CCP W19G mit korrekt gesetzten Wassermolekülen . . . . .	101
5.11. Normalisierte B-Faktoren . . . . .	105
5.12. Wasserdeskriptoren aus Relibase+ . . . . .	106
6.1. Kristallwassermoleküle in MD-Simulationen . . . . .	111
6.2. Daten über die durchgeführten MD-Simulationen . . . . .	112
6.3. RMSD-Werte und Fluktuationen ausgewählter Wassermoleküle . . . . .	121
6.4. Lokale Maxima der Hydratationsdichte . . . . .	124
6.5. Bindetaschenvolumina . . . . .	125
6.6. Lokale Maxima der Hydratationsdichte für MD-Simulationen mit Zusatz- kräften . . . . .	139
6.7. Ergebnisse des Dockings in die ionenfreie Modellbindetasche der CCP W19G mit vorselektierten Wassermolekülen . . . . .	145
7.1. Einzelbeiträge zu $\Delta G$ für die Transformation von 2- zu 4-Aminopyridin . .	155
7.2. Einzelbeiträge zu $\Delta G$ für die Transformation von 4- zu 3-Aminopyridin . .	155
7.3. Einzelbeiträge zu $\Delta G$ für die Transformation von 3- zu 2-Aminopyridin . .	156
7.4. Beiträge der Unterregionen zur Solvens-Solvens Entropie . . . . .	168
7.5. Vergleich der Wassernummerierung von 1CPE Trajektorie und Kristallstruktur	169
7.6. Thermodynamische Beiträge von einzelnen geordneten Wassermolekülen in den Komplexen . . . . .	173

7.7. Thermodynamische Beiträge von einzelnen geordneten Wassermolekülen im System 1CPE . . . . .	174
B.1. Sonden in GRID . . . . .	195

# Abbildungsverzeichnis

1.1.	Verdrängung von Wasser aus der Bindetasche von CA. II . . . . .	4
1.2.	Aktives Zentrum der Cytochrom <i>c</i> Peroxidase . . . . .	13
3.1.	Isotherme Titrationskalorimetrie . . . . .	40
3.2.	Thermodynamischer Kreisprozess . . . . .	49
5.1.	Kristallstrukturen der untersuchten CCP W191G Komplexe . . . . .	70
5.2.	RMSD vom Modell der CCP W191G . . . . .	74
5.3.	Modell der CCP W191G mit Hydratationsdichte aus MD Simulation . . . . .	75
5.4.	GRID-Analyse der 1CPE Tasche für K <sup>+</sup> . . . . .	76
5.5.	GRID-Analyse der CCP W191G Bindetasche für die Wasser-Sonde . . . . .	77
5.6.	Docking in Kristallstruktur 1CPE . . . . .	94
6.1.	Auswertung zur MD-Simulation von 1AA4 . . . . .	115
6.2.	RMSD-Diagramme der MD-Simulationen von CCP W191G . . . . .	116
6.3.	Anteil verschiedener Sekundärstrukturelemente während der MD-Simulation . . . . .	117
6.4.	Repräsentative Strukturen der Trajektorien zusammen mit Hydratationsdichtekarten . . . . .	123
6.5.	Besetzungsgrad von Wasserpositionen in Komplexen . . . . .	127
6.6.	Besetzungsgrad von Wasserpositionen in ligandfreien Strukturen . . . . .	128
6.7.	Elektronendichte von PDB-Struktur 1AEF . . . . .	130
6.8.	Neue Kristallstruktur des Komplexes von CCP W191G mit 3-Aminopyridin . . . . .	131
6.9.	Torsionswinkel in Arg48 . . . . .	133
6.10.	Charakteristische Abstände der Ligandbewegung von 3-Aminopyridin . . . . .	133
6.11.	RMSD-Diagramme der MD-Simulationen von CCP W191G mit Zusatzkräften . . . . .	136
6.12.	Anteil verschiedener Sekundärstrukturelemente während der MD-Simulation mit Zusatzkräften . . . . .	137
6.13.	Repräsentative Strukturen der Trajektorien zusammen mit Hydratationsdichtekarten für MD-Simulationen mit Zusatzkräften . . . . .	138
6.14.	RMSD-Werte aus der MD-Simulation des kationenfremen Bindetaschenmodells . . . . .	141
6.15.	Loop-Reorganisation während der MD-Simulation . . . . .	141
6.16.	Hydratationsdichteverteilung für Strukturmodell der CCP W191G . . . . .	142
6.17.	Besetzungsgrad von Wasserpositionen in der Modelltasche . . . . .	144
7.1.	Ergebnisse der TI-Berechnungen für die Verschiebung des aromatischen N-Atoms in 2- und 3-Aminopyridin . . . . .	152

7.2. Ergebnisse der TI-Berechnungen für das Entfernen und Erzeugen einer Aminogruppe . . . . .	153
7.3. Ti-Kreisprozess für die Transformation von 2-,3- und 4-Aminopyridin . . .	154
7.4. Abhängigkeit der Entropie von der Winkelverteilung . . . . .	166
7.5. Wassernummerierung in 1CPE Trajektorie . . . . .	169
7.6. Separation einzelner hydratisierter Unterregionen um ein geordnetes Wassermolekül in Komplexen von CCP W191G . . . . .	171
7.7. Separation einzelner hydratisierter Unterregionen um ein geordnetes Wassermolekül im System 1CPE . . . . .	172

# Abkürzungsverzeichnis

<b>2AP</b>	2-Aminopyridin
<b>3AP</b>	3-Aminopyridin
<b>4AP</b>	4-Aminopyridin
<b>Abb.</b>	Abbildung
<b>AMBER</b>	assisted model building with energy refinement
<b>Anz.</b>	Anzahl
<b>Asp</b>	Aspartat
<b>CA. II</b>	Carboanhydrase II
<b>CCP</b>	Cytochrom <i>c</i> Peroxidase
<b>CSD</b>	Cambridge structural database
<b>DA1</b>	2,5-Diaminopyridin
<b>DNA</b>	Deoxyribonucleic acid bzw. Desoxyribonukleinsäure
<b>ED</b>	Elektronendichte
<b>et al.</b>	et alii
<b>FF</b>	Kraftfeld
<b>GA</b>	Genetischer Algorithmus
<b>Gl.</b>	Gleichung
<b>Gly188</b>	Glycin 188
<b>His</b>	Histidin
<b>HIV</b>	Human immunodeficiency virus
<b>i.A.</b>	im Allgemeinen
<b>I76A</b>	Mutation von Isoleucin 76 zu Alanin
<b>ITC</b>	Isotherme Titrationskalorimetrie
<b>Kap.</b>	Kapitel
<b>L94A</b>	Mutation von Leucin 94 zu Alanin
<b>Leu</b>	Leucin
<b>LG4</b>	2-Amino-5-picolin
<b>MD</b>	Molekulardynamik
<b>MIF</b>	Molecular Interaction Fields
<b>MM</b>	Molekulare Mechanik
<b>M102Q</b>	Mutation von Methionin 102 zu Glutamin
<b>NMR</b>	Nuclear Magnetic Resonance
<b>PBSA</b>	Poisson-Boltzmann surface area
<b>PDB</b>	Protein Datenbank
<b>PGGAAN</b>	Prolin-Glycin-Glycin-Alanin-Alanin-Asparagin

<b>PTTI</b>	Parallel Tempering Thermodynamische Integration
<b>QSAR</b>	Quantitative structure activity relationship bzw. quantitative Struktur-Wirkungs-Beziehung
<b>RDF</b>	Radial Distribution Function bzw. Radiale Verteilungsfunktion
<b>RESP</b>	Restrained electrostatic potential
<b>RETI</b>	Replica Exchange Thermodynamische Integration
<b>RMSD</b>	Root mean square deviation bzw. Quadratwurzel der mittleren quadratischen Abweichung
<b>RNA</b>	Ribonucleic acid bzw. Ribonucleinsäure
<b>SANDER</b>	simulated annealing using NMR-derived energy restraints
<b>SFB</b>	Sonderforschungsbereich
<b>SFC</b>	Scoring function consortium
<b>Tab.</b>	Tabelle
<b>TI</b>	Thermodynamische Integration
<b>UV</b>	ultraviolett
<b>W191G</b>	Mutation von Tryptophan 191 zu Glycin



# Literaturverzeichnis

- [1] Chaplin, M.: *Water structure and science*. <http://www.lsbu.ac.uk/water/>.
- [2] Holm, L. and C. Sander: *Evaluation of protein models by atomic solvation preference*. *J. Mol. Biol.*, 225:93–105, 1992.
- [3] Frank, H. S. and M. W. Evans: *Free Volume and Entropy in Condensed Systems III. Entropy in Binary Liquid Mixtures; Partial Molal Entropy in Dilute Solutions; Structure and Thermodynamics in Aqueous Electrolytes*. *J. Chem. Phys.*, 13:507–531, 1945.
- [4] Nemethy, G. and H. A. Scheraga: *Structure of Water and Hydrophobic Bonding in Proteins. I. A Model for the Thermodynamic Properties of Liquid Water*. *J. Chem. Phys.*, 36:3382–3400, 1962.
- [5] Tanford, C.: *The hydrophobic effect: Formation of micelles and biological membranes*. John Wiley & Sons Inc, 2nd edition, 1980.
- [6] Israelachvili, J. and H. Wennerström: *Role of hydration and water structure in biological and colloidal interactions*. *Nature*, 379:219–225, 1996.
- [7] Chandler, D.: *Interfaces and the driving force of hydrophobic assembly*. *Nature*, 437:640–647, 2005.
- [8] Ladbury, J. E.: *Just add water! The effect of water on the specificity of protein-ligand binding sites and its potential application to drug design*. *Chem. Biol.*, 3:973–980, 1996.
- [9] Klebe, G.: *Wirkstoffdesign*. Spektrum, Heidelberg, 2. Auflage, 2009.
- [10] Steuber, H., M. Zentgraf, C. LaMotta, S. Sartini, A. Heine, and G. Klebe: *Evidence for a Novel Binding Site Conformer of Aldose Reductase in Ligand-Bound State*. *J. Mol. Biol.*, 369:186–197, 2007.
- [11] Lam, P. Y. S., P. K. Jadhav, C. J. Eyermann, C. N. Hodge, Y. Ru, L. T. Bachelier, J. L. Meek, M. J. Otto, M. M. Rayner, Y. N. Wong, C. H. Chang, P. C. Weber, D. A. Jackson, T. R. Sharpe, and S. Erickson-Viitanen: *Rational design of potent, bioavailable, nonpeptide cyclic ureas as HIV protease inhibitors*. *Science*, 263:380–384, 1994.

- [12] Lam, P. Y. S., Y. Ru, P. K. Jadhav, P. E. Aldrich, G. V. DeLucca, C. J. Eyermann, C. H. Chang, G. Emmett, E. R. Holler, W. F. Daneker, L. Li, P. N. Confalone, R. J. McHugh, Q. Han, R. Li, J. A. Markwalder, S. P. Seitz, T. R. Sharpe, L. T. Bacheler, M. M. Rayner, R. M. Klabe, L. Shum, D. L. Winslow, D. M. Kornhauser, D. A. Jackson, S. Erickson-Viitanen, and C. N. Hodge: *Cyclic HIV Protease Inhibitors: Synthesis, Conformational Analysis, P2/P2' Structure-Activity Relationship, and Molecular Recognition of Cyclic Ureas*. J. Med. Chem., 39:3514–3525, 1996.
- [13] Lu, Y., R. Wang, C. Y. Yang, and S. Wang: *Analysis of Ligand-Bound Water Molecules in High-Resolution Crystal Structures of Protein-Ligand Complexes*. J. Chem. Inf. Model., 47:668–675, 2007.
- [14] Günther, J.: *Entwicklung einer Datenbank und wissensbasierter Vorhersagemethoden zur Untersuchung von Wassermolekülen in Proteinstrukturen sowie ihrer Rolle in der Protein-Ligand-Bindung*. Dissertation, Philipps-Universität Marburg, Marburg, 2003.
- [15] Li, Z. and T. Lazaridis: *Water at biomolecular binding interfaces*. Phys. Chem. Chem. Phys., 9:573–581, 2007.
- [16] Homans, Steve W.: *Water, water everywhere - except where it matters?* Drug Discovery Today, 12:534–539, 2007.
- [17] Matthews, B. W. and L. Liu: *A review about nothing: Are apolar cavities in proteins really empty?* Protein Sci., 18:494–502, 2009.
- [18] Berman, H. M., 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., 28:235–242, 2000.
- [19] Kleywegt, G. J., M. R. Harris, J. Zou, T. C. Taylor, A. Wählby, and T. A. Jones: *The Uppsala Electron-Density Server*. Acta Crystallogr., Sect. D: Biol. Crystallogr., 60:2240–2249, 2004.
- [20] Davis, A. M., S. J. Teague und G. J. Kleywegt: *Anwendung und Grenzen kristallographischer Daten im strukturbezogenen Liganden- und Wirkstoff-Design*. Angew. Chem., 115:2822–2841, 2003.
- [21] Fields, B. A., H. H. Bartsch, H. D. Bartunik, F. Cordes, J. M. Guss, and H. C. Freeman: *Accuracy and precision in protein crystal structure analysis: two independent refinements of the structure of poplar plastocyanin at 173 K*. Acta Crystallogr., Sect. D: Biol. Crystallogr., 50:709–730, 1994.
- [22] Ohlendorf, D. H.: *Accuracy of refined protein structures. ii. comparison of four independently refined models of human interleukin 1 $\beta$* . Acta Crystallogr., Sect. D: Biol. Crystallogr., 50:808–812, 1994.

- 
- [23] Levitt, M. and B. H. Park: *Water: now you see it, now you don't*. Structure, 1:223–226, 1993.
- [24] Denisov, V. P., J. Peters, H. D. Hörlein, and B. Halle: *Using buried water molecules to explore the energy landscape of proteins*. Nat. Struct. Biol., 3:505–509, 1996.
- [25] Denisov, V. P., K. Venu, J. Peters, H. D. Hörlein, and B. Halle: *Orientalional Disorder and Entropy of Water in Protein Cavities*. J. Phys. Chem. B, 101:9380–9389, 1997.
- [26] Takano, K., J. Funahashi, Y. Yamagata, S. Fujii, and K. Yutani: *Contribution of water molecules in the interior of a protein to the conformational stability*. J. Mol. Biol., 274:132–142, 1997.
- [27] Fischer, S. and C. S. Verma: *Binding of buried structural water increases the flexibility of proteins*. Proc. Natl. Acad. Sci. USA, 96:9613–9615, 1999.
- [28] Kirchmair, J., G. M. Spitzer, and K. R. Liedl: *Consideration of water and solvation effects in virtual screening*. In Sotriffer, C. (editor): *Virtual Screening: Principles, Challenges, and Practical Guidelines*. Wiley-VCH, Weinheim, 2011.
- [29] Pitt, W. R. and J. M. Goodfellow: *Modelling of solvent positions around polar groups in proteins*. Protein Eng., 4:531–537, 1991.
- [30] Goodford, P. J.: *A Computational Procedure for Determining Energetical Favorable Binding Sites on Biologically Important Macromolecules*. J. Med. Chem., 28:849–857, 1985.
- [31] Miranker, A. and M. Karplus: *Functionality maps of binding sites: a multiple copy simultaneous search method*. Proteins: Struct., Funct., Bioinf., 11:29–34, 1991.
- [32] Verdonk, M. L., J. C. Cole, and R. Taylor: *SuperStar: A knowledge-based approach for identifying interaction sites in proteins*. J. Mol. Biol., 289:1093–1108, 1999.
- [33] Kortvelyesi, T., S. Dennis, M. Silberstein, L. Brown III, and S. Vajda: *Algorithms for computational solvent mapping of proteins*. Proteins: Struct., Funct., Bioinf., 51:340–351, 2003.
- [34] Gohlke, H., M. Hendlich, and G. Klebe: *Knowledge-based scoring function to predict protein-ligand interactions*. J. Mol. Biol., 295:337–356, 2000.
- [35] Velec, H. F. G., H. Gohlke, and G. Klebe: *DrugScore(CSD)-knowledge-based scoring function derived from small molecule crystal data with superior recognition rate of near-native ligand poses and better affinity prediction*. J. Med. Chem., 48:6296–6303, 2005.

- [36] Halgren, T.: *New Method for Fast and Accurate Binding-site Identification and Analysis*. Chem. Biol. Drug Des., 69:146–148, 2007.
- [37] Halgren, T. A.: *Identifying and Characterizing Binding Sites and Accessing Druggability*. J. Chem. Inf. Model., 49:377–389, 2009.
- [38] Kellogg, G. E., S. F. Semus, and D. J. Abraham: *HINT: a new method of empirical hydrophobic field calculation for CoMFA*. J. Comput. Aided Mol. Des., 5:545–552, 1991.
- [39] Abraham, D. J. and G. E. Kellogg: *The effect of physical organic properties on hydrophobic fields*. J. Comput. Aided Mol. Des., 8:41–49, 1994.
- [40] Kellogg, G. E. and D. J. Abraham: *Hydrophobicity: is  $\log P_{o/w}$  more than the sum of its part?* Eur. J. Med. Chem., 35:651–661, 2000.
- [41] Amadasi, A., F. Spyrakis, P. Cozzini, D. J. Abraham, G. E. Kellogg, and A. Mozzarelli: *Mapping the Energetics of Water–Protein and Water–Ligand Interactions with the "Natural" HINT Forcefield: Predictive Tools for Characterizing the Roles of Water in Biomolecules*. J. Mol. Biol., 358:289–309, 2006.
- [42] Amadasi, A., J. A. Surface, F. Spyrakis, P. Cozzini, A. Mozzarelli, and G. E. Kellogg: *Robust Classification of "Relevant" Water Molecules in Putative Protein Binding Sites*. J. Med. Chem., 51:1063–1067, 2008.
- [43] Schymkowitz, J. W. H., F. Rousseau, I. C. Martins, J. Ferkinghoff-Borg, F. Stricher, and L. Serrano: *Prediction of water and metal binding sites and their affinities by using the Fold-X force field*. Proc. Natl. Acad. Sci. USA, 102:10147–10152, 2005.
- [44] Raymer, M. L., P. C. Sanschagrín, W. F. Punch, S. Venkataraman, E. D. Goodman, and L. A. Kuhn: *Predicting conserved water-mediated and polar ligand interactions in proteins using a K-nearest-neighbors genetic algorithm*. J. Mol. Biol., 265:445–464, 1997.
- [45] García-Sosa, A. T., R. L. Mancera, and P. M. Dean: *WaterScore: a novel method for distinguishing between bound and displaceable water molecules in the crystal structure of the binding site of protein-ligand complexes*. J. Mol. Model., 9:172–182, 2003.
- [46] Villacañas, O., S. Madurga, E. Giralt, and I. Belda: *Explicit treatment of water molecules in protein-ligand docking*. Curr. Comput.-Aided Drug Des., 5:145–154, 2009.
- [47] Rarey, M., B. Kramer, and T. Lengauer: *The Particle Concept: Placing Discrete Water Molecules During Protein-Ligand Docking Predictions*. Proteins: Struct., Funct., Bioinf., 34:17–28, 1999.

- [48] Rarey, M., B. Kramer, T. Lengauer, and G. Klebe: *A fast flexible docking method using an incremental construction algorithm*. J. Mol. Biol., 261:470–489, 1996.
- [49] Jones, G., P. Willett, and R. C. Glen: *Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation*. J. Mol. Biol., 245:45–53, 1995.
- [50] Jones, G., P. Willett, R. C. Glen, A. R. Leach, and R. Taylor: *Development and Validation of a Genetic Algorithm for Flexible Docking*. J. Mol. Biol., 267:727–748, 1997.
- [51] Verdonk, M. L., J. C. Cole, M. J. Hartshorn, C. W. Murray, and R. D. Taylor: *Improved protein-ligand docking using GOLD*. Proteins: Struct., Funct., Bioinf., 52:609–623, 2003.
- [52] Verdonk, M. L., G. Chessari, J. C. Cole, M. J. Hartshorn, C. W. Murray, J. W. M. Nissink, R. D., and R. Taylor: *Modeling Water Molecules in Protein-Ligand Docking Using GOLD*. J. Med. Chem., 48:6504–6515, 2005.
- [53] Korb, O., T. Stützle, and T. E. Exner: *PLANTS: Application of Ant Colony Optimization to Structure-Based Drug Design*. In Dorigo, M., L. M. Gambardella, M. Birattari, A. Martinoli, R. Poli, and T. Stützle (editors): *Ant Colony Optimization and Swarm Intelligence*, pages 247–258, 2006.
- [54] Korb, O., T. Stützle, and T. E. Exner: *An Ant Colony Optimization Approach to Flexible Protein-Ligand Docking*. Swarm Intelligence, 1:115–134, 2007.
- [55] Korb, O.: *Efficient Ant Colony Optimization Algorithms for Structure- and Ligand-Based Drug Design*. PhD thesis, University of Konstanz, 2008.
- [56] Korb, O., T. Stützle, and T. E. Exner: *Empirical Scoring Functions for Advanced Protein-Ligand Docking with PLANTS*. J. Chem. Inf. Model., 49:84–96, 2009.
- [57] Exner, T. E., O. Korb, and T. ten Brink: *New and improved features of the docking software plants*. In *4th German Conference on Chemoinformatics*, 2009.
- [58] Schneck, V. and L. A. Kuhn: *Virtual screening with solvation and ligand-induced complementarity*. Perspect. Drug Discov. Des., 20:171–190, 2000.
- [59] Corbeil, C. R., P. Englebienne, and N. Moitessier: *Docking Ligands into Flexible and Solvated Macromolecules. 1. Development and Validation of FITTED 1.0*. J. Chem. Inf. Model., 47:435–449, 2007.
- [60] Friesner, R. A., J. L. Banks, R. B. Murphy, T. A. Halgren, J. J. Klicic, D. T. Mainz, M. P. Repasky, E. H. Knoll, M. Shelley, J. K. Perry, D. E. Shaw, P. Francis, and P. S. Shenkin: *Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy*. J. Med. Chem., 47:1739–1749, 2004.

- [61] Brenk, R., L. Naerum, U. Grädler, H. D. Gerber, G. A. García, K. Reuter, M. T. Stubbs, and G. Klebe: *Virtual Screening for Submicromolar Leads of tRNA-guanine Transglycosylase Based on a New Unexpected Binding Mode Detected by Crystal Structure Analysis*. *J. Med. Chem.*, 46:1133–1143, 2003.
- [62] Lloyd, D. G., A. T. Garcia-Sosa, I. L. Alberts, N. P. Todorov, and R. L. Mancera: *The effect of tightly bound water molecules on the structural interpretation of ligand-derived pharmacophore models*. *J. Comput. Aided Mol. Des.*, 18:89–100, 2004.
- [63] Lounnas, V. and B. M. Pettitt: *A Connected-cluster of hydration around myoglobin: Correlation between molecular dynamics simulations and experiment*. *Proteins: Struct., Funct., Bioinf.*, 18:133–147, 1994.
- [64] Henchman, R. H. and J. A. McCammon: *Structural and dynamic properties of water around acetylcholinesterase*. *Protein Sci.*, 11:2080–2090, 2002.
- [65] Henchman, R. H. and J. A. McCammon: *Extracting hydration sites around proteins from explicit water simulations*. *J. Comput. Chem.*, 23:861–869, 2002.
- [66] Henchman, R. H., K. Tai, T. Shen, and J. A. McCammon: *Properties of Water Molecules in the Active Site Gorge of Acetylcholinesterase from Computer Simulation*. *Biophys. J.*, 82:2671–2682, 2002.
- [67] Åqvist, J., C. Medina, and J. F. Samuelsson: *A new method for predicting binding affinity in computer-aided drug design*. *Protein Eng.*, 7:385–391, 1994.
- [68] Hansson, T., J. Marelus, and J. Åqvist: *Ligand binding affinity prediction by linear interaction energy methods*. *J. Comput. Aided Mol. Des.*, 12:27–35, 1998.
- [69] Srinivasan, J., J. Miller, P. A. Kollman, and D. A. Case: *Continuum solvent studies of the stability of RNA hairpin loop and helices*. *J. Biomol. Struct. dyn.*, 16:671–682, 1998.
- [70] Vorobjev, Y. N., J. C. Almagro, and J. Hermans: *Discrimination between native and intentionally misfolded conformations of proteins: ES/IS, a new method for calculating conformational free energy that uses both dynamics simulations with an explicit solvent and an implicit solvent continuum model*. *Proteins: Struct., Funct., Bioinf.*, 32:399–413, 1998.
- [71] Jayaram, B., D. Sprous, M. A. Young, and D. L. Beveridge: *Free Energy Analysis of the Conformational Preferences of A and B Forms of DNA in Solution*. *J. Am. Chem. Soc.*, 120:10629–10633, 1998.
- [72] Kollman, P. A., I. Massova, C. Reyes, B. Kuhn, S. Huo, L. Chong, M. Lee, T. Lee, Y. Duan, W. Wang, O. Donini, P. Cieplak, J. Srinivasan, D. A. Case, and III T. E. Cheatham: *Calculating Structures and Free Energies of Complex Molecules:*

- Combining Molecular Mechanics and Continuum Models.* Acc. Chem. Res., 33:889–897, 2000.
- [73] Huang, N., 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., 46:243–253, 2006.
- [74] Yu, Z., M. P. Jacobson, and R. A. Friesner: *What role do surfaces play in GB models? A new-generation of surface-generalized born model based on a novel gaussian surface for biomolecules.* J. Comput. Chem., 27:72–89, 2006.
- [75] Young, T., 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. USA, 104:3808–3813, 2007.
- [76] Abel, R., 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., 130:2817–2831, 2008.
- [77] Beuming, T., R. Farid, and W. Sherman: *High-energy water sites determine peptide binding affinity and specificity of PDZ domains.* Protein Sci., 18:1609–1619, 2009.
- [78] Young, T., R. Abel, R. A. Friesner, and B. J. Berne: *Methods of calculating differences of binding affinities between congeneric pairs of ligands by way of a displaced solvent functional.* Patent, July 2010. [http://www.patentlens.net/patentlens/patent/US\\_7756674/en/](http://www.patentlens.net/patentlens/patent/US_7756674/en/), US 7756674.
- [79] Lazaridis, T.: *Inhomogeneous Fluid Approach to Solvation Thermodynamics. 1. Theory.* J. Phys. Chem. B, 102:3531–3541, 1998.
- [80] Lazaridis, T.: *Solvent Reorganization Energy and Entropy in Hydrophobic Hydration.* J. Phys. Chem. B, 104:4964–4979, 2000.
- [81] Li, Z. and T. Lazaridis: *Thermodynamic Contributions of the Ordered Water Molecule in HIV-1 Protease.* J. Am. Chem. Soc., 125:6636–6637, 2003.
- [82] Li, Z. and T. Lazaridis: *The Effect of Water Displacement on Binding Thermodynamics: Concanavalin A.* J. Phys. Chem. B, 109:662–670, 2005.
- [83] Li, Z. and T. Lazaridis: *Thermodynamics of Buried Water Clusters at a Protein-Ligand Binding Interface.* J. Phys. Chem. B, 110:1464–1475, 2006.
- [84] Robinson, D. D., W. Sherman, and R. Farid: *Understanding Kinase Selectivity Through Energetic Analysis of Binding Site Waters.* ChemMedChem, 5:618–627, 2010.

- [85] Guimarães, C. R. W. and A. M. Mathiowetz: *Addressing Limitations with the MM-GB/SA Scoring Procedure using the WaterMap Method and Free Energy Perturbation Calculations*. J. Chem. Inf. Model., 50:547–559, 2010.
- [86] Higgs, C., T. Beuming, and W. Sherman: *Hydration Site Thermodynamics Explain SARs for Triazolylpurines Analogues Binding to the A2A Receptor*. ACS Med. Chem. Lett., 1:160–164, 2010.
- [87] Chrencik, J. E., A. Patny, I. K. Leung, B. Korniski, T. L. Emmons, T. Hall, R. A. Weinberg, J. A. Gormley, J. M. Williams, J. E. Day, J. L. Hirsch, J. R. Kiefer, J. W. Leone, H. D. Fischer, C. D. Sommers, H. C. Huang, E.J. Jacobsen, R. E. Tenbrink, A. G. Tomasselli, and T. E. Benson: *Structural and Thermodynamic Characterization of the TYK2 and JAK3 Kinase Domains in Complex with CP-690550 and CMP-6*. J. Mol. Biol., ASAP, 2010.
- [88] Pearlstein, R. A., Q Y. Hu, J. Zhou, D. Yowe, J. Levell, B. Dale, V. K. Kaushik, D. Daniels, S. Hanrahan, W. Sherman, and R. Abel: *New hypotheses about the structure-function of proprotein convertase subtilisin/kexin type 9: Analysis of the epidermal growth factor-like repeat A docking site using WaterMap*. Proteins: Struct., Funct., Bioinf., 78:2571–2586, 2010.
- [89] Kraut, J.: *Cytochrome c peroxidase*. <http://chem-faculty.ucsd.edu/kraut/projects.html>.
- [90] Fitzgerald, M. M., M. J. Churchill, D. E. McRee, and D. B. Goodin: *Small Molecule Binding to an Artificially Created Cavity at the Active Site of Cytochrome c Peroxidase*. Biochemistry, 33:3807–3818, 1994.
- [91] Fitzgerald, M. M., R. A. Musah, D. E. McRee, and D. B. Goodin: *A ligand-gated, hinged loop rearrangement opens a channel to a buried artificial protein cavity*. Nat. Struct. Biol., 3:626–631, 1996.
- [92] Fitzgerald, M. M., M. L. Trester, G. M. Jensen, D. E. Mcree, and D. B. Goodin: *The role of aspartate-235 in the binding of cations to an artificial cavity at the radical site of cytochrome c peroxidase*. Protein Sci., 4:1844–1850, 1995.
- [93] Fitzgerald, M. M., R. A. Musah, D. E. McRee, and D. B. Goodin: *Variation in strength of an unconventional C-H to O hydrogen bond in an engineered protein cavity*. J. Am. Chem. Soc., 119:626–631, 1997.
- [94] Musah, R. A. and D. B. Goodin: *Introduction of Novel Substrate Oxidation into Cytochrome c Peroxidase by Cavity Complementation: Oxidation of 2-Aminothiazole and Covalent Modification of the Enzyme*. Biochemistry, 36:11665–11674, 1997.
- [95] Musah, R. A., G. M. Jensen, S. W. Bunte, R. J. Rosenfeld, and D. B. Goodin: *Artificial Protein Cavities as Specific Ligand-binding Templates: Characterization of*



- an Engineered Heterocyclic Cation-binding Site that Preserves the Evolved Specificity of the Parent Protein.* J. Mol. Biol., 315:845–857, 2002.
- [96] Brenk, R., S. Vetter, S. E. Boyce, D. B. Goodin, and B. K. Shoichet: *Probing molecular docking in a charged model binding site.* J. Mol. Biol., 357:1449–1470, 2006.
- [97] Graves, A. P., D. M. Shivakumar, S. E. Boyce, M. P. Jacobson, D. A. Case, and B. K. Shoichet: *Rescoring Docking Hit Lists for Model Cavity Sites: Predictions and Experimental Testing.* J. Mol. Biol., 377:914–934, 2008.
- [98] Rosenfeld, R. J., D. S. Goodsell, R. A. Musah, G. M. Morris, D. B. Goodin, and A. J. Olson: *Automated docking of ligands to an artificial active site: augmenting crystallographic analysis with computer modeling.* J. Comput. Aided Mol. Des., 17:525–536, 2003.
- [99] Deng, W. and C. L. M. J. Verlinde: *Evaluation of Different Virtual Screening Programs for Docking in a Charged Binding Pocket.* J. Chem. Inf. Model., 48:2010–2020, 2008.
- [100] Baron, R. and J. A. McCammon: *Dynamics, Hydration, and Motional Averaging of a Loop-Gated Artificial Protein Cavity: The W191G Mutant of Cytochrome c Peroxidase in Water as Revealed by Molecular Dynamics Simulations.* Biochemistry, 46:10629–10642, 2007.
- [101] Baron, R. and J. A. McCammon: *(Thermo)dynamic Role of Receptor Flexibility, Entropy, and Motional Correlation in Protein-Ligand Binding.* ChemPhysChem, 9:983–988, 2008.
- [102] Amaro, R. E., R. Baron, and J. A. McCammon: *An improved relaxed complex scheme for receptor flexibility in computer-aided drug design.* J. Comput. Aided Mol. Des., 22:693–705, 2008.
- [103] Kim, K. H.: *Outliers in SAR and QSAR: Is unusual binding mode a possible source of outliers?* J. Comput. Aided Mol. Des., 21:63–86, 2007.
- [104] Miller, M. A., G. W. Hand, and J. Kraut: *A cation binding motif stabilizes the compound I radical of cytochrome c peroxidase.* Proc. Natl. Acad. Sci. USA, 91:11118–11122, 1994.
- [105] Putnam, A. M. Hays, Y. T. Lee, and D. B. Goodin: *Replacement of an electron transfer pathway in cytochrome c peroxidase with a surrogate peptide.* Biochemistry, 48:1–3, 2009.
- [106] Gunsteren, W. F. van und H. J. C. Berendsen: *Moleküldynamik-Computersimulationen; Methodik, Anwendung und Perspektiven in der Chemie.* Angew. Chem., 102:1020–1055, 1990.

- [107] Gunsteren, W. F. van, D. Bakowies, R. Baron, I. Chandrasekhar, M. Christen, X. Daura, P. Gee, D. P. Geerke, A. Glättli, P. H. Hünenberger, M. A. Kastenholz, C. Oostenbrink, M. Schenk, D. Trzesniak, N. F. A. van der Vegt und H. B. Yu: *Biomolekulare Modellierung: Ziele, Probleme, Perspektiven*. *Angew. Chem.*, 118:4168–4198, 2006.
- [108] Jorgensen, W. L. and J. Tirado-Rives: *The OPLS Force Field for Proteins. Energy Minimizations for Crystals of Cyclic Peptides and Crambin*. *J. Am. Chem. Soc.*, 110:1657–1666, 1988.
- [109] Jorgensen, W. L., D. S. Maxwell, and J. Tirado-Rives: *Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids*. *J. Am. Chem. Soc.*, 118:11225–11236, 1996.
- [110] MacKerel Jr., A. D., C. L. Brooks III, L. Nilsson, B. Roux, Y. Won, and M. Karplus: *CHARMM: The Energy Function and Its Parameterization with an Overview of the Program*, volume 1 of *The Encyclopedia of Computational Chemistry*, pages 271–277. John Wiley & Sons: Chichester, 1998.
- [111] MacKerell, A. D., D. Bashford, M. 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. Wiorkiewicz-Kuczera, D. Yin, and M. Karplus: *All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins*. *J. Phys. Chem. B*, 102:3586–3616, 1998.
- [112] MacKerell Jr., A. D., N. Banavali, and N. Foloppe: *Development and current status of the CHARMM force field for nucleic acids*. *Biopolymers*, 56:257–265, 2001.
- [113] Brooks, B. R., R. E. Bruccoleri, D. J. Olafson, D. J. States, S. Swaminathan, and M. Karplus: *CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations*. *Journal of Computational Chemistry*, 4:187–217, 1983.
- [114] Gunsteren, W. F. van and H. J. C. Berendsen: *Groningen Molecular Simulations (GROMOS) Library Manual*. University of Groningen, The Netherlands, 1987.
- [115] Case, D.A., III T.E. Cheatham, T. Darden, H. Gohlke, R. Luo, Jr. K.M. Merz, A. Onufriev, C. Simmerling, B. Wang, and R. Woods: *The Amber biomolecular simulation programs*. *J. Computat. Chem.*, 26:1668–1688, 2005.
- [116] Gunsteren, W. F. van, X. Daura, and A. E. Mark: *GROMOS force field*, pages 1211–1216. *The Encyclopedia of Computational Chemistry*. John Wiley & Sons: Chichester, 1998.

- [117] Schuler, L. D., X. Daura, and W. F. van Gunsteren.: *An improved GROMOS96 force field for aliphatic hydrocarbons in the condensed phase.* J. Comput. Chem., 22:1205–1218, 2001.
- [118] Ponder, J. W. and D. A. Case: *Force Fields for Protein Simulations.* Adv. Protein Chem., 66:27–85, 2003.
- [119] Wang, J., P. Cieplak, and P. A. Kollman: *How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules?* J. Comput. Chem., 21:1049–174, 2000.
- [120] Bayly, C. I., P. Cieplak, W. D. Cornell, and P. A. Kollman: *A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges: The RESP Model.* J. Phys. Chem., 97:10269–10280, 1993.
- [121] University of California, San Francisco: *Amber 10: User's manual*, 2008.
- [122] Berendsen, H. J. C., J. P. M. Postma, W. F. van Gunsteren, and J. Hermans: *Intermolecular Forces*, page 331. Reidel, Dordrecht, 1981.
- [123] Berendsen, H. J. C., J. R. Grigera, and T. P. Straatsma: *The missing term in effective pair potentials.* J. Phys. Chem., 91:6269–6271, 1987.
- [124] Jorgensen, W. L., J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein: *Comparison of Simple Potential Functions for Simulating Liquid Water.* J. Chem. Phys., 79:926–935, 1983.
- [125] Leach, A. R.: *Molecular Modelling: Principles and Applications.* Pearson, Harlow, 2nd edition, 2001.
- [126] Ryckaert, J. P., G. Ciccotti, and H. J. C. Berendsen: *Numerical integration of the cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes.* J. Comput. Phys., 23:327–341, 1977.
- [127] Darden, T., D. York, and L. Pedersen: *Particle mesh Ewald: An  $N \log(N)$  method for Ewald sums in large systems.* J. Chem. Phys., 98:10089–10092, 1993.
- [128] Friesner, R. A., R. B. Murphy, M. P. Repasky, L. L. Frye, J. R. Greenwood, T. A. Halgren, P. C. Sanschagrin, and D. T. Mainz: *Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein-Ligand Complexes.* J. Med. Chem., 49:6177–6196, 2006.
- [129] Klebe, G. und H. Gohlke: *Ortsaufgelöste Identifizierung günstiger Protein-Ligand-Wechselwirkungen in Bindetaschen.* Angew. Chem., 114:2764–2798, 2002.
- [130] Kirkwood, J. G.: *Statistical Mechanics of Fluid Mixtures.* J. Chem. Phys., 3:300–313, 1935.

- [131] Zwanzig, R. Z.: *High-Temperatur Equation of State by a Perturbation Method. I. Nonpolar Gases.* J. Chem. Phys., 22:1420–1426, 1954.
- [132] Ladbury, J. E. and M. L. Doyle (editors): *Biocalorimetry 2.* John Wiley & Sons, 2004.
- [133] Wiseman, T., S. Williston, J. F. Brandts, and L. N. Lin: *Rapid measurement of binding constants and heats of binding using a new titration calorimeter.* Anal. Biochem., 179:131–137, 1989.
- [134] Olsson, T. S. G., M. A. Williams, R. W. Pitt, and J. E. Ladbury: *The Thermodynamics of Protein–Ligand Interaction and Solvation: Insights for Ligand Design.* J. Mol. Biol., 384:1002–1017, 2008.
- [135] Böhm, H. J. and M. Stahl: *The Use of Scoring Functions in Drug Discovery Applications.* In Lipkowitz, K. B. and D. B. Boyd (editors): *Reviews in Computational Chemistry*, volume 18, pages 41–87. Wiley-VCH, New York, USA, 2002.
- [136] Rajamani, R. and A. C. Good: *Ranking poses in structure-based lead discovery and optimization: Current trends in scoring function development.* Curr. Opin. Drug Disc. Dev., 10:308–315, 2007.
- [137] Sottriffer, C. A. and H. Matter: *The challenge of affinity prediction: Scoring functions for structure-based virtual screening.* In Sottriffer, C. (editor): *Virtual Screening: Principles, Challenges, and Practical Guidelines.* Wiley-VCH, Weinheim, 2011.
- [138] Stahl, M. and M. Rarey: *Detailed Analysis of Scoring Functions for Virtual Screening.* J. Med. Chem., 44:1035–1042, 2001.
- [139] Wang, R., 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., 44:2114–2125, 2004.
- [140] Ferrara, P., H. Gohlke, D. J. Price, G. Klebe, and C. L. Brooks III: *Assessing Scoring Functions for Protein-Ligand Interactions.* J. Med. Chem., 47:3032–3047, 2004.
- [141] Cheng, T., X. Li, Y. Li, Z. Liu, and R. Wang: *Comparative Assessment of Scoring Functions on a Diverse Test Set.* J. Chem. Inf. Model., 49:1079–1093, 2009.
- [142] Eldridge, M. D., 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.
- [143] Baxter, C. A., 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, 33:367–382, 1998.

- 
- [144] Wang, R., L. Lai, and S. Wang: *Further development and validation of empirical scoring functions for structure-based binding affinity prediction.* J. Comput.-Aided Mol. Des., 16:11–26, 2004.
- [145] Sotriffer, C. A., P. Sanschagrin, H. Matter, and G. Klebe: *SFCscore: Scoring functions for affinity prediction of protein-ligand complexes.* Proteins: Struct., Funct., Bioinf., 73:395–419, 2008.
- [146] Pfeffer, P. and H. Gohlke: *DrugScore<sup>RNA</sup> - Knowledge-Based Scoring Function To Predict RNA-Ligand Interactions.* J. Chem. Inf. Model., 47:1868–1876, 2007.
- [147] Muegge, I. and Y. C. Martin: *A General and Fast Scoring Function for Protein-Ligand Interactions: A Simplified Potential Approach.* J. Med. Chem., 42:791–804, 1999.
- [148] Muegge, I.: *PMF Scoring Revisited.* J. Med. Chem., 49:5895–5902, 2006.
- [149] Mitsutake, A., Y. Sugita, and Y. Okamoto: *Generalized-ensemble algorithms for molecular simulations of biopolymers.* Biopolymers, 60:96–123, 2001.
- [150] Nymeyer, H., S. Gnanakaran, and A. E. Garcia: *Atomic Simulations of Protein Folding, Using the Replica Exchange Algorithm.* Meth. Enzymol., 383:119–149, 2004.
- [151] Cheng, X., G. Cui, V. Hornak, and Carlos Simmerling: *Modified Replica Exchange Simulation Methods for Local Structure Refinement.* J. Phys. Chem. B, 109:8220–8230, 2005.
- [152] Kumar, S., J. M. Rosenberg, D. Bouzida, R. H. Swendsen, and P. A. Kollman: *THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method.* J. Comput. Chem., 13:1011–1021, 1992.
- [153] Kumar, S., J. M. Rosenberg, D. Bouzida, R. H. Swendsen, and P. A. Kollman: *Multidimensional free-energy calculations using the weighted histogram analysis method.* J. Comput. Chem., 16:1339–1350, 1995.
- [154] Roux, B.: *The calculation of the potential of mean force using computer simulations.* Comput. Phys. Comm., 91:275–282, 1995.
- [155] Woods, C. J., M. A. King, and J. W. Essex: *The Development of Replica-Exchange-Based Free-Energy Methods.* J. Phys. Chem. B, 107:13703–13710, 2003.
- [156] Woods, C. J., M. A. King, and J. W. Essex: *Enhanced Configurational Sampling in Binding Free-Energy Calculations.* J. Phys. Chem. B, 107:13711–13718, 2003.
- [157] Tembe, B. L. and J. A. McCammon: *Ligand-receptor interactions.* Comput. Chem., 8:281–283, 1982.

- [158] Pranata, J. and W. L. Jorgensen: *Monte Carlo simulations yield absolute free energies of binding for guanine–cytosine and adenine–uracil base pairs in chloroform*. *Tetrahedron*, 47:2491–2501, 1991.
- [159] Gilson, M. K., J. A. Given, B. L. Bush, and J. A. McCammon: *The Statistical-Thermodynamic Basis for Computation of Binding Affinities: A Critical Review*. *Biophys. J.*, 72:1047–1069, 1997.
- [160] Mobley, D., J. Chodera, and M. Shirts: *Alchemy wiki*. <http://www.alchemistry.org/wiki/index.php>.
- [161] Steinbrecher, T., D. L. Mobley, and D. A. Case: *Non-linear scaling schemes for Lennard-Jones interactions in free energy calculations*. *J. Chem. Phys.*, 127:214108, 2007.
- [162] Åqvist, J., V. B. Luzhkov, and B. O. Brandsdal: *Ligand Binding Affinities from MD Simulations*. *Acc. Chem. Res.*, 35:358–365, 2002.
- [163] Pearlman, D. A.: *Evaluating the Molecular Mechanics Poisson-Boltzmann Surface Area Free Energy Method Using a Congeneric Series of Ligands to p38 MAP Kinase*. *J. Med. Chem.*, 48:7796–7807, 2005.
- [164] Lazaridis, T. and M. E. Paulaitis: *Entropy of Hydrophobic Hydration: A New Statistical Mechanical Formulation*. *J. Phys. Chem.*, 96:3847–3855, 1992.
- [165] Lazaridis, T. and M. E. Paulaitis: *Reply on the Comment on “Entropy of Hydrophobic Hydration: A New Statistical Mechanical Formulation”*. *J. Phys. Chem.*, 97:5789–5790, 1993.
- [166] Lazaridis, T. and M. E. Paulaitis: *Simulation Studies of the Hydration Entropy of Simple, Hydrophobic Solutes*. *J. Phys. Chem.*, 98:635–642, 1994.
- [167] Morita, T. and K. Hiroike: *A New Approach to the Theory of Classical Fluids. III*. *Prog. Theor. Phys.*, 25:537–578, 1961.
- [168] Peter, C., C. Oostenbrink, A. van Dorp, and W. F. van Gunsteren: *Estimating entropies from molecular dynamics simulations*. *J. Chem. Phys.*, 120:2652–2661, 2004.
- [169] Allen, M. P. and D. J. Tildesley: *Computer Simulation of Liquids*. Oxford University Press, Oxford, 2002.
- [170] Frenkel, D. and B. Smit: *Understanding Molecular Simulation: From Algorithms to Applications*. Academic Press, San Diego, 1996.
- [171] Bronstein, I. N., K. A. Semendjajew, G. Musiol und H. Mühlig: *Taschenbuch der Mathematik*. Verlag Harri Deutsch, Thun und Frankfurt am Main, 5., überarbeitete und erweiterte Auflage, 2000.

- [172] Lazaridis, T. and M. Karplus: *Orientational correlations and entropy in liquid water*. J. Chem. Phys., 105:4294–4316, 1996.
- [173] Wang, L., R. Abel, R. A. Friesner, and B. J. Berne: *Thermodynamic Properties of Liquid Water: An Application of a Nonparametric Approach to Computing the Entropy of a Neat Fluid*. J. Chem. Theory Comput., 5:1462–1473, 2009.
- [174] Abel, R., L. Wang, R. A. Friesner, and B. J. Berne: *A Displaced-Solvent Functional Analysis of Model Hydrophobic Enclosures*. J. Chem. Theory Comput., ASAP, 2010.
- [175] Shoichet, B. K.: *Virtual screening of chemical libraries*. Nature, 432:862–865, 2004.
- [176] Klebe, G.: *Virtual ligand screening: strategies, perspectives and limitations*. Drug Discovery Today, 11:580–594, 2006.
- [177] Sotriffer, C. (editor): *Virtual Screening: Principles, Challenges, and Practical Guidelines*. Wiley-VCH, Weinheim, 2011.
- [178] Chemical Computing Group: *Molecular Operating Environment (MOE), version 2007.09*. Quebec, Kanada, 2007.
- [179] Bergner, A., J. Günther, M. Hendlich, G. Klebe, and M. Verdonk: *Use of Relibase for Retrieving Complex 3D Interaction Patterns Including Crystallographic Packing Effects*. Biopolymers (Nucleic Acid Sci.), 61:99–110, 2002.
- [180] Hendlich, M., A. Bergner, J. Günther, and G. Klebe: *Relibase - Design and Development of a Database for Comprehensive Analysis of Protein-Ligand Interactions*. J. Mol. Biol., 326:607–620, 2003.
- [181] Günther, J., A. Bergner, M. Hendlich, and G. Klebe: *Utilising Structural Knowledge in Drug Design Strategies - Applications Using Relibase*. J. Mol. Biol., 326:621–636, 2003.
- [182] Sanschagrin, P. C. and L. A. Kuhn: *Cluster Analysis of Consensus Water Sites in Thrombin and Trypsin Shows Conservation Between Serine Proteases and Contributions to Ligand Specificity*. Protein Sci., 7:2054–2064, 1998.
- [183] Nayal, M. and E. Di Cera: *Valence Screening of Water in Protein Crystals Reveals Potential Na<sup>+</sup> Binding Sites*. J. Mol. Biol., 256:228–234, 1996.
- [184] Gohlke, H., M. Hendlich, and G. Klebe: *Knowledge-Based Scoring Function to Predict Protein-Ligand Interactions*. J. Mol. Biol., 295:337–356, 2000.
- [185] Moitessier, N., 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*. British Journal of Pharmacology, 153:S7–S26, 2008.

- [186] McGann, M. R., H. R. Almond, A. Nicholls, J. A. Grant, and F. K. Brown: *Gaussian docking functions*. *Biopolymers*, 63:76–90, 2002.
- [187] Oshiro, C. M., I. D. Kuntz, and J. Scott Dixon: *Flexible ligand docking using a genetic algorithm*. *J. Comput. Aided Mol. Des.*, 9:113–130, 1995.
- [188] Knegtela, R. M. A., I. D. Kuntza, and C. M. Oshiro: *Molecular docking to ensembles of protein structures*. *J. Mol. Biol.*, 266:424–440, 1997.
- [189] Kang, X., R. H. Shafer, and I. D. Kuntz: *Calculation of ligand-nucleic acid binding free energies with the generalized-born model in DOCK*. *Biopolymers*, 73:192–204, 2003.
- [190] Moustakas, D. T., P. T. Lang, S. Pegg, E. Pettersen, I. D. Kuntz, N. Brooijmans, and R. C. Rizzo: *Development and validation of a modular, extensible docking program: Dock 5*. *J. Comput. Aided Mol. Des.*, 20:10–11, 2006.
- [191] Morris, G. M., 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.*, 19:1639–1662, 1998.
- [192] Österberg, F., G. M. Morris, M. F. Sanner, A. J. Olson, and D. S. Goodsell: *Automated docking to multiple target structures: Incorporation of protein mobility and structural water heterogeneity in autodock*. *Proteins: Struct., Funct., Bioinf.*, 46:34–40, 2002.
- [193] Sherman, W., T. Day, M. P. Jacobson, R. A. Friesner, and R. Farid: *Novel Procedure for Modeling Ligand/Receptor Induced Fit Effects*. *J. Med. Chem.*, 49:534–553, 2006.
- [194] Cozzin, P., G. E. Kellogg, F. Spyrakis, D. J. Abraham, G. Costantino, A. Emerson, F. Fanelli, H. Gohlke nad L. A. Kuhn, G. M. Morris, M. Orozco, T. A. Pertinhez, M. Rizzi, and C. A. Sotriffer: *Target Flexibility: An Emerging Consideration in Drug Discovery and Design*. *J. Med. Chem.*, 51:6237–6255, 2008.
- [195] Ritschel, T., P. C. Kohler, G. Neudert, A. Heine, F. Diederich, and G. Klebe: *How to Replace the Residual Solvation Shell of Polar Active Site Residues to Achieve Nanomolar Inhibition of tRNA-Guanine Transglycosylase*. *ChemMedChem*, 4:2012–2023, 2009.
- [196] Wade, R. C., K. J. Clark, and P. J. Goodford: *Further development of hydrogen bond functions for use in determining energetically favorable binding sites on molecules of known structure. 1. Ligand probe groups with the ability to form two hydrogen bonds*. *J. Med. Chem.*, 36:140–147, 1993.
- [197] Wade, R. C. and P. J. Goodford: *Further development of hydrogen bond functions for use in determining energetically favorable binding sites on molecules of known*



- structure. 2. *Ligand probe groups with the ability to form more than two hydrogen bonds*. J. Med. Chem., 36:148–156, 1993.
- [198] Carosati, E., S. Sciabola, and G. Cruciani: *Hydrogen Bonding Interactions of Covalently Bonded Fluorine Atoms: From Crystallographic Data to a New Angular Function in the GRID Force Field*. J. Med. Chem., 47:5114–5125, 2004.
- [199] DeLano, W. L.: *The PyMOL Molecular Graphics System*. <http://www.pymol.org>, 2002.
- [200] *Sybyl 8.0*. Tripos Inc., 1699 South Hanley Rd., St. Louis, Missouri, 63144, USA.
- [201] Tronrud, D. E.: *Knowledge-Based B-Factor Restraints for the Refinement of Proteins*. J. Appl. Crystallogr., 29:100–104, 1996.
- [202] Frauenfelder, H. and G. A. Petsko: *Structural dynamics of liganded myoglobin*. Biophys. J., 32:465–478, 1980.
- [203] Wang, W., O. Donini, C. M. Reyes, and P. A. Kollman: *Biomolecular Simulations: Recent Developments in Force Fields, Simulations of Enzyme Catalysis, Protein-Ligand, Protein-Protein, and Protein-Nucleic Acid Noncovalent Interactions*. Annu. Rev. Biophys. Biomol. Struct., 30:211–243, 2001.
- [204] Chemical Computing Group: *Molecular Operating Environment (MOE), version 2008.10*. Quebec, Kanada, 2008.
- [205] Krause, R., J. E. Nielsen, and G. Vriend: *What if*. <http://swift.cmbi.ru.nl/>.
- [206] Case, D. A., T. A. Darden, III T. E. Cheatham, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, K. M. Merz, D. A. Pearlman, M. Crowley, R. C. Walker, W. Zhang, B. Wang, S. Hayik, A. Roitberg, G. Seabra, K. F. Wong, F. Paesani, X. Wu, S. Brozell, V. Tsui, H. Gohlke, L. Yang, C. Tan, J. Mongan, V. Hornak, G. Cui, P. Beroza, D. H. Mathews, C. Schafmeister, W. S. Ross, and P. A. Kollman: *AMBER 9*. University of California, San Francisco, 2006.
- [207] Case, D. A., T. A. Darden, III T. E. Cheatham, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, M. Crowley, R. C. Walker, W. Zhang, K. M. Merz, B. Wang, S. Hayik, A. Roitberg, G. Seabra, I. Kollosvary, K. F. Wong, F. Paesani, J. Vanicek, X. Wu, S. R. Brozell, T. Steinbrecher, H. Gohlke, L. Yang, C. Tan, J. Mongan, V. Hornak, G. Cui, P. Beroza, D. H. Mathews, M. G. Seetin, C. Sagui, V. Babin, and P. A. Kollman: *AMBER 10*. University of California, San Francisco, 2008.
- [208] Frisch, M. J., G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, Jr., T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda,

- J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, and J. A. Pople: *Gaussian 03, Revision D.01*, 2004. Gaussian, Inc., Wallingford, CT.
- [209] Wang, J., W. Wang, P. A. Kollman, and D. A. Case: *Automatic atom type and bond type perception in molecular mechanical calculations*. J. Mol. Graphics Modell., 25:247–260, 2006.
- [210] Winter, M.: *WebElements™ Periodic Table*. United Kingdom: University of Sheffield. <http://www.webelements.com>, 2007. english.
- [211] *Amber Parameter Database*. <http://pharmacy.man.ac.uk/amber/>, 2007.
- [212] Giammona, D. A.: *An Examination of Conformational Flexibility in Porphyrins and Bulky-Ligand Binding in Myoglobin*. PhD thesis, University of California, Davis, 1984.
- [213] Wang, J., R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case: *Development and testing of a general Amber force field*. J. Comput. Chem., 25:1157–1174, 2004.
- [214] Onufriev, A., D. Bashford, and D. A. Case: *Exploring protein native states and large-scale conformational changes with a modified generalized born model*. Proteins: Struct., Funct., Bioinf., 55:383–394, 2004.
- [215] Harvey, S. C., R. K. Tan, and III T. E. Cheatham: *The flying ice cube: Velocity rescaling in molecular dynamics leads to violation of energy equipartition*. J. Comput. Chem., 19:726–740, 1998.
- [216] Humphrey, W., A. Dalke, and K. Schulten: *VMD – Visual Molecular Dynamics*. J. Mol. Graphics, 14:33–38, 1996.
- [217] Baker, N. A., D. Sept, S. Joseph, M. J. Holst, and J. A. McCammon: *Electrostatics of nanosystems: Application to microtubules and the ribosome*. Proc. Natl. Acad. Sci. USA, 98:10037–10041, 2001.
- [218] Miller, M. A., G. W. Han, and J. Kraut: *A cation binding motif stabilizes the compound I radical of cytochrome c peroxidase*. Proc. Natl. Acad. Sci. USA, 91:11118–11122, 1994.

- [219] Ross, W. S. and C. C. Hardin: *Ion-Induced Stabilization of the G-DNA Quadruplex: Free Energy Perturbation Studies*. J. Am. Chem. Soc., 116:6070–6080, 1994.
- [220] Jeffrey, G. A. and W. Saenger: *Hydrogen Bonding in Biological Structures*. Springer, Berlin, 1994.
- [221] Creighton, T. E.: *Proteins: Structures and Molecular Properties*. W. H. Freeman, New York, 2nd edition, 1992.
- [222] Böhm, H. J., G. Klebe und H. Kubinyi: *Wirkstoffdesign*. Spektrum, Heidelberg, 1996.
- [223] Hamelryck, T. and B. Manderick: *PDB file parser and structure class implemented in Python*. Bioinformatics, 19:2308–2310, 2003.
- [224] *Biopython*. [http://biopython.org/wiki/Main\\_Page](http://biopython.org/wiki/Main_Page).
- [225] Chemical Computing Group: *Molecular Operating Environment Manual*, 2008.
- [226] Wahlström, R.: *Validation of docking performance in the context of a structural water molecule using model system*. Master's thesis, Linköping University, 2009.
- [227] Wade, R. C., M. H. Mazar, J. A. McCammon, and F. A. Quiñocho: *A Molecular Dynamics Study of Thermodynamic and Structural Aspects of the Hydration of Cavities in Proteins*. Biopolymers, 31:919–931, 1991.
- [228] Helms, V. and R. C. Wade: *Thermodynamics of water mediating protein-ligand interactions in cytochrome P450cam: a molecular dynamics study*. Biophys. J., 69:810–824, 1995.
- [229] Roux, B., M. Nina, R. Pomes, and J. C. Smith: *Thermodynamic stability of water molecules in the bacteriorhodopsin proton channel: a molecular dynamics free energy perturbation study*. Biophys. J., 71:670–681, 1996.
- [230] Hamelberg, D. and J. A. McCammon: *Standard Free Energy of Releasing a Localized Water Molecule from the Binding Pockets of Proteins: Double-Decoupling Method*. J. Am. Chem. Soc., 126:7683–7689, 2004.
- [231] Lu, Y., C. Y. Yang, and S. Wang: *Binding Free Energy Contributions of Interfacial Waters in HIV-1 Protease/Inhibitor Complexes*. J. Am. Chem. Soc., 128:11830–11839, 2006.
- [232] Olano, L. R. and S. W. Rick: *Hydration Free Energies and Entropies for Water in Protein Interiors*. J. Am. Chem. Soc., 126:7991–8000, 2004.
- [233] Yu, H. and S. W. Rick: *Free Energies and Entropies of Water Molecules at the Inhibitor-Protein Interface of DNA Gyrase*. J. Am. Chem. Soc., 131:6608–6613, 2009.

- [234] Barillari, C., J. Taylor, R. Viner, and J. W. Essex: *Classification of Water Molecules in Protein Binding Sites*. J. Am. Chem. Soc., 129:2577–2587, 2007.
- [235] Sarkisov, G. N., V. G. Dashevsky, and G. G. Malenkov: *The thermodynamics and structure of liquid water: The Monte Carlo method*. Mol. Phys., 27:1249–1269, 1974.
- [236] Ben-Naim, A. and Y. Marcus: *Solvation thermodynamics of nonionic solutes*. J. Chem. Phys., 81:2016–2027, 1984.
- [237] Hermans, J., A. Pathiaseril, and A. Anderson: *Excess Free Energy of Liquids from Molecular Dynamics Simulations. Application to Water Models*. J. Am. Chem. Soc., 110:5982–5986, 1988.
- [238] Jorgensen, W. L., J. F. Blake, and J. K. Buckner: *Free energy of TIP4P water and the free energies of hydration of CH<sub>4</sub> and Cl<sup>-</sup> from statistical perturbation theory*. Chemical Physics, 129:193–200, 1989.
- [239] Cappel, D.: *Simulationmethoden zur Untersuchung von Wasserbeiträgen in Protein-Ligand Wechselwirkungen: HIV-1 Protease und Cytochrom c Peroxidase als Testsysteme*. Diplomarbeit, Universität Marburg, 2006.
- [240] Lazaridis, T. [http://www.charmm.org/ubbthreads/ubbthreads.php?ubb=showflat&Number=1759&site\\_id=1#import](http://www.charmm.org/ubbthreads/ubbthreads.php?ubb=showflat&Number=1759&site_id=1#import).
- [241] GNU Fortran. <http://gcc.gnu.org/fortran/>.
- [242] PGI Fortran. <http://www.pgroup.com/>.