

# **Modellierung von Protein-Ligand-Komplexen: Validierung und Vorhersage von ausgewählten Fällen**

## **Dissertation**

zur

Erlangung des Doktorgrades (Dr. rer. nat.)

der

Mathematisch-Naturwissenschaftlichen Fakultät

der

Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von

Astrid Maaß

aus

Karlsruhe

Bonn, März 2003

Angefertigt mit Genehmigung der Mathematisch-Naturwissenschaftlichen  
Fakultät der Rheinischen Friedrich-Wilhelms-Universität Bonn

1. Referent: Prof. Dr. Christa E. Müller
2. Referent: Prof. Dr. Michael Gütschow

Tag der Promotion: 7. April 2003

Die vorliegende Arbeit wurde in der Zeit von Februar 1999 bis Dezember 2002 unter der Leitung von Frau Prof. Dr. Christa E. Müller am Pharmazeutischen Institut der Universität Bonn in Poppelsdorf sowie an der ehemaligen Gesellschaft für Mathematik und Datenverarbeitung, inzwischen Fraunhofer-Institutszentrum Birlinghoven in Sankt Augustin, angefertigt.

Mein besonderer Dank gilt Frau Prof. Dr. Christa E. Müller für die Ermöglichung dieser Arbeit, die wissenschaftliche Betreuung und ihre Unterstützung.

Ebenso möchte ich mich bei Herrn Prof. Dr. Michael Gütschow für die freundliche Übernahme des Korreferats bedanken.



## Vorwort des Institutsleiters

Die Suche nach Wirkstoffen in der wissenschaftlichen und pharmazeutischen Forschung erfolgt heute zunehmend nicht nur experimentell, sondern auch rechnergestützt mit bioinformatischen Methoden. Eine konsequente Begleitung von Labor-Experimenten durch Computer für den gesamten Prozess der Wirkstoffsuche ist jedoch noch nicht verwirklicht. Das Beispiel der molekularen Modellierung konkreter Probleme macht den möglichen Nutzen eines engen Zusammenspiels beider Zugänge deutlich.

Diese Dissertation entstand am Fraunhofer Institut für Algorithmen und Wissenschaftliches Rechnen (SCAI) in der Abteilung für Bioinformatik sowie am Pharmazeutischen Institut der Rheinischen Friedrich-Wilhelms-Universität Bonn. Das Fraunhofer Institut ist unter anderem auf die Entwicklung bioinformatischer Werkzeuge zur Analyse experimenteller Daten, zur Vorhersage von Protein- bzw. Komplex-Strukturen und zur allgemeinen Simulation von Molekülen ausgerichtet.

Bei dem experimentell orientierten Pharmazeutischen Institut liegt der Schwerpunkt auf dem Entwurf, der Synthese und der Untersuchung von potentiellen Wirkstoffen. In dieser Arbeit konnten also die Erfahrungen und Ressourcen aus beiden Bereichen genutzt werden.

So sind hier dreidimensionale Modelle von experimentell nicht fassbaren Komplexen aus Proteinen und Substraten unmittelbar vor der enzymatischen Reaktion berechnet worden, ebenso Komplexe aus Liganden mit Rezeptoren, deren Strukturen bislang nicht bestimmbar sind. In SCAI entwickelte Software, insbesondere das Docking-Programm FLEXX und der Poisson-Boltzmann-Solver in CamLab, wurde durch die Anwendung auf konkrete Systeme an den experimentellen Daten validiert. Es zeigte sich, dass erst die Kombination der verschiedenen Programme bzw. der ihnen zugrunde liegenden Modelle und Methoden zuverlässige Aussagen über das modellierte System zulässt.

Die Rechnungen demonstrieren die Leistungsfähigkeit, aber auch die Grenzen der heutigen Simulationsprogramme. Vor allem aber verschaffen solche Simulationen dem Experimentator Zugang zu sonst nicht beobachtbaren molekularen Verhältnissen. Mit dieser gewonnenen Information – sowie der weiterhin unerlässlichen menschlichen Fantasie und Expertise – eröffnen sich neue Ansatzpunkte zur Planung weiterer Erfolg versprechender Experimente.

Ulrich Trottenberg



## Kurzfassung

Die vorliegende Arbeit behandelt die Rechner-gestützte Modellierung ausgewählter Protein-Ligand-Komplexe. Ausgangspunkt für jede Modellierung sind Röntgen-Kristall-Strukturen.

Im ersten Fall werden Komplexe der HIV-Protease hinsichtlich der Bindestärke zu verschiedenen Liganden untersucht. MD-Simulationen eignen sich zwar prinzipiell zur Bestimmung von Bindestärken, sind aber sehr aufwändig. In der Hoffnung, einen Weg zur schnellen Abschätzung von Bindestärken zwischen verschiedenen Liganden und der HIV-Protease zu finden, wird ein stark vereinfachender Ansatz gewählt (*continuum solvent*-Modell, relative Bindestärken). Die molekularen Strukturen und Zusammenhänge lassen sich zwar qualitativ gut nachzeichnen, quantitative Aussagen sind jedoch nicht mit der erwünschten Genauigkeit zu erzielen.

Im zweiten Fall wird die Arbeitsweise zweier verwandter Enzyme (Phenylalanin-Hydroxylase und Tyrosin-Hydroxylase) durch Erzeugung von Substrat- und Inhibitor-Komplexen studiert. Die Ergebnisse ermöglichen die Interpretation vieler, teils widersprüchlicher experimenteller Befunde (Röntgen-Struktur-Daten, NMR-Studien und kinetische Messungen) nach einem einheitlichen Konzept. Auf Grundlage der berechneten Strukturen können konkrete Reaktionsmechanismen vorgeschlagen werden, die praktisch überprüft werden können.

Zuletzt werden drei-dimensionale Modelle von zwei G-Protein-gekoppelten Adenosin-Rezeptoren erstellt, jeweils ausgehend von der Kristall-Struktur des eng verwandten Rhodopsins. Diese Homologie-basierte Modellierung der beiden Rezeptoren ist notwendig, da sie als integrale Membran-Proteine strukturell nur sehr schwer charakterisierbar sind. Die erzeugten Strukturen passen gut zu den verfügbaren experimentellen Daten. Gleichzeitig liefern die Modelle Hinweise auf die Ursachen für die trotz der engen Verwandtschaft beider Rezeptoren große Diskrepanz in ihren Ligand-Affinitäten. Auf der Grundlage dieser Modelle können nicht nur Crosslinker-Moleküle entworfen, sondern auch für Mutagenese-Studien potentiell relevante Aminosäuren identifiziert werden, sodass die Qualität der Modelle experimentell überprüfbar ist.





# Inhaltsverzeichnis

<b>1. Molekularbiologische Grundlagen</b>	<b>1</b>
1.1. Einleitung . . . . .	1
1.1.1. Zellen . . . . .	1
1.2. Proteine . . . . .	3
1.2.1. Aufbau und Eigenschaften . . . . .	3
1.2.2. Räumliche Struktur . . . . .	6
1.2.3. Exkurs: Die wässrige Umgebung . . . . .	8
1.3. Wirkstoff-Design . . . . .	9
1.4. Verfahren zur Aufklärung von Strukturen . . . . .	9
1.4.1. Röntgen-Kristall-Struktur-Analyse . . . . .	10
1.4.2. NMR-Analyse . . . . .	11
1.4.3. Crosslinking und Massenspektroskopie . . . . .	12
1.4.4. Mutagenese-Studien . . . . .	12
1.5. Verfahren zur Vorhersage von 3D-Strukturen . . . . .	12
1.5.1. Struktur-Vergleiche . . . . .	14
1.5.2. Homologie-basierte Struktur-Vorhersage . . . . .	15
1.5.3. Konstruktion von Komplexen: Docking . . . . .	17
<b>2. Theoretische Grundlagen</b>	<b>19</b>
2.1. Proteine in wässriger Lösung . . . . .	19
2.1.1. Moleküle und Zustände . . . . .	19
2.1.2. Wechselwirkungen zwischen Molekülen . . . . .	20
2.1.3. Verhalten von Molekülen im Kollektiv . . . . .	21
2.1.4. Komplexierungsreaktionen . . . . .	21
2.1.5. Absolute Bindestärke eines Komplexes . . . . .	22
2.2. Simulation von Molekülen als klassische Teilchen . . . . .	23
2.2.1. Das Kraftfeld-Modell . . . . .	24
2.3. Simulation des Wassers als homogenes Medium . . . . .	29
2.3.1. Die Poisson-Gleichung . . . . .	30
2.4. Statischer Ansatz: Verzicht auf Dynamik . . . . .	34
2.4.1. Relative Bindestärke von Komplexen . . . . .	35
2.5. Zusammenfassung . . . . .	36
2.5.1. Konzept . . . . .	38

<b>3. HIV-Protease-Inhibitoren</b>	<b>39</b>
3.1. Einleitung . . . . .	39
3.2. Allgemeines über HIV . . . . .	40
3.2.1. Vermehrungszyklus . . . . .	40
3.2.2. Behandlungsstrategien . . . . .	42
3.3. Zielenzym HIV-1-Protease . . . . .	44
3.3.1. Eigenschaften und Eigenarten . . . . .	44
3.3.2. Substratbindung . . . . .	45
3.3.3. Modellierung von Inhibitor-Komplexen . . . . .	46
3.4. Studie eines NMR-Ensembles . . . . .	48
3.4.1. Durchführung . . . . .	48
3.4.2. Ergebnisse . . . . .	50
3.4.3. Diskussion . . . . .	58
3.4.4. Schlussfolgerungen . . . . .	60
3.5. Komplex-Konstruktion . . . . .	62
3.5.1. Durchführung . . . . .	62
3.5.2. Ergebnisse . . . . .	64
3.5.3. Diskussion . . . . .	74
3.5.4. Schlussfolgerungen . . . . .	75
3.6. Relative Bindestärken für eine Serie von 10 Liganden . . . . .	76
3.6.1. Durchführung . . . . .	76
3.6.2. Ergebnisse . . . . .	79
3.6.3. Diskussion . . . . .	85
3.6.4. Zusammenfassung und Ausblick . . . . .	95
<b>4. Tyrosin-Hydroxylase und Phenylalanin-Hydroxylase</b>	<b>97</b>
4.1. Allgemeines . . . . .	97
4.1.1. Die Enzyme PAH und TH . . . . .	98
4.1.2. Fragestellungen . . . . .	106
4.2. Durchführung . . . . .	107
4.2.1. Komplex-Konstruktion . . . . .	109
4.2.2. Auswertung . . . . .	113
4.3. Ergebnisse . . . . .	114
4.3.1. Validierung . . . . .	114
4.3.2. Vorhersage . . . . .	122
4.4. Diskussion . . . . .	130
4.4.1. Validierung . . . . .	130
4.4.2. Vorhersage . . . . .	135
4.4.3. Biologische Konsequenzen . . . . .	139
4.5. Zusammenfassung . . . . .	145

<b>5. GPCRs: Adenosin-Rezeptoren</b>	<b>147</b>
5.1. Allgemeines . . . . .	147
5.1.1. G-Protein gekoppelte Rezeptoren . . . . .	147
5.2. Adenosin-Rezeptoren . . . . .	153
5.2.1. Extrazelluläres Adenosin . . . . .	153
5.2.2. Eigenschaften der Rezeptoren . . . . .	153
5.2.3. Motivation . . . . .	159
5.3. Homologie-basierte Modellierung von AR-Komplexen . . . . .	161
5.3.1. Multiples Sequenz-Alignment . . . . .	161
5.3.2. Threading . . . . .	163
5.3.3. Vervollständigung der 3D-Modelle . . . . .	164
5.3.4. Diskussion . . . . .	167
5.4. Komplex-Konstruktion . . . . .	171
5.4.1. Ergebnisse . . . . .	174
5.4.2. Diskussion . . . . .	177
5.5. Zusammenfassung und Ausblick . . . . .	180
<b>A. Anhang</b>	<b>185</b>
A.1. Abkürzungsverzeichnis . . . . .	185
A.2. Sequenz-Alignments . . . . .	187
A.2.1. Multiples Sequenz-Alignment . . . . .	187
A.2.2. Editiertes Sequenz-Alignment . . . . .	195
<b>Literaturverzeichnis</b>	<b>197</b>
<b>Danksagung</b>	<b>207</b>
<b>Lebenslauf</b>	<b>209</b>



# 1. Molekularbiologische Grundlagen

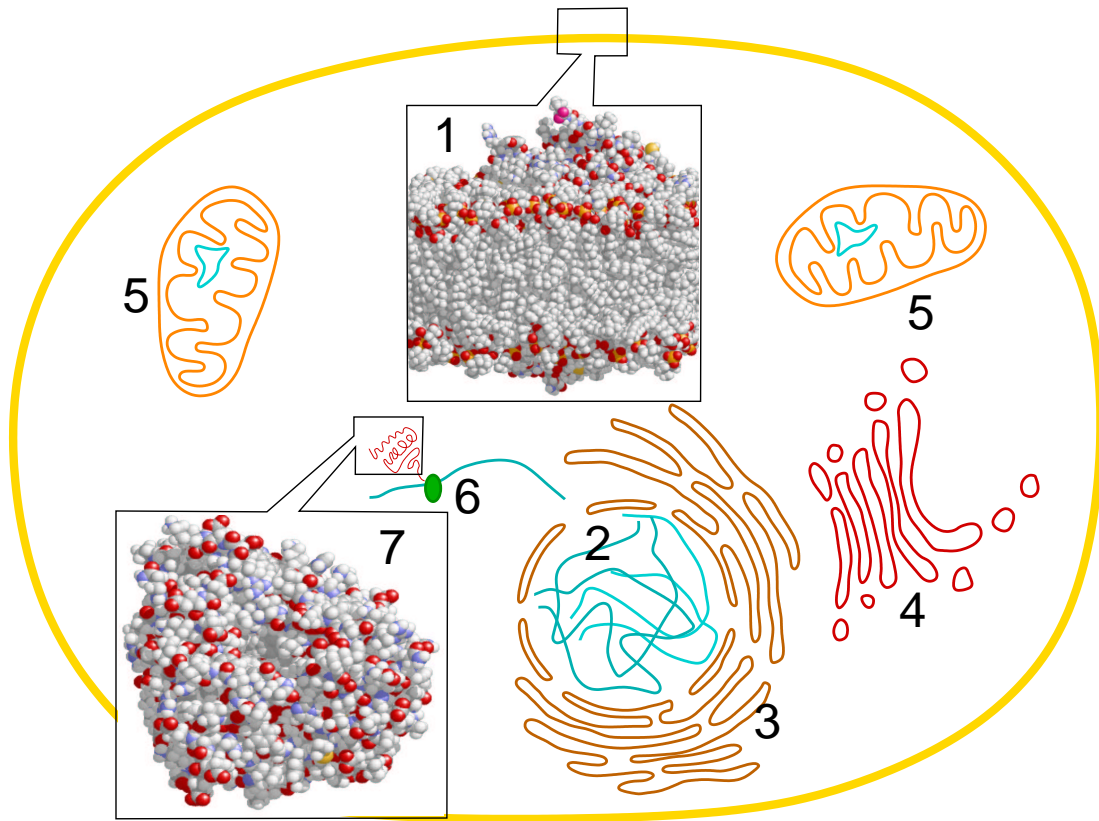
## 1.1. Einleitung

Diese Arbeit behandelt Proteine und ihre Wechselwirkungen mit anderen Molekülen auf atomarer Ebene. Proteine sind grundlegend an allen Vorgängen des Lebens beteiligt. Daher werden zunächst die molekularbiologischen und theoretischen Grundlagen ihrer Wirkungsweise vorgestellt.

### 1.1.1. Zellen

Die kleinste lebende Einheit ist die Zelle. Eine Zelle enthält verschiedene organische Verbindungen, die sich zu komplexen Strukturen, den Organellen, organisieren. Die zelluläre Organisation erhält sich selbst unter kontinuierlichem Austausch von Stoffen und Energie mit der Umgebung aufrecht. Zellen können mit dem Mikroskop beobachtet werden, wobei auch subzelluläre Strukturen, Organellen, erkennbar sind (s. Abb. 1.1).

- Eine *Lipid-Membran* von ca. 40 nm Dicke umschließt den Zellinhalt, der im wesentlichen aus Erbsubstanz und Stoffwechsel-Apparat besteht. Über diese Membran werden Stoffe und Informationen zwischen Innen- und Außenraum ausgetauscht. Solche Lipid-Membranen unterteilen aber auch das Zellinnere in verschiedene Kompartimente:
  - Der Zellkern enthält das in eine Doppelmembran eingehüllte Erbgut.
  - Das endoplasmatische Retikulum ist ein Netz aus feinen, membranumgrenzten Kanälen, die mit der Kernhülle in Verbindung stehen.
  - Der Golgi-Apparat besteht aus Stapeln von membranumschlossenen, Flüssigkeits-gefüllten Scheiben, von deren Rand sich kleine Bläschen abschnüren.
  - Mitochondrien sind „Zellen in einer Zelle“ und als solche von zwei Membranen umhüllt, von der eigenen und von einer der Wirtszelle (Endosymbionten-Hypothese).
- Das *Erbgut* speichert die Informationen zum Aufbau von Proteinen in Form langer DNS-Stränge. Diese Stränge sind Ketten aus Nucleinsäuren. Es kommen vier verschiedene Nucleinsäure-Bausteine vor; die jeweilige Reihenfolge ist die genetische Information. Vor einer Zellteilung wird das Erbgut verdoppelt, sodass jede Tochterzelle



*Abbildung 1.1.:* Prinzipieller Aufbau der Zelle: Eine Doppelschicht aus Lipiden umschließt als Zellmembran (1) den gesamten Inhalt. Der Zellkern (2) enthält das Erbgut, seine Kern-Membran steht in Verbindung mit dem Endoplasmatischen Retikulum (3). Im Golgi-Apparat (4) findet die Nachbearbeitung von Proteinen statt. In den Mitochondrien (5) läuft ein großer Teil des Energie-Stoffwechsels ab. An den Ribosomen (6) des Cytosols werden die Proteine (7) hergestellt. Die Ausschnitte zeigen submikroskopische Details aus der Membran (Lipid-Doppelschicht mit integralem Protein) [59] und dem Cytosol (Enzym in wässriger Lösung) [53].

wieder die gleichen Proteine daraus ableiten kann wie die Stammzelle. Bei jedem Kopiervorgang können Veränderungen im neuen DNS-Strang entstehen, die permanent sind und wiederum kopiert werden.

- Die Erbinformation wird genutzt für die Synthese von Proteinen. Dazu werden von Abschnitten der DNS kurzlebige Kopien (Boten-RNS) angefertigt und an den *Ribosomen* in Proteine übersetzt. Diese Proteine üben definierte Wirkungen auf ihre unmittelbare Umgebung aus.
- Katalytisch aktive *Proteine* bilden den Stoffwechsel-Apparat der Zelle. Als Stoffwechsel oder Metabolismus bezeichnet man die chemische Umwandlung von Stoffen innerhalb der Zelle. Metabolite sind niedermolekulare organische Verbindungen, von

denen entweder die chemische Energie (aus dem Abbau zu  $\text{CO}_2$  und  $\text{H}_2\text{O}$ ) oder das Molekül-Gerüst zum Erhalt der Zell-Struktur genutzt werden. Viele chemische Reaktionen laufen unter physiologischen Bedingungen nur mittels Katalyse durch Proteine ab. Andere Proteine wiederum ermöglichen den Informationsfluss innerhalb und zwischen Zellen.

Ihre Eigenschaften als Katalysatoren oder Rezeptoren machen Proteine zu Ansatzpunkten für eine medikamentöse Therapie, da man über sie in das Stoffwechsel-Geschehen eines Organismus eingreifen kann.

## 1.2. Proteine

### 1.2.1. Aufbau und Eigenschaften

Proteine sind kettenförmige Polymere aus  $\alpha$ -Aminosäuren (AS) mit diversen Substituenten (Seitenketten,  $\text{R}_1$ ,  $\text{R}_2$ ,  $\text{R}_3$ ) am  $\text{C}_\alpha$ -Atom (s. Abb. 1.2 bzw. 1.3).

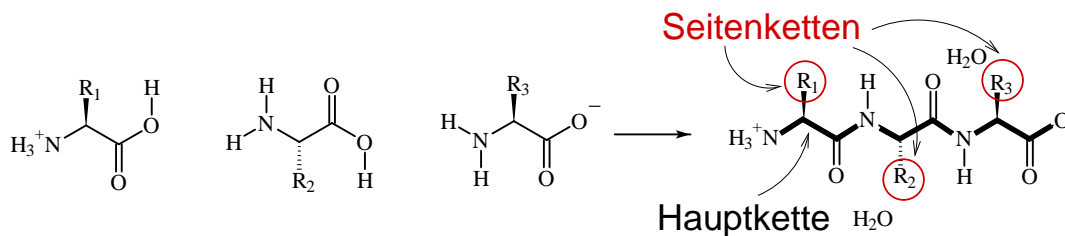


Abbildung 1.2.: Unter Abspaltung von Wasser können AS zu langen Peptid-Ketten verknüpft werden. Die Seitenketten stehen von der Hauptkette ab.

Die Amino-Gruppe eines Monomers wird jeweils mit der Säure-Gruppe des nächsten Monomers über eine sogenannte peptidische Bindung verknüpft. Die Zahl der auf diese Weise verknüpften AS bestimmt die Länge der Hauptkette. Die Hauptkette eines Proteins besitzt einen sog. N-Terminus mit einer freien  $\text{NH}_3^+$ -Gruppe und entsprechend einen C-Terminus mit freier  $\text{CO}_2^-$ -Gruppe. Die AS-Sequenz wird vereinbarungsgemäß vom N-Terminus her angegeben, an dem auch die Biosynthese beginnt. Die Abfolge der miteinander verketteten AS ist in der Nucleinsäure-Sequenz des Erbguts festgelegt. Der genetische Code umfasst ein Repertoire von 20 verschiedenen AS. Ein Triplet von Nucleinsäuren in der DNS codiert jeweils für eine bestimmte AS.

Je nachdem, welcher DNS-Abschnitt übersetzt worden ist, unterscheiden sich die Proteine voneinander in der Anzahl und in der Reihenfolge der miteinander verknüpften AS. Folglich unterscheiden sie sich auch in ihrer Struktur und Funktion. Proteine, die nur in einigen wenigen AS voneinander abweichen, und bei denen dieser Unterschied auf einem

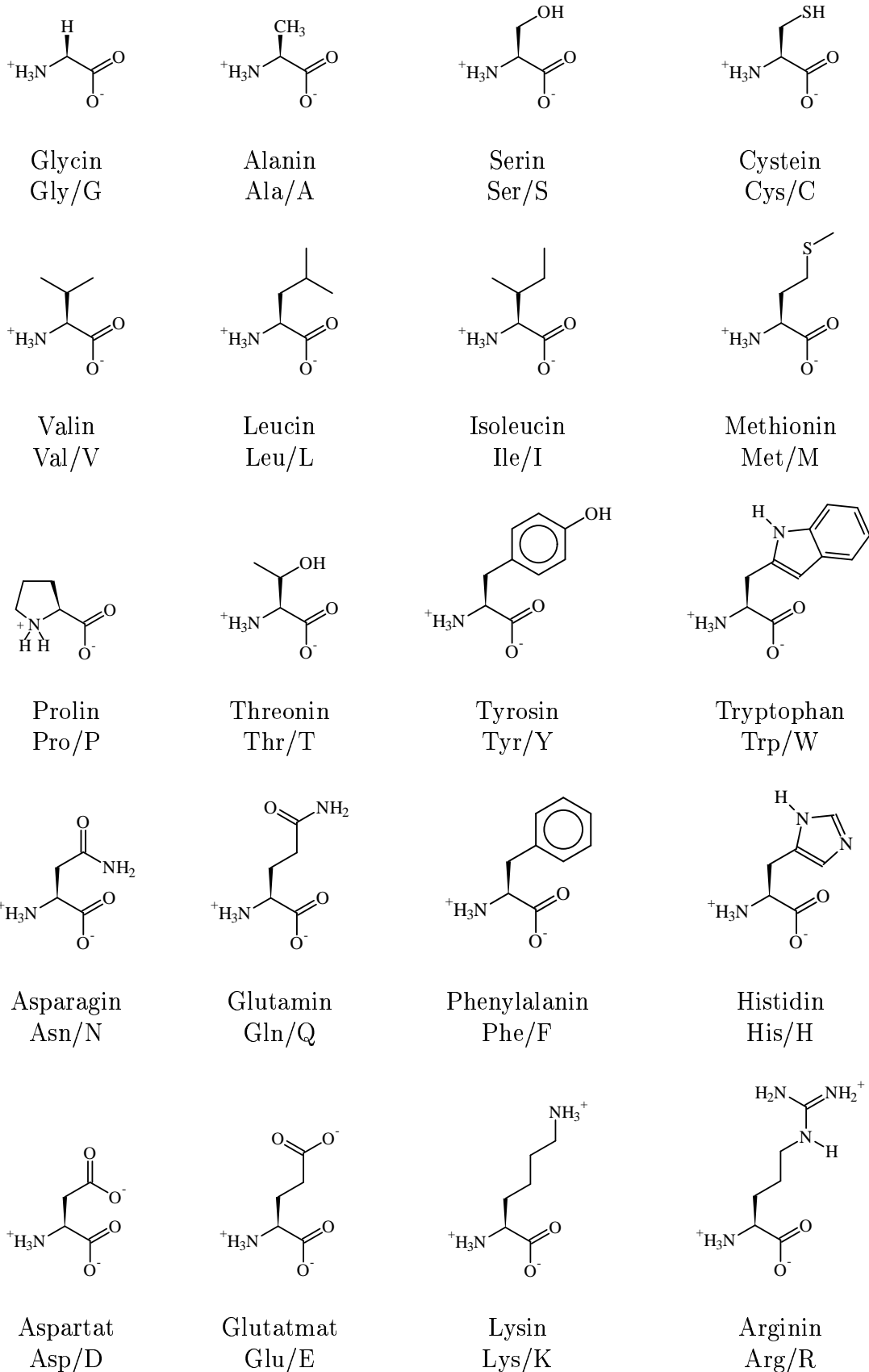


Abbildung 1.3.: Die 20 natürlichen Aminosäuren (mit Abkürzungen im Drei- bzw. Ein-Buchstaben-Code)



Kopierfehler bei der DNS-Verdopplung beruht, bezeichnet man als *Mutanten*. Eine Mutation wird in Bezug zur Sequenz des Wild-Typs (Referenz) angegeben durch die ursprüngliche AS, ihre Position innerhalb der Sequenz und die neue AS.

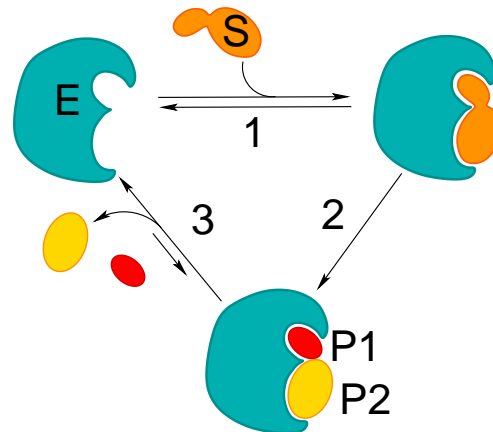
Trotz ihres primär linearen Aufbaus sind Proteine massiv und kompakt geformt. Die Hauptkette windet und faltet sich, sodass sich Abschnitte der Hauptkette und/oder der Seitenketten untereinander berühren. Der räumliche *Verlauf der Hauptkette* legt die globale Struktur des Proteins fest. Die Positionen der Seitenketten bestimmen die lokale Oberflächen-Struktur und können oft ohne Einfluss auf den Verlauf der Hauptkette variieren.

## Funktion

Unter physiologischen Bedingungen liegt ein Protein in einer ganz bestimmten globalen Struktur vor. Diese sogenannte *native Struktur* beruht auf dem individuellen Faltungsprozess und auf der AS-Sequenz; sie bedingt die einmalige Funktion eines Proteins in der Zelle: Proteine treten mit anderen Molekülen in Kontakt, entfalten ihre spezielle Wirkung und lösen sich wieder von ihrem Gegenüber (s. Abb. 1.4).

Abbildung 1.4:

1. Das Enzym (E) bindet reversibel sein Substrat (S).
  2. Im Komplex mit dem Enzym findet die meist irreversible chemische Umsetzung des Substrates statt.
  3. Die Produkte (P1 und P2) lösen sich wieder vom Protein.
- Ein Teil der Oberfläche des Enzyms ist dem Substrat komplementär (Bindetasche).



Sie bilden also erst einen *Komplex*, bevor eine chemische Reaktion stattfindet, d. h. sie gehen reversible, nicht kovalente Bindungen mit speziellen Partner-Molekülen ein. Die kleinere Komponente eines solchen Komplexes wird generell als *Ligand* bezeichnet. Falls ein oder mehrere Ligand-Moleküle durch das Protein chemisch umgewandelt werden, handelt es sich bei dem Protein um ein *Enzym* und bei den Liganden um *Substrate*; falls eine solche Reaktion durch den Liganden verhindert wird, bezeichnet man ihn als *Inhibitor* des Enzyms.

## Protein-Ligand-Wechselwirkungen

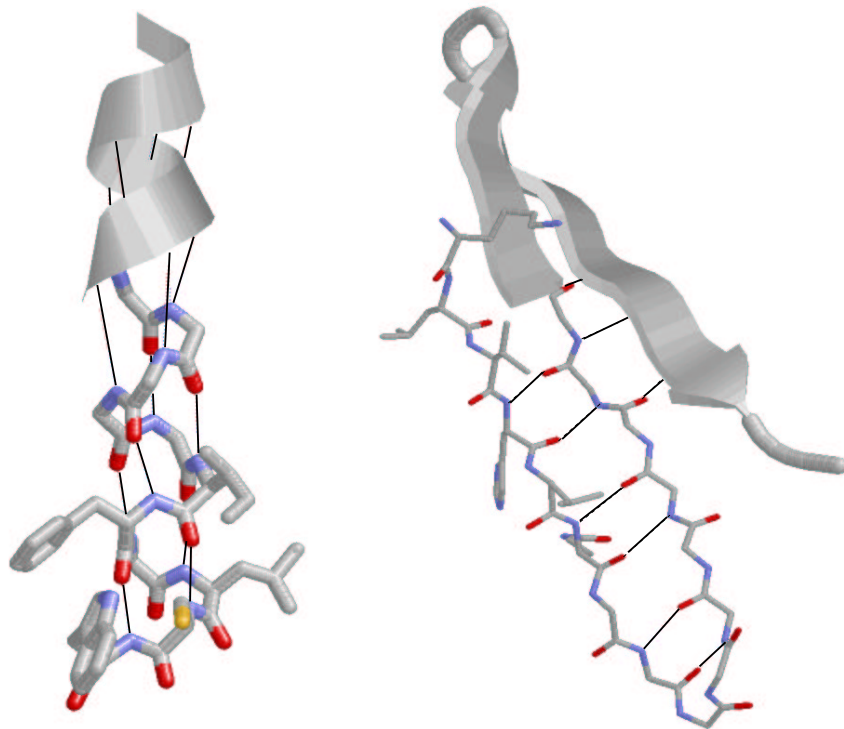
Die bindenden Kräfte zwischen Ligand und Protein sind schwach, da keine kovalenten Bindungen ausgebildet werden. Um dennoch einen stabilen Komplex zu bilden, müssen sich die anziehenden Wechselwirkungen über eine möglichst große gemeinsame Kontaktfläche

zwischen den Komponenten summieren. Meist sind die Komponenten eines Komplexes einander räumlich und elektrostatisch komplementär (Schlüssel-Schloss-Hypothese). Jede Abweichung von der gegenseitigen Entsprechung verringert die Stärke der Bindung zwischen den Komponenten. Die *Bindetasche* ist diejenige Region eines Proteins, deren Oberflächenstruktur zu einem speziellen Liganden passt. Oft ist die Bindetasche von unpolaren Resten ausgekleidet. Innerhalb der Bindetasche eines Enzyms befindet sich das *katalytische/aktive Zentrum*. Dabei handelt es sich um eine Gruppe von Atomen, die einen Weg eröffnet für eine sonst unmögliche, in Gegenwart des Proteins aber spontan ablaufende Umwandlung der Substrate. Die Oberfläche der Bindetasche ist meist so gestaltet, dass sie besonders den Übergangszustand, den die Substrate während der Reaktion durchlaufen, stabilisiert. Dadurch, dass nur Moleküle, die genau in die Bindetasche eines Enzyms passen, umgesetzt werden, bleiben andere Verbindungen auch trotz funktioneller Ähnlichkeiten unverändert. So können die verschiedenen Moleküle aus dem zellulären Stoffgemisch selektiv metabolisiert werden.

### 1.2.2. Räumliche Struktur

Die Haupt- und Seitenketten eines Proteins sind wegen der vorherrschenden Einfach-Bindungen, mit denen die Atome untereinander verknüpft sind, leicht drehbar. Ein Protein kann daher theoretisch fast unendlich viele räumliche Strukturen einnehmen. Trotzdem faltet sich die Hauptkette eines Proteins in der Zelle zu seiner speziellen natürlichen Struktur.

*Abbildung 1.5:* Sekundär-Struktur-Elemente werden durch H-Brücken (schwarze Linien) zwischen den Hauptketten stabilisiert. Die Seitenketten stehen bei  $\alpha$ -Helices radial von der Helix-Achse ab und bei  $\beta$ -Faltblättern senkrecht von der Faltblatt-Ebene (alternierend).



Der räumliche Aufbau eines Proteins wird durch den Verlauf der Hauptkette beschrieben (s. Abb. 1.5). Deren individuelle AS-Sequenz wird als *Primär-Struktur* bezeichnet. Kürzere Abschnitte der Hauptkette organisieren sich unter gleichförmiger Anordnung unmittelbar aufeinander folgender Peptid-Gruppen zu *Sekundär-Struktur-Elementen*. Bei einer  $\alpha$ -*Helix* windet sich die Hauptkette um eine imaginäre Achse, wobei die Carbonyl-Gruppe einer AS eine H-Brücke zur NH-Gruppe der jeweils dritt- oder viertnächsten AS ausbildet. Die Seitenketten stehen etwa senkrecht von der Helix-Achse ab. Bei einem  $\beta$ -*Faltblatt* ist die Hauptkette gestreckt, sodass die Seitenketten abwechselnd von der einen oder anderen Seite weg weisen. Die parallele oder antiparallele Anordnung zweier solcher Stränge wird durch H-Brücken zwischen den Hauptketten stabilisiert. Innerhalb eines solchen Sekundär-Struktur-Elements treten bestimmte AS besonders häufig auf. Sequenz-Abschnitte, die nicht zu diesen gut definierbaren Elementen gehören, z. B. Übergänge zwischen zwei definierten Bereichen oder längere *Schlaufen* oder Knäuel, werden im Fachjargon mit den englischen Begriffen *loops* oder *coils* bzw. *random coils* bezeichnet.

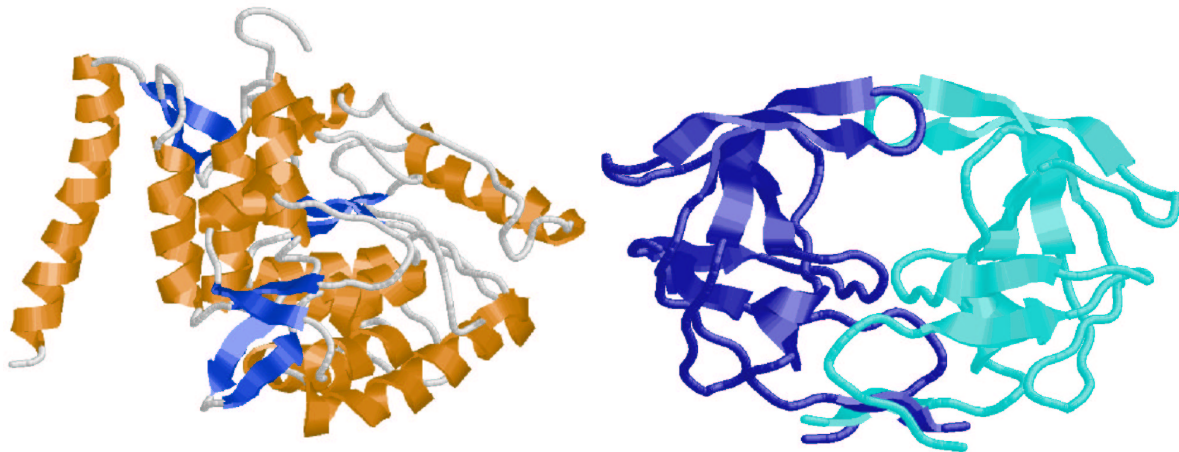


Abbildung 1.6.: Die Tertiär-Struktur ist die Gesamt-Architektur des Proteins aus Sekundär-Struktur-Elementen (links am Beispiel der Tyrosin-Hydroxylase [53]). Die Quartär-Struktur bezeichnet die Zusammensetzung aus mehreren Monomeren (rechts, am Beispiel der dimeren HIV-Protease [1]).

Die *Tertiär-Struktur* gibt den räumlichen Verlauf der gesamten Hauptkette an und schließt die Anordnung von Abschnitten mit Sekundär-Struktur-Elementen ein. Häufig weisen Proteine trotz unterschiedlicher Sequenzen ähnliche Tertiär-Strukturen auf. Die Sekundär-Struktur-Elemente organisieren sich zu zusammenhängenden *Domänen*, d. h. funktionellen Untereinheiten innerhalb des Proteins. Ein Protein besitzt mindestens eine Domäne. Bilden zwei oder mehr Hauptketten einen funktionellen Komplex, spricht man von *Quartär-Struktur*. Ein solcher Komplex besitzt meist mehrere Domänen.

Da Struktur und Funktion eng miteinander gekoppelt sind, werden in Abschnitt 1.4 und 1.5 Methoden zur Struktur-Aufklärung und -Vorhersage behandelt.

### 1.2.3. Exkurs: Die wässrige Umgebung

Die Faltung der Hauptkette eines Proteins geht zurück auf Wechselwirkungen der Haupt- und Seitenketten untereinander und auf Wechselwirkungen mit der Umgebung. Proteine können in der unpolaren Zellmembran, in der wässrigen Phase des Cytosols oder an der Grenzfläche zwischen diesen beiden Medien angesiedelt sein. Wasser-Moleküle sind polar und leicht beweglich und können sich daher im Feld einer elektrischen Ladung immer in der energetisch günstigsten Orientierung ausrichten. Das heißt, dicht neben einer Ladung wird sich der gegensinnig geladene Pol des Wasser-Moleküls platzieren. Daher herrschen zwischen polaren Molekülen und Wasser zusätzlich zur universellen, aber schwachen van-der-Waals-Anziehung (s. Abschnitt 2.1.2) starke elektrostatische, immer *anziehende* Wechselwirkungen. Diese herrschen ebenso zwischen den Wasser-Molekülen untereinander; darauf beruht die hohe beobachtbare Kohäsion des Wassers. Die O-Atome zweier Wasser-Moleküle sind immer über ein H-Atom miteinander verbrückt: Ein Wasser-Molekül, der *Donor*, stellt sein H-Atom dem zweiten Wasser-Molekül, dem *Akzeptor*, anteilig zur Verfügung. Durch seine zwei H-Atome und die zwei freien Elektronen-Paare am O-Atom kann ein Wasser-Molekül vier solcher *H-Brücken* gleichzeitig ausbilden. In flüssigem Wasser unterliegen die H-Brücken einem ständigen Auf- und Abbau, da ihre Bindungsenergie nur maximal 10 kcal/mol beträgt und die Moleküle sich ständig bewegen [116].

Ein gelöstes Molekül (allg.: Solut) ist umgeben von Lösungsmittel-Molekülen (allg.: Solvens). Die Schicht von Lösungsmittel-Molekülen, die unmittelbaren Kontakt zum Solut hat, wird als Solvat-Käfig bezeichnet. Bei Wasser als Lösungsmittel können die Moleküle innerhalb des Solvat-Käfigs ihre H-Brücken untereinander nicht beliebig in alle Raumrichtungen ausbilden, sondern nur entlang der Kontaktfläche und zum restlichen Wasser hin, nicht aber zum Solut. Dadurch wird die Beweglichkeit der Moleküle der Hydrat-Hülle eingeschränkt, denn es werden mangels Alternativen weniger H-Brücken gelöst (notwendige Voraussetzung für eine Bewegung und eine erneute Knüpfung) als in freiem Wasser. Ein Solut verringert somit die Bewegungsfreiheit der Solvens-Moleküle, und zwar proportional zur Größe seiner zu benetzenden Oberfläche. Dieser Verlust der Entropie des Wassers wird minimiert, indem unpolare Moleküle oder Molekülteile in Wasser zusammengedrängt werden, da sich so die zu benetzende Oberfläche verringert. Bei geladenen Molekülen kann die elektrostatische Anziehung den Entropie-Verlust überkompensieren, dadurch werden sie monomer dispers gelöst.

Die Umgebung geladener Gruppen durch eine Schicht von Wasser-Dipolen bewirkt zweierlei: Erstens, das elektrostatische Feld der geladenen Gruppe wird abgeschirmt, die Kraft der Ladung hat also eine verringerte Reichweite. Zweitens, die relativ feste Bindung der Hydratschicht erschwert eine räumliche Annäherung einer weiteren geladenen Gruppe an die erstere. Das heißt, Wasser erschwert Kontakte geladener Molekülen zueinander. Daher finden die klassischen, über Ladungen gesteuerten organisch-chemischen Reaktionen (nucleophile/elektrophile Substitutionen) in Wasser bei Raum- oder Körper-Temperatur kaum statt. Wechseln die Reaktionspartner aus dem wässrigen Medium in ein unpolares, findet eine Reaktion leichter statt. Auf diesem Prinzip beruht die enzymatische Katalyse: Innerhalb eines Proteins, d. h. unter Ausschluss von Wasser, sind die elektrostatisch gesteuerten Reaktionen wieder leicht möglich.

## 1.3. Wirkstoff-Design

Zur Behandlung von Infektionskrankheiten wird die Achilles-Ferse des Erregers als Ansatzpunkt für eine Blockade anvisiert. Dabei handelt es sich meist um ein Schlüssel-Enzym in einem seiner Stoffwechselwege. Es muss ein passender Wirkstoff eingesetzt werden, der dort (und nur dort) bindet und damit das Funktionieren dieses Enzyms verhindert. Es gibt zwei Strategien zur Identifizierung von Wirkstoffen [16]:

- Beim empirisch/heuristischen Ansatz wird (blind) erprobt, welchen Effekt ein beliebiges, verfügbares Molekül auf das Ziel-Protein bzw. den Ziel-Organismus ausübt. Aus einer großen Zahl von Molekülen können auf diese Weise nur zufällig Verbindungen identifiziert werden, die die gewünschten Eigenschaften aufweisen.
- Bei einem wissensbasierten Ansatz hingegen werden vorhandene Informationen genutzt, um entweder im Vorfeld der erwähnten Reihenuntersuchungen die Brauchbarkeit verfügbarer Moleküle als Wirkstoffe abzuschätzen oder um passende Wirkstoffe zu entwerfen.

Beim Entwurf von Wirkstoffen orientiert man sich zunächst häufig am Substrat bzw. am Übergangszustand und versucht ein Molekül zu konstruieren, das eine ähnliche Struktur besitzt, aber inert ist und nicht reagiert. Substrat und Wirkstoff nutzen die selben Ankerpunkte innerhalb des Proteins. Substrat-Imitate sind jedoch in zweierlei Hinsicht problematisch: Erstens müssen sich die natürlichen Produkte wieder vom Enzym lösen, die Wechselwirkungen zwischen Enzym und Substrat sind also von Natur aus im Hinblick auf die größte Umsatzrate optimiert, aber nicht unbedingt im Hinblick auf die größte Bindestärke. Zweitens kommen in Naturstoffen häufig stereogene Zentren vor, die aufwändige stereoselektive Synthesen bzw. Ausbeute mindernde Reinigungsschritte erfordern.

In der Praxis werden heuristische und wissensbasierte Methoden kombiniert, indem Leit-Moleküle, die erwiesenermaßen fest binden, gezielt im Hinblick auf noch größere Affinität modifiziert werden. Die Suche nach Leit-Molekülen und die Optimierung dieser Leit-Strukturen für eine Blockade des Ziel-Enzyms wird in zunehmendem Maße mit Hilfe rechnergestützter Methoden durchgeführt.

Eine wesentliche Grundlage für die rationale Wirkstoff-Entwicklung ist dabei die Aufklärung der Struktur des Ziel-Proteins und des Bindemodus der Leit-Verbindung innerhalb des Proteins. Nur diese Kenntnisse machen es möglich, Bindestärken nach einem realitätsgetreuen Modell *in silico* abzuschätzen und anschließend *in vitro* zu bestätigen.

## 1.4. Verfahren zur Aufklärung von Strukturen

Ausgehend von einem lebenden Gewebe, das eine Vielzahl von Proteinen enthält, bis hin zum Verständnis der Funktionsweise eines einzelnen Proteins aus einem solchen Gemisch, ist es ein weiter Weg. Zur Aufklärung seiner Struktur muss das Protein rein und in ausreichenden Mengen gewonnen werden. Physikalische Methoden zur Struktur-Aufklärung

(Röntgen-Kristall-Struktur-Analyse [6, 16], siehe auch <http://www-structure.llnl.gov/Xray/101index.html>, Kern-Resonanz-Spektroskopie [20], siehe auch <http://www.ch.ic.ac.uk/local/organic/nmr.html>) erfordern zwar große Mengen an gereinigtem Protein, liefern dafür aber die zuverlässigsten Atom-Positionen. Gentechnik-basierte Methoden nutzen einzelne Proteine oder ganze Zellen als Werkzeuge für die Analyse anderer Proteine [105, 135], siehe auch <http://www.zum.de/Gentechnik/>. Sie benötigen weniger Ausgangsmaterial, können aber nur Aussagen über bestimmte Bezugspunkte innerhalb des Proteins liefern. Diese Methoden werden durch theoretische Verfahren wie Molekular Modelling und Struktur-Vorhersage (s. Abschnitt 1.5) gewinnbringend ergänzt: Anhand experimentell gesicherter Eckpunkte wird die vorhergesagte räumliche Struktur eines theoretischen Modells bestätigt oder widerlegt. Lassen sich die experimentell und theoretisch erzielten Ergebnisse zu einem stimmigen Gesamtbild vereinen, so liegen Atom-Koordinaten für eine plausible Protein-Konformation vor, aus der man Struktur-Funktions-Beziehungen ableiten kann.

### 1.4.1. Röntgen-Kristall-Struktur-Analyse

Die Methode der Wahl zur Struktur-Aufklärung von Proteinen oder anderen Makromolekülen (z. B. DNS) ist die Röntgen-Kristall-Struktur-Analyse (RKSA). Das Ziel-Protein wird isoliert und in Lösung angereichert, sodass daraus ein Protein-Kristall gezüchtet werden kann. Aus dem Beugungsmuster von Röntgen-Strahlen an diesem Kristall lässt sich die räumliche Anordnung der Protein-Atome ableiten.

Röntgen-Strahlen werden an der Elektronen-Hülle jedes Atoms gestreut. Je höher die Elektronen-Dichte dort ist, desto stärker werden sie gestreut (H-Atome lassen sich daher kaum beobachten). Bei einer regelmäßigen Anordnung der Atome im Kristall-Gitter überlagern sich die gestreuten Strahlen ebenfalls regelmäßig und bilden ein Interferenz-Muster. Aus diesem Interferenz-Muster lässt sich auf die räumliche Verteilung der Elektronen-Dichte innerhalb des Kristalls rückschließen. Den lokalen Schwerpunkten dieser berechneten Elektronen-Dichte-Verteilung werden die Atom-Positionen zugeordnet, deren Gesamtheit die räumliche Struktur des Proteins ergibt.

Protein-Kristalle enthalten neben dem Protein immer auch Wasser und Pufferionen [22]. Daher nimmt man an, dass auch im Kristall die natürliche Konformation des Proteins in Lösung vorliegt. Dabei handelt es sich aber nicht um eine einzige Konformation, sondern ein ganzes *Ensemble von Konformationen*, zwar mit weitgehend übereinstimmendem Verlauf der Hauptkette, aber mit oft abweichenden Positionen der Seitenketten. Zudem schwingen die Atome durch ihre thermische Eigenbewegung um ihre Gleichgewichtspositionen. Insofern kommen die Verhältnisse im Kristall (Wasser-Gehalt, Konformationsvielfalt und Vibrations-Bewegungen) den Verhältnissen in Lösung recht nahe. Allerdings schränken gerade diese Punkte die erreichbare Genauigkeit des Verfahrens ein, das ein möglichst ideal geordnetes Kristall-Gitter voraussetzt.

Die Qualität der Kristalle bestimmt die *Auflösung*, d.h. welche Details der Elektronen-Dichte-Verteilung noch ermittelt werden können. In einigen Fällen können trotz gut aufgelöster Elektronen-Dichte bestimmte Atome nicht eindeutig zugeordnet werden: Bei freien Amid-Gruppen beispielsweise kann das vorliegende Rotamer nur anhand von hypo-

thetischen H-Brücken-Mustern bestimmt werden, denn sowohl das N-Atom, als auch das O-Atom der Amid-Gruppe besitzen ungefähr gleiche Streufaktoren und die N-gebundenen H-Atome sind ohnehin nicht „sichtbar“. Weiterhin führen Fehlorderungen, Konformationsvielfalt und Vibrationen dazu, dass nur eine mittlere Elektronen-Dichte-Verteilung registriert werden kann. Daher sind in der Regel nur Auflösungen zwischen 1.5-3.0 Å erreichbar. Sie liegen damit in der Größenordnung von Bindungslängen (zum Vergleich: C-C-Einfachbindung 1.54 Å). Oft können daher die Positionen von besonders variablen Regionen (einzelne Seitenketten oder Teile der Hauptkette) nicht angegeben werden. Die Zuverlässigkeit, mit der eine bestimmte Atom-Position ermittelt werden kann, wird über den sog. kristallographischen *Temperatur-Faktor* und die Besetzungszahl angegeben.

Nachteilig ist, dass diese Methode relativ große Mengen des zu untersuchenden Enzyms erfordert. Außerdem setzt das Verfahren voraus, dass das zu untersuchende Protein überhaupt kristallisierbar ist, was in der Praxis eine ernstzunehmende Einschränkung darstellen kann. Kristallisation bedeutet, dass sich eine große Zahl identischer Proteine so organisiert, dass alle Einzel-Proteine gleichartig ausgerichtet sind. Doch Proteine sind meist so konzipiert, dass sie sich (unter physiologischen Bedingungen) gerade *nicht* zusammenlagern, weder zu geordneten Kristallen, noch zu ungeordneten Aggregaten, um ihre Funktionsfähigkeit zu erhalten.

Traditionell werden in der Röntgen-Kristallographie Abstände in der Größenordnung von Bindungslängen (handlicher Werte wegen) in der Einheit Ångström [Å] angegeben. Dies wird im folgenden beibehalten.

### 1.4.2. NMR-Analyse

Der Gesamtspin eines Atom-Kerns ergibt sich aus der Summe der Spins der einzelnen Kern-Bestandteile. Sind Neutronen- und/oder Protonen-Zahl ungerade, resultiert ein messbarer Gesamtspin und damit ein nach außen wirksames magnetisches Moment [120]. Proteine lassen sich also auch mit magnetischer Kernspin-Resonanz-Spektroskopie (nuclear magnetic resonance, NMR) beobachten. Das hat den Vorteil, dass man ein unkristallisierbares Protein auch in wässriger Lösung untersuchen und seine Bewegungen verfolgen kann. So können mehrere Konformationen, die im dynamischen Gleichgewicht miteinander stehen, parallel registriert werden.

Die Proben-Lösung wird dazu in ein starkes Magnetfeld gebracht. Das magnetische Dipolmoment von Atom-Kernen mit messbarem Spin (z. B.  $^1\text{H}$ ,  $^{13}\text{C}$ ) kann mit zusätzlich eingestrahlten elektromagnetischen Wellen in Resonanz treten. Das erhältliche Absorptionsspektrum liefert Informationen über die Art der Atom-Kerne und ihre unmittelbare chemische Umgebung, denn die genaue Resonanz-Frequenz ist abhängig von der Art der Atom-Kerne und dem lokalen Magnetfeld, welches die Atom-Kerne umgibt. Das lokale Magnetfeld um einen Atom-Kern setzt sich zusammen aus einem äußeren, gerätebedingten Anteil und einem inneren, moleküleigenen Anteil, der auf Effekten der Elektronenhülle beruht. Die Struktur der Elektronenhülle wiederum ist abhängig von der Zahl und Art der benachbarten Atom-Kerne. So kann jedem NMR-aktiven Kern eines Moleküls eine auf der Konfiguration beruhende sog. magnetische Verschiebung zugeordnet werden. Schwierig ist

allerdings die Zuordnung von einzelnen Signalen bei großen Molekülen mit sich wiederholenden funktionellen Gruppen. Außerdem benötigt auch dieses Verfahren große Mengen des gereinigten Proteins, da die Empfindlichkeit durch die niedrige Ansprechrate der aktiven Kerne und deren geringes natürliches Vorkommen (bes. bei  $^{13}\text{C}$ -NMR) begrenzt ist.

### 1.4.3. Crosslinking und Massenspektroskopie

Bei bekannter Sequenz eines Proteins kann man über die Vernetzung von Seitenketten mit sog. „Crosslinkern“ die Entfernung zwischen diesen AS abschätzen. Crosslinker sind gestreckte Moleküle, die reaktive Enden besitzen, mit denen sie an bestimmte Gruppen auf der Protein-Oberfläche binden können. Die Länge des Crosslinkers ist bekannt. Um festzustellen, *welche* Seitenketten innerhalb der Sequenz verbunden worden sind, werden vernetzte und unvernetzte Proteine enzymatisch zerlegt. Das resultierende Fragment-Gemisch wird massenspektroskopisch analysiert. Die Fragmente lassen sich aufgrund ihrer molekularen Masse und der Spezifität der verwendeten Enzyme eindeutigen Sequenz-Abschnitten zuordnen. So kann ermittelt werden, welche Abschnitte miteinander vernetzt wurden. Vernetzte Seitenketten müssen demnach in der nativen Konformation des Proteins in einer Entfernung vorliegen, die zur Länge des Crosslinkers passt.

### 1.4.4. Mutagenese-Studien

Inzwischen sind Mutagenese-Studien üblich, um Struktur-Funktions-Beziehungen bei Proteinen zu untersuchen, von denen nur wenig strukturelle Informationen verfügbar sind [80].

Welche Seitenketten für die Funktion eines Proteins relevant sind, lässt sich daran erkennen, wie sich der absichtliche Austausch einzelner AS auf die Funktionsfähigkeit auswirkt. Trotz der Veränderung einzelner Seitenketten behält ein Protein in der Regel die native Struktur der Ursprungs-Sequenz (Wild-Typ). Vergleicht man die kinetischen Eigenschaften des Wild-Typ-Proteins mit denen der Mutanten, findet man meist, dass einige Mutationen keinen beobachtbaren Effekt auf die Funktionsfähigkeit ausüben, während andere diese drastisch beeinflussen. Die letzteren Seitenketten liegen sehr wahrscheinlich in der Bindestelle des Proteins und in enger Nachbarschaft zu katalytisch relevanten Seitenketten oder zum Substrat.

## 1.5. Verfahren zur Vorhersage von 3D-Strukturen

Alle Eigenschaften eines Moleküls beruhen auf seiner Zusammensetzung und der aktuellen räumlichen Struktur. Die Atome sind über kovalente Bindungen zwar eindeutig und fest, aber nicht starr miteinander verknüpft. Je mehr frei drehbare Bindungen vorkommen, desto größer ist die Zahl der möglichen Konformationen, die ein Molekül einnehmen kann. Die spezielle Funktion, die es im Körper ausübt, beruht aber meist auf einer bzw. einigen wenigen Grundstrukturen. Welche der sehr vielen *möglichen* Konformationen ein Protein tatsächlich einnimmt, unterliegt physikochemischen Gesetzmäßigkeiten. Statt den



natürlichen Faltungsprozess unter großem Aufwand zu modellieren, kann man die wichtigsten *Auswirkungen* dieser physikochemischen Prinzipien auf Protein-Strukturen durch die Analyse bekannter Strukturen ermitteln. So leitet man empirische Kriterien ab, mit denen eine Konformation schnell und effektiv einzuschätzen ist. Dadurch wird die riesige Zahl der potentiellen Konformationen reduziert auf wenige, wahrscheinliche Möglichkeiten.

Die Atomkoordinaten von über 15500 Proteinen sind zur Zeit (Stand April 2002) in der Protein Data Bank (PDB) verfügbar (<http://www.rcsb.org/pdb/> [40]). Sie wurden meist durch RKSA oder NMR ermittelt, es sind aber auch einige berechnete 3D-Modelle vorhanden. Auf dieser Basis wird nach Zusammenhängen zwischen Primär-, Sekundär- und Tertiär-Strukturen gesucht. Informationen, die man aus diesen Struktur-Daten gewinnen kann, sind z. B. die Häufigkeiten, mit denen bestimmte AS in Sekundär-Struktur-Elementen auftauchen, die häufigsten Rotamere von Seiten- oder Hauptketten, oder welche Faltungsklassen auftreten. Als Faltungs-Klasse werden Proteine mit vergleichbarer nativer Struktur, also ähnlichem Hauptketten-Verlauf, bezeichnet. Aus dem Quervergleich der Strukturen lassen sich empirische Regeln ableiten, nach denen vorhergesagt werden kann, welche Konformation ein Sequenz-Abschnitt mit hoher Wahrscheinlichkeit einnehmen wird.

Ob die für ein Molekül vorhergesagte räumliche Struktur tatsächlich zutrifft, hängt ab von der Größe des Moleküls und den Informationen, die zusätzlich zur bloßen chemischen Struktur verfügbar sind. Um Einsicht in die Wechselwirkungen zwischen Proteinen und ihren Liganden zu erhalten, müssen möglichst genaue Struktur-Informationen über ihre Komplexe vorliegen. Bei der computergestützten Modellierung von Komplexen aus Proteinen und Liganden müssen zuerst für das Protein Atom-Koordinaten vorhanden sein, bevor man dann einen Liganden in dessen Bindetasche platzieren kann. Man unterscheidet daher zwei Gebiete der Struktur-Vorhersage: Die *Vorhersage von Protein-Strukturen* und die *Konstruktion von Protein-Ligand-Komplexen*. Die jeweilige Problemstellung bestimmt die verwendeten Methoden und die erreichbare Genauigkeit der Vorhersagen.

- **Protein-Struktur-Vorhersage:** Ist nur die AS-Sequenz eines Proteins bekannt, vergleicht man sie mit den Sequenzen von Proteinen mit bereits bekannter Struktur (vergleichende Modellierung/Homologie-Modellierung). Bei guter Übereinstimmung mit einer Struktur-bekanntem Sequenz überträgt man deren Verlauf der Hauptkette auf die fragliche Sequenz. Der genaue Verlauf der Hauptkette und die Positionen der Seitenketten sind ungewiss; oft bleiben auch ganze Sequenz-Abschnitte, für die kein passender Vergleich zu finden ist, undefiniert. Eine anschließende Struktur-Optimierung (s. Abschnitt 2.2.1) ist erforderlich, um eine Konformation für das fragliche Protein zu erhalten, die nicht mehr von der Vorlage dominiert wird. Die Qualität der Vorhersagen hängt sehr stark vom Grad der Übereinstimmungen zwischen Struktur-Vorlage und fraglicher Sequenz ab. Existiert keine geeignete Struktur-Vorlage, können nur qualitative Vorhersagen (Abfolge von Sekundärstruktur-Elementen) über den möglichen Verlauf der Hauptkette gemacht werden.
- **Docking:** Bei der Konstruktion von Komplexen wird nach der Konformation eines bestimmten Liganden innerhalb der Bindetasche eines gegebenen Proteins gesucht,

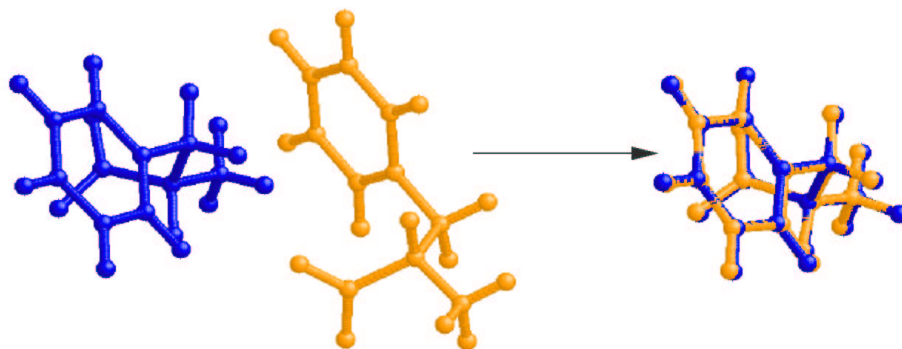
um die Bindestärke zwischen beiden abzuschätzen. Der Komplex wird ausgehend von der atomar aufgelösten Struktur eines Proteins und der chemischen Struktur eines Liganden aufgebaut. Die Konformation der Bindestasche reduziert den Konformationspielraum des Liganden erheblich, denn die aktuelle Konformation des Liganden muss zur Protein-Oberfläche passen. Durch diese Einschränkungen der Konformations-Möglichkeiten können Vorhersagen mit atomarer Auflösung gewagt werden.

Eine Übersicht über die verschiedenen Ansätze und Algorithmen zur Komplex-Konstruktion findet sich bei Taylor et al. [126]. Im folgenden werden nur diejenigen Verfahren zur Erzeugung von drei-dimensionalen Strukturen, die in dieser Arbeit eingesetzt wurden, erläutert.

### 1.5.1. Struktur-Vergleiche

Ob eine berechnete Struktur tatsächlich für Funktionsuntersuchungen brauchbar ist, hängt davon ab, ob sich das berechnete Ergebnis mit der Realität deckt. Um also zu prüfen, ob die verwendeten Verfahren Strukturen gut vorhersagen können, müssen die Koordinaten bereits bekannter Strukturen „nachträglich vorhergesagt“, d.h. reproduziert werden. Dabei wird eine hohe Übereinstimmung der berechneten mit den experimentellen Koordinaten erwartet. Die durchschnittliche Abweichung *rmsd* (*root mean square deviation*) zwischen den  $N$  Atomen, die einander in den Strukturen  $A$  und  $B$  entsprechen, ist das Maß zur Beurteilung der Übereinstimmung.

$$rmsd = \frac{1}{N} \cdot \sum_{i=1}^N \sqrt{(x_{A,i} - x_{B,i})^2 + (y_{A,i} - y_{B,i})^2 + (z_{A,i} - z_{B,i})^2} \quad (1.1)$$



*Abbildung 1.7.:* Um zwei im Raum verschobene Moleküle zu vergleichen, wird eine Struktur auf die Koordinaten der anderen als Referenz-Struktur abgebildet. Die einander entsprechenden Atome werden so um den Schwerpunkt rotiert, dass die Summe ihrer *rmsd*-Abweichungen minimal wird [73].

### 1.5.2. Homologie-basierte Struktur-Vorhersage

Homologie-basierte Struktur-Vorhersage beruht auf der Beobachtung, dass Proteine mit verwandten Sequenzen auch ähnliche Strukturen aufweisen. Um von der bekannten Struktur eines verwandten Proteins zur Struktur des fraglichen Proteins zu kommen, sind folgende Schritte notwendig:

**Sequenz-Alignment:** Zunächst wird ein Protein gesucht, das der fraglichen Sequenz ähnlich ist, dessen Struktur bekannt ist und als Vorlage dienen kann. Daher werden Protein-Sequenzen miteinander verglichen, indem sich möglicherweise entsprechende AS einander gegenübergestellt werden. Die Übereinstimmungen der fraglichen Sequenz mit derjenigen der Struktur-Vorlage sollte möglichst groß sein.

**Threading:** Die Positionen der Hauptketten-Atome werden von der Struktur-Vorlage für die korrespondierenden AS der fraglichen Sequenz übernommen.

**Loop-Modellierung:** Fehlende Abschnitte der Hauptkette müssen ergänzt werden. Da keine Struktur-Informationen vorliegen, ist dies ein kritischer Schritt, der nur für kurze Abschnitte einfach zu lösen ist.

**Seitenketten-Platzierung:** Die Seitenketten werden nach Möglichkeit konservativ ergänzt, d.h. unter weitgehendem Beibehalt der durch die Vorlage gegebenen Konformationen. Dennoch sind Kollisionen unvermeidlich, die eine anschließende Struktur-Optimierung notwendig machen.

Wieweit sich diese Schritte zufriedenstellend durchführen lassen und welche Zuverlässigkeit der vorhergesagten Struktur mit diesem Verfahren erreichbar ist, hängt ab vom Grad der Sequenz-Identität zwischen dem fraglichen Protein und der gewählten Vorlage. Bei lediglich 10 – 30 % Übereinstimmung kann man den groben Verlauf der Hauptkette abschätzen, bei 20 – 60 % Sequenz-Identität können Sequenz-Abschnitte – meist mit definierten Sekundär-Struktur-Elementen – einander zugeordnet werden. Im Bereich von 45 – 80 % Sequenz-Identität sind die Lücken so klein, dass sie eventuell durch die Modellierung von Verbindungssequenzen geschlossen werden können. Bei 65 – 95 % Sequenz-Identität ist die globale Struktur so eindeutig, dass auch Seitenketten mit hinreichender Zuverlässigkeit platziert werden können.

#### Sequenz-Alignment

Das Alignment dient dazu, den Grad der Identität und das mögliche Bestehen einer Homologie zwischen Sequenzen festzustellen und gegebenenfalls Stammbäume zu erstellen. Mit Homologie bezeichnet man Ähnlichkeiten zwischen Sequenzen, die auf der Abstammung von einem gemeinsamen Vorfahren beruhen. Homologe Proteine verschiedener Spezies bilden Familien mit gemeinsamer, charakteristischer Architektur. Die für die Funktion essentiellen Seitenketten oder Struktur-Elemente einer Familie sind nicht variabel, sondern

kommen in allen Mitgliedern vor; sie sind konserviert. Andere Abschnitte variieren zwischen den individuellen Vertretern der Familie. Mit Hilfe eines Sequenz-Profiles aus einem multiplen Sequenz-Alignment lassen sich bei einer Familie von verwandten Proteinen variable Regionen von konstanten leicht unterscheiden: Das Profil listet für alle Positionen innerhalb der Sequenz die Häufigkeit auf, mit der jede einzelne der 20 natürlichen AS dort vertreten ist. Unter dem Alignment versteht man den Vorgang, die AS zweier oder mehrerer Sequenzen einander so gegenüber zu stellen, dass eine maximale Übereinstimmung erreicht wird.

Bei der Gegenüberstellung zweier AS muss bewertet werden, ob diese einander entsprechen oder nicht. Das ist nur dann eindeutig zu beurteilen, wenn es sich um dieselbe AS handelt. Dies ist aber nicht immer der Fall. Die 20 AS unterscheiden sich in Größe und Polarität ihrer Seitenketten (s. Abb. 1.3). Für jede mögliche Kombination sind daher in einer Bewertungs-Matrix (BloSum-Matrix, Dayhoff-Matrix [24, 60]) empirische Parameter aufgelistet, die jeder möglichen Zuordnung einen bestimmten Wert zuweisen. Diese Parameter sind so gewählt, dass Ähnlichkeiten hoch, und Abweichungen niedrig bewertet werden, ihre exakten Werte lassen sich aber nicht biologisch begründen. In Falle deutlicher Unterschiede bleiben in einer Sequenz ggf. einige AS ohne Entsprechung; dann werden Lücken in das Alignment eingefügt. Dies kann notwendig sein, wenn die Sequenzen unterschiedlich lang sind und/oder Abschnitte signifikant voneinander abweichen.

Beim Vergleich zweier Sequenzen der Längen  $n$  und  $m$  gibt es  $n \cdot m$  Möglichkeiten, diese einander gegenüberzustellen. Eine gegebene Zuordnung der vollständigen Sequenzen wird bewertet nach der Summe aller Einzelwerte aus der Bewertungs-Matrix. Die Einführung von Lücken wird von der Bewertungs-Funktion „bestraft“. Aus der Menge der möglichen Alignments wird dasjenige mit dem besten Endwert gewählt. Die Qualität eines Alignments hängt ab von den Parametern der Bewertungs-Matrix. Diese Werte werden aus der statistischen Auswertung bekannter Protein-Klassen gewonnen, daher sind sie auf neue Protein-Klassen unter Umständen nur eingeschränkt übertragbar. Der Artikel von C. Notredame [106] schildert die Vor- und Nachteile verschiedener Algorithmen.

## Threading

Nach dem Alignment wird die Sequenz in die Struktur-Vorlage „eingefädelt“, d.h. die Koordinaten der einander zugeordneten AS werden von der Vorlage für die fragile Struktur übernommen. Lücken im Alignment haben keine strukturelle Entsprechung. Daher tauchen in der neuen Struktur entweder Unterbrechungen in der Hauptkette auf, oder aber es werden Segmente nahtlos aneinander gefügt, auch wenn zwischen ihnen zusätzliche, „überzählige“ AS liegen.

## Loop-Modellierung

Bei der Modellierung von Hauptketten-Abschnitten mit unbekanntem Verlauf zwischen strukturell festgelegten Bereichen ist die Länge der ungewissen Übergangsstücke entscheidend. Bei kurzen Übergängen schränken die Randbereiche die Möglichkeiten des Haupt-

ketten-Verlaufs so stark ein, dass wenig Spielraum für Struktur-Alternativen existiert und sichere Vorhersagen leicht möglich sind. Längere Abschnitte kann man mosaikartig aus Strukturen verschiedener Proteine zusammensetzen. Nach der Rosetta-Methode [36] wird für kurze Sequenz-Fragmente (jeweils 3 und 9 AS) mit lokaler Ähnlichkeit zu einer bekannten Struktur der jeweils häufigste Hauptketten-Verlauf aus der PDB übernommen. Mit diesem Verfahren gelingt es zwar, extrem unwahrscheinliche Konformationen auszuschließen, gesicherte Informationen über den Verlauf der Hauptkette sind jedoch nur bedingt zu erzielen. Das gleiche gilt auch für alternative Verfahren [125], sodass das Modellieren unbekannter Loops den kritischen Schritt bei der Homologie-basierten Struktur-Vorhersage darstellt.

### Seitenketten-Platzierung

Bereits im zweiten Schritt, beim Threading, werden von der Struktur-Vorlage auch die mit der fraglichen Sequenz übereinstimmenden Seitenketten-Positionen übernommen, soweit die Seitenketten einander entsprechen. Für kleinere oder gleich große Seitenketten steht daher immer genügend Raum zur Verfügung. Ist die fragliche Seitenkette jedoch größer als ihre Vorlage, ist zum einen ihre Orientierung, zum anderen ein ausreichendes Platzangebot völlig ungewiss. Bei der vollautomatischen Ergänzung durch CHARMM [18] (s. Abschnitt 2.2) wird eine gestreckte Standard-Konformation ausgehend vom  $C_{\alpha}$ -Atom erzeugt, ungeachtet möglicher Kollisionen mit benachbarten Teilen des Proteins. Andere Verfahren versuchen auf Grundlage einer Rotamer-Datenbank plausible Seitenketten-Konformationen zu erzeugen [17]. Die dennoch mögliche gegenseitige Durchdringung von Seitenketten macht in jedem Fall eine anschließende Struktur-Optimierung erforderlich (s. Abschnitt 2.2.1).

### 1.5.3. Konstruktion von Komplexen: Docking

Mit dem Programm FLEXX [115] lassen sich flexible Liganden in die Bindetasche eines Proteins einpassen oder „docken“. Dabei geht man von der chemischen Struktur des Liganden und von der Kristall-Struktur des Proteins aus. Die Region der Bindetasche muss eigens festgelegt werden. Verschiedene Möglichkeiten für die Ligand-Konformation innerhalb der Bindetasche werden parallel erzeugt und bewertet, sodass man schließlich einige hundert Konformationsvorschläge erhält.

Der Ligand wird von FLEXX an jeder frei drehbaren Einfachbindung in Fragmente unterteilt. Bindungslängen und -winkel werden beibehalten. Jedem Atom innerhalb eines Fragmentes wird entsprechend seines Typs (Element und lokale chemische Umgebung) eine bestimmte Wechselwirkungsart und -geometrie zugewiesen. Für einen H-Brücken-Donor beispielsweise erwartet man in einem Abstand von ca. 2 Å in der Verlängerung seiner Bindungsachse einen Akzeptor und umgekehrt. Um ein erstes, sog. Basis-Fragment in der Bindetasche zu platzieren, werden jeweils drei Atome dieses Fragmentes zusammengefasst zu einem Dreieck. Ebenso lassen sich die Atome der Bindetasche „triangulieren“. Dann werden zueinander passende Dreiecke gesucht. Dabei wird nicht nur die Kantenlänge, sondern auch die chemische Komplementarität der Eckpunkte berücksichtigt. Es werden mehrere

Fragmente eines Liganden parallel als Basis-Fragmente verwendet. Lassen sich dennoch keine übereinstimmenden Atom-Tripel finden, werden ersatzweise Paare von Ligand- und Protein-Atomen miteinander verglichen. Das Basis-Fragment kann aber auch manuell auf eine gewünschte Referenz-Position gesetzt werden. Ist ein solches Fragment einmal platziert, wird jeweils das nächste daran angehängt unter Berücksichtigung passender Wechselwirkungen zwischen Ligand- und Protein-Gruppen. Jede Fragment-Platzierung wird bewertet nach einer einfachen und darum schnell berechenbaren Bewertungs-Funktion (LUDI [15]). Da die Fragmente gegeneinander drehbar sind (Rotation in 30°-Schritten), gibt es für ein Fragment mehrere alternative Platzierungen. Nur die günstigsten davon werden für den nächsten Schritt des Ligand-Aufbaus berücksichtigt. So erhält man einige hundert Vorschläge, wie der Ligand in der Bindetasche lokalisiert sein könnte, geordnet nach größter Wahrscheinlichkeit im Sinne der verwendeten Bewertungs-Funktion. Diese Bewertungs-Funktion verzichtet allerdings auf eine detaillierte Modellierung physikochemischer Eigenschaften. Durch die diskontinuierliche Rotation der Fragmente kann eine tatsächlich günstige Position oft knapp verfehlt werden; entweder, weil Wechselwirkungspartner zu weit voneinander entfernt sind, oder weil sie miteinander kollidieren. Um solch qualitativ gute, aber lokal gespannte Platzierungen nicht fälschlicherweise zu verwerfen, wird auch eine gewisse Durchdringung von Atomen zugelassen. Die Platzierung eines Liganden mit einer *rmsd* von 1.0–1.5 Å zu seiner Referenz-Position wird als Treffer bewertet, bei der der gedockte Ligand den gleichen Bindemodus wie die Referenz-Konformation besitzt. Die Ungenauigkeiten der Bewertungsfunktion und der Atom-Positionen machen je nach Anwendungszweck eine anschließende Struktur-Optimierung erforderlich.

## 2. Theoretische Grundlagen

### 2.1. Proteine in wässriger Lösung

Im Folgenden wird beschrieben, wie die Wechselwirkungen zwischen Proteinen und Liganden auf molekularer Ebene im Rahmen dieser Arbeit modelliert werden.

Gleichgültig, ob Untersuchungen an einem Protein *in vitro*, also naßchemisch im Labor, oder *in silico*, d. h. als Simulation im Rechner, durchgeführt werden: Bei beiden Ansätzen wird das Protein isoliert von den anderen Bestandteilen einer Zelle beobachtet. Somit gehen alle messbaren Effekte nur von dem ausgewählten Protein aus. Bei solchen Experimenten ist das zu untersuchende *System* also ein winzig kleiner Ausschnitt aus einer Zelle, ein Würfel oder ein Tropfen, der eben nur diese eine Sorte von Protein enthält. Die Zahl der Atome bzw. Moleküle bleibt während der Untersuchung konstant, ebenso die Gesamtenergie dieses abgeschlossenen Systems. Es herrschen physiologische Bedingungen, also Körper-Temperatur und Normal-Druck (37°C, 1013 mbar).

#### 2.1.1. Moleküle und Zustände

Ein *einzelnes Molekül* ist zusammengesetzt aus Atomen definierter Zahl und Sorte. Ein *Atom* wiederum besteht aus einem Atom-Kern mit Element-spezifischer (positiver) Kernladungszahl und ebenso vielen negativ geladenen Elektronen. Die Atom-Kerne eines Moleküls sind durch die Elektronen in einer *kovalenten Bindung* miteinander verknüpft. Die Verteilung der Elektronen fixiert die Atom-Kerne und ihre nächsten Nachbarn in einer lokalen Geometrie mit definierten Bindungslängen und -winkeln. Das kovalente Verknüpfungsmuster wird als *Konfiguration* des Moleküls bezeichnet. Trotz der eindeutig festgelegten Konfiguration kann ein Molekül in verschiedenen *Konformationen* vorkommen, d. h. die Atome nehmen unterschiedliche Positionen relativ zueinander ein. Der aktuelle *Zustand* des Moleküls ist charakterisiert durch seine jeweilige Konformation und die damit verbundene potentielle Energie. Jedes Molekül hat eine *Grund-Konformation mit minimalem Energie-Inhalt*.

Durch die thermischen Bewegungen schwanken innerhalb eines Moleküls die Abstände zwischen Nachbar-Atomen (Bindungslängen und -winkel) um einen Gleichgewichtswert (*Vibration*). Außerdem drehen sich Untergruppen innerhalb des Moleküls gegeneinander (*Rotation* um Verbindungs-Achse). Die Zahl der möglichen Konformationen eines Moleküls wächst exponentiell mit der Zahl der frei drehbaren Bindungen. Die

Moleküle bewegen sich auch als Ganzes (*Translation*). Auf begrenztem Raum prallen sie häufig aufeinander. Bei einem Zusammenstoß bremst die rapide zunehmende potentielle Energie eine weitere gegenseitige Annäherung ab, sodass die Teilchen sich verformen und/oder anschließend in neue Richtungen auseinander getrieben werden.

### 2.1.2. Wechselwirkungen zwischen Molekülen

Unabhängig von den gerichteten, kovalenten Bindungen herrschen zwischen Atomen zusätzlich – intra- und intermolekular – ungerichtete, *elektrostatische Wechselwirkungen*, die vom Coulomb-Gesetz (Gl. 2.1) beschrieben werden. Nach Gl. 2.1 stoßen sich gleichsinnige Ladungen ab, gegensinnige ziehen sich an. Die aufeinander ausgeübte Kraft ist umso stärker, je größer der Betrag der Ladungen und je kürzer ihr Abstand voneinander ist. Anziehung entspricht einer niedrigen potentiellen Energie, Abstoßung einer hohen potentiellen Energie.

$$F = \frac{q_1 \cdot q_2}{4\pi\epsilon_0\epsilon \cdot r_{1,2}^2} \quad (2.1)$$

$q_{1,2}$  sind die Beträge und  $r_{1,2}$  der Abstand zweier Punkt-Ladungen,  $\epsilon_0$  ist die Permeabilitätskonstante des Vakuums und  $\epsilon$  eine stoffspezifische, *makroskopische* Dielektrizitätskonstante (im Vakuum 1). Diese Stoff-Konstante wird aufgrund der Beobachtung eingeführt, dass die tatsächlich messbare Kraft zwischen zwei Ladungen von der Natur der Umgebung abhängt, in der die Ladungen eingebettet sind. Es handelt sich dabei um eine Größe, die den *durchschnittlichen* Einfluss der umgebenden Moleküle auf die Kraft zwischen den Ladungen wiedergibt. Mit dem geometrischen Faktor  $4\pi$  wird die Kugelsymmetrie des elektrischen Feldes einer Punktladung beschrieben.

Innerhalb eines Moleküls treten lokal Ladungen auf, wenn entweder der Betrag der negativen und positiven Ladungen nicht gleich ist oder die Schwerpunkte der Ladungsverteilungen nicht übereinstimmen. Diese lokalen Ladungen können permanent sein oder fluktuieren.

- Überschussladungen sind permanent und ganzzahlig. Bei Molekülen verteilt sich die Ladung anteilig auf mehrere Atome innerhalb einer anionischen bzw. kationischen Gruppe.
- Unterschiedlich geladene Atom-Kerne (verschiedene Elemente) verzerren die ansonsten gleichmäßig verteilte Elektronen-Dichte zwischen den Atomen. Dadurch entstehen permanente Partial-Ladungen an bestimmten sog. polaren Gruppen im Molekül.
- Fluktuierende Partial-Ladungen beruhen auf der Dynamik der Elektronen-Hüllen: Zwei benachbarte Elektronen-Hüllen stoßen sich ab und verschieben dadurch ihre Ladungs-Schwerpunkte, sodass kurzzeitig benachbarte, einander anziehende Dipole entstehen. Diese Dipol-Dipol-Wechselwirkungen haben zwar nur eine sehr kurze Reichweite (proportional  $1/r^6$ ), doch üben alle Moleküle *immer* diese schwache *van-der-Waals-Anziehung* aufeinander aus.

Die gegenseitige Annäherung bzw. Durchdringung der Moleküle durch die van-der-Waals-Anziehung wird begrenzt durch die gleichzeitige gegenseitige Abstoßung der Elektronen-Hüllen: Nach dem Pauli-Prinzip können sich Elektronen mit gleichen Quanten-



Zuständen nicht im selben Raumgebiet aufhalten. Daher steigt die potentielle Energie sehr stark an, wenn Atom-Kerne (bzw. ihre Hüllen) einen gewissen Minimal-Abstand unterschreiten. Dieser Abstand gilt nicht bei Atom-Kernen in einer gemeinsamen kovalenten Bindung. Auf dieser sog. *Pauli-Abstoßung* beruht der Raumbedarf eines Atoms oder Moleküls.

### 2.1.3. Verhalten von Molekülen im Kollektiv

Moleküle befinden sich in thermischer Bewegung und wechseln daher ständig ihre Zustände. Auf engem Raum beeinflussen sie sich gegenseitig. Die vielen möglichen Zustände aller Moleküle des Systems multiplizieren sich zu der sehr großen Zahl der möglichen Zustände des gesamten Systems. Welche der prinzipiell möglichen Zustände die Moleküle dann tatsächlich einnehmen, hängt von der jeweiligen potentiellen Energie  $\epsilon_i$  des Zustands und der mittleren kinetischen Energie der Teilchen innerhalb des Systems, der Temperatur, ab. Die Häufigkeiten, mit denen bestimmte Zustände eingenommen werden, beschreibt die **Boltzmann-Verteilung** (Gl. 2.2):

$$\frac{N_i}{N_{\text{gesamt}}} = \frac{g_i e^{-\frac{\epsilon_i}{kT}}}{\sum g_i e^{-\frac{\epsilon_i}{kT}}} \quad (2.2)$$

$N_i$  ist die Häufigkeit, mit der ein Zustand  $i$  eingenommen wird,  $g_i$  ist der Entartungsgrad, für den Fall dass der Zustand  $i$  mehrfach besetzt werden kann.  $\epsilon_i$  ist die Energie des Zustands.  $N_{\text{gesamt}}$  ist die Zahl aller möglichen Zustände des Systems, der Nenner auf der rechten Seite der Gleichung ist die Zustandssumme des Systems.  $k$  ist die Boltzmann-Konstante und  $T$  die aktuelle Temperatur des Systems.

Die meisten Teilchen nehmen bei tiefen Temperaturen Zustände mit möglichst niedriger potentieller Energie ein, bei höherer Temperatur verschiebt sich das Maximum der Häufigkeitsverteilung nicht nur zu höheren Werten der potentiellen Energie, auch die Bandbreite der besetzten Zustände wird größer. Da die Zahl der besetzten Zustände insgesamt gleich bleibt, wird die Verteilungskurve flacher. Während eines Experiments liegt demnach kein singulärer System-Zustand vor, vielmehr erfasst man bei der Untersuchung das *durchschnittliche Verhalten* einer Menge von Molekülen, eines *Ensembles*, gemittelt über den Zeitraum der Untersuchung.

### 2.1.4. Komplexierungsreaktionen

Im Hinblick auf eine spätere medizinische Anwendung sind *Reaktionen* von Proteinen  $P$  mit speziellen Liganden  $L$  interessant. Im Rahmen dieser Arbeit werden nur physikalische Reaktionen behandelt, d.h. die reversible Zusammenlagerung von wenigen molekularen Komponenten zu einem supramolekularen Komplex (s. Gl. 2.3). Die jeweiligen Komponenten behalten dabei ihren individuellen molekularen Aufbau, eine chemische Reaktion findet *nicht* statt.



Das zu betrachtende System umfasst eine gewisse Menge von Protein- und Ligand-Molekülen in wässriger Lösung. Jedes Molekül nimmt die unter den gegebenen Bedingungen erreichbaren Zustände ein. Im Gleichgewicht liegen  $P$  und  $L$  nebeneinander in zwei Formen vor, nämlich gebunden als Komplex ( $P \cdot L$ ) oder ungebunden als freie Komponenten ( $P + L$ ). Nie kommt nur eine einzige Form vor; durch die thermischen Bewegungen nehmen die einzelnen Komponenten im Laufe der Zeit wechselnd sowohl die gebundene als auch die ungebundene Form ein.

Alle Moleküle innerhalb der Lösung haben durch ihre ständigen Bewegungen zwar enge, aber nur kurzzeitige Kontakte zueinander. Von diesen zufälligen Berührungen zu unterscheiden ist die physikalische Bindung als eine deutlich längere Kontaktphase zwischen zwei Komponenten in einem definierten *Bindemodus*, d. h. einer bestimmten Orientierung zueinander (Komplementarität, s. Abschnitt 1.2.1).

Treten Moleküle miteinander in Kontakt, ändert sich die energetische Reihenfolge ihrer Zustände: Einige Zustände werden energetisch abgesenkt und dadurch häufiger eingenommen; umgekehrt werden andere ungünstiger und daher seltener besetzt. Wenige der vom freien Molekül bevorzugten Zustände sind mit einer Komplexierung vereinbar. Dies bedeutet einen *Entropie-Verlust für die gebundenen Komponenten*, denn sie können sich nur noch abhängig voneinander bewegen und ihre internen Rotations- und Vibrationsfreiheitsgrade sind eingeschränkt. Da sich aber gleichzeitig die zu benetzende Oberfläche der Komponenten im Komplex um die beiden Kontaktflächen verringert, wird die Bindung andererseits begünstigt durch den *Entropie-Gewinn des Wassers bei der Komplexierung*, (s. Kap. 1.2.3). Welcher dieser gegenläufigen Effekte überwiegt, legt die resultierende energetische Abfolge der Zustände des Systems fest.

Ein Molekül aus der unmittelbaren Umgebung des Komplexes kann den Liganden nur dann aus seiner Position am Protein verdrängen, wenn es genug Schwung bzw. kinetische Energie besitzt, um den Liganden so zu beschleunigen, dass er aus der Kontaktregion herausgeschoben wird. Wie viele Moleküle innerhalb der Lösung dazu in der Lage sind, hängt davon ab, wie hoch die mittlere kinetische Energie der Teilchen, bzw. die Temperatur des Systems ist, und wie sehr der Komplex stabilisiert wird. Je mehr Teilchen genug kinetische Energie besitzen, um den Liganden vom Protein weg zu schieben, desto kürzer ist die Dauer des Kontaktes zwischen den Komponenten. Umgekehrt liegt die gebundene Form umso länger und damit auch häufiger vor, je mehr der Komplex stabilisiert ist, d. h. je niedriger die durchschnittliche potentielle Energie der gebundenen Form im Vergleich zur freien Form ist.

### 2.1.5. Absolute Bindestärke eines Komplexes

Aus der energetischen Differenz zwischen der dissoziierten und der assoziierten Form einer Komplex-Verbindung resultieren die Mengenverhältnisse zwischen den freien und gebundenen Komponenten (bei gegebener Temperatur  $T$ ).

Nach dem Massenwirkungsgesetz bestimmt man über die messbaren Konzentrationen (in eckigen Klammern) die Gleichgewichtskonstante  $K$  einer beliebigen Reaktion (Gl. 2.4). Bei Komplezierungsreaktionen wird diese experimentelle Größe auch als Dissoziationskonstante  $K_D$  oder Inhibitionskonstante  $K_i$  bezeichnet und zur Beurteilung der Bindestärke herangezogen.

$$K = \frac{[P] \cdot [L]}{[P \cdot L]} \quad (2.4)$$

Aus der messbaren Gleichgewichtskonstanten  $K$  lässt sich nach Gleichung 2.5 auf die energetische Differenz der beiden Formen, die *Reaktionsenthalpie* bzw. *absolute Bindestärke*  $\Delta G$  rückschließen:

$$\Delta G = -RT \ln K \quad (2.5)$$

Dabei ist  $R$  die allgemeine Gaskonstante und  $T$  die aktuelle Temperatur des Systems. Die Reaktionsenthalpie  $\Delta G$  gibt die Differenz der freien Enthalpie zwischen beiden Formen (assoziiert oder dissoziiert) für den (Standard-)Fall konstanter Druckverhältnisse an.

Statt über die Konzentrationen kann  $\Delta G$  nach der Gibbs-Gleichung auch über die Energie- und Entropie-Beiträge aller beteiligten Verbindungen ausgedrückt werden:

$$\begin{aligned} \Delta G &= G_{P \cdot L} - (G_P + G_L) \\ &= (H_{P \cdot L} - TS_{P \cdot L}) - ((H_P - TS_P) + (H_L - TS_L)) \\ &= H_{P \cdot L} - (H_P + H_L) - TS_{P \cdot L} + (TS_P + TS_L) \\ &= \Delta H - T\Delta S \end{aligned} \quad (2.6)$$

$H$  ist die Enthalpie eines Moleküls,  $S$  seine Entropie. Für Reaktionen, bei denen wie im Fall einer Komplexierung keine Volumenarbeit geleistet wird, ist  $H$  gleichzusetzen mit der inneren Energie  $E$ .

Um einen Liganden  $L$  auf seine Brauchbarkeit als Wirkstoff am Protein  $P$  zu prüfen, bietet sich alternativ zur experimentellen Bestimmung von  $K$  die Berechnung der Reaktionsenthalpie  $\Delta G$  aus den energetischen und entropischen Einzelbeiträgen der Komponenten an.

## 2.2. Simulation von Molekülen als klassische Teilchen

Um die Reaktionsenthalpie  $\Delta G$  allein auf Grundlage der chemischen Zusammensetzung der Komponenten berechnen zu können, müssen Molekül-Strukturen mit passenden Energie-Werten korreliert werden. Dieser Abschnitt behandelt die Modellierung von Molekülen im Computer.

Wie erläutert, ist die Energie eines Moleküls abhängig von der Verteilung der Atomkerne und Elektronen im Raum. Die korrekte Methode zur Simulation von Molekülen wäre

die Lösung der Schrödinger-Gleichung. Aus prinzipiellen Gründen ist diese Lösung für große Moleküle jedoch unmöglich und auch andere quantenmechanische Berechnungsverfahren sind bei der Größe von Protein-Molekülen überfordert.

Für die Beschreibung von Komplexierungsreaktionen wird es als ausreichend erachtet, Moleküle als klassisch-mechanische Objekte zu modellieren.

### 2.2.1. Das Kraftfeld-Modell

Dem theoretischen Modell eines Kraftfeldes liegt die Born-Oppenheimer-Näherung zugrunde: Wegen des großen Massenunterschiedes zwischen Atom-Kernen und Elektronen können sich die Elektronen fast augenblicklich auf jede Änderung der Kern-Position einstellen. Elektronen- und Kern-Bewegungen können demnach getrennt voneinander behandelt werden.

Man nimmt an, dass Atome Kugeln sind, die über elastische Federn statt kovalenter Bindungen miteinander verknüpft sind. Weiterhin geht man davon aus, dass die Bindungen und Winkel im „ungestörten System“ natürliche Größen haben und Moleküle ihre Geometrie so einrichten, dass diese natürlichen Größen so gut wie möglich realisiert werden. Ist dies aus sterischen Gründen nicht möglich, wird das Molekül auf eine voraussagbare Weise deformiert, wobei Spannungsenergien auftreten, die berechnet und minimiert werden können. Zusätzlich werden auch Wechselwirkungen zwischen den Atomen, die nicht über kovalente Bindungen vermittelt werden, berücksichtigt.

#### Die Energie-Funktion

Kraftfeld-Rechnungen bestimmen den Energie-Inhalt eines Moleküls in einer gegebenen Konformation nach einer empirischen Energie-Funktion. Die Energie-Funktion wird an jedes Molekül, seiner Konfiguration entsprechend, angepasst. Diese individuelle Energie-Funktion berücksichtigt unter anderem Terme für kovalente Bindungen, die über Zuordnung von Atom-Typen und spezielle Parameter angemessen gewichtet werden. Durch Einführung von Atom-Typen zusätzlich zur Element-Bezeichnung lassen sich die Bindungsverhältnisse detaillierter beschreiben. Für die jeweiligen Kombinationen von Atom-Typen werden passende Parameter bereitgestellt. Diese Parameter werden an bestimmten Verbindungsklassen geeicht. Man setzt voraus, dass sie von Molekül zu Molekül übertragbar sind, wenn die gleiche lokale Umgebung vorliegt. Eine Kraftfeld-Rechnung ist daher nur für einen begrenzten Anwendungsbereich gültig (abhängig vom „Eichsatz“). Im Rahmen dieser Arbeit wurde das Kraftfeld-Programm CHARMM [18] verwendet, das gute und weithin akzeptierte Parametersätze für Proteine und andere organische Verbindungen beinhaltet. Im einzelnen setzt sich die Energie-Funktion  $E^{\text{Kraftfeld}}(\vec{r}_1, \dots, \vec{r}_n)$  für ein Molekül mit  $n$  Atomen an den Orten  $\vec{r}_1, \dots, \vec{r}_n$  zusammen aus Beiträgen für kovalente Bindungen (Gl. 2.8 – 2.11), elektrostatische Wechselwirkungen (Gl. 2.12), sowie Lennard-Jones-Wechselwirkungen (LJ-Wechselwirkungen), die die van-der-Waals-Anziehung und die Pauli-Wechselwirkungen modellieren (Gl. 2.13). Für bestimmte Anwendungen, z.B. bei der Energie-Minimierung, kann die Energie-Funktion um bestimmte Terme erweitert oder verkürzt werden.

$$E^{\text{Kraftfeld}}(\vec{r}_1, \dots, \vec{r}_n) = \quad (2.7)$$

$$\frac{1}{2} \sum_i k_{r,i} (r_i - r_{0,i})^2 \quad \text{Bindungslängen} \quad (2.8)$$

$$+ \frac{1}{2} \sum_j k_{\alpha,j} (\alpha_j - \alpha_{0,j})^2 \quad \text{Bindungswinkel} \quad (2.9)$$

$$+ \sum_k k_{\phi,k} (1 - \cos(n_k(\phi - \phi_{0,k}))) \quad \text{Dieder - Winkel} \quad (2.10)$$

$$+ \sum_l k_{\theta,l} (\theta_l - \theta_{0,l})^2 \quad \text{uneigentliche Dieder} \quad (2.11)$$

$$+ \sum_{p>q} \frac{q_p q_q}{4\pi\epsilon\epsilon_0 r_{pq}} \quad \text{Coulomb - Energie} \quad (2.12)$$

$$+ \sum_{m>n} 4E_{0,mn} \left( \left( \frac{\sigma_{mn}}{r_{mn}} \right)^{12} - \left( \frac{\sigma_{mn}}{r_{mn}} \right)^6 \right) \quad \text{Lennard - Jones} \quad (2.13)$$

**Bindungslängen:** Kovalente Bindungen werden nach Gl. (2.8) als harmonische Federn mit der Kraftkonstanten  $k_r$  und der Gleichgewichtslänge  $r_0$  beschrieben. Beide Parameter hängen von der Art der Bindungspartner und der Art der Bindung ab. Aus der Abweichung der aktuellen Bindungslänge  $r_i$  vom Gleichgewichtswert  $r_0$  wird ihr Beitrag zur Kraftfeld-Energie ermittelt. Die Kraftkonstante  $k_r$  bestimmt, wie groß der Einfluss einer Längenänderung ist. Kovalente Bindungen lassen sich wenig strecken oder stauchen, die Kraftkonstante ist also relativ groß, sodass schon eine geringe Abweichung des Abstands vom Gleichgewichtswert die Energie des Moleküls beträchtlich erhöht.

**Bindungswinkel:** Auch bei Bindungswinkeln werden die Atome als harmonisch-elastisch verbundene Massepunkte behandelt (Gl. 2.9).  $k_\alpha$  ist ebenfalls abhängig von den an der Bindung beteiligten Elementen und der Bindungsart. Die entspannte Position wird durch den Gleichgewichtswinkel  $\alpha_0$  definiert.  $k_\alpha$  ist kleiner als  $k_r$  und erlaubt bei gleicher potentieller Energie größere Auslenkungen aus der entspannten Lage.

**Diederwinkel:** Unter einem Diederwinkel versteht man den Winkel zwischen den beiden durch jeweils drei von vier kovalent verknüpften Atomen definierten Ebenen (s. Abb. 2.1a).

Das Torsionsprofil der Drehung einer Atom-Gruppe um eine Achse wird durch die Verwendung der Cosinus-Funktion in Gl. 2.10 beschrieben. Dabei wird über  $n_k$  die interne Symmetrie der rotierenden Gruppe und damit die Zahl der Minima berücksichtigt.  $\phi_0$  ist der Gleichgewichtsdiederwinkel. Die Kraftkonstante  $k_\phi$  bestimmt die Höhe der Torsionsbarrieren.

**Uneigentliche Diederwinkel:** Bei uneigentlichen Diederwinkeln sind die zu berücksichtigenden Atome nicht linear aneinander gereiht, sondern an ein Zentralatom gebunden (s. Abb. 2.1b). Durch das harmonische Potential (Gl. 2.11) wird die Inversion chiraler Gruppen oder eine zu starke Verdrillung planarer Gruppen verhindert. Dabei sind  $\theta_0$  und  $k_\theta$  der Gleichgewichtswinkel beziehungsweise die Kraftkonstante.

**Lennard-Jones:** Das Lennard-Jones-Potential (Gl. 2.13) umfasst einen attraktiven (proportional  $-1/r^6$ ) und einen repulsiven (proportional  $+1/r^{12}$ ) Anteil, der die van-der-Waals-Anziehung und Pauli-Abstoßung zusammen modelliert. Für große Abstände zwischen zwei Atomen geht das Potential gegen Null. Der sterische Anspruch der Atome wird über den Kollisionsparameter  $\sigma_{m,n}$ , der ein Maß für die Größe der Atome  $m$  und  $n$  ist, berücksichtigt.  $E_0$  ist die Tiefe des Potenzialtopf, d. h. die minimal mögliche Energie zwischen den Atomen.



Abbildung 2.1.: Diederwinkel  $\Phi$  und uneigentlicher Diederwinkel  $\theta$ .

**Coulomb-Energie:** Die elektrostatischen Wechselwirkungen der atomaren Partial-Ladungen werden durch das Coulomb-Gesetz (Gl. 2.1) beschrieben. Die Partial-Ladungen werden in dieser Arbeit dem CHARMM Kraftfeld entnommen [18] bzw. nach einer in der Quanta-Software verfügbaren Methode berechnet (charge templates [70]).

### Energie-Minimierung

Mit der beschriebenen Energie-Funktion (Gl. 2.8 – 2.13) besteht die Möglichkeit, für jede Konformation eines Moleküls die entsprechende Kraftfeld-Energie  $E^{\text{Kraftfeld}}$  zu berechnen. Durch die Form der Energie-Terme, besonders bei Bindungslängen und LJ-Wechselwirkungen, kann schon eine kleine Abweichung vom Gleichgewichtsabstand zu einer hohen potentiellen Energie führen. Oft lässt sich die Energie eines Moleküls mit einer lokal ungünstigen Anordnung von Atomen ohne nennenswerte Veränderung der allgemeinen Konformation drastisch senken, wenn die Atome geringfügig verschoben werden. Um also relevante Zustände nicht fälschlicherweise auszuschließen bzw. um nicht physikalisch unsinnige Konformationen beizubehalten, werden Energie-Minimierungen unter Veränderung der Atom-Positionen durchgeführt.

Solange sich die Energie eines Systems verringern lässt, werden die Atom-Positionen schrittweise verändert, bis ein Minimum der verwendeten Energie-Funktion erreicht ist (s. Abb. 2.2). Die Energie-Funktion eines Systems ist abhängig von den Raum-Koordinaten aller seiner Atome, die resultierende Energie-Landschaft ist also sehr komplex strukturiert. Man unterscheidet innerhalb dieser Landschaft ein niedrigstes, globales Minimum der Energie (A) von verschiedenen höheren, lokalen Minima (B, C). Der Übergang des Systems aus einem Minimum in ein anderes ist zunächst immer mit einer Energie-Zunahme verbunden (energetische Barriere). Welches Minimum der Kraftfeld-Energie erreicht werden kann, hängt von dem Ausgangspunkt zu Beginn der Minimierung, von der Start-Konformation ab. Wegen der energetischen Barrieren kann ein Molekül trotz Energie-Minimierung in dem lokalen Minimum einer ungünstigen Konformation gefangen bleiben (Fall C in Abb. 2.2). In jedem Fall erhält man nach einer Energie-Minimierung eine einzelne Molekül-Konformation.

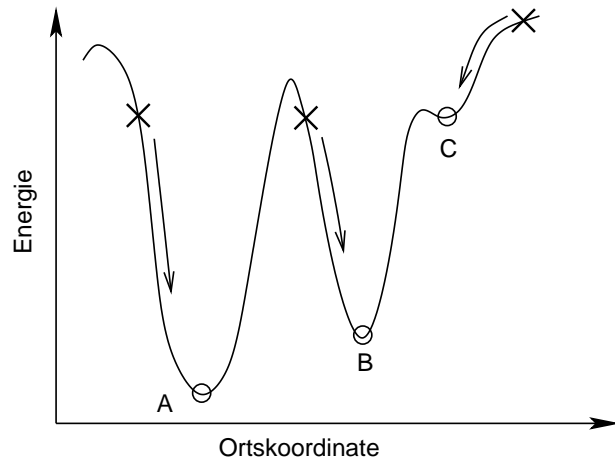


Abbildung 2.2: 1-dimensionale Energie-Landschaft. Abhängig von den verschiedenen Ausgangspunkten  $\times$  für eine Minimierung werden verschiedene Minima  $\circ$  erreicht: A) globales Minimum, B) und C) lokale Minima, wobei C) einen Problem-Fall illustriert!

Dies entspricht einer „Moment-Aufnahme“ vom Protein, das eigentlich ständig in Bewegung ist. In der Praxis verwendet man bei der Energie-Minimierung oft einschränkende Randbedingungen, die festlegen, welche Atom-Positionen innerhalb des Systems beweglich sein sollen und welche fixiert bleiben sollen.

### Molekulardynamik: Simulation thermischer Bewegungen

Das Kraftfeld-Modell erlaubt auch, thermische Bewegungen innerhalb eines Systems zu simulieren (Molekulardynamik-Simulation, MD-Simulation). Über die Energie-Funktion  $E^{\text{Kraftfeld}}(\vec{r}_1, \dots, \vec{r}_n)$  (Gln. 2.8–2.13) kann für jede Konstellation der  $n$  Atome die aktuelle potentielle Energie des gesamten Systems berechnet werden. Daraus lassen sich die Richtungsvektoren der Kräfte  $\vec{F}_i$  ableiten, in deren Feld sich das Atom  $i$  befindet:

$$\vec{F}_i = m_i \frac{d^2}{dt^2} \vec{r}_i = - \left( \frac{d}{dx_i}, \frac{d}{dy_i}, \frac{d}{dz_i} \right)^T E^{\text{Kraftfeld}}(\vec{r}_1, \dots, \vec{r}_n) \quad (2.14)$$

Dabei ist  $\vec{F}_i$  die Kraft auf das  $i$ -te Atom der Masse  $m_i$ . Mit  $\vec{F}_i = m_i \cdot \vec{a}_i$  kann für jedes Atom der Masse  $m_i$  eine Beschleunigung  $\vec{a}_i$  bzw. eine aktuelle, für den kurzen Zeitraum  $\Delta t$  geltende Geschwindigkeit  $\vec{v}_i = \vec{a}_i \cdot \Delta t$  angegeben werden. Die Temperatur  $T$  ist konstant.

Man kann über den Zeitschritt  $\Delta t$  die individuellen Bewegungen der Atome verfolgen und dann ihre Positionen sowie die entsprechenden potentiellen Energien und Kraftvektoren erneut berechnen.  $\Delta t$  ist so zu wählen, dass die schnellsten Bewegungen innerhalb eines Systems noch detailliert wiedergegeben werden. Die schnellsten Bewegungen in Molekülen sind Streckschwingungen kovalenter Bindungen mit Schwingungsdauern von ca. 10 fs. Das erlaubt Zeitschritte  $\Delta t$  von maximal einem Zehntel der kürzesten Schwingungsdauer, also etwa 1 fs, nach denen jeweils die Energien, Positionen und Geschwindigkeiten neu zu berechnen sind. Die Streckschwingungen sind selten von Interesse, sodass die MD-Simulation auch häufig mit fixierten Bindungslängen durchgeführt werden. Die während der Simulation durchlaufenen Zustände des Systems werden registriert und in ihrer Gesamtheit als

Trajektorie bezeichnet. Mit der Trajektorie lässt sich die zeitliche Entwicklung des Systems verfolgen.

Durch Zuordnung einer Geschwindigkeit  $v_i$  für jedes Atom kann man über

$$E_i^{\text{kin}} = \frac{1}{2} m_i \cdot v_i^2 \quad (2.15)$$

neben der potentiellen Energie  $E^{\text{Kraftfeld}}$  auch die kinetische Energie der Teilchen bestimmen. Diese mittlere kinetische Energie eines Teilchens  $\bar{E}^{\text{kin}}$  ist über die Boltzmann-Konstante  $k$  mit der makroskopisch messbaren Temperatur  $T$  verknüpft:

$$\bar{E}_i^{\text{kin}} = \frac{3}{2} kT \quad (2.16)$$

Während der Simulation müssen bestimmte Größen erhalten bleiben: Dies sind die Teilchenzahl  $N$ , das Volumen  $V$  oder der Druck  $p$  innerhalb des Systems, sowie die Gesamtenergie  $E^{\text{Gesamt}} = E^{\text{kin}} + E^{\text{Kraftfeld}}$  oder die Temperatur  $T$ . Je nach gewählter Temperatur können energetische Barrieren überwunden werden, sodass das System mit der Zeit verschiedene Konformationen einnimmt.

Um den Rechenaufwand zu begrenzen, umfasst das System statt eines Ensembles vieler identischer Protein-Moleküle nur ein Exemplar. Man setzt voraus, dass das dadurch erhältliche Zeitmittel nach einer angemessenen Simulationsdauer dem makroskopisch messbaren Ensemble- (und Zeit-)Mittel entspricht. Der Energie-Inhalt des Systems ist zu Beginn einer Simulation zufällig, bedingt durch die vorgegebene Startkonformation und die willkürliche Zuweisung von atomaren Geschwindigkeiten. Daher muss dem System zuerst in einer sog. *Äquilibrierungsphase* Energie zu- oder abgeführt werden, bis die gewünschte Temperatur eingestellt ist. Erst nach der Äquilibrierungsphase kann die eigentlich interessante Phase der Simulation durchgeführt werden. Simulationen können unter periodischen oder aperiodischen Randbedingungen durchgeführt werden. Diese Verfahren verwenden unterschiedliche Algorithmen zur Lösung der Bewegungsgleichungen. Periodische Randbedingungen erlauben einzelnen Teilchen, aus den räumlichen Begrenzungen des Systems heraus- bzw. hineinzutreten. Das System wird simuliert als Ausschnitt aus einem Verband aus unendlich vervielfältigten Untereinheiten. Bei aperiodischen Randbedingungen bleibt das System isoliert und auf ein definiertes Raumgebiet beschränkt; kein Teilchen kann weg- oder hineinzutreten.

Die praktische Durchführung einer MD-Simulation benötigt viel Zeit und Speicherplatz. Besonders die Berechnung der Coulomb-Wechselwirkungen ist in diesem Sinne teuer, da für jedes der  $n$  Atome mit jedem anderen die Wechselwirkungen bestimmt werden müssen ( $\frac{n^2}{2}$  Beiträge zur elektrostatischen Energie). Die maximale Simulationsdauer für ein System aus etwa 10.000 Atomen beträgt etwa 1-10 ns und benötigt sogar auf einem Hochleistungsrechner wie der CRAY T3D-Maschine 120 Stunden reine Rechenzeit [88]. Dies ist zu kurz, um biologische *Prozesse*, z.B. das Herantreten und Binden eines Liganden an ein Protein zu verfolgen. Stattdessen werden MD-Simulationen für die isolierten Komponenten und den Komplex durchgeführt. Aus der Differenz der jeweiligen mittleren Energien geht die



gesuchte Reaktionsenthalpie  $\Delta G$  direkt hervor, da durch die Dynamik entropische Effekte schon berücksichtigt sind.

## 2.3. Simulation des Wassers als homogenes Medium

Bisher ist mit dem Kraftfeld-Modell eine Methode geschildert worden, die die molekularen Strukturen innerhalb des Systems wiedergibt. Protein-, Ligand- und Wasser-Moleküle werden gleichermaßen in ihrer atomaren Struktur beschrieben. Um die Energie des in Wasser gelösten Proteins zu bestimmen, müssen also neben den gesuchten Beiträgen des Proteins selbst und den Wechselwirkungsbeiträgen mit seiner Hydrat-Hülle die mengenmäßig bei weitem überwiegenden, aber eigentlich uninteressanten Wechselwirkungen der restlichen Wasser-Moleküle untereinander berechnet werden. Das kostet während einer Energie-Minimierung oder MD-Simulation Zeit und Speicherplatz, zwei Aspekte, die die maximale Größe bzw. die Simulationsdauer eines Systems beschränken.

Der sogenannte „*continuum-solvent*“-Ansatz [70] umgeht das explizite Modellieren der Wasser-Wasser-Wechselwirkungen. Das Protein wird weiterhin atomar nach dem Kraftfeld-Modell behandelt, doch vom umgebenden Wasser wird nur die Abschirmung von Ladungen und die Oberflächen-Spannung berücksichtigt. Die wässrige Umgebung wird als homogenes Medium ohne molekulare Struktur angenommen, unter der Voraussetzung, dass die Menge der kleinen und sehr beweglichen Wasser-Moleküle sich jeder Konformation des Proteins ohne Widerstand anpasst. Die Gesamtenergie für das gelöste Molekül setzt sich dann zusammen aus der Kraftfeld-Energie des isolierten Moleküls und seiner Solvatationsenergie (Gl. 2.17).

$$E_{\text{Molekül}+\text{H}_2\text{O}}^{\text{Kraftfeld}} = E_{\text{Molekül}}^{\text{Kraftfeld}} + E_{\text{Molekül}}^{\text{Solvatation}} = E_{\text{Molekül}}^{\text{Gesamt}} \quad (2.17)$$

Die Solvatationsenergie setzt sich zusammen aus einem elektrostatischen Beitrag  $E^{\text{Elektrostatik}}$  und einem Beitrag für die Oberflächen-Spannung  $E^{\text{Oberfläche}}$ . (Gl. 2.18).

$$E_{\text{Molekül}}^{\text{Solvatation}} = E_{\text{Molekül}}^{\text{Elektrostatik}} + E_{\text{Molekül}}^{\text{Oberfläche}} \quad (2.18)$$

Wie in Abschnitt 1.2.3 erwähnt, schwächt Wasser die Wechselwirkungen zwischen Ladungen durch seine große Beweglichkeit und hohe Polarität ab. Es wird daher eine hohe Dielektrizitätskonstante  $\epsilon_{\text{H}_2\text{O}}$  von 78.5 beobachtet. Zum Vergleich: Die Dielektrizitätskonstante  $\epsilon$  des Vakuums ist 1; die von völlig unpolaren Stoffen, d.h. Stoffen ohne Dipol-Moment, bei denen nur die Elektronen-Hüllen geringfügig auf elektrische Ladungen reagieren können, ist etwa 2. Im Gegensatz zu den kleinen Wasser-Dipolen können sich die an der Hauptkette verankerten Seitenketten eines Proteins nur schlecht nach dem Feld einer elektrischen Ladung ausrichten. Zudem sind viele Seitenketten unpolar, also gänzlich ohne abschirmenden Einfluss auf eine benachbarte Ladung. Diese Eigenschaften führen zu einer niedrigen Dielektrizitätskonstante  $\epsilon_{\text{Protein}}$ . Proteine sind aber heterogen aufgebaut und weisen eine große Vielfalt auf.

Mit der Dielektrizitätskonstante  $\varepsilon_{\text{Protein}}$  wird jedoch nur das *durchschnittliche* Verhalten einer großen Menge von Molekülen berücksichtigt. Pauschal wird für die Dielektrizitätskonstante von Proteinen häufig  $\varepsilon_{\text{Protein}} = 4$  verwendet, da sich damit die beste Übereinstimmung mit experimentellen Befunden erzielen lässt [70].

Je größer die Dielektrizitätskonstante ist, desto schwächer ist die Kraft zwischen zwei Ladungen (Gl. 2.1). In Wasser wird eine Anziehung oder Abstoßung also ca. 20 mal stärker abgeschwächt als im Inneren eines Proteins.

Zur Bestimmung des *elektrostatistischen Beitrags zur Solvatationsenergie* eines Moleküls  $E_{\text{Molekül}}^{\text{Elektrostatik}}$  nach diesem Ansatz wird der „Körper“ des gelösten Moleküls aufgefasst als eine Region mit niedriger Dielektrizitätskonstante  $\varepsilon_{\text{Molekül}} = \varepsilon_{\text{intern}}$ , in der die (Partial-)Ladungen an den Atom-Positionen eingebettet sind. Diese Region wird umgeben von einem Medium mit hoher Dielektrizitätskonstante  $\varepsilon_{\text{Medium}} = \varepsilon_{\text{extern}}$ . Die Grenze zwischen diesen beiden Medien ist die molekulare Oberfläche, die von den van-der-Waals-Radien der Atome bestimmt wird (s. Abschnitt 2.3.1).

### 2.3.1. Die Poisson-Gleichung

Ein in Wasser gelöstes Protein besitzt im Inneren eine niedrigere Dielektrizitätskonstante als seine wässrige Umgebung. Um elektrostatische Wechselwirkungen innerhalb des geschilderten *heterogenen* dielektrischen Systems zu berechnen, kann nicht mehr die einfache Coulomb-Gleichung (Gl. 2.1) verwendet werden, denn sie gilt nur für den *homogenen* Fall. Daher muss die allgemeinere Poisson-Gleichung (Gl. 2.19) angewendet werden, mit der jede räumliche Verteilungen von Dielektrizitätskonstanten  $\varepsilon(\vec{r})$  behandelt werden kann:

$$-\nabla \cdot [\varepsilon(\vec{r}) \nabla \Phi_i(\vec{r})] = \frac{\rho}{\varepsilon_0} . \quad (2.19)$$

Die (Partial-) Ladungen des Proteins werden dabei als Punktladungen an den Atom-Positionen angenommen. Am Ort einer Punktladung  $q_i$  hat das elektrostatische Potential jeweils den Wert  $\Phi_i$ . Die gesamte Verteilung aller Ladungen im Raum  $\rho$  erzeugt ein elektrisches Potential  $\Phi(\vec{r})$ , das aus der Lösung der Poisson-Gleichung hervorgeht [51, 70].

Die elektrostatische Energie  $E^{\text{Elektrostatik}}$  dieser Ladungsverteilung ist gegeben durch:

$$E^{\text{Elektrostatik}} = \frac{1}{2} \sum_i q_i \Phi_i = \frac{1}{2} \int \rho \Phi dr \quad (2.20)$$

Die Poisson-Gleichung kann *analytisch* nur für Systeme mit einfachen Grenzflächen wie Kugeloberflächen oder Ebenen zwischen den dielektrischen Medien gelöst werden. Molekulare Oberflächen haben aber unregelmäßige Einstülpungen und Ausbuchtungen, sodass die Berechnung über ein *numerisches* Verfahren erfolgen muss; hier nach der Methode der finiten Differenzen mit dem Programm CAMLab [65] auf einer Sun-Fire-4800-Maschine mit 8 Prozessoren mit einer Taktfrequenz von je 750 MHz durchgeführt.

Die Verteilung der Punkt-Ladungen  $\rho(\vec{r})$  zur Lösung der Poisson-Gleichung wird „gerastert“ bzw. diskretisiert, d. h. auf ein drei-dimensionales kubisches Gitter abgebildet (s. Abb. 2.3). Das Protein befindet sich dann in einem hypothetischen Kasten, der sich aus Elementarwürfeln der Kantenlänge  $h$  zusammensetzt. Die Würfecken sind die Gitterpunkte, auf die die benachbarten atomaren Partial-Ladungen verteilt werden. Jeder Verbindungsachse zweier Ladungspunkte wird ein Wert für die jeweilige Dielektrizitätskonstante  $\varepsilon$  zugewiesen. Dieses sogenannte  $\varepsilon$ -Gitter zur Beschreibung der Verteilung der dielektrischen Medien hat damit dreimal mehr Werte als das Ladungsgitter. Durch die Rasterung können statt der infinitesimalen Ableitungen der Differential-Gleichung finite Differenzen zur iterativen Lösung der Poisson-Gleichung (Gl. 2.19) verwendet werden.

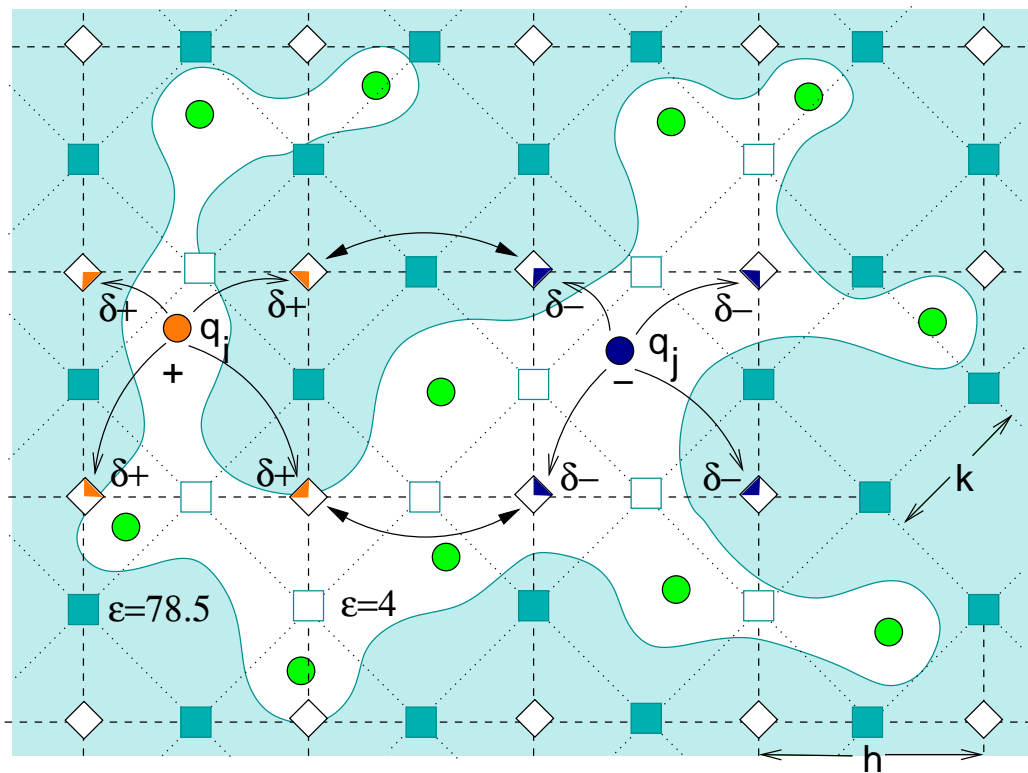


Abbildung 2.3.: 2D-Ladungs- und  $\varepsilon$ -Gitter: Nach dem Kraftfeld-Modell sind innerhalb eines Moleküls den Atom-Positionen (Kreise) Punktladungen  $q_i, q_j$  zugewiesen. Die Ladungen werden anteilig auf die nächsten Gitterpunkte (Rauten) des Ladungsgitters mit der Kantenlänge  $h$  verteilt. Die molekulare Oberfläche ist die Grenzfläche zwischen den elektrischen Medien mit  $\varepsilon = 4$  innerhalb des Moleküls (hell) und  $\varepsilon = 78.5$  außerhalb (dunkel). Jeder Kante des Ladungsgitters wird ein Wert für  $\varepsilon$  zugewiesen, je nachdem, in welchem Medium sich der Mittelpunkt der Kante befindet (Quadrate,  $\varepsilon$ -Gitter, Kantenlänge  $k$ ). Die elektrostatischen Wechselwirkungen zwischen  $q_i$  und  $q_j$  beispielsweise werden berechnet als die Beiträge zweier Paare von Gitterpunkten des Ladungsgitters (Doppelpfeile).

Die Verwendung finiter Differenzen zwischen *wenigen Punkten* statt Ableitungen an *jedem Raumpunkt* zur Lösung der Poisson-Gleichung führt zu systematischen Fehlern in der berechneten elektrostatischen Energie, die von der Auflösung des Gitters (Maschenweite  $h$ ) abhängen [87]. Um ein präzises Ergebnis zu erhalten, muss die Rechnung mit möglichst feiner Auflösung durchgeführt werden. Es hat sich gezeigt, dass eine Maschenweite von 0.4 Å optimal ist [87]. Da die Zahl der Punkte mit der dritten Potenz der Kantenlänge eingeht, ist die Auflösung aus praktischen Gründen (Rechenzeit und Speicherplatz) auf Gitter mit etwa 1 Million Punkten begrenzt.

### Elektrostatischer Beitrag zur Solvatationsenergie

Auch bei möglichst feiner Auflösung kommt es durch Gittereffekte zu numerischen Fehlern. Daher ist die exakte Berechnung des Beitrags der elektrostatischen Energie über Gl. 2.19 und Gl. 2.20 nicht direkt möglich. Stattdessen berechnet man Energiedifferenzen, wodurch der systematische Diskretisierungsfehler weitgehend eliminiert wird.

Zuerst wird Gleichung 2.19 für das dielektrisch heterogene System gelöst. Aus den erhaltenen Potential-Werten der Gitterpunkte wird nach Gl. 2.20 die Energie der Ladungsverteilung für  $\epsilon_{\text{extern}} = 78.5 \neq \epsilon_{\text{intern}} = 4$  ermittelt. Dann wird die gleiche Rechnung für den homogenen Fall mit  $\epsilon_{\text{extern}} = \epsilon_{\text{intern}} = 4$  durchgeführt. Gitter-Orientierung und -Auflösung sind bei beiden Rechnungen gleich. Daher sind auch die Wechselwirkungen der Ladungen *innerhalb* der molekularen Oberfläche für beide Systeme identisch. Ihre Beiträge zur elektrostatischen Energie entfallen bei der Differenzbildung und müssen daher im Rahmen der Kraftfeld-Rechnung berücksichtigt werden. Die Differenz der berechneten elektrostatischen Energien entspricht dann dem elektrostatischen Anteil der Solvatationsenergie für das Molekül  $E_{\text{Molekül}}^{\text{Solv, Elektrostatik}}$  [70].

$$E_{\text{Molekül}}^{\epsilon_{\text{ext}} \neq \epsilon_{\text{int}}} - E_{\text{Molekül}}^{\epsilon_{\text{ext}} = \epsilon_{\text{int}}} = E_{\text{Molekül}}^{\text{Solv, Elektrostatik}} \quad (2.21)$$

### Beitrag der Oberflächen-Spannung zur Solvatationsenergie

Die Oberflächen-Spannung des Wassers bestimmt den nicht-polaren Anteil der Solvatationsenergie  $E_{\text{Molekül}}^{\text{Solvatation}}$  (s. Abschnitt 1.2.3). Die Effekte der van-der-Waals-Anziehungen und des Entropie-Verlusts werden dominiert von den Wasser-Molekülen, die unmittelbarem Kontakt zum Solut haben. Deshalb ist der Beitrag der Oberflächen-Spannung in guter Näherung proportional zur benetzbaren Oberfläche (solvent accessible surface, SAS):

$$E_{\text{Molekül}}^{\text{Oberfläche}} \approx c + a \cdot \text{SAS}. \quad (2.22)$$

Dabei ist  $a$  die Proportionalitätskonstante und  $c$  der Achsenabschnitt. Die Konstanten  $a$  und  $c$  werden experimentell bestimmt über die Verteilungskoeffizienten von Alkanen zwischen einem unpolaren Lösungsmittel und einem polaren Lösungsmittel. Die Konstante  $c$  ist kleiner als 1 kcal/mol und wird im folgenden vernachlässigt. Die Konstante  $a$  in Gl. 2.22 ist positiv, das heißt die Oberflächen-Spannung wirkt dem Wechsel eines Moleküls aus einem unpolaren Medium in eine wässrige Umgebung

immer entgegen. In der Literatur werden für  $a$  Werte zwischen  $5 \text{ cal}/(\text{mol}\cdot\text{\AA}^2)$  und  $70 \text{ cal}/(\text{mol}\cdot\text{\AA}^2)$  angegeben. Hier wird ein Wert von  $20 \text{ cal}/(\text{mol}\cdot\text{\AA}^2)$  gewählt, da man damit gute Übereinstimmung mit dem Experiment erzielt, wenn die van-der-Waals-Radien des CHARMM-Kraftfeldes verwendet werden [18].

### Bestimmung von Oberflächen:

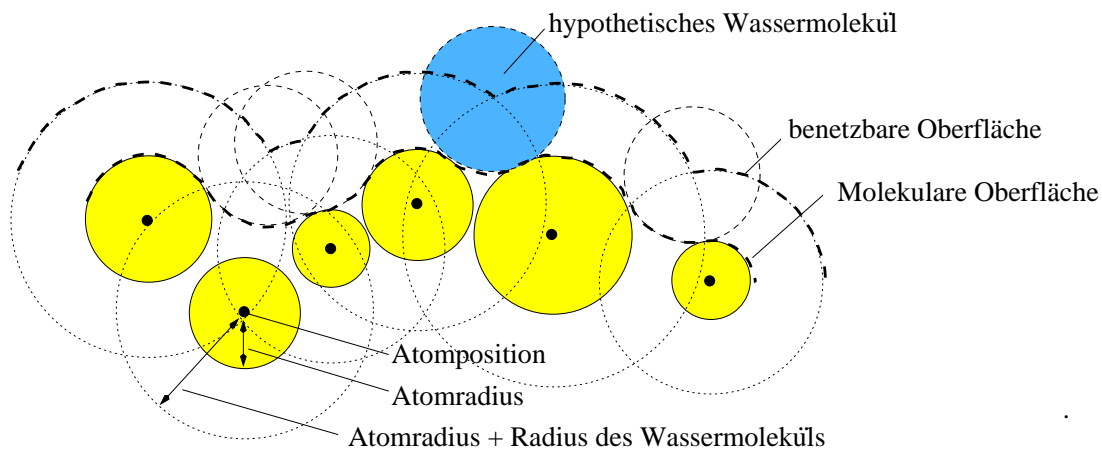
- Die *zu benetzende Oberfläche* eines Moleküls (solvent accessible surface, SAS) ist die Fläche, die die Schwerpunkte der Wasser-Moleküle im Kontakt mit dem gelösten Molekül einnehmen. Zur Bestimmung dieser Fläche behandelt man die Atome des Moleküls als Kugeln mit dem Radius  $r_{\text{Hydrat}}$ .

$$r_{\text{Hydrat}} = r_i + r_{\text{H}_2\text{O}}$$

Dabei ist  $r_i$  der van-der-Waals-Radius des  $i$ -ten Atoms nach den CHARMM-Parametern und  $r_{\text{H}_2\text{O}}$  der Radius der als kugelförmig betrachteten Wasser-Moleküle.

Alle Punkte von Kugeloberflächen, die nicht innerhalb einer benachbarten Kugel liegen, gehören zur zu benetzenden Oberfläche.

- Die *molekulare Oberfläche* (molecular surface, MS) hingegen ist die Grenzfläche zwischen den Gebieten mit unterschiedlichen Dielektrizitätskonstanten,  $\epsilon_{\text{intern}}$  und  $\epsilon_{\text{extern}}$ . Zur Bestimmung der molekularen Oberfläche ordnet man jedem Punkt der zu benetzenden Oberfläche eine Kugel vom Radius  $r_{\text{H}_2\text{O}}$  zu. Dann bilden alle Punkte dieser Kugeloberflächen mit Kontakt zum Molekül die Grenzfläche zwischen der wässrigen Umgebung und dem Molekül, also die sogenannte molekulare Grenzfläche. Die molekulare Oberfläche ist dadurch konkav und konvex gewölbt, weist aber keine Grate auf (ist also überall stetig differenzierbar). Allen Gitterpunkte des  $\epsilon$ -Gitters, die innerhalb der molekularen Oberfläche liegen, wird der Wert  $\epsilon_{\text{intern}}$  zugewiesen, denen außerhalb  $\epsilon_{\text{extern}}$  (s. Abschnitt 2.3.1).



*Abbildung 2.4:* Molekulare und zu benetzende Oberfläche. Die Atome des Moleküls sind als helle Kugeln dargestellt. Die Methode entspricht dem Abrollen eines hypothetischen, kugelförmigen Wassermoleküls (dunkel) vom Radius  $r_{\text{H}_2\text{O}}$  über die Protein-Oberfläche. Dabei ist die zu benetzende Oberfläche die Fläche, auf der sich der Mittelpunkt dieser Kugel bewegt. Die Fläche, die von der Kugel abgerollt wird, ist die molekulare Oberfläche. Im gezeigten Fall liegt das unterste Atom soweit im Inneren des Moleküls, dass es weder zur molekularen, noch zur zu benetzenden Oberfläche beiträgt.

## 2.4. Statischer Ansatz: Verzicht auf Dynamik

Die Reaktionsenthalpie  $\Delta G$  resultiert aus der Energie-Differenz zwischen den Edukten und den Produkten. Tatsächlich messbar ist aber nicht die aktuelle (zufällige) Energie eines jeden Reaktionsteilnehmers, sondern jeweils nur die über einen Zeitraum *gemittelte* Energie (s. Kap. 2.1.3):

$$\Delta G = \bar{E}_{P \cdot L} - (\bar{E}_P + \bar{E}_L)$$

Um die Reaktionsenthalpie  $\Delta G$  einer Komplexbildungsreaktion bzw. Bindestärke des Komplexes zu bestimmen, ist prinzipiell für jeden Reaktanden ( $P$ ,  $L$  und  $P \cdot L$ ) eine MD-Simulation durchzuführen und jeweils die mittlere Gesamtenergie zur Bestimmung der Reaktionsenthalpie  $\Delta G$  heranzuziehen. Dabei geht man davon aus, dass das berechenbare Zeit-Mittel dem experimentell zugänglichen Zeit- und Ensemble-Mittel entspricht [50]. Dieser Ansatz ist zwar realitätsgetreu, aber wegen der Größe von Proteinen sehr teuer im Sinne des Rechenaufwands.

Schneller und einfacher ist es, statt eines ganzen Ensembles jeweils nur eine einzige Konformation als Repräsentant für den jeweiligen Reaktionsteilnehmer zu verwenden. Das ist aber nur dann möglich, wenn sich die mittlere Gesamtenergie des Ensembles und Gesamtenergie des Repräsentanten entsprechen. Die folgende Überlegung dient dazu, abzuschätzen, ob diese Voraussetzung prinzipiell erfüllbar ist:

Zwei Konformationen mit den Energieinhalten  $\epsilon_1$  und  $\epsilon_2$  kommen bei der Temperatur  $T$  in den Häufigkeiten  $N_1$  und  $N_2$  vor:

$$\frac{N_1}{N_2} = e^{-\frac{(\epsilon_1 - \epsilon_2)}{kT}} \quad (2.23)$$

Demnach liegen bei 37°C zwei Konformationen des selben Moleküls bei einem angenommenen Energie-Unterschied von 1 kcal/mol in Mengenverhältnissen von 5:1 vor. Die Konformationen mit niedriger Energie tragen aufgrund ihrer größeren Häufigkeit im Ensemble zur mittleren Energie wesentlich mehr bei als diejenigen mit höherer Energie. Die Besetzung von Zuständen geht jeweils von der moleküleigenen Minimalenergie aus. Bei einem fiktiven Energie-Inhalt von 100 kcal/mol für  $\epsilon_1$  beträgt der Mittelwert der Gesamtenergie für dieses Ensemble 100.17 kcal/mol. Bei  $\Delta E = 2$  kcal/mol beträgt das Mengenverhältnis fast 26:1 und der Mittelwert 100.04 kcal/mol. In diesem Rechenbeispiel weichen die Mittelwerte demnach um 0.17 bzw. 0.04 kcal/mol von der minimalen Energie ab. Tatsächlich kommen in einem Ensemble wesentlich mehr als nur zwei verschiedene Zustände vor. Mengenmäßig relevant sind aber nur die Zustände mit den niedrigsten Energien, sodass anzunehmen ist, dass sich diese energetisch wesentlich weniger als 1 kcal/mol von dem Zustand mit minimaler Energie unterscheiden.

Demnach entspricht bei Körpertemperatur die mittlere Gesamtenergie eines Ensembles der minimalen Gesamtenergie der günstigsten Konformation bis auf einen möglichen Fehler von etwa 0.2 kcal/mol. Es reicht also voraussichtlich aus, von jedem Reaktionspartner nur die Konformation mit minimalem Energie-Inhalt als Repräsentante eines ganzen Ensembles zu berücksichtigen.

$$\Delta G = \bar{E}_{P,L} - (\bar{E}_P + \bar{E}_L) = E_{P,L}^{min} - (E_P^{min} + E_L^{min})$$

Statt der aufwendigen MD-Simulation genügt es also, nur die Konformation mit niedrigster Gesamtenergie für ein Molekül zu verwenden. Dieser „statische Ansatz“, der mit nur einer einzigen Konformation der Beweglichkeit der Moleküle keine Rechnung tragen kann, birgt somit auch Fehlerquellen:

- Die Konformation mit niedrigster Gesamtenergie muss gefunden werden. Mit Kristallstrukturen (KS) sind Informationen über die tatsächlich relevanten Konformationen verfügbar. Man kann daher erwarten, dass man mit einer KS wenn schon nicht die günstigste, dann doch eine hinreichend gute Konformation als Repräsentante für ein Ensemble besitzt.
- Entropische Effekte, die bei einer MD-Simulation implizit enthalten sind, werden bei diesem Ansatz ignoriert. Es steht auch kein geeignetes Modell bzw. keine empirische Beziehung zur Berechnung von Entropie-Änderungen zur Verfügung.

Da der wichtige Entropie-Beitrag zur Reaktionsenthalpie (s. Gl.2.6) systematisch fehlt, können nach diesem statischen Ansatz prinzipiell *keine Aussagen über die absolute Bindestärke*  $\Delta G$  eines Liganden an ein Protein gemacht werden.

Doch der geschilderte Ansatz bietet eine andere Möglichkeit: Um aus verschiedenen Kandidaten denjenigen Liganden auszuwählen, der am stärksten an ein gegebenes Protein bindet, bestimmt man die *relative Bindestärke*  $\Delta\Delta G$  in Bezug auf einen Komplex mit bekannter Bindestärke.

### 2.4.1. Relative Bindestärke von Komplexen

Bei der Suche nach guten Inhibitoren geht man meist von einem Referenz-Liganden  $L_o$  aus, der bereits eine nennenswerte Affinität  $\Delta G_o$  zum Zielenzym hat (Entwurf-Strategien s. Abschnitt 1.3). Man versucht, durch Variation von Substituenten neue Verbindungen mit größerer Bindestärke zu produzieren. Daher möchte man zuvor abschätzen können, welche der möglichen Modifikationen Erfolg versprechen. Die Frage lautet dann nicht mehr, wie groß der absolute Betrag der Bindestärke  $\Delta G_i$  einer neuen Verbindung  $L_i$  ist, sondern, ob  $\Delta G_i$  größer oder kleiner als die Bindestärke  $\Delta G_o$  des Referenz-Liganden  $L_o$  ist.

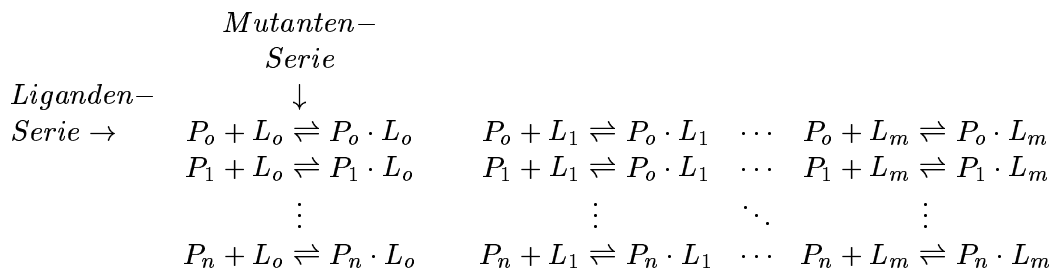
Angenommen, die Modifikationen am Grundgerüst des Referenz-Liganden sind gering, dann werden die entropischen Beiträge zur Reaktionsenthalpie  $\Delta G$  der jeweiligen Komplexierung von Referenz-Molekül  $L_o$  und dem Derivat  $L_i$  kaum beeinflusst werden (vgl. Gl 2.6).

Im Vergleich zwischen beiden Reaktionen sind dann nur die Beiträge der inneren Enthalpien  $\Delta H$  relevant, die mit Energie-Differenzen  $\Delta E$  gleichgesetzt werden:

$$\begin{aligned}
 \Delta\Delta G &= \Delta G_o - \Delta G_i \\
 &= \Delta H_o - \Delta H_i - T(\Delta S_o - \Delta S_i) && \text{für } \Delta S_o \simeq \Delta S_i \\
 &\simeq \Delta H_o - \Delta H_i && \text{wird } \Delta S_o - \Delta S_i = 0 \\
 &= (E_{P \cdot L_o} - (E_P + E_{L_o})) - (E_{P \cdot L_i} - (E_P + E_{L_i})) \\
 &= (E_{P \cdot L_o} - E_{L_o}) - (E_{P \cdot L_i} - E_{L_i}) && E_P - E_P = 0 \\
 &= \Delta E_{P \cdot L} - \Delta E_L
 \end{aligned} \tag{2.24}$$

Vergleicht man also zwei analoge Reaktionen miteinander, werden identische Beiträge und systematische Fehler eliminiert. Sogar die Bestimmung der Gesamtenergie des freien Proteins erübrigt sich. Stärkere oder schwächere Bindung des Kandidaten im Vergleich zur Referenz sind am Vorzeichen erkennbar.

Es bietet sich an, mehrere Reaktionen mit derselben Referenz-Reaktion zu vergleichen. In einer Serie von ähnlichen Reaktionen kann statt der Ligand-Komponente auch die Protein-Komponente variieren. Um relative Bindestärken einer beliebigen Kombination zu bestimmen, müssen folglich entweder Serien von verschiedenen Mutanten mit demselben Inhibitor oder Serien von unterschiedlichen Ligand-Derivaten mit dem selben Protein modelliert werden. Im Idealfall sind die erhaltenen Tabellen von relativen Bindestärken identisch, d.h. zwischen den Serien übertragbar.



## 2.5. Zusammenfassung

Um eine wirksame Therapie für eine Erkrankung zu finden, werden Liganden gesucht, die fest an ein ausgewähltes Ziel-Protein binden. Jede in dieser Arbeit erzeugte Protein-Konformation (frei oder im Komplex) basiert auf einer KS. Die Ligand-Konformationen im Komplex werden auf dieser Grundlage, nämlich der experimentell gesicherten Struktur des Proteins, konstruiert (s. Kap. 1.5). Die korrekte Konformation der Komplexe ist eine der Voraussetzungen für die zuverlässige Abschätzung von (relativen) Bindestärken. Dabei werden folgende Näherungen in Kauf genommen:

1. Ligand- und Protein-Moleküle werden als klassische Teilchen behandelt.
2. Die wässrige Umgebung von Ligand- bzw. Protein-Molekülen wird als homogene Masse ohne molekulare Struktur simuliert.



3. In die Berechnung von Bindestärken gehen nur die Grundkonformationen der Reaktionspartner ein.
4. Entropische Beiträge werden als konstant angenommen. Vergleicht man zwei Reaktionen, erübrigt sich die Berechnung entropischer Beiträge zur Bestimmung der relativen Bindestärke  $\Delta\Delta G$ .

Diese Vereinfachungen ermöglichen zwar schnelle Resultate, bergen aber auch folgende Fehlerquellen:

**Zu 1.:** Im klassischen Kraftfeld-Modell bleiben elektronische Effekte unberücksichtigt. Chemische Reaktionen können also nicht simuliert werden, nicht einmal einfache (De-)Protonierungsreaktionen. Doch gerade der Protonierungszustand saurer oder basischer Gruppen entscheidet über Anziehung oder Abstoßung durch die elektrostatischen Wechselwirkungen. Dieser Umstand macht die sorgfältige Kontrolle des Protonierungszustands erforderlich.

Andere spezielle Wechselwirkungen wie zum Beispiel die zwischen Aromaten, die sich bevorzugt entweder parallel übereinander schichten („ $\pi$ -stacking“) oder senkrecht zueinander anordnen, werden ignoriert. Sie werden lediglich als ungerichtete van-der-Waals-Kontakte behandelt.

**Zu 2.:** Beim auf Seite 29 eingeführten „continuum-solvent“-Ansatz gehen detaillierte Informationen über gerichtete Wechselwirkungen zwischen Protein- und Wassermolekülen verloren. Einzelne Wassermoleküle vermitteln oft über H-Brücken wichtige Wechselwirkungen zwischen zwei Komplex-Komponenten. Während einer Energie-Minimierung ohne explizite Wasser-Teilchen entfällt dieser Struktur-stabilisierende Effekt.

**Zu 3.:** Die Grundkonformation bzw. das globale Minimum der Gesamtenergie muss gefunden werden. Dazu muss entweder der Konformationsraum hinreichend gründlich abgesucht werden (nur praktikabel für kleine Moleküle), um sichergehen zu können, dass eine „ausreichend gute“ Konformation als Grundkonformation vorliegt, oder eine natürliche, experimentell belegte Konformation gegeben sein.

**Zu 4.:** Streng genommen gilt die Annahme identischer Entropie-Beiträge nur für isostere Moleküle, d.h. bei Verbindungen, die nicht über verschiedene Freiheitsgrade der Rotation verfügen und die gleiche Größe besitzen. In der Praxis werden jedoch meistens möglichst diverse Liganden als potentielle Wirkstoffe erprobt, sodass sich die Entropie-Beiträge unvorhersagbar ändern können.

**Fehlerbetrachtung:** Die erreichbare Genauigkeit der relativen Bindestärken  $\Delta\Delta G$  hängt von der Präzision der  $\Delta E_L$  bzw.  $\Delta E_{P,L}$ -Beiträge (hier am Beispiel einer Serie von Liganden) ab, d. h. davon, wie gut die Unterschiede zwischen den vergleichbaren Molekülen reproduzierbar sind. Bei einem angenommenen Fehler von 0.2 kcal/mol für die  $\Delta E_L$  bzw.  $\Delta E_{P,L}$ -Beiträge ist für die relativen Bindestärken  $\Delta\Delta G$  mit einer Unsicherheit von 0.4 kcal/mol

zu rechnen. Dissoziationskonstanten  $K_i$  zweier Komplexe, die sich um den Faktor 2 unterscheiden, sind also nicht mehr auflösbar. In der Praxis sind jedoch Unterschiede in den  $K_i$ -Werten bis zu einem Faktor von maximal 1:1000 relevant. Dies entspricht dem Vergleich eines sehr guten mit einem sehr schlechten Inhibitor, deren Bindestärken folglich um 4.3 kcal/mol abweichen. Demnach ist zu erwarten, dass die mit diesem Verfahren erreichbare Genauigkeit ausreicht, um zumindest qualitativ gute von schlechten Inhibitoren zu unterscheiden.

### 2.5.1. Konzept

Wegen der genannten Fehlerquellen ist zunächst zu prüfen, ob die gesuchten Bindestärken mit dieser Methode tatsächlich exakt genug berechenbar sind. Bevor Bindestärken blind vorhergesagt werden können, ist also eine sorgfältige *Validierung der Methode an experimentellen Daten* notwendig.

Da im Rahmen des statischen Ansatzes auf MD-Simulation zur Struktur-Optimierung verzichtet wird, ist die Energie-Minimierung entscheidend. Das Ergebnis einer Energie-Minimierung lässt sich durch die Vorgabe von Randbedingungen beeinflussen. Um geeignete Randbedingungen zu finden, werden die Resultate verschiedener Varianten mit tatsächlich gemessenen Daten abgeglichen. Ein geeignetes Verfahren zur Energie-Minimierung soll universell auf alle Vertreter einer Serie anwendbar und auch effizient sein. Das zum Abgleich benötigte experimentelle Referenz-Material muss den folgenden Anforderungen genügen:

- Neben umfangreichen Bindungsstudien auf deren Grundlage experimentelle Bindestärken zum Vergleich berechnet werden können, müssen auch entsprechende Struktur-Informationen vorliegen.
- Gemessene  $K_i$ -Werte sollten aus einer einzigen Messreihe, mindestens jedoch aus demselben Labor stammen, um unterschiedliche systematische Fehler zwischen den Messungen auszuschließen.
- Die Bindestärken sollten sich so deutlich voneinander unterscheiden, dass die Differenzen nicht unter die erreichbare Genauigkeit der relativen Bindestärke  $\Delta\Delta G$  fallen.
- Gleichzeitig müssen sich die chemischen Strukturen der Reaktanden bei einer Serie von Reaktionen möglichst ähnlich sein, damit sich systematische Rechenfehler durch schwankende Entropie-Beiträge und Ungenauigkeiten im Modell eliminieren.

In den folgenden experimentellen Kapiteln werden die geschilderten Methoden zur Konstruktion von Komplexen, zur Energie-Minimierung, zur MD-Simulation und zur Modellierung von Protein-Strukturen auf konkrete Fragestellungen angewendet. Die Resultate werden in Bezug zu experimentellen Daten gesetzt. Auf dieser Grundlage wird die Brauchbarkeit der Verfahren bewertet.

# 3. HIV-Protease-Inhibitoren

## 3.1. Einleitung

Um die tatsächliche Brauchbarkeit der im vorigen Kapitel geschilderten Methode zur Modellierung von Komplexierungsreaktionen auszuloten, sollen Bindungsstudien der HIV-Protease als Referenz herangezogen werden. Die einzelnen, zur Berechnung von relativen Bindestärken notwendigen Schritte sollen ausgeführt und besonders im Hinblick auf die erreichbare Genauigkeit der Werte für die relative Bindestärke  $\Delta\Delta G$  geprüft werden. Insbesondere sind Komplexe und freie Liganden zu modellieren.

### Komplex-Modellierung

- Ausgehend von einem *Stamm-Komplex*, der Kristall-Struktur eines Protein-Ligand-Komplexes, werden manuell einzelne Substituenten verändert<sup>1</sup>. Es resultieren *Konstrukte* mit gemeinsamem Bindemodus, aber mit regional willkürlicher und wahrscheinlich ungünstiger Konformation.
- Die Struktur jedes Konstrukts wird mittels *Energie-Minimierung* optimiert. Um die Vergleichbarkeit zwischen Stamm-Komplex und Konstrukt zu gewährleisten, wird auch die Energie der Stamm-Struktur mit denselben Randbedingungen minimiert.
- Für alle *Abkömmlinge*, also für die resultierenden struktur-optimierten Stamm-Komplexe und Konstrukte, wird die Gesamtenergie prinzipiell nach Gl. 2.17 und 2.18 berechnet.

**Ligand-Modellierung** Für die freien Liganden liegen im Gegensatz zu den Proteinen keine experimentellen Informationen vor, die ihre günstigste Konformation in Lösung belegen. Die Ligand-Moleküle sind aber so klein, dass die relevanten Konformationen durch kurze MD-Simulationen erzeugt werden können. Nach Berechnung der Gesamtenergien der Liganden werden die relativen Bindestärken  $\Delta\Delta G$  bestimmt. Die Brauchbarkeit dieses Verfahrens zur Vorhersage von Bindungseigenschaften wird nach der Übereinstimmung der berechneten mit den experimentellen Werten beurteilt.

---

<sup>1</sup>Alternativ zum Editieren könnten die verschiedenen Komplexe durch Docking erzeugt werden (s. Kap. 1.5.3). Davon wird jedoch abgesehen, da es aufwändiger ist und zudem die Gefahr abweichender Ligand-Platzierungen birgt, die systematische Fehler stärker als notwendig schwanken lassen.

## 3.2. Allgemeines über HIV

Ein Virus-Partikel besteht aus Erbgut in einer Schutzhülle. Meist verfügt es auch über einige wenige Enzyme; im Gegensatz zu einer Zelle fehlt ihm aber ein eigener Energie- und Stoffwechsel-Apparat. Es kann sich daher nur vermehren, wenn es den Stoffwechsel-Apparat einer Zelle zum Aufbau einer neuen Generation von Viren zweckentfremden kann. Dazu muss das Virus-Partikel zunächst eine intakte Zelle infizieren, d. h. es muss sein Erbgut und seine wenigen Enzyme durch die Zellmembran in die Zelle schleusen. Die Strukturen von Virus- und Zell-Oberfläche entscheiden darüber, ob ein Virus im Kontakt mit einer bestimmten Zelle diese infizieren kann. Viren sind daher auf verschiedene Tier- oder Pflanzen-Arten bzw. Zell-Typen spezialisiert.

Das humane Immunschwäche Virus (*HIV*) ist ein *Retrovirus*, d. h. sein Erbgut liegt als RNS statt als DNS vor. Durch Blut-Kontakt mit einer Virus-haltigen Körperflüssigkeit (besonders Blut und Sperma) wird es von Mensch zu Mensch übertragen. Es befällt bestimmte Blutzellen (*CD4-Lymphozyten*, sog. T-Helfer-Zellen) und zerstört sie bei seiner Vermehrung. Die Aufgabe der CD4-Lymphozyten im Körper ist es, bei Kontakt mit körperfremdem Material Alarmstoffe auszusenden, die andere Zellen der Immunabwehr anlocken, aktivieren und zur Vermehrung anregen. Über diesen indirekten Weg zur Fremdkörper-Bekämpfung wird die Immunantwort kaskadenartig verstärkt. Bei einer Zerstörung der CD4-Lymphozyten durch HIV bleibt daher im Fall einer Infektion durch einen *zweiten* Erreger die volle Wucht der Abwehr gegen diesen aus, sodass eine sonst harmlose Infektion heftig, häufig auch tödlich verläuft. Eine HIV-Infektion äußert sich daher (oft erst nach langer Inkubationszeit) als erworbene, d. h. nicht angeborene, allgemeine Abwehrschwäche (acquired immuno deficiency syndrome, AIDS).

Solange sich die Virus-Partikel frei im Blut befinden, sind sie selbst den Angriffen der körpereigenen Abwehr ausgesetzt; es werden Antikörper gegen die Proteine an ihrer Außenseite gebildet. Diese Oberflächen-Proteine unterscheiden sich aber durch die *hohe Mutationsrate* der Viren von Generation zu Generation. Durch die kurze Generationsdauer des Virus ist die vollständige Entfernung von Virus-Partikeln aus dem Blutplasma durch die Körper-Abwehr kaum möglich.

Um die Viren medikamentös zu unterdrücken ohne gleichzeitig auch den Wirtsorganismus zu schädigen, kommen vor allem virale Strukturen als Ziel für eine Blockade durch geeignete Wirkstoffe in Frage. Die Strategie zur Bekämpfung einer HIV-Infektion muss also auf der molekularen Organisation des Virus bzw. auf dem Ablauf seines Vermehrungszyklus beruhen.

### 3.2.1. Vermehrungszyklus

**Aufbau:** Ein infektiöses Virus-Partikel besitzt eine *Membran-Hülle*, die von der ehemaligen Wirtszelle seines Vorfahren stammt und von viralen Proteinen durchsetzt ist. Ein *Transmembran-Protein* (TM) verankert das *Oberflächen-Protein* (surface: SU) an der Außenseite der Membran. Die Innenseite wird bedeckt von *Matrix-Proteinen* (MA). Die Membran-Hülle hat einen Durchmesser von 1000 – 3000 Å und schließt zusammen mit

einigen zellulären Proteinen eine virale *Protease* (PR) sowie das konisch geformte *Capsid* (CA) ein. Das Capsid ist ein Komplex aus den beiden einzelnen *RNS*-Strängen und einigen Proteinen: *Nucleocapsid* (NC), *Integrase* (IN) und *Reverse Transkriptase* (RT) [131]. Die Produktion einer neuen Generation von Viren verläuft über die folgenden Schritte (s. Abb. 3.1):

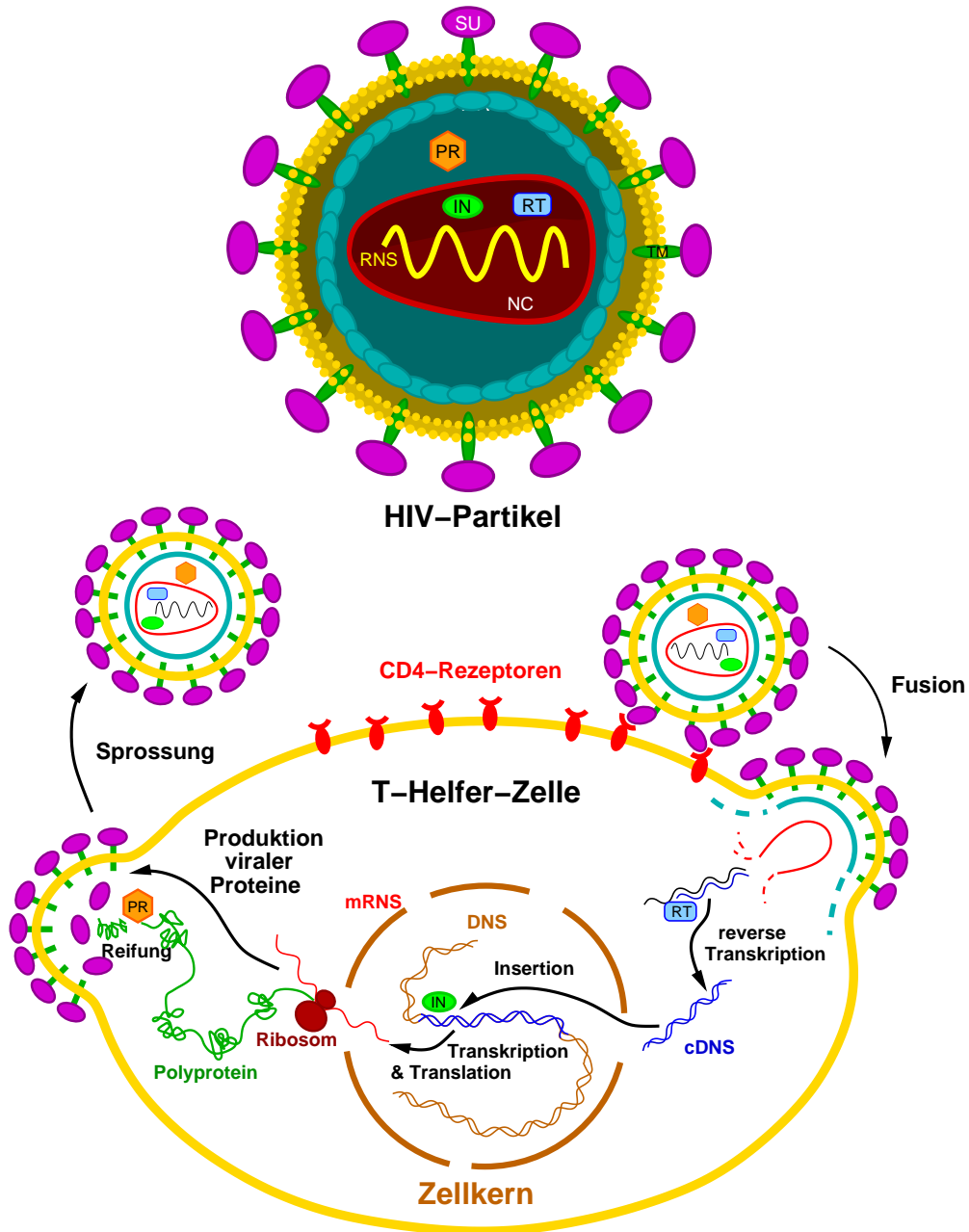


Abbildung 3.1.: HIV-Aufbau und Lebenszyklus

- 1. Fusion mit der Wirtszelle:** Bei Kontakt mit einer CD4-Zelle verschmelzen Zell- und Virus-Membran über einen komplizierten Prozess miteinander, sodass der CA-RNS-Komplex und die viralen Enzyme in die Zelle geschleust werden. Eine Schlüsselrolle für die Fusion spielt der Konformationswechsel von TM. Einmal in der Zelle, ist das Virus vor Angriffen der Körper-Abwehr geschützt.
- 2. Insertion des viralen Erbguts ins Wirtsgenom:** Die virale RT erzeugt aus zelleigenen Nukleotiden zunächst einen zur RNS komplementären DNS-Einzelstrang. Dieses RNS/DNS-Hybrid wird unter Abbau der RNS in doppelsträngige DNS überführt. Die RT arbeitet recht ungenau, ungefähr eines von 20.000 Nukleotiden wird falsch kopiert. Daher ist die Mutationsrate der HI-Viren so hoch [105]. Die IN fügt den DNS-Doppelstrang in den Doppelstrang eines Chromosoms der Wirtszelle ein. In dieser Form ruht die virale DNS innerhalb der Wirtszelle. Diese wird davon nicht beeinträchtigt.
- 3. Produktion viraler Proteine:** Werden die viralen Gene aktiviert, erzeugt der Transkriptions-Apparat der Wirtszelle eine Vielzahl von RNS-Kopien davon. Dieses Primärtranskript wird in zwei Teile gespalten, die den zelleigenen Ribosomen als Vorlage zum Aufbau der Polyproteine A und B dienen. In den Polyproteinen sind die späteren viralen Einzelproteine noch nahtlos aneinander gereiht. Polyprotein A ist das Vorläufer-Molekül von MA, CA und NC, Polyprotein B von MA, CA, NC, PR, RT und IN. Ein weiteres Polyprotein enthält die SU- und TM-Domänen.
- 4. Sprossung und Reifung der viralen Partikel:** Die Polyproteine sammeln sich zu Hunderten bis Tausenden zusammen mit neuer viraler RNS an der Innenseite der Zellwand. Die sprossenden Partikel stülpen die Zell-Membran aus, schnüren sich davon ab und gelangen so mit ihrer Membran-Hülle ins Blutplasma. Nach dieser Sprossung spaltet die PR die Polyproteine A und B in die funktionsfähigen Domänen (Reifung). Dadurch erst werden die Partikel infektiös. Die massenhafte Produktion von Virus-Partikeln innerhalb der Wirtszelle verbraucht ihre Vorräte. Die Wirtszelle geht dabei zugrunde.

### 3.2.2. Behandlungsstrategien

Aus dem geschilderten Vermehrungszyklus ergeben sich die folgende Angriffspunkte zur Bekämpfung der Viren innerhalb des Körpers nach einer Infektion.

1. Membran-Fusion verhindern → Konformationswechsel von TM blockieren.
2. Erbgut-Kopie verhindern → RT blockieren.
3. Partikel-Reifung verhindern → PR blockieren.
4. Brutkammern vernichten → Apoptose der Wirtszelle auslösen.

- Zu 1.:** Mit dem synthetischen Peptid T-20, welches an das Transmembran-Protein des Virus bindet, kann die Fusion der viralen und zellulären Membranen erfolgreich verhindert werden [75].
- Zu 2.:** Zwei Wirkstoff-Gruppen gegen die RT sind im klinischen Einsatz, sechs verfügbare *Nucleosid-Analoga* und drei *Non-Nucleosid-Analoga*. Nucleosid-Analoga sind verstümmelte Substrate, die regulär in eine entstehende DNS-Kette eingebaut werden, aber eine weitere Ketten-Verlängerung verhindern. Diese Verbindungen werden aber auch in zelluläre DNS-Stränge eingebaut. Dadurch wird die normale Zellteilung beeinträchtigt, was schwere Nebenwirkungen zur Folge hat. Non-Nucleosid-Analoga inhibieren im Gegensatz dazu nur die virale RT und verursachen dadurch weniger Nebenwirkungen. Probleme bereitet jedoch das häufige Auftreten resistenter RT-Mutanten [88, 32].
- Zu 3.:** Fünf *Protease-Blocker* stehen der HIV-Therapie zur Verfügung. Auch bei dieser Wirkstoff-Gruppe besteht die Gefahr der Resistenz-Entwicklung [81, 84].
- Zu 4.:** Eine infizierte T-Helfer-Zelle muss früher oder später zu Grunde gehen. Dieser Verlust ist für den Gesamt-Organismus unvermeidlich, aber leichter zu verschmerzen, wenn sie dabei keine weiteren Helfer-Zellen durch die Freisetzung einer neuen Generation von Virus-Partikeln gefährdet. Daher soll gezielt bei infizierten Zellen der Selbstmord durch die Einschleusung des Apoptose-Promotors TAT-Caspase 3 ausgelöst werden. Dieser Ansatz von der Gruppe um S.F. Dowdy [134] aus dem Jahr 1999 befindet sich noch nicht in einer klinischen Testphase.

Bisher werden in der klinischen Praxis HIV-Infektionen meist mit einer Kombination von bis zu drei Wirkstoffen behandelt, die parallel auf die RT und die PR abzielen.

### **Therapeutische Ansätze, Probleme und Ziele**

Eine erfolgreiche medikamentöse Therapie senkt den Pegel der im Blut vorhandenen Virus-Partikel unter die Nachweisgrenze. Die aktue Bedrohung des Patienten ist dann zwar gebannt, doch ruhende Viren, deren Erbgut innerhalb der Wirts-DNS geschützt ist, stellen weiterhin eine potentielle Gefahr dar: Immer wieder werden solche "Schläfer" aktiviert und treten in die Phase der Partikel-Produktion ein. Unter dem Einfluss einer Kombination von wirksamen Inhibitoren gegen die PR und die RT werden jedoch entweder keine infektiösen Partikel gebildet oder die Produktion von viraler DNS wird blockiert, sodass eine weitere Ausbreitung ausbleibt.

**Resistenz-Bildung unter Selektionsdruck:** Aufgrund der hohen Fehlerrate der RT unterscheidet sich die DNS-Sequenz einer neuen Viren-Generation meist in mindestens einem Basen-Paar von ihrer Vorlage. Daher kommen innerhalb einer Viren-Population viele *Mutanten* neben dem ursprünglichen *Wild-Typ* vor. Manche Mutanten besitzen Enzyme mit geringerer Affinität zum jeweiligen Inhibitor als der Wild-Typ. Kann ein solches Enzym

trotz der Mutation auch seine Funktion im Vermehrungszyklus erfüllen, wird das Wachstum *dieses* Stammes nicht mehr durch das zuvor wirksame Mittel gebremst. So entsteht ein resistenter Stamm,<sup>2</sup> der schnell mengenmäßig innerhalb der Viren-Population vorherrscht.

Der Erfolg einer Kombinationstherapie wird regelmäßig kontrolliert, vor allem auch durch Sequenzierung der viralen RNS bei *niedriger* Viruslast im Blut.

Es wird angestrebt, sofort bei einem Wechsel des vorherrschenden Virus-Stammes auf gegen den neuen Stamm wirksamere Medikamente umzustellen, *bevor* die Viruslast steigt. Dazu jedoch muss man die Affinität von verschiedenen Inhibitoren an die verschiedenen Mutanten eines Proteins vorhersagen können.

Nach einem statistisch-empirischen Ansatz [10] kann die Therapie anhand von Sequenz-Daten und Resistenz-Profilen optimiert werden. Eine Modellierung von Komplexen mit dem Ziel der Vorhersage von Bindestärken, d. h. der Vorhersage der potentiellen Wirksamkeit einzelner Wirkstoffe, könnte diese Bemühungen wertvoll ergänzen.

Von den für Modellierungsversuche in Frage kommenden Ziel-Proteinen wird hier die PR ausgewählt, da sie relativ klein ist, nur ein einzelnes Substrat umsetzt und experimentell sowie Röntgen-kristallographisch gut charakterisiert ist.

### 3.3. Zielenzym HIV-1-Protease

#### 3.3.1. Eigenschaften und Eigenarten

Die HIV-PR ist ein *Dimer* aus zwei identischen Untereinheiten und besitzt  $C_2$ -Symmetrie. Die Monomere A und B sind jeweils 99 AS lang [111]. Die Hauptmassen der Monomere bilden einen relativ starren Grundkörper mit einer quer verlaufenden Rinne [1]. Zwei bewegliche  $\beta$ -Faltblatt-Schleifen verschließen diese Rinne flügelartig zu einem Kanal. Wie bei einer Schere müssen sich die Flügel öffnen, damit das Polyprotein ans aktive Zentrum gelangen kann, und wieder schließen, damit die Spaltung des Polyproteins stattfinden kann. NMR-Untersuchungen und die Temperatur-Faktoren aus der RKSA zeigen, dass das ganze Enzym und besonders die *Flügel sehr beweglich* sind [71]. Nach Scott et al. [124] rollen sich die Flügel an den Spitzen ein und ermöglichen so dem Polyprotein den Zutritt zum aktiven Zentrum.

Gebunden an die PR ist die Hauptkette des Polyproteins gestreckt und reicht mit 7–8 AS durch die PR hindurch. H-Brücken werden hauptsächlich zur immer gleichen Hauptkette gebildet. Im Boden des ansonsten unpolaren Kanals liegen die katalytisch aktiven Gruppen der Protease, die Aspartate 25A und 25B (*Aspartat-Diade*). Ihre  $\text{CO}_2^-$ -Gruppen stehen einander gegenüber und fixieren über H-Brücken ein Wasser-Molekül. Eine der  $\text{CO}_2^-$ -Gruppen ist deprotoniert, sodass das Sauerstoff-Atom des Wasser-Moleküls stark negativ polarisiert wird. Die Hydrolyse des Polyproteins verläuft Basen-katalysiert über dieses sog. *lytische Wasser-Molekül*, welches eine Carbonyl-Gruppe der Hauptkette des Substrates nucleophil angreift (s. Abb. 3.2).

---

<sup>2</sup>Von den vielen Mutanten innerhalb einer Population zu unterscheiden sind die beiden Virus-Typen 1 bzw. 2.



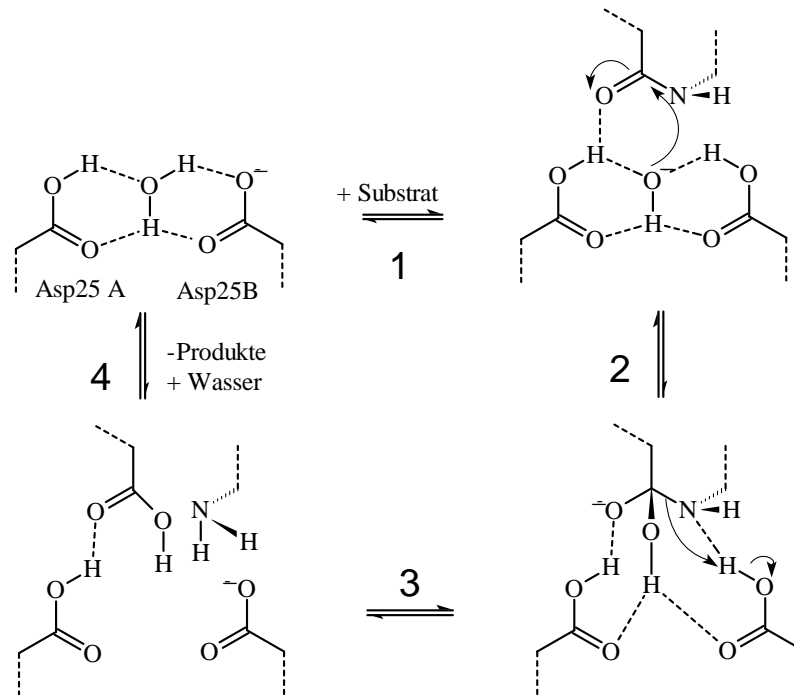


Abbildung 3.2.: Mechanismus der Peptid-Hydrolyse:

1. Das Substrat bindet an das aktive Zentrum.
2. Das lytische Wasser-Molekül greift die Carbonyl-Gruppe nucleophil an.
3. Das entstandene Intermediat eliminiert die Amino-Gruppe.
4. Die Produkte lösen sich vom aktiven Zentrum.

Die PR spaltet die viralen Polyproteine an 9 verschiedenen Stellen [124], weist also eine *geringe Substrat-Spezifität* auf; Sequenzen mit überwiegend unpolaren Seitenketten werden jedoch leicht bevorzugt. Die Oberfläche der PR passt sich dem jeweiligen Substrat an, indem die Seitenketten des Polyproteins (bezeichnet mit P1–P4 bzw. P1'–P4' ausgehend von der zu spaltenden Bindung) jeweils von kleineren Nischen (*subsites*, entsprechend mit S1–4 bzw. S1'–S4' bezeichnet) aufgenommen werden. Die PR toleriert ein breites Spektrum verschiedener Mutationen, bleibt also häufig trotz der Mutation aktiv genug, um Partikel reifen zu lassen.

### 3.3.2. Substratbindung

Die Affinität der HIV-PR zu ihren Substraten wird folgendermaßen gesteuert:

Enthalpisch: Die katalytischen Aspartate sind mit der zu spaltenden Peptid-Gruppe über *H-Brücken* verbunden. Weitere H-Brücken werden zwischen den Atomen der Substrat-Hauptkette und der PR ausgebildet [54].

Entropisch: Die mit Wasser zu benetzenden Oberflächen des ausgestreckten Substrats und die des Bindekanals werden bei der Komplexbildung verringert. Die *Verringerung der zu benetzenden Oberfläche* trägt wesentlich zur Bindung bei [86]. Andererseits stehen der *Flexibilitätsverlust des Substrats* und der beweglichen Flügel der PR der Komplexbildung entgegen und erleichtern die Dissoziation der Produkte vom Enzym [86].

### 3.3.3. Modellierung von Inhibitor-Komplexen

Um das Enzym blockieren zu können, sucht man Moleküle, die an schon bewährte Ankerpunkte binden (s. Kap. 1.3). Die verfügbaren Medikamente (s. Abb. 3.3) imitieren folglich das Substrat: Es sind Oligopeptid-ähnliche Verbindungen, bei denen die zentrale, normalerweise spaltbare Peptid-Gruppe durch eine nicht hydrolysierbare sekundäre Alkohol-Gruppe bzw. eine Glykol-Gruppe ersetzt ist.

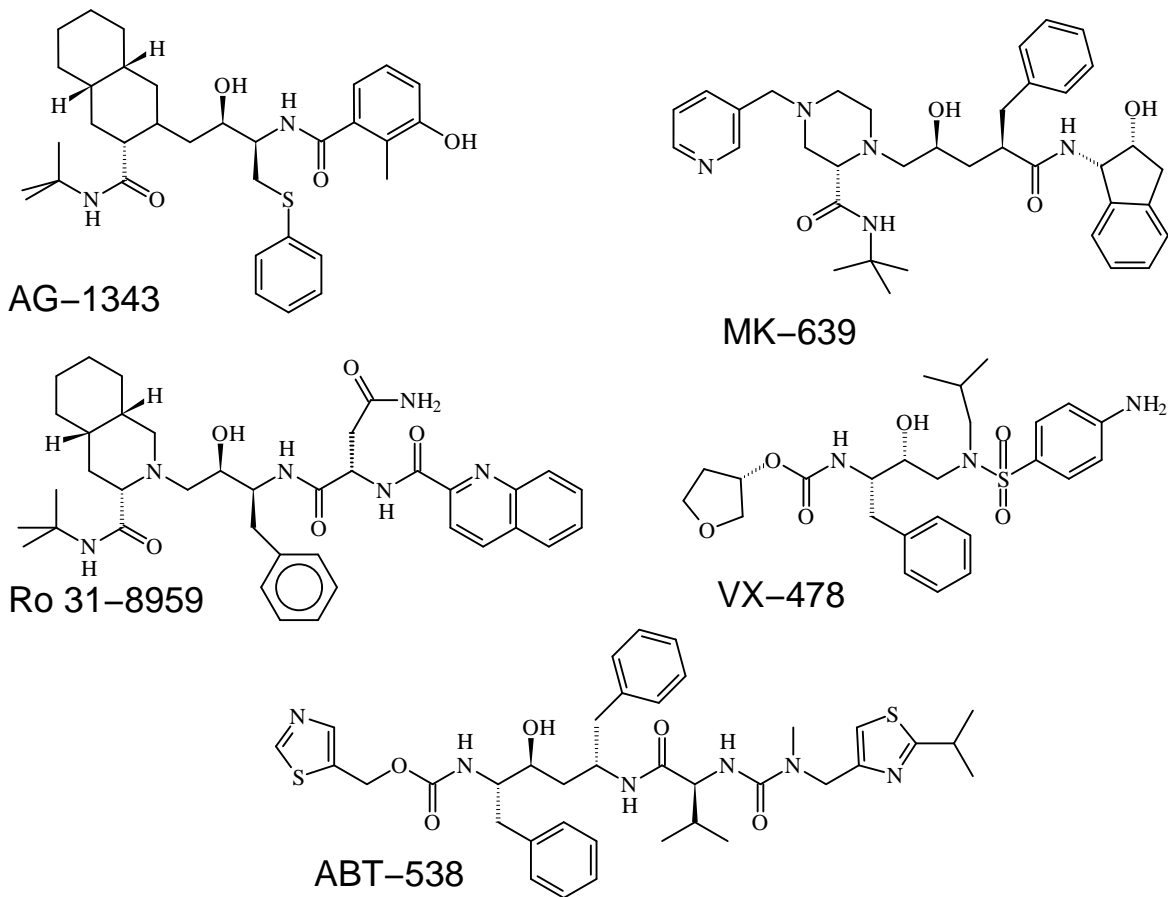


Abbildung 3.3.: Medikamente

Für eine möglichst effektive Therapie der HIV-Infektion wäre es wichtig, die Bindestärke dieser Medikamente an neu auftretende HIV-PR-Mutanten vorhersagen zu können. Um die in Kap. 2 skizzierten Modellierungsmethoden zu erproben und auf Brauchbarkeit zu überprüfen, ist diese Gruppe der klinischen Wirkstoffe als Test-Beispiel jedoch ungeeignet und zwar aus folgenden Gründen: Diese Art von Liganden besitzt viele frei drehbare Bindungen. Diese Flexibilität ist im Hinblick auf die Modellierung ein Nachteil wegen des unkalkulierbaren Entropie-Verlustes bei der Komplexbildung. Außerdem sind diese Inhibitoren strukturell sehr unterschiedlich. Der Bindemodus ist im Einzelfall ungewiss, denn es existiert lediglich eine einzige KS von einem Ro-31-8959-Komplex mit der HIV-PR. Von dieser Struktur aus müssten die anderen Komplexe abgeleitet werden. Für das Sammeln erster Erfahrungen mit der Modellierung von Komplexen, scheidet diese Gruppe von Liganden daher aus.

**Auswahl eines Test-Systems:** Die nächste Generation von Inhibitoren hat sich vom Substrat als Struktur-Vorlage gelöst. Die Verbindungen sind meist unpolar und klein bzw. rigide, um den unerwünschten Entropie-Verlust bei der Komplexbildung zu minimieren. Speziell für solche Liganden scheint der statische Ansatz zur Bestimmung relativer Bindestärken gut geeignet zu sein. Für zwei Ligand-Familien, die zyklischen Harnstoff-Derivate (CHD) und die Cyclooctylpyranon-Derivate (COPS), sind sowohl Bindungsstudien, als auch KS verfügbar. Sie bieten sich also aufgrund ihrer Eigenschaften und der Datenlage für Modellierungsstudien an.

Doch zuvor sind folgende Punkte zu klären:

1. Von welcher Qualität sind die gegebenen Struktur-Daten; mit welchen Energie-Inhalten korrelieren sie?
2. Wie gut können Komplexe ausgehend von einer KS allein durch Energie-Minimierung modelliert werden?
3. Wie sind die freien Liganden zu behandeln?
4. Lassen sich relative Bindestärken für eine Serie von Liganden reproduzieren?

**Zu 1.:** Für den Liganden DMP-323 im Komplex mit der Mutante C95A sind NMR-Daten verfügbar. Struktur-Daten aus NMR-Experimenten enthalten Informationen über die Vielfalt der in Lösung relevanten Konformationen. Ein solches Ensemble von Konformationen bietet sich an zur Untersuchung der natürlichen Schwankungsbreite der Komplex-Energie in Lösung, die die Genauigkeit der zu berechnenden Bindestärke  $\Delta G$  begrenzt (vgl. erreichbare Genauigkeit Abschnitt 2.4 bzw. 2.5). Gleichzeitig können Erfahrungen mit verschiedenen Verfahren zur Energie-Minimierung gesammelt werden.

**Zu 2.:** Von den Liganden DMP-323 und DMP-450 im Komplex mit der HIV-PR existieren zwei KS. Indem von jeder KS aus der jeweils andere Komplex erzeugt wird, kann überprüft werden, inwieweit die relativen Komplex-Energien  $\Delta E_{P,L}$  abhängig von

der Stamm-Struktur sind (s. Gl. 2.24). Dasjenige Verfahren der Energie-Minimierung, nach welchem sich die beste Übereinstimmung erzielen lässt, muss ermittelt werden.

**Zu 3.:** Anders als bei den Komplexen, liegen für die Liganden keine Hinweise auf die in Lösung relevante Konformation vor; lediglich die Konformation des Liganden *im Komplex* mit der HIV-PR ist bekannt. Die jeweils repräsentative Konformation ist durch MD-Simulation zu erzeugen.

**Zu 4.:** Für eine Serie von 10 Cyclooctylpyranon-Derivaten sollen relative Bindestärken berechnet werden. Die Liganden sind so ausgewählt, dass sie sich strukturell nur wenig, in der Bindestärke aber deutlich voneinander unterscheiden. Dann sind Binde-  
modus und Flexibilitätsverlust voraussichtlich gleich und dann werden die relativen Bindestärken von (berechenbaren) energetischen Effekten dominiert, sodass man erwarten kann, die Affinitäten tendenziell gut wiedergeben zu können.

### 3.4. Studie eines NMR-Ensembles

Ein Protein nimmt in wässriger Lösung auch im Komplex mit einem Liganden viele – mit NMR-Methoden unterscheidbare – Konformationen ein. Jede individuelle Konformation besitzt ihren eigenen Energie-Inhalt. Die Untersuchung eines solchen Konformationsensembles soll folgende Fragen beantworten:

- Wie sehr unterscheiden sich die Konformationen des Ensembles in Bezug auf ihre Atom-Positionen und ihren Energie-Inhalt?
- Lassen sich die verschiedenen Konformationen bzw. ihre Energie-Inhalte durch eine Energie-Minimierung vereinheitlichen?
- Sind die Unterschiede zwischen den Konformationen innerhalb des Komplexes zufällig oder systematisch verteilt?
- Lassen sich relevante Unterschiede von irrelevanten separieren?

#### 3.4.1. Durchführung

Der PDB-Eintrag 1bve enthält die Koordinaten von 28 verschiedenen Konformationen einer HIV-PR-Mutante (I3V, C95A) im Komplex mit dem zyklischen Harnstoff-Derivat (DMP-323) als Liganden (s. Abb.3.4). Unter 1bvg findet man eine gemittelte Konformation dieses Ensembles. Alle Konformationen werden einer Energie-Minimierung unterzogen. Als Referenz für strukturelle und energetische Vergleiche wird jeweils die gemittelte Struktur verwendet.

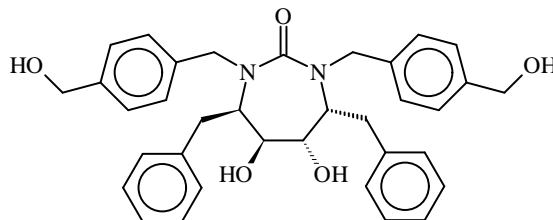


Abbildung 3.4: Das zyklische Harnstoff-Derivat DMP-323.

**Struktur-Optimierung:** Zwei Protokolle zur Energie-Minimierung werden alternativ mit verschiedenen Randbedingungen durchgeführt:

**Alternative 1** Eine Minimierung unter elastischer Bindung aller Atome an ihre gegebenen Positionen (Kraftkonstante  $8 \text{ kcal/mol} \cdot \text{\AA}^2$ ) soll die individuellen Züge jeder Konformation erhalten. Lediglich kleine Atomverschiebungen, die ungünstige Bindungslängen oder Kollisionen korrigieren, werden zugelassen. Die Bindung an die Start-Positionen wird zudem massengewichtet, d.h. H-Atome können leichter verschoben werden als C-, N- und O-Atome.

**Alternative 2** Eine Minimierung ohne jede Fixierung der Atome soll die zu Beginn verschiedenen Konformationen entweder in eine einzige *gemeinsame* Konformation konvergieren lassen oder *noch unterschiedlichere* Konformationen erzeugen.

Für die Berechnung der elektrostatischen Wechselwirkungen wird in beiden Fällen  $\varepsilon = 2r$  verwendet, wobei  $r$  der Abstand der jeweiligen Ladungen ist. Mit diesem Wert werden die Wechselwirkungen von eng benachbarten Ladungen stärker gewichtet als diejenigen in größerer Entfernung ( $F_{Coulomb} \propto 1/r^2$ , s. Gl. 2.1). Dadurch soll der Einfluss einer wässrigen Umgebung während der Energie-Minimierung des Komplexes ohne explizite Wasser-Moleküle nachgeahmt werden. Der sog. Cutoff-Wert wird so gewählt, dass elektrostatische Wechselwirkungen nur für Atom-Paare berücksichtigt werden, die maximal  $99 \text{ \AA}$  voneinander entfernt sind. Im Fall der HIV-Protease mit einem maximalen Durchmesser von ca.  $60 \text{ \AA}$  sind dies alle möglichen Atom-Paare.

Die Minimierungen werden nach dem *conjugated-gradient*-Verfahren bis zur Konvergenz durchgeführt.

### Struktur-Vergleiche:

- Jede resultierende Konformation wird mit der unter denselben Bedingungen minimierten Referenz-Struktur ( $rmsd^{ref}$ ) bzw. mit ihrer jeweiligen Start-Konformation ( $rmsd_{Min}^{Start}$ ) verglichen.
- Um abzuschätzen, wie sehr der Bindemodus des Liganden festgelegt ist, werden zwischen zwei Strukturen radiale  $rmsd$ -Verteilungen mit dem Liganden als Zentrum des Komplexes ermittelt. Dazu werden für zunehmende Entfernungen vom Liganden die

mittleren *rmsd*-Werte der Atome in 2 Å-Intervallen bestimmt. H-Atome werden hierbei nicht berücksichtigt, denn ihre Positionen sind weitgehend durch die schwereren Atome festgelegt.

**Energie-Rechnungen:** Die Gesamtenergie eines Komplexes wird als Summe aus Kraftfeld-Energie und Solvatationsenergie berechnet (vgl. Kap.2.3). Für Energie-Rechnungen werden  $\varepsilon_{\text{Kraftfeld}} = \varepsilon_{\text{intern}} = 4$  und  $\varepsilon_{\text{extern}} = \varepsilon_{\text{H}_2\text{O}} = 78.5$  eingesetzt.

Einige Teilbeiträge zur Gesamtenergie werden gesondert aufgeführt. Um die Beiträge des Liganden und seiner Coulomb- bzw. Lennard-Jones-Wechselwirkungen mit dem Protein zu isolieren, wird der Einfluss peripherer und daher möglicherweise irrelevanter Konformationsunterschiede ausgeblendet, indem jeweils der Beitrag des Protein-Anteils von dem des Komplexes abgezogen wird:

$$E_{L+WW}^{\text{Kraftfeld}} = E_{P.L}^{\text{Kraftfeld}} - E_P^{\text{Kraftfeld}} \quad (3.1)$$

$$E_{L+WW}^{\text{Solvation}} = E_{P.L}^{\text{Solvation}} - E_P^{\text{Solvation}} \quad (3.2)$$

Diese abgeleiteten Energie-Beiträge  $E_{L+WW}^{\text{Kraftfeld}}$  und  $E_{L+WW}^{\text{Solvation}}$  liefern zusammen den Beitrag des Liganden zur Gesamtenergie des Komplexes:

$$E_{L+WW}^{\text{Gesamt}} = E_{L+WW}^{\text{Kraftfeld}} + E_{L+WW}^{\text{Solvation}} \quad (3.3)$$

### 3.4.2. Ergebnisse

#### Zustand vor Energie-Minimierung

Tabelle 3.1 enthält die *rmsd*- und Energie-Daten zu den Start-Konformationen *vor* einer Minimierung.

**Struktur-Vergleich:** Die 28 einzelnen Konformationen des NMR-Ensembles weichen im Durchschnitt um nur 1.49 Å von der gemittelten bzw. Referenz-Struktur ab. Berücksichtigt man lediglich die  $C_\alpha$ -Atome des Proteins, sind es sogar nur 0.93 Å. In Abb. 3.5 sind die  $C_\alpha$ -Atome jeder Konformation nach ihrem Abstand zum jeweiligen Referenz-Atom gefärbt.

Deutlich ist eine bei allen fast deckungsgleiche, dunkle Region im Inneren des Komplexes zu erkennen, während in der Peripherie hellere Farbtöne vorkommen, die größeren strukturellen Abweichungen entsprechen. Die Bestimmung radialer *rmsd*-Werte für alle nicht-H-Atome (s. Tab. 3.5 und Abb. 3.6) bestätigt, dass die Atom-Positionen des Liganden selbst und die seiner unmittelbaren Nachbarschaft eindeutiger festgelegt sind, als die Peripherie des Proteins. Die durchschnittliche Abweichung der einzelnen Konformationen von der Referenz-Struktur beträgt hier sogar nur 0.94 Å, was zeigt, dass die Positionen der H-Atome wesentlich stärker variieren als die Gerüst-Atome (C, N, O, S).

**Energie-Vergleich:** Trotz der großen strukturellen Gemeinsamkeiten schwanken die Energie-Inhalte der einzelnen Konformationen sehr stark. Sie erstrecken sich bei der Kraftfeld-Energie über einen Bereich von fast 570 kcal/mol, bei der Solvatationsenergie über eine Spanne von ca. 100 kcal/mol und bei der Gesamtenergie um ca. 400 kcal/mol bei einer Standard-Abweichung von  $\pm 126$  kcal/mol.

**Protein-Ligand-Wechselwirkungen:** Wird der Beitrag der Protein-Konformation ausgeblendet, verbessert sich die Genauigkeit für die Gesamtenergie fast um den Faktor 5 auf eine Standard-Abweichung von  $\pm 26$  kcal/mol. Diese Beobachtung deckt sich mit dem Befund aus den radialen *rmsd*-Vergleichen (s. oben), nach dem die Binde-tasche des Proteins, die über Lennard-Jones- und Coulomb-Wechselwirkungen mit dem Liganden in Kontakt steht, wesentlich besser definiert ist als die Peripherie.

### Resultate nach Minimierung 1

Tabelle 3.2 zeigt die *rmsd*- und Energie-Daten nach dem Minimierungsverfahren 1.

**Struktur-Vergleich:** Mit einer Energie-Minimierung unter harmonischer Fixierung der Atome an ihre Start-Positionen erreicht man allgemein geringfügige Verschiebungen, im Mittel um  $0.4 \pm 0.25$  Å relativ zur jeweiligen Start-Struktur. Die Konformationen ändern sich kaum, im Vergleich zur Referenz-Struktur bleibt die durchschnittliche mittlere Abweichung der Atom-Positionen von zuvor 1.49 bei nun 1.50 Å fast unverändert. Das Profil der radialen *rmsd*-Verteilung folgt demjenigen vor der Minimierung, allerdings um ca. 0.15 Å verschoben zu geringeren Werten (s. Tab. 3.5 bzw. Abb. 3.6).

**Energie-Vergleich:** Der Beitrag der Kraftfeld-Energie wird durch diese durchweg geringen Atom-Verschiebungen drastisch um etwa 1650 kcal/mol auf durchschnittlich  $-1067 \pm 67$  kcal/mol gesenkt, die Standard-Abweichung wird gegenüber dem Zustand vor Minimierung fast halbiert. Gleichzeitig wird aber der Beitrag der Solvatationsenergie ( $-372 \pm 17$  kcal) um ca. 310 kcal/mol ungünstiger, ohne aber die Schwankungsbreite signifikant zu verändern. Insgesamt verringert sich die Standard-Abweichung bei der Gesamtenergie dennoch fast um den Faktor 2 ( $-1439 \pm 69$  kcal/mol).

**Protein-Ligand-Wechselwirkungen:** Ohne Berücksichtigung der Protein-Konformation, reduziert sich die Standard-Abweichung der Protein-Ligand-Wechselwirkungen wiederum merklich, wobei auch hier die stärkeren Schwankungen des Kraftfeld-Beitrags die erreichbare Genauigkeit auf  $3.2 \pm 9.0$  kcal/mol begrenzen.

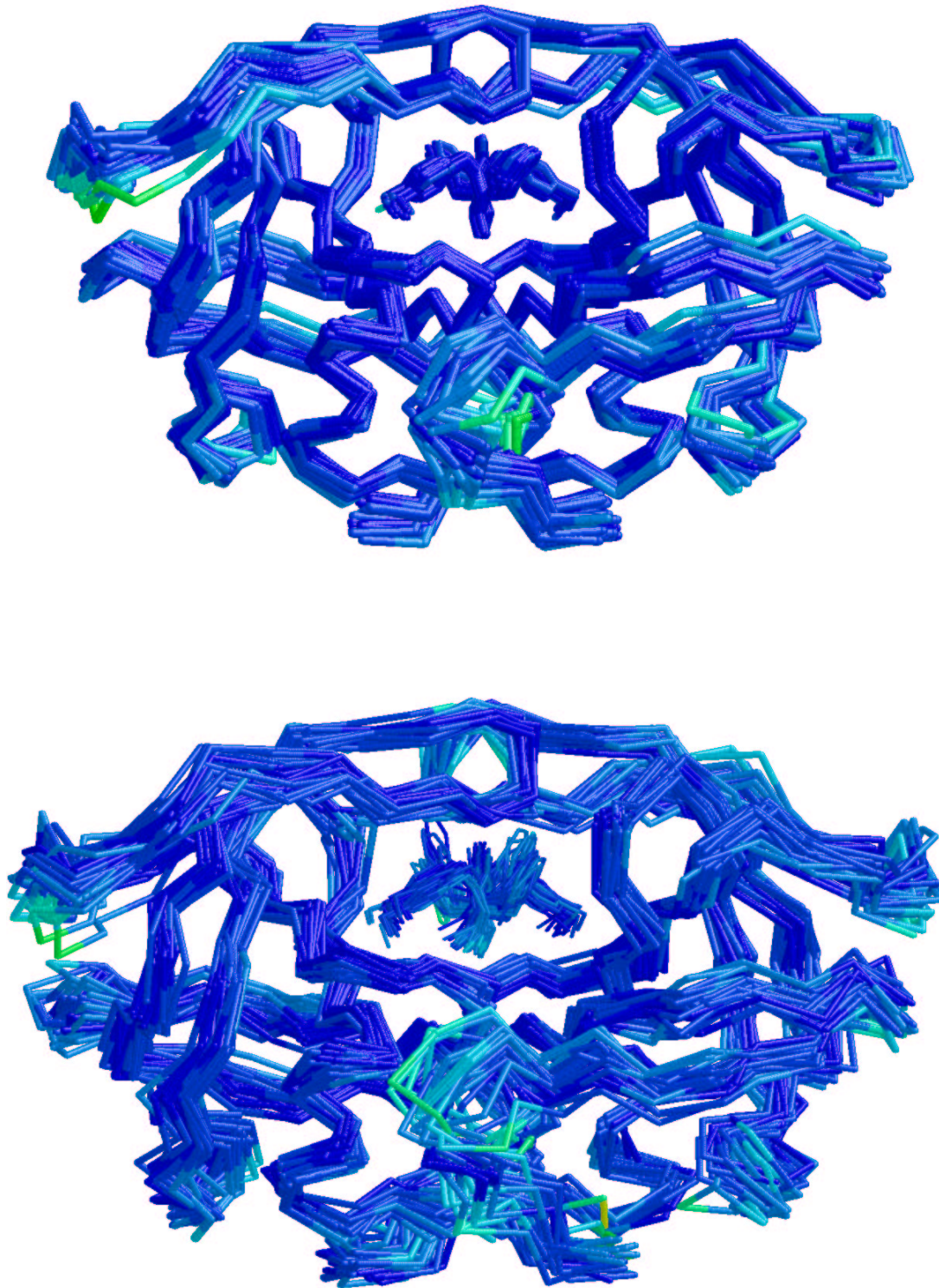


Abbildung 3.5.: Überlagerte Hauptketten ( $C_{\alpha}$ -Atome) vor (oben) und nach (unten) Minimierung 2. Die Atome sind entsprechend ihrer *rmsd*-Werte bzgl. der Referenz-Struktur gefärbt. Dunkle Bereiche stimmen gut, hellere schlechter überein. Deutlich ist zu erkennen, dass die Ligand-Konformationen sich während der Minimierung stark ändern.



Table 3.1.: Struktur- und Energie-Vergleiche der Ausgangsdaten *vor* der Minimierung. Minimal- und Maximal-Werte sind hervorgehoben. *rmsd*-Werte werden bezogen auf die Referenz-Struktur.

	$rmsd^{ref}$	$rmsd_{C\alpha}^{ref}$	$E_{P.L}^{Kraftfeld}$	$E_{P.L}^{Solv.}$	$E_{P.L}^{Gesamt}$	$E_{L+WW}^{Kraftfeld}$	$E_{L+WW}^{Solv.}$	$E_{L+WW}^{Gesamt}$
	[Å]	[Å]	Energiebeiträge [kcal/mol]					
1	1.51	0.90	793.5	-666.4	<b>127.1</b>	94.2	-292.5	-198.3
2	1.65	1.10	473.8	-658.1	-184.3	59.7	-270.2	-210.5
3	1.71	1.14	478.5	-697.8	-219.3	66.8	-294.6	-227.8
4	1.38	0.79	449.8	-697.5	-247.6	74.8	-286.8	-212.0
5	1.49	0.89	616.1	-642.6	-26.4	<b>119.4</b>	-269.2	-149.8
6	1.64	0.81	702.9	-643.2	59.7	90.3	-267.5	-177.2
7	<b>1.15</b>	<b>0.62</b>	540.6	-663.8	-123.2	40.2	-266.3	-226.0
8	1.38	0.80	574.6	-663.2	-88.6	56.3	-265.9	-209.6
9	1.57	0.87	491.3	-615.4	-124.1	<b>35.7</b>	-251.7	-216.0
10	1.67	0.84	596.7	-649.3	-52.5	66.0	-262.7	-196.6
11	1.80	1.23	576.2	-635.7	-59.4	60.9	<b>-232.3</b>	-171.4
12	1.38	0.87	539.2	-656.9	-117.6	83.4	-272.4	-188.9
13	1.71	1.32	516.2	<b>-613.0</b>	-96.7	64.7	-256.8	-192.0
14	1.26	0.69	536.3	-647.1	-110.7	78.6	-279.3	-200.7
15	1.50	0.84	599.1	-682.7	-83.6	63.8	-285.9	-222.0
16	1.56	0.97	580.8	-621.9	-41.0	85.8	-256.7	<b>-170.8</b>
17	<b>2.33</b>	<b>1.86</b>	596.6	-684.2	-87.5	54.9	-300.1	-245.1
18	1.53	0.89	<b>971.3</b>	-657.7	313.6	93.4	-268.8	-175.3
19	1.32	0.64	620.8	-621.5	-0.7	58.4	-241.2	-182.7
20	1.66	0.92	455.1	-663.5	-208.3	43.4	-301.3	<b>-257.8</b>
21	1.63	0.89	609.6	-637.0	-27.3	50.8	-235.0	-184.2
22	1.36	0.80	467.6	-710.9	-243.2	67.4	-291.0	-223.5
23	1.47	0.84	693.2	-643.3	49.9	88.5	-245.7	-157.2
24	1.31	0.69	553.5	-665.3	-111.8	85.9	-298.7	-212.7
25	1.46	0.77	455.9	-652.7	-196.8	44.3	-275.4	-231.1
26	1.39	0.71	726.7	-699.9	26.8	72.1	<b>-315.5</b>	-243.4
27	1.69	1.07	701.8	-666.3	35.5	65.0	-292.1	-227.1
28	1.71	1.23	724.5	<b>-716.3</b>	8.1	67.5	-302.8	-235.2
Ref.	0	0	<b>403.7</b>	-700.0	<b>-296.2</b>	88.4	-291.2	-202.7
Ø	1.49	0.93	581.39	-661.2	-73.3	69.7	-274.9	-205.1
±σ	0.36	0.25	122.2	28.3	126.7	19.0	21.7	27.1

Tabelle 3.2.: Struktur- und Energie-Vergleich nach Minimierung 1. *rmsd*-Werte werden sowohl in Bezug auf die Referenz-Struktur, als auch in Bezug auf die Start-Konformation berechnet.

	$rmsd^{ref}$	$rmsd_{Min}^{Start}$	$E_{P.L}^{Kraftfeld}$	$E_{P.L}^{Solv.}$	$E_{P.L}^{Gesamt}$	$E_{L+WW}^{Kraftfeld}$	$E_{L+WW}^{Solv.}$	$E_{L+WW}^{Gesamt}$
	[Å]	[Å]	Energiebeiträge [kcal/mol]					
1	1.52	0.35	-949.0	-361.4	-1310.4	<b>13.5</b>	7.9	<b>21.4</b>
2	1.66	0.36	-1115.9	-376.7	-1492.7	-7.4	7.5	0.0
3	1.72	0.29	-1081.8	-388.0	-1469.9	0.1	11.3	11.5
4	1.38	0.39	-1129.2	-400.9	-1530.1	4.0	<b>6.1</b>	10.1
5	1.50	0.32	-1075.1	-360.8	-1436.0	4.1	7.7	11.9
6	1.65	0.39	-1053.7	-364.7	-1418.4	-9.2	8.8	-0.3
7	<b>1.16</b>	0.29	-1116.9	-381.1	-1498.1	-16.8	12.5	-4.2
8	1.38	0.38	-1084.1	-383.1	-1467.3	-13.5	9.6	-3.9
9	1.58	0.39	-1129.6	-349.4	-1479.0	-18.3	8.7	-9.6
10	1.67	0.32	-1064.9	-373.8	-1438.8	-12.5	6.8	-5.7
11	1.81	0.34	-1009.1	-385.6	-1394.7	-5.4	<b>12.8</b>	7.4
12	1.39	0.30	-1038.0	-371.4	-1409.4	7.1	8.1	15.3
13	1.72	0.34	-1081.7	<b>-342.8</b>	-1424.6	-3.9	6.0	2.0
14	1.27	0.40	-1066.5	-347.3	-1413.9	-2.8	12.7	9.8
15	1.51	0.36	-1113.8	-380.2	-1494.1	-1.2	8.6	7.4
16	1.57	0.36	-1072.3	-350.5	-1422.8	1.9	8.7	10.7
17	<b>2.34</b>	0.43	-1090.0	-371.8	-1461.8	<b>-21.8</b>	6.0	<b>-15.8</b>
18	1.54	0.30	<b>-831.7</b>	-376.4	<b>-1208.1</b>	12.4	8.6	21.1
19	1.32	0.39	-1021.0	-366.5	-1387.6	-7.0	9.6	2.5
20	1.67	<b>0.24</b>	-1141.7	-347.5	-1489.2	-13.0	8.4	-4.5
21	1.63	0.38	-1073.2	-385.0	-1458.3	-16.9	10.8	-6.1
22	1.37	0.35	<b>-1158.2</b>	<b>-403.5</b>	<b>-1561.7</b>	-16.2	11.3	-4.9
23	1.48	0.42	-1039.7	-384.3	-1424.0	-5.0	9.8	4.8
24	1.32	0.30	-1092.2	-352.3	-1444.6	-11.4	10.2	-1.2
25	1.46	0.32	-1114.0	-365.0	-1479.0	-15.0	9.1	-5.9
26	1.48	<b>1.67</b>	-1017.6	-368.0	-1385.7	-7.1	10.6	3.5
27	1.69	0.33	-1092.2	-352.3	-1444.6	-11.4	10.2	-1.2
28	1.71	0.35	-969.5	-397.7	-1367.3	-4.0	11.7	7.6
ref	0	0.44	-1119.2	-392.7	-1511.9	-4.3	12.2	7.9
Ø	1.50	0.40	-1067.0	-371.8	-1438.8	-6.3	9.4	3.2
±σ	0.36	0.25	66.8	17.1	68.6	9.1	2.0	9.0

## Resultate nach Minimierung 2

Tabelle 3.3 zeigt die *rmsd*- und Energie-Daten für das Minimierungsverfahren 2.

*Tabelle 3.3.:* Struktur- und Energie-Vergleich nach Minimierung 2. *rmsd*-Werte werden sowohl in Bezug auf die Referenz-Struktur, als auch in Bezug auf die Start-Konformation berechnet.

	$rmsd^{ref}$	$rmsd_{Min}^{Start}$	$E_{P.L}^{Kraftfeld}$	$E_{P.L}^{Solv.}$	$E_{P.L}^{Gesamt}$	$E_{L+WW}^{Kraftfeld}$	$E_{L+WW}^{Solv.}$	$E_{L+WW}^{Gesamt}$
	[Å]	[Å]	Energiebeiträge [kcal/mol]					
1	2.16	1.86	-3046.0	-191.3	-3237.3	-42.4	8.3	-34.0
2	2.25	<b>1.71</b>	-3054.8	-176.8	-3231.7	-43.7	9.1	-34.5
3	2.49	1.79	-2996.3	<b>-242.1</b>	-3238.5	-45.9	6.9	-38.9
4	2.10	1.86	-3078.3	-179.1	-3257.4	-61.3	10.7	<b>-50.6</b>
5	2.25	2.01	-3023.4	-206.1	-3229.6	-51.1	8.4	-42.6
6	2.12	1.82	-2996.3	<b>-242.1</b>	-3238.5	-45.9	6.9	-38.9
7	<b>1.97</b>	1.78	-3060.7	-194.7	-3255.4	-49.5	7.8	-41.7
8	2.12	1.89	-3031.9	-180.8	-3212.8	-43.6	8.2	-35.4
9	2.29	1.95	-3063.6	-177.0	-3240.6	-55.9	11.8	-44.0
10	2.35	1.84	-3020.2	-186.7	-3207.0	<b>-35.1</b>	7.9	<b>-27.1</b>
11	2.18	2.04	-3044.7	-209.5	-3254.2	-41.2	8.6	-32.6
12	2.04	2.13	-3045.3	-177.4	-3222.7	-42.3	8.2	-34.0
13	2.39	1.98	-3037.8	-177.0	-3214.9	-44.7	7.7	-37.0
14	2.25	2.09	-3071.3	<b>-151.0</b>	-3222.4	-50.1	12.4	-37.6
15	2.32	1.96	-3052.7	-191.7	-3244.4	-49.4	7.3	-42.0
16	2.17	1.81	-3044.7	-162.0	-3206.7	-45.1	9.6	-35.5
17	2.52	<b>2.35</b>	-3032.0	-182.4	-3214.4	-43.2	9.9	-33.3
18	2.26	1.73	-3000.8	-192.6	<b>-3193.4</b>	-51.8	<b>5.9</b>	-45.8
19	2.07	1.75	<b>-2989.9</b>	-213.8	-3203.8	-44.0	8.3	-35.6
20	2.30	1.97	-3015.5	-186.4	-3201.9	-45.5	8.5	-37.0
21	2.30	2.26	-3030.3	-180.5	-3210.9	-55.1	8.0	-47.1
22	2.21	2.16	-3030.3	-180.5	-3210.9	-55.1	8.0	-47.1
23	2.44	2.01	<b>-3098.1</b>	-158.4	-3256.5	-48.1	9.1	-39.0
24	<b>1.96</b>	1.99	-3094.9	-178.6	<b>-3273.6</b>	<b>-63.4</b>	<b>17.0</b>	-46.4
25	2.23	1.93	-3043.4	-187.2	-3230.7	-42.1	10.4	-31.6
26	1.96	1.85	-3088.9	-174.7	-3263.6	-52.2	7.3	-44.9
27	2.14	1.96	-3043.4	-187.2	-3230.7	-42.1	10.4	-31.6
28	2.35	2.23	-3062.3	-180.9	-3243.2	-48.1	7.1	-40.9
ref		1.92	-3044.9	-209.4	-3254.3	-41.3	8.5	-32.7
∅	2.23	1.96	-3042.9	-188.2	-3231.1	-47.6	9.0	-38.7
±σ	0.15	0.16	28.3	20.6	21.1	6.3	2.1	5.8

**Struktur-Vergleich:** Ohne Randbedingungen verschieben sich die Atome um durchschnittlich fast  $2.0 \pm 0.16 \text{ \AA}$  gegenüber der Start-Konformation und weichen damit um durchschnittlich  $2.23 \pm 0.15 \text{ \AA}$  gegenüber der ebenfalls Energie-minimierten Referenz-Struktur ab. Das radiale *rmsd*-Profil zeigt deutlich, dass sich besonders die Konformation des Liganden und die der Bindetasche stärker verändert hat als der periphere Rest des Proteins. Ab einem Abstand von etwa  $10\text{--}12 \text{ \AA}$  zum Liganden folgt das radiale *rmsd*-Profil wieder demjenigen vor der Minimierung im Abstand von etwa  $0.4 \text{ \AA}$  (s. Tab. 3.5 bzw. Abb. 3.6). Abbildung 3.5 zeigt die Verteilung der Positionsverschiebungen innerhalb des Komplexes für die  $C_\alpha$ -Atome. Die mittlere *rmsd* für  $C_\alpha$ -Atome gegenüber der jeweiligen Startkonformation ist mit  $1.40 \pm 0.29 \text{ \AA}$  kleiner als die *rmsd* für alle Atome mit  $2.23 \pm 0.15 \text{ \AA}$ . Das zeigt, dass sich die Seitenketten deutlicher verändern als der globale Verlauf der Hauptkette.

**Energie-Vergleich:** Trotz der offensichtlichen Divergenz der Konformationen reduzieren sich die Schwankungen der Kraftfeld-Energie. Der durchschnittliche Beitrag der Kraftfeld-Energie ist um etwa  $3600 \text{ kcal/mol}$  gegenüber den Start-Konformationen abgesenkt worden auf  $-3043 \pm 28 \text{ kcal/mol}$ . Damit ist die Standard-Abweichung gegenüber dem Zustand ohne Minimierung etwa um den Faktor 4 verbessert worden. Die Beiträge der Solvatationsenergie haben sich, wie auch bei der Energie-Minimierung nach dem ersten Verfahren, erhöht; hier um etwa  $450 \text{ kcal/mol}$  auf  $-188 \pm 21 \text{ kcal/mol}$ . Die Zuverlässigkeit der Solvatationsenergien ist demnach bei beiden Minimierungs-Verfahren vergleichbar.

**Protein-Ligand-Wechselwirkungen:** Obwohl sich gerade die Kontakt-Zone zwischen Ligand und Protein während dieser Energie-Minimierung am deutlichsten verändert hat, haben sich die Schwankungen des Ligand- und Wechselwirkungsbeitrags reduziert auf nunmehr fast  $6 \text{ kcal/mol}$  in der Gesamtenergie ( $-39 \pm 5.8 \text{ kcal/mol}$ ).

### Vergleich der Ergebnisse

Tabelle 3.4 zeigt, wie sich die Energie-Minimierungen auf die einzelnen Beiträge zur Kraftfeld-Energie auswirken. Zu den großen Schwankungen der Kraftfeld-Energie vor einer Minimierung tragen mit einer Standard-Abweichung von fast  $111 \text{ kcal/mol}$  hauptsächlich die LJ-Wechselwirkungen bei, gefolgt von den elektrostatischen Wechselwirkungen und den Beiträgen der Torsions-Winkel mit Standard-Abweichungen von jeweils ca.  $20 \text{ kcal/mol}$ .

**Minimierung 1** – nur durchgeführt zur Korrektur ungünstiger Bindungslängen und LJ-Kontakte – liefert das erwartete Ergebnis, nämlich um ca.  $390 \text{ kcal/mol}$  verbesserte Beiträge der Bindungslängen und um ca.  $1110 \text{ kcal/mol}$  verbesserte Beiträge der LJ-Wechselwirkungen, sogar mit auf  $\pm 50 \text{ kcal/mol}$  halbiertes Standard-Abweichung. Alle anderen Beiträge behalten in etwa ihre Größenordnung und auch ihre Genauigkeit.

**Minimierung 2** lässt größere Konformationsänderungen zu. Wesentliche Änderungen der Energiebeiträge liefern immer noch die Bindungslängen ( $-430 \text{ kcal/mol}$ ) und van-

Tabelle 3.4.: Mittlere Teilbeiträge zur Gesamtenergie in Abhängigkeit vom Minimierungsverfahren

	Bindungen	Winkel	Dieder- winkel	uneigentl. Diederw.	LJ- Energie	elektrostat. Energie	Solv. Energie
Energiebeiträge [kcal/mol]							
vor	459.8	425.5	997.7	9.5	303.5	-1608.3	-661.2
Min.	$\pm 2.3$	$\pm 2.3$	$\pm 19.7$	$\pm 0.4$	$\pm 110.8$	$\pm 21.0$	$\pm 28.3$
nach	69.5	370.6	922.5	4.4	-815.0	-1618.9	-371.8
Min. 1	$\pm 3.0$	$\pm 6.6$	$\pm 18.8$	$\pm 1.2$	$\pm 50.3$	$\pm 20.9$	$\pm 17.1$
nach	31.7	359.6	502.9	3.0	-2014.2	-1925.9	-188.2
Min. 2	$\pm 0.9$	$\pm 7.5$	$\pm 12.1$	$\pm 0.8$	$\pm 15.4$	$\pm 22.0$	$\pm 20.6$

der-Waals-Beiträge (-2300 kcal/mol), nun aber auch Torsions- und Coulomb-Wechselwirkungen (-500 kcal/mol und -300 kcal/mol). Gleichzeitig wird eine erhebliche Reduktion der Energie-Schwankungen und damit auch eine gewisse Vereinheitlichung des Energie-Beitrages erreicht.

Zur Berechnung der radialen *rmsd*-Verteilungen ist auf die H-Atome verzichtet worden. Der Vergleich der resultierenden durchschnittlichen *rmsd*-Werte mit den zuvor für alle Atome berechneten durchschnittlichen *rmsd*-Werte zeigt, dass besonders die H-Atome zur allgemeinen Abweichung der mittleren Atom-Verschiebung in Bezug auf die Referenz-Struktur beitragen. Dies ist teilweise durch ihre schon von Beginn an abweichenden Positionen, teilweise aber auch durch ihre höhere Beweglichkeit während der Energie-Minimierung bedingt (Massenwichtung).

Abbildung 3.6: Radiale *rmsd*-Verteilungen vor bzw. nach der Energie-Minimierung. Sowohl vor Minimierung, als auch nach Minimierung (1) ist deutlich zu erkennen, dass der dem Liganden benachbarte Bereich zwischen allen Komplexen besser übereinstimmt als die Peripherie des Proteins. Nach freier Minimierung (2) finden strukturelle Veränderungen gerade in dem zuvor so gut definierten Bereich um den Liganden statt.

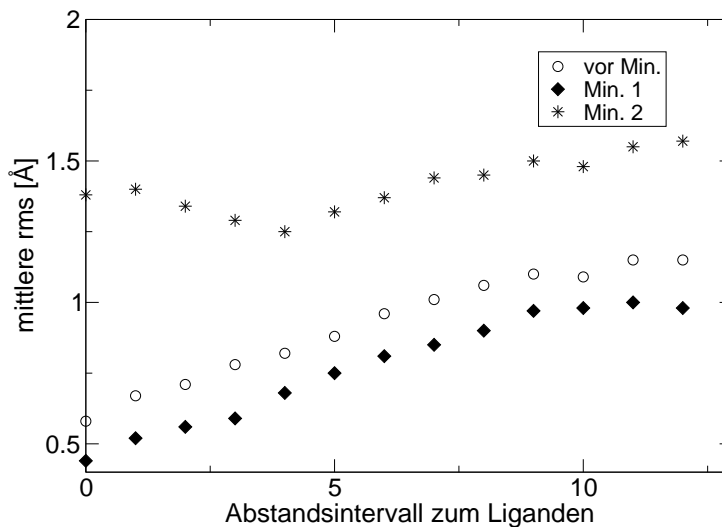


Tabelle 3.5.: Radiale *rmsd*-Verteilungen

Intervall	Abstand zum Liganden [ $\text{\AA}$ ]	radiale <i>rmsd</i> [ $\text{\AA}$ ]		
		vor Min.	Min.1	Min.2
0	0-2	$0.58 \pm 0.21$	$0.44 \pm 0.20$	$1.38 \pm 0.41$
1	2-4	$0.67 \pm 0.15$	$0.52 \pm 0.15$	$1.40 \pm 0.17$
2	4-6	$0.71 \pm 0.14$	$0.56 \pm 0.14$	$1.34 \pm 0.13$
3	6-8	$0.78 \pm 0.14$	$0.59 \pm 0.13$	$1.29 \pm 0.11$
4	8-10	$0.82 \pm 0.15$	$0.68 \pm 0.15$	$1.25 \pm 0.26$
5	10-12	$0.88 \pm 0.15$	$0.75 \pm 0.16$	$1.32 \pm 0.11$
6	12-14	$0.96 \pm 0.17$	$0.81 \pm 0.18$	$1.37 \pm 0.11$
7	14-16	$1.01 \pm 0.19$	$0.85 \pm 0.25$	$1.44 \pm 0.12$
8	16-18	$1.06 \pm 0.19$	$0.90 \pm 0.27$	$1.45 \pm 0.12$
9	18-20	$1.10 \pm 0.20$	$0.97 \pm 0.22$	$1.50 \pm 0.12$
10	20-22	$1.09 \pm 0.29$	$0.98 \pm 0.22$	$1.48 \pm 0.31$
11	22-24	$1.15 \pm 0.21$	$1.00 \pm 0.22$	$1.55 \pm 0.13$
12	24-26	$1.15 \pm 0.21$	$0.98 \pm 0.29$	$1.57 \pm 0.13$
13	26-28	$1.16 \pm 0.21$	$1.10 \pm 0.16$	$1.51 \pm 0.32$
1-13	$\emptyset$	$0.94 \pm 0.20$	$0.80 \pm 0.21$	$1.42 \pm 0.10$

### 3.4.3. Diskussion

Die NMR-Strukturen des Komplexes einer HIV-PR-Mutante (I3V, C95A) mit dem Liganden DMP-323 sind untersucht worden. 28 unterscheidbare Konformationen liegen in Lösung nebeneinander vor. Die einzelnen Konformationen dieses Ensembles (1bve) weichen um durchschnittlich  $1.49 \text{ \AA}$  von ihrer gemittelten Struktur (1bvg) ab. Diese Abweichungen sind im Bereich des Liganden mit  $0.58 \text{ \AA}$  am geringsten und nehmen zur Peripherie hin auf  $1.1 \text{ \AA}$  zu (s. Tab. 3.5).

Obwohl die Koordinaten der verschiedenen Konformationen sehr ähnlich sind, schwanken die berechneten Gesamtenergien unerwartet stark, nämlich um über  $\pm 100 \text{ kcal/mol}$ . Dies beruht vermutlich auf der Qualität der veröffentlichten Atom-Koordinaten, die (technisch bedingt) begrenzt ist.

Eine Energie- bzw. Struktur-Optimierung ist also auf jeden Fall notwendig. Daher werden Energie-Minimierungen unter verschiedenen Randbedingungen durchgeführt, um zu prüfen, ob man eine Vereinheitlichung der Energie-Werte aller Konformationen erreichen kann.

Eine Energie-Minimierung, die nur geringfügige Atom-Verschiebungen zulässt, erlaubt, die Konformationen insgesamt beizubehalten und gleichzeitig Unstimmigkeiten durch ungenaue Koordinaten-Angaben zu verringern. Die Molekül-Gerüste gleichen sich einander ein wenig an: Die mittlere *rmsd* beträgt nach der Minimierung nur  $0.80 \text{ \AA}$  im Vergleich zu

0.94 Å davor, während sich die Positionen der H-Atome deutlicher gegeneinander verschieben. Die berechneten Energie-Werte streuen trotzdem mit einer Standard-Abweichung von rund  $\pm 70$  kcal/mol noch sehr stark.

Eine alternative Energie-Minimierung, die allen Atomen die maximale Bewegungsfreiheit einräumt, baut weitere Spannungen innerhalb des Komplexes ab und damit auch die Bandbreite der berechneten Energie-Werte. Doch selbst nach diesem Verfahren lässt sich die Standard-Abweichung nur auf  $\pm 20$  kcal/mol begrenzen. Während der Minimierung verändern besonders die Atome des Liganden und seiner nächsten Umgebung ihre Positionen, sodass der zuvor sehr eindeutige Bindemodus des Liganden nun weniger festgelegt erscheint. In diesem Zusammenhang erscheint die Verringerung der Standard-Abweichung der Energie-Beiträge überraschend, denn man erwartet, dass größere strukturelle Veränderungen die systematischen Fehler beeinflussen und mit zusätzlichen Energie-Schwankungen einhergehen. Ligand und Bindetasche bestehen jedoch überwiegend aus unpolaren Gruppen, die keine gerichteten Wechselwirkungen eingehen, was die geringe Empfindlichkeit gegenüber unterschiedlichen Konformationen in diesem Bereich erklären könnte.

Die Gegenüberstellung der einzelnen Energie-Beiträge (s. Tab. 3.4) verdeutlicht, welche Terme die großen Schwankungen verursachen. Bindungslängen und uneigentliche Torsionswinkel sind mit durchschnittlichen Schwankungen von unter 1 kcal/mol fast konstant. Dies sind die Beiträge, die mit den höchsten Kraftkonstanten versehen sind und gleichzeitig für ein gegebenes Atom mit höchstens vier Termen in die Energie-Funktion eingehen. Im Gegensatz dazu kann ein einzelnes Atom (je nach Konfiguration) an bis zu 18 einfachen Winkeln und 18 Torsions-Winkeln beteiligt sein. Diese Terme werden mit kleineren Kraftkonstanten gewichtet, sodass Abweichungen von der natürlichen Größe häufiger vorkommen. Elektrostatische und LJ-Wechselwirkungen eines Atoms werden jeweils zu *allen* anderen Atomen berechnet. Diese „weichen“ Terme dominieren mengenmäßig in der individuellen Energie-Funktion, sodass sich die durchweg kleinen Abweichungen vom Idealwert zu großen Werten summieren können. Eine Energie-Minimierung korrigiert zuerst die „harten“ Energie-Terme. Daher kann eine Konformation, beispielsweise trotz Spannung eines Torsionswinkels, in einem lokalen Minimum „gefangen“ bleiben, weil die energetische Barriere, über die eine Entspannung erreicht werden könnte, beim Verfahren der *Energie-Minimierung* unüberwindbar ist.

Tabelle 3.4 zeigt deutlich, dass beide Minimierungsverfahren vor allem ungünstige LJ-Wechselwirkungen und Bindungslängen korrigieren. Energie-Minimierung 2 führt darüber hinaus noch zu einer wesentlichen Absenkung ungünstiger Torsions- und elektrostatischer Beiträge. Mit dieser Verringerung der Teilenergien aus elektrostatischen und LJ-Wechselwirkungen nimmt der Beitrag der Solvatationsenergie zu. Da sich die Konformationen mit durchschnittlich 1.96 Å *rmsd*-Verschiebung gegenüber der Start-Konformation nicht wesentlich verändert haben, scheidet eine drastische Zunahme der benetzbaren Oberfläche als Ursache für diese Beobachtung aus. Im Gegenteil: Eine Energie-Minimierung ohne Wassermoleküle fördert die Annäherung von Atomen (und damit eine Verringerung der benetzbaren Oberfläche), getrieben durch die schwachen LJ-Anziehungen untereinander und durch die unverminderte Anziehung zwischen gegensinnigen Ladungen. Diese Paarung von gegensinnigen Ladungen während der Energie-Minimierung verringert allerdings das effektive

elektrische Feld um den nun entstandenen Dipol und damit den Gewinn an elektrostatischer Solvatationsenergie des Wassers. Der Gewinn an elektrostatischer Kraftfeld-Energie des Proteins kompensiert nur teilweise den Verlust der Solvatationsenergie. An dieser Stelle wird deutlich, wie sehr sich die Energie-Funktionen während der Energie-Minimierung (nur Kraftfeld-Energie mit  $\varepsilon = 2r$ ) und bei der Energie-Rechnung (Kraftfeld-Energie + Solvatationsenergie mit  $\varepsilon = 4$ ) unterscheiden.

Von den 3200 Atomen des gesamten Komplexes sind möglicherweise nur die 80 Atome des Liganden und die ihm direkt benachbarten Atome des Proteins relevant, da nur diese miteinander im entscheidenden van-der-Waals- bzw. Coulomb-Kontakt stehen. Subtrahiert man von der Gesamtenergie eines Komplexes den Beitrag des Proteins, reduziert sich die Standard-Abweichung entsprechend um den Faktor 3–7 auf 9.0 kcal/mol nach Minimierung 1 bzw. 5.8 kcal/mol nach Minimierung 2. Dieser Befund verdeutlicht, wie sehr der über die Bindetasche hinausgehende Protein-Anteil die energetische Schwankungsbreite beeinflusst.

#### 3.4.4. Schlussfolgerungen

Die Analyse des NMR-Ensembles hat gezeigt, dass der Bindemodus des Liganden an das Protein gut definiert ist und dass vor allem die eigentlich uninteressante Peripherie des Proteins energetische Schwankungen verursacht. Gleichzeitig tolerieren die überwiegenden, ungerichteten LJ-Kontakte zwischen Ligand und Protein einen gewissen Spielraum ohne die Wechselwirkungsenergien wesentlich zu verändern. Diese Befunde lassen den statischen Ansatz zur Bestimmung relativer Bindestärken tatsächlich praktikabel erscheinen.

So feine Unterschiede zwischen Konformationen wie bei der NMR-Analyse sind mit RKSA nicht differenzierbar, eben weil innerhalb eines Protein-Kristalls eine ähnliche Vielfalt an Konformationen vorliegt (Fehlordnung im Kristall). Die Auflösung in der RKSA beträgt bei Proteinen etwa 2.0 Å. Man erhält aus der Kristallographie also Atom-Koordinaten, deren Qualität etwa der einer gemittelten NMR-Konformation entspricht, allerdings ohne H-Atome. Diese müssen dem Molekül-Gerüst entsprechend ergänzt werden. Die örtliche Ungenauigkeit der automatisch platzierten H-Atome ist unbedeutend, da die H-Atome im Zuge einer Energie-Minimierung ohnehin ihre Positionen stark ändern.

Die Gesamtenergien der gemittelten NMR-Struktur, die hier als Referenz-Konformation dient, sind nach der Energie-Minimierung jeweils günstiger als die mittlere Gesamtenergie aller Konformationen. Die Differenz zwischen der Energie der Referenz-Struktur und der mittleren Energie aller übrigen Konformationen liegt jeweils in der Größenordnung der Standard-Abweichung. Es ist also eher zulässig, Koordinaten zu mitteln als Energie-Werte. Demnach ist eine gemittelte Konformation, wie z. B. eine KS eine plausible Repräsentante für ein Ensemble.

Die Schwankungen in der Gesamtenergie innerhalb des Ensembles sind jedoch selbst nach Energie-Minimierung im besten Fall mit über 20 kcal/mol um den Faktor 100 größer als erwartet. Die Annahme, dass die nach diesem Verfahren berechenbare minimale Gesamtenergie innerhalb eines Ensembles mit der mittleren näherungsweise gleichgesetzt werden kann (s. Kap. 2.4), ist damit also widerlegt.



Die großen energetischen Schwankungen werden wahrscheinlich verursacht durch ungenaue Start-Koordinaten einerseits und die Diskrepanz zwischen Minimierungs- und Bewertungsfunktion andererseits, denn die lokalen Minima der Kraftfeld-Funktion müssen nicht notwendigerweise lokalen Minima der Gesamtenergie-Funktion entsprechen. Demnach ergibt sich durch die Energie-Minimierung ohne wässrige Umgebung ein größerer systematischer Fehler.

Im Hinblick auf die Berechnung von Bindestärken der HIV-PR zu verschiedenen Liganden ist die Streuung der Energie-Werte innerhalb des Ensembles belanglos, solange dieser Fehler bei allen Komplex-Repräsentanten systematisch gleich ist. Für die Modellierung einer Serie von Komplexen ist daher sicherzustellen, dass die Protein-Konformationen während einer Energie-Minimierung nicht auseinander driften, eben um einheitliche systematische Fehler zu gewährleisten. Die Struktur-Optimierung sollte daher auf die Region um den Liganden beschränkt werden.

### 3.5. Komplex-Konstruktion

Zur relativen Bindestärke  $\Delta\Delta G$  zweier Komplexierungsreaktionen mit verschiedenen Liganden tragen die Energiedifferenzen zwischen den Liganden  $\Delta E_L$  und die Energiedifferenzen  $\Delta E_{P.L}$  zwischen den Komplexen bei (s. Reaktionen-Schema Abschnitt 2.4.1 und Gl. 2.24).

Um  $\Delta E_{P.L}$  zu bestimmen, wird aus der KS des Referenz-Komplexes der Vergleichskomplex konstruiert; beide werden durch Energie-Minimierung Struktur-optimiert. Der Beitrag  $\Delta E_{P.L}$  muss unabhängig von der jeweiligen Ausgangskonformation des Referenz-Komplexes sein. Um dies zu prüfen, wird  $\Delta E_{P.L}$  für zwei äquivalente Paare von Komplexen berechnet. Ausgehend von zwei verschiedenen KS wird aus jeder einzelnen der jeweils andere Komplex abgeleitet und einer Energie-Minimierung unterzogen (s. Abb. 3.7). Folgende Punkte sollen geklärt werden:

- Wie verändern sich die Konformationen während der Energie-Minimierung?
- Ist der Energie-Beitrag  $\Delta E_{P.L}$  unabhängig von der jeweiligen Stamm-Konformation?
- Mit welchem Verfahren zur Energie-Minimierung lassen sich übereinstimmende  $\Delta E_{P.L}$ -Beiträge erzielen?

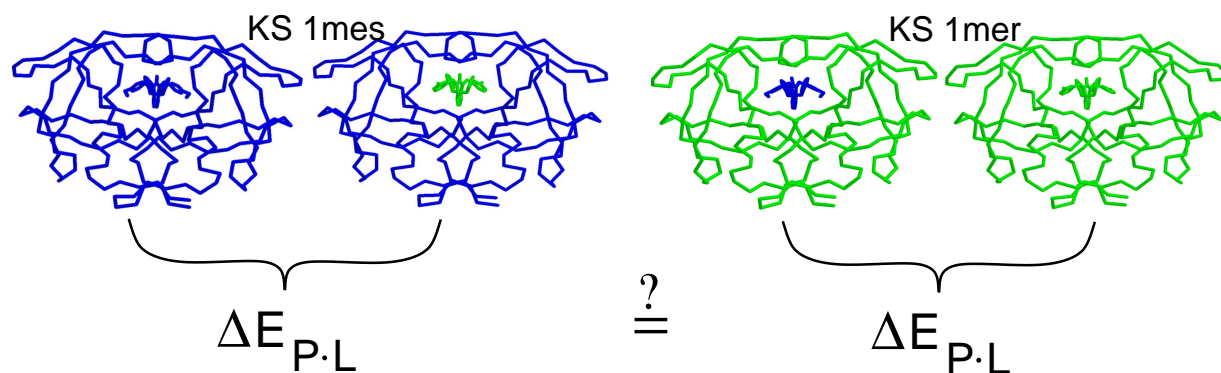


Abbildung 3.7.: Ausgehend von jeder der beiden KS (dunkel: 1mes und mittelgrau: 1mer) wird der jeweils andere Komplex erzeugt. Die Stamm-Komplexe sind also ein- die Konstrukte zweifarbig. Stamm-Komplexe und Konstrukte werden einer Energie-Minimierung unterzogen, für die resultierenden Strukturen werden die Gesamtenergien bestimmt. Die Differenzen  $\Delta E_{P.L}$  für die äquivalenten Paare sollen unabhängig von der Stamm-Struktur sein.

#### 3.5.1. Durchführung

Von einer HIV-PR-Mutante (I3V, I84V) existieren zwei KS mit sehr ähnlichen Liganden, den beiden zyklischen Harnstoff-Derivaten DMP-450 und DMP-323 (PDB-Codes 1mer und 1mes [1]). Die KS dienen als Stamm-Konformationen zur Konstruktion des jeweils anderen Komplexes. Die Proteine haben identische Sequenzen, die Liganden sind wie das dimer Enzym  $C_2$ -symmetrisch und besitzen vier aromatische Ringe, von denen jeweils zwei unterschiedlich substituiert sind (s. Abb. 3.8).

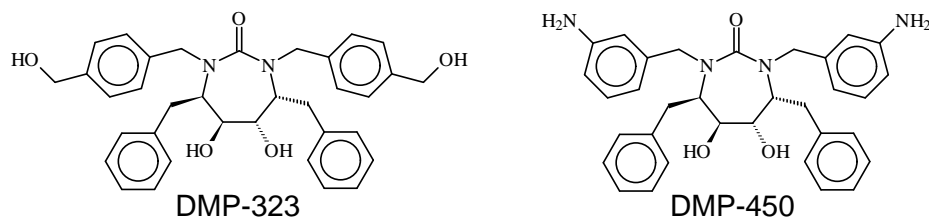


Abbildung 3.8.: Die beiden zyklischen Harnstoff-Derivate DMP-323 aus KS 1mes und DMP-450 aus KS 1mer.

**Komplex-Konstruktion:** Ausgehend von KS 1mer, werden von DMP-450 die N-Atome in *meta*-Position entfernt und HO-CH<sub>2</sub>-Gruppen in *para*-Position eingefügt, sodass der neue Ligand die Struktur von DMP-323 besitzt. Bei 1mes wird entsprechend umgekehrt verfahren. Die in der KS fehlenden H-Atome werden automatisch von CHARMM ergänzt. Die Konstrukte besitzen die chemische Struktur des jeweils anderen Stamm-Komplexes, aber nicht dessen Konformation. Sie dienen ebenso wie die originalen KS als Start-Konformationen für die sich anschließenden Energie-Minimierungen (s. Abb. 3.7).

**Struktur-Optimierung:** Die Wahl der Randbedingungen ist entscheidend für das Ergebnis einer Energie-Minimierung, denn die Randbedingungen legen fest, welche Atome in welchem Ausmaß beweglich sein sollen. Nach einer ersten, vorbereitenden Struktur-Optimierung werden verschiedene alternative Randbedingungen während der Energie-Minimierungen erprobt:

**Vorstufe, vorbereitende Struktur-Optimierung:** Zunächst sind nur die neu hinzugefügten Atome, also die H-Atome und die der Liganden, beweglich, um grobe Unstimmigkeiten durch die Manipulation der Strukturen zu beseitigen. Die hieraus resultierenden Konformationen werden als Start-Strukturen für verschiedene weitere Energie-Minimierungen verwendet.

**Alternative 1:** Alle Atome werden elastisch an ihre gegebenen Start-Positionen gebunden mit einer Kraftkonstanten von 10 kcal/mol·Å<sup>2</sup> (Diese Anbindung ist massengewichtet, d. h. H-Atome sind beweglicher als C-, N- und O-Atome).

**Alternative 2:** Nur die C<sub>α</sub>-Atome werden mit einer Kraftkonstanten von 10 kcal/mol·Å<sup>2</sup> an ihre Start-Positionen gebunden, die Seitenketten und Liganden sind frei beweglich.

**Alternative 3:** Alle H-Atome und alle Atome im Radius von 14 Å um das Carbonyl-C-Atom des jeweiligen Liganden sind frei beweglich, alle anderen fixiert.

**Alternative 4:** Alle H-Atome und alle Atome im Radius von 14.0 Å um das Carbonyl-C-Atom des jeweiligen Liganden sind frei beweglich, alle anderen elastisch mit einer Kraftkonstanten von 10 kcal/mol·Å<sup>2</sup> an ihre Start-Positionen gebunden.

**Alternative 5:** Alle Atome sind frei beweglich. Um den Zeitbedarf zu begrenzen werden zunächst 300 Schritte nach dem *steepest descent*-Verfahren durchgeführt, dann der aufwändigere *conjugated gradient*-Algorithmus eingesetzt.

Bei allen eingesetzten Verfahren zur Energie-Minimierung wird der Schwellenwert für die zu berücksichtigenden Atom-Atom-Kombinationen mit 99 Å so groß gewählt, dass *alle* möglichen Atom-Paare in die Berechnung der elektrostatischen und LJ-Energie eingehen (Durchmesser der HIV-PR: ca. 60 Å). Wieder wird  $\varepsilon = 2r$  zur stärkeren Gewichtung der Beiträge nah benachbarter Ladungen benutzt. Die Kraftfeld-Energie wird in allen Fällen nach dem *conjugated gradient*-Verfahren bis zur Konvergenz minimiert.

### Struktur-Vergleiche:

**rmsd<sub>min</sub><sup>start</sup>:** Von den einzelnen Komplexen werden die Strukturen vor und nach der Energie-Minimierung miteinander verglichen.

**rmsd<sub>L</sub><sup>L</sup>:** Für jedes Paar von Komplexen mit *verschiedenen Liganden*, aber *gleicher Start-Konformation des Proteins*, wird die radiale *rmsd*-Verteilung und ihr Mittelwert  $\bar{\varnothing}$  ermittelt. Dabei werden jeweils nur die gemeinsamen Atome beider Komplexe herangezogen, d. h. die Atome, in denen sich die Liganden DMP-323 und DMP-450 unterscheiden, gehen nicht in die Berechnung ein.

**Energie-Rechnungen:** Die Gesamtenergie der Komplexe wird wie zuvor als Summe aus Kraftfeld- und Solvations-Energie berechnet mit  $\varepsilon_{\text{Kraftfeld}} = \varepsilon_{\text{intern}} = 4$  und  $\varepsilon_{\text{extern}} = \varepsilon_{\text{H}_2\text{O}} = 78.5$  (vgl. Abschnitt 2.3). Für eine detaillierte Analyse werden auch die Teilbeiträge der Komplex-Komponenten und die elektrostatischen und LJ-Wechselwirkungen zwischen den Liganden und den beiden Monomeren A und B der HIV-PR angegeben.

## 3.5.2. Ergebnisse

### Zustand vor Minimierung

**Struktur-Vergleich:** Die beiden KS 1mes und 1mer weisen mit 1.9 Å eine recht hohe Auflösung auf. Die Temperaturfaktoren stimmen weitgehend überein (Korrelationskoeffizient bei linearer Regression: 0.91). Den Temperaturfaktoren zufolge sind die Atom-Position im Bereich der Bindetasche und der Liganden besser definiert als in der Peripherie. Dieser Befund deckt sich mit den Ergebnissen der Untersuchung des NMR-Ensembles (s. Kap. 3.4).

Die äquivalenten Atome der beiden KS (vollständiger Protein-Anteil und gemeinsames Grundgerüst der Liganden) zeigen nach Überlagerung eine durchschnittliche Positionsabweichung von 1.13 Å. Die größten Abweichungen betreffen periphere Seitenketten, denn die C<sub>α</sub>-Atome der Hauptketten weichen um nur 0.23 Å voneinander ab. Die Positionen der Atome, die beiden Liganden gemeinsam sind, liegen durchschnittlich 0.36 Å auseinander. Die Konformationen des Grundgerüsts sind also fast

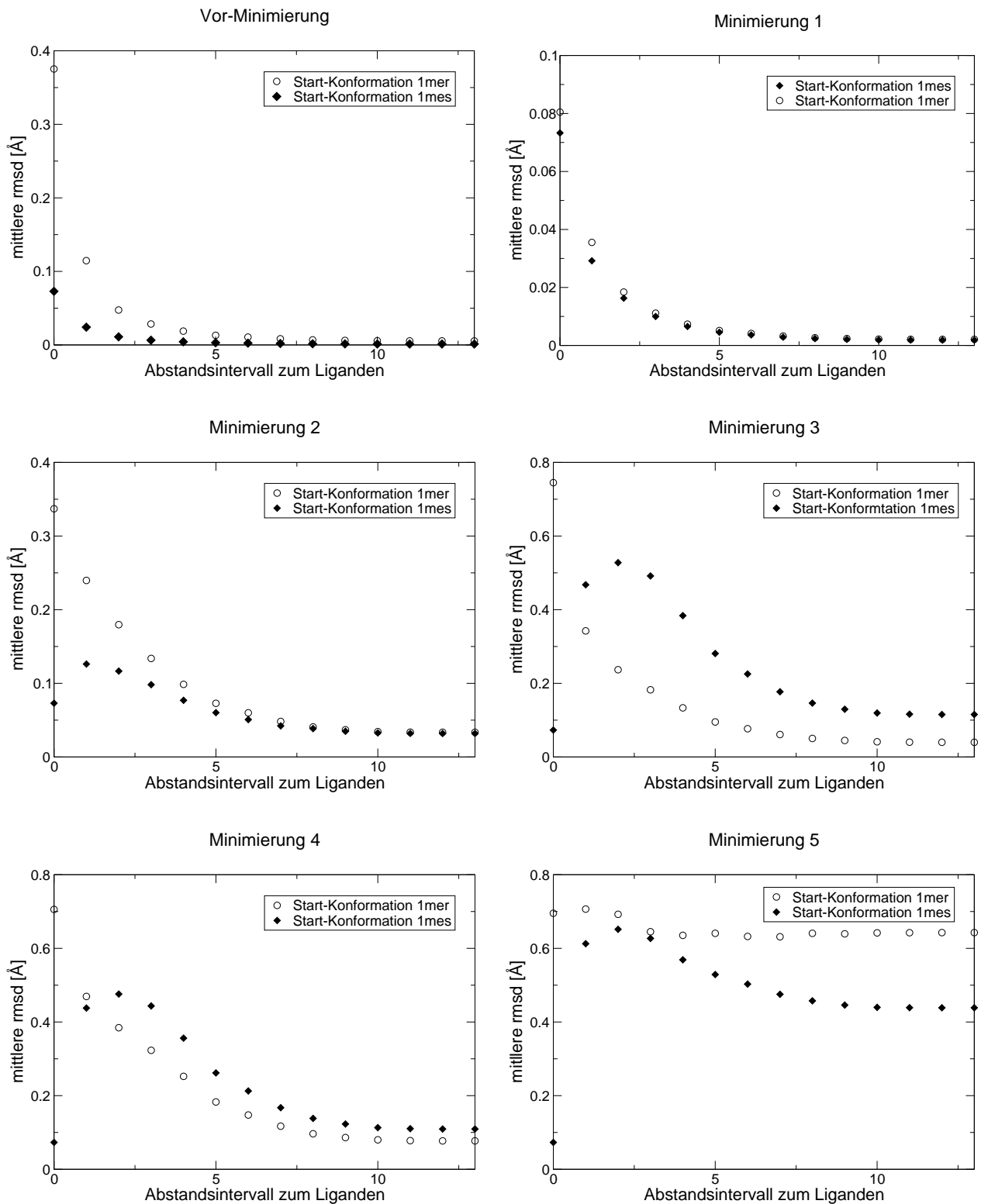


Abbildung 3.9.: Radiale *rmsd*-Verteilungen innerhalb der Komplexe mit den verschiedenen Liganden, jeweils ausgehend von Stammkonformation 1mes bzw. 1mer nach alternativen Verfahren der Energie-Minimierung.

identisch, lediglich die unterschiedlich substituierten aromatischen Ringe sind etwas gegeneinander verdrillt.

**Energie-Vergleich:** Auf die Bestimmung der verschiedenen Energie-Beiträge für die Start-Strukturen wird verzichtet, da die vorangehenden Untersuchungen zum NMR-Ensemble gezeigt haben, dass Strukturen ohne Energie-Minimierung energetisch kaum vergleichbar sind. Außerdem müssen zumindest die Positionen der automatisch hinzugefügten H-Atome vor einer Energie-Rechnung optimiert werden. Zudem haben die manipulierten Liganden willkürliche Konformationen, denn auch wenn sich die Substituenten in Größe und Polarität nur wenig voneinander unterscheiden, kann ihre Verschiebung aus der *meta*- in die *para*-Position (bzw. umgekehrt) zur Kollision mit der Protein-Umgebung führen.

### Zustand nach vorbereitender Energie-Minimierung

Die Energie-Minimierungen unter Fixierung der schweren Atome des Proteins (1662 von 3196 Atomen insgesamt bei 1mes+DMP-323) benötigen etwa zwei Stunden reine Rechenzeit.

*Tabelle 3.6.:* Ergebnisse nach vorbereitender Energie-Minimierung. Die Schweratome des Proteins sind fixiert, die Ligand- und H-Atome frei beweglich. Die Stamm-Komplexe sind mit \* markiert.

		Stamm-Komplex 1mes			Stamm-Komplex 1mer		
		DMP-323*	DMP-450	$\Delta$	DMP-323	DMP-450*	$\Delta$
$\bar{\emptyset} \text{ rmsd}_{L'}^L$	[Å]	0.004			0.037		
$E_{P.L}^{\text{Kraftfeld}}$	[kcal/mol]	-1921.72	-1916.70	-5.02	-1902.16	-1927.14	24.92
$E_P^{\text{Kraftfeld}}$	[kcal/mol]	-1851.63	-1851.57	0.06	-1850.99	-1852.88	1.89
$E_L^{\text{Kraftfeld}}$	[kcal/mol]	12.72	8.17	4.45	27.75	6.85	20.90
$E_{WWA}^{\text{Kraftfeld}}$	[kcal/mol]	-39.58	-38.65	-0.93	-35.97	-41.48	5.49
$E_{WWB}^{\text{Kraftfeld}}$	[kcal/mol]	-41.52	-40.26	-1.26	-37.33	-41.34	4.01
$E_{P.L}^{\text{Solvation}}$	[kcal/mol]	-300.00	-298.42	-1.58	-300.30	-300.67	0.37
$E_P^{\text{Solvation}}$	[kcal/mol]	-312.37	-312.31	0.06	-317.82	-317.76	0.06
$E_{P.L}^{\text{Gesamt}}$	[kcal/mol]	-2221.73	-2215.12	<b>-6.61</b>	-2202.46	-2227.80	<b>25.02</b>

**Struktur-Vergleich:** Die Komplex-Strukturen, die auf die KS 1mes zurückgehen unterscheiden sich nach der Energie-Minimierung um nur 0.004 Å im Durchschnitt. Die

radiale Verteilung der *rmsd*-Werte (s. Abb. 3.9) zeigt, dass dieser Wert auf die Einflüsse der beiden Liganden zurückgeht, deren gemeinsame Atome um 0.07 Å verschoben sind. Obwohl das Protein-Gerüst fixiert ist, pflanzen sich die unterschiedlichen Einflüsse der Liganden auf die H-Atome noch etwa 8 Å durch den Komplex hindurch fort.

Bei den Komplex-Strukturen, die auf der KS 1mer basieren, sind die strukturellen Unterschiede größer: Die gemeinsamen Ligand- und H-Atome liegen um durchschnittlich 0.037 Å auseinander. Wiederum gehen die Abweichungen auf ein Auseinanderdriften der Ligand-Konformationen während der Energie-Minimierung zurück. Die resultierenden Ligand-Konformationen stimmten mit einer mittleren *rmsd* von 0.38 Å schlechter miteinander überein als bei dem anderen Komplex-Paar.

**Energie-Vergleich:** Die unterschiedliche strukturelle Anpassung der manipulierten Liganden an die jeweils fremde Protein-Konformation macht sich im Energie-Inhalt bemerkbar (s. Tab. 3.6): Ein Vergleich der Kraftfeld-Energien der minimierten Stamm-Komplexe und derjenigen der konstruierten Komplexe zeigt, dass die Originale mit -1921.7 kcal/mol für Komplex [1mes·DMP-323] bzw. -1927.1 kcal/mol für Komplex [1mer·DMP-450] signifikant (19 bzw. 10 kcal/mol) günstiger sind, als die Konstrukte mit -1902.2 kcal/mol für [1mer·DMP-323] bzw. -1916.7 kcal/mol für [1mes·DMP-450]. Die Differenzen gehen zurück auf die sehr unterschiedlichen Beiträge der Liganden selbst und ihrer Wechselwirkungen mit dem Enzym. Die Kraftfeld-Beiträge der bloßen Protein-Anteile sind fast konstant.

Der Ligand DMP-323 besitzt in der minimierten Stamm-Struktur eine um fast 15 kcal/mol günstigere Kraftfeld-Energie als im Konstrukt. Gleichzeitig sind auch die Wechselwirkungsenergien zwischen dem Liganden und den beiden Monomeren beim Original um fast 4 kcal/mol günstiger. Bei DMP-450 ist der Unterschied zwischen Original-Struktur und Konstrukt weniger ausgeprägt, da sich die Kraftfeld-Energien für die isolierten Liganden nur um 1.3 kcal/mol unterscheiden. Die Wechselwirkungsenergien sind allerdings wie zuvor beim Konstrukt um ca. 5 kcal/mol ungünstiger als beim Original. Die Beiträge der Solvatationsenergie sind abhängig vom Stamm-Komplex, stimmen paarweise aber sowohl für den Komplex, als auch besonders für den isolierten Protein-Anteil gut überein. Insgesamt resultieren Differenzen  $\Delta E_{P,L}$  der Gesamtenergie der Komplexe, die mit -6.6 bzw. +25 kcal/mol für die zu vergleichenden Paare um fast 32 kcal/mol auseinander liegen und somit nicht vergleichbar sind.

### Zustand nach Energie-Minimierung 1

Die Energie-Minimierungen unter elastischer Anbindung aller 3196 bzw. 3200 Atome der einzelnen Komplexe an die Start-Positionen dauern knapp eine Stunde reine Rechenzeit.

**Struktur-Vergleich:** Die Atome der vor-optimierten Strukturen verschieben sich während der Energie-Minimierung gegenüber ihrer jeweiligen Start-Position um 0.10–0.14 Å

*Tabelle 3.7.:* Ergebnisse nach Energie-Minimierung 1: Alle Atome sind massengewichtet und elastisch an ihre Start-Positionen gebunden mit einer Kraftkonstanten von 10 kcal/Å<sup>2</sup>. Die Stamm-Komplexe sind mit \* markiert.

		Stamm-Komplex 1mes		Δ	Stamm-Komplex 1mer		Δ
		DMP-323*	DMP-450		DMP-323	DMP-450*	
rmsd <sub>min</sub> <sup>start</sup>	[Å]	0.100	0.103		0.144	0.132	
Ø rmsd <sub>L</sub> <sup>L</sup>	[Å]	0.010			0.012		
$E_{P.L}^{\text{Kraftfeld}}$	[kcal/mol]	-2157.41	-2159.59	2.18	-2170.02	-2216.23	46.21
$E_P^{\text{Kraftfeld}}$	[kcal/mol]	-2116.54	-2117.09	0.55	-2159.31	-2161.59	2.28
$E_L^{\text{Kraftfeld}}$	[kcal/mol]	29.02	18.09	10.93	31.35	17.13	14.22
$E_{WWA}^{\text{Kraftfeld}}$	[kcal/mol]	-35.23	-30.15	-5.08	-12.39	-35.23	22.84
$E_{WWB}^{\text{Kraftfeld}}$	[kcal/mol]	-34.65	-30.44	-4.21	-29.67	-36.54	6.87
$E_{P.L}^{\text{Solvation}}$	[kcal/mol]	-302.67	-300.03	-2.64	-304.05	-300.97	-3.08
$E_P^{\text{Solvation}}$	[kcal/mol]	-315.74	-315.68	0.06	-317.78	-318.33	-0.55
$E_{P.L}^{\text{Gesamt}}$	[kcal/mol]	-2460.09	-2459.62	<b>-0.47</b>	-2474.08	-2517.21	<b>43.13</b>

im Mittel. Dabei entwickeln sich die Komplexe, die jeweils von derselben Protein-Konformation abstammen, paarweise offensichtlich in übereinstimmende Richtungen, denn die Abkömmlinge einer KS unterscheiden sich jeweils nur um 0.01 Å. Die größten strukturellen Änderungen erfahren die Liganden mit 0.08 bzw 0.73 Å, gefolgt von den Ligand-nahen Protein-Anteilen bis zu einem Abstand von ca. 10 Å (s. Abb. 3.9).

**Energie-Vergleich:** Tabelle 3.7 zeigt die berechneten Energie-Beiträge der Komplexe. Die Kraftfeld-Energie der Protein-Anteile wird gegenüber der jeweiligen Start-Konformation deutlich gesenkt, allerdings auf Kosten der Ligand-Beiträge und der Wechselwirkungen mit dem Liganden. Verglichen mit den vor-optimierten Strukturen nimmt die Kraftfeld-Energie der Liganden um 3–17 kcal/mol zu, und die der Wechselwirkungen verringert sich um 5–6 kcal/mol. Dabei treten deutliche Unterschiede zwischen den Komplex-Paaren auf: Die Abkömmlinge von 1mes werden in Bezug auf die Kraftfeld-Energie des isolierten Protein-Anteils um ca. 365 kcal/mol günstiger, die 1mer-Abkömmlinge um etwa 400 kcal/mol. Die Kraftfeld-Energien von DMP-323 betragen 29.0 bzw 31.1 kcal/mol, die von DMP-450 18.1 bzw. 17.1 kcal/mol und stimmen paarweise also ebenso gut überein wie die vergleichbaren Protein-Anteile. Die Wechselwirkungsenergien zwischen Ligand und den Monomeren des Enzyms bewegen sich bei den Original-Strukturen zwischen -34.6 und -36.5 kcal/mol, bei den Konstrukten um -30 kcal/mol (bis auf eine Ausnahme: Für DMP-323 beträgt die



Wechselwirkungsenergie mit Monomer A nur -12.4 kcal/mol). Die Solvatationsenergien unterscheiden sich bei dem 1mes-Komplex-Paar um -2.6 kcal/mol, beim 1mer-Paar um -3.1 kcal/mol, sind also durchaus vergleichbar. Die Solvatationsenergien der isolierten Proteine sind mit Differenzen von 0.03 bzw. 0.06 kcal/mol fast konstant. Das zeigt, dass die Liganden die Solvatationsenergien der Komplexe in charakteristischer Weise beeinflussen. Die starken Abweichungen der Wechselwirkungsbeiträge verursachen allerdings die große Diskrepanz von fast 44 kcal/mol für die entscheidenden  $\Delta E_{P,L}$ -Beiträge.

## Zustand nach Energie-Minimierung 2

Die Energie-Minimierungen unter elastischer Anbindung der  $C_\alpha$ -Atome dauern etwa 2.5 Stunden reine Rechenzeit.

*Tabelle 3.8:* Ergebnisse nach Energie-Minimierung 2. Nur die  $C_\alpha$ -Atome sind mit einer Kraftkonstanten von 10 kcal/mol-Å<sup>2</sup> elastisch an ihre Start-Positionen gebunden, die Seitenketten und Liganden sind frei beweglich. Die Stamm-Komplexe sind mit \* markiert.

		Stamm-Komplex 1mes			Stamm-Komplex 1mer		
		DMP-323*	DMP-450	$\Delta$	DMP-323	DMP-450*	$\Delta$
$rmsd_{min}^{start}$	[Å]	0.730	0.736		0.799	0.787	
$\emptyset rmsd_L^L$	[Å]	0.107			0.158		
$E_{P,L}^{Kraftfeld}$	[kcal/mol]	-2913.05	-2913.14	0.09	-2967.21	-2968.47	1.26
$E_P^{Kraftfeld}$	[kcal/mol]	-2840.35	-2843.61	3.26	-2895.52	-2894.56	-0.96
$E_L^{Kraftfeld}$	[kcal/mol]	16.59	8.33	8.26	18.15	7.70	10.45
$E_{WWA}^{Kraftfeld}$	[kcal/mol]	-42.73	-39.55	-3.18	-45.25	-41.87	-3.38
$E_{WWB}^{Kraftfeld}$	[kcal/mol]	-46.55	-38.31	-8.24	-44.58	-39.74	-4.84
$E_{P,L}^{Solvatation}$	[kcal/mol]	-207.74	-205.41	-2.33	-181.79	-183.65	1.86
$E_P^{Solvatation}$	[kcal/mol]	-218.60	-216.99	1.61	-192.72	-196.67	-3.95
$E_{P,L}^{Gesamt}$	[kcal/mol]	-3120.80	-3118.55	<b>-2.25</b>	-3149.00	-3152.12	<b>3.12</b>

**Struktur-Vergleich:** Im Vergleich zur jeweiligen Start-Konformation verschieben sich die Atome bei den 1mes-Abkömmlingen um 0.73 bzw. 0.74 Å, bei den aus 1mer erzeugten Komplexen um 0.79 bzw. 0.80 Å. Paarweise unterscheiden sich die resultierenden Strukturen mit durchschnittlich 0.11 bzw. 0.16 Å wesentlich weniger voneinander, haben sich also abhängig vom Stamm-Komplex jeweils gleichartig verändert. Die radiale *rmsd*-Verteilung (Abb.3.9) zeigt, dass sich die Ligand-Konformationen bei den

Abkömmlingen von KS 1mes um nur 0.07 Å gegeneinander verschoben, während sich die benachbarten Protein-Anteile dem veränderten Liganden anpassen. Umgekehrt driften die Konformationen der Liganden bei den Abkömmlingen von KS 1mer mit 0.34 Å auseinander, doch auch die Ligand-nahen Protein-Anteile verformen sich. Insgesamt pflanzen sich die Konformations-Änderungen, ausgelöst durch die ausgetauschten Substituenten an den Liganden, durch die gesamten Komplexe hindurch fort.

**Energie-Vergleich:** Tabelle 3.8 zeigt, dass die Komplex-Energien der 1mes- und der 1mer-Abkömmlinge paarweise mit Differenzen von 3.3 bzw. -1.0 kcal/mol jeweils gut übereinstimmen; die Paare mit gleichen Liganden allerdings unterscheiden sich um deutliche 50 kcal/mol voneinander. Dieser Unterschied beruht auf den Beiträgen der isolierten Protein-Anteile und ist somit abhängig von der Stamm-Konformation des Proteins.

Die Kraftfeld-Energien für die Liganden sind weitgehend unabhängig von den Protein-Anteilen und mit 16.6 bzw. 18.1 kcal/mol für DMP-323 und 8.3 bzw. 7.7 kcal/mol für DMP-450 jeweils vergleichbar. Auch die Wechselwirkungsenergien zwischen den Liganden und den beiden Monomeren des Enzyms werden dominiert von den Einflüssen der Liganden: Mit -42.7 und -46.6 bzw. -45.3 und -44.6 kcal/mol sind die Wechselwirkungen zwischen der HIV-PR und DMP-323 günstiger als die zwischen der HIV-PR und DMP-450 mit -39.6 und -38.3 bzw. -41.9 und -39.7 kcal/mol. Die Solvatationsenergien der Komplexe sind wiederum stark abhängig von den Protein-Anteilen und verringern das energetische Gefälle zwischen den Komplex-Paaren auf ca. 30 kcal/mol. Die entscheidenden Differenzen  $\Delta E_{P,L}$  betragen -2.3 bzw. +3.1 kcal/mol, liegen also um 5.4 kcal/mol auseinander.

### Zustand nach Energie-Minimierung 3

Eine Energie-Minimierung unter Fixierung aller Schweratome mit mehr als 14 Å-Abstand zum Carbonyl-C-Atom des Liganden (1075 von 3196 Atomen) benötigt etwa zwei Stunden reine Rechenzeit. Alle Ligand- und H-Atome sind frei beweglich, ebenso die Protein-Atome bis zu einer Entfernung von ca. 14 Å zum Liganden.

**Struktur-Vergleich:** Die Atome verschieben sich während der Minimierung um durchschnittlich 0.37 – 0.42 Å relativ zur Start-Konformation. Die Konformationen der 1mer-Abkömmlinge gehen konform (mittlere *rmsd* zwischen den Komplexen nur 0.09 Å), während die Abkömmlinge von 1mes stärker auseinander driften (mittlere *rmsd*: 0.36). Der Vergleich der radialen *rmsd*-Verteilungen (Abb. 3.9) zeigt, dass die Ligand-Konformationen bei den 1mes-Komplexen erhalten bleiben (*rmsd* 0.08 Å), während die frei beweglichen Protein-Anteile innerhalb der Kugel um das Carbonyl-C-Atom sich stärker verschieben. Bei den 1mer-Komplexen ist es umgekehrt: Die Ligand-Atome verschieben sich um 0.74 Å und die Protein-Anteile behalten ihre Positionen weitgehend bei.

*Tabelle 3.9.:* Ergebnisse nach Energie-Minimierung 3. Die Schweratome mit einem Abstand von mehr als 14 Å zum Carbonyl-C-Atom des jeweiligen Liganden bleiben fixiert, während die Ligand- und H-Atome, sowie alle Protein-Atome innerhalb dieses Bereiches frei beweglich sind. Die Stamm-Komplexe sind mit \* markiert.

		Stamm-Komplex 1mes			Stamm-Komplex 1mer		
		DMP-323*	DMP-450	$\Delta$	DMP-323	DMP-450*	$\Delta$
$rmsd_{min}^{start}$	[Å]	0.420	0.366		0.411	0.417	
$\emptyset rmsd_{L'}^L$	[Å]	0.359			0.094		
$E_{P.L}^{Kraftfeld}$	[kcal/mol]	-2182.36	-2176.35	-6.01	-2194.18	-2196.57	2.39
$E_P^{Kraftfeld}$	[kcal/mol]	-2107.90	-2103.02	-4.88	-2119.02	-2118.42	0.60
$E_L^{Kraftfeld}$	[kcal/mol]	17.28	5.48	11.80	19.50	9.49	10.49
$E_{WWA}^{Kraftfeld}$	[kcal/mol]	-45.88	-40.37	-5.51	-49.71	-47.97	-1.74
$E_{WWB}^{Kraftfeld}$	[kcal/mol]	-45.86	-38.44	-7.42	-44.95	-39.67	-5.28
$E_{P.L}^{Solvatation}$	[kcal/mol]	-276.41	-278.39	1.98	-276.49	-278.42	1.93
$E_P^{Solvatation}$	[kcal/mol]	-288.96	-292.08	3.12	-288.19	-290.14	1.95
$E_{P.L}^{Gesamt}$	[kcal/mol]	-2458.78	-2454.75	<b>-4.03</b>	-2470.68	-2474.99	<b>4.31</b>

**Energie-Vergleich:** In Tabelle 3.9 sind die Energie-Beiträge aufgeführt. Die Differenzen in der Kraftfeld-Energie der Komplexe betragen  $-6.0$  kcal/mol für die Abkömmlinge von KS 1mes und  $+2.4$  kcal/mol für die Abkömmlinge von KS 1mer, sind also deutlich verschieden. Beim 1mes-Komplex-Paar unterscheiden sich die Protein-Anteile um  $4.9$  kcal/mol; beim 1mer-Paar um nur  $0.6$  kcal/mol. Diese Energie-Differenzen zwischen den Paaren korrelieren mit entsprechenden *rmsd*-Werten von  $0.36$  bzw.  $0.09$  Å. Die Ligand-Beiträge stimmen mit  $17.3$  bzw.  $19.5$  kcal/mol für DMP-323 und  $5.5$  bzw.  $9.5$  kcal/mol für DMP-450 wesentlich besser überein als die Wechselwirkungsenergien zum Protein. Diese unterscheiden sich sowohl zwischen den Komplex-Paaren, als auch zwischen den Komplexen mit gleichem Liganden und auch innerhalb eines Komplexes zwischen beiden Monomeren stark voneinander und bewegen sich im Bereich von  $-38.4$  –  $-49.7$  kcal/mol. Eine klare Abhängigkeit von der Stamm-Konformation oder der Art des Liganden lässt sich nicht erkennen. Die Solvatationsenergien sind mit  $-276.4$  bzw.  $-276.5$  kcal/mol für die DMP323-Komplexe und  $-278.39$  bzw.  $-278.42$  kcal/mol für die DMP-450-Komplexe fast deckungsgleich. Dennoch unterscheiden sich die  $\Delta E_{P.L}$ -Werte zwischen den Komplex-Paaren mit  $-4.0$  bzw.  $+4.3$  kcal/mol um über  $8$  kcal/mol voneinander.

### Zustand nach Energie-Minimierung 4

Die Energie-Minimierungen unter elastischer Anbindung aller Schweratome mit mehr als 14 Å Entfernung zum Carbonyl-C-Atom des jeweiligen Liganden dauern etwa 3.5 Stunden reine Rechenzeit. Dabei sind die Ligand- und H-Atome sowie Protein-Atome in einem Bereich von etwa 8 Å um den Liganden frei beweglich.

*Tabelle 3.10.:* Ergebnisse nach Energie-Minimierung 4: Die Schweratome mit einem Abstand von mehr als 14 Å zum Carbonyl-C-Atom des Liganden sind elastisch an ihre Ausgangspositionen gebunden, die Ligand- und H-Atome, sowie alle Protein-Atome innerhalb dieses Bereiches sind frei beweglich. Die Stamm-Liganden sind mit \* markiert.

		Stamm-Komplex 1mes			Stamm-Komplex 1mer		
		DMP-323 *	DMP-450	Δ	DMP-323	DMP-450*	Δ
rmsd <sup>start</sup> <sub>min</sub>	[Å]	0.431	0.412		0.430	0.441	
Ø rmsd <sup>L</sup> <sub>L'</sub>	[Å]	0.349			0.238		
$E_{P.L}^{\text{Kraftfeld}}$	[kcal/mol]	-2598.02	-2597.07	-0.95	-2650.91	-2650.54	-0.37
$E_P^{\text{Kraftfeld}}$	[kcal/mol]	-2523.40	-2520.87	-2.53	-2574.56	-2571.36	-3.20
$E_L^{\text{Kraftfeld}}$	[kcal/mol]	17.25	6.33	10.92	19.16	8.34	10.82
$E_{WWA}^{\text{Kraftfeld}}$	[kcal/mol]	-46.26	-39.16	-7.10	-50.22	-46.30	-3.92
$E_{WWB}^{\text{Kraftfeld}}$	[kcal/mol]	-45.62	-43.37	-2.25	-45.29	-41.22	-4.07
$E_{P.L}^{\text{Solvatation}}$	[kcal/mol]	-273.94	-273.01	-0.93	-273.71	-277.18	3.47
$E_P^{\text{Solvatation}}$	[kcal/mol]	-287.04	-287.06	0.02	-284.56	-288.98	4.42
$E_{P.L}^{\text{Gesamt}}$	[kcal/mol]	-2871.96	-2870.08	<b>-1.88</b>	-2924.62	-2927.72	<b>3.10</b>

**Struktur-Vergleich:** Die Atom-Positionen der einzelnen Komplexe verschieben sich während der Energie-Minimierung um 0.41 – 0.44 Å. Gleichzeitig verformen sich die Konformationen der 1mes-Abkömmlinge mit einer Abweichung von 0.35 Å untereinander stärker als die der 1mer-Abkömmlinge mit nur 0.24 Å. Wieder verhalten sich beide Komplex-Paare unterschiedlich: Während sich bei den 1mer-Abkömmlingen die Ligand-Konformationen deutlicher unterscheiden, sind es bei den 1mes-Abkömmlingen die Bindetaschen, die stärker voneinander abweichen als die entsprechenden Regionen bei den Vergleichskomplexen (s. Abb. 3.9).

**Energie-Vergleich:** Tabelle 3.10 zeigt die verschiedenen Energie-Beiträge. Trotz der strukturellen Unterschiede der Protein-Anteile haben ihre Kraftfeld-Beiträge mit Differenzen von -2.53 kcal/mol für das 1mes-Komplex-Paar und -3.20 kcal/mol für das

1mer-Komplex-Paar jeweils die gleiche Größenordnung und auch das gleiche Vorzeichen. Auch die Ligand-Beiträge stimmen gut überein; die Differenzen zwischen den Komplex-Paaren betragen 10.9 bzw. 10.8 kcal/mol für die 1mes- bzw. 1mer-Abkömmlinge. Die Wechselwirkungsenergien jedoch schwanken stärker; im Vergleich sind die Beziehungen von DMP-323 zum Protein bei beiden Komplex-Paaren günstiger als diejenigen von DMP-450. Mit  $-273.0 - -273.9$  kcal/mol sind die Solvatationsenergien für drei von den vier Komplexen fast konstant, lediglich für die Original-Struktur von 1mer mit DMP-450 liegt sie mit  $-277.2$  kcal/mol etwa 3 kcal/mol niedriger. Die Gesamtenergien beider Komplex-Paare  $\Delta E_{P,L}$  unterscheiden sich mit Werten von  $-1.9$  bzw.  $3.1$  kcal/mol um 5.0 kcal/mol voneinander.

### Zustand nach Energie-Minimierung 5

Alle 3196 bzw. 3200 Atome sind frei beweglich. Obwohl für die Energie-Minimierung 300 Schritte des schnelleren *steepest-descent*-Verfahrens dem *conjugated-gradient*-Verfahren bis zur Konvergenz vorgeschaltet werden, benötigt ein Durchlauf etwa 5.4 Stunden reine Rechenzeit.

Tabelle 3.11.: Ergebnisse nach Energie-Minimierung 5. Alle 3200 Atome sind frei beweglich. Die Stamm-Liganden sind mit \* markiert.

		Stamm-Komplex 1mes			Stamm-Komplex 1mer		
		DMP-323*	DMP-450	$\Delta$	DMP-323	DMP-450*	$\Delta$
$rmsd_{min}^{start}$	[Å]	1.393	1.362		1.398	1.340	
$\emptyset rmsd_L^L$	[Å]	0.561			0.794		
$E_{P,L}^{Kraftfeld}$	[kcal/mol]	-3156.05	-3151.70	-4.35	-3200.44	-3196.95	-3.49
$E_P^{Kraftfeld}$	[kcal/mol]	-3074.86	-3072.17	-2.69	-3116.63	-3110.26	-6.37
$E_L^{Kraftfeld}$	[kcal/mol]	17.36	6.24	11.12	16.31	6.30	10.01
$E_{WWA}^{Kraftfeld}$	[kcal/mol]	-52.29	-45.44	-6.85	-52.69	-48.77	-3.92
$E_{WWB}^{Kraftfeld}$	[kcal/mol]	-46.26	-40.33	-5.93	-47.43	-44.21	-3.23
$E_{P,L}^{Solvatation}$	[kcal/mol]	-170.67	-169.32	-1.35	-169.91	-169.04	-0.87
$E_P^{Solvatation}$	[kcal/mol]	-180.79	-182.50	1.71	-181.11	-181.82	0.71
$E_{P,L}^{Gesamt}$	[kcal/mol]	-3326.73	-3321.03	<b>-5.70</b>	-3370.36	-3366.00	<b>-4.36</b>

**Struktur-Vergleich:** Während der Energie-Minimierung verschieben sich die Atome aller Komplexe um  $1.34 - 1.40$  Å. Paarweise unterscheiden sich die Strukturen im Fall der

1mes-Abkömmlinge um 0.56 Å, im Fall der 1mer-Komplexe um 0.79 Å. Die strukturellen Unterschiede zwischen den 1mer-Abkömmlingen erstrecken sich fast konstant über den gesamten Komplex, bei den 1mes-Abkömmlingen fast ausschließlich nur über den Protein-Anteil und dort besonders über die Ligand-nahen Regionen (s. Abb.3.9).

**Energie-Vergleich:** Die einzelnen Energie-Beiträge werden in Tabelle 3.11 aufgeführt. Die weitreichenden Struktur-Veränderungen innerhalb der Komplexe senken die Kraftfeld-Energien um ca. 1330 bzw 1380 kcal/mol gegenüber den vor-optimierten Start-Strukturen ab. Es resultieren Energie-Differenzen für die Protein-Anteile von -2.7 bzw. -6.4 kcal/mol. Die Differenzen der Ligand-Beiträge stimmen in beiden Komplex-Paaren mit 11.1 bzw. 10.0 kcal/mol gut überein. Für die Komplexe, die DMP-323 als Liganden enthalten, sind die LJ und Coulomb-Beiträge vergleichbar (-52.3 und -46.3 kcal/mol gegenüber -52.7 und -47.4 kcal/mol). Die Wechselwirkungsenergien der DMP-450-Komplexe betragen -45.4 und -40.3 kcal/mol bzw. -48.8 und -44.2 kcal/mol und sind damit übereinstimmend ungünstiger als diejenigen für DMP-323, weichen aber dennoch stark voneinander ab. Die Solvatationsenergien sind mit einem Unterschied von -1.4 kcal/mol für die 1mes-Komplexe und -0.9 kcal/mol für die 1mer-Komplexe fast konstant und zeigen übereinstimmende Vorzeichen. Trotz der Schwankungen stimmen die resultierenden  $\Delta E_{P.L}$ -Werte mit -5.7 bzw. -4.4 kcal/mol in Größenordnung und Vorzeichen bis auf eine Abweichung von 1.3 kcal/mol überein ( $\bar{\Delta} = 5.03 \pm 0.95$  kcal/mol).

### 3.5.3. Diskussion

Es wurde untersucht, ob der Beitrag  $\Delta E_{P.L}$  zur relativen Bindestärke  $\Delta\Delta G$  bei der Konstruktion von Komplexen unabhängig von der Stamm-Konformation ist, bzw. mit welchen Randbedingungen sich  $\Delta E_{P.L}$  durch eine Energie-Minimierung vereinheitlichen lässt. Dazu wurden parallel an zwei äquivalenten Paaren von Komplexen fünf verschiedene alternative Randbedingungen erprobt.

Die geringe Differenz der Gesamtenergien bei den Komplexen mit identischer Konfiguration von 13–20 kcal/mol nach der vorbereitenden Energie-Minimierung zeigt, dass Kristall-Strukturen prinzipiell gute Ausgangsdaten sind. Nach der vorangegangenen Studie des NMR-Ensembles hätte man erwarten können, dass die Energien der nur wenig struktur-optimierten Komplexe wesentlich schlechter übereinstimmen. Im Gegenteil: Energetische Unterschiede treten erst im Zuge weiterer Energie-Minimierungen deutlicher hervor.

Die Randbedingungen sind so gewählt, dass den Atomen zunehmend mehr Bewegungsspielraum eingeräumt wird. Am Vergleich der radialen *rmsd*-Verteilungen ist zu erkennen, wie die Start-Konformation die strukturellen Veränderungen während der Energie-Minimierung beeinflusst: Die KS 1mes enthält mit DMP-323 und seiner HO-CH<sub>2</sub>-Gruppe am Aromaten den größeren Substituenten im Vergleich zur NH<sub>2</sub>-Gruppe von DMP-450. Bei der Konstruktion des DMP-450-Komplexes ausgehend von KS 1mes wird also der Ligand insgesamt verkleinert. Der neue Ligand kollidiert offenbar kaum mit Protein-Atomen der Bindetasche, die an den größeren Liganden DMP-323 angepasst sind. Umgekehrt scheint

die Konstruktion des DMP-323-Komplexes aus der KS 1mer schwieriger zu sein, denn hierbei sind die Energie-Beiträge des Liganden durchgängig höher als in der Original-Struktur. Die *rmsd*-Verteilungen zeigen, dass besonders dieser Ligand sich während der Energie-Minimierungen strukturell reorganisieren muss.

Die Einzelbeiträge zur Gesamtenergie geben Aufschluss über die Art der Protein-Ligand-Beziehungen. Die Beiträge der isolierten Komplex-Komponenten stimmen erwartungsgemäß meist gut überein: Die Liganden mit gleicher Konfiguration und die Proteine mit gleicher Start-Konformation zeigen jeweils vergleichbare Energie-Beiträge. Die Wechselwirkungsenergien zwischen Protein und Ligand jedoch schwanken unvorhersehbar. Trotz der prinzipiell  $C_2$ -symmetrischen Komplexe sind die Wechselwirkungen zwischen den Liganden und den einzelnen Monomeren des Proteins oft sehr verschieden. Das deutet darauf hin, dass der Bindemodus der aromatischen Substituenten der Liganden in der unpolaren Bindetasche nicht eindeutig festgelegt ist.

Wie schon bei der NMR-Studie, ist zu beobachten, dass sich die  $\Delta E_{P,L}$ -Beiträge einander umso mehr angleichen, je deutlicher sich die Komplexe strukturell voneinander unterscheiden. Die beste Übereinstimmung von  $\Delta E_{P,L}$ -Beiträgen wird mit der langwierigen Energie-Minimierung ohne Randbedingungen erzielt; dennoch bleibt eine Unsicherheit von fast 1 kcal/mol für den  $\Delta E_{P,L}$ -Wert. Außerdem bleibt unklar, ob die gute Korrelation dieser letzten  $\Delta E_{P,L}$ -Werte in anderen Anwendungsbeispielen reproduzierbar ist, denn die einzelnen Energie-Beiträge und die Strukturen der Komplexe weichen stark voneinander ab.

#### 3.5.4. Schlussfolgerungen

Diese Studie zur Komplex-Konstruktion belegt, dass selbst ohne Randbedingungen während der Energie-Minimierung die strukturellen Grundzüge erhalten bleiben. Auf detaillierte Strukturanalysen kann also verzichtet werden, ebenso auf die Bestimmung der einzelnen Teilbeiträge zur Gesamtenergie, denn die Ligand- und Wechselwirkungsenergien schwanken im Vergleich stärker als die Gesamtenergien. Während einer Struktur-Optimierung müssen die Atome von Ligand und Protein mindestens innerhalb des Kontakt-Bereichs beweglich sein, damit die Moleküle sich einander anpassen können.

Die Reproduzierbarkeit der  $\Delta E_{P,L}$ -Beiträge ist mit einer Standard-Abweichung von 1 kcal/mol schlechter als angenommen und somit höchstens für grob qualitative Aussagen zur relativen Bindestärke  $\Delta\Delta G$  ausreichend. Da mit zwei Komplexen allerdings nur eine „Minimal-Serie“ vorliegt, besteht die Möglichkeit, dass die Resultate nur in diesem Fall unbefriedigend sind und auf der zwangsläufigen Kollision mit dem Protein bei der Vergrößerung des Liganden beruhen. Dieses Problem entfällt bei gleichgroßen oder kleineren Liganden. Um abschließend zu beurteilen, ob der statische Ansatz zur Bestimmung relativer Bindestärken brauchbar ist, muss eine größere Serie von Komplexen herangezogen werden.

## 3.6. Relative Bindestärken für eine Serie von 10 Liganden

Bindestärken von Wirkstoffen an Proteinen vorhersagen zu können, ist therapeutisch von großem Wert. In Kapitel 2 wird eine Methode vorgestellt, nach der relative Bindestärken zu berechnen sind. Um diese Modellierungsmethode zu validieren, sollen die relativen Bindestärken  $\Delta\Delta G$  für eine Serie von 10 Cyclooctylpyranon-Liganden (s. Tab. 3.12) danach berechnet und mit experimentellen Werten verglichen werden.

Die repräsentativen Konformationen der Komplexe werden wie bereits beschrieben (s. Kap. 3.5) von einer KS abgeleitet und einer Struktur-Optimierung unterzogen. Die repräsentativen Konformationen der freien Liganden werden durch MD-Simulation, ausgehend von der jeweiligen Konformation des Liganden im Komplex mit der HIV-PR, erzeugt. Aus den  $\Delta E_{P-L}$ - und  $\Delta E_L$ -Beiträgen werden die  $\Delta\Delta G$ -Werte ermittelt. Es gilt, die folgenden Fragen zu beantworten:

- Welche speziellen Randbedingungen sind am besten für die Struktur-Optimierung der Komplexe geeignet?
- Wie erhält man repräsentative Konformationen für den freien Liganden in Lösung?
- Wie gut ist das experimentelle  $\Delta\Delta G$ -Profil reproduzierbar?

### 3.6.1. Durchführung

**Komplex-Konstruktion:** Der als KS 7upj vorliegende Komplex der HIV-PR mit COP1 als Ligand dient als Stamm-Struktur zur Erzeugung neun weiterer Komplexe. Tabelle 3.12 zeigt das den Liganden gemeinsame Grundgerüst und die variablen Substituenten, die entsprechenden Inhibitionskonstanten [37] sowie die daraus hervorgehenden relativen Bindestärken  $\Delta\Delta G_{\text{exp}}$  (s. Gl. 2.5). Unter Quanta werden die Liganden editiert, entsprechende Topologie-Dateien erstellt und nach der *charge-templates*-Methode atomare Ladungen zugewiesen. Die in der KS fehlenden H-Atome werden automatisch von CHARMM ergänzt.

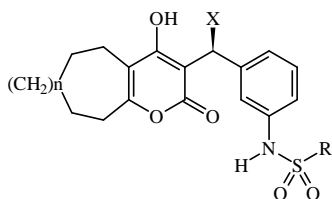
**Struktur-Optimierung:** Die 10 Komplexe werden parallel mehreren Energie-Minimierungen mit verschiedenen, jeweils allgemein geltenden Randbedingungen bis zur Konvergenz unterzogen.

**Alternative 1:** Alle Protein-Atome bleiben fixiert in der Konformation des Stamm-Komplexes, lediglich die Ligand-Atome sind frei beweglich. Elektrostatische Wechselwirkungen werden mit  $\varepsilon = 4$  berechnet.

**Alternative 2:** Alle Atome im Radius von 14 Å um das stereogene C-Atom der Liganden sowie alle H-Atome sind frei beweglich, die restlichen sind fixiert. Elektrostatische Wechselwirkungen werden mit  $\varepsilon = 4$  berechnet.



Tabelle 3.12.: Liganden und experimentelle Daten [81]



Ligand	$n$	X	R	$K_i$ [nM]	$\Delta\Delta G_{\text{exp}}$ [kcal/mol]
COP1	2			3.2	0
COP2	2			32	1.37
COP3	2			0.2	-1.65
COP4	2			33	1.39
COP5	2			4.1	0.12
COP6	2			0.4	-1.24
COP7	1		-CH <sub>3</sub>	120	2.16
COP8	2		-CH <sub>3</sub>	11	0.74
COP9	2			2.5	-0.15
COP10	2			18	1.03

**Alternative 3:** Alle Atome im Radius von 14 Å um das stereogene C-Atom der Liganden sowie alle H-Atome sind frei beweglich, die restlichen sind fixiert. Elektrostatische Wechselwirkungen werden mit  $\varepsilon = 4r$  berechnet.

**Alternative 4:** Alle Atome im Radius von 14 Å um das stereogene C-Atom der Liganden sind frei beweglich, die restlichen elastisch mit einer Kraftkonstanten von 10 kcal/mol·Å<sup>2</sup> an ihre Positionen gebunden. Elektrostatische Wechselwirkungen werden mit  $\varepsilon = 2r$  berechnet.

**Alternative 5:** Alle Seitenketten im Abstand von 4 Å zum Liganden, sowie die Liganden selbst sind frei beweglich. Die restlichen Atome sind elastisch mit einer Kraftkonstanten von 10 kcal/mol·Å<sup>2</sup> an ihre Positionen gebunden. Elektrostatische Wechselwirkungen werden mit  $\varepsilon = 2r$  berechnet.

**Alternative 6:** Alle Atome sind elastisch mit einer Kraftkonstanten von 10 kcal/mol·Å<sup>2</sup> an ihre Positionen gebunden. Elektrostatische Wechselwirkungen werden mit  $\varepsilon = 2r$  berechnet.

Bei allen Energie-Minimierungen wird der Schwellenwert für die zu berücksichtigenden Atom-Atom-Kombinationen mit 99 Å so groß gewählt, dass alle möglichen Paare in die Berechnungen der elektrostatischen und LJ-Wechselwirkungen eingehen.

Um sicherzugehen, dass der generelle Bindemodus während einer Minimierung erhalten bleibt, werden jeweils die *rmsd*-Werte der äquivalenten Atome der verschiedenen Liganden einer Serie mit COP1 als Referenz bestimmt.

**Liganden:** Um für jeden einzelnen Liganden die eine, repräsentative Konformation zu ermitteln, die der in Lösung günstigsten Konformation entspricht, wird eine MD-Simulation ausgehend von der Ligand-Konformation im Komplex mit der HIV-PR durchgeführt. Zunächst wird ein „Wärmebad“ mit einer Temperatur von 10 000 K simuliert, indem einzelne Atome zufällig ausgewählt und verschoben werden. Die hohe Temperatur des Wärmebads gewährleistet, dass die verschiedensten Konformationen erzeugt werden können (wird die Temperatur zu niedrig gewählt, können Energiebarrieren verhindern, dass die Konformation mit der niedrigsten Energie eingenommen werden kann).

Die nach 3000 Zeitschritten vorliegende, „heiße“ Konformation eines Liganden wird einer Energie-Minimierung bis zur Konvergenz unterzogen (max. 2000 Iterationsschritte mit  $\varepsilon = 2r$ ) und abgespeichert. Die Kraftfeld-Energie dieser „abgekühlten“ Konformation eines lokalen Minimums wird berechnet. Gleichzeitig dient sie als Start-Konformation für eine erneute MD-Simulation nach gleichem Muster. Auf diese Weise werden 2000 Konformationen für den ersten Liganden, COP1, erzeugt. Nach Auswertung der Kraftfeld-Energien und ihrer Häufigkeiten wird die notwendige Größe des zu erzeugenden Ensembles für die neun weiteren Liganden bestimmt.

Exemplarisch werden für alle Konformationen des Stammliganden COP1 die Solvatationsenergien berechnet. Anhand dieser Daten wird ermittelt, ob irrelevante Konformationen schon auf Grundlage der Kraftfeld-Energie ausgeschlossen werden können, sodass sich der

Rechenaufwand für die Bestimmung der Solvatationsenergien reduziert. Bei den restlichen Liganden werden nur noch für die potentiell relevanten Konformationen die Solvatationsenergien bestimmt.

**Energie-Rechnungen:** Die Gesamtenergie der Komplexe und Liganden wird als Summe aus Kraftfeld- und Solvatations-Energie berechnet mit  $\varepsilon_{\text{Kraftfeld}} = \varepsilon_{\text{intern}} = 4$  und  $\varepsilon_{\text{extern}} = 78.5$  (vgl. Abschnitt 2.3).

Um eventuell störende, Protein-bedingte Schwankungen in der Gesamt-Energie auszublenken, wird der Beitrag des isolierten Proteins wie zuvor abgezogen und nur der Beitrag des Liganden und der Wechselwirkungsenergien im Komplex verwendet (s. Kap. 3.4.1, Gl. 3.3).

$$E_{L+WW}^{\text{Gesamt}} = E_{L+WW}^{\text{Kraftfeld}} + E_{L+WW}^{\text{Solvatation}}$$

### 3.6.2. Ergebnisse

#### Komplexe

Die Rechenzeiten zur Energie-Minimierung sind abhängig von den gewählten Randbedingungen. Sie betragen für die verschiedenen Verfahren zwischen 3 Minuten und 2 Stunden:

1. ca. 3 Minuten
2. ca. 30 Minuten
3. ca. 30 Minuten
4. ca. 40 Minuten
5. ca. 2 Stunden
6. ca. 40 Minuten

Gemäß den vorgegebenen Randbedingungen resultieren nach einer Energie-Minimierung aus einer Start-Struktur unterschiedliche Konformationen. Tabelle 3.13 zeigt die Gesamtenergien der einzelnen Komplexe bzw. die vom Protein-Anteil bereinigten Energie-Beiträge des Liganden und seiner Wechselwirkungen mit dem Protein, abhängig vom jeweiligen Minimierungsverfahren.

Die Beiträge der verschiedenen Komplexe zur relativen Bindestärke  $\Delta\Delta G$  werden als Energie-Differenz zum Referenz-Liganden COP1 berechnet. Tabelle 3.14 bzw. die Abbildungen 3.12(a) und 3.12(b) zeigen die Differenz-Profile, die abhängig von den Varianten der Energie-Minimierung resultieren, sowie zum Vergleich die aus dem Experiment bekannten  $\Delta\Delta G_{\text{exp}}$ -Werte [37].

Tabelle 3.15 illustriert, in welchem Ausmaß sich die Ligand-Konformationen während der verschiedenen Energie-Minimierungen verändern. Die jeweils nach dem Verfahren 6

Tabelle 3.13.: Gesamtenergien bzw. Ligand- und Wechselwirkungsenergien (s. Gl. 3.3) der Komplexe nach Energie-Minimierung unter verschiedenen Randbedingungen

Gesamtenergie der Komplexe [kcal/mol] je nach Minimierungsverfahren						
	1	2	3	4	5	6
COP1	-2323.64	-2505.21	-2503.94	-2854.40	-3325.17	-2627.11
COP2	-2325.78	-2506.17	-2506.41	-2862.83	-3330.23	-2580.51
COP3	-2317.17	-2505.47	-2495.12	-2841.86	-3306.59	-2616.93
COP4	-2324.25	-2507.54	-2505.35	-2853.59	-3333.41	-2606.59
COP5	-2318.95	-2501.12	-2497.55	-2851.15	-3318.73	-2585.66
COP6	-2331.84	-2515.96	-2510.61	-2863.95	-3335.19	-2595.11
COP7	-2320.44	-2503.86	-2500.44	-2846.79	-3327.74	-2605.92
COP8	-2318.68	-2503.99	-2501.33	-2849.93	-3320.79	-2621.85
COP9	-2329.62	-2510.77	-2510.12	-2860.75	-3336.84	-2610.82
COP10	-2331.31	-2514.08	-2513.40	-2807.10	-3338.05	-2583.65
∅	-2324.2	-2507.4	-2504.4	-2849.2	-3327.3	-2603.4
	±5.4	±4.7	±5.9	±16.4	±9.8	±16.5

Ligand- und Wechselwirkungsenergien [kcal/mol]						
	1	2	3	4	5	6
COP1	326.11	322.70	332.99	-32.54	230.09	315.50
COP2	328.25	331.78	341.93	-25.25	236.78	277.75
COP3	319.64	320.66	324.85	-44.31	213.17	305.41
COP4	326.72	325.79	335.51	-29.94	223.77	295.08
COP5	321.42	322.35	331.91	-31.20	228.62	279.96
COP6	334.31	335.69	344.02	-15.79	242.76	289.40
COP7	322.91	318.64	326.19	-43.58	222.18	294.50
COP8	321.15	319.88	327.30	-39.58	209.41	310.33
COP9	332.09	329.22	339.67	-26.58	233.73	299.29
COP10	333.78	332.22	342.48	-79.70	236.56	272.15
∅	326.6±5.4	325.9±6.0	334.7±7.1	-36.8±17.4	227.7±10.6	293.9±14.3

resultierenden Komplexe dienen dabei als Referenz-Strukturen, da diese den Ausgangskonformationen am nächsten kommen. Mit *rmsd*-Werten von 0.6–2.0 Å ist belegt, dass die Liganden ihren allgemeinen Bindemodus trotz struktureller Veränderungen beibehalten.

*Tabelle 3.14.:* Aus den Gesamt-Energien bzw. den Ligand- und Wechselwirkungsenergien der Komplexe werden die Komplex-Beiträge  $\Delta E_{P-L}$  bzw.  $\Delta E_{L+WW}$  berechnet:

$$\Delta E_{P-L} = E_{gesamt}^{COP1} - E_{gesamt}^{COPi} \text{ bzw. } \Delta E_{L+WW} = E_{L+WW}^{COP1} - E_{L+WW}^{COPi}$$

Komplex-Beiträge $\Delta E_{P-L}$ aus den Gesamtenergien [kcal/mol] nach Minimierungsverfahren							
	1	2	3	4	5	6	$\Delta\Delta G_{exp}$
COP1	0.00	0.00	0.00	0.00	0.00	0.00	0.00
COP2	2.13	0.95	2.46	8.43	3.46	-46.60	1.37
COP3	-6.47	0.25	-8.81	-12.54	-6.39	-10.18	-1.65
COP4	0.60	2.33	1.40	-0.81	-4.24	-20.52	1.39
COP5	-4.69	-4.09	-6.38	-3.24	-7.28	-41.45	0.12
COP6	8.20	10.75	6.67	9.55	7.25	-32.00	-1.24
COP7	-3.20	-1.35	-3.50	-7.61	-4.5	-21.19	2.16
COP8	-4.96	-1.22	-2.61	-4.46	-4.39	-5.26	0.74
COP9	5.98	5.55	6.18	6.35	7.71	-16.29	-0.15
COP10	7.66	8.86	9.46	-47.30	12.45	-43.46	1.03

Wechselwirkungsbeiträge $\Delta E_{L+WW}$ [kcal/mol]							
	1	2	3	4	5	6	$\Delta\Delta G_{exp}$
COP1	0.00	0.00	0.00	0.00	0.00	0.00	0.00
COP2	2.14	9.08	8.94	7.29	6.69	-37.75	1.37
COP3	-6.47	-2.04	-8.14	-11.77	-16.92	-10.09	-1.65
COP4	0.61	3.09	2.52	2.60	-6.32	-20.42	1.39
COP5	-4.69	-0.35	-1.08	1.34	-1.47	-35.54	0.12
COP6	8.20	12.99	11.03	16.75	12.67	-26.10	-1.24
COP7	-3.20	-4.06	-6.80	-11.04	-7.91	-21.00	2.16
COP8	-4.96	-2.82	-5.69	-7.04	-20.68	-5.17	0.74
COP9	5.98	6.52	6.68	5.96	3.64	-16.21	-0.15
COP10	7.67	9.52	9.49	-47.16	6.47	-43.35	1.03

## Liganden

Man benötigt ca. 15 Stunden reine Rechenzeit, um 2000 Konformationen des Liganden COP1 (63 Atome) mittels der geschilderten MD-Prozedur zu erzeugen (s. Durchführung).

*Tabelle 3.15.:* *rmsd*-Werte der Liganden nach Struktur-Optimierung in Bezug zu ihrer jeweiligen Konformation aus dem Komplex nach Energie-Minimierung 6.

	<i>rmsd</i> -Werte [Å]				
	1	2	3	4	5
COP1	0.64	0.73	0.60	0.80	1.38
COP2	1.06	1.30	1.38	1.29	1.71
COP3	0.91	1.20	1.15	1.11	1.23
COP4	0.83	1.61	1.34	0.95	1.75
COP5	0.86	0.75	0.87	1.06	1.39
COP6	0.88	0.87	1.00	1.16	1.43
COP7	0.81	1.42	1.51	1.12	1.99
COP8	0.68	1.24	1.16	0.92	1.40
COP9	0.64	0.70	0.72	0.85	1.28
COP10	0.82	0.94	1.05	0.94	1.34

Abbildung 3.10 zeigt, wie häufig Konformationen mit bestimmten Kraftfeld-Energien vorkommen. Demnach gibt es zwei große Energie-Bereiche bzw. zwei Gruppen von Konformationen, die lokalen Minima entsprechen und durch eine Lücke von gut 20 kcal/mol von einander getrennt sind. Konformationen mit niedriger Energie treten wesentlich häufiger auf als Konformationen mit hoher Energie. Die Häufigkeitsverteilung innerhalb der Gruppe der günstigen Konformationen hat ihr Maximum im mittleren Bereich, es ist weder deutlich zu den niedrigen noch zu den höheren Energien verschoben. Der niedrigste berechnete Wert beträgt 34.32 kcal/mol und tritt nur ein einziges Mal in dem Ensemble der 2000 Konformationen auf. Um abzuschätzen, wie viele Konformationen ein Ensemble umfassen muss, um mindestens eine Konformation zu enthalten, die dem globalen Minimum nahe genug kommt, werden die niedrigsten Werte unterschiedlich großer Mengen aus diesen 2000 Konformationen verglichen (s. Tab. 3.16)

*Tabelle 3.16.:* Abhängigkeit des Wertes für die niedrigste Kraftfeld-Energie von der Größe des betrachteten Ensembles.

Ensemble-Größe	2000	1000	500	333
Kraftfeld-	34.32	34.40	35.05	35.10
Energie [kcal/mol]		34.32	34.40	34.40
			34.47	34.47
			34.32	34.73
				34.32
				35.05

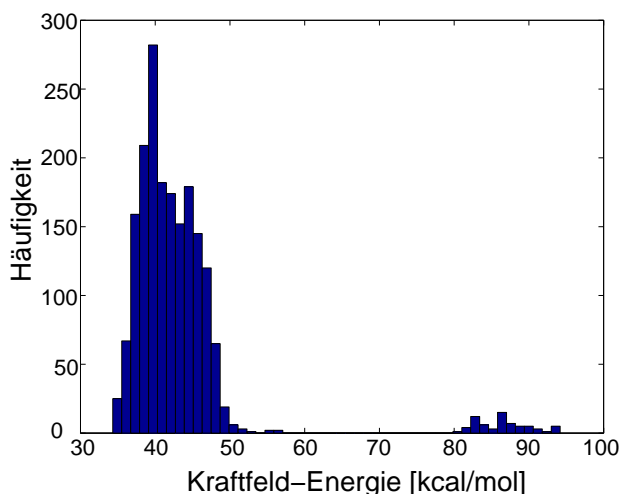


Abbildung 3.10: Häufigkeitsverteilung für die verschiedenen Konformationen des gelösten Liganden COP1, abhängig von der Kraftfeld-Energie.

Ein mögliches „Verfehlen“ des globalen Minimums um  $+0.2$  kcal/mol wird als noch akzeptabel betrachtet (s. Kap. 2.5); demnach ist eine Ensemble-Größe von jeweils 1000 Konformationen ausreichend. Für die übrigen neun Liganden werden also jeweils nur 1000 Konformationen erzeugt. Die Solvatationsenergien für COP1 sind mit  $6.7 \pm 0.60$  kcal/mol fast konstant, die Profile der Gesamtenergien werden also von den Kraftfeld-Beiträgen dominiert. Um die Konformation mit niedrigster Gesamtenergie zu bestimmen, genügt es folglich, die Solvatationsenergien nur für diejenigen Konformationen zu berechnen, die maximal 5 kcal/mol mehr als die Konformation mit minimaler Kraftfeld-Energie besitzen. Der Wert von 5 kcal/mol soll auch für die anderen Liganden sicherstellen, dass alle relevanten Konformationen erfasst werden, gleichzeitig wird damit die Zahl der durchzuführenden Solvatationsrechnungen etwa halbiert.

Für alle Liganden erhält man nach der MD-Simulation die schon für COP1 typische Häufigkeitsverteilung von Konformationen bzw. Energie-Inhalten (s. Abb. 3.11). Tabelle 3.17 listet jeweils die niedrigsten Gesamtenergien, sowie die mit dem Liganden COP1 als Referenz berechneten  $\Delta E_L$ -Beiträge auf. Es wird auch gezeigt, wie sich die niedrigste Gesamtenergie der Liganden aus Solvatations- und Kraftfeld-Beiträgen zusammensetzt. Die Werte liegen durchweg in vergleichbaren Größenordnungen. Lediglich bei COP5 fällt der hohe Beitrag der Kraftfeld-Energie auf und bei COP6 der niedrige Beitrag der Solvatationsenergie.

### Relative Bindestärke

Der Stamm-Ligand COP1 wird als Referenz zur Berechnung der relativen Bindestärken  $\Delta\Delta G$  verwendet. Tabelle 3.18 bzw. die Abbildungen 3.14(a) und 3.14(b) zeigen die Profile der relativen Bindestärken  $\Delta\Delta G$ , die abhängig von den jeweiligen Randbedingungen aus der Energie-Minimierung der Komplexe und der MD-Simulation der Liganden resultieren.

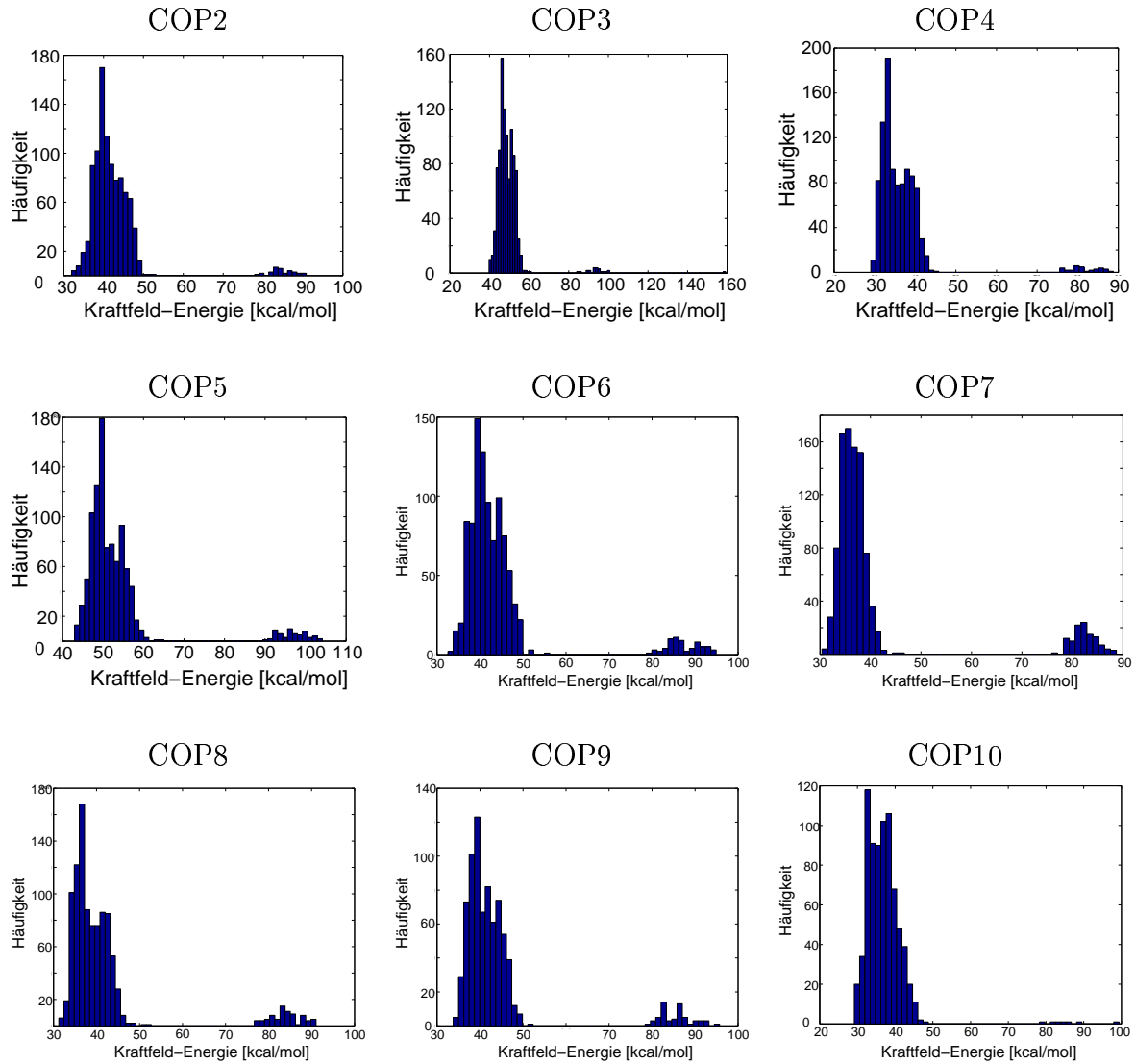


Abbildung 3.11.: Häufigkeitsverteilung für die verschiedenen Konformationen der gelösten Liganden, abhängig von der Kraftfeld-Energie.



*Tabelle 3.17.:* Energie-Werte für die nach MD-Simulation jeweils günstigste Konformation der einzelnen Liganden in kcal/mol.

	$E^{\text{Kraftfeld}}$	$E^{\text{Solvatation}}$	$E^{\text{Gesamt}}$	$\Delta E_{\text{L}}$
COP1	34.40	6.54	40.94	0.00
COP2	31.98	7.35	39.32	1.62
COP3	39.75	6.18	45.93	-4.98
COP4	29.57	5.25	34.82	6.13
COP5	43.01	6.78	49.79	-8.84
COP6	32.54	3.62	36.16	4.79
COP7	30.39	4.47	34.86	6.79
COP8	31.23	4.73	35.96	4.98
COP9	33.77	6.52	40.29	0.65
COP10	29.61	6.15	35.76	5.18

### 3.6.3. Diskussion

Um die bekannten relativen Bindestärken  $\Delta\Delta G$  für eine Serie von zehn Komplexen zu reproduzieren, sind die Gesamtenergien der Protein-Ligand-Komplexe bzw. der freien Liganden in wässriger Lösung zu berechnen. Die Genauigkeit der daraus zu berechnenden  $\Delta\Delta G$ -Werte hängt besonders davon ab, ob die Energie-Werte aller Mitglieder einer Serie (von Komplexen bzw. von freien Liganden) neben den für das Bindeverhalten charakteristischen Unterschieden, einheitliche systematische Fehler besitzen. Zusätzliche, *zufällige* Abweichungen verschlechtern die Korrelation der berechneten mit den experimentellen  $\Delta\Delta G$ -Werten. Im folgenden werden die Energie-Werte und die Werte der daraus abgeleiteten Größen interpretiert.

**Komplexe:** Tabelle 3.13 listet die Gesamtenergien bzw. die vom Protein-Anteil bereinigten Ligand- und Wechselwirkungsenergien aller Komplexe auf, die nach den verschiedenen Alternativen zur Energie-Minimierung resultieren. Aus den durchschnittlichen Energie-Werten der verschiedenen Serien minimierter Komplexe geht hervor, wie unterschiedlich sich die allgemeinen Konformationen je nach Wahl der Randbedingungen entwickeln: Die Mittelwerte über die Gesamtenergien von Komplexen mit verschiedenen Liganden haben zwar keine physikalische Bedeutung, veranschaulichen aber, dass die Energie der Komplexe während einer Minimierung etwa in dem Maße abgesenkt wird, wie viele Atome beweglich sind. Der Wert der Standard-Abweichung ist ein Indiz dafür, wie viel Spielraum die jeweiligen Randbedingungen der Struktur-Optimierung lassen. Je größer die Standard-Abweichung, desto mehr sind die Möglichkeiten zur Reduktion von Spannungen innerhalb des Komplexes eingeschränkt: Dann können nur wenige Komplexe der Serie Konformationen mit niedriger Gesamtenergie einnehmen, während andere in relativ dazu ungünstigen lokalen Minima gefangen bleiben. Bei den Minimierungsverfahren 4–6 sind die Standard-

Tabelle 3.18.: Relative Bindestärken in Abhängigkeit vom verwendeten Verfahren zur Erzeugung der Komplexe in kcal/mol.

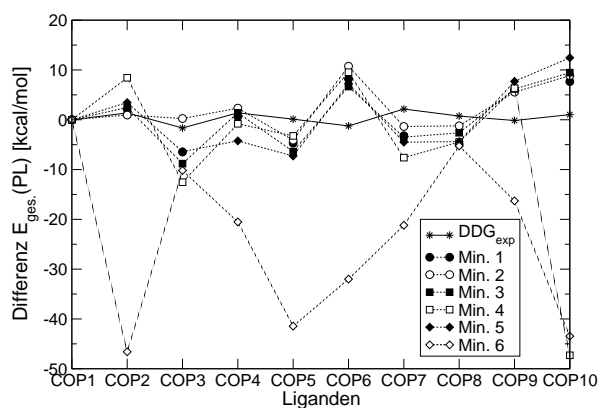
$$\Delta\Delta G = \Delta E_{PL} - \Delta E_L$$

$\Delta\Delta G$ basierend auf den Beiträgen der Gesamtenergie [kcal/mol] Minimierungsverfahren							
	1	2	3	4	5	6	exp
COP1	0.00	0.00	0.00	0.00	0.00	0.00	0.00
COP2	0.51	-0.67	0.84	6.81	1.84	-48.22	-1.37
COP3	-1.49	5.23	-3.83	-7.56	-1.41	-5.20	1.65
COP4	-5.53	-3.80	-4.73	-6.94	-10.37	-26.65	-1.39
COP5	4.15	4.75	2.46	5.60	1.56	-32.61	-0.115
COP6	3.41	5.96	1.88	4.76	2.46	-36.79	1.241
COP7	-9.99	-8.14	-10.29	-14.40	-11.29	-27.98	-2.163
COP8	-9.94	-6.20	-7.59	-9.44	-9.37	-10.24	-0.737
COP9	5.33	4.90	5.53	5.70	7.06	-16.94	0.147
COP10	2.48	3.68	4.28	-52.48	7.27	-48.64	-1.031

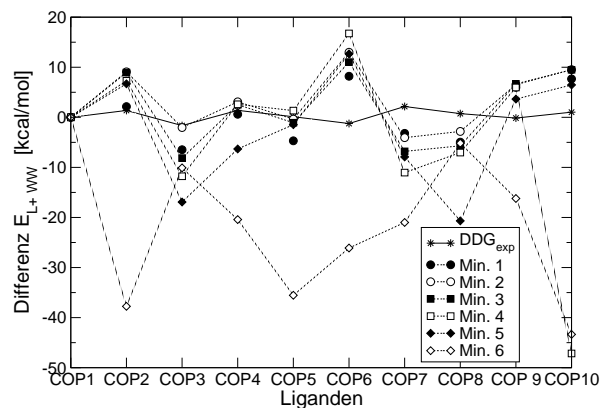
$\Delta\Delta G$ basierend auf den Beiträgen der Ligand- und Wechselwirkungsenergien [kcal/mol]							
	1	2	3	4	5	6	exp
COP1	0.00	0.00	0.00	0.00	0.00	0.00	0.00
COP2	0.52	7.46	7.32	5.67	5.07	-39.37	-1.37
COP3	-1.49	2.94	-3.16	-6.79	-11.94	-5.11	1.65
COP4	-5.52	-3.04	-3.61	-3.53	-12.45	-26.55	-1.39
COP5	4.15	8.49	7.76	10.18	7.37	-26.70	-0.115
COP6	3.41	8.20	6.24	11.96	7.88	-30.89	1.241
COP7	-9.99	-10.85	-13.59	-17.83	-14.70	-27.79	-2.163
COP8	-9.94	-7.80	-10.67	-12.02	-25.66	-10.15	-0.737
COP9	5.33	5.87	6.03	5.31	2.99	-16.86	0.147
COP10	2.49	4.34	4.31	-52.34	1.29	-48.53	-1.031

Abweichungen zwei- bis dreimal größer als bei den Verfahren 1–3, die ersten drei Verfahren scheinen also zuverlässigere Ergebnisse zu liefern. Um störende, Protein-bedingte Schwankungen zu eliminieren, wird von der Gesamtenergie der Komplexe wie schon zuvor (s. Kap. 3.4, Gl. 3.3) der Anteil des isolierten Proteins subtrahiert. Trotz Abzug des Protein-Anteils bleiben die Standard-Abweichungen groß; die Schwankungen der Komplex-Energien beruhen also auf den für die relative Bindestärke relevanten Beiträgen des Liganden und seiner Wechselwirkungen mit dem Protein.

**Komplex-Profil:** Aus den Gesamtenergien der Komplexe bzw. aus den Ligand- und Wechselwirkungsenergien werden die Beiträge  $\Delta E_{P-L}$  bzw.  $\Delta E_{L+WW}$  zur relativen Bindestärke  $\Delta\Delta G$  berechnet (s. Tab. 3.14 und Abb. 3.12(a) und 3.12(b)). Dabei wird COP1 als Referenz-Struktur verwendet, dessen natürliche Konformation im Komplex durch die KS belegt ist. Die Profile der Verfahren 1–5 verlaufen recht einheitlich, lediglich der  $\Delta E_{P-L}$ - bzw.  $\Delta E_{L+WW}$ -Wert für COP10 bei der Energie-Minimierung nach dem Verfahren 4 und die meisten der nach Verfahren 6 berechneten Werte weichen stark von den anderen ab. Trotz ihrer generell guten Übereinstimmungen sind die Unterschiede zwischen den Profilen relativ groß; der exakte  $\Delta E_{P-L}$ - bzw.  $\Delta E_{L+WW}$ -Beitrag eines Liganden ist also nicht nur vom Liganden selbst, sondern auch empfindlich vom jeweiligen Minimierungsverfahren abhängig. Die nach den Verfahren 1–5 resultierenden Profile werden teilweise gemittelt (s. Tab. 3.19). Mit Standard-Abweichungen der  $\overline{\Delta E_{P-L}}$ - bzw.  $\overline{\Delta E_{L+WW}}$ -Beiträge von bis zu 5.75 kcal/mol sind die Profile wesentlich ungenauer als für die Reproduktion von  $\Delta\Delta G$ -Werten erforderlich (vgl. Kap. 2.5).



(a)  $\Delta E_{P-L}$ -Profile der Komplexe aus den Gesamtenergien nach den verschiedenen Minimierungsverfahren im Vergleich mit dem erwarteten  $\Delta\Delta G$ -Profil



(b)  $\Delta E_{L+WW}$ -Profile der Ligand- und Wechselwirkungsenergien nach den verschiedenen Minimierungsverfahren im Vergleich mit dem erwarteten  $\Delta\Delta G$ -Profil

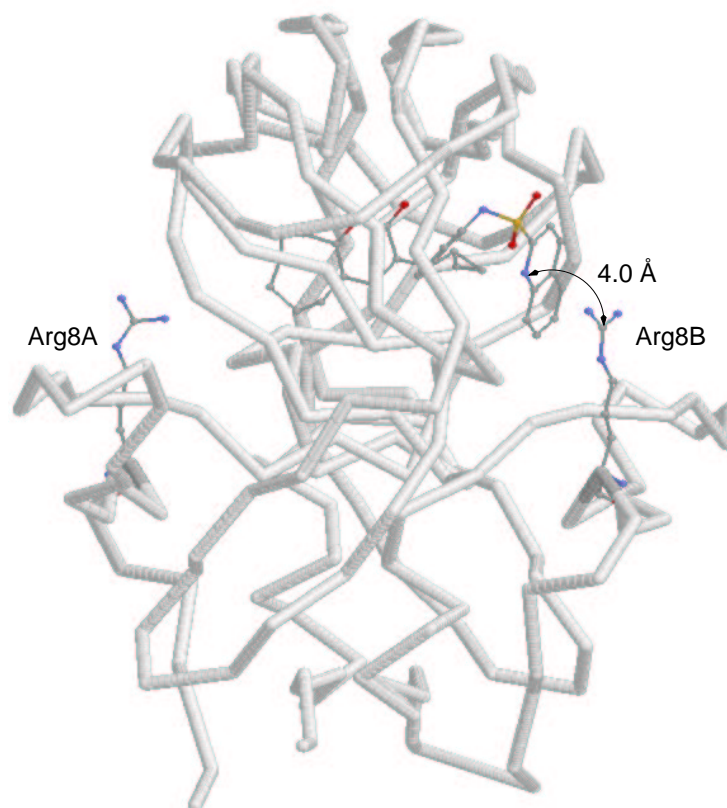
Abbildung 3.12.: Komplex-Beiträge zur relativen Bindestärke vor bzw. nach Bereinigung vom Protein-Anteil.

Tabelle 3.19.: Durchschnittliche Komplex-Beiträge nach den Verfahren 1–5

	$\overline{\Delta E}_{P-L}$	$\overline{\Delta E}_{L+WW}$
COP1	0.00	0.00
COP2	2.90±2.96	6.82±2.82
COP3	-6.79±4.66	-9.06±5.61
COP4	-0.14±2.56	0.41±3.51
COP5	-5.13±1.66	-1.04±2.03
COP6	8.48±1.67	10.27±5.75
COP7	-4.04±2.30	-6.60±3.14
COP8	-3.52±1.56	-8.23±7.12
COP9	6.35±0.81	5.75±1.22
COP10	9.60±2.03	8.28±1.48

Die verschiedenen Komplexe werden in etwa abhängig von der Ähnlichkeit des jeweiligen Liganden mit dem Stamm-Liganden COP1 bewertet. Komplexe, die über dieses Maß hinaus im Vergleich zum Referenz-Komplex ungünstig abschneiden, erhalten systematisch sehr niedrige relative Beiträge  $\Delta E_{P-L}$ . Besonders stark negative  $\Delta E_{P-L}$ -Beiträge weisen also darauf hin, dass einzelne Liganden während der Minimierung ein ungünstiges, lokales Minimum nicht verlassen können und in einer gespannten Konformation verharren. So bleibt der Ligand COP10 wohl durch die bei Alternative 4 stärkere Gewichtung der elektrostatischen Wechselwirkungen ( $\varepsilon = 2r$  statt  $\varepsilon = 4$  oder  $4r$ ) auf Grund von elektrostatischen Effekten in einer ungünstigen Konformation gefangen. Weiterhin schränken die Randbedingungen von Verfahren 6, also die elastische Anbindung aller Atome an ihre Start-Positionen, wie schon bei den CHD-Komplexen (s. Kap. 3.5) die Beweglichkeit der manipulierten Liganden übermäßig ein. Die Resultate dieser Serie werden von weiteren Betrachtungen ausgeschlossen.

Im Gegensatz dazu fallen die Komplex-Beiträge von COP6 durchweg als überraschend günstig auf. COP6 besitzt einen Chinolin-Substituenten, der bioisoster zu dem Naphthalin-Rest des Liganden COP5 ist. COP5 ist ebenfalls eng verwandt mit dem Phenyl-Ring des Stamm-Liganden COP1 und die Komplexe werden ähnlich bewertet (s. Tab. 3.12). Der unerwartet große Energie-Unterschied von 9–15 kcal/mol zwischen COP5 und COP6 beruht demnach auf dem Vorhandensein des N-Atoms im Chinolin-System, welchem mit -0.74 Coulomb eine stark negative Partial-Ladung zugeordnet ist. Dadurch wird bei COP6, im Gegensatz zu den anderen Liganden, eine anziehende Wechselwirkung zur unmittelbar benachbarten, positiv geladenen Guanidinium-Gruppe von Arg8 berechnet, die wahrscheinlich zu hoch ist (der Abstand zwischen Chinolin-N-Atom und Arginin-C $\zeta$ -Atom beträgt je nach Verfahren 4.0–4.6 Å, Beispiel s. Abb. 3.13).



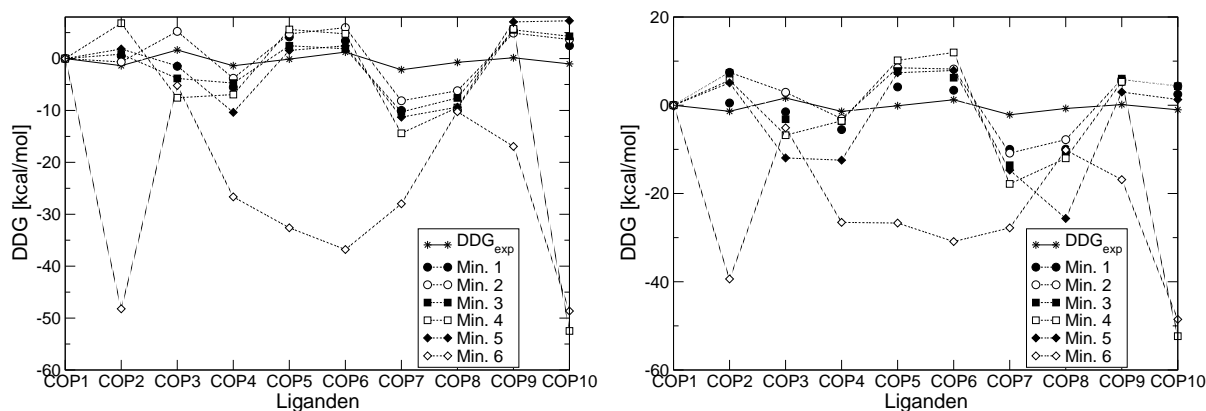
**Abbildung 3.13:** Seitenansicht der HIV-PR im Komplex mit COP6 nach Energie-Minimierung mit Alternative 2: Das Chinolin-Ringsystem liegt parallel zur kationischen Guanidinium-Gruppe von Arg8 aus Monomer B mit einem Abstand von nur 4.0 Å zwischen dem zentralen C<sub>γ</sub>-Atom des Arginins und des Chinolin-N-Atoms.

**Liganden:** Tabelle 3.17 zeigt die Gesamtenergien bzw. die  $\Delta E_L$ -Beiträge der freien Liganden in wässriger Lösung. Die berechneten Gesamtenergien variieren zwischen 35–50 kcal/mol. Diese Spannweite geht vor allem zurück auf die sehr unterschiedlichen Kraftfeld-Beiträge zwischen 32–43 kcal/mol, während die Solvatationsenergien mit 3.6–7.3 kcal/mol recht einheitlich (und weitgehend von der Atomzahl abhängig) sind. Die Liganden sind größtenteils unpolar. Günstige, elektrostatische Effekte können also nicht zur Solvatationsenergie beitragen; vielmehr dominiert der Entropie-Verlust des Wassers, daher sind die Solvatationsenergien durchweg positiv.

COP6 besitzt mit 3.62 kcal/mol den niedrigsten Beitrag, was wie bei den Komplexen auf der hohen Partial-Ladung von -0.74 Coulomb am Chinolin-N-Atom beruht. Bei den Kraftfeld-Beiträgen der Liganden besitzt COP5 mit einem Abstand von ca. 9 kcal/mol zu den anderen Werten (29.6–34.4 kcal/mol) die höchste moleküleigene Energie (43.01 kcal/mol). Dies ist in sofern überraschend, als der isostere Ligand COP6 um 10 kcal/mol günstiger bewertet wird. Um zu prüfen, ob die MD-Simulation bei der Suche nach der günstigsten Konformation versagt hat, wird die für COP6 ermittelte, beste Konformation auf den strukturell identischen Liganden COP5 übertragen. Die Kraftfeld-Energie von COP5 in der Minimum-Konformation von COP6 beträgt jedoch 45.20 kcal/mol und ist demnach um 2.19 kcal/mol ungünstiger als die zuvor bestimmte Konformation. Die MD-Simulation hat bei der Suche nach günstigen Konformationen also nicht versagt. Solange jedoch der

Grund für die Diskrepanz der Energie-Werte unklar ist, sollte auch der für COP5 berechnete  $\Delta E_L$ -Beitrag aus der Serie ausgeschlossen werden.

**Relative Bindestärken**  $\Delta\Delta G$ -Werte spiegeln als Differenz zwischen den  $\Delta E_L$ - und  $\Delta E_{P,L}$ -Beiträgen die energetischen Verhältnisse zwischen den gelösten und komplexierten Formen der Komponenten wider. Sie sind nur dann aussagekräftig, wenn die  $\Delta E_L$ - bzw.  $\Delta E_{P,L}$ -Beiträge zuverlässig sind, denn Fehler in den Differenz-Profilen der Ligand- oder der Komplex-Beiträge summieren sich bei der Berechnung der  $\Delta\Delta G$ -Werte. Wie Tabelle 3.18 und die Abbildungen 3.14(a) und 3.14(b) belegen, stimmen die berechneten  $\Delta\Delta G$ -Werte kaum mit den experimentell bestimmten  $\Delta\Delta G$ -Werten überein, was wohl vor allem auf den ungenauen Komplex-Beiträgen  $\Delta E_{P,L}$  bzw.  $\Delta E_{L+WW}$  beruht. Um die Ursachen für das Versagen der verwendeten Methode zu finden, werden die Daten auf Zusammenhänge zwischen den Differenz-Beiträgen der Komplexe bzw. der Liganden mit der relativen Bindestärke  $\Delta\Delta G_{\text{exp}}$  untersucht. Eine Korrelation der berechneten  $\Delta E_{P,L}$ - bzw.  $\Delta E_L$ -Beiträge mit der relativen Bindestärke  $\Delta\Delta G$  ist zwar nicht zwingend, kann aber Hinweise auf die Stimmigkeit der berechneten Ergebnisse liefern.



(a) Profil der relativen Bindestärken  $\Delta\Delta G$  berechnet aus den Beiträgen der freien Liganden  $\Delta E_L$  und der Gesamtenergie der Komplexe  $\Delta E_{P,L}$ .

(b) Profil der relativen Bindestärken  $\Delta\Delta G$  berechnet aus den Beiträgen der freien Liganden  $\Delta E_L$  und der Ligand- und Wechselwirkungsenergie der Komplexe  $\Delta E_{L+WW}$ .

Abbildung 3.14.:  $\Delta\Delta G$ -Profile ohne bzw. mit Bereinigung vom Protein-Anteil.

**Profil-Vergleiche** Der Vergleich der Energie-Profile der Liganden  $\Delta E_L$ , der Komplexe  $\Delta E_{P,L}$  bzw.  $\Delta E_{L+WW}$  und der erwarteten relativen Bindestärken  $\Delta\Delta G_x$  zeigt auffällige Parallelen (Abb. 3.16 enthält die für die Verfahren 1–5 gemittelten Profile der Komplexe).

**Komplex- und Ligand-Profil:** Demnach sind die  $\Delta E_{P,L}$ - bzw.  $\Delta E_{L+WW}$ -Profile erwartungsgemäß vor allem abhängig von der Art des Liganden. Damit ist auch die

stillschweigende Annahme belegt, dass die Liganden tatsächlich einen gemeinsamen Bindemodus aufweisen. Andernfalls wären deutlichere Unterschiede in den Energie-Profilen zu erwarten. Die Übereinstimmung der Ligand- und Komplex-Profile verdeutlicht auch, dass die Energie-Minimierungen in erster Näherung in den meisten Fällen erfolgreich waren. Nur die Komplex-Beiträge von COP7 und COP8 weichen stark von den entsprechenden  $\Delta E_L$ -Werten ab: Sie liegen zu niedrig, d. h. die Strukturen dieser beiden Komplexe sind offensichtlich systematisch nicht ausreichend optimiert. COP7 und COP8 besitzen innerhalb dieser Serie von Liganden mit Methyl-Gruppen die kleinsten Substituenten an der Sulfonamid-Gruppe. Sie können also nur in wesentlich geringerem Umfang als die anderen Liganden unpolare Kontakte zur Bindetasche der HIV-PR ausbilden. Ihre Komplexe werden vermutlich deshalb ungünstig bewertet, weil sich während einer Energie-Minimierung die Seitenketten des Proteins nur bedingt an einen kleinen Liganden anpassen und bestehende Lücken, die zu hohen Solvatationsenergien führen, schließen können. Hier ist ein Rearrangement der Seitenketten von zumindest Leu23, Pro81 und Val82 erforderlich. Für diesen Fall der verkleinerten Liganden wäre sicherlich eine MD-Simulation des Komplexes notwendig, um die benötigte Anpassungsfähigkeit des Protein-Anteils berücksichtigen bzw. ausnutzen zu können.

**Komplex-/ $\Delta\Delta G_{\text{exp}}$ -Profile:** Trotz der teilweise passenden Profile, korrelieren die nach den Verfahren 1–5 berechneten Komplex-Beiträge schlecht mit der experimentell bestimmten relativen Bindestärke  $\Delta\Delta G$ ; Tabelle 3.20 weist Korrelationskoeffizienten von -2.4–0.13 aus. Eliminiert man jedoch die Liganden COP6, COP7 und COP8 als Ausreißer aus der Reihe der Komplexe, ist die Korrelation mit den  $\Delta E_{P,L}$ -Werten für alle Verfahren (außer Verfahren 6) deutlich verbessert: Die Werte steigen auf signifikantere 0.29 – 0.71. Verwendet man die vom Protein-Anteil bereinigten Komplex-Beiträge  $\Delta E_{L+WW}$ , findet man (unter Ausschluss der Ausreißer) sogar noch deutlichere Zusammenhänge, die durch Korrelationskoeffizienten von 0.58 – 0.81 belegt werden. Demnach ist, übereinstimmend mit der Theorie, die Bindestärke umso geringer, je höher, also je ungünstiger die Energie eines Komplexes ist.

**Ligand-/ $\Delta\Delta G_{\text{exp}}$ -Profile:** Bei den Ligand-Beiträgen zeichnen sich die Zusammenhänge mit der Bindestärke noch deutlicher ab: Lässt man für einen Vergleich der  $\Delta\Delta G_{\text{exp}}$ - und  $\Delta E_L$ -Profile die Liganden COP5 und COP6 als Ausreißer außer acht, erhält man eine Korrelation von 91% für die übrigen 8 Werte-Paare (inklusive COP6 beträgt der Korrelationskoeffizient 0.68). Demnach ist, ebenfalls wie erwartet, die Bindung eines Liganden an die HIV-PR umso stärker, je höher die Gesamtenergie des Liganden in Lösung ist.

**$\Delta\Delta G$ -/ $\Delta\Delta G_{\text{exp}}$ -Profile:** Nachdem die Komplexe mit den Liganden COP6, COP7 und COP8 (bei Verfahren 4 auch COP10) als Ausreißer identifiziert sind, und bei den freien Liganden COP5 und COP6 suspekt erscheinen, lassen sich mit den verbleibenden 6 bzw. 5 berechneten  $\Delta\Delta G$ -Werten Korrelationskoeffizienten von -0.75 bis

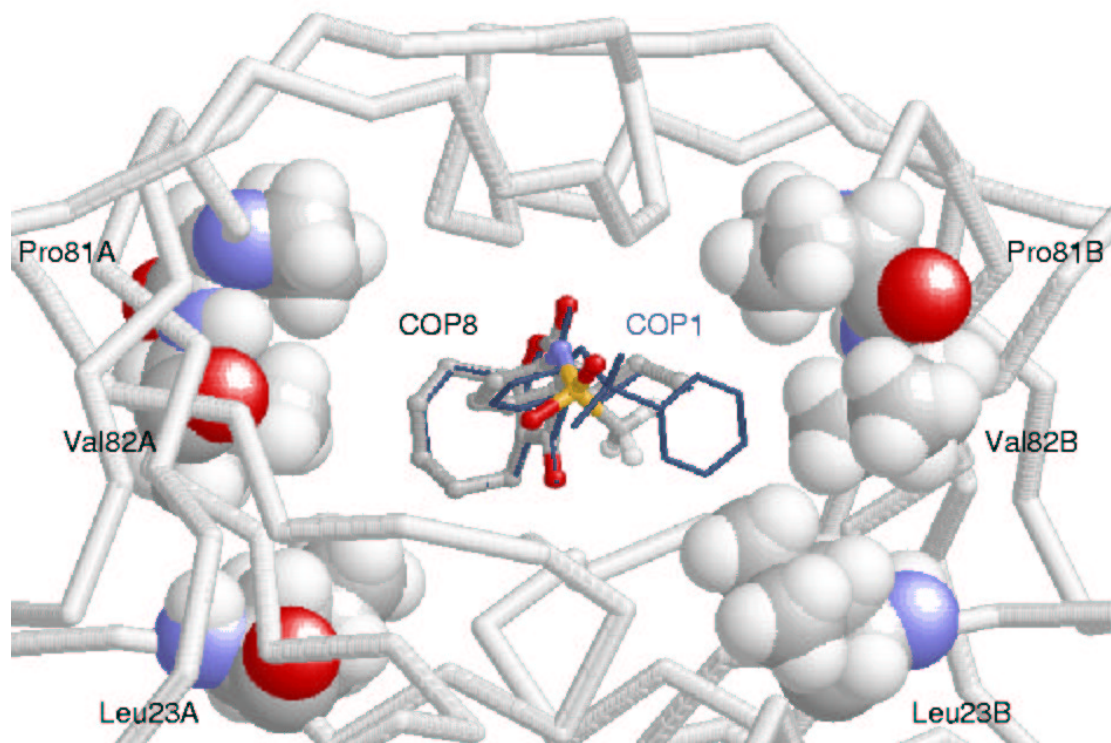


Abbildung 3.15.: Ausschnitt aus der Bindetasche der HIV-PR mit den Liganden COP1 (dunkel) und COP8 nach Minimierungsverfahren 1. Der Methyl-Rest von COP8 kann nicht annähernd die Lücke schließen, die zwischen den an den Phenyl-Ring von COP1 angepassten Seitenketten von Leu23, Pro81 und Val82 klafft.

immerhin 0.36 erreichen. Die beste Übereinstimmung der berechneten mit den gemessenen  $\Delta\Delta G$ -Werten erzielt man also mit Struktur-Optimierungen nach den Verfahren 3 und 5, sowie unter Abzug der jeweiligen Protein-Anteile. Die erreichbare Korrelation von 36 bzw. 33% ist jedoch zu ungenau, um sichere Vorhersagen machen zu können. Zudem müssen 4 von 10 Liganden als Ausreißer behandelt werden, was die allgemeine Anwendbarkeit und Zuverlässigkeit der Methode stark beeinträchtigt.

Unabhängig von dem nur schlecht wiedergegebenen Zusammenhang zwischen experimentellen und berechneten  $\Delta\Delta G$ -Werten, ist es überraschend, dass die Steigungen der Regressionsgeraden in allen Fällen größer als 1 sind. Demnach wird das beobachtbare  $\Delta\Delta G_{\text{exp}}$ -Profil überzeichnet: Die Werte der experimentell bestimmten relativen Bindestärken  $\Delta\Delta G$  (s. Tab. 3.18) liegen zwischen -2.2 bis + 1.6 kcal/mol, also in einem Intervall von knapp 4 kcal/mol. Die berechneten Werte erstrecken sich über einen gut fünfzehn mal größeren Bereich von ca. 60 kcal/mol. In einer anderen Arbeit, die sich mit der Berechnung von  $\Delta\Delta G$ -Werten beschäftigt, konnte (mit MD-Simulationen sowohl für die freien Liganden, als auch für die Komplexe) eine sehr gute Korrelation von 89.9% erreicht werden [13]. Die absoluten Werte jedoch waren



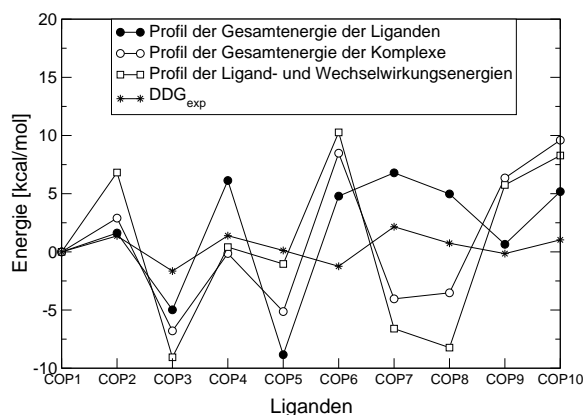


Abbildung 3.16: Durchschnittliche Energie-Profile der Komplexe im Vergleich mit dem Energie-Profil der freien Liganden und dem experimentellen  $\Delta\Delta G$ -Profil. Bei der Mittelung werden nur die Minimierungsverfahren 1–5 berücksichtigt.

Tabelle 3.20.: Korrelation der Komplex-Beiträge  $\Delta E_{P,L}$  bzw.  $\Delta E_{L+WW}$  und der berechneten  $\Delta\Delta G$ -Werte mit den experimentellen Bindestärken  $\Delta\Delta G_{\text{exp}}$ .

	Minimierungsverfahren						
	1	2	3	4	5	6	Ø
$\Delta E_{P,L}$ (alle)	-0.03	-0.26	0.13	-0.17	-0.04	-2.4	0.02
ohne Ausreißer	<b>0.58</b>	0.29	<b>0.63</b>	<b>0.71</b>	0.38	-0.58	-0.56
$\Delta E_{L+WW}$ (alle)	-0.03	-0.16	-0.02	0.27	-0.04	-0.29	0.04
ohne Ausreißer	<b>0.58</b>	<b>0.69</b>	<b>0.79</b>	<b>0.81</b>	<b>0.68</b>	-0.59	-0.75
$\Delta\Delta G$							
mit $\Delta E_{P,L}$ -Beiträgen (alle)	-0.52	-0.77	-0.37	-0.31	-0.44	-0.40	0.39
ohne Ausreißer	-0.15	-0.70	0.15	<b>0.34</b>	-0.12	-0.76	-0.05
mit $\Delta E_{L+WW}$ -Beiträgen (alle)	-0.51	-0.53	-0.37	-0.39	-0.28	-0.45	0.33
ohne Ausreißer	-0.15	-0.05	<b>0.36</b>	-0.48	<b>0.33</b>	-0.75	-0.26

ebenfalls deutlich überschätzt (Faktor 10); die Autoren führen dies auf den Umgang mit dem Cutoff für die van-der-Waals- und Coulomb-Wechselwirkungen innerhalb des verwendeten Programms zurück.

**Bewertung der alternativen Minimierungsverfahren:** Die nach den Verfahren 1–5 zur Energie-Minimierung berechneten Profile der Gesamtenergien der Komplexe verlaufen weitgehend konform; allen ist gemeinsam, dass nur der Protein-Anteil, nicht der jeweilige Ligand (mehr oder weniger weitgehend) fixiert wird. Die Struktur-Optimierung nach Verfahren 6 unter Anbindung *aller* Atome an ihre Ausgangspositionen, liefert davon völlig abweichende Ergebnisse. Wie der Vergleich der nach Alternative 1 bzw. 6 erzielten Ergebnisse zeigt, müssen in erster Linie die Liganden während der Energie-Minimierung frei beweglich sein, um einigermaßen plausible Konformationen der Liganden im Komplex mit dem Protein zu erzeugen. Der Einfluss der während der Minimierung verwendeten Dielektri-

zitätskonstanten scheint (bis auf eine Ausnahme, COP10 bei Verfahren 4) nicht wesentlich für das Ergebnis zu sein. Im Endeffekt schneiden die Alternativen 3 und 5 am vorteilhaftesten ab, wobei nicht sicher ist, ob dieser Befund für andere Liganden reproduzierbar wäre. In beiden Fällen ist jedoch sowohl der Ligand, als auch seine unmittelbare Nachbarschaft beweglich, was nach den vorigen Überlegungen sinnvoll erscheint.

**Fazit:** Die Gründe für die Schwierigkeiten bei der Reproduktion von  $\Delta\Delta G$ -Werten nach der gewählten Methode sind vielfältig.

Zunächst ist die Berechnung von  $\Delta\Delta G$ -Werten aus  $\Delta E_L$ - und  $\Delta E_{P,L}$ -Beiträgen prinzipiell sehr anfällig gegenüber Ungenauigkeiten der Beiträge, da Fehler sich (sofern sie nicht systematisch sind) addieren. Außerdem sind die zu erwartenden Differenzbeträge mit maximal ca. 4 kcal/mol sehr klein, sodass Fehler schnell relativ groß werden (vgl. Fehlerbetrachtung Kap. 2.5).

Dass die berechneten  $\Delta\Delta G$ -Werte sich so schlecht mit den experimentellen decken, beruht vor allem auf den unsicheren Beiträgen der Komplexe, denn die Komplex-Beiträge korrelieren schlechter mit den erwarteten  $\Delta\Delta G_{\text{exp}}$ -Werten als die Ligand-Beiträge. Dies belegt, zusätzlich zu der bekanntermaßen eingeschränkten Eignung von Energie-Minimierungen zur Auffindung globaler Minima, dass Energie-Minimierungen allein zur Optimierung der Komplex-Strukturen versagen. Die Möglichkeiten zur Optimierung der Protein- und Ligand-Wechselwirkungen, die die Flexibilität der HIV-PR birgt, können ohne ein aufwändigeres Abtasten des Konfigurationsraumes des Proteins nicht genutzt werden. Ebenso wie für die freien Liganden sollte also auch eine (möglicherweise auf bestimmte Regionen beschränkte) MD-Simulation der Komplexe durchgeführt werden. Davon könnten vor allem Komplexe mit verkleinerten Liganden (wie zum Beispiel COP7 und COP8) profitieren.

Da ein großer Teil der Wechselwirkungen zwischen den Komplex-Komponenten auf elektrostatischen Effekten beruht, ist auch die Zuordnung von Partial-Ladungen kritisch. Bei Ligand COP6 hat sich gezeigt, dass die verwendete *charge-templates*-Methode einem aromatischen N-Atom mit -0.74 Coulomb eine sehr hohe Partial-Ladung zuweist, während nach einem alternativen Verfahren (*Gasteiger*-Methode, Quanta) diesem N-Atom nur -0.234 Coulomb zugeordnet werden. Die *Gasteiger*-Methode hat jedoch andere Schwächen (stark alternierende Ladungsverteilungen bei Carboxylat- oder Ammonium-Gruppen), sodass im Einzelfall andere, quantenmechanische Verfahren zur Berechnung der atomaren Ladungen erforderlich sein können.

Der statische Ansatz ignoriert weitgehend entropische Effekte, lediglich Entropie-Änderungen des Wassers sind implizit über die Bestimmung der benetzbaren Oberflächen berücksichtigt. Wie experimentelle Arbeiten [129] gezeigt haben, ist die Balance zwischen enthalpischen und entropischen Effekten gerade bei der HIV-PR kritisch: Unter natürlichen Bedingungen kann die HIV-PR nicht als eine einzige Struktur, sondern nur als statistisches Ensemble vieler Konformationen behandelt werden, die entsprechend ihres Energie-Inhalts mehr oder weniger häufig eingenommen werden. Besonders die beweglichen Flügel der HIV-PR haben kaum strukturelle Stabilität. Die Bindung von Liganden bevorzugt energetisch selektiv solche Zustände des Proteins, bei denen die Kontaktfläche zum Liganden

stabilisiert wird. Das bedeutet, dass je nach Ligand verschiedene bzw. verschieden viele Seitenketten am Kontakt beteiligt sind, was sehr unterschiedliche Folgen für die entropischen Beiträge des Komplexes nach sich ziehen kann. Möglicherweise eignet sich gerade das Beispiel der HIV-PR nur schlecht zur Erprobung der Leistungsfähigkeit eines statischen Ansatzes.

Für qualitative Aussagen, z. B. über den Bindemodus, sind die erzeugten Strukturen sicher ausreichend, wie die durchaus vorhandenen Korrelationen zeigen; d. h. Energie-Minimierungen sind geeignet zur konservativen Struktur-Optimierung, die die grundsätzlichen Züge einer Konformation nicht antastet. Für quantitative Aussagen reicht die damit erreichbare Genauigkeit der Energie-Profile jedoch wie gezeigt nicht aus. MD-Simulationen der Komplexe (evtl. beschränkt auf Ausschnitte) wie bei den freien Liganden sind also doch erforderlich, wie andere, erfolgreiche Arbeiten [13] gezeigt haben. Damit ist die „schnelle“ Methode gescheitert.

Auf Energie-Minimierung ganz ohne Randbedingungen wird verzichtet, da im vorigen Beispiel der CHD-Komplexe danach die Protein-Beiträge und die Wechselwirkungsenergien stark schwanken und zu erwarten ist, dass sich abhängig von der jeweiligen Start-Konformation alle Komplexe stark auseinander entwickeln.

#### 3.6.4. Zusammenfassung und Ausblick

Die Berechnung von relativen Bindestärken nach dem statischen Ansatz ist nicht mit der erforderlichen Zuverlässigkeit möglich. Auf die aufwändige MD-Simulation von Komplexen oder alternative Verfahren zum Absuchen des Konformationsraumes kann also nicht verzichtet werden. Die systematischen Fehler bei den nach dem statischen Ansatz erzeugten Serien von Komplexen sind zwar jeweils sehr ähnlich, aber nicht identisch. Die gegenseitige Anpassung der Komplex-Komponenten aneinander setzt der Struktur-Optimierung allein durch Energie-Minimierung Grenzen.

Diese Studie zeigt aber auch, dass im Fall der COP-Derivate das Editieren des Stammliganden zur Erzeugung der anderen Komplex-Strukturen ausreichend und eigenes Docking nicht erforderlich ist.

Außerdem zeichnen sich die beiden Komponenten des statischen Ansatzes, die  $\Delta E_L$ - und  $\Delta E_{P-L}$ -Beiträge, durch signifikante Korrelationen mit der Bindestärke aus. So hat sich gezeigt, dass die relative Bindestärke der COP-Derivate an die unpolare Bindetasche der HIV-PR unerwartet gut korreliert mit dem Profil der Gesamtenergie der freien Liganden. Um die Eignung verwandter Moleküle als Liganden für die HIV-PR grob abzuschätzen, kann die Bestimmung des  $\Delta E_L$ -Profils bereits ausreichen.

Man muss daraus schließen, dass die Brauchbarkeit von Energie-Minimierungen allein zur Struktur-Optimierung stark abhängig von der Start-Konformation ist, dass man also mehr Aufwand in die Erzeugung alternativer Start-Strukturen investieren muss, um aus einer Menge von Konformationen die besten auszuwählen. Außerdem ist vor allem die Beweglichkeit des Liganden kritisch für ein passables Profil von Komplex-Beiträgen.

Die verwendeten Methoden der Energie-Minimierung und Gesamtenergie-Berechnung eignen sich in Kombination also vor allem zur Beurteilung eines Ensembles von verschiedenen Ligand-Konformationen im Komplex mit einem möglichst starren Protein. Daher sollen sie zur Vorhersage von anderweitig nur schwer beobachtbaren Enzym-Substrat-Komplexen eingesetzt werden.

# 4. Tyrosin-Hydroxylase und Phenylalanin-Hydroxylase

## 4.1. Allgemeines

Die in Kap. 2 beschriebenen Methoden zur Struktur-Optimierung und Energie-Bewertung sollen nun zusammen mit den bei der Modellierung von HIV-PR-Komplexen gewonnenen Erkenntnissen verwendet werden, um Komplex-Strukturen eines strukturell gut definierten, d. h. wenig flexiblen Proteins vorherzusagen. Mit der Tyrosin-Hydroxylase (TH) wird ein Enzym ausgewählt, das erstens eine relativ starre katalytische Einheit besitzt und zweitens als Schlüssel-Enzym in der Biosynthese von Catecholaminen gut charakterisiert ist. Im Rahmen dieser Arbeit sollen Einzelheiten zur Position von Liganden innerhalb des aktiven Zentrums untersucht werden.

Die TH produziert Dopamin (DA), einen wichtigen Neurotransmitter [74]. Ein Mangel daran in bestimmten Hirn-Regionen hat die von der Parkinsonschen Erkrankung bekannten Symptome zur Folge [45]. Vor diesem Hintergrund besteht die medizinische Absicht nicht darin, das körpereigene Enzym durch Inhibitoren zu blockieren (wie z. B. im Fall der HIV-PR), sondern es detailliert auf seine Funktionsweise hin zu untersuchen. Zu diesem Zweck muss man den Komplex eines Proteins mit seinem Substrat betrachten. Da dieser Komplex jedoch naturgemäß schnell abreagiert, ist er experimentell kaum fassbar. Wohl aber sind solche Komplexe modellierbar und somit eine wertvolle Ergänzung der experimentellen Daten. Sie zeigen die Konstellation der Reaktionspartner unmittelbar vor der Reaktion; die Reaktion selbst kann mit Molecular-Modelling-Methoden allerdings nicht verfolgt werden. Ein modellierter Substrat-Komplex könnte aber als Ausgangspunkt für weitergehende, quantenmechanische Rechnungen dienen.

Die Konstruktion von Komplexen erfolgt nun in zwei Stufen, da der Bindemodus der Liganden teilweise unbekannt ist. Zunächst werden die Liganden mit FLEXX in der Bindetasche platziert, anschließend werden die Strukturen durch Energie-Minimierung optimiert und neu bewertet. Neben der gewünschten *Vorhersage* von Komplex-Strukturen muss gleichzeitig sichergestellt werden, dass dieses Verfahren bekannte Komplexe tatsächlich korrekt *reproduziert*. Zur Validierung des Verfahrens wird das zur TH homologe, strukturell aber umfassender charakterisierte Enzym PAH herangezogen. Für die als Leber-Enzym leicht verfügbare Phenylalanin-Hydroxylase (PAH) existiert auch eine große Menge ergänzender experimenteller Daten zur Enzym-Funktion.

### 4.1.1. Die Enzyme PAH und TH

Säugetiere besitzen zur Hydroxylierung der aromatischen AS Phe, Tyr und Trp jeweils ein eigenes Enzym (s. Abb. 4.1). Phenylalanin-Hydroxylase (PAH, EC 1.14.16.1), Tyrosin-Hydroxylase (TH, EC 1.14.16.2) und Tryptophan-Hydroxylase (TPH, EC 1.14.16.4) sind, der Ähnlichkeit der Reaktionen entsprechend, auch strukturell eng miteinander verwandt. Sie alle fügen unter Verbrauch von  $O_2$  mit Hilfe eines Cofaktors OH-Gruppen in die aromatische Struktur der jeweiligen AS ein (s. Abb. 4.1). Der Cofaktor Tetrahydrobiopterin ( $BH_4$ ) wird intrazellulär wieder aufbereitet [74, 43].

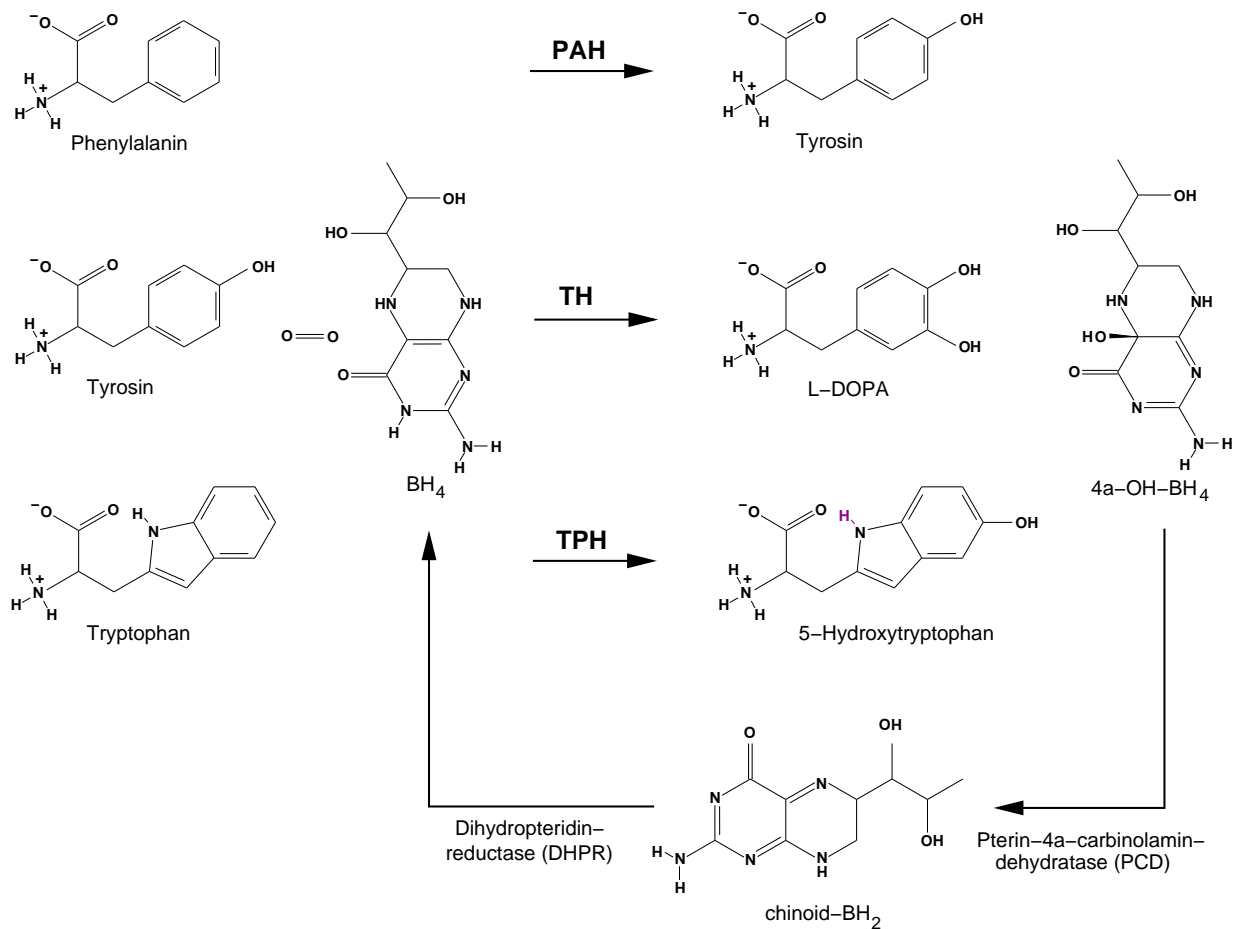
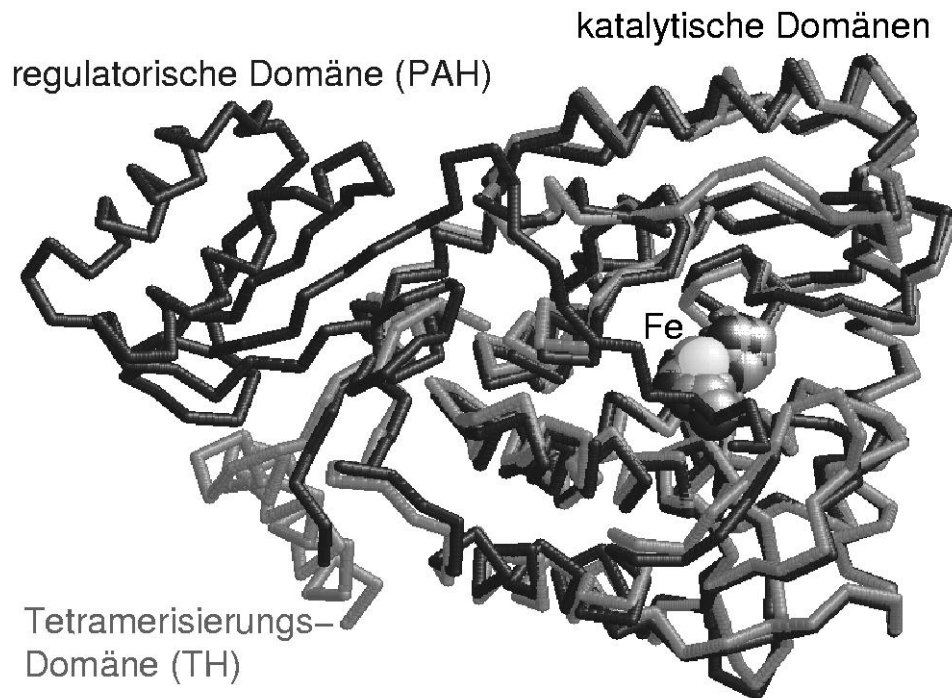


Abbildung 4.1.: Die enzymatischen Reaktionen von PAH, TH und TPH gleichen sich darin, dass alle drei eine aromatische AS unter Verbrauch von  $O_2$  und dem Cosubstrat  $BH_4$  hydroxylieren. Das ebenfalls hydroxylierte  $BH_4$  wird innerhalb der Zelle erneuert.

Die Enzyme besitzen einen gemeinsamen Aufbau aus drei Untereinheiten (s. Abb. 4.2):

- *N-terminale, regulatorische Domäne* zur Steuerung der Aktivität
- *zentrale, katalytische Domäne* mit gebundenem Eisen-Atom
- *C-terminale Kontakt-Region* zur Dimerisierung (PAH) bzw. Tetramerisierung (TH) [43].



*Abbildung 4.2.:* Die Überlagerung der KS 2toh und 2phm veranschaulicht den dreiteiligen Aufbau von TH und PAH. Im Boden der Bindetasche befindet sich das essentielle Eisen-Atom (Fe), das dort jeweils von zwei Histidin- und einer Glutamat-Seitenkette (raumfüllend dargestellt) fixiert wird.

Von den drei Hydroxylasen werden in dieser Arbeit nur die TH und PAH einander gegenübergestellt, da diese beiden einander besonders ähnlich sind (s. Abb. 4.3) und nur von diesen KS zur Verfügung stehen.<sup>1</sup> Ihre katalytischen Domänen gleichen sich sehr; sie besitzen 65% Sequenz-Identität und 80% Sequenz-Homologie [45]. Entsprechend lassen sich die C<sub>α</sub>-Atome der katalytischen Domänen (d. h. die Sequenz-Abschnitte 117–410 (PAH) bzw. 163–456 (TH)) mit einer durchschnittlichen Abweichung von nur 0.5 Å zur Deckung bringen. Dagegen unterscheiden sich die regulatorischen Domänen – gemäß ihrer jeweiligen Funktion – sowohl in Sequenz und Länge, als auch in ihrer Reaktion auf Signal-Stoffe deutlich von einander [74].

<sup>1</sup>Eine KS zur TPH wurde erst kürzlich veröffentlicht [136].

1pah	117	----- <u>TVPWFPRTIQELDRFANQILSYGAELDADHPGFKDPVYRARRKQFADIAYN</u> <u>YR</u>
2toh	156	<u>VRSAREDKVPWFPRKVSELDKCHHLVTKFDPDLDLHPGFSDQVYRQRRLIAEIAFQYK</u>
1pah	170	<u>HGQPIPRVEYMEEEEKKTWGTVFKTLKSLYKTHACYEYNHIFP</u> <u>LLEKYCGFHEDNIPQLED</u>
2toh	216	<u>HGEPIPHVEYTAEEIATWKEVYVTLKGLYATHACREHLEGFQLLERYCGYREDSIPQLED</u>
1pah	230	<u>VSQFLQTCTGFRLRPVAGLLSSRDFLGGLAFRVFHCTQYIRHGSKPMYTP</u> <u>EPDICHHELLG</u>
2toh	276	<u>VSRFLKERTGFQLRPVAGLLSARDFLASLAFRVFQCTQYIRHASSPMHS</u> <u>PEPDCHELLG</u>
1pah	290	<u>HVPLFSDRSFAQFSQEIFGLASLGAPDEYIEKLATIIYWFTVEFGLCKQ</u> <u>GDSIKAYGAGLLS</u>
2toh	336	<u>HVPMLADRTFAQFSQDIGLASLGASDEEIEKLSTVYWFTVEFGLCKQ</u> <u>NGELKAYGAGLLS</u>
1pah	410	<u>FSVRYDPYTQR</u> <u>IEVL</u>
2toh	456	<u>FSVKFDPYTLA</u> <u>IDVL</u>

Abbildung 4.3.: Das Alignment der katalytischen Domäne zeigt die weitgehende Übereinstimmung zwischen den Proteinen in diesem Sequenz-Abschnitt (identische AS sind unterstrichen).

Das *zweiwertige Eisen-Atom* wird von einer konservierten Triade aus zwei Histidin- und einer Glutamat-Seitenkette in der Bindetasche jeder katalytischen Domäne fixiert [90, 74]. Es ist für die Aktivität der Hydroxylasen essentiell.<sup>2</sup> In den meisten KS [35, 34, 33, 4, 52] ist das Eisen-Atom annähernd oktaedrisch durch die Ligand-Atome der Triade und weiterer Ligand- bzw. Wasser-Moleküle umgeben (s. Abb. 4.4). Es zeigt sich, dass das Eisen-fixierende Glutamat 376 der TH ein andere andere Konformation besitzt, als das entsprechende Glutamat 330 in den meisten KS der PAH. Dies ist möglicherweise ein wichtiger Unterschied in den aktiven Zentren der TH und PAH, da die exakte Position der CO<sub>2</sub><sup>-</sup>-Gruppe die Geometrie der übrigen Eisen-Liganden wesentlich mitbestimmt.<sup>3</sup> Bei der TH liegt ein O-Atom der Glutamat-Gruppe dicht am Oktaeder-Platz *trans* zu His331 und drängt das dort gebundene Wasser-Molekül von der durch das Eisen-Atom und den His331-Liganden gegebenen Achse weg.

## Vorkommen und Funktion

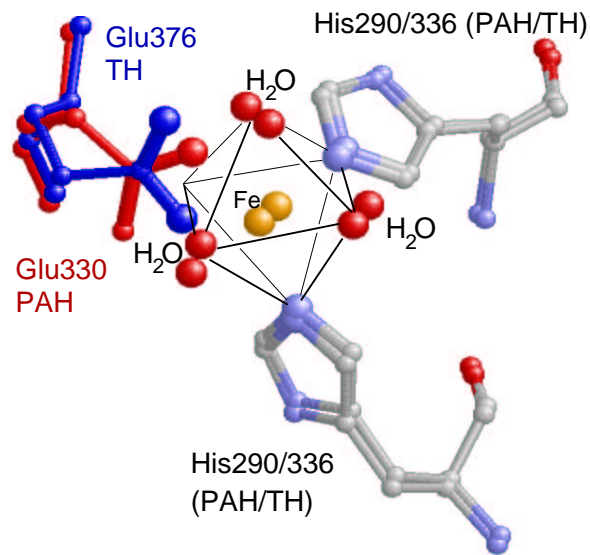
Das Leber-Enzym PAH führt den ersten Schritt in der *Entgiftung* (d. h. beim Abbau) von überschüssigem Phenylalanin (Phe) zu CO<sub>2</sub> und Wasser aus. Dabei entsteht das Vorläufer-Molekül der Catecholamine, die AS Tyrosin (Tyr). Eine hohe Konzentration von Phe im Blut (Hyperphenylalaninämie) stört die normale Entwicklung des Gehirns während der

<sup>2</sup>Ein Austausch gegen andere zweiwertige Kationen oder auch nur die Oxidation des Eisens führt zur totalen Inaktivierung des Enzyms [56].

<sup>3</sup>Da in *einer* KS der PAH (2phm) auch das sonst nur bei den KS der TH (1toh und 2toh) vorkommende Glu-Rotamer auftaucht, ist fraglich, ob es sich bei diesen Rotameren um feste Eigenschaften der katalytischen Untereinheiten handelt. Möglicherweise sind es alternative Konformationen, die bei beiden Enzymen vorkommen können.



Abbildung 4.4: Ausschnitt aus den überlagerten KS 1toh (TH) und 1pah (PAH). Eine sogenannte Triade aus zwei Histidin-Ringen und einer Glutamat-Gruppe fixiert das Eisen-Atom in den Bindetaschen. Zusammen mit den im Kristall gebundenen Wasser-Molekülen (hier nur als O-Atome sichtbar) erkennt man eine fast regelmäßige oktaedrische Umgebung des Eisens in der ersten Liganden-Sphäre. Die Rotamere der Glutamat-Seitenkette weichen voneinander ab.



frühen Kindheit. Ein eventueller Mangel an PAH-Aktivität wird anhand der Ausscheidung des Phe-Abbauproduktes Phenylpyruvat im Urin (Phenylketonurie) diagnostiziert [45].

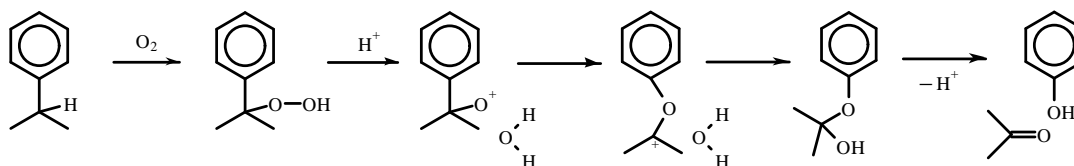
Die TH kommt in bestimmten Regionen des Gehirns und im Nebennierenmark vor und setzt Tyr zu L-DOPA um. Damit katalysiert sie den Schlüssel-Schritt in der *Catecholamin-Biosynthese*. Catecholamine sind wichtige Botenstoffe: Das im Gehirn entstehende Dopamin (DA) ist wichtig für die Koordination von Bewegungen; die unter anderem von den Nebennieren freigesetzten Folgeprodukte Adrenalin und Noradrenalin sind ebenfalls wichtige Neurotransmitter.

Ein Mangel an DA in bestimmten Hirn-Regionen (*substantia nigra*) ist verknüpft mit den Symptomen der Parkinsonschen Erkrankung [45]. Statt DA findet man zu Beginn der Erkrankung in der Hirn-Flüssigkeit von Parkinson-Patienten neurotoxische Abkömmlinge von DA. Durch (unkontrollierte) Kondensation von „ordnungsgemäß“ produziertem DA mit Formaldehyd und anschließende Methylierung am N-Atom entstehen in den dopaminergen Neuronen selbst Dihydroxy-tetrahydroisochinolin-Derivate (*dihydroxy-tetrahydroisochinoline*; DTIQ [98]). In den späteren Stadien der Krankheit gehen die DA-produzierenden Zellen, geschädigt durch oxidativen Stress innerhalb der Zellen (Produktion von Sauerstoff-Radikalen), zugrunde [122, 98, 57, 46]. Auch ständiger Kontakt zu Rotenon, einem gängigen Pestizid (mit Catechol-Substruktur, s. Abb. 4.8), scheint Morbus Parkinson-Symptome zu auslösen zu können [39].

## Katalytische Reaktion

Aromatische Ringe sind reaktionsträge. Neue Gruppen lassen sich im allgemeinen nur durch *elektrophile aromatische Substitution* am Ring einführen<sup>4</sup>. Auch O<sub>2</sub> ist im Grundzustand wenig reaktiv [82], er wird jedoch aktiviert durch das essentielle Eisen-Atom<sup>5</sup>. Gleichzeitig mit der *aromatischen AS* wird auch ein *Cosubstrat*, Tetrahydrobiopterin (BH<sub>4</sub>) hydroxyliert, denn dieses übernimmt das überzählige O<sub>2</sub>-Atom und ermöglicht damit dem anderen den elektrophilen Angriff [92]. Damit die Reaktion ablaufen kann, müssen alle Substrate (BH<sub>4</sub>, O<sub>2</sub>, AS) am Enzym gebunden sein [43]. Erst dann kann (im geschwindigkeitsbestimmenden Schritt) das O<sub>2</sub>-Molekül reduktiv durch BH<sub>4</sub> aktiviert werden (*initiale Hydroxylierung*), BH<sub>4</sub> wird dabei zum 4a-Hydroxytetrahydropterin (4a-OH-BH<sub>4</sub>) hydroxyliert [47]. Außerdem verbleibt ein hoch reaktives Sauerstoff-Atom der Oxidationsstufe ±0 in der Nähe des Eisen-Atoms. Dieses Intermediat konnte bislang nicht experimentell abgefangen und charakterisiert werden, doch ist davon auszugehen, dass das O-Atom in Analogie zur klassischen Synthese den Aromaten schnell und elektrophil angreift [21]. Die beiden Produkte, OH-BH<sub>4</sub> und die oxygenierte AS, werden schlagartig freigesetzt [43]. Durch radioaktive Markierung an der zu hydroxylierenden Position des Aromaten [43] kann man verfolgen, über welchen Weg sich das primäre Addukt aus Aromat und O-Atom umlagert (NIH-Shift, s. Abb. 4.5): Der ursprüngliche Substituent wird vom O-Atom verdrängt. Dabei übernimmt er das bindende Elektronen-Paar und wandert als anionisches Fragment zum benachbarten C-Atom mit Elektronen-Lücke [135]. Der schnelle Ablauf der katalytischen Reaktion deutet darauf hin, dass die Substrate sowohl dem essentiellen Eisen-Atom, als auch einander eng benachbart sind.

<sup>4</sup>Bei der klassischen Phenol-Synthese nach Hock [12] entsteht aus Cumol und Luft-Sauerstoff erst Cumolhydroperoxid, welches unter saurer Katalyse in Wasser und ein Intermediat mit Elektronen-Mangel am O-Atom zerfällt. Das O-Atom greift die hohe π-Elektronen-Dichte des Rings an und drängt sich zwischen die bestehende C-C-Bindung, sodass sich der Elektronen-Mangel zum benachbarten C-Atom verschiebt. Nachdem sich ein Wasser-Molekül wieder angelagert hat, zerfällt das Halbacetal in Phenol und Aceton.



<sup>5</sup>Üblicherweise liegt O<sub>2</sub> als Triplett vor, d. h. es hat im Grundzustand zwei ungepaarte Elektronen mit gleich gerichtetem Spin. Im Gegensatz dazu befinden sich organische Verbindungen normalerweise in einem Singulett-Zustand, d. h. alle ihre Elektronen sind mit antiparallelen Spins gepaart. Organische Moleküle sind an der Luft metastabil, weil ihre (eigentlich stark exotherme) Verbrennung durch Triplett-O<sub>2</sub> „Spin-verboten“, also kinetisch gehemmt ist. Eine Komplexbildung von O<sub>2</sub> mit Übergangsmetallen kann den Spin-Zustand ändern, sodass Reaktionen mit organischen Verbindungen „Spin-erlaubt“ sind [74, 82].

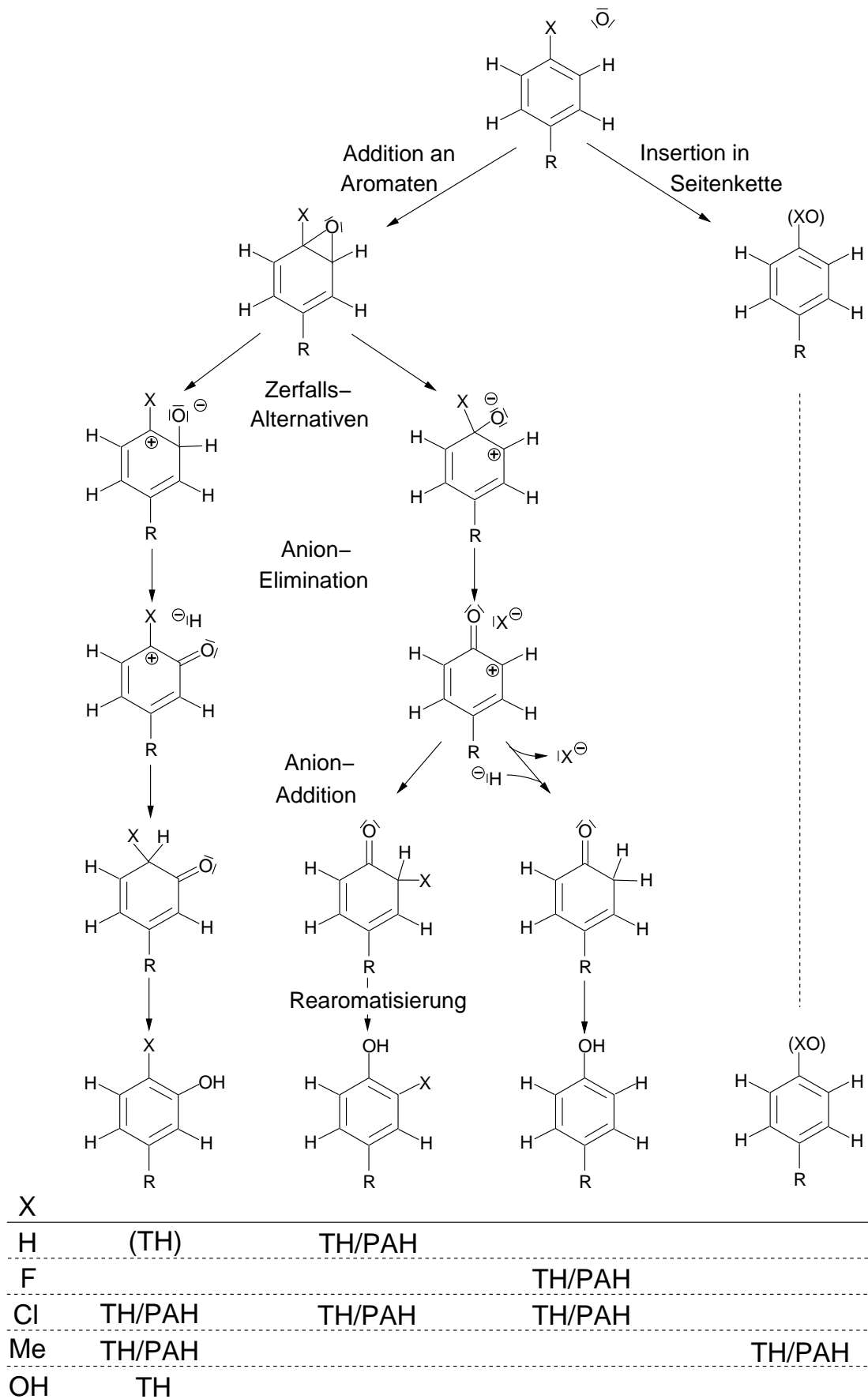
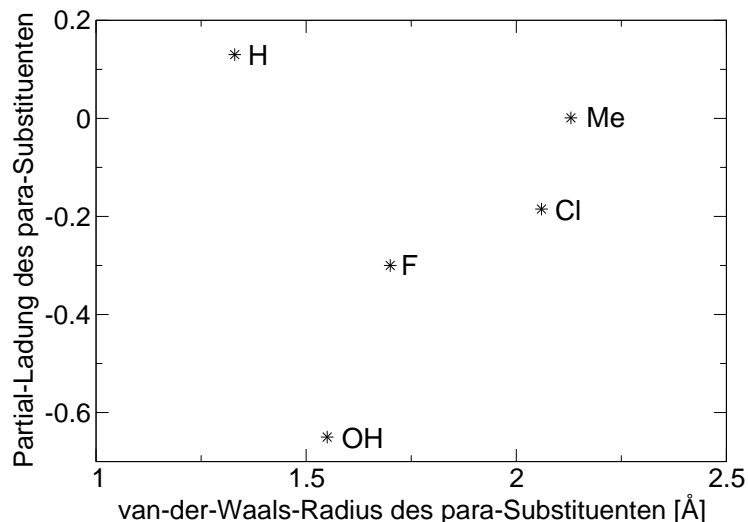


Abbildung 4.5.: NIH-Shift &amp; Verteilung der Produkte

### Substrat-Spezifitäten

PAH und TH können nicht nur ihre speziellen AS – Phe bzw. Tyr – sondern auch einige andere, verwandte Moleküle verarbeiten (wenn auch nur mit geringerer Effizienz). Ist nach der initialen Hydroxylierung erst einmal ein reaktives Intermediat erzeugt worden, kann es auf verschiedene *AS-Substrate*, auch mit unnatürlichen Substituenten in 4-Position (wie z. B. 4-F-Phe, 4-Cl-Phe, 4-Me-Phe, s. Abb. 4.6, links unten) übertragen werden [43]. Abhängig vom Substituenten entstehen verschiedene Produkt-Gemische (s. Abb. 4.5). Dies beruht auf den speziellen Eigenschaften des jeweiligen Substituenten in Bezug auf die Möglichkeiten für einen Einschub des O-Atoms in eine Bindung<sup>6</sup> bzw. in Bezug auf seine Stabilität als freies Anion<sup>7</sup>, sowie auf der Richtung, in der das primäre Addukt zerfällt (s. Abb. 4.5, unten).

Abbildung 4.6: Native Substrate und einige Substrat-Analoga: Vergleich der Van-der-Waals-Radien (nach CHARMM-Parametern) und Partial-Ladungen (berechnet nach der *charge-templates*-Methode von Quanta) der Substituenten in para-Position.



Bei diesen sehr ähnlichen AS überrascht es, dass die TH aus dem Vorläufer ihres eigenen Substrates (Phe) das Substrat (Tyr) selbst erzeugen kann, die PAH dagegen neben all den anderen Derivaten einzig ihr Produkt Tyr nicht weiter umsetzt<sup>8</sup>. Schon die isolierte katalytische Untereinheit zeigt diese strenge Selektivität; sie wird demnach nicht von der regulatorischen Domäne verursacht [23].

<sup>6</sup>Bei 4-Methyl-Phenylalanin konkurrieren die Methyl-Gruppe und der Ring um das O-Atom: Statt den Aromaten zu hydroxylieren, kann sich das O-Atom auch in eine C-H-Bindung der Methyl-Gruppe einschleichen, sodass eine benzyliche OH-Gruppe entsteht.

<sup>7</sup>Methyl-Anionen oder Hydrid-Ionen sind als Anion wenig stabil und binden sofort wieder an den Ring. Im Gegensatz dazu sind Halogene, besonders Fluor, als Anionen so stabil, dass sie sich nach Verdrängung durch das O-Atom nicht unbedingt wieder am Aromaten anlagern. Dann ist ein Hydrid-Ion als Ersatz zur Vervollständigung der Reaktion notwendig. Dieses Hydrid-Ion stammt vermutlich von einem weiteren Cosubstrat-Molekül; so erklärt sich der erhöhte Verbrauch an Cosubstrat bei halogenierten Phenylalaninen. Mitunter dirigiert ein Halogen-Atom den Angriff eines O-Atoms auch an die Nachbar-Position, sodass neben 4-OH-3-Cl-Phenylalanin auch 4-Cl-3-OH-Phenylalanin entsteht.

<sup>8</sup>Dieses Merkmal der PAH ist physiologisch sehr vorteilhaft; denn ansonsten würde die Leber mit L-DOPA unkontrolliert Transmitter- bzw. Hormon-Vorstufen und vor allem einen starken Inhibitor in großen Mengen produzieren können.

Platzmangel innerhalb der Bindetasche scheidet als Ursache dafür aus, da die zuvor genannten Substituenten in 4-Position durchweg größer sind als die OH-Gruppe (s. Abb. 4.6). Prinzipiell sollte eine zweite Hydroxylierung des Rings sogar leichter möglich sein, denn die schon vorhandene OH-Gruppe macht den Aromaten für eine weitere Substitution besonders empfänglich. Von den genannten AS-Derivaten zeigen Phe und Tyr in der Partialladung des Substituenten in 4-Position die größte Differenz (s. Abb. 4.6). Demnach könnte das Unvermögen der PAH auf unterschiedlichen elektrostatischen Wechselwirkungen mit den beiden AS beruhen.

## Inhibition

Die Hemmung der TH-Aktivität durch ihr eigenes Produkt L-DOPA bzw. seine weiteren Abkömmlinge Dopamin, Noradrenalin und Adrenalin verläuft in *zwei Stufen*: Die Catecholamine konkurrieren mit dem Cosubstrat  $\text{BH}_4$  um die Bindung am aktiven Zentrum [2, 113, 45]. Trotz der großen Stabilität von Catecholamin-Komplexen ist diese Blockade aber zunächst *reversibel*. In Gegenwart von Catecholaminen wird das zweiwertige Eisen-Atom allerdings schnell zur dreiwertigen Form oxidiert, sodass das Enzym selbst nach Dissoziation des Catecholamins *irreversibel* inaktiviert zurückbleibt [56, 45]. Erst durch Reduktion des dreiwertigen Eisens mit stöchiometrischen Mengen von  $\text{BH}_4$  wird die TH wieder funktionsfähig [89]. PAH zeigt *in vitro* das gleiche Verhalten [34].

Kristall-Strukturen der PAH zeigen, dass Catecholamine über die beiden phenolischen OH-Gruppen einen *bidentaten Chelat-Komplex* mit dem Eisen-Atom bilden [34]. Beide Enzyme bilden mit Catecholaminen blau-grün gefärbte *charge-transfer*-Komplexe ( $\lambda_{max}$  695–700 nm bei TH bzw. 698 nm bei PAH [34, 74]) und besitzen außerdem ähnliche Raman- und Elektronen-Spin-Resonanz-Spektren<sup>9</sup> [95]. Die experimentellen Befunde sind deutliche Hinweise darauf, dass Catechol-Derivate in beiden Enzymen den gleichen Bindemodus aufweisen. Von der TH existiert allerdings keine KS, die diesen Schluss belegt.

## Verfügbare Kristall-Strukturen

Vollständig lassen sich die Enzyme TH und PAH nicht kristallisieren, daher fehlt meist die regulatorische Untereinheit oder ist (wie bei 2phm [83]) nur teilweise vorhanden. Struktur-Informationen zu den prinzipiell wichtigeren, katalytischen Untereinheiten sind jedoch verfügbar. Im Hinblick auf funktionelle Fragen sind weniger die freien Enzyme, sondern ihre Komplexe mit verschiedenen Liganden interessant. Von der PAH existieren Komplexe mit dem Cosubstrat-Analog 7,8-Dihydrobiopterin (7,8-BH<sub>2</sub>, [33]) bzw. auch mit dem Cosubstrat  $\text{BH}_4$  selbst [4] und mit Catechol-Inhibitoren [34]. Von der TH ist lediglich das freie Enzym und ein Komplex mit dem Cosubstrat-Analog 7,8-BH<sub>2</sub> bekannt [52, 53].

---

<sup>9</sup>Mit Noradrenalin zeigt die PAH ein ESR-Spektrum, das dreiwertigem high-spin-Eisen in einem axialen Liganden-Feld entspricht (Gesamtspin  $S = 5/2$ ,  $g = 7.0, 5.2, 1.9$  [34]); die TH zeigt nach Zugabe von DA ein EPR-Spektrum mit ebenfalls  $S = 5/2$  und  $g = 7.1, 4.8, 1.9$  [123].

### 4.1.2. Fragestellungen

Aufgrund der auffälligen Ähnlichkeiten von Substraten, Sequenzen und Strukturen zwischen den katalytischen Domänen der PAH und TH würde man annehmen, dass sie einen gemeinsamen Katalyse-Mechanismus besitzen [43]. Tatsächlich aber gibt es einige Diskrepanzen zwischen den experimentellen Befunden zu beiden Enzymen und einige offene Fragen:

- Die KS 1dmw [33] und 2toh [53] zeigen das Cosubstrat-Analog in den Bindetaschen von PAH und TH. Bindet das aktive Cosubstrat  $\text{BH}_4$  an dieselbe Stelle, wie das Analog 7,8- $\text{BH}_2$ , liegt schon die erste Stufe eines produktiven Substrat-Komplexes vor. Übereinstimmend weist das Cosubstrat-Analog 7,8- $\text{BH}_2$  in beiden Fällen jeweils mit dem Carbonyl-O-Atom zum Eisen hin, während das Ring-System annähernd parallel zum Phenyl-Ring von Phe254 (PAH) bzw. Phe300 (TH) liegt. Abweichend jedoch ist das 7,8- $\text{BH}_2$ -Molekül in der TH (TH-Bindemodus) um  $180^\circ$  gegenüber der Orientierung in PAH gedreht (PAH-Bindemodus, s. Abb. 4.7). Damit ist unklar, ob beide Enzyme den gleichen Katalyse-Ablauf besitzen, und falls ja, welche der beiden Orientierungen der produktiven Position entspricht.
- Zur Position der AS im produktiven Substrat-Komplex gibt es lediglich Hinweise für Phe im Komplex mit 7,8- $\text{BH}_2$ -PAH aus NMR- und Docking-Experimenten [127]. Informationen über die Struktur eines entsprechenden Substrat-Komplexes der TH bzw. über die Übertragbarkeit von Ligand-Positionen zwischen den beiden Enzymen liegen nicht vor.
- Bei der PAH ist die Reihenfolge, in der die drei Substrate Phe,  $\text{O}_2$  und  $\text{BH}_4$  binden, gleichgültig für eine produktive Reaktion [43]. Bei der TH hingegen müssen die Substrate streng in der Reihenfolge erstens  $\text{BH}_4$ , dann  $\text{O}_2$  und schließlich Tyr an das aktive Zentrum binden, um umgesetzt werden zu können [44]. Die Ursache für dieses Verhalten ist bislang unklar.
- Speziell bei der TH ist (besonders im Hinblick auf die Parkinsonsche Erkrankung) der Bindemodus der kürzlich entdeckten Neurotoxine vom Tetrahydroisochinolin-Typ interessant, und ebenso, ob möglicherweise eine Parallele zu dem pflanzlichen Herbizid und Parkinson-Auslöser Rotenon besteht.

Daher werden im Folgenden einige Komplexe von PAH und TH mit Substraten bzw. Catechol-Derivaten konstruiert, um die zugrundeliegenden Prinzipien zu untersuchen.

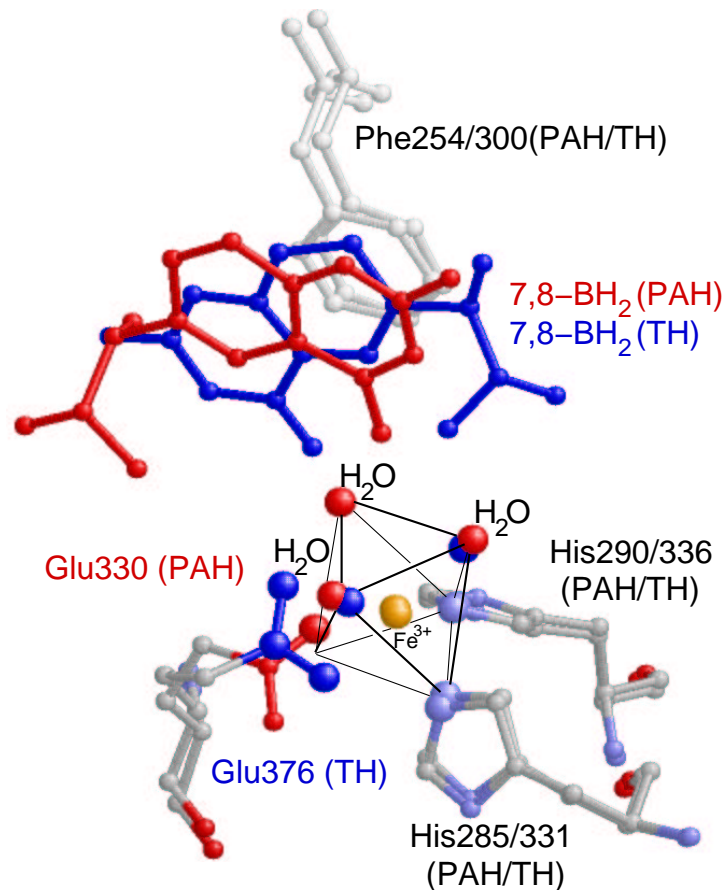


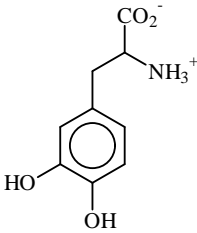
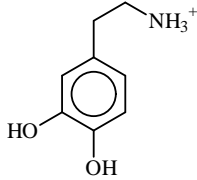
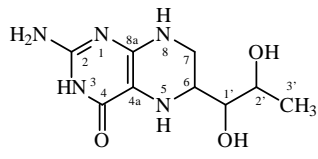
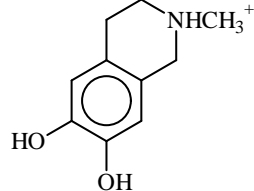
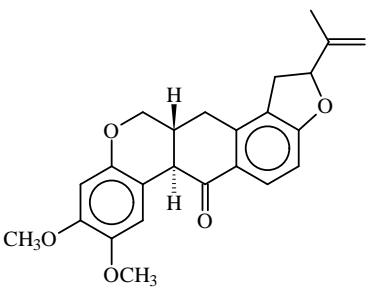
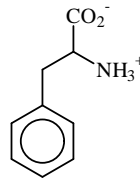
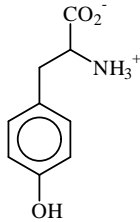
Abbildung 4.7: Ausschnitt aus den überlagerten KS 2toh (TH) und 1dmw (PAH), beide haben 7,8-BH<sub>2</sub> gebunden, allerdings in jeweils entgegengesetzten Orientierungen (TH-Bindemodus dunkel, PAH-Bindemodus hell).

## 4.2. Durchführung

Die Komplexe von PAH und TH werden ausgehend von den Kristall-Strukturen der Proteine in zwei Stufen [64] erzeugt, da der Bindemodus der Liganden teilweise unbekannt ist. Zuerst wird durch Docking mit FLEXX eine Vorauswahl der wahrscheinlichsten Ligand-Konformationen innerhalb der Bindetasche getroffen. Danach werden die verschiedenen Struktur-Vorschläge durch Energie-Minimierung optimiert und nach ihrer Gesamtenergie bewertet. Diejenigen Komplexe mit den niedrigsten Energie-Werten müssten theoretisch den tatsächlich vorkommenden Strukturen entsprechen. Die auf diese Weise ausgewählten Komplexe werden auf Stimmigkeit mit verschiedenen experimentellen Befunden überprüft.

Hoffmann et al. [63] haben diese Methode bereits erfolgreich angewendet, um mehrere Komplexe bekannter Struktur zu reproduzieren. In dem speziellen Fall der TH und PAH ist dennoch eine wiederholte Validierung der Vorgehensweise notwendig, da das aktive Zentrum beider Enzyme – anders als bei der Mehrzahl aller Proteine – ein Eisen-Atom enthält, welches z. B. bei FLEXX ursprünglich nicht vorgesehen ist. Um die Verlässlichkeit der Methode zu prüfen, werden die folgenden Komplexe erzeugt und mit den entsprechenden Referenz-KS verglichen (s. Abb. 4.8, obere Reihe):

Abbildung 4.8.: Übersicht über die Liganden, aus denen mit TH bzw. PAH Komplexe konstruiert werden.

	Catechol-Derivate		Substrate			
Rekonstruktion						
	L-DOPA	DA	BH <sub>4</sub>			
	TH/PAH	TH/PAH	TH/PAH			
Vorhersage						
	MDTIQ	Rotenon	Phe	Tyr		
	TH	TH	TH/PAH	TH/PAH		

**Komplexe der PAH mit Catechol-Derivaten** zur Klärung, ob Docking und Energie-Bewertung auch bei Eisen-haltigen Proteinen funktioniert und ob die Ergebnisse von individuellen Protein-Konformationen unabhängig sind:

- DA·5pah (Referenz: 5pah)
- L-DOPA·6pah (Referenz: 6pah)
- DA·6pah (Referenzen: 5pah und 6pah)
- L-DOPA·5pah (Referenzen: 6pah und 5pah)



**Komplexe beider Enzyme mit dem Cosubstrat** für Hinweise auf den relevanten Bindemodus von BH<sub>4</sub>:

- BH<sub>4</sub>·2toh (Referenz: 2toh)
- BH<sub>4</sub>·5pah (Referenz: 1dmw)

Nach erfolgreicher Validierung sollen, um Aussagen über die Funktionsweise der Enzyme ableiten zu können, weitere Komplexe modelliert werden (s. Abb. 4.8, untere Reihe):

**TH-Komplexe mit Catechol-Derivaten** zur Untersuchung, ob sich der für spezielle Catecholamine in PAH belegte Bindemodus allgemein für Catechol-Derivate auf die TH übertragen lässt:

- DA·2toh
- L-DOPA·2toh
- MDTIQ·2toh (N-Methyl-Dihydroxytetrahydroisochinolin)
- Rotenon·2toh

**Substrat-Komplexe** von TH bzw. PAH für Hinweise auf den Ablauf der katalytischen Reaktion und die Ursachen der beobachteten Substrat-Spezifitäten beider Enzyme:<sup>10</sup>

- Tyr·BH<sub>4</sub>·2toh
- Phe·BH<sub>4</sub>·2toh
- Tyr·BH<sub>4</sub>·5pah
- Phe·BH<sub>4</sub>·5pah

### 4.2.1. Komplex-Konstruktion

Die Komplexe werden durch flexibles Docking der Liganden in die Bindetasche des Proteins erzeugt. Dazu sind die folgenden Arbeitsschritte notwendig:

#### Vorbereitung des Protein-Anteils als Rezeptor

**Spezielle Modifikation:** Das in der KS der TH (2toh) angegebene *meta*-Tyrosin300 ist ein Kristallisationsartefakt [31]. Daher wird es wieder in die Standard-AS Phenylalanin umgewandelt.

---

<sup>10</sup>Auf die explizite Modellierung von O<sub>2</sub> wird wegen mangelnder Parametrisierung verzichtet. Wassermoleküle in der Liganden-Sphäre des Eisens dienen stattdessen als Platzhalter und sollen „Fehlplatzierungen“ der anderen Substrate verhindern.

**Überlagerung der Protein-Strukturen:** Die verschiedenen KS sind nicht einheitlich ausgerichtet. Um die Komplexe der TH und PAH später besser miteinander vergleichen zu können, werden sie aufeinander abgebildet; die KS 2toH dient dabei als Referenz. Zur Erzeugung geeigneter Rotationsmatrizen werden die den Proteinen jeweils gemeinsamen C $\alpha$ -Atome überlagert (TH: 163–179 und 188–470, PAH: 117–133 und 142–424<sup>11</sup>). Mit Hilfe der entsprechenden Rotationsmatrix werden auch die Ligand-Koordinaten transformiert, sodass ihre Positionen innerhalb der Bindetaschen zwischen den beiden Enzymen übertragbar sind.

**Gebundene Partikel:** Neben den Protein-Atomen gehören vor allem das Eisen-Atom und einige Wasser-Moleküle zur Bindetasche.

- In den KS 5pah, 6pah und 2toH liegt das Eisen-Atom in seiner dreiwertigen, enzymatisch inaktiven Form vor [74]. Es wird jedoch als zweiwertiges Kation für FLEXX nach dem Muster des Zink-Atoms mit dem Ionen-Radius von 0.75 Å (passend für *high-spin*-Fe<sup>II</sup> in oktaedrischer Umgebung [68]) definiert. Auch im Rahmen der Kraftfeld-Rechnungen unter CHARMM wird es als zweiwertiges Atom behandelt. Die Ladung wird in einer eigenen RTF-Datei angegeben.
- Die meisten KS der PAH zeigen eine regelmäßig oktaedrische Anordnung von Liganden in der ersten Koordinationssphäre des Eisens. Um die Erzeugung irrelevanter Konformationen einzuschränken, wird die Zahl der für die Ligand-Platzierung zur Verfügung stehenden Koordinationsstellen am Eisen reduziert, indem spezielle, in der KS vorhandene Wasser-Moleküle im Raum der Bindetasche beibehalten werden. Tabelle 4.1 zeigt, welche Wasser-Moleküle aus der KS 1pah bzw. 2toH jeweils berücksichtigt werden. Die Koordinationsstelle *trans* zu His285(PAH) bzw. His331(TH) bleibt immer frei. Bei der Konstruktion der Substrat-Komplexe stellen explizite Wasser-Moleküle sicher, dass an deren Stelle in der ersten Liganden-Sphäre des Eisens bei einem Austausch gegen ein Sauerstoff-Molekül genug Platz dafür bleibt.

Um explizite Wasser-Moleküle im Rahmen von FLEXX in die Bindetasche einbinden zu können, müssen die O-Atome erst um H-Atome ergänzt und optimiert werden. Dazu wird mit CHARMM eine kurze Struktur-Optimierung der ausgewählten Wasser-Moleküle unter Fixierung aller anderen Atome vorgenommen (*conjugated gradient*, 100 Schritte mit  $\varepsilon = 2r$ ).

**Rezeptor-Beschreibung:** Für jedes Protein wird eine eigene Datei angelegt (*receptor description file*), in der die von FLEXX zu verwendende PDB-Datei des Proteins, sowie die Liste der zu berücksichtigenden Partikel (Eisen-Atom und Wasser-Moleküle, Auswahl s. o.) angegeben wird.

<sup>11</sup>In der Sequenz der KS 2toH fehlen die AS183–185, der korrespondierende oberflächliche Loop ist also kürzer als derjenige bei der KS der PAH

*Tabelle 4.1.:* Besetzungsmuster der Oktaeder-Plätze um das Eisen-Atom mit Wasser-Molekülen je nach zu platzierendem Liganden.

Ligand	H <sub>2</sub> O <i>trans</i> zu His290/336 (PAH/TH)	H <sub>2</sub> O <i>trans</i> zu Glu330/376 (PAH/TH)
Catechol-Derivate	–	+
BH <sub>4</sub>	+	+
Tyr	–	+
Phe	+	+

### Vorbereitung der Liganden

**Editieren:** Ausgehend von den in den KS 6pah und 2toh enthaltenen Liganden L-DOPA bzw. 7,8-BH<sub>2</sub> werden die Strukturen der übrigen Liganden abgeleitet (s. Abb. 4.8). Ihre Atom-Koordinaten und die zur Kraftfeld-Rechnung benötigten Topologie-Dateien werden mittels der *molecular editor*-Funktion unter Quanta (<http://www.csc.fi/chem/progs/msi.html> [70]) erstellt.

Aus KS gehen die Protonierungszustände titrierbarer Gruppen prinzipiell nicht hervor; sie sind aber entscheidend für eine korrekte Platzierung. Die Catecholamin-Komplexe der PAH (KS 3pah – 6pah) zeigen, dass die Seitenketten trotz Bindung zum Protein solvatisiert bleiben, daher wird für L-DOPA, DA und auch die AS Phe und Tyr die in Lösung überwiegende, zwitterionische Form eingesetzt (die *pK*-Werte betragen ca. 2 für die CO<sub>2</sub><sup>–</sup>-Gruppe und ca. 9 für die NH<sub>3</sub><sup>+</sup>-Gruppe [135]). In Homologie zum DA wird auch die tertiäre Amino-Gruppe von MDTIQ in der protonierten Form verwendet. Bei BH<sub>4</sub> wird die auch bei den verwandten Nucleinbasen und neutralem pH vorherrschende Lactam-Form statt der alternativen Lactim-Form verwendet [105].

**Struktur-Optimierung:** Alle Liganden werden einer Energie-Minimierung unter den Standard-Einstellungen von Quanta (*conjugated gradient*, 200 Schritte mit  $\varepsilon = 1$ ) unterzogen, um Bindungslängen zu optimieren.

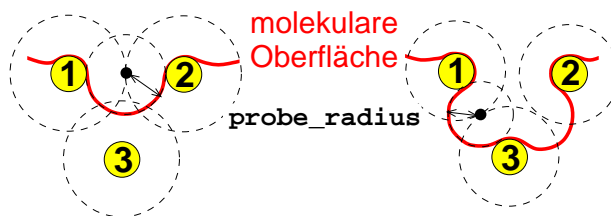
**Umwandlung in FLEXX-kompatibles Eingabe-Format:** FLEXX benötigt für jeden Liganden die Atom-Koordinaten, Atom-Typen, Formal-Ladungen und das Verknüpfungsmuster in einer einzigen Datei. Die im (für FLEXX unlesbaren) PDB-Format vorliegenden Koordinaten jedes Liganden werden in das *molecular-modeling*-Programm Sybyl (Tripos <http://www.tripos.com/>) eingelesen, welches diese auch im benötigten Mol2-Format ausgeben kann. Über eine entsprechende Editor-Funktion werden die dem Mol2-Kraftfeld entsprechenden Atom-Typen und Formal-Ladungen zugeordnet (den beiden O-Atomen der CO<sub>2</sub><sup>–</sup>-Gruppe 2×-0.5, dem N-Atom der NH<sub>3</sub><sup>+</sup>-Gruppe +1).

### Docking mit FLEXX:

**Definition der Bindetaschen:** Die Auswahl der relevanten Atome erfolgt interaktiv: Alle Partikel bzw. alle AS im Radius von 8.0 Å um den in der jeweiligen KS gebundenen Liganden (DA in 5pah, L-DOPA in 6pah und 7,8-BH<sub>2</sub> in 2toh) werden bei der Validierung als zur Bindetasche gehörig betrachtet. Zur Vorhersage der AS-Substrat-Positionen wird die Bindetasche auf alle AS im Radius von 14.0 Å um die Referenz-Liganden erweitert.

Nur die Atome dieser Auswahl, die zur molekularen Oberfläche beitragen, werden von FLEXX als potentielle Wechselwirkungspartner aufgefasst. Um auch das wichtige Eisen-Atom trotz seiner vielen Liganden als Wechselwirkungspartner erfassen zu können, muss der Parameter `probe_radius` vom vorgegebenen Wert (0.8 Å) auf 0.6 Å verringert werden (vgl. Kap. 2.3.1 und Abb. 4.9). Die molekulare Oberfläche des Proteins ist dann feiner strukturiert.

Abbildung 4.9: Bei größerem `probe_radius`-Wert (links) tragen von den drei Atomen nur die ersten beiden zur molekularen Oberfläche bei. Erst wenn dieser Parameter verringert wird, kann auch Atom 3 als potentieller Wechselwirkungspartner von FLEXX berücksichtigt werden (rechts).



### Platzierung der Liganden:

- Der Parameter `max_overlap_vol` gibt das erlaubte Maß der Durchdringung von Protein- und Ligand-Atomen an. Um auch Platzierungen mit kleineren Kollisionen, die bei der anschließenden Struktur-Optimierung leicht zu beheben sind, beizubehalten, wird der Wert von den vorgegebenen 2.5 Å<sup>3</sup> auf 5.0 Å<sup>3</sup> erhöht.
- Die Liganden werden automatisch in Fragmente zerlegt, die Basis-Fragmente werden ebenfalls automatisch ausgewählt.
- Die Platzierung der Basis-Fragmente erfolgt grundsätzlich automatisch, d. h. nach dem Triangulierungsalgorithmus (s. Kap. 1.5.3) und ohne Zusatz-Information über die erwartete Position des Liganden innerhalb der Bindetasche. Nur BH<sub>4</sub> wird zusätzlich auch mit dem *perturbate*-Modus auf eine vorgegebene Referenz-Position (BH<sub>4</sub> im TH-Bindemodus) platziert: Dabei wird das automatisch festgelegte Basis-Fragment auf die entsprechende Gruppe der Referenz-Struktur gelegt und nur für die restlichen Fragmente sucht FLEXX passende Positionen.
- Aufbauend auf der gegebenen Platzierung des Basis-Fragmentes wird der Rest des Moleküls ergänzt. Die maximale Zahl alternativer Platzierungen des Basis-Fragmentes beträgt in jedem Schritt 30.

## Struktur-Optimierung

Jeder FLEXX-Vorschlag wird einer Energie-Minimierung mit  $\varepsilon = 2r$  und einem Cutoff von 15.0 Å unterzogen (600 Schritte nach der *conjugated-gradient*-Methode). Nur die Ligand-Atome sind frei beweglich.

### 4.2.2. Auswertung

**Energie-Rechnungen** Die Gesamtenergie der Komplexe wird als Summe aus Kraftfeld- und Solvatationsenergie berechnet (Gl. 2.17), mit  $\varepsilon = 4$ . Der Cutoff-Wert für die Bestimmung der Coulomb-Wechselwirkungen im Rahmen der Kraftfeld-Rechnung liegt bei 100 Å.

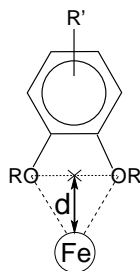
Um Rechenleistung während der Berechnung der Solvatationsenergien zu sparen, wird die Gesamtenergie nur für einen Ausschnitt aus dem Protein berechnet: Vom Protein werden nur die AS im Umkreis von 15.0 Å um das Eisen-Atom berücksichtigt. Damit sind *alle* Komplexe zusätzlich zu den Modell-bedingten Fehlern (s. Abschnitte 2.2, 2.3, 2.3.1 und 2.4) mit einem sehr großen, aber konstanten Fehler behaftet. Aus diesem Grund werden nicht absolute, sondern relative Energie-Werte angegeben, die bezogen werden auf die minimale Energie des jeweiligen Ensembles verschiedener Komplex-Konformationen.

**rmsd-Werte** zwischen äquivalenten Atomen werden zum Vergleich verschiedener Ligand-Platzierungen ermittelt. Bei den Catechol-Derivaten bzw. beim Cosubstrat werden *rmsd*-Werte für die Ring-Systeme bzw. Seitenketten getrennt ermittelt, um besser zwischen relevanten und irrelevanten Abweichungen von Molekülteilen unterscheiden zu können.

**Referenz-Strukturen** werden aus den KS übernommen. Bei 5pah und 6pah werden nur die Catechol-Ringe, nicht die Seitenketten von DA bzw. L-DOPA verwendet. Das Molekülgerüst von 7,8-BH<sub>2</sub> aus 2toh bzw. 1dmw dient als Vorlage für die Platzierung von BH<sub>4</sub>.

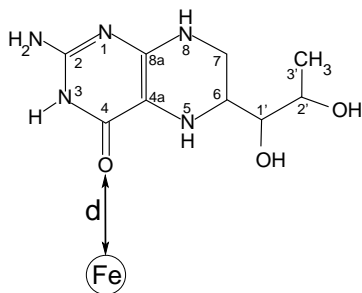
**Intermolekulare Abstände** zwischen bestimmten Bezugspunkten werden zur qualitativen Einschätzung von Platzierungen verwendet. Durch die Größe der Bindetasche und die fehlenden regulatorischen Domänen sind Platzierungen (besonders im Fall der AS-Substrate) in der Peripherie der Bindetasche zwar sehr häufig, aber irrelevant, ungeachtet ihrer energetischen Bewertung. Bei den Catechol-Derivaten dient das geometrische Zentrum zwischen den OH-Gruppen bzw. das N-Atom der NH<sub>3</sub><sup>+</sup>-Gruppe als Bezugspunkt, beim Cosubstrat BH<sub>4</sub> das Carbonyl-O-Atom und bei den AS Phe und Tyr das Zentrum des aromatischen Rings (s. Abb. 4.10). Entscheidend ist jeweils die Entfernung des jeweiligen Bezugspunktes zum zentralen Eisen-Atom. Als relevant werden jeweils nur Platzierungen erachtet, die Abstände von weniger als 5.0 Å im Fall der Catechol-Derivate, 4.5 Å bei BH<sub>4</sub> und 8.0 Å bei den AS aufweisen. Die Werte beinhalten einen gewissen Sicherheitsabstand, sodass alle potentiell relevanten Platzierungen erfasst werden.

R= -CH<sub>3</sub>: Rotenon  
 R= -H: DA, L-DOPA,  
 MDTIQ  
 R': vgl. Abb. 4.8



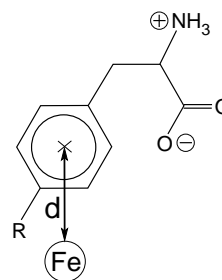
$$d_{max} = 3.0 \text{ \AA}$$

BH<sub>4</sub>



$$d_{max} = 4.5 \text{ \AA}$$

R= -H: Phe  
 R= -OH: Tyr



$$d_{max} = 8.0 \text{ \AA}$$

Abbildung 4.10.: Der Abstand des Eisen-Atoms zum geometrischen Zentrum der O-Atome der Catechol-Ringe bzw. zum Carbonyl-O-Atom von BH<sub>4</sub> bzw. zum Zentrum der aromatischen Ringe dient als Kriterium zur ersten Einschätzung von Struktur-Vorschlägen.

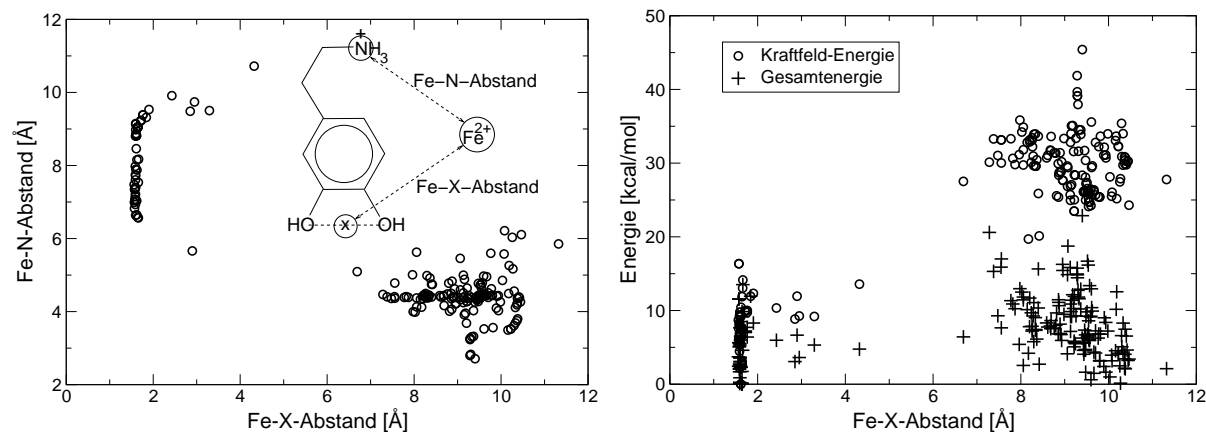
## 4.3. Ergebnisse

### 4.3.1. Validierung

#### Catechol-PAH-Komplexe

**DA-5pah** Für das Redocking von DA in die Bindetasche der PAH liefert FLEXX 213 Struktur-Vorschläge. Das Diagramm in Abb. 4.11(a) gibt anhand der intermolekularen Eisen-Catechol- bzw. Eisen-NH<sub>3</sub><sup>+</sup>-Abstände einen Überblick über die vorhergesagten Konformationen. Man erkennt zwei große Gruppen, nämlich entweder mit kleinen Eisen-Catechol-Abständen und großen Eisen-NH<sub>3</sub><sup>+</sup>-Abständen (1.6–2.3 Å bzw. 6.5–11 Å) oder umgekehrt (2.5–6.3 Å bzw. 6.5–11.5 Å). Außerdem gibt es noch eine weitere, kleinere Gruppe von Konformationen mit jeweils mittleren Abständen. Nach den experimentellen Befunden [34, 74, 123, 95] ist nur die erste Gruppe von Konformationen relevant; sie zeigen den erwarteten bidentaten Bindemodus der Catechol-Gruppe von DA an das Eisen-Atom. Die zweite Gruppe von Vorschlägen enthält Platzierungen, die monodentat, d. h. nur über *eine* OH-Gruppe am Eisen verankert sind. Bei der dritten Gruppe von Konformationen, ohne jede Koordination zum Eisen, sind die NH<sub>3</sub><sup>+</sup>-Gruppen dem Eisen-Atom jedoch unerwartet nah, mit Abständen von teilweise nur 3.0 Å. Die NH<sub>3</sub><sup>+</sup>-Gruppe ist in diesen Fällen über zwei H-Brücken, zur OH-Gruppe von Tyr325 einerseits und zur CO<sub>2</sub><sup>-</sup>-Gruppe von Glu330 andererseits, dicht beim Eisen-Atom verankert, während die OH-Gruppen des DA weitere H-Brücken zur phenolischen OH-Gruppe von Tyr377 ausbilden. Ungeachtet der Abstoßung zwischen dem zweiwertigen Eisen-Atom und der kationischen NH<sub>3</sub><sup>+</sup>-Gruppe, gehören nicht

nur ca. 75% der Vorschläge zu dieser letzten Gruppe, sondern werden nach der FLEXX-Bewertungsfunktion auch als sehr wahrscheinlich eingestuft.



(a) Überblick über die Vielfalt der Konformationen: Meist ist das Dopamin-Molekül über einen der beiden Bezugspunkte (Catechol- oder  $\text{NH}_3^+$ -Gruppe) in der Bindetasche verankert.

(b) Bewertung der Komplex-Konformationen nach dem Abstand der Catechol-Gruppe zum Eisen-Atom und nach der Kraftfeld- bzw. Gesamtenergie.

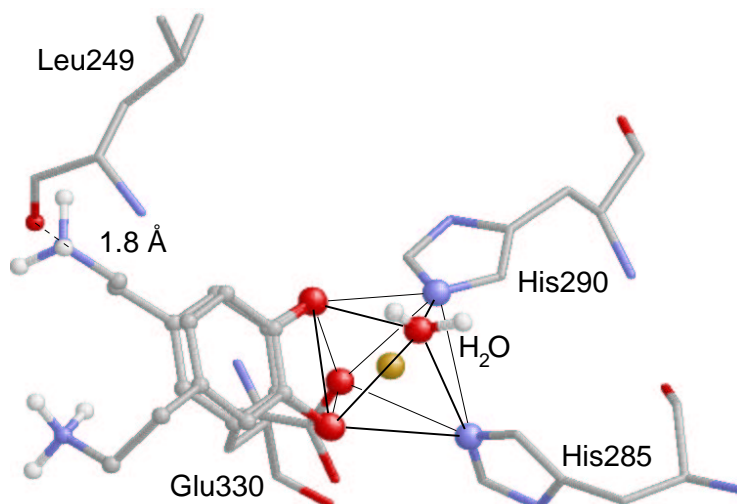
Abbildung 4.11.: Redocking von DA in 5pah: Die bidentaten Komplexe werden nach der Kraftfeld-Energie deutlich günstiger bewertet als die anderen Platzierungen; nach der Gesamtenergie lassen sich die Gruppen kaum von einander trennen.

Überraschend ist auch, dass diese beiden Gruppen der bidentaten bzw. über die  $\text{NH}_3^+$ -Gruppe verankerten Komplexe auch nach der Gesamtenergie ebenfalls etwa gleich beurteilt werden (s. Abb. 4.11(b)). Nach dem Kriterium der Kraftfeld-Energie allerdings, d. h. ohne Berücksichtigung der Solvatationsenergie, wird eindeutig die (korrekte) Gruppe, mit bidentater Koordination zum Eisen, bevorzugt.

Nach der Gesamtenergie wird der Komplex mit FLEXX-Rang 16 als beste bidentate Konformation bewertet. Bezogen auf diese Struktur werden die übrigen 53 bidentaten Konformationen untersucht. Es zeigt sich, dass die Catechol-Ringe mit Eisen-Catechol-Abständen von 1.56 – 1.90 Å ( $\bar{\text{O}}$ :  $1.63 \pm 0.07$  Å) einheitlich platziert sind ( $\text{rmsd}$ : 0.00 – 1.42 Å,  $\bar{\text{O}}$ :  $0.65 \pm 0.44$  Å), während die Positionen der Seitenketten stärker abweichen und  $\text{rmsd}$ -Werte zwischen 0.005–5.40 Å aufweisen ( $\bar{\text{O}}$ :  $2.33 \pm 1.72$  Å).

Die große Vielfalt der Seitenketten-Konformationen beruht darauf, dass die planare Catechol-Einheit von DA in zwei Orientierungen einen bidentaten Komplex mit dem Eisen-Atom bilden kann, ohne dass sich dadurch die unmittelbare Umgebung des Eisens verändert: Bei einer Variante befindet sich die *para*-OH-Gruppe *trans* zu His285 und die *meta*-OH-Gruppe *trans* zu His290, bei der anderen, um 180° gedrehten Variante ist es umgekehrt. So können sich die Catechol-Ringe decken, während die  $\text{NH}_3^+$ -Seitenketten in unterschiedliche Richtungen weisen (s. Abb. 4.12).

Abbildung 4.12: Alternativen des bidentaten Bindemodus von DA in PAH (5pah) gemäß der Bewertung nach der Gesamtenergie. Dargestellt sind nur die Liganden des Eisen-Atoms, sowie Leu249 als Wechselwirkungspartner des Dopamins. Die Catechol-Gruppen beider Varianten decken sich fast vollständig, lediglich die  $\text{NH}_3^+$ -Seitenketten sind unterschiedlich orientiert. H-Atome sind nicht dargestellt.

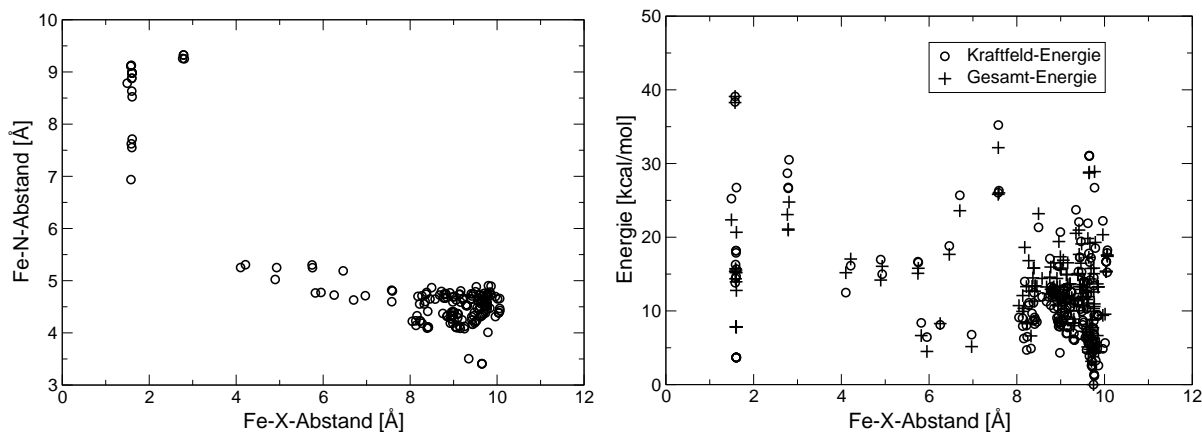


Die Gesamtenergien der 10 günstigsten bidentaten Konformationen liegen zwischen 0.0 – 0.1 kcal/mol, die genaue Position der Seitenkette spielt also offensichtlich keine Rolle für die Energie-Bewertung. Es kommt vor, dass die  $\text{NH}_3^+$ -Seitenkette völlig frei bleibt oder eine H-Brücke ausbildet zu einer passenden Akzeptor-Gruppe des Proteins, wie z.B. dem O-Atom von Leu249 oder der OH-Gruppe von Tyr377.

**L-DOPA-6pah** Beim Redocking von L-DOPA in die KS 6pah der PAH schlägt FLEXX 197 Varianten vor. Dabei sind die OH-Gruppen von L-DOPA wieder bidentat, monodentat oder gar nicht am Eisen-Atom fixiert (s. Abb. 4.13(a)). Die Gruppe der bidentaten Konformationen wird sowohl nach der Gesamtenergie, als auch nach der Kraftfeld-Energie deutlich schlechter bewertet als die große Gruppe der *nicht* über die Catechol-Gruppe verankerten Konformationen (s. Abb. 4.13(b)). Dies beruht darauf, dass in diesen Fällen die anionische  $\text{CO}_2^-$ -Gruppe an das kationische Eisen-Atom bindet und die OH-Gruppen H-Brücken zu den schon bekannten Partnern Ser251 und Leu249 ausbilden können. Die  $\text{NH}_3^+$ -Gruppe bleibt ohne Wechselwirkungspartner.

Von den 19 Vorschlägen mit dem erwarteten Bindemodus konvergieren sieben zu einer einzigen, besonders günstigen Struktur. In dieser sitzt die *para*-OH-Gruppe *trans* zu His290 und die *meta*-OH-Gruppe *trans* zu His285; die  $\text{NH}_3^+$ -Gruppe bildet mit einem Abstand von 1.85 Å eine H-Brücke zur Carbonyl-Gruppe von Leu249 und die unpolaren Seitenketten von Phe254 und Leu248 umgeben die  $\text{CO}_2^-$ -Gruppe. Die Positionen der übrigen Catechol-Ringe weichen nur um  $0.44 \pm 0.46$  Å von der günstigsten Platzierung ab. Die Seitenketten dagegen sind um bis zu 4.9 Å ( $\emptyset 2.04 \pm 2.26$ ) verschoben, was mit um 5 – 30 kcal/mol bzw. 10 – 36 kcal/mol höheren Werten in der Gesamt- bzw. Kraftfeld-Energie bestraft wird. Demnach wäre der Bindemodus von L-DOPA an PAH auf die beschriebene Konformation festgelegt.





(a) Überblick über die Vielfalt der Konformationen: Das L-DOPA-Molekül ist entweder über eine bzw. beide OH-Gruppen oder über die  $\text{CO}_2^-$ -Gruppe am Eisen-Atom verankert.

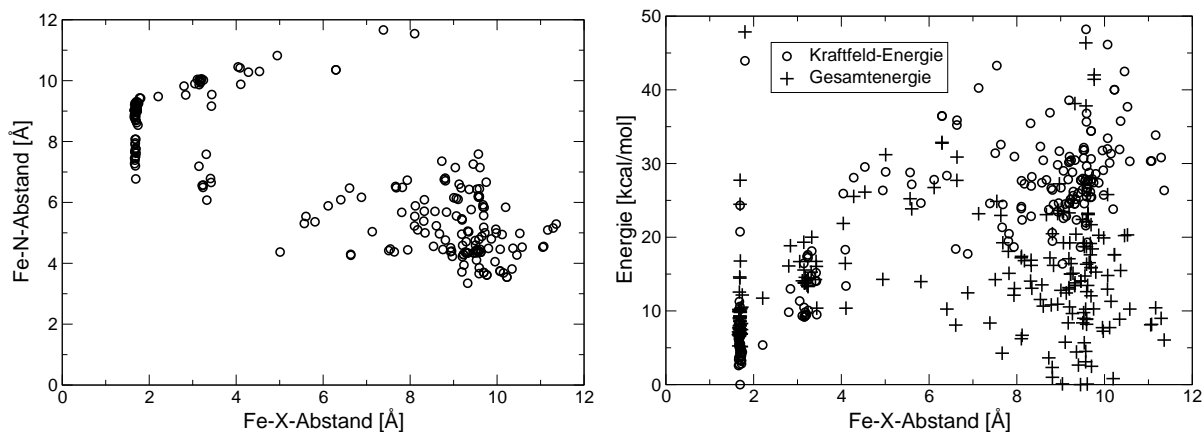
(b) Bewertung der Komplex-Konformationen nach dem Abstand der Catechol-Gruppe zum Eisen-Atom und nach der Kraftfeld- bzw. Gesamtenergie.

Abbildung 4.13.: Redocking von L-DOPA in 6pah: Die Gruppe der über die  $\text{CO}_2^-$ -Gruppe angebotenen Konformationen wird sowohl nach der Kraftfeld-, als auch nach der Gesamtenergie gegenüber der Catechol-Anbindung bevorzugt.

**DA-6pah** Die KS 6pah enthält den Komplex der PAH mit dem Liganden L-DOPA. FLEXX liefert für das Docking von DA in diese, eigentlich an L-DOPA angepasste Binde-tasche 228 Vorschläge. Die Abbildung 4.14(a) zeigt wieder die Eisen-Catechol- bzw. Eisen- $\text{NH}_3^+$ -Abstände aller Komplexe. Demnach tauchen die drei, schon vom Redocking des DA in die KS 5pah bekannten Gruppen von Konformationen auf, die bidentaten, monodentaten und anderen Bindemodi des DA an die PAH entsprechen. 63 Konformationen (ca. 30% der Vorschläge) besitzen die erwartete bidentate Koordination des DA zum Eisen mit Abständen von weniger als 2.0 Å zwischen dem Eisen-Atom und der Catechol-Gruppe.

Die Bewertung der verschiedenen Konformationen nach ihrem Energie-Inhalt ist hier jedoch weniger eindeutig als beim Redocking: Bei der Platzierung von DA in 6pah werden nach der Gesamtenergie Komplexe ohne jede Verankerung der Catechol-Gruppe am Eisen als besonders günstig bewertet (s. Abb. 4.14(b)). Nach der Kraftfeld-Energie hingegen werden wiederum eindeutig die bidentaten Varianten als die günstigsten identifiziert.

Unter den bidentaten Komplexen kommen wie zuvor die alternativen Orientierungen des DA-Moleküls vor. Als Referenz zum Struktur-Vergleich dient die nach der Gesamtenergie günstigste bidentate Konformation. Deren *para*-OH-Gruppe ist *trans* zu His290 gebunden, die *meta*-OH-Gruppe *trans* zu His285 und die  $\text{NH}_3^+$ -Gruppe bildet mit einem Abstand von 2.5 Å eine H-Brücke zur OH-Gruppe von Ser251. Alle Catechol-Gruppen sind einheitlich, mit *rmsd*-Werten von  $1.05 \pm 0.37$  Å zu dieser Struktur platziert, während die Seitenketten weniger festgelegt sind und *rmsd*-Werte von  $2.32 \pm 0.85$  Å (0.98 – 5.98 Å) aufweisen.



(a) Überblick über die Vielfalt der Konformationen: Wie zuvor ist das DA-Molekül entweder über die Catechol- oder  $\text{NH}_3^+$ -Gruppe in der Bindetasche verankert.

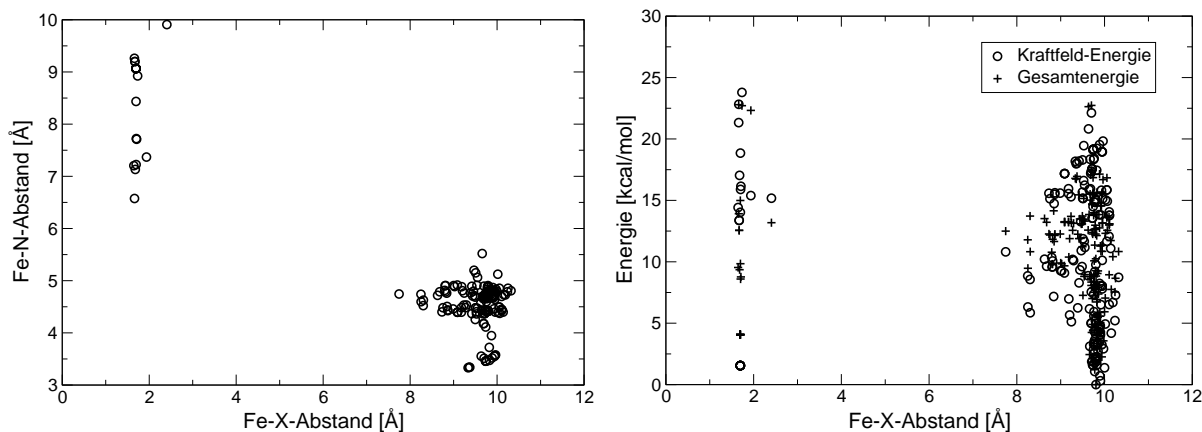
(b) Bewertung der Komplex-Konformationen nach dem Abstand der Catechol-Gruppe zum Eisen-Atom und nach der Kraftfeld- bzw. Gesamtenergie.

Abbildung 4.14.: Docking von DA in 6pah: Die Kraftfeld-Energie bewertet bidentate Platzierungen deutlich günstiger als alle anderen. Nach der Gesamtenergie werden irrelevante Konformationen bevorzugt.

Im Energie-Vergleich der 10 günstigsten Konformationen unterscheidet die Gesamtenergie kaum zwischen DA-Konformationen mit oder ohne Wechselwirkungen zum Protein. Es kommen gleichermaßen Formen mit H-Brücken alternativ zur OH-Gruppe von Ser251 ( $r=2.5\text{--}2.9\text{ \AA}$ ), zur OH-Gruppe von Tyr377 ( $r=2.4\text{--}2.8\text{ \AA}$ ) oder zur Carbonyl-Gruppe von Leu249 ( $r=1.8\text{ \AA}$ ) vor, wie auch Formen, bei denen die  $\text{NH}_3^+$ -Seitenkette völlig frei aus der Bindetasche herausragt.

**L-DOPA-5pah** Für die Platzierung von L-DOPA in die an DA angepasste Bindetasche der PAH der KS 5pah macht FLEXX 189 Vorschläge, von denen 10% dem erwarteten bidentaten Bindemodus entsprechen (s. Abb. 4.15(a)). Die übrigen werden sowohl nach der Gesamt- als auch nach der Kraftfeld-Energie günstiger bewertet; sie weisen durchweg Koordinationen der  $\text{CO}_2^-$ -Gruppe über eines oder beide O-Atome an das Eisen auf (s. Abb. 4.15(b)).

Die günstigste bidentate Konformation ist gekennzeichnet durch die Position des *para*-O-Atoms *trans* zu His290 bzw. des *meta*-O-Atoms *trans* zu His 285 und eine H-Brücke zwischen der  $\text{NH}_3^+$ -Gruppe und Leu249. Neun FLEXX-Vorschläge konvergieren zu dieser Struktur, die nach der Gesamt- bzw. Kraftfeld-Energie um mindestens 4 bzw. 12 kcal/mol günstiger bewertet wird als die übrigen bidentaten Konformationen. Bezüglich dieser günstigsten Konformation weisen die Catechol-Ringe *rmsd*-Werte bis zu  $1.64\text{ \AA}$  auf (im Mittel nur  $0.35 \pm 0.43\text{ \AA}$ ), während die Seitenketten-Orientierungen stärker voneinander abweichen mit *rmsd*-Werten zwischen  $0.84\text{--}5.73\text{ \AA}$  ( $\text{Ø}$ :  $1.53 \pm 1.75\text{ \AA}$ ).



(a) Überblick über die Vielfalt der Konformationen: Das L-DOPA-Molekül ist entweder über eine bzw. beide OH-Gruppen oder über die  $\text{CO}_2^-$ -Gruppe am Eisen-Atom verankert.

(b) Bewertung der Komplex-Konformationen nach dem Abstand der Catechol-Gruppe zum Eisen-Atom und nach der Kraftfeld- bzw. der Gesamtenergie.

Abbildung 4.15.: Docking von L-DOPA in 5pah: Die über die  $\text{CO}_2^-$ -Gruppe am Eisen angelagerten Konformationen werden gegenüber bidentaten Platzierungen nach der Kraftfeld-Energie weniger deutlich favorisiert als nach der Gesamtenergie.

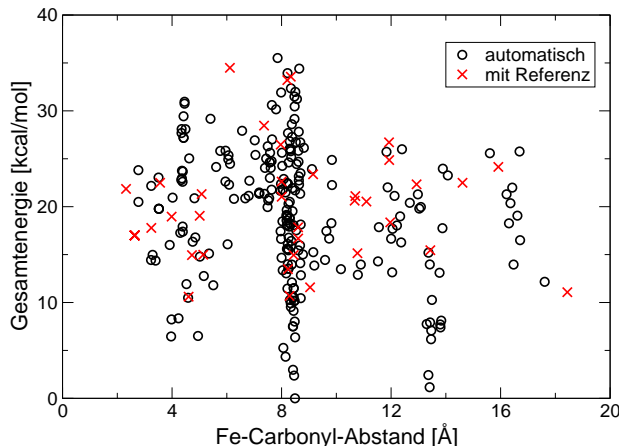
### $\text{BH}_4$ -Komplexe mit PAH bzw. TH

**$\text{BH}_4$ -5pah** Beim Docking von  $\text{BH}_4$  in die an DA angepasste Bindetasche der PAH (KS 5pah) liefert FLEXX 261 Struktur-Vorschläge, von denen 36 einen Eisen-Carbonyl-Abstand von weniger als 4.5 Å besitzen (s. Abb.4.16). 25 davon haben gute *rmsd*-Werte für das Pterin-Ringsystem von weniger als 2.0 Å bezüglich der Referenz-Position des  $\text{BH}_2$  aus der KS 1dmw. Die übrigen  $\text{BH}_4$ -Platzierungen mit Eisen-Carbonyl-Abständen von mehr als 4.5 Å befinden sich in der Peripherie der Bindetasche. Wieder unterscheiden sich die Bewertungen der einzelnen Konformationen nach Gesamt- und Kraftfeld-Energie: Nach der Kraftfeld-Energie werden innerhalb der relevanten Gruppe von Konformationen diejenigen mit den kürzeren Eisen-Carbonyl-Abständen (2.3–3.5 Å) bevorzugt. Die Gesamtenergie scheint die Komplexe jedoch zutreffender zu bewerten, denn danach ist eine Konformation günstiger, deren Pterin-Ringsystem zur Referenz (7,8- $\text{BH}_2$  aus 1dmw) eine *rmsd*-Verschiebung von nur 1.40 Å aufweist (s. Abb. 4.27). Der Eisen-Carbonyl-Abstand beträgt 3.9 Å. Die Position der Seitenkette weicht mit durchschnittlich 3.23 Å deutlich von der Referenz-Struktur ab<sup>12</sup>. In der günstigsten berechneten Position ist das  $\text{BH}_4$ -Molekül über drei H-Brücken in der PAH-Bindetasche fixiert: Vom N2-gebundenen H-Atom zum Carbonyl-O-Atom von Gly247, vom N9-gebundenen H-Atom zum Carbonyl-O-Atom von Leu249 und von der C2'-

<sup>12</sup>Allerdings stimmen auch die Ringsysteme der 7,8- $\text{BH}_2$  bzw.  $\text{BH}_4$ -Moleküle im Kristall mit der PAH (1dmw und 1j8u) mit durchschnittlich  $0.31 \pm 0.17$  Å besser überein als die Seitenketten mit  $2.55 \pm 1.28$  Å. Die genaue Position der Seitenkette scheint also weniger entscheidend zu sein, als die des Pterin-Teils.

OH-Gruppe zur Seitenkette von Ser251. Außerdem besteht ein enger hydrophober Kontakt zwischen dem Ring von Phe254 und einem Teil des  $\text{BH}_4$ -Moleküls. Die Ringebenen sind mit einem Neigungswinkel von nur  $17.9^\circ$  zwar fast parallel, aber gegeneinander verschoben. Das Eisen-Atom ist mit  $3.96 \text{ \AA}$  zu weit für eine Bindung entfernt, der Abstand stimmt jedoch gut mit den jeweils  $3.8 \text{ \AA}$  in den beiden Referenz-Strukturen überein.

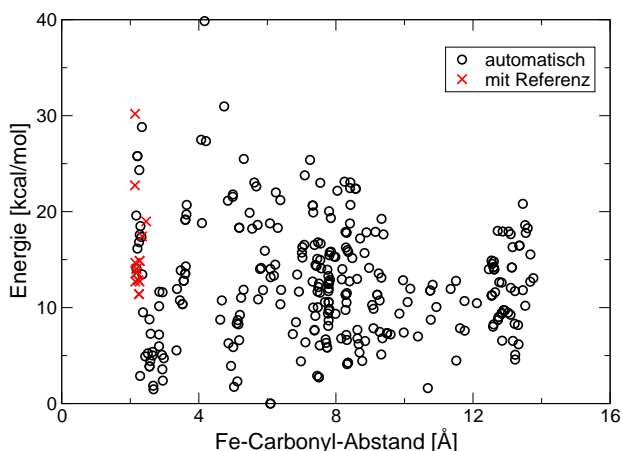
Abbildung 4.16: Docking von  $\text{BH}_4$  in die KS der PAH (5pah): Bewertung der Komplex-Konformationen nach dem Abstand der Carbonyl-Gruppe zum Eisen-Atom und der Gesamtenergie. Drei Arten der Platzierung sind besonders häufig und energetisch günstig; nur eine hat jedoch entsprechende Abstände von weniger als  $4.5 \text{ \AA}$  zum Eisen-Atom. Die automatisch erzeugten Konformationen werden gegenüber den manuell erzeugten bevorzugt.



Konformationen, die dem TH-Bindemodus entsprechen, werden während eines automatischen FLEXX-Durchlaufs nicht erzeugt. Bei der erzwungenen Platzierung von  $\text{BH}_4$  auf das 7,8- $\text{BH}_2$ -Molekül in dieser Orientierung erhält man 44 weitere Konformationen. Die günstigste von diesen zeigt einen Eisen-Carbonyl-Abstand von  $2.63 \text{ \AA}$  und wiederum eine zu Phe254 annähernd parallele Ringebene des  $\text{BH}_4$ -Moleküls (Neigungswinkel:  $16.8^\circ$ ). Zwischen dem  $\text{C1}'\text{-OH}$ -Atom und dem Carbonyl-O-Atom von His264 könnte noch eine (stark gestreckte) H-Brücke zur Bindung an das Protein beitragen. Im übrigen hat die  $\text{BH}_4$ -Seitenkette nur Kontakte zu unpolaren Protein-Teilen. Diese nachträglich erzeugten Konformationen schneiden nach der Gesamtenergie allerdings ungünstiger als die besten Vertreter der automatisch erzeugten Platzierungen ab (s. Abb. 4.16). Beim Vergleich der Energie-Beiträge der jeweils besten Kandidaten mit TH- bzw. PAH-Bindemodus zeigt die Solvatationsenergie einen um  $11.4 \text{ kcal/mol}$  ungünstigeren Beitrag beim TH-Bindemodus, während die übrigen Beiträge nur um höchstens  $2 \text{ kcal/mol}$  voneinander abweichen.

**$\text{BH}_4$ ·2toh** Das automatische Docking von  $\text{BH}_4$  in die an 7,8- $\text{BH}_2$  angepasste KS der TH (2toh) liefert 302 Komplex-Konformationen (s. Abb. 4.17); davon kommen 56 Vorschläge aufgrund ihrer Eisen-Carbonyl-Abstände von weniger als  $4.5 \text{ \AA}$  als plausibel in Betracht. Zur Referenz-Position des 7,8- $\text{BH}_2$  in 2toh weisen die entsprechenden Vorschläge jedoch *rmsd*-Werte von  $3.04$  bis über  $10 \text{ \AA}$  auf – der kristallographisch belegte TH-Bindemodus wird also nicht wiedergefunden, wohl aber der PAH-Bindemodus: Die sowohl nach der Gesamtenergie, als auch nach der FLEXX-Bewertung favorisierte Platzierung weist für das Pterin-Ringsystem eine *rmsd* von nur  $1.52 \text{ \AA}$  zu der aus der KS 1dmw entnommenen Referenz-Position von 7,8- $\text{BH}_2$  in PAH auf. Das  $\text{BH}_4$ -Ringsystem liegt dem aromatischen Ring von Phe300 fast parallel auf (Neigungswinkel:  $15.4^\circ$ ), während das Carbonyl-O-Atom

Abbildung 4.17: Docking von  $\text{BH}_4$  in die KS der TH (2toh): Bewertung der Komplex-Konformationen nach dem Abstand der Carbonyl-Gruppe zum Eisen-Atom und der Gesamtenergie. In der relevanten Gruppe von Vorschlägen mit Abständen kürzer als 4.5 Å werden die automatisch erzeugten Strukturen günstiger bewertet als die manuell erzeugten.



mit 2.68 Å relativ dicht am Eisen-Atom sitzt. Das N8-gebundene H-Atom ist 2.74 Å von der Carbonyl-Gruppe von Leu295 entfernt, also gerade noch in der Reichweite einer H-Brücke, ebenso wie die an C2 sitzende  $\text{NH}_2$ -Gruppe, deren H-Atom einen Abstand von 2.64 Å zur  $\text{CO}_2^-$ -Gruppe von Glu332 aufweist. Die Seitenkette bildet keine H-Brücken zum Protein.  $\text{BH}_4$  wird häufig in ähnlichen Positionen platziert: 15 weitere Vorschläge weisen für die Ringsysteme *rmsd*-Werte von bis zu 1.05 Å zum günstigsten Vorschlag auf ( $\emptyset$ :  $0.58 \pm 0.30$  Å). Die Seitenketten-Positionen sind weniger festgelegt, sie weichen um bis zu 2.89 Å von der besten Konformation ab (im Durchschnitt:  $1.47 \pm 0.85$  Å).

Um auch Komplexe mit der für die TH beobachteten Orientierung des  $\text{BH}_4$ -Moleküls zu erzeugen, wird es auf die Referenz-Position des 7,8- $\text{BH}_2$  in TH platziert. Daraus resultieren 17 weitere Konformationen mit entsprechend niedrigen *rmsd*-Werten von 1.13 – 2.18 Å zur Referenz-Struktur. Sie werden jedoch nach der Kraftfeld-Energie um mindestens 5 kcal/mol, nach der Gesamtenergie sogar um mindestens 10 kcal/mol ungünstiger bewertet als die Komplexe mit der alternativen Orientierung.

Beim günstigsten Vertreter dieser Gruppe beträgt der Neigungswinkel zwischen den Ringebenen von  $\text{BH}_4$  und Phe300 36.3°. Die 2'-OH-Gruppe der  $\text{BH}_4$ -Seitenkette ist durch eine H-Brücke mit dem N-Atom von Leu294 verbunden. Die Carbonyl-Gruppe ist im Abstand von 2.26 Å am Eisen-Atom verankert, das N4-gebundene Proton befindet sich nur 2.24 Å von einem der  $\text{CO}_2^-$ -O-Atome von Asp376 entfernt.

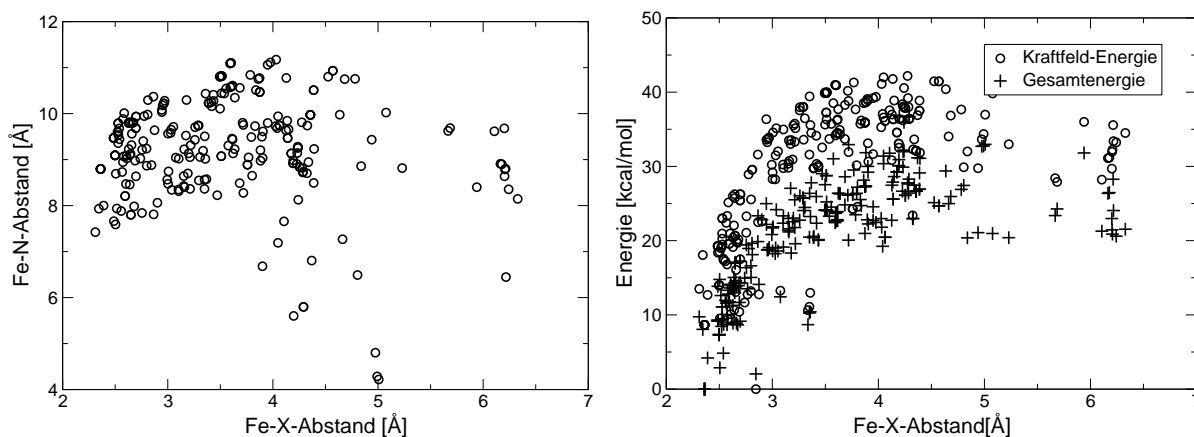
Abbildung 4.28 zeigt die berechneten Alternativen für die  $\text{BH}_4$ -Platzierung im Vergleich mit der Referenz-Position des in der KS gebundenen 7,8- $\text{BH}_2$ . Die Energie-Differenz zwischen der besten berechneten Platzierung und der günstigsten erwarteten Platzierung geht zurück auf die Beiträge für Torsionswinkel, LJ- und Coulomb-Wechselwirkungen, sowie auf die Solvatationsenergien. Torsionswinkel, LJ-Wechselwirkungen und die Solvatationsenergie sind im erwarteten TH-Bindemodus ungünstiger als im PAH-Bindemodus.

### 4.3.2. Vorhersage

#### Catechol-TH-Komplexe

**DA-2toh** FLEXX gibt 256 Vorschläge von DA-Platzierungen in der Bindetasche der TH aus. Es werden nur verzerrt bidentate Strukturen mit Abständen von mindestens 1.94 Å zwischen dem Eisen-Atom und der Catechol-Einheit erzeugt. Bei diesen „bidentaten“ Strukturen vergrößert sich der Eisen-Catechol-Abstand während der Energie-Minimierungen sogar noch: Vor der Struktur-Optimierung weisen 100 Komplexe Eisen-Catechol-Abstände von weniger als 3.0 Å auf, danach nur noch 84. Das Diagramm in Abb. 4.18(a) zeigt, dass die Vorschläge sich über einen Bereich mit Eisen-Catechol-Abständen bis 6.5 Å und Eisen-N-Abständen bis ca. 11 Å verteilen, statt Gruppen verwandter Konformationen zu bilden. Die Strukturen mit kurzen Eisen-Catechol-Abständen werden sowohl nach der Kraftfeld- als auch nach der Gesamtenergie günstiger bewertet als die übrigen Vorschläge (s. Abb. 4.18(b)).

Sieben FLEXX-Vorschläge konvergieren in das gleiche, nach der Gesamtenergie bevorzugte Energie-Minimum mit verzerrt bidentater Koordination. Dabei sitzt die *meta*-OH-Gruppe im Abstand von 2.14 Å zum Eisen-Atom *trans* zu His336 und die *para*-OH-Gruppe im Abstand von 3.23 Å *trans* zu His331. Die NH<sub>3</sub><sup>+</sup>-Seitenkette bildet keine Kontakte zum Protein aus. Die Catechol-Gruppen aller bi- bzw. monodentaten Konformationen zeigen bezüglich dieser Platzierung *rmsd*-Werte von 0.78±0.71 Å, während die Seitenketten im Durchschnitt um 2.81±1.59 Å davon abweichen (s. Abb. 4.29).



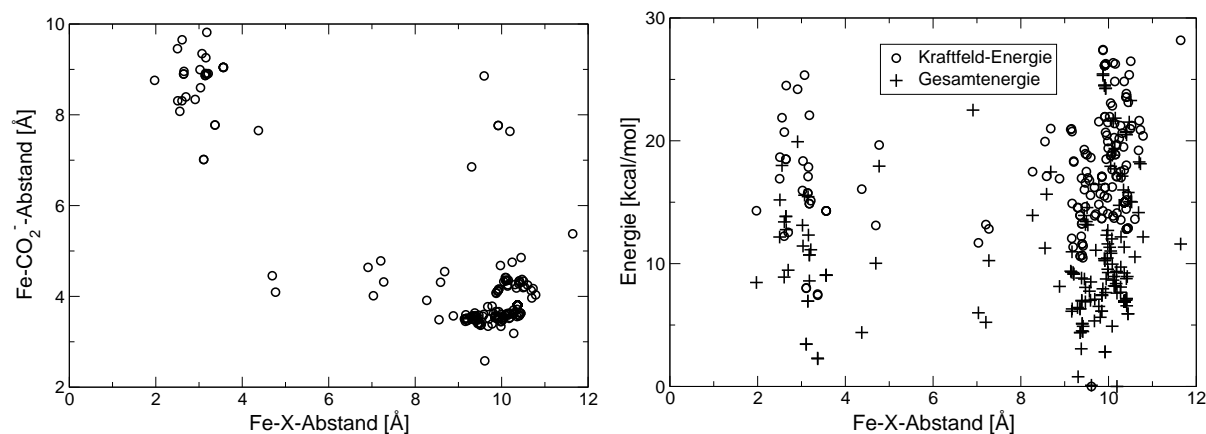
(a) Überblick über die Vielfalt der Konformationen: Das DA-Molekül wird in sehr verschiedenen Positionen platziert, es entstehen keine unterscheidbaren Gruppen von Konformationen mit ähnlichem Bindemodus.

(b) Bewertung der Komplex-Konformationen nach dem Abstand der Catechol-Gruppe zum Eisen und nach der Kraftfeld- bzw. Gesamtenergie.

*Abbildung 4.18.*: Docking von DA in 2toh: Strukturen mit kurzen Eisen-Catechol-Abständen werden nach beiden Energie-Kriterien gegenüber solchen mit größerem Abstand bevorzugt.

**L-DOPA·2toh** Beim Docking von L-DOPA in die KS 2toh werden 165 Struktur-Vorschläge erzeugt, von denen allerdings keine einzige einen Eisen-Catechol-Abstand von weniger als 1.9 Å aufweist. 29 FLEXX-Vorschläge sind mit Eisen-Catechol-Abständen von bis zu 3.0 Å noch als verzerrt bidentat bzw. monodentat anzusehen, *nach* der Struktur-Optimierung besitzen allerdings nur noch 11 Komplexe Abstände von weniger als 3.0 Å (s. Abb. 4.19(a)). Der Hauptanteil der Struktur-Vorschläge entfällt wiederum auf Komplexe, bei denen die CO<sub>2</sub><sup>-</sup>-Gruppe von L-DOPA an das Eisen-Atom bindet. Wie bei den L-DOPA·PAH-Komplexen wird auch bei der TH diese letztere Gruppe sowohl nach der Gesamt- als auch nach der Kraftfeld-Energie als die günstigere Variante bewertet (s. Abb. 4.19(b)).

Im einzigen, vom Bindemodus her in Betracht kommenden Komplex sitzt die *meta*-OH-Gruppe im Abstand von 2.1 Å *trans* zu His331, die *para*-OH-Gruppe *trans* zu His336 mit einer gestreckten Bindung von 2.6 Å Länge (s. Abb. 4.29). Die NH<sub>3</sub><sup>+</sup>-Seitenkette bildet eine H-Brücke zu einem der beibehaltenen expliziten Wasser-Moleküle und die CO<sub>2</sub><sup>-</sup>-Gruppe befindet sich ungünstig nah neben dem O-Atom von Leu295 (3.1 Å).



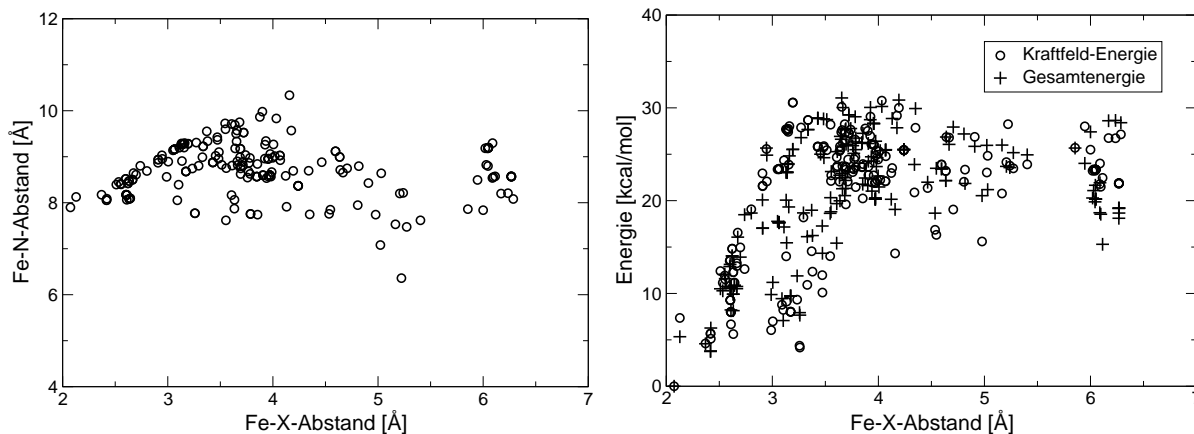
(a) Überblick über die Vielfalt der Konformationen: Das L-DOPA-Molekül wird meist entweder über die OH-Gruppen oder über die CO<sub>2</sub><sup>-</sup>-Gruppe am Eisen-Atom verankert.

(b) Bewertung der Komplex-Konformationen nach dem Abstand der Catechol-Gruppe zum Eisen und nach der Kraftfeld- bzw. Gesamtenergie.

*Abbildung 4.19.:* Docking von L-DOPA in 2toh: Die CO<sub>2</sub><sup>-</sup>-Eisen-Bindung wird gegenüber der Catechol-Eisen-Bindung nach der Kraftfeld-Energie leicht, nach der Gesamtenergie deutlicher bevorzugt.

**MDTIQ·2toh** Für MDTIQ in der Bindetasche der TH werden 206 alternative Platzierungen erzeugt, von denen 38 Eisen-Catechol-Abstände von weniger als 3.0 Å haben. Konformationen, bei denen das N-Atom, wie z. B. bei L-DOPA, in der Nähe des Eisen-Atoms sitzt, kommen nicht vor (s. Abb. 4.20(a)). Die Konformation mit dem kleinsten Eisen-Catechol-Abstand (2.07 Å) wird sowohl nach der Kraftfeld- als auch nach der Gesamtenergie als

die günstigste aller Strukturen bewertet (s. Abb. 4.20(b)). Die OH-Gruppen des Moleküls sind in diesem Fall 2.13 und 2.78 Å vom Eisen entfernt, dabei ist die Bindung *trans* zu His336 gestreckt. Das positiv geladene N-Atom geht jedoch keine Wechselwirkungen mit den Seitenketten des Proteins ein (s. Abb. 4.29).



(a) Überblick über die Vielfalt der Konformationen: Das MDTIQ-Molekül wird in vielen, schlecht voneinander abgrenzbaren Varianten platziert.

(b) Bewertung der Komplex-Konformationen nach dem Abstand der Catechol-Gruppe zum Eisen-Atom und nach der Kraftfeld- bzw. Gesamtenergie.

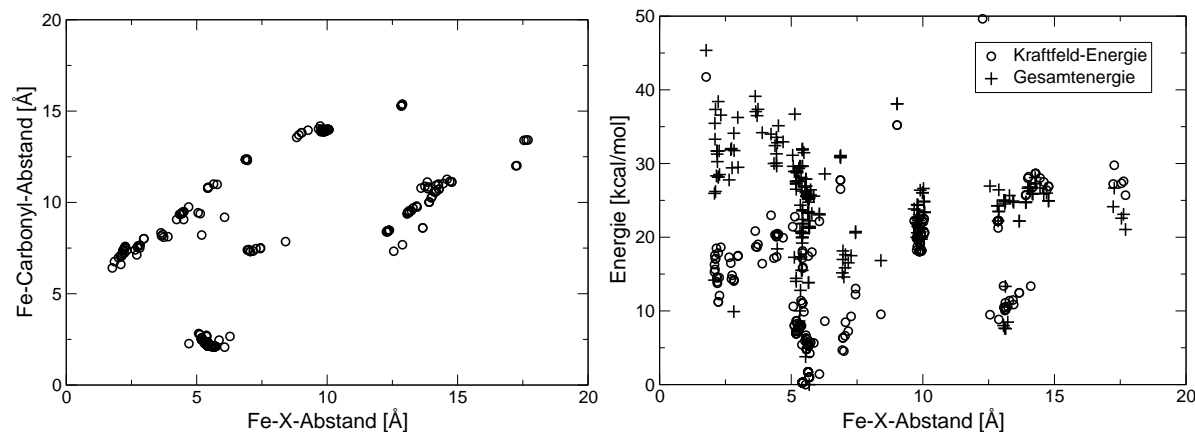
Abbildung 4.20.: Docking von MDTIQ in 2toh: Nach beiden Energie-Kriterien wird ein bidentater Bindemodus als günstigste Struktur bewertet.

**Rotenon-2toh** Für das große Rotenon-Molekül gehen beim Docking 317 verschiedene Positionen innerhalb der Bindetasche hervor. Das Diagramm von Abbildung 4.21(a) zeigt, dass sich jeweils mehrere Vorschläge zu Gruppen mit ähnlichem Bindemodus sammeln. 30 Konformationen besitzen die gesuchten Eisen-Catechol-Abstände von weniger als 3.0 Å. Die nach der Gesamtenergie günstigste Konformation dieser Gruppe mit annähernd bidentatem Bindemodus weist zwei gestreckte Eisen-Sauerstoff-Bindungen von 2.4 bzw. 2.5 Å Länge auf; das weitgehend planare und starre Molekül ragt dabei ein wenig aus der Bindetasche heraus (s. Abb. 4.29). Es hat zu den AS Leu294, Phe300, Phe309, Pro327, His331, Tyr371, Trp372 und Tyr423 überwiegend hydrophobe Kontakte.

Zwei alternative Bindemodi mit mehr als 3.0 Å Abstand zwischen Catechol-Gruppe und Eisen-Atom sind energetisch günstiger als der bidentate Bindemodus (s. Abb. 4.21(b)): Bei der ersten Alternative ist das Rotenon über seine Carbonyl-Gruppe *trans* zu His 336 am Eisen-Atom gebunden, sodass das unpolare Molekül tief in der Bindetasche liegt und sogar von Asp425 und Gln426 bedeckt wird<sup>13</sup>. Bei der zweiten Alternative, die nur 2.5 kcal/mol

<sup>13</sup>Dieser Umstand stellt die tatsächliche Relevanz dieses Bindemodus in Frage, denn das Docking mit FLEXX berücksichtigt nur den in der Bindetasche zur Verfügung stehenden Raum, nicht aber, ob der Zugang zur Bindetasche möglicherweise versperrt ist.





(a) Überblick über die Vielfalt der Konformationen: Das Rotenon-Molekül wird in mehreren, voneinander abgrenzbaren Varianten platziert.

(b) Bewertung der Komplex-Konformationen nach dem Abstand der Catechol-Gruppe zum Eisen-Atom und nach der Kraftfeld- bzw. Gesamtenergie.

*Abbildung 4.21.*: Docking von Rotenon in 2toh: Nach den Energie-Kriterien wird ein bidentater Bindemodus ungünstiger bewertet als eine Verankerung über die Carbonyl-Gruppe oder eine der bidentaten Variante entgegengesetzt orientierte Platzierung.

günstiger ist als die bidentate Variante, liegt der Dihydrofuran-Ring in der Nähe des Eisen-Atoms, allerdings ohne Kontakt des O-Atoms zum Eisen. Trotz der sechs potentiellen H-Brücken-Akzeptoren wird keine einzige H-Brücke zum Protein ausgebildet, es herrschen wie bei den anderen Platzierungen unpolare Wechselwirkungen vor.

### Substrat-Komplexe mit PAH bzw. TH

Zusätzlich zum zuvor in der PAH bzw. TH platzierten  $\text{BH}_4$  soll das zweite Substrat, eine der AS Phe oder Tyr, in die Bindetasche eingepasst werden. Zur Untersuchung der Substrat-Spezifitäten beider Enzyme werden auch entsprechende Komplexe mit vertauschten AS erzeugt. Um bei der TH ebenfalls einen möglichen Einfluss der Cosubstrat-Orientierung zu studieren, wird  $\text{BH}_4$  zuerst im TH-Bindemodus eingesetzt, nach erfolgter AS-Platzierung aber (parallel) auch im PAH-Bindemodus zur Energie-Minimierung verwendet.

**Tyr- $\text{BH}_4$ -2toh** Zum Docking des natürlichen Substrats Tyr in den binären Komplex der TH mit  $\text{BH}_4$  in TH-Orientierung wird die KS 2toh unter Beibehaltung des *trans* zu Glu376 gebundenen Wasser-Moleküls verwendet. Die Vorschläge werden mit  $\text{BH}_4$  sowohl im TH-Bindemodus, als auch im PAH-Bindemodus optimiert. Die energetische Reihenfolge der Platzierungen ist (fast) unabhängig von der Orientierung des Cosubstrates, wie der Korrelationskoeffizient von 0.95 für die Gesamtenergien aller Komplexe mit alternativen  $\text{BH}_4$ -Orientierungen zeigt. Von den jeweils 318 Strukturen sind 170 potentiell relevant, da

sie Abstände des Ring-Zentrums zum Eisen-Atom zwischen 4.7–8.0 Å besitzen (s. Abb. 4.23(a)). Die übrigen Vorschläge liegen mit bis zu 16.7 Å für einen plausiblen Substrat-Komplex zu weit vom aktiven Zentrum entfernt. Die relevanten Vorschläge bilden drei Gruppen mit jeweils eigenem Bindemodus (s. Abb. 4.22).

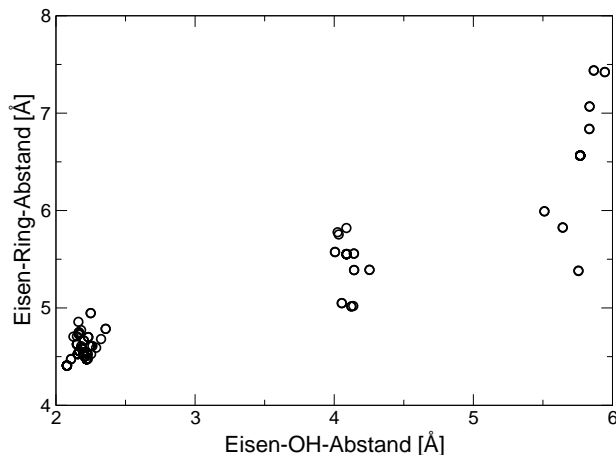


Abbildung 4.22: Docking von Tyr in TH: Die relevanten Platzierungen mit Eisen-Ring-Abständen von weniger als 8.0 Å bilden drei voneinander gut abgrenzbare Gruppen von Konformationen.

- Die energetisch günstigste Variante ist die zu erwartende Koordination der OH-Gruppe zum Eisen. 116 Konformationen zeigen entsprechende Abstände von 2.08–2.36 Å. Bei der Orientierung des  $\text{BH}_4$  im TH-Modus bildet der beste Vertreter mit einem Eisen-OH-Abstand von 2.16 Å eine H-Brücke der  $\text{NH}_3^+$ -Gruppe zur  $\text{CO}_2^-$ -Gruppe von Asp425.

Die günstigste Tyr-Platzierung mit  $\text{BH}_4$  in PAH-Orientierung ist mit 2.09 Å etwas dichter am Eisen-Atom gebunden, während die  $\text{CO}_2^-$ -Gruppe von Asp425 von der  $\text{NH}_3^+$ - und der  $\text{CO}_2^-$ -Gruppe flankiert wird.

Die Ringebenen in der TH- bzw. PAH-Variante stehen etwa senkrecht zu einander. Insgesamt sind die aromatischen Ringe bei OH-Koordination mit einer *rmsd* von durchschnittlich 2.38 Å recht einheitlich platziert.

Abhängig von der Orientierung des Cosubstrates  $\text{BH}_4$  unterscheiden sich die Komplexe um fast 10 kcal/mol in der Gesamtenergie: Der Komplex mit  $\text{BH}_4$  im PAH-Bindemodus ist (wie auch ohne AS) günstiger.

- Die nächste Gruppe von Konformationen ist um mindestens 14.8 bzw. 18.7 kcal/mol ( $\text{BH}_4$  in TH- bzw. PAH-Bindemodus) ungünstiger als die vorherige. Bei dem – unabhängig von der Cosubstrat-Orientierung – besten Vertreter liegt der Tyrosin-Ring (im Abstand von 4.95 Å zum Eisen) etwa parallel zum Imidazol-Ring von His331: Der Abstand der Ring-Zentren beträgt 4.0 Å und die Ringebenen sind um ca. 22° gegeneinander geneigt. Die anionische  $\text{CO}_2^-$ -Gruppe liegt zwischen der kationischen Guanidinium-Gruppe von Arg316 und der Seitenkette von Ser396. Die positiv geladene  $\text{NH}_3^+$ -Gruppe wird über eine H-Brücke zum Carbonyl-O-Atom von Ser324 am

Protein verankert, während die  $\text{CO}_2^-$ -Gruppe mit dem N-Atom von Ser324 ebenfalls eine H-Brücke bildet.

Die Ring-Positionen von weiteren 23 Platzierungen mit ähnlichen Eisen-Ring-Abständen weichen von der beschriebenen um durchschnittlich  $1.67 \pm 0.51 \text{ \AA}$ , die Seitenketten-Positionen jedoch um  $4.35 \pm 2.41 \text{ \AA}$  ab. Damit wird der aromatische Ring wieder recht einheitlich platziert, die Ringebene liegt jedoch nicht mehr parallel, sondern fast senkrecht zur Ringebene von His331. Nach der Gesamtenergie wird wiederum der Komplex mit dem PAH-Bindemodus des  $\text{BH}_4$  um  $5.5 \text{ kcal/mol}$  vor demjenigen mit TH-Bindemodus bevorzugt.

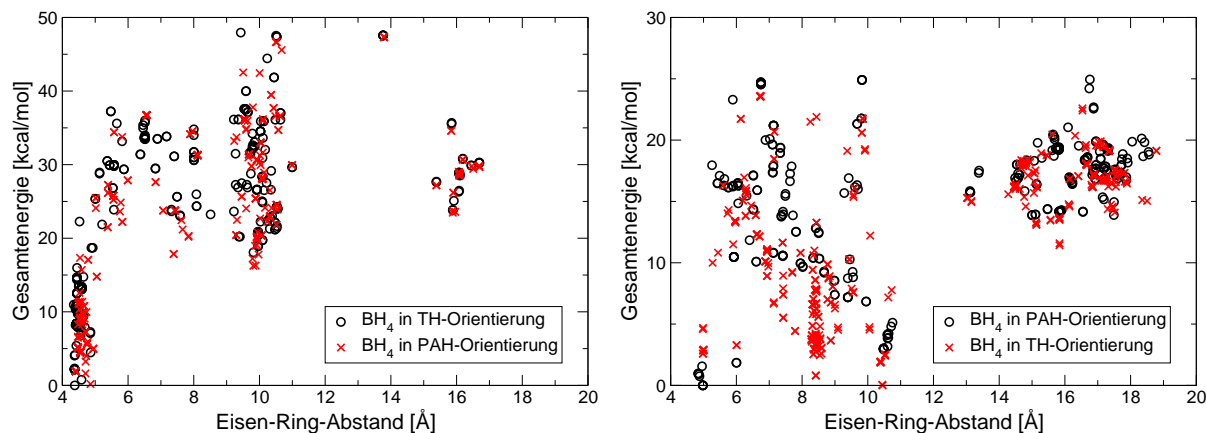
- Der Bindemodus der dritten, relativ zur OH-Koordination energetisch um mindestens  $21.5$  bzw.  $26.0 \text{ kcal/mol}$  höheren Gruppe von 45 Platzierungen ist weniger einheitlich als bei den vorherigen. Bei dem besten Vertreter mit H-Bindemodus ist die  $\text{CO}_2^-$ -Gruppe wiederum über Arg316 verankert, jedoch steht der aromatische Ring der AS senkrecht, sowohl zur Ebene des  $\text{BH}_4$ -Moleküls, als auch zur Seitenkette von His331, d. h. er weist mit seiner Kante zum Eisen-Atom. Der Abstand des Rings zum Eisen ist mit  $5.6 \text{ \AA}$  größer als zuvor. Zwischen der  $\text{NH}_3^+$ -Gruppe des Tyrosins und der Carbonyl-Gruppe von Gly392 existiert eine H-Brücke, ebenso zwischen der  $\text{CO}_2^-$ -Gruppe und der Guanidinium-Gruppe von Arg316.

Beim wiederum um ca.  $5 \text{ kcal/mol}$  günstigeren Komplex mit  $\text{BH}_4$  im PAH-Bindemodus wird eine Tyr-Platzierung bevorzugt, bei der der Abstand des Ringzentrums zum Eisen  $7.1 \text{ \AA}$  beträgt. Die phenolische OH-Gruppe des Tyrosins bildet eine H-Brücke zur  $\text{CO}_2^-$ -Gruppe von Asp425, und seine  $\text{CO}_2^-$ -Gruppe eine H-Brücke zur Seitenkette von Ser296 aus. Gleichzeitig hat die  $\text{NH}_3^+$ -Gruppe Kontakt zum Carbonyl-Atom von Ser324.

**Phe· $\text{BH}_4$ ·2toh** Statt des eigentlichen Substrates Tyrosin wird auch die Vorstufe Phe in die schon  $\text{BH}_4$  enthaltende Bindetasche der TH platziert. Die Oktaeder-Ecken *trans* zu His 336 und *trans* zu Glu376 sind von Wasser-Molekülen besetzt, das  $\text{BH}_4$ -Molekül liegt in der TH-Orientierung vor. Passende Eisen-Ring-Abstände von maximal  $8.0 \text{ \AA}$  weisen nur 89 der insgesamt 379 Vorschläge auf (s. Abb. 4.23(b)). Nach Austausch des  $\text{BH}_4$  in TH-Orientierung gegen PAH-Orientierung (s. S. 106) kollidieren einige Phe-Platzierungen mit der neuen  $\text{BH}_4$ -Orientierung. Abgesehen von diesen Fällen bleibt die energetische Bewertung weitgehend erhalten (Korrelationskoeffizient: 0.87); generell sind die Komplexe mit PAH-Orientierungen günstiger als solche mit  $\text{BH}_4$  in TH-Orientierung. Folgende Varianten der Platzierung kommen aufgrund des Eisen-Ring-Abstandes in Betracht:

- In der (vom  $\text{BH}_4$ -Bindemodus unabhängig) günstigsten Gruppe von Konformationen (11 Vertreter mit max.  $1.2 \text{ \AA}$  *rmsd* zwischen den Ring-Positionen und Eisen-Ring-Abständen von  $5.0$ – $6.0 \text{ \AA}$ ) ist die  $\text{CO}_2^-$ -Gruppe durch Arg316 verankert, während die  $\text{NH}_3^+$ -Gruppe eine H-Brücke zum Carbonyl-Atom von Ser324 bildet. Die Fläche des Phenylrings ist dem Eisen-Atom zugewandt, sodass das para-C-Atom direkten Kontakt zum Wasser-Molekül *trans* zu His336 hat.

- Mehrere Platzierungen (25) flankieren mit der  $\text{NH}_3^+$ -Gruppe die eine oder andere Seite der  $\text{CO}_2^-$ -Gruppe von Asp425 (in einem Abstand von 3.0–4.0 Å). Der Phenylrest ist jeweils unterschiedlich orientiert. Ihre relativen Energie-Werte liegen zwischen 5.0–22.0 kcal/mol.
- Weitere 18 Konformationen im Energie-Bereich von 6.7–13.2 kcal/mol sind über die  $\text{CO}_2^-$ -Gruppe am Eisen-Atom verankert, während der Phenylrest in verschiedene Richtungen weist.
- Bei den restlichen 35 Konformationen liegt der Phenylring in der Mitte der Bindetasche, nahe bei den aromatischen Seitenketten von Phe377 und Trp 372, wohingegen die Seitenketten stark voneinander abweichend platziert sind. Die korrespondierenden Gesamtenergien betragen 4.4–23.5 kcal/mol.



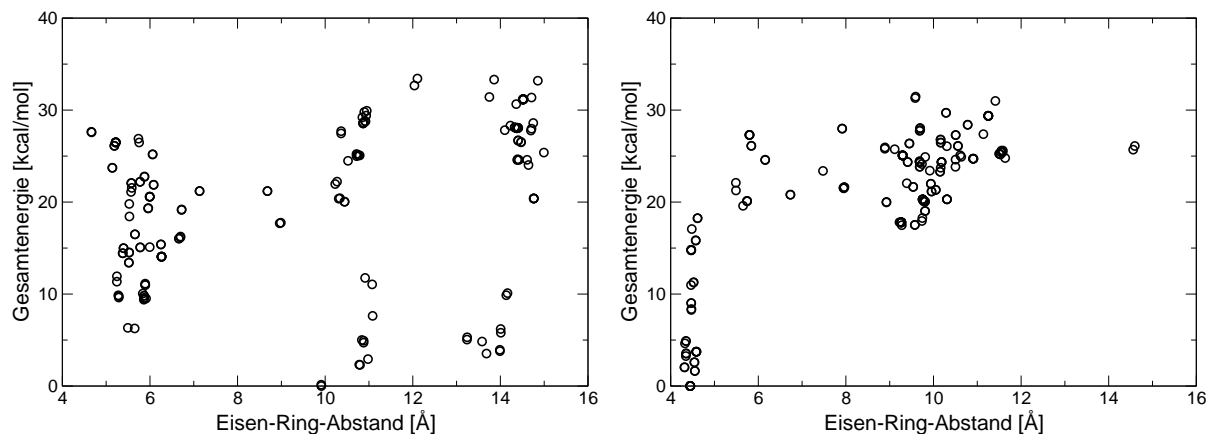
(a) Tyr·BH<sub>4</sub>·2toh: Bewertung der AS-Platzierungen nach dem Abstand des Ring-Zentrums zum Eisen-Atom und nach der Gesamtenergie.

(b) Phe·BH<sub>4</sub>·2toh: Bewertung der AS-Platzierungen nach dem Abstand des Ring-Zentrums zum Eisen-Atom und nach der Gesamtenergie.

Abbildung 4.23.: Substrat-Komplexe der TH: Bei Tyr wird die Koordination der OH-Gruppe an das Eisen-Atom bevorzugt. Phe nimmt verschiedene Positionen ein.

**Phe·BH<sub>4</sub>·5pah** Wird das natürliche Substrat der PAH in ihre mit BH<sub>4</sub> und zwei Wasser-Molekülen besetzte Bindetasche gedockt, resultieren 179 Vorschläge. BH<sub>4</sub> liegt dabei im berechneten PAH-Bindemodus vor, die Wasser-Moleküle befinden sich an den Oktaeder-Ecken *trans* zu His290 und *trans* zu Glu330. 80 Strukturen weisen nach der Optimierung Eisen-Ring-Abstände von weniger als 8.0 Å auf (s. Abb. 4.24(a)).

Die nach der Gesamtenergie günstigste Konformation aus dieser Gruppe zeigt eine enge Nachbarschaft zwischen den aromatischen Ringen der AS und der Seitenkette von His285: Die Ringzentren liegen nur 3.78 Å auseinander und die Ringebenen sind um knapp 10°



(a) Phe·BH<sub>4</sub>·5pah: Bewertung der AS-Platzierungen nach dem Abstand des Ring-Zentrums zum Eisen-Atom und nach der Gesamtenergie.

(b) Tyr·BH<sub>4</sub>·5pah: Bewertung der AS-Platzierungen nach dem Abstand des Ring-Zentrums zum Eisen-Atom und nach der Gesamtenergie.

**Abbildung 4.24.:** Substrat-Komplexe der PAH: Bei Phe liegen die günstigsten Platzierungen außerhalb des aktiven Zentrums in der Bindetasche der PAH. Trotz der großen Energie-Spanne repräsentiert die relevante Gruppe mit weniger als 6.5 Å Eisen-Ring-Abstand einen einheitlichen Bindemodus. Bei Tyr wird die direkte Koordination der OH-Gruppe an das Eisen-Atom bevorzugt.

gekippt. Die NH<sub>3</sub><sup>+</sup>- und die CO<sub>2</sub><sup>-</sup>-Gruppe von Phe sind über H-Brücken mit den Gegengruppen von Thr278 verknüpft. Das C-Atom der CO<sub>2</sub><sup>-</sup>-Gruppe ist 6.3 Å vom zentralen C-Atom der Guanidinium-Gruppe von Arg270 entfernt. Bezüglich dieser Konformation sind zumindest die Ringe von 57 weiteren Vorschlägen mit bis zu 1.82 Å *rmsd*-Abweichung sehr ähnlich platziert ( $\bar{\sigma}$ : 1.42 ± 0.78 Å). Die um mehr als 2.0 Å von der Referenz-Platzierung abweichenden Vorschläge sind mindestens 20 kcal/mol ungünstiger und werden daher vernachlässigt.

**Tyr·BH<sub>4</sub>·5pah** Bei der Platzierung ihres Produktes (Tyr) in der Bindetasche von PAH werden 179 Konformationen erzeugt (s. Abb. 4.24(b)). Die Gruppe der 65 relevanten Konformationen mit Eisen-Ring-Distanzen von weniger als 8.0 Å sind entweder über die OH-Gruppe oder über die CO<sub>2</sub><sup>-</sup>-Gruppe am Eisen-Atom oder über die NH<sub>3</sub><sup>+</sup>-Gruppe in der Peripherie der Bindetasche verankert.

- In der sowohl nach der Gesamt-, als auch Kraftfeld-Energie weitaus günstigeren Gruppe ist das Tyrosin-Molekül über die OH-Gruppe gebunden: Beim besten der insgesamt 41 Vertreter ist das O-Atom nur 2.08 Å vom Eisen-Atom entfernt, sodass das daran benachbarte H-Atom simultan eine H-Brücke zum Carbonyl-O-Atom des BH<sub>4</sub> bilden kann. Zusätzlich ist die NH<sub>3</sub><sup>+</sup>-Gruppe über eine H-Brücke mit dem Carbonyl-O-Atom von Pro279 verknüpft.

- Zwanzig andere, um mindestens 19.6 kcal/mol ungünstigere Platzierungen zeigen, dass auch die  $\text{CO}_2^-$ -Gruppe im Abstand von 2.5–3.4 Å an das Eisen-Atom binden kann. Bei der günstigsten Variante bildet die  $\text{NH}_3^+$ -Gruppe eine H-Brücke zur C1'-OH-Gruppe des  $\text{BH}_4$ -Moleküls. Die Anziehung zwischen der  $\text{CO}_2^-$ -Gruppe und dem zweiwertigen Eisen ist zwar sehr vorteilhaft, gleichzeitig besteht aber auch eine starke Abstoßung durch Glu330.
- Einige wenige Vorschläge sind über die  $\text{NH}_3^+$ -Gruppe an Glu353 in der Peripherie der Bindetasche verankert. Der Phenyl-Ring ist weiterhin zum Eisen gerichtet, das Ring-Zentrum ist aber mit ca. 7.0 Å schon recht weit vom Eisen entfernt. Auch diese Gruppe von Konformationen ist um mindestens 20.8 kcal/mol ungünstiger als die Koordination über die OH-Gruppe.

Eine Platzierung des Tyrosins, die der bevorzugten Position des Phenylalanins in PAH etwa entspricht, taucht nicht auf.<sup>14</sup>

## 4.4. Diskussion

### 4.4.1. Validierung

#### Catechol-PAH-Komplexe

**Strukturen** In allen Fällen werden beim Docking von DA und L-DOPA in 5pah bzw. 6pah Konformationen erzeugt, die nach Energie-Minimierung mit den Referenz-Strukturen aus den KS sehr gut übereinstimmen (s. Abb. 4.25). Die *rmsd*-Werte zwischen den Catechol-Ringen betragen 0.28–0.91 Å bezüglich der Catechol-Einheit von L-DOPA aus der KS 6pah (die Catechol-Einheiten der beiden KS 5pah und 6pah weichen um nur 0.26 Å voneinander ab). Die Ergebnisse sind unabhängig von den geringfügigen Unterschieden der Protein-Konformationen (mittlere *rmsd*-Abweichung zwischen den Protein-Anteilen der KS 5pah und 6pah: nur 0.23 Å).

Viele energetisch äquivalente Seitenketten-Konformationen werden jeweils für DA berechnet; ein Befund, der sich experimentell darin ausdrückt, dass die Position der Seitenkette anhand der Elektronendichte nicht bestimmbar ist [34]. Für L-DOPA dagegen scheint die Position der Seitenkette nach den Berechnungen eindeutiger festgelegt zu sein, als kristallographisch feststellbar ist [34].

**FLEXX-Bewertung** Die Reihenfolge, in der die Struktur-Vorschläge ausgegeben werden, entspricht der Einschätzung von FLEXX, wie zutreffend die jeweilige Platzierung ist. Im Idealfall sollten unter den ersten Vorschlägen gute Ligand-Positionen vertreten sein. Als

<sup>14</sup>Ein einziger, allerdings um 20.8 kcal/mol ungünstiger und um durchschnittlich 2.92 Å abweichender Vorschlag kommt der Phe-Platzierung einigermaßen nahe. Dabei steht der aromatische Ring jedoch senkrecht zur Vergleichsstruktur und auch die Seitenkette wird durch andere Gruppen fixiert: Die  $\text{NH}_3^+$ -Gruppe bildet H-Brücken zur Seitenkette von Glu353 und zum Carbonyl-O-Atom von Gly346.

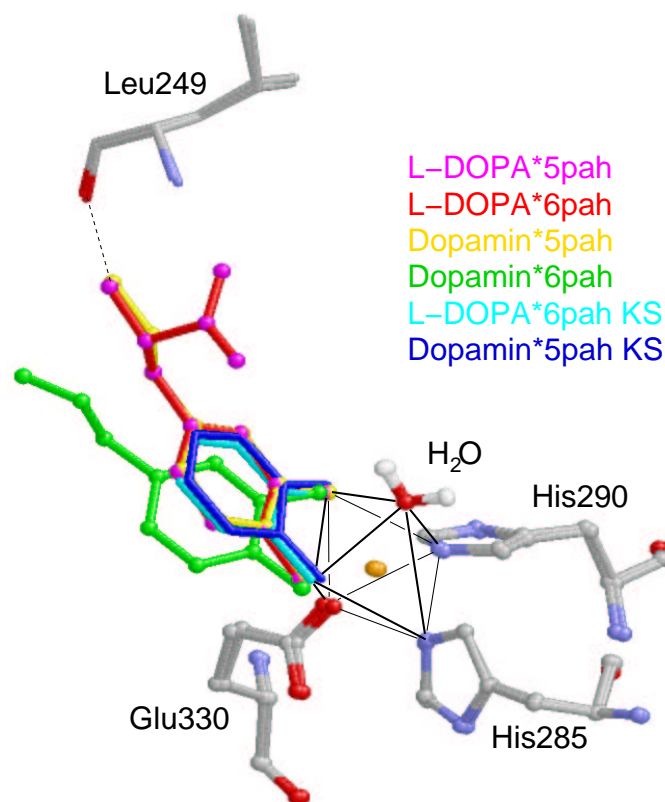
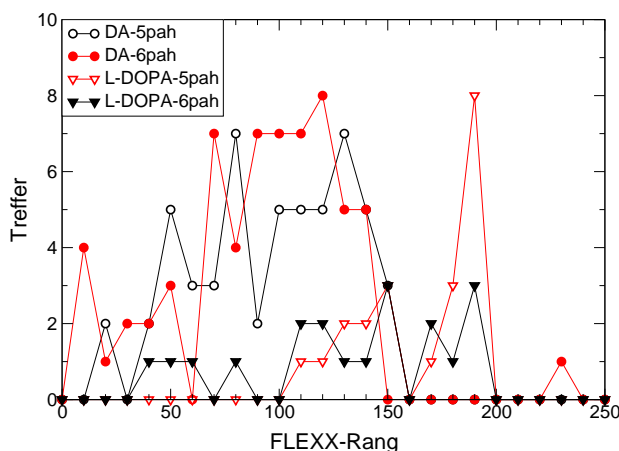


Abbildung 4.25: Die besten Docking-Ergebnisse im Vergleich mit den Catechol-Einheiten aus den KS 5pah bzw. 6pah. Die Positionen der Catechol-Einheiten decken sich sehr gut. Die erste Liganden-Sphäre bildet einen sehr regelmäßigen Oktaeder um das Eisen-Atom. H-Atome sind (außer bei dem Wasser-Molekül) nicht dargestellt.

genau zutreffend werden Platzierungen betrachtet, die *rmsd*-Werte von weniger als 1.0 Å besitzen; bis ca. 2.0 Å ist der Bindemodus noch relativ gut wieder gegeben. Abbildung 4.26 zeigt für das Docking und Redocking von DA bzw. L-DOPA die Häufigkeit von Treffern für Gruppen von jeweils zehn Vorschlägen. Von der Menge der FLEXX-Vorschläge kommen allerdings jeweils nur 10–28% als sinnvolle Konformationen tatsächlich in Betracht und zudem tauchen diese „guten“ Platzierungen teilweise erst spät in der Liste der Vorschläge auf. Das bedeutet, die FLEXX-interne Bewertung der Catechol-PAH-Komplexe ist fehlerhaft und eine energetische Neubewertung in der Tat sinnvoll, um ein zuverlässigeres Kriterium für die Auswahl der relevanten Konformationen zu erhalten.

**Energie-Bewertung** Es zeigt sich, dass bei DA und L-DOPA die Bewertung nach der Gesamtenergie ebenfalls problematisch ist, da selbst mit diesem Kriterium der richtige Bindemodus nicht sicher von falschen Platzierungen unterschieden werden kann. Im Fall des DA ist die Beurteilung der Komplexe nach der Kraftfeld-Energie, also ohne Berücksichtigung der Solvatationsenergie, deutlich besser geeignet, um relevante Konformationen zu identifizieren. Dies geht vermutlich zurück auf den Gewinn an Solvatationsenergie bei den „falschen“ Platzierungen, wenn das Eisen-Atom nicht durch andere Liganden von der wässrigen Umgebung abgeschirmt wird. Im Fall des L-DOPA wird das Eisen-Atom zwar auch durch die CO<sub>2</sub><sup>-</sup>-Gruppe vom Solvens abgeschirmt, die Coulomb-Wechselwirkungen zwischen dieser anionischen Gruppe und dem Eisen-Atom überwiegen jedoch die entspre-

Abbildung 4.26: Die Reihenfolge der Vorschläge macht Aussagen über die interne Bewertung von Konformationen durch FLEXX: Entgegen der Erwartung (viele Treffer mit niedrigem Rang, wenige Treffer bei den höheren Rängen) tauchen die plausiblen Strukturen mit Eisen-Catechol-Abständen von weniger als 3.0 Å häufig erst bei hohen Rängen auf.



chenden Anziehungen der OH-Gruppen (bei vergleichbaren Eisen-O-Abständen von ca. 2.2 Å im Fall der  $\text{CO}_2^-$ -Gruppe bzw. ca. 2.1 Å bei den OH-Gruppen betragen die Partial-Ladungen  $-0.57 \text{ C}$  für jedes  $\text{CO}_2^-$ -O-Atom und nur  $-0.25 \text{ C}^{15}$  für jede OH-Gruppe).

Möglicherweise ist die Diskrepanz zwischen den erwarteten und berechneten Ergebnissen eine Folge der gewählten Ladungsverteilungen innerhalb des Eisen-Komplexes. Das nach dem HSAB-Prinzip (*hard/soft acids/bases* von Pearson [108]) „weiche“ Eisen-Atom bildet im Gegensatz zu den „harten“ Kationen der ersten beiden Hauptgruppen-Elemente auch kovalente Bindungen zu seinen Liganden aus. Es ist in der Lage, in gewissem Umfang Elektronen-Dichte aufzunehmen, d. h. die angenommene Ladung von  $+2.0 \text{ C}$  für das Eisen-Atom ist möglicherweise unrealistisch hoch. Für den gesamten Komplex wäre also eine exaktere, quantenmechanische Berechnung der Ladungsverteilung angemessen. Bei einem niedrigeren Betrag der Punktladung am Eisen ist mit einer Verringerung des Gewinns der Solvatationsenergie bei freien Koordinationsstellen und auch mit einer Verringerung der attraktiven Wechselwirkungen zur  $\text{CO}_2^-$ -Gruppe, allerdings auch zu den OH-Gruppen, zu rechnen. Um die Bindungsverhältnisse innerhalb von Übergangsmetall-Komplexen richtig zu beschreiben, wäre zusätzlich noch die Erfassung kovalenter Bindungsanteile zu den Liganden notwendig. Für solche Wechselwirkungen sind im Rahmen des CHARMM-Kraftfeldes jedoch keine validierten Parameter verfügbar. Zur Auswahl der korrekten Ligand-Platzierungen sind also unbedingt Zusatzinformationen notwendig. In diesem Fall genügt glücklicherweise das in Abschnitt 13 definierte Abstandskriterium, um die richtige Gruppe von Konformationen auszuwählen, innerhalb derer der Kandidat mit der niedrigsten Gesamtenergie zu suchen ist.

#### **BH<sub>4</sub>-Komplexe mit PAH bzw. TH**

**Strukturen** Kristallographisch sind für 7,8-BH<sub>2</sub> unterschiedliche Bindemodi in den Bindetaschen von TH bzw. PAH belegt [53, 33]. In beiden Fällen besteht kein unmittelbarer Kontakt des BH<sub>4</sub>-Moleküls zum Eisen-Atom. Die automatischen Rechnungen liefern für

<sup>15</sup> C (Coulomb) ist die Einheit für Ladungen. Die Elementar-Ladung eines Protons oder Elektrons beträgt  $+1$  bzw.  $-1 \text{ C}$ .



beide Enzyme Platzierungen, die dem PAH-Bindemodus entsprechen. Die *rmsd*-Werte zur Referenz-Position des 7,8-BH<sub>2</sub> in der KS 1dmw für BH<sub>4</sub> in PAH bzw. TH betragen 2.1 bzw. 2.3 Å, decken sich also recht gut mit dem experimentellen Befund (s. Abb. 4.27). Die beste Platzierung von BH<sub>4</sub> in TH (im PAH-Bindemodus) stimmt strukturell auch mit dem von Almås et al. [3] berechneten Ensemble von BH<sub>4</sub>-Konformationen in der Bindetasche der TH (KS: 1toh) gut überein (durchschnittliche *rmsd*-Werte bei 30 Konformationen:  $1.75 \pm 0.25$  Å; die Platzierungen unterscheiden sich vor allem in den Konformationen der Seitenkette).

**FLEXX-Bewertung** Beim automatischen Docking werden sowohl für die PAH, als auch für die TH u. a. Konformationen erzeugt, die einheitlich dem PAH-Bindemodus entsprechen. Diese Konformationen sind bereits unter den ersten FLEXX-Vorschlägen; die FLEXX-interne Bewertung von Konformationen ist also besser auf Wechselwirkungen zwischen organischen Verbindungen, d. h. reine Protein-Ligand-Wechselwirkungen ohne Kontakt über Schwermetall-Ionen, abgestimmt.

Dass der TH-Bindemodus nicht unter den FLEXX-Vorschlägen auftaucht, kann entweder daher rühren, dass die Platzierung der ersten Fragmente nach der Bewertungsfunktion so ungünstig ist, dass der weitere Aufbau des vollständigen Moleküls in der Bindetasche frühzeitig abgebrochen wird, oder daher, dass für die Platzierung des Basis-Fragments in der gewünschten Orientierung keine weiteren geeigneten Wechselwirkungspartner zur Verfügung stehen. In der Tat ist das besondere Merkmal beider Bindemodi ein so genannter  $\pi$ -Stapel, d. h. ein enger Kontakt zwischen planaren Ringen mit konjugierten  $\pi$ -Systemen, die parallel, aber gegeneinander verschoben liegen. Durch die Wechselwirkungen der besetzten und unbesetzten Molekül- $\pi$ -Orbitale der Ring-Systeme bevorzugen die Ringe definierte Positionen relativ zueinander [93]. Solche  $\pi$ -Stapel sind jedoch in der verwendeten FLEXX-Version nicht mit der entsprechenden Wechselwirkungsgeometrie und Energie-Bewertung implementiert; bei der Platzierung von BH<sub>4</sub> in der Nähe des Phenylrings von Phe254 bzw. Phe300 können also nur *ungerichtete* hydrophobe Kontakte zwischen ihnen berücksichtigt werden. Offensichtlich bietet der PAH-Bindemodus mehr Möglichkeiten, geeignete Wechselwirkungspartner zu finden.

Bei beiden Enzymen lässt sich der TH-Bindemodus durch Docking auf die gewünschte Referenz-Struktur erzwingen. Die FLEXX-interne Bewertung beider Bindemodi ist sogar sehr ähnlich; demnach sind also wirklich die fehlenden Ankerpunkte für die Platzierung des Pterin-Ringsystems als Basis-Fragment dafür verantwortlich, dass der TH-Bindemodus nicht automatisch reproduzierbar ist.

**Energie-Bewertung** Nach den Energie-Kriterien wird von beiden Enzymen der PAH-Bindemodus bevorzugt, wenn auch aus verschiedenen Gründen: Bei PAH ist die Solvationsenergie für BH<sub>4</sub> im TH-Bindemodus höher als im PAH-Bindemodus; bei TH ist die von Proteinteilen eingeengte Seitenkette des BH<sub>4</sub> ungünstig verdrillt. Aufgrund der restriktiven Randbedingungen während der Struktur-Optimierung können die Seitenketten des Proteins nicht nachgeben, sodass möglicherweise die hohe Torsionsenergie der Seitenkette

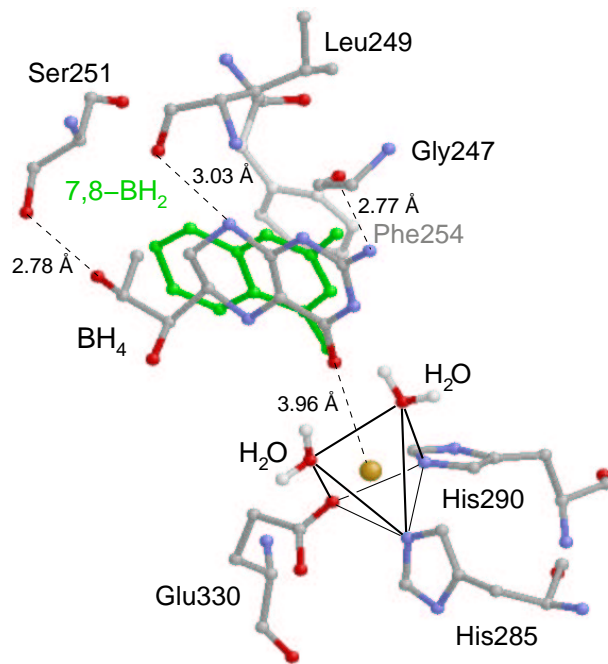


Abbildung 4.27: Das beste Docking-Ergebnis für die Platzierung von  $\text{BH}_4$  in der Bindetasche der PAH im Vergleich mit der Referenz-Struktur des 7,8- $\text{BH}_2$  (grün) aus der KS 1dmw. H-Atome sind (außer bei Wasser-Molekülen) nicht dargestellt.

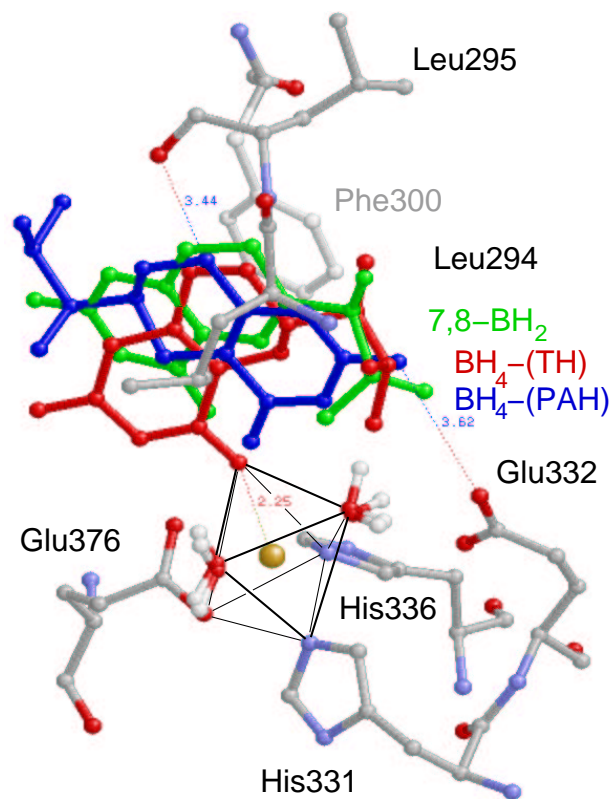


Abbildung 4.28: Die besten Ergebnisse des manuellen (rot) bzw. automatischen (blau) Dockings von  $\text{BH}_4$  in die Bindetasche der TH im Vergleich mit der Referenz-Struktur des 7,8- $\text{BH}_2$  (grün) aus der KS 2toh. H-Atome sind (außer bei Wasser-Molekülen) nicht dargestellt.

die realistische Bewertung der Konformationen verfälscht. Bei der TH sollten also beide Orientierungen als potentiell äquivalent in Betracht gezogen werden. Im Gegensatz zu den Energie-Beiträgen der Torsionswinkel spricht die Solvatationsenergie weniger empfindlich auf geringfügige Atom-Verschiebungen an, daher ist bei PAH nur der kristallographisch belegte Bindemodus tatsächlich relevant.

#### 4.4.2. Vorhersage

##### Catechol-TH-Komplexe

**Strukturen** Alle Catechol-Gruppen der Liganden L-DOPA, DA, MDTIQ und Rotenon lassen sich übereinstimmend mit *rmsd*-Werten von höchstens 0.97 Å in der Bindetasche der TH platzieren. Abbildung 4.29 zeigt die besten, nach den Kriterien der Gesamtenergie und des Eisen-Catechol-Abstands ausgewählten Platzierungen. Ebenso wie bei der PAH ist die Bindung der OH-Gruppen zum Eisen-Atom die einzige definierte Wechselwirkung zwischen den Liganden und der TH; es gibt keine Kontakt-Gruppen des Proteins, die allgemeine Ankerpunkte für Catechol-Derivate wären. Der bidentate Bindemodus ist jedoch leicht verzerrt, bedingt durch das in der TH vorliegende Glu-Rotamer. Während in der PAH eindeutig nur eines der O-Atome der CO<sub>2</sub><sup>-</sup>-Gruppe zur ersten Liganden-Sphäre des Eisens gehört, sitzt bei der TH auch das zweite, mit 2.67 Å relativ dicht am Eisen. Dadurch wird die bei der PAH fast ideale oktaedrische Ligand-Umgebung des Eisens in der TH verzerrt und einzelne Liganden werden zur Seite gedrängt.

**FLEXX-Leistung** Der Einfluss des Glutamat-Rotamers zeigt sich schon in der Art der von FLEXX vorgeschlagenen Konformationen: Nur für L-DOPA werden die vom Docking der Catechol-Derivate in die PAH bekannten Bindemodi (über Catechol- oder CO<sub>2</sub><sup>-</sup>-Gruppe) ebenfalls erzeugt; DA und MDTIQ in TH zeigen eine zwar untereinander ähnliche, aber von der für das Docking von DA in PAH abweichende Verteilung von Platzierungen. Positionierungen von DA, die der Anlagerung der NH<sub>3</sub><sup>+</sup>-Gruppe zwischen Tyr325 und Glu330, und der Verankerung der Catechol-Gruppen an Tyr377 in der PAH entsprechen, fehlen. Durch das andere Glu-Rotamer ist in der TH die Flankierung der Ammonium-Gruppe zwischen der OH-Gruppe von Tyr371 und Glu376 offensichtlich nicht mehr so günstig. Zudem ist, bedingt durch die unterschiedlichen Referenz-Liganden, die zur Definition der Bindetasche verwendet worden sind (s. Abschnitt 13), bei der TH die AS Tyr423 (entspricht Tyr377 in PAH) nicht erfasst. Je größer die Bindetasche gewählt ist, desto sicherer kann zwar eine korrekte Platzierung gefunden werden; gleichzeitig wird aber auch die Zahl der irrelevanten Vorschläge größer. Insofern war die Wahl der Bindetasche bei der TH glücklicher als bei der PAH.

**Vergleich von Docking und Redocking** Übereinstimmend werden die Catechol-Gruppen sowohl bei der Reproduktion, als auch bei der Vorhersage von Komplexen, bidentat am Eisen-Atom platziert. Diese berechneten Positionen passen sehr gut zu dem auch kristallographisch belegten Bindemodus. In den KS sind die Positionen der Seitenkette der

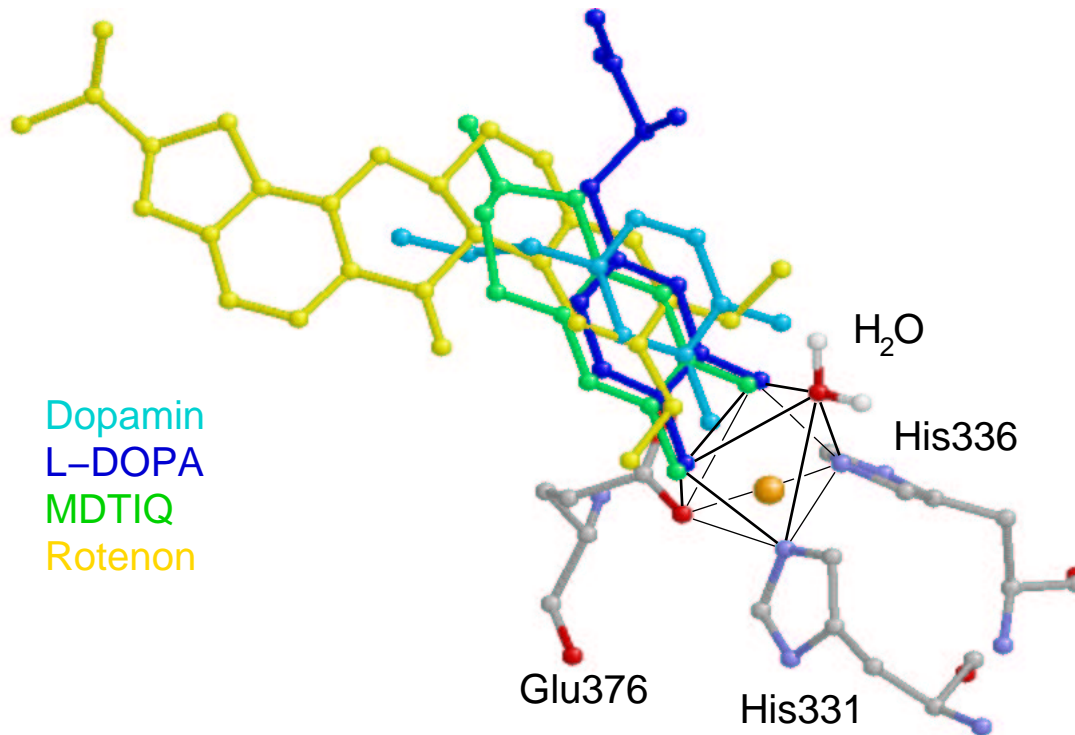


Abbildung 4.29.: Die besten Docking-Ergebnisse der Catechol-Derivate an das aktive Zentrum der TH (KS 2toh). Die Positionen der Catechol-Einheiten decken sich weitgehend. Die erste Liganden-Sphäre bildet einen verzerrten Oktaeder um das Eisen-Atom. H-Atome sind nicht dargestellt.

Catechol-Derivate Adrenalin, Noradrenalin, DA und L-DOPA (3pah, 4pah, 5pah und 6pah [34]) schlecht definiert, d. h. offensichtlich wird keine spezielle Seitenketten-Konformation bevorzugt; ein Befund, der sich mit den Rechnungen zu DA deckt. Nur bei L-DOPA in TH und PAH werden sehr wenige relevante Seitenketten-Konformationen ermittelt, obwohl die entsprechende KS der PAH zeigt, dass die L-DOPA-Seitenkette strukturell genauso wenig festgelegt ist, wie diejenige von DA. Die bei der Berechnung fehlende Vielfalt von Konformationen ist vermutlich methodisch bedingt: FLEXX verankert Liganden grundsätzlich an möglichst vielen Wechselwirkungspartnern gleichzeitig. Durch das Vorhandensein zweier funktioneller Gruppen in der Seitenkette von L-DOPA ( $\text{NH}_3^+$ - und  $\text{CO}_2^-$ -Gruppe) sind die Möglichkeiten, für beide gleichzeitig eine Gegengruppe zu finden, stark eingeschränkt; daher resultieren viele ähnliche Vorschläge, die bei der weiteren Struktur-Optimierung in das gleiche Energie-Minimum konvergieren. Da völlig kontaktlose, gänzlich solvatisierte Konformationen der Seitenkette prinzipiell nicht vorgeschlagen werden (FLEXX *sucht* Wechselwirkungspartner), scheint der L-DOPA-Bindemodus nach den Berechnungen eindeutiger festgelegt zu sein, als es tatsächlich zutrifft.

Abgesehen davon, wird dieser L-DOPA-Platzierung in TH im Gegensatz zu PAH eine recht hohe Energie zugewiesen, was auf der Abstoßung zwischen der (bei DA und MDTIQ fehlenden) anionischen  $\text{CO}_2^-$ -Gruppe und der  $\text{CO}_2^-$ -Gruppe von Asp425 beruht. Bei PAH

entspricht diese AS dem ungeladenen Val379, sodass die Gegenwart oder das Fehlen einer  $\text{CO}_2^-$ -Gruppe bei L-DOPA bzw. DA keinen Unterschied bewirkt. Tatsächlich ist *in vitro* messbar, dass die L-DOPA-Menge, die zur Hemmung rekombinanter, menschlicher TH notwendig ist, 10 – 15fach größer ist, als die dazu notwendige DA-Konzentration [2]. Bei der PAH wird im Vergleich dazu nur die dreifache L-DOPA-Menge benötigt [96]. Allgemein sind bei der TH die Dissoziationskonstanten für Catecholamin-Derivate mit  $\text{CO}_2^-$ -Gruppen 100 – 1000fach größer als für solche ohne  $\text{CO}_2^-$ -Gruppen [114]. Die negative Ladung von Asp425 verringert also offensichtlich (im Vergleich zu DA) die Hemmung der TH durch ihr eigenes Produkt L-DOPA.

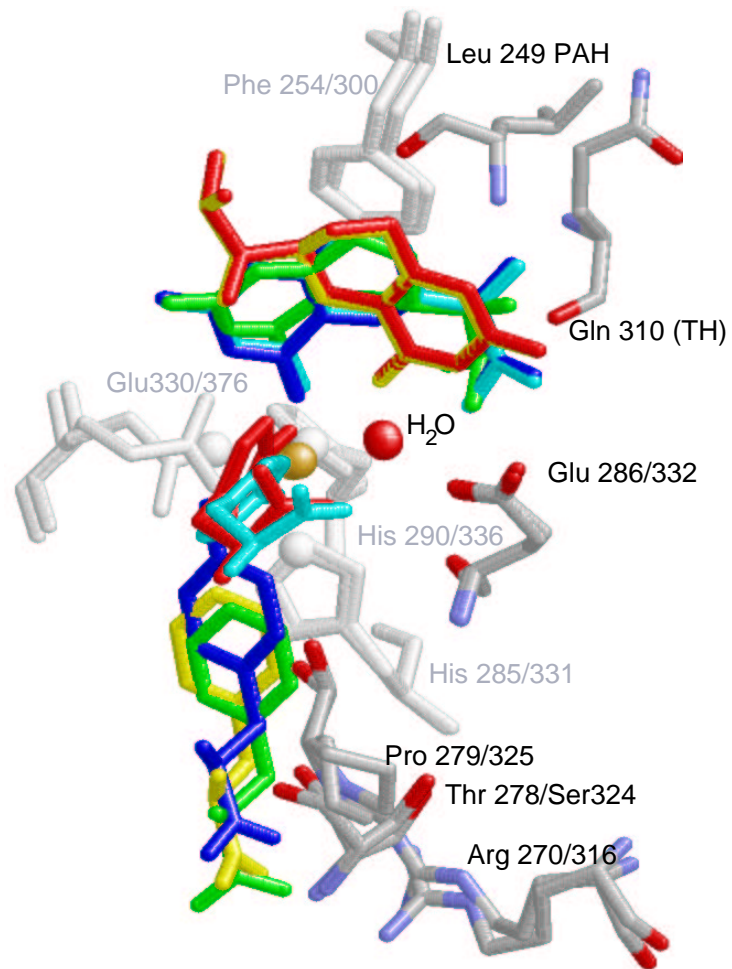
### Substrat-Komplexe von PAH bzw. TH

**Strukturen** Den Berechnungen zufolge bevorzugen Tyr und Phe (unabhängig von der Orientierung des Cosubstrates  $\text{BH}_4$ ) sehr unterschiedliche Positionen innerhalb der Bindetaschen von PAH und TH. Dies ist insofern zu erwarten, als Tyr eine OH-Gruppe besitzt, die an das Eisen-Atom binden kann und Phe nicht. Für Tyr in TH wird zusätzlich noch eine weitere Platzierung gefunden, die den Phe-Positionen entspricht. Abbildung 4.30 fasst die besten Platzierungen von Phe und Tyr in den beiden Bindetaschen zusammen; die Tabellen 4.2 und 4.3 zeigen die einzelnen LJ- und Coulomb-Beiträge zur Komplex-Energie zwischen den best-platzierten AS und den benachbarten Seitenketten des jeweiligen Proteins.

Demnach zeigt Phe übereinstimmend einen Kontakt zwischen der  $\text{CO}_2^-$ -Gruppe und der Seitenkette von Arg316 (TH) bzw. Arg270 (PAH), während der Ring mit seiner Fläche zum aktiven Zentrum weist (s. Tab. 4.2); der *rmsd*-Wert zwischen den beiden besten Platzierungen in den überlagerten Bindetaschen beträgt 2.15 Å. Dagegen ist Tyrosin bevorzugt über seine OH-Gruppe am Eisen verankert, während der Ring mit seiner Kante zum Eisen gedreht ist. Der *rmsd*-Wert zwischen den Tyr-Platzierungen in den beiden Enzymen ist mit 1.11 Å noch geringer. Für *Tyrosin in TH* wird zusätzlich noch eine dem Phe in PAH bzw. TH entsprechende Platzierung mit jeweils 1.49 bzw. 1.92 Å *rmsd*-Abweichung gefunden (s. Tab. 4.2), jedoch *nicht für Tyr in PAH*.

NMR-unterstütztes Modelling des Substrat-Komplexes von Phe·7,8- $\text{BH}_2$ ·PAH [127] zeigt eine ganz ähnliche AS-Platzierung (mit *rmsd*-Werten von 2.25, 1.13 bzw. 1.90 Å für Phe· $\text{BH}_4$ ·PAH, Phe· $\text{BH}_4$ ·TH und Tyr· $\text{BH}_4$ ·TH). Erst seit Januar 2003 ist die KS des  $\text{BH}_4$ ·PAH-Komplexes mit Thienylalanin als AS-Analog öffentlich verfügbar. Den Autoren zufolge liegt der aromatische Ring des Thienylalanins wie bei Tyr· $\text{BH}_4$ ·TH bzw. Phe· $\text{BH}_4$ ·PAH fast planar auf His285 auf und die Seitenkette ist wie beschrieben über Arg270 und Thr278 fixiert [5].

Als Ursache für das Unvermögen der PAH, Tyr im produktiven Bindemodus zu platzieren kommen sterische Gründe, wie von Andersen et al. [5] vermutet, kaum in Frage, denn F-, Cl- und Me-Phe werden prinzipiell verarbeitet (vgl. Kap. 7). Auch ein testweiser Transfer der produktiven Tyr-Position von TH auf PAH zeigt, dass das Molekül sehr wohl in diesem Bindemodus in der Bindetasche der PAH Platz findet. Ausschlaggebend für das Ergebnis der Berechnungen ist demnach, dass FLEXX keine entsprechenden Struktur für



*Abbildung 4.30:* AS-Substrate nach Docking in die Komplexe von PAH bzw. TH mit gebundenem  $\text{BH}_4$ . Die besten Konformationen von Phe (gelb) und Tyr (rot) in  $\text{BH}_4$ -PAH, sowie von Phe (grün) und Tyr (türkis bzw. dunkelblau) in  $\text{BH}_4$ -TH sind überlagert dargestellt. Die wichtigsten Ankerpunkte sind Arg270/316, Thr278/Ser324 und His285/331 für die Phe-Platzierungen, sowie das Eisen-Atom für die Tyr-Platzierungen.

Tyr in PAH vorschlägt. Da sich Phe erwiesenermaßen produktiv in der PAH platzieren lässt, sind sicher die Wechselwirkungen der phenolischen OH-Gruppe mit ihrer nächsten Umgebung, d. h. der Liganden-Sphäre des Eisens, entscheidend. Von der Liganden-Sphäre der TH, bei der die Platzierung von Tyr im produktiven Bindemodus auch belegt ist, unterscheidet sich diejenige der PAH (den KS zufolge) nur im Rotamer des Eisen-fixierenden Glutamats.  $\text{CO}_2^-$ -Gruppen fungieren als Elektronen-Paar-Donoren, entweder für H-Brücken oder für koordinative Bindungen zu Übergangsmetall-Kationen. OH-Gruppen ihrerseits können sowohl als H-Brücken-Donoren, als auch als -Akzeptoren bzw. Elektronen-Paar-Donoren für koordinative Bindungen, fungieren. Vermutlich bewirkt die spezielle Anordnung eines Elektronen-Paar-Akzeptors (Eisen-Atom) dicht neben einem Elektronen-Paar-Donor (Glutamat), dass wegen der ambivalenten Funktionalität der OH-Gruppe ihre Eigenschaft als Elektronen-Donor stärker bewertet wird, als ihre Eigenschaft als H-Brücken-Donor. Bedingt durch das Glu-Rotamer ist bei der PAH eine regelmäßiger oktaedrische Liganden-Sphäre am Eisen möglich als bei der TH. Auch dies könnte ein Grund dafür sein, dass die direkte Koordinierung des Tyrosins über die OH-Gruppe bei der Platzierung bevorzugt wird.

FLEXX Viele Konformationen werden doppelt vorgeschlagen, da rotationssymmetrische Gruppen, wie einfache oder para-substituierte Phenylringe oder  $\text{CO}_2^-$ -Gruppen, in den beiden äquivalenten Möglichkeiten platziert und als verschiedene Vorschläge ausgegeben werden, obwohl sie chemisch bzw. energetisch nicht unterscheidbar sind. Die mitunter sehr großen Zahlen von Vorschlägen gehen auch darauf zurück, dass die Bindetaschen (besonders für die Platzierung von Phe und Tyr) sehr groß gewählt sind.

*Tabelle 4.2.: Energie-Beiträge der Lennard-Jones-, Coulomb- und H-Brücken-Wechselwirkungen zwischen den AS der Bindetaschen von PAH bzw. TH mit dem AS-Substrat im bevorzugten Phe-Bindemodus [kcal/mol].*

PAH/TH	Phe·BH <sub>4</sub> ·PAH			Phe·BH <sub>4</sub> ·TH			Tyr·BH <sub>4</sub> ·TH		
	$E_{\text{LJ}}$	$E_{\text{Coul.}}$	$E_{\text{H-Br.}}$	$E_{\text{LJ}}$	$E_{\text{Coul.}}$	$E_{\text{H-Br.}}$	$E_{\text{LJ}}$	$E_{\text{Coul.}}$	$E_{\text{H-Br.}}$
<b>Arg270/316</b>	<b>-1.12</b>	<b>-6.04</b>	<b>-0.9</b>	<b>-0.34</b>	<b>-10.02</b>	-0.68	<b>-1.06</b>	<b>-4.78</b>	
Met276/322	-0.74	1.16		-0.14	0.22		-0.36	0.68	
His277/323	-1.45	-1.10		-0.57	-0.71		-2.19	-1.85	
<b>Thr278/Ser324</b>	<b>-2.65</b>	<b>-4.41</b>	<b>-3.0</b>	<b>-1.57</b>	<b>-3.88</b>	<b>-3.31</b>	<b>-2.51</b>	<b>-4.81</b>	<b>-4.90</b>
Pro279/325	-0.50	0.23		-1.01	-1.40		-1.88	-1.44	
Glu280/326	-1.31	-1.36		-1.52	-0.08		-1.31	-0.74	
Pro281/327	-1.73	-0.09		-2.45	-0.48		-1.85	-0.24	
Asp282/328	-0.37	0.57		-0.49	3.68		-0.37	1.33	
<b>His285/331</b>	<b>-3.16</b>	-0.08		<b>-2.97</b>	-0.75		<b>-2.87</b>	0.44	
Glu330/376	-1.08	-1.80		-1.64	-0.57		-1.62	0.39	
Phe331/377	-0.85	0.09		-1.11	0.10		-1.10	0.12	
Gly346/392	-1.50	0.09		-1.16	0.16		-1.27	0.14	
Ser349/395	-1.85	1.08		-2.38	1.00		-1.32	0.73	
Ser350/396	-1.41	0.10		-0.86	-0.37		-1.33	-1.04	
<b>Val379/Asp425</b>	<b>-0.12</b>	<b>-0.38</b>		<b>-0.33</b>	<b>-3.17</b>		<b>-0.19</b>	<b>-3.90</b>	
Eisen	-0.04	2.90		-0.05	-0.20		-2.49	-4.26	

### 4.4.3. Biologische Konsequenzen

#### Catechol-Komplexe

**Produkt-Hemmung:** Die Produkt-Hemmung der TH und PAH durch Catecholamine umfasst zwei Stufen:

1. In einer reversiblen Reaktion konkurrieren Catecholamine mit  $\text{BH}_4$  um die Bindestelle am aktiven Zentrum [2, 113, 45].
2. In einer anschließenden, irreversiblen Reaktion wird das zunächst zweiwertige Eisen im Komplex mit einem Catechol-Derivat oxidiert zur dreiwertigen Form [56, 45]. Die Reaktivierung des Enzyms muss durch Reduktion des Eisens mit  $\text{BH}_4$  erfolgen [89].

*Tabelle 4.3.:* Energie-Beiträge der Lennard-Jones-, Coulomb- und H-Brücken-Wechselwirkungen zwischen den AS der Bindetaschen von PAH bzw. TH mit Tyr im bevorzugten Bindemodus [kcal/mol].

PAH/TH	Tyr·BH <sub>4</sub> ·PAH			Tyr·BH <sub>4</sub> ·TH		
	$E_{LJ}$	$E_{Coul.}$	$E_{H-Br.}$	$E_{LJ}$	$E_{Coul.}$	$E_{H-Br.}$
Leu248/294	-0.75	-0.21		-2.57	-0.23	
<b>Pro279/325</b>	-1.38	-2.13	-2.53	-1.52	-0.16	
Glu280/326	-1.54	-0.28		-1.53	-0.54	
Pro281/327	-3.20	-0.16		-2.91	-0.03	
His385/331	-1.13	0.48		-0.72	0.17	
Trp326/372	-1.48	0.07		-1.53	0.04	
Glu330/376	-1.92	2.20		-1.27	1.85	
<b>Val379/Asp425</b>	-1.28	-0.39		-0.40	<b>-13.12</b>	<b>-1.00</b>
Eisen	<b>-17.32</b>	<b>-15.86</b>		<b>-15.62</b>	<b>-12.89</b>	
BH <sub>4</sub>	<b>-2.78</b>	<b>-1.62</b>	<b>-2.69</b>	<b>-2.95</b>	1.00	

**zu 1:** Aus den KS der PAH und den konstruierten Komplexen der TH geht übereinstimmend hervor, dass die Positionen der Catechol-Derivate und der Cosubstrate innerhalb der Bindetaschen beider Enzyme überlappen. Die berechneten Komplexe sind also im Einklang mit den experimentellen Befunden über die reversible, kompetitive Produkt-Hemmung, bei der die Catecholamine mit dem Cosubstrat BH<sub>4</sub> um die Bindung konkurrieren [2].

**zu 2:** Die Beobachtung, dass das zweiwertige Eisen-Atom der katalytischen Untereinheit in Gegenwart von Catecholaminen *oxidiert* wird, überrascht, denn Catecholamine sind *Reduktionsmittel*, die ihrerseits selbst oxidiert werden (zu Chinon-Derivaten, in Gegenwart von Luft-Sauerstoff findet auch Autoxidation statt [91]).

Die Oxidation des Eisens in Gegenwart von Catecholaminen erfordert also ein weiteres Molekül als Oxidationsmittel. Es liegt nahe, anzunehmen, dass dieses fehlende Oxidationsmittel O<sub>2</sub> ist, welches auch bei der enzymatischen Reaktion an das Eisen bindet [45]. Das O<sub>2</sub>-Molekül würde dann das letzte, *trans* zur Glutamat-Gruppe sitzende Wasser-Molekül ersetzen<sup>16</sup> (s. Abb. 4.31).

Analog zur initialen BH<sub>4</sub>-Oxidation während der enzymatischen Reaktion (s. Abschnitt 5) könnte ein Hydrid-Ion vom Catecholamin auf das distale Atom des O<sub>2</sub>-Moleküls übertragen werden, sodass ein protoniertes *o*-Chinon-Derivat, ein Hydroxyl-Anion und das

<sup>16</sup>Eine entsprechende Position nimmt das O<sub>2</sub>-Analog NO in der KS 1blz der Isopenicillin-N-Synthase ein [38], die ihr Eisen-Atom ebenfalls über zwei Histidin-Ringe und eine CO<sub>2</sub><sup>-</sup>-Gruppe fixiert.



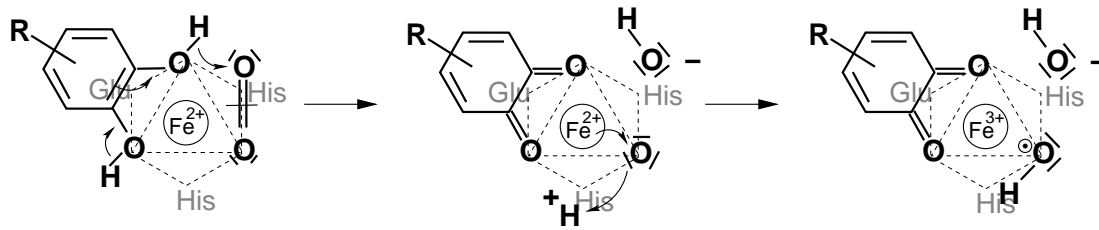


Abbildung 4.31.: Hypothese zur Toxizität von Catechol-Derivaten: Catecholamine könnten daran beteiligt sein,  $O_2$  teilweise zu reduzieren, sodass ein Intermediat entsteht, welches dem Übergangszustand während der enzymatischen Katalyse entspricht. Es zerfällt dann allerdings in ein  $^{\circ}OH$ -Radikal, ein Hydroxyl-Ion und dreiwertiges Eisen.

bekannte Eisen-Sauerstoff-Intermediat entstehen. Ohne Nachbarschaft zu einem elektronenreichen Aromaten zerfiel das  $Fe^{(II)}-O$ -Intermediat unter Protonierung in dreiwertiges Eisen und ein  $^{\circ}OH$ -Radikal.

Nach dieser Hypothese hängt die Bildung von  $^{\circ}OH$ -Radikalen vom stöchiometrischen Gleichgewicht zwischen der TH, dem Cosubstrat  $BH_4$  und den Catecholamin-Endprodukten ab. Eine klinisch wichtige Verschiebung dieses Gleichgewichts könnte auftreten, wenn Patienten systemisch mit L-DOPA behandelt werden. L-DOPA wird innerhalb des Gehirns zu DA decarboxyliert. *Post mortem*-Untersuchungen an den Gehirnen von Patienten und Tiermodellen haben gezeigt, dass die untergehenden Neurone in der *substantia nigra* (bedingt durch eine Schwächung der Abwehr-Systeme gegen oxidativen Stress, wie z. B. Glutathion) besonders verletzlich gegenüber reaktiven Sauerstoff-Verbindungen sind [14]. Die L-DOPA-Therapie könnte als unerwünschten Nebeneffekt zusätzlich oxidativen Stress verursachen, indem das beschriebene Gleichgewicht in Richtung TH-Hemmung und Eisen-Oxidation geschoben wird.

Reaktionen nach diesem Muster könnten ein Grund für die Toxizität von Catechol-Verbindungen wie MDTIQ oder Rotenon sein. Von MDTIQ ist nachgewiesen, dass es  $^{\circ}OH$ -Radikale produziert. Als Oxidationsprodukt entsteht dabei das N-Methyl-dihydroisochinoly-Kation ( $MDIQ^+$ ), welches aus dem primär entstanden Chinon-Derivat über den in Abbildung 4.32 gezeigten Mechanismus hervorgehen könnte [99].

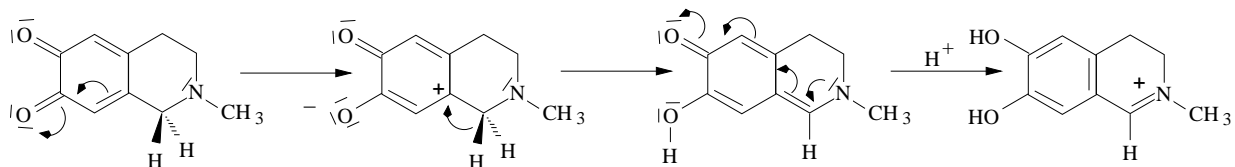


Abbildung 4.32.: Aus dem Chinon-Derivat des MDTIQ entsteht das nachweisbare MDIQ-Kation.

Rotenon kann unter Ausdehnung des aromatischen Systems zwei Reduktionsäquivalente abgeben, prinzipiell also auch  $O_2$  reaktiv aktivieren (s. Abb. 4.33).

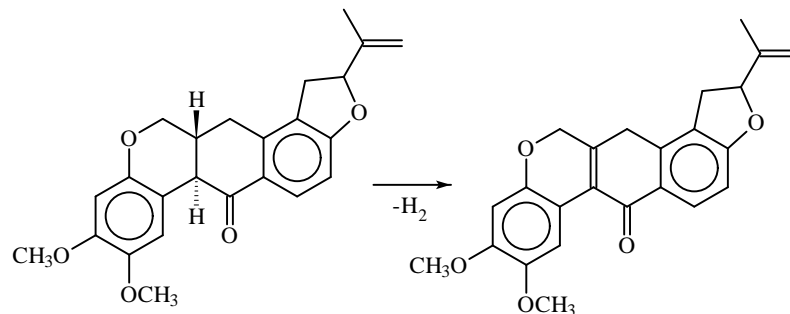


Abbildung 4.33.: Auch Rotenon kann unter Ausdehnung des konjugierten  $\pi$ -Systems ein  $H_2$ -Molekül abspalten.

### Substrat-Komplexe

**Produktiver  $BH_4$ -Bindemodus** Sowohl bei PAH, als auch bei TH wird der PAH-Bindemodus des Cosubstrates favorisiert. Die berechneten Ergebnisse werden dadurch bestätigt, dass sowohl bei TH als auch bei PAH das S-Enantiomer des 4a-OH- $BH_4$  entsteht [74]. Man kann davon ausgehen, dass während der initialen Hydroxylierung das distale Sauerstoff-Atom an der leichter zugänglichen Seite des  $BH_4$ -Moleküls gebunden wird, d. h. an der Seite, die dem Ring von Phe254/300 abgewandt ist. Folglich trifft für beide Enzyme der PAH-Bindemodus als produktiver Bindemodus zu. Für die TH ist eine gewisse Ambivalenz bei der Platzierung des Cosubstrates feststellbar, da der PAH-Bindemodus nur geringfügig günstiger ist. Offensichtlich wird abhängig von den experimentellen Bedingungen die eine oder andere Orientierung bevorzugt.

**Produktiver AS-Bindemodus** Da sich in drei von vier Fällen die AS Phe und Tyr übereinstimmend platzieren lassen, wird zunächst diese Variante auf ihre Eignung als produktive Platzierung überprüft. Alle Vorschläge weisen eine ionische Wechselwirkung zu Arg270/316 und H-Brücken zu Hauptketten-Gruppen von Thr278/Ser324 auf. Mutagenese-Experimente an Arg316 (TH) belegen passenderweise, dass diese Seitenkette essentiell für eine produktive Bindung des AS-Substrates ist: Ein Austausch von Arg316 gegen Lysin führt zu einer signifikanten Verschlechterung der Tyrosin-Bindung ( $K_{Tyr}$  steigt um den Faktor 400 [21]). Auch ein Austausch von Asp328, das eine stabilisierende Salz-Brücke zu Arg316 ausbildet, gegen Serin verringert die Fähigkeit des Enzyms, Tyr effektiv zu verarbeiten ( $K_{Tyr}$  steigt in diesem Fall um den Faktor 60)<sup>17</sup>. Genauso zeigen Experimente an der PAH, dass das

<sup>17</sup>Offensichtlich fixiert Asp328 die kationische Guanidinium-Gruppe von Arg316 an der Position, in der sie die  $CO_2^-$ -Gruppe der jeweiligen AS durch elektrostatische Anziehung verankert.

Enzym nach Mutation von Arg270<sup>18</sup> keinerlei Aktivität besitzt [28]. Ein weiteres Indiz für diese Platzierung als produktiver Bindemodus für AS ist das Konkurrenz-Verhalten der Substrate und der Catechol-Derivate: Die Catechol-Derivate konkurrieren wie beschrieben mit  $\text{BH}_4$  und nicht mit ihren Vorläufer-Molekülen um die Bindung am aktiven Zentrum. Passend zu diesem Befund überlappen sich die berechneten Positionen von Catechol-Derivaten und AS in diesem Bindemodus nicht.

**Hypothetischer Zustand vor der Reaktion** Obwohl der aromatische Ring im Phe· $\text{BH}_4$ ·TH-Komplex nicht planar auf His331 aufliegt, scheint dieser Komplex dennoch besonders geeignet, die Verhältnisse innerhalb des reaktiven Substrat-Komplexes vorzustellen, denn die reaktiven Zentren liegen hier in einer idealen Konstellation zueinander (s. Abb. 4.34).

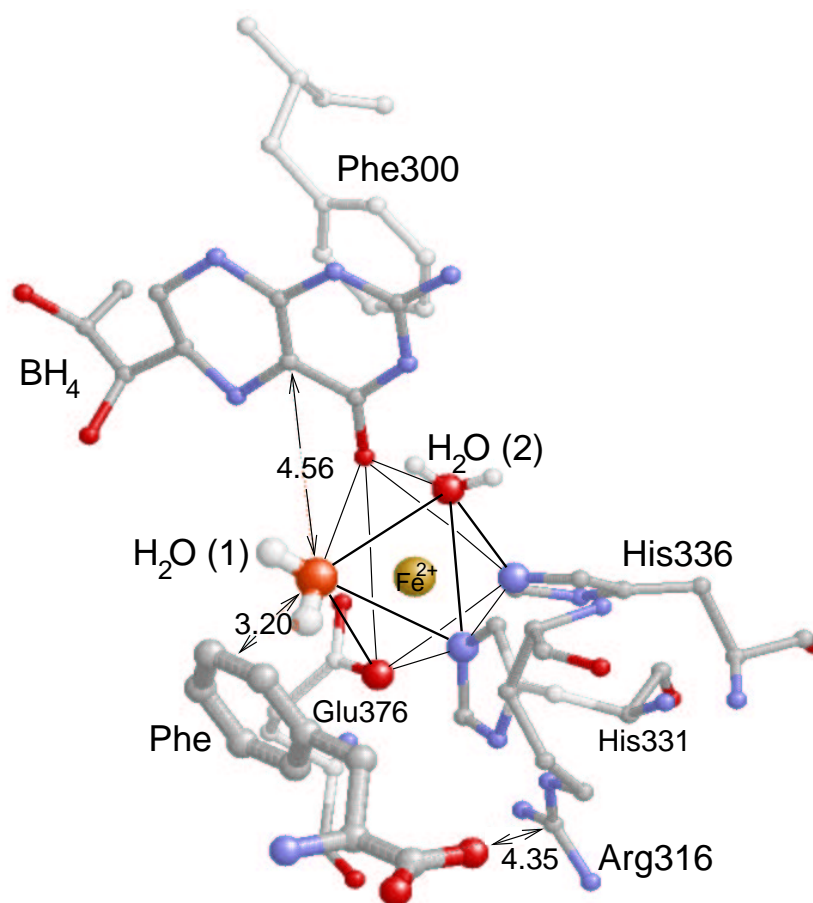


Abbildung 4.34: Der Phe· $\text{BH}_4$ ·TH-Komplex als Beispiel für einen produktiven Substrat-Komplex. Das fehlende  $\text{O}_2$ -Molekül ist an der Position von Wasser-Molekül 1 anzunehmen. Entfernungen sind in Å angegeben, H-Atome der besseren Übersichtlichkeit wegen weggelassen.

Für einen produktiven Komplex ist das (bisher fehlende)  $\text{O}_2$ -Molekül zu ergänzen. Es muss einerseits an das Eisen-Atom gebunden sein (Steigerung der Reaktivität durch Spin-Umkehr bei der Komplexbildung [74]), andererseits muss es in der Nähe der beiden zu hydroxylierenden Zentren von  $\text{BH}_4$  und AS liegen (einstufige Reaktion ohne Zeit für große

<sup>18</sup>Ebenso wie bei der TH bildet Asp282 eine Salz-Brücke zu Arg270.

„Wanderungen“ von Zwischen-Produkten [43]). Beide Bedingungen sind nur dann erfüllbar, wenn es an der Oktaeder-Ecke *trans* zu His336 sitzt, die im konstruierten Komplex von einem Wasser-Molekül besetzt wird [127]. Das direkt an das Eisen-Atom bindende, proximale O-Atom würde die Stelle des Wasser-O-Atoms einnehmen, während das freie, distale O-Atom sich in Richtung des BH<sub>4</sub>-Moleküls orientieren könnte. Die Entfernung des proximalen O-Atoms zum C4-Atom des Phenyl-Rings betrüge dann nur 3.2 Å und es läge damit ideal, um mit den  $\pi$ -Elektronen des Aromaten zu reagieren. Die Entfernung zum C4a-Atom des BH<sub>4</sub> wäre mit 4.56 Å zwar größer, doch ist zu berücksichtigen, dass die O-Atome des O<sub>2</sub>-Moleküls einen Bindungsabstand von 1.21 Å aufweisen [119]. Das distale O-Atom wäre also sicherlich in Reichweite des C4a-Atoms des BH<sub>4</sub>. Die Reaktion könnte wie in Abschnitt 5 beschrieben über die initiale BH<sub>4</sub>-Hydroxylierung und den elektrophilen Angriff des O-Fe-Intermediats am aromatischen Ring ablaufen.

**Unproduktiver AS-Bindemodus** In beiden Enzymen bevorzugt Tyrosin (ähnlich wie Catechol-Derivate) für seine OH-Gruppe ausgerechnet die hypothetisch vom O<sub>2</sub>-Molekül zu besetzende Oktaeder-Ecke *trans* zu His290/336. Die für die O<sub>2</sub>-Bindung alternative Oktaeder-Ecke *trans* zu Glu330/376 ist mit 4.40 Å zwar ähnlich weit vom C4a-Atom des BH<sub>4</sub> entfernt, liegt jedoch in einem sehr ungünstigen, spitzen Winkel zur Molekül-Ebene des BH<sub>4</sub>, sodass kaum mit einer initialen Hydroxylierung zu rechnen ist. Außerdem ist die Oktaeder-Ecke *trans* zu Glu330/376 mit 4.82 – 4.95 Å recht weit von den zu hydroxylierenden C3-Atomen der Tyr-Ringe entfernt und befindet darüber hinaus etwa in den Ring-Ebenen, hat also keinen Kontakt zu den  $\pi$ -Elektronen des Aromaten. Aus diesen Gründen wird die Koordination der phenolischen OH-Gruppe zum Eisen als unproduktive AS-Platzierung betrachtet.

**AS-Substrat-Spezifität** Mit der Existenz von produktiven und unproduktiven AS-Platzierungen lassen sich die Substrat-Spezifitäten der katalytischen Domänen von PAH und TH einfach erklären: Da sich Tyr in PAH nicht in einem produktiven Bindemodus platzieren lässt, sondern zudem die optimale Bindung des O<sub>2</sub>-Moleküls blockiert, kann die PAH Tyr nicht verarbeiten.

**Reihenfolge der Substrat-Bindung** Obwohl die unproduktive Platzierung des Tyr-Moleküls auch von der TH bevorzugt wird, kann den Rechnungen zufolge innerhalb der Bindetasche auch eine zwar vergleichsweise ungünstige, aber produktive Position eingenommen werden. Wenn die Oktaederstelle *trans* zu His336 in der TH frei ist, wird Tyr zwangsläufig unproduktiv binden. Sollte diese Oktaeder-Position allerdings bereits besetzt sein und sich dieser Ligand nur schlecht verdrängen lassen, wäre das Tyr-Molekül gezwungen, die nächst beste Alternative, den produktiven Bindemodus einzunehmen. Falls das O<sub>2</sub>-Molekül tatsächlich ein solch starker Ligand am Eisen-Atom der TH ist, wäre damit zu begründen, warum für die katalytische Reaktion O<sub>2</sub> vor Tyr binden muss<sup>19</sup> [44, 109]). Tyr

<sup>19</sup>Bei PAH taucht die Notwendigkeit der vorherigen O<sub>2</sub>-Bindung nicht auf, da Phe nicht an das Eisen-Atom koordinieren kann.

ist bei hohen Konzentrationen auch ein kompetitiver Inhibitor gegenüber dem  $\text{BH}_4$ -Analog 6-Methyltetrahydropterin [44]. Es liegt nahe, dass es wie die Catechol-Derivate über die OH-Gruppe auch *trans* zu His331 an das Eisen binden kann und dann mit  $\text{BH}_4$  um die Bindung konkurriert. Beides würde erklären, warum die katalytische Untereinheit der TH mit ihrem eigenen Substrat fünffach niedrigere Umsatzraten aufweist als die PAH [23], weil nämlich Phe freiwillig eine produktive Position einnimmt, während Tyr die erforderliche Platzierung der anderen Substrate ( $\text{BH}_4$  und  $\text{O}_2$ ) erschwert und nur gezwungenermaßen auf den produktiven Bindemodus ausweicht.

## 4.5. Zusammenfassung

Nach der Abhandlung der verschiedenen Einzel-Aspekte können sie zu einem übergeordneten Bild vereinigt werden. Die Abbildung 4.35 beinhaltet sowohl Ergebnisse aus kinetischen, kristallographischen und NMR-Untersuchungen, als auch berechnete und hypothetisch ergänzte Strukturen. Als Beispiel für die Enzym-Funktionen von PAH und TH dient der Katalyse- und Inhibitionszyklus der TH: Die Substrate binden in der festgelegten Reihenfolge beginnend mit dem Cosubstrat  $\text{BH}_4$ , gefolgt von  $\text{O}_2$  und dem AS-Substrat Tyr. Während des katalytischen Schrittes reagiert ein O-Atom mit  $\text{BH}_4$  unter Bildung von 4a-OH- $\text{BH}_4$ . Das zweite O-Atom bildet kurzzeitig ein Eisen-Oxo-Intermediat, bevor die Hydroxylierung des aromatischen Rings stattfindet. Nach der Freisetzung der Produkte aus der Bindetasche wird L-DOPA von der aromatischen L-Aminosäure-Decarboxylase (EC 4.1.1.28) zu DA decarboxyliert. DA kann bidentat an das Eisen-Atom binden und ein nachfolgend gebundenes  $\text{O}_2$ -Molekül analog zur reduktiven Aktivierung während des Katalyse-Zyklus aktivieren. Als Resultat entsteht ein ortho-Chinon-Derivat und ein hochreaktives Intermediat, welches ohne ein benachbartes, Elektronen-reiches Nucleophil in ein oxidiertes Eisen-Ion und ein hoch reaktives  $^{\circ}\text{OH}$ -Radikal zerfällt.  $\text{BH}_4$  wird benötigt, um das dreiwertige Eisen wieder zu reduzieren und dadurch die Enzym-Aktivität wieder herzustellen.

Die Rechnungen zu den Substrat- und Inhibitor-Komplexen von PAH und TH zeigen, dass die verwendeten Methoden sehr gut geeignet sind, Strukturen zu erzeugen, die sich mit experimentellen Untersuchungen decken. Dies ist um so erfreulicher, als die Verhältnisse in den Substrat-Komplexen von PAH und TH mit bis zu drei Liganden ( $\text{BH}_4$ , AS und Wasser-Molekül) um ein zentrales Eisen-Atom sehr kompliziert sind. Die scheinbare Diskrepanz zwischen dem kristallographischen Befund zum  $\text{BH}_4$ -Bindemodus in TH und den berechneten Ergebnissen konnte durch gezieltes, manuelles Docking aufgelöst werden; gleichzeitig wird die Leistungsfähigkeit des automatischen Verfahrens bestätigt. Dass für die verwandten Enzyme PAH und TH übertragbare Strukturen resultieren, zeigt nebenbei auch, wie bedeutend das Homologie-Konzept bei der Modellierung von Proteinen und ihren Komplexen ist.

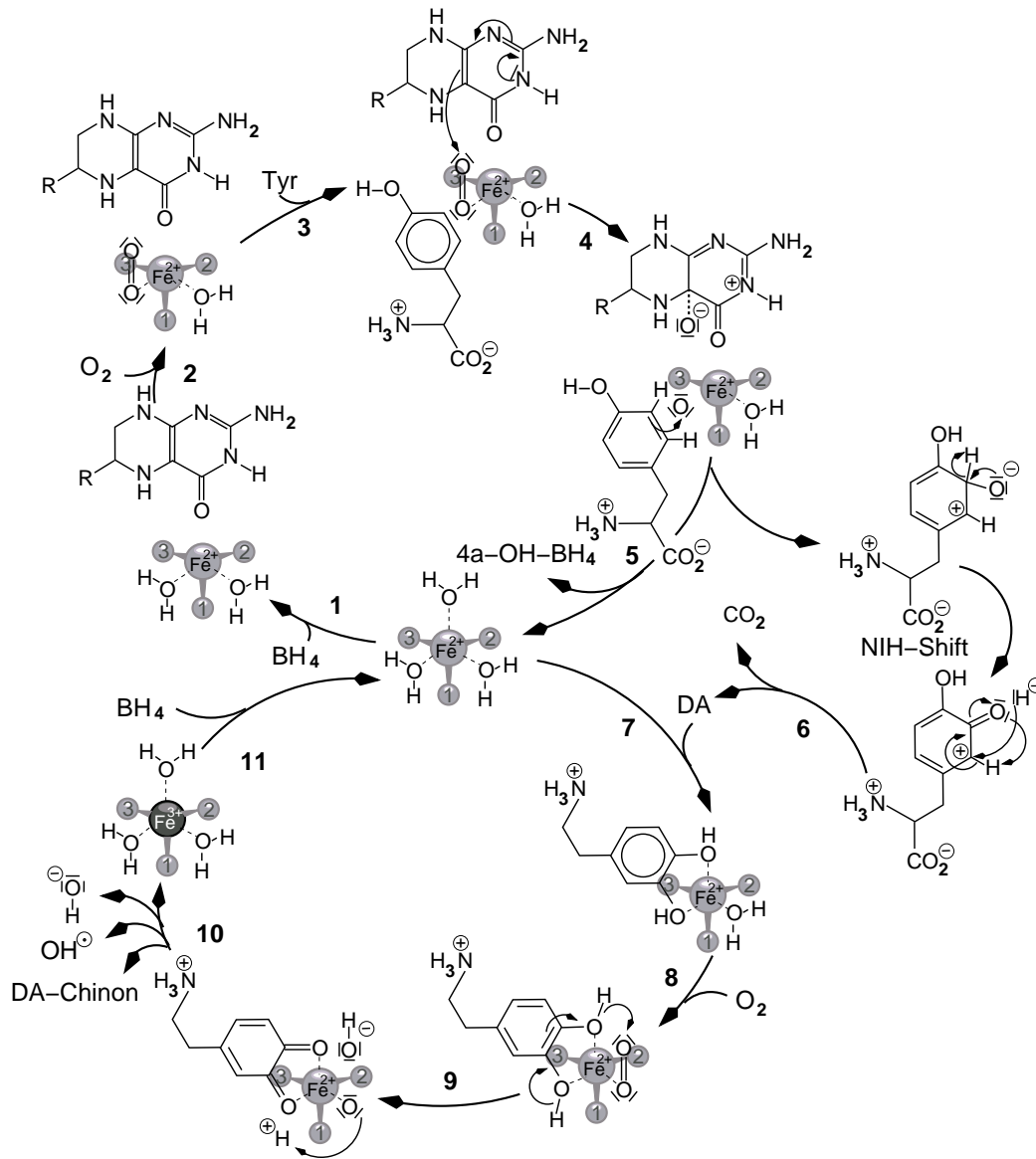


Abbildung 4.35.: Zusammenfassende Übersicht über den Katalyse- und Inhibitionszyklus der TH Katalyse-Zyklus:

1–3: Anlagerung der Substrate an das aktive Zentrum in der erforderlichen Reihenfolge.

4: Reduktive Aktivierung des Sauerstoffs

5: Elektrophiler Angriff am Aromaten und Umlagerung des primären Addukts

6: Spontane oder enzymatische Decarboxylierung des Produktes L-DOPA

Inhibitionszyklus:

7+8: Reversible Anlagerung von DA und  $O_2$

9: Reduktive Aktivierung des Sauerstoffs analog zu Schritt 4

10: Zerfall in die Produkte, insbesondere in ein  $^{\ominus}OH$ -Radikal und dreiwertiges Eisen

11: Reaktivierung des Enzyms durch Reduktion des Eisens.

# 5. GPCRs: Adenosin-Rezeptoren

## 5.1. Allgemeines

G-Protein gekoppelte Rezeptoren (GPCRs) vermitteln Signale aus der Umgebung von Zellen an ihr Inneres. Sensorische Rezeptoren zum Beispiel, die auf Geschmacks- und Geruchsstoffe oder sogar auf Licht ansprechen, sind GPCRs [42]. Andere, nicht-sensorische GPCRs regulieren elementare physiologische Prozesse: Suchtverhalten, Stimmungskontrolle und Erinnerungen sind nur einige der vielen Vorgänge, an denen GPCRs wesentlich beteiligt sind. Ihrer großen Bedeutung und dem allgegenwärtigen Vorkommen entsprechend wirken etwa 40% aller Medikamente über GPCRs [128].

Bedingt durch die technischen Schwierigkeiten im Umgang mit Membran-Proteinen sind trotz ihrer immensen Bedeutung nur wenige detaillierte Struktur-Informationen über GPCRs verfügbar [128]. Um Beziehungen zwischen der Struktur und der Funktionsweise von GPCRs herstellen zu können, konnten lange Zeit nur Affinitäts- und Mutagenese-Studien durchgeführt werden. Seit Sommer 2000 ist jedoch die Kristall-Struktur des Rinder-Rhodopsins verfügbar, die exemplarisch für die allgemeine Architektur von GPCRs ist. Nun ist die Homologie-basierte Modellierung beliebiger GPCRs möglich [107, 85, 134]. Die konstruierten 3D-Modelle können anhand der vorhandenen Affinitäts- und Mutagenese-Daten sowie anhand von Crosslinking-Experimenten validiert werden (s. Kap. 1.5). Damit eröffnen sich neue Möglichkeiten zum Verständnis der Funktionsweise der diversen GPCRs und zur gezielten Manipulation ihrer Funktion durch maßgeschneiderte Liganden.

Extrazelluläres Adenosin moduliert über GPCRs eine Reihe verschiedener physiologischer Funktionen [110, 103]. Wegen des damit verbundenen therapeutischen Potenzials sind Adenosin-Rezeptoren (AR) von großem Interesse und sollen deshalb im Rahmen dieser Arbeit modelliert werden.

### 5.1.1. G-Protein gekoppelte Rezeptoren

#### Vorkommen und Funktion

Zellen müssen auf Signale von außerhalb der Zellmembran reagieren können. Rezeptoren in der Membran vermitteln die Weiterleitung der äußeren Reize ins Innere. Wie Enzyme sind Rezeptoren Proteine, die nur ganz bestimmte Liganden binden. Im Gegensatz zu den Enzymen ändert sich im Komplex mit dem Liganden jedoch nicht das Ligand-, sondern das Protein-Molekül: Der Rezeptor verformt sich. Diese Konformationsänderung löst direkt

oder indirekt eine Reaktion in der Zelle aus [105].

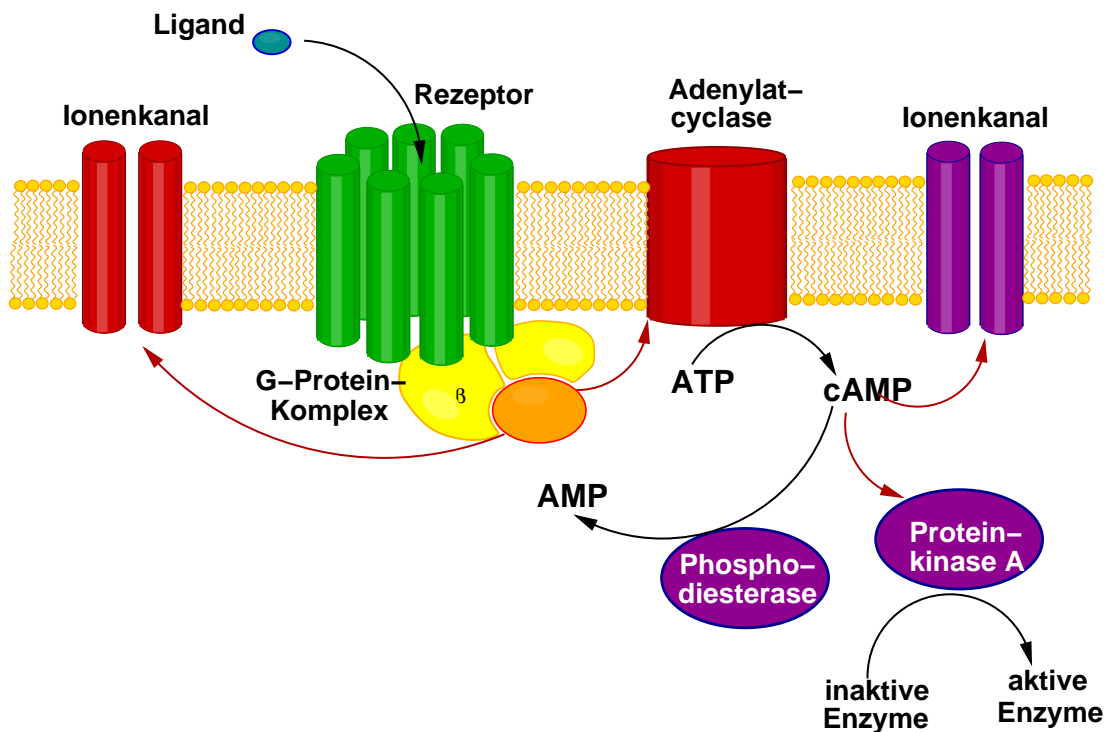


Abbildung 5.1.: Reiz-Weiterleitung über G-Protein gekoppelte Rezeptoren

Bei GPCRs läuft die Signal-Weiterleitung folgendermaßen ab (s. Abb. 5.1):

1. Der *Rezeptor* reicht durch die Membran hindurch. Von der Außenseite her können sich Liganden daran anlagern. An der Innenseite der Zellmembran ist er mit einem so genannten G-Protein assoziiert. Bei der Bindung eines speziellen Liganden, des *Agonisten*, geht ein GPCR aus dem Ruhezustand in eine aktive Form über. *Antagonisten* hingegen besetzen den Rezeptor, ohne ihn zu aktivieren [132]. Agonisten stabilisieren also die aktive Rezeptor-Konformation, während Antagonisten dies nicht tun.
2. Das GTP-bindende Protein (*G-Protein*) besteht aus drei Untereinheiten ( $G_{\alpha}$ ,  $G_{\beta}$  und  $G_{\gamma}$ ). Wenn sich der Rezeptor im Komplex mit einem Agonisten verformt, löst sich die  $G_{\alpha}$ -Untereinheit von dem Komplex aus Rezeptor und G-Protein. Die dissoziierte  $\alpha$ -Untereinheit enthält zunächst GDP, gibt es aber bald ab und bindet stattdessen GTP. So aktiviert, bewegt sie sich zu einem weiteren Protein, dem Effektor-System.
3. Das *Effektor-System*, ist ebenfalls ein membranständiges Protein, entweder ein Enzym (z. B. Adenylat-Cyclase) oder ein Ionenkanal. Wenn ein Effektor-System durch den Kontakt mit einer freien, GTP-haltigen  $\alpha$ -Untereinheit des G-Proteins stimuliert bzw. davon nicht inhibiert wird, erhöht es die Konzentration eines intrazellulären Botenstoffs (*second messenger*, z. B. cAMP).



4. Je nach enzymatischer Ausstattung der Zelle werden durch die *second messengers* typische Enzyme (z. B. Proteinkinase A) oder Ionenkanäle als nachgeschaltete Effektor-Systeme aktiviert oder deaktiviert, sodass sich der Stoffwechsel innerhalb der Zelle kurzfristig ändert.
5. Die Amplitude und Dauer des GPCR-Signals wird durch Arrestin-Proteine kontrolliert: Diese binden an aktivierte GPCRs und beenden das primäre Signal dadurch, dass sie weitere Wechselwirkungen des aktivierten Rezeptors mit G-Proteinen verhindern [105]. Das an der  $\alpha$ -Untereinheit des G-Proteins gebundene GTP, aber auch die *second messengers* werden nach einiger Zeit zersetzt, sodass die Dauer der Aktivierung begrenzt bleibt und das ganze System in den Ruhezustand zurückkehren kann [105].

Ein Vorteil dieser vielstufigen Kaskade ist einerseits ihr verstärkender Effekt, der die Zellen schon auf sehr schwache äußere Signale empfindlich reagieren lässt. Andererseits sind die Zellen durch Kombination verschiedener Rezeptoren, G-Proteine und Effektor-Systeme, sowie Zelltyp-spezifischer Stoffwechsellzyme dazu in der Lage, parallel und sehr differenziert auf ein breites Spektrum von Signalen zu reagieren.

### Grundsätzliche Architektur

GPCRs sind Monomere mit drei elementaren Regionen:

- *Extrazelluläre Region* zur Ligand-Erkennung
- *Transmembran-Region* mit zentraler Ligand-Bindestelle
- *Intrazelluläre Region* zur Kopplung an G-Proteine

Abbildung 5.2 zeigt die typische Unterteilung der GPCR-Sequenzen in 15 strukturelle Abschnitte: Ein extrazellulärer N-Terminus (ET), drei extrazelluläre Loops (EL1–3), ein Bündel von sieben Transmembran-Helices (TM1–7), drei intrazelluläre Loops (IL1–3) und ein intrazellulärer C-Terminus (IT) [133, 100]. Die zentrale Bindetasche befindet sich etwa im äußeren Drittel des Helix-Bündels [104, 133].

Die einzelnen GPCRs sind spezialisiert auf ihre jeweiligen extrazellulären Liganden und die intrazellulär angekoppelten G-Proteine. Bedingt dadurch unterscheiden sich die Rezeptoren vor allem in den entsprechenden Kontakt-Regionen hinsichtlich der Länge und der Sequenz der einzelnen Hauptketten-Abschnitte [107]. Das Bündel der Helices hingegen, die Transmembran-Region, ist gut konserviert. Einzelne AS in jeder Helix sind bei allen GPCRs so konstant, dass sie als Referenz-Punkte beim Vergleich verschiedener Sequenzen dienen [133]. Nach der Nomenklatur von Ballesteros und Weinstein [8] wird jede AS zusätzlich zu ihrer Position in der jeweiligen Sequenz mit der Nummer der TM-Helix und der Position relativ zur Referenz-AS bezeichnet. Die Referenz-AS ist jeweils die am höchsten konservierte AS dieser Helix und erhält konventionsgemäß die Nummer 50 [8, 85].

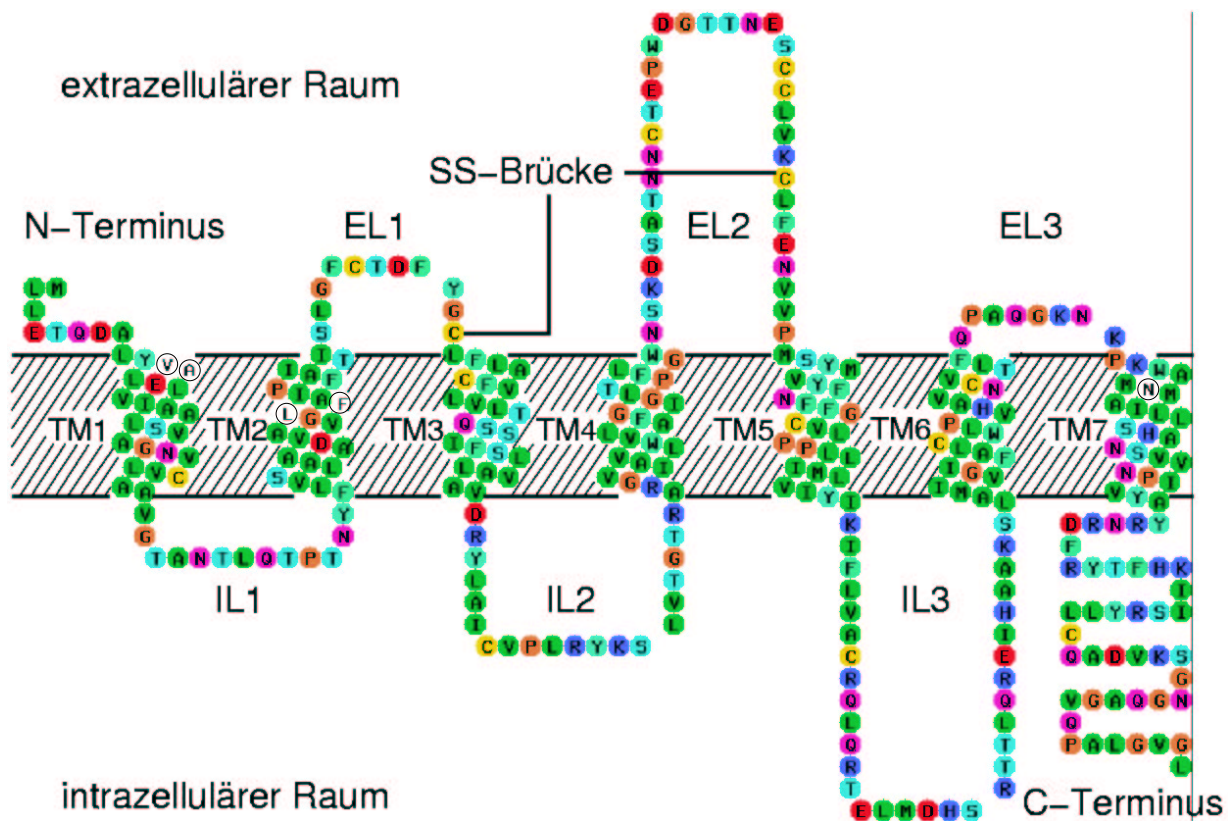


Abbildung 5.2.: So genanntes Schlangen-Diagramm der Sequenz des humanen  $A_{2B}$ -Rezeptors, exemplarisch für den allgemeinen Aufbau von GPCRs. Alle GPCRs besitzen sieben Transmembranhelices (TM1–7), eine extrazelluläre Region aus dem N-Terminus und drei Loops, sowie eine intrazelluläre Region aus drei Loops und dem C-Terminus. Die Disulfid-Brücke zwischen TM3 und EL2 ist hoch konserviert.

Alle GPCRs besitzen ein *E/DRY-Motiv* am Ende von TM3, welches für die Kopplung an G-Proteine notwendig ist, sowie eine *konservierte Disulfid-Brücke*, die den EL2 mit dem Beginn von TM3 verknüpft [133]. Sehr charakteristisch ist auch ein *NPXXY-Motiv* am Ende von TM7 und ein *Fettsäureester* am IT, der die Hauptkette zusätzlich an der Innenseite der Membran verankert. Extrazelluläre AS sind häufig glykosyliert [133].

### Exemplarisch für alle GPCRs: Rhodopsin

Rinder-Rhodopsin ist der erste GPCR, für den eine 3D-Struktur mit relativ guter Auflösung vorliegt (2.8 Å). Mit 348 AS ist Rhodopsin ein GPCR mittlerer Größe. Die KS 1f88<sup>1</sup> [107] zeigt es im dunkel-adaptierten Grundzustand mit gebundenem 11-*cis*-Retinal

<sup>1</sup>Eine neuere KS mit einer Auflösung von ebenfalls 2.8 Å (1hzx [128]) enthält verbesserte Strukturdaten; die mittlere *rmsd* zu den Atom-Koordinaten der älteren KS 1f88 beträgt 1.4 Å. Darüber hinaus ist eine NMR-Struktur (1jfp [138]) verfügbar. Sie ist zusammengesetzt aus den Strukturen einzelner, über-

(s. Abb. 5.3). Die Strukturen der sieben TM-Helices und der extrazellulären Region aus dem N-Terminus und den drei Loops (EL1–3) sind vollständig, nur von der intrazellulären Region aus den drei Loops (IL1–3) und dem C-Terminus fehlen kurze Abschnitte. Da es als Vorlage für die Modellierung anderer GPCRs dienen soll, wird es hier kurz beschrieben.

**Transmembran-Region:** Das Bündel der sieben Helices formt einen unregelmäßigen Zylinder, der senkrecht in die Ebene der Lipid-Membran eingelassen ist. Die Länge dieses Zylinders beträgt ca. 41 Å, durchspannt also den etwa 30 Å breiten unpolaren Kern einer Zell-Membran [135] vollständig. Die einzelnen Helices unterscheiden sich in Länge und Neigungswinkel gegenüber der gemeinsamen Hauptachse; TM1, TM4, TM6 und TM7 sind jeweils an den mit Prolin besetzten Positionen gebogen. Diese strukturellen Anomalien sind durch andere Experimente bestätigt, es handelt sich also nicht um kristallographische Artefakte [128]. Die Helices TM3, TM5, TM6 und TM7 haben Kontakt zum 11-*cis*-Retinal.

**Extrazelluläre Region:** Der N-Terminus liegt dicht beim EL3 und besitzt zwei antiparallele Stränge,  $\beta_1$  und  $\beta_2$ , die eine typische  $\beta$ -Faltblatt-Haarnadel-Schleife bilden. Ein Disaccharid an Asn15(ET) ragt in den extrazellulären Raum hinein. Während EL1 und EL3 mit jeweils acht AS recht kurz sind, faltet sich ein Teil des längeren EL2 mitten zwischen die Helices. EL2 besitzt ebenfalls zwei antiparallele Stränge,  $\beta_3$  und  $\beta_4$ , die zusammen ein  $\beta$ -Faltblatt formen. Der Strang  $\beta_4$  berührt das 11-*cis*-Retinal, gehört also mit zur Bindetasche des 11-*cis*-Retinals. Cys187(EL2) und Cys110(3.25) bilden eine Disulfid-Brücke, die auch bei den meisten anderen GPCRs vorhanden ist. Das an Lys296(7.43) gebundene 11-*cis*-Retinal liegt im extrazellulären Drittel der Transmembran-Region und ist unmittelbar von AS der Helices TM3, TM5, TM6 und TM7 umgeben.

**Intrazelluläre Domäne:** Die Verbindungssequenz zwischen den Helices TM1 und TM2 (IL1) ist mit sechs AS gerade so lang, dass sie eine Helix-Unterbrechung mit einer Biegung um ca. 180° erlaubt. Der Abschnitt zwischen den Helices TM3 und TM4 (IL2, 10 AS) ist auch relativ kurz. Deutlich länger sind IL3 (20 AS) und der intrazelluläre C-Terminus mit 37 AS. Ein kurzer, helikaler Abschnitt des IT liegt etwa parallel zur anzunehmenden Membran-Ebene. Obwohl keine Fettsäuren im derzeitigen Modell enthalten sind, kann man davon ausgehen, dass die Seitenketten von Cys322 und/oder Cys323 palmitoyliert sind, da sie zur Membran-Ebene hin orientiert sind.

**Bindetasche:** Rhodopsin hat seinen Liganden 11-*cis*-Retinal (im Gegensatz zu anderen GPCRs) *kovalent* über eine Schiff'sche Base an Lys296(7.43) gebunden. Die folgenden AS befinden sich im Abstand von bis zu 4.0 Å zum 11-*cis*-Retinal: Glu113(3.28), Ala117(3.32), Thr118(3.33), Gly121(3.36), Glu122(3.37), Ser186(EL2), Cys187(EL2),

---

lappender Protein-Fragmente, und zeigt eine vollständige intrazelluläre Domäne (bei allerdings unvollständiger extrazellulärer Domäne). Die Sekundär-Struktur-Elemente sind hier undeutlicher erkennbar als bei den beiden KS; Diese 3D-Struktur erscheint als „Mosaik“ zur Repräsentation der allgemeinen GPCR-Architektur weniger geeignet zu sein als die KS.

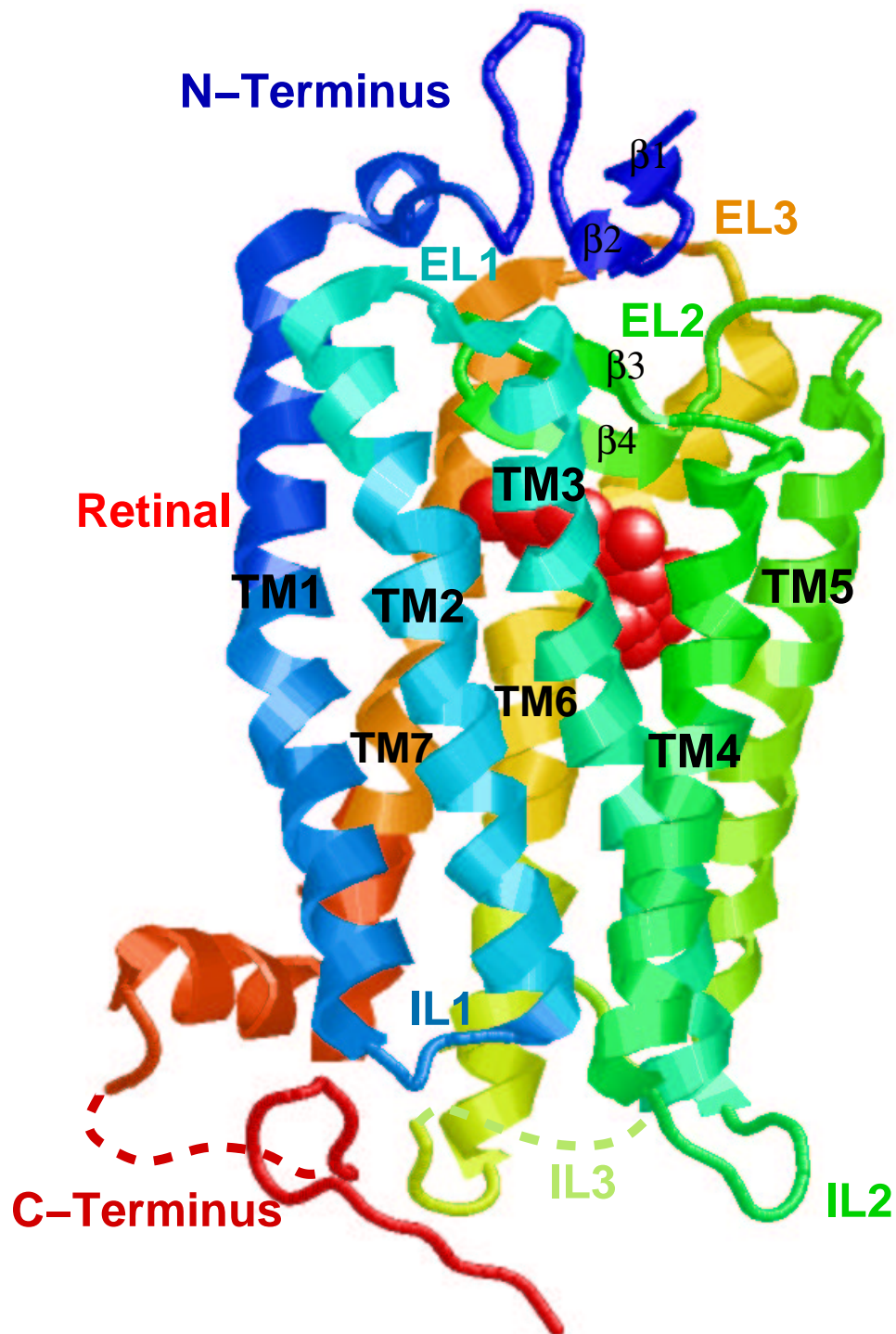


Abbildung 5.3.: KS des Rinder-Rhodopsins (1f88 [107]). Nur der Verlauf der Hauptkette ist gezeigt. Die Helices und  $\beta$ -Stränge sind bandförmig dargestellt, das in der zentralen Bindetasche gebundene 11-*cis*-Retinal als Raum-füllendes Modell. Die fehlenden Abschnitte sind als gestrichelte Linien ergänzt.

Ile189(EL2), Met207(5.42), Phe212(5.47), Phe261 (6.44), Trp265(6.48) und Ala269(6.52), Ala292(7.39). Das unpolare 11-*cis*-Retinal ist also überwiegend von ebenfalls unpolaren Seitenketten umgeben.

**Funktion:** Rhodopsin ist beteiligt an der Umwandlung von Licht-Reizen in neuronale Signale, die letztlich bis ins Sehzentrum des Gehirns weitergeleitet werden [128]. Die Stäbchen- und Zapfenzellen der Netzhaut des Auges besitzen viele Membranscheiben bzw. -einfaltungen. Die Rhodopsin-Moleküle sind regelmäßig in die Lipid-Doppelschicht dieser Membran-Stapel eingelagert [121]. Der Rezeptor ist aus dem GPCR Opsin und dem kovalent daran gebundenen 11-*cis*-Retinal zusammengesetzt. Bei Absorption eines Photons isomerisiert das 11-*cis*-Retinal zur all-*trans*-Form, woraufhin sich auch die Konformation des Opsins ändert.

In dieser Form stimuliert es das gekoppelte G-Protein Transducin, welches die cGMP-Phosphodiesterase anschaltet [105]. Das Rhodopsin bleibt nur kurz aktiviert, denn die Schiff'sche Base mit dem all-*trans*-Retinal wird hydrolysiert. Das all-*trans*-Retinal wird freigesetzt und der isolierte Protein-Anteil Opsin bleibt zurück. Es wird durch Verknüpfung mit frisch synthetisiertem 11-*cis*-Retinal aus epithelialen Nachbarzellen wieder zu aktivierbarem Rhodopsin regeneriert [107].

## 5.2. Adenosin-Rezeptoren

### 5.2.1. Extrazelluläres Adenosin

Adenosin ist das einzige Nucleosid, für das Membran-Rezeptoren bekannt sind [100, 110]. Viele Zellen produzieren im Cytosol Adenosin, meistens als Reaktion auf physiologischen Stress [76]. Es wird durch erleichterte Diffusion über ein spezifisches Nucleosid-Transporter-Protein der Membran aus der Zelle hinaus bzw. auch wieder in die Zelle hinein geschleust. Die Halbwertszeit von im Blut zirkulierendem Adenosin beträgt einige Sekunden, d. h. es wirkt vorwiegend lokal. Aus dem extrazellulären Raum wird es durch Transport in die Zellen (passiv oder aktiv) oder durch Abbau zu Inosin durch die Adenosin-desaminase bzw. durch Phosphorylierung zu AMP durch die Adenosin-Kinase wieder entfernt [110]. Adenosin kann auch durch extrazellulären Abbau von Adenin-Nucleotiden (z. B. ATP, ADP oder AMP) gebildet werden.

### 5.2.2. Eigenschaften der Rezeptoren

Bei Säugern lassen sich vier Subtypen von Adenosin-Rezeptoren (AR) differenzieren ( $A_1$ ,  $A_{2A}$ ,  $A_{2B}$ ,  $A_3$ ). Sie sind organ- bzw. gewebespezifisch verteilt (s. Tab. 5.1). Die Rezeptor-Subtypen  $A_1$  und  $A_{2A}$  schützen Zellen vor Schäden bei kurzfristigem Sauerstoff-Mangel: Über den  $A_1$ -Rezeptor wird der Sauerstoff-Bedarf des Arbeitsgewebes gedrosselt. Bei Stimulation des  $A_{2A}$ -AR von Gefäß-Zellen wird das Sauerstoff-Angebot für das umgebende

Arbeitsgewebe erhöht. Diese beiden Subtypen, A<sub>1</sub> und A<sub>2A</sub> reagieren schon auf sehr geringe extrazelluläre Adenosin-Mengen. Der A<sub>3</sub>-AR wird meist erst durch höhere Adenosin-Konzentrationen aktiviert aufgrund der geringen Rezeptor-Dichte in der Zellmembran. Der Subtyp A<sub>2B</sub> dagegen spricht in der Regel erst bei sehr hohen extrazellulären Adenosin-Konzentrationen an [58]. Von dem A<sub>2B</sub>-AR vermutet man daher, dass er bei akuter Bedrohung der Zellen die üblichen Schutzmechanismen zugunsten eines „Notprogramms“ außer Kraft setzt [110]. Auch die Rolle des zuletzt entdeckten A<sub>3</sub>-Rezeptors ist nicht vollständig geklärt. Es wird eine Beteiligung an entzündlichen bzw. allergischen Reaktionen und eine cardioprotektive Wirkung diskutiert [110].

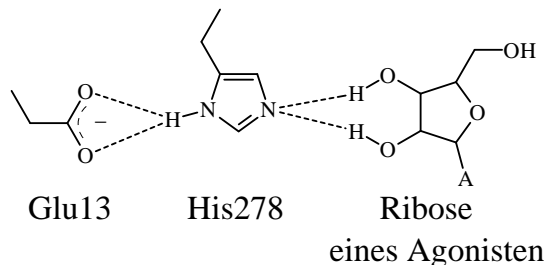
Bedingt durch die breite, aber Zell-spezifische Verteilung der AR im Körper und die prinzipielle Möglichkeit, sie durch Agonisten zu stimulieren oder durch Antagonisten zu blockieren, ist das therapeutische Potenzial der AR-Liganden groß: Es umfasst die Behandlung von neurodegenerativen Erkrankungen, von Epilepsie, Morbus Alzheimer, Morbus Parkinson (Adenosin ist der Gegenspieler von Dopamin [49]), den Einsatz als Beruhigungsmittel, als Muskel-Relaxantien und Schmerzmittel, ebenso die Anwendung bei Herzrhythmusstörungen und anderen Herzbeschwerden, zur Unterstützung der Nierenfunktion und zur Beeinflussung von entzündlichen Reaktionen, von Asthma bronchiale und anderen allergischen Reaktionen [110].

Um die therapeutischen Möglichkeiten der AR erschließen und nutzen zu können, sind hochaffine und Subtypen-selektive Agonisten und Antagonisten erforderlich. Die enge Verwandtschaft zwischen den verschiedenen Subtypen jedoch erschwert die Entwicklung selektiver Liganden [110]. Dennoch sind Abweichungen vorhanden, die die AR-Subtypen auf eine Reihe von Wirkstoffen unterschiedlich reagieren lassen. Jeder Subtyp ist daher charakterisierbar über sein Agonisten- bzw. sein Antagonisten-Profil (s. Tab. 5.1 [100, 110, 48]).

## Agonisten

Agonisten der ARs sind immer *Derivate des Adenosins*. Für die agonistische Wirkung eines Liganden ist die an die Adenin-Base angeknüpfte, intakte Ribose-Einheit entscheidend [110]. Die Ribose-Einheit ist vermutlich über His(7.43) und Glu(1.39) (in Analogie zur Triade der Serin-Proteasen [69]) verankert (s. Abb. 5.4).

**Abbildung 5.4:** Mutagenese-Experimenten zufolge ist die Ribose-Einheit der Agonisten über H-Brücken an dem Histidin-Ring von His(7.43) in der zentralen Bindestelle verankert. Die CO<sub>2</sub><sup>-</sup>-Gruppe von Glu(1.39) fixiert dabei den Histidin-Ring in der geeigneten tautomeren Form und Ausrichtung [49, 69].



N<sup>6</sup>-H und N7 sind bei Adenosin vermutlich das für die Bindung an den Rezeptor essentielle Paar aus einem *H-Brücken-Donor* vicinal zu einem *H-Brücken-Akzeptor* im Abstand

Tabelle 5.1.: Subtypen des humanen Adenosin-Rezeptors [102, 61, 100, 110, 48]

	A <sub>1</sub>	A <sub>2A</sub>	A <sub>2B</sub>	A <sub>3</sub>
<b>Vorkommen:</b>				
• viel	Hippocampus, Cortex	Striatum, N. accumbens, Riechkolben, Blutplättchen	Caecum, Dünndarm, Blase	Leber, Lunge
• wenig	Herz, Niere, Lunge, Hoden, Fettzellen	Leber, Lunge, Herz, Niere	Hirn, Fibroblasten, Lunge	Herz, Hirn, Herz, Hoden, Niere
<b>Kopplung:</b>				
• G-Proteine	G <sub>i(1-3)</sub> , G <sub>o</sub> , G <sub>q/11</sub>	G <sub>s</sub>	G <sub>s</sub>	G <sub>i(2,3)</sub> , G <sub>q/11</sub>
• Effektor-Systeme	AC: cAMP↓ PLC: IP <sub>3</sub> ↑, Ca <sup>2+</sup> ↑ Kanäle: K <sup>+</sup> ↑, Ca <sup>2+</sup> ↓	AC: cAMP ↑	AC: cAMP ↑ PLC: IP <sub>3</sub> ↑, Ca <sup>2+</sup> ↑	AC: cAMP↓ PLC: IP <sub>3</sub> ↑, Ca <sup>2+</sup> ↑
<b>Sequenz-Länge:</b>	326	412	332	318
<b>Homologie zu:</b>				
• A <sub>1</sub>	-	51 %	46 %	50%
• A <sub>2A</sub>	.	-	61 %	43%
• A <sub>2B</sub>	.	.	-	40 %
• Rho	49%	46 %	53 %	53 %
<b>selektiv<sup>e</sup> Agonisten:</b>	CPA, CCPA, CHA	CG-21680, HENECA, CV-1808, CV-1674, ATL146e	-	CI-IB-MECA
<b>selektiv<sup>e</sup> Antagonisten:</b>	DPCPX, 8-Cyclopentyltheophyllin, WRC0571, ADPEP, DPEAP	SCH58261, MSX-2; bedingt selektiv: ZM241385, KF 17387, CSC	MRS1754, Enprophyllin, Alloxazin, PSB-1115	MRS1220, MRE3008-F20, MRS1191, MRS1523
<p>Legende: G<sub>i</sub>: inhibitorische G-Proteine; G<sub>s</sub>: stimulatorische G-Proteine; AC: Adenylat-Cyclase; cAMP: zyklisches Adenosinmonophosphat; PLC: Phospholipase C; IP<sub>3</sub>: Inositoltriphosphat Systematische Namen der Agonisten und Antagonisten: s. Tab. 5.2</p>				

Tabelle 5.2.: Systematische Bezeichnungen der Agonisten und Antagonisten aus Tabelle 5.1.

**Agonisten**

CPA:	N <sup>6</sup> -(Cyclopentyl)adenosin
CCPA:	2-Chlor-N <sup>6</sup> -(Cyclopentyl)adenosin
CHA:	N <sup>6</sup> -(Cyclohexyl)adenosin
CGS-21680:	2-[4-[(2-Carboxyethyl)phenylethylamino]-5'-ethylcarboxamidoadenosin
HENECA:	2-Hex-1-ynyl-5'-N-ethylcarboxamidoadenosin
CV-1808:	2-Phenylaminoadenosin
CV-1674:	2-Phenyl-4-methoxy-adenosin
ATL146e:	4-3-[6-Amino-9-(5-ethylcarbonyl-3,4-dihydroxy-tetrahydro-furan-2-yl)-9 <i>H</i> -purin-2-yl]-prop-2-ynyl-cyclohexan-carboxylmethylester
Cl-IB-MECA:	2-Chlor-N <sup>6</sup> -(3-iodobenzyl)adenosin

**Antagonisten**

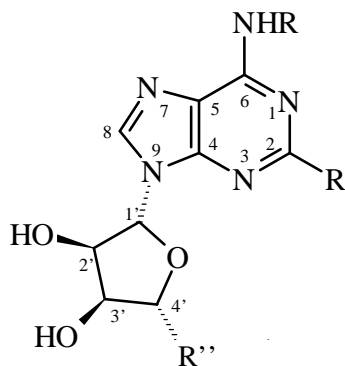
DPCPX:	1,3-Dipropyl-8-cyclopentylxanthin
WRC0571:	C <sup>8</sup> -( <i>N</i> -Methylisopropyl)-amino-N <sup>6</sup> -(5'-endohydroxy)endonorboman-2-yl-9-methyladenin
ADPEP:	( <i>R</i> )-7-1-(Methylbenzyl)-2-phenylpyrrolo[2,3- <i>d</i> ]pyrimidin-4-amin
DPEAP:	( <i>R</i> )- <i>N</i> -(1-Phenylethyl)-4-amino-5,6-dimethyl-2-phenyl-7 <i>H</i> -pyrrolo[2,3- <i>d</i> ]pyrimidin
SCH58261:	5-Amino-7-(2-phenylethyl)-2-(d-furyl)-pyrazolo[4,3- <i>e</i> ]-1,2,4-triazolo[1,5- <i>c</i> ]pyrimidin
MSX-2:	3-(3-Hydroxypropyl)-8-( <i>m</i> -methoxystyryl)-1-propargylxanthin
ZM-241385:	4-(2-[7-Amino-2-(2-furyl)-[1,2,4]-triazol[1,5- <i>a</i> ][1,3,5]triazin-5-yl-amino]ethyl)phenol
KF17387:	1,3-Dipropyl-8-(3,4-dimethoxystyryl)-7-methylxanthin
CSC:	8-(3-Chlorstyryl)coffein
MRS1754:	N-(4-Cyano-phenyl)-2-[4-(2,6-dioxo-1,3-dipropyl-2,3,4,5,6,7-hexahydro-1 <i>H</i> -purin-8-yl)-phenoxy]acetamid
Enprofyllin:	3,7-Dihydro-3-propyl-1 <i>H</i> -purin-2,6-dion
Alloxazin:	1,3 <i>H</i> -Benzopteridin-2,4-dion
PSB-1115:	1-Propyl-8- <i>p</i> -sulfophenylxanthin
MRS1220:	9-Chloro-2-(2-furanyl)-5-[(phenylacetyl)amino][1,2,4]-triazolo[1,5- <i>c</i> ]chinazolin
MRE3008-F20:	5-[[[(4-Methoxyphenyl)amino]carbonyl]amino-8-ethyl-2-(2-furyl)-pyrazolo[4,3- <i>e</i> ]1,2,4-triazolo[1,5- <i>c</i> ]pyrimidin
MRS1191:	3-Ethyl-5-benzyl-2-methyl-4-phenylethynyl-6-phenyl-1,4-(±)-dihydropyridin-3,5-dicarboxylat
MRS1523:	2,3-Diethyl-4,5-propyl-6-phenylpyridin-3-thiocarboxylat-5-carboxylat



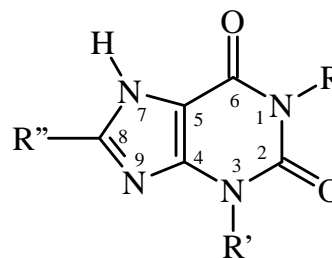
von ca. 3.0 Å [100]. Ringschluss-Experimente zwischen Base und Ribose haben gezeigt, dass das Adenosin im Komplex mit dem A<sub>2B</sub>-Rezeptor in *anti*-Konformation, d.h. mit einem Torsionswinkel von 15 – 60° zwischen den Atomen C8, N9, C1' und O, gebunden ist [19].

Für die AR-Subtypen A<sub>1</sub>, A<sub>2A</sub> und A<sub>3</sub> existieren potente und selektive Agonisten, die sich Rezeptor-abhängig im bevorzugten Substitutionsmuster des Adenosin-Grundgerüsts unterscheiden (s. Tab. 5.1).

- Die A<sub>1</sub>- und A<sub>3</sub>-AR haben eine hohe Affinität zu N<sup>6</sup>-substituierten Adenosin-Analoga [117, 41];
- Der A<sub>2A</sub>-AR hingegen hat eine niedrige Affinität zu den N<sup>6</sup>-substituierten Derivaten und bevorzugt stattdessen Substituenten am C2- und 5'-Atom [117, 41].
- Für den A<sub>2B</sub>-AR ist kein selektiver Agonist bekannt; am wirksamsten ist das bei allen Subtypen aktive NECA<sup>2</sup> [25]. Mono-Substitution am N<sup>6</sup>-Atom von NECA verringert die Affinität, disubstituierte Verbindungen sind vollständig inaktiv, ebenso C8- und C2-substituierte NECA-Derivate [25].



Adenosin: R= R'= H  
R''=CH<sub>2</sub>OH



Xanthin: R=R'=R''= H

Abbildung 5.5.: Grundstrukturen von Agonisten und Antagonisten an Adenosin-Rezeptoren

---

<sup>2</sup>5'-N-Ethylcarboxamidoadenosin

## Antagonisten

Antagonisten sind meist planare, aromatische oder zumindest  $\pi$ -elektronenreiche Heterozyklen wie *Xanthin-Derivate* mit einem ähnlichen H-Brücken-Donor- und -Akzeptor-Paar wie Adenosin, allerdings *ohne* Ribose-Einheit [100, 110]. Einfache Xanthine wirken nicht selektiv auf die verschiedenen AR-Subtypen und zeigen neben dem kompetitiven Antagonismus zu Adenosin auch Interaktionen mit anderen Rezeptoren und verschiedenen Enzymen [100]. Wie bei den Agonisten lassen sich Affinität und Selektivität aber durch Variation der Substituenten optimieren.

- 8-Cycloalkyl-substituierte 1,3-Dipropylxanthine sind starke und selektive Antagonisten für den  $A_1$ -AR [100, 27].
- Xanthin-Derivate mit hoher Affinität und Selektivität zum  $A_{2A}$ -Rezeptor besitzen einen voluminösen oder aromatischen Substituenten in 8-Position, einen essentiellen H-Brücken Akzeptor ( $O^6$ -Atom) und einen H-Brücken-Donor ( $N^7$ -H-Atom).<sup>3</sup>
- Anders als bei den Agonisten existieren auch für den  $A_{2B}$ -Rezeptor potente Antagonisten. Der  $A_{2B}$ -Rezeptor bevorzugt Xanthin-Derivate mit Substituenten, die an Position 1 größer sind als an Position 3 [78]. Generell sind Alkyl-Substituenten an Position 1 und ein H-Atom oder eine Methyl-Gruppe an Position 3 günstig; 1,3-Dipropylxanthine sind jedoch wie schon beim  $A_1$ -Rezeptor ebenfalls wirksam [78]. Ein Phenyl-Ring am C8-Atom verstärkt die Affinität; der dort gebundene Ring sollte möglichst elektronenreich sein [78]. Kürzlich ist mit 1-Propyl-8-*p*-sulfophenylxanthin (PSB-1115) die Entwicklung eines sehr potenten und selektiven  $A_{2B}$ -Antagonisten gelungen [58]. Es sind auch Non-Xanthin-Antagonisten wirksam, z.B. das Triazolotriazin-Derivat ZM241385, welches an beiden  $A_2$ -Rezeptoren eine hohe Affinität aufweist [26].

## Bindetasche

Es gilt als gesichert, dass Agonisten und Antagonisten die gleiche Kontakt-Region des Rezeptors besetzen [110, 29]. Aus den Affinitäts-Profilen für die Agonisten und Antagonisten des jeweiligen Rezeptor-Subtyps (s. Tab. 5.1) kann man auf den Aufbau der Bindetaschen rückschließen, denn hochaffine Liganden müssen der jeweiligen Bindetasche sterisch und elektrostatisch weitgehend komplementär sein. Eine Überlagerung der Ligand-Strukturen lässt ihre entscheidenden Gemeinsamkeiten deutlicher hervortreten, quasi als „Negativ-Abdruck“ der Bindetasche.

---

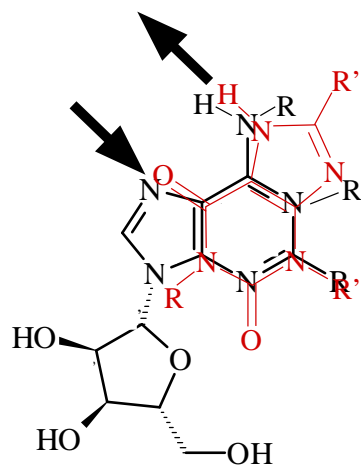
<sup>3</sup>Dieser darf bei 8-Styrylxanthinen fehlen [101].

Demnach besitzen die Bindetaschen aller Subtypen drei verschiedene Regionen:

- Eine Kontakt-Region für die Ribose-Einheit der Agonisten,
- Eine zentrale Kontakt-Region, die das planare,  $\pi$ -Elektronen-reiche Purin- bzw. Xanthin-Ringsystem aufnimmt. Hier muss sich außerdem ein H-Brücken-Donor-/Akzeptor-Paar in passendem Abstand zu den Donor- und Akzeptor-Atomen der Liganden in der zentralen Kontakt-Region befinden.
- Eine Region, die Kontakt zu einem voluminösen, unpolaren Substituenten aufnimmt.

Aus der typischen Anordnung des *H-Brücken-Donor-* und *-Akzeptor-Paares* der Liganden kann man ableiten, dass sich die Positionen der Doppelringe von Adenosin und Xanthin-Derivaten trotz der passenden Fünf- und Sechsring-Struktur in der Bindetasche gerade *nicht* decken [100, 110]: Beim  $A_1$ -AR sind die Liganden in etwa so zu überlagern, dass die  $N^6$ - bzw.  $N7$ -Atome der Agonisten mit den  $N7$ - bzw.  $O^6$ -Atomen der Antagonisten übereinstimmen ( $N^6$ - $N7$ -Modell, s. Abb. 5.6). Dann liegen auch die Substituenten am  $N^6$ -Atom der Agonisten bzw. die Substituenten am C8-Atom der Antagonisten an ähnlichen Positionen [29].

Abbildung 5.6: Überlagerung der Agonist- bzw. Antagonist-Grundstrukturen nach dem  $N^6$ - $N7$ -Modell [29]. Wichtig ist, dass die H-Brücken-Donor- und Akzeptor-Gruppen übereinstimmen (Pfeile).



### 5.2.3. Motivation

Mit Kenntnissen über die Sequenzen, die allgemeine Architektur, das jeweilige Verhalten gegenüber Agonisten und Antagonisten, sowie Mutagenese-Daten sind die hochaffinen Subtypen der AR,  $A_1$ ,  $A_{2A}$  und  $A_3$ , bereits gut charakterisiert. Vor allem die potenten und selektiven Agonisten und Antagonisten ermöglichen es, die physiologische Funktion des jeweiligen AR im Tierversuch zu untersuchen [110]. Dies ist die Voraussetzung für die Entwicklung von Medikamenten, die später auch für die Behandlung von Menschen zugelassen werden sollen.

Für den  $A_{2B}$ -Rezeptor hingegen ist kein selektiver Agonist verfügbar und auch Adenosin bindet nur mit sehr schwacher Affinität [11]. Daher ist seine physiologische Charakterisierung schwierig; bislang müssen die empfindlicheren  $A_1$  und  $A_{2A}$ -Rezeptoren erst *selektiv blockiert* werden, bevor sich durch NECA am  $A_{2B}$ -AR ausgelöste Effekte beobachten lassen [41]. Für eine mögliche therapeutische Nutzung der  $A_{2B}$ -Rezeptoren ist diese Methode der selektiven Blockade jedoch unbrauchbar; denn das Ziel einer Therapie wäre es, den  $A_{2B}$ -AR *unter Erhalt der übrigen Körperfunktionen* zu stimulieren. Der  $A_{2B}$ -Rezeptor ist also trotz seiner engen Verwandtschaft zu den anderen AR physiologisch und strukturell schlecht charakterisiert.

## Konzept

Die Suche nach potenten Liganden wird allgemein erleichtert, wenn ein 3D-Modell des Rezeptors vorhanden ist, welches die Beziehungen zwischen Struktur und Funktion erklärt, sodass Vorhersagen über die Wechselwirkungen zwischen Ligand und Protein möglich werden [16]. Daher soll nun die 3D-Struktur des  $A_{2B}$ -AR auf Grundlage der KS des Rhodopsins modelliert werden. Besonders interessant – im Hinblick auf die Entwicklung von Medikamenten — sind vor allem die Regionen, die für die Bindung von Liganden entscheidend sind, also die extrazelluläre und die Transmembran-Region. Parallel dazu soll die Zuverlässigkeit der Methode bzw. der resultierenden 3D-Struktur validiert werden durch die gleichzeitige Modellierung des  $A_{2A}$ -Rezeptors, der unter den anderen Subtypen mit dem  $A_{2B}$ -Rezeptor und dem Rinder-Rhodopsin am engsten verwandt ist (s. Tabelle 5.1). Für den  $A_{2A}$ -Rezeptor sind Mutagenese-Daten vorhanden, die die Validierung des erstellten 3D-Modell erlauben.

Die Bedeutung einzelner Mutationen für die Ligand-Affinität wird häufig erst dann klar, wenn direkte Wechselwirkungen des Rezeptors mit dem Liganden erkennbar sind. Um die Modelle anhand der Mutagenese-relevanten AS zu validieren, müssten also auch Liganden im Rezeptor platziert, d. h. Komplexe erzeugt werden. Aufgrund der (methodisch bedingten) Unzuverlässigkeit<sup>4</sup> der erzeugten Atom-Positionen des Proteins scheidet das flexible Docking der Liganden zur Konstruktion der Komplexe von vornherein aus. Man kennt allerdings die Rotamer-Form des Adenosins und seine Verankerung über die Ribose an His(7.43), sodass in diesem Fall die gezielte Platzierung des starren Adenosin-Moleküls an dem genannten Ankerpunkt manuell möglich scheint. Aus einem solchermaßen erzeugten Komplex wären im Erfolgsfall die Gegengruppen des H-Donor-/H-Akzeptor-Paares des Agonisten abzulesen, sodass wiederum Antagonisten an diesen Gegengruppen platziert werden könnten.

Gelingt auf diese Weise eine passable Platzierung von Liganden, erhalten die Mutagenese-Daten durch die dann nachvollziehbaren Nachbarschaftsbeziehungen stärkeres Gewicht bei der Beurteilung, wie realistisch die generierten Modell-Strukturen sind. Jeder Schritt

---

<sup>4</sup>Das Homologie-basierte 3D-Modell eines Proteins entspricht zunächst nicht seiner eigenen nativen Form, sondern derjenigen der Struktur-Vorlage. Die Atom-Koordinaten der hier verwendeten Vorlage sind zudem mit einer Auflösung der KS von 2.8 Å relativ unsicher. Aus beiden Gründen ist eine Verfeinerung der Modell-Struktur sinnvoll.

des skizzierten Weges birgt allerdings viele Unsicherheiten, sodass die Zuverlässigkeit der erzeugten Komplexe sehr kritisch zu beurteilen ist. Da Röntgen-kristallographische Methoden bei GPCRs so schlecht anwendbar sind, bieten sich zur Bestätigung der 3D-Modelle Crosslinking-Experimente zur Überprüfung der Modelle an. Diese Untersuchungen setzen wiederum geeignete Crosslinker voraus, die aber nach dem 3D-Modell entworfen werden könnten.

## 5.3. Homologie-basierte Modellierung von AR-Komplexen

Auf der Grundlage der KS des Rinder-Rhodopsins und der Sequenzen der  $A_{2A}$ - und  $A_{2B}$ -Rezeptoren sollen für beide dreidimensionale Modelle konstruiert werden. Dadurch sollen folgende Fragen geklärt werden:

- Können plausible 3D-Modelle für den  $A_{2A}$ - und  $A_{2B}$ -Rezeptor erzeugt werden?
- Welche AS kleiden jeweils die Bindetasche aus?
- Worauf könnten die unterschiedlichen Affinitäten von  $A_{2A}$ - und  $A_{2B}$ -Rezeptor beruhen?

### 5.3.1. Multiples Sequenz-Alignment

#### Durchführung

Um sicher die wesentlichen Übereinstimmungen zwischen der Rhodopsin-Vorlage und den AR-Sequenzen zu erfassen, wird ein multiples Sequenz-Alignment durchgeführt, statt die Sequenzen nur paarweise miteinander zu vergleichen. Über die GPCR-Datenbank (<http://www.gpcr.org/7tm/> [67, 66]) werden mehrere Rezeptoren aus der Gruppe der Rhodopsine und der Gruppe der AR ausgewählt (s. Tab.5.3). Die Sequenzen werden aus der Swiss-Prot-Datenbank (<http://www.expasy.ch/sprot/sprot-top.html> [7, 30]) entnommen. Mit dem Programm ClustalW [62, 112] wird ein vorläufiges Alignment erstellt, welches mit dem Programm Jalview (<http://www2.ebi.ac.uk/~michele/jalview/contents.html>) im Hinblick auf das spätere Threading und experimentelle Hinweise manuell nachbearbeitet wird.

#### Ergebnis

**Alignment 1 (s. Anhang A.2.1)** zeigt das vollständige multiple Sequenz-Alignment, das mit ClustalW erzeugt wurde. Die Sequenz-Abschnitte sind gemäß der Sekundär-Struktur der KS des Rhodopsins als Helix- und Loop-Bereiche gekennzeichnet. Die Helices sind meist durchgängig und mit exakter Übereinstimmung der Referenz-AS (s. Abschnitt 5.1.1) aliniert. Generell sind die intra- und extrazellulären Loops der Rhodopsine und ARs von

*Tabelle 5.3.:* Über die GPCR-Datenbank und die SwissProt-Datenbank erhältliche Adenosin-Rezeptor-Sequenzen mit den Positionen, die eventuell modifiziert sind.

Rezeptor	SwissProt-Eintrag	Disulfid-Brücke	Lipid an Cys	Zucker	Länge
<b>Rhodopsin-Rezeptoren</b>					
OPSD_PIG	O18766	110:187	322, 323	2, 15	348
OPSD_TURTR	O62798	110:187	322, 323	2, 15	348
OPSD_BOVIN	P02699	110:187	322, 323	2, 15	348
OPSD_SHEEP	P02700	110:187	322, 323	2, 15	348
OPSD_HUMAN	P08100	110:187	322, 323	2, 15	348
OPSD_RABIT	P49912	110:187	322, 323	2, 15	348
OPSD_MOUSE	P15409	110:187	322, 323	2, 15	348
OPSD_RAT	P51489	110:187	322, 323	2, 15	348
OPSD_ALLMI	P52202	110:187	322, 323	2, 15	352
OPSD_CHICK	P22328	110:187	322, 323	2, 15	351
OPSD_AMBTI	Q90245	110:187	322, 323	2, 15	354
OPSD_ANGAN	Q90214	110:187	323	2, 15, 300	352
<b>A<sub>1</sub> Adenosin-Rezeptoren</b>					
AA1R_CAVPO	P47745	80:169	309	159	326
AA1R_HUMAN	P30542	80:169	309	159	326
AA1R_RAT	P25099	80:169	309	148, 159	326
AA1R_BOVIN	P28190	80:169	309	159	326
AA1R_CANFA	P11616	80:169	309	159	326
AA1R_RABIT	P34970	80:169	309	159	328
AA1R_CHICK	P49892	80:169	309	159	324
<b>A<sub>2</sub> Adenosin-Rezeptoren</b>					
AA2A_CANFA	P11617	77:166	-	145, 154	412
AA2A_HUMAN	P29274	77:166	-	154	412
AA2A_CAVPO	P46616	74:163	-	142, 151	409
AA2A_MOUSE	Q60613	74:161	-	142, 149	410
AA2A_RAT	P30543	74:161	-	142, 149	410
AA2B_MOUSE	Q60614	78:171	311	153, 163	332
AA2B_RAT	P29276	78:171	311	153, 163	332
AA2B_HUMAN	P29275	78:171	311	153, 163	332
AA2B_CHICK	O13076	76:171	310	151, 154	340
<b>A<sub>3</sub> Adenosin-Rezeptoren</b>					
AA3R_CANFA	Q28309	83:166	303	4, 12, 160	314
AA3R_HUMAN	P33765	83:166	303	3, 4, 12	318
AA3R_RABIT	O02667	84:167	304	4	319
AA3R_RAT	P28647	85:168	305	4, 5	320

vergleichbarer Länge. Lediglich die folgenden Unstimmigkeiten müssen vor dem anschließenden Threading durch Editieren des Alignments beseitigt werden:

1. Die N-Termini der Rhodopsine sind ca. 20 AS länger als die der Adenosin-Rezeptoren. Dementsprechend tauchen in diesem Bereich Lücken im Alignment der ARs auf.
2. Die Helix TM1 ist bei den ARs unterbrochen; mit 3 AS fehlt etwa eine Helix-Windung.
3. Zu Beginn der Helix TM4 taucht in der Gruppe der Rhodopsine eine Lücke von einer AS auf.
4. Vor allem im Bereich des EL2 unterscheidet sich die Länge der miteinander verglichenen Sequenzen stark. EL2 ist bei den Rhodopsinen generell kürzer als bei den AR. Die an der konservierten Disulfid-Brücke beteiligten Cystein-Reste des EL2 sind einander zwar innerhalb der Rhodopsin- bzw. AR-Gruppen jeweils übereinstimmend zugeordnet, jedoch zwischen den Gruppen versetzt.

**Alignment 2 (s. Anhang A.2.2)** zeigt nur die Sequenzen der humanen AR und der Struktur-Vorlage Rinder-Rhodopsin nach dem Editieren. Die Zuordnung der AS ist so geändert, dass Helix-Abschnitte durchgängig sind und Längen-Differenzen eher in die ohnehin variablen Loop-Regionen verlegt werden.

**zu 1 + 2:** Die Lücken bei den AR in den Bereichen des N-Terminus und der ersten Helix werden geschlossen, sodass ihre N-Termini nur noch 2-8 AS kurz sind (vgl. Rinder-Rhodopsin: 33 AS).

**zu 3:** Die bei den AR in Helix TM4 überzählige AS wird in den Loop-Bereich vorverlegt, sodass bei gleicher Helix-Länge IL2 bei den AR nun 11 statt 10 AS lang ist.

**zu 4:** Um die an der Disulfid-Brücke zu TM3 beteiligten Cysteine des EL2 mit dem entsprechenden Cystein-Rest des Rinder-Rhodopsins zur Deckung zu bringen, muss unmittelbar vor dem Beginn von TM5 eine Lücke in die Sequenz der ARs eingefügt werden. Dadurch bleibt im Fall des  $A_{2A}$ - bzw.  $A_{2B}$ -Rezeptors ein Abschnitt von 11 bzw. 15 AS ohne Entsprechung in der Struktur-Vorlage.

### 5.3.2. Threading

#### Durchführung

Die Sequenzen des  $A_{2A}$ - und des  $A_{2B}$ -Rezeptors werden mit Ali2pdb aus dem Topalign-Paket (<http://cartan.gmd.de/> [94]) gemäß dem editierten Alignment 2 (Fasta-Format) auf den PDB-Eintrag 1f88A<sup>5</sup> übertragen. Soweit eine Entsprechung vorhanden ist, werden die Koordinaten der Haupt- und Seitenketten-Atome aus der Vorlage in die neuen 3D-Modelle übernommen.

<sup>5</sup>Die kristallographisch asymmetrische Einheitszelle enthält zwei Rhodopsin-Moleküle, A und B.

## Ergebnis

Es resultieren 3D-Strukturen der  $A_{2A}$ - und  $A_{2B}$ -Rezeptoren mit Lücken im Hauptketten-Verlauf (wegen Längen-Differenzen zwischen Modell- und Vorlage-Sequenz). Auch Seitenketten sind teilweise unvollständig. Die Hauptkette muss an den folgenden Stellen überarbeitet werden:

1. **EL1** ist bei der Vorlage des Rinder-Rhodopsins zwei AS länger als bei den  $A_{2A}$  und  $A_{2B}$ -Rezeptoren. Diese Lücken in den Modellen sind zu schließen.
2. **EL2**: Sequenz-aufwärts von dem an der Disulfid-Brücke beteiligten Cys(5.27) sind beim  $A_{2A}$ -Rezeptor 11, beim  $A_{2B}$ -Rezeptor 15 AS zu ergänzen. Sequenz-abwärts von dem konservierten Cystein ist die Vorlage um 5 AS länger; dieser Abschnitt muss also verkürzt werden.
3. **IL1** ist bei den AR um eine AS länger als bei der Rhodopsin, muss also geringfügig erweitert werden.
4. **IL3** ist beim  $A_{2A}$ -AR 7 AS, beim  $A_{2B}$  3 AS länger als bei der Rhodopsin-Vorlage; dort ist er allerdings unvollständig, sodass in den Modellen für diesen Loop 11 bzw. 7 AS eingefügt werden müssen.
5. **EL3** ist beim  $A_{2A}$  um eine, beim  $A_{2B}$ -AR um zwei AS länger als bei der Rhodopsin-Vorlage, muss also auch verlängert werden.
6. **IT**: Die C-Termini von Rhodopsin und  $A_{2B}$ -Rezeptor sind gleich lang, die (unvollständige) Vorlage kann insofern leicht übernommen werden. Der C-Terminus des  $A_{2A}$ -Rezeptors ist jedoch um 82 AS länger. Auf die Modellierung des IT von  $A_{2A}$ - und  $A_{2B}$ -AR wird verzichtet, da der Aufwand gegenüber dem zu erwartenden Erkenntnis-Gewinn nicht gerechtfertigt zu sein scheint (Das Hauptaugenmerk liegt auf extrazellulärer und Transmembran-Domäne).

### 5.3.3. Vervollständigung der 3D-Modelle

#### Durchführung

Die rohen Modelle werden um die fehlenden Atome ergänzt und gleichzeitig strukturell optimiert. Einzelne AS werden in die Hauptkette eingefügt, indem die Koordinaten-Dateien zuerst editiert werden und dann (ggf. mehrmals) einer Energie-Minimierung unterzogen werden (jeweils 200 Schritte *conjugated gradient* mit  $\varepsilon = 2r$ ). Während der Energie-Minimierung wird entweder der gesamte Loop oder ein Abschnitt von mindestens 5 AS Sequenz-aufwärts und -abwärts von der kritischen Stelle beweglich gelassen. Der Rest des Proteins bleibt fixiert.



**Loop-Modellierung** Falls im Modell eine in der Vorlage vorhandene AS fehlt, sind die beiden Enden der dann unterbrochenen Hauptkette nun in der Sequenz unmittelbar benachbart. Ihre Peptid-Bindung ist folglich stark gedehnt und wird automatisch während der Struktur-Optimierung eines größeren Abschnitts verkürzt.

Im Fall einer Verlängerung werden die zusätzlichen AS nacheinander hinzugefügt, indem zunächst jeweils ein Glycin eingeschoben wird.<sup>6</sup> Dazu werden die drei Hauptketten-Atome N, C<sub>α</sub> und Carbonyl-C des Glycins zwischen zwei miteinander verbundenen AS platziert, die restlichen Glycin-Atome werden von CHARMM automatisch ergänzt (s. unten). Während der anschließenden Energie-Minimierung des Loops dehnen sich die Atom-Abstände auf ihr natürliches Maß aus und der Loop erweitert sich insgesamt. Anschließend wird die entsprechende Seitenkette gemäß der eigentlichen AS-Bezeichnung eingefügt und strukturell optimiert. Diese Prozedur wird für jede weitere einzufügende AS durchgeführt.

Geringfügige Verkürzungen oder Verlängerungen von relativ kurzen Loops ändern nur wenig am Verlauf ihrer Hauptkette und die resultierende Struktur gibt die tatsächlichen Verhältnisse sicherlich vernünftig wieder. Dies trifft zu für EL1 (8 → 6 AS), EL3 (8 → 9/10 AS), IL2 (10 → 11 AS) und IL3 (20 → 27/23 AS). Von den zu modellierenden Loop-Abschnitten ist aufgrund der großen Längendifferenz zwischen der Vorlage und den fraglichen Sequenzen nur EL2 kritisch.

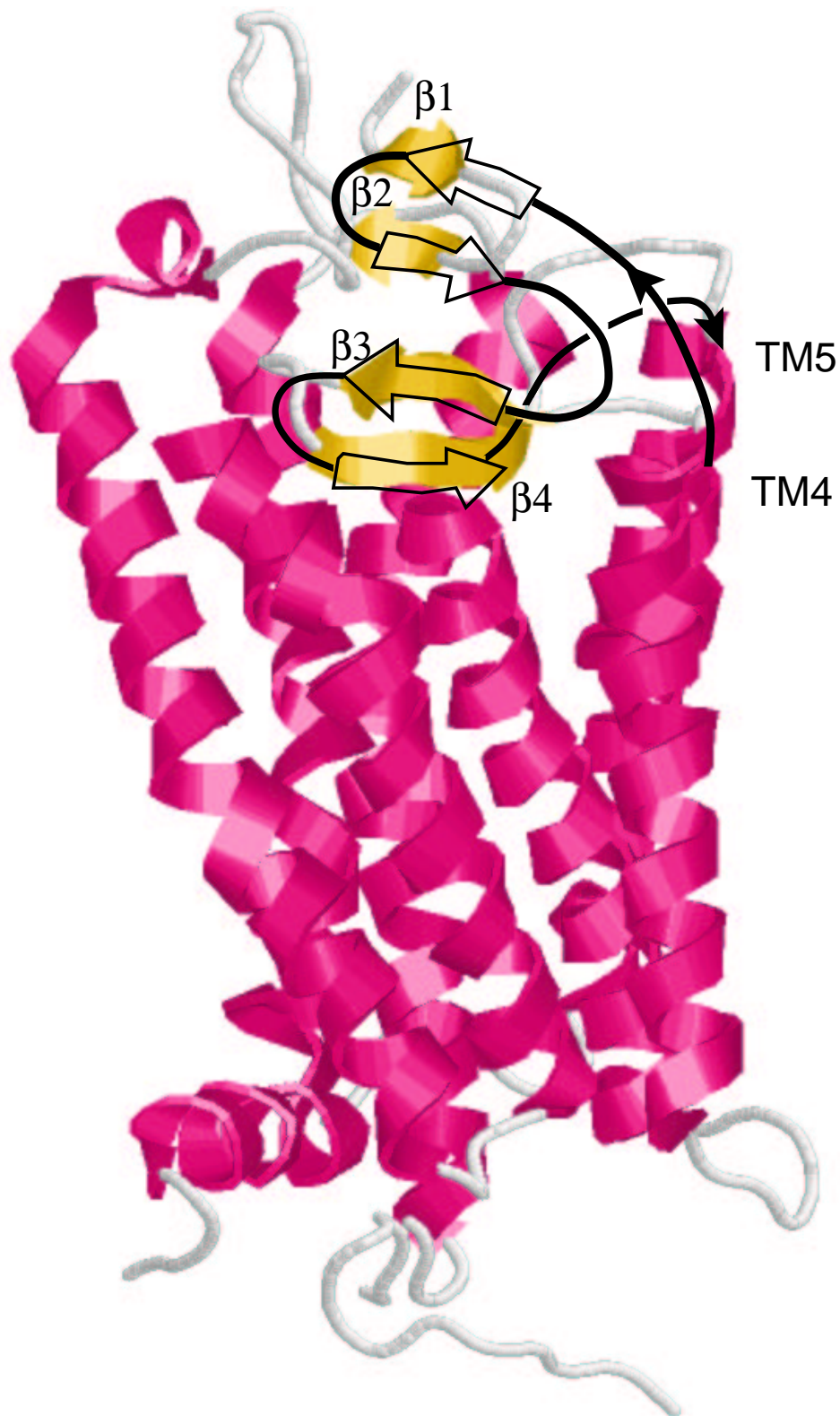
Die Konformation von EL2 ist gänzlich unbekannt, lediglich die konservierte Disulfid-Brücke zu TM3 ist ein Anhaltspunkt für den Verlauf der Hauptkette. Als *Arbeitshypothese* zur Modellierung dieses Loops wird angenommen, dass neben dem Helix-Bündel nicht nur die Disulfid-Brücke, sondern auch die viersträngige  $\beta$ -Faltblatt-Formation der extrazellulären Region für die allgemeine GPCR-Architektur charakteristisch ist (s. Abb. 5.3). Von der Vorlage werden also die  $\beta$ -Stränge  $\beta 3$  (Tyr178 – Glu181) und  $\beta 4$  (Ser186 – Ile189) als definierte Struktur-Elemente einschließlich des konservierten Brücken-Cysteins, dem editierten Alignment entsprechend, übernommen. Zwischen dem C-terminalen Ende von TM4 und  $\beta 3$  ist dann eine Sequenz von 11 bzw. 15 AS beim A<sub>2A</sub>- bzw. A<sub>2B</sub>-Rezeptor einzufügen, während das Verbindungsstück von  $\beta 4$  nach TM5 um fünf AS zu verkürzen ist. Der Hauptketten-Verlauf der verbleibenden 11 bzw. 15 AS wird ausgehend von TM4 und einmündend vor  $\beta 3$  entlang der  $\beta$ -Stränge 1 und 2 *des N-Terminus der Vorlage* geführt. Es soll also das Struktur-Motiv der vier  $\beta$ -Stränge unabhängig von der Sequenz-Abfolge nachgebildet werden (s. Abb. 5.7).

**Seitenketten-Modellierung:** Die Seitenketten der aus dem Threading übernommenen AS-Rümpfe bzw. die Seitenketten der hinzugefügten Glycin-Stellvertreter werden gemäß ihrer eigentlichen Bezeichnung automatisch von CHARMM ergänzt, allerdings mit fixen Torsionswinkeln,<sup>7</sup> die keine Rücksicht auf bevorzugte Rotamere oder Kollisionen mit benachbarten Atom-Gruppen Rücksicht nehmen. Daher kommt es häufig zu Kollisionen, die durch Energie-Minimierung nicht behoben werden können. In dem Fall bricht die Struktur-

---

<sup>6</sup>Wegen seiner kleinen Seitenkette kann man die Hauptkette zunächst mit Glycin erweitern, ohne dass Kollisionen zwischen der neuen Seitenkette und der Umgebung stören.

<sup>7</sup>Die Torsionswinkel betragen jeweils 180°, entsprechend der *anti*-Konformation gestreckter Alkylketten.



*Abbildung 5.7.:* Arbeitshypothese zur Modellierung von EL2: Die  $\beta$ -Stränge aus der Vorlage (hell) werden in das Modell übernommen (schwarze Pfeile und Linien). Der vor  $\beta 3$  befindliche Loop-Abschnitt der Adenosin-Rezeptoren ohne Struktur-Entsprechung wird entlang der  $\beta$ -Stränge 1 und 2 des N-Terminus des Rhodopsins geführt. Das Verbindungsstück zwischen  $\beta 4$  und TM5 wird gegenüber der Vorlage gekürzt. Der Übersichtlichkeit wegen ist EL1 hier nicht dargestellt.

Optimierung durch CHARMM automatisch ab. Dann wird die entsprechende AS durch die nächst kleinere Variante ersetzt, bis die umgebenden Seitenketten sich so aneinander angepasst haben, dass auch für die endgültige AS genug Platz zur Verfügung steht. Notfalls werden Rotamere manuell adjustiert, bis das gesamte Protein kollisionsfrei ist.

## Ergebnis

Das resultierende 3D-Modell des A<sub>2B</sub>-Rezeptors ist vollständig; beim A<sub>2A</sub>-Modell fehlen lediglich der intrazelluläre C-Terminus und vier AS des IL3 (siehe Abb. 5.8). Obwohl die Verbindungssequenz zwischen  $\beta$ 4 und TM5 in den Modellen gegenüber der Vorlage verkürzt ist, wird der durch die Disulfid-Brücke fixierte  $\beta$ -Strang nicht verzerrt. In dem frei modellierten Abschnitt von EL2 unterscheiden sie sich in Länge und Sequenz.

### 5.3.4. Diskussion

#### Mutagenese-Daten

Bevor auf Grundlage der erzeugten 3D-Modelle Komplexe des Rezeptors mit Liganden aufgebaut werden, müssen die Strukturen auf Übereinstimmung mit experimentellen Daten abgeglichen werden. Nach der allgemeinen Architektur der GPCRs sollte die Lage der Bindetaschen im Rinder-Rhodopsin und den AR ungefähr übereinstimmen. Also werden die AS der unmittelbaren Umgebung des 11-*cis*-Retinals (s. Seite 151) den für die Ligand-Bindung relevanten, aus Mutagenese-Studien für die Subtypen A<sub>1</sub>, A<sub>2A</sub> und A<sub>3</sub> bekannten AS gegenübergestellt (s. Tab. 5.4). Für den A<sub>2B</sub>-Rezeptor ist bislang erst eine solche Studie durchgeführt worden [11]. In Alignment 2 (Anhang A.2.2) sind diese AS jeweils farblich hervorgehoben.

Das 11-*cis*-Retinal hat in der KS des Rhodopsins zu 12 AS der Helices TM3, TM5, TM6 und TM7 unmittelbaren Kontakt. Passend dazu sind sieben AS mindestens bei einem der Subtypen A<sub>1</sub>, A<sub>2A</sub> bzw. A<sub>3</sub> für die Bindung von Liganden relevant, und zwar an den Positionen 3.32, 3.33, 3.36, 3.37, 6.52, 7.39 und 7.43. Berücksichtigt man zusätzlich, dass (bedingt durch die Ganghöhe von 3.6 AS pro Windung) bei  $\alpha$ -Helices jeweils die dritt- und viertnächsten Sequenz-Nachbarn auf der gleichen Seite liegen, befinden sich auch noch fünf weitere, ebenfalls Mutagenese-relevante AS (5.43, 6.55, 6.59, 7.35 und 7.36) im Bereich der Bindetasche. Der Boden der Bindetasche befindet sich etwa auf halber Höhe der Transmembran-Helices und Ser281(7.46) markiert ihren tiefsten Punkt. Dass so viele Mutagenese-relevante AS bei den 3D-Modellen an der Innenseite der Helix-Bündel liegen, bestätigt die modellierte Struktur der Transmembran-Region.

Nach Arbeiten von Kim et al. [76] und Gao et al. [139] ist auch der EL2 an Wechselwirkungen mit Liganden beteiligt: Innerhalb der AR-Familie sind negative Ladungen im EL2 konserviert. Glu151(EL2) ist beim A<sub>2A</sub> absolut erforderlich für hohe Affinität zu Liganden; Glu169(EL2) ist ebenfalls wichtig, an dieser Stelle wird aber auch Glutamin toleriert. Im A<sub>2A</sub>-Modell liegt dieses Glu169 vor dem Beginn von Helix TM5 am Ende des vierten  $\beta$ -Strangs, d. h. in einem aus der Vorlage übernommenen Struktur-Abschnitt. Die Seitenket-

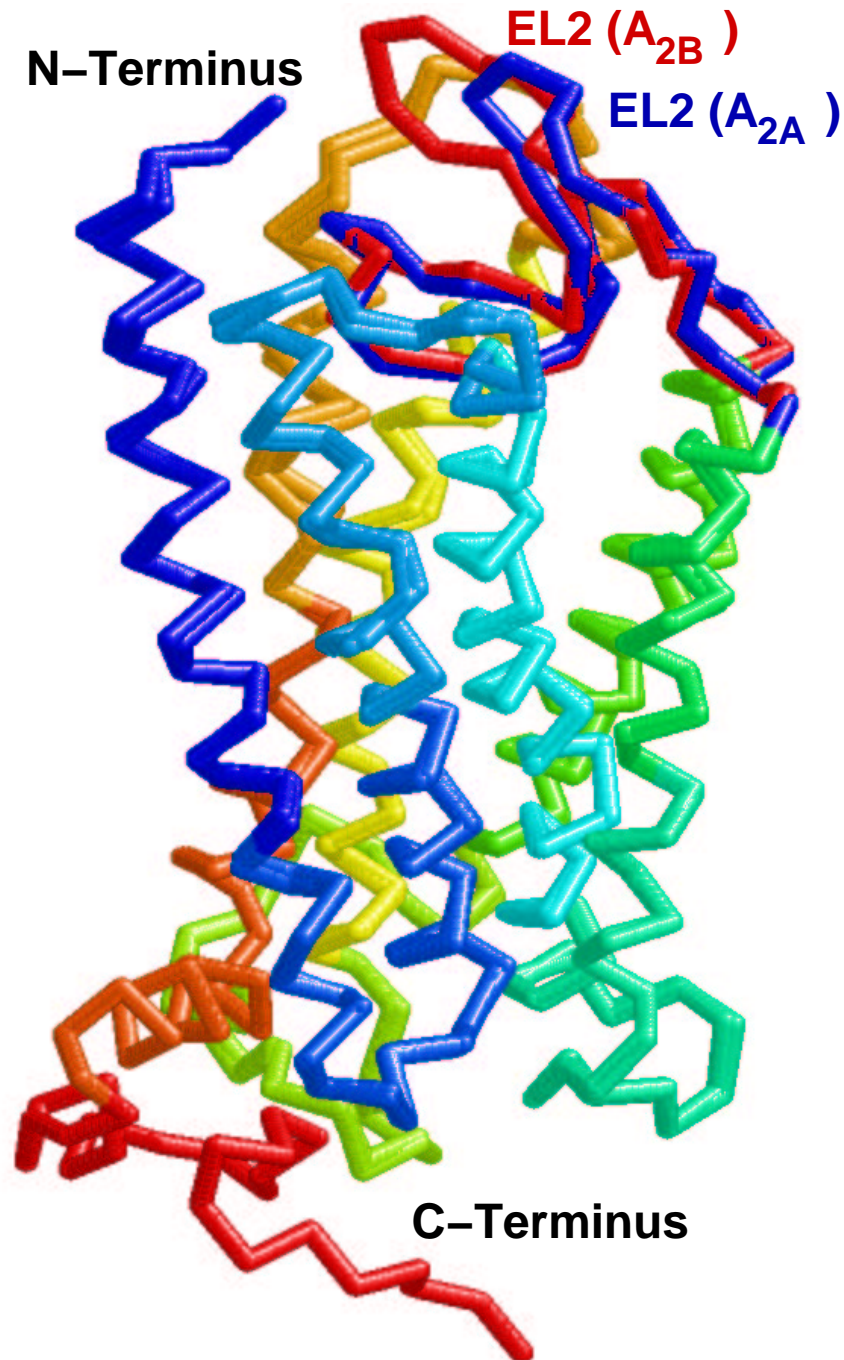


Abbildung 5.8.: Verlauf der Hauptketten von A<sub>2A</sub>- bzw. A<sub>2B</sub>-Rezeptor im 3D-Modell. EL2 ist gesondert hervorgehoben: Er ist beim A<sub>2B</sub>-Rezeptor länger als beim A<sub>2A</sub>-Rezeptor. Der C-Terminus des A<sub>2A</sub>-Rezeptors fehlt ganz.

*Tabelle 5.4.:* Übersicht über die AS, die nach Affinitäts- und Mutagenese-Studien für die Ligand-Bindung bei den verschiedenen Subtypen relevant sind. Wenn bekannt, sind auch die mutmaßlichen Wechselwirkungsgruppen der Agonisten bzw. Xanthin-Antagonisten angegeben.

	A <sub>1</sub>	A <sub>2A</sub>	A <sub>3</sub>		A <sub>1</sub>	A <sub>2A</sub>	A <sub>3</sub>
<b>1.37</b>	Gly14	Thr11	Thr17	<b>6.55</b>	Asn254	Asn253	Asn250
Agonist	[118]				N <sup>6</sup> -H [100]	[77]	[139]
Antagon.					N7-H [100]	[77]	[139]
<b>1.39</b>	Glu16	Glu13	Glu19	<b>6.59</b>	Leu258	Phe257	Tyr254
Agonist	[100]	[69, 49]				[77]	
Antagon.						[77]	
<b>2.50</b>	Asp55	Asp52	Asp58	<b>7.35</b>	Thr270	Met270	Leu264
Agonist	[100]		[139]		N <sup>6</sup> [130]		
Antagon.	[100]		[139]		C8 [130]		
<b>3.32</b>	Val87	Val84	Leu90	<b>7.36</b>	Tyr271	Tyr271	Tyr265
Agonist		[72]			[100]	[77]	
Antagon.		[72]			[100]	[77]	
<b>3.33</b>	Leu88	Leu85	Leu91	<b>7.39</b>	Ile274	Ile274	Ile268
Agonist	N <sup>6</sup> [118]						[77]
Antagon.	[118]						[77]
<b>3.36</b>	Thr91	Thr88	Thr94	<b>7.42</b>	Thr277	Ser277	Ser271
Agonist	N <sup>6</sup> [118]				Rib. [130]	Rib.[77]	
Antagon.	[118]						
<b>3.37</b>	Gln92	Gln89	His95	<b>7.43</b>	His278	His278	His272
Agonist	N <sup>6</sup> [118]		[139]		Rib.[130, 100]	Rib.[49, 72]	
Antagon.	[118]		[139]		[130]	[72]	
<b>3.39</b>	Ser94	Ser91	Ser97	<b>7.46</b>	Ser281	Ser281	Ser275
Agonist	[100]				[100]	Rib.[77]	
Antagon.	N1/N3[9]					[77]	
<b>5.43</b>	Phe185	Phe182	Phe182	<b>7.50</b>	Pro285	Pro285	Pro279
Agonist	[100]	[77]			[100]		
Antagon.	[100]	[77]					
<b>6.52</b>	His251	His250	Ser247				
Agonist		N <sup>6</sup> [118, 77]					
Antagon.	[100]	[77]					

te weist zur Bindetasche. Glu169 entspricht der Position 190 im Rinder-Rhodopsin und befindet sich daher direkt neben Ile189, welches in der KS Kontakt zum 11-*cis*-Retinal hat. Dies lässt den von der Vorlage übernommenen Teil von EL2 plausibel erscheinen.

Glu151 hingegen liegt im 3D-Modell des A<sub>2A</sub>-Rezeptors im „U“ zwischen den frei modellierten  $\beta$ -Strängen 1 und 2, ist also dem extrazellulären Raum zu- und der Bindetasche abgewandt. Der Verlauf dieses Loop-Abschnitts bleibt also im Fall des A<sub>2A</sub>-Rezeptors fraglich.

### Potentielle Glykosylierungsstellen

Tabelle 5.3 zeigt unter anderem, dass die A<sub>2A</sub>- bzw. A<sub>2B</sub>-Rezeptoren im EL2 an Asn154 bzw. Asn153 und Asn163 glykosyliert sein können. Mit Hilfe enzymatischer Behandlung ist zwar nicht nachzuweisen, dass A<sub>2B</sub>-Rezeptoren in epithelialen Zellen mit Zucker-Resten verknüpft sind; es ist jedoch auch nicht auszuschließen, denn im Striatum von Hunden kommt die Glykosylierung an A<sub>2B</sub>-Rezeptoren tatsächlich vor [41]. Das als Vorlage dienende Rinder-Rhodopsin hingegen ist am N-Terminus über Asn15 mit einem Disaccharid (N-Acetyl-D-Glucosamin und  $\alpha$ -D-Mannose) verbunden [107]. Im 3D-Modell des A<sub>2B</sub>-Rezeptoren ist die Amid-Gruppe von Asn153 im frei modellierten Abschnitt von EL2 nur 4.4 Å von der Amid-Gruppe des Asn15 der Vorlage entfernt und weist ebenfalls in den extrazellulären Raum, sodass eine Glykosylierung dort gut möglich erscheint, ein Indiz für eine passende Struktur des frei modellierten Abschnitts von EL2.

### Zwischenbilanz

Insgesamt stimmen die 3D-Modelle von A<sub>2A</sub>- und A<sub>2B</sub>-Rezeptor im Bereich der Transmembran-Helices mit den verfügbaren, leider nur indirekten Struktur-Informationen (aus dem Vergleich der KS des Rinder-Rhodopsins mit Mutagenese-Experimenten über die ARs) sehr gut überein. Auch die Modellierung des  $\beta 3/\beta 4$ -Abschnitts von EL2 einschließlich der konservierten Disulfid-Brücke erscheint zutreffend. Der frei modellierte  $\beta 1/\beta 2$ -Abschnitt scheint beim A<sub>2B</sub>-Rezeptor besser gelungen zu sein als beim A<sub>2A</sub>-AR. Da aber offenbar hauptsächlich der  $\beta 3/\beta 4$ -Abschnitt zur Bindetasche gehört, ist die Unsicherheit im frei modellierten Teil für die Platzierung von Liganden in der zentralen Bindetasche vermutlich von untergeordneter Bedeutung. Der Zugang zu dieser Bindestelle von der extrazellulären Seite aus erscheint als enger Kanal, der geeignet erscheint, die Hydrat-Hülle vom Liganden abzustreifen.

### Meta-Bindestellen

Die Bindung eines Liganden an die zentrale Bindestelle ist wahrscheinlich ein mehrstufiger Prozess [97]. Das bedeutet, dass ein Ligand nicht nur zur zentralen Bindestelle, sondern auch zu anderen Bereichen der Rezeptor-Öffnung eine merkliche Affinität aufweisen muss. Solche so genannten Meta-Bindestellen werden in der Literatur für verwandte Purinozeptoren beschrieben [97]. Meta-Bindestellen können durch kombinierte Docking- und Energie-Rechnungen identifiziert werden. Dies setzt eine sehr hohe Zuverlässigkeit der 3D-Modelle

voraus. Durch die hohe anzunehmende Beweglichkeit extrazellulärer Loops und eventuelle Glykosylierungen in diesem Bereich ist die genaue Struktur dieser Region jedoch ungewiss. Aus diesem Grund wird davon abgesehen, im Rahmen dieser Arbeit potentielle Meta-Bindestellen zu untersuchen.

### Vergleich mit 3D-Modell aus der GPCR-Datenbank

Mit dem hier erzeugten 3D-Modell des A<sub>2B</sub>-Rezeptors liegt eine Struktur vor, die mit den experimentellen Anhaltspunkten wesentlich besser im Einklang steht, als das über die GPCR-DB verfügbare, automatisch generierte Modell von G. Vriend [55]. Neben einigen Abweichungen in der Sequenz fällt gravierend auf, dass gegenüber dem individuell modellierten Rezeptor bei dem mit WHATIF erzeugten Modell die an der Ligand-Bindung auch beteiligte fünfte Helix um 25 AS verschoben ist. EL2 wird damit sehr kurz und IL2 sehr lang; die an der Bindetasche beteiligten AS von TM5 sind völlig verschoben. Darüber hinaus sind die AS der für die Ligand-Bindung wichtigen siebten Helix um eine Position versetzt zugeordnet. Das bedeutet, dass diese Helix in jenem Modell um ca. 100° gegenüber dem selbst erzeugten Modell verdreht ist und somit die Mutagenese-relevanten AS an die interhelikalen Kontakt-Regionen geraten statt zur Bindetasche zu weisen.

Während der einzelnen Arbeitsschritte sind experimentelle Hinweise eingeflossen. Es ist daher kaum verwunderlich, dass das hier erstellte Modell mit diesen besser übereinstimmt, als ein ohne diese Zusatzinformationen automatisch generiertes Modell. Man könnte argumentieren, die gute Übereinstimmung sei unter diesen Bedingungen keine Gewähr für die tatsächliche Qualität der Modelle. Dem ist entgegenzusetzen, dass das dem Threading zugrunde liegende Alignment in den wesentlichen Bereichen, d. h. bei den Helices, unverändert beibehalten wird (bis auf TM1<sup>8</sup>). Nur die peripheren Bereiche sind gezielt manipuliert worden, ohne jedoch den Charakter des ursprünglichen Alignments anzutasten.

## 5.4. Komplex-Konstruktion

Da der Transmembran-Kern der Adenosin-Rezeptoren zutreffend modelliert zu sein scheint, sollen nun Liganden innerhalb der Bindetasche platziert werden, um genauere Informationen über die Wechselwirkungen zwischen den Rezeptoren und ihren jeweiligen Liganden zu gewinnen. Insbesondere sollen die AS der Bindetasche den speziellen Kontakt-Regionen (s. S. 158) zugeordnet werden. Als Liganden (s. Tab. 5.9) werden kleine und große Agonisten (Adenosin, CGS-21680) bzw. Antagonisten (ZM-241385, Xanthin, 1<sup>9</sup> Xanthin 2<sup>10</sup>) ausgewählt; die Agonisten deswegen, weil auch die Komplexe unter Verwendung von experimentellen Hinweisen konstruiert werden müssen,<sup>11</sup> und nur vom natürlichen Agonisten

<sup>8</sup>Wie sehr dieser Lückenschluss im Alignment bei TM1 berechtigt ist, zeigt sich darin, dass in den Modellen Glu(1.39) nur dann, wie von IJzerman et al. [69] postuliert, in die Nachbarschaft von His(7.43) gelangt.

<sup>9</sup>1-Propyl-8-phenylxanthin

<sup>10</sup>1-Butyl-8-[4-([4-benzyl]piperazino-2-oxoethoxy)phenyl]xanthin

<sup>11</sup>Versuche, sie automatisch durch Docking mit FLEXX zu erzeugen sind fehlgeschlagen. Sämtliche Struktur-Vorschläge waren nicht im Zentrum der sieben Helices, sondern an ihrer Außenseite lokalisiert.

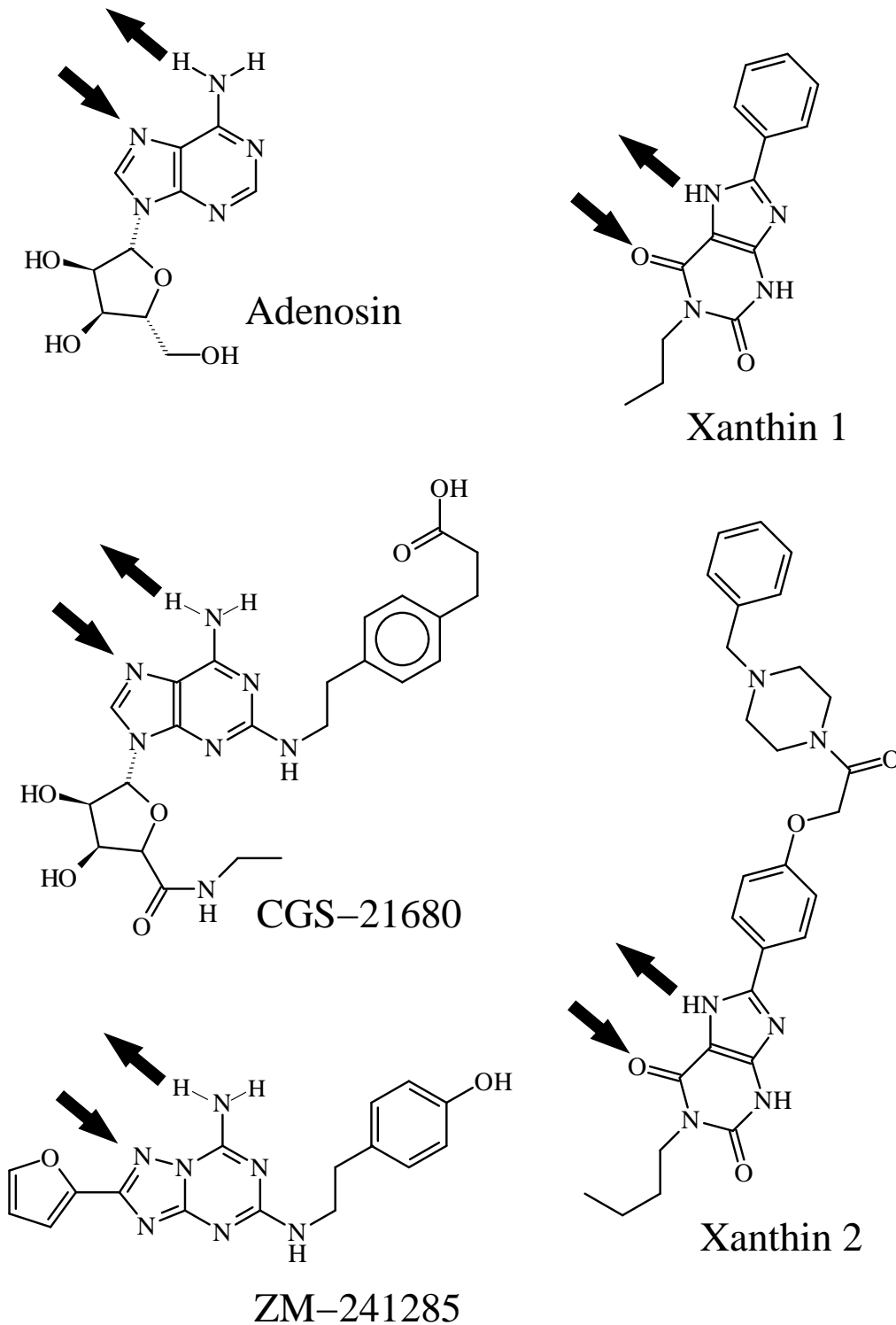


Abbildung 5.9.: Übersicht über die Liganden, die in der zentralen Bindetasche der 3D-Modelle des A<sub>2A</sub>- bzw. A<sub>2B</sub>-Rezeptors platziert werden sollen.



Adenosin das relevante Rotamer und eine Hypothese zum Bindemodus (s. S. 154) bekannt ist. Antagonisten wiederum sollen platziert werden, weil sie, im Gegensatz zu Agonisten, keine Konformationsänderungen des Rezeptors auslösen und demzufolge besonders gut in die dem Ruhezustand eines GPCRs entsprechenden Modelle passen sollten. Zu beiden Xanthin-Derivaten, besonders aber zu Xanthin 2 besitzt der  $A_{2B}$ -Rezeptor eine deutlich höhere Affinität als der  $A_{2A}$ -Rezeptor [58], sodass aus den Komplex-Strukturen eventuell die Ursachen dafür erkennbar werden. Auf Grundlage der Xanthin-Antagonisten können außerdem später eventuell Crosslinker entworfen werden, die die Validierung der Vorhersagen erlauben. Agonisten und Antagonisten werden dem  $N^6$ - $N7$ -Modell (s. S. 159) entsprechend überlagert.

### Durchführung

Alle Manipulationen an Atom-Positionen werden durch Editieren der Atom-Koordinaten bzw. durch Translations- und Rotations-Operationen vorgenommen. Nach jeder Änderung werden die Strukturen durch eine Energie-Minimierung optimiert (gegebenenfalls mehrfach: 200 Schritte *conjugated gradient* mit  $\epsilon=2r$ ). Treten unvermeidbare Kollisionen mit dem Protein auf, werden Seitenketten, wie zuvor geschildert (s. S. 27), entweder vorübergehend verkleinert oder andere Rotamere eingestellt, bis der Komplex kollisionsfrei ist.

**Vorbereitung des Protein-Anteils** Unter der Annahme, dass die Ribose-Einheit des Adenosins tatsächlich, wie in Abbildung 5.4 gezeigt, über H-Brücken an His(7.43) verankert ist, wird zunächst diese Seitenkette passend zu Glu(1.39) ausgerichtet.

**Vorbereitung der Liganden** Die Molekül-Strukturen der Liganden und ihre Ladungsverteilungen werden unter Quanta erstellt.

**Einpassen der Liganden in die Bindetasche** Zuerst wird die Ribose-Einheit am Grund der Bindetasche des  $A_{2A}$ -Rezeptors in der Nähe von His(7.43) platziert und zwar so, dass beim *anti*-Rotamer des Adenosins die Adenin-Einheit später in Richtung der Bindetaschen-Öffnung zu liegen kommt. Es zeigt sich, dass der Abstand der  $C_\alpha$ -Atome des His(7.43) zum Mutagenese-relevanten Asn(6.55) mit 15.9 Å so groß ist, dass das Adenosin-Molekül in *anti*-Konformation gut zwischen den beiden AS platziert werden kann. Durch manuelle Ausrichtung der Asn-Seitenkette lassen sich Komplexe konstruieren, bei denen das Adenosin-Molekül sowohl über die von IJzerman [69] vorgeschlagene His-Glu-Kette mit der Ribose-Einheit, als auch über ein H-Brücken-Paar mit der Adenin-Gruppe am Rezeptor zu verankern ist (s. Abb. 5.10). Da das Adenosin-Molekül als starr betrachtet wird, und die Bindetasche recht eng ist, kommen keine alternativen Platzierungen unter gleichzeitiger Beachtung experimenteller Anhaltspunkte in Betracht.

Nach dem Einpassen des Adenosins werden die einander entsprechenden Atome der anderen Liganden auf die erhaltenen Adenosin-Koordinaten übertragen. Im Fall der Xanthin-Derivate sind dies nach dem  $N^6$ - $N7$ -Modell nur die Atome des vicinalen H-Brücken-Donor/Akzeptor-Paares, bei ZM-241385 ist es das zentrale Ring-System.

### 5.4.1. Ergebnisse

**Komplexe mit Adenosin** Der erzeugte  $A_{2A}$ -Komplex zeigt in sehr guter Übereinstimmung mit den Erwartungen [100] mehrere H-Brücken zwischen dem Adenosin und den folgenden AS des Rezeptors: Von der 2'-OH-Gruppe zum amidischen H-Atom von Cys171(EL2), von der 3'-OH-Gruppe zum N $\epsilon$ 2-Atom des His278(7.43), von der N<sup>6</sup>-Amino-Gruppe zum Amid-O-Atom von Asn253(6.55) und von der NH<sub>2</sub>-Gruppe desselben Asn253(6.55) zum Adenin-N7-Atom zurück. Die 5'-OH-Gruppe der Ribose ist nur 2.9 Å von der OH-Gruppe von Ser277(7.42) entfernt, allerdings ohne dass in diesem Modell eine H-Brücke zwischen diesen Gruppen vorliegt (was prinzipiell aber möglich zu sein scheint, s. Abb. 5.10). Alle an H-Brücken zum Adenosin-Molekül beteiligten Seitenketten sind durch Mutagenese-Studien als essentiell für die Agonist-Bindung identifiziert worden (s. Tabelle 5.4). Die Ribose-Einheit ist außerdem dicht umgeben von Ile80(3.28), Val84(3.32), Trp246(6.48), Leu249(6.51) und Ile274(7.39). Die Adenin-Einheit wird flankiert von Seiten- und Hauptketten-Atomen von Gly158(EL2), Gly160(EL2), Ala165(EL2), Cys166(EL2), Leu167(EL2), Leu249(6.51), Thr256(6.58) und Ile274(7.39).

In den  $A_{2B}$ -Rezeptor kann Adenosin fast ebenso hineingesetzt werden (*rmsd*-Wert zur Position im  $A_{2A}$ -Rezeptor nach der Struktur-Optimierung: 2.28 Å). Das Adenosin-Molekül ist dann ebenfalls über ein H-Brücken-Paar ausgehend von der NH<sub>2</sub>-Gruppe des Asn254(6.55) zum N7-Atom des Adenins und von der N<sup>6</sup>-Amino-Gruppe des Adenins zum Amid-O-Atom des Asn254(6.55), sowie von der 2'-OH-Gruppe zu His280(7.43) an das Protein gebunden. Die ursprünglich vorgegebene H-Brücke der Ribose-Einheit zu His280(7.43) besteht nach der Energie-Minimierung nicht mehr, wohl aber eine enge Nachbarschaft der 5'-OH-Gruppe zur Seitenkette von Ser279(7.42). Die Ribose wird umgeben von Val85(3.32), Cys246(6.47), Trp247(6.48), Ser279(7.42), Asn282(7.45) und Ile276(7.39). Ile276(7.39) hat auch Kontakt zur Adenin-Gruppe, ebenso wie Ser165(EL2), Cys171(EL2), Leu172(EL2), Val250(6.51) und Met272(7.35).

**Komplexe mit CGS-21680** Das CGS-21680-Molekül mit seinem großen Substituenten am C2-Atom ist manuell nur schwierig in der Bindetasche unterzubringen und entfernt sich während der Energie-Minimierung von der Ausgangsposition. Im  $A_{2A}$ -Rezeptor bildet es eine H-Brücke über die CO<sub>2</sub><sup>-</sup>-Gruppe zu Glu261(6.64). Die vorgegebene H-Brücke der 2'-OH-Gruppe zu His278(7.43) bleibt erhalten, eine weitere entsteht zwischen der 3'-OH-Gruppe und dem Carbonyl-O-Atom von Ser277(7.42). Das ursprünglich vorhandene H-Brücken-Paar zu Asn253(6.55) wird während der Struktur-Optimierung gelöst.

Beim  $A_{2B}$ -Rezeptor hat CGS-21680 über die CO<sub>2</sub><sup>-</sup>-Gruppe Kontakt zur NH<sub>3</sub><sup>+</sup>-Gruppe von Lys267(EL3). Außerdem besteht eine H-Brücke zwischen seiner 3'-OH-Gruppe und dem Carbonyl-O-Atom von Ser279(7.42), ansonsten existieren nur ungerichtete Kontakte.

**Komplexe mit ZM-241385** Trotz der großen Übereinstimmung der molekularen Strukturen von Adenosin und ZM-241385 verliert dieser Ligand während der Struktur-Optimierung in beiden Rezeptoren den anfangs bestehenden Kontakt zu Asn253/254(6.55), vermutlich durch Kollision des Phenyl-Rings mit den  $\beta$ 3- und  $\beta$ 4-Strängen von EL2. Da ZM-241385

über mehrere H-Brücken-Donor und -Akzeptor-Funktionen verfügt, ist allerdings nicht sicher, ob das Molekül tatsächlich in der angenommenen Orientierung am Rezeptor bindet.

**Komplexe mit Xanthin 1** Die Komplexe mit dem kleinen 1-Propyl-8-phenyl-xanthin behalten ihre Start-Konformation ungefähr bei. Die optimale Ausrichtung bezüglich des H-Brücken-Paares geht allerdings während der Struktur-Optimierung beim  $A_{2A}$ -Rezeptor weitgehend verloren, im  $A_{2B}$ -Rezeptor hingegen bleibt sie gut erhalten. Bei beiden besteht zusätzlich ein schwacher Kontakt des N7-gebundenen H-Atoms zum Hauptketten-O-Atom von Gly163(EL2,  $A_{2A}$ ) bzw. zur OH-Gruppe von Ser165(EL2,  $A_{2B}$ ). Die Region um das N3-Atom ist in beiden Rezeptoren von der Wurzel einer Seitenkette aus dem  $\beta 4$ -Abschnitt von EL2 benachbart; beim  $A_{2A}$ -AR handelt es sich um die kurze Seitenkette von Ala165, im Fall des  $A_{2B}$ -AR um die lange Seitenkette von Lys170.

**Komplexe mit Xanthin 2** Das große Molekül von Xanthin 2 besetzt in beiden Rezeptoren etwa die gleiche Region (*rmsd* zwischen beiden Platzierungen 2.0 Å). Der endständige Phenyl-Ring liegt in beiden Fällen zwischen dem frei modellierten Loop-Abschnitt von EL2 und EL3 und ragt damit fast über den Eingang der Bindetasche hinaus. Im  $A_{2A}$ -Rezeptor bewegt sich das Xanthin-Ringsystem während der Energie-Minimierung unter Verlust des H-Brücken-Paares mit Asn253(6.55) von seiner Ausgangsposition weiter weg als im  $A_{2B}$ -Rezeptor. Demzufolge herrschen im Komplex mit dem  $A_{2A}$ -Modell nur noch ungerichtete Kontakte zwischen Ligand und Protein. Im Gegensatz dazu bleibt im Komplex mit dem  $A_{2B}$ -Rezeptor eine schwache Wechselwirkung des N7-H-Atoms zum Amid-O-Atom von Asn254(6.55) bestehen. Außerdem bildet hier die  $NH_3^+$ -Gruppe von Lys267(EL3) eine H-Brücke zum Ether-O-Atom des Liganden aus. Der endständige Phenyl-Ring im  $A_{2B}$ -Rezeptor wird von der unpolaren Seitenkette von Cys154(EL2) flankiert, während sich im  $A_{2A}$ -Rezeptor die kationische Seitenkette von Lys153(EL2) darüberlegt, die gleichzeitig mit dem gegenüber liegenden Ser263(EL3) eine H-Brücke bildet. Der Piperazin-Ring von Xanthin 2 wird im  $A_{2B}$ -Rezeptor von Pro157(EL2) umgeben, während er im  $A_{2A}$ -Rezeptor (bedingt durch den kürzeren EL2) frei, d. h. solvatisiert, bleibt.

### Strukturelle Gemeinsamkeiten

Tabelle 5.5 zeigt zusammenfassend die AS, die im jeweiligen Komplex dicht, d. h. im Abstand von maximal 4.0 Å, am betreffenden Liganden liegen. Die AS Val84/85-(3.32), A165/K170(EL2) C166/171(EL2), L167/172(EL2), W246/247(6.46), I274/276-(7.39), S277/279(7.42), H278/280(7.43) befinden sich im Bereich der Ribose-Gruppen, während die AS Leu249/Val250(6.51), M270/272(7.35), N253/254(6.55), I274/276(7.39), C166/171(EL2), L167/172(EL2) G160/Ser165(EL2) die zentrale Bindestelle für aromatische Ringe bilden.

Tabelle 5.5.: AS im Abstand von max. 4.0 Å zum jeweiligen Liganden. Die in beiden Rezeptor-Typen unterschiedlichen AS sind hervorgehoben. Mutagenese-relevante AS sind mit ! markiert.

	Adenosin		CGS-21680		ZM-241385		Xanthin1		Xanthin2	
	A <sub>2A</sub>	A <sub>2B</sub>	A <sub>2A</sub>	A <sub>2B</sub>	A <sub>2A</sub>	A <sub>2B</sub>	A <sub>2A</sub>	A <sub>2B</sub>	A <sub>2A</sub>	A <sub>2B</sub>
TM3					C77	C78				
TM3					L78	L79				
TM3					I80	L81				
TM3					A81	A82				
TM3!	V84	V85	V84	V85	V84	V85	V84	V85	V84	V85
TM3					L85			TM3		
EL2									<b>G152</b>	<b>N153</b>
EL2									<b>K153</b>	<b>C154</b>
EL2										P157
EL2										W158
EL2									<b>N154</b>	<b>D159</b>
EL2					<b>Q157</b>	<b>T162</b>				
EL2	G160			N163				N163		N163
EL2	<b>G162</b>	<b>S165</b>		S165						
EL2	<b>A165</b>	<b>K170</b>			<b>A165</b>	<b>K170</b>	<b>A165</b>	<b>K170</b>	<b>A165</b>	<b>K170</b>
EL2	C165	C171	C165	C171	C165	C171				
EL2	L166	L172	L166	L172	L166	L172	L166	EL2	L172	L166
EL2			F168	F173						
EL2			<b>D170</b>	<b>N175</b>			<b>D170</b>	<b>N175</b>	<b>D170</b>	<b>N175</b>
EL2									V171	V176
TM6			L244	L245		L245				
TM6			C245	C246	C245	C246				
TM6	W246	W247	W246	W247	W246	W247	W246	W247	W246	W247
TM6			P248	P249	P248	P249				
TM6	<b>L249</b>	<b>V250</b>	<b>L249</b>	<b>V250</b>	<b>L249</b>	<b>V250</b>	<b>L249</b>	<b>V250</b>	<b>L249</b>	<b>V250</b>
TM6			<b>I252</b>	<b>V253</b>			<b>I252</b>	<b>V253</b>	<b>I252</b>	<b>V253</b>
TM6!	N253	N254	N253	N254			N253	N254	N253	N254
TM6!			T256	T257			T256	T257	T256	T257
TM6			E261				E261		E261	A262
EL3									<b>C262</b>	<b>G264</b>
EL3									<b>S263</b>	<b>K265</b>
EL3										N266
EL3				K267				K267	<b>A265</b>	<b>K267</b>
EL3										P268
TM7							<b>L269</b>	<b>A271</b>		
TM7!	M270	M272	M270	M272		M272	M270	M272	M270	M272
TM7			A273	A275	A273	A275	A273	A275	A273	A275
TM7!	I274	I276	I274	I276	I274	I276	I274	I276	I274	I276
TM7			L276	L278	L276	L278				
TM7!	S277	S279	S277	S279	S277	S279	S277	S279	S277	S279
TM7!	H278	H280	H278	H280			H278	H280	H278	H280
TM7			N280	N282						

### 5.4.2. Diskussion

Bedingt durch die gewählte Methode des manuellen, starren Platzierens der Liganden, erlauben die erzeugten Komplex-Strukturen nur qualitative Aussagen. Die Komplexe zeigen, ob sich ein Ligand tatsächlich im vermuteten Bindemodus platzieren lässt; eine Bewertung der Platzierung nach der Energie oder sogar nach der Bindestärke ist nicht möglich. Die Komplex-Strukturen liefern Hinweise darauf, welche AS in der Nachbarschaft der Liganden liegen könnten und aus diesem Grund für weitere Untersuchungen interessant sind. Bei den Antagonisten ist zu beachten, dass abgesehen von einer Platzierung nach dem N<sup>6</sup>-N7-Modell auch alternative Bindemodi in Betracht kommen können, hier aber nicht erfasst sind.<sup>12</sup>

**Komplex-Strukturen** Adenosin und Xanthin1 sind im Vergleich zu CGS-21680, ZM-241385 und Xanthin2 kleine Liganden, größtmäßig gut mit dem in der Vorlage gebundenen 11-*cis*-Retinal vergleichbar. Sie lassen sich erfolgreich im erwarteten Bindemodus in den Bindetaschen beider Rezeptoren einpassen. Das Adenosin-Molekül hat in beiden Modellen Kontakt zu TM3, TM6, TM7 und EL2. Abbildung 5.10 illustriert am A<sub>2A</sub>-Modell, wie Adenosin im Rezeptor liegt; demnach wäre Asn253(6.55) der Gegenpart zum H-Brücken-Donor/-Akzeptor-Paar der Liganden (für den A<sub>1</sub>-Rezeptor ist die Nähe dieser Seitenkette zur N<sup>6</sup>H<sub>2</sub>-Gruppe der Agonisten bzw. zur N7H-Gruppe der Antagonisten belegt [100]). Die 2'-OH-Gruppe der Ribose bildet eine H-Brücke zur Hauptkette von Cys171(EL2), die 3'-OH-Gruppe eine H-Brücke zur Seitenkette von His278(7.43)<sup>13</sup> und auch die 5'-OH-Gruppe liegt in H-Brücken-Reichweite zur OH-Gruppe von Ser277(7.42). Die Ribose-Einheit ist demnach über drei Kontakt-Gruppen in der Bindetasche verankert. So erklärt sich, dass für die Agonist-Bindung eine intakte Ribose-Einheit erforderlich ist, am 5'-C-Atom aber auch Substituenten toleriert werden. Die große Zahl der H-Brücken zu Mutagenese-relevanten AS zeigt, dass dieses Modell die zentrale Bindetasche zuverlässig wiedergibt, insbesondere auch im Bereich des  $\beta$ 4-Strangs des EL2, dessen Zuordnung im Alignment manuell nachbearbeitet ist.

Xanthin1 lässt sich dem N<sup>6</sup>-N7-Modell entsprechend gut in beiden Rezeptoren unterbringen; jedoch können mangels geeigneter Donor- bzw. Akzeptor-Gruppen neben dem H-Brücken-Paar mit Asn(6.55) keine weiteren nennenswerten, gerichteten Wechselwirkungen eingegangen werden. Die Platzierung der größeren Liganden erscheint weniger zuverlässig, da das manuelle Einfügen starrer Moleküle der Flexibilität dieser Liganden nicht gerecht wird. Abbildung 5.11 zeigt dennoch neben Adenosin und Xanthin 1 die Lage von Xanthin 2 im Modell des A<sub>2B</sub>-Rezeptors. Man erkennt, dass der voluminöse Substituent von Xan-

<sup>12</sup>Das N<sup>6</sup>-N7-Modell setzt voraus, dass die Ankergruppe für das H-Brücken-Donor/-Akzeptor-Paar der Liganden starr ist. Lässt sich die Amid-Gruppe von Asn(6.55) aber um 180° drehen, könnten Ringsysteme von Agonisten und Xanthin-Antagonisten trotz der dann vertauschten Donor/-Akzeptor-Funktionen dennoch deckungsgleich binden. Dies hätte eine gewisse Ambivalenz im Substitutionsmuster von Antagonisten zur Folge.

<sup>13</sup>Bemerkenswert ist, dass sowohl bei der Vorlage Rhodopsin, als auch bei den AR der AS an Position 7.43 eine entscheidende Rolle bei der Fixierung des Agonisten zukommt. Möglicherweise ist diese Position entscheidend an der Reiz-Aufnahme zum Konformationswechsel beteiligt.

thin2 zwischen EL3 und dem frei modellierten Teil von EL2 liegt, also den Eingang zur Bindetasche ausfüllt. Bedingt durch die beim  $A_{2A}$ -Rezeptor im Vergleich zum  $A_{2B}$ -Rezeptor kürzere Schlaufe des EL2 ist die Öffnung zur Bindetasche im  $A_{2A}$ -Modell weiter als beim  $A_{2B}$ -Modell.

Interessanterweise behält das Xanthin-Ringsystem der beiden Xanthin-Derivate jeweils beim  $A_{2B}$ -Rezeptor die vorgegebene Start-Struktur mit dem H-Brücken-Paar zu Asn(6.55) besser bei, als beim  $A_{2A}$ -Modell. Beide Verbindungen haben zum  $A_{2B}$ -Rezeptor eine höhere Affinität als zum  $A_{2A}$ -Rezeptor [58].

**Selektivitäten** Als Verursacher für die bemerkenswerten Unterschiede in den Affinitäten der beiden Rezeptor-Subtypen zum Agonisten Adenosin kommen zunächst die voneinander abweichenden AS in der nächsten Umgebung von Adenosin bzw. Xanthin1 in Betracht (s. Tab. 5.5). Dies sind A165/K170 (EL2), D170/N175 (EL2), L249/V250 (TM6), I252/V253 (TM6), Ala265/Lys267 (EL3) und Leu269/Ala271 (TM7). Die unpolaren AS Val, Leu, Ile und Ala sind vermutlich kaum an den Selektivitäten der Rezeptoren beteiligt, da Val, Leu, Ile und Ala als unpolare AS keine gerichteten Wechselwirkungen mit ihrer Umgebung eingehen und sich einem Liganden daher leichter anpassen können als durch H-Brücken fixierbare Seitenketten. Zudem weist in allen Fällen der  $A_{2B}$ -Rezeptor mit Valin eine kleinere Seitenkette als der  $A_{2A}$ -Rezeptor auf, seine Bindetasche ist also im Vergleich zum  $A_{2A}$ -Rezeptor nicht eingengt. Der Austausch von Asp170 gegen Asn175 im konservierten  $\beta$ 4-Abschnitt des EL2 könnte eher bedeutsam sein [76, 139], da negative Ladungen im EL2 des  $A_{2A}$ -AR offensichtlich konserviert sind. Allerdings wird vom  $A_{2A}$ -AR auch Gln toleriert, sodass auch damit die so stark unterschiedlichen Affinitäten nicht erklärt werden können. Die anionische Carboxylat-Gruppe von Asp170 könnte jedoch für die niedrigere Affinität des  $A_{2A}$ -AR zu den Xanthin-Derivaten 1 und 2 mit verantwortlich sein, da sie Kontakt zum C8-gebundenen Phenyl-Ring dieser beiden Verbindungen hat. Das im  $A_{2B}$ -AR an dieser Position vorhandene Asn175 wird leichter desolvatisiert und eignet sich daher besser als Nachbar für den hydrophoben Phenyl-Ring.

Die bisher genannten AS-Paare sind einander in Größe und Polarität recht ähnlich. Ala165 und Lys170 in EL2 bzw. Ala265 und Lys267 in EL3 hingegen unterscheiden sich in dieser Beziehung deutlich voneinander. In der Literatur wird allerdings nur das Paar Leu267( $A_{2A}$ ) bzw. Lys269( $A_{2B}$ ) als mit verantwortlich für den Affinitätsunterschied zwischen  $A_{2A}$ - und  $A_{2B}$ -Rezeptor diskutiert [79]. Zum Adenosin-Molekül haben in den Modell-Komplexen jedoch weder Ala265/Lys267 noch Leu267/Lys269 Kontakt.

Doch Ala165( $A_{2A}$ ) und Lys170( $A_{2B}$ ) liegen jeweils nahe am C2-Atom des Adenins bzw. am N3-Atom von Xanthin 1. Ihre Positionen sind durch die unmittelbare Nachbarschaft zu Cys166/171(EL2), welches die Disulfid-Brücke zu TM3 ausbildet, weitgehend festgelegt.<sup>14</sup> Die kationische Kopfgruppe von Lys170(EL2) bildet im Modell des  $A_{2B}$ -AR mit der Carbonyl-Gruppe von Ala64(2.61) eine tief im Protein-Inneren verborgene H-Brücke. Der  $A_{2A}$ -AR bevorzugt Agonisten mit Substituenten am C2-Atom [41, 117]. Der  $A_1$ -AR dagegen, der

<sup>14</sup>Auch das entsprechende Ser186 des Rinder-Rhodopsins hat in der KS Kontakt zum Liganden 11-*cis*-Retinal.

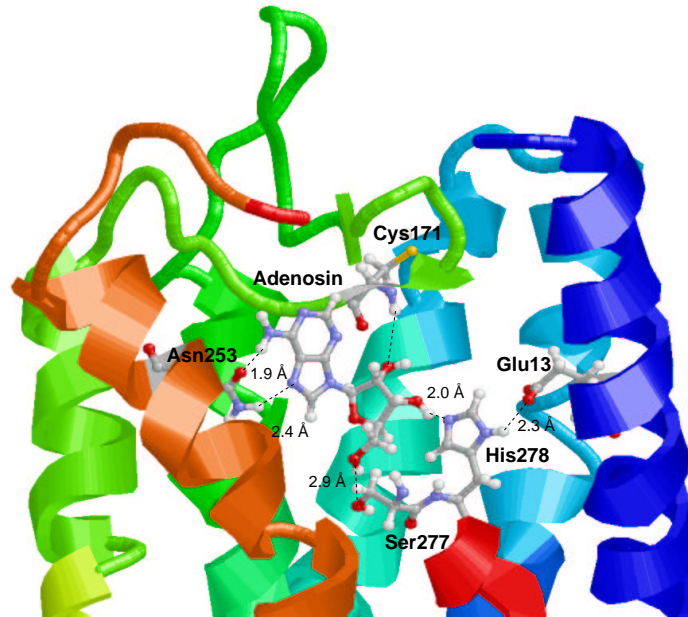


Abbildung 5.10.: Einblick in die Bindetasche des A<sub>2A</sub>-Rezeptors mit manuell platziertem Adenosin nach Energie-Minimierung.

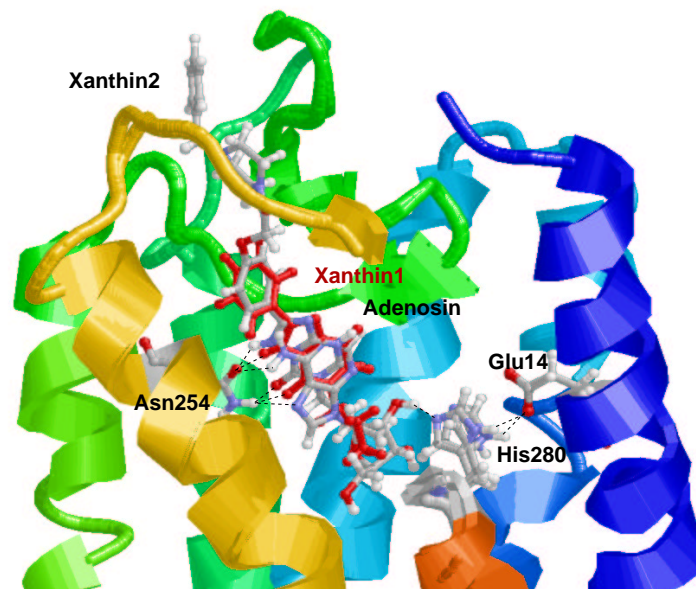


Abbildung 5.11.: Einblick in die Bindetasche des A<sub>2B</sub>-Rezeptors mit Adenosin und den beiden Xanthin-Derivaten nach Energie-Minimierung (Xanthin1 ist rot markiert, Helix TM6 teilweise nicht dargestellt).

wie der  $A_{2B}$ -Rezeptor an der fraglichen Position einen Lysin-Rest besitzt, bindet eher  $N^6$ -substituierte Agonisten [117]. Bei Antagonisten wird eine Bevorzugung von Xanthin-Derivaten ohne N3-Substituenten durch den  $A_{2B}$ -AR beobachtet [58]. Das Paar Ala165( $A_{2A}$ ) bzw. Lys170( $A_{2B}$ ) könnte also das unterschiedliche Verhalten der beiden Subtypen gegenüber C2-substituierten Agonisten bzw. N3-substituierten Xanthin-Antagonisten bedingen.

Abgesehen von direkten Protein-Ligand-Wechselwirkungen scheint auch ein Unterschied in der Kontakt-Fläche zwischen den Helices TM1 und TM7 an den Affinitätsunterschieden beteiligt zu sein: Die Gruppe von Jacobson stellte fest, dass ein Austausch von Tyr271(7.36) gegen Ala beim  $A_{2A}$ -Rezeptor sowohl die Agonist- als auch die Antagonist-Bindung stört (weder Bindung von CGS-21680 noch von XAC<sup>15</sup>); ein Austausch gegen einen anderen aromatischen Rest (Phe oder His) erhält jedoch die Affinität zu Agonisten und Antagonisten [77].

Im 3D-Modell liegt dieser Tyrosin-Ring fast parallel zum Ring des konservierten Tyr9(1.32) (s. Abb. 5.12). Auf der entgegengesetzten Seite könnte Tyr9(1.32) zusätzlich vom aromatischen Ring-System von Trp268(7.33) bedeckt werden. So wäre es in einem dreifachen  $\pi$ -Stapel von zwei Seiten durch aromatische Ringe flankiert.<sup>16</sup>

Beim  $A_{2B}$ -Rezeptor befindet sich statt eines Tyrosin-Rings die Seitenkette von Asn273(7.36) in der interhelikalen Kontakt-Region zu TM1. Tyr10(1.32) kann also nicht mehr wie beim  $A_{2A}$  zwischen zwei Aromaten eingeschoben werden, sodass eine Lockerung des Helix-Helix-Kontaktes zu erwarten ist. Passenderweise erhöht umgekehrt der Austausch von Asn273(7.36) gegen Tyrosin die Affinität des  $A_{2B}$ -Rezeptors zu CGS-21680 [11]. Da besonders Ser(7.42) und His(7.43) aus TM7 essentiell an der Bindung der Ribose-Einheit beteiligt sind, ist vermutlich eine Destabilisierung der Gesamt-Struktur für die niedrigen Affinitäten und das schlechtere Ansprechverhalten des  $A_{2B}$ -Rezeptors auf Agonisten zurückzuführen.

## 5.5. Zusammenfassung und Ausblick

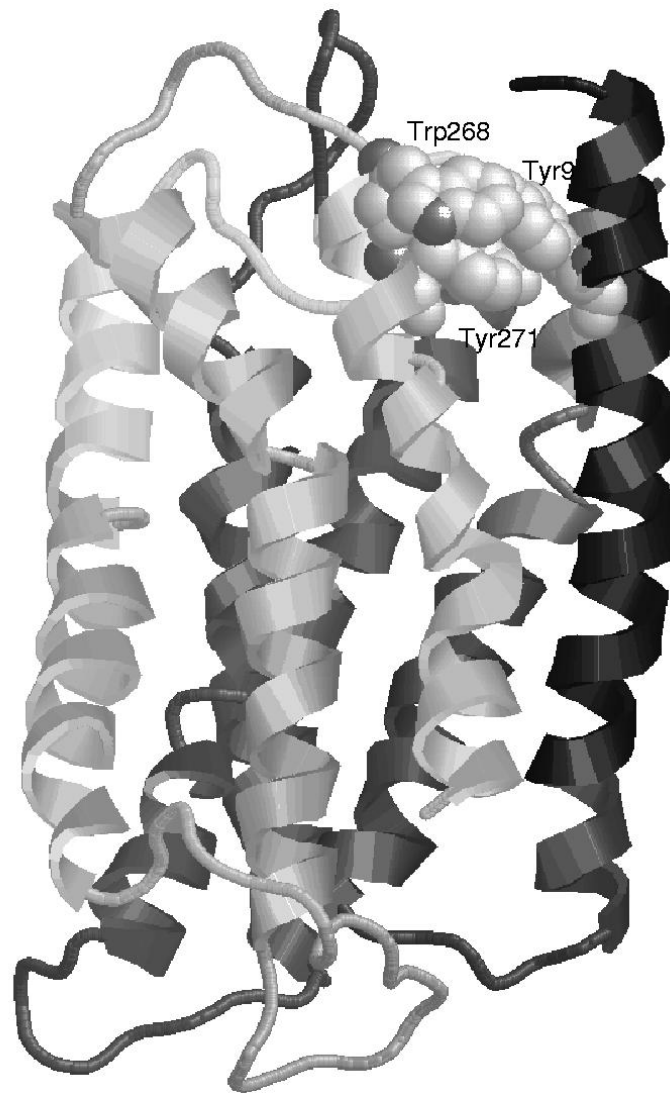
Es sind 3D-Modelle erzeugt worden, die sich gut mit den experimentellen Befunden decken. Dies zeigt, wie fruchtbar sich das Homologie-Konzept bei hinreichender Ähnlichkeit zwischen Vorlage und Modell<sup>17</sup> mit den angewendeten Verfahren umsetzen lässt. Die 3D-Modelle liefern Hinweise auf die oberflächlich gelegenen Seitenketten und damit auf mögliche Wechselwirkungspartner zwischen Ligand und Rezeptor. Ohne Korrelation mit experimentellen Befunden bleiben die auf Grundlage der 3D-Strukturen gezogenen Schlussfolgerungen jedoch weitgehend spekulativ. Die berechneten Positionen einzelner Seitenketten sollten für gezielte Crosslinking-Experimente und/oder Mutagenese-Studien genutzt werden.

<sup>15</sup>Xanthin-amin-congener

<sup>16</sup>Die 3D-Struktur lehnt sich insgesamt noch sehr stark an die Vorlage an, sodass diese gegenseitige Stabilisierung der Helix-Positionen nur als Hypothese formuliert werden kann. Bei Rhodopsin herrschen zwischen den Helices TM1 und TM7 nur einfache hydrophobe Kontakte zwischen zwar unpolaren, aber nicht aromatischen Seitenketten. Möglicherweise ist aber die Gesamt-Struktur durch Kontakte des längeren N-Terminus zu EL2 und EL3 stabilisiert.

<sup>17</sup>Im vorliegenden Fall: jeweils 53% Sequenz-Homologie und ca. 20% Sequenz-Identität





*Abbildung 5.12:* Die aromatischen Ringe von Trp268(7.33), Tyr9(1.32) und Tyr271(7.36) liegen einander so nahe, dass sie sich zu einem dreifachen  $\pi$ -Stapel übereinander legen könnten und so die Helices TM1 (dunkel, rechts) und TM7 (hellgrau, mitte) miteinander verzahnen.

### Vorschläge für Crosslinker

Die zentrale Bindetasche des  $A_{2A}$ -Adenosin-Rezeptors ist durch Mutagenese-Studien bereits gut bekannt. Die Charakterisierung der extrazellulären Region, die den Eingang zur Bindetasche bildet, ist aber nur schwer über Affinitäts- bzw. Mutagenese-Studien zugänglich. Eine Validierung der vorhergesagten Modelle in diesem Bereich scheint jedoch durch Crosslinking-Experimente gut möglich (s. Abschnitt 1.4.3). Im folgenden werden (ungeachtet möglicher Synthesewege) einige Moleküle vorgeschlagen, mit deren Hilfe ein Rezeptor zu vermessen ist. Die Vorschläge orientieren sich an den beiden bereits beschriebenen Xanthin-Derivaten [58]. Xanthin 1 besetzt die zentrale Bindetasche, das wesentlich längere Xanthin 2 ragt über diesen Bereich weit hinaus (die terminale Phenyl-Gruppe liegt zwischen EL2 und EL3). Generell sollten die Crosslinker-Moleküle zwei photoaktivierbare Gruppen tragen, um einerseits spezifische Bindung von unspezifischer unterscheiden zu können und

andererseits Aussagen über Entfernungen innerhalb des Moleküls machen zu können. Es werden jeweils Moleküle unterschiedlicher Länge konstruiert, um alle "Etagen" der Bindetasche bzw. ihres Eingangs erreichen zu können. Tabelle 5.6 zeigt die Crosslinker und die AS, die gemäß der berechneten 3D-Modelle von A<sub>2A</sub>- bzw. A<sub>2B</sub>-Rezeptor innerhalb der Reichweite der reaktiven Gruppen liegen, sowie die Abstände zwischen den beiden reaktiven Gruppen innerhalb der Crosslinker.

Allen vorgeschlagenen Strukturen gemeinsam ist eine Verankerung tief in der zentralen Bindetasche über eine Diazocarbonyl-Gruppe an N3<sup>18</sup>. Die reaktive Gruppe liegt dann in unmittelbarer Nachbarschaft von Gly160/Ser165(EL2), Val164/169(EL2), Ala165/Lys170(EL2), Cys166/171(EL2), Leu167/172(EL2) und Ile274/276(7.39). Da die Seitenkette von Lys170 im A<sub>2B</sub>-Rezeptor die Bindung von Antagonisten mit einem solchen Substituenten an N3 vermutlich stört [58], sind solche Crosslinking-Experimente möglicherweise mit einer Mutante des A<sub>2B</sub>-AR, die dort eine kleinere Seitenkette trägt, leichter durchführbar (s. unten).

### Vorschläge für Mutagenese-Studien

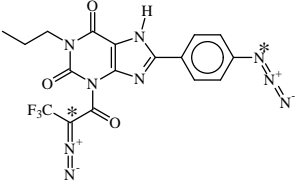
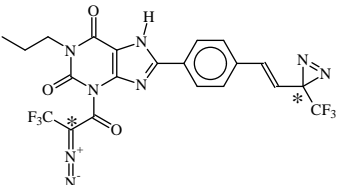
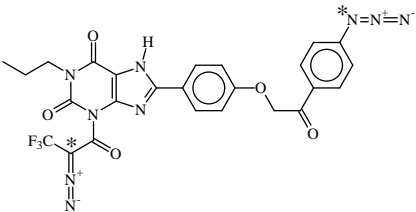
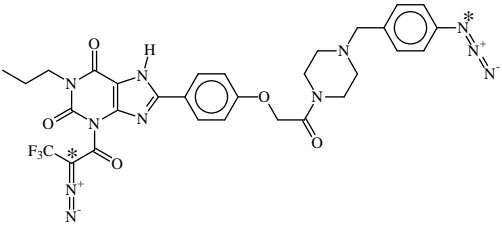
Die Abweichungen in den Sequenzen von A<sub>2A</sub>- und A<sub>2B</sub>-AR verursachen vermutlich über zwei Mechanismen die unterschiedlichen Rezeptor-Eigenschaften: Zunächst beeinflussen einzelne AS der Bindetasche direkt über den unmittelbaren Kontakt zu den Liganden die Affinität. Darüber hinaus sind periphere AS durch die Beeinflussung der Gesamt-Struktur des Rezeptors indirekt an der Gestaltung der Bindetasche beteiligt. Vergleichende Mutagenese-Studien zwischen A<sub>2A</sub>- und A<sub>2B</sub>-AR könnten beitragen, diese Wirkungsweisen voneinander zu separieren und die die Eigenschaften determinierenden AS zu identifizieren.

**Veränderung der Bindetasche** Vermutlich die lange Seitenkette von Lys170 stört beim A<sub>2B</sub>-Rezeptor die Bindung Xanthin-Derivaten mit N3-Substituenten bzw. von Adenosin-Derivaten mit C2-Substituenten. Daher könnte ein Austausch dieses Lysins gegen Glycin oder Alanin die Bindung von Xanthin-Derivaten mit größeren Substituenten in 3-Position erlauben. Ein solcher Austausch wäre vermutlich hilfreich, um für Crosslinking-Experimente eine photoaktivierbare Gruppe tief in die zentrale Bindetasche bringen zu können (s. oben).

**Veränderung des interhelikalen Kontaktes zwischen TM1 und TM7** Die Hypothese der über einen dreifachen Aromaten-Stapel (Tyr271(7.36)–Tyr9(1.32)–Trp268(7.33)) miteinander verzahnten Helices kann geprüft werden, indem beim A<sub>2A</sub>-AR neben dem bereits untersuchten Tyr271(7.36) auch Tyr9(1.32) oder Trp268(7.33) gegen nicht aromatische Seitenketten ausgetauscht werden. Jede solche Veränderung von Tyr9(1.32) oder Trp268(7.33) müsste die Affinität des Rezeptors zu Agonisten und Antagonisten ähnlich wie die bekann-

<sup>18</sup>An der alternativen N1-Position werden nur n-Propyl- bzw. n-Butyl-Substituenten toleriert, nicht aber Carbonyl-Gruppen.

Tabelle 5.6.: Abstände zwischen den reaktiven (\*) Gruppen der Crosslinker und Aminosäuren in Reichweite der reaktiven Gruppe des Substituenten an C8.

Crosslinker	Abstand	AS(A <sub>2A</sub> )	AS(A <sub>2B</sub> )
	ca. 10 Å	Asp170(EL2) Ile252(6.54) Asn253(6.55) Thr256(6.58) Glu261(EL3) Met270(7.35)	Asn163(EL2) Asn175(EL2) Val176(EL2) Thr257(6.58) Lys267(EL3) Met272(7.35)
	ca. 13 Å	Lys153(EL2) Val171(EL2) Thr256(6.58) Glu261(EL3) Ala265(EL3) Met270(7.35)	Pro157(EL2) Lys267(EL3)
	ca. 16 Å	Lys153(EL2) Asn154(EL2) Val171(EL2) Glu261(EL3) Cys262(EL3) Ser263(EL3) Ala265(EL3)	Cys158(EL2) Asp159(EL2) Val176(EL2) Ala262(EL3) Gln263(EL3) Gly264(EL3) Lys265(EL3) Lys267(EL3)
	ca. 21 Å	Gly152(EL2) Lys153(EL2) Asn154(EL2) Val171(EL2) Cys262(EL3)	Asn153(EL2) Cys154(EL2) Thr155(EL2) Asp159(EL2) Gly264(EL3) Lys265(EL3)

te Tyr271Ala-Mutation [77] verringern; falls nicht, würde damit ein direkter Kontakt von Tyr271 zum Liganden belegt.

Die N273Y-Mutante des A<sub>2B</sub>-Adenosin-Rezeptors zeigt eine höhere Affinität zu CGS-21680 als der Wild-Typ [11]. Die schon für den A<sub>2A</sub>-AR vorgeschlagenen Veränderungen, also der Austausch von Tyr10(1.32) bzw. Trp270(7.33) an dieser Mutante könnten zeigen, inwieweit sich die beiden Rezeptoren (WT-A<sub>2A</sub> bzw. N273Y-A<sub>2B</sub>) parallel verhalten und welche Rolle dieser interhelikale Kontakt für die Ligand-Bindung bzw. die Rezeptor-Struktur tatsächlich spielt.

Falls die aus den 3D-Modellen abgeleiteten Überlegungen zutreffen, kann durch den Austausch von Lys170(EL2) gegen Ala und Asn273(7.36) gegen Tyr vom A<sub>2B</sub>-Rezeptor eine Doppel-Mutante erzeugt werden, deren zentrale Bindestelle ähnliche Eigenschaften aufweist, wie die des A<sub>2A</sub>-Rezeptors.

# A. Anhang

## A.1. Abkürzungsverzeichnis

AC	Adenylatcyclase
AIDS	acquired immune deficiency syndrome
AR	Adenosin-Rezeptor
AS	Aminosäure
ATP	Adenosin-triphosphat
CA	Capsid
cAMP	zyklisches Adenosin-monophosphat
CHD	zyklische Harnstoff-Derivate
COP	Cyclooctylpyranon-Derivate
DA	Dopamin
DAG	Diacylglycerin
DNS	Desoxy-Ribonucleinsäure
EL	extrazellulärer Loop
ER	endoplasmatisches Reticulum
GDP	Guanosin-diphosphat
GMP	Guanosin-monophosphat
GPCR	G-Protein gekoppelter Rezeptor
GTP	Guanosin-triphosphat
HIV	humanes Immunschwäche-Virus
HIV-PR	HIV-Protease
IL	intrazellulärer Loop
IN	Integrase
IP <sub>3</sub>	Inositol-1,4,5-trisphosphat
KS	Kristall-Struktur
LJ	Lennard-Jones
L-DOPA	L-Dihydroxyphenylalanin
MA	Matrix-Protein
MD	Molekulardynamik
MS	Massenspektroskopie
NC	Nucleocapsid
NMR	nuclear magnetic resonance (Kernspin-Resonanz-Spektroskopie)

PAH	Phenylalanin-Hydroxylase
PLC	Phospholipase C
PR	Protease
RKSA	Röntgen-Kristall-Struktur-Analyse
RNS	Ribonucleinsäure
RT	Reverse Transkriptase
SAS	solvent accessible surface
SU	surface protein
TH	Tyrosin-Hydroxylase
TM	Transmembran
TPH	Tryptophan-Hydroxylase
WT	Wild-Typ
WW	Wechselwirkung

## A.2. Sequenz-Alignments

### A.2.1. Multiples Sequenz-Alignment

Alignment 1: Multiples Sequenz-Alignment, erzeugt mit ClustalW (1.8). Referenz-AS und Cysteine, die die hoch konservierte Disulfid-Brücke bilden, sind unterstrichen. Mit '\*', ':' und '.' sind konstante bzw. mehr oder weniger konservativ veränderte AS gekennzeichnet.

```

OPSD_PIG      MNGTEGPNFYVPFSNKTGVVRSUPFUEYUPQYYLAE|PWQFSMLAAYMFLLIVLGFPINUFLUTLY
OPSD_TURTR   MNGTEGLNFYVPFSNKTGVVRSUPFUEYUPQYYLAE|PWQFSVLAAYMFLLIVLGFPINUFLUTLY
OPSD_BOVIN   MNGTEGPNFYVPFSNKTGVVRSUPFUEAPUQYYLAE|PWQFSMLAAYMFLLIMLGFPINUFLUTLY
OPSD_SHEEP   MNGTEGPNFYVPFSNKTGVVRSUPFUEAPUQYYLAE|PWQFSMLAAYMFLLIVLGFPINUFLUTLY
OPSD_HUMAN   MNGTEGPNFYVPFSNATGVVRSUPFUEYUPQYYLAE|PWQFSMLAAYMFLLIVLGFPINUFLUTLY
OPSD_RABIT   MNGTEGPDFYIPMSNQTGVVRSUPFUEYUPQYYLAE|PWQFSMLAAYMFLLIVLGFPINUFLUTLY
OPSD_MOUSE   MNGTEGPNFYVPFSNVTGVGRSUPFUEQUPQYYLAE|PWQFSMLAAYMFLLIVLGFPINUFLUTLY
OPSD_RAT     MNGTEGPNFYVPFSNITGVVRSUPFUEQUPQYYLAE|PWQFSMLAAYMFLLIVLGFPINUFLUTLY
OPSD_ALLMI   MNGTEGPDFYIPFSNKTGVVRSUPFUEYUPQYYLAE|PWKYSALAAAYMFLLIILGFPINUFLUTLY
OPSD_CHICK   MNGTEGQDFYVPMSNKTGVVRSUPFUEYUPQYYLAE|PWKFSALAAAYMFLLILLGFPVUNFLUTLY
OPSD_AMBTI   MNGTEGPNFYVPFSNKSUGVVRSUPFUEYUPQYYLAE|PWQYSVLAAYMFLLILLGFPVUNFLUTLY
OPSD_ANGAN   MNGTEGPNFYIPMSNITGVVRSUPFUEYUPQYYLAE|PWAYTILAAAYMFTLILLGFPVUNFLUTLY
AA1R_CAVPO   -----MUPHSVS----AF|QAAYIGIEVLIALVSVPG---NUVLVIW
AA1R_HUMAN   -----MUPPSIS----AF|QAAYIGIEVLIALVSVPG---NUVLVIW
AA1R_RAT     -----MUPPYIS----AF|QAAYIGIEVLIALVSVPG---NUVLVIW
AA1R_BOVIN   -----MUPPSIS----AF|QAAYIGIEVLIALVSVPG---NUVLVIW
AA1R_CANFA   -----MUPPAIS----AF|QAAYIGIEVLIALVSVPG---NUVLVIW
AA1R_RABIT   -----MUPPSIS----AF|QAAYIGIEVLIALVSVPG---NUVLVIW
AA1R_CHICK   -----MUAQSVT----AF|QAAYISIEVLIALVSVPG---NUILVIW
AA2A_CANFA   -----MUSTMG----SW|--VYITVELAIAVLAILG---NUVLVCW
AA2A_HUMAN   -----MUPMIG----SS|--VYITVELAIAVLAILG---NUVLVCW
AA2A_CAVPO   -----MUS-----SS|--VYITVELVIAVLAILG---NUVLVCW
AA2A_MOUSE   -----MUMG-----SS|--VYIMVELAIAVLAILG---NUVLVCW
AA2A_RAT     -----MUMG-----SS|--VYITVELAIAVLAILG---NUVLVCW
AA2B_MOUSE   -----MUQLETQ----DA|--LYVALELVIAALAVAG---NUVLVCA
AA2B_RAT     -----MUQLETQ----DA|--LYVALELVIAALAVAG---NUVLVCA
AA2B_HUMAN   -----MULLETQ----DA|--LYVALELVIAALSVAG---NUVLVCA
AA2B_CHICK   -----MUMNTM----KT|--TYIVLELIIAVLSIAG---NUVLVCW
AA3R_CANFA   -----MUAVNGTALLLAN|-VTYITVEILIGLCAIVG---NUVLVIW
AA3R_HUMAN   -----MUPNNSTALSLAN|-VTYITMEIFIGLCAIVG---NUVLVIC
AA3R_RABIT   -----MUPDNSTTLFLAI|RASYIVFEIVIGCAVVG---NUVLVIW
AA3R_RAT     -----MUKANNTTTSALWL|QITYITMEEAIGLCAVVG---NUMULVUW
                                                    |   :   :   : *   *.
<-----N-Terminus----->|<-----Helix I----->
```

```

OPSD_PIG      VTVQ|HKKLRT|PLNYILLNLAVADLFMVFGGFTTTLYTS|LHGYFVFGPT|GCNLEGFFATLG
OPSD_TURTR   VTVQ|HKKLRT|PLNYILLNLAVANLFMVFGGFTTTLYTS|LHAYFVFGPT|GCNLEGFFATLG
OPSD_BOVIN   VTVQ|HKKLRT|PLNYILLNLAVADLFMVFGGFTTTLYTS|LHGYFVFGPT|GCNLEGFFATLG
OPSD_SHEEP   VTVQ|HKKLRT|PLNYILLNLAVADLFMVFGGFTTTLYTS|LHGYFVFGPT|GCNLEGFFATLG
OPSD_HUMAN   VTVQ|HKKLRT|PLNYILLNLAVADLFMVLGGFTSTLYTS|LHGYFVFGPT|GCNLEGFFATLG
OPSD_RABIT   VTVQ|HKKLRT|PLNYILLNLAVADLFMVLGGFTTTLYTS|LHGYFVFGPT|GCNVEGFFATLG
OPSD_MOUSE   VTVQ|HKKLRT|PLNYILLNLAVADLFMVFGGFTTTLYTS|LHGYFVFGPT|GCNLEGFFATLG
OPSD_RAT     VTVQ|HKKLRT|PLNYILLNLAVADLFMVFGGFTTTLYTS|LHGYFVFGPT|GCNLEGFFATLG
OPSD_ALLMI   VTVQ|HKKLRS|PLNYILLNLAVADLFMVLGGFTTTLYTS|MNGYFVFGVT|GCYFEGFFATLG
OPSD_CHICK   VTIQ|HKKLRT|PLNYILLNLVVDLFMVFGGFTTTMYTS|MNGYFVFGVT|GCYIEGFFATLG
OPSD_AMBTI   VTIQ|HKKLRT|PLNYILLNLAFANHFMVFGGFPVTMYSS|MHGYFVFGQT|GCYIEGFFATMG
OPSD_ANGAN   VTIE|HKKLRT|PLNYILLNLAVANLFMVFGGFTTTVYTS|MHGYFVFGPT|GCNLEGYFATLG
AA1R_CAVPO   AVKV|NQALRD|ATFCFIASLAVADVAVGALVIPLAILIN|IG--PQTYFH|TCLMVACPVLIL
AA1R_HUMAN   AVKV|NQALRD|ATFCFIVSLAVADVAVGALVIPLAILIN|IG--PQTYFH|TCLMVACPVLIL
AA1R_RAT     AVKV|NQALRD|ATFCFIVSLAVADVAVGALVIPLAILIN|IG--PQTYFH|TCLMVACPVLIL
AA1R_BOVIN   AVKV|NQALRD|ATFCFIVSLAVADVAVGALVIPLAILIN|IG--PRTYFH|TCLKVACPVLIL
AA1R_CANFA   AVKV|NQALRD|ATFCFIVSLAVADVAVGALVIPLAILIN|IG--PRTYFH|TCLMVACPVLIL
AA1R_RABIT   AVKV|NQALRD|ATFCFIVSLAVADVAVGALVIPLAILIN|IG--PETYFH|TCLMVACPVLIL
AA1R_CHICK   AVKM|NQALRD|ATFCFIVSLAVADVAVGALVIPLAIIN|IG--PQTEFY|SCLMMACPVLIL
AA2A_CANFA   AVWL|NSNLQN|VTNYFVVS�AAADIAAVGVLAIPFAITIS|TG--FCAACH|NCLFFACFVLVL
AA2A_HUMAN   AVWL|NSNLQN|VTNYFVVS�AAADIAAVGVLAIPFAITIS|TG--FCAACH|GCLFIACFVLVL
AA2A_CAVPO   AVWI|NSNLQN|VTNYFVVS�AAADIAAVGVLAIPFAITIS|TG--FCAACH|GCLFFACFVLVL
AA2A_MOUSE   AVWI|NSNLQN|VTNFFVVS�AAADIAAVGVLAIPFAITIS|TG--FCAACH|GCLFIACFVLVL
AA2A_RAT     AVWI|NSNLQN|VTNFFVVS�AAADIAAVGVLAIPFAITIS|TG--FCAACH|GCLFFACFVLVL
AA2B_MOUSE   AVGA|SSALQT|PTNYFLVSLATADVAVGLFAIPFAITIS|LG--FCTDFH|GCLFLACFVLVL
AA2B_RAT     AVGA|SSALQT|PTNYFLVSLATADVAVGLFAIPFAITIS|LG--FCTDFH|SCLFLACFVLVL
AA2B_HUMAN   AVGT|ANTLQT|PTNYFLVSLAAADAAVAVGLFAIPFAITIS|LG--FCTDFY|GCLFLACFVLVL
AA2B_CHICK   AVAI|NSTLKN|ATNYFLVSLAVADIAAVGLLAIPFAITIS|IG--FQVDFH|SCLFFACFVLVL
AA3R_CANFA   VVKL|NPSLQT|TTFYFIVSLALADIAAVGVLMPLAIVIS|LG--ITIQFY|NCLFMTCLLLIF
AA3R_HUMAN   VVKL|NPSLQT|TTFYFIVSLALADIAAVGVLMPLAIVVS|LG--ITIHFY|SCLFMTCLLLIF
AA3R_RABIT   VIKL|NPSLKT|TTFYFIVSLALADIAAVGFLMPLAIVIS|LG--ITIGFY|SCLVMSCLLLVF
AA3R_RAT     VVKL|NRTLRT|TTFYFIVSLALADIAAVGVLVIPLAIAVS|LE--VQMHFY|ACLFMSCVLLVF
      | * |      :: .* * :      :. :. |      | *      :
---->|<-IL1>|<-----Helix II----->|<--EL1-->|<-----

```



```

OPSD_PIG      GEIALWLSLVVLAIERYVVV | CKPMSNFRFG | EN-HAIMGLALTWVMALACAAPPLV | GWS---
OPSD_TURTR    GEIALWLSLVVLAIERYVVV | CKPMSNFRFG | EN-HAIMGLALTWIMAMACAAAPLV | GWS---
OPSD_BOVIN    GEIALWLSLVVLAIERYVVV | CKPMSNFRFG | EN-HAIMGVAFTWVMALACAAPPLV | GWS---
OPSD_SHEEP    GEIALWLSLVVLAIERYVVV | CKPMSNFRFG | EN-HAIMGVAFTWVMALACAAPPLV | GWS---
OPSD_HUMAN    GEIALWLSLVVLAIERYVVV | CKPMSNFRFG | EN-HAIMGVAFTWVMALACAAPPLA | GWS---
OPSD_RABIT    GEIALWLSLVVLAIERYVVV | CKPMSNFRFG | EN-HAIMGVAFTWIMALACAAPPLV | GWS---
OPSD_MOUSE    GEIALWLSLVVLAIERYVVV | CKPMSNFRFG | EN-HAIMGVVFTWIMALACAAPPLV | GWS---
OPSD_RAT      GEIGLSLVVLAIERYVVV | CKPMSNFRFG | EN-HAIMGVAFTWVMALACAAPPLV | GWS---
OPSD_ALLMI    GEVALWCLVVLAIERYIVV | CKPMSNFRFG | EN-HAIMGVVFTWIMALTCAAPPLV | GWS---
OPSD_CHICK    GEIALWLSLVVLAVERYVVV | CKPMSNFRFG | EN-HAIMGVAFSWIMAMACAAPPLF | GWS---
OPSD_AMBTI    GEIALWLSLVVLAIERYVVV | CKPMSNFRFG | EN-HAIMGVMMTWIMALACAAPPLF | GWS---
OPSD_ANGAN    GEISLWLSLVVLAIERWVVV | CKPMSNFRFG | EN-HAIMGLAFTWIMANSCAMPPLF | GWS---
AA1R_CAVPO    TQSSILALLAIAVDRYLRV | KIPLRYKTVV | TPRRAAVAIAGCWILSLVVGLTPMF | GWNN--
AA1R_HUMAN    TQSSILALLAIAVDRYLRV | KIPLRYKMVV | TPRRAAVAIAGCWILSFVVGLTPMF | GWNN--
AA1R_RAT      TQSSILALLAIAVDRYLRV | KIPLRYKTVV | TQRRAAVAIAGCWILSLVVGLTPMF | GWNN--
AA1R_BOVIN    TQSSILALLAMAVDRYLRV | KIPLRYKTVV | TPRRAVAITGCWILSFVVGLTPMF | GWNN--
AA1R_CANFA    TQSSILALLAIAVDRYLRV | KIPLRYKTVV | TPRRAAVAIAGCWILSFVVGLTPLF | GWNR--
AA1R_RABIT    TQSSILALLAIAVDRYLRV | KIPLRYKAVV | TPRRAAVAIAGCWILSLVVGLTPMF | GWNN--
AA1R_CHICK    TESSILALLAIAVDRYLRV | KIPVRYKSVV | TPRRAAVAIACCWIVSFLVGLTPMF | GWNN--
AA2A_CANFA    TQSSIFSLLAIAIDRYIAI | RIPLRYNGLV | TGTRAKGIIACWVLSFAIGLTPML | GWN---
AA2A_HUMAN    TQSSIFSLLAIAIDRYIAI | RIPLRYNGLV | TGTRAKGIIACWVLSFAIGLTPML | GWN---
AA2A_CAVPO    TQSSIFSLLTITIDRYIAI | RIPLRYNGLV | TCTRAKGIIACWVLSFAIGLTPML | GWN---
AA2A_MOUSE    TQSSIFSLLAIAIDRYIAI | RIPLRYNGLV | TGMKAKGIIACWVLSFAIGLTPML | GWN---
AA2A_RAT      TQSSIFSLLAIAIDRYIAI | RIPLRYNGLV | TGVRAKGIIACWVLSFAIGLTPML | GWN---
AA2B_MOUSE    TQSSIFSLLAVAVDRYLAI | RVPLRYKGLV | TGTRARGIIAVLWVLAFGIGLTPFL | GWNSKD
AA2B_RAT      TQSSIFSLLAVAVDRYLAI | RVPLRYKGLV | TGTRARGIIAVLWVLAFGIGLTPFL | GWNSKD
AA2B_HUMAN    TQSSIFSLLAVAVDRYLAI | CVPLRYKSLV | TGTRARGVIAVLWVLAFGIGLTPFL | GWNSKD
AA2B_CHICK    TQSSIFSLLAVAIDRYLAI | KIPLRYNSLV | TGKRARGLIAVLWLSFVIGLTPML | GWNKAM
AA3R_CANFA    THASIMSLLAIAVDRYLRV | KLTVRYRRVT | TQRRIWLALGCWLVSFVGLTPMF | GWN---
AA3R_HUMAN    THASIMSLLAIAVDRYLRV | KLTVRYKRVT | THRRIWLALGCWLVSFVGLTPMF | GWN---
AA3R_RABIT    THASIMSLLAIAVDRYLRV | KLTVRYRRVT | TQRRIWLALGCWVVSLLVGFTPMF | GWN---
AA3R_RAT      THASIMSLLAIAVDRYLRV | KLTVRYRTVT | TQRRIWLFLGCWLVSFVGLTPMF | GWN---
. . : .*: : : : * : : : | . : . | : : * : : . * : | ** .
Helix III----->|<---IL2--->|<-----Helix IV----->|<-----

```

```

OPSD_PIG      -----R-YIPEGLQCSCGIDYYTLKPEVN | NESFVIYMFVVHFSIPLV I IFFCYGQ | L
OPSD_TURTR   -----R-YIPEGMQCSCGIDYYTSRQEVN | NESFVIYMFVVHFTIPLV I IFFCYGQ | L
OPSD_BOVIN   -----R-YIPEGMQCSCGIDYYTPHEETN | NESFVIYMFVVHF I IPLV I IFFCYGQ | L
OPSD_SHEEP   -----R-YIPQGMQCSCGALYFTLKPEIN | NESFVIYMFVVHFSIPLV I IFFCYGQ | L
OPSD_HUMAN   -----R-YIPEGLQCSCGIDYYTLKPEVN | NESFVIYMFVVHFTIPLM I I IFFCYGQ | L
OPSD_RABIT   -----R-YIPEGMQCSCGIDYYTLKPEVN | NESFVIYMFVVHFTIPL I I I IFFCYGQ | L
OPSD_MOUSE   -----R-YIPEGMQCSCGIDYYTLKPEVN | NESFVIYMFVVHFTIPLM I V I IFFCYGQ | L
OPSD_RAT     -----R-YIPEGMQCSCGIDYYTLKPEVN | NESFVIYMFVVHFTIPLM I V I IFFCYGQ | L
OPSD_ALLMI   -----R-YIPEGMQCSCGVDYYTLKPEVN | NESFVIYMFVVHFAIPLAV I IFFCYGR | L
OPSD_CHICK   -----R-YIPEGMQCSCGIDYYTLKPEIN | NESFVIYMFVVHFMIPLAV I IFFCYGN | L
OPSD_AMBTI   -----R-YIPEGMQCSCGVDYYTLKPEVN | NESFVIYMFVHVFTIPLM I I IFFCYGR | L
OPSD_ANGAN   -----R-YIPEGMQCSCGVDYYTLKPEVN | NESFVIYMFIVHFSVPLT I I I SFCYGR | L
AA1R_CAVPO   -----LSKIEMAWAANGSVGEPVIKCFEFKVIS | MEYMVYFNFFVWVLPPLLLM VLIYLE | V
AA1R_HUMAN   -----LSAVERAWAANGSMGEPVIKCFEFKVIS | MEYMVYFNFFVWVLPPLLLM VLIYLE | V
AA1R_RAT     -----LSVVEQDWRANGSVGEPVIKCFEFKVIS | MEYMVYFNFFVWVLPPLLLM VLIYLE | V
AA1R_BOVIN   -----LSAVERDWLANGSVGEPVIECFEFKVIS | MEYMVYFNFFVWVLPPLLLM VLIYME | V
AA1R_CANFA   -----LGEAQRAWAANGSGGEPVIKCFEFKVIS | MEYMVYFNFFVWVLPPLLLM VLIYLE | V
AA1R_RABIT   -----LREVQRAWAANGSVGEPVIKCFEFKVIS | MEYMVYFNFFVWVLPPLLLM VLIYLE | V
AA1R_CHICK   -----LNKVLGTRDLNVSHSEFVIKCFEFKVIS | MEYMVYFNFFVWVLPPLLLM LLIYLE | V
AA2A_CANFA   -----NCSQPKEGRNYSQCGEGQVACLFEDEVVP | MNYMVYFNFFAFVLPPLLLM LGVYLR | I
AA2A_HUMAN   -----NCGQPKEGKNHSQCGEGQVACLFEDEVVP | MNYMVYFNFFACVLPPLLLM LGVYLR | I
AA2A_CAVPO   -----NCSQPKGDKNHSESCDEGQVTCLEFDEVVP | MNYMVYFNFFAFVLPPLLLM LGIYLR | I
AA2A_MOUSE   -----NCSQTDE--NSTKTCGEGRVTCLFEDEVVP | MNYMVYFNFFAFVLLPPLLLM LAIYLR | I
AA2A_RAT     -----NCSQKDG--NSTKTCGEGRVTCLFEDEVVP | MNYMVYFNFFAFVLLPPLLLM LAIYLR | I
AA2B_MOUSE   SATSNCTELG---DGIANKSCCPLTCLFENVVP | MSYMVYFNFFGCVLPPLLI MLVIYIK | I
AA2B_RAT     RATSNCTEPG---DGI TNKSCCPVKCLFENVVP | MSYMVYFNFFGCVLPPLLI MMVIYIK | I
AA2B_HUMAN   SATNNCTEPW---DGT TNESCCLVKCLFENVVP | MSYMVYFNFFGCVLPPLLI MLVIYIK | I
AA2B_CHICK   SGCPNSTNETGA-DHGAGHHGCFISCLFENVVT | MSYMVYFNFFGCVLLPLLI MLGIYIK | I
AA3R_CANFA   -----MKLTSEHRNVT--FLSCQFSSVMR | MDYMVYFSFTWIL I PLVVMCAIYLD | I
AA3R_HUMAN   -----MKLTSEHRNVT--FLSCQFVSVMR | MDYMVYFSFLTWIF I PLVVMCAIYLD | I
AA3R_RABIT   -----MKPTLESARNYS--DFQCKFDSVIP | MEYMVFFSFTWIL I PLLLMCALYVY | I
AA3R_RAT     -----RKVTLELSQNSS--TLSCHFRSVVG | LDYMVFFSFITWIL I PLVVMCI IYLD | I
                                     | . : * : * . . * : : : * | :
-----EL2----->|<-----Helix V----->|<

```

```

OPSD_PIG      VFTVKEAAAQQQE-----SATTQK|AEKEVTRMVIIMVVAFLICWLPYASVAFYIFT|HQ
OPSD_TURTR    VFTVKEAAAQQQE-----SATTQK|AEKEVTRMVIIMVVAFLICWVPYASVAFYIFT|HQ
OPSD_BOVIN    VFTVKEAAAQQQE-----SATTQK|AEKEVTRMVIIMVIAFLICWLPYAGVAFYIFT|HQ
OPSD_SHEEP    VFTVKEAAAQQQE-----SATTQK|AEKEVTRMVIIMVIAFLICWLPYAGVAFYIFT|HQ
OPSD_HUMAN    VFTVKEAAAQQQE-----SATTQK|AEKEVTRMVIIMVIAFLICWVPYASVAFYIFT|HQ
OPSD_RABIT    VFTVKEAAAQQQE-----SATTQK|AEKEVTRMVIIMVIAFLICWVPYASVAFYIFT|HQ
OPSD_MOUSE    VFTVKEAAAQQQE-----SATTQK|AEKEVTRMVIIMVIFFLICWLPYASVAFYIFT|HQ
OPSD_RAT      VFTVKEAAAQQQE-----SATTQK|AEKEVTRMVIIMVIFFLICWLPYASVAMYIFT|HQ
OPSD_ALLMI    VCTVKEAAAQQQE-----SATTQK|AEKEVTRMVIIMVVSFLICWVPYASVAFYIFS|NQ
OPSD_CHICK    VCTVKEAAAQQQE-----SATTQK|AEKEVTRMVIIMVIAFLICWVPYASVAFYIFT|NQ
OPSD_AMBTI    VCTVKEAAAQQQE-----SATTQK|AEKEVTRMVIIMVVAFLICWVPYASVAFYIFS|NQ
OPSD_ANGAN    VCTVKEAAAQQQE-----SETTQR|AEREVTRMVIMVIAFLVCWVPYASVAVYIFT|HQ
AA1R_CAVPO    FYLIRKQLSKKVSASS--GDPQKYYG|KELKIAKSLALILFLFALSWLPLHLILNCITLF|CP
AA1R_HUMAN    FYLIRKQLNKKVSASS--GDPQKYYG|KELKIAKSLALILFLFALSWLPLHLILNCITLF|CP
AA1R_RAT      FYLIRKQLNKKVSASS--GDPQKYYG|KELKIAKSLALILFLFALSWLPLHLILNCITLF|CP
AA1R_BOVIN    FYLIRKQLSKKVSASS--GDPQKYYG|KELKIAKSLALILFLFALSWLPLHLILNCITLF|CP
AA1R_CANFA    FYLIRRQLGKKVSASS--GDPQKYYG|KELKIAKSLALILFLFALSWLPLHLILNCITLF|CP
AA1R_RABIT    FYLIRRQLSKKASASS--GDPHKYYG|KELKIAKSLALILFLFALSWLPLHLILNCVTLF|CP
AA1R_CHICK    FNLIRTQLNKKVSSSS--NDPQKYYG|KELKIAKSLALVLFLFALSWLPLHLILNCITLF|CP
AA2A_CANFA    FLAARRQLKQMESQPLPGERARSTLQ|KEVHAAKSLAIIVGLFALCWLPLHIINCFTFF|CP
AA2A_HUMAN    FLAARRQLKQMESQPLPGERARSTLQ|KEVHAAKSLAIIVGLFALCWLPLHIINCFTFF|CP
AA2A_CAVPO    FLAARRQLKQMESQPLPGERTRSTLQ|KEVHPAKSLAIIVGLFALCCLPLNIINCFTFF|CP
AA2A_MOUSE    FLAARRQLKQMESQPLPGERTRSTLQ|KEVHAAKSLAIIVGLFALCWLPLHIINCFTFF|CS
AA2A_RAT      FLAARRQLKQMESQPLPGERTRSTLQ|KEVHAAKSLAIIVGLFALCWLPLHIINCFTFF|CS
AA2B_MOUSE    FMVACKQLQSMELMDH----SRTTLQ|REIHAAKSLAMIVGIFALCWLPVHAINCITLF|HP
AA2B_RAT      FMVACKQLQHMELEH----SRTTLQ|REIHAAKSLAMIVGIFALCWLPVHAINCITLF|HP
AA2B_HUMAN    FLVACRQLQRTELMDH----SRTTLQ|REIHAAKSLAMIVGIFALCWLPVHAVNCVTLF|QP
AA2B_CHICK    FMVACKQLHQIELMGN----SRTTLQ|KEVHAAKSLAIIVGLFAFCWLPLHLILNCITHF|HE
AA3R_CANFA    FYVIRNKLQNFS-SS--KETGAFYG|REFKTAKSLFLVLFLFAFSWLPLSIINCITYF|H-
AA3R_HUMAN    FYIIRNKLSNLS-NS--KETGAFYG|REFKTAKSLFLVLFLFALSWLPLSIINCIIYF|N-
AA3R_RABIT    FYIIRNKLVSFS-SF--KETGAFYR|REFKTAKSLFLVLLALFAGCWLPLSIINCIVTYF|K-
AA3R_RAT      FYIIRNKLSQNLT-GF--RETRAFYG|REFKTAKSLFLVLFLFALCWLPLSIINFVSYF|N-
.              .          | * . :: : ::: * . :* :          |
-----IL3----->|<-----Helix VI----->|<-

```

```

OPSD_PIG      G--SDFGP|IFMTIPAFFAKSASIYNPVIIYIMM|NKQFRNCMLTTLCCGKNPLGDD--EA--
OPSD_TURTR   G--SDFGP|IFMTIPSFFAKSSSIYNPVIIYIMM|NKQFRNCMLTTLCCGRNPLGDD--EA--
OPSD_BOVIN   G--SDFGP|IFMTIPAFFAKTSAVYNPVIIYIMM|NKQFRNCMVTTLCCGKNPLGDD--EA--
OPSD_SHEEP   G--SDFGP|IFMTIPAFFAKSSSVYNPVIIYIMM|NKQFRNCMLTTLCCGKNPLGDD--EA--
OPSD_HUMAN   G--SNFGP|IFMTIPAFFAKSAAIYNPVIIYIMM|NKQFRNCMLTTICCGKNPLGDD--EA--
OPSD_RABIT   G--SNFGP|IFMTIPAFFAKSSSIYNPVIIYIMM|NKQFRNCMLTTICCGKNPLGDD--EA--
OPSD_MOUSE   G--SNFGP|IFMTLPAFFAKSSSIYNPVIIYIML|NKQFRNCMLTTLCCGKNPLGDD--DA--
OPSD_RAT     G--SNFGP|IFMTLPAFFAKTASIYNPIIYIMM|NKQFRNCMLTSLCCGKNPLGDD--EA--
OPSD_ALLMI   G--SDFGP|VFMTIPAFFAKSSAIYNPVIIYIVM|NKQFRNCMITTLCCGKNPLGDD--ET--
OPSD_CHICK   G--SDFGP|IFMTIPAFFAKSSAIYNPVIIYIVM|NKQFRNCMITTLCCGKNPLGDE--DT--
OPSD_AMBTI   G--TDFGP|IFMTVPAFFAKSSAIYNPVIIYIVL|NKQFRNCMITTICCGKNPFQDD--ET--
OPSD_ANGAN   G--STFGP|VFMTVPSFFAKSSAIYNPLIYICL|NSQFRNCMITTLFCGKNPFQEE--EG--
AA1R_CAVPO   T--CHKPT|ILTYIAIFLTHGNSAMNPIVYAFR|IQKFRVTFMKIWNDFHRCQPEP-----
AA1R_HUMAN   S--CHKPS|ILTYIAIFLTHGNSAMNPIVYAFR|IQKFRVTFMKIWNDFHRCQPAP-----
AA1R_RAT     T--CQKPS|ILTYIAIFLTHGNSAMNPIVYAFR|IHKFRVTFMKIWNDFHRCQPKP-----
AA1R_BOVIN   S--CHMPR|ILTYIAIFLSHGNSAMNPIVYAFR|IQKFRVTFMKIWNDFHRCQPAP-----
AA1R_CANFA   S--CRKPS|ILMYIAIFLTHGNSAMNPIVYAFR|IQKFRVTFMKIWNDFHRCQPTP-----
AA1R_RABIT   S--CQKPS|ILVYTAIFLTHGNSAMNPIVYAFR|IHKFRVTFMKIWNDFHRCRPAP-----
AA1R_CHICK   S--CKTPH|ILTYIAIFLTHGNSAMNPIVYAFR|IKKFRTAFLQIWNQYFCCKTNK-----
AA2A_CANFA   E-CSHAPL|WLMYLTIVLSHTNSVVNPFIYAYR|IREFRQTFRKIIRSHVLRREPFKAGGT
AA2A_HUMAN   D-CSHAPL|WLMYLAIIVLSHTNSVVNPFIYAYR|IREFRQTFRKIIRSHVLRQEPFKAAGT
AA2A_CAVPO   E-CDHAPP|WLMYLTIIILSHGNSVVNPLIYAYR|IREFRQTFRKIIRSHILRRRELKAGGT
AA2A_MOUSE   T-CQHAPP|WLMYLAIILSHSNSVVNPFIYAYR|IREFRQTFRKIIRTHVLRREQEPFRAGGS
AA2A_RAT     T-CRHAPP|WLMYLAIILSHSNSVVNPFIYAYR|IREFRQTFRKIIRTHVLRREQEPFQAGGS
AA2B_MOUSE   ALAKDKPK|WVMNVAILLSHANSVVNPIVYAYR|NRDFRYSFHKIISRYVLCQAE-----
AA2B_RAT     ALAKDKPK|WVMNVAILLSHANSVVNPIVYAYR|NRDFRYSFHRIISRYVLCQTD-----
AA2B_HUMAN   AQGKNKPK|WAMNMAILLSHANSVVNPIVYAYR|NRDFRYTFHKIISRYLLCQAD-----
AA2B_CHICK   EFSKSKPE|WVMYVAIILSHANSVINPIIYAYR|IRDFRYTFHKIISK-ILCKTDDFPK---
AA3R_CANFA   ---GEVPQ|IILYLGILLSHANSMMNPIVYAYK|IKKFKETYLLIFKTYMICQSSD-----
AA3R_HUMAN   ---GEVPQ|LVLYMGILLSHANSMMNPIVYAYK|IKKFKETYLLILKACVVCHPSD-----
AA3R_RABIT   ---CKVPD|VLLLVGILLSHANSMMNPIVYACK|IQKFKETYLLIFKARVTCQPSD-----
AA3R_RAT     ---VKIPE|IAMCLGILLSHANSMMNPIVYACK|IKKFKETYFVILRACRLCQTS-----
      |           .:::  :  **.*  |  .*:
-EL3-->|<-----Helix VII----->|<-----C-Terminus----->

```

```

OPSD_PIG      -----STT-----TSKTETSQVAPA-----
OPSD_TURTR   -----STT-----ASKTETSQVAPA-----
OPSD_BOVIN   -----STT-----VSKTETSQVAPA-----
OPSD_SHEEP   -----STT-----VSKTETSQVAPA-----
OPSD_HUMAN   -----SAT-----VSKTETSQVAPA-----
OPSD_RABIT   -----SAT-----ASKTETSQVAPA-----
OPSD_MOUSE   -----SAT-----ASKTETSQVAPA-----
OPSD_RAT     -----SAT-----ASKTETSQVAPA-----
OPSD_ALLMI   -----ATGSK----TETSSVSTSQVSPA-----
OPSD_CHICK   -----SAG-K----TETSSVSTSQVSPA-----
OPSD_AMBTI   -----TSAATSK--TEASSVSSSQVSPA-----
OPSD_ANGAN   -----ASTTASK--TEASSVSS--VSPA-----
AA1R_CAVPO   -----PID-----EDLPEEKVDD-----
AA1R_HUMAN   -----PID-----EDLPEERPDD-----
AA1R_RAT     -----PID-----EDLPEEKAED-----
AA1R_BOVIN   -----PID-----EDAPAERPDD-----
AA1R_CANFA   -----PVD-----EDPPEEAPHD-----
AA1R_RABIT   -----AGDGD---EDLPEEKPND-----
AA1R_CHICK   -----SSS-----SS-TAETVN-----
AA2A_CANFA   SARALAAHGSDGEQISLRLNGHPPGVWANGSAPHPERRPNGYTLGLVSGGIAPESHGDMG
AA2A_HUMAN   SARVLAAHGSDGEQVSLRLNGHPPGVWANGSAPHPERRPNGYALGLVSGGSAQESQNTG
AA2A_CAVPO   SARASAAHSPEGEQVSLRLNGHPPGVWANGSALRPEQRPNGYVTLGLVSGRSAQRSHGDAS
AA2A_MOUSE   SAWALAAHSTEGEQVSLRLNGHPLGVWANGSAPHSGRRPNGYTLGPGGGGSTQGSPG---
AA2A_RAT     SAWALAAHSTEGEQVSLRLNGHPLGVWANGSATHSGRRPNGYTLGLGGGGSAQGSPR---
AA2B_MOUSE   -----TKGG-----SGQAGAQ---STLSLGL-----
AA2B_RAT     -----TKGG-----SGQAGGQ---STFSLSL-----
AA2B_HUMAN   -----VKSG-----NGQAGVQ---PALGVGL-----
AA2B_CHICK   -----C-----TTDN-----NQHLTVTINVNAPAASVTI-----
AA3R_CANFA   -----SLD-----SS-TE-----
AA3R_HUMAN   -----SLD-----TS-IEKNSE-----
AA3R_RABIT   -----SLD-----PS-SEQNSE-----
AA3R_RAT     -----SLD-----SN-LEQTTE-----

-----C-Terminus-----

```

OPSD\_PIG -----  
 OPSD\_TURTR -----  
 OPSD\_BOVIN -----  
 OPSD\_SHEEP -----  
 OPSD\_HUMAN -----  
 OPSD\_RABIT -----  
 OPSD\_MOUSE -----  
 OPSD\_RAT -----  
 OPSD\_ALLMI -----  
 OPSD\_CHICK -----  
 OPSD\_AMBTI -----  
 OPSD\_ANGAN -----  
 AA1R\_CAVPO -----  
 AA1R\_HUMAN -----  
 AA1R\_RAT -----  
 AA1R\_BOVIN -----  
 AA1R\_CANFA -----  
 AA1R\_RABIT -----  
 AA1R\_CHICK -----  
 AA2A\_CANFA LPDVELLSHELKGACPESPGLEGLAQDGAGVS-----  
 AA2A\_HUMAN LPDVELLSHELKGVCPPEPGLDDPLAQDGAGVS-----  
 AA2A\_CAVPO LSDVELLSHEHKGTCPESPSLEDPPAHGGAGVS-----  
 AA2A\_MOUSE --DVELLTQEHQ-EGQEHPGLGDHLAQGRVGTASWSSEFAPS  
 AA2A\_RAT --DVELPTQERQ-EGQEHPGLRGHLVQARVGASSWSSEFAPS  
 AA2B\_MOUSE -----  
 AA2B\_RAT -----  
 AA2B\_HUMAN -----  
 AA2B\_CHICK -----  
 AA3R\_CANFA -----  
 AA3R\_HUMAN -----  
 AA3R\_RABIT -----  
 AA3R\_RAT -----

-----C-Terminus----->

## A.2.2. Editiertes Sequenz-Alignment

Alignment 2: Die Referenz-AS und Cysteine der konservierten Disulfid-Brücke sind mit '\*' markiert. Die AS, die im Fall des Rinder-Rhodopsins im Abstand von 4.0 Å zum 11-*cis*-Retinal liegen, bzw. die bei den Adenosin-Rezeptoren laut Mutagenese-Befund für die Bindung von Liganden relevant sind, sind unterstrichen.

```

OPSD_BOVIN  MNGTEGPNFY VPFSNKTGVV RSPFEAPQYY LAE|PWQFSM LAAYMFLIM 49
AA1R_HUMAN  -----MP PSI|SAFQAA YIGIEVLIAL 21
AA2A_HUMAN  -----MP|IMGSSV YITVELAIAV 18
AA2B_HUMAN  -----MLL|ETQDAL YVALELVIAA 19
AA3R_HUMAN  -----MPNNS TAL|SLANVT YITMEIFIGL 24

-----N-Terminus----->|<-----Helix I-----

OPSD_BOVIN  LGFPINFLTL YVTVQ|HKKL RT|PLNYILL NLAVADLFMV FGGFTTTLYT 97
AA1R_HUMAN  VSVPGNVLVI WAVKV|NQAL RD|ATFCFIV SLAVADVAVG ALVIPLAILI 69
AA2A_HUMAN  LAILGNVLVC WAVWL|NSNL QN|VTNYFVV SLAAADIAVG VLAIPFAITI 66
AA2B_HUMAN  LSVAGNVLVC AAVGT|ANTL QT|PTNYFLV SLAAADVAVG LFAIPFAITI 67
AA3R_HUMAN  CAIVGNLVI CVVKL|NPSL QT|TTFYFIV SLALADIAVG VLVMPAIVV 72
          *                               *
--Helix I----->|<-IL1->|<-----Helix II-----

OPSD_BOVIN  S|LHGYFVFG PT|GCNLEGF FATLGGEIAL WSLVLAIER YVVV|CKPMS 144
AA1R_HUMAN  N|IG--PQTY FH|TCLMVAC PVLILTQSSI LALLAIAVDR YLRV|KIPLR 114
AA2A_HUMAN  S|TG--FCAA CH|GCLFIAC FVLVLTQSSI FSLLAIAIDR YIAI|RIPLR 111
AA2B_HUMAN  S|LG--FCTD FY|GCLFLAC FVLVLTQSSI FSLLAIVADR YLAI|CVPLR 112
AA3R_HUMAN  S|LG--ITIH FY|SCLFMT C LLLIFTHASI MSLLAIAVDR YLRV|KLTVR 117
          *                               *
>|<--EL1----->|<-Helix III----->|<-IL2

OPSD_BOVIN  NFRFG-|ENH AIMGVAFTWV MALACAAPPL V|GWS----- 176
AA1R_HUMAN  YKMOVV|PRR AAVAIAGCWI LSFVVGLTPM F|GWNN----- ---LSAVERA 155
AA2A_HUMAN  YNGLVT|GTR AKGIIAICWV LSFAIGLTPM L|GWNNCGQP KEGK----- 153
AA2B_HUMAN  YKSLVT|GTR ARGVIAVLWV LAFGIGLTPF L|GWNSKDSA TNNCTEPW-- 158
AA3R_HUMAN  YKRVT|HRR IWLALGLCWL VSFLVGLTPM F|GWN----- -----MKLT 154
          *
-IL2->|<-----Helix IV----->|<-----EL2-----

OPSD_BOVIN  ---RYIPEGM QCSCGIDYYT PHEETN|NES FVIYMFVVHF IIPLIVIFFC 222
AA1R_HUMAN  WAANGSMGEP VIKCEFEKVI S-----|MEY MUYFNFFVWV LPPLLLMVLI 199
AA2A_HUMAN  -NHSQCGEG QVACLFEDVV P-----|MNY MUYFNFFACV LVPLLLMLGV 196
AA2B_HUMAN  -DGTNESCC LVKCLFENVV P-----|MSY MUYFNFFGCV LPPLLLMLVI 201
AA3R_HUMAN  SEYHRNVT-- FLSCQFVSVM R-----|MDY MUYFSFLTWI FIPLVVMCAI 196
          *                               *
-----EL2----->|<-----Helix V-----

```

```

OPSD_BOVIN YGQ|LVFTVK EAAAQQQE-- -----SATTQ K|AEKEVTRM VIIMVIAFLI 263
AA1R_HUMAN YLE|VFYLIR KQLNKKVSAS S--GDPQKYY G|KELKIAKS LALILFLFAL 245
AA2A_HUMAN YLR|IFLAAR RQLKQMESQP LPGERARSTL Q|KEVHAAKS LAIIVGLFAL 244
AA2B_HUMAN YIK|IFLVAC RQLQRTELMD H----SRTTL Q|REIHAASKS LAMIVGIFAL 245
AA3R_HUMAN YLD|IFYIIR NKLSLNLS-N S--KETGAFY G|REFKTAKS LFLVLFLFAL 241

```

```

-->|<-----IL3----->|<---Helix VI----->

```

```

OPSD_BOVIN CWLPYAGVAF YIFT|HQG-- SDFGP|IFMT IPAFFAKTSA VYNPVIIYIMM| 309
AA1R_HUMAN SWLPLHILNC ITLF|CPS-- CHKPS|ILTY IAIFLTHGNS AMNPIVYAFR| 291
AA2A_HUMAN CWLPLHIINC FTFF|CPD-C SHAPL|WLMY LAIVLSHTNS VVNPFIYAYR| 291
AA2B_HUMAN CWLPVHAVNC VTLF|QPAQG KNKPK|WAMN MAILLSHANS VVNPIVYAYR| 293
AA3R_HUMAN SWLPLSIINC IIYF|N---- GEVPQ|LVLY MGILLSHANS MMNPIVYAYK| 285

```

```

* *
----->|<--EL3---->|<-----Helix VII----->|

```

```

OPSD_BOVIN |NKQFRNCMV TTLCCGKNPL GDDEASTTVS KTETSQVAPA ----- 348
AA1R_HUMAN |IQKFRVTFL KIWNDFRCQ PAPPIDEDLP EERPDD---- ----- 326
AA2A_HUMAN |IREFRQTFR KIIRSHVLRQ QEPFKAAGTS ARVLAHGS GEQVSLRLNG
AA2B_HUMAN |NRDFRYTFH KIISRYLLCQ ADVKSGNGQA GVQPALGVGL ----- 332
AA3R_HUMAN |IKKFKETYL LILKACVVCH PSDSLDTSIE KNSE----- ----- 318
|<-----C-Terminus ----->

```

```

AA2A_HUMAN HPPGVWANGS APHPERRPNG YALGLVSGGS AQESQNTGL PDVELLSHEL
AA2A_HUMAN KGVCPEPPGL DDPLAQDGAG VS----- 412

```



# Literaturverzeichnis

- [1] P.J. Ala, E.E. Houston, R.M. Klabe, D.D. McCabe, J.L. Duke, C.J. Rizzo, B.D. Korant, R.J. Deloskey, P.Y. Lam, C.N. Hodge und C.H. Chang. Molecular basis of HIV-1 protease drug resistance: structural analysis of mutant proteases complexed with cyclic urea inhibitors. *Biochemistry* **1998**, *36*, 1537.
- [2] B. Almås, B. Le Bourdelles, T. Flatmark, J. Mallet und J. Haavik. Regulation of recombinant human tyrosine hydroxylase isozymes by catecholamine binding and phosphorylation. *Eur. J. Biochem.* **1992**, *209*, 249–255.
- [3] B. Almås, K. Toska, K. Teigen, V. Groehn, W. Pfeleiderer, A. Martinez, T. Flatmark und J. Haavik. A kinetic and conformational study on the interaction of tetrahydropteridines with tyrosine hydroxylase. *Biochemistry* **2000**, *39*, 13676–13686.
- [4] O.A. Andersen, T. Flatmark und E. Hough. High resolution crystal structures of the catalytic domain of human phenylalanine hydroxylase in its catalytically active form and binary complex with tetrahydrobiopterine. *J. Mol. Biol.* **2001**, *314*, 266.
- [5] O.A. Andersen, T. Flatmark und E. Hough. Crystal structure of human phenylalanine hydroxylase with tetrahydrobiopterin and 3-(2-thienyl)-L-alanine, and its implications for the mechanism of catalysis and substrate activation. *J. Mol. Biol.* **2002**, *320*, 1095–1108.
- [6] P.W. Atkins. *Physikalische Chemie*. VCH Verlagsgesellschaft, Weinheim **1988**.
- [7] A. Bairoch. *Proteome databases (In) Proteome research, new frontiers in functional genomics*. Springer Verlag, Heidelberg **1997**.
- [8] J. A. Ballesteros und H. Weinstein. Integrated methods for the construction of three dimensional models and computational probing of structure-function relations in G-protein coupled receptors. *Methods Neurosci.* **1995**, *25*, 366–428.
- [9] H. Barbhayia, R. McClain, A. P. IJzermann und S. A. Rivkees. Site-directed mutagenesis of the human A1 adenosine receptor: influences of acidic and hydroxy residues in the first four transmembrane domains on ligand binding. *Mol. Pharmacol.* **1996**, *50*, 1635–1642.
- [10] N. Beerenwinkel, B. Schmidt, H. Walter, R. Kaiser, T. Lengauer, D. Hoffmann, K. Korn und J. Selbig. Diversity and complexity of HIV-1 drug resistance: A bioinformatics approach to predicting phenotype from genotype. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 8271–8276.
- [11] M. W. Beukers, H. den Dulk, E. W. van Tilburg und J. Brouwer und A. P. IJzerman. Why are A<sub>2B</sub> receptors low-affinity adenosine receptors? Mutation of Asn273 to Tyr increases

- affinity of human A2B receptor for 2-(1-hexynyl) adenosine. *Mol. Pharmacol.* **2000**, *58*, 1349–1356.
- [12] H. Beyer und W. Walter. *Lehrbuch der Organischen Chemie*. S. Hirzel Verlag, Stuttgart **1991**.
- [13] A.-M. Bianucci, M.U. Bigi, G. Biagi, I. Giorgi, O. Livi und V. Scartoni. A 3D Model of the Human A1 Adenosine Receptor. An Evaluation of the Binding Free Energy with Ligands. *Drug Design and Discovery* **1998**, *15*, 149–156.
- [14] D. Blum, S. Torch, N. Lamberg, M. Nissou, A.L. Beabid, R. Sadoul und J. M. Verna. Molecular Pathways involved in the neurotoxicity of 6-OHDA, dopamine and MPTP, contribution to the apoptotic theory in Parkinson's disease. *Progr. Neurobiol.* **2001**, *65*, 135–172.
- [15] H.-J. Böhm. The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional structure. *J. Comput. Aided Mol. Des.* **1994**, *8*, 243–256.
- [16] H.-J. Böhm, G. Klebe und H. Kubinyi. *Wirkstoffdesign*. Spektrum Akademischer Verlag, Heidelberg **1996**.
- [17] M. J. Bower, F. E. Cohen und Jr. R. L. Dunbrack. Prediction of protein sidechain conformations from a backbone-dependent rotamer library: A new homology modeling tool for Proteins, Application to Side-chain prediction. *J. Mol. Biol.* **1997**, *267*, 1268–1282.
- [18] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, und M. Karplus. CHARMM: A Program for Macromolecular Energy, Minimization and Dynamics Calculation. *J. Comp. Chem.* **1983**, *4*, 187–217.
- [19] R.F. Bruns. Adenosine receptor activation in human fibroblasts: nucleoside agonists and antagonists. *Can. J. Physiol. Pharmacol.* **1979**, *58*, 673–691.
- [20] J. Cavanagh, W.J. Fairbrother, A.G. Palmer III und N.J. Skelton. *Protein NMR Spectroscopy*. Academic Press **1996**.
- [21] C. Daubner und P. Fitzpatrick. Site-directed mutants of charged residues in the active site of Tyrosine Hydroxylase. *Biochemistry* **1999**, *38*, 44448–44454.
- [22] T.E. Creighton. *Proteins: Structures and Molecular Properties*. W. H. Freeman, New York **1993**.
- [23] S.C. Daubner, P.J. Hillas und P. F. Fitzpatrick. Characterization of chimeric pterin-dependent hydroxylases: Contributions of the regulatory domains of tyrosine and phenylalanine hydroxylase to substrate specificity. *Biochemistry* **1997**, *36*, 11574–11582.
- [24] M. O. Dayhoff, R. M. Schwartz und B. C. Orcutt. A model of evolutionary change in proteins. matrices for detecting distant relationships. *Atlas of protein sequence and structure* **1978**, *5*, 345–358.

- [25] M. de Zwart, R. Link, J.K von Frijtag Drabbe Künzel, G. Cristalli, K.A. Jacobson, A. Townsend-Nicolson und A.P. IJzerman. A functional screening of adenosine analogues at the adenosine A<sub>2B</sub> receptor: A search for potent agonists. *Nucleosides & Nucleotides* **1998**, *17*, 969–985.
- [26] M. de Zwart, R.C. Vollinga, M.W. Beukers, D.F. Slegers, J.K. von Frijtag Drabbe Künzel, M. De Groot und A.P. IJzerman. Potent antagonists for the human adenosine A<sub>2a</sub> receptor. Derivatives of the triazolotriazine adenosine receptor antagonist ZM241385 with high affinity. *Drug. Dev. Res.* **1999**, *48*, 95–103.
- [27] E.M Van der Wenden, Philip J.M Van Galen, A. P. IJzerman und W. Soudijn. Mapping the xanthine C8-region of the adenosine A1 receptor with computer graphics. *Eur. J. Pharmacol.* **1991**, *206*, 315–323.
- [28] P.W. Dickson, I.G. Jennings und R.G.H. Cotton. Delineation of the catalytic core of phenylalanine hydroxylase and identification of glutamate 286 as a critical residue for pterin function. *J. Biol. Chem.* **1994**, *269*, 20369–20375.
- [29] I. Doytchinova. CoMFA-based comparison of two models of binding site on adenosine A<sub>1</sub> receptor. *J. Comput. Aided Mol. Design* **2001**, *15*, 29–39.
- [30] E.Gasteiger, Jung E und A. Bairoch. SWISS-PROT: Connecting biological knowledge via a protein database. *Curr. Issues Mol. Biol.* **2001**, *3*, 47–55.
- [31] H.R. Ellis, S. C. Daubner, R. I. McCulloch und P.F. Fitzpatrick. Phenylalanine residues in the active site of tyrosine hydroxylase: Mutagenesis of Phe300 and Phe309 to alanine and metal ion-catalyzed hydroxylation of Phe300. *Biochemistry* **1999**, *38*, 10909–10914.
- [32] M. A. L. Eriksson, J. Pitera und P. A. Kollmann. Prediction of the binding free energies of new TIBO-like HIV-1 reverse transcriptase inhibitors using a combination of PROFEC, PB/SA, CMC/MD and free energy calculations. *J. Med. Chem.* **1999**, *42*, 868–881.
- [33] H. Erlandsen, T. Flatmark und R.C. Stevens. Crystal structure and site-specific mutagenesis of Pterin-bound human Phenylalanine Hydroxylase. *Biochemistry* **2000**, *39*, 2208–2217.
- [34] H. Erlandsen, T. Flatmark, R.C. Stevens und E. Hough. Crystallographic analysis of the human phenylalanine hydroxylase catalytic domain with bound catechol inhibitors at 2.0 Å Resolution. *Biochemistry* **1998**, *37*, 15638–15646.
- [35] H. Erlandsen, F. Fusetti, A. Martinez, E. Hough, T. Flatmark und R.C. Stevens. Crystal structure of the catalytic domain of human phenylalanine hydroxylase reveals the structural basis for phenylketonuria. *Nature Structural Biology* **1997**, *4*, 995–1000.
- [36] E.M. Marcotte et. al. Detecting protein function and protein-protein interactions from genome sequences. *Science* **1999**, *282*, 751–753.
- [37] H. I. Skulnick et al. Structure-based design of nonpeptidic hiv protease inhibitors: The sulfonamide-substituted cyclooctopyranones. *J. Med. Chem.* **1997**, *40*, 1149–1164.
- [38] P.L.Roach et al. Structure of Isopenicillin-N-Synthase complexed with substrate and the mechanism of penicillin formation. *Nature* **1997**, *387*, 827.

- [39] R. Betarbet et al. Chronic systemic pesticide exposure reproduces features of Parkinson's disease. *Nature Neuroscience* **2000**, *3*, 1 301–1306.
- [40] F.C. Bernstein, T.F. Koetzle, G. J. B. Williams, E.F. Meyer, M. D. Brice, J. R. Rogers, O. Kennard, T. Shimanouchi und M. Tasumi. The protein data bank: a computer based archival file for macromolecular structures. *J. Mol. Biol.* **1977**, *112*, 535–542.
- [41] I. Feoktistov und I. Biaggioni. Adenosine A<sub>2B</sub>-Receptors. *Pharmacol. Reviews* **1997**, *49*, 381–402.
- [42] S. Firestein. How the olfactory system makes sense of scents. *Nature* **2001**, *413*, 211–217.
- [43] Paul F. Fitzpatrick. Tetrahydropterin-dependent amino acid hydroxylases. *Ann. Rev. Biochem.* **1999**, *68*, 355–381.
- [44] P.F. Fitzpatrick. Steady-state kinetic mechanism of rat tyrosine hydroxylase. *Biochemistry* **1991**, *30*, 3658–3662.
- [45] T. Flatmark und R. Stevens. Structural insight into the aromatic amino acid hydroxylases and their disease-related mutant forms. *Chem. Rev.* **1999**, *99*, 2137–2160.
- [46] P. Foley und P. Riederer. Influence of neurotoxins and oxidative stress on the onset and progression of Parkinson's disease. *J. Neurol.* **2000**, *247*, 82–94.
- [47] W.A. Francisco, C. Tian, P. Fitzpatrick und J.P. Klinman. Oxygen-18 Kinetic isotope effect studies of the tyrosine hydroxylase reaction: Evidence of rate limiting oxygen activation. *J. Amer. Chem. Soc.* **1997**, *120*, 4057–4062.
- [48] B.B. Fredholm, A.P. IJzerman, K. A. Jacobson, K.-N. Klotz und J. Linden. International Union of Pharmacology. XXV. Nomenclature and Classification of Adenosine Receptors. *Pharmacol. Rev.* **2001**, *53*, 527–552.
- [49] Z.G. Gao, Q. Jiang, K.A. Jacobson und A.P. IJzerman. Site-directed mutagenesis studies of human A<sub>2A</sub> Adenosine receptors. *Biochem. Pharmacol.* **2000**, *60*, 661–668.
- [50] M. Gastreich. *Werkzeuge zur Modellierung von Siliciumbornitrid-Keramiken*. Dissertation, Univ. Bonn **2001**.
- [51] Christian Gerthsen. *Physik*. Springer, Berlin **1997**.
- [52] K.E. Goodwill, C. Sabatier, C. Marks, R. Raag, P.F. Fitzpatrick und R. C. Stevens. Crystal structure of tyrosine hydroxylase at 2.3 Å and its implications for inherited neurodegenerative diseases. *Nature structural biology* **1997**, *4*, 578–585.
- [53] K.E. Goodwill, C. Sabatier und R.C. Stevens. Crystal structure of tyrosine hydroxylase with bound cofactor analogue and iron at 2.3 Å resolution: Self-hydroxylation of Phe 300 and the Pterin-binding site. *Biochemistry* **1998**, *37*, 13437–13445.
- [54] A. Gustchina und I. Weber. Comparison of inhibitor binding in hiv-1 protease and in non-viral aspartic proteases: the role of the flaps. *FEBS Letters* **1990**, *269*, 269–272.

- [55] G.Vriend. WHAT IF: A molecular modeling and drug design program. *J. Mol. Graph.* **1990**, *8*, 52–56.
- [56] J. Haavik, Martinez A, S. Olafsdottir, J. Mallet und T. Flatmark. The incorporation of divalent metal ions into recombinant human tyrosine hydroxylase apoenzymes studied by intrinsic fluorescence and H-NMR spectroscopy. *Eur. J. Biochem.* **1992**, *210*, 23–31.
- [57] J. Haavik, B. Almås und T.Flatmark. Generation of reaktive oxygen species by tyrosine hydroxylase: A possible contribution to the degeneration of dopaminergic neurons? *J. Neurochem.* **1997**, *68*, 328–332.
- [58] A.M. Hayallah, J. Sandoval-Ramirez, U. Reith, U. Schobert, B. Preiss, B. Schumacher, J.W. Daly und C.E. Müller. 1,8-Disubstituted xanthine derivatives: Synthesis of potent A<sub>2B</sub>-selective adenosine receptor antagonists. *J. Med. Chem.* **2002**, *45*, 1500–1510.
- [59] Helmut Heller. *Simulation einer Lipidmembran auf einem Parallelrechner*. Dissertation, TU München **1994**.
- [60] S. Henikoff und J. G. Henikoff. Amino acid substitution matrices from protein blocks. *Proc. Natl. Acad. Sci.* **1992**, *89*, 10915–10919.
- [61] S. Hess, C.E. Müller, W. Frobenius, U. Reith, K.-N. Klotz und K. Eger. 7-Deazaadenines bearing polar substituents: Structure – activity relationships of new A<sub>1</sub> and A<sub>3</sub> adenosine receptor antagonists. *J. Med. Chem.* **2000**, *43*, 4636 – 4646.
- [62] D. Higgins, J. Thompson, T. Gibson, J.D. Thompson, D.G. Higgins und T.J. Gibson. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **1994**, *22*, 4673–4680.
- [63] D. Hoffmann, B. Kramer, T. Washio, T. Steinmetzer, M. Rarey und T. Lengauer. Two-stage method for protein-ligand docking. *J. Med. Chem.* **1999**, *42*, 4422–4433.
- [64] D. Hoffmann, T. Washio, K. Gessler und J Jacob. Tackling concrete problems in molecular biophysics using Monte Carlo and related methods: Glycosylation, folding, solvation. In P. Grassberger, G. Barkema und W. Nadler, editors, *Proceedings of the workshop on Monte Carlo approach to Biopolymers and Protein Folding* **1998**, 153–170. ISBN 981-02-3658-1.
- [65] D. Hoffmann, T. Washio und K. Oosterlee. CAMLab: Computational analysis and modelling laboratory. *GMD-SCAI and NEC, St. Augustin, Germany* **1997**.
- [66] F. Horn, G. Vriend und F.E. Cohen. Collecting and harvesting biological data: The GP-CRDB & NucleaRDB databases. *Nucleic Acids Res.* **2001**, *29*, 346–349.
- [67] F. Horn, J. Weare, W.M. Beukers, S. Hörsch, A. Bairoch, W. Chen, Ø. Edvardsen, F. Campaigne und G. Vriend. GPCRDB: an information system for G protein-coupled receptors. *Nucleic Acids Res.* **1998**, *26*, 277–281.
- [68] James E. Huheey. *Anorganische Chemie*. Berlin; New York: Walter de Gruyter **1988**.

- [69] A.P. IJzerman, J.K. Von Frijtag Drabbe Künzel, J. Kim, Q. Jiang und K.A. Jacobson. Site-directed mutagenesis of the human adenosine  $A_{2A}$  receptor. Critical involvement of Glu<sup>13</sup> in agonist recognition. *Eur. J. Pharmacology* **1996**, *310*, 269–272.
- [70] Molecular Simulations Inc. Delphi and solvation. *Manual* **1993**.
- [71] R. Ishima, D.I. Freedberg, Y.X. Wang, J.M.Louis und D.A. Torchia. Flap opening and dimer-interface flexibility in the free and inhibitor-bound HIV protease, and their implications for function. *Structure Fold. Des.* **1999**, *7*, 1047–1055.
- [72] Q. Jiang, B.X. Lee, M. Glashofer, A.M. van Rhee und K. A. Jacobson. Mutagenesis reveals structure–activity parallels between human  $A_{2A}$  adenosine receptors and biogenic amine G protein-coupled receptors. *J. Med. Chem.* **1997**, *40*, 2588–2595.
- [73] Wolfgang Kabsch. A solution for the best rotation to relate two sets of vectors. *Acta Cryst.* **1976**, *A32*, 922–923.
- [74] T. Joseph Kappock und John P. Caradonna. Pterin-dependent amino acid hydroxylases. *Chem. Rev.* **1996**, *96*, 2659–2756.
- [75] M. Kilby. Potent suppression of HIV-1 replication in humans by T-20, a peptide inhibitor of gp41-mediated virus entry. *Nature Medicine* **1998**, *4*, 1302.
- [76] J. Kim, Q. Jiang, M. Glashofer, S. Yehle, J. Wess und K.A. Jacobson. Glutamate residues in the second extracellular loop of the human  $A_{2A}$  adenosine receptor are required for ligand recognition. *Mol. Pharmacol.* **1996**, *49*, 683–691.
- [77] J. Kim, J. Wess, A.M. van Rhee, T. Schöneberg und K.A. Jacobson. Site-directed mutagenesis identifies residues involved in ligand recognition in the human  $A_{2A}$  adenosine receptor. *J. Biol. Chem.* **1995**, *270*, 13987–13997.
- [78] S. Kim, M.A. Marshall, N. Melman, H.S. Kim, C.E. Müller und K.A. Jacobson. Structure activity relationships at human and rat  $A_{2B}$  adenosine receptors of xanthine derivatives substituted at the 1-, 3-, 7- and 8-positions. *J. Med. Chem.* **2002**.
- [79] Y.-C. Kim, M. de Zwart, L. Chang, S. Moro, J.K. von Frijtag Drabbe Künzel, N. Melman, A.P. IJzermann und K.A. Jacobson. Derivatives of triazoloquinazoline adenosine antagonist (CGS15943) having high potency at the human  $A_{2B}$  and  $A_3$  receptor subtypes. *J. Med. Chem* **1998**, *41*, 2835–2845.
- [80] F.D. King und S. Wilson. Recent advances in 7-transmembrane receptor research. *Current Opinion in Drug Discovery & Development* **1999**, *2*, 83–95.
- [81] R.M. Klabe, L.T. Bachelere, P.J. Ala, S. Erickson-Viitanen und J.L. Meek. Resistance to HIV protease inhibitors: A comparison of enzyme inhibition and antiviral potency. *Biochemistry* **1998**, *37*, 8735–8742.
- [82] J. Klinman. Life as aerobes: Are there simple rules for activation of dioxygen by enzymes? *J. Biol. Inorg. Chem.* **2001**, *6*, 1–13.

- [83] B. Kobe, I.G. Jennings, C.M. House, B.J. Michell, K.E. Goodwill, B.D.Santarsiero, R.C. Stevens, R.G. H. Cotton und B.C. Kemp. Structural basis of autoregulation of phenylalanine hydroxylase. *Nature Structural Biology* **1999**, *6*, 442–448.
- [84] K. Lee und C. K. Chu. Molecular modeling approach to understanding the mode of action of L-nucleosides as antiviral agents. *Antimicrobial Agents and Chemotherapy* **2001**, *45*, 138–144.
- [85] M. Lopez-Rodriguez, M. Murcia, B. Benhamu, M. Olivella, M. Campillo und L. Pardo. Computational model of the complex between GR113808 and the 5-TH<sub>4</sub> receptor guided by site-directed mutagenesis and the crystal structure of rhodopsin. *Journal of Computer-Aided Molecular Design* **2001**, *15*, 1025–1033.
- [86] I. Luque, M.J. Todd, J. Gomez, N. Semo und E. Freire. Molecular basis of resistance to hiv-1 protease inhibition: A plausible hypothesis. *Biochemistry* **1998**, *37*, 5791–5797.
- [87] A. Maaß. Berechnung des Einflusses der Elektrostatik auf Protein-Ligand-Wechselwirkungen. Diplomarbeit, FWU Bonn **1999**.
- [88] M. Madrid, K. Jacobo-Molina, J. Ding und E. Arnold. Major subdomain rearrangement in HIV-1-reverse transcriptase simulated by molecular dynamics. *Proteins* **1999**, *35*, 332–337.
- [89] J. Marota und R. Shimann. Stoichiometric reduction of phenylalanine hydroxylase by its cofactor: A requirement for enzymatic activity. *Biochemistry* **1984**, *23*, 1303–1311.
- [90] A. Martinez, K. K. Anderson und J. Haavik. EPR and 1H-NMR spectroscopic studies on the paramagnetic iron at the active site of phenylalanine hydroxylase and its interaction with substrates and inhibitors. *Eur. J. Biochem.* **1991**, *198*, 675–682.
- [91] A. Martinez, J. Haavik und T. Flatmark. Cooperative homotropic interaction of L-noradrenaline with the catalytic site of phenylalanine 4-monoxygenase. *Eur. J. Biochem.* **1990**, *193*, 211–219.
- [92] A. Martinez, P.M. Knappskog und J. Haavik. A structural approach into human tryptophan hydroxylase and its implications for the regulation of serotonin biosynthesis. *Curr. Med. Chem.* **2001**, *8*, 1077–1091.
- [93] G. B. McGaughey, M. Gagne und A. K. Rappe.  $\pi$ -Stacking interactions. *J. Biol. Chem.* **1998**.
- [94] H. Mevissen, R. Thiele, R. Zimmer und T. Lengauer. The ToPLign software environment - Toolbox for protein alignment. *Bioinformatik, Jena, IMB - Institut für molekulare Biotechnologie* **1994**.
- [95] I. Michaud-Soret, K.K. Andersson, L. Jr. Que und J. Haavik. Resonance Raman studies of catecholate and phenolate complexes of recombinant human tyrosine hydroxylase. *Biochemistry* **1995**, *34*, 5504–5510.
- [96] M. Miller und R. Shiman. Reversible inactivation of phenylalanine hydroxylase by catecholamines in cultured hepatoma cells. *J. Biol. Chem.* **1976**, *251*, 3671–3676.

- [97] S. Moro und K.A. Jacobson. Molecular modeling as a tool to investigate molecular recognition in P2Y receptors. *Current Pharmaceutical Design* **2002**, *8*, 2401–2413.
- [98] A. Moser, J. Scholz, F. Nobbe, P. Vieregge, V. Bohme und H. Bamberg. Presence of N-Methyl-norsalsolinol in the CSF: correlations with dopamine metabolites of patients with Parkinson's disease. *J. Neurol. Science* **1995**, *2*, 183–189.
- [99] A. Moser, F. Siebecker, F. Nobbe und V. Bohme. Rotational behaviour and neurochemical changes in unilateral N-methyl-norsalsolinol and 6-hydroxy-dopamine lesioned rats. *Exp. Brain Res.* **1996**, *112*, 89–95.
- [100] C.E. Müller. A<sub>1</sub>-Adenosine receptor antagonists. *Exp. Opin. Ther. Patents* **1997**, *7*, 419–440.
- [101] C.E. Müller. A<sub>2a</sub>-adenosine receptor antagonists - future drugs for Parkinson's disease. *Drugs of the future* **2000**, *25*, 1043–1052.
- [102] C.E. Müller, U. Geis, B. Grahner, W. Lanzner und K. Eger. Chiral Pyrrolo[2,3-*d*]pyrimidine and pyrimido[4,5*b*]indole derivatives: Structure–activity relationships of potent, highly stereoselective A<sub>1</sub>-adenosine receptor antagonists. *J. Med. Chem.* **1996**, *39*, 2482–2491.
- [103] C.E. Müller und T. Scior. Adenosine receptors and their modulators. *Pharmaceutica Acta Helveticae* **1993**, *68*, 77–11.
- [104] G. Müller. Towards 3D structures of G protein-coupled receptors: A multidisciplinary approach. *Curr. Med. Chem.* **2000**, *7*, 861–888.
- [105] D.L. Nelson und M.M.Cox. *Lehninger Biochemie*. Springer-Verlag Berlin Heidelberg New York **2001**.
- [106] C. Notredame. Recent progress in multiple sequence alignment: a survey. *Pharmacogenomics* **2002**, *3*, 131–144.
- [107] K. Palczewski, T. Kumasaka, T. Hori, C. A. Behnke, H. Motoshima, B. A. Fox, I. Le Trong, D. C. Teller, T. Okada, R. E. Stenkamp, M. Yamamoto und M. Miyano. Crystal structure of rhodopsin: A G-protein-coupled receptor. *Science* **2000**, *289*, 739–745.
- [108] R.G. Parr und R.G. Pearson. HSAB-Concept. *J. Am. Chem. Soc.* **1983**, *105*, 7512.
- [109] P.F.Fitzpatrick. Studies of the rate-limiting step in the tyrosine hydroxylase reaction: alternate substrates, solvent isotope effects, and transition-state analogs. *Biochemistry* **1991**, *30*, 6386–6391.
- [110] S.-A. Poulsen und R.J.Quinn. Adenosine receptors: New opportunities for future drugs. *Bioorganic and Medicinal Chemistry* **1998**, *6*, 619–641.
- [111] J.P. Priestle, A. Fässler, J. Rösel, M. Tintelnot-Blomley, P. Strop, und M.G. Grütter. Comparative analysis of the x-ray structures of hiv-1 and hiv-2 proteases in complex with cgp53820, a novel pseudosymmetric inhibitor. *Structure* **1995**, *3*, 381–389.



- [112] Services Programme R. Lopez und A. Lloyd. ClustalW WWW Service at the European Bioinformatics Institute. The ClustalWWW server at the EBIembnet.news **1997**, 4.2.
- [113] A. Ramsey und P. Fitzpatrick. Effects of phosphorylation of serine 40 of tyrosine hydroxylase on binding of catecholamines: Evidence for a novel regulatory mechanism. *Biochemistry* **1998**, 37, 8980–8986.
- [114] A.J. Ramsey und P.F. Fitzpatrick. Effects of phosphorylation on binding of catecholamines to tyrosine hydroxylase: Specificity and thermodynamics. *Biochemistry* **2000**, 39, 773–778.
- [115] M. Rarey, B. Kramer, T. Lengauer und G. Klebe. Predicting receptor-ligand interactions by an incremental construction algorithm. *J. Mol. Biol.* **1996**, 261, 470–489.
- [116] E. Riedel. *Anorganische Chemie*. De Gruyter Verlag **1999**.
- [117] S. A. Rivkees, M. E. Lasbury und H. Barbhaiya. Identification of domains of the human A<sub>1</sub> adenosine receptor that are important for binding receptor subtype-selective ligands using chimeric A<sub>1</sub>/A<sub>2a</sub> adenosine receptors. *J. Biol. Chem.* **1995**, 270, 20485–20490.
- [118] S.A. Rivkees, H. Barbahiya und A. P. IJzerman. Identification of the adenine binding site of the human A<sub>1</sub> adenosine receptor. *J. Biol. Chem.* **1999**, 274, 3617–3621.
- [119] Bernd Ross. *Allgemeine und physikalische Chemie für Pharmazeuten und Naturwissenschaftler*. Thieme Verlag, Stuttgart **1988**.
- [120] G. Rücker, M. Neugebauer und G.G. Willems. *Instrumentelle pharmazeutische Analytik*. Wissensch. VG., Stuttgart **1992**.
- [121] R.F. Schmidt und G. Thews. *Physiologie des Menschen*. Springer-Verlag Berlin Heidelberg New York, 27. Auflage **1997**.
- [122] J. Scholz, H. Bamberg und A. Moser. N-Methyl-norsalsolinol, an endogenous neurotoxin, inhibits tyrosine hydroxylase activity in the rat brain nucleus accumbens in vitro. *Neurochem. Int.* **1997**, 6, 845–849.
- [123] V. Schünemann. Charakterisierung der Tyrosinhydroxylase mit Mößbauer- und EPR-Spektroskopie. persönliche Mitteilung **1999**.
- [124] W.R.P. Scott und C.A. Schiffer. Curling of flap tips in HIV-1 protease as a mechanism for substrate entry and tolerance of drug resistance. *Structure* **2000**, 8, 1259–1265.
- [125] D. Dengler, T. Fichteler und U. Schomburg. Prediction of protein three-dimensional structures in insertion and deletion regions: A procedure for searching data bases of representative protein fragments using geometric criteria. *J. Mol. Biol.* **1995**, 253, 114–131.
- [126] R.D. Taylor, P.J. Jewsbury und J.W. Essex. A review of protein-small molecule docking methods. *J. of Computer-Aided Mol. Design* **2002**, 16, 151–166.
- [127] K. Teigen, N. Frøystein und A. Martinez. The structural basis of the recognition of phenylalanine and pterin cofactors by phenylalanine hydroxylase: Implications for the catalytic mechanism. *J. Mol. Biol.* **1999**, 294, 807–823.

- [128] D. C. Teller, T. Okada, C.A. Behnke, K. Palczewski und R. E. Stenkamp. Advances in determination of a high-resolution three-dimensional structure of rhodopsin, a model of G-protein-coupled receptors. *Biochemistry* **2001**, *40*, 7761.
- [129] M.J. Todd und E. Freire. The effect of inhibitor binding on the structural stability and cooperativity of the HIV-1 protease. *Proteins* **1999**, *36*, 147–156.
- [130] A.L. Tucker, A.S. Robeval, H.E. Taylor, D. Holetton, Bockner, K.R. Lynch und J. Linden. A<sub>1</sub> Adenosine receptors. *J. Biol. Chem.* **1994**, *269*, 27900–27906.
- [131] T. Wilk und S.D. Fuller. Towards the structure of the human immunodeficiency virus: divide and conquer? *Current Opinion in Structural Biology* **1999**, *9*, 231–243.
- [132] G. Kuschinski und H. Lüllmann. *Kurzes Lehrbuch der Pharmakologie und Toxikologie*. Georg Thieme Verlag Stuttgart · New York, 12. Auflage **1989**.
- [133] A.M. van Rhee und K.A. Jacobson. Molecular architecture of G protein-coupled receptors. *Drug Development Research* **1996**, 1–38.
- [134] A.M. Vocero-Akbani, N. Vander Heyden, N.A. Lissy, L. Ratner und S.F. Dowdy. Killing hiv-infected cells by transduction with an hiv protease-activated caspase-3 protein. *Nature Medicine* **1999**, *5*, 29–33.
- [135] D. Voet und J. G. Voet. *Biochemie*. VCH Verlagsgesellschaft, Weinheim **1992**.
- [136] L. Wang, H. Erlandsen, J. Haavik, P.M. Knappskog und R.C. Stevens. Three-dimensional structure of human tryptophan hydroxylase and its implications for the biosynthesis of the neurotransmitters serotonin and melatonin. *Biochemistry* **2002**, *41*, 12569.
- [137] I.T. Weber, J. Wu, J. Adomat, R.W. Harrison, A.R. Kimmel, E.M. Wondrak und J.M. Louis. Crystallographic analysis of human immunodeficiency virus 1 protease with an analog of the conserved CA-p2 substrate. *Eur. J. Biochem.* **1997**, *249*, 523–530.
- [138] P. L. Yeagle, G. Choi und A. D. Albert. Studies on the structure of the G-protein-coupled receptor rhodopsin including the putative G-protein binding site in the unactivated and activated forms. *Biochemistry* **2001**, *40*, 11932–11937.
- [139] Z.-G.-Gao, A.Chen, D. Barak, S. Kim, C.E. Müller und K.A. Jacobson. Identification by site-directed mutagenesis of residues involved in ligand recognition and activation of the human A<sub>3</sub> adenosine receptor. *J. Biol. Chem.* **2002**.

# Danksagung

Mein größter Dank gilt all denen, die mir diese Arbeit ermöglicht haben: Dies sind vor allem die Omas & Opas, die sich während meiner Abwesenheiten regelmäßig und liebevoll um meine Tochter Sarah gekümmert haben. Im Notfall sind auch Anne Lücker, Ines Brieger und Ulla Schneider als Babysitter für mich eingesprungen.

Ohne das Doktoranden-Programm der ehemaligen GMD hätte ich diese Arbeit ebenfalls nicht durchführen können. Daher danke ich Prof. Thomas Lengauer, der mir den Start in seiner Abteilung an SCAI ermöglicht hat, sowie Dr. Martin Hofmann und Dr. Johannes Linden, die mich auch später unterstützt haben.

Viele andere haben noch zum Gelingen dieser Arbeit beigetragen: Die Kollegen an SCAI haben mir mit Rat und Tat zur Seite gestanden und mir das Leben am Computer erträglich gemacht. Vor allem Patrick Fricker, Marcus Gastreich, Ingolf Sommer, Niklas von Oehsen, Daniel Hanisch, Theo Mevissen, Eberhard Schrüfer und Jürgen Klein habe ich des öfteren in Anspruch genommen. Mit Dorothee Liebig habe ich jahrelang und gern das Zimmer geteilt (Alles Gute für die Zukunft!). Auch bei den Kollegen an der Pharmazie in Poppelsdorf habe ich mich bei jedem meiner seltenen Besuche sehr wohl gefühlt. Besonders meiner dortigen „Bezugsperson“ Andrea Herzog habe ich einige Liter Tee und einen als „Externe“ guten Kontakt zur Gruppe zu verdanken. Sehr erbaulich waren auch die botanischen Exkursionen mit Karl Vollmann.

Ohne die Erfahrungen und den persönlichen Einsatz von Joachim Scholz, wäre nichts aus unserem Paper geworden, zu dem auch Marcus, Jannis Apostolakis und Joachim Selbig wesentlich beigetragen haben. Sandra Lilley hat geholfen, aus dem ersten Kauderwelsch verständliches Englisch zu machen.

Zu guter letzt haben sich Ines und Ulla seitenweise durch wirre Manuskripte gekämpft, bis diese lesbar wurden. Mit Patricks Hilfe sind sie nun auch ansehnlich geworden.

Schließlich muss ich ganz besonders anerkennen, wie ungern, aber tapfer Sarah mich Zeit ihres Lebens entbehrt hat.

Allen ein herzliches Dankeschön!



# Lebenslauf

*Astrid Berta Katharina Maaß,*

geboren am 12.01.1970 in Karlsruhe, deutsch, ledig, eine Tochter (\*15.07.97)

## Schulausbildung

1976 – 1979	Grundschule Köln-Flittard
1980	Grundschule Bergisch Gladbach-Hand
1980 – 1989	Nicolaus-Cusanus-Gymnasium Bergisch Gladbach
18.05.1989	<b>Allgemeine Hochschulreife</b>

## Studium

seit WS 1989/90	Studium an der Friedrich-Willhelms-Universität in Bonn
Juli 1992	Vordiplom Chemie
April 1993	Zwischenprüfung Lebensmittelchemie
Mai 1995	<b>1. Staatsexamen Lebensmittelchemie</b>
Juli 1997	<b>Diplom Chemie: Mündliche Prüfungen</b> Nebenfächer: Physik, Biochemie, Mikrobiologie, Nuklearchemie, theoretische Chemie
April 1998 - Jan.1999	<b>Diplom Chemie: Praktische Arbeit</b> am Institut für Theoretische Chemie in Bonn und am GMD-Forschungszentrum Informationstechnik GmbH in Sankt Augustin unter der Leitung von Frau Priv. Doz. Dr. C. Marian und der fachlichen Anleitung von Dr. D. Hoffmann zum Thema "Berechnung des Einflusses der Elektrostatik auf Protein-Ligand-Wechselwirkungen"
seit März 1999	<b>Dissertation</b> am Pharmazeutischen Institut der Rheinischen Friedrich-Wilhelms-Universität in Bonn und dem ehem. GMD-Zentrum, nun Fraunhofer Institut für Algorithmen und Wissenschaftliches Rechnen, im Rahmen des GMD-DoktorandInnen-Programms unter der Leitung von Frau Prof. C.E. Müller und der zeitweiligen fachlichen Betreuung von Dr. D. Hoffmann.
März 2001	<b>Physikum Human-Medizin</b>

## Publikation

Astrid Maaß, Joachim Scholz, Andreas Moser:  
"Modeled ligand-protein complexes elucidate the origin of substrate specificity and provide insight into catalytic mechanisms of phenylalanine hydroxylase and tyrosine hydroxylase", *Eur. J. Biochem.* **2003**, *270*, 1065–1075