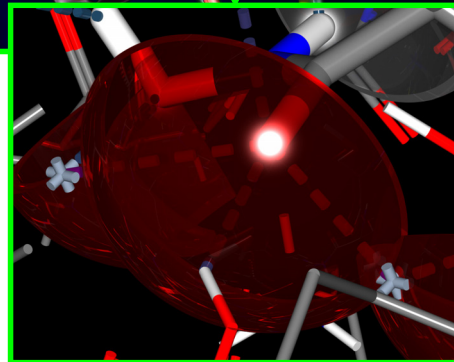
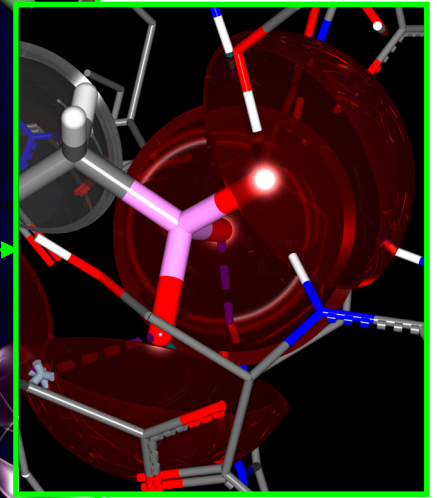
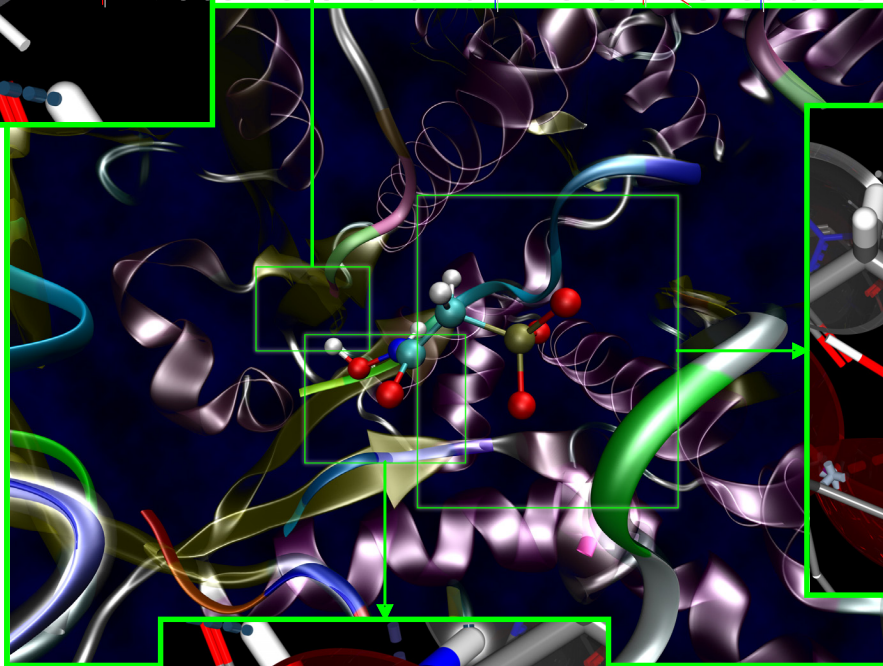
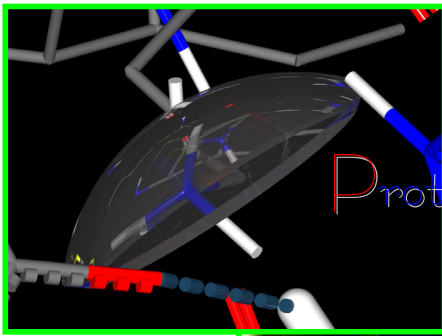


Entwicklung einer Bewertungsfunktion für das Docking von Liganden

in

Proteine und artifizielle Rezeptoren



INAUGURAL-DISSERTATION

ZUR ERLANGUNG DES DOKTORGRADES
DER MATHEMATISCH-NATURWISSENSCHAFTLICHEN FAKULTÄT
DER HEINRICH-HEINE-UNIVERSITÄT DÜSSELDORF

vorgelegt von
Stephan Raub
aus Düsseldorf

2007

Aus dem Institut für Theoretische Chemie und Computerchemie
der Heinrich-Heine-Universität Düsseldorf

Gedruckt mit der Genehmigung der
Mathematisch-Naturwissenschaftlichen Fakultät der
Heinrich-Heine-Universität Düsseldorf

Referentin: Frau Prof. Dr. Christel M. Marian
Korreferent: Herr Prof. Dr. Walter Thiel
Tag der mündlichen Prüfung: 29.06.2007

Zusammenfassung

In dieser Arbeit stelle ich die Entwicklung einer neuen Bewertungsfunktion für FLEXX vor, genannt AISCORE. AISCORE erweitert die ursprüngliche Bewertungsfunktion um einen chemisch diversen Satz von Wechselwirkungsgeometrien und entsprechenden Wechselwirkungsenergien, die von aufwendigen quantenchemischen *ab initio*-Rechnungen abgeleitet wurden.

Die Wechselwirkungsenergien sind an die quantenchemischen *ab initio*-Wechselwirkungsenergien auf Niveau der Møller-Plesset-Störungstheorie zweiter Ordnung (MP2) von 81 kleinen wasserstoffverbrückten Komplexen angepasst worden. Die Komplexe wurden so gewählt, dass die wichtigsten Arten von Donor–Akzeptor-Paaren, wie sie in biologischen Systemen gefunden werden, repräsentiert sind. Durch die Berücksichtigung von CH-aziden H–Donoren konnte eine lineare Korrelation zwischen den MP2-Energien und dem Inkrementmodell mit einem Korrelationskoeffizienten von $r^2 = 0,994$ erreicht werden.

Die *ab initio*-Inkmente wurden dann durch einen einzigen Faktor skaliert, um die Erniedrigung der Wechselwirkungsenergie in Lösung zu modellieren. Geladene und resonanzassistierte Wechselwirkungsenergien wurden an experimentelle Daten angepasst. Dafür habe ich einen Satz von 100 Protein–Ligand-Komplexen mit bekannter experimenteller Bindungsaffinität aus der Protein Database (PDB) zusammengestellt. Dabei galt das Hauptaugenmerk der chemisch korrekten Zuweisung der Atomtypen und die Berücksichtigung von gebundenen Wassermolekülen in der Bindetasche. Des weiteren habe ich einen Algorithmus zur Behandlung von mehrfach verbrückten Wasserstoffbrücken entwickelt.

Für den Komplexsatz korrelieren die mit AISCORE vorhergesagten freien Bindungsenergien mit den experimentellen Werten mit einem $r^2 = 0,76$. Eine zehnfache Kreuzvalidierung erreichte ein $q^2 = 0,59$. Verglichen mit der originalen Bewertungsfunktion von FLEXX wurde eine signifikante Verbesserung der Vorhersage der freien Bindungsenergien von Kristallstrukturen erreicht.

AISCORE wurde darüber hinaus auf das Docken von flexiblen Liganden in flexible artifizielle Rezeptoren angewendet. Für den Koffeinrezeptor von WALDVOGEL *et al.* konnten mit Hilfe von AISCORE nativ-ähnliche Strukturen von einer Reihe von Oxypurin-Komplexen generiert werden. Außerdem korrelierten die vorhergesagten Energien gut mit den experimentellen Werten.

Das letzte Kapitel beschreibt die Parametrisierung von Kraftfeldparametern für die Eisen–Stickstoff-Bindung in Cytochromen durch die Anpassung an quantenchemische Potentialflächen.

Abstract

In this work I present the development of a novel scoring function for FLEXX named AISCORE. AISCORE extends the original scoring function by means of a chemically diverse set of interaction geometries and corresponding interaction energies that are derived from extensive quantum chemical *ab initio* calculations.

The interaction energies have been fitted to quantum chemical *ab initio* interaction energies of 81 small hydrogen-bonded complexes determined at the level of second-order Møller-Plesset perturbation theory (MP2). The complexes have been chosen such that the most important types of donor–acceptor pairs found in biological systems are represented. By taking CH acidic H donors into account, a linear correlation between MP2 energies and the increment model with a coefficient of correlation of $r^2 = 0.994$ has been accomplished.

The *ab initio* increments were scaled by a single factor in order to model the loss of interaction energy in solution. Charged and resonance assisted interaction energies were fitted to experimental data. Therefore I assembled a set of 100 protein–ligand complexes with known experimental binding affinities from the protein database (PDB). A chemically correct atom type assignment and the incorporation of tightly bound water molecules in the active site was a main focus while processing the complexes. Furthermore I developed an algorithmical extension to the scoring function that introduces a way to treat multifurcated hydrogen bonds.

For this set the predicted binding free energies of AISCORE correlate with an $r^2 = 0.76$ to the experimental energies. In ten-fold cross validation I achieved a q^2 value of 0.59. Compared to the original scoring function of FLEXX a significant improvement was obtained with respect to the prediction of binding free energies of the crystal structure.

AISCORE has also been applied to score the dockings of flexible ligands into flexible synthetic receptors. For the caffeine receptor by WALDVOGEL *et al.* AISCORE has helped to generate near-native structures of a series of oxypurine complexes. Furthermore the predicted energies correlate well with the experimental values.

The last chapter describes the parameterisation of force field parameters for the iron–nitrogen-bond in cytochromes by fitting them to quantumchemical potential surfaces.

Danksagung

Natürlich wäre diese Arbeit nie ohne die Unterstützung von Familie, Freunden und Kollegen möglich gewesen. Ihnen möchte ich deswegen an dieser Stelle danken.

Frau Prof. Dr. Christel Marian möchte ich für das Vertrauen danken, das sie in mich gesetzt hat als sie mir dieses Thema anvertraute, und für die stetige Bereitschaft zur Diskussion. Außerdem bin ich ihr für die Betreuung und Unterstützung dankbar, für die hilfreichen Kommentare zur Erstellung dieser Dissertation und auch dafür, dass sie die Finanzierung einer Stelle in ihrem Institut immer wieder ermöglicht hat.

Herrn Prof. Dr. Walter Thiel danke ich für die Übernahme des Korreferats und für die damit verbundene Arbeit.

Meiner Frau Sandra danke ich für die Unterstützung außerhalb der Universität, für den seelischen Rückhalt und ihre stärkende Liebe. Ohne sie hätte ich das nie geschafft.

Andreas Steffen vom Max-Planck-Institut für Informatik in Saarbrücken ist mir durch unsere gemeinsame Arbeit ein wertvoller Freund geworden. Er hat mir oft wissenschaftlich zur Seite gestanden und mir immer wieder wertvolle Anregungen gegeben. Darüber hinaus hat er den XFURCATE-Algorithmus in FLEXX implementiert.

Andreas Kämper am Wilhelm-Schickard-Institut für Computerwissenschaften hat mich in FLEXX eingewiesen und mir während der ganzen Zeit immer wieder mit Rat und Tat zur Seite gestanden. Er hat außerdem den Algorithmus zur Optimierung der Wasserstoff-Diederwinkel geschrieben.

Martin Kleinschmidt hat lange das Büro mit mir geteilt und mich in dieser Zeit in die Geheimnisse und Tiefen von Linux eingeweiht. Ohne dieses Wissen, das ich von ihm erhalten habe, wäre vieles in dieser Arbeit nicht so ohne Weiteres möglich gewesen.

Meinen Eltern danke ich dafür, dass sie mir den Weg bis zur Universität überhaupt erst möglich gemacht haben. Leicht habe ich es ihnen wahrlich nicht gemacht. Um so höher schätze ich ihren grundsätzlichen Anteil an dieser Arbeit.

Allen Leuten im Institut möchte ich für eine lockere und freundschaftliche Atmosphäre danken. In dieser Gemeinschaft ist nie jemand mit einem Problem auf sich allein gestellt.

Der Deutschen Forschungsgemeinschaft danke ich für die Finanzierung dieser Arbeit durch das Projekt MA 1051/6.

Nicht zuletzt danke ich BioSolveIT (St. Augustin) dafür, dass für diese Arbeit eine freie Entwicklerversion der FLEXX-Software zur Verfügung gestellt wurde. Auch für den kompetenten Support und die gute Anwenderbetreuung möchte ich allen Mitarbeitern von BioSolveIT danken.

Inhaltsverzeichnis

1	Einleitung	1
1.1	Aufbau und Ziel dieser Arbeit	4
2	Quantenchemische Rechnungen an Wasserstoffbrücken	5
2.1	Der Ursprung der Wasserstoffbrücke	9
2.2	Quantenchemische Grundlagen	14
2.2.1	Die Schrödinger-Gleichung	14
2.2.2	Slater-Determinanten	16
2.2.3	HF – Die Hartree–Fock-Methode	17
2.2.4	MP2 – Møller–Plesset-Störungstheorie zweiter Ordnung	21
2.2.5	DFT – Dichte-Funktional-Theorie	24
2.2.5.1	Austausch–Korrelations-Energiefunktionale	26
2.2.6	AIM – Atome in Molekülen	27
2.3	Quantenchemische Rechnungen	32
2.3.1	Der Molekülsatz	32
2.3.2	Quantenchemische Methoden	38
2.4	Quantenchemische Ergebnisse	41
2.4.1	DFT- und MP2-Geometrien	41
2.4.2	Energetische und strukturelle Eigenschaften des Trainingsatzes	42
2.4.2.1	Korrelationen der Wasserstoffbrückenenergien	43
2.4.2.2	Donor- und Akzeptor Energien	45
2.4.3	Validierung des Modells	53
3	Chemisch diverses Scoring für FlexX	65
3.1	Molekulares Docking	65
3.2	Bewertungsfunktionen	66
3.2.1	Empirische Bewertungsfunktionen	68
3.2.2	Wissensbasierte Bewertungsfunktionen	68
3.3	FlexX	69
3.3.1	Wechselwirkungen in FlexX	69
3.3.2	Flexibles Docking in FlexX	73
3.4	AIScore – Chemisch diverses Scoring	76
3.4.1	Chemisch diverse Subgraphen	76

3.4.2	Optimieren der Parameter	76
3.4.3	Mehrfach gegabelte Wasserstoffbrücken	80
3.4.4	Ein Datensatz von Protein-Ligand-Komplexen	82
3.4.4.1	Bindetaschen	83
3.4.4.2	Protonierung des Liganden und Zuweisung der Atomtypen	85
3.4.4.3	Das Tool WaterGen	87
3.5	AIScore vs. FlexXScore – Ergebnisse	93
3.5.1	Freie Bindungsenergien	93
3.5.1.1	AIScore	93
3.5.1.2	FlexXScore	95
3.5.1.3	Optimiertes und erweitertes FlexXScore	96
3.5.2	Kreuzvalidierung	104
3.5.3	Redocking	109
3.5.3.1	Kumulative Erfolgsraten	109
3.5.3.2	HIV1-Proteasen und die Frage nach dem RMSD-Wert als Kriterium	113
3.6	Anwendung auf artifizielle Rezeptoren	117
3.6.1	FlexR	117
3.6.2	AIScore und der Koffeinrezeptor	117
3.6.2.1	Datenaufbereitung und Docking	118
3.6.2.2	Ergebnisse und Diskussion	119
3.7	GeomGen5	121
3.7.1	Eingabedateien	121
3.7.2	Bewertungsmodi	122
3.7.3	Optimierungsalgorithmen	123
3.7.4	Geplante Erweiterungen	124
4	Quantenchemische Bestimmung von Kraftfeldparametern für CYP51	125
4.1	Die Biochemie von CYP-51	125
4.2	Quantenmechanische Bestimmung von Kraftfeldparametern	127
4.2.1	Modellsystem und quantenchemische Methoden	127
4.2.2	Rechnungen und Ergebnisse	128
4.2.3	Anpassung der Kraftfeldparameter	129
5	Zusammenfassung und Ausblick	134
6	Aus dieser Arbeit hervorgegangene Veröffentlichung	137
	Literatur	138

Abbildungsverzeichnis

1.1	Beim experimentellen <i>High-Throughput Screening</i> werden hochautomatisierte Robotersysteme eingesetzt.	1
2.1	Siedepunkte in Abhängigkeit von der Molmasse <i>a)</i> für Hydride der Elemente der Gruppe 4 und <i>b)</i> für Elemente der Gruppen 5, 6 und 7.	5
2.2	Modell einer Wasserstoffbrückenbindung zwischen einem Donor $X - H$ und einem Akzeptor B mit dem Bindungskritischen Punkt BCP.	6
2.3	Schematischer Aufbau der DNA und Wasserstoffbrücken der Basenpaare.	6
2.4	Phasendiagramm von Wasser.	7
2.5	Isoflächen der Elektronendichte verschiedener Moleküle. Die Farbe der Fläche entspricht dem elektrischen Potential an dieser Stelle (rot entspricht dabei einer negativen Ladung, blau einer positiven). Die Zahlen in den Atomen entsprechen der jeweiligen Elektronegativität.	9
2.6	Die jeweils günstigsten Orientierungen für verschiedene Multipole (in Anlehnung an [1]).	10
2.7	Logarithmische Elektronendichte von Wasser (Schnitt durch die Molekülebene) mit den bindungskritischen Punkten (<i>BCPs</i>) und den interatomaren Separatrizes (<i>IASs</i>).	29
2.8	<i>AIM</i> - und <i>MULLIKEN</i> -Partialladungen für ein nicht-komplexiertes Wasserstoffatom im Wasserdimer für verschiedene atomare Basissätze.	30
2.9	Hierarchischer Molekülfragment-Baum der Wasserstoffbrückendonoren und -akzeptoren, die Schwefel als Akzeptor oder Donor enthalten.	34
2.10	Hierarchischer Molekülfragment-Baum der Wasserstoffbrückendonoren, die Stickstoff als Donor enthalten.	35
2.11	Hierarchischer Molekülfragment-Baum der Wasserstoffbrückendonoren, die Sauerstoff als Donor enthalten.	35

2.12	Hierarchischer Molekülfragment-Baum der Wasserstoffbrückenakzeptoren, die Stickstoff als Akzeptor enthalten.	36
2.13	Hierarchischer Molekülfragment-Baum der Wasserstoffbrückenakzeptoren, die Sauerstoff als Akzeptor enthalten.	37
2.14	<i>AIM</i> -Partiellladungen der Akzeptoratome (<i>oben</i>) und der Protonen in den Donorgruppen (<i>unten</i>) in Abhängigkeit der Wechselwirkungsenergie des entsprechenden Komplexes.	44
2.15	Korrelation zwischen der geschätzten Coulomb-Wechselwirkung und der <i>MP2</i> -Wechselwirkungsenergie.	45
2.16	Korrelation zwischen der Wasserstoffbrückenlänge und der <i>MP2</i> -Wechselwirkungsenergie.	46
2.17	Minimumsstruktur der Komplexe mit einer zweiten konventionellen Wasserstoffbrücke.	47
2.18	Korrelation zwischen der <i>MP2</i> Wechselwirkungsenergie $\Delta E^{\text{MP2}}(\text{Don} \cdots \text{Acc})$ und $\Delta E^{\text{HBInc}}(\text{Don} \cdots \text{Acc})$	48
2.19	Minimumsstrukturen der Komplexe mit einer zweiten Wasserstoffbrücke, die von einer CH-gruppe gebildet wird.	50
2.20	Vergleich der Korrelationen zwischen $\Delta E^{\text{MP2}}(\text{Don} \cdots \text{Acc})$ und $\Delta E^{\text{HBInc}}(\text{Don} \cdots \text{Acc})$ von Komplexen, die eine CH-Brücke aufweisen, <i>a</i>) mit und <i>b</i>) ohne CH-Energieterm.	52
2.21	Minimumsstruktur des Komplexes $\text{MeNH}_2 \cdots \text{H}_2\text{CO}$, in dem eine CH-Wasserstoffbrücke erwartet wird, aber kein bindungskritischer Punkt gefunden wurde.	53
2.22	Korrelation zwischen der <i>MP2</i> Wechselwirkungsenergie $\Delta E^{\text{MP2}}(\text{Don} \cdots \text{Acc})$ und $\Delta E^{\text{HBInc}}(\text{Don} \cdots \text{Acc})$ mit zusätzlichen Termen für schwache CH-Wasserstoffbrücken.	54
2.23	Minimumsstrukturen der Komplexe des Testsatzes. Die Abkürzungen sind in den Fußnoten zu Tabelle 2.7 erklärt.	55
2.24	Korrelation zwischen der <i>MP2</i> -Wechselwirkungsenergie $\Delta E^{\text{MP2}}(\text{Don} \cdots \text{Acc})$ und $\Delta E^{\text{HBInc}}(\text{Don} \cdots \text{Acc})$ für die Komplexe des Testsatzes.	56
2.25	Korrelation zwischen der <i>MP2</i> -Wechselwirkungsenergie $\Delta E^{\text{MP2}}(\text{Don} \cdots \text{Acc})$ und $\Delta E^{\text{HBInc}}(\text{Don} \cdots \text{Acc})$ für die Komplexe des Testsatzes mit zyklischen Wasserstoffbrückenmustern.	57
2.26	Vereinfachte DNA-Modelle für Watson–Crick-Adenin–Thymin (M1) und Watson–Crick-Guanin–Cytosin (M2) (siehe Referenz [2]). DNA-M1: Komplex von <i>N</i> -Methylen-Methan-Diamin mit <i>N</i> -Formyl-Formamid. DNA-M2: Komplex von <i>N</i> -(Aminomethyl)-Formamid mit (<i>E</i>)- <i>N</i> -(Aminomethylen)-Formamid.	58
3.1	LUDI-Wechselwirkungsmodell am Beispiel des Wasserdimers.	69
3.2	Die Straffunktionen von LUDI.	71
3.3	Inkrementeller Aufbau	74

3.4	<i>Pose Clustering</i> in FLEXX. Ein Dreieck aus Wechselwirkungsgeometrien des Liganden (grün) wird mit einer geeigneten Transformation mit einem Dreieck aus kompatiblen Wechselwirkungsgeometrien des Proteins (blau) überlagert. Aus algorithmischen Gründen werden die Wechselwirkungsflächen des Proteins durch diskrete Punkte angenähert.	75
3.5	Vergleich der Wechselwirkungsgeometrien, wie sie bei diesem hypothetischen Molekül von FLEXXSCORE und von AISCORE zugewiesen werden.	77
3.6	Verschiedene Arten von bifurkaten Wasserstoffbrücken.	80
3.7	MP2-Wechselwirkungsenergien von Formaldehyd komplexiert mit <i>trans</i> -1,-Ethendiol (<i>links</i>) und mit <i>cis</i> -1,2-Ethendiol (<i>rechts</i>).	81
3.8	Schematische Darstellung der Unterschiede in der Bewertung von bifurkaten Wasserstoffbrücken. ΔG_O ist das Energieinkrement für einen sp^2 -Sauerstoff, ΔG_{XH} und ΔG_{YH} sind die H-Brückendonorkomponenten für $X-H$ and $Y-H$	82
3.9	Komplex 8ABP jeweils vor und nach der Optimierung der Diederwinkel der Wasserstoffatome.	86
3.10	Lineare Korrelation zwischen der experimentellen freien Bindungsenergie der Komplexe ΔG_{Exp} und der berechneten Bindungsenergie von AISCORE a) mit XFURCATE und b) ohne XFURCATE.	98
3.11	Lineare Korrelation zwischen der experimentellen freien Bindungsenergie der Komplexe ΔG_{Exp} und der berechneten Bindungsenergie von FLEXXSCORE a) ohne XFURCATE und b) mit XFURCATE.	99
3.12	Lineare Korrelation zwischen der experimentellen freien Bindungsenergie der Komplexe ΔG_{Exp} und der berechneten Bindungsenergie von FLEXXSCORE a) nach der Optimierung und b) mit zusätzlich Wechselwirkungsgeometrien und anschließender Optimierung.	100
3.13	Darstellung der aktiven Zentren der im Text diskutierten Komplexe 1A08, 1A1C, 1ADB.	101
3.14	Darstellung der aktiven Zentren der im Text diskutierten Komplexe 1EBG, 2QWD, 5YAS.	102
3.15	Darstellung der aktiven Zentren der im Text diskutierten Komplexe 6RNT, 6TIM.	103
3.16	Schematische Illustration der Kreuzvalidierung.	104
3.17	Lineare Korrelationskoeffizienten r^2 des Trainings und der anschließenden Vorhersage der Bindungsenergie für die zehn Subsets der Kreuzvalidierung.	105
3.18	Lineare Korrelationen für Training und Vorhersage der zehnfachen Kreuzvalidierung.	106
3.19	Lineare Korrelationen für Training und Vorhersage der zehnfachen Kreuzvalidierung. <i>Fortsetzung</i>	107

3.20	Lineare Korrelationen für Training und Vorhersage der zehnfachen Kreuzvalidierung.	108
3.21	Lineare Korrelationen für Training und Vorhersage der zehnfachen Kreuzvalidierung.	109
3.22	Vergleich der kumulativen Erfolgsraten in Abhängigkeit vom entsprechenden RMSD-Wert für a) Platzierungen auf Rang 1 und b) für die jeweils beste Platzierung mit einem beliebigen Rang.	111
3.23	Vergleich der nativen Struktur (<i>dünn</i>) und der von AISCORE auf Rang 1 bewerteten Struktur (<i>dick</i>) des Liganden im Komplex 1BV7.	114
3.24	Vergleich der nativen Struktur (<i>dünn</i>) und der von AISCORE auf Rang 1 bewerteten Struktur (<i>dick</i>) des Liganden im Komplex 1MEU.	115
3.25	Vergleich der nativen Struktur (<i>dünn</i>) und der von AISCORE auf Rang 1 bewerteten Struktur (<i>dick</i>) des Liganden im Komplex 1HIV in der Oberflächendarstellung des Proteins.	116
3.26	Die verschiedenen Liganden die WALDVOGEL <i>et al.</i> in ihren synthetischen Koffeinrezeptor komplexiert haben.	118
3.27	Der Koffeinrezeptor von WALDVOGEL <i>et al.</i>	119
3.28	Vergleich der linearen Korrelationen.	120
4.1	Strukturdarstellung des CYP51 im Komplex 1E9X mit dem hervorgehobenen HEM im aktiven Zentrum.	125
4.2	Strukturen einiger Antimykotika auf Azolbasis.	126
4.3	Vergrößerung von <i>candida albicans</i> Präparaten.	127
4.4	Modellsystem (mit Cystein als proximalen Liganden) und die geometrischen Parameter der Imidazol–Eisen–Wechselwirkung.	128
4.5	Eingeschränkter Minimumsenergiepfad für Variation des Eisen–Stickstoff-Abstands $d_{\text{FeN}^{\text{Im}}}$ für a) Cystein als proximalen Liganden und b) für MeS. Die gestrichelten Kurven sind die angepassten Morse-Potenziale.	132
4.6	a) Eingeschränkter Minimumsenergiepfad für die Torsion des Imidazols im Komplex FeP(Im)MeS. Die rote Kurve ist das angepasste Cosinuspotential. b) Eingeschränkte Minimumsenergiepfade für die Verkippung des Imidazols im Komplex FeP(Im)MeS bei jeweils $\theta = 0^\circ$ und $\theta = 45^\circ$	133

Tabellenverzeichnis

2.1	SYBYL-Atomtypen	33
2.2	Vergleich der MP2-Wechselwirkungsenergie $\Delta E_{\text{MP2}}^{\text{CPC}}$ (in kJ Mol^{-1}) und der Wasserstoffbrückenlänge $d_{\text{H-B}}$ (in \AA) für DFT-optimierte Geometrien und für Geometrien die auf MP2-Level optimiert wurden.	41
2.3	Donor- und Akzeptorinkremente, deren Energien (in kJ Mol^{-1}) und die Häufigkeit ihres Vorkommens im Trainingssatz der 81 Komplexe.	59
2.4	Ergebnisse für wasserstoffverbrückte Komplexe, die ein Sauerstoffatom als Akzeptor oder Donor enthalten. Die MP2-Bindungsenergie $\Delta E_{\text{MP2}}^{\text{CPC}}$ [kJ Mol^{-1}] beinhaltet <i>Counterpose-</i> <i>Korrektur</i> . $d_{\text{H-B}}$ [\AA] ist die Wasserstoffbrückenlänge, $d_{\text{BCP-B}}$ [\AA] ist der Abstand zwischen dem bindungskritischen Punkt und dem Akzeptoratom. Die Partialladungen des Wasserstoffatoms ($q^{\text{AIM}}(\text{H})$) und die des Akzeptors ($q^{\text{AIM}}(\text{B})$) sind durch AIM-Analyse des wasserstoffverbrückten Komplexes bestimmt. . . .	60
2.5	Ergebnisse für wasserstoffverbrückte Komplexe mit Stickstoff als Donor oder Akzeptor. Für weitere Erläuterungen siehe Tabelle 2.4.	62
2.6	Ergebnisse für wasserstoffverbrückte Komplexe mit Schwefel als Donor oder Akzeptor. Für weitere Erläuterungen siehe Tabelle 2.4.	63
2.7	Ergebnisse für die wasserstoffverbrückten Komplexe des Testsatzes. Die Energien sind in kJ Mol^{-1} angegeben.	64
3.1	Wechselwirkungen in FLEXX. Die Wechselwirkungsenergie- Inkremente ΔG_{ω} sind in kJ Mol^{-1} angegeben, die idealen Bindungsradii r_0 in \AA	72
3.2	Parameter, die optimiert werden.	78
3.3	Komplexe des Protein-Ligand-Datensatzes.	89
4.1	Geometrische Parameter der vollständig optimierten Modellsysteme und der Kristallstruktur.	129
4.2	Bindungslängen verschiedener Hem(Fe)-Stickstoff-Bindungen. . . .	130

Einleitung

Die Entwicklung eines neuen Wirkstoffes für ein Medikament kann je nach Komplexität des Projektes bis zu 15 Jahre dauern. In dieser Zeit entstehen Kosten von bis zu einer Milliarde Euro. [3, 4, 5]

In vielen Fällen sind die Wirkstoffe kleine Moleküle, die für einen Krankheitsverlauf relevante Proteine gezielt an- oder ausschalten und so den gewünschten therapeutischen Effekt erzielen. Man spricht in diesem Zusammenhang von den Molekülen auch als *Effektoren* oder *Liganden* der Proteins.

Am Anfang der Entwicklung eines neuen Wirkstoffes steht deshalb meistens die Suche nach einer neuen *Leitstruktur*. Gemeint ist damit ein Molekül, das zu einem gegebenen Protein eine nachweisbare Affinität besitzt. [6] Diese Leitstruktur wird anschließend in vielen Schritten optimiert, z.B. um die Affinität zu dem



Abbildung 1.1. Beim experimentellen *High-Throughput Screening* werden hochautomatisierte Robotersysteme eingesetzt.

Rezeptor zu erhöhen oder um rein pharmakokinetische Eigenschaften, wie die Aufnahme des Wirkstoffes, die Verteilung im Körper oder Ausscheidung und Toxizität zu verbessern. [7, 8, 9]

Aufgrund der hohen Kosten und der Zeit, die eine Entwicklung eines neuen Wirkstoffes benötigt, ist es elementar wichtig, in einer sehr frühen Phase eines Projektes vielversprechende Leitstrukturen zu identifizieren und ungeeignete so früh wie möglich auszuschließen, um die ohnehin beschränkten Ressourcen nicht unnötig zu binden.

In der traditionellen Wirkstoffforschung beginnt die Suche nach neuen Leitstrukturen mit dem sog. „*Screening*“. Dabei werden ganze Bibliotheken von synthetischen und natürlichen Molekülen auf eine mögliche Aktivität bezüglich eines Proteins experimentell getestet. [10, 11]. Diese Methoden sind in den letzten Jahren ausgiebig verbessert worden. Beim *High-Throughput Screening* werden z.B. Synthese- und Analyse-Robotersysteme [12] (Abbildung 1.1) eingesetzt, um automatisiert die experimentelle Suche nach aktiven Leitstrukturen durchzuführen. Doch trotz dieser Fortschritte bleibt der experimentelle Ansatz sehr kostenintensiv. Alle zu testenden Stoffe müssen synthetisiert und dann experimentell gegen ein entsprechendes Protein, das isoliert und aufgereinigt werden muss, auf Aktivität getestet werden.

Um noch vor den ersten experimentellen Tests die Anzahl der Kandidaten gezielt und sinnvoll einzuschränken, werden heute immer mehr Computermethoden genutzt. [13, 14] In den späten 90ern wurde der Begriff „*Virtual Screening*“ geprägt. *Virtual Screening* orientiert sich in seiner Vorgehensweise am traditionellen Screening: eine Vielzahl von Stoffen wird auf ihre Aktivität gegenüber einem gegebenen Zielprotein getestet. [15] Dabei werden *hochentwickelte Algorithmen* verwendet, um aus der Menge der Kandidaten geeignete Leitstrukturen herauszufiltern. Da eine ständig wachsende Zahl von dreidimensionalen Proteinstrukturen durch Röntgenstrukturanalyse, NMR-Spektroskopie und Homologie-Modeling verfügbar ist, sind Softwaretools, die diese Informationen nutzen, in der Wirkstoffentwicklung immer wichtiger geworden. [16, 17]

Grundsätzlich gibt es zwei Arten von *Virtual Screening*. Beim *ähnlichkeitsbasierten* Ansatz werden neue Moleküle gesucht, die charakteristische Merkmale (z.B. Ladung, Form oder bestimmte funktionelle Gruppen) mit einem bereits bekannten Wirkstoff gleich haben. Dagegen wird bei einer *strukturbasierten* Suche die mögliche Aktivität eines Kandidaten nur aus der räumlichen Struktur des Moleküls im Komplex mit dem Zielprotein bestimmt.

Zu diesem Zweck braucht *Virtual Screening* ein Verständnis davon, welche Wechselwirkungen dafür verantwortlich sind, dass ein bestimmtes Molekül an einen bestimmten Rezeptor bindet. [14] Somit ist *Virtual Screening* ein *wissensbasierter* Ansatz und der Erfolg hängt stark von Qualität und Quantität der Informationen ab, die über das zu untersuchende System bekannt sind. Dies steht im klaren Gegensatz zum experimentellen *Screening*, welches rein phänomenologisch ist.

Trotz dieser Voraussetzungen hat sich *Virtual Screening* zur Entdeckung neuer Leitstrukturen etabliert. [18, 19, 20] Die Vorteile des *Virtual Screenings* liegen auf der Hand: Die zu untersuchenden Substanzen müssen nicht notwendigerweise existieren und zudem verbraucht der Test auf ihre Aktivität keine materiellen Ressourcen. Während die experimentelle Bestimmung der Aktivität durch schlechte Löslichkeit, Stabilität oder welche Faktoren auch immer gestört werden kann, hat dies keinen Einfluss auf das *Virtual Screening*.

Durch *Virtual Screening*-Methoden wurden bereits erfolgreich Leitstrukturen für Thrombin [21, 22], Thymidylatsynthase [23, 24], DHFR-Enzyme [25] und für die HIV-Protease [26, 27] gefunden.

Eine Schlüsseltechnologie für *Virtual Screening* ist das *Docking* – die algorithmische Vorhersage der dreidimensionalen Struktur des Komplexes aus einem Protein und einem Liganden. Von maßgeblicher Bedeutung ist dabei eine schnelle Berechnung der *freien Bindungsenergie* des Komplexes aus Ligand und Rezeptor. Nur wenn diese Energie hinreichend genau und verlässlich berechnet werden kann, können durch Computermethoden neue Leitstrukturen in der Masse der Verbindungen einer Komponentenbibliothek heraus entdeckt werden. [6]

Da *Wasserstoffbrücken* (H-Brücken) einen signifikanten Einfluss auf chemische und biologische Systeme haben, wird jeder Ansatz, die Wechselwirkungsenergie solcher Systeme vorherzusagen, misslingen, wenn die Vielfältigkeit von H-Brücken nicht hinreichend berücksichtigt wird.

FLEXX [28, 29, 30] ist ein Programm, welches mittels Docking-Methoden die geometrische Struktur und die freie Bindungsenergie von Protein-Ligand-Komplexen vorhersagt. In einer großen, vergleichenden Studie von 11 Docking-Programmen, konnte es die Geometrie von 90% der getesteten Protein-Ligand-Komplexe richtig berechnen. [31] Doch fehlt FLEXX eine differenzierte Behandlung von H-Brücken, so dass es trotz richtiger Komplexgeometrie nicht in der Lage ist, die freie Bindungsenergie der Komplexe richtig vorherzusagen. [32, 33]

Eine chemisch diverse Behandlung von Wasserstoffbrücken erscheint deshalb ein vielversprechender Ansatz, die Berechnung der freien Bindungsenergie in FLEXX zu verbessern.

1.1 Aufbau und Ziel dieser Arbeit

Ziel dieser Arbeit war es, eine chemisch diverse Bewertungsfunktion für Rezeptor–Ligand-Komplexe zu entwickeln – AISCORE. Fokus war dabei eine differenzierte Behandlung von Wasserstoffbrücken, die den chemischen Kontext berücksichtigt.

Als erstes werden deshalb die Eigenschaften von Wasserstoffbrücken zusammengefasst (Kapitel 2.1). In Kapitel 2.2 werden quantenchemische Methoden vorgestellt, die geeignet sind, Wechselwirkungsenergien wasserstoffverbrückter Komplexe zu berechnen. Ein Satz von 81 wasserstoffverbrückten Komplexen kleiner Moleküle wurde erstellt, um die Wechselwirkungsenergien verschiedener Wasserstoffbrücken in unterschiedlichen chemischen Umgebungen zu untersuchen (Kapitel 2.3.1). In einem anschließenden Schritt, wurde aus diesen Energien ein additives Inkrementsystem für Wasserstoffbrücken abgeleitet (Kapitel 2.4.2.2), das HBINC genannt wurde.

Die Inkremente von HBINC wurden dann für die Parametrisierung von AISCORE benutzt (Kapitel 3.4.2). AISCORE basiert auf dem LUDI-Modell von HANS-JOACHIM BÖHM [34, 35, 36, 37], das in Kapitel 3.3.1 vorgestellt wird. Das ursprüngliche LUDI-Konzept wurde um einen Algorithmus zur Behandlung *mehrfach gegabelter H-Brücken* erweitert (XFURCATE, Kapitel 3.4.3).

Für Optimierung und Validierung von AISCORE wurde ein Satz von 100 Protein–Ligand-Komplexen, für die experimentelle freie Bindungsenergien verfügbar waren, zusammengestellt (Kapitel 3.4.4).

AISCORE und XFURCATE wurden in das Docking-Programm FLEXX implementiert und für die Vorhersage der freien Bindungsenergie dieser Protein–Ligand-Komplexe verwendet (Kapitel 3.5).

Am Beispiel eines synthetischen Rezeptors für Koffein mit unterschiedlichen alkylierten Oxypurinen, wurde AISCORE auch in dem Tool FLEXR getestet (Kapitel 3.6).

Quantenchemische Rechnungen an Wasserstoffbrücken

Hydride sind die Verbindungen des Wasserstoffs mit anderen Elementen. Betrachtet man die Siedepunkte der Hydride mit Elementen der vierten Hauptgruppe, so findet man, dass sich mit steigender Periode die Siedepunkte durch die Zunahme der VAN-DER-WAALS-Wechselwirkungen zwischen den Molekülen erhöhen (Abbildung 2.1a).

Ein völlig anderes Verhalten der Siedepunkte zeigen die Hydride der Elemente der Gruppen 5, 6 und 7 (Abbildung 2.1b). Bei diesen Hydriden sind die Schmelzpunkte der Hydride der jeweils ersten Elemente einer Gruppe abnormal hoch. Hier muss also eine zusätzliche Kraft zwischen den Molekülen wirken, die deutlich mehr Energie nötig macht, um die Anziehungskräfte zwischen den Molekülen

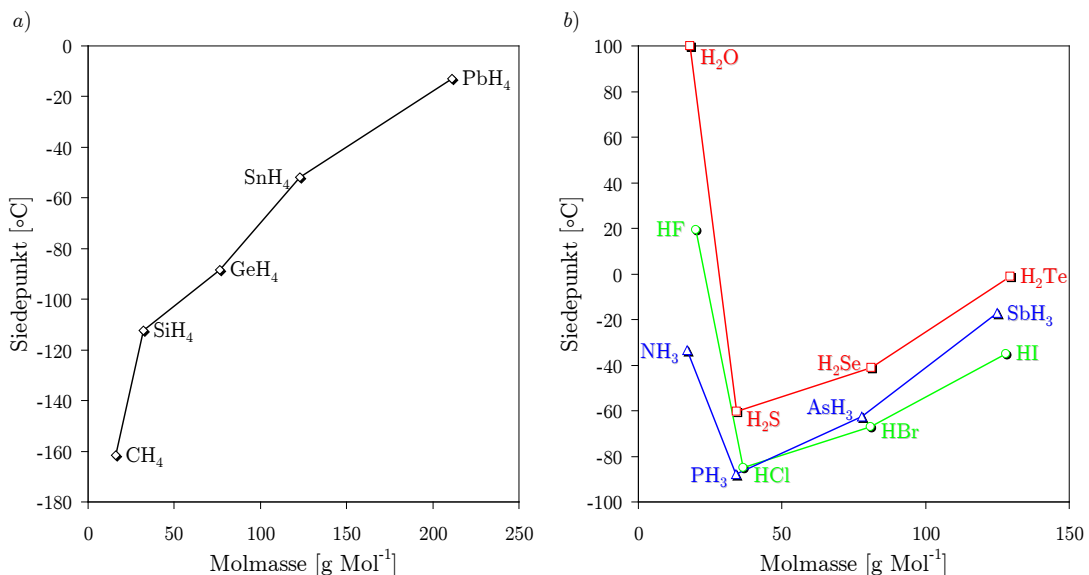


Abbildung 2.1. Siedepunkte in Abhängigkeit von der Molmasse a) für Hydride der Elemente der Gruppe 4 und b) für Elemente der Gruppen 5, 6 und 7.

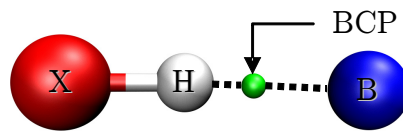


Abbildung 2.2. Modell einer Wasserstoffbrückenbindung zwischen einem Donor X – H und einem Akzeptor B mit dem Bindungskritischen Punkt BCP.

im Verlauf des Siedens zu brechen. Bei dieser zusätzlichen Kraft handelt es sich um *Wasserstoffbrücken*.

Die Wasserstoffbrücke gilt auch heute noch, viele Jahre nachdem sie in den Laboren von G. N. LEWIS in den 20er Jahren des 20. Jahrhunderts das erste Mal vorgeschlagen wurde [38], als eine der außergewöhnlichsten Bindungen, deren physikalische Natur immer noch Gegenstand aktueller Forschung ist. [39] Mit einer typischen Bindungslänge von 1.97 Å ist sie deutlich länger als eine kovalente Bindung. Ferner weist sie nur bis zu etwa einem Zehntel der Bindungsenergie einer kovalenten Wechselwirkung auf. Tatsächlich werden H-Brücken oft als langreichweitige VAN-DER-WAALS-Wechselwirkungen gesehen. Jedoch sind Wasserstoffbrücken gerichtet und viel stärker als reine elektrostatische Wechselwirkungen. Die interatomaren Abstände der an der H-Brücke beteiligten Atome

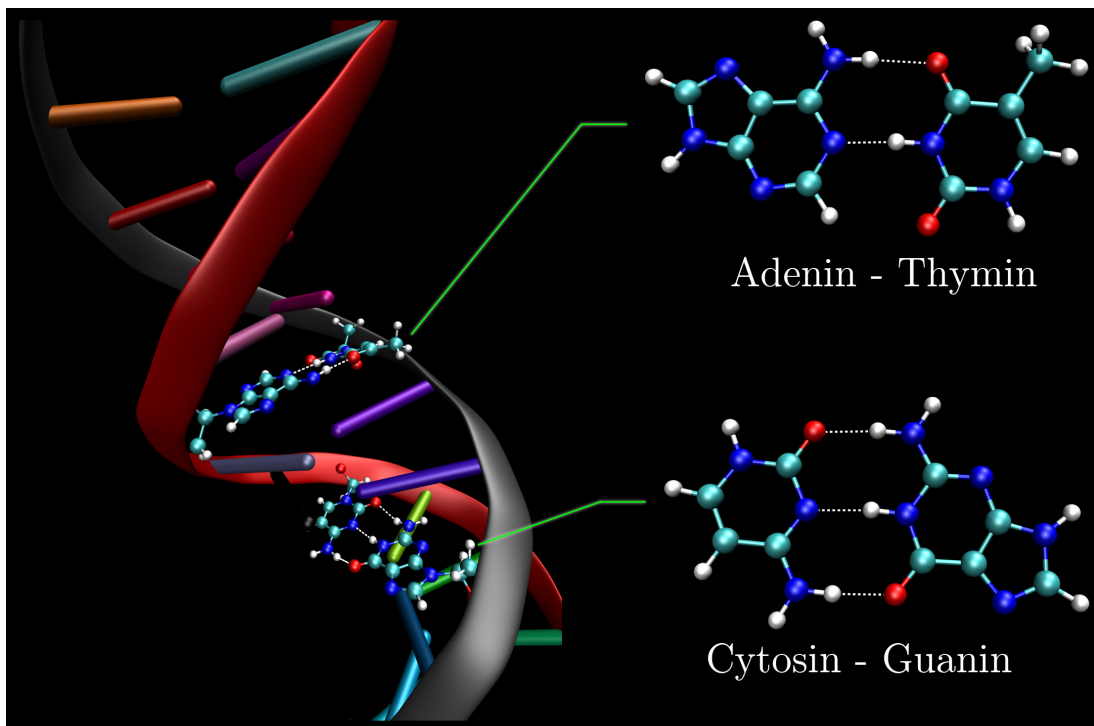


Abbildung 2.3. Schematischer Aufbau der DNA und Wasserstoffbrücken der Basenpaare.

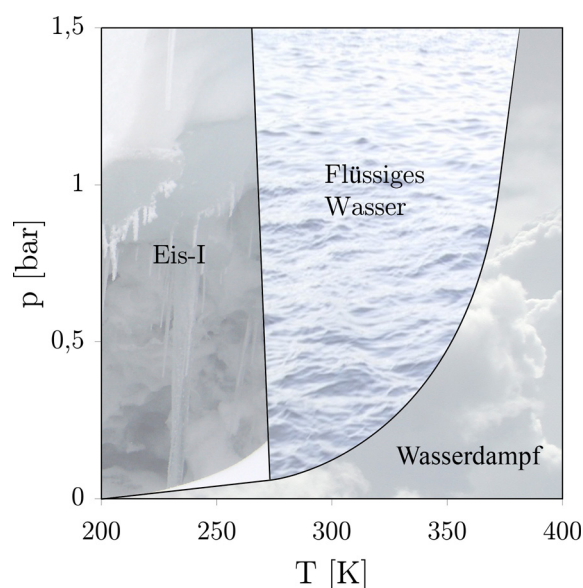


Abbildung 2.4. Phasendiagramm von Wasser.

sind kürzer als die Summe der VAN-DER-WAALS-Radien (der sog. “*penetration effect*“, [40, 41]). Darüber hinaus sind für gewöhnlich immer nur eine *begrenzte Anzahl* an Wechselwirkungspartnern involviert, was man als eine Art von *Valenz* ansehen kann. All dies sind Eigenschaften, die man eher kovalenten Bindungen zuschreiben würde.

Wasserstoffbrücken spielen eine große Rolle in den meisten chemischen und biologischen Systemen. [42] Rein phänomenologisch ist eine Wasserstoffbrücke eine attraktive Wechselwirkung zwischen einem *Protonendonator* $X - H$ und einem *Protonenakzeptor* B im selben oder in verschiedenen Molekülen (Abbildung 2.2). [43]

H-Brücken sind *hoch spezifisch* und hinsichtlich ihrer Geometrie *hochgradig restriktiv*, wodurch sie eine signifikante Rolle bei der Stabilisierung von Makromolekülen einnehmen. [44]

Die DNA verdankt ihre außergewöhnliche Struktur [45] den verschiedenen Wasserstoffbrücken zwischen den vier Nukleinbasen Adenin, Cytosin, Guanin und Thymin (Abbildung 2.3). Die DNA-Replikation ist nur deswegen derart genau, weil diese Wasserstoffbrücken so hoch spezifisch sind. [46, 47, 48, 49]

Wasser zeigt ein ungewöhnliches Verhalten beim Schmelzen: das Volumen nimmt ab (Abbildung 2.4). Das hat zur Folge, dass bei Druckerhöhung Wasser die flüssige Form bevorzugt. Grund für die Volumenabnahme ist die recht lockere Packung der Wassermoleküle im Eis, für die Wasserstoffbrücken verantwortlich zeichnen. [50]

H-Brücken sind auch für die molekulare Erkennung in artifiziiellen Systemen außerhalb des Körpers die wichtigsten Wechselwirkungen. Ein vollständiges Wissen über die optimalen H-Brücken zwischen dem Rezeptor und dem Liganden (oder Gastmolekül) ist eine entscheidende Voraussetzung für eine gezielte Entwicklung solcher Systeme. [51]

Sauerstoff, Stickstoff und Fluor sind die Elemente mit der höchsten Elektronegativität. Mit wenigen Ausnahmen sind mehrheitlich diese Elemente an Wasserstoffbrücken beteiligt. Nach diesem klassischen Bild kommt es zu einer Wasserstoffbrücke durch die Wechselwirkung eines elektronegativen Atoms, welches mindestens ein freies Elektronenpaar (*lone pair*) besitzt, und einem Proton, das an ein weiteres elektronegatives Atom gebunden ist. [52, 53, 54, 55] Doch auch sog. schwache H-Brücken, die weniger elektronegative Elemente enthalten, sind für die Struktur und die Funktion von biologischen Makromolekülen wichtig. CH-azide Protonen können H-Brücken ausbilden, welche auch eine wichtige Rolle bei der Proteinfaltung spielen. [56, 57] Schwefel wird von dem Docking-Programm FLEXX nicht als H-Brückenpartner behandelt. Es ist aber schon allein deswegen von Interesse, weil es im Körper reichlich vorhanden ist – immerhin enthalten zwei unserer Aminosäuren Schwefel. [42]

Durch die Bildung der Wasserstoffbrücke $X - H \cdots B$ erhöht sich die Polarität der $X - H$ -Bindung. Das wiederum kann zu einer signifikanten Erhöhung der Bindungsstärke von weiteren Wasserstoffbrücken führen (Kooperativer Effekt).

2.1 Der Ursprung der Wasserstoffbrücke

Ein H-Atom kann als Wasserstoffbrücken-*Donor* fungieren, wenn es an ein Atom gebunden ist, welches eine höhere Elektronegativität als das H-Atom selber besitzt. Das elektronegrativere Atom zieht dabei Elektronendichte vom Wasserstoff an, dezentralisiert sie je nach seiner *Elektronegativität* (Abbildung 2.5). Dadurch bleibt am Wasserstoff eine leicht positive Ladung zurück, die durch die geringe Größe des Wasserstoffs zu einer hohen *Ladungsdichte* am Wasserstoffatom führt. Das positive Wasserstoffatom bindet dann an eine elektronenreiche Region des Wasserstoffbrücken-*Akzeptors*.

Eine Beschreibung der Wasserstoffbrücke als reine Dipol-Dipol-Wechselwirkung wäre aber zu kurz gegriffen, da sich gezeigt hat, dass es nur geringe oder gar

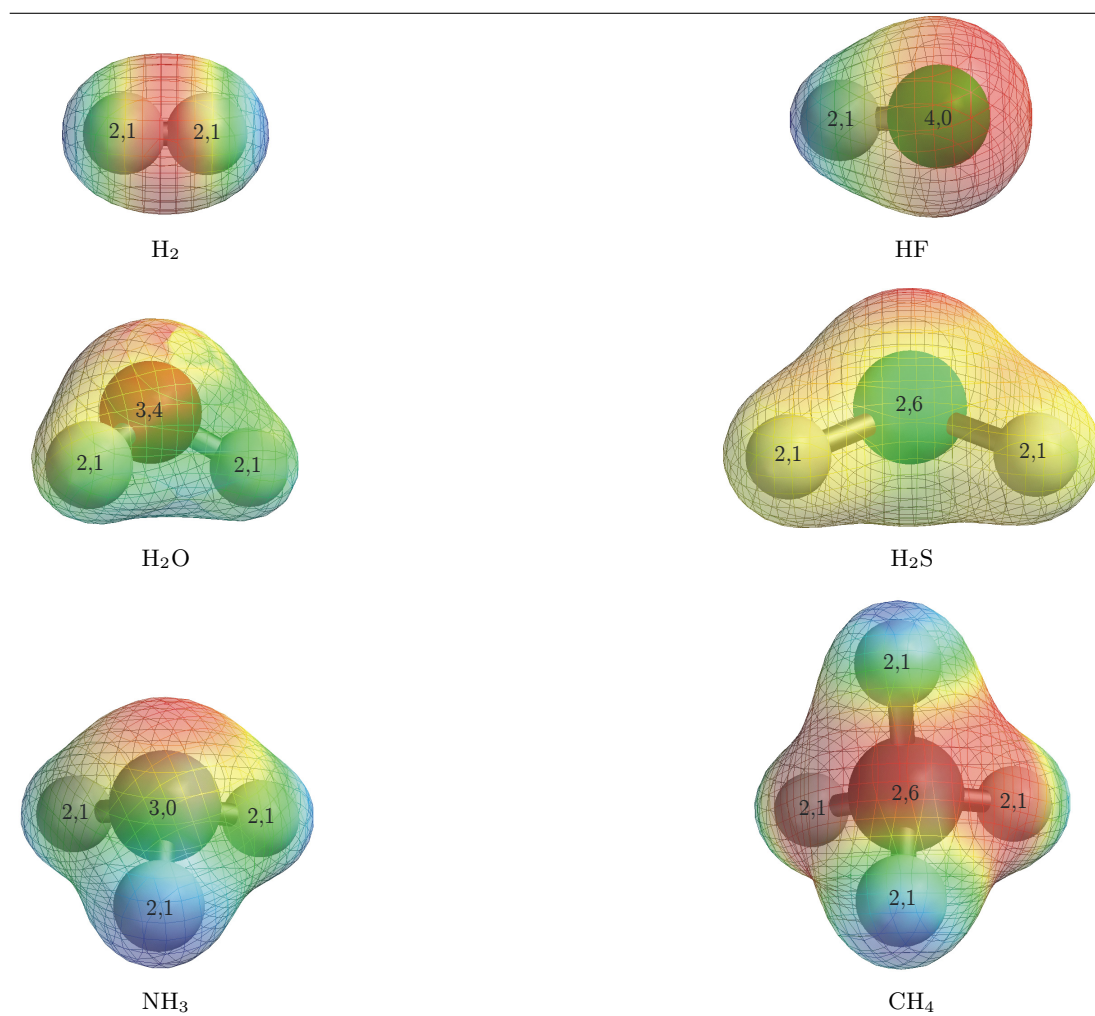


Abbildung 2.5. Isoflächen der Elektronendichte verschiedener Moleküle. Die Farbe der Fläche entspricht dem elektrischen Potential an dieser Stelle (rot entspricht dabei einer negativen Ladung, blau einer positiven). Die Zahlen in den Atomen entsprechen der jeweiligen Elektronegativität.

keine Korrelation zwischen der H-Brücken-Stärke und dem Dipolmoment des H-Akzeptors B gibt. [43] Vielmehr spielen die Beiträge mehrerer Wechselwirkungen zusammen.

Elektrostatische Wechselwirkungen

Elektrostatische Wechselwirkungen ergeben sich durch permanente Punktladungen, Dipole, Quadrupole und höhere Multipole, die über ein Molekül verteilt sind. Die elektrostatische Wechselwirkung von zwei nicht-kovalent gebundenen Molekülen oder Molekülteilen ist dann die Summe über Wechselwirkungen zwischen Paaren dieser Multipole. (Abbildung 2.6).

Die Topologie des intermolekularen Potentials für wasserstoffverbrückte Moleküle wird von den elektrostatischen Wechselwirkungen dominiert. [58] Diese Wechselwirkungen hängen stark von der Orientierung ab. Aus diesem Grund sind sie maßgeblich dafür verantwortlich, dass Wasserstoffbrücken bestimmte Regionen am Akzeptor bevorzugen und bestimmen somit die geometrische Struktur wasserstoffverbrückter Moleküle.

Induktion

In einem externen Feld, was z.B. durch benachbarte Moleküle erzeugt werden kann, richtet sich die Verteilung der Elektronen in einem Molekül gemäß dieses

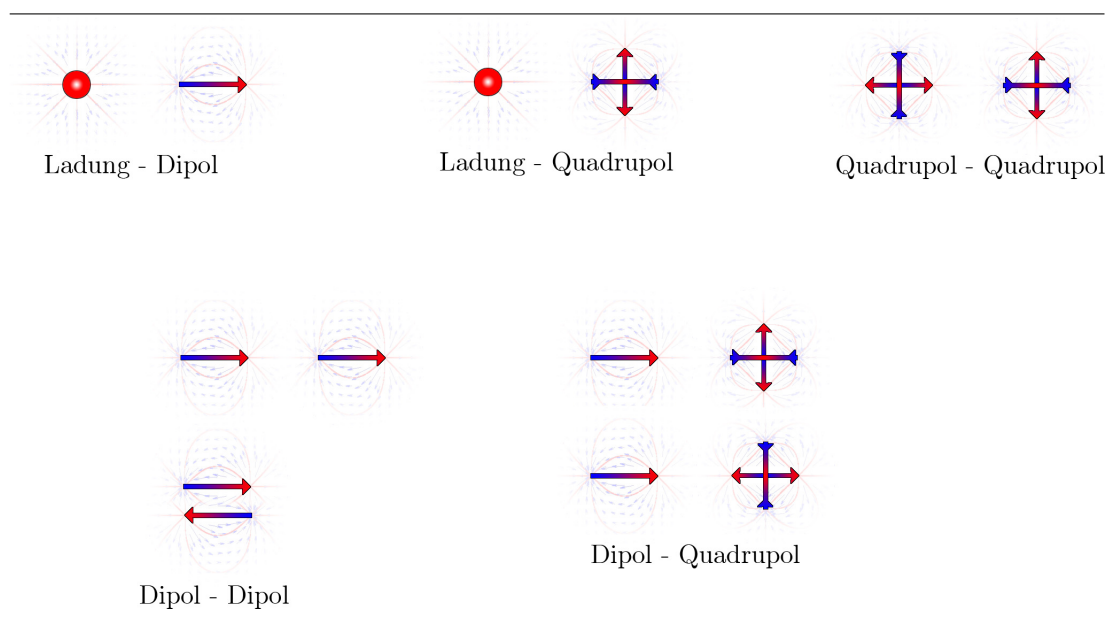


Abbildung 2.6. Die jeweils günstigsten Orientierungen für verschiedene Multipole (in Anlehnung an [1]).

Feldes aus. Dadurch entsteht in diesem Molekül ein zusätzlicher Dipol. Die Größe des induzierten Dipols ist proportional zum elektrischen Feld. Durch diese induzierten Dipole kommt es, zusätzlich zu eventuellen permanenten elektrostatischen Wechselwirkungen, zu Wechselwirkungen mit dem externen Feld.

Natürlich beeinflusst ein erzeugter Dipol in einem Molekül die Ladungsverteilung in einem anderen Molekül, welche im Gegenzug wieder die Ladungsverteilung im ersten Molekül beeinflusst.

Die Induktionsenergie ist ein sehr wichtiger Bestandteil der Wechselwirkung in Wasserstoffbrücken, obwohl sie die Wechselwirkungsenergie nicht dominiert. [58, 59]

Dispersion

Dispersionskräfte ergeben sich durch *spontane Dipole* in Molekülen oder Atomen, die durch Fluktuationen in der Elektronendichte entstehen. Selbst in einem sphärisch-symmetrischen Atom ist die Elektronendichte nur im *zeitlichen Mittel* gleichförmig verteilt. In Wirklichkeit sind die Elektronen hoch mobil und es kann jederzeit zu einer temporären lokalen Anhäufung von Elektronen kommen. Dadurch entsteht eine temporäre negative Ladung. Gleichzeitig ist eine andere Region im Atom um eben diese Elektronen ärmer und weist temporär eine entsprechend positive Ladung auf. Durch diese Fluktuationen in der Elektronendichte entstehen temporäre Dipole.

Diese temporären Dipole können ihrerseits wieder Dipole in benachbarten Atomen oder Molekülen induzieren.

Bei schwachen Wasserstoffbrücken und bei H-Donoren und -Akzeptoren mit gut polarisierbaren Elementen (z.B. Schwefel) können die dispersiven Wechselwirkungen der bestimmende Faktor für die Bindungsstärke werden. [42, 60] Auch in DNA-Basenpaaren, die sogar besonders starke Wasserstoffbrücken aufweisen, wurde bereits ein größerer Einfluss dispersiver Effekte auf die H-Brückenlänge gefunden, als man ursprünglich gedacht hatte. [61]

Repulsion

Die Repulsion kann als Effekt des Pauli-Prinzips verstanden werden, und wird durch Elektronen gleichen Spins verursacht, weswegen man auch von Austauschwechselwirkung spricht. [62]

Durch den Austausch wird es jedem beliebigen Paar von Elektronen verboten, dieselbe intermolekulare Raumregion einzunehmen. Dadurch wird Elektronen-

dichte aus den Regionen zwischen den Molekülen abgezogen. Als Folge werden die Atomkerne nur noch unvollständig abgeschirmt und stoßen sich ab.

Die Repulsion ist in guter Näherung proportional zum Überlapp der Ladungsdichten der beteiligten Moleküle oder Atome. [63]

Ladungstransfer

Die *Ladungstransfer*-Energie beschreibt die Stabilisierung durch den Elektronentransfer von einem Elektronendonator zu einem Akzeptor, wie er auch in Wasserstoffbrücken vorkommt. [58]. In den Watson–Crick-Komplexen der DNA-Basen ist der Ladungstransfer (*charge transfer*, *CT*) zwischen den einsamen Elektronenpaaren am Sauerstoff oder Stickstoff der einen Base zu der N – H-Bindung der jeweils anderen Base von vergleichbarer Stärke wie die elektrostatischen Wechselwirkungen. [64, 65, 66] Neuere Untersuchungen beziffern den Anteil der *CT*-Stabilisierung an der Gesamtwechselwirkungsenergie der DNA-Basenpaare auf 25–30%. [2]

Resonanz

Eine moderne Sicht auf Wasserstoffbrücken schreibt diesen auch einen kovalenten Charakter zu, nicht zuletzt deswegen, weil Wasserstoffbrücken gerichtet sind. [67] GILLI *et al.* haben dafür den Begriff „*resonanzunterstützte Wasserstoffbrücke*“ (RAHB = *resonance-assisted hydrogen bond*) geprägt. [68, 69, 70, 71, 72] Nach diesem Modell ist der eine Grenzfall der extrem schwachen Wasserstoffbrücke eine rein elektrostatische Wechselwirkung und tatsächlich sind schwache Wasserstoffbrücken fast gar nicht von reinen VAN-DER-WAALS-Wechselwirkungen zu unterscheiden. [73]. Doch je stärker die H–Brücke wird, desto größer wird ihr kovalenter Charakter, bis im anderen Grenzfall eine sehr starke Wasserstoffbrücke letztlich eine Dreizentren–Vierelektronen-Bindung ist.

Obwohl GILLI *et al.* ursprünglich das Konzept der RAHBs nur für intramolekulare Wasserstoffbrücken in β -Diketonen und β -Dienaminonen formuliert haben, kann es auch auf intermolekulare Wechselwirkungen zwischen Carbonsäuren, DNA-Paaren und Proteinen angewendet werden. [61, 2]

Im intra- wie im intermolekularen Fall wird durch die Existenz von verschiedenen Resonanzstrukturen des Protonendonors dessen Säurekonstante geändert. Am Protonenakzeptor kommt es durch unterschiedliche Resonanzstrukturen dort zu einer Änderung der Protonenaffinität. Beides zusammen führt zu einer Verstärkung der Wasserstoffbrücke.

Eine alternative Sicht auf RAHBs ist, dass es, unterstützt durch die Delokalisierung der beteiligten π -Elektronen, zu zusätzlichen, günstigen Dipolen in den Molekülen kommt. Danach würde es sich bei den RAHBs um eine rein elektrostatische Wechselwirkung handeln. [67]

2.2 Quantenchemische Grundlagen

Im Folgenden sollen die Grundlagen der quantenchemischen Methoden kurz erläutert werden, die ich für meine Rechnungen an wasserstoffverbrückten Komplexen benutzt habe. Dabei steht im Vordergrund, die Konzepte der einzelnen Methoden zu erläutern, so dass deren Eigenheiten erkennbar werden. Dies ist später wichtig, wenn es darum geht, welche der Methoden für die jeweiligen Anwendungen geeignet ist. Aus diesem Grund wird auf komplette Herleitungen verzichtet und nur wesentliche Formeln, an denen die Grundstrukturen erkennbar sind, werden gezeigt.

Es wird das in der Quantenchemie übliche *atomare Einheitensystem* verwendet:

$$\begin{aligned}\frac{1}{4\pi\epsilon_0} &= 1 \\ \hbar &= 1 \\ m_e &= 1 \\ e &= 1\end{aligned}$$

Soweit möglich werden die Ausdrücke durch die Diracsche „*bra-ket*“-Nomenklatur verkürzt; für die Notation der atomaren und molekularen Integrale siehe die übliche Literatur (z.B. Referenz [74]).

2.2.1 Die Schrödingergleichung

In der Quantenmechanik beschreibt man den Zustand eines Systems durch seine Wellenfunktion Ψ . Im Falle eines molekularen Systems mit N Elektronen und M Atomkernen hängt die Wellenfunktion von den Koordinaten \mathbf{r} aller Elektronen ($\mathbf{r} = \mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N$) und den Koordinaten \mathbf{R} aller Kerne ($\mathbf{R} = \mathbf{R}_1, \mathbf{R}_2, \dots, \mathbf{R}_M$) ab.

Der Wellenfunktion eines Systems kommt in der Quantenmechanik eine zentrale Bedeutung zu, da eines der wesentlichen Postulate der Quantenmechanik besagt, dass alle messbaren Eigenschaften eines Systems berechenbar sind, wenn nur die Wellenfunktion bekannt ist.

Die Wellenfunktion $\Psi(\mathbf{r}, \mathbf{R})$ eines molekularen Systems, dessen Eigenschaften zeitunabhängig sind, ist eine Eigenfunktion der zeitunabhängigen Schrödingergleichung: [75, 76, 77, 78]

$$\hat{\mathcal{H}}\Psi(\mathbf{r}, \mathbf{R}) = \mathcal{E}_n \Psi(\mathbf{r}, \mathbf{R}). \quad (2.1)$$

Dabei sind \mathcal{E}_n die Energieeigenwerte des Systems und $\hat{\mathcal{H}}$ der *nicht-relativistische Hamiltonoperator*. Letzterer ist der Operator der Wechselwirkungen von Elektronen und Kernen in dem System und setzt sich aus der *kinetischen Energie* \hat{T} und der *potenziellen Energie* \hat{V} der Elektronen und Kerne zusammen:

$$\hat{\mathcal{H}} = \hat{T} + \hat{V} \quad (2.2)$$

Der Operator \hat{T} für die kinetische Energie besteht seinerseits wieder aus der kinetischen Energie der Elektronen \hat{T}_e und der kinetischen Energie der Kerne \hat{T}_n :

$$\begin{aligned} \hat{T} &= \hat{T}_e + \hat{T}_n \\ &= -\frac{1}{2} \sum_i^N \nabla_i^2 - \frac{1}{2} \sum_A^M \frac{1}{m_A} \nabla_A^2 \end{aligned} \quad (2.3)$$

$$\nabla_i^2 = \left(\frac{\partial^2}{\partial x_i^2} + \frac{\partial^2}{\partial y_i^2} + \frac{\partial^2}{\partial z_i^2} \right) \quad (2.4)$$

Hier ist m_A die Masse des Atoms A .

Die potentielle Energie aller Coulomb-Wechselwirkungen besteht aus der Anziehung der Elektronen und der Kerne \hat{V}_{en} , der Elektronen–Elektron-Abstoßung \hat{V}_{ee} und der Kern–Kern-Abstoßung \hat{V}_{nn} :

$$\begin{aligned} \hat{V} &= \hat{V}_{ne} + \hat{V}_{ee} + \hat{V}_{nn} \\ &= -\sum_i^N \sum_A^M \frac{Z_A}{r_{i,A}} + \sum_i^N \sum_{j>i}^N \frac{1}{r_{i,j}} + \sum_A^M \sum_{B>A}^M \frac{Z_A Z_B}{r_{A,B}} \end{aligned} \quad (2.5)$$

wobei Z_A die Kernladungszahl des Atoms A , $r_{i,A}$ der Abstand zwischen Atom A und Elektron i , $r_{i,j}$ der Abstand zwischen den beiden Elektronen i und j und $r_{A,B}$ der Abstand zwischen den Atomen A und B ist.

Systeme, die chemisch interessant sind, haben so viele Elektronen und Kerne, dass Gleichung (2.1) nicht mehr exakt lösbar ist. In der *Born–Oppenheimer-Näherung* lassen sich Kern- und Elektronenbewegung näherungsweise separieren, da die Elektronen ungleich leichter sind als die Kerne. Die kleinere Masse erlaubt es den Elektronen, auf jede (langsame) Kernbewegung fast instantan zu reagieren. Man beschreibt dann die Elektronenbewegung im entsprechenden Feld einer ruhenden Kernkonfiguration. Die Gesamtwellenfunktion $\Psi(\mathbf{r}, \mathbf{R})$ wird angesetzt als Produkt einer elektronischen Wellenfunktion Ψ_E und einer Kernwellenfunktion Ψ_N :

$$\Psi(\mathbf{r}, \mathbf{R}) = \Psi_E(\mathbf{r}) \cdot \Psi_N(\mathbf{R}) \quad (2.6)$$

Für die Energie der Elektronen im ruhenden Feld der Kerne ergibt sich der *elektronische Hamiltonoperator*

$$\begin{aligned} \hat{\mathcal{H}}_E &= \hat{h}_i + \hat{V}_{ee} \\ &= \sum_i^N \left(-\frac{1}{2} \nabla_i^2 - \sum_A^M \frac{Z_A}{r_{i,A}} \right) + \sum_i^N \sum_{j>i}^N \frac{1}{r_{i,j}}, \end{aligned} \quad (2.7)$$

der nur noch *parametrisch* von den Kernkoordinaten \mathbf{R} abhängt. In die Gesamtenergie des Moleküls bei ruhender Kernkonfiguration geht außerdem noch die Kern–Kern–Abstoßung \hat{V}_{nn} mit ein.

2.2.2 Slater-Determinanten

Für die Lösung der elektronischen Schrödingergleichung hat sich eine Separation der Vielelektronenwellenfunktion $\Psi_E(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N)$ in einzelne Einelektronenwellenfunktionen als sinnvoll erwiesen. Diese Einelektronenwellenfunktionen sind die *Molekülorbitale (MO)*. Ein *Raumorbital* $\phi(\mathbf{r})$ beschreibt die *räumliche* Verteilung eines Elektrons, die *Spinfunktion* $\sigma(s)$ gibt den *Spinfreiheitsgrad* des Elektrons an. Das *Spinorbital* $\chi(\mathbf{x})$ eines Elektrons beschreibt seinen gesamten Zustand und ist das Produkt aus:

$$\chi(\mathbf{x}) = \phi(\mathbf{r}) \cdot \sigma(s) \quad (2.8)$$

mit

$$\begin{aligned}\sigma = \alpha & : \text{Spin } \frac{1}{2} \\ \sigma = \beta & : \text{Spin } -\frac{1}{2}\end{aligned}$$

Für $\Psi_E(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N)$ setzt man jetzt eine Determinante aus orthonormalen Spinorbitalen $\chi_i(\mathbf{x}_i)$ an:

$$\Phi(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \chi_1(\mathbf{x}_1) & \chi_2(\mathbf{x}_1) & \cdots & \chi_N(\mathbf{x}_1) \\ \chi_1(\mathbf{x}_2) & \chi_2(\mathbf{x}_2) & \cdots & \chi_N(\mathbf{x}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \chi_1(\mathbf{x}_N) & \chi_2(\mathbf{x}_N) & \cdots & \chi_N(\mathbf{x}_N) \end{vmatrix} \quad (2.9)$$

$$= |\chi_1 \chi_2 \cdots \chi_N\rangle \quad (2.10)$$

Diese *Slater-Determinante* ist antisymmetrisch bezüglich der Vertauschung zweier Spalten oder zweier Zeilen. Da beide Operationen jeweils die Vertauschung zweier Elektronen darstellen, erfüllt die Slater-Determinante die nach dem *Pauli-Prinzip* [79] geforderte Antisymmetrie der Gesamtwellenfunktion bei Permutation der Elektronen.

Dieser Ansatz beschreibt ein System nicht-wechselwirkender Elektronen, berücksichtigt aber die Austauschkorrelation von Elektronen mit parallelem Spin.

2.2.3 HF – Die Hartree–Fock-Methode

Die HARTREE–FOCK-Methode dient der näherungsweisen Lösung der Schrödingergleichung. Dafür wird zunächst die Wellenfunktion als Slater-Determinante des Grundzustandes angesetzt:

$$\Psi_0^{\text{HF}} = |\chi_1 \chi_2 \cdots \chi_N\rangle \quad (2.11)$$

Durch Optimierung des Erwartungswertes $E[\Psi_0^{\text{HF}}]$ des elektronischen Hamiltonoperators über Variation der Spinorbitale χ_i wird die optimale Funktion Ψ_0^{HF} bestimmt. Es ist

$$E[\Psi_0^{\text{HF}}] = \frac{\langle \Psi_0^{\text{HF}} | \hat{\mathcal{H}}_E | \Psi_0^{\text{HF}} \rangle}{\langle \Psi_0^{\text{HF}} | \Psi_0^{\text{HF}} \rangle} \quad (2.12)$$

Entsprechend den Slater–Condon-Regeln [80, 81, 82] ist

$$E_0^{\text{HF}} = \sum_i^N h_{ii} + \frac{1}{2} \sum_i^N \sum_j^N (J_{ij} - K_{ij}) \quad (2.13)$$

$$= \sum_i^N [i|\hat{h}|i] + \frac{1}{2} \sum_i^N \sum_j^N ([ii|jj] - [ij|ji]). \quad (2.14)$$

Dabei sind $h_{ij} = [\chi_i|\hat{h}|\chi_j]$ die Eielektronenintegrale. J_{ij} beschreibt die Coulomb-Wechselwirkungen und K_{ij} die Austausch-Wechselwirkungen.

Durch *Lagrange-Multiplikation* mit ϵ_{ij} variiert man χ_i unter der Nebenbedingung des Erhalts der Orthonormalität der Spinorbitale:

$$L = \frac{\langle \Psi_0^{\text{HF}} | \hat{\mathcal{H}}_E | \Psi_0^{\text{HF}} \rangle}{\langle \Psi_0^{\text{HF}} | \Psi_0^{\text{HF}} \rangle} - \sum_i^N \sum_j^N \epsilon_{ij} (\langle \chi_i | \chi_j \rangle - \delta_{ij}) \quad (2.15)$$

Minimierung und Diagonalisierung der Matrix der Lagrange-Multiplikatoren liefern die sog. *kanonische HARTREE–FOCK-Gleichung*

$$\hat{F}(\mathbf{x}_1) \chi_i(\mathbf{x}_1) = \epsilon_i \chi_i(\mathbf{x}_1), \quad (2.16)$$

mit der sich nun die Spinorbitale χ_i bestimmen lassen. \hat{F} ist der *Fockoperator*

$$\hat{F} = \hat{h} + \underbrace{\sum_i^N (\hat{J}_i - \hat{K}_i)}_{\hat{v}^{\text{HF}}}, \quad (2.17)$$

wobei \hat{h} die kinetische Energie des Elektrons in einem Spinorbital und die Wechselwirkung mit den positiven Atomkernen beschreibt.

$$\hat{h} = -\frac{1}{2}\nabla^2 - \sum_A^M \frac{Z_A}{r_A}. \quad (2.18)$$

\hat{v}^{HF} beschreibt die gemittelte Wechselwirkung eines Elektrons mit allen anderen Elektronen. Der Coulomb-Operator \hat{J}_i hat die Form

$$\hat{J}_i \chi_j = \left\langle \chi_i \left| \frac{1}{r_{12}} \right| \chi_i \right\rangle \chi_j, \quad (2.19)$$

der Austausch-Operator \hat{K}_i ist

$$\hat{K}_i \chi_j = \left\langle \chi_i \left| \frac{1}{r_{12}} \right| \chi_j \right\rangle \chi_i. \quad (2.20)$$

Die Eigenwerte ϵ_i in Gleichung (2.16) sind die Orbitalenergien der entsprechenden Spinorbitale.

$$\epsilon_i = h_{ii} + \sum_j^N ([ii|jj] - [ij|ji]) \quad (2.21)$$

Spinsummation nach $\langle \sigma | \tau \rangle = \delta_{\sigma\tau}$ liefert uns für den geschlossenschaligen Fall

$$\hat{F}_i = \hat{h}_i + \sum_j^{N/2} (2\hat{J}_{ij} + \hat{K}_{ij}) \quad (2.22)$$

Die Grundzustandsenergie und die Orbitalenergien der spinfreien *Raumorbitale* ψ_i sind dann

$$E_0^{\text{RHF}} = 2 \sum_i^{N/2} h_{ii} + \sum_i^{N/2} \sum_j^{N/2} (2(ii|jj) - (ij|ji)) \quad (2.23)$$

$$\epsilon_i = h_{ii} + \sum_j^{N/2} (2(ii|jj) - (ij|ji)). \quad (2.24)$$

Der LCAO-Ansatz (*linear combination of atomic orbitals*) ist die Näherung der Raumorbitale des Moleküls als Linearkombination einer endlichen Anzahl von Basisfunktionen.

$$\psi_i(\mathbf{r}) = \sum_j^n c_{pi} \phi_j(\mathbf{r}) \quad (2.25)$$

In der Praxis werden als Basisfunktionen Atomorbitale verwendet.

Einsetzen der LCAO-Entwicklung (2.25) in (2.22) führt zu den ROOTHAAN-HALL-Gleichungen [83, 84]

$$\mathbf{F}\mathbf{c} = \mathbf{S}\mathbf{c}\epsilon \quad (2.26)$$

mit der Fock-Matrix F_{ij}

$$F_{ij} = \langle \phi_i | \hat{\mathcal{F}} | \phi_j \rangle \quad (2.27)$$

und der Überlappungsmatrix der nicht orthogonalen Atomorbitale S_{ij}

$$S_{ij} = \langle \phi_i | \phi_j \rangle. \quad (2.28)$$

\mathbf{c} ist die Matrix der MO-Koeffizienten, ϵ die Diagonalmatrix der Orbitalenergien. Der explizite Ausdruck für Fock-Matrixelemente ist

$$F_{ij} = H_{ij}^{\text{core}} + \sum_r^K \sum_s^K P_{rs} \left((ij | rs) - \frac{1}{2} (ir | sj) \right) \quad (2.29)$$

mit P_{rs} der Dichtematrix für den geschlossenschaligen Fall

$$\begin{aligned} P_{rs} &= \sum_i^N c_{ri} c_{si}^* \\ &= 2 \sum_i^{N/2} c_{ri} c_{si}^* \end{aligned} \quad (2.30)$$

und H_{ij}^{core} der Hamilton-Matrix des Einelektronenteils

$$H_{ij}^{\text{core}} = \langle \phi_i | \hat{h} | \phi_j \rangle. \quad (2.31)$$

Damit ist die Fock-Matrix vollständig durch AO-Integrale und MO-Koeffizienten bestimmt.

Für die Aufstellung der Fock-Matrix (2.29) muss zuvor die Dichtematrix (2.30) bekannt sein. Die ROOTHAAN–HALL-Gleichungen werden deshalb iterativ bis zur Selbstkonsistenz gelöst (*SCF* = „*self-consistent field*“).

In der HARTREE–FOCK-Theorie wird die Elektron–Elektron-Abstoßung immer nur gemittelt berücksichtigt. Doch in Realität sind die Bewegungen der Elektronen über die *Fermikorrelation* (zwischen Elektronen gleichen Spins) und der *Coulombkorrelation* (zwischen Elektronen ungleichen Spins) korreliert. Davon wird die Coulombkorrelation in der HARTREE–FOCK-Methode nicht mit einbezogen. Die Differenz zwischen der HARTREE–FOCK-Energie mit gemittelter Elektron–Elektron-Wechselwirkung und der wahren nichtrelativistischen Energie wird *Korrelationsenergie* genannt.

2.2.4 MP2 – Møller–Plesset-Störungstheorie zweiter Ordnung

Die Störungstheorie auf Grundlage der HARTREE–FOCK-Methode geht auf Arbeiten von Møller und Plesset zurück. [85] Bei Störungsrechnungen wird der Hamilton-Operator in einen ungestörten Teil $\hat{\mathcal{H}}^{(0)}$, von dem man die Lösung kennt, und dem Operator einer kleinen Störung $\hat{\mathcal{H}}^{(1)}$, deren Stärke durch einen Störparameter $0 \leq \lambda \leq 1$ bestimmt wird, zerlegt.

$$\hat{\mathcal{H}} = \hat{\mathcal{H}}^{(0)} + \lambda \hat{\mathcal{H}}^{(1)} \quad (2.32)$$

Die Idee dahinter ist, ein System durch einen vereinfachten, aber lösbaren Hamilton-Operator zu beschreiben. Den Fehler, den man durch die vereinfachte Beschreibung macht, fasst man dann als die Störung der vereinfachten Beschreibung auf.

In der MØLLER–PLESSET-Variante der Störungstheorie wird der ungestörte Operator $\hat{\mathcal{H}}^{(0)}$ als Summe der Einteilchen-Fock-Operatoren \hat{f} angesetzt

$$\hat{\mathcal{H}}^{(0)} = \sum_i \hat{f}_i \quad (2.33)$$

Also ist die HARTREE–FOCK-Wellenfunktion Eigenfunktion des ungestörten Systems. Die Elektronenkorrelation wird in der HARTREE–FOCK-Methode nur gemittelt durch den Operator \hat{v}^{HF} beschrieben. Die Störung ist damit die Differenz zwischen der exakten Elektronenwechselwirkung und der gemittelten Wechselwirkung aus dem HARTREE–FOCK-Verfahren.

$$\hat{\mathcal{H}}^{(1)} = \left(\sum_i^N \sum_{j>i}^N \frac{1}{r_{12}} \right) - \sum_i^N \hat{v}_i^{\text{HF}} \quad (2.34)$$

Entwicklung von Wellenfunktion und Energie nach λ ergibt

$$\Psi = \Psi^{(0)} + \lambda\Psi^{(1)} + \lambda^2\Psi^{(2)} + \dots \quad (2.35)$$

$$E = E^{(0)} + \lambda E^{(1)} + \lambda^2 E^{(2)} + \dots \quad (2.36)$$

Hier sind $\Psi^{(n)}$ und $E^{(n)}$ die Beiträge n -ter Ordnung; beim sog. *MP2*-Verfahren z.B. erfolgt die Störungsentwicklung bis $n = 2$.

Setzt man die Gleichungen (2.32), (2.35) und (2.36) in die elektronische Schrödingergleichung ein, ordnet nach Potenzen von λ und normiert intermediär die Wellenfunktion nach

$$\langle \Psi^{(0)} | \Psi^{(0)} \rangle + \lambda \langle \Psi^{(0)} | \Psi^{(1)} \rangle + \lambda^2 \langle \Psi^{(0)} | \Psi^{(2)} \rangle + \dots = 1, \quad (2.37)$$

so erhält man $\langle \Psi^{(0)} | \Psi^{(n)} \rangle = \delta_{0n}$ und damit

$$\hat{\mathcal{H}}^{(0)} | \Psi^{(0)} \rangle = E^{(0)} | \Psi^{(0)} \rangle \quad (2.38)$$

$$\hat{\mathcal{H}}^{(0)} | \Psi^{(1)} \rangle + \hat{\mathcal{H}}^{(1)} | \Psi^{(0)} \rangle = E^{(0)} | \Psi^{(1)} \rangle + E^{(1)} | \Psi^{(0)} \rangle \quad (2.39)$$

$$\hat{\mathcal{H}}^{(0)} | \Psi^{(2)} \rangle + \hat{\mathcal{H}}^{(1)} | \Psi^{(1)} \rangle = E^{(0)} | \Psi^{(2)} \rangle + E^{(1)} | \Psi^{(1)} \rangle + E^{(2)} | \Psi^{(0)} \rangle \quad (2.40)$$

...

Multiplikation mit $\langle \Psi^{(0)} |$ ergibt

$$E^{(0)} = \langle \Psi^{(0)} | \hat{\mathcal{H}}^{(0)} | \Psi^{(0)} \rangle \quad (2.41)$$

$$E^{(1)} = \langle \Psi^{(0)} | \hat{\mathcal{H}}^{(1)} | \Psi^{(0)} \rangle \quad (2.42)$$

$$E^{(2)} = \langle \Psi^{(0)} | \hat{\mathcal{H}}^{(1)} | \Psi^{(1)} \rangle \quad (2.43)$$

...

Zur Berechnung der Störung erster Ordnung ist also die Korrekturwellenfunktion $\Psi^{(n \geq 1)}$ nicht notwendig, sondern erst bei Energien höherer Ordnung müssen auch die Korrekturwellenfunktionen bekannt sein.

Im Falle der MØLLER–PLESSET-Störungstheorie, in der die Elektronenkorrelation als Störung der HARTREE–FOCK-Wellenfunktion beschrieben wird, ergibt sich für die Energiekorrektur erster Ordnung

$$E_0^{(1)} = \langle \Psi_0^{(0)} | \hat{\mathcal{H}}^{(1)} | \Psi_0^{(0)} \rangle \quad (2.44)$$

$$= \left\langle \Psi^{\text{HF}} \left| \sum_i^N \sum_{j>i}^N \frac{1}{r_{12}} \right| \Psi^{\text{HF}} \right\rangle - \left\langle \Psi^{\text{HF}} \left| \sum_i^N \hat{v}_i^{\text{HF}} \right| \Psi^{\text{HF}} \right\rangle \quad (2.45)$$

$$= \frac{1}{2} \sum_i^N \sum_j^N \left(\left\langle ij \left| \frac{1}{r_{12}} \right| ij \right\rangle - \left\langle ij \left| \frac{1}{r_{12}} \right| ji \right\rangle \right) - \sum_i^N \langle i | \hat{v}^{\text{HF}} | i \rangle \quad (2.46)$$

$$= -\frac{1}{2} \sum_i^N \sum_j^N \left(\left\langle ij \left| \frac{1}{r_{12}} \right| ij \right\rangle - \left\langle ij \left| \frac{1}{r_{12}} \right| ji \right\rangle \right) \quad (2.47)$$

Es ist also

$$E^{\text{HF}} = E_0^{(0)} + E_0^{(1)}, \quad (2.48)$$

was bedeutet, dass die Energiekorrektur erster Ordnung lediglich die Doppelzählung der Elektron–Elektron-Wechselwirkung, die man bei der Aufsummierung der Orbitalenergien erhält, beseitigt, aber noch keine Verbesserung der Beschreibung des Systems gegenüber HARTREE–FOCK bringt. Dafür sind höhere Korrekturen der Energie nötig. Für deren Berechnung müssen aber außer $\Psi_0^{(0)}$ noch Wellenfunktionen angeregter Zustände bekannt sein. Tatsächlich tragen nur

zweifach angeregte Zustände bei, da Einfachanregungen als Folge von Brillouins Theorem wegfallen und höhere als Zweifachanregungen wegen des Zweikörper-Charakters der Störung gar keinen Beitrag liefern. So erhält man

$$E_0^{(2)} = \sum_a^{N_{\text{occ}}} \sum_{b>a}^{N_{\text{occ}}} \sum_r^{N_{\text{virt}}} \sum_{s>r}^{N_{\text{virt}}} \frac{\left| \langle \Psi_0 \left| \sum_i^N \sum_{j>i}^N \frac{1}{r_{12}} \right| \Psi_{ab}^{rs} \rangle \right|^2}{\epsilon_a + \epsilon_b - \epsilon_r - \epsilon_s} \quad (2.49)$$

$$= \sum_a^{N_{\text{occ}}} \sum_{b>a}^{N_{\text{occ}}} \sum_r^{N_{\text{virt}}} \sum_{s>r}^{N_{\text{virt}}} \frac{\left| \langle ab \left| \frac{1}{r_{12}} \right| rs \rangle - \langle ab \left| \frac{1}{r_{12}} \right| sr \rangle \right|^2}{\epsilon_a + \epsilon_b - \epsilon_r - \epsilon_s} \quad (2.50)$$

Die Energiekorrekturen dritter Ordnung und höher lassen sich ähnlich bestimmen.

Eine wichtige Eigenschaft der MØLLER–PLESSET-Störungstheorie ist ihre *Größenkonsistenz*. Das bedeutet, dass die Energie, die sich bei der Berechnung eines Supersystems von zwei nicht-miteinander wechselwirkenden gleichen Molekülen ergibt, gleich sein muss mit der doppelten berechneten Energie eines einzelnen Moleküls. Jedoch sind MØLLER–PLESSET-Energien *keine* oberen Schranken der wahren Werte, wie dies bei variationellen Verfahren der Fall ist. Die Konvergenz der Störungsentwicklung mit der Ordnung ist vielfach oszillatorisch. [86]

2.2.5 DFT – Dichte-Funktional-Theorie

1964 konnten HOHENBERG und KOHN zeigen, dass das externe Potential $v(\mathbf{r})$ eines Systems eindeutig ein Funktional der Elektronendichte $\rho(\mathbf{r})$ ist.

$$\rho(\mathbf{r}) = N \int \cdots \int |\Psi(\mathbf{x})|^2 d\mathbf{x}_1 d\mathbf{x}_2 \cdots d\mathbf{x}_N \quad (2.51)$$

Damit ist auch der Hamilton-Operator ein Funktional der Dichte. Diese fundamentale Erkenntnis ist das *erste Hohenberg–Kohn-Theorem*. [87] Die elektronische Energie kann man somit auch als Funktional der Dichte schreiben:

$$E^{\text{DFT}}[\rho] = T_e[\rho] + V_{\text{ne}}[\rho] + V_{\text{ee}}[\rho] \quad (2.52)$$

Nach dem *zweiten Hohenberg–Kohn-Theorem* [87] kann das Variationsprinzip in der Form $E^{\text{DFT}}[\rho] \geq E_0^{\text{DFT}}[\rho_0]$ für den Grundzustand angewendet werden.

Damit hatten HOHENBERG und KOHN bewiesen, dass es möglich ist, die Grundzustandsenergie eines Systems über die Elektronendichte zu berechnen. Jedoch sind aus den Theoremen keine Aussagen über die Form der Funktionale abzuleiten. Erst die Arbeiten von KOHN und SHAM [88] ein Jahr später ermöglichten die Anwendung der DFT-Methode für chemische Systeme.

KOHN und SHAM setzen an dem Problem an, dass die kinetische Energie bislang nur unzureichend behandelt werden konnte, da eben exakte Funktionale unbekannt waren. Sie formulierten die kinetische Energie als Funktional der Dichte $T_s[\rho]$ unter der Annahme von *nicht-wechselwirkenden* Elektronen (die es natürlich nicht gibt) in den Orbitalen $\psi_i(\mathbf{r})$.

$$T_s[\rho] = -\frac{1}{2} \sum_i^N \langle \psi_i | \nabla^2 | \psi_i \rangle \quad (2.53)$$

Die Dichte $\rho(\mathbf{r})$ ist in diesem Fall

$$\rho(\mathbf{r}) = \sum_i^N |\psi_i(\mathbf{r})|^2 \quad (2.54)$$

Das Funktional der Gesamtenergie $E_s^0[\rho]$ für dieses hypothetische System aus nicht-wechselwirkenden Elektronen ist die Summe aus der kinetischen Energie $T_s[\rho]$, der Elektron–Kern-Wechselwirkung $E_{ne}[\rho]$ und der Coulomb-Abstoßung $E_{ee}[\rho]$:

$$E_s^0[\rho] = \underbrace{-\frac{1}{2} \sum_i^N \langle \psi_i | \nabla^2 | \psi_i \rangle}_{T_s[\rho]} + \underbrace{\sum_a^M \int -\frac{Z_a \rho(\mathbf{r})}{r_A} d\mathbf{r}}_{E_{ne}[\rho]} + \underbrace{\frac{1}{2} \int \int \frac{\rho(\mathbf{r}_1) \rho(\mathbf{r}_2)}{r_{12}} d\mathbf{r}_1 d\mathbf{r}_2}_{E_{ee}[\rho]} \quad (2.55)$$

Für das hypothetische System nicht-wechselwirkender Elektronen ist das *exakt*. KOHN und SHAM wendeten diese Gleichung aber auf reale Systeme mit Elektronen an, die sehr wohl miteinander in Wechselwirkung sind, führten dafür aber einen Term $E_{xc}[\rho]$ ein, der die *Abweichung* von ihrem Modellsystem *ohne* Wechselwirkung zu dem realen System *mit* Wechselwirkung beschreibt.

$$E_0^{\text{DFT}}[\rho] = T_s[\rho] + E_{\text{ne}}[\rho] + E_{\text{ee}}[\rho] + E_{\text{xc}}[\rho] \quad (2.56)$$

Alle nicht-klassischen Effekte sind in dem *Austausch-Korrelations-Energiefunktional* $E_{\text{xc}}[\rho]$ enthalten: die Austausch-Energie als Folge des Pauli-Prinzips und die Elektronenkorrelationsenergie aufgrund der elektrostatischen Abstoßung der Elektronen untereinander. Das *Austausch-Korrelations-Potential* erhält man durch Funktionalableitung von $E_{\text{xc}}[\rho]$:

$$v_{\text{xc}}(\mathbf{r}) = \frac{\delta E_{\text{xc}}[\rho]}{\delta \rho(\mathbf{r})} \quad (2.57)$$

Die *Kohn-Sham-Orbitale* ψ_i erhält man durch Lösen des Variationsproblems, ähnlich wie in der HARTREE-FOCK-Theorie (siehe Kapitel 2.2.3):

$$\left(-\frac{1}{2}\nabla^2 + v(\mathbf{r}) + \int \frac{\rho(\mathbf{r}')}{d_{r,r'}} d\mathbf{r}' + v_{\text{xc}}(\mathbf{r}) \right) \psi_i(\mathbf{r}) = \epsilon_i \psi_i(\mathbf{r}) \quad (2.58)$$

Sie werden wie in den ROOTHAAN-HALL-Gleichungen (2.26) in einem Basissatz entwickelt und die Einteilchengleichungen bis zur Selbstkonsistenz in $\rho(\mathbf{r})$ und ϵ_i iteriert.

2.2.5.1 Austausch-Korrelations-Energiefunktionale

Die Dichtefunktionaltheorie ist – einen vollständigen Basissatz vorausgesetzt – eigentlich *exakt*. Nur kennt man das exakte Austausch-Korrelations-Energiefunktional nicht. Im Laufe der Zeit sind eine Reihe von Ansätzen für dieses Funktional gemacht worden.

Die einfachsten Funktionale sind *local density approximations* (LDA) oder *local spin density approximations* (LSDA)

$$E_x^{\text{LSD(A)}}[\rho] = \int f(\rho(\mathbf{r})) d\mathbf{r}, \quad (2.59)$$

wobei für $f(\rho(r))$ gut bekannte Werte eines gleichverteilten Elektronengas entsprechender Dichte bzw. entsprechender Spindichte eingesetzt werden. Dahinter steckt die Annahme, dass sich die Elektronendichte nur langsam bezüglich

des Raumes ändert und deshalb als homogenes Elektronengas behandelt werden kann. Das Korrelationsfunktional von VOSKO, WILK und NUSAIR E_c^{VWN} [89] ist ein Vertreter dieser Funktionalklasse.

Genauer sind die *generalized gradient approximations* (GGA)

$$E_{\text{xc}}^{\text{GGA}}[\rho] = \int f(\rho(\mathbf{r}), \nabla\rho(\mathbf{r})), \quad (2.60)$$

die durch die Abhängigkeit vom *Dichtegradienten* auch einen Teil der Inhomogenität der Elektronendichte berücksichtigen. Das Austauschfunktional von BECKE E_x^{B} [90] gehört genauso zu dieser Klasse von Funktionalen wie das Korrelationsfunktional von LEE, YANG und PARR E_c^{LYP} . [91] Das BECKE-Austauschfunktional ist von einem empirischen Parameter abhängig, der an Daten von HARTREE–FOCK-Rechnungen an sechs Edelgasatomen angepasst ist.

Durch die HARTREE–FOCK-Theorie kann die nichtlokale Austauschenergie exakt berechnet werden. Diesen Umstand nutzte BECKE [92] für ein Hybridfunktional aus, in welchem exakter HARTREE–FOCK-Austausch und Austausch–Korrelations-Funktionale kombiniert werden:

$$E_{\text{xc}}^{\text{B3LYP}} = a_1 E_x^{\text{LSD}} + (1 - a_1) E_x^{\text{HF}} + a_2 E_x^{\text{B}} + (1 - a_3) E_c^{\text{VWN}} + a_3 E_c^{\text{LYP}} \quad (2.61)$$

wobei a_1 , a_2 und a_3 an experimentelle Daten angepasst sind. Hier wird deutlich, dass das B3LYP–Funktional semiempirischen Charakter hat.

2.2.6 AIM – Atome in Molekülen

Ein Großteil heutiger *ab initio* QM-Rechnungen gilt der (immer besser werden) näherungsweise Lösung der Schrödinger-Gleichung. Obwohl man damit gute Ergebnisse für Energien und Wellenfunktionen erhält, ist der breite chemische Erfahrungsschatz in quantenchemischen Ergebnissen nicht direkt wiederzufinden. Im Verlaufe der Entwicklungen in der Chemie erkannte man, dass sich die Eigenschaften vieler Moleküle in erster Näherung aus charakteristischen atomaren Eigenschaften sowie Bindungseigenschaften zusammensetzen lassen – wobei die angesprochenen atomaren Eigenschaften allerdings in zweiter Näherung wiederum von der molekularen Umgebung abhängen.

Die Theorie der *Atome in Molekülen* (*AIM*) [93, 94, 95, 96] ist das geistige Kind von RICHARD F. W. BADER und soll die Lücke zwischen anerkannten und nützlichen Begriffen wie Atome und Bindungen auf der einen und den Lösungen der Schrödinger-Gleichung auf der anderen Seite schließen. [97].

Wenn isolierte Atome sich zu einem Molekül verbinden, so gehen die meisten Chemiker implizit davon aus, dass diese individuellen Atome mehr oder weniger als solche existent bleiben und durch Wechselwirkungen untereinander einen Verbund bilden – das Molekül. Aber ist das wirklich so? Genauso gut könnte es sein, dass die entsprechenden Atome zu einem homogenen Objekt namens Molekül zusammenfließen und damit vollständig ihre individuelle Identität verlieren. Die Wellenfunktion eines Molekül „weiß“ *a priori* nichts von Untereinheiten, also Atomen.

(*AIM*) benutzt die Einelektronendichte $\rho(\mathbf{r})$ um das Molekül durch mathematische Methoden in mehrere Subräume zu zerlegen. BADER definiert anhand von Eigenschaften der Elektronendichte zwei wichtige Eckpunkte der anschaulichen Chemie: Atome und Bindungen.

Die Topologie von $\rho(\mathbf{r})$ wird typischerweise durch die *Atomkerne* bestimmt (Abbildung 2.7), wodurch der Schluss nahe liegt, dass die Eigenschaften eines Moleküls vollständig von den Eigenschaften der Atome, aus denen es besteht, dominiert werden. In Abbildung 2.7 kann man gut den Peak des Sauerstoffatoms erkennen. Die Peaks der Wasserstoffatome sind etwas flacher, da ihr einziges Elektron an der OH-Bindung beteiligt ist.

Jede topologische Eigenschaft in $\rho(\mathbf{r})$ ist verbunden mit einem *kritischen Punkt* an der Stelle \mathbf{r}_{cp} , so dass gilt

$$\nabla\rho(\mathbf{r}_{cp}) = 0 \tag{2.62}$$

Da die Atomkerne echte Maxima in $\rho(\mathbf{r})$ darstellen, muss sich auf der Verbindungslinien zwischen jeweils zwei benachbarten Maxima ein Sattelpunkt befinden (Abbildung 2.7). Bader betrachtet zwei Kerne als *chemisch verbunden*, wenn sie durch ein Paar von Trajektorien verbunden sind, welches in einem Sattelpunkt zwischen diesen Kerne ihren Ursprung hat und dann entlang eines Pfades maximaler Dichte verläuft und in den Dichtemaxima der Kerne endet. Folglich nennt man diese Sattelpunkte *Bindungskritische Punkte* (*BCPs*). Sie liegen zwischen kovalent gebunden Kernen, aber auch zwischen wasserstoffverbrückten Zentren.

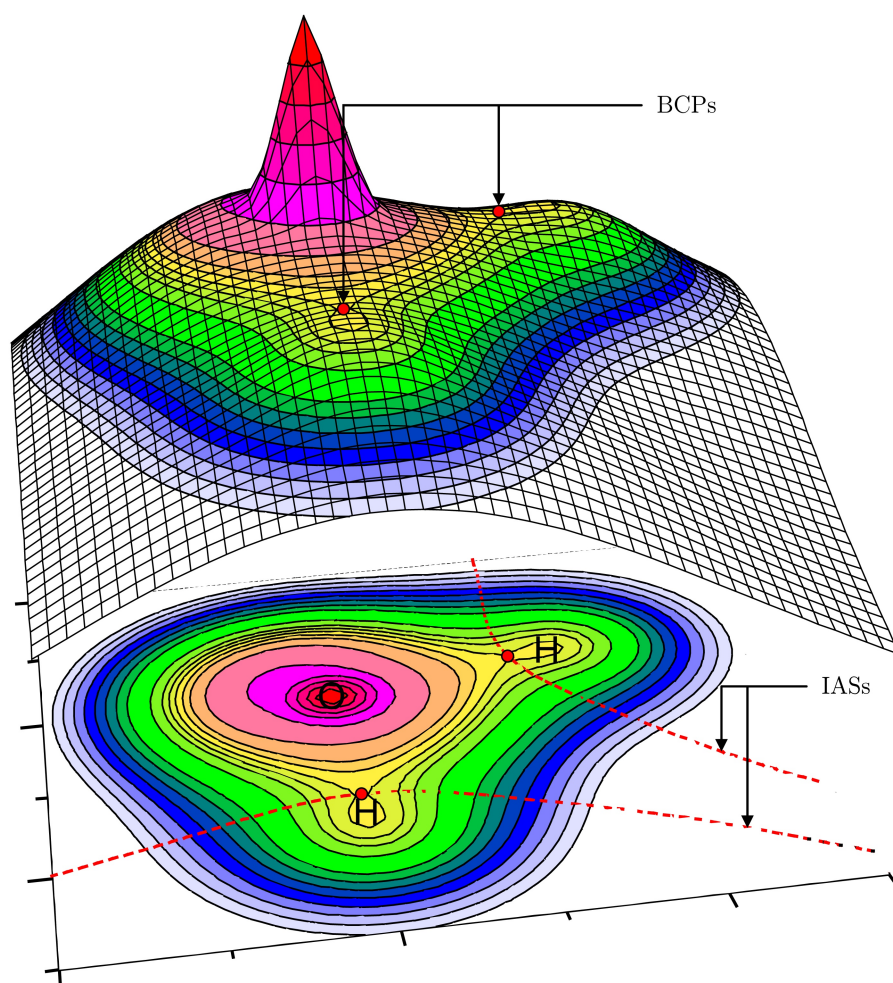


Abbildung 2.7. Logarithmische Elektronendichte von Wasser (Schnitt durch die Molekülebene) mit den bindungskritischen Punkten (*BCPs*) und den interatomaren Separatrizes (*IASs*).

In der *AIM* ist ein atomares Bassin die Menge aller Punkte auf allen Gradientenpfaden in $\rho(\mathbf{r})$, die im selben Maximum (dem Atom des Bassins) enden. Zwei Bassins Ω_a und Ω_b sind durch *interatomare Separatrizes* (*IAS*) S_{ab}^ρ getrennt (Abbildung 2.7). Das sind die Flächen, die niemals von einem Gradientenpfad geschnitten werden:

$$\nabla\rho(\mathbf{r}) \cdot \mathbf{n}_{ab}^\rho(\mathbf{r}) = 0 \quad \forall \mathbf{r} \in S_{ab}^\rho \quad (2.63)$$

mit $\mathbf{n}_{ab}^\rho(\mathbf{r})$ dem Normalenvektor senkrecht zur Ebene S_{ab}^ρ an der Stelle \mathbf{r} .

Der R^3 wird durch die *IASs* also in klar von einander abgegrenzte disjunkte Gebiete unterteilt, die man als die Atome in dem Molekül identifiziert. Somit

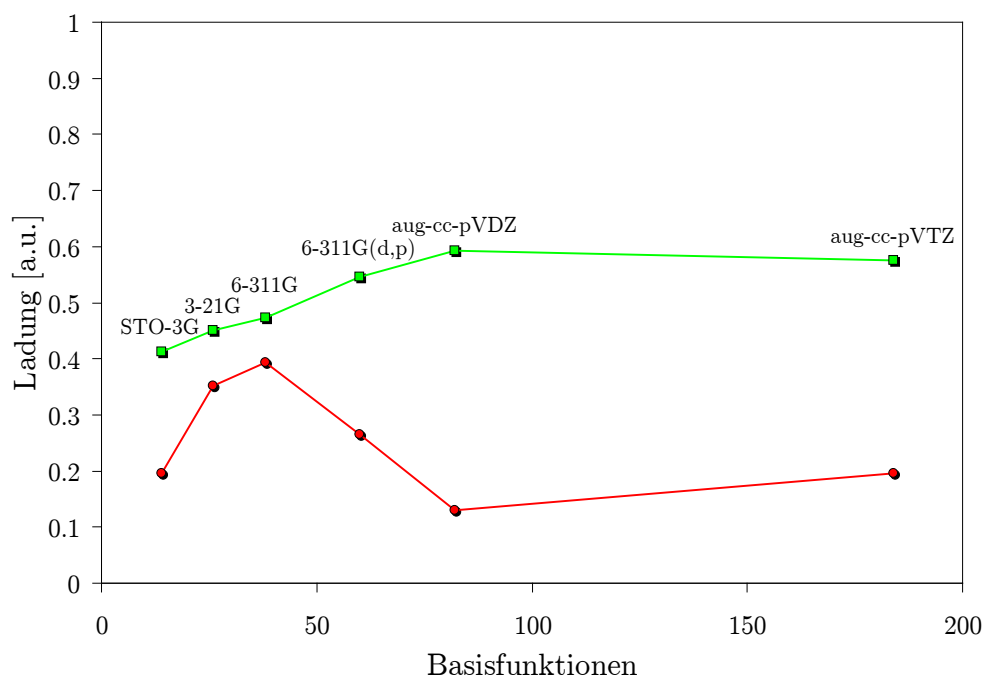


Abbildung 2.8. *AIM*- und *MULLIKEN*-Partialladungen für ein nicht-komplexiertes Wasserstoffatom im Wasserdimer für verschiedene atomare Basissätze.

sind jedem Atom ein definierter Raum und ein definierter Anteil an der Elektronendichte zugeteilt.

Die Integration von $\rho(\mathbf{r})$ innerhalb eines Bassins liefert die mittlere Elektronenpopulation dieses Bassins. Allgemein ist sie gegeben als

$$\bar{N}_s = \bar{N}(\Omega_s) = \int_{\Omega_s} \rho(\mathbf{r}) d\tau. \quad (2.64)$$

Verglichen mit der Anzahl der Elektronen des ungebundenen Atoms ergibt sich die formale Ladung als Differenz der beiden Größen.

So erlaubt *AIM* eine verlässliche Zuweisung von atomaren Partialladungen. Werden für die Berechnung der *AIM*-Partialladungen Einelektronendichten aus quantenchemischen Programmen verwendet, so zeigt sich, dass die *AIM*-Partialladungen erheblich unabhängiger vom gewählten atomaren Basissatz sind als die Ergebnisse einer *MULLIKEN*-Populationsanalyse (Abbildung 2.8). [98] Im Gegensatz zu *MULLIKEN*-Ladungen reproduzieren *AIM*-Partialladungen darüber hinaus andere, genauer definierte molekulare Eigenschaften wie Dipole oder Quadrupole. [99]

Alternativ können atomare Partialladungen auch durch *ESP-Fitting* (*electrostatic potential fitting*) zugewiesen werden. Dabei werden an einer Reihe von Punkten um die Atome herum das elektrische Potential in Abhängigkeit der atomaren Partialladungen berechnet und durch Variation der Partialladungen die Differenz zum elektrischen Potential der Wellenfunktion minimiert. [99] Auch diese atomaren Partialladungen können molekulare Eigenschaften reproduzieren. [100]

2.3 Quantenchemische Rechnungen

Trotz erheblicher Fortschritte in der Bestimmung von Struktureigenschaften wasserstoffverbrückter Systeme, etwa durch NMR-Spektroskopie und Röntgenbeugungsexperimente (oder selten auch durch Neutronenbeugungsexperimente), bleibt es experimentell hoch anspruchsvoll, verlässliche Informationen über einzelnen Wasserstoffbrücken zu sammeln. Von großen Systemen, wie Protein–Ligand-Komplexen, liegen in den meisten Fällen nur Röntgenstrukturdaten vor. Bei dieser Art der Strukturbestimmung können Protonen nur sehr schwer detektiert werden und somit sind auch Informationen über eventuelle Wasserstoffbrücken nicht direkt zugänglich.

Aus diesem Grund haben sich quantenchemische Methoden weithin für detaillierte Untersuchungen von strukturellen und energetischen Eigenschaften von wasserstoffverbrückten Systemen etabliert.

In diesem Teil stelle ich quantenchemische *ab initio*-Rechnungen vor, mit denen ich die Wechselwirkungsenergie eines Satzes aus 81 wasserstoffverbrückten Komplexen aus kleinen organischen Molekülen (siehe Kapitel 2.3.1) berechnet habe. Aus diesen Wechselwirkungsenergien leite ich dann die Inkremente für verschiedene Wasserstoffbrückendonoren und -akzeptoren ab. Das ganze Inkrementsystem heißt im Folgenden HBINC. Die Theorien hinter den verwendeten quantenchemischen Methoden wurden bereits erläutert (Kapitel 2.2), doch gehe ich in Kapitel 2.3.2 kurz erneut auf Eigenschaften ein, die die Anwendbarkeit der verschiedenen Methoden bestimmen.

Die Ergebnisse und deren Diskussion werden in Kapitel 2.4 vorgestellt. Sie sind in [98] veröffentlicht.

2.3.1 Der Molekülsatz

Für die Zusammenstellung eines Satzes von kleinen, wasserstoffverbrückten Systemen habe ich einen *kombinatorischen Ansatz* gewählt, der im Folgenden erläutert wird. Die Ergebnisse der quantenchemischen Rechnungen sollen in einem anschließenden Schritt zur Verbesserung der FLEXX-Bewertungsfunktion genutzt werden. Da FLEXX intern Atome als SYBYL-Atomtypen verwaltet, erschien es mir sinnvoll, von Anfang an auch in der Systematik dieser Atomtypen zu bleiben, damit der Transfer nachher reibungsloser geschehen kann.

Typische Atome, die als Wasserstoffbrückendonor und auch als -akzeptor agieren können sind Sauerstoff, Stickstoff und Schwefel (siehe auch Kapitel 2.1). Ich habe zunächst von jedem dieser drei Atome alle möglichen SYBYL-Atomtypen als

Startpunkt für den kombinatorischen Ansatz gewählt. Die unterschiedlichen SYBYL-Typen für ein Atom repräsentieren unterschiedliche Hybridisierungen oder besondere chemische Umgebungen, wie z.B. ein Stickstoff in einer Amidgruppe. Eine Übersicht über die SYBYL-Atomtypen findet sich in Tabelle 2.1.

Ich habe danach an jede freie Valenz der Heteroatome andere passend hybridisierte Atome gesetzt. Führte das zu einem chemisch sinnvollen molekularen Fragment, habe ich ein möglichst kleines Molekül gesucht, das dieses molekulare Fragment enthält. Die so entstehenden Baumstrukturen sind in den Abbildungen 2.9, 2.10, 2.11, 2.12 und 2.13 dargestellt.

Ein wichtiger Schwerpunkt war die Untersuchung von jeweils gleichen Donoren oder gleichen Akzeptoren in unterschiedlicher chemischer Umgebung. Deshalb habe ich zudem Moleküle ausgewählt, in denen eine besonders wichtige Gruppe an möglichst verschiedene SYBYL-Atomen gebunden ist. So entstehen z.B. Reihen wie Ethanol, Ethenol, Ethinol oder primäres, sekundären und tertiäres Amin.

Insgesamt habe ich bevorzugt funktionelle Gruppen, die in biologisch und pharmakologisch relevanten Systemen häufig vorkommen, ausgewählt. Darunter sind Alkohole, Carbonsäuren, Ester, Phosphate, Sulfate, Amine, Nitroverbindungen, Mercaptane, Disulfide, usw. Als Modell für die wichtigen Peptide ist *N*-Methylacetamid im Molekülsatz vertreten. Zusätzlich habe ich die hydrophilen Seitenketten von Aminosäuren mit aufgenommen. Dazu gehören Pyrrolidin (Prolin), Imidazol (Histidin) und Indol (Tryptophan). Die Seitenketten von Serin, Cystein, Methionin, Tyrosin, Arginin, Lysin, Asparagin, Glutamin, Asparaginsäure und Glutaminsäure sind bereits durch die Auswahl an kleinen organischen Komponenten abgedeckt.

Tabelle 2.1. SYBYL-Atomtypen

SYBYL	Beschreibung	SYBYL	Beschreibung
C.3	sp ³ -Kohlenstoff	S.3	sp ³ -Schwefel
C.2	sp ² -Kohlenstoff	S.2	sp ² -Schwefel
C.1	sp-Kohlenstoff	S.O	Schwefel in Sulfoxiden
C.ar	aromatischer Kohlenstoff	S.O2	Schwefel in Sulfonen
N.3	sp ³ -Stickstoff	P.3	sp ³ -Phosphor
N.2	sp ² -Stickstoff	F	Fluor
N.1	sp-Stickstoff	Cl	Chlor
N.ar	aromatischer Stickstoff	Br	Brom
N.am	amidischer Stickstoff	I	Iod
N.pl3	trigonal-planarer Stickstoff	H	Wasserstoff
N.4	positiv geladener sp ³ -Stickstoff		
O.3	sp ³ -Sauerstoff	Na	Natrium
O.2	sp ² -Sauerstoff	K	Kalium
O.co2	Carboxylat-Sauerstoff	Ca	Calcium
		Li	Lithium

Aus diesen Molekülen habe ich dann ein Trainingssatz mit 81 wasserstoffverbrückten Komplexen gebaut, wobei ich immer nur einen Donor und einen Akzeptor pro Komplex benutzt habe. Die Komponenten sind dabei nicht notwendigerweise in ihrer energetisch niedrigsten Konformation. Wenn möglich, wurden energetisch günstige Konformere gewählt, die nur eine einzige Wasserstoffbrücke ausbilden. Das war nicht in jedem Komplex möglich, so dass auch Systeme mit mehreren Wasserstoffbrücken vertreten sind (siehe auch Abschnitt 2.4.2.2).

Es wurden nicht alle möglichen Donor–Akzeptor-Kombinationen aus allen Molekülen gebildet. Da Wasser sowohl der wichtigste Wasserstoffbrückendonator und auch -akzeptor ist, wurden zunächst alle Moleküle mit Wasser (als Donor und/oder Akzeptor) komplexiert. Danach wurden noch besonders repräsentative Kombinationen zusätzlich gebildet. Alle Moleküle und ihre Komplexe im Komplexsatz sind in den Tabellen 2.4, 2.5 und 2.6 zu finden.

Für die Validierung des aus dem Trainingssatz abgeleiteten Inkrementsystems HBINC wurde ein Testsatz zusammengestellt. Er enthält 11 wasserstoffverbrückte Komplexe und beinhaltet neue Donor–Akzeptor-Kombinationen von Molekülen aus dem Trainingsset, aber auch Komplexe von größeren Molekülen, die vorher nicht berücksichtigt wurden, z.B. der Ringzucker Fucose und die Nukleinbase Adenin. Zusätzlich wurden DNA-Basenpaare und chemisch ähnliche, zyklische wasserstoffverbrückte System aus dem Hobza-Satz [101] mit aufgenommen. Die Komplexe des Testsatzes sind in Tabelle 2.7 zusammengestellt.

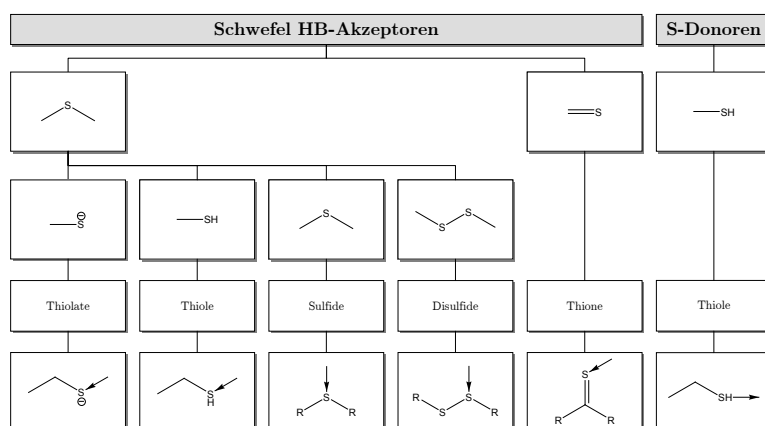


Abbildung 2.9. Hierarchischer Molekülfragment-Baum der Wasserstoffbrückendonoren und -akzeptoren, die Schwefel als Akzeptor oder Donor enthalten.

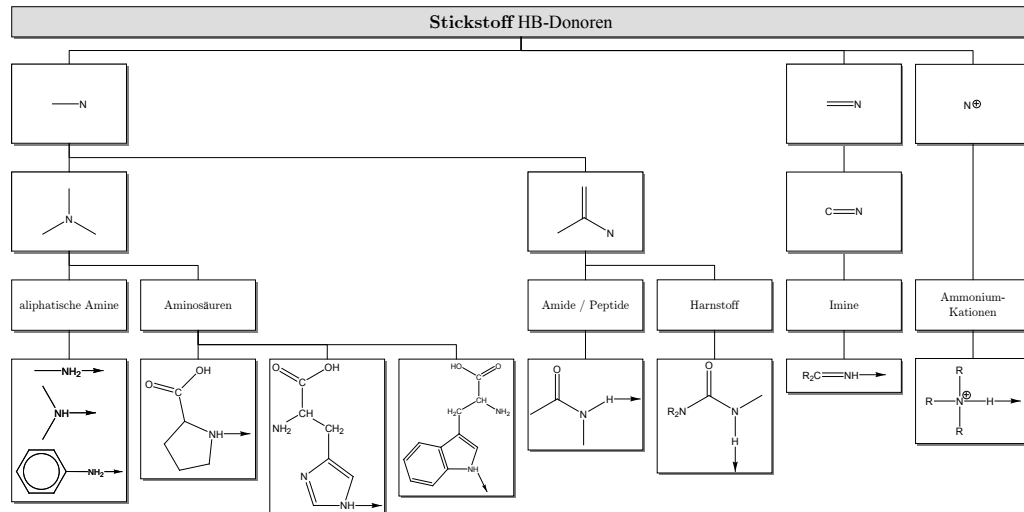


Abbildung 2.10. Hierarchischer Molekülfragment-Baum der Wasserstoffbrückendonoren, die Stickstoff als Donor enthalten.

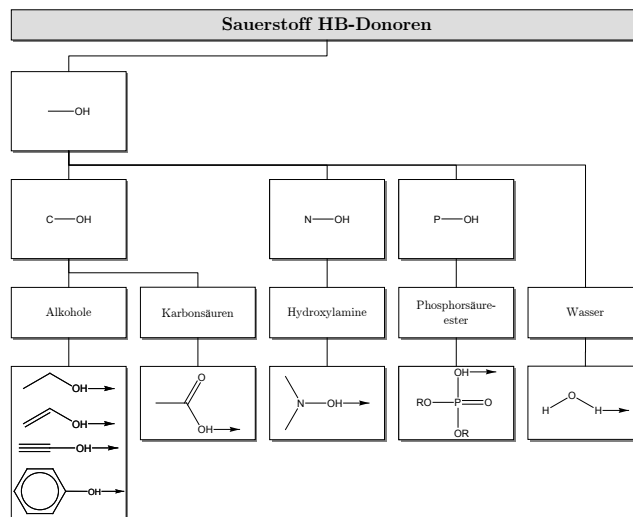


Abbildung 2.11. Hierarchischer Molekülfragment-Baum der Wasserstoffbrückendonoren, die Sauerstoff als Donor enthalten.

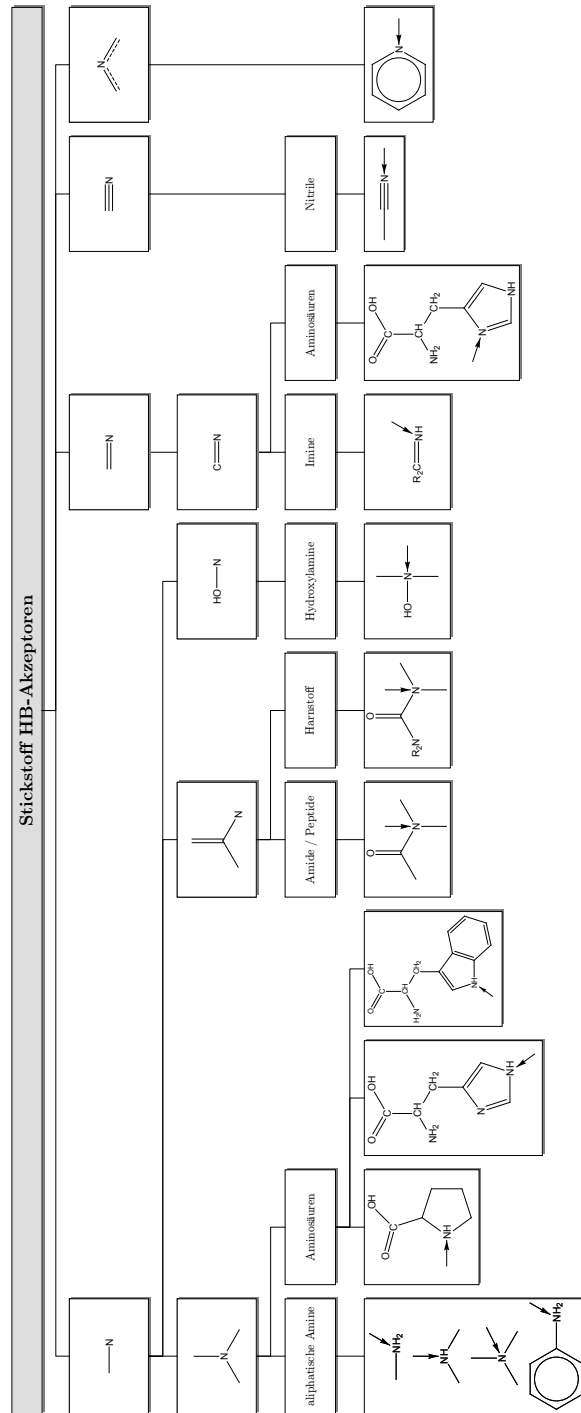


Abbildung 2.12. Hierarchischer Molekülfragment-Baum der Wasserstoffbrückenakzeptoren, die Stickstoff als Akzeptor enthalten.

Sauerstoff HB-Akzeptoren

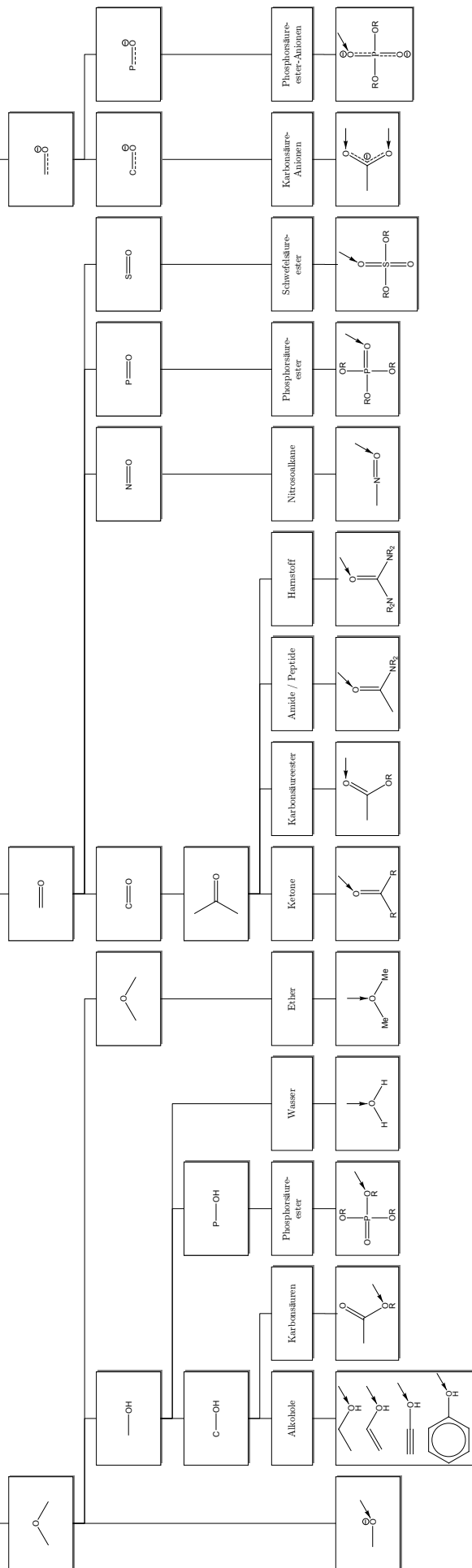


Abbildung 2.13. Hierarchischer Molekülfragment-Baum der Wasserstoffbrückenakzeptoren, die Sauerstoff als Akzeptor enthalten.

2.3.2 Quantenchemische Methoden

Wie in Kapitel 2.1 bereits dargelegt, spielen bei Wasserstoffbrücken neben elektrostatischen Wechselwirkungen auch dispersive Kräfte, Induktion und Ladungstransfer-Wechselwirkungen eine wichtige Rolle. [58] In *schwachen Wasserstoffbrücken* können Induktion und Dispersionskräfte sogar die bestimmenden Wechselwirkungen werden. [60]

Dispersion ist ein Effekt der *Elektronenkorrelation*. Deshalb sind für quantenmechanische Rechnungen an wasserstoffverbrückten Systemen hochentwickelte Methoden zur Behandlung der Elektronenkorrelation unverzichtbar, um diese Wechselwirkung mit der notwendigen Genauigkeit erfassen zu können. Auch sind große und ausgedehnte atomare Basissätze erforderlich, da Wasserstoffbrücken *langreichweitige* Wechselwirkungen sind.

DFT-Rechnungen (siehe Kapitel 2.2.5) mit dem *B3LYP*-Austausch-Korrelationsfunktional (Gleichung (2.61), Abschnitt 2.2.5.1) können die *Geometrien* von wasserstoffverbrückten Komplexen gut vorhersagen. [102, 103, 104, 105, 106, 107] Jedoch ist *B3LYP* genau wie andere aktuell verbreitete Austausch-Korrelationsfunktionale nicht in der Lage, die langreichweitigen Dispersionswechselwirkungen richtig abzubilden. [108] Trotz verlässlicher Geometrie wird deshalb die Wechselwirkungsenergie eines wasserstoffverbrückten Komplexes *systematisch unterschätzt* – gerade bei schwachen Wasserstoffbrücken, bei denen die Dispersion das bestimmende Element werden kann.

Zurzeit gelten *Coupled Cluster*-Rechnungen mit Einfach- und Zweifachanregungen und ergänzt durch nicht-iterative Korrekturen für die Dreifachanregungen (*CCSD(T)*) [109] als die beste quantenchemische Methode zur verlässlichen Berechnung von Dispersionsenergien. Jedoch sind *CCSD(T)*-Rechnungen selbst für kleine Systeme mit einem enormen Aufwand an Rechenzeit verbunden, so dass sie für solche Serienrechnungen, wie hier vorgestellt, ungeeignet sind.

MØLLER–PLESSET-Störungsrechnungen (siehe Kapitel 2.2.4) sind in der Lage, Elektrostatik, Induktion und Dispersion zu handhaben. TSUZUKI konnte sogar zeigen, dass MØLLER–PLESSET-Rechnungen zweiter Ordnung (*MP2*) für H-Brückenenergien Ergebnisse liefern, die nahe an denen der hochpräzisen *CCSD(T)*-Rechnungen liegen. [110] *MP2*-Rechnungen erfordern weniger Computerressourcen als *CCSD(T)*-Rechnungen, aber immer noch weit mehr als DFT-Rechnungen.

Da die schnellen DFT-Rechnungen die Geometrie eines wasserstoffverbrückten Komplexes verlässlich berechnen können, wurden die Geometrien aller 81

Komplexe des Trainingsatzes auf *B3LYP*-Niveau optimiert. An diesen DFT-Geometrien wurden dann *MP2*-Wechselwirkungsenergien berechnet. So konnte eine gute Balance zwischen Genauigkeit und Anforderungen an die Computerressourcen erreicht werden.

Technische Details

Die Geometrie jedes wasserstoffverbrückten Komplexes wurde mit dem Programmpaket TURBOMOLE [111, 112] optimiert. Die Optimierung wurde auf DFT-Niveau [113, 114] unter Verwendung des *B3LYP*-Austausch-Korrelationsfunktional [92, 91] durchgeführt. Jedem nicht-Wasserstoffatom wurde eine *Triple Zeta-Basis* mit einer *Polarisationsfunktion* (TZVP) [115] zugewiesen, jedem Wasserstoffatom eine *Split Valence-Basis* (SP) [116]. Für Schwefelatome war eine TZVPP-Basis [115] notwendig. Um sicher zu sein, dass es sich jeweils bei der gefundenen Geometrie um ein Minimum handelt, wurden die zweiten Ableitungen berechnet und untersucht.

Zur Berechnung der Wechselwirkungsenergie wurden *aug-cc-pVTZ*-Basen [117, 118] an *alle* Atome gesetzt. An der DFT-optimierten Geometrie wurden einzelne *MP2*-Rechnungen mit den Programmen MOLPRO [119] und TURBOMOLE [120, 121] durchgeführt.

Die Wasserstoffbrückenenergie wurde angesetzt als die Differenz zwischen der Gesamtenergie des relaxierten Komplexes $A \cdots B$ und den Energien der individuellen Komponenten A und B in den Geometrien, die sie im Komplex innehaben:

$$\Delta E(A \cdots B) = E(A \cdots B) - E(A) - E(B) \quad (2.65)$$

Dieser Ansatz ist jedoch anfällig für Basissatz-Superpositions-Fehler (*BSSE*), wenn große, aber eben unvollständige Basissätze benutzt werden. Der *BSSE* führt zu einer zusätzlichen Stabilisierung des Komplexes durch die größere Anzahl der Basisfunktionen im Vergleich zu den Monomeren. Um diese zu korrigieren, wurde die *Counterpoise-Korrektur* von BOYS und BERNARDI angewendet. [122] Dabei werden die Energien der Monomere in der Basis des Komplexes berechnet, um bei jeder Energieberechnung die gleiche Anzahl an Freiheitsgraden durch die Basisfunktionen zu haben. Um Rechenzeit zu sparen, habe ich die *BSSE*-Korrektur nur auf die Bindungsenergie des Komplexes angewendet, nicht auf seine Geometrie. Die Geometrien sowohl des Komplexes als auch der einzelnen Komponenten wurden von der DFT-Optimierung übernommen.

An den DFT-optimierten Geometrien wurden mit meinem selbst entwickelten Programm XELF2 [123, 124] die bindungskritischen Punkte (*BCPs*, siehe Kapitel 2.2.6) bestimmt. *BCPs* zwischen nicht-kovalent gebundenen Atomen sind ein Hinweis auf eine Wasserstoffbrücke zwischen diesen beiden Atomen, gerade bei Komplexen, bei denen die Lage der H-Brücken nicht intuitiv erkennbar ist. Die Anzahl und Lage der H-Brücken in den Komplexen kann somit quantitativ bestimmt werden. Außerdem wurden die atomaren Partialladungen durch Integration der Dichte innerhalb der atomaren *AIM*-Subräume berechnet.

XELF2 benötigt als Input lediglich die Wellenfunktion des Komplexes im *wfn*-Format, wie es die Softwaresuite GAUSSIAN [125] exportiert. Die Lage der *BCPs* und die *AIM*-Partialladungen sind weitgehend unabhängig vom verwendeten Basissatz, weshalb, wiederum um Rechenzeit zu sparen, *rB3LYP*-Rechnungen mit der *STO-3G*-Basis für die Berechnung der Wellenfunktion durchgeführt wurden.

2.4 Quantenchemische Ergebnisse

Im Folgenden werden die Ergebnisse der quantenchemischen *ab initio*-Rechnungen vorgestellt. Das Hauptaugenmerk gilt den *MP2*-Wechselwirkungsenergien von Wasserstoffbrücken in unterschiedlichen chemischen Kontexten. Als erstes werden mögliche Korrelationen zwischen Eigenschaften der Komplexe und deren Wechselwirkungsenergie diskutiert. Anschließend erläutere ich, wie ich aus diesen Energien das Inkrementsystem HBINC zur schnellen Berechnung von Wasserstoffbrückenstärken ableite.

Die Ergebnisse der quantenchemischen Rechnungen für die 81 wasserstoffverbrückten Komplexe des Trainingsatzes sind in den Tabellen 2.4, 2.5 und 2.6 aufgeführt. Zusätzlich zu der Komplexbindungsenergie sind noch verschiedene andere Eigenschaften wie Wasserstoffbrückenlängen und Partialladungen gezeigt. Die Ergebnisse für den Testsatz finden sich in Tabelle 2.7.

In den Tabellen und im Text wird für die wasserstoffverbrückten Komplexe folgende Notation verwendet: Ist nur eine einzige Wasserstoffbrücke im Komplex vorhanden, ist die Notation *Donor*···*Akzeptor*. Sollten mehr als eine Wasserstoffbrücke gefunden worden sein, zeigt ein kleiner Pfeil von dem Teil des Komplexes, der als Donor agiert, zu dem anderen Teil, welcher dann als Akzeptor agiert.

2.4.1 DFT- und MP2-Geometrien

Um den Fehler abzuschätzen, der dadurch entsteht, dass die Geometrie der Komplexe auf DFT-Niveau optimiert wurde (und nicht wie die Energieberechnung mit *MP2*), habe ich fünf wasserstoffverbrückte Komplexe zum Vergleich auf DFT- und auf *MP2*-Niveau geometrieoptimiert, wobei ich jeweils die gleichen Basissätze verwendet habe, wie für die DFT-Optimierung (siehe Kapitel 2.3.2).

Tabelle 2.2. Vergleich der *MP2*-Wechselwirkungsenergie $\Delta E_{\text{MP2}}^{\text{CPC}}$ (in kJ Mol^{-1}) und der Wasserstoffbrückenlänge $d_{\text{H-B}}$ (in Å) für DFT-optimierte Geometrien und für Geometrien die auf *MP2*-Level optimiert wurden.

Komplex	DFT Geometrie		<i>MP2</i> Geometrie		$\Delta d_{\text{H-B}}^{\text{MP2-DFT}}$		$\Delta \Delta E_{\text{MP2-DFT}}$	
	$d_{\text{H-B}}$	$\Delta E_{\text{MP2}}^{\text{CPC}}$	$d_{\text{H-B}}$	$\Delta E_{\text{MP2}}^{\text{CPC}}$	absolut	relativ	absolut	relativ
PhOH···H ₂ O	1,81	-28,48	1,83	-28,61	0,02	%1,20	-0,14	0,50
MeCOOH···MeOH	1,71	-50,13	1,73	-48,50	0,01	%0,80	1,62	3,20
←	1,99		2,09		0,10	5,20		
H ₂ O···Imidazol	1,89	-32,65	1,93	-32,31	0,05	2,40	0,34	1,10
MeNH ₂ ···Imidazol	2,31	-22,19	2,26	-23,07	-0,05	2,00	-0,88	4,00
←	2,67		2,58		-0,09	3,30		
MeOH···MeSMe	2,37	-23,54	2,39	-24,44	0,01	0,60	-0,90	3,80

An beiden Sätzen von Komplexen wurde wie bereits beschrieben die *MP2*-Wechselwirkungsenergie berechnet.

Die resultierenden Wasserstoffbrückenlängen $d_{\text{H-B}}$ und die *counterpoise*-korrigierten *MP2*-Wechselwirkungsenergien $\Delta E_{\text{MP2}}^{\text{CPC}}$ sind in Tabelle 2.2 aufgelistet.

Im Falle von starken H-Brücken, mit Sauerstoff oder Stickstoff als Protonendonator oder -akzeptor, ist die Abweichung in der H-Brückenbindungslänge in der Größenordnung von 2%. Schwächere und damit auch längere Wasserstoffbrücken, bei denen Schwefel der Akzeptor ist, oder eine CH-Gruppe der Donor, werden von der einfacheren DFT-Methode mit einer Abweichung von bis zu 5% gegenüber der *MP2*-optimierten Geometrie reproduziert. Wie bereits mehrfach erwähnt, spielen für die Geometrie dieser Wasserstoffbrücken Dispersionswechselwirkungen, die mit dem *B3LYP*-Funktional nicht richtig erfasst werden, eine große Rolle.

Die *MP2*-Wechselwirkungsenergien der DFT-Geometrien und der *MP2*-Geometrien unterscheiden sich jeweils um weniger als 2 kJ Mol⁻¹.

Der Ansatz, die Geometrie der Komplexe mit der weniger aufwendigen DFT-Methode zu optimieren und dann an dieser Geometrie aufwendigere *MP2*-Energieberechnungen durchzuführen, erscheint im Lichte dieser geringen Abweichungen und der massiv geringeren Computerleistung, die für die DFT-Rechnungen nötig sind, als gerechtfertigt.

2.4.2 Energetische und strukturelle Eigenschaften des Trainingsatzes

Betrachtet man die *MP2*-Wechselwirkungsenergien aller Komplexe, ist ein augenfälliges Ergebnis der weite Bereich, in dem die Wasserstoffbrücken-Bindungsenergien für Donor-Akzeptor-Paare mit jeweils gegebenen SYBYL-Atomtypen liegen. Betrachten wir z.B. alle Komplexe, in denen eine HO.3-Donorgruppe eine Wasserstoffbrücke mit einem O.2-Akzeptoratom bildet (Tabelle 2.4), so liegen die Wechselwirkungsenergien für diese Komplexe in einem Bereich von -15 kJ Mol⁻¹ bis zu -35 kJ Mol⁻¹. Keine dieser Wasserstoffbrücken ist geladen oder resonanzunterstützt, so dass konventionelle Bewertungsfunktionen die Bindungsenergien all dieser Komplexe auf den gleichen Wert geschätzt hätten.

Ein weiteres erwähnenswertes Ergebnis ist die hohe Akzeptorstärke von Schwefel (Tabelle 2.6), welche vergleichbar zu der von Sauerstoff ist. Diese Wechselwirkung wird von FLEXSCORE gar nicht berücksichtigt. Die Donorstärke von Schwefel ist hingegen sehr gering.

Genauso wenig wie Schwefel behandelt FLEXSCORE CH–azide Gruppen als Protonendonoren. Doch wie sich später zeigen wird, ist der Beitrag, den diese Gruppen zu der Gesamtwechselwirkungsenergie beitragen, alles andere als gering.

2.4.2.1 Korrelationen der Wasserstoffbrückenenergien

Elektrostatik

Ein erster offensichtlicher Schritt, um eine Parametrisierung für Wasserstoffbrückenstärken zu finden, ist, nach Korrelationen mit *elektrostatischen* Wechselwirkungen zu suchen (siehe auch Abschnitt 2.1).

Aus den *AIM*-Rechnungen sind die partiellen Ladungen von allen an der Wasserstoffbrücke beteiligten Atomen im Komplex bekannt. Trägt man jedoch die Komplexbindungsenergie und die partiellen *AIM*-Ladungen von Donoren und Akzeptoren gegeneinander auf, so ergibt sich keine erfolgreiche Korrelation (Abbildung 2.14).

Die *elektrostatische Anziehung* zwischen dem Proton des Donors und dem Akzeptor kann zu

$$\Delta E_{\text{pClmb}}(\text{XH} \cdots \text{B}) = \frac{q^{\text{AIM}}(\text{H})q^{\text{AIM}}(\text{B})}{d_{\text{HB}}} \quad (2.66)$$

abgeschätzt werden. Hier sind $q^{\text{AIM}}(\text{H})$ und $q^{\text{AIM}}(\text{B})$ jeweils die partiellen *AIM*-Ladungen des Protons und des Akzeptoratoms. Doch wie Abbildung 2.15 zeigt, kann auch hier keine signifikante Korrelation gefunden werden. Gute Gegenbeispiele sind Wasserstoffbrücken mit Schwefel als Akzeptor. Die Schwefelatome haben in diesen Komplexen eine so gut wie neutrale Partiaalladung. Somit ist die elektrostatische Wechselwirkung nahe Null. Jedoch ergibt z.B. die Komplexbildung von Dimethylsulfid mit Essigsäure eine starke Wasserstoffbrücke mit einer Wechselwirkungsenergie von -28 kJ Mol^{-1} .

Genau umgekehrt verhält es sich bei dem Komplex aus Wasser und 3,3-Dimethyl-but-1-yn-1-ol. Die Partiaalladung am Wasserstoff der Donorgruppe beläuft sich auf 0,59 atomare Einheiten, der alkoholische Sauerstoff trägt eine Partiaalladung von $-1,16$. Doch trotz dieser starken elektrostatischen Anziehung erweist sich die Wasserstoffbrücke als eher schwach (-13 kJ Mol^{-1}).

Diese Beispiele zeigen, dass *AIM*-Partiaalladungen oder damit verwandte Eigenschaften nicht allein als ein Maß für die Stärke einer Wasserstoffbrückenbin-

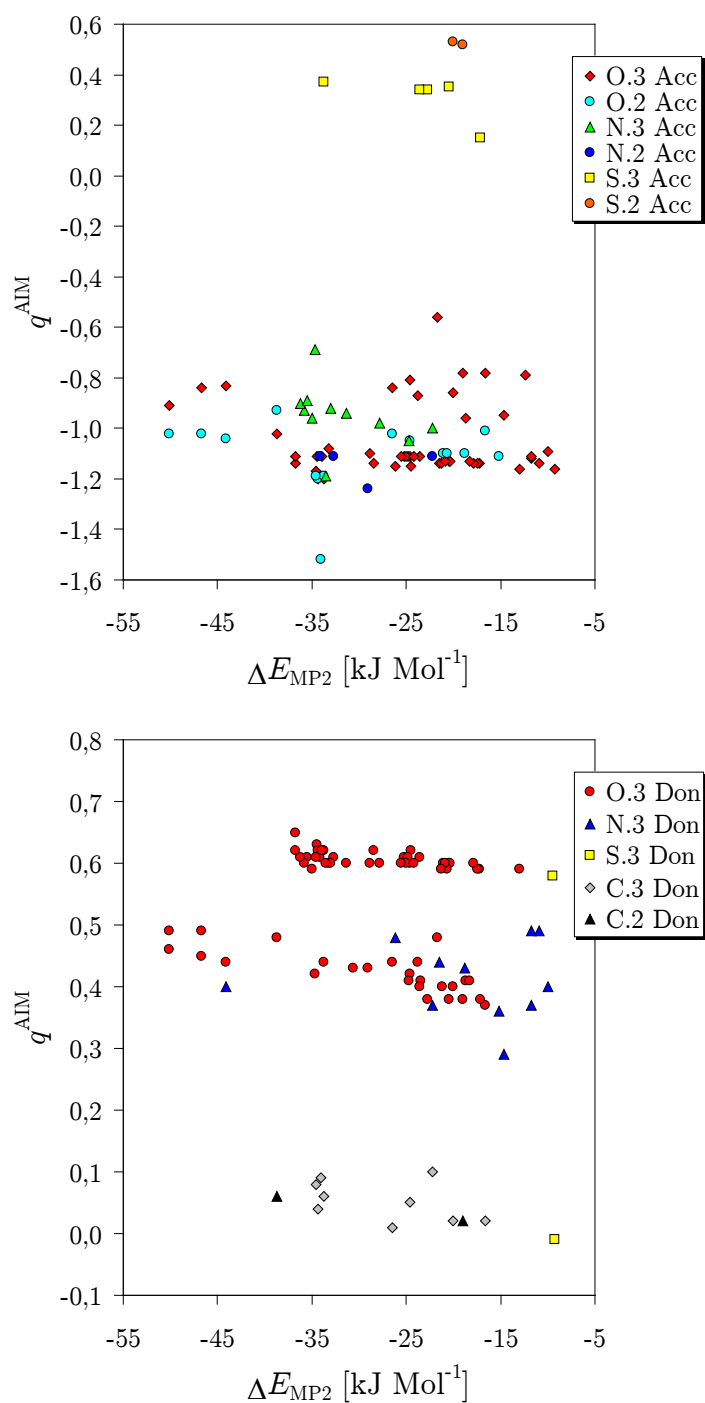


Abbildung 2.14. AIM-Partialladungen der Akzeptoratome (*oben*) und der Protonen in den Donorgruppen (*unten*) in Abhängigkeit der Wechselwirkungsenergie des entsprechenden Komplexes.

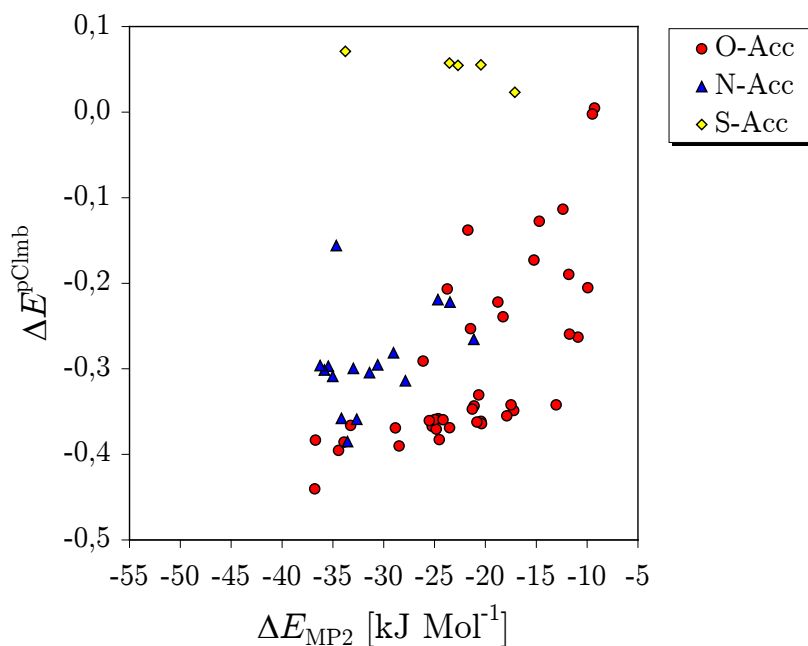


Abbildung 2.15. Korrelation zwischen der geschätzten Coulomb-Wechselwirkung und der $MP2$ -Wechselwirkungsenergie.

dung dienen können. Vielmehr scheint es, dass die einzelnen Beiträge von elektrostatischer Wechselwirkung, Dispersion, Induktion und Ladungstransfer bei einer H-Brücke zu stark miteinander verflochten sind, als dass man einen einfacher Ansatz zur Aufstellung einer empirischen Bewertungsfunktion finden könnte.

Bindungslänge

Wie in Abbildung 2.16 gezeigt, korreliert die Wechselwirkungsenergie mit der Länge der Wasserstoffbrücke für einen gegebenen Akzeptor. Starke Wasserstoffbrücken sind bekanntermaßen kürzer als schwache. Trotzdem überraschen die hohen linearen Korrelation für Sauerstoff ($r^2 = 0,8$), Stickstoff ($r^2 = 0,90$) und Schwefel ($r^2 = 0,94$).

Obwohl diese Korrelation bemerkenswert ist, hilft sie nicht, die Stärke einer Wasserstoffbrücke *a priori* zu bestimmen, da der Abstand zwischen dem Proton des Donors und dem Akzeptoratom nicht im Voraus bekannt ist.

2.4.2.2 Donor- und Akzeptor Energien

Anstatt die zuvor beschriebenen physikalisch motivierten Terme einfach aufzaddieren, führe ich hier das Inkrementensystem HBINC ein. Hiernach werden in

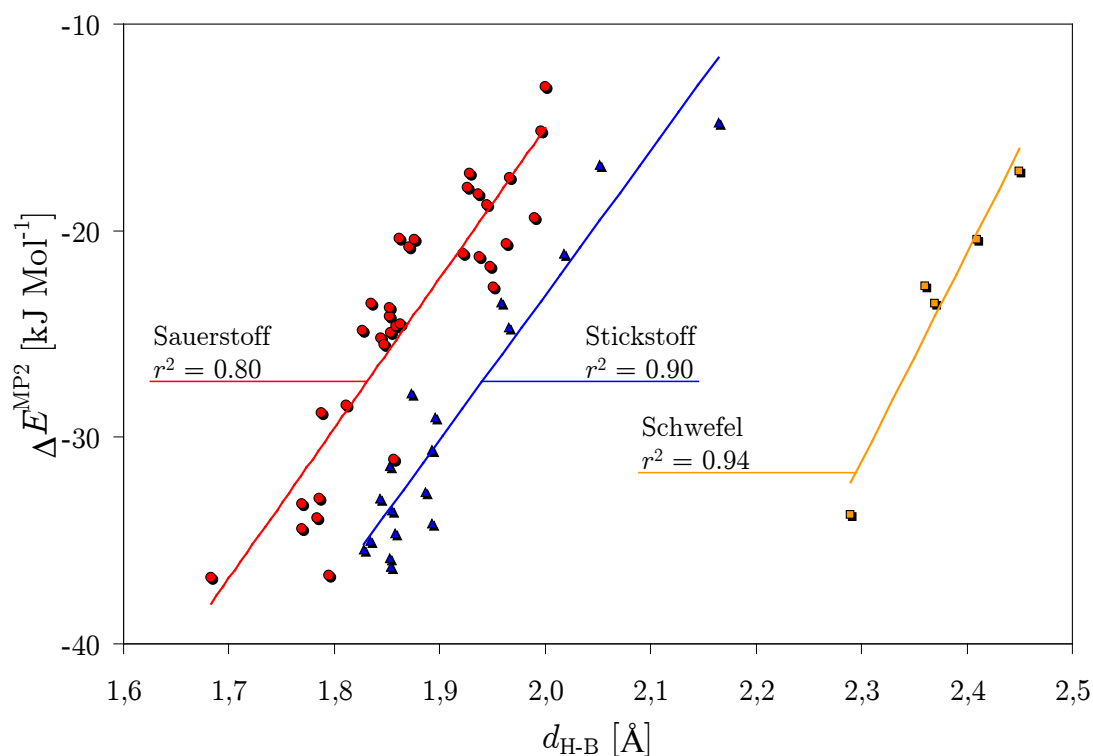


Abbildung 2.16. Korrelation zwischen der Wasserstoffbrückenlänge und der $MP2$ -Wechselwirkungsenergie.

den Inkrementtermen alle Beiträge zur Wechselwirkungsenergie einparametrisiert. Entsprechend ihres chemischen Kontextes werden Sauerstoff-, Stickstoff- und Schwefelatomen unterschiedliche Akzeptorenergie-Terme zugewiesen. Entsprechend werden polaren Wasserstoffatomen verschiedene Donorenergie-Terme zugewiesen in Abhängigkeit der funktionellen Gruppe, zu denen dieses Wasserstoffatom gehört.

Besteht eine Wasserstoffbrücke, so wird ihre Wechselwirkungsenergie angenähert durch die einfache Summe der entsprechenden Donor- und Akzeptorenergie-Terme:

$$\Delta E^{\text{HBInc}}(X - H \cdots B) = \Delta E_{\text{X-H}}^{\text{HBInc}} + \Delta E_{\text{B}}^{\text{HBInc}} \quad (2.67)$$

HBINC ist per Design der LUDI-Bewertungsfunktion so ähnlich wie möglich (siehe auch Kapitel 3.3.1). So können die Inkremente von HBINC nachher problemlos zur Erweiterung des LUDI-Modells herhalten.

In einigen Komplexen wurden mehr als ein nicht-kovalenter *BCP* gefunden, was darauf hinweist, dass es mehr als eine Wasserstoffbrücke in diesem Komplex gibt. Die entsprechenden Minimumsstrukturen dieser Komplexe, einschließlich der *BCPs* dieser H-Brücken sind in Abbildung 2.17 gezeigt. Als Beispiel sei der Komplex von Essigsäure (MeCOOH) und Wasser (H₂O) (Abbildung 2.17, **2**) näher betrachtet. Hier gibt es zwei nicht-kovalente *BCPs*, von denen der erste eine Wasserstoffbrückenbindung zwischen der OH-Gruppe der Essigsäure und dem Sauerstoff des Wassers anzeigt. Der zweite *BCP* weist auf eine weitere H-Brücke zwischen der OH-Gruppe des Wassers und des Carbonylsauerstoffs der Säure hin. In einem solchen Ringkomplex müssen jetzt natürlich die energetischen Beiträge *beider* Wasserstoffbrücken berücksichtigt werden.

Die Bindungsenergie der ersten H-Brücke lässt sich aufteilen in die Donorenergie der Hydroxylgruppe in MeCOOH ($\Delta E_{O.3H(COOH)}^{HBInc}$) und die Akzeptorenergie des Wassersauerstoffs ($\Delta E_{O.3(HOH)}^{HBInc}$). Die zweite Wasserstoffbrücke besteht aus der Donorenergie des Wassers ($\Delta E_{O.3H(HOH)}^{HBInc}$) und der Akzeptorenergie des Carbonylsauerstoffs ($\Delta E_{O.2(COOH)}^{HBInc}$).

$$\begin{aligned} \Delta E^{HBInc} (\text{MeCOOH} \cdots \text{H}_2\text{O}) &= \Delta E_{O.3H(COOH)}^{HBInc} + \Delta E_{O.3(HOH)}^{HBInc} \\ &+ \Delta E_{O.3H(HOH)}^{HBInc} + \Delta E_{O.2(COOH)}^{HBInc} \end{aligned} \quad (2.68)$$

Entsprechend ergibt sich die geschätzte Wechselwirkungsenergie des Komplexes von Wasser und Harnstoff (Abkürzung *Ur*) (Abbildung 2.17, **1**) zu

$$\begin{aligned} \Delta E^{HBInc} (\text{H}_2\text{O} \cdots \text{Ur}) &= \Delta E_{N.amH(urea)}^{HBInc} + \Delta E_{O.3(HOH)}^{HBInc} \\ &+ \Delta E_{O.3H(HOH)}^{HBInc} + \Delta E_{O.2(urea)}^{HBInc} \end{aligned} \quad (2.69)$$

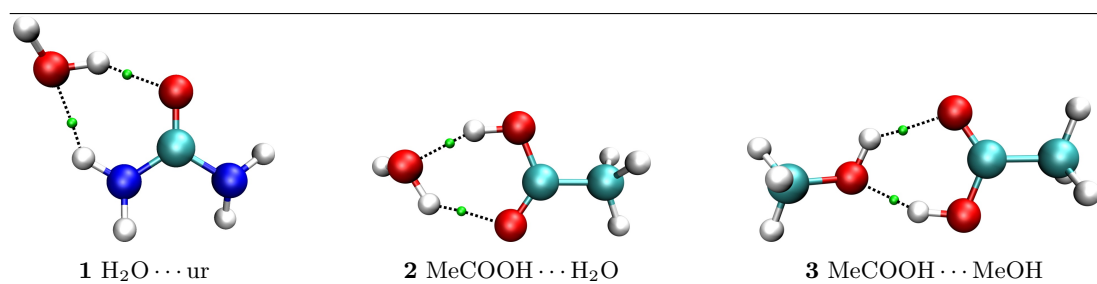


Abbildung 2.17. Minimumsstruktur der Komplexe mit einer zweiten konventionellen Wasserstoffbrücke.

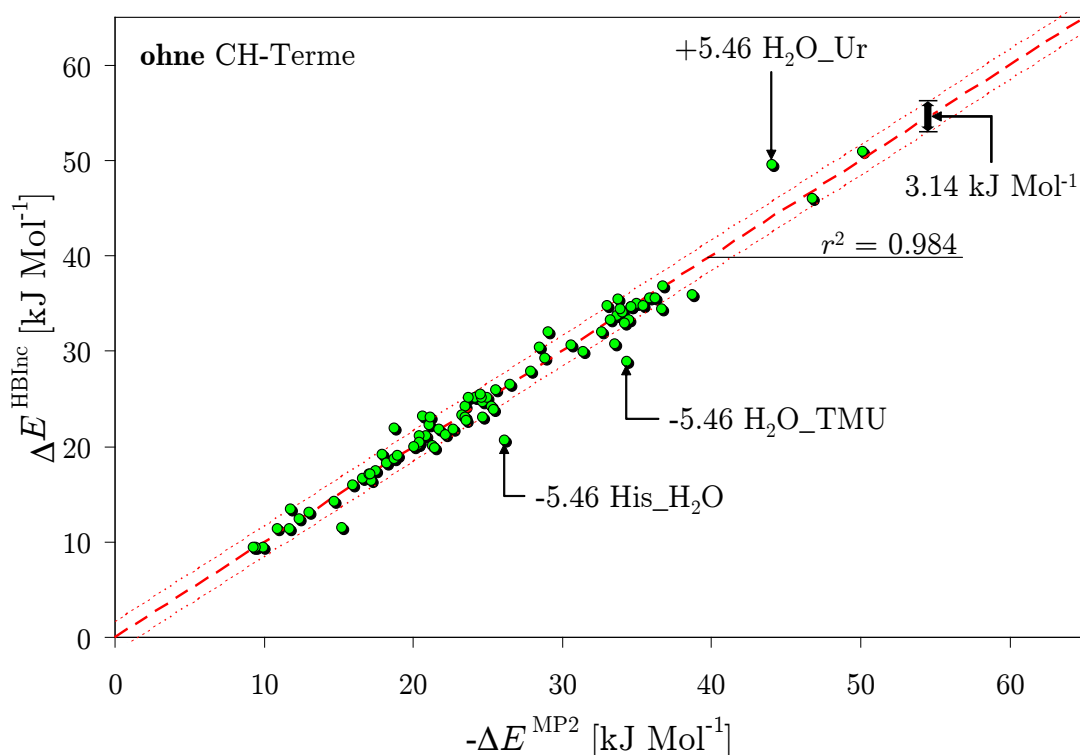


Abbildung 2.18. Korrelation zwischen der MP2 Wechselwirkungsenergie $\Delta E^{\text{MP2}}(\text{Don} \cdots \text{Acc})$ und $\Delta E^{\text{HBInc}}(\text{Don} \cdots \text{Acc})$.

und die des Komplexes von Essigsäure und Methanol (Abbildung 2.17, **3**) zu

$$\begin{aligned} \Delta E^{\text{HBInc}}(\text{MeCOOH} \cdots \text{MeOH}) &= \Delta E_{\text{O.3}(\text{COOH})}^{\text{HBInc}} + \Delta E_{\text{O.3}(\text{anol})}^{\text{HBInc}} \\ &+ \Delta E_{\text{O.3H}(\text{anol})}^{\text{HBInc}} + \Delta E_{\text{O.2}(\text{COOH})}^{\text{HBInc}}. \end{aligned} \quad (2.70)$$

Die einzelnen Beiträge jeder Donor- und Akzeptorgruppe sind durch eine *Partial Least Squares*-Regressionsanalyse (*PLS*) bestimmt worden. Hierbei werden die einzelnen Beiträge ΔE^{HBInc} darauf hin optimiert, dass die Summe der Fehlerquadrate von *MP2*-Energie $\Delta E^{\text{MP2}}(\text{Don} \cdots \text{Acc})$ und der mit HBINC geschätzten Energie $\Delta E^{\text{HBInc}}(\text{Don} \cdots \text{Acc})$ ein Minimum ist.

Insgesamt wurden so 12 Donortermine und 28 Akzeptortermine an die *MP2*-Wechselwirkungsenergien von 81 Komplexen mit insgesamt 97 Wasserstoffbrücken angepasst. Eine *signifikante* lineare Korrelation zwischen den *ab initio*-Energien $\Delta E^{\text{MP2}}(\text{Don} \cdots \text{Acc})$ und den HBINC-Energien $\Delta E^{\text{HBInc}}(\text{Don} \cdots \text{Acc})$ mit einem quadratischen Korrelationskoeffizienten von $r^2 = 0,984$ und einer Ausgleichsgeraden mit einer Steigung von $m = 0,977$ konnte so erreicht werden (siehe

Abbildung 2.18). Die Standardabweichung beträgt $1,57 \text{ kJ Mol}^{-1}$. Trotz dieses bemerkenswerten Resultats lohnt es sich, die wenigen Ausreißer näher zu analysieren.

Es fällt auf, dass die Ausreißer immer paarweise auftreten. Zum Beispiel wird der Komplex $\text{H}_2\text{O} \cdots \text{TMU}$ durch HBINC *unterschätzt*, während der Komplex $\text{H}_2\text{O} \cdots \text{Ur}$, der dieselbe Kombination aus Donor- und Akzeptorgruppe besitzt, um etwa den gleichen Betrag *überschätzt* wird. Die Tatsache, dass die Wechselwirkungsenergie von $\text{H}_2\text{O} \cdots \text{TMU}$ unterschätzt wird, deutet darauf hin, dass es in diesem Komplex noch weitere Wechselwirkungen neben der einzelnen H-Brücke gibt, die zur Stabilisierung des Komplexes beitragen. Die Überschätzung der Bindungsenergie von $\text{H}_2\text{O} \cdots \text{Ur}$ ist dann ein Artefakt der Parameteroptimierung. Da offensichtlich Wechselwirkungsterme für $\text{H}_2\text{O} \cdots \text{TMU}$ fehlen, werden die Werte der bereits vorhandenen Terme unnatürlich erhöht, um doch noch annähernd die *MP2*-Energie erreichen zu können. Komplexe, die die gleichen Inkremente benutzen, aber bereits vollständig ohne zusätzliche Terme beschrieben werden können, werden durch die unnatürlich hohen Werte dieser Inkremente entsprechend überschätzt. Deren Unterschätzung und die damit verbundene Abweichung wirken einer Überhöhung der Parameter entgegen. So entstehen die Paare von Abweichungen.

Erst eine detailliertere Analyse der Elektronendichte der Komplexe enthüllte in 12 Komplexen (**4** – **15**, Abbildung 2.19) zusätzliche nicht-kovalente *BCPs* zwischen jeweils einer CH-Gruppe und einem konventionellen Wasserstoffbrückenakzeptor. Diese *BCPs* sind deutlich schwerer zu finden als die einer konventionellen H-Brücke. Die Elektronendichte weist um die *BCPs* einer CH-Wasserstoffbrücke eine deutlich flachere Topologie auf. Dadurch sind für die numerische Bestimmung viel engere Konvergenzkriterien und sehr viel mehr Schritte notwendig. [124]

Die CH-aziden Wechselwirkungen werden durch die Einführung von zwei verschiedenen CH-Termen, einen für sp^3 -CH-Donoren (C.3H) und einen weiteren für sp^2 -Kohlenstoffatome (C.2H), berücksichtigt. Mit dieser Ergänzung ergibt sich die gesamte Bindungsenergie des wasserstoffverbrückten Komplexes von Essigsäure und Formaldehyd (Komponente **12** in Abbildung 2.19) zu:

$$\begin{aligned} \Delta E^{\text{HBInc}}(\text{MeCOOH} \cdots \text{H}_2\text{CO}) &= \Delta E_{\text{O.3H}(\text{COOH})}^{\text{HBInc}} + \Delta E_{\text{O.2}(\text{keto})}^{\text{HBInc}} \\ &+ \Delta E_{\text{C.2H}}^{\text{HBInc}}. \end{aligned} \quad (2.71)$$

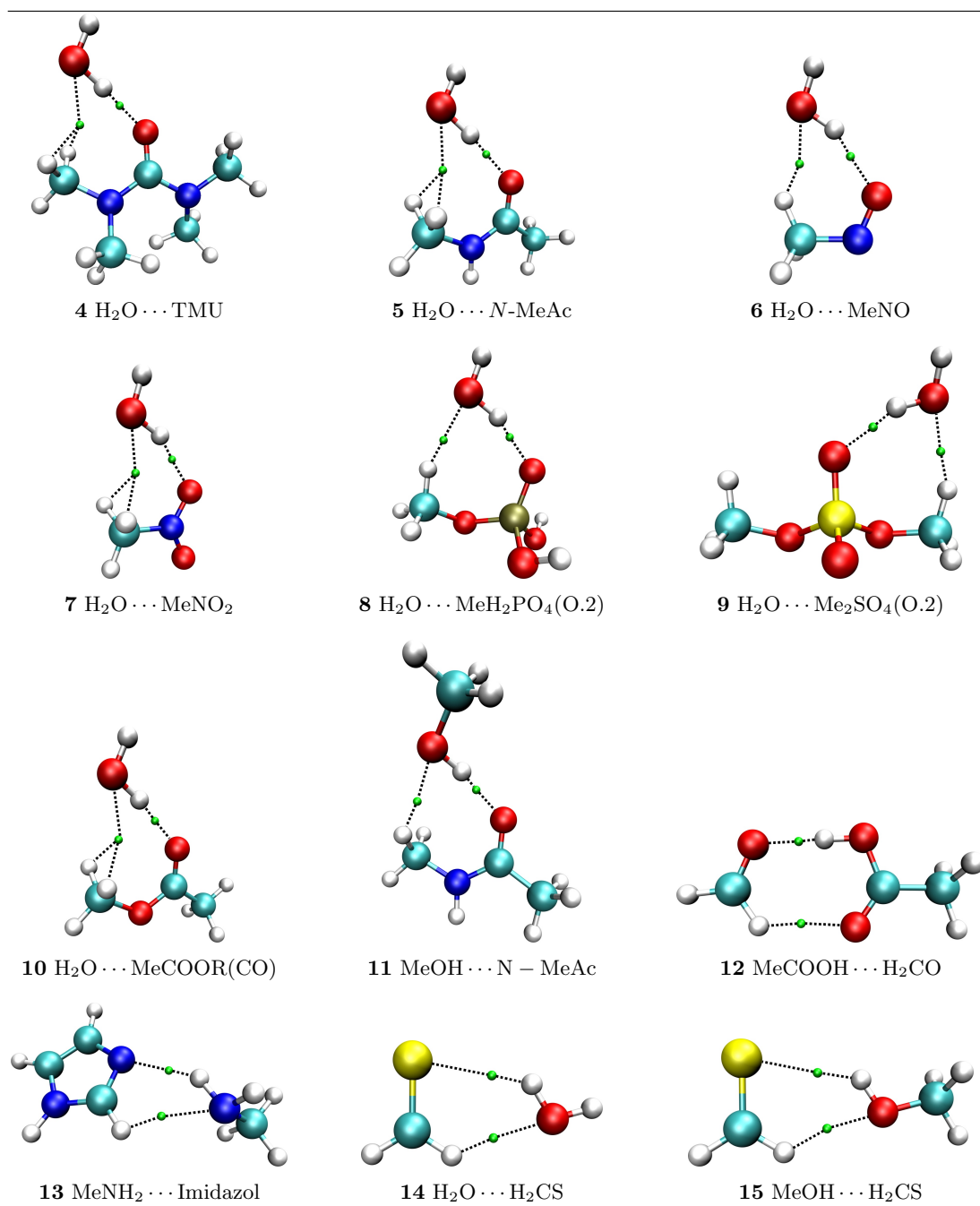


Abbildung 2.19. Minimumstrukturen der Komplexe mit einer zweiten Wasserstoffbrücke, die von einer CH-gruppe gebildet wird.

Die CH–Terme sind also keine normalen Donorsterme, da sie in der Summe kein entsprechendes Akzeptor-Pendant haben, sondern der Beitrag ist für jede CH–Brücke, entsprechend ihrer Hybridisierung, immer gleich und zwar explizit unabhängig von der Akzeptorgruppe.

In Abbildung 2.20 ist die Korrelation zwischen der *MP2*-Bindungsenergie der Komplexe, die eine CH–Donorgruppe besitzen, und den HBINC-Energien mit und ohne den neuen CH–Termen verglichen. Werden die CH–aziden Wechselwirkungen nicht berücksichtigt (Abbildung 2.20a), wird die fehlende Wechselwirkungsenergie durch eine Erhöhung der entsprechenden Inkremente konventioneller Donor- und Akzeptorgruppen kompensiert. Durch die Einführung der CH–Terme (Abbildung 2.20b) verschwinden die vormaligen Ausreißer während die Bewertung der Komplexe, deren Bindungsenergie auch ohne die CH–Terme schon gut abgebildet wurde, so gut wie unverändert bleibt.

Doch es gibt Ausnahmen. Die Bindungsenergie des MeNH₂–Imidazol-Komplexes wird deutlich überschätzt, wenn der C.2H-Term berücksichtigt wird, wohingegen HBINC ohne CH–Terme am Imidazol die *MP2*-Wechselwirkungsenergie dieses Komplexes mit einem Fehler von nicht mal $\Delta\Delta E = 1,0 \text{ kJ Mol}^{-1}$ wiedergibt. Der Grund für diese Abweichung liegt vermutlich in einer Unstimmigkeit in der Spezifikation der Atomtypen. In der SYBYL-Repräsentation des Imidazols sind *per definitionem* alle Kohlenstoffatome C.2 (also sp²-hybridisiert) und alle Stickstoffatome N.2 (ebenfalls sp²-hybridisiert). Da Imidazol aber ein *aromatischer* Heterozyklus ist, ist das Kohlenstoffatom, das hier als H–Brückendonor fungiert, eher von Typ C.ar anstatt C.2. Da in dem Komplex ein *BCP* für diese CH–Bindung gefunden wurde (**13**, Abbildung 2.19), kann man vermuten, dass aromatische Kohlenstoffatome nur sehr schwache Wasserstoffbrücken bilden. Das grundsätzliche Vorhandensein einer H–Brücke an dieser Stelle wird durch den *BCP* indiziert. Diese Sichtweise wird durch den Umstand bekräftigt, dass keine CH–aziden H–Brücken in den Komplexen von Phenol (Tabelle 2.4) oder Phenylamin (Tabelle 2.5) mit Wasser oder aliphatischen Alkoholen gefunden wurden. In diesen Fällen wurde nicht einmal ein *BCP* zwischen dem *ortho*-Wasserstoff des Phenylrings und dem Akzeptorsauerstoff in der unmittelbaren Nähe gefunden.

Eine weitere Ausnahme ist der Komplex MeNH₂ ··· H₂CO (**16**, Abbildung 2.21). Hier wurde kein *BCP* für eine CH–Wasserstoffbrücke gefunden, obwohl es Gründe gibt anzunehmen, dass eine C.2H–Wasserstoffbrücke in dem Komplex existiert. Diese Vermutung gründet sich darauf, dass die HBINC-Energie von diesem Kom-

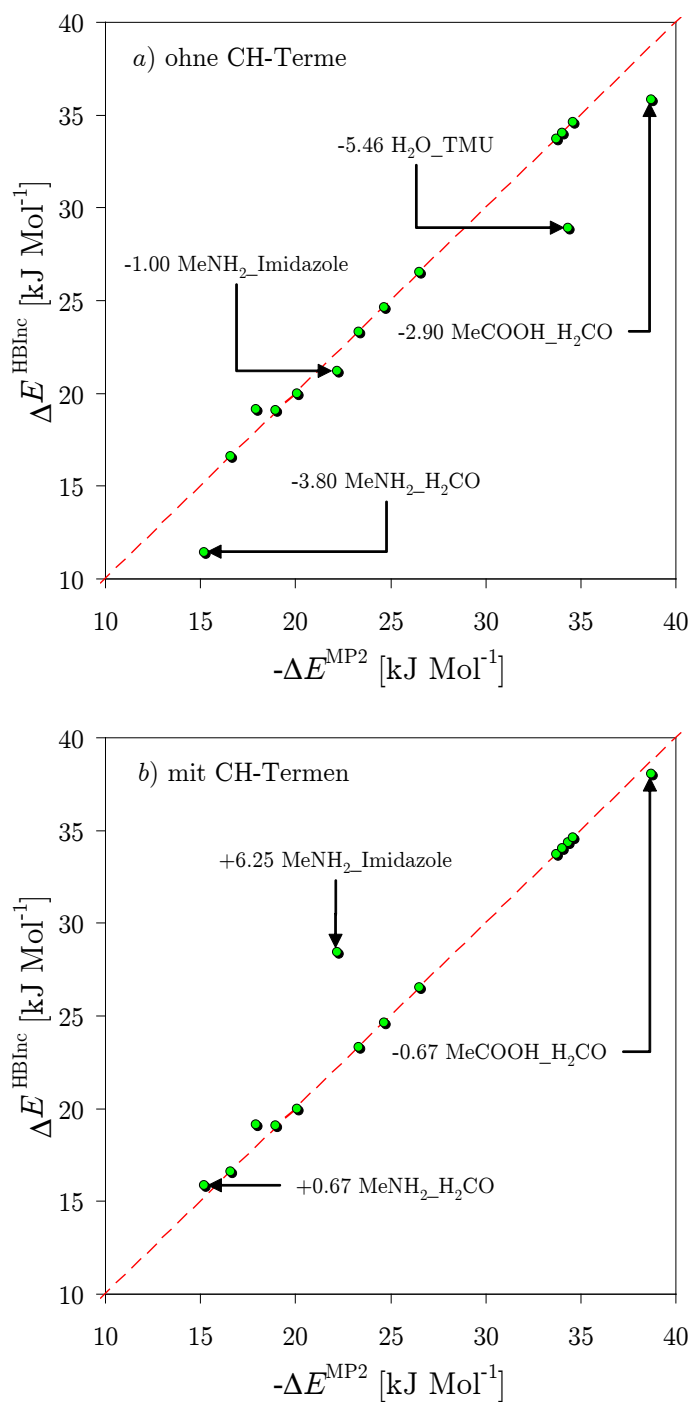


Abbildung 2.20. Vergleich der Korrelationen zwischen $\Delta E^{\text{MP2}}(\text{Don} \cdots \text{Acc})$ und $\Delta E^{\text{HBInc}}(\text{Don} \cdots \text{Acc})$ von Komplexen, die eine CH-Brücke aufweisen, a) mit und b) ohne CH-Energierme.

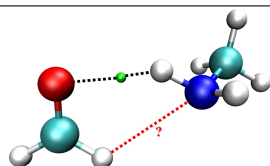
plex ohne einen CH-Term zu tief ist und gleichzeitig andere Komplexe mit einer O.2(keto)-Akzeptorgruppe überbewertet werden.

Die lineare Regression zwischen der MP2-Energie aller 81 Komplexe und der endgültigen HBINC-Funktion mit 13 Donortermen, 2 aliphatischen CH-Termen und 28 Akzeptortermen ist in Abbildung 2.22 gezeigt. Eine exzellente lineare Korrelation mit einem quadratischen Korrelationskoeffizienten von $r^2 = 0,994$ und einer Steigung von $m = 0,999$ wurde erreicht. Die Standardabweichung ist $0,96 \text{ kJ Mol}^{-1}$. Obwohl die Einführung der CH-Terme nur eine leichte Verbesserung der ohnehin schon guten Korrelation gebracht hat, wurden dadurch die augenscheinlichen Ausreißer des ersten Ansatzes korrigiert. Die maximale Abweichung zwischen HBINC und MP2-Energie wurde für den Komplex $\text{H}_2\text{O} \cdots \text{MeN}=\text{CH}_2$ gefunden und beträgt $3,2 \text{ kJ Mol}^{-1}$.

2.4.3 Validierung des Modells

Um die Leistungsfähigkeit von HBINC in der Vorhersage der Wechselwirkungsenergien von Komplexen, die *nicht* in dem Trainingssatz enthalten sind, zu testen, wurden die DFT-Geometrien und die entsprechenden MP2-Wechselwirkungsenergien von 11 weiteren Komplexen bestimmt (Tabelle 2.7 und Abbildung 2.23). Zusätzlich wurden die MP2-Bindungsenergien von verschiedenen starken zyklischen Wasserstoffbrücken aus der Literatur [44] zur Validierung verwendet. Dies sind das Ameisensäuredimer, das Formamiddimer, der 2-Pyridoxin-2-Aminopyridin-Komplex und verschiedene Nukleinbasenpaare (Tabelle 2.7).

Der Vergleich der MP2-Bindungsenergien der Komplexe, die in Abbildung 2.23 dargestellt sind, mit den Wechselwirkungsenergien, die mit HBINC berechnet wurden, zeigt eine gute lineare Korrelation mit einem quadratischen Korrelationskoeffizienten von $r^2 = 0,94$ (Abbildung 2.24). Die H-Brückenenergie von sieben der elf Komplexe wurden mit einem Fehler von weniger als $2,7 \text{ kJ Mol}^{-1}$ abgeschätzt.



16 $\text{MeNH}_2 \cdots \text{H}_2\text{CO}$

Abbildung 2.21. Minimumsstruktur des Komplexes $\text{MeNH}_2 \cdots \text{H}_2\text{CO}$, in dem eine CH-Wasserstoffbrücke erwartet wird, aber kein bindungskritischer Punkt gefunden wurde.

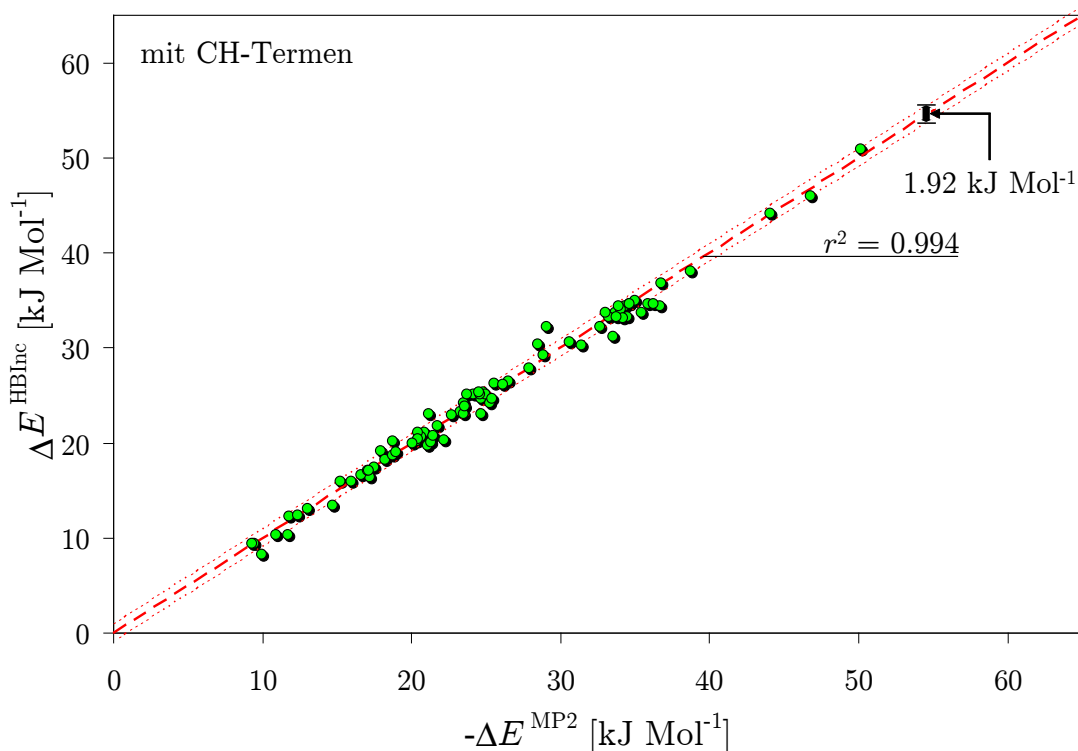


Abbildung 2.22. Korrelation zwischen der MP2 Wechselwirkungsenergie $\Delta E^{\text{MP2}}(\text{Don} \cdots \text{Acc})$ und $\Delta E^{\text{HBinc}}(\text{Don} \cdots \text{Acc})$ mit zusätzlichen Termen für schwache CH-Wasserstoffbrücken.

Im Fall des Komplexes von Fucose und Wasser ($\text{FCA} \cdots \text{H}_2\text{O}$, siehe **19**, Abbildung 2.23) überschätzt HBINC die MP2-Energie deutlich um mehr als $5,8 \text{ kJ Mol}^{-1}$. Während in den anderen Wasser–Fucose Komplexen das Wasser an eine externe OH-Gruppe des Zuckers bindet, ist einer der Wasserstoffbrückenakzeptoren in Komplex **19** ein sp^3 -Sauerstoff vom Ether-Typ ($\text{O.3}(\text{COC})$). Die H–Brücke zwischen der OH-Gruppe des Wassers und des Ether-O.3 des Fucose-Moleküls ist *länger* ($d_{\text{OH}} = 1.9 \text{ \AA}$) als es für diese Art von Donor und Akzeptor üblich ist ($d_{\text{OH}} = 1.8 \text{ \AA}$). Lässt man jedoch den Beitrag dieser Wasserstoffbrücke einfach weg, wird die Wechselwirkungsenergie des Komplexes um mehr als 19 kJ Mol^{-1} *unterschätzt*.

Obwohl kein *BCP* in der Elektronendichte zwischen Donor und Akzeptor gefunden werden konnte, habe ich den Beitrag dieser H–Brücke zum Ether-Sauerstoff des Zuckers mit berücksichtigt. Da die Ringzucker in biologischen Systemen eine wichtige Rolle einnehmen, sind auf Grund dieser Unstimmigkeiten weitere Untersuchungen an dieser Wechselwirkung von Nöten.

Systeme mit zwei Wasserstoffbrücken, deren zwei Donorzentren und zwei Akzeptorzentren einen 8-Ring bilden, sind im aktuellen Inkrementmodell von HBINC

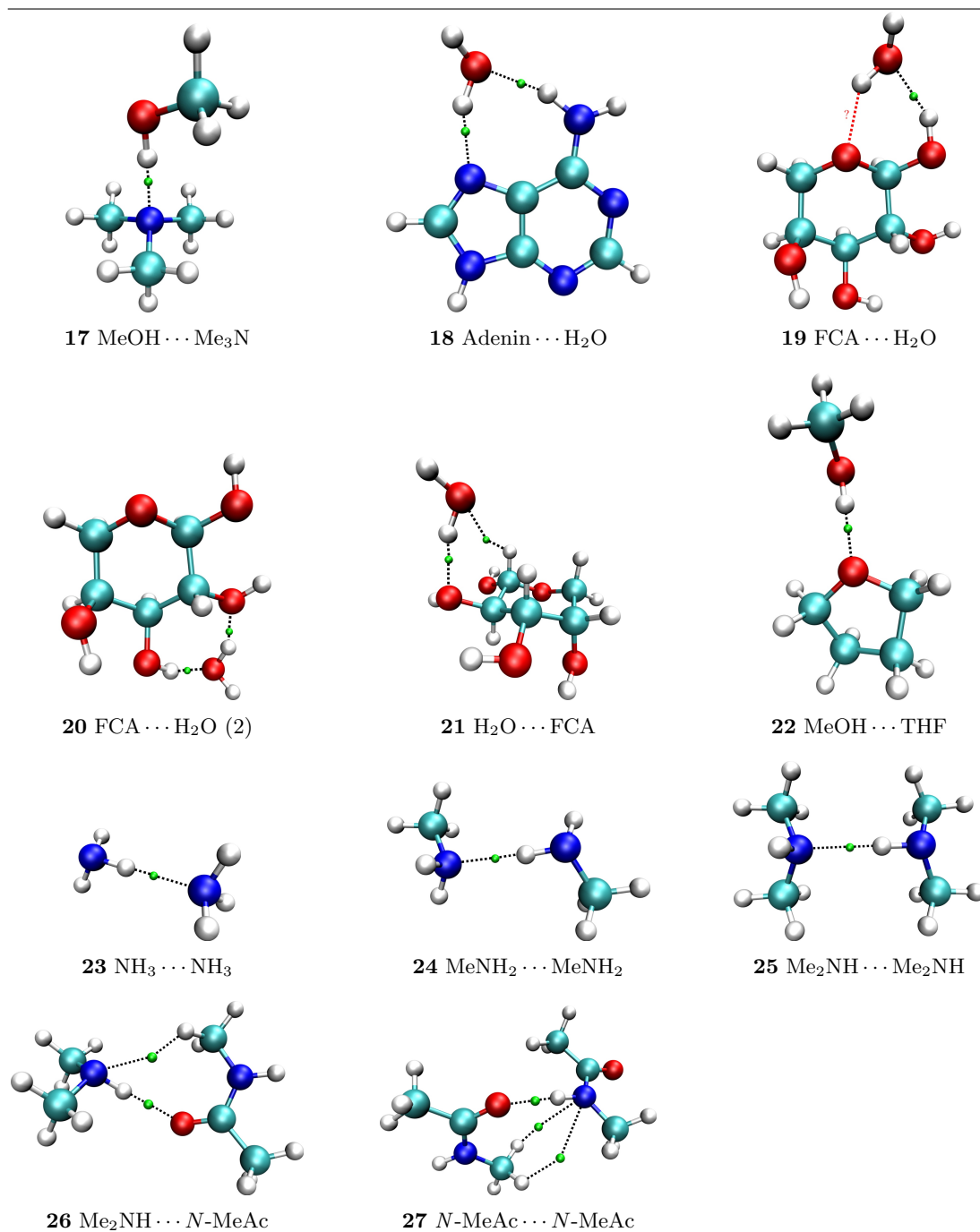


Abbildung 2.23. Minimumsstrukturen der Komplexe des Testsatzes. Die Abkürzungen sind in den Fußnoten zu Tabelle 2.7 erklärt.

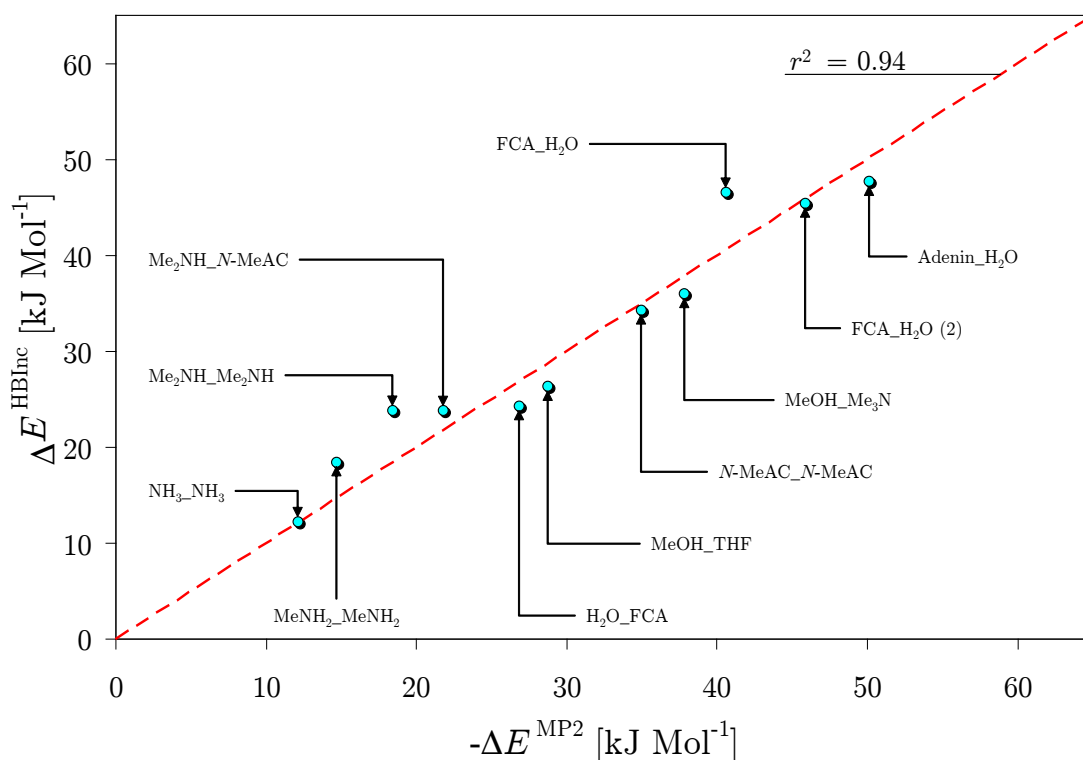


Abbildung 2.24. Korrelation zwischen der MP2-Wechselwirkungsenergie $\Delta E^{\text{MP2}}(\text{Don} \cdots \text{Acc})$ und $\Delta E^{\text{HBINC}}(\text{Don} \cdots \text{Acc})$ für die Komplexe des Testsatzes.

nicht gesondert vertreten. Die Struktur eines 8-Rings ist wegen ihrer Geometrie besonders günstig für Wasserstoffbrücken, da Donor, Proton und Akzeptor ungefähr auf einer Linie liegen.

Die zyklischen Komplexe in Tabelle 2.7 haben alle diese Struktur. Ihre Wechselwirkungsenergien werden von HBINC um 30 – 40% *unterschätzt* (Abbildung 2.25). Die lineare Korrelation zwischen der MP2-Wechselwirkungsenergie dieser Komplexe und der mit HBINC bestimmten Energie ist trotz dieser starken Abweichung mit $r^2 = 0,95$ überraschend groß. Die resultierende Ausgleichsgerade hat jedoch eine Steigung $m = 0,59$, ist also zu flach. Das deutet darauf hin, dass in allen betroffenen Komplexen eine Wechselwirkung systematisch unterschätzt oder überhaupt gar nicht erfasst wird. Besonders interessant ist, dass der Anteil, den diese fehlende Wechselwirkung ausmacht, bei allen Komplexen ungefähr gleich groß ist, was die Vermutung nahe legt, dass es sich jedes Mal um die *gleiche* Wechselwirkung handelt.

Eine detaillierte Analyse der DNA-Basenpaare von Mo [2] hat gezeigt, dass in diesen Komplexen die elektrostatischen Wechselwirkungen und die Polarisation

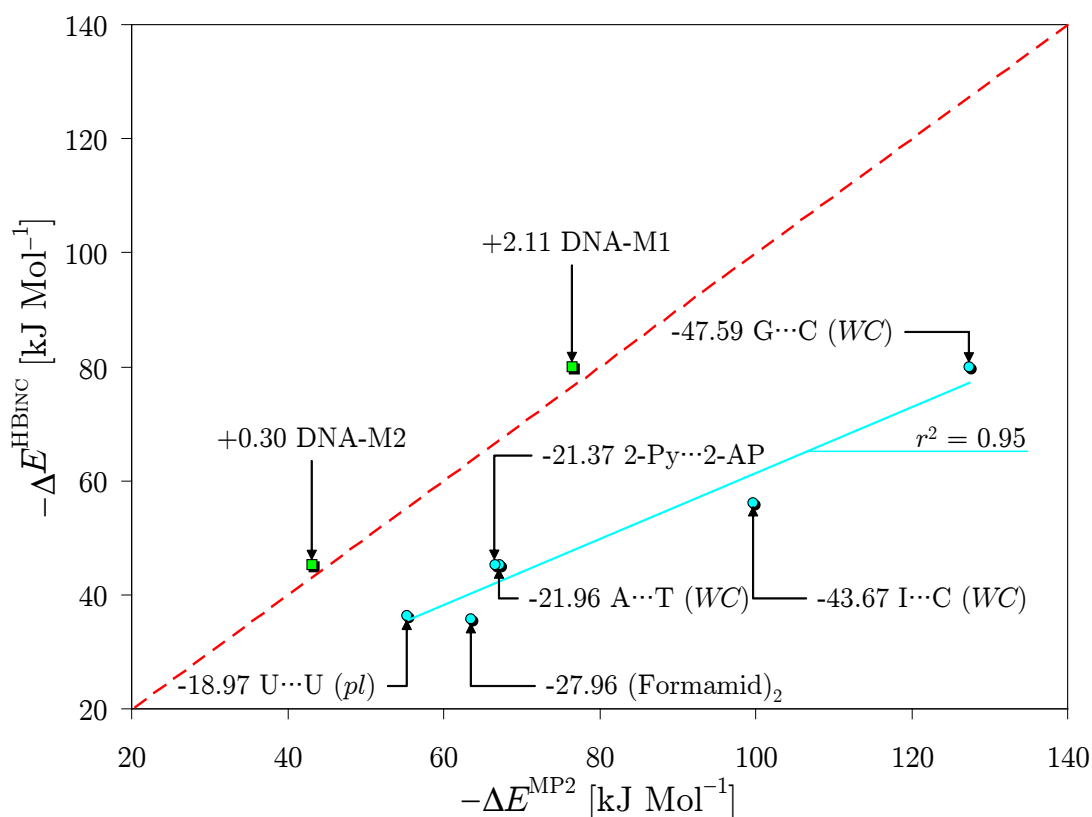


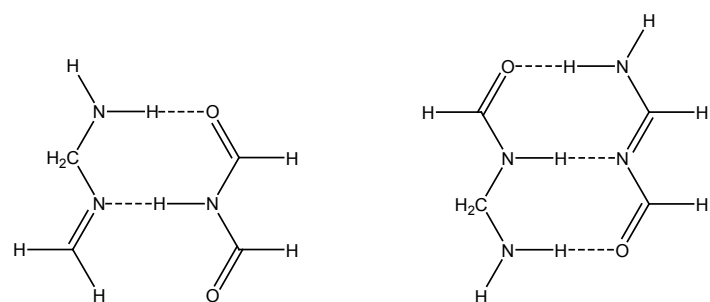
Abbildung 2.25. Korrelation zwischen der MP2-Wechselwirkungsenergie $\Delta E^{\text{MP2}}(\text{Don} \cdots \text{Acc})$ und $\Delta E^{\text{HBINC}}(\text{Don} \cdots \text{Acc})$ für die Komplexe des Testsatzes mit zyklischen Wasserstoffbrückenmustern.

durch *Resonanz* verstärkt werden, was für die außergewöhnlich hohe Bindungsenergie verantwortlich ist.

Mo hat darüber hinaus einfache Modellsystem vorgestellt, die das *Interface* (also den Teil der Komplexe, in denen sich die H-Brücken ausbilden) der Basenpaare nachahmt, jedoch ohne die (aromatischen) Ringsysteme (Abbildung 2.26). Damit ist dem System die Möglichkeit genommen, zwitterionische Resonanzstrukturen zu bilden und dadurch kann es in diesen Modellen der DNA-Basenpaaren auch nicht zu einer Verstärkung der Elektrostatik oder der Polarisierung kommen.

Die berechneten Wechselwirkungsenergien von DNA-M1 und DNA-M2 sind deutlich niedriger als die Energien der entsprechenden DNA-Basenpaare (Tabelle 2.7) und werden auch von HBINC sehr gut reproduziert (Tabelle 2.7, Abbildung 2.25).

Unglücklicherweise lassen sich solche Resonanz-Effekte nur schwerlich in einfache Terme, wie HBINC sie verwendet und in denen die Energie *per definitionem* additiv ist, einbauen. In solchen Fällen erscheint es angezeigt, für Donor-



28 DNA-M1: Model of WC-AT

29 DNA-M2: Model of WC-GC

Abbildung 2.26. Vereinfachte DNA-Modelle für Watson-Crick-Adenin-Thymin (M1) und Watson-Crick-Guanin-Cytosin (M2) (siehe Referenz [2]). DNA-M1: Komplex von *N*-Methylen-Methan-Diamin mit *N*-Formyl-Formamid. DNA-M2: Komplex von *N*-(Aminomethyl)-Formamid mit (*E*)-*N*-(Aminomethylen)-Formamid.

und Akzeptorgruppen, die aufgrund von Resonanz verstärkte H-Brücken bilden, gesonderte Inkremente einzuführen und anzupassen. Genau das werde ich in Kapitel 3 für die Parametrisierung einer Bewertungsfunktion für *Docking* und *High-Throughput Screening* vorstellen.

Tabelle 2.3. Donor- und Akzeptorinkremente, deren Energien (in kJ Mol^{-1}) und die Häufigkeit ihres Vorkommens im Trainingssatz der 81 Komplexe.

Inkrement	ΔE	Häufigkeit
HO.3(H ₂ O)	8,56	39
HO.3(anol)	9,48	16
HO.3(enol)	17,59	4
HO.3(inol)	25,13	1
HO.3(c,2ar)	18,78	7
HN.3(prim)	-3,34	5
HN.3(sec)	-1,34	2
HN.3(planar)	14,47	2
HN.3(amid)	9,07	3
HN.3(C.ar)	5,40	1
HN.2(cmn)	0,72	1
HS.3(cmn)	-2,28	2
O.3(H ₂ O)	11,65	17
O.3(anol)	15,63	13
O.3(enol)	7,79	2
O.3(inol)	4,46	1
O.3(phenol)	10,56	2
O.3(COC)	16,78	3
O.3(COOH)(OH)	8,90	1
O.3(COOR)(OR)	10,17	1
O.3(NO)	13,15	1
O.3(PO)	9,68	1
O.3(SO)	7,36	1
O.2(keto)	11,14	5
O.2(COOH)(CO)	6,93	2
O.2(COOR)(CO)	-5,00	1
O.2(amid)	9,08	4
O.2(NO)	-8,29	2
O.2(PO)	9,08	1
O.2(SO)	-0,30	1
N.3(cmn)	19,29	1
N.3(prim)	21,67	2
N.3(sec/tert)	25,78	5
N.3(planar)	14,52	3
N.3(NOH)	26,08	1
N.2ar(cmn)	23,85	5
S.3(thiol)	11,87	1
S.3(sulfid)	14,38	3
S.3(disulfid)	8,54	1
S.2(cmn)	2,41	2
HC.3	16,38	7
HC.2	8,09	4

Tabelle 2.4: Ergebnisse für wasserstoffverbrückte Komplexe, die ein Sauerstoffatom als Akzeptor oder Donor enthalten. Die MP2-Bindungsenergie $\Delta E_{\text{MP2}}^{\text{CPC}}$ [kJ Mol⁻¹] beinhaltet *Counterpose-Korrektur*. $d_{\text{H-B}}$ [Å] ist die Wasserstoffbrückenlänge, $d_{\text{BCP-B}}$ [Å] ist der Abstand zwischen dem bindungskritischen Punkt und dem Akzeptoratom. Die Partialladungen des Wasserstoffatoms ($q^{\text{AIM}}(\text{H})$) und die des Akzeptors ($q^{\text{AIM}}(\text{B})$) sind durch AIM-Analyse des wasserstoffverbrückten Komplexes bestimmt.

Komplexe	Donor	Akzeptor	$\Delta E_{\text{MP2}}^{\text{CPC}}$	$d_{\text{H-B}}$	$d_{\text{BCP-B}}$	$q^{\text{AIM}}(\text{H})$	$q^{\text{AIM}}(\text{B})$
H ₂ O...H ₂ O	O.3H	O.3	-20,38	1,86	1,22	0,60	-1,13
H ₂ O...MeOH	O.3H	O.3	-23,52	1,84	1,20	0,61	-1,11
H ₂ O...EtOH	O.3H	O.3	-25,23	1,84	1,21	0,61	-1,11
H ₂ O...Ethenol	O.3H	O.3	-17,21	1,93	1,25	0,59	-1,14
H ₂ O... <i>m</i> -Butinol ¹	O.3H	O.3	-13,04	2,00	1,28	0,59	-1,16
H ₂ O...PhOH	O.3H	O.3	-17,90	1,93	1,25	0,60	-1,14
H ₂ O...Me ₂ N – OH	O.3H	O.3	-21,73	1,95	1,24	0,48	-0,56
H ₂ O...MeH ₂ PO ₄	O.3H	O.3	-18,26	1,94	1,18	0,41	-1,13
H ₂ O...Me ₂ SO ₄	O.3H	O.3	-15,93	2,21	1,29	—	—
H ₂ O...MeOMe	O.3H	O.3	-24,83	1,83	1,20	0,61	-1,11
H ₂ O...MeCOOH	O.3H	O.3	-17,47	1,97	1,26	0,59	-1,14
H ₂ O...MeCOOMe	O.3H	O.3	-18,74	1,95	1,19	0,41	-0,96
MeOH...H ₂ O	O.3H	O.3	-20,83	1,87	1,22	0,60	-1,13
MeOH...MeOH	O.3H	O.3	-24,17	1,85	1,22	0,60	-1,11
MeOH...EtOH	O.3H	O.3	-24,97	1,85	1,22	0,60	-1,11
MeOH...PhOH	O.3H	O.3	-21,28	1,94	1,25	0,59	-1,14
MeOH...MeOMe	O.3H	O.3	-25,51	1,85	1,21	0,60	-1,11
EtOH...H ₂ O	O.3H	O.3	-20,43	1,88	1,23	0,60	-1,13
EtOH...MeOH	O.3H	O.3	-23,75	1,85	1,56	0,44	-0,87
EtOH...EtOH	O.3H	O.3	-24,63	1,86	1,22	0,60	-1,11
Ethenol...H ₂ O	O.3H	O.3	-28,85	1,79	1,17	0,60	-1,10
Ethenol...MeOH	O.3H	O.3	-33,24	1,77	1,16	0,60	-1,08
Ethenol...EtOH	O.3H	O.3	-34,44	1,77	1,18	0,63	-1,11
Ethenol...Ethenol	O.3H	O.3	-24,53	1,86	1,22	0,62	-1,15
Ethinol...H ₂ O	O.3H	O.3	-36,78	1,68	1,14	0,65	-1,14
PhOH...H ₂ O	O.3H	O.3	-28,48	1,81	1,20	0,62	-1,14
PhOH...MeOH	O.3H	O.3	-33,90	1,78	1,19	0,62	-1,11
PhOH...EtOH	O.3H	O.3	-36,71	1,80	1,19	0,62	-1,11
MeCOOH...MeOH	O.3H	O.3	-50,13	1,71	1,10	0,49	-0,91
←	O.3H	O.2		1,99	1,21	0,46	-1,02
H ₂ O...H ₂ CO	O.3H	O.2	-21,11	1,92	1,25	0,60	-1,10
H ₂ O...Harnstoff	O.3H	O.2	-44,98	1,81	1,13	0,44	-1,04
←	N, amH	O.3		2,02	1,22	0,40	-0,83
H ₂ O...TMU ²	O.3H	O.2	-34,36	1,81	1,19	0,62	-1,20
←	C.3H	O.3		2,69	1,64	0,04	-1,20
H ₂ O... <i>N</i> -MeAc ³	O.3H	O.2	-33,70	1,82	1,17	0,62	-1,19
←	C.3H	O.3		2,69	1,52	0,06	-1,20
H ₂ O...MeNO	O.3H	O.2	-16,58	2,01	1,22	0,37	-1,01
←	C.3H	O.3		2,46	1,42	0,02	-0,78
H ₂ O...MeNO ₂	O.3H	O.2	-23,30	1,96	1,17	—	—
←	C.3H	O.3		2,47	1,52	—	—
H ₂ O...MeH ₂ PO ₄	O.3H	O.2	-34,03	1,83	1,20	0,62	-1,52
←	C.3H	O.3		2,34	1,40	0,09	-1,19
H ₂ O...Me ₂ SO ₄	O.3H	O.2	-24,65	1,94	1,18	0,42	-1,05
←	C.3H	O.3		2,31	1,33	0,05	-0,81

...

¹ 3,3-Dimethyl-but-1-yn-1-ol² *N,N,N',N'*-Tetramethyl-Harnstoff³ *N*-Methyl-Acetamid als Modellsystem für Peptide

Fortsetzung ...

Komplexe	Donor	Akzeptor	$\Delta E_{\text{MP2}}^{\text{CPC}}$	$d_{\text{H-B}}$	$d_{\text{BCP-B}}$	$q^{\text{AIM}}(\text{H})$	$q^{\text{AIM}}(\text{B})$
H ₂ O...MeCOOH	O.3H	O.2	-46,74	1,73	1,11	0,49	-0,84
←	O.3H	O.3		1,94	1,19	0,45	-1,02
H ₂ O...MeCOOMe	O.3H	O.2	-26,51	1,94	1,19	0,44	-1,02
←	C.3H	O.3		2,83	1,66	0,01	-0,84
MeOH...H ₂ CO	O.3H	O.2	-20,66	1,96	1,27	0,59	-1,10
MeOH...N-MeAc	O.3H	O.2	-34,61	1,84	1,20	0,61	-1,19
←	C.3H	O.3		2,49	1,50	0,08	-1,17
MeCOOH...H ₂ CO	O.3H	O.2	-38,73	1,79	1,14	0,48	-0,93
←	C.2H	O.3		2,43	1,37	0,06	-1,02

Tabelle 2.5: Ergebnisse für wasserstoffverbrückte Komplexe mit Stickstoff als Donor oder Akzeptor. Für weitere Erläuterungen siehe Tabelle 2.4.

Komplexe	Donor	Akzeptor	$\Delta E_{\text{MP2}}^{\text{CPC}}$	$d_{\text{H-B}}$	$d_{\text{BCP-B}}$	$q^{\text{AIM}}(\text{H})$	$q^{\text{AIM}}(\text{B})$
$\text{H}_2\text{O} \cdots \text{NH}_3$	O.3H	N.3	-27,87	1,87	1,25	0,60	-0,98
$\text{H}_2\text{O} \cdots \text{MeNH}_2$	O.3H	N.3	-31,40	1,85	1,25	0,60	-0,94
$\text{H}_2\text{O} \cdots \text{Me}_2\text{NH}$	O.3H	N.3	-32,98	2,04	1,24	0,60	-0,92
$\text{H}_2\text{O} \cdots \text{Me}_3\text{N}$	O.3H	N.3	-35,01	1,83	1,24	0,59	-0,96
$\text{H}_2\text{O} \cdots \text{PhNH}_2$	O.3H	N.pl3	-24,67	1,97	1,24	0,41	-1,05
$\text{H}_2\text{O} \cdots \text{Me}_2\text{N} - \text{OH}$	O.3H	N.3	-34,66	1,86	1,20	0,42	-0,69
$\text{H}_2\text{O} \cdots \text{H}_2\text{C} = \text{CH} - \text{NH}_2$	O.3H	N.pl3	-23,49	1,96	1,24	0,41	-1,06
$\text{H}_2\text{O} \cdots \text{Pyrrolidin}$	O.3H	N.3	-35,46	1,83	1,23	0,51	-0,89
$\text{MeCOOH} \cdots \text{MeNH}_2$	O.3H	N.3	-60,07	1,66	1,12	0,63	-0,99
	← N.3H	O.2		2,56	1,44	0,37	-1,20
$\text{MeOH} \cdots \text{MeNH}_2$	O.3H	N.3	-33,56	1,85	1,25	0,60	-1,19
$\text{MeOH} \cdots \text{Me}_2\text{NH}$	O.3H	N.3	-35,84	1,85	1,25	0,60	-0,93
$\text{MeOH} \cdots \text{Pyrrolidin}$	O.3H	N.3	-36,26	2,05	1,25	0,61	-0,90
$\text{H}_2\text{O} \cdots N\text{-MeAc}$	O.3H	N.am	-21,14	2,02	1,26	0,40	-1,34
$\text{H}_2\text{O} \cdots \text{Imidazol}$	O.3H	N.2	-32,65	1,89	1,25	0,61	-1,11
$\text{H}_2\text{O} \cdots \text{MeN} = \text{CH}_2$	O.3H	N.2	-29,04	1,90	1,21	0,43	-1,24
$\text{H}_2\text{O} \cdots \text{Pyridin}$	O.3H	N.ar	-30,59	1,89	1,21	0,43	-1,30
$\text{MeOH} \cdots \text{Imidazol}$	O.3H	N.2	-34,17	1,89	1,26	0,61	-1,11
$\text{MeNH}_2 \cdots \text{H}_2\text{O}$	N.3H	O.3	-9,93	2,12	1,32	0,40	-1,09
$\text{MeNH}_2 \cdots \text{MeOH}$	N.3H	O.3	-11,79	2,16	1,35	0,37	-1,11
$\text{MeNH}_2 \cdots \text{MeOMe}$	N.3H	O.3	-14,69	2,16	1,27	0,29	-0,95
$\text{Me}_2\text{NH} \cdots \text{H}_2\text{O}$	N.3H	O.3	-10,89	2,12	1,32	0,49	-1,14
$\text{PhNH}_2 \cdots \text{H}_2\text{O}$	N.pl3H	O.3	-17,05	2,02	1,23	—	—
$\text{Pyrrolidin} \cdots \text{H}_2\text{O}$	N.3H	O.3	-11,72	2,11	1,32	0,49	-1,12
$\text{Imidazol} \cdots \text{H}_2\text{O}$	N.pl3H	O.3	-26,12	1,90	1,23	0,48	-1,15
$\text{MeNH}_2 \cdots \text{H}_2\text{CO}$	N.3H	O.2	-15,21	2,31	1,42	0,36	-1,11
$\text{CH}_2 = \text{NH} \cdots \text{H}_2\text{O}$	N.2H	O.3	-12,37	2,09	1,25	0,30	-0,79
$N\text{-MeAc} \cdots \text{H}_2\text{O}$	N.amH	O.3	-21,46	1,98	1,26	0,44	-1,14
$N\text{-MeAc} \cdots \text{MeOH}$	N.amH	O.3	-25,40	1,97	1,21	—	—
$N\text{-MeAc} \cdots \text{H}_2\text{CO}$	N.amH	O.2	-18,77	2,13	1,38	0,43	-1,10
$\text{MeNH}_2 \cdots \text{Imidazol}$	N.3H	N.2	-22,19	2,31	1,44	0,37	-1,11
	← C.2H	N.3		2,67	1,60	0,10	-1,00

Tabelle 2.6: Ergebnisse für wasserstoffverbrückte Komplexe mit Schwefel als Donor oder Akzeptor. Für weitere Erläuterungen siehe Tabelle 2.4.

Komplexe	Donor	Akzeptor	$\Delta E_{\text{MP2}}^{\text{CPC}}$	$d_{\text{H-B}}$	$d_{\text{BCP-B}}$	$q^{\text{AIM}}(\text{H})$	$q^{\text{AIM}}(\text{B})$
H ₂ O···EtSH	O.3H	S.3	-20,44	2,41	1,55	0,38	0,35
H ₂ O···MeSMe	O.3H	S.3	-22,70	2,36	1,52	0,38	0,34
H ₂ O···MeSSMe	O.3H	S.3	-17,11	2,45	1,57	0,38	0,15
H ₂ O···H ₂ CS	O.3H	S.2	-18,96	2,47	1,58	0,38	0,52
	← C.2H	O.3		2,53	1,47	0,02	-0,78
MeOH···MeSMe	O.3H	S.3	-23,54	2,37	1,54	0,40	0,34
MeOH···H ₂ CS	O.3H	S.2	-20,08	2,49	1,59	0,40	0,53
	← C.2H	O.3		2,58	1,50	0,02	-0,86
MeCOOH···MeSMe	O.3H	S.3	-33,78	2,29	1,50	0,44	0,37
MeSH···H ₂ O	S.3H	O.3	-9,47	2,41	1,37	0,58	-0,01
EtSH···H ₂ O	S.3H	O.3	-9,26	2,42	1,46	-0,01	-1,16

Tabelle 2.7: Ergebnisse für die wasserstoffverbrückten Komplexe des Testsatzes. Die Energien sind in kJ Mol^{-1} angegeben.

Komplex	Donor	Akzeptor	$\Delta E_{\text{MP2}}^{\text{CFC}}$	ΔE^{HBInc}	Abs. Fehler	Rel. Fehler
$\text{H}_2\text{O} \cdots \text{FCA}^4$	O.3H	O.3	-26,90	-24,21	2,69	%10
	← C.3H	O.3H				
$\text{MeOH} \cdots \text{THF}^5$	O.3H	O.3	-28,74	-26,25	2,49	% 9
$\text{FCA}^1 \cdots \text{H}_2\text{O}$	O.3H	O.3	-40,59	-46,47	5,88	%14
	← O.3H	O.3				
$\text{FCA}^1 \cdots \text{H}_2\text{O}$	O.3H	O.3	-45,91	-45,33	0,57	% 1
	← O.3H	O.3				
$\text{MeOH} \cdots \text{Me}_3\text{N}$	O.3H	N.3	-37,85	-35,92	1,93	% 5
$\text{NH}_3 \cdots \text{NH}_3$	N.3H	N.3	-12,17	-12,17	0,00	% 0
$\text{MeNH}_2 \cdots \text{MeNH}_2$	N.3H	N.3	-14,74	-18,33	3,59	%24
$\text{Me}_2\text{NH} \cdots \text{Me}_2\text{NH}$	N.3H	N.3	-18,44	-23,77	5,33	%29
$\text{Me}_2\text{NH} \cdots \text{N-MeAc}$	N.3H	N.pl3	-21,85	-23,79	1,94	% 9
$\text{N-MeAc} \cdots \text{N-MeAc}$	N.pl3H	O.2	-34,99	-34,20	0,79	% 2
$\text{Adenin} \cdots \text{H}_2\text{O}$	N.3H	O.3	-50,14	-47,64	2,50	% 5
	← O.3H	N.ar				
$\text{G} \cdots \text{C WC}^{67}$	N.pl3H	O.2	-127,45 ⁸	-79,85	-47,59	%37
	→ N.amH	N.2				
	← N.pl3H	O.2				
$\text{A} \cdots \text{T WC}^{67}$	N.pl3H	O.2	-67,20 ⁸	-45,24	-21,96	%33
	← N.amH	N.ar				
$\text{I} \cdots \text{C WC pl}^{67}$	N.amH	N.2	-99,65 ⁸	-55,98	-43,67	%44
	← N.pl3H	O.2				
$\text{U} \cdots \text{U pl}^{67}$	N.amH	O.2	-55,27 ⁸	-36,29	-18,97	%34
	← N.amH	O.2				
$2\text{-Py} \cdots 2\text{-AP}^{910}$	N.amH	N.ar	-66,61 ¹¹	-45,24	-21,37	%32
	← N.pl3H	O.2				
Formamid Dimer	N.amH	O.2	-63,60 ¹²	-35,64	-27,96	%44
	← N.amH	O.2				
DNA-M1 ¹³	N.pl3H	O.2	-43,12 ¹⁴	-45,24	2,11	% 5
	← N.amH	N.ar				
DNA-M2 ¹⁵	N.pl3H	O.2	-79,55 ¹⁴	-79,85	0,30	< % 1
	→ N.amH	N.2				
	← N.pl3H	O.2				

⁴ FCA ist *D*- α -Fucose⁵ THF ist Tetrahydrofuran⁶ A ist Adenin, C ist Cytosin, G ist Guanin, I ist Inosin, T ist Thymin, U ist Uracil⁷ WC steht für *Watson-Crick*-Struktur, pl für *planare Geometrie*⁸ Ref [101], MP2, aug-cc-VTZ Basis⁹ 2-Py steht für 2-Pyridoxin¹⁰ 2-AP ist 2-Aminopyridin¹¹ Ref [44], MP2, aug-cc-VTZ Basis¹² Ref [44], MP2, aug-cc-VQZ Basis¹³ Komplex von *N*-Methylen-Methan-Diamin und *N*-Formyl-Formamid¹⁴ Ref [2]¹⁵ Komplex von *N*-(Aminomethyl)-Formamid und (*E*)-*N*-(Aminomethylen)-Formamid

Chemisch diverses Scoring für FlexX

In diesem Kapitel stelle ich die Entwicklung und Validierung der Bewertungsfunktion AISCORE vor, die H-Brücken in Abhängigkeit des chemischen Kontextes berücksichtigt. Die Untersuchungen zur Vorhersagekraft dieser neuen Bewertungsfunktion sind in einer Arbeit zusammengefasst, die bei *Journal of Medicinal Chemistry* zur Veröffentlichung eingereicht wurde. [126].

AISCORE ist für das Docking-Programm FLEXX entwickelt worden, weshalb ich auf die Algorithmen und Besonderheiten von FLEXX in Kapitel 3.3 eingehe. Grundlage für AISCORE ist das LUDI-Modell, welches im Zusammenhang mit der FLEXX-Bewertungsfunktion FLEXXSCORE beschrieben wird (Kapitel 3.3.1).

Da der Code von FLEXX seit nunmehr 11 Jahren immer weiterentwickelt wird, sollte dieser soweit möglich unangetastet bleiben. Deswegen wird die chemische Diversität in der Bewertungsfunktion über neue molekulare Subgraphen und dazu gehörige Wechselwirkungsgeometrien implementiert. Die quantenchemischen *ab initio*-Ergebnisse aus Kapitel 2.4 sind die Basis für die diversen Wechselwirkungsenergie-Inkrementen. Die einzige Änderung am Code von FLEXX besteht in einem Algorithmus für die Behandlung von multifurkaten Wasserstoffbrücken (Kapitel 3.4.3).

Für die Entwicklung und Validierung war ein Satz von Protein-Ligand-Komplexen mit bekannter experimenteller Struktur und freien Bindungsenergie nötig. Seine Erstellung wird in Kapitel 3.4.4 beschrieben.

Ein Vergleich von AISCORE und der ursprünglichen Bewertungsfunktion von FLEXX findet sich in Kapitel 3.5.

3.1 Molekulares Docking

Molekulares Docking ist eine algorithmische Methode, um die dreidimensionale Struktur eines Komplexes aus einer Proteinbindetasche und einem Liganden vor-

herzusagen. [16] Die Freiheitsgrade des Docking-Problems sind die relative Orientierung des Liganden zum Rezeptor und die Konformationen, die die beteiligten Moleküle bei der Komplexierung einnehmen. Neben der Konformationsgenerierung ist auch die energetische Bewertung wichtig. Hierfür werden *Bewertungsfunktionen* eingesetzt, die auf Basis einer dreidimensionalen Komplexstruktur energetische Aussagen machen (siehe unten, Kapitel 3.2).

Die meisten Programme behandeln den Rezeptor als rigide und modellieren lediglich die konformationelle Beweglichkeit des Liganden. Moderne Docking-Programme sind heute in der Lage, die Geometrie von Protein–Ligand-Komplexen mit einer guten Genauigkeit vorherzusagen und dabei auch die Flexibilität des Liganden voll zu berücksichtigen.

Aktuelle Docking-Algorithmen unterscheiden sich z.B. darin, wie der Konformationsraum abgetastet wird, in der Beschreibung der molekularen Wechselwirkungen, wie die Platzierung des Ligand im Rezeptor bestimmt wird und natürlich in der Rechenzeit, die pro Ligand benötigt wird.

Bei den *Rigid-Body*-Methoden wird der Konformationsraum von Protein und Ligand fix gehalten. Dadurch reduziert sich das Docking-Problem darauf, diejenige relative Lage der beiden starren Körper zu finden, die der niedrigsten Energie entspricht. Die erste Docking-Software überhaupt war das Programm DOCK. [127] Es basiert auf der Idee, durch einen rein geometrischen Algorithmus einen starren Liganden in eine starre Proteinbindetasche einzuplatzieren.

Um unterschiedliche Konformationen eines Liganden beim Docking zu berücksichtigen, kann man die oben beschriebenen *Rigid-Body*-Methoden auf einen Satz von verschiedenen starren Konformationen des Liganden anwenden. Da die Anzahl der möglichen Konformere selbst bei kleinen Molekülen mit wenigen konformationellen Freiheitsgraden schnell mehrere hundert Millionen erreichen kann, sind diesem Ansatz enge Grenzen gesetzt.

Ein sehr eleganter Weg, die konformationelle Flexibilität schon direkt beim Docken zu berücksichtigen, ist die *Fragmentierung* des Liganden. Der Ligand wird dabei in verschiedene *Fragmente* zerteilt, die jedes für sich als starr behandelt werden.

3.2 Bewertungsfunktionen

Der Schwachpunkt aktueller Programme sind nicht die Docking-Algorithmen zur Berechnung der geometrischen Struktur eines Komplexes selbst, sondern ist die

Ungenauigkeit der *Bewertungsfunktionen*, die benutzt wird, um die Wechselwirkung von Ligand und Protein abzuschätzen. Diese Bewertungsfunktionen sind die Zielfunktionen, deren Minimum beim Docking gefunden werden soll. [128]

Typische Vertreter von Docking-Programmen erzeugen in aller Regel einen Satz von möglichen Bindungsmodi. Die Bewertungsfunktion hat deswegen zwei Funktionen:

- Während des Dockings werden verschiedene Orientierungen des *selben* Liganden im *selben* Rezeptor bewertet. Die Bewertungsfunktion muss die nativ-ähnlichste Orientierung unter allen anderen identifizieren.
- Für diese Orientierung soll die Bewertungsfunktion einen Wert liefern, der der experimentellen Bindungsenergie nahe kommt.

Selbst, wenn das Docking das richtige Konformer und die richtige Platzierung vorhersagt, versagt der ganze Docking-Ansatz, wenn die Bewertungsfunktion nicht korrekte Platzierungen von falschen unterscheiden kann und somit „echte“ Liganden nicht identifiziert werden können. [129]

Aktuelle Bewertungsfunktionen sind noch weit davon entfernt, die freie Bindungsenergie akkurat berechnen zu können. Die Bindung eines Liganden im Komplex ist durch eine Kombination von *enthalpischen* und *entropischen* Effekten bestimmt. Das ist oft ein konzeptionelles Problem, denn die meisten Bewertungsfunktionen sind mehr auf die Erfassung von enthalpischen denn entropischen Effekten ausgerichtet. Die Entropie wird in der Regel durch sehr starke Vereinfachung approximiert. In FLEXXSCORE z.B. wird die Anzahl an Torsionswinkel zur Beschreibung des *Entropieverlustes* während der Komplexierung benutzt. Schwierig ist es auch *Solvations-Effekte* zu quantifizieren. Entweder werden hier größtenteils Abschätzungen gemacht, oder diese Beiträge werden direkt ganz ignoriert.

Zusätzlich müssen die Bewertungsfunktionen recht fehlertolerant bezüglich kleiner geometrischer Abweichungen sein, da schnelle Docking-Algorithmen die Struktur eines Liganden im Komplex nur sehr grob berechnen. Eine vollständige Geometrieoptimierung wäre jedoch zu aufwendig, als dass ein solches Prozedere dann noch schnell genug wäre, um bei *High-Throughput*-Studien anwendbar zu sein.

Zurzeit wird an zwei Arten von Bewertungsfunktionen geforscht: *empirische* Bewertungsfunktionen und *wissensbasierte* Bewertungsfunktionen.

3.2.1 Empirische Bewertungsfunktionen

Empirische Bewertungsfunktionen [130] versuchen die maßgeblichen Elemente der freien Bindungsenergie als eine Summe verschiedener Terme auszudrücken. Die Wichtungen der einzelnen Terme werden entweder an experimentelle Daten durch multivariate Regressionsmethoden angepasst oder durch physikalische Erwägungen bestimmt.

Die erste Funktion dieser Art wurde von BÖHM für das *de novo*-Programm LUDI [34, 35] entwickelt. Eine Weiterentwicklung davon kommt in FLEXX zur Anwendung und bildet die Grundlage für diese Arbeit. Deswegen wird in Kapitel 3.3.1 weiter unten noch genauer darauf eingegangen.

Die Leistung einer empirischen Bewertungsfunktion hängt stark von Qualität und Quantität der Daten ab, an die die verschiedenen Parameter angepasst wurden. Eines der Hauptprobleme bei der Entwicklung einer Bewertungsfunktion ist die Verfügbarkeit *diverser* und *akkuratere* experimenteller Daten. Ein Weg, dieses Problem zu beheben, ist es, die experimentellen Daten durch Ergebnisse quantenchemischer *ab initio*-Rechnungen zu erweitern.

3.2.2 Wissensbasierte Bewertungsfunktionen

Wissensbasierte Bewertungsfunktionen werden aus statistischen Analysen von experimentellen Daten abgeleitet.

Die Grundidee hinter wissensbasierten Bewertungsfunktionen ist die Annahme, dass Eigenschaften, die häufig in Strukturen beobachtet werden, energetisch günstig sind. Bei den drei am weitesten verbreiteten wissensbasierten Bewertungsfunktionen sind diese Eigenschaften Atom–Atom-Abstände. [131, 132, 133, 134, 135] Aus einem Satz von experimentellen Daten wird für jedes mögliche Paar von Atomtypen das Vorkommen in Abhängigkeit vom Abstand gezählt. Die resultierenden Histogramme werden dann über die inverse BOLTZMANN-Gleichung in eine Energiefunktion konvertiert. Für die Bewertung einer Ligand–Rezeptor-Konformation werden dann die Werte dieser Energiefunktion für alle interatomaren Paare aufaddiert.

3.3 FlexX

FLEXX ist ein Computerprogramm zur Voraussage von Protein–Ligand–Wechselwirkungen. FLEXX kombiniert einen sehr schnellen Docking-Algorithmus auf Basis des *inkrementellen Aufbaus* mit einer einfachen Bewertungsfunktion.

3.3.1 Wechselwirkungen in FlexX

Als Modell für die molekularen Wechselwirkungen wurde bei FLEXX das LUDI-Modell [34, 35, 36, 136, 37] adaptiert. Eine *Wechselwirkungsgruppe* besteht hierbei aus einem *Wechselwirkungszentrum* und einer *Wechselwirkungsfläche* (Abbildung 3.1).

Zwei Wechselwirkungsgruppen verschiedener Moleküle gehen eine Bindung ein, wenn das Wechselwirkungszentrum des einen Moleküls ungefähr auf der Wechselwirkungsfläche des anderen liegt und *vice versa* (Abbildung 3.1). Jedoch müssen die beiden Wechselwirkungsgruppen *kompatible* Wechselwirkungstypen sein, z.B. eine Gruppe repräsentiert einen Wasserstoffbrückendonator und die andere ist ein Wasserstoffbrückenakzeptor.

Jeder Wechselwirkungsgruppen ω ist ein Wechselwirkungsenergie-Inkrement ΔG_ω zugewiesen. Für eine bestehende Bindung besteht die *Match Score* ΔG_{ia} aus der Summe aller an dieser Bindung beteiligten Inkremente ΔG_ω des Typs ia.

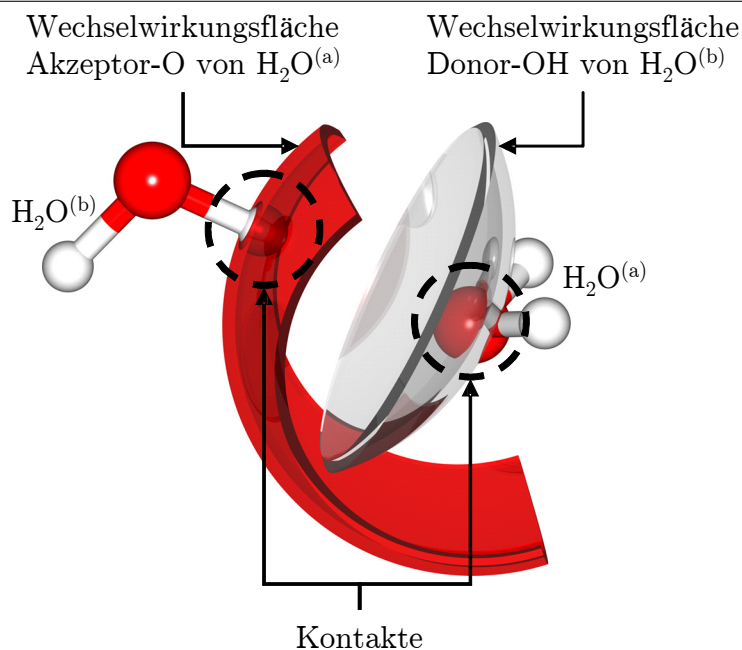


Abbildung 3.1. LUDI-Wechselwirkungsmodell am Beispiel des Wasserdimers.

Die Bindung kann eine Wasserstoffbrücke sein ($ia = hb$), eine ionische Bindung ($ia = io$), oder es kann sich um eine lipophile Wechselwirkung handeln ($ia = lipo$). Für FLEXX wurde das Modell um Terme für aromatische ($ia = aro$) und metallische ($ia = metal$) Wechselwirkungen erweitert. Außerdem wurden hydrophobe Kontakte eingeführt. [30]

Die freie Bindungsenergie eines ganzen Komplexes ist die Summe aller individuellen *Match Scores* ΔG_{ia} und des *Entropieverlustes*. [36]

$$\begin{aligned}
\Delta G &= \Delta G_0 + \Delta G_{\text{rot}} N_{\text{rot}} \\
&+ \Delta G_{\text{hb}} \sum_{\text{HBonds}} f(\Delta r, \Delta \alpha) \\
&+ \Delta G_{\text{io}} \sum_{\text{ion.int.}} f(\Delta r, \Delta \alpha) \\
&+ \Delta G_{\text{aro}} \sum_{\text{arom.int.}} f(\Delta r, \Delta \alpha) \\
&+ \Delta G_{\text{metal}} \sum_{\text{metal.int.}} f(\Delta r, \Delta \alpha) \\
&+ \Delta G_{\text{lipo}} \sum_{\text{lipo.int.}} f_1(\Delta r) \\
&+ \Delta G_{\text{ambig}} \sum_{\text{ambig}} f_1(\Delta r) \\
&+ \Delta G_{\text{clash}} \sum_{\text{clash}} f_2(\Delta r)
\end{aligned} \tag{3.1}$$

mit

$$f(\Delta r, \Delta \alpha) = f_1(\Delta r) \cdot f_1(\Delta \alpha) \tag{3.2}$$

$$\Delta r = r - r_0 \tag{3.3}$$

$$\Delta \alpha = \alpha - \alpha_0 \tag{3.4}$$

Beim LUDI-Modell beschreiben ΔG_0 und $\Delta G_{\text{rot}} N_{\text{rot}}$ den Entropieverlust des Liganden. Der erste Term trägt dem Einfrieren translatorischer und rotatorischer Freiheitsgrade des Liganden Rechnung. Der zweite Term berücksichtigt die Festsetzung der freidrehbaren Einfachbindungen des Liganden im Komplex. ΔG_{hb} und ΔG_{io} , ΔG_{aro} und ΔG_{metal} sind die *Match Scores* ΔG_{ia} jeweils von Wasserstoffbrücken, ionischen, aromatische und metallische Wechselwirkungen. Die entsprechenden Energieinkremente ΔG_{ω} sind an experimentelle Bindungsenergien angepasst. Bei geladenen Wasserstoffbrücken wird ein zusätzlicher Faktor

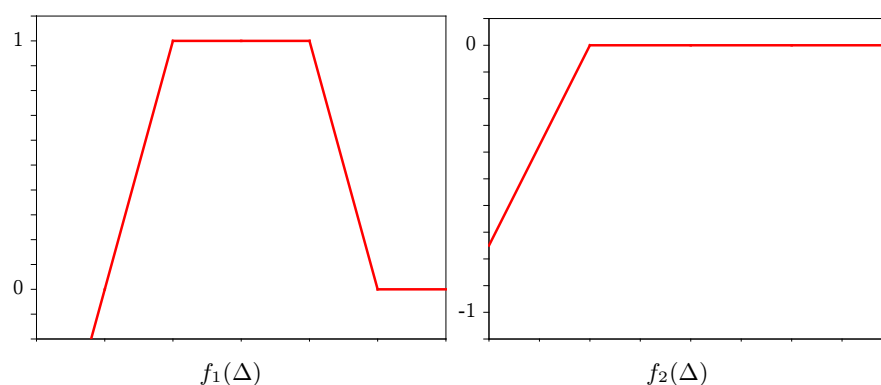


Abbildung 3.2. Die Straffunktionen von LUDI.

von 1,667 für jeden Kontakt mit einer Gruppe, deren Ladung einen bestimmten Grenzwert überschritten hat, auf ΔG_{hb} multipliziert.


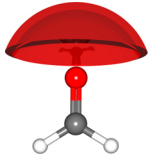
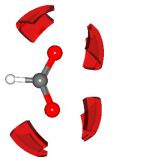
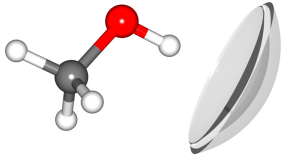
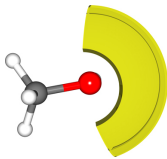
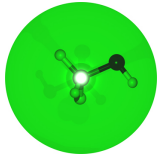
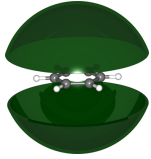
Die letzten drei Terme mit ΔG_{lipo} , ΔG_{ambig} und ΔG_{clash} schließlich bewerten Atom–Atom-Kontakte zwischen Rezeptor und Ligand. ΔG_{lipo} berücksichtigt dabei nur Paare von unpolaren Atomen, ΔG_{ambig} nur Paare aus einem polaren und einem unpolaren Atom. ΔG_{clash} bestraft Überlappungen von Protein und Ligand.

All diese Bindungen haben typische Bindungslängen, Wasserstoffbrücken sind sogar sehr stark abstandabhängig und hochgradig gerichtet (siehe Kapitel 2.1). Um diesem Umstand Rechnung zu tragen, benutzt das LUDI-Modell eine Straffunktion $f(\Delta r, \Delta \alpha)$, die Abweichungen Δr vom idealen Bindungsabstand r_0 und Abweichungen $\Delta \alpha$ vom idealen Bindungswinkel α_0 bestraft. f_1 und f_2 sind in Abbildung 3.2 dargestellt.

Alle Wechselwirkungstypen, die in FLEXX implementiert sind, sind in Tabelle 3.1 aufgelistet. Um funktionellen Gruppen eine passende Wechselwirkungsgeometrie zuzuweisen, wird in einer *Graphdarstellung* des Moleküls (mit den SYBYL-Atomen als Knoten und den Bindungen als Kanten) nach charakteristischen Subgraphen gesucht. In einer Liste sind die entsprechenden Subgraphen und die passenden Wechselwirkungsgeometrien abgelegt. Nach dem Einlesen des Moleküls werden dann entsprechend der gefundenen Subgraphen die Wechselwirkungsgeometrien zugewiesen.

In der ursprünglichen Version von FLEXXSCORE sind alle Wechselwirkungsenergie-Inkrementen ΔG_{ω} *identisch* für alle Wasserstoffbrückendonoren- und akzeptoren ($\Delta G_{\omega} = -2,35 \text{ kJ Mol}^{-1}$), völlig unabhängig von ihrem jeweiligen chemischen Kontext. Wie jedoch die quantenchemischen *ab initio*-Ergebnisse aus Kapitel 2.4 gezeigt haben, variiert die Wechselwirkungsenergie von Wasserstoffbrücken in einem sehr weiten Bereich.

Tabelle 3.1. Wechselwirkungen in FLEXX. Die Wechselwirkungsenergie-Inkremente ΔG_ω sind in kJ Mol^{-1} angegeben, die idealen Bindungsradii r_0 in Å.

	Wechselwirkungs-Typ	ia	ΔG_ω	r_0
	H-Brücken-Akzeptor	hb io	-2,35 -4,15	1,9
	sp ² -H-Akzeptor	hb io	-2,35 -4,15	1,9
	COO ⁻ -H-Akzeptor	io	-4,7	1,9
	H-Brücken-Donor	hb io	-2,35 -4,15	1,9
	Metall-Akzeptor	metal	-2,35	2,0
	lipophiler Kontakt	lipo	-0,35	4,0 – 4,8
	aromatische Wechselwirkung arom		-0,35	4,5

Auch wurde bereits gezeigt, dass Schwefel ein H-Brückenakzeptor von vergleichbarer Stärke wie Sauerstoff ist. Dieser Tatsache wird im Bewertungsschema von FLEXX nicht Rechnung getragen, da Schwefel hier erst gar nicht als Wechselwirkungszentrum behandelt wird.

3.3.2 Flexibles Docking in FlexX

FLEXX berücksichtigt die Flexibilität des Liganden, in dem es ihn in Fragmente zerteilt und diese durch *inkrementellen Aufbau* im Rezeptor zu einer bindenden Konformation zusammensetzt (Abbildung 3.3). Der Ligand wird an allen azyklischen rotierbaren Einfachbindungen zerteilt. Einfachbindungen zu Methyl- und Amingruppen werden als frei rotierbar definiert. Ringsysteme werden nicht zerlegt.

Als Ankerpunkt für den inkrementellen Aufbau, werden ein oder mehrere Basisfragmente benötigt. Ein geeignetes Basisfragment muss mindestens zwei, besser drei Wechselwirkungszentren besitzen und nur wenige Konformationen. Alle geeigneten Basisfragmente werden unabhängig voneinander in das aktive Zentrum des Proteins platziert und anschließend, ausgehend von allen Platzierungsvorschläge, die jeweils übrigen Fragmente zu verschiedenen Konformationen des Liganden inkrementell im aktiven Zentrum aufgebaut.

Das Platzieren der Basisfragmente geschieht durch einen *Pose Clustering*-Algorithmus. [137] *Pose Clustering* kommt aus der Mustererkennung und wurde ursprünglich entwickelt, um Modelle in Szenen mit unbekannter Kameraposition zu erkennen. Der für FLEXX angepasste Algorithmus sucht Übereinstimmungen von jedem Dreieck aus Wechselwirkungsgeometrien des Liganden mit jedem Dreieck aus Wechselwirkungsgeometrien des Proteins (Abbildung 3.4).

Jedes neue Fragment wird der bereits aufgebauten Teilstruktur mit jeweils mehreren unterschiedlichen Torsionswinkeln an azyklischen Einfachbindungen hinzugefügt. Diese Winkel sind ein diskreter Satz von niedrig-energetischen Torsionswinkel aus der MIMUMBA-Datenbank. [138, 139] Für die Generierung unterschiedlicher niedrig-energetischer Konformationen von Ringsystemen ruft FLEXX das externe Programm CORINA [140, 141, 142, 143] auf.

Nach jedem Anfügen eines Fragmentes in all seinen möglichen Torsionswinkeln liegt ein Satz von Lösungen vor. Das nächste Fragment müsste jetzt an alle Elemente dieses Lösungssatzes wieder mit unterschiedlichen Torsionswinkeln addiert werden. Um den inkrementellen Aufbau noch effizient fortsetzen zu können, wird nicht der volle Lösungssatz von einer Iteration in die nächste übernommen.

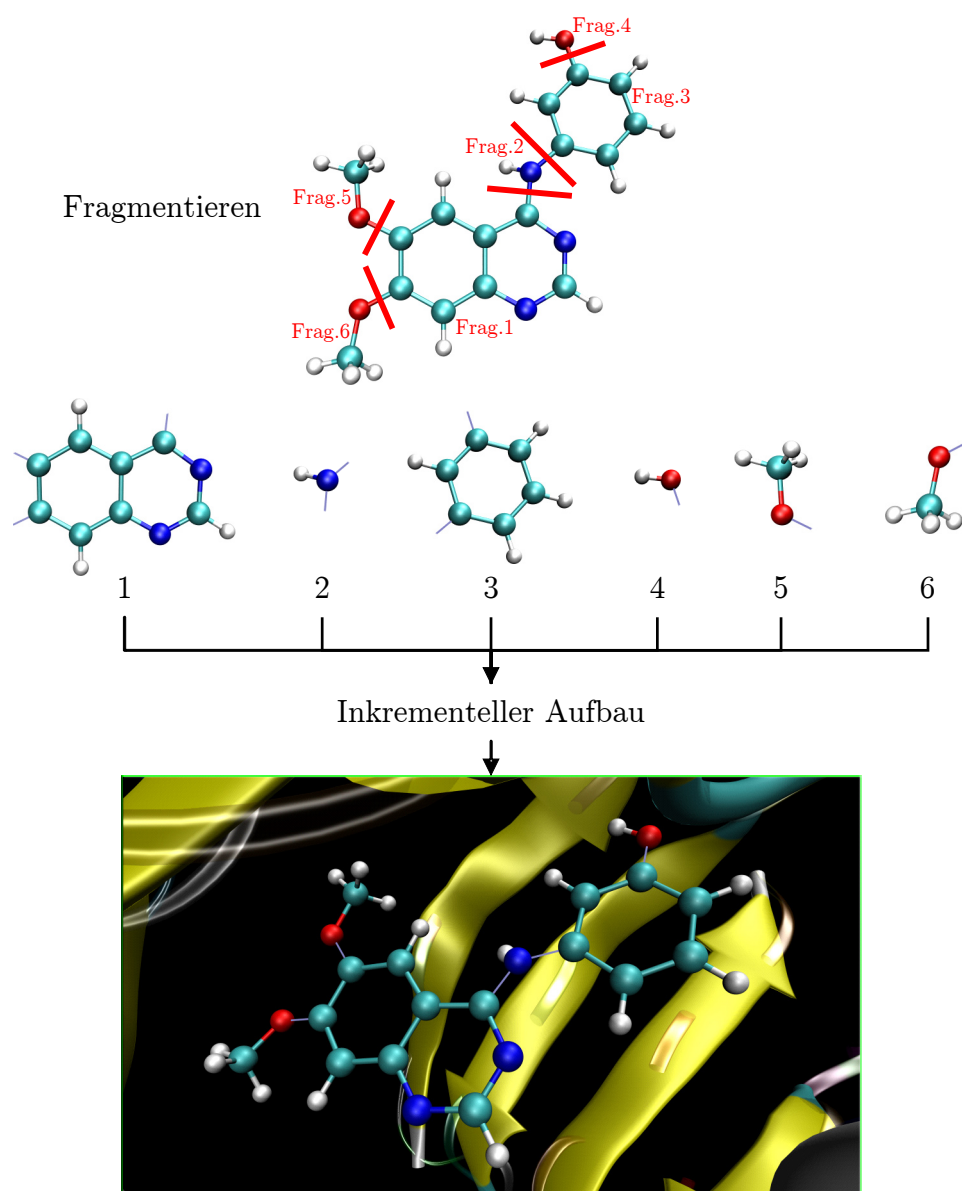


Abbildung 3.3. Inkrementeller Aufbau

FLEXX verwendet eine *Greedy*-Strategie, bei der immer nur die besten k Lösungen im Lösungssatz verbleiben. Zusätzlich werden die 100 besten Lösungen pro Basisfragment mit übernommen.

Durch dieses Konformations-Modell können für ein Molekül von einer Größe, die typisch für pharmazeutische Leitstrukturen ist, einige Millionen Konformationen generiert und bewertet werden.

Durch die gleichzeitige Verwendung von verschiedenen Basisfragmenten als Ausgangspunkt für unterschiedliche inkrementelle Konstruktionen tut sich ein subtiles Problem auf. In einem Lösungssatz sind alle Teilplatzierungen ausge-

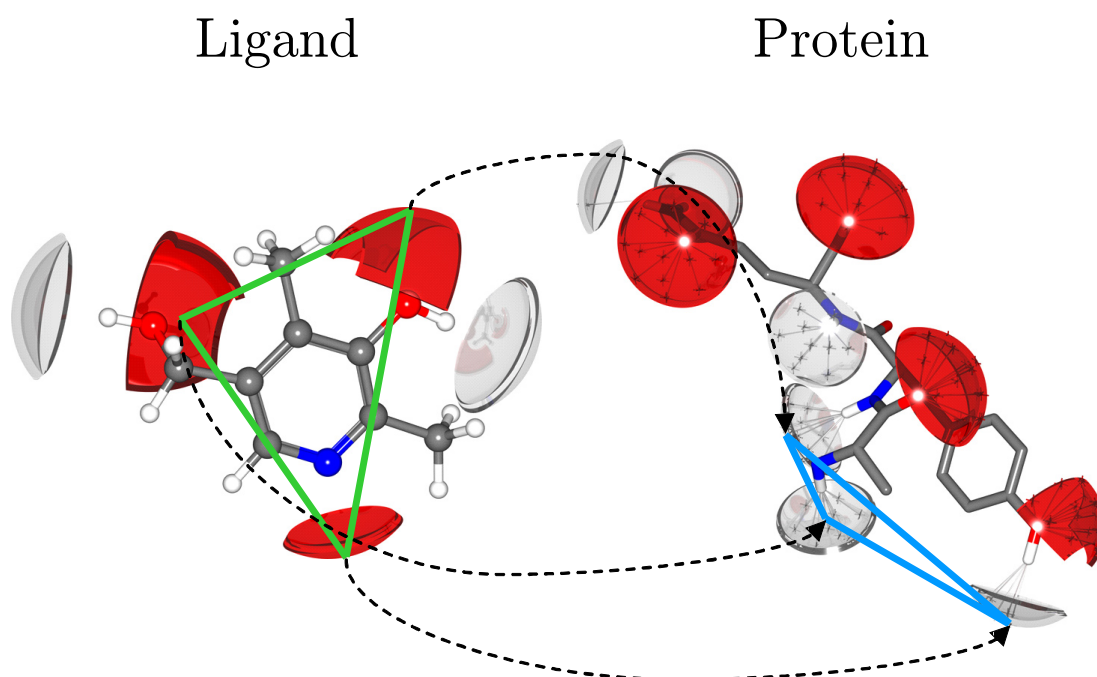


Abbildung 3.4. *Pose Clustering* in FLEXX. Ein Dreieck aus Wechselwirkungsgeometrien des Liganden (grün) wird mit einer geeigneten Transformation mit einem Dreieck aus kompatiblen Wechselwirkungsgeometrien des Proteins (blau) überlagert. Aus algorithmischen Gründen werden die Wechselwirkungsflächen des Proteins durch diskrete Punkte angenähert.

hend von allen Basisfragmenten enthalten. Da aber die Teilplatzierungen unterschiedlich lang sein können und auch unterschiedliche *Matches* haben können, sind deren Energien nicht direkt vergleichbar. Um dennoch ein Kriterium für die *Greedy*-Strategie zu haben, schätzt FLEXX die Energie, die von den noch nicht platzierten Fragmenten zu erwarten ist, ab. Dafür wird geschaut, ob das betreffende Fragment in einer Teilplatzierung eines Basisfragmentes bereits bewertet wurde. Die *Match Score* für diese Platzierung wird dann zur Abschätzung der zu erwartenden Wechselwirkungsenergie für dieses Fragment verwendet.

Ist das Fragment in keiner Platzierung bisher vorgekommen, wird dessen potenzieller Beitrag durch Aufsummieren aller möglichen *Match Scores* aller Wechselwirkungsgruppen des Fragments angenommen.

3.4 AIScore – Chemisch diverses Scoring

Im Folgenden sollen die Ideen hinter AIScore vorgestellt werden.

3.4.1 Chemisch diverse Subgraphen

Eine Verbesserung des bestehenden FLEXXSCORE muss eine größere chemische Diversität zulassen. Somit liegt es auf der Hand, mehr unterschiedliche molekulare Subgraphen und somit auch mehr verschiedene Wechselwirkungsgeometrien zu implementieren.

In dieser Arbeit wurde in Kapitel 2.3.1 bereits ein Verfahren vorgestellt, kombinatorisch unterschiedliche chemische Umgebungen für Wasserstoffbrückendonoren und -akzeptoren zu generieren. Genau dieselben Baumstrukturen habe ich verwendet um molekulare Subgraphen für FLEXX zu generieren, mit denen ich den Wechselwirkungszentren in unterschiedlichen chemischen Kontexten auch unterschiedliche Wechselwirkungsflächen und Wechselwirkungsenergien zuordnen kann.

Jeder einzelne dieser Subgraphen weist einer funktionellen Gruppe eine unterschiedliche Wechselwirkungsgeometrie zu. So kann berücksichtigt werden, dass Wasserstoffbrückenenergien eben so unterschiedlich sein können. Dabei wurden den Subgraphen von allgemeineren Strukturen (oben in der Baumstruktur) niedrigere Prioritäten zugewiesen als solche, die hochspezifische chemische Kontexte repräsentieren. Einer funktionellen Gruppe wird diejenige Wechselwirkungsgeometrie zugewiesen, deren molekularer Subgraph die höchste Priorität besitzt.

Abbildung 3.5 veranschaulicht die Zuweisung der Wechselwirkungsgeometrien einmal, wie sie das originale FLEXXSCORE vornimmt, und einmal, wie durch das neue AIScore geschehen. Man erkennt deutlich, wie die Diversität der Wechselwirkungstypen erhöht wird.

3.4.2 Optimieren der Parameter

Die Vielfalt der Subgraphen erlaubt eine feinere Differenzierung der Wechselwirkungsgruppen und somit auch der Wechselwirkungsenergie-Inkrementen der Bewertungsfunktion. Dafür müssen nun neue Parameter für Akzeptor- und Donorkontakte angepasst werden.

In Kapitel 2.4.2.2 wurde ein Verfahren vorgestellt, um die Gesamtwechselwirkungsenergie einer singulären Wasserstoffbrücke in additive Donor- und Akzeptorenergie-Inkrementen für das Wasserstoffbrücken-Inkrementensystem HBINC

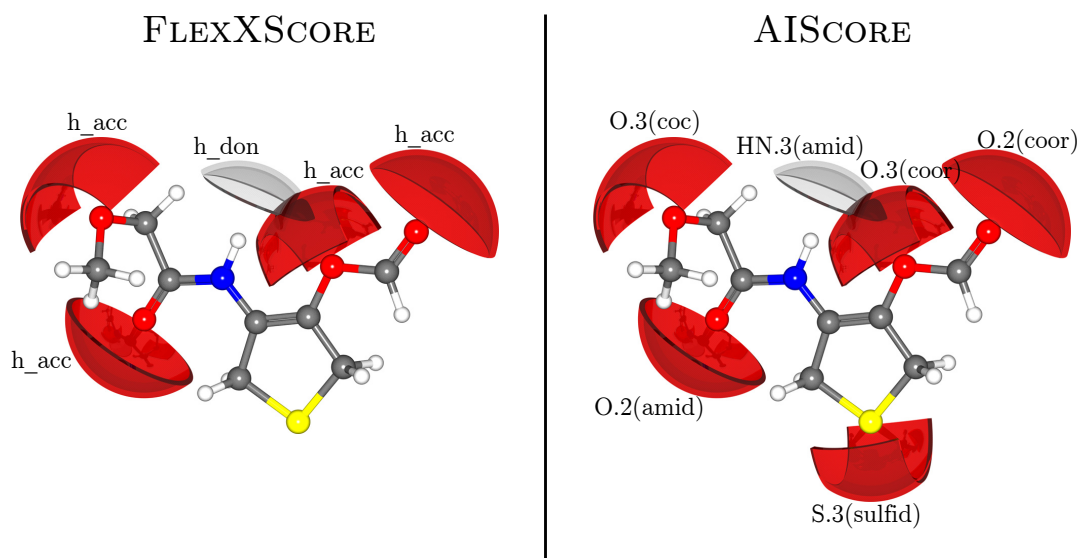


Abbildung 3.5. Vergleich der Wechselwirkungsgeometrien, wie sie bei diesem hypothetischen Molekül von FLEXSCORE und von AISCORE zugewiesen werden.

aufzuteilen. Im Prinzip können diese Inkremente von HBINC direkt zur Berechnung der *Match Score* für eine gegebene Wechselwirkung dienen. Jedoch sind diese Inkremente von *ab initio*-Rechnungen *in vacuo* abgeleitet worden – Solvatationseffekte sind hier nicht berücksichtigt.

In der Realität sind sowohl das Protein als auch der Ligand solvatisiert und müssen erst teilweise desolvatisiert werden bevor der Protein–Ligand-Komplex gebildet werden kann. Die Energie die dafür benötigt wird, führt typischerweise zu einer Verringerung der Gesamtkomplexbindungsenergie.

In dieser Arbeit wird angenommen, dass diese Verringerung für ein gegebenes Lösungsmittel (z.B. Wasser) linear mit der Wasserstoffbrückenstärke skaliert. Mit dieser Annahme können die Solvatationseffekte auf die Wasserstoffbrückendonoren und Wasserstoffbrückenakzeptoren mit einem einzigen Faktor f_s , der alle *in vacuo*-Inkremente skaliert, modelliert werden. Dieser Faktor wird an die experimentellen Bindungsaffinitäten des Trainingsatzes (siehe Kapitel 3.4.4) angepasst.

Die Bindungsenergie von starken Wasserstoffbrücken in konjugierten molekularen Systemen, die durch RAHBs charakterisiert sind (z.B. DNA-Basen, siehe Kapitel 2.1 und 2.4.3), werden durch das additive HBINC systematisch unterschätzt. Aus diesem Grund und wegen der Wichtigkeit dieser Wasserstoffbrücken in biologischen Systemen, wurden spezielle Wechselwirkungsenergien für solche funktionellen Gruppen auch an die experimentellen Werte angepasst.

Die Wechselwirkungsenergien von Salzbrücken können mit quantenchemischen Methoden *in vacuo* nicht berechnet werden, da *Zwitterionen* in Abwesenheit eines polaren Lösungsmittels nicht stabil sind. Deshalb wurden die Wechselwirkungsenergien von vier Typen von ionischen Wechselwirkungen (siehe Tabelle 3.5) ebenfalls optimiert, um die experimentellen Daten des Komplexsatzes wiedergeben zu können.

Lipophile Wechselwirkungen sind erheblich schwächer als Wasserstoffbrücken und ionische Wechselwirkungen. Ihr Ursprung liegt hauptsächlich im *entropischen Effekt* – die entsprechenden Gruppen werden vom Lösungsmittel ausgeschlossen. Um diese Effekte *ab initio* zu berechnen, wären explizit ein Lösemittel und eine Konformationsmittelung nötig. Mit heutigen Computer-Ressourcen ist das für den gesamten Datensatz außerhalb unserer Reichweite.

Obwohl lipophile Wechselwirkungen so schwach sind, dürfen sie nicht vernachlässigt werden, da sie in biologischen Systemen reichlich vorhanden sind. Da AISCORE mit völlig anderen Wasserstoffbrückenstärken als FLEXX arbeitet, wurden zwar die aliphatischen und aromatischen Wechselwirkungsgeometrien

Tabelle 3.2. Parameter, die optimiert werden.

Name	Beschreibung	Häufigkeit
scaling		
f_s	Solvatationsfaktor	482
Spezielle Wasserstoffbrücken		
O.2(peptide)	sp ² -Akzeptor-O im Peptid-Gerüst	79
O.2(po)	sp ² -Akzeptor-O mit einer partiellen Ladung an einem P	31
N.2(HIS)	aromatisches N im Imidzolring von Histidin	2
HN.3(DNA)	H-Donor des Amid-N in Guanin-ähnlichen Ringen	5
HN.3(peptide)	H-Donor des Amid-N im Peptid-Gerüst	108
Ionische Wechselwirkungen		
O.co2(coo-)	negativ geladenes O in Carboxylatgruppen	151
O.co2(po)	negativ geladenes O an einem Phosphor	15
HN+	positiv geladenes Donor-N	12
HN+(ARG)	positiv geladener H-Donor von Arginin	78
Metallische Wechselwirkungen		
metal_d	Übergangsmetall-Atome	5
metal_s	Alkali- und Erdalkali-Atome	4
metal_sp2	sp ² -Metall-Akzeptor	5
metal_ arom	aromatischer Metall-Akzeptor	3
Lipophile Wechselwirkungen		
arom1	π -System eines aromatischen Rings	636
arom2	C-Atom eines aromatischen Rings	108
CH	CH-Gruppe beliebiger Hybridisierung	528

übernommen, ihre Wechselwirkungsenergien hingegen wurden zusammen mit den anderen Parametern an die experimentellen Daten des Komplexsatzes angepasst.

Grundsätzlich sind auch Metall–Ligand-Wechselwirkungen mit quantenchemischen *ab initio* Methoden berechenbar, wenn ein Großteil der Elektronenkorrelation mitberücksichtigt wird (siehe Kapitel 4.2.1). Jedoch neigen Metallatome und -ionen dazu, mit multiplen Liganden gleichzeitig zu komplexieren und eine sinnvolle Aufteilung der Gesamtwechselwirkungsenergie ist nicht mehr offensichtlich. Deswegen wurden auch diese Inkremente an die experimentellen Werte gefittet.

Um die oben beschriebenen Parameter, die nicht durch *ab initio*-Rechnungen bestimmt wurden, sowie den Solvatationsfaktor f_s an die experimentellen freien Bindungsenergien der 100 Komplexe im Protein–Ligand-Datensatz anzupassen, wurde folgendes Optimierungsverfahren gewählt:

Der Ligand wird in der Bindungsmode bewertet, die er in der Kristallstruktur einnimmt (Referenzstruktur). Für alle Komplexe des Trainingssatzes wird die lineare Korrelation r^2 zwischen der Bewertungsfunktion und den experimentellen Bindungsenergien als Funktion der freien Parameter berechnet und r^2 bezüglich dieser Parameter maximiert.

$$\begin{aligned}
 r^2(\Delta G_{\text{HN.3(peptid)}}, \Delta G_{\text{O.2(peptid)}}, \Delta G_{\text{O.2(po)}}, \Delta G_{\text{HN.3(DNA)}}, \Delta G_{\text{N.2(HIS)}}, \\
 \Delta G_{\text{O.co2(coo-)}}, \Delta G_{\text{HN+(ARG)}}, \Delta G_{\text{O.co2(po)}}, \Delta G_{\text{HN+}}, \\
 \Delta G_{\text{metal_sp}^2}, \Delta G_{\text{metal_d}}, \Delta G_{\text{metal_s}}, \Delta G_{\text{metal_arom}}, \\
 \Delta G_{\text{arom1}}, \Delta G_{\text{CH}}, \Delta G_{\text{arom2}}, \\
 f_s) \stackrel{!}{=} \text{Max.}
 \end{aligned} \tag{3.5}$$

Eine Aufstellung, welche Parameter optimiert werden, findet sich in Tabelle 3.5.

Die Optimierung startet mit zehn schnellen *Grid*-Suchläufen auf einem volldimensionalen Gitter mit einer Schrittweite von $0,05 \text{ kJ Mol}^{-1}$ und einem zufälligen Startvektor. So wird eine Abschätzung für das globale Maximum der linearen Korrelation erhalten. Anschließend wird diese Abschätzung mit einem *konjugierten Gradientenverfahren* verfeinert. [144] Die ersten Ableitungen von r^2 werden dabei numerisch bestimmt.

Es werden nur die Energie-Inkremente ΔG_ω der Wechselwirkungsgeometrien während der Optimierung geändert. Alle anderen Parameter von der FLEXX-Bewertungsfunktion (z.B. für die Abschätzung des Entropieverlustes oder die

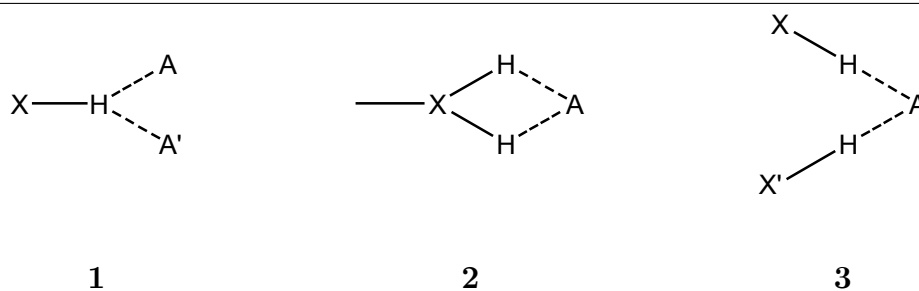


Abbildung 3.6. Verschiedene Arten von bifurkaten Wasserstoffbrücken.

Clash-Inkrementen) sind so belassen worden wie in der Originalversion von FLEXX-SCORE.

3.4.3 Mehrfach gegabelte Wasserstoffbrücken

Mehrfach gegabelte Wasserstoffbrücken, die ein Atom gemeinsam haben, wurden 1939 von ALBRECHT und COREY [145] das erste Mal für möglich gehalten. Die Autoren schlugen eine „bifurkate Bindung“ (siehe **1** in Abbildung 3.6) vor, um die Kristallstruktur von α -Glycin zu erklären. Später wurden dann der Terminus „bifurkate Wasserstoffbrücke“ auch für andere Konfigurationen verwendet (siehe **2** und **3** in Abbildung 3.6), die z.B. in bestimmten Wasserdimeren oder -trimeren vorkommen. Gegabelten H-Brücken werden heute zur Erklärung einer großen Anzahl biologischer Strukturen und vieler Wechselwirkungen in ihnen benutzt und sie finden sich in ca. 70% der Kristallstrukturen von zwitterionischen Aminosäuren.

FLEXX behandelt solche Wasserstoffbrücken nicht gesondert. Es werden alle Wechselwirkungszentren, die auf derselben kompatiblen Wechselwirkungsfläche liegen (nach Skalierung durch $f(\Delta r, \Delta \alpha)$) voll zur Gesamtenergie hinzuaddiert. In Extremfällen wird eine gegabelte H-Brücke wie mehrere vollständige, separate Kontakte bewertet.

Um zu illustrieren, wie wichtig eine gesonderte Behandlung bifurkater Wasserstoffbrücken ist, wurden quantenchemische *ab initio*-Rechnungen an zwei Modellsystem durchgeführt. Methoden und Basissätze sind dabei die gleichen wie bei den Rechnungen in Kapitel 2.3. Ein Vergleich der Ergebnisse von Formaldehyd komplexiert einmal mit der *trans*-Form von 1,2-Ethendiol (eine H-Brücke, Abbildung 3.7 *links*) und einmal mit der *cis*-Form (bifurkate H-Brücke, Abbildung 3.7 *rechts*) zeigt, dass eine gegabelte Wasserstoffbrücke eine deutliche geringen Bindungsenergie liefert als zwei separate H-Brücken.

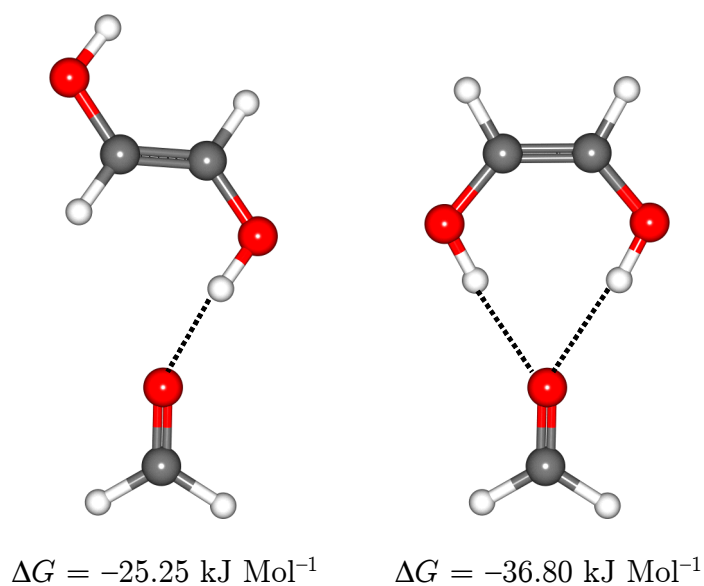


Abbildung 3.7. MP2-Wechselwirkungsenergien von Formaldehyd komplexiert mit *trans*-1,-Ethanediol (*links*) und mit *cis*-1,2-Ethanediol (*rechts*).

Die undifferenzierte Behandlung bi- und multifurkater Wasserstoffbrücken führt in FLEXX dazu, dass während des inkrementellen Aufbaus solche partiellen Positionierungen der Liganden mit vielen H-Brücken unrealistisch hoch bewertet und damit bevorzugt werden. Auf der anderen Seite kommt es selbst bei korrekter Geometrie des Komplexes so zu einer Überschätzung der Bindungsenergie.

Um multifurkate Wasserstoffbrücken in FLEXX berücksichtigen zu können, wird die Annahme gemacht, dass jede Wechselwirkungsgeometrie ihre Wechselwirkungsenergie immer nur *einmal* zur Gesamtbindungsenergie beisteuern kann, unabhängig davon, wie viele Kontakte sie in einem gegebenen Komplex hat. Hat ein Wechselwirkungszentrum mit der *match score* E_{match} also n -viele Kontakte, so trägt sie zu jedem Kontakt nur E_{match}/n bei. So wird auf Rezeptor- und Ligandseite verfahren.

Eine schematische Erklärung dieses Modells (XFURCATE genannt) für eine bifurkate Wasserstoffbrücke und ein Vergleich mit dem originären Bewertungsschema von FLEXX ist in Abbildung 3.8 zu finden.

Das XFURCATE-Modell kann einfach auf multifurkate Wasserstoffbrücken erweitert werden.

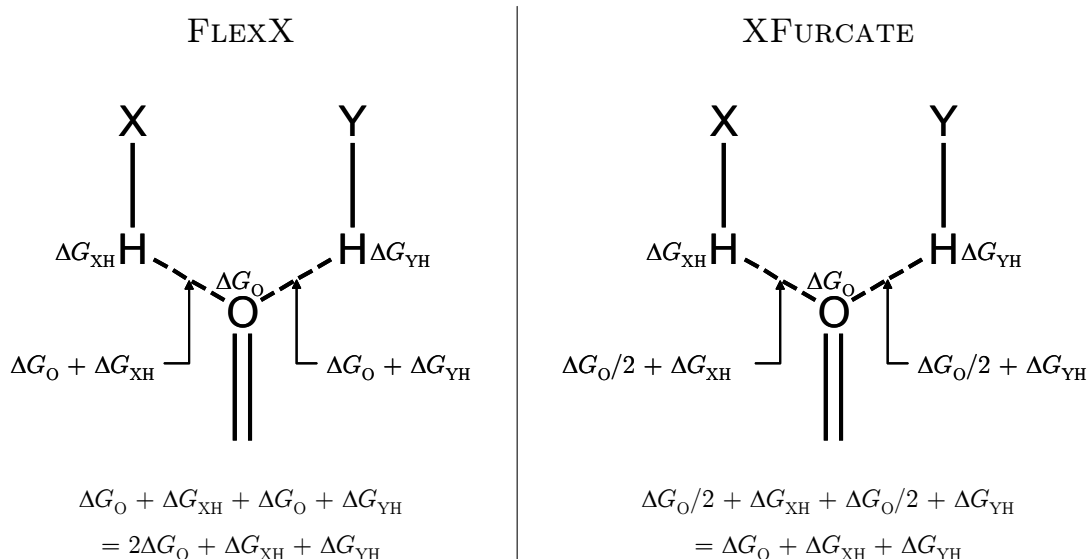


Abbildung 3.8. Schematische Darstellung der Unterschiede in der Bewertung von bifurkaten Wasserstoffbrücken. ΔG_O ist das Energieinkrement für einen sp^2 -Sauerstoff, ΔG_{XH} und ΔG_{YH} sind die H-Brückendonor-Inkremete für X – H and Y – H.

3.4.4 Ein Datensatz von Protein–Ligand-Komplexen

Für die Optimierung und Validierung von AISCORE (siehe Kapitel 3.4.2) wird ein Satz von Protein–Ligand-Komplexen mit bekannter experimenteller Bindungsenergie benötigt. Für einen solchen Satz wurden 100 Komplexe aus der *Protein-Database* (PDB) [146] ausgewählt, für die experimentelle Bindungsenergien zu finden waren (für eine vollständige Auflistung aller Komplexe siehe Tabelle 3.3). Unter diesen Komplexen sind chemisch diverse Liganden wie Alkohole, Ketone, Phosphate, Sulfate, Amine, Sulfonamide, Carbonsäuren, Zucker, Heterozyklen, Nukleinbasen, Peptide, Harnstoffderivate, uvam. Die Proteine decken pharmakologisch relevante Klassen ab (Hydrolasen, Oxidoreduktasen, Transferasen, Proteinkinasen, Lyasen, Proteinasen, Ligasen, HIV-1 Proteasen, Immunoglobuline, Thrombine, verschiedene Bindeproteine, Transportproteine, uvam.)

Wenn möglich, wurden nur Strukturen mit aufgenommen, die mit einer Auflösung von 2,5 Å oder besser bestimmt wurden. Jedoch mussten einige Ausnahmen für Komplexe mit seltenen, aber wichtigen Liganden gemacht werden.

In den PDB-Einträgen sind immer einige Fehler vorhanden und auch automatische Korrekturalgorithmen können nur bedingt aus dem Zusammenhang eine besondere funktionelle Gruppe erkennen. Die PDB-Syntax kennt nur Atomtypen, die in den Aminosäuren, deren Seitenketten und in typischen organischen Molekülen vorkommen. [135] Die Atomtypen der Aminosäuren beziehen sich dabei

rein auf eine Nomenklatur der Atome und geben daher nur indirekt Auskunft über Hybridisierung und Protonierung. Im Unterschied zu FLEXXSCORE ist AISCORE von einer exakten Zuweisung der Atomtypen abhängig. Kleine Fehler würden für FLEXXSCORE keine Rolle spielen, werden doch ohnehin für jede Wasserstoffbrücke dieselben Energie-Inkrementen genommen. Da AISCORE je nach chemischem Kontext stark unterschiedliche Wechselwirkungsgeometrien zuweist, ist es sehr wichtig, dass die Atomtypen in dem Protein–Ligand-Satz chemisch korrekt sind.

3.4.4.1 Bindetaschen

Als Bindetasche wurden alle vollständigen Aminosäuren in einem Radius von 10 Å um den Liganden herum definiert. Für eine sinnvolle Protonierung der Atome in der Bindetasche wurden die *Amino Acid Templates* von FLEXX verwendet. Jede Bindetasche wurde anschließend am Bildschirm überprüft und wenn nötig Atomtypen korrigiert. Cofaktoren wurden extrahiert, protoniert und mit chemisch korrekten Atomtypen in SYBYL-Dateien gespeichert. Sie können in FLEXX einfach zum aktiven Zentrum hinzugefügt werden.

Wassermoleküle in der Bindetasche

Typischerweise enthalten die Kristallstrukturen von Proteinen bis zu 70% Wasser. [147, 148] In experimentellen Studien mit Kristallstrukturen wurde über vielfältige Beispiele von *wasser-vermittelten* Wechselwirkungen zwischen Protein und Ligand berichtet. [149, 150, 151] Die Wassermoleküle können Platzierungen des Liganden, die ansonsten sterisch oder elektrostatisch ungünstig wären, stabilisieren oder überhaupt erst ermöglichen. Die *lokale Dieletrizitätskonstante* wird durch Wassermoleküle ebenfalls beeinflusst. [152] In solchen Fällen müssen die Wassermoleküle als integraler Bestandteil der Bindetasche angesehen werden. Bei gleichen Proteinen sind oft immer die gleichen Stellen mit Wassermolekülen besetzt [149] selbst unter unterschiedlichen kristallographischen Bedingungen [153] und/oder mit anderen Liganden [154, 155].

Die meisten Wassermoleküle in einem Proteinkristall sind derart beweglich, dass sie kristallographisch überhaupt nicht aufgelöst werden können. [156] Nur Wassermoleküle in unmittelbarer Nähe zur Proteinoberfläche (*erste Hydratationsschale*) und die wenigen internen Wassermoleküle (ganz im Inneren des Proteins) können detektiert werden. [157] Die Verschmierung der Elektronendichte durch die Bewegung der Wassermoleküle wird durch eine isotrope, harmonische

Oszillation modelliert. Der *isotrope Temperaturfaktor* (*B*-Faktor) eines Atoms *j* ist ein Maß für die rein harmonische thermische Bewegung gleicher Größe in alle Richtungen:

$$B_j = 8\pi^2 \{u_j^2\}. \quad (3.6)$$

Dabei ist $\{u_j^2\}$ die gemittelte Verschiebung des Atoms von seiner Ruheposition.

Doch auch bei guter Qualität der gemessenen Dichte, können Verwechslungen mit anderen Zehn-Elektronen-Teilchen vorkommen (NH_4^+ , Na^+ , Mg^{++}).

Wenn Wassermoleküle zusätzliche Wechselwirkungen zwischen Ligand und Protein ermöglichen, so tragen sie direkt zu der freien Bindungsenergie des Komplexes bei. Ein Versuch, die Parameter einer empirischen Bewertungsfunktion an die experimentellen Energien solcher Komplexe anzupassen, *ohne* diese zusätzlichen Wechselwirkungen zu berücksichtigen, würde dazu führen, dass die Parameter in dem Versuch, den experimentellen Wert mit zu wenigen Wechselwirkungen zu erreichen, unnatürlich hoch werden. Das würde jedoch für Komplexe, in denen *keine* Wassermoleküle Wechselwirkungen vermitteln, zu einer Überschätzung der Bindungsenergie führen. Deshalb ist es so wichtig, dass in einem Komplexsatz, der zur Kalibrierung einer empirischen Funktion erhalten soll, möglichst alle Wechselwirkungen, die die Modellfunktion abbilden kann, auch erfasst werden können.

Die Positionen der Wasserstoffatome der Wassermoleküle in einer Protein-Bindetasche werden für gewöhnlich mit Röntgenstrukturanalyse *nicht* aufgelöst. Es wurden deshalb in einem ersten Schritt an alle Sauerstoffatome, die zu Wassermolekülen innerhalb eines Radius von 3,5 Å um den Liganden herum gehören, Wasserstoffatome addiert. Die genauen Positionen dieser Wasserstoffatome wurden dann durch partielle Geometrieoptimierung (nur die Koordinaten der Wasserstoffe wurden optimiert) mit dem MMFF94-Kraftfeld [158] in Anwesenheit des Liganden (in der Lage wie in der Kristallstruktur) und des Proteins bestimmt. Danach wurde überprüft, ob neue Wechselwirkungen mit den Wassermolekülen die Positionen der Protonen des Liganden beeinflussen. In keinem der hier verwendeten Komplexe war dies der Fall.

Um zu entscheiden, ob eines dieser Wassermoleküle als ein Teil der Bindetasche angesehen werden muss oder ob es entfernt werden muss, wurde ein Bewertungsschema aufgestellt, das in einigen Belangen ähnlich zu WATERSCORE [149] ist.

Ein Wassermolekül wird als integraler Bestandteil der Bindetasche definiert, wenn drei Kriterien erfüllt sind:

- Das Wassermolekül muss mindestens eine Wasserstoffbrücke mit dem Liganden ausbilden.
- Das Wassermolekül muss an die Oberfläche des aktiven Zentrums über eine oder mehrere Wasserstoffbrücken gebunden sein.
- Das Wassermolekül muss immobil innerhalb der Bindetasche sein.

Die meisten Kristallographen fordern mindestens eine Wasserstoffbrücke mit der Bindetasche für ein Wassermolekül, damit es ins Strukturmodell übernommen wird. [159, 160, 161] Als erstes werden deshalb alle Wassermoleküle in der Bindetasche mit einer rekalierten Version von FLEXXSCORE bewertet. Diese Version von FLEXX wurde so kalibriert, dass sie für ein Wasserdimer in idealer Geometrie $\Delta G = -3,1 \text{ kJ Mol}^{-1}$ liefert. Wenn ein Wassermolekül mit $\Delta G > -4,1 \text{ kJ Mol}^{-1}$ bewertet wird, wird davon ausgegangen, dass es nicht stark genug an das aktive Zentrum gebunden ist, und sofort verworfen. Nur Wassermoleküle, die zum Protein eine stärkere Wasserstoffbrücke ausbilden, als sie im Wasserdimer existiert, werden berücksichtigt.

Da hauptsächlich die Beiträge interessieren, die die Wassermoleküle zur Bindungsenergie des Liganden im Komplex beitragen, müssen diese auch eine hinreichend starke Wechselwirkung mit dem Liganden eingehen. Deswegen wurden die Wechselwirkungen der einzelnen Wassermoleküle mit dem Liganden ebenfalls mit der oben benannten Bewertungsfunktion abgeschätzt. Wassermoleküle mit $\Delta G < -3,1 \text{ kJ Mol}^{-1}$ wurden in die Bindetasche mit aufgenommen.

Nur Wassermoleküle, deren Sauerstoffatom einen isotropen Temperaturfaktor von $B_{\text{O}} < 30 \text{ \AA}^2$ haben, werden als immobil genug angesehen, um so fest gebunden zu sein, dass sie zur Bindetasche gehören.

3.4.4.2 Protonierung des Liganden und Zuweisung der Atomtypen

Die Koordinaten der nicht-Wasserstoff-Atome der Liganden wurden direkt von den PDB-Files übernommen. Mit dem Programm CORINA wurden den Strukturen passende SYBYL-Atomtypen zugeordnet und diese Strukturen dann protoniert. CORINA hat teilweise Schwierigkeiten, aromatische Gruppen und Ringe zu erkennen, wenn nicht von Anfang an die richtigen Bindungstypen vorhanden

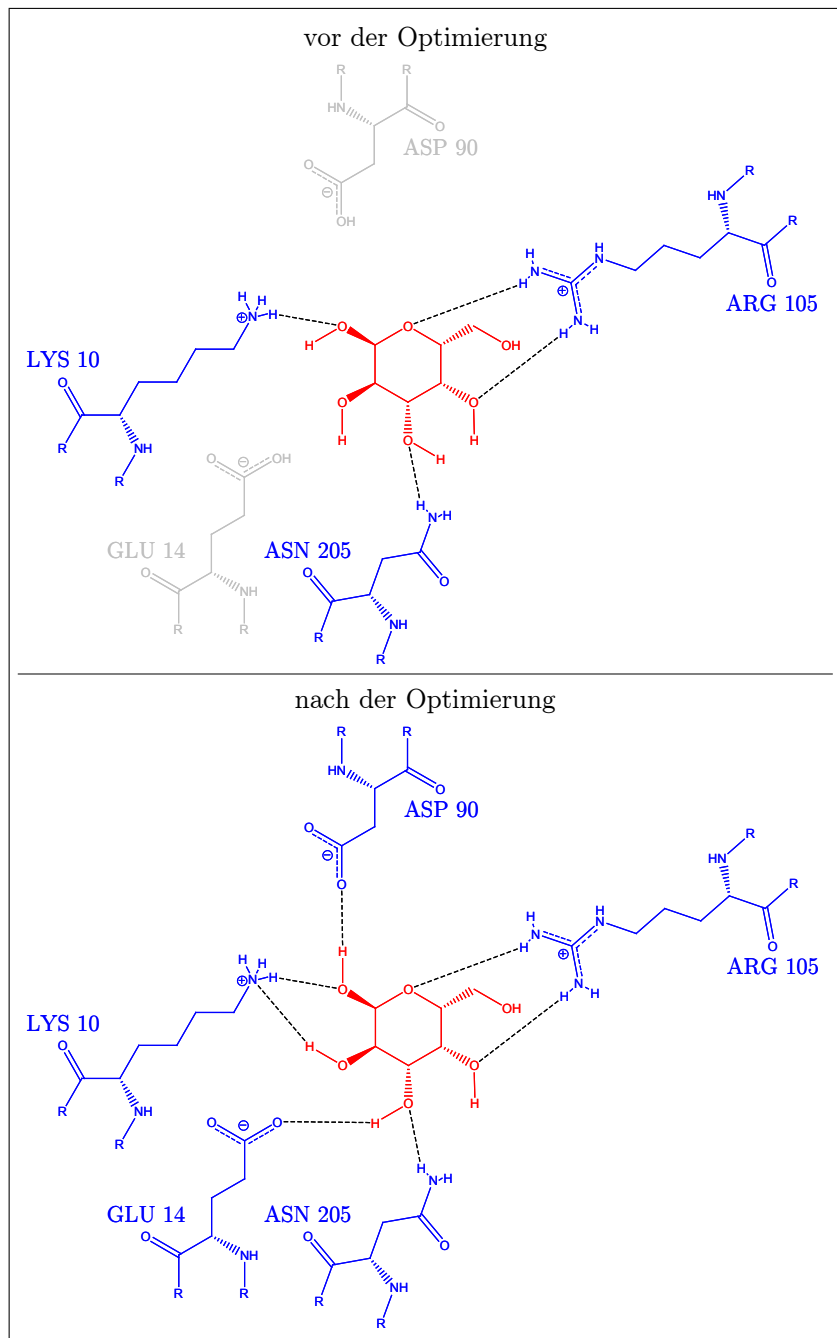


Abbildung 3.9. Komplex 8ABP jeweils vor und nach der Optimierung der Diederwinkel der Wasserstoffatome.

sind. Deswegen wurde die Zuweisung der Atomtypen und die Protonierung aller Liganden manuell überprüft und, falls nötig, korrigiert.

Die Diederwinkel der Protonen setzt CORINA entsprechend einer internen Datenbank. Jedoch werden sich die Protonen während der Komplexierung so ausrichten, dass sie maximal viele Wechselwirkungen eingehen können. Durch die richtige Ausrichtung der Protonen kann je nach Komplex die freie Bindungs-

energie massiv erhöht werden, weil so zusätzlich Wasserstoffbrücken entstehen. Zur Verdeutlichung ist in Abbildung 3.9 schematisch der Ligand und die mit ihm wechselwirkenden Aminosäuren des PDB-Eintrags 8ABP gezeigt. Im oberen Teil sind die Diederwinkel der Protonen des Zuckers so belassen, wie sie von CORINA berechnet wurden. Mit dieser Konfiguration können insgesamt vier Wasserstoffbrücken zu drei Aminosäuren ausgebildet werden. Durch die Veränderung der Diederwinkel von nur drei Protonen können jedoch sieben H-Brücken zu fünf Aminosäuren gebildet werden, was entsprechend eine Erhöhung der Bindungsenergie ergibt (unterer Teil von Abbildung 3.9).

Es ist daher extrem wichtig, die Position der Protonen des Liganden zu optimieren. Ohne diese Optimierung würden Wechselwirkungen fehlen. Während der Parameteroptimierung von AISCORE würden dann, durch die fehlenden Wechselwirkungen, andere Energie-Inkrementen unnatürlich groß werden, um trotzdem die experimentelle Bindungsenergie erreichen zu können.

Die Positionen der Wasserstoffatome eines Liganden, die eine Wasserstoffbrücke ausbilden können, werden deshalb in zwei Schritten optimiert. Diese Optimierung ist Teil des Tools FLEXR (siehe auch Kapitel 3.6) [162, 163]. Im ersten Schritt werden Ausgangswerte für die Positionen der einzelnen Hydroxyl- und Thiolgruppen bestimmt. Dies geschieht, indem die einzelnen Wasserstoffe um die HO – R- bzw. HS – R-Bindungen in 15°-Schritten gedreht werden. Für jede Position wird die Wechselwirkungsenergie mit dem Protein mit dem TRIPOS-Kraftfeld [164] berechnet. Diejenige Struktur mit der größten Bindungsenergie wird beibehalten.

Im zweiten Schritt werden die Koordinaten von *allen polaren* Wasserstoffatomen des Liganden simultan in Anwesenheit der Bindetasche optimiert. Hierfür kommt eine *quasi-Newton*-Minimierung zum Einsatz.

3.4.4.3 Das Tool WaterGen

Da ein mittelfristiges Ziel die Erstellung eines chemisch korrekten Protein-Ligand-Datensatzes mit deutlich mehr als 100 Komplexen ist, habe ich hochflexible Algorithmen programmiert, die sich wiederholende Schritte automatisieren. Das Tool WATERGEN übernimmt, ausgehend von einem PDB-File, das Extrahieren des Liganden und ruft anschließend CORINA für die Zuweisung der Atomtypen und die Protonierung auf. Anschließend extrahiert es die Bindetasche und generiert die für FLEXX nötigen Inputfiles. Danach übernimmt es voll

automatisch die Protonierung, Optimierung und Bewertung der Wassermoleküle. WaterGen ist in Fortran95 geschrieben.

Tabelle 3.3: Komplexe des Protein–Ligand-Datensatzes.

PDB Code	Protein	Ligand	ΔG_{exp} [kJ Mol ⁻¹]	Auflösung [Å]	# H ₂ O ¹
1MFE	IGG1-λ SE155-4 FAB	GAL-MAN-ABE	-30.29	2.0	2
1ACJ	Acetylcholinesterase	Tacrine	-41.65	2.8	1
1ADD	Adenosine Deaminase	1-Deaza-Adenosine	-38.45	2.4	
1ADB	Alcohol Dehydrogenase	CNAD	-47.92	2.4	
1BCU	α-Thrombin	Proflavin	-28.93	2.0	
1DWB	α-Thrombin	Benzamidine	-16.79	3.2	
1DWD	α-Thrombin	NAPAP	-46.67	3.0	1
1TMT	α-Thrombin	CGP 50,856	-36.09	2.2	
9AAT	Aspartate Aminotransferase	Pyridoxal-5'-Phosphate	-46.90	2.2	
3PTB	β-trypsin	Benzamidine	-27.42	1.7	
1YDR	CAMP-Dependent Protein Kinase	H7 Protein Kinase Inhibitor	-31.82	2.2	
1CBX	Carboxypeptidase A	L-Benzylsuccinate	-36.23	2.0	
2CTC	Carboxypeptidase A	L-phenyl lactate	-22.19	1.4	
3CPA	Carboxypeptidase A	Glycyl-L-Tyrosine	-22.82	2.0	
1CPS	Carboxypeptidase A	[L-(-)-2-carboxy-3-phenylpropyl]methyl-sulfodiimine	-38.00	2.3	
1GHI	Cell Division Protein Kinase 2	CDK4 Inhibitor 1	-40.61	2.8	
1DI8	Cyclin-Dependent Protein Kinase 2	DTQ	-34.71	2.2	
1EIV	Cyclin-Dependent Protein Kinase 2	NU2058	-28.43	1.9	
1EIX	Cyclin-Dependent Protein Kinase 2	NU6027	-34.04	1.9	1
2GBP	D-Galactose/D-Glucose Binding Protein	Beta-D-Glucose	-43.36	1.9	1
1DIH	Dihydrodipicolinate Reductase	NADPH	-32.77	2.2	
1DHF	Dihydrofolate Reductase	Folate	-42.22	2.3	2
1DRF	Dihydrofolate Reductase	Folate	-42.45	2.0	1
3DFR	Dihydrofolate Reductase	Methotrexate	-58.76	1.7	1
4DFR	Dihydrofolate Reductase	Methotrexate	-49.18	1.7	1
					...

Fortsetzung ...

PDB Code	Protein	Ligand	ΔG_{exp} [kJ Mol ⁻¹]	Resoluton [Å]	# H ₂ O
7DFR	Dihydrofolate Reductase	Folate	-42.21	2.5	1
1EBG	Enolase	Phosphonoacetohydroxamate	-61.73	2.1	2
1ETR	ϵ -Thrombin	MQPA	-43.92	2.2	2
1D7H	FK506-Binding Protein	DMSO	-9.84	1.9	1
1D7I	FK506-Binding Protein	DSS	-20.85	1.9	
1FKB	FK506-Binding Protein	Rapamycin	-56.10	1.7	1
1FKF	FK506-Binding Protein	FK506	-56.10	1.7	
2GPB	Glycogen Phosphorylase B	Alpha-D-Glucose	-15.80	2.3	1
3AID	HIV Protease (Q7K)	AQ148	-39.69	2.5	1
1AAQ	HIV-1 Protease	Hydroxyethylene Isostere	-46.01	2.5	1
1DMP	HIV-1 Protease	DMP450 of Dupont-Merck	-54.34	2.0	
1HBV	HIV-1 Protease	SB202127	-36.84	2.3	1
1HHH	HIV-1 Protease	CGP 53820	-45.93	2.2	1
1HIV	HIV-1 Protease	U75875	-51.37	2.0	1
1HPV	HIV-1 Protease	VX-478	-53.93	1.9	1
1HPX	HIV-1 Protease	KNI-272	-48.86	2.0	2
1HSG	HIV-1 Protease	Indinavir L-735,524	-53.59	2.0	3
1HVI	HIV-1 Protease	A77003 (R,S)	-57.53	1.8	1
1HVJ	HIV-1 Protease	A78791 (S,-)	-60.50	2.0	1
1HVK	HIV-1 Protease	A76928 (S,S)	-58.53	1.8	1
1HVL	HIV-1 Protease	A76889 (R,R)	-52.13	1.8	1
1HVR	HIV-1 Protease	XK263 of Dupont-Merck	-54.60	1.8	
4HVP	HIV-1 Protease	MVT-101	-35.34	2.3	1
4PHV	HIV-1 Protease	L-700,417(R,R)	-52.38	2.1	1
7HVP	HIV-1 Protease	JG-365	-57.19	2.4	3
9HVP	HIV-1 Protease	A74704	-47.65	2.8	1
1MES	HIV-1 Protease (I84V)	DMP323	-44.55	1.9	

...

Fortsetzung ...

PDB Code	Protein	Ligand	ΔG_{exp} [kJ Mol ⁻¹]	Resolution [Å]	# H ₂ O
1BDR	HIV-1 Protease (T31S, L33V, E34T, E35G, M36I, SB203386 S37E)		-44.38	2.8	1
1HVS	HIV-1 Protease (V82A)	A77003 (R,S)	-58.78	2.3	
1BV7	HIV-1 Protease (V82F)	XV638	-57.86	2.0	
1MEU	HIV-1 Protease (V82F, I84V)	DMP323	-34.71	1.9	
3YAS	Hydroxynitrile Lyase	Acetone	-8.67	1.9	
5YAS	Hydroxynitrile Lyase	Hexafluoro-Acetone Hydrate	-18.55	2.2	
1AJ7	Immunoglobulin 48G7 FAB	5-(para-nitrophenyl phosphonate)-pentanoic acid	-22.08	2.1	
1DD7	Inducible Nitric Oxide Synthase	IPM	-49.39	2.3	1
1ABE	L-Arabinose-Binding Protein	L-Arabinose	-37.60	1.7	1
1ABF	L-Arabinose-Binding Protein	D-Fucose	-30.06	1.9	1
5ABP	L-Arabinose-Binding Protein	D-Galactose	-38.48	1.8	2
6ABP	L-Arabinose-Binding Protein (M108L)	L-Arabinose	-40.49	1.7	
7ABP	L-Arabinose-Binding Protein (M108L)	D-Fucose	-31.23	1.7	1
8ABP	L-Arabinose-Binding Protein (M108L)	D-Galactose	-46.35	1.5	2
1BAP	L-Arabinose-Binding Protein (P254G)	L-Arabinose	-39.90	1.8	1
1NNB	Neuraminidase	DANA	-30.65	2.8	
2QWC	Neuraminidase (R292K)	Neu5Ac2en	-20.52	1.6	1
2QWD	Neuraminidase (R292K)	4-Amino-Neu5Ac2en	-28.05	2.0	1
1KV1	p38 MAP Kinase	Inhibitor-1	-34.33	2.5	
1APU	Penicillopepsin	Pepstatine-IsoVS	-43.33	1.8	
1PPM	Penicillopepsin	Phosphoanalog 2	-33.54	1.7	
2WEA	Penicillopepsin	MCP1P31a	-35.25	1.3	1
2PHH	p-Hydroxybenzoate Hydroxylase	p-Hydroxy-Benzolate	-26.63	2.7	
1RBP	Plasma Retinol-Binding Protein Precursor	Retinol	-38.34	2.0	1
1ULB	Purine Nucleoside Phosphorylase	Guanine	-25.11	2.8	
1ERB	Retinol Binding Protein	N-Ethyl Retinamide	-40.07	1.9	

...

Fortsetzung ...

PDB Code	Protein	Ligand	ΔG_{exp} [kJ Mol ⁻¹]	Resoluton [Å]	# H ₂ O
1FEL	Retinol Binding Protein	Feretinide	-38.48	1.8	1
6APR	Rhizopuspepsin	Pepstatin	-44.97	2.5	2
6RNT	Ribonuclease T1	2-AMP	-13.69	1.8	1
1NVQ	Serine/Threonine-Pprotein Kinase Chk1	UCN-01	-47.73	2.0	2
1NVR	Serine/Threonine-Pprotein Kinase Chk1	Staurosporine	-46.89	1.8	1
1NVS	Serine/Threonine-Pprotein Kinase Chk1	SB218078	-45.26	1.8	1
1A08	SH2 of C-SRC Tyrosine Kinase	DF-PTR	-32.49	2.2	1
1A1C	SH2 of C-SRC Tyrosine Kinase	MCP-PTR	-37.01	2.4	3
2TSC	Thymidylate Synthase	CB3717	-48.61	2.0	1
2YPI	Triosephosphate Isomerase	2-Phospho-Glycolate	-27.88	2.5	1
4TIM	Triosephosphate Isomerase	2-Phospho-Glycerate	-12.48	2.4	1
6TIM	Triosephosphate Isomerase	Glycerol-3-Phosphate	-18.59	2.2	1
7TIM	Triosephosphate Isomerase	Phospho-Glycolo-Hydroxamate	-31.23	1.9	2
1PPH	Trypsin	3-TAPAP	-36.05	1.9	1
1TNH	Trypsin	4-Fluoro-Benzylamine	-19.51	1.8	1
1TNJ	Trypsin	2-Phenyl-Ethylamine	-11.35	1.8	1
1TNK	Trypsin	3-Phenyl-Propylamine	-8.50	1.8	1
1TNL	Trypsin	Tranylcyproline	-10.89	1.9	1
1BRA	Trypsin (D189G, G226D)	Benzamidine	-10.59	2.2	1
3TPI	Trypsinogen	Ile-Val	-24.53	1.9	2
4TS1	Tyrosyl-tRNA Synthetase	Tyrosine	-32.01	2.5	2
1EJN	Urokinase-Type Plasminogen Activator	N-(1-Adamantyl)-N'-(4-Guanidinobenzyl)Urea	-32.07	1.8	1

3.5 AIScore vs. FlexXScore – Ergebnisse

Im Folgenden werden AISCORE und FLEXXSCORE miteinander verglichen. Im Vordergrund steht dabei die Fähigkeit, die freie Bindungsenergie eines Komplexes in der Konfiguration, wie sie in der Kristallstruktur vorliegt, richtig vorherzusagen zu können (Kapitel 3.5.1).

3.5.1 Freie Bindungsenergien

Ein gutes Kriterium für die Güte einer Bewertungsfunktion in Hinblick auf ihre Zuverlässigkeit bei der Vorhersage von Bindungsenergien ist die lineare Korrelation zwischen experimentellen und berechneten Werten. Wichtig ist dabei auch die Bandbreite der berechneten Energien, denn eine hohe lineare Korrelation ist wertlos, wenn sie nur dadurch erreicht wird, dass die berechneten Energien einen nur so schmalen Bereich abdecken, dass zwar große Abweichungen zur genäherten Gerade gar nicht auftreten können, aber alle Komplexe ungefähr die gleiche Bindungsenergie haben.

Zunächst werden die linearen Korrelationen des Protein–Ligand-Satzes vorgestellt und Ausreißer und Besonderheiten diskutiert. Als Bewertungsfunktionen wurden dabei AISCORE mit und ohne XFURCATE-Algorithmus untersucht (Abbildung 3.10) sowie die ursprüngliche FLEXXSCORE-Implementation (Abbildung 3.11*a*) und FLEXXSCORE, erweitert um den XFURCATE-Algorithmus (Abbildung 3.11*b*).

Zum Vergleich wurde FLEXXSCORE dem gleichen Optimierungs-Prozedere wie AISCORE unterworfen (Abbildung 3.12*a*), um den Einfluss des Datensatzes und der Parameteroptimierung untersuchen zu können. Doch besitzt FLEXXSCORE sehr viel weniger Parameter als AISCORE. Deshalb wurde FLEXXSCORE um die Wechselwirkungsgeometrien für spezielle Wasserstoffbrücken und ionische Wechselwirkungen (vgl. Tabelle 3.2) erweitert und anschließend optimiert (Abbildung 3.12*b*).

3.5.1.1 AIScore

Die verschiedenen Korrelationsplots (Abbildungen 3.10, 3.11, 3.12) zeigen deutlich, dass AISCORE mit seinen chemisch diversen Inkrementen und dem XFURCATE-Algorithmus zur Behandlung von multifurkaten Wasserstoffbrücken am besten die freie Bindungsenergie der Protein–Ligand-Komplexe in ihrer experimentell Konformation berechnen kann (Abbildung 3.10*a*). Die lineare Korrelation

ist $r^2 = 0,76$, die Standardabweichung ist $9,53 \text{ kJ Mol}^{-1}$. Der Bereich, der von den abgeschätzten Bindungsenergien abgedeckt wird, ist für 99 der 100 Komplexe vergleichbar demjenigen Bereich, der von den entsprechenden experimentell bestimmten Bindungsenergien eingenommen wird ($\Delta G_{\text{AIScore}} = -1,5 \dots -72,9 \text{ kJ Mol}^{-1}$, $\Delta G_{\text{exp}} = -8,5 \dots -61,8 \text{ kJ Mol}^{-1}$). Nur der Komplex mit dem PDB-Code 1BV7 fällt aus der Reihe ($\Delta G_{\text{AIScore}} = -81,32 \text{ kJ Mol}^{-1}$, $\Delta G_{\text{exp}} = -57,86 \text{ kJ Mol}^{-1}$).

Ohne den XFURCATE-Algorithmus (Abbildung 3.10b) verringert sich die lineare Korrelation auf $r^2 = 0,66$. Der ehemals einzige Ausreißer 1BV7 bleibt durch die Vernachlässigung von XFURCATE mehr oder weniger unverändert, jedoch sind *fünf neue* Ausreißer hinzugekommen.

Im Komplex 1EBG (Enolase komplexiert mit Phosphono-Aceto-Hydroxamat, Abbildung 3.14d) existiert eine bifurkate Wasserstoffbrücke zwischen dem Amid-Sauerstoff des Liganden und GLN-167 und HOH-222. Außerdem ist dieses Sauerstoffatom noch für *beide* Magnesiumatome Akzeptor. Der XFURCATE-Algorithmus skaliert davon nur die Wasserstoffbrücken, die metallischen Wechselwirkungen bleiben unangetastet. Doch ohne den XFURCATE-Algorithmus werden die beiden Wasserstoffbrücken einfach aufaddiert, was zu einer Überbewertung der Bindungsenergie führt.

In 1ADB (Alkohol-Dehydrogenase komplexiert mit CNAD, Abbildung 3.13c) existieren hauptsächlich lipophile und aromatische Wechselwirkungen. Diese werden offensichtlich im Zuge der Optimierung *ohne* XFURCATE im Verhältnis zu den anderen Wechselwirkungsenergie-Inkrementen überbewertet, sodass es bei solchen Komplexen, in denen sehr viele dieser Wechselwirkungen vorkommen, zu einer Überschätzung der Bindungsenergie kommt.

1A08 ist die SH2-Domäne der Tyrosinkinase komplexiert mit DF-PTR (Abbildung 3.13a). Hier zeigt das Atom mit der Kennung O3P der PO_3 -Gruppe eine dreifach gegabelte Wasserstoffbrücke zu drei positiv geladenen NH^+ -Gruppen und zwar zum Atom NH_2 von ARG-178 und zu den Atomen NH_2 und NE von ARG-158. Zusätzlich bildet das Atom O2 der Amidgruppe eine bifurkate Wasserstoffbrücke mit ARG-178 und ARG-158. Insofern ist es nicht überraschend, dass die Vernachlässigung von XFURCATE zu einer zu hohen Bindungsenergie führt. Auf der anderen Seite wird die Bindungsenergie von 1A08 etwas unterschätzt, wenn XFURCATE verwendet wird.

6RNT (Ribonuklease T1 mit 2-AMP, Abbildung 3.15g) besitzt eine dreifach gegabelte H-Brücke von der PO_4H -Gruppe zu den beiden geladenen NH^+ -

Gruppen von ARG-77 und zu HIS-92. Auch dies führt ohne XFURCATE zu einer Überschätzung der freien Bindungsenergie dieses Komplexes.

Im Komplex 6TIM (Triosephosphat-Isomerase komplexiert mit Glycerol-3-Phosphat, Abbildung 3.15h) wird die ohne XFURCATE berechnete freie Bindungsenergie *unterschätzt*, ist sogar *repulsiv*. Der Grund liegt darin, dass das Energie-Inkrement der positiv geladenen Sauerstoffatome in Phosphatgruppen zu klein ist. Dieses Inkrement ist nämlich so oft an multifurkaten H-Brücken beteiligt, dass es während der Optimierung ohne XFURCATE erniedrigt wird, um die lineare Korrelation zu maximieren.

3.5.1.2 FlexXScore

Abbildung 3.11a zeigt die Korrelation zwischen den experimentellen Bindungsenergien und der original FLEXSCORE. Der quadrierte lineare Korrelationskoeffizient von $r^2 = 0,24$ zeigt, dass die Korrelation sehr viel geringer als bei AISCORE ist. Der Bereich der Energien, die von einem Großteil der mit FLEXSCORE bewerteten Komplexe abgedeckt wird, ist recht schmal verglichen mit den experimentellen Werten ($\Delta G_{\text{FlexXScore}} = -46,02 \dots + 1,29 \text{ kJ Mol}^{-1}$). Die Erweiterung um den XFURCATE-Algorithmus bewirkt eine Verbesserung der linearen Korrelation auf $r^2 = 0,33$.

In beiden Fällen gibt es einige hervorstechende Ausreißer, von denen 1ADB der extremste ist. Seine Bindungsenergie wird sehr stark überschätzt, was sich auch nicht ändert, wenn XFURCATE benutzt wird. In dem Komplex gibt es, zusätzlich zu den Wasserstoffbrücken viele aromatische Kontakte. Es scheint deshalb so, dass das undifferenziert Energieinkrement von $-2,35 \text{ kJ Mol}^{-1}$, das FLEXSCORE für Wasserstoffbrückendonoren und -akzeptoren verwendet, gegenüber den Energieinkrementen der aromatischen Wechselwirkungen zu unbalanciert ist.

Im Gegensatz zu diesem Fall, wird die Vorhersage der freien Bindungsenergie des Komplexes 1EBG durch den XFURCATE-Algorithmus verbessert, genau wie das bei AISCORE auch der Fall war.

Ein weiterer Ausreißer von FLEXSCORE ist 1A1C (SH2-Domäne der Tyrosinkinase komplexiert mit MCP-PTR, Abbildung 3.13b), in dem es einige bi- und trifurkate Wasserstoffbrücken zwischen geladenen Sauerstoffatomen einer PO_4 -Gruppe und den NH^+ -Gruppen von ARG-178 und ARG-158 gibt. Die Abweichung zum linearen Fit wird wie erwartet durch die Anwendung von XFURCATE verringert.

Die Bindungsenergien der Komplexe 2QWD (Neuraminidase (R292K-Mutation) komplexiert mit 4-Amino-Neu-6-Ac-2-en, Abbildung 3.14e) und 1EJN (Plasminogenaktivator Urokinase-Typ mit AGB-Harnstoff) sind dermaßen weit entfernt von der Ausgleichsgeraden, weil ihre häufigsten Wechselwirkungen geladene Wasserstoffbrücken sind, deren Wechselwirkungsenergie-Inkrementen zu unangewogen sind, so wie bei den aromatischen Wechselwirkungen im Falle von 1ADB (siehe oben).

Der Komplex 5YAS (Hydroxynitrillyase im Komplex mit Hexyflur-Acetonhydrat, Abbildung 3.14f) wird repulsiv bewertet. Seine einzigen beiden Wasserstoffbrücken werden durch das Standard-Inkrement von $-2,35 \text{ kJ Mol}^{-1}$ zu gering bewertet, als dass sie den Entropieverlust und den *Clash* von $1,75 \text{ kJ Mol}^{-1}$ kompensieren könnten.

Zusammenfassend kann man sagen, dass die fehlende chemische Diversität der Energie-Inkrementen von FLEXXSCORE schon an dem schmalen Bereich erkennbar ist, der von den vorhergesagten Bindungsenergien abgedeckt wird. Durch dieses Fehlen an Diversität ist für einige Komplexe sogar das Verhältnis der Energie-Inkrementen der einzelnen Wechselwirkungstypen untereinander zu unbalanciert, so dass es zu extremen Ausreißern kommt.

3.5.1.3 Optimierte und erweiterte FlexXScore

Die Optimierung der ursprünglichen fünf Energie-Inkrementen der original FLEXXSCORE hat eine Erhöhung des quadrierten linearen Korrelationskoeffizienten auf $r^2 = 0,42$ zur Folge (Abbildung 3.12a). Die berechneten Energien umfassen einen Bereich von $\Delta G_{\text{FlexXScore}} = -10,5 \text{ kJ Mol}^{-1}$ bis $\Delta G_{\text{FlexXScore}} = -679 \text{ kJ Mol}^{-1}$. Auf einer absoluten Skala betrachtet, werden also die Bindungsenergien absolut unrealistisch überschätzt, obwohl die pure Korrelation zwischen berechneten und experimentellen Bindungsenergien besser geworden ist. Da die lineare Korrelation die Zielfunktion während der Optimierung ist, verwundert dieses Ergebnis nicht. Jedoch führt die Optimierung zu sieben extremen Ausreißern (1EBG, 1BV7, 1NVQ, 1NVR, 1NVS, 1ADB). Es sei trotzdem nochmals darauf hingewiesen, dass die Bewertungsfunktion FLEXXSCORE nur fünf variierbare Parameter enthält, im Vergleich zu den 17 Freiheitsgraden von AISCORE.

Der extreme Bereich der Bindungsenergien ist ein Artefakt der Optimierung auf maximale lineare Korrelation ohne weitere Randbedingungen. In Fällen, in denen in einem einzigen Komplex, viele verschiedene Wasserstoffbrücken vorkommen, kann das einzige Energie-Inkrement für Wasserstoffbrücken kaum mehr

optimiert werden. Durch Variation nach oben würde es den schwächeren Wasserstoffbrücken nicht gerecht werden. Eine Verringerung des Inkrements hätte eine Unterschätzung der starken H-Brücken zur Folge. Dieser Parameter befindet sich somit schon in einem Optimum, jede Veränderung würde die Zielfunktion (die lineare Korrelation) verschlechtern. In diesen Fällen bleiben lediglich die Inkremente für aromatische und aliphatische Wechselwirkungen als Freiheitsgrade übrig, um eine weitere Verbesserung der linearen Korrelation zu erreichen. In den meisten Komplexen kommen nur wenige solcher Kontakte zustande, weshalb eine Erhöhung dieser Inkremente für den ganzen Satz betrachtet tatsächlich zu einer Verbesserung von r^2 führt. Jedoch führt diese Erhöhung bei Komplexen mit vielen solcher aromatischen Kontakte zu einer drastischen Überschätzung der freien Bindungsenergie.

Die Erweiterung von FLEXXSCORE um die zusätzlichen Wechselwirkungsgeometrien (und damit auch um die Subgraphen und die Energie-Inkremente) für geladene und resonanzassistierte Wasserstoffbrücken gibt der Funktion 9 Freiheitsgrade zusätzlich für die Optimierung. Das bringt eine Erhöhung der linearen Korrelation auf $r^2 = 0,46$ (Abbildung 3.12). Dies ist eine nur geringe Verbesserung im Vergleich zur optimierten Version von FLEXXSCORE. Daran, dass AISCORE eine signifikant bessere lineare Korrelation liefert, zeigt, wie wichtig die chemisch diversen H-Brücken-Energieinkremente für eine verlässliche Vorhersage der Protein-Ligand-Komplexbindungsenergie ist.

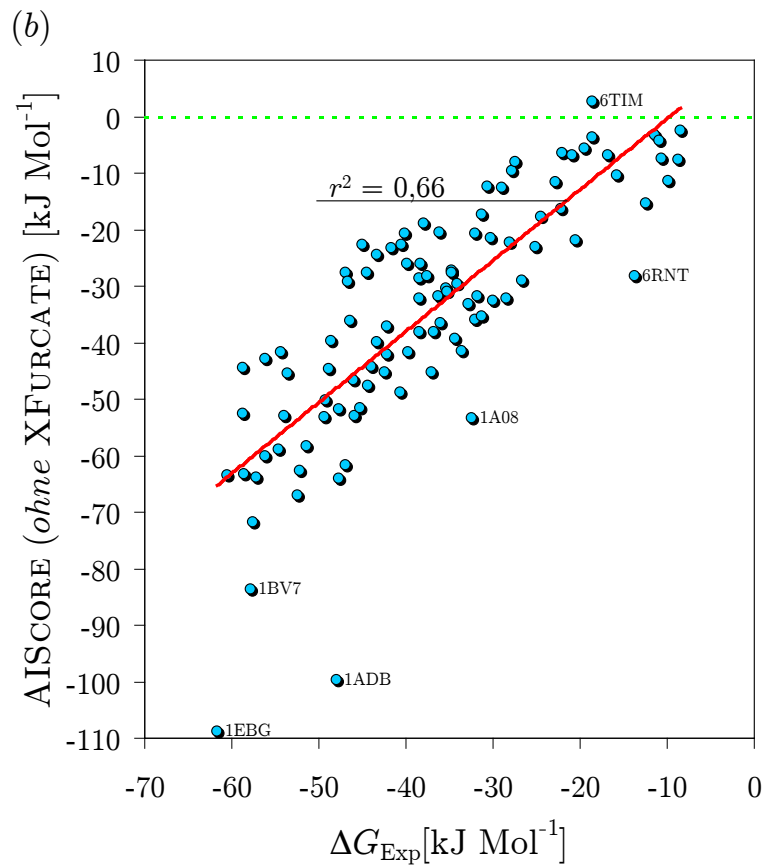
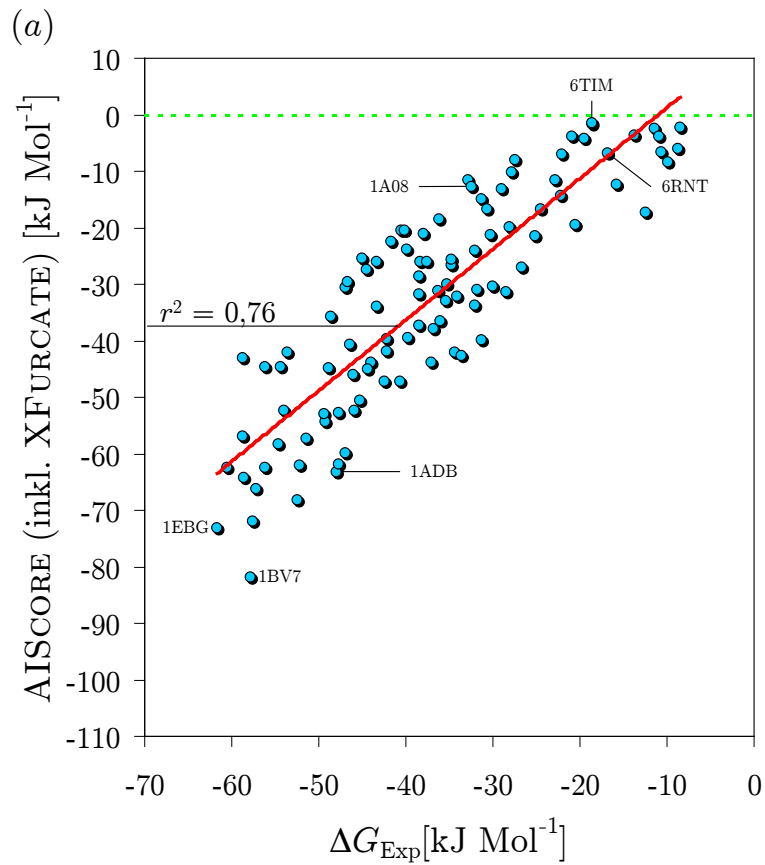


Abbildung 3.10. Lineare Korrelation zwischen der experimentellen freien Bindungsenergie der Komplexe ΔG_{Exp} und der berechneten Bindungsenergie von AISCORE a) mit XFURCATE und b) ohne XFURCATE.

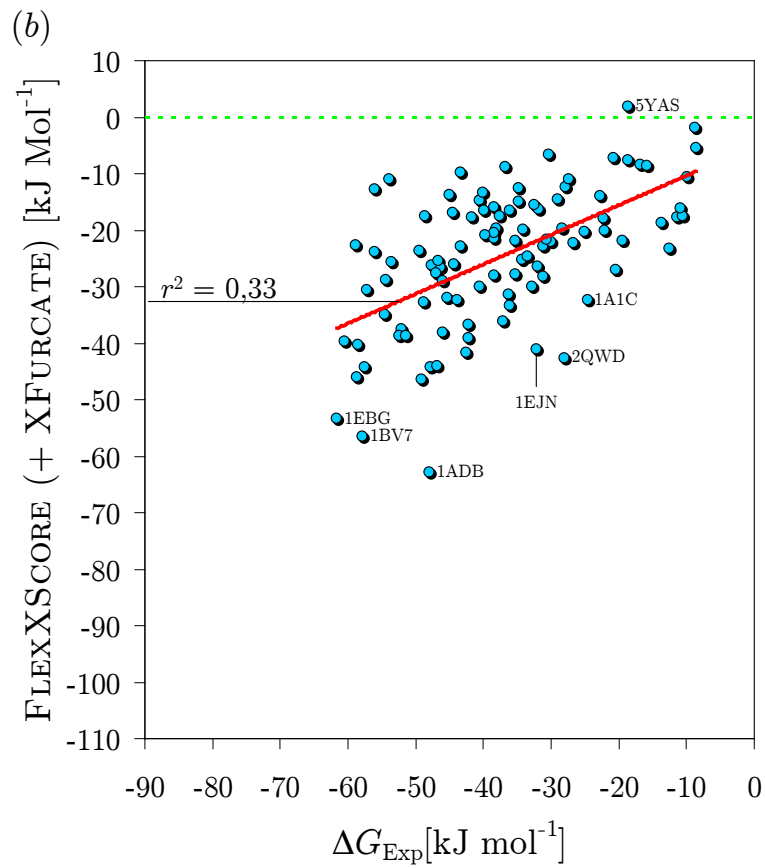
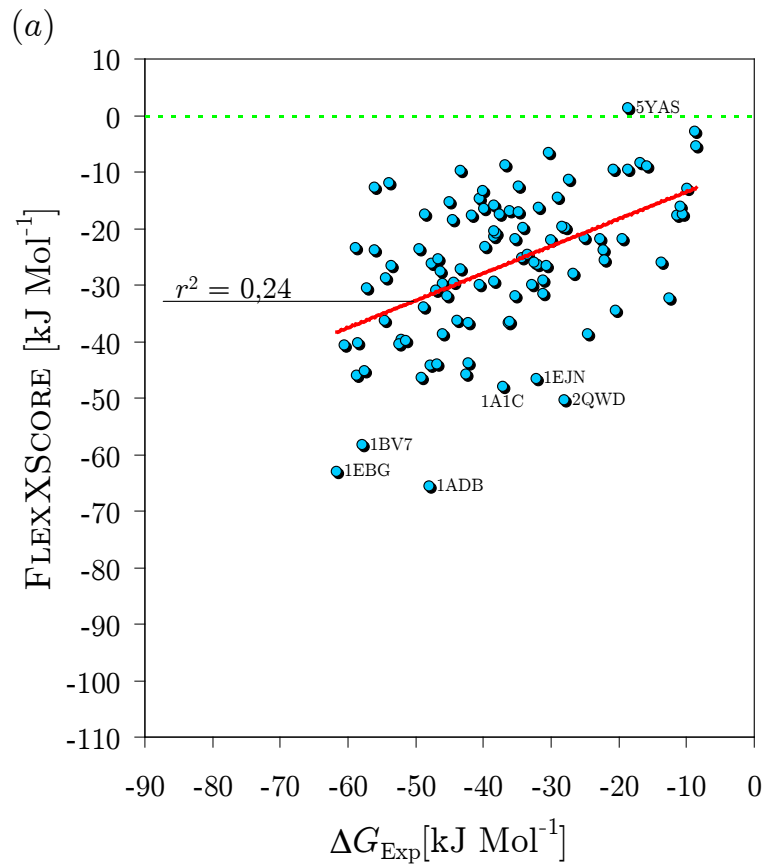


Abbildung 3.11. Lineare Korrelation zwischen der experimentellen freien Bindungsenergie der Komplexe ΔG_{Exp} und der berechneten Bindungsenergie von FLEXSCORE a) ohne XFURCATE und b) mit XFURCATE.

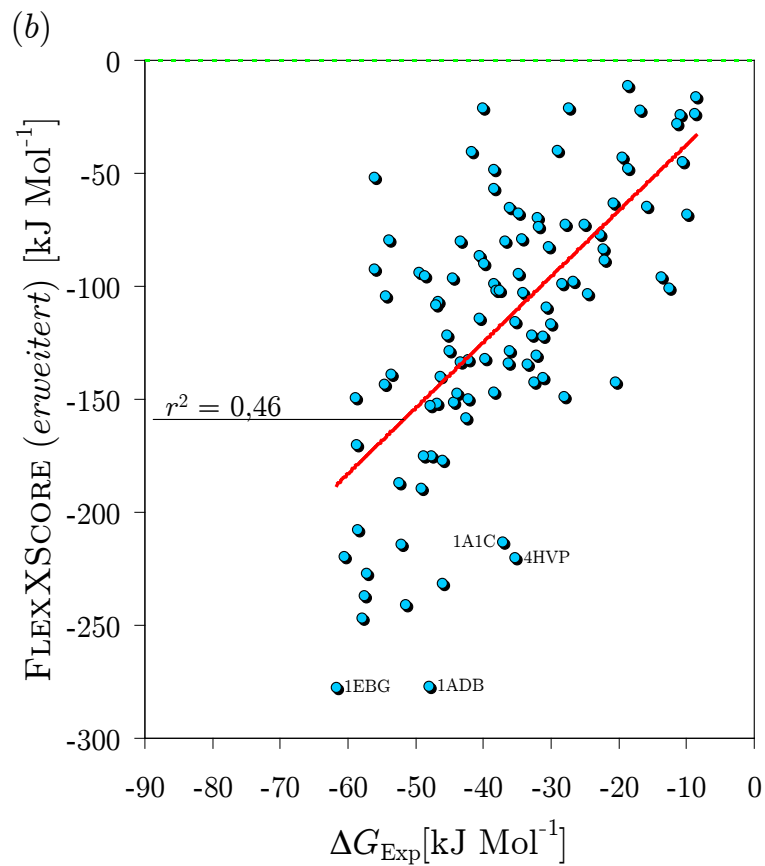
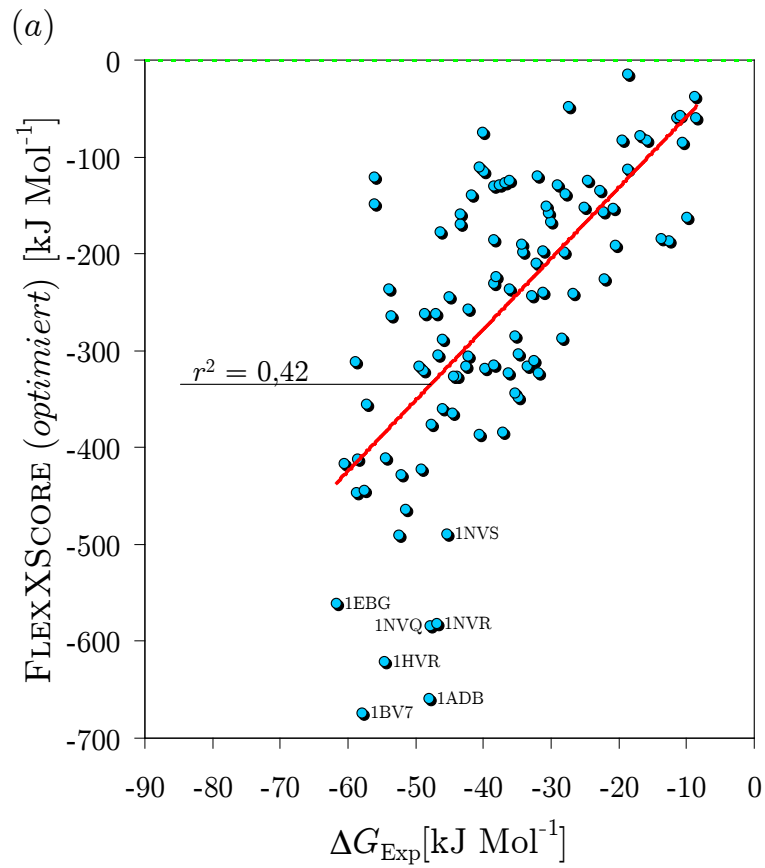
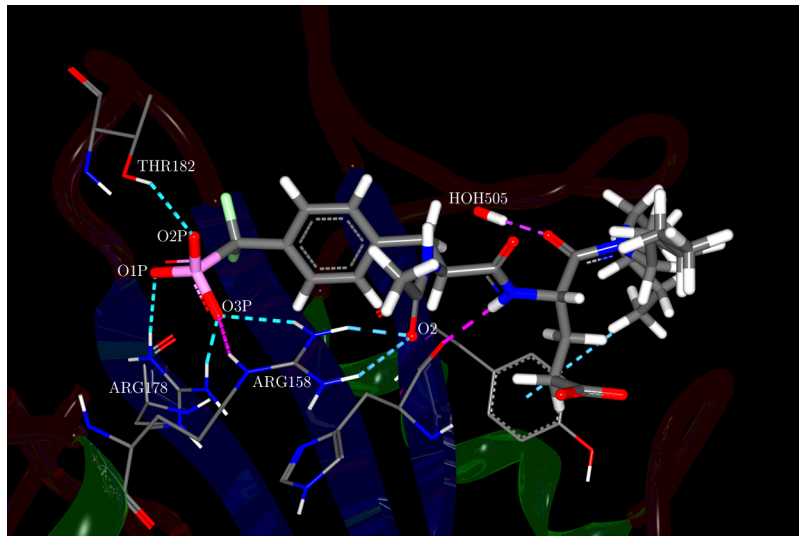
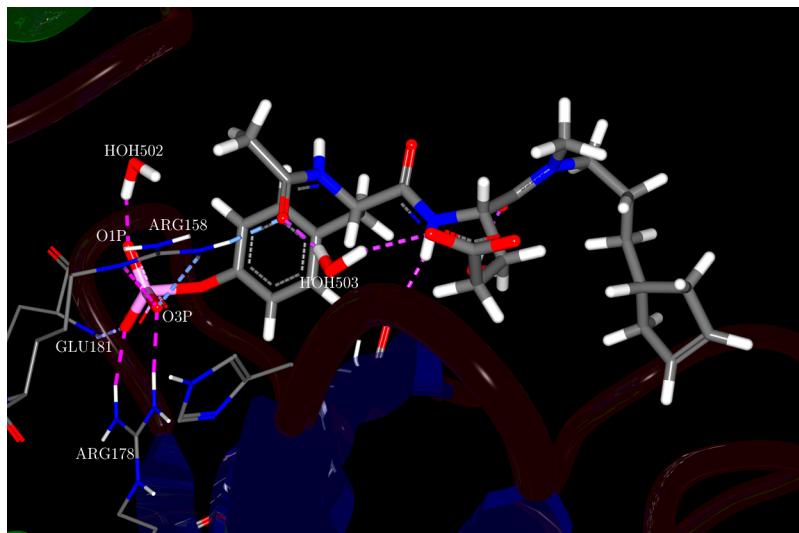


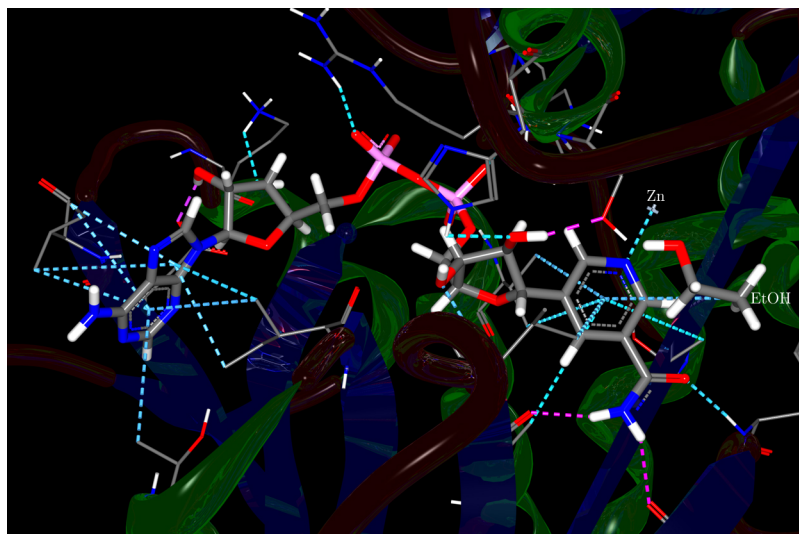
Abbildung 3.12. Lineare Korrelation zwischen der experimentellen freien Bindungsenergie der Komplexe ΔG_{Exp} und der berechneten Bindungsenergie von FLEXSCORE a) nach der Optimierung und b) mit zusätzlich Wechselwirkungsgeometrien und anschließender Optimierung.



a) 1A08

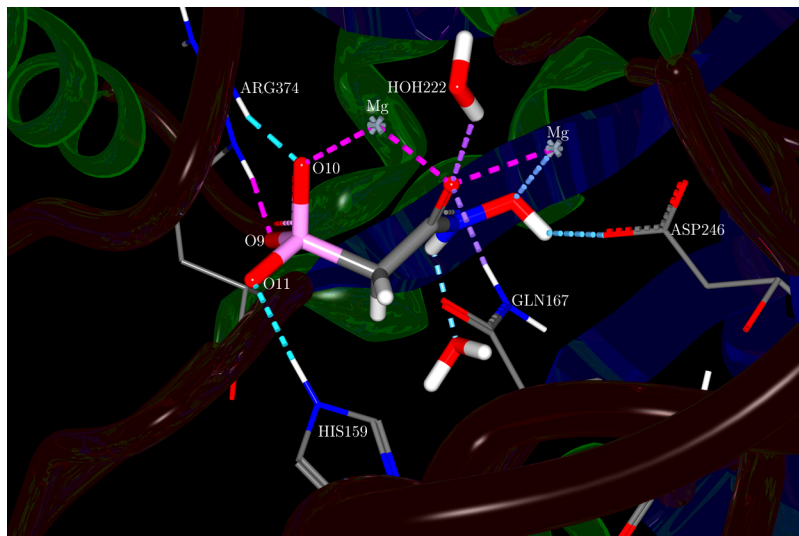


b) 1A1C

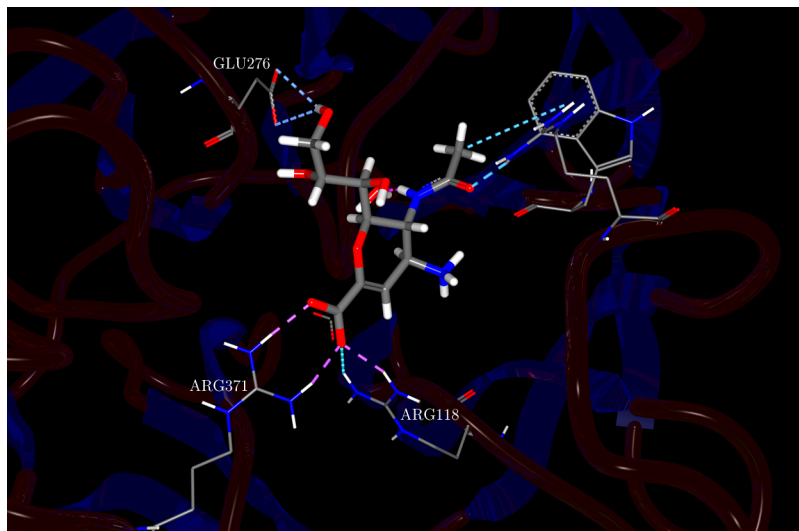


c) 1ADB

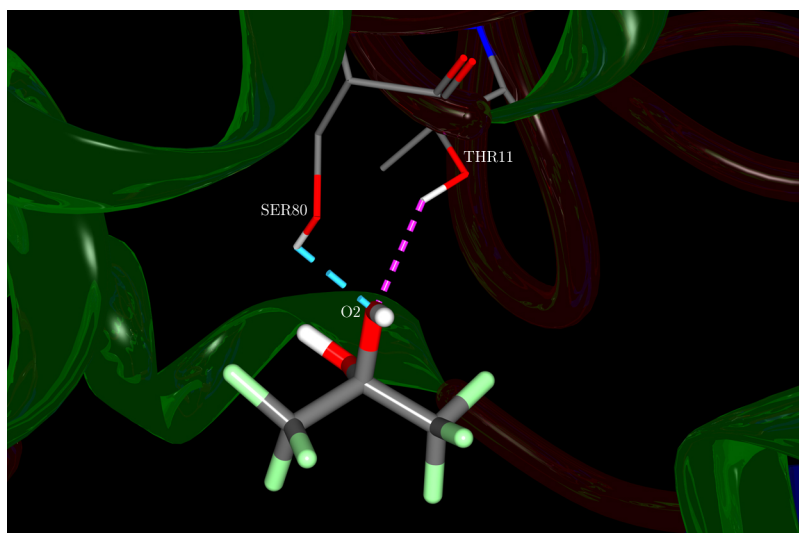
Abbildung 3.13. Darstellung der aktiven Zentren der im Text diskutierten Komplexe 1A08, 1A1C, 1ADB.



d) 1EBG

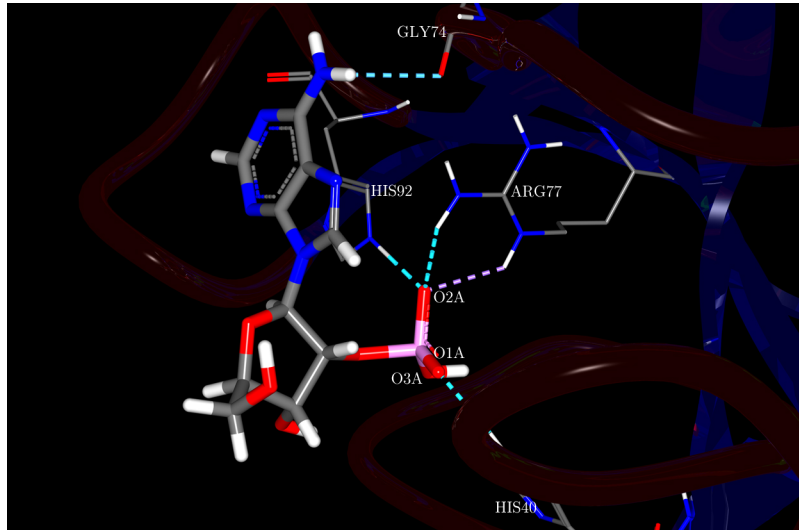


e) 2QWD

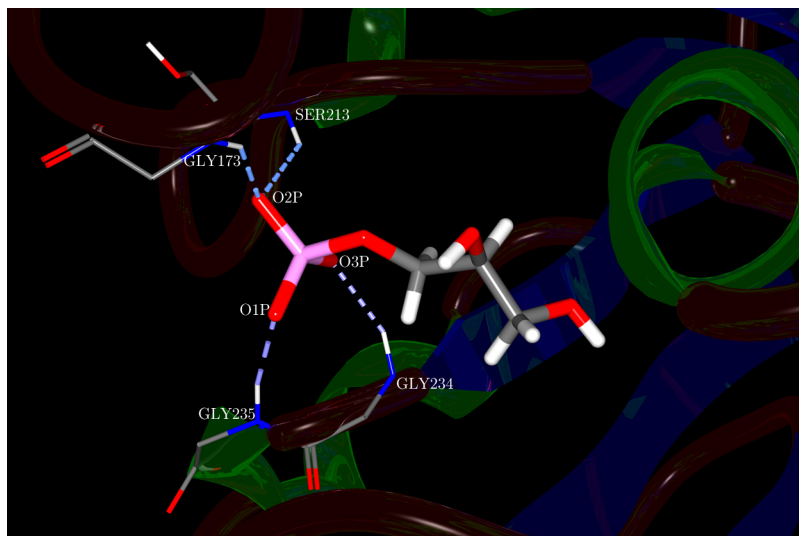


f) 5YAS

Abbildung 3.14. Darstellung der aktiven Zentren der im Text diskutierten Komplexe 1EBG, 2QWD, 5YAS



g) **6RNT**



h) **6TIM**

Abbildung 3.15. Darstellung der aktiven Zentren der im Text diskutierten Komplexe 6RNT, 6TIM

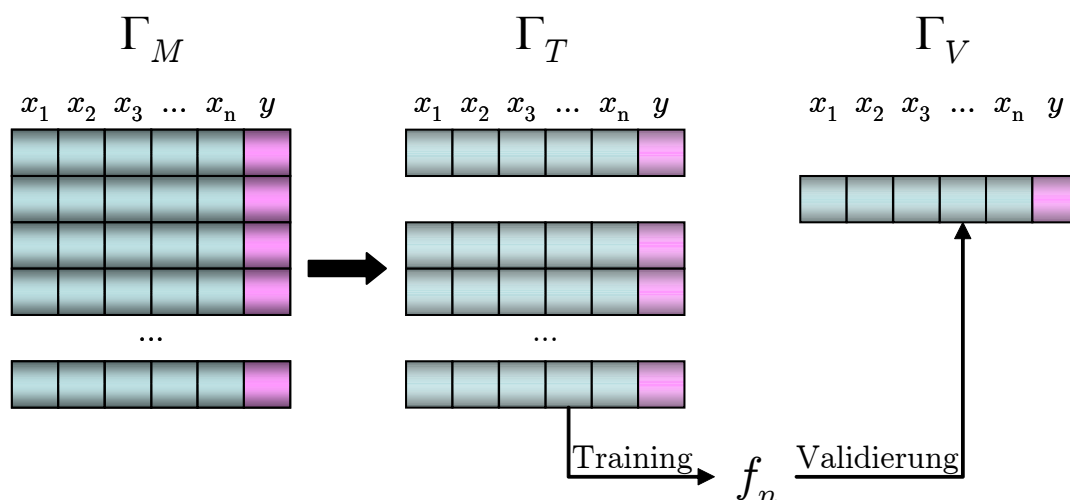


Abbildung 3.16. Schematische Illustration der Kreuzvalidierung.

3.5.2 Kreuzvalidierung

Bei der Entwicklung multivarianter Modelle wie AISCORE ist es sehr wichtig, deren Validität zu überprüfen. Eine Methode, die Gültigkeit eines Modells zu bewerten, ist die *Kreuzvalidierung* (*Cross Validation* oder auch *Boot Strapping*). Dabei wird die Menge der Modelldaten Γ_M (in unserem Fall also der Satz aus Protein–Ligand-Komplexen) in zwei sich gegenseitig ausschließende Teilmengen zerlegt: ein großer *Trainingssatz* Γ_T und eine kleiner *Testsatz* Γ_V , so dass gilt:

$$\Gamma_T \subset \Gamma_M \quad (3.7)$$

$$\Gamma_V \subset \Gamma_M \quad (3.8)$$

$$\Gamma_V \notin \Gamma_T \wedge \Gamma_V \cup \Gamma_T = \Gamma_M \quad (3.9)$$

Das zu überprüfende Modell wird jetzt mit Γ_T aufgestellt. Es wird also eine Version der Modellfunktion f_T mit einer Teilmenge der Daten parametrisiert. Die Werte der Parameter von f_T werden sich dadurch von denen, die mit dem vollständigen Datensatz bestimmt wurden, unterscheiden.

Anschließend wird die Gültigkeit von f_T anhand von Γ_V überprüft und bewertet (Abbildung 3.16). Dieser Prozess wird mit verschiedenen Γ_T und Γ_V wiederholt, bis jedes Objekt in Γ_M einmal in einer Testmenge Γ_V war.

Für die Kreuzvalidierung von AISCORE wurde der Protein–Ligand-Satz zufällig in zehn Paare von Trainingssatz Γ_{T_n} (mit 90% der Komplexe) und Testsatz Γ_{V_n}

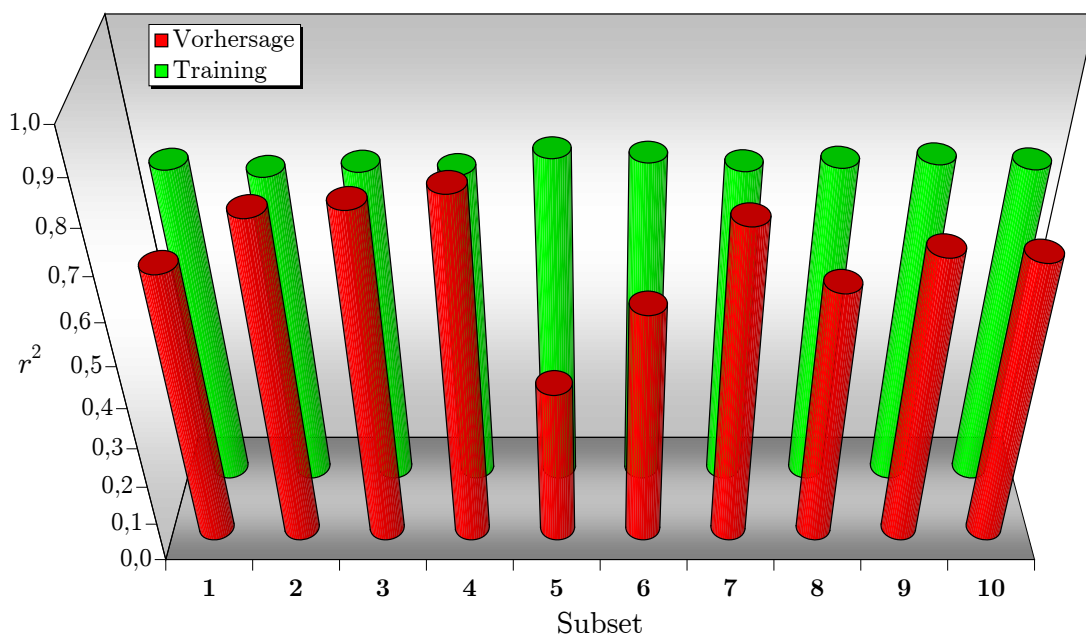


Abbildung 3.17. Lineare Korrelationskoeffizienten r^2 des Trainings und der anschließenden Vorhersage der Bindungsenergie für die zehn Subsets der Kreuzvalidierung.

(mit den restlichen 10% der Komplexe) aufgeteilt („Subsets“, $n = 1 \dots 10$). Für jeden Satz Γ_{Tn} wurde eine Version von AISCORE f_n bestimmt. Die Bindungsenergien der 10 Komplexe in den Γ_{Vn} wurden mit der entsprechenden Bewertungsfunktion f_n vorhergesagt.

Eine Übersicht der einzelnen linearen Korrelationskoeffizienten des Trainings und der Vorhersage für die 10 Subsets ist graphisch in Abbildung 3.17 dargestellt.

Der *leave-ten-out*-Korrelationskoeffizient q^2 ist dann

$$q^2 = 1 - \frac{\sum_i (\Delta G_{\text{score}}(i) - \Delta G_{\text{Exp}}(i))^2}{\sum_i (\Delta G_{\text{Exp}}(i) - \Delta G_{\text{Exp}}^{\text{mean}})^2} \quad (3.10)$$

$$(3.11)$$

wobei $\Delta G_{\text{score}}(i)$ die mit dem entsprechenden f_n geschätzten Bindungsenergie für dasjenige n ist, in dessen Γ_{Vn} sich der Komplex i befindet. $\Delta G_{\text{Exp}}^{\text{mean}}$ ist der Mittelwert der experimentellen Bindungsenergien.

Im Falle von AISCORE ist $q^2 = 0,59$.

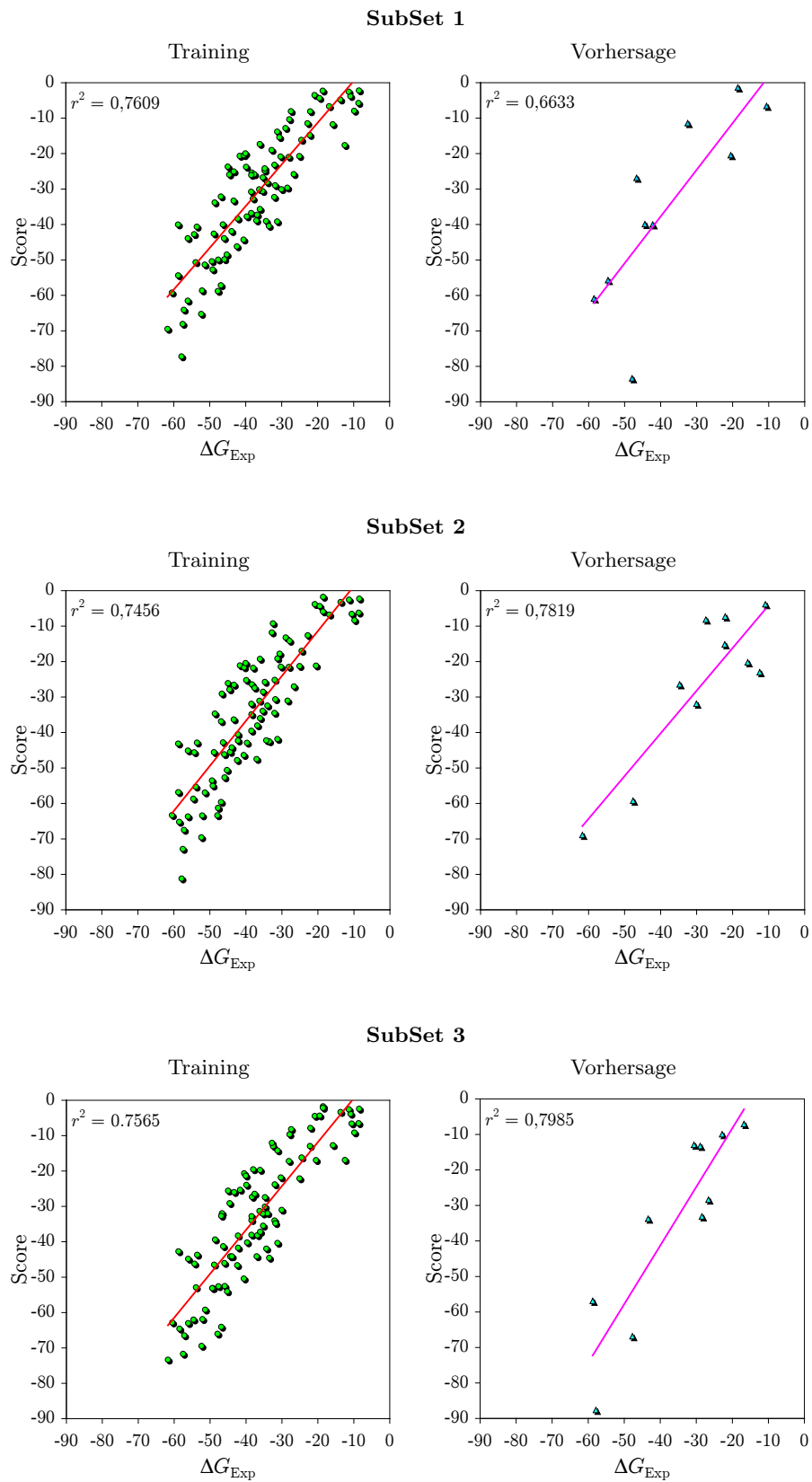
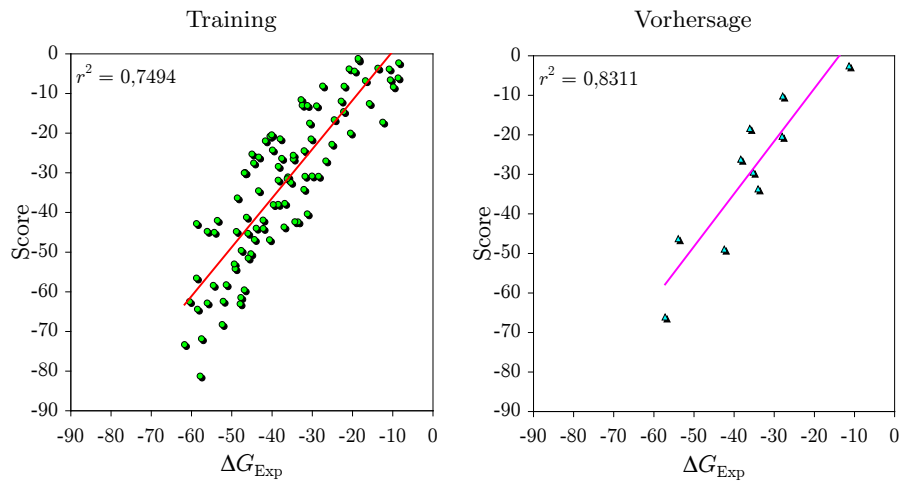
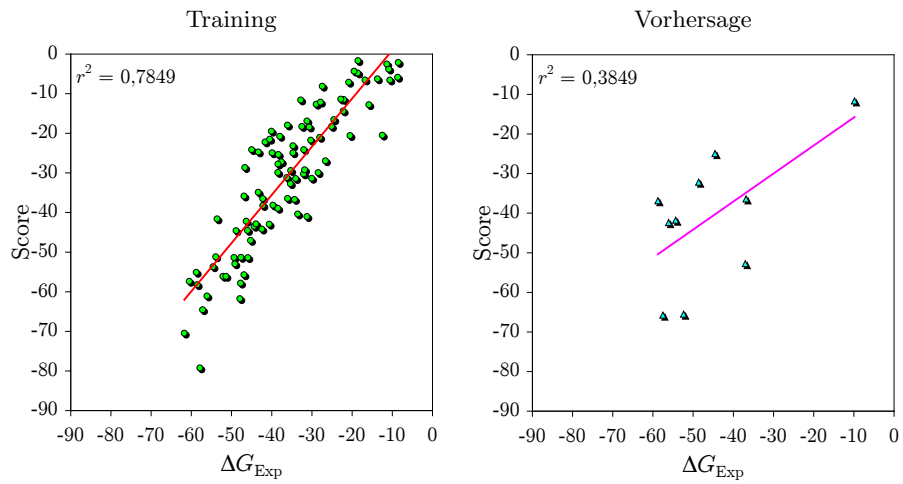


Abbildung 3.18. Lineare Korrelationen für Training und Vorhersage der zehnfachen Kreuzvalidierung.

SubSet 4



SubSet 5



SubSet 6

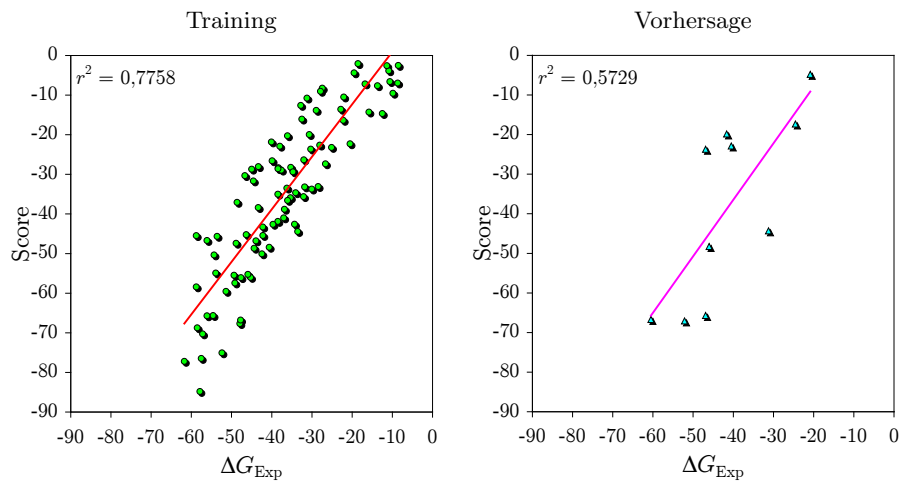
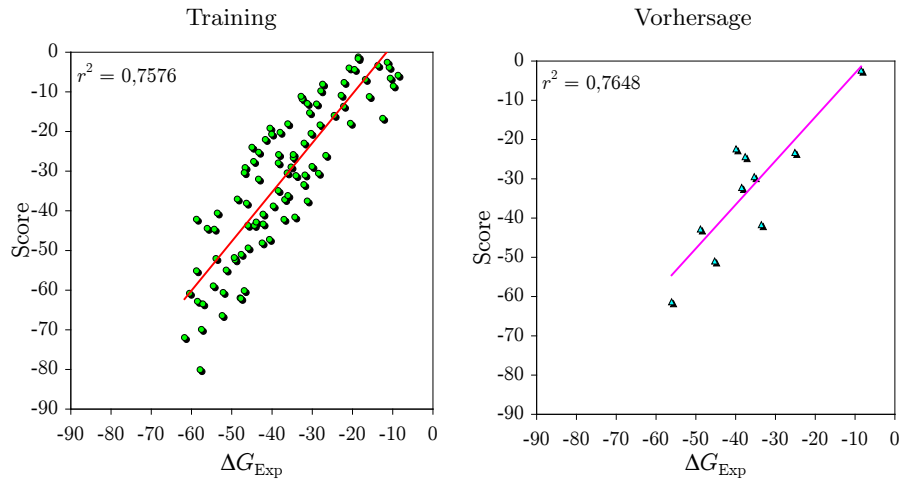
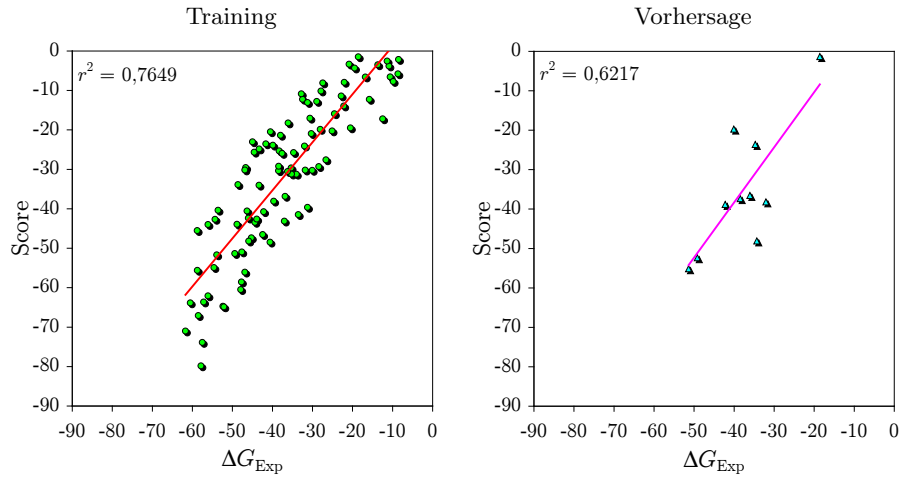


Abbildung 3.19. Lineare Korrelationen für Training und Vorhersage der zehnfachen Kreuzvalidierung. *Fortsetzung...*

SubSet 7



SubSet 8



SubSet 9

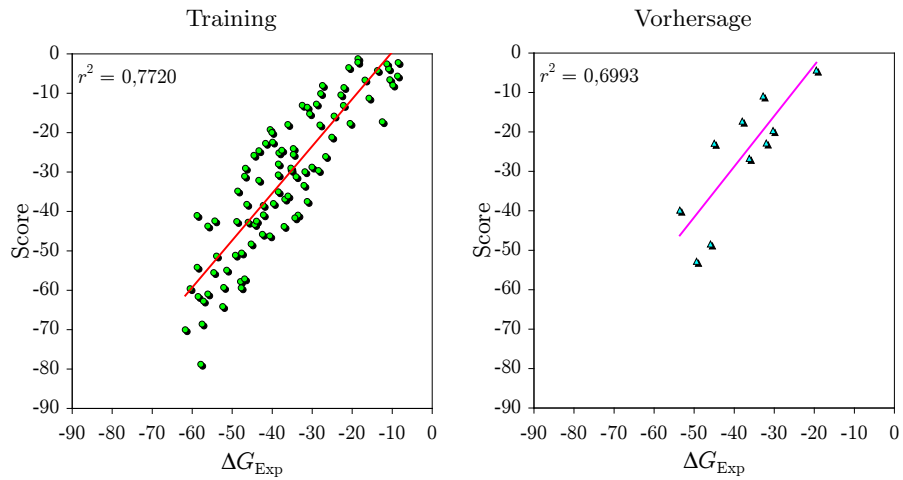


Abbildung 3.20. Lineare Korrelationen für Training und Vorhersage der zehnfachen Kreuzvalidierung.

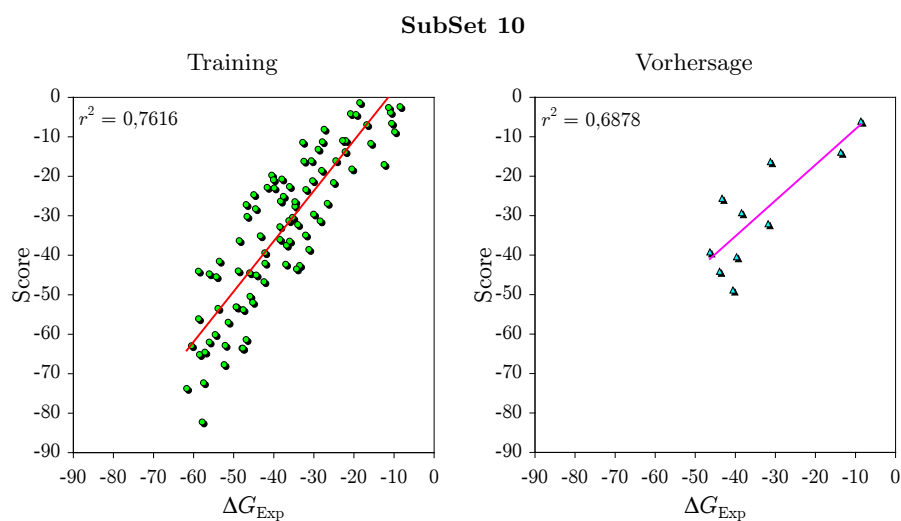


Abbildung 3.21. Lineare Korrelationen für Training und Vorhersage der zehnfachen Kreuzvalidierung.

3.5.3 Redocking

Die Bewertungsfunktion wird in FLEXX nicht nur zur Vorhersage der freien Bindungsenergie des kompletten Liganden im Komplex benutzt. Auch während des inkrementellen Aufbaus werden die Teilstrukturen einer Iteration energetisch bewertet und nur die besten Lösungen kommen in die nächste Iteration (*Greedy*-Strategie, vgl. Kapitel 3.3.2). Damit greift die Bewertungsfunktion massiv in den inkrementellen Aufbau ein und lenkt ihn dadurch, dass bestimmte Teilfragmente höher bewertet werden als andere. Ob eine Konfiguration, die der nativen Form des Liganden im Komplex nahe kommt, überhaupt generiert wird, hängt deswegen auch von der Bewertungsfunktion ab. Schlussendlich dient die Bewertungsfunktion FLEXX auch dazu, die unterschiedlichen Platzierungen eines Liganden in einer Bindetasche untereinander zu bewerten. Selbst, wenn also die eine sinnvolle Platzierung generiert wurde, muss die Bewertungsfunktion sie aus der Menge aller Platzierung herausfinden.

3.5.3.1 Kumulative Erfolgsraten

Um zu sehen, inwieweit die neue Bewertungsfunktion AISCORE die Docking-Fähigkeiten von FLEXX beeinflusst, wurden *Redocking-Experimente* durchgeführt, d.h. für alle 100 Komplexe wurde der Ligand der Kristallstruktur mittels der Docking-Algorithmen von FLEXX zurück in die Bindetasche einplatziert. Als

Kriterium für die Leistung wird die RMSD (*root mean square deviation*) der gedockten Konformation im Vergleich zur Konformation in der Kristallstruktur verwendet.

Aus dem RMSD lässt sich die *kumulative Erfolgsrate* $s(r)$ ableiten. Das ist der Anteil derjenigen Komplexe im kompletten Satz, die eine Lösung mit einem $\text{RMSD} < r$ aufweisen. Hier werden noch einmal zwei Arten der kumulativen Erfolgsrate unterschieden: $s^{(1)}(r)$ bezeichnet die kumulative Erfolgsrate der Lösungen auf *Rang 1*, $s^{(n)}(r)$ bezieht sich auf die Lösung mit dem kleinsten RMSD unter *allen* generierten Platzierungen eines Komplexes.

In Abbildung 3.22 sind die kumulativen Erfolgsraten $s^{(1)}(r)$ und $s^{(n)}(r)$ in Abhängigkeit von r für die verschiedenen Bewertungsfunktionen gezeigt. Für den Graphen im oberen Teil von Abbildung 3.22 wurden nur Platzierungen auf Rang 1 berücksichtigt ($s^{(1)}(r)$), für den Graphen im unteren Teil die jeweils beste Platzierung ($s^{(n)}(r)$). Beide Plots sind ähnlich, weshalb nur die Platzierungen mit Rang 1 diskutiert werden sollen.

FLEXSCORE erreicht eine gute kumulative Erfolgsrate von 52% bei 2,5 Å, was sogar noch auf 55% verbessert werden kann, wenn der XFURCATE-Algorithmus zusätzlich verwendet wird. Im Vergleich dazu ist Redocking-Leistung von AISCORE schlechter. Bei 2,5 Å können mit AISCORE 40% erreicht werden, mit XFURCATE 44%.

Die Redocking-Fähigkeiten des erweiterten FLEXSCORE liegen etwa zwischen denen von AISCORE und dem originalen FLEXSCORE. Das zeigt, dass die Optimierung der Parameter der Bewertungsfunktion allein auf die Maximierung der linearen Korrelation nicht der einzige Grund für die eher mittelmäßige Leistung von AISCORE im Redocking sein kann. Dass AISCORE trotz der dramatisch besseren Leistung gegenüber FLEXSCORE in der Berechnung der freien Bindungsenergie beim Redocking schlechter ist als FLEXSCORE, legt eher die Vermutung nahe, dass eine physikochemische Bewertung von Teilplatzierungen im Verlauf des inkrementellen Aufbaus ungeeignet ist.

Eine gründliche Analyse zeigt, dass die chemisch diversen Energieinkremente zu einer Zunahme von Abbrüchen während des inkrementellen Aufbaus führen. Ein Grund für einen solchen Abbruch ist, dass die aktuelle Teilplatzierung so ungünstig liegt, dass jedes weitere Fragment zu großen *Clashes* mit der Bindeta-sche führt und alle Erweiterungen des Lösungssatzes somit abgelehnt werden.

In diesen Fällen hat die Bewertungsfunktion den inkrementellen Aufbau in eine falsche Richtung dirigiert. Für dieses Verhalten gibt es zwei vorstellba-

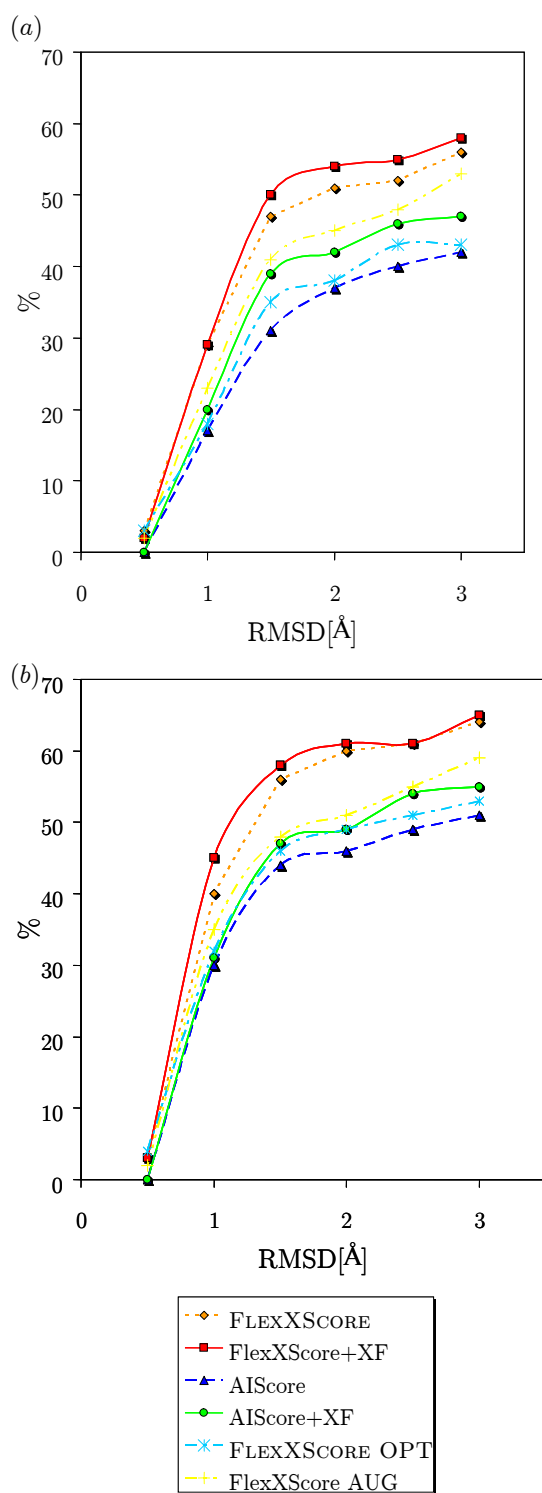


Abbildung 3.22. Vergleich der kumulativen Erfolgsraten in Abhängigkeit vom entsprechenden RMSD-Wert für *a*) Platzierungen auf Rang 1 und *b*) für die jeweils beste Platzierung mit einem beliebigen Rang.

re Gründe: Ein offensichtlicher Grund ist, dass die Beiträge, die schwache H-Brücken zur *Match Score* beitragen zu gering sind, um das Auswahlkriterium des *Greedy-Algorithmus* zu erfüllen. Durch die hohe Diversität der AISCORE-Inkrementen für die Wasserstoffbrücken-Donoren und -Akzeptoren unterscheiden sich die *Match Scores* um bis zu 60 kJ Mol^{-1} . Dadurch kann es dazu kommen, dass nicht-native Platzierungen mit starken Wasserstoffbrücken native Platzierungen mit schwächeren Wasserstoffbrücken verdrängen. Dieser Effekt tritt bei FLEXSCORE weit weniger drastisch auf, da dort alle H-Brücken nahezu gleich bewertet werden und somit auch die *Match Scores* der verschiedenen Teilplatzierungen nahe beieinander liegen.

Ein zweiter Grund für die schlechtere Performance von AISCORE beim Redocking ist, dass der Algorithmus zur Abschätzung der noch zu erwartenden Wechselwirkungsbeiträge der noch nicht gedockten Fragmente, an der Diversität der AISCORE-Inkrementen scheitert. Das kann an drei Fällen gezeigt werden:

Fragmente mit schwachen H-Brücken

Wenn dieses Fragment bisher noch in keinem Lösungssatz einplatziert wurde, wird sein möglicher Beitrag anhand seiner Wechselwirkungsgeometrien abgeschätzt. Handelt es sich dabei um schwache Inkrementen, so wird auch nur ein kleiner möglicher Beitrag angenommen, obwohl mit einem Wechselwirkungspartner auf Proteinseite, der ein starkes Inkrement besitzt, eine hohe *Match Score* resultieren könnte. Doch durch die Unterschätzung kann es dazu kommen, dass dieser Lösungssatz verdrängt wird.

Fragmente mit starken H-Brücken

Für ein Fragment mit starken H-Brücken, das in keinem anderen Lösungssatz bisher einplatziert wurde, wird der zu erwartende Beitrag zur Komplexbindungsenergie anhand *aller* Wechselwirkungsgeometrien dieses Fragmentes abgeschätzt. Das führt sowohl im Falle von AISCORE als auch FLEXSCORE zu einer Überbewertung, da im Komplex selten alle Wechselwirkungsgeometrien *Matches* bilden. Diese Überbewertung findet natürlich auch bei FLEXSCORE statt, wird jedoch bei AISCORE durch Inkrementen, die Werte bis zu -30 kJ Mol^{-1} haben, verschärft.

Fragmente aus anderen Lösungssätzen

Wird ein noch zu platzierendes Fragment in einem anderen Lösungssatz gefunden, so wird die dort erreichte *Match Score* zur Abschätzung herangezogen. Durch

die neue Diversität der Inkremente in AISCORE, ist die *Match Score* sehr sensibel auf den chemischen Kontext der Bindungspartner. Deshalb erscheint eine Abschätzung der potentiellen Beiträge zur Wechselwirkungsenergie anhand von anderen Platzierungen nicht geeignet.

Ohne eine vergleichbare Abschätzung der Bindungsenergie für die Platzierungen ausgehend von allen Basisfragmenten wird die Sinnhaftigkeit der *Greedy*-Strategie in Frage gestellt. Hier muss eine Anpassung an die Diversität der Wechselwirkungsenergien erfolgen.

3.5.3.2 HIV1-Proteasen und die Frage nach dem RMSD-Wert als Kriterium

Es besteht ein grundsätzliches Problem bei dem Vergleich von Platzierungen, was am Beispiel der Komplexe der HIV1-Protease erläutert werden soll.

Innerhalb des Protein-Ligand-Datensatzes ist die Gruppe der HIV1-Protease-Komplexe mit 22 Vertretern die größte. Beim Redocking konnte FLEXX unter Verwendung von FLEXXSCORE für 7 Komplexe keinerlei Platzierung generieren (1HIH, 1HIV, 1HVJ, 1HVK, 1HVL, 9HVP, 1HVS).

Im Falle von 1HIH konnte jedoch mit AISCORE eine sehr gute Platzierung auf Rang 1 mit einem RMSD von 1,33 Å gefunden werden. Die AISCORE-Energie dieser Platzierung ist mit $\Delta G_{\text{AIScore}} = -49,58 \text{ kJ Mol}^{-1}$ nahe am experimentellen Wert von $\Delta G_{\text{Exp}} = -45,93 \text{ kJ Mol}^{-1}$.

Für die Komplexe 1HIV, 1HVL, 9HVP und 1HVS konnten mit AISCORE zwar Platzierungen gefunden werden, jedoch weisen diese Platzierungen sehr große RMSD-Werte auf, was im weiteren Verlauf noch diskutiert wird. Für die Komplexe 1HVJ, 1HVK, 4HVP und 1MES konnte mit AISCORE gar keine Platzierung gefunden werden.

Insgesamt zeigen sowohl für AISCORE als auch für FLEXXSCORE die Platzierungen auf Rang 1 große RMSD-Werte. Auch die Platzierungen mit der jeweils geringsten Abweichung zeigen RMSD-Werte von deutlich über 9 Å. Man kann drei große Gruppen von Abweichungen unterscheiden: *falsche Lage*, *falsche Orientierung* und *vertauschte Gruppen*.

Bei einer *falschen Lage* ist der Ligand als Ganzes an einer ganz anderen Stelle in der Bindetasche einplatziert worden, als es in der Kristallstruktur beobachtet werden kann. Hier werden bereits die Basisfragmente mit anderen Wechselwirkungsgruppen überlagert als in der nativen Struktur und der Ligand von dieser Basisfragmentplatzierung ausgehend weiter aufgebaut. Dabei können durchaus

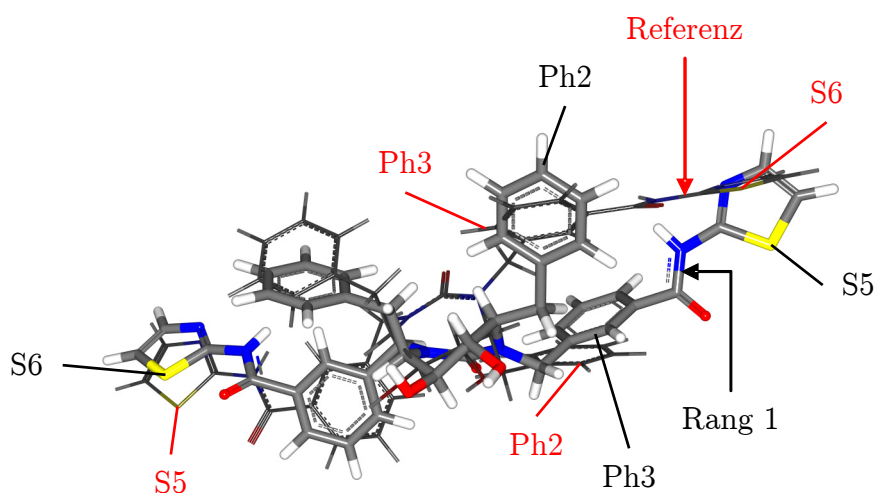


Abbildung 3.23. Vergleich der nativen Struktur (*dünn*) und der von AIScore auf Rang 1 bewerteten Struktur (*dick*) des Liganden im Komplex 1BV7.

sinnvolle und stark bindende Konfigurationen erzeugt werden, die aber nicht der nativen Struktur entsprechen. Unter Verwendung von AIScore ist das für die Komplexe 1AAQ, 1HBV und 1MEU der Fall.

Liganden, die einen Großteil der auch im Kristall beobachteten Wechselwirkungen ausbilden, aber verdreht sind oder teilweise eine andere Konformation einnehmen, bezeichne ich mit „*falsche Orientierung*“. Im Extremfall werden bei einem symmetrischen Liganden alle Wechselwirkungen richtig ausgebildet, aber der Ligand ist entsprechend seiner Symmetrie gedreht. Ein Beispiel dafür ist die Lösung auf Rang 1 von 1BV7 (Abbildung 3.23). Ein Vergleich mit der nativen Position des Liganden zeigt eine gute Übereinstimmung, jedoch ist der Ligand um 180° gedreht, sodass FlexX einen sehr großen RMSD berechnet, da die „falschen“ Atome für die Berechnung verglichen werden. Zusätzlich sind die Phenylyringe Ph2 und Ph3 vertauscht gegenüber der Kristallstruktur.

Ein ähnlicher Fall liegt bei der Platzierung von 1MEU (Abbildung 3.24) vor. Ein Teil des Moleküls wurde richtig einplatziert, aber an einer Amidgruppe die falsche Konformation gewählt. Dadurch werden im Vergleich zur nativen Struktur ein Phenylyring und ein Methoxyphenylyring miteinander vertauscht. Da aber die dadurch vertauschten Gruppen sehr ähnlich sind, sind die beiden Möglichkeiten rein energetisch nur schwer zu unterscheiden. Beide Lösungen erscheinen gleich plausibel. Auch hier wird ein großer RMSD-Wert berechnet.

Das Problem übermäßig hoher RMSD-Werte bei symmetrischen Liganden kann mit einer neuen Routine für die RMSD-Berechnung von U. KORBEL und

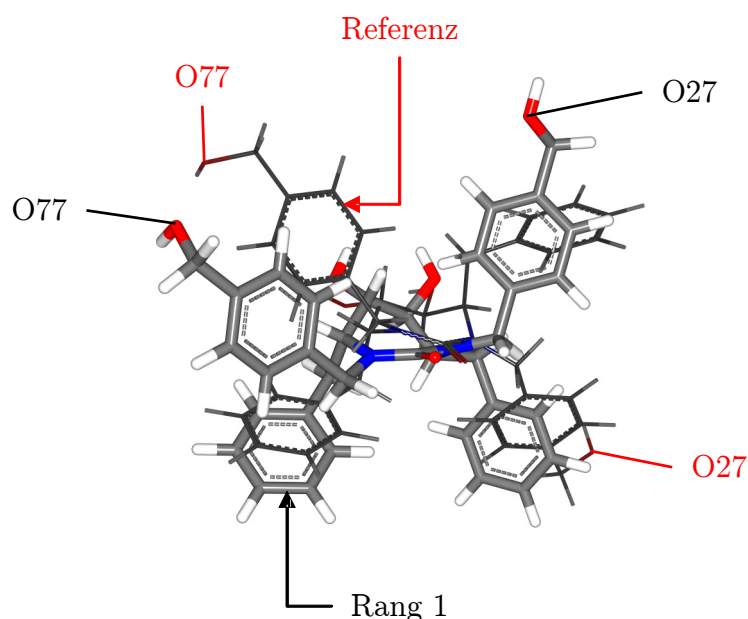


Abbildung 3.24. Vergleich der nativen Struktur (*dünn*) und der von AIScore auf Rang 1 bewerteten Struktur (*dick*) des Liganden im Komplex 1MEU.

A. KÄMPER, die bald in FlexX implementiert werden wird, gelöst werden. Diese neue Routine kann lokale und globale Symmetrie von beliebigen Molekülen und Komplexen handhaben. [165]

Bei vielen Komplexen der HIV1-Protease kommen solche vollständigen oder teilweisen Verdrehungen des Liganden beim Einplatzieren vor. Von den 22 HIV1P-Komplexen findet FLEXX unter Verwendung von AIScore für 6 Komplexe Platzierungen, bei denen so gut wie alle Wechselwirkungszentren richtig überlagert werden, aber die globale Orientierung des Liganden teilweise oder vollständig falsch ist.

Bei den restlichen Komplexen werden die richtigen Wechselwirkungszentren der Proteinbindetasche besetzt, jedoch von anderen Gruppen des Liganden als in der nativen Struktur. Dadurch liegt der Ligand grob an der richtigen Stelle, aber in einer anderen Orientierung. Als Beispiel sei der Komplex 1HIV (Abbildung 3.25) erwähnt. Hier werden im Vergleich zur nativen Platzierung der Hexyl-Ring und eine der beiden Amidgruppen gut getroffen und der Imidazolring in einer verkippten Variante platziert. Aber die Lage der übrigen Gruppen ist von der nativen Platzierung verschieden.

Vor allem bei den großen, aber ungerichteten Wechselwirkungsgeometrien aromatischer Wechselwirkungen können Vertauschungen energetisch kaum unterschieden werden, da es nur 2 verschiedene aromatische Energieinkremente für

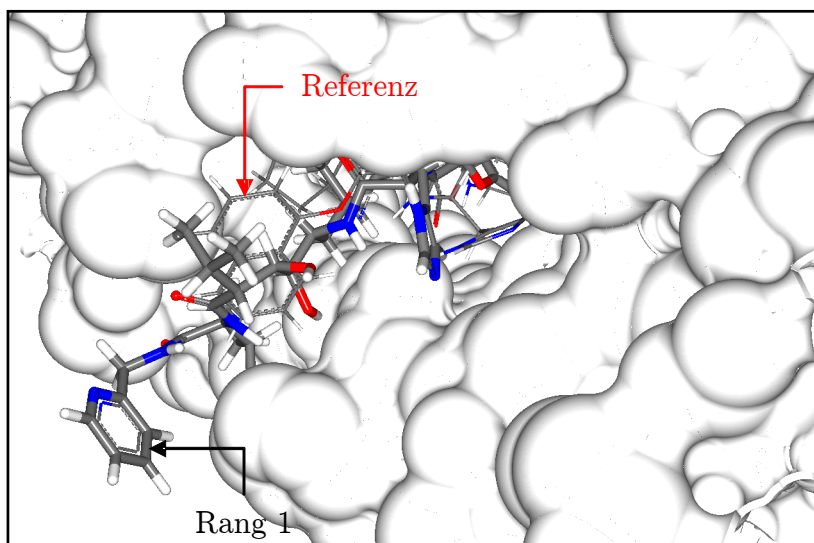


Abbildung 3.25. Vergleich der nativen Struktur (*dünn*) und der von AIScore auf Rang 1 bewerteten Struktur (*dick*) des Liganden im Komplex 1HIV in der Oberflächendarstellung des Proteins.

aromatische Wechselwirkungen in AIScore gibt – bei FLEXSCORE sogar nur eins. Gleichzeitig werden starre Ringe, mit Heteroatomen als Zentren für gerichtete Wechselwirkungen, bevorzugt als Basisfragmente ausgewählt. Offensichtlich finden solche Basisfragmente sehr viele mögliche Platzierungen. Im Zuge der *Greedy*-Strategie werden dann aber in vielen Fällen die nativ-ähnlichen Platzierungen zugunsten von stärker bindenden Teilplatzierungen an anderer Stelle verworfen.

Insofern stellt sich die Frage, ob der RMSD-Wert allein überhaupt ein geeignetes Kriterium für die Beurteilung der Docking-Fähigkeiten eines Ansatzes ist!

In einer voluminösen Bindetasche kann ein Ligand sicher an mehreren Stellen platziert werden. Alle Platzierungen können energetisch sehr ähnlich sein. Ohne weitere Informationen wäre eine Platzierung so gut wie jede andere. Dass ein Docking-Ansatz unter den vielen energetisch so gut wie gleichen Lösungen dann diejenige identifiziert, die auch im Kristall beobachtet wurde, erscheint im Lichte dieser Überlegungen eher zufällig.

Als Alternative könnte vielmehr die Anzahl und Art der ausgebildeten Wechselwirkungen verglichen werden. An einer Berechnung der Ähnlichkeit einer Platzierung im Vergleich zur nativen Struktur anhand der ausgebildeten Kontakte arbeite ich zurzeit.

3.6 Anwendung auf artifizielle Rezeptoren

Das computergestützte *de novo*-Design hoch-spezifischer synthetischer Rezeptoren für einen gegebenen Liganden läuft auf das Konstruieren eines Rezeptors um diesen Liganden herum hinaus. Es ist quasi die Umkehrung des Dockens.

3.6.1 FlexR

FLEXR [162] basiert auf den grundlegenden Algorithmen von FLEXX und kann flexible Liganden in flexible Rezeptoren docken. In Komplexen aus zwei kleinen Molekülen sind die Rollen von Rezeptor und Ligand vertauschbar. Deshalb erlaubt FLEXR dass der Rezeptor inkrementell um den Liganden herum aufgebaut wird. Dadurch eröffnet sich die Möglichkeit des *Screening* nach potenziellen artifiziellen Rezeptoren.

Die konformationelle Flexibilität *beider* Komplex-Partner wird in FLEXR dadurch gelöst, dass der Ligand flexibel nacheinander in verschiedene mögliche Konformere des Rezeptors gedockt wird oder umgekehrt. Die Generierung der Rezeptor-Konformere geschieht, ähnlich dem inkrementellen Aufbau in FLEXX, durch Sätze diskreter Torsionswinkel aus der MIMUMBA-Datenbank an azyklischen Einfachbindungen und für Ringsysteme mit CORINA. Das eigentliche Docken nutzt die FLEXX-Algorithmen für den inkrementellen Aufbau. Der Ligand wird nacheinander in jede der Rezeptorkonformere gedockt. Zum Schluss werden die Lösungen der verschiedenen Docking-Läufe zusammengefasst und energetisch bewertet.

3.6.2 AIScore und der Koffeinrezeptor

Alkylierte Oxypurine sind die am häufigsten konsumierten Alkaloidwirkstoffe. [166] Der häufigste überhaupt ist Koffein (**1**, Abbildung 3.26). WALDVOGEL *et al.* [167] haben einen artifiziellen Rezeptor synthetisiert, der alkylierte Oxypurine dreidimensional als C_3 -symmetrische Klammer umgreift (Abbildung 3.27). Die Methylgruppen des Koffeins passen dabei genau in die Freiräume zwischen den Ankergruppen.

WALDVOGEL *et al.* haben für den Komplex aus Koffein und ihrem Koffeinrezeptor durch NMR-Titration die Komplexbildungskonstante bestimmt und darüber hinaus noch die von verschiedenen Substratvariationen (Abbildung 3.26). Außer von dem Komplex mit Koffein sind jedoch keine experimentellen Strukturen bekannt.

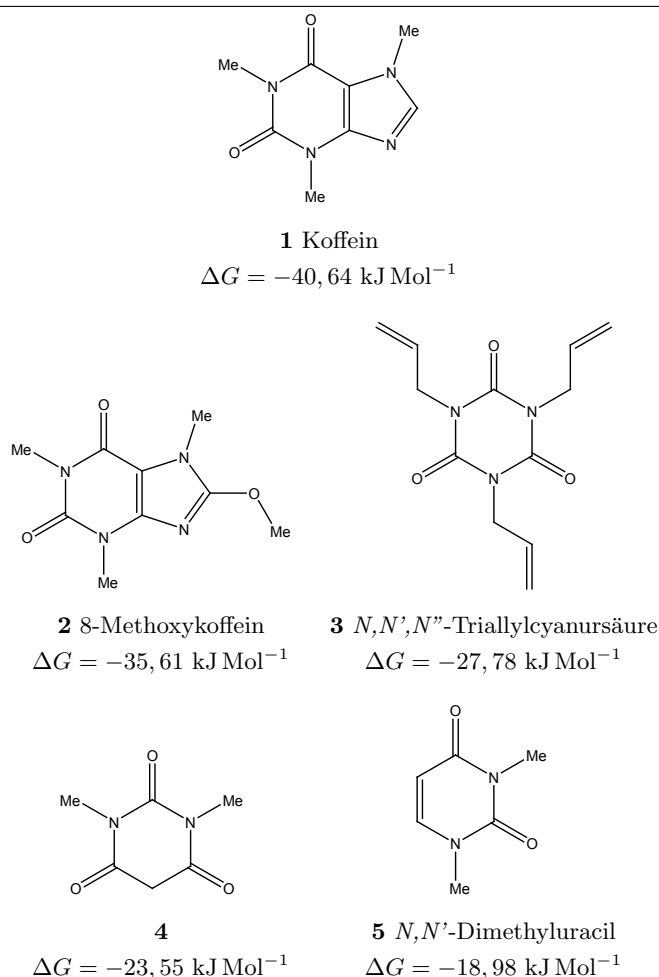


Abbildung 3.26. Die verschiedenen Liganden die WALDVOGEL *et al.* in ihren synthetischen Koffeinrezeptor komplexiert haben.

Trotzdem ist dieser Satz von ähnlichen Liganden im selben Rezeptor ideal als Test für AISCORE in Hinblick auf die Anwendbarkeit auf artifizielle Rezeptoren.

3.6.2.1 Datenaufbereitung und Docking

Der Rezeptor wurde aus der publizierten Kristallstruktur [167] übernommen und die aliphatischen C_6H_{13} -Reste abgeschnitten. Danach wurden manuell chemisch sinnvolle SYBYL-Atomtypen zugewiesen. In Ermangelung von Kristallstrukturen wurden die Liganden per Hand erstellt und jeweils mit dem MMFF94-Kraftfeld [158] geometrieoptimiert.

AISCORE konnte direkt in FLEXR übernommen werden.

Die Liganden wurden jeweils mit FLEXR in den Koffeinrezeptor gedockt und die Energien der Platzierungen auf Rang 1 ausgewertet.

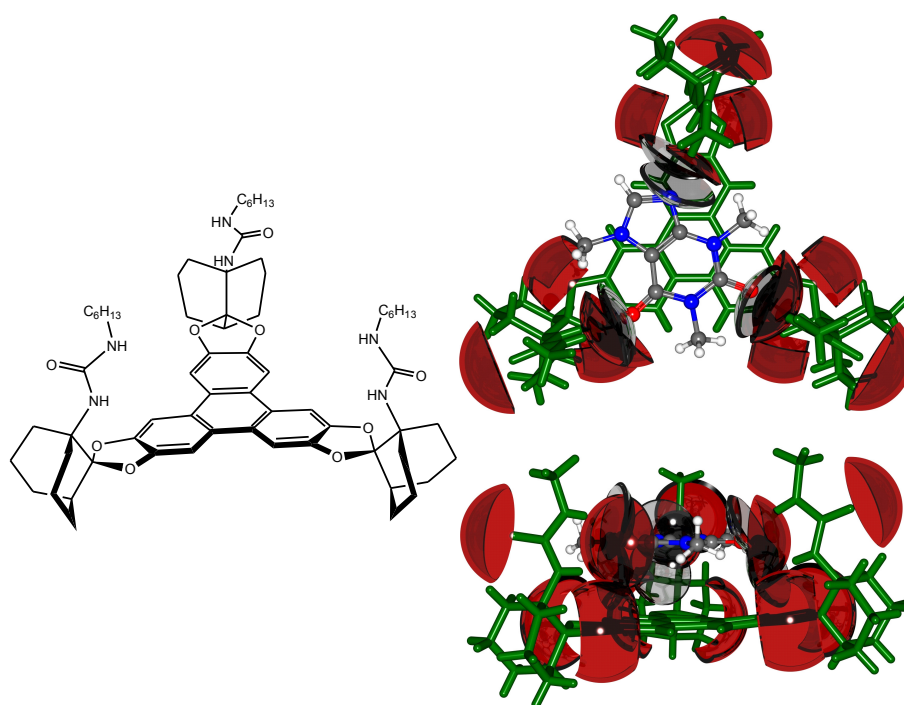


Abbildung 3.27. Der Koffeinrezeptor von WALDVOGEL *et al.*

3.6.2.2 Ergebnisse und Diskussion

Abbildung 3.28 zeigt die linearen Korrelationen zwischen den experimentellen Bindungsenergien der verschiedenen Liganden im Komplex mit dem Koffeinrezeptor und den berechneten Energie jeweils von FLEXXSCORE und AISCORE.

In beiden Fällen fällt der Ausreißer **3** auf. Die drei Allylketten bilden selbst in der besten Platzierung große *Clashes* mit dem Liganden, die von den attraktiven *Match Scores* nicht vollständig kompensiert werden können. Dies widerspricht dem Experiment, scheint aber ein Effekt des inkrementellen Aufbaus zu sein. Der Sechsring wird zum Basisfragment erklärt und für sich einplatziert, entsprechend seiner Wechselwirkungsgeometrien. Von dieser Einplatzierung ausgehend, können die Allylketten aber immer nur so aufgebaut werden, dass sie mit dem Rezeptor kollidieren. Es wäre für die Basisfragmentplatzierung eine energetisch etwas ungünstigere Platzierung, bei der der Ring verdreht liegt, nötig. Der Verlust an Bindungsenergie zur idealen Basisfragmentplatzierung würde im späteren Verlauf des inkrementellen Aufbaus durch die Vermeidung der *Clashes* kompensiert, was zum Zeitpunkt der Basisfragmentplatzierung aber nicht bekannt ist. Hier scheint einer der wenigen Fälle zu sein, wo gerade das unabhängige Platzieren des Basisfragmentes zu ungewünschten Lösungen führt.

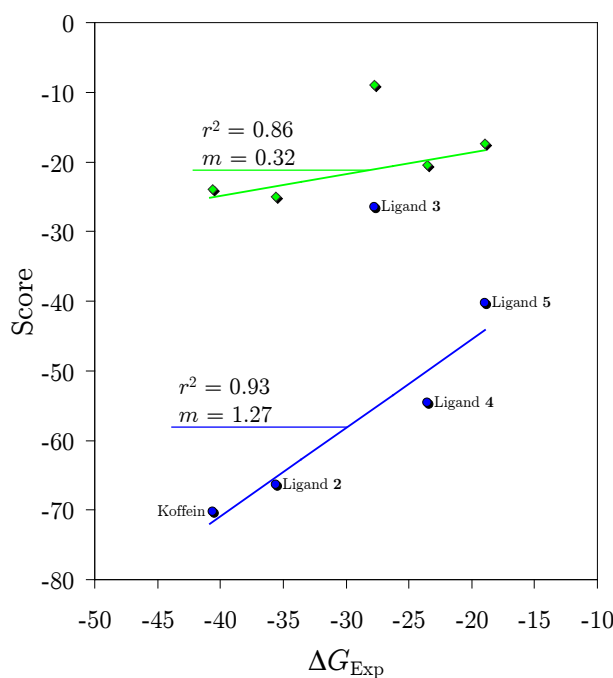


Abbildung 3.28. Vergleich der linearen Korrelationen.

Für die Betrachtungen der Korrelationen wurde deshalb **3** ignoriert. Sowohl FLEXXSCORE als auch AISCORE erzielen mit $r^2 = 0,86$ bzw. $r^2 = 0,93$ sehr gute lineare Korrelationen. Jedoch ist die Steigung der Ausgleichsgeraden im Falle von FLEXXSCORE sehr flach (Steigung $m = 0,32$). Die berechneten Bindungsenergien sind alle mehr oder weniger gleich.

Im Falle von AISCORE ist die Ausgleichsgeraden mit $m = 1,27$ fast ideal. Der Trend der experimentellen Bindungsenergien wird richtig wiedergegeben. Die Energien liegen jedoch insgesamt gut $20 - 30 \text{ kJ Mol}^{-1}$ zu tief, was jedoch nicht weiter überraschend ist. Die Wechselwirkungsenergie-Inkrementen von AISCORE wurden an experimentelle Werte von Protein-Ligand-Komplexen *in wässriger Lösung* angepasst. Die Messungen von WALDVOGEL *et al.* jedoch wurden in CDCl_3 durchgeführt.

In diesem ersten Test konnte AISCORE somit für einen artifiziellen Rezeptoren sowohl den inkrementellen Aufbau richtig lenken als auch anschließend die Bindungsenergien verschiedener Liganden im selben Rezeptor *im Verhältnis zueinander* gut berechnen. Die Energien sind aufgrund des von Wasser verschiedenen Lösungsmittel zu tief. Weitere Tests und eine Rekalibrierung auf das andere Lösungsmittel sind nötig, doch diese ersten Ergebnisse sind ermutigend.

3.7 GeomGen5

Im Zuge meiner Arbeit an AISCORE ist das Programm GEOMGEN5 entstanden.

GEOMGEN5 ist ein hochflexibles Programm zur multivarianten Anpassung von empirischen Bewertungsfunktionen vom LUDI-Typ an einen beliebigen Satz von Komplexen. Für die Optimierung der Parameter der Bewertungsfunktion stehen drei verschiedene numerische Algorithmen zur Verfügung. Die Parameter können jeder für sich optimiert oder eingefroren werden.

GEOMGEN5 besitzt eine eigene Implementierung einer allgemeinen LUDI-Bewertungsfunktion und kann mit beliebig vielen Wechselwirkungsgeometrien und Wechselwirkungstypen umgehen. Außerdem kann der XFURCATE-Algorithmus nach Wunsch dazu geschaltet werden.

Zusätzlich besitzt GEOMGEN5 die Möglichkeit, FLEXX aufzurufen und die Ergebnisse für Energien und Platzierungen einzusammeln. Dabei verteilt GEOMGEN5 alle notwendigen FLEXX-Aufrufe an beliebig viele Prozessoren und startet diese völlig *parallel*. Es können dabei sowohl *homogene* wie *inhomogene* Cluster genutzt werden.

Zu guter Letzt kann GEOMGEN5 Analysen durchführen, wie *n*-fache Kreuzvalidierung, Korrelationen und Redocking.

Der Name „GeomGen“ zeugt noch von der Herkunft des Programms: entstanden ist es ursprünglich als sehr simples Tool, das aus einem Vektor von Parametern ein `geometry.dat`-Files für FLEXX erzeugt. Nach und nach habe ich GEOMGEN immer mehr erweitert, entsprechend der Aufgabenstellungen, die sich im Zuge meiner Arbeit an AISCORE ergaben. Dabei habe ich von Anfang an alle Algorithmen und I/O-Routinen so allgemein wie möglich gehalten.

GEOMGEN5 ist in Fortran95 geschrieben und arbeitet mit mehreren allgemein gehaltenen *Modulen*, die für die *Datenkapselung* und die logische Bündelung von Funktionen verantwortlich sind.

Alle Optimierungen und Analysen, die hier vorgestellt wurden, sind mit GEOMGEN5 durchgeführt worden.

3.7.1 Eingabedateien

GEOMGEN5 benötigt mindestens zwei Eingabefiles: Eine Liste von Komplexen und ein `geometry.dat`-File für FLEXX.

Wechselwirkungsgeometrien

GEOMGEN5 kann `geometry.dat`-Files, wie sie von FLEXX verwendet werden, direkt einlesen. Für jede gefundene Wechselwirkungsgeometrie wird ein Objekt mit Name, den Parametern, einem Zähler für die Häufigkeit und einem Status-Feld angelegt. Als erlaubte Status gibt es `opt`, `freeze` und `scale`. `opt` erlaubt eine freie Optimierung, `freeze` lässt den Parameter unverändert und `scale` skaliert ihn mit dem aktuellen Wert des Solvatationsfaktors f_s . Die jeweiligen Schlüsselwörter werden als Kommentarzeile innerhalb der Definition der jeweiligen Wechselwirkungsgeometrie im `geometry.dat`-File eingefügt.

Die Komplexliste

In der Komplexliste stehen die *Namen* der Komplexe und deren *experimentelle Bindungsenergien*. GEOMGEN5 liest die Liste ein und alloziert den benötigten Speicher. Für jeden Eintrag in der Liste wird ein Objekt erstellt, das den Namen, die Energie und je einen Zeiger auf eine *Wechselwirkungsstruktur* und eine Liste von *Geometriestrukturen* beinhaltet. Außerdem stehen Funktionen zur Verwaltung, Abfrage und Berechnung wichtiger Größen zur Verfügung.

Nachdem die Komplexliste eingelesen ist, ruft GEOMGEN5 für jeden Komplex FLEXX auf und holt sich jeweils alle Kontakte eines Komplexes. Die Kontakte werden mit den Objekten für die bekannten Wechselwirkungsgeometrien verglichen und entsprechende Zeiger auf die Wechselwirkungsgeometrien in die Wechselwirkungsstruktur geschrieben. Außerdem werden die *Häufigkeiten* der Wechselwirkungsgeometrien nachgehalten. Die FLEXX-Aufrufe für alle Komplexe können auf beliebig viele Prozessoren verteilt werden.

IA-File

Ein IA-File beinhaltet alle Komplexe, deren experimentelle Energien, alle Kontakte und alle Wechselwirkungsgeometrien. IA-Files werden komprimiert im Binärformat geschrieben und sind somit klein und können schnell eingelesen werden. Ein IA-File ersetzt die Angabe von `geometry.dat`-File und Komplexliste. Außerdem sind keine FLEXX-Aufrufe zur Ermittlung der Kontakte mehr erforderlich. Ist kein IA-File angegeben, wird eines erzeugt, sobald alle nötigen Informationen eingelesen und berechnet worden sind.

3.7.2 Bewertungsmodi

GEOMGEN5 kennt zwei Modi für die Berechnung der Bewertungsfunktion. Im Modus `ref` wird mit der internen Bewertungsfunktion die Referenzstruktur im

Komplex bewertet. Dieser Modus ist sehr schnell aber nicht parallelisiert. Zurzeit ist auch keine Parallelisierung geplant, da in einer Sekunde auf einem 2,4 GHz Xeon-Prozessor ca. 10 Millionen Funktionsaufrufe möglich sind.

Im Modus `dock` ruft GEOMGEN5 für jeden Komplex FLEXX auf. Dabei wird vor dem Aufruf eine aktuelle `geometry.dat` für FLEXX entsprechend dem aktuellen Vektor an Parametern geschrieben. Diese Aufrufe erfolgen auf beliebig vielen Prozessoren parallel und benötigen untereinander keinerlei Kommunikation.

GEOMGEN5 überwacht die Vollendung der FLEXX-Läufe und sammelt danach für jeden Komplex die Energien und RMSDs für dessen erste n Platzierungen ein (n ist frei wählbar). Diese Daten stehen für die Zielfunktion, auf die die Parameter optimiert werden, zur Verfügung.

3.7.3 Optimierungsalgorithmen

GEOMGEN5 besitzt 3 Optimierungsalgorithmen: `GRID`, `DownHill` und `ConjGrd`.

`GRID` ist eine Abtastung der Parameter auf einem *volldimensionalen Gitter* beliebiger Schrittweite. Dabei wird immer eine Dimension auf dem Gitter abgetastet unter Beibehaltung der aktuellen Werte für alle anderen Dimensionen. Wird ein neues Minimum entlang dieses eindimensionalen Schnittes gefunden, wird der entsprechende Wert für diese Dimension festgehalten. So verfährt der Algorithmus nacheinander für jede Dimension. Danach wird wieder mit der ersten Dimension gestartet, denn durch die neuen Werte aller Dimensionen kann sich das jeweilige Minimum auf allen eindimensionalen Schnitten verändert haben. Deswegen werden alle Dimensionen solange nacheinander auf neue Minima abgetastet, bis kein neues Minimum mehr gefunden wird. In meiner bisherigen Erfahrung konnte so immer das globale Minimum gefunden werden. Dieser *brute force*-Ansatz benötigt sehr viele Aufrufe der Bewertungsfunktion und eignet sich deswegen nur in Kombination mit der internen Bewertungsfunktion.

`DownHill` basiert auf immer kleiner werdenden Schrittweiten. Für eine Dimension wird unter Beibehaltung der Werte für alle anderen Dimensionen ein Schritt der Länge s_l gemacht. Führt das zu einer Erniedrigung der Zielfunktion, geht der Algorithmus zur nächsten Dimension über. Andernfalls wird ein Schritt $-s_l$ versucht. Führt dies zu einer Minimierung wird $-s_l$ als neue Schrittweite für diese Dimension gespeichert und zur nächsten Dimension übergegangen. Sollte auch hier keine Erniedrigung der Zielfunktion möglich gewesen sein, wird $s_l = s_l/2$ und die nächste Dimension angegangen. Die Anzahl der bereits erfolgten Halbierun-

gen der Schrittweite wird für jede Dimension nachgehalten. Sind alle Dimensionen bereits n -mal halbiert worden, stoppt `DownHill`.

Sowohl `GRID` als auch `DownHill` benötigen keinerlei Ableitungen der Zielfunktion.

`ConjGrd` (Verfahren der konjugierten Gradienten) ist der einzige Algorithmus, den ich nicht selber geschrieben habe, sondern stammt aus dem bekannten Buch „Numerical Recipes“ von PRESS *et al.* [144]. Die hierfür benötigten numerischen Ableitungen werden ebenfalls mit einem Algorithmus aus diesem Buch berechnet.

Vor und nach jeder Optimierung wird eine Analyse der aktuellen Ergebnisse der Komplexe gemacht und ausgegeben. Analysiert werden die lineare Korrelation zwischen Bewertungsfunktion und experimenteller Bindungsenergie (Korrelationskoeffizient, Standardabweichung, extreme Ausreißer) und die absoluten Abweichungen zwischen Bewertungsfunktion und experimenteller Bindungsenergie. Sind Docking-Ergebnisse vorhanden, wird die kumulative Erfolgsrate in Abhängigkeit des RMSD ausgewertet und Komplexe, für die gar keine Lösung gefunden werden konnte, gezählt.

3.7.4 Geplante Erweiterungen

Geplant ist eine Erweiterung, die für jeden Komplex *alle* von `FLEXX` generierten Platzierungen in einer Liste speichert und dann nur noch alle Platzierungen aller Komplexe mit der internen Bewertungsfunktion berechnet. Dadurch kann auch eine effiziente Optimierung auf die *Platzierung* der Komplexe erfolgen, wodurch auch die Redocking-Fähigkeiten von `AIScore` verbessert werden können. Dabei wird vernachlässigt, dass im Extremfall selbst kleine Änderungen der Parameter nicht nur die Reihenfolge aller erzeugten Platzierungen eines Komplexes ändern, sondern dass auch die Platzierungen selbst beeinflusst werden. Denkbar wäre eine Hybridverfahren, bei dem all i Optimierungsschritte mit der gerade aktuellen Parametrisierung der Bewertungsfunktion neue Platzierungen generiert werden.

Quantenchemische Bestimmung von Kraftfeldparametern für CYP51

In diesem Teil stelle ich Arbeiten vor, die in Zusammenarbeit mit BERND RUPP (damals im Arbeitskreis von Prof. Dr. Hans-Dieter Höltje am Institut für Pharmazeutische Chemie an der Heinrich–Heine-Universität Düsseldorf) entstanden sind. Ziel war die Parametrisierung eines Kraftfeldes speziell für die Eisen–Stickstoff-Bindung in der 14α -Demethylase (CYP51) anhand von quantenchemischen Potenzialflächen.

4.1 Die Biochemie von CYP-51

Cytochrom P450 Sterol-14 α -Demethylase (CYP51) (Abbildung 4.1) katalysiert die oxidative Abspaltung der 14α -Methylgruppe von Sterol-Precursoren in Hefen und Säugetieren und spielt damit eine zentrale Rolle in der vitalen *Sterolsynthese*. [168] Von Azolen ist bekannt, dass sie hochpotente Inhibitoren der CYP51 in Pilzen sind, weshalb Azole wie z.B. Clotrimazol, Ketoconazol und Fluconazol (Abbildung 4.2) als *Antimykotika* eingesetzt werden. Sie inhibieren die CYP51

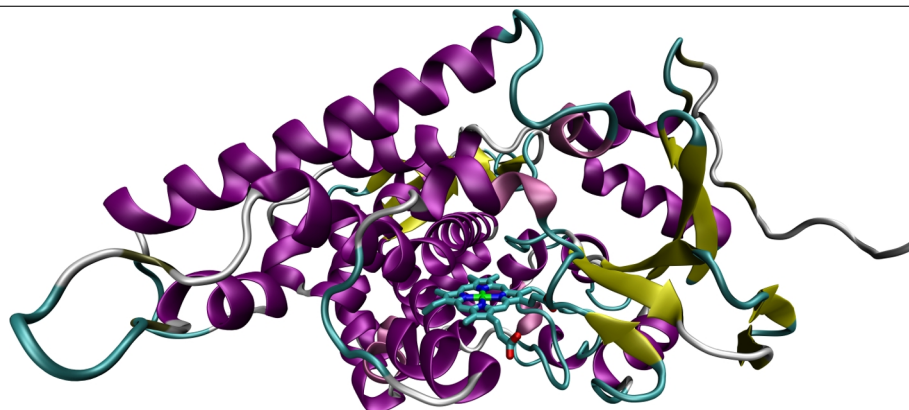


Abbildung 4.1. Strukturdarstellung des CYP51 im Komplex 1E9X mit dem hervorgehobenen HEM im aktiven Zentrum.

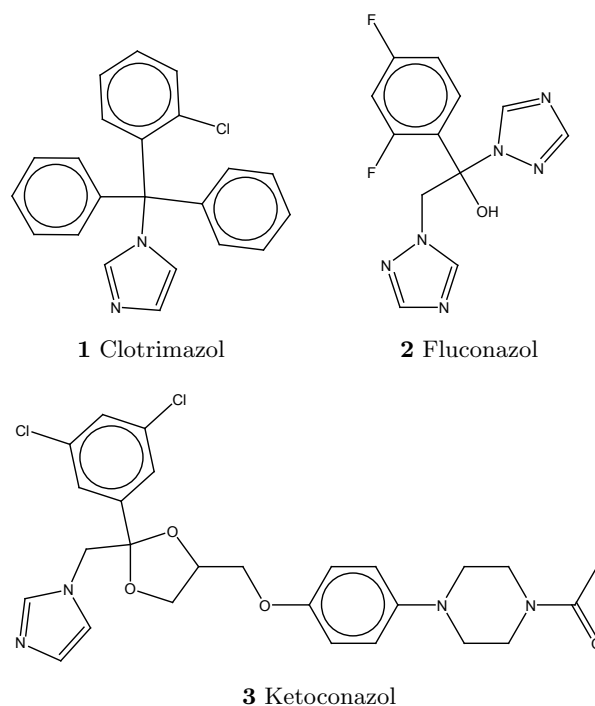


Abbildung 4.2. Strukturen einiger Antimykotika auf Azolbasis.

durch die Bindung des Azolrings an das zentrale Eisenatom der HEM-Gruppe des aktiven Zentrums. Ist das Enzym blockiert, erschöpft sich das Sterol und die *Precursor* werden nicht mehr abgebaut. Dadurch verändert sich die Struktur der Plasmamembran des Pilzes, was ihn damit verwundbar macht. [169, 170]

Azole werden in der Therapie von pathogenen Mykosen und lebensbedrohlichen systemischen Pilzinfektionen erfolgreich eingesetzt. Jedoch besitzen sie eine gewisse Toxizität, da die Azole ebensogut an die HEM-Gruppen von *körpereigenen* Cytochromen binden können.

Struktur–Aktivitäts-Studien haben gezeigt, dass der Azolring die wichtigste pharmakophore Eigenschaft der bisherigen Wirkstoffe ist, dass er aber genauso für die Toxizität eben dieser Wirkstoff verantwortlich ist. [171] Um die Selektivität der Wirkstoffe durch rationales Design zu verbessern (um so auch die Toxizität zu verringern), bedarf es eines größeren Verständnisses der Wechselwirkungen von Ligand und aktivem Zentrum.

Bernd Rupp untersuchte die Unterschiede der CYP51 aus dem Pilz *candida albicans* (Abbildung 4.3) und menschliches CYP51 durch molekulardynamische Rechnungen mit dem Programm GROMACS [172, 173]. Im GROMACS-Kraftfeld sind für koordinative Metall–Ligand-Bindungen jedoch keine Parameter

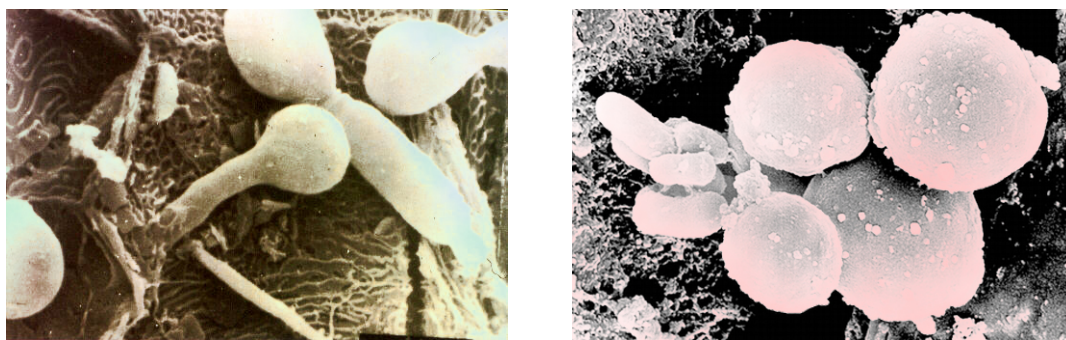


Abbildung 4.3. Vergrößerung von *candida albicans* Präparaten.

vorhanden. Diese Kraftfeldparameter galt es, durch quantenchemische *ab initio*-Rechnungen zu bestimmen.

Mit Hilfe dieser Parameter sind Bernd Rupp molekulardynamische Simulationen gelungen, durch die die Enantioselektivität des Pilzenzyms abgeschätzt werden konnte. [168]

4.2 Quantenmechanische Bestimmung von Kraftfeldparametern

4.2.1 Modellsystem und quantenchemische Methoden

Das CYP51-Enzym ist viel zu groß, um es komplett mit *ab initio*-Methoden zu behandeln. Für die Berechnung der Wechselwirkungsparameter wurde deshalb ein Modellsystem konstruiert, das nur das aktive Zentrum des Enzyms abbildet. Das aktive Zentrum besteht aus einer $\text{Fe}^{\text{IV}}\text{L}$ -Gruppe (L ist der Ligand) eingebettet in einem Protoporphyrin IX-Ring. [174] Ein unsubstituiertes Porphyrin ersetzt im Modellsystem das Protoporphyrin IX. Der Porphyrin-Eisen-Komplex wird im Folgenden mit „FeP“ bezeichnet.

Im CYP51 ist der sechste Ligand des Eisens (der *proximale Ligand*) ein Cystein des Proteins. Es ist bekannt [175], dass der proximale Ligand (in diesem Fall das Cystein) einen starken Einfluss auf die Eisen-Stickstoff-Bindung hat. Um diesen Effekt zu untersuchen, wurden zwei verschiedene Thiolate als proximaler Ligand des Eisens verwendet: Methylthiolat und die vollständige Aminosäure Cystein.

Der Azol-Ligand ist im Modellsystem unsubstituiertes Imidazol (Im). Die Position des Imidazols relativ zum aktiven Zentrum kann durch drei Parameter beschrieben werden (Abbildung 4.4): der Bindungsabstand $d_{\text{FeN}^{\text{Im}}}$ zwischen dem Eisenatom und der N_1 -Position des Imidazols, der Torsionswinkel θ (Diederwinkel $\text{N}_2^{\text{P}} - \text{Fe}_1 - \text{N}_1^{\text{Im}} - \text{C}_2^{\text{Im}}$) und der Kippwinkel φ .

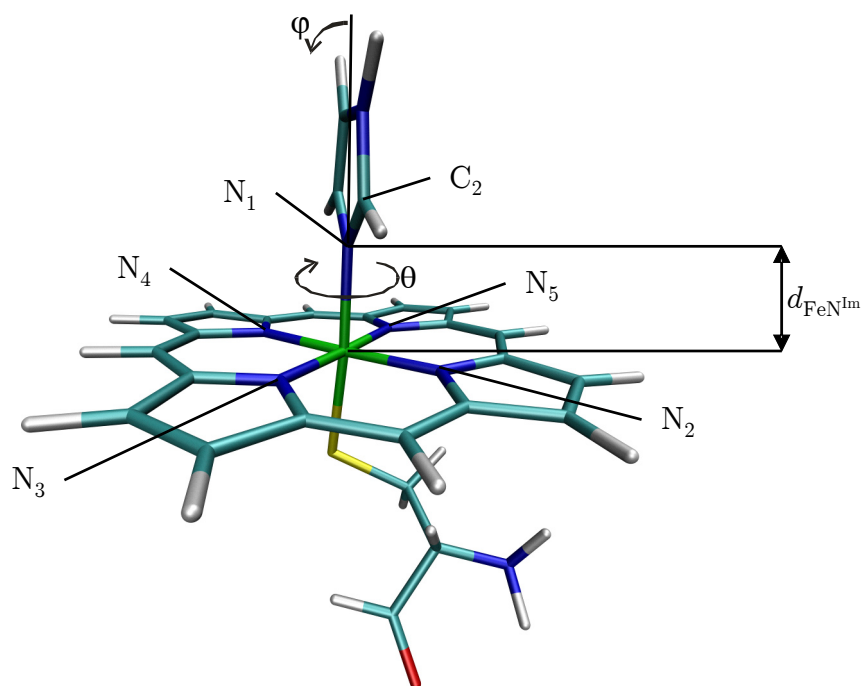


Abbildung 4.4. Modellsystem (mit Cystein als proximalen Liganden) und die geometrischen Parameter der Imidazol-Eisen-Wechselwirkung.

DFT (siehe Kapitel 2.2.5) hat sich als machtvolles Werkzeug für die Untersuchung von Übergangsmetallkomplexen erwiesen. [176]. Für diese Rechnungen wurde das BLYP-Funktional (siehe Kapitel 2.2.5.1) im Programm TURBOMOLE verwendet. An jedes Wasserstoffatom wurde eine SV-P-Basis gesetzt, an alle anderen Atome eine TZVP-Basis.

4.2.2 Rechnungen und Ergebnisse

Als erstes wurde die Geometrie des Modellsystems in allen Kernfreiheitsgraden optimiert. Von diesem Optimum aus wurden eingeschränkte Minimumenergiepfade für jeweils feste Werte der Parameter $d_{\text{FeN}^{\text{Im}}}$, θ und φ berechnet. Um Rechenzeit zu sparen, wurden alle CH- und NH-Bindungen auf ihre Werte in der Optimumsgeometrie eingefroren, alle anderen Freiheitsgrade wurden nachrelaxiert.

In Tabelle 4.1 sind die geometrischen Parameter der Optimumsgeometrie des Modellsystems aufgelistet und mit experimentellen Daten der Röntgenstrukturanalyse des PDB-Eintrags 1E9X (Cytochrom P450 14 α -Sterol-Demethylase aus *mycobacterium tuberculosis* komplexiert mit 4-Phenyl-Imidazol) verglichen. Es fällt auf, dass die geometrischen Parameter offensichtlich recht unsensitiv ge-

genüber den proximalen Liganden sind. Vergleicht man die Bindungslängen der Kristallstruktur und der quantenchemischen Rechnungen, fällt jedoch auf, dass der Abstand zwischen dem zentralen Eisenatom und dem Stickstoff des Liganden im Kristall deutlich größer ist als die Werte aus der Geometrieoptimierung. Auf der anderen Seite ist die *out-of-plane*-Verschiebung des Eisenatoms im Modellsystem größer als im Kristall.

Um diesen Umstand besser bewerten zu können, wurden aus der PDB die Kristallstrukturen von Komplexen mit einer HEM-Gruppe im aktiven Zentrum und einer koordinativen Eisen-Stickstoff-Bindung gesucht (Tabelle 4.2).

Genau wie in den theoretischen Ergebnissen zeigen die meisten Strukturen einen Fe – N-Abstand zwischen 0,21 nm und 0,22 nm. 1EA1 weist hingegen einen deutlich längeren Abstand auf. Grund könnte die sterische Repulsion der 2,4-Difluoro-Benzol-Gruppe des Fluconazols, die dem Protoporphyrin sehr nahe kommt, sein.

Der größere Verkippungswinkel in 1E9X gegenüber dem optimierten Modellsystem ist wahrscheinlich auch auf sterische Effekte zurückzuführen. Das Histidin-259 im I-Loop des Proteingerüsts verhindert, dass sich das 4-Phenyl-Imidazols senkrecht zum Porphyrinring orientieren kann.

4.2.3 Anpassung der Kraftfeldparameter

Die eingeschränkten Minimumsenergiepfade für $d_{\text{FeN}^{\text{Im}}}$ der Komplex mit Cystein und mit Methylthiolat als proximalen Liganden (Abbildung 4.5) sind nahezu perfekte Morsepotentiale.

$$V_{\text{Morse}}(d_{\text{FeN}^{\text{IM}}}) = D_{\text{FeN}^{\text{IM}}}(1 - \exp(\beta_{\text{FeN}^{\text{IM}}}(d_{\text{FeN}^{\text{IM}}} - b_{\text{FeN}^{\text{IM}}})) \quad (4.1)$$

Tabelle 4.1. Geometrische Parameter der vollständig optimierten Modellsysteme und der Kristallstruktur.

	FeP(Im)Cys	FeP(Im)MeS	1E9X
$d_{\text{FeN}^{\text{Im}}}$ [nm]	0,214	0,215	0,240
θ [°]	43,900	44,100	68,000
ϕ [°]	1,500	2,200	10,300
d_{FeN_2} [nm]	0,203	0,202	0,202
d_{FeN_3} [nm]	0,202	0,203	0,202
d_{FeN_4} [nm]	0,202	0,202	0,204
d_{FeN_5} [nm]	0,203	0,202	0,199
$d_{\text{FeN}}^{\text{OutOfPlane}}$ [nm]	0,004	0,003	0,001
d_{FeS} [nm]	0,251	0,238	0,239

Die Parameter in Gleichung (4.1) werden durch multivariante Regressionsanalyse gegen die DFT-Kurve bestimmt. Im Falle von FeP(Im)Cys liefert die Regression eine Potentialtopftiefe von $D_{\text{FeNIm}} = 54 \text{ kJ Mol}^{-1}$, einen Gleichgewichtsabstand $b_{\text{FeNIm}} = 0,213 \text{ nm}$ und ein $\beta_{\text{FeNIm}} = 22,8 \text{ nm}^{-1}$. Für FeP(Im)MeS erhält man Parameter, die weitgehend ähnlich sind, bis auf eine deutliche geringere Potentialtopftiefe ($D_{\text{FeNIm}} = 44 \text{ kJ Mol}^{-1}$, $b_{\text{FeNIm}} = 0,215 \text{ nm}$, $\beta_{\text{FeNIm}} = 22,5 \text{ nm}^{-1}$). Obwohl also der proximale Ligand einen nur geringen Einfluss auf die Geometrie des Modellsystems hat, hat er dennoch einen großen Einfluss auf die Bindungsenergie. Mit Methylthiolat als proximalen Liganden ist die Eisen–Schwefel-Bindung ein wenig kürzer, was im Gegenzug zu einer Schwächung der Eisen–Imidazol-Wechselwirkung führt.

In Abbildung 4.6a ist der eingeschränkte Minimumsenergiepfad für die Rotation um die Fe – N-Bindung θ für FeP(Im)MeS bei einer optimalen Fe – N-Bindungslänge dargestellt. Die höchste Energiebarriere für die Rotation von $\Delta E = 4,5 \text{ kJ Mol}^{-1}$ kann bei den Temperaturen in lebenden Zellen leicht überwunden werden. Das Energieprofil entlang dieser Koordinate wurde an ein periodisches Potential

$$V(\theta) = k_{\theta} \cdot \left(1 + \cos\left(n \cdot \theta - \theta^{(0)}\right)\right) \quad (4.2)$$

angepasst. Obwohl die Maxima in dem berechneten Energieprofil nicht gleich sind, ist die Abweichung des einfachen Cosinusterns nur gering. Mit $n = 4$ erhält man Kraftkonstanten von $k_{\theta} = 1,86 \text{ kJ Mol}^{-1}$ und $\theta^0 = 0^{\circ}$.

Zum Schluss wurde der Energieverlust der Eisen–Azol-Bindung durch Verkipfung des Azolrings bestimmt. Unterschiedliche Verkipfungen sind typisch für Azolringe die im aktiven Zentrum eines Cytochroms gebunden sind. An den bei-

Tabelle 4.2. Bindungslängen verschiedener Hem(Fe)–Stickstoff-Bindungen.

PDB Code	Liganden	d_{FeN} [nm]	$d_{\text{Fe-Proximal}}$ [nm]	Auflösung [Å]
1EA1	Fluconazol, Cystein	0,234	0,231	0,221
1EGY	9-Amino-Phenantren, Cystein	0,210	0,210	0,235
1F4T	4-Phenyl-Imidazol, Cystein	0,213	0,264	0,193
1JIN	Ketoconazol, Cystein	0,204	0,230	0,204
1MBI	Imidazol, Histidin	0,214	0,204	0,200
1PHA	PFZ, Cystein	0,185	0,205	0,163
1PHD	4-Phenyl-Imidazol, Cystein	0,228	0,232	0,160
1PHF	4-Phenyl-Imidazol, Cystein	0,222	0,207	0,160
1PHG	Metyrapon, Cystein	0,216	0,220	0,160

den Potentialen in Abbildung 4.6*b* kann man sehen, dass die Beiträge von Torsion und Verkippung zur Bindungsenergie annähernd additiv sind. Die obere Kurve wurde bei einem Torsionswinkel von $\theta = 0^\circ$ berechnet und verläuft so gut wie parallel zu der Kurve, die in der ekliptischen Orientierung ($\theta = 45^\circ$, Minimum) erhalten wurde, was die Unabhängigkeit dieser zwei Bewegungen demonstriert.

Mein Teil des Projekts war damit beendet. Doch gerade mit den Ergebnissen über das Potential der Verkippung konnten die biologischen Daten der verschiedenen Inhibitoren erklärt werden. Diese Verkippung ist es, die den größten Einfluss auf die Bindungsstärke des Inhibitors hat. Indem man diese Verkippung als einen Strukturparameter der Wechselwirkung nutzt, sollte es möglich sein, gezielt neue und selektivere Inhibitoren zu entwickeln.

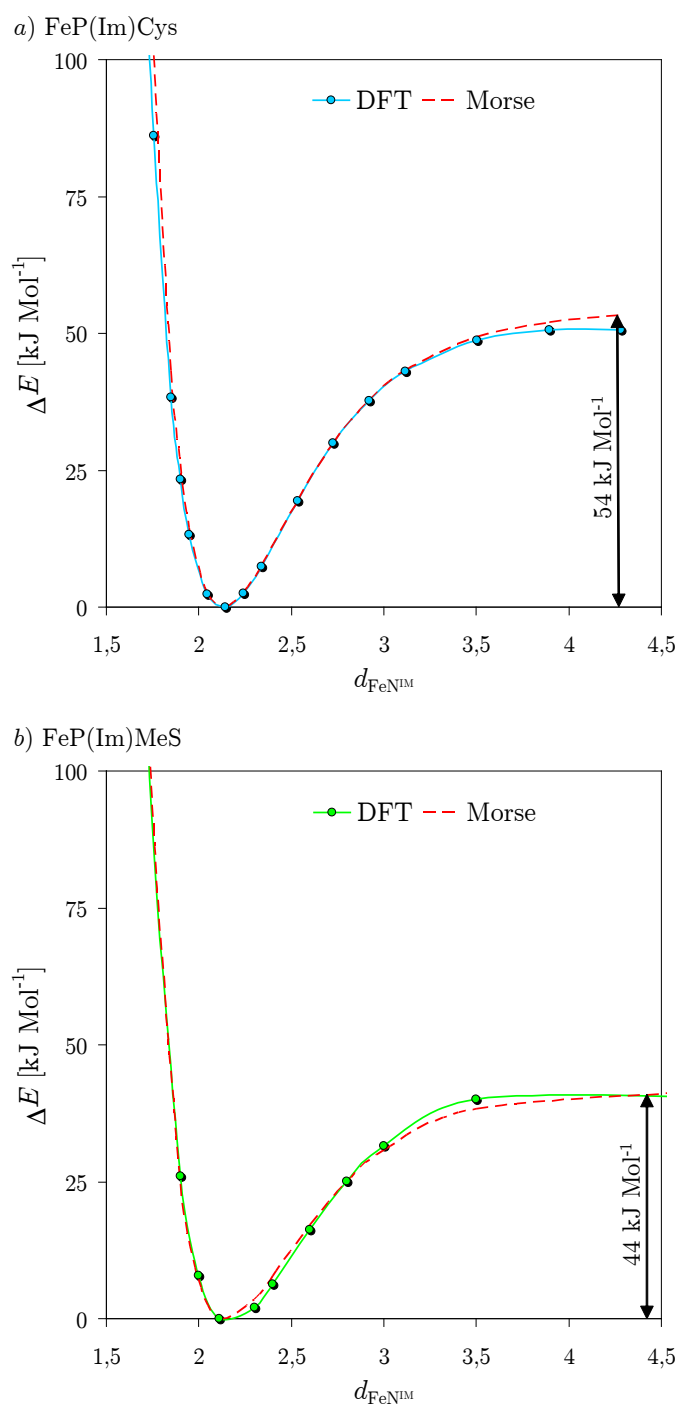


Abbildung 4.5. Eingeschränkter Minimumenergiepfad für Variation des Eisen–Stickstoff-Abstands d_{FeNIm} für a) Cystein als proximalen Liganden und b) für MeS. Die gestrichelten Kurven sind die angepassten Morse-Potenziale.

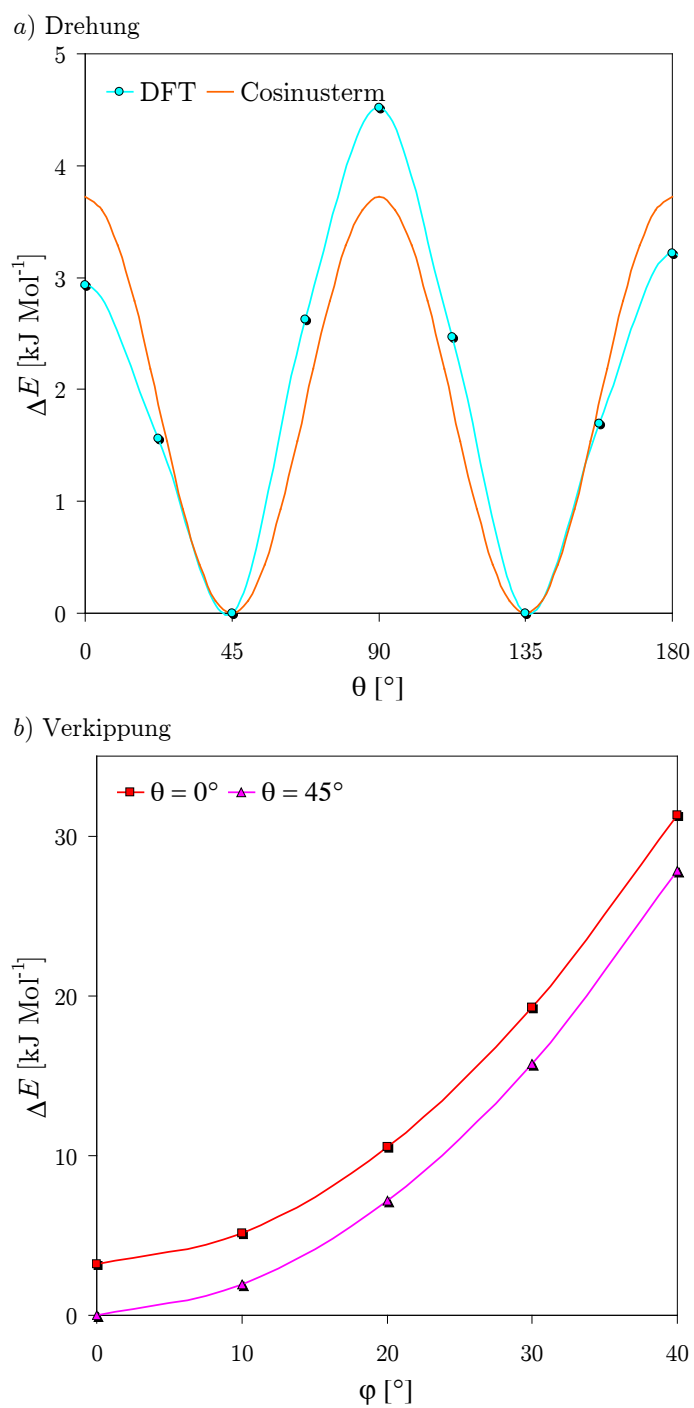


Abbildung 4.6. a) Eingeschränkter Minimumenergiepfad für die Torsion des Imidazols im Komplex FeP(Im)MeS. Die rote Kurve ist das angepasste Cosinuspotential. b) Eingeschränkte Minimumenergiepfade für die Verkipfung des Imidazols im Komplex FeP(Im)MeS bei jeweils $\theta = 0^\circ$ und $\theta = 45^\circ$.

Zusammenfassung und Ausblick

Ich habe in dieser Arbeit die Entwicklung der neuen Bewertungsfunktion AISCORE für das Docking-Programm FLEXX vorgestellt. AISCORE erweitert die bekannte LUDI-Bewertungsfunktion um eine differenzierte Behandlung von Wasserstoffbrücken nach ihrem chemischen Kontext. Außerdem wurde ein Algorithmus zur Behandlung von mehrfach gegabelten Wasserstoffbrücken implementiert.

Für die Parametrisierung von AISCORE wurden zunächst quantenchemische *ab initio*-Rechnungen an 81 wasserstoffverbrückten Komplexen aus kleinen organischen Molekülen durchgeführt. Mit Hilfe von Dichte-Funktional-Methoden und der MØLLER-PLESSET-Störungstheorie zweiter Ordnung wurden die Wechselwirkungsenergien dieser 81 Komplexe berechnet und anschließend in Donor- und Akzeptorenergien für das additive Inkrementensystem HBINC aufgeteilt. Es konnte eine große Bandbreite an möglichen Wasserstoffbrückenenergien gezeigt werden, ebenso die Wichtigkeit von Schwefel als Wasserstoffbrückendonator und -akzeptor. Eine exzellente lineare Korrelation von $r^2 = 0,994$ zwischen HBINC-Energien und *MP2*-Energien konnte erreicht werden.

Die Inkremente von HBINC wurden dann als Grundlage für die Parametrisierung von AISCORE verwendet. Ein Solvatationsfaktor skaliert alle Inkremente gleichermaßen um die Erniedrigung der Wechselwirkungsenergie in einem Lösungsmittel gegenüber der Energien im Vakuum zu modellieren. Für resonanz-assistierte und geladene Wasserstoffbrücken, sowie für lipophile, aromatische und metallische Wechselwirkungen wurden eigene Parameter eingeführt. Diese Parameter sind zusammen mit dem Solvatationsfaktor an einen Datensatz von 100 Protein-Ligand-Komplexen mit bekannter Struktur und bekannter experimenteller freien Bindungsenergie angepasst worden.

Der Vergleich von AISCORE mit FLEXXSCORE zeigt, dass AISCORE die freien Bindungsenergien der Protein-Ligand-Komplexe signifikant besser abschätzen kann. Dabei spielen sowohl die chemisch diverse Behandlung der Wasserstoff-

brückenenergien als auch die Behandlung der gegabelten Wasserstoffbrücken eine Rolle. Selbst eine nachoptimierte Version von FLEXXSCORE, bei der zusätzlich die neue Funktionalität für die Behandlung multifurkater H-Brücken zum Einsatz kam, konnte die lineare Korrelation von AISCORE nicht erreichen.

Durch eine zehnfache *leave-ten-out* Kreuzvalidierung wurde die Möglichkeit einer Übertrainierung von AISCORE auf die 100 Komplexe des Datensatzes diskutiert.

Trotz der signifikanten Verbesserung in der Vorhersage der freien Bindungsenergie, ist AISCORE im Redocking mit einer kumulativen Erfolgsrate von 44% (bei 2,5 Å) schlechter als FLEXXSCORE (55%). Als mögliche Ursache wurde die Abschätzung der Gesamtkomplexierungsenergie während des inkrementellen Aufbaus diskutiert.

Auf Grundlage des im Zuge dieser Arbeit entwickelten Programms GEOMGEN wäre eine Parametrisierung, die Redocking-Ergebnisse mit berücksichtigt, möglich. Dafür ist eine Erweiterung von GEOMGEN nötig sowie ein größerer Satz von Protein-Ligand-Komplexen. Damit ein solcher Ansatz erfolgreich ist, bedarf es vorher einer Anpassung des Algorithmus, der für die Abschätzung der zu erwartenden Beiträge von noch nicht platzierten Fragmenten zur freien Bindungsenergie verwendet wird. Alternativ kann FLEXX mit wenig Aufwand so umprogrammiert werden, dass jeder inkrementelle Aufbau ausgehend von den Basisfragmenten separat erfolgt, so dass eine solche Abschätzung nicht benötigt wird.

Um die Bewertungsfunktion auch auf die Erkennung möglicher Leitstrukturen zu trainieren, sind *Enrichment Tests* unerlässlich. Dabei werden einige zig-Tausend Moleküle in ein Protein gedockt. Von den Molekülen ist jeweils bekannt, ob sie eine Aktivität gegenüber dem Protein haben oder nicht. Solche Datensätze für *Enrichment Tests* sind verfügbar und erfordern nur wenige Erweiterungen im Programm GEOMGEN.

AISCORE konnte bereits in FLEXR implementiert werden. Mit FLEXR können flexible Liganden in flexible artifizielle Rezeptoren gedockt werden. Im Falle des Koffeinrezeptors von WALDVOGEL *et al.* konnte mit AISCORE von alkylierten Oxypurinen im Komplex mit diesem Rezeptor die richtige Struktur generiert werden und die freien Bindungsenergien der Komplexe untereinander in der richtigen Reihenfolge berechnet werden. AISCORE ist jedoch für Protein-Ligand-Wechselwirkungen in Wasser parametrisiert. Die berechneten Energien der Oxypurin-Komplexe wurden durchweg überschätzt, da die experimentelle

Energiebestimmung in CDCl_3 durchgeführt worden ist. Es ist deswegen eine Rekalibrierung von AISCORE für artifizielle Rezeptoren notwendig. Dafür braucht es einen größeren Satz an Komplexen mit bekannter experimenteller Bindungsenergie und bekannter Komplexgeometrie.

In einem letzten Kapitel habe ich die Parametrisierung von Kraftfeldparametern für eine ganz spezielle Bindung mit Hilfe von quantenchemischen *ab initio*-Rechnungen beschrieben. An einem Modellsystem wurden die Potentiale für eine koordinative Eisen–Stickstoff-Bindung wie sie in Cytochromen vorkommt, berechnet und Parameter für Kraftfelder bestimmt. Mit Hilfe dieser Parameter konnten molekulardynamische Rechnungen durchgeführt werden, die die Enantioselektivität von Cytochromen in pathogenen Pilzen abschätzen konnten.

Aus dieser Arbeit hervorgegangene Veröffentlichung

Bernd Rupp, Stephan Raub, Christel M. Marian, Hans-Dieter Höltje

„Molecular design of two 14 α -demethylase homology models and their interactions with the azole antifungals ketoconazole and bifonazole“

Journal of Computer-Aided Molecular Design (2005), 19: 149 – 163

Stephan Raub, Christel M. Marian

„Quantum Chemical Investigation of Hydrogen-Bond Strengths and Partitioning into Donor and Acceptor Contributions“

Journal of Computational Chemistry (2007), DOI: 10.1002/jcc.20673

Stephan Raub, Andreas Steffen, Andreas Kämper, Christel M. Marian

„Chemically Diverse Scoring in FlexX“

Journal of Medicinal Chemistry

Eingereicht.

Literatur

1. BUCKINGHAM, A. D.: Molecular Quadrupole Moments. In: *Quarterly Reviews of the Chemical Society* 13 (1959), S. 183–214
2. MO, Y.: Probing the nature of hydrogen bonds in DNA base pairs. In: *J. Mol. Model* 12 (2006), S. 665–672
3. PESCO, G. A.: In: *Nature* 384 (1996), S. 7–9
4. DREWS, J.: Genomic sciences and the medicine of tomorrow. In: *Nat. Biotechnol.* 14 (1996), Nr. 11, S. 1516–1518
5. DREWS, J.: Quo vadis, biotech? (Part 1). In: *Drug Discovery Today* 5 (2000), Nr. 12, S. 2–4
6. GOHLKE, H. ; KLEBE, G.: Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptors. In: *Angew. Chem. Int. Ed.* 41 (2002), Nr. 15, S. 2645–2676
7. CHAN, O. H. ; STEWART, B. H.: Physicochemical and drug-delivery considerations for oral drug bioavailability. In: *Drug Discovery Today* 1 (1996), Nr. 11, S. 461–473
8. SMITH, D. A. ; WATERBEEMD, H. van d.: Pharmacokinetics and metabolism in early drug discovery. In: *Curr. Opin. Chem. Biol.* 3 (1999), Nr. 4, S. 373–378
9. WATT, A. P. ; MORRISON, D. ; EVANS, D. C.: Approaches to higher-throughput pharmacokinetics (HTPK) in drug discovery. In: *Drug Discovery Today* 5 (2000), Nr. 1, S. 17–24
10. GORDON, E. M. ; GALLOP, M. A. ; PATEL, D. V.: Strategy and tactics in combinatorial organic synthesis. Applications to drug discovery. In: *Acc. Chem. Res.* 29 (1996), Nr. 3, S. 144–154
11. CARELL, T. ; WINTNER, E. A. ; REBEK, J.: A solution-phase screening-procedure for the isolation of active compounds from a library of molecules. In: *Angew. Chem. Int. Ed. Engl.* 33 (1994), Nr. 20, S. 2061–2064
12. HOUSTON, J. G. ; BANKS, M.: The chemical-biological interface: Developments in automated and miniaturised screening technology. In: *Curr. Opin. Biotechnol.* 8 (1997), Nr. 6, S. 734–740
13. BAJORATH, F.: Integration of virtual and high-throughput screening. In: *Nature Rev. Drug Design* 1 (2002), Nr. 11, S. 882–894
14. KLEBE, G.: Virtual ligand screening: Strategies, perspectives and limitations. In: *Drug Disc. Today* 11 (2006), Nr. 13–14, S. 580–594
15. LENGAUER, T. ; RAREY, M.: Computational methods for biomolecular docking. In: *Curr. Opin. Struct. Biol.* 6 (1996), Nr. 3, S. 402–406
16. RAREY, M.: Protein–Ligand Docking in Drug Design. In: LENGAUER, T. (Hrsg.): *Bioinformatics – From Genomes to Drugs*. 1. Weinheim : Wiley-VCH, S. 315–360
17. FERRARA, P. ; GOHLKE, H. ; PRICE, D. J. ; KLEBE, G. ; BROOKS III, C. L.: Assessing scoring functions for protein-ligand interactions. In: *J. Med. Chem.* 47 (2004), S. 3032–3047
18. RASTELLI, G. ; PACCHIONI, S. ; SIRAWARAPORN, W. ; SIRAWARAPORN, R. ; PARENTI, M. D. ; FERRARI, A. M.: Docking and database screening reveal new classes of plasmodium falciparum dihydrofolate reductase inhibitors. In: *J. Med. Chem.* 46 (2003), Nr. 14, S. 2834–2845

19. DOMAN, T. N. ; MCGOVERN, S. L. ; WITHERBEE, B. J. ; KASTEN, T. P. ; KURUMBAIL, R. ; STALLINGS, W. C. ; CONNOLLY, D. T. ; SHOICHET, B. K.: Molecular docking and high-throughput screening for novel inhibitors of protein tyrosine phosphatase-1B. In: *J. Med. Chem.* 45 (2002), Nr. 11, S. 2213–2221
20. MOZZICONACCI, J. C. ; ARNOULT, E. ; BERNARD, P. ; DO, Q. T. ; MAROT, C. ; MORIN-ALLORY, L.: Optimization and validation of a docking-scoring protocol; Application to virtual screening for COX-2 inhibitors. In: *J. Med. Chem.* 48 (2005), Nr. 4, S. 1055–1068
21. MASSOVA, I. ; MARTIN, P. ; BULYCHEV, A. ; DOYLE, R. Kocz M. ; EDWARDS, B. F. P. ; MOBASHERY, S.: Templates for design of inhibitors for serine proteases: Application of the program DOCK to the discovery of novel inhibitors for thrombin. In: *S. Bioorg. Med. Chem. Lett.* 8 (1998), Nr. 18, S. 2463–2466
22. BURKHARD, P. ; TAYLOR, P. ; WALKINSHAW, M. D.: An example of a protein ligand found by database mining: Description of the docking method and its verification by a 2.3 Ångstrom X-Ray structure of a thrombin-ligand complex. In: *J. Mol. Biol.* 277 (1998), Nr. 2, S. 449–466
23. SHOICHET, B. K. ; STROUD, R. M. ; SANTI, D. V. ; KUNTZ, I. D. ; PERRY, K. M.: Structure-based discovery of inhibitors of thymidylate synthase. In: *Science* 259 (1993), Nr. 5100, S. 1445–1450
24. TOYODA, T. ; BROBEY, R. K. B. ; SANO, G. ; HORII, T. ; TOMIOKA, N. ; ITAI, A.: Lead discovery of inhibitors of the dihydrofolate reductase domain of plasmodium falciparum dihydrofolate reductase-thymidylate synthase. In: *Biochem. Biophys. Res. Commun.* 235 (1997), Nr. 3, S. 515–519
25. GSCHWEND, D. A. ; SIRAWARAPORN, W. ; SANTI, D. V. ; KUNTZ, I. D.: Specificity in structure-based drug design: Identification of a novel, selective inhibitor of pneumocystis carinii dihydrofolate reductase. In: *PROTEINS: Structure, Function and Genetics* 29 (1997), Nr. 1, S. 59–67
26. DESJARLAIS, R. L. ; DIXON, J. S.: A shape-based and chemistry-based docking method and its use in the design of HIV-1 protease inhibitors. In: *J. Comput.-Aided Mol. Des.* 8 (1994), Nr. 3, S. 231–242
27. FRIEDMAN, S. H. ; GANAPATHI, P. S. ; RUBIN, Y. ; KENYON, G. L.: Optimizing the binding of fullerene inhibitors of the HIV-1 protease through predicted increases in hydrophobic desolvation. In: *J. Med. Chem.* 41 (1998), Nr. 13, S. 2424–2429
28. RAREY, M. ; LENGAUER, T. ; KLEBE, G.: A fast flexible docking method using an incremental construction algorithm. In: *J. Mol. Biol.* 261 (1996), S. 470–489
29. RAREY, M. ; KRAMER, B. ; LENGAUER, T.: Multiple automatic base selection: Protein-ligand docking based on incremental construction without manual intervention. In: *J. Comput.-Aided Mol. Des.* 11 (1997), S. 369–384
30. KRAMER, B. ; RAREY, M. ; LENGAUER, T.: Evaluation of the FlexX incremental construction algorithm for protein-ligand docking. In: *PROTEINS: Structure, Function and Genetics* 37 (1999), S. 228–241
31. WANG, R. X. ; LU, Y. P. ; WANG, S. M.: Comparative evaluation of 11 scoring functions for molecular docking. In: *J. Med. Chem.* 46 (2002), Nr. 12, S. 2287–2303
32. SATO, H. ; SHEWCHUCK, L. M. ; TANG, J.: Prediction of multiple binding modes of the CDK2 inhibitors, anilinopyrazoles, using the automated docking programs GOLD, FlexX, and LigandFit: An evaluation of performance. In: *J. Chem. Inf. Model.* 46 (2006), Nr. 6, S. 2552–2562
33. HUANG, S.-Y. ; ZOU, X. Q.: An iterative knowledge-based scoring function to predict protein-ligand interactions: II. Validation of the scoring function. In: *J. Comput. Chem.* 27 (2006), Nr. 15, S. 1876–1882
34. BÖHM, H.-J.: The computer-program LUDI - A new method for the denovo design of enzyme-inhibitors. In: *J. Comput.-Aided Mol. Des.* 6 (1992), Nr. 1, S. 61–78
35. BÖHM, H.-J.: LUDI - Rule-based automatic design of new substituents for enzyme-inhibitor leads. In: *J. Comput.-Aided Mol. Des.* 6 (1992), Nr. 6, S. 593–606

36. BÖHM, H.-J.: The development of a simple empirical scoring function to estimate the binding constant for a protein ligand complex of known 3-dimensional structure. In: *J. Comput.-Aided Mol. Des.* 8 (1994), Nr. 3, S. 243–256
37. BÖHM, H.-J.: Prediction of binding constants of protein ligands: A fast method for the prioritization of hits obtained from de novo design or 3D database search programs. In: *J. Comput. Aided Mol. Des.* 12 (1998), Nr. 4, S. 309–323
38. LATIMER, W. M. ; RODEBUSH, W. H.: Polarity and ionization from the standpoint of the lewis theory of valence. In: *J. Am. Chem. Soc.* 42 (1920), S. 1410
39. PENDÁS, A. M. ; BLANCO, M. A. ; FRANCISCO, E.: The nature of hydrogen bond: A synthesis from the interacting quantum atoms picture. In: *J. Chem. Phys.* 125 (2006), S. 184112
40. KING, B. F. ; WEINHOLD, F.: Structure and spectroscopy of (HCN)(N) clusters - Cooperative and electronic delocalization effects in C-H...N hydrogen-bonding. In: *J. Chem. Phys.* 103 (1995), Nr. 1, S. 333–347
41. THOMPSON, W. H. ; HYNES, J. T.: Frequency shifts in the hydrogen-bonded OH stretch in halide-water clusters. The importance of charge transfer. In: *J. Am. Chem. Soc.* 122 (2000), Nr. 26, S. 6278–6286
42. WENNMOHS, F. ; STAEMMLER, V. ; SCHINDLER, M.: Theoretical investigation of weak hydrogen bonds to sulfur. In: *J. Chem. Phys.* 119 (2003), Nr. 6, S. 3208–3218
43. BUCKINGHAM, A. D.: The Hydrogen Bond. In: XANTHEAS, S. S. (Hrsg.): *Recent Theoretical and Experimental Advances in Hydrogen Bonded Clusters*. Kluwer Academic Publishers, Boston, 2000, S. 1–24
44. JUREČKA, P. ; ŠPONER, J. ; ČERNÝ, J. ; HOBZA, P.: Benchmark database of accurate (MP2 and CCSD(T) complete basis set limit) interaction energies of small model complexes, DNA base pairs, and amino acid pairs. In: *Phys. Chem. Chem. Phys.* 8 (2006), S. 1985–1993
45. WATSON, J. D. ; CRICK, F. H. C.: Molecular structure of nucleic acids - A structure for deoxyribose nucleic acid. In: *Nature* 171 (1953), Nr. 4356, S. 737–738
46. STRYER, L. ; BERG, J. M. ; TYMOCZKO, J. L.: *Biochemistry*. Palgrave Macmillan, 2006
47. KORNBERG, A. ; BAKER, T. A.: *DNA Replication*. University Science Books, U.S., 2005
48. GUCKIAN, K. M. ; KRUGH, T. R. ; KOOL, E. T.: Solution structure of a nonpolar, non-hydrogen-bonded base pair surrogate in DNA. In: *J. Am. Chem. Soc.* 122 (2000), Nr. 29, S. 6841–6847
49. KOOL, J. C. Morales E. T.: Varied molecular interactions at the active sites of several DNA polymerases: Nonpolar nucleoside isosteres as probes.
50. ATKINS, P. W.: *Physikalische Chemie*. Wiley-VCH, 2001
51. NEUHEUSER, T. ; HESS, B. A. ; REUTEL, C. ; WEBER, E.: Ab initio calculations of supramolecular recognition modes. Cyclic versus noncyclic hydrogen bonding in the formic acid/formamide system. In: *J. Phys. Chem.* 98 (1994), S. 6459–6467
52. JEFFREY, G. A.: *An Introduction to Hydrogen Bonding*. Oxford University Press, New York, 1997
53. SCHEINER, S.: *Hydrogen Bonding: A Theoretical Perspective*. Oxford University Press, New York, 1997
54. DEL BENE, J. E.: Hydrogen Bonding. In: R. SCHLEYER, P. v. (Hrsg.) ; ALLINGER, N. L. (Hrsg.) ; CLARK, T. (Hrsg.) ; GASTEIGER, J. (Hrsg.) ; KOLLMAN, P. A. (Hrsg.) ; III., H. F. S. (Hrsg.) ; SCHREINER, P. R. (Hrsg.): *Encyclopedia of Computational Chemistry*. John Wiley & Sons, Chichester, UK, 1998, S. 1263–1271
55. DESIRAJU, G. R. ; STEINER, T.: *The Weak Hydrogen Bond in Structural Chemistry and Biology*. Oxford University Press: New York, 1999
56. JEFFREY, G. A. ; SAENGER, W.: *Hydrogen Bonding in Biological Structures*. Springer-Verlag: Berlin, Heidelberg, 1991
57. WAHL, M. C. ; SUNDARALINGAM, M.: C-H...O hydrogen bonding in biology. In: *Trends in Biochem. Sci.* 22 (1997), Nr. 3, S. 97–102

58. STONE, A. J.: Universal Models of Hydrogen Bonding. In: XANTHEAS, S. S. (Hrsg.): *Recent Theoretical and Experimental Advances in Hydrogen Bonded Clusters*. Kluwer Academic Publishers, Boston, 2000, S. 25–34
59. HODGES, M. P. ; STONE, A. J. ; XANTHEAS, S. S.: Contribution of many-body terms to the energy for small water clusters: A comparison of ab initio calculations and accurate model potentials. In: *J. Phys. Chem. A* 101 (1997), Nr. 48, S. 9163–9168
60. RAPPE, A. K. ; BERNSTEIN, E. R.: Ab initio calculation of nonbonded interactions: Are we there yet? In: *J. Phys. Chem. A* 104 (2000), Nr. 26, S. 6117–6128
61. ŠPONER, J. ; JUREČKA, P. ; HOBZA, P.: Accurate interaction energies of hydrogen-bonded nucleic acid base pairs. In: *J. Am. Chem. Soc.* 126 (2004), Nr. 32, S. 10142–10151
62. LEACH, A. R.: *Molecular Modelling. Principles and Applications*. Prentice Hall, 2001
63. WHEATLY, R. J. ; PRICE, S. L.: An overlap model for estimating the anisotropy of repulsion. In: *Molec. Phys.* 69 (1990), Nr. 3, S. 507–533
64. GUERRA, C. F. ; BICKELHAUPT, F. M.: Charge transfer and environment effects responsible for characteristics of DNA base pairing. In: *Angew. Chem. Int. Ed. Engl.* 38 (1999), Nr. 19, S. 2942–2945
65. GUERRA, C. F. ; BICKELHAUPT, F. M. ; SNIJDERS, J. G. ; BAERENDS, E. H.: The nature of the hydrogen bond in DNA base pairs: The role of charge transfer and resonance assistance. In: *Chem. Eur. J.* 5 (1999), Nr. 12, S. 3581–3594
66. GUERRA, C. F. ; BICKELHAUPT, F. M. ; SNIJDERS, J. G. ; BAERENDS, E. H.: Hydrogen bonding in DNA base pairs: Reconciliation of theory and experiment. In: *J. Am. Chem. Soc.* 122 (2000), Nr. 17, S. 4117–4128
67. BECK, J. F. ; MO, Y.: How resonance assists hydrogen bonding interactions: An energy decomposition analysis. In: *J. Comput. Chem.* 28 (2007), S. 455–466
68. GILLI, G. ; BELLUCCI, F. ; FERRETTI, V. ; BERTOLASI, V.: Evidence for resonance-assisted hydrogen-bonding from crystal-structure correlations on the enol form of the β -diketone fragment. In: *J. Am. Chem. Soc.* 111 (1989), Nr. 3, S. 1023–1028
69. GILLI, P. ; BERTOLASI, V. ; FERRETTI, V. ; GILLI, G.: Covalent nature of the strong homonuclear hydrogen-bond - Study of the O-H \cdots O system by crystal-structure correlation methods. In: *J. Am. Chem. Soc.* 116 (1994), Nr. 3, S. 909–915
70. GILLI, P. ; BERTOLASI, V. ; FERRETTI, V. ; GILLI, G.: Evidence for intramolecular N-H \cdots O resonance-assisted hydrogen bonding in β -enaminones and related heterodienes. A combined crystal-structural, IR and NMR spectroscopic, and quantum-mechanical investigation. In: *J. Am. Chem. Soc.* 122 (2000), Nr. 42, S. 10405–10417
71. BERTOLASI, V. ; GILLI, P. ; FERRETTI, V. ; GILLI, G.: Evidence for resonance-assisted hydrogen-bonding - 2. Intercorrelation between crystal-structure and spectroscopic parameters in 8 intramolecularly hydrogen-bonded 1,3-diaryl-1,3-propanedione enols. In: *J. Am. Chem. Soc.* 113 (1991), Nr. 13, S. 4917–4925
72. GILLI, P. ; BERTOLASI, V. ; PRETTO, L. ; FERRETTI, V. ; GILLI, G.: Covalent versus electrostatic nature of the strong hydrogen bond: Discrimination among single, double, and asymmetric single-well hydrogen bonds by variable-temperature X-Ray crystallographic methods in β -diketone enol RAHB systems. In: *J. Am. Chem. Soc.* 126 (2004), Nr. 12, S. 3845–3855
73. PARTHASARATHI, R. ; SUBRAMANIAN, V. ; SATHYAMURTHY, N.: Hydrogen bonding without borders: An atoms-in-molecules perspective. In: *J. Phys. Chem. A* 110 (2006), Nr. 10, S. 3349–3351
74. SZABO, A. ; OSTLUND, N. S.: *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*. Mineola, New York : Dover Publication, 1996
75. SCHRÖDINGER, E.: Quantisierung als Eigenwertproblem (Erste Mitteilung). In: *Ann. Phys.* 79 (1926), Nr. 4, S. 361–376

76. SCHRÖDINGER, E.: Quantisierung als Eigenwertproblem (Zweite Mitteilung). In: *Ann. Phys.* 79 (1926), Nr. 4, S. 489–527
77. SCHRÖDINGER, E.: Quantisierung als Eigenwertproblem (Dritte Mitteilung: Störungstheorie, mit Anwendung auf den Starkeffekt der Balmerlinien). In: *Ann. Phys.* 80 (1926), Nr. 4, S. 437–490
78. SCHRÖDINGER, E.: Quantisierung als Eigenwertproblem (Vierte Mitteilung). In: *Ann. Phys.* 81 (1926), Nr. 4, S. 109–139
79. PAULI, W.: The relation between the completion of the electron groups of the atom and the complex structure of the spectra. In: *Z. Phys.* 31 (1925), S. 735–786
80. SLATER, J. C.: The theory of complex spectra. In: *Phys. Rev.* 34 (1929), S. 1293–1323
81. SLATER, J. C.: Molecular energy levels and valence bonds. In: *Phys. Rev.* 38 (1931), S. 1109–1144
82. CONDON, E. U.: The theory of complex spectra. In: *Phys. Rev.* 36 (1930), S. 1121–1133
83. Roothaan, C. C. J.: New developments in molecular orbital theory. In: *Rev. Mod. Phys.* 23 (1951), Nr. 2, S. 69–89
84. HALL, G. G.: The molecular orbital theory of chemical valency: 8. A method of calculating ionization potentials. In: *Proc. Roy. Soc. A* 205 (1951), Nr. 1083, S. 541–552
85. MøLLER, C. ; PLESSET, M. S.: Note on an approximation treatment for many-electron systems. In: *Phys. Rev.* 46 (1934), Nr. 7, S. 618–622
86. DUNNING, JR., T. H. ; PETERSON, K. A.: Use of Møller–Plesset perturbation theory in molecular calculations: Spectroscopic constants of first row diatomic molecules. In: *J. Chem. Phys.* 108 (1998), Nr. 12, S. 4761–4771
87. HOHENBERG, P. ; KOHN, W.: Inhomogeneous electron gas. In: *Phys. Rev. B* 136 (1964), Nr. 3, S. 864
88. KOHN, W. ; SHAM, L.: Self-consistent equations including exchange and correlation effects. In: *Phys. Rev. A* 140 (1965), Nr. 4, S. 1133
89. VOSKO, S. H. ; WILK, L. ; NUSAIR, M.: Accurate spin-dependent electron liquid correlation energies for local spin-density calculations - A critical analysis. In: *C. J. Phys* 58 (1980), Nr. 8, S. 1200–1211
90. BECKE, A. D.: Density-functional exchange-energy approximation with correct asymptotic-behavior. In: *Phys. Rev. A* 38 (1988), Nr. 6, S. 3098–3100
91. LEE, C. ; YANG, W. ; PARR, R. G.: Development of the Colle-Salvetti correlation-energy formula into a functional of the electron-density. In: *Phys. Rev. B* 37 (1988), Nr. 2, S. 785–789
92. BECKE, A. D.: Density-functional thermochemistry – 3. The role of exact exchange. In: *J. Chem. Phys.* 98 (1993), Nr. 7, S. 5648–5652
93. BADER, R. F. W. ; NGUYEN-DANG, T. T. ; TAL, Y.: A topological theory of molecular-structure. In: *Rep. Prog. Phys.* 44 (1981), S. 893–948
94. BADER, R. F. W.: *Atoms in Molecules*. Oxford : Oxford University Press, 1990
95. BADER, R. F. W.: A Quantum Theory of Molecular Structure and Its Applications. In: *Chem. Rev.* 91 (1991), S. 893–928
96. BADER, R. F. W.: 1997 Polanyi Award Lecture - Why are there atoms in chemistry? In: *Can. J. Chem.* 76 (1998), S. 973–988
97. POPELIER, P. L. A. ; AICKEN, F. M. ; O'BRIEN, S. E.: In: *Chemical Modelling: Applications and Theory, Volume 1*. The Royal Society of Chemistry, 2000, Kapitel 3 – Atoms in Molecules
98. RAUB, S. ; MARIAN, C.: Quantum chemical investigation of hydrogen-bond strengths and partitioning into donor and acceptor contributions. In: *J. Comput. Chem.* (2007), S. DOI: 10.1002/jcc.20673
99. CHIRLIAN, L. E. ; FRANCL, M. M.: Atomic charges derived from electrostatic potentials: A detailed study. In: *J. Comput. Chem.* 8 (1987), Nr. 6, S. 894–905
100. COX, S. R. ; WILLIAMS, D. E.: Representation of the molecular electrostatic potential by a net atomic charge model. In: *J. Comput. Chem.* 2 (1981), Nr. 3, S. 304–323

101. JUREČKA, P. ; HOBZA, P.: True stabilization energies for the optimal planar hydrogen-bonded and stacked structures of guanine· · · cytosine, adenine· · · thymine, and their 9- and 1-methyl derivatives: Complete basis set calculations at the MP2 and CCSD(T) levels and comparison with experiment. In: *J. Am. Chem. Soc.* 125 (2003), S. 15608–15613
102. RABLEN, P. R. ; LOCKMAN, J. W. ; JORGENSEN, W. L.: Ab initio study of hydrogen-bonded complexes of small organic molecules with water. In: *J. Phys. Chem. A* 102 (1998), Nr. 21, S. 3782–3797
103. DEL BENE, J. E. ; PERSON, W. B. ; SZCZEPANIAK, K.: Properties of hydrogen-bonded complexes obtained from the B3LYP functional with 6-31G(d,p) and 6-31+G(d,p) basis sets - comparison with MP2/6-31+G(d,p) results and experimental data. In: *J. Phys. Chem.* 99 (1995), Nr. 27, S. 10705–10707
104. ELSTNER, M. ; HOBZA, P. ; FRAUENHEIM, T. ; SUHAI, S. ; KAXIRAS, E.: Hydrogen bonding and stacking interactions of nucleic acid base pairs: A density-functional-theory based treatment. In: *J. Chem. Phys.* 114 (2001), Nr. 12, S. 5149–5155
105. TUMA, C. ; BOESE, A. D. ; HANDY, N. C.: Predicting the binding energies of H-bonded complexes: A comparative DFT study. In: *Phys. Chem. Chem. Phys.* 1 (1999), Nr. 17, S. 3939–3947
106. KIENINGER, M. ; SUHAI, S.: Conformational and energetic properties of the ammonia dimer - comparison of post-hartree-fock and density functional methods. In: *J. Comput. Chem.* 17 (1996), Nr. 13, S. 1508–1519
107. NOVOA, J. J. ; SOSA, C.: Evaluation of the density-functional approximation on the computation of hydrogen-bond interactions. In: *J. Phys. Chem.* 99 (1995), Nr. 43, S. 15837–15845
108. GRIMME, S.: Semiempirical hybrid density functional with perturbative second-order correlation. In: *J. Chem. Phys.* 124 (2006), Nr. 2, S. 034108
109. POPLE, J. A. ; HEAD-GORDON, M. ; RAGHAVACHARI, K.: Quadratic configuration-interaction - A general technique for determining electron correlation energies. In: *J. Chem. Phys.* 87 (1987), Nr. 10, S. 5968–5975
110. TSUZUKI, S. ; UCHIMARU, T. ; MATSUMURA, K. ; MIKAMI, M. ; TANABE, K.: Effects of basis set and electron correlation on the calculated interaction energies of hydrogen bonding complexes: MP2/cc-pV5Z calculations of H₂O· · · MeOH, H₂O· · · Me₂O, H₂O· · · H₂CO, MeOH· · · MeOH, and HCOOH· · · HCOOH complexes. In: *J. Chem. Phys.* 110 (1999), Nr. 24, S. 11906–11910
111. AHLRICHS, R. ; BÄR, M. ; BARON, H.-P. ; BAUERNSCHMITT, R. ; BÖCKER, S. ; DEGLMANN, P. ; EHRIG, M. ; EICHKORN, K. ; ELLIOTT, S. ; FURCHE, F. ; HAASE, F. ; HÄSER, M. ; HÄTTIG, C. ; HORN, H. ; HUBER, C. ; HUNJAR, U. ; M.KATTANNEK ; KÖHN, A. ; KÖLMEL, C. ; KOLLWITZ, M. ; MAY, K. ; OCHSENFELD, C. ; ÖHM, H. ; SCHÄFER, A. ; SCHNEIDER, U. ; SIERKA, M. ; TREUTLER, O. ; UNTERREINER, B. ; ARNIM, M. von ; WEIGEND, F. ; WEIS, P. ; WEISS, H.: *TURBOMOLE (Vers. 5.6)*, 2002
112. AHLRICHS, R. ; BÄR, M. ; HÄSER, M. ; HORN, H. ; KÖLMEL, C.: Electronic structure calculations on workstation computers: The program system TurboMole. In: *Chem. Phys. Lett.* 162 (1989), Nr. 3, S. 165–169
113. TREUTLER, O. ; AHLRICHS, R.: Efficient molecular numerical integration schemes. In: *J. Chem. Phys.* 102 (1995), Nr. 1, S. 346–354
114. EICHKORN, K. ; WEIGEND, F. ; TREUTLER, O. ; AHLRICHS, R.: Auxiliary basis sets for main row atoms and transition metals and their use to approximate Coulomb potentials. In: *Theor. Chem. Acc.* 97 (1997), Nr. 1–4, S. 119–124
115. SCHÄFER, A. ; HUBER, C. ; AHLRICHS, R.: Fully optimized contracted Gaussian basis set of triple zeta valence quality for atoms Li to Kr. In: *J. Chem. Phys.* 100 (1994), Nr. 8, S. 5829–5835
116. SCHÄFER, A. ; HORN, H. ; AHLRICHS, R.: Fully optimized contracted Gaussian basis set for atoms Li to Kr. In: *J. Chem. Phys.* 97 (1992), S. 2571

117. DUNNING, JR., T. H.: Gaussian-basis sets for use in correlated molecular calculations: 1. The atoms boron through neon and hydrogen. In: *J. Chem. Phys.* 90 (1989), Nr. 2, S. 1007–1023
118. WOON, D. E. ; DUNNING, JR., T. H.: Gaussian-basis sets for use in correlated molecular calculations: 3. The atoms aluminum through argon. In: *J. Chem. Phys.* 98 (1993), Nr. 2, S. 1358–1371
119. AMOS, R. D. ; BERNHARDSSON, A. ; BERNING, A. ; CELANI, P. ; COOPER, D. L. ; DEEGAN, M. J. O. ; DOBBYN, A. J. ; ECKERT, F. ; HAMPEL, C. ; HETZER, G. ; KNOWLES, P. J. ; KORONA, T. ; LINDH, R. ; LLOYD, A. W. ; MCNICHOLAS, S. J. ; MANBY, F. R. ; MEYER, W. ; MURA, M. E. ; NICKLASS, A. ; PALMIERI, P. ; PITZER, R. ; RAUHUT, G. ; SCHÜTZ, M. ; SCHMANN, U. ; STOLL, H. ; STONE, A. J. ; TARRONI, R. ; THORSTEINSSON, T. ; WERNER, H.-J.: *MOLPRO, a package of ab initio programs designed by H.-J. Werner and P. J. Knowles*, 2003
120. HÄTTIG, C. ; WEIGEND, F.: CC2 excitation energy calculations on large molecules using the resolution of the identity approximation. In: *J. Chem. Phys.* 113 (2000), Nr. 13, S. 5154–5161
121. HÄTTIG, C. ; HALD, K.: Implementation of RI-CC2 triplet excitation energies with an application to trans-azobenzene. In: *Phys. Chem. Chem. Phys.* 4 (20052), Nr. 11, S. 2111–2118
122. BOYS, S. F. ; BERNARDI, F.: Calculation of small molecular interactions by differences of separate total energies - Some procedures with reduced errors. In: *Mol. Phys.* 19 (1970), Nr. 4, S. 553
123. RAUB, S. ; JANSEN, G.: A quantitative measure of bond polarity from the electron localization function and the theory of atoms in molecules. In: *Theor. Chem. Acc.* 106 (2001), Nr. 3, S. 223–232
124. RAUB, S.: *Quantitative Charakterisierung von Bindungspolaritäten mittels der Elektronenlokalisationsfunktion und der Theorie von Atomen in Molekülen, Diplomarbeit, Heinrich-Heine-Universität Düsseldorf*. 2001
125. FRISCH, M. J. ; TRUCKS, G. W. ; SCHLEGEL, H. B. ; SCUSERIA, G. E. ; ROBB, M. A. ; CHEESEMAN, J. R. ; ZAKRZEWSKI, V. G. ; MONTGOMERY, JR., J. A. ; STRATMANN, R. E. ; BURANT, J. C. ; DAPPRIICH, S. ; MILLAM, J. M. ; DANIELS, A. D. ; KUDIN, K. N. ; STRAIN, M. C. ; FARKAS, O. ; TOMASI, J. ; BARONE, V. ; COSSI, M. ; CAMMI, R. ; MENNUCCI, B. ; POMELLI, C. ; ADAMO, C. ; CLIFFORD, S. ; OCHTERSKI, J. ; PETERSSON, G. A. ; AYALA, P. Y. ; CUI, Q. ; MOROKUMA, K. ; MALICK, D. K. ; RABUCK, A. D. ; RAGHAVACHARI, K. ; FORESMAN, J. B. ; CIOSLOWSKI, J. ; ORTIZ, J. V. ; STEFANOV, B. B. ; LIU, G. ; LIASHENKO, A. ; PISKORZ, P. ; KOMAROMI, I. ; GOMPERTS, R. ; MARTIN, R. L. ; FOX, D. J. ; KEITH, T. ; AL-LAHAM, M. A. ; PENG, C. Y. ; NANAYAKKARA, A. ; GONZALEZ, C. ; CHALLACOMBE, M. ; GILL, P. M. W. ; JOHNSON, B. ; CHEN, W. ; WONG, M. W. ; ANDRES, J. L. ; GONZALEZ, C. ; HEAD-GORDON, M. ; REPLOGLE, E. S. ; POPLE, J. A.: *Gaussian 98, Revision A.4*. Gaussian, Inc., Pittsburgh PA, 1998
126. RAUB, S. ; STEFFEN, A. ; KÄMPER, A. ; MARIAN, C. M.: Chemically diverse scoring in FlexX. In: *J. Med. Chem. submitted...*
127. EWING, T. J. ; MAKINO, S. ; SKILLMAN, A. G. ; KUNTZ, I. D.: DOCK 4.0: Search strategies for automated molecular docking of flexible molecule databases. In: *J. Comput.-Aided Mol. Des.* 15 (2001), S. 411–428
128. STAHL, M. ; RAREY, M.: Detailed analysis of scoring functions for virtual screening. In: *J. Med. Chem.* 44 (2001), Nr. 7, S. 1035–1042
129. KITCHEN, D. B. ; DECORNEZ, H. ; FURR, J. R. ; BAJORATH, J.: Docking and scoring in virtual screening for drug discovery: Methods and applications. In: *Nat. Rev. Drug Discovery* 3 (2004), Nr. 11, S. 935–949
130. BÖHM, H.-J. ; STAHL, M.: Rapid empirical scoring functions in virtual screening applications. In: *Med. Chem. Res.* 9 (1999), Nr. 7–8, S. 445–462
131. MITCHELL, J. B. O. ; LASKOWSKI, R. A. ; ALEX, A. ; THORNTON, J. M.: BLEEP - Potential of mean force describing protein-ligand interactions: I. Generating potential. In: *J. Comput. Chem.* 20 (1999), Nr. 11, S. 1165–1176

132. MUEGGE, I. ; MARTIN, Y. C.: A general and fast scoring function for protein-ligand interactions: A simplified potential approach. In: *J. Med. Chem.* 42 (1999), Nr. 5, S. 791–804
133. MUEGGE, I. ; MARTIN, Y. C. ; HAJDUK, P. H. ; FESIK, S. W.: Evaluation of PMF scoring in docking weak ligands to the FK506 binding protein. In: *J. Med. Chem.* 42 (1999), Nr. 14, S. 2498–2503
134. GOHLKE, H. ; HENDLICH, M. ; KLEBE, G.: Knowledge-based scoring function to predict protein-ligand interactions. In: *J. Mol. Biol.* 295 (2000), Nr. 2, S. 337–356
135. VELEC, H. F. ; GOHLKE, H. ; KLEBE, G.: DrugScore^{CSD} – Knowledge-based scoring function derived from small molecule crystal data with superior recognition rate of near-native ligand poses and better affinity prediction. In: *J. Med. Chem.* 48 (2005), S. 6296–6303
136. KLEBE, G.: The use of composite crystal-field environments in molecular recognition and the de-novo design of protein ligands. In: *J. Mol. Biol.* 237 (1994), Nr. 2, S. 221–235
137. LINNAINMAA, S. ; HARWOOD, D. ; DAVIS, L. S.: Pose determination of a three-dimensional object using triangle pairs. In: *IEEE Trans. Pattern Anal. Mach. Intell.* 10 (1988), Nr. 5, S. 634–647
138. KLEBE, G. ; MIETZNER, T.: A fast and efficient method to generate biologically relevant conformations. In: *J. Comput.-Aided Mol. Des.* 8 (1994), Nr. 5, S. 583–606
139. KLEBE, G.: Toward a more efficient handling of conformational flexibility in computer-assisted modelling of drug molecules. In: MÜLLER, K. (Hrsg.): *De Novo Design* Bd. 3. Leiden : Escom, 1995, S. 99–114
140. HILLER, C. ; GASTEIGER, J.: *Ein automatisierter Molekülbaukasten*. Springer Berlin, 1987
141. SADOWSKI, J. ; RUDOLPH, C. ; GASTEIGER, J.: The generation of 3D-models of host guest complexes. In: *Anal. Chim. Acta* 265 (1992), S. 233–241
142. SADOWSKI, J. ; GASTEIGER, J.: Polygon patterns for the generation of conformations of large rings. In: ZIESSOW, D. (Hrsg.): *Software Development in Chemistry* Bd. 7. Gesellschaft Deutscher Chemiker, Frankfurt am Main, 1993, S. 65–76
143. SCHÖNBERGER, H. ; SCHWAB, C. H. ; HIRSCH, A. ; GASTEIGER, J.: Molecular modeling of fullerene dendrimers. In: *J. Mol. Model* 6 (2000), Nr. 3, S. 379–395
144. PRESS, W. H. ; TEUKOLSKY, S. A. ; VETTERLING, W. T. ; FLANNERY, B. P.: *Numerical Recipes*. 2nd. Cambridge University Press, 1992
145. ALBRECHT, G. ; COREY, R. B.: The crystal structure of glycine. In: *J. Am. Chem. Soc.* 61 (1939), S. 1087–1103
146. BERMAN, H. M. ; WESTBROOK, J. ; FENG, Z. ; GILILAND, G. ; BAHT, T. N. ; WEISSIG, H. ; SHINDYALOV, I. N. ; BOURNE, P. E.: The Protein Data Bank. In: *Nucleic Acid Res.* 28 (2000), Nr. 1, S. 234–242
147. LEVITT, M. ; PARK, B. H.: Water - now you see it, now you don't. In: *Structure* 1 (1993), Nr. 4, S. 223–226
148. CARUGO, O. ; BORDO, D.: How many water molecules can be detected by protein crystallography? In: *Acta Crystallogr. Sect. D* 55 (1999), S. 479–483
149. GARCÍA-SOSA, A. T. ; MANCERA, R. L. ; DEAN, P. M.: WaterScore: A novel method for distinguishing between bound and displaceable water molecules in the crystal structure of the binding site of protein-ligand complexes. In: *J. Mol. Model* 9 (2003), Nr. 3, S. 172–182
150. GIACOVAZZO, C. ; MONACO, H. L. ; ARTIOLI, G. ; VITERBO, D. ; FERRARIS, G. ; GILLI, G. ; ZANOTTI, G. ; CATTI, M.: *Fundamentals of Crystallography*. Oxford University Press, 1992
151. MCPHALEN, C. A. ; JAMES, M. N. G.: Structural comparison of 2 serine proteinase protein inhibitor complexes – Eeglin-C-Subtilisin Carlsberg and CI-2-subtilisin Novo. In: *Biochemistry* 27 (1988), Nr. 17, S. 6582–6598
152. FRANKS, F.: *Water – A Comprehensive Treatise*. Bd. 6: *Recent Advances*. Plenum Press, New York, London, 1979

153. NAKASAKO, M.: Large-scale networks of hydration water molecules around bovine β -trypsin revealed by cryogenic X-Ray crystal. In: *J. Mol. Bio.* 289 (1999), Nr. 3, S. 547–564
154. FAERMAN, C. H. ; KARPLUS, P. A.: Consensus preferred hydration sites in 6 FKBP12 drug complexes. In: *PROTEINS* 23 (1995), Nr. 1, S. 1–11
155. SCHWABE, J. W. R.: The role of water in protein DNA interactions. In: *Curr. Opin. Struct. Biol.* 7 (1997), Nr. 1, S. 126–134
156. BAKER, E. N. ; HUBBARD, R. E.: Hydrogen bonding in globular proteins. In: *Prog. Biophys. Mol. Biol.* 44 (1984), Nr. 2, S. 97–179
157. RASHIN, A. A. ; IOFIN, M. ; HONIG, B.: Internal cavities and buried waters in globular-proteins. In: *Biochemistry* 25 (1986), Nr. 12, S. 3619–3625
158. HALGREN, T. A.: Merck Molecular Force Field: II. MMFF94 van der Waals and electrostatic parameters for intermolecular interactions. In: *J. Comput. Chem.* 17 (1996), S. 520–552
159. FUJINAGA, M. ; DELBAERE, L. T. ; BRAYER, G. D. ; JAMES, M. N.: Refined structure of alpha-lytic protease at 1.7 Å resolution - Analysis of hydrogen-bonding and solvent structure. In: *J. Mol. Biol.* 184 (1985), Nr. 3, S. 479–502
160. LORIS, R. ; LANGHORST, U. ; VOS, S. D. ; DECANNIERE, K. ; BOUCKAERT, J. ; MAES, D. ; TRANSUE, T. R. ; STEYAERT, J.: Conserved water molecules in a large family of microbial ribonucleases. In: *PROTEINS: Structure, Function and Genetics* 36 (1999), Nr. 1, S. 117–134
161. FIELDS, B. A. ; GOLDBAUM, F. A. ; DALLACQUA, W. ; MALCHIODI, E. L. ; CAUERHFF, A. ; SCHWARZ, F. P. ; YSERN, X. ; POLJAK, R. J. ; MARIUZZA, R. A.: Hydrogen bonding and solvent structure in an antigen-antibody interface. Crystal structures and thermodynamic characterization of three Fv mutants complexed with lysozyme. In: *Biochemistry* 35 (1996), Nr. 48, S. 15494–15503
162. KÄMPER, A. ; APOSTOLAKIS, J. ; RAREY, M. ; MARIAN, C. M. ; LENGAUER, T.: Fully automated flexible docking of ligands into flexible synthetic receptors using forward and inverse docking strategies. In: *J. Chem. Inf. Model.* 46 (2006), Nr. 2, S. 903–911
163. STEFFEN, A. ; KÄMPER, A. ; LENGAUER, T.: Flexible docking of ligands into synthetic receptors using a two-sided incremental construction algorithm. In: *J. Chem. Inf. Model.* 46 (2006), Nr. 4, S. 1695–1703
164. CLARK, M. ; CRAMER III, R. D. ; VAN OPDENBOSCH, N.: Validation of the general purpose Tripos 5.2 force field. In: *J. Comput. Chem.* 10 (1989), S. 982–1012
165. KORBEL, U. S.: *Identification of Symmetries in Molecules and Complexes*. Saarbrücken, Deutschland, Saarland Universität, Diplomarbeit, 2006
166. WOODWARD, H. W.: *Teas Of The World*. New York : MacMillan, 1980
167. WALDVOGEL, S. R. ; FRÖHLICH, R. ; SCHALLEY, C. A.: Erster künstlicher Koffeinrezeptor – ein neues Konzept zur Komplexierung alkylierter Oxopurine. In: *Angew. Chem.* 112 (2000), Nr. 14, S. 2280–2583
168. RUPP, B. ; RAUB, S. ; MARIAN, C. ; HÖLTJE, H.-D.: Molecular design of two sterol 14- α -demethylase homology models and their interactions with the azole antifungals ketoconazole and bifonazole. In: *J. Comput.-Aided Mol. Des.* 19 (2005), Nr. 3, S. 149–163
169. GEORGOPAPADAKOU, N. H. ; WALSH, T. J.: Antifungal agents: Chemotherapeutic targets and immunologic strategies. In: *Antimicrob. Agents Chemother.* 40 (1996), Nr. 2, S. 279–291
170. LUPETTI, A. ; DANESI, R. ; CAMPA, M. ; DELTACCA, M. ; KELLY, S.: Molecular basis of resistance to azole antifungals. In: *Trends Mol. Med.* 8 (2002), Nr. 2, S. 76–81
171. FROMTLING, R. A.: Overview of medically important antifungal azole derivatives. In: *Clin. Microbiol. Rev.* 1 (1988), Nr. 2, S. 187–217
172. BERENDSEN, H. J. C. ; VAN DER SPOEL, D. ; VAN DRUNEN, R.: GROMACS: A message-passing parallel molecular dynamics implementation. In: *Comp. Phys. Comm.* 91 (1995), Nr. 1–3, S. 43–56

173. LINDAHL, E. ; HESS, B. ; VAN DER SPOEL, D.: GROMACS 3.0: A package for molecular simulation and trajectory analysis. In: *J. Mol. Mod.* 7 (2001), Nr. 8, S. 306–317
174. SONO, M. ; ROACH, M. P. ; COULTER, E. D. ; DAWSON, J. H.: Heme-containing oxygenases. In: *Chem. Rev.* 96 (1996), Nr. 7, S. 2841–2887
175. FILATOV, M. ; HARRIS, N. ; SHAIK, S.: A theoretical study of electronic factors affecting hydroxylation by model ferryl complexes of cytochrome P-450 and horseradish peroxidase. In: *J. Chem. Soc., Perk. Trans.* 2 (1999), Nr. 3, S. 399–410
176. JONAS, V. ; THIEL, W.: Theoretical study of the vibrational-spectra of the transition-metal carbonyls $M(CO)_6$ ($M=Cr, Mo, W$), $M(CO)_5$ ($M=Fe, Ru, Os$), and $M(CO)_4$ ($M=Ni, Pd, Pt$). In: *J. Chem. Phys.* 102 (1995), Nr. 21, S. 8474–8484

Die hier vorgelegte Dissertation habe ich eigenständig und ohne unerlaubte Hilfe angefertigt. Die Dissertation wurde in der vorgelegten oder in ähnlicher Form noch bei keiner anderen Institution eingereicht. Ich habe bisher keine erfolglosen Promotionsversuche unternommen.

Düsseldorf, den

(Stephan Raub)