

**Ermittlung, Charakterisierung  
und effiziente Verarbeitung  
von Oberflächenparametern  
für die Simulation von  
molekularen Wechselwirkungen der Proteine**

Inaugural - Dissertation

zur

Erlangung des Doktorgrades  
der Mathematisch-Naturwissenschaftlichen Fakultät  
der Universität zu Köln

vorgelegt von

**Peter Krämer**

aus Bergisch Gladbach

Köln 2001

Berichterstatter:

Prof. Dr. Dietmar Schomburg

Priv.-Doz. Dr. Sabine Waffenschmidt

Tag der mündlichen Prüfung:

13. November 2001

My heartfelt condolences  
to the families and friends  
of any victim  
of national or international terrorism.

11. September 2001



## Danksagung

Meinem Betreuer und Doktorvater Herrn Professor Dr. D. Schomburg danke ich für die Gelegenheit, dieses Thema, das von Anfang bis Ende sehr interessant war und auch noch weiter voller Entwicklungsmöglichkeiten ist, in seiner Arbeitsgruppe und unter seiner Anleitung bearbeiten zu dürfen.

Priv.-Doz. Dr. Sabine Waffenschmidt hat sich ein großes Lob verdient, nach der Diplomarbeit auch noch für diese Ausführungen das Coreferat zu übernehmen.

Andrea und Jan haben mir in den Jahren die Möglichkeit gegeben, in einer wunderbaren Arbeitsatmosphäre mit lieben Leuten richtig gut arbeiten zu können. Ich bin sehr froh, bei Euch gelandet zu sein und werde die Spätnachmittagspausen vermissen.

Ein Dankeschön an Jan und Klaus sowie Andrea, Helge und Sasha, die sich beim Korrekturlesen tapfer durch meinen Schreibstil und das ungewohnte und abstrakte Thema gebissen haben.

Allen weiteren derzeitigen und ehemaligen Post-Docs, Doktoranden, Diplomanden und Praktikanten der Arbeitsgruppe danke ich auch ohne Namensnennung ganz besonders für ihre Hilfsbereitschaft, Denkanstöße und weitere, mehr oder weniger direkte Beiträge zum Zustandekommen dieser Ausführung.

Meine geliebten Eltern, meine geliebte Schwester und alle meine Freundinnen und Freunde, die ich so gerne hier einzeln aufführen würde, haben es eigentlich in keinster Weise verdient, erst an dieser letzten Stelle angesprochen zu werden. Diese Arbeit ist Euch gewidmet - auch wenn das überwältigende Leben in der Welt gerade die entsprechende Seite einnimmt. Ich muss und will mich auf das Herzlichste für die großartige Unterstützung in vielerlei Lebenslagen bedanken, die ich von Euch erfahren habe. Einige Personen haben ganz besondere Bedeutungen für mich aber alle haben natürlich zu meinem bisherigen Leben mit seinen spannenden Hochs und Tiefs beigetragen.

love, take care   yours pk



## **Erklärung**

Ich versichere hiermit, dass ich die von mir vorgelegte Dissertation selbstständig angefertigt, die benutzten Quellen und Hilfsmittel vollständig angegeben und die Stellen der Arbeit - einschließlich Tabellen und Abbildungen -, die anderen Werken im Wortlaut oder dem Sinn nach entnommen sind, in jedem Einzelfall als Entlehnung kenntlich gemacht habe; dass diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; dass sie - abgesehen von unten angegebenen Teilpublikationen - noch nicht veröffentlicht worden ist sowie, dass ich eine solche Veröffentlichung vor Abschluss des Promotionsverfahrens nicht vornehmen werde. Die Bestimmungen dieser Promotionsordnung sind mir bekannt.

Köln, 19. September 2001

Peter Krämer

Berichterstatter:	Prof. Dr. Dietmar Schomburg Priv.-Doz. Dr. Sabine Waffenschmidt
Tag der mündlichen Prüfung:	13. November 2001
Teilpublikationen:	keine





# Inhaltsverzeichnis

## Vorspann

Danksagung .....	iii
Erklärung .....	v
Inhaltsverzeichnis .....	vii
Tabellarischer Lebenslauf .....	xiii

<b>I</b>	<b>Einleitung</b>	<b>1</b>
1	Proteinkomplexe und Proteindocking .....	1
1.1	Kenntnisse über Proteinkomplexe .....	1
1.2	Leistungen im Proteindocking .....	4
2	Motivation und Lösungsstrategie der Arbeit .....	9
2.1	Ideen und Ziele der Arbeit .....	9
2.2	Komponenten des vorgestellten Docking-Verfahrens .....	10
<b>II</b>	<b>Theoretische und methodische Grundlagen</b>	<b>15</b>
3	Wasserstoffbrückenbindungen .....	15
4	Methoden der transformations-invarianten Parametrisierung und des <i>Geometric Hashing</i> .....	21
5	Räumliche Transformationen und Clusteranalysen von Transformationsangaben .....	25
5.1	Eindeutige Transformationsbeschreibung im dreidimensionalen Raum ..	25
5.2	Eulerwinkel mit X-Konvention .....	26
5.3	Gruppierung und Clusterung gleichartiger Transformationen .....	27
5.4	Regelmäßige Unterteilung des Rotationsraums .....	28
<b>III</b>	<b>Material und Methoden</b>	<b>31</b>
6	Hilfsmittel .....	31
6.1	Programmiersprachen, Rechensysteme, Betriebssysteme und Compiler ..	31
6.2	Quelltexte und Computerprogramme .....	32
6.3	Proteinstrukturdaten .....	33
7	Erstellung der Dockingrepräsentation einzelner Proteinmoleküle .....	34

7.1	Qualifizierung und Charakterisierung bindungsfähiger Atome . . . . .	35
7.1.1	Qualifizierung bindungsfähiger Atome . . . . .	35
7.1.2	Charakterisierung bindungsfähiger Atome . . . . .	36
7.1.3	Disqualifizierung bindungsfähiger Atome aufgrund intramolekularer Wasserstoffbrückenbindungen . . . . .	37
7.1.4	Zusammenfassung der Repräsentation bindungsfähiger Atome . .	38
7.2	Monomolekulare transformations-invariante Eigenschaften eines Paares bindungsfähiger Atome . . . . .	39
7.3	Monomolekulare transformations-invariante Eigenschaften eines Triplets bindungsfähiger Atome zuzüglich Triplett-Vektordivergenz und Triplett- Bindungsseite . . . . .	42
7.4	Erstellung der Molekülrepräsentationen unter Selektion monomolekularer transformations-invarianter Eigenschaften . . . . .	44
8	Docking zweier Proteinmoleküle . . . . .	46
8.1	Bimolekulare transformations-invariante Eigenschaften der Kombination zweier Paare bindungsfähiger Atome . . . . .	47
8.2	Bimolekulare transformations-invariante Eigenschaften der Kombination zweier Triplets bindungsfähiger Atome . . . . .	50
8.3	Molekülorientierung und Erstellung der Transformationscluster . . . . .	51
8.3.1	Orientierungsmodell der Molekülanordnung . . . . .	52
8.3.2	Erstellung der Transformationsvorschriften . . . . .	55
8.3.3	Bildung der finalen Transformationscluster . . . . .	57
8.4	Qualifizierung und Sortierung der Transformationscluster . . . . .	58
8.5	Gesamtablauf eines Dockings unter Erläuterung aller möglichen Anordnungen einer Paar-Paar- bzw. Triplett-Triplett-Kombination . . . . .	59
8.5.1	Übersicht über die Arbeitsschritte . . . . .	59
8.5.2	Ausführungen zu den möglichen Anordnungen einer Paar-Paar- bzw. Triplett-Triplett-Kombination . . . . .	62
9	Auswahl der Test-Datensätze . . . . .	64
10	Festlegung der Ensembles von Kriteriengrenzwerten und Identifikation der Referenzergebnisse . . . . .	67
10.1	Real existente Werte für die transformations-invarianten Eigenschaften .	67
10.2	Zusammenstellung der verwendeten Ensembles von Kriteriengrenzwerten . . . . .	67
10.3	Erstellung und Wiederfindung der Referenzergebnisse . . . . .	70

10.4	Bedeutungen und Auswirkungen zu differenzierender Typen von Selektionskriterien .....	70
<b>IV</b>	<b>Ergebnisse und Diskussion</b>	<b>73</b>
11	Festlegung selektierender Grenzwerte .....	75
11.1	Identifikation bindungsfähiger Atome .....	76
11.1.1	Wasserstoffbrückenbindungsfähige Atomtypen .....	76
11.1.2	Nähe zur lösungsmittelzugänglichen Moleküloberfläche .....	76
11.1.3	Ausrichtung der Bindungscharakteristik .....	77
11.1.4	Disqualifikation aufgrund intramolekularer Wasserstoffbrückenbindungen .....	78
11.2	Auswahl mono- und bimolekularer transformations-invarianter Eigenschaften .....	80
11.2.1	Verteilung unselektiert auftretender Atomabstände .....	80
11.2.2	Grundlegende Tendenzen der Werteverteilungen unter Vorselektion der Atomabstände von 2 Å bis 20 Å .....	81
11.2.3	Die drei Ensembles der Selektionsgrenzwerte .....	85
11.2.4	Triplett-Vektordivergenz und Triplett-Bindungsseite .....	90
11.3	Orientierungsabhängige Nachbewertung und Qualifizierung kriterienkonformer Kandidaten .....	91
11.3.1	Nachbewertung multipler Rotationsbestimmungen eines Atomensembles .....	91
11.3.2	Translationsabhängige Erhaltung in der Ergebniscusterung .....	94
11.4	Tabellarische Zusammenstellung aller verwendeten Selektionswerte .....	97
12	Beschreibung und Eignungsbestimmung der Test-Datensätze für die vorgestellten Dockingmethoden und Selektionsgrenzwerte .....	98
12.1	Ausdehnung der Kontaktflächen .....	98
12.2	Anzahl intermolekularer Wasserstoffbrückenbindungen .....	100
12.3	Anzahl erzielbarer Referenzergebnisse .....	101
12.4	Eignung der Test-Datensätze für unterschiedliche Methoden und Grenzwertensembles .....	102
12.5	Strukturelle Grundlagen der Unterschiede obligatorischer und fakultativer Proteinkomplexe .....	104

12.6	Überlegungen zu einem präselektiven Eignungstest einzelner Proteinmoleküle für die vorgestellten Methoden und Grenzwertensembles	106
13	Algorithmen der Kombinatorik und die Spezifität erzielbarer Suchergebnisse	109
13.1	Kombinatorik der Suchmethode	109
13.2	Spezifität von Ergebnissen	114
14	Evaluierungs- und Wiederfindungsquoten	117
14.1	Erzielte Evaluierungsquoten	118
14.2	Erzielte Wiederfindungsquoten	121
14.3	Diskussion der Evaluierungs- und Wiederfindungsquoten	122
15	Orientierungsabhängige Erhaltungsquoten	124
15.1	Einflüsse und Qualität der Rotationsbestimmung	124
15.2	Qualität der Translationsbestimmung	128
15.3	Diskussion der Rotationsbestimmung und Überlegungen zu alternativen Orientierungsmethoden	130
16	Ergebnisqualität vollständiger Dockingexperimente	134
16.1	Docking obligatorischer Proteinkomplexe	135
16.1.1	Ergebnistabellen	136
16.1.2	Spezifität richtiger Ergebnisse	143
16.1.3	Qualifizierung richtig rotierter Transformationscluster	148
16.1.4	Qualität der Translationsbestimmung	152
16.2	Docking fakultativer Proteinkomplexe	153
16.2.1	Eignung fakultativer Proteinkomplexe	153
16.2.2	Sinnvolle Durchführbarkeit von Dockingexperimenten	155
16.2.3	Spezifität, Qualifizierung und Translationsbestimmung	156
16.3	Diskussion der Ergebnisqualität	160
16.4	Enger positionierte Rotationsrepräsentanten und alternative statische und dynamische Qualifizierungsmethoden	165
17	Programmlaufzeiten in Abhängigkeit von den kombinatorischen Effekten	172
17.1	Erstellung der Molekülrepräsentation	172
17.2	Docking zweier Moleküle	173
17.2.1	Programmlaufzeiten	174
17.2.2	Abhängigkeiten der Dockingdauer	177
17.3	Diskussion der Programmlaufzeiten	183

---

<b>V</b>	<b>Abschlussdiskussion, Ausblick und Zusammenfassungen</b>	<b>185</b>
18	Zusammenfassende Diskussion und Ausblick .....	185
19	Kurzzusammenfassung .....	197
20	Abstract .....	198
<b>Anhang</b>		<b>199</b>
A1	Abkürzungen, Symbole und Begriffe .....	199
A2	Wasserstoffbrückenbindungsatome, van-der-Waals Radien .....	200
A3	Übersicht über die Positionsbestimmung der Protonen .....	202
A4	Verwendete Test-Datensätze .....	204
A5	Beschreibungen und Eignungsbestimmung der Test-Datensätze .....	206
A6	Differenzierte Zulassung der fakultativer Protein-Protein-Komplexe .....	209
A7	Evaluierungs-, Wiederfindungs und Erhaltungsquoten .....	210
A8	Die Translationsbestimmung obligatorischer Komplexe .....	212
A9	Ergebnistabellen des Dockings fakultativer Komplexe .....	213
A9.1	Docking mit engem Grenzwertensemble .....	213
A9.2	Docking mit mittlerem Grenzwertensemble .....	216
A9.3	Docking mit weitem Grenzwertensemble .....	219
A10	Ab2/Ab3-Werte, Evaluierungsaufrufe, Kandidatenzahlen und Dockinglaufzeiten .....	223
<b>Literaturverzeichnis</b>		<b>227</b>



## Tabellarischer Lebenslauf

19. 02. 1973 geboren in Bergisch Gladbach, NRW  
deutsche Staatsangehörigkeit
- 1979 - 1983 Katholische Grundschule in Bergisch Gladbach-Paffrath
- 1983 - 1992 Nicolaus Cusanus Gymnasium in Bergisch Gladbach
12. 06. 1992 Abitur mit den Leistungskursen Biologie und Geographie
- 1993 Beginn des Biologiestudiums an der Universität zu Köln
- 1995 Vordiplomprüfung
- 1997 Abschluss des Hauptstudiums  
Diplomprüfungen in Biochemie, Genetik und Pharmakologie
- 1997 - 1998 Diplomarbeit bei Herrn Prof. Dr. D. Schomburg  
am Institut für Biochemie der Universität zu Köln  
*“Kriterien zur Vorhersage von Protein-Docking -  
Die Verwendung von Wasserstoffbrückenbindungen“*
02. 10. 1998 Prüfungszeugnis der Diplomprüfung in Biologie
- 1998 - 2001 Doktorarbeit bei bei Herrn Prof. Dr. D. Schomburg  
am Institut für Biochemie der Universität zu Köln  
als Stipendiat im Graduierten Kolleg *“Scientific Computing“*  
*“Ermittlung, Charakterisierung und effiziente Verarbeitung von  
Oberflächenparametern für die Simulation von molekularen  
Wechselwirkungen der Proteine“*





# **I Einleitung**

Diese kurze Einführung soll eine Positionierung der durchgeführten Arbeiten innerhalb der Themengebiete Proteinstrukturen und Proteininteraktionen liefern. Eine Darstellung der Ideen und Zielsetzungen wird die konkrete Motivation verdeutlichen bevor eine Vorstellung des Verfahrens die Übersicht über Sinn und Zweck nachfolgender Ausführungen ermöglicht.

## **1 Proteinkomplexe und Proteindocking**

### **1.1 Kenntnisse über Proteinkomplexe**

Die gezielten Interaktionen von biologischen Molekülen, insbesondere auch der Proteine, gehören zu den fundamentalsten Vorgängen des Lebens. Man unterscheidet Proteininteraktionen mit niedermolekularen Liganden und mit makromolekularen Partnern wie anderen Proteinen oder Nukleinsäuren. Hochspezifische Protein-Protein-Wechselwirkungen als Gegenstand dieser Arbeit sind notwendig, um Strukturen zu bilden oder zu zerlegen, um Stoffwechsel-, Signal- und Genaktivität zu regulieren, um Moleküle zu markieren oder auch um aus mehreren Untereinheiten ein funktionstüchtiges Protein zu bilden. Aus einer Vielzahl von Gründen ist es daher für die Biologie, Biochemie, Pharmazie und Medizin von Interesse, nicht nur Kenntnis über die Strukturen einzelner Proteine, sondern auch Einblicke in den Aufbau und die bedingten Eigenschaften ihrer funktionellen Komplexe zu erlangen [Lengauer *et al.*, 1996].

Mit den Methoden der Röntgenkristallographie und Kernresonanzspektroskopie lassen sich einzelne Biomoleküle und ihre funktionelle Komplexe strukturell bis zu atomarer Auflösung aufklären. Dieses ermöglicht die Beschreibung und Analyse ihrer Funktions- und Interaktionsweisen. Das Wissen über die Grundlagen produktiver Protein-Protein-Interaktionen entspringt der Betrachtung entsprechender Komplexe, deren Strukturen experimentell gelöst und veranschaulicht gemacht wurden.

Eine grundlegende Kategorisierung der bekannten Protein-Protein-Komplexe liegt darin, heterooligomere Komplexe von Immunoglobulinen, heterooligomere Komplexe von Nichtimmunoglobulinen und homooligomere Komplexe von Nichtimmunoglobulinen zu trennen. Immunoglobuline unterscheiden sich deutlich in dem Zustandekommen ihrer Bindungsspezifität von Nichtimmunoglobulinen, da sie im Gegensatz zu diesen nicht evolutiv

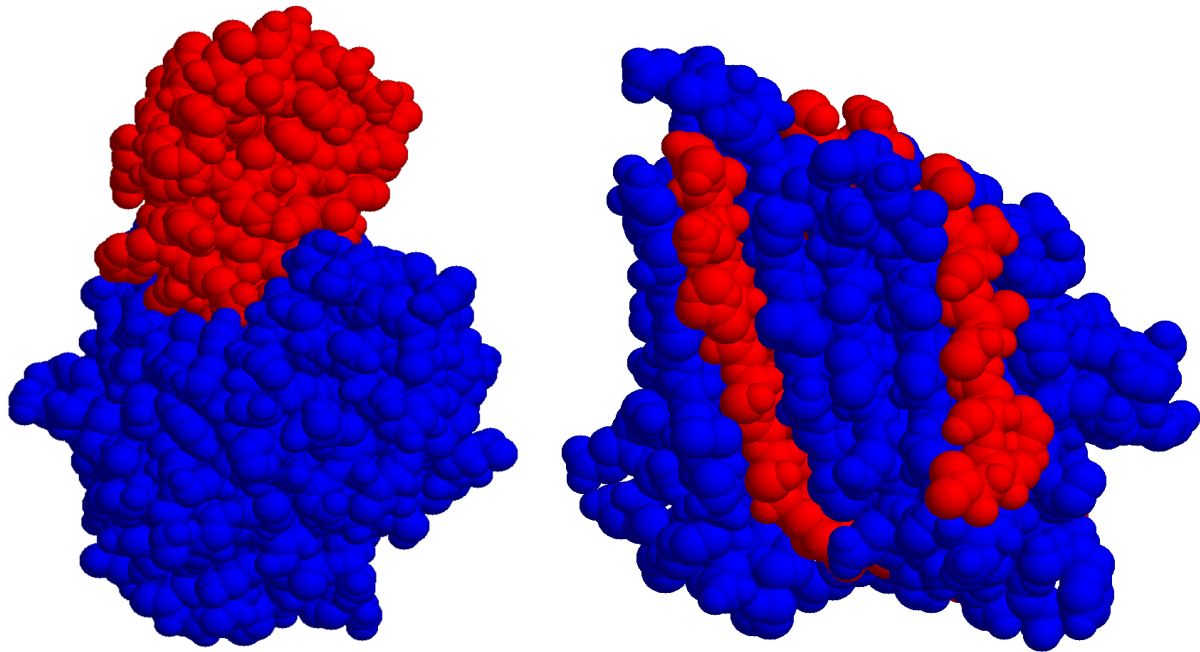
mit einem Bindungspartner verknüpft sind, sondern individuell und zufallsgesteuert Bindungseigenschaften ausbilden [Wilson & Stanfield, 1993]. Hetero- und Homooligomere unterscheiden sich tendenziell in der Zusammensetzung und den Eigenschaften ihrer Kontaktflächen. Erstere bilden auffallend mehr polare Interaktionen aus, bei Homooligomeren haben hydrophobe Interaktionen eine größere Bedeutung [Chothia & Janin, 1975] [Jones & Thornton, 1996 und 1997] [Larsen, Goodsell, 1998].

Für heterooligomere Nichtimmunoglobulin-Komplexe soll hier eine weitere, für die vorliegende Arbeit bedeutende Unterscheidung getroffen werden:

- **Fakultative Protein-Protein-Komplexe** werden von zwei Proteinmolekülen gebildet, die einzeln betrachtet mehr oder weniger globuläre Formen aufweisen, weshalb sie auch unabhängig voneinander in einer Plasmaumgebung stabil beständig sein können. Die Oberflächen sind frei zugänglich und ihre spezifische Interaktion kann ohne externe Hilfestellung stattfinden, wenn die Moleküle aufeinander treffen.
- **Obligatorische Protein-Protein-Komplexe** werden mit der Hilfe von Chaperonen, die u.a. unterstützend bei der richtigen Proteinfaltung eingreifen, direkt während der Proteinbiosynthese zum Zeitpunkt der koordinierten Translation aus den beteiligten Peptidketten gebildet [Tsai *et al.*, 1998]. Diese enge, überaus stabile Verbundenheit geben sie nicht wieder auf und als Einzelmoleküle besitzen die Ketten keinerlei Wirkung und Bedeutung, da ihre Strukturen nicht ausreichend stabil sind und eine spätere Komplexbildung nicht möglich ist. Weitere Vertreter dieses Typs innerhalb dieser Arbeit sind zwei- oder mehrkettige Komplexe, welche als eine einzige, durchgängige Kette synthetisiert und richtig gefaltet wurden. Erst zu einem späteren Zeitpunkt entstanden durch posttranslationale Proteolyse die einzelnen Peptidketten des Heterooligomers. Dieses bleibt nun ebenfalls derart bestehen, da es bei einem Kettenverlust ebenfalls funktionslos würde oder es muss sich von seinem speziellen Propeptidabschnitt trennen, um überhaupt seine Funktionalität zu erhalten. Letzteres beschreibt den klassischen Fall der proteolytischen Aktivierung vieler Proteasen.

Abbildung I-1 soll den generellen Unterschied zweier Vertreter dieser Unterklassifizierung verdeutlichen. Einerseits eine typische Serin-Protease mit Inhibitor-molekül, andererseits ein Molekül, das proteolytisch aufgetrennt wurde. Deutlich ist die unterschiedliche Form der Einzelmoleküle und das grundverschiedene Erscheinungsbild der Komplexe ersichtlich. Die eng umschlungene Anordnung der Ketten des obligatorischen Komplexes vermittelt einen Eindruck, wie instabil die umgebende Kette allein stehend sein wird, dass sie diese Konformation alleine

in Lösung nicht aufrecht erhalten kann und dass die Erzeugung einer derart komplexen Anordnung mit dem Dockingpartner ohne Hilfestellung nicht vorstellbar ist.



**Abbildung I-1:** Beispiele für einen fakultativen (links, 1cse\_EI) und einen obligatorischen (rechts, 2ltn\_AB) Protein-Protein-Komplex.

Da die strukturellen Grundlagen der Unterschiede dieser Komplexe, ihre Auswirkungen und Bedeutungen innerhalb dieser Arbeit erst anhand von Untersuchungsergebnissen während der mehrjährigen Methodenentwicklung auffällig wurden, sollen sie erst in der Ergebnisdiskussion bei der Beschreibung der verschiedenen Test-Datensätze im Zusammenhang dargelegt werden.

Die Grundlagen der molekularen Interaktionen sind in großem Maße vergleichbar, ob es sich um Protein-Protein-Wechselwirkungen oder Proteinfaltung handelt. Um einen spezifischen Protein-Protein-Kontakt ausbilden zu können, bedarf es insbesondere bei Nichtimmunoglobulinen einer intermolekularen Kontaktfläche, die ein ausreichendes, oft hochgradiges Maß komplementärer Oberflächentopographie aufweist [Connolly, 1986] [Katchalski-Katzir *et al.*, 1992] [Norel *et al.*, 1994, 1995 und 1999].

Ebenso bedarf es hier auch einer gewissen Anzahl qualifizierter, intermolekularer Wechselwirkungen, die Bindungsstabilität und Bindungsspezifität vermitteln [Fersht, 1984] [Honig & Yang, 1995]. Im Einzelnen sind die Bedeutung von Wasserstoffbrückenbindungen sowie gleichfalls nichtkovalenter elektrostatischer [Chau, Dean, 1994] [McCoy *et al.*, 1997] und

van-der-Waals-Wechselwirkungen [Roth *et al.*, 1996] [Tsai *et al.*, 1997] vielfach nachgewiesen, in das gängige biochemische Wissen [Stryer, 1975] übernommen und durch aktuellere Untersuchungen verifiziert [Xu *et al.*, 1997] worden. Derartige reversible Bindungen besitzen unterschiedliche Eigenschaften in Geometrie, Stärke sowie Spezifität. Sie spielen auch bedeutende Rollen in der Faltung der Proteine, wo hydrophobe Bindungseffekte und posttranslationale kovalente Verknüpfungen über Disulfidbrücken jedoch zusätzliche Bedeutung erlangen.

Fakultative Molekülkomplexe liegen generell nicht in ihrem globalen Bindungsenergie-Minimum vor, sind daher metastabil und können somit überhaupt reversibel binden und sich wieder trennen [Lin *et al.*, 1995] [Xu *et al.*, 1997]. Der Grund hierfür liegt darin, dass die beiden Einzelmoleküle sich unabhängig voneinander falten und somit auch unabhängig die Beschaffenheit ihrer zukünftigen Kontaktflächen determinieren. Sie befinden sich selbst im globalen Minimum der Faltungsenergie. Kommt es zur Komplexbildung, so ändern sich die Umgebungen für die Aminosäuren dieser Oberflächenbereiche von Lösungsmittelmolekülen im ungebundenen zu Proteinoberflächenbestandteilen des Komplexpartners im gebundenen Zustand. Es werden eine Vielzahl von intermolekularen Bindungen ausgebildet, welche jedoch aufgrund der eingeschränkten Beweglichkeit der Aminosäureseitenketten in der Regel nicht den optimalen Bindungsgeometrien gehorchen können. Kleine Änderungen zur Verbesserung der Oberflächenkonformationen (*induced fit*) können jedoch stattfinden [Norel *et al.*, 1997].

Vielfach lassen sich Proteinkomplexe im Gegensatz zu den einzelnen Molekülen nicht problemlos oder qualitativ ausreichend aufbereiten, um sie nachfolgend mit experimentellen Methoden zu untersuchen. Auch mag die Notwendigkeit zunehmen, mögliche Wechselwirkungen für molekularbiologisch veränderte, vielleicht sogar *de novo* synthetisierte Einzelmoleküle oder auch für Molekülmodelle, wie sie bei Modellierungen nach homologen Molekülmodellen erzeugt werden, vorherzusagen. Somit stellt sich die Aufgabe, die hypothetisch funktionellen Komplexe *in silico* zu erzeugen und qualitativ zu verfeinern [Lengauer *et al.*, 1996].

## 1.2 Leistungen im Proteindocking

Eigentlich wäre es wünschenswert, eine echte Dockingsimulation durchführen zu können, um die Orientierung zu ermitteln, welche zwei Moleküle in ihrem funktionellen Komplex einnehmen. Eine derartige Simulation bestände daraus, die natürlichen Vorgänge nachzuempfinden, welche resultierend aus den verschiedenartigen Bindungskräften bei der

räumlichen Annäherung der Partner in Abhängigkeit der sich verändernden Positionen ihre selektierenden, positionierenden und stabilisierenden Wirkungen ausüben. Derartiges stellt sich jedoch für die Anwendung auf zwei Makromoleküle bislang nicht als praktikabel heraus, da es bereits an dem Wissen über die unterschiedlichen Vorgänge fehlt und derartige Berechnungen zum momentanen Zeitpunkt unverhältnismäßige Rechenzeiten verlangen würden [Lengauer *et al.*, 1996].

Es ist hingegen möglich, verschiedene Faktoren für gegebene Molekülerorientierungen zu bewerten, welche die Interaktion beeinflussen und vermitteln. Hieraus folgt die Möglichkeit, aus den Einzelbewertungen eines umfassenden Kollektivs an Orientierungen die Anordnungen zu isolieren, die aufgrund guter Bindungseigenschaften des gerade verwendeten Bewertungsmodells wahrscheinlich auch gute Orientierungskandidaten sind [Norel *et al.*, 1999]. Hierbei werden also Bindungsfaktoren bewertet und nicht Bindungsvorgänge simuliert.

Größte Bedeutung muss somit der Repräsentation der Moleküloberfläche zugemessen werden. Verschiedenste Modelle werden hierfür verwendet. Beginnend mit einer vollständigen, in ihrer Dichtedichte variablen Punktoberfläche, in der Regel als van-der-Waals- oder lösungsmittelzugängliche Oberfläche, auch Connolly-Oberflächen genannt [Connolly, 1983], welche zunehmend auf spezielle geometrische Extrema (*sparse critical points*) reduziert werden kann [Lin *et al.*, 1994] [Norel *et al.*, 1995], kann auch eine generell Abstraktion durch gitterbasierte Repräsentationen [Vakser, Aflalo, 1994] [Ackermann *et al.*, 1995] durchgeführt werden. Diese kann wiederum Verfeinerungen erfahren durch die Projektion der Eigenschaften auf z.B. zylindrische Oberflächen [Helmer-Citterich, Tramontano, 1994] oder eine Oberflächenbeschreibung durch Kugelflächenfunktionen (*spherical harmonics*) [Ritchie, Kemp, 1999 und 2000].

Die hierbei betrachteten, repräsentierenden Eigenschaften können rein geometrischer Natur sein oder auch die interessanten physikochemischen Eigenschaften der Elektrostatik, Hydrophobizität oder Wasserstoffbrückenbindungsfähigkeit in beliebiger Kombination berücksichtigen [Norel *et al.*, 1999].

Verwendete Methoden, die potentiellen Interaktionspositionen zu ermitteln sind vor allem der bekannte und sehr variabel einzusetzende DOCK-Algorithmus [Kuntz *et al.*, 1982] [Shoichet *et al.*, 1992 und 1993], der *Geometric Hashing*-Algorithmus [Lamdan, Wolfson, 1988] mit seinen Bearbeitungen [Nussinov & Wolfson, 1991] [Fischer *et al.*, 1995][Lin *et al.*, 1996] und die

Verwendung der *Fast Fourier* Transformation [Vakser, Aflalo, 1994] [Gabb *et al.*, 1997]. Systematische, gitterartige Suchmethoden verbieten sich bei derart großen Molekülen aufgrund der kombinatorischen Komplexität [Jiang & Kim, 1991]. Weitere Anstrengungen in der Methodenentwicklung werden unternommen für die Verwendung genetischer Algorithmen und der Clique-Detection, bei der informationstragende Punkte (*feature points*) eine Struktur repräsentieren und zwischen Vergleichsrepräsentationen gemeinsame Subgraphen gesucht werden.

Letztendlich bedarf es immer noch einer Bewertung und Einordnung der vorgeschlagenen Orientierungen und der damit erzielbaren Qualität. In der Regel erfolgt eine Betrachtung der Geometrischen Fitness der vorgeschlagenen Lösung, indem eben die verschiedenen Bindungsfaktoren [Jones & Thornton, 1996], vom Solvens durch den Komplexpartner abgeschirmte Oberfläche (*buried surface*) [Hubbard, 1992] [Kabsch & Sander, 1983], Volumen zwischen den beiden Interfaceflächen (*gap volume*) [Laskowski, 1995], Geometrie der verschiedenen intermolekularen Bindungen [McDonald & Thornton, 1994] [CCP4, 1994] in der jeweiligen Molekülanordnung qualifiziert werden. Eine ausschließliche Bewertung “richtiger“ oder “falscher“ Anordnungen durch Berechnung des Energiegehalts der jeweiligen Bindungen kann mangels Bindungsenergie-Minimums irreführend sein, wie zuvor ausgeführt wurde [Finkelstein *et al.*, 1995].

Die Wahl der betrachteten Eigenschaften, der verwendeten Such- und Selektions-Methoden und der Bewertungsfunktionen lässt einige Freiheitsgrade in der Zusammenstellung einer Suchstrategie zu [Lengauer *et al.*, 1996]. Die bislang besten Ergebnisse liefern für Nichtimmunoglobuline mit Kontaktflächen von 500 Å<sup>2</sup> und mehr die Analysen der geometrischen Oberflächenkomplementarität innerhalb der jeweiligen Interfaceflächen bei hochauflösenden Oberflächenrepräsentationen mittels der *Geometric Hashing*- [Fischer *et al.*, 1995] oder *Fast Fourier* Transformations-Methoden [Meyer, Wilson, Schomburg, 1996] mit einer postselektiven Qualitätsbewertung anhand der resultierenden physikochemischen Gegebenheiten. Derartige Methoden bereiten jedoch teilweise einen Rechenaufwand von durchschnittlich etwa 60 Minuten pro Molekülpaarung [Zimmermann, 1997] bei einem initialen Suchraster über 2600 Rotationen. Bei einem Datenbestand von nur 1000 nicht vorselektierten Strukturlösungen von Einzelmolekülen, entspräche diese bereits dem Bedarf von 42 Tagen Prozessor-Nutzung um ein einzelnen Dockingkandidaten auszutesten. Gilt es also, den zeitlichen Aufwand zu verringern, so muss eine Oberflächenrepräsentation angestrebt werden, die zwar

präzise ist, jedoch nur eine ausreichende aber nichtübermäßige, redundante Anzahl an Vergleichspunkten bietet, wie es bei einer unselektierten Connolly-Oberfläche der Fall ist.

Erschwerend zur Problematik kommt der oben beschriebene Umstand, dass sich bei Proteinen fakultativer Komplexe Abschnitte der Moleküloberflächen im ungebunden und gebundenen Zustand mehr oder weniger in der Orientierung einzelner Aminosäureseitenketten unterscheiden können. Dieser Effekt gewinnt für die Simulation im *in silico* Proteindocking noch an Bedeutung, bedenkt man die gegebenenfalls Artefakt-erzeugende Wirkung einer Proteinkristallisation an den Molekülrandbereichen. Möchte man nun derartige Verzerrungen der Oberflächentopologie in den angesprochenen Methoden eliminieren, so ist die Beschaffenheit der Oberfläche nicht als absolut anzusehen, sondern muss variabel gehalten werden, indem den betroffenen Seitenketten Bewegungsfreiheiten eingeräumt werden. Hierzu könnten Bestrafungswerte bei einer van-der-Waals-Überlappung derartiger Randbereiche herabgesetzt werden oder es werden explizite, sehr rechenaufwendige Moleküldynamik-Simulationen eingesetzt, um alternative Oberflächenbeschaffenheiten im ungebundenen oder gebundenen Zustand zu erzeugen. Man unterscheidet daher das *Rigid Body Docking*, bei dem die Oberflächen der Einzelmoleküle gänzlich unbeweglich gehalten werden und das *Flexible Docking*, bei dem Konformationsänderungen simuliert werden. Da diese Oberflächenänderungen nachgewiesen sind, spricht man bei derartigen Simulationen auch vom *Real World-Problem*.

Die Notwendigkeit der ausreichend feinmaschigen Suche im gesamte, sechsdimensionalen Konformationsraum aus Translation und Rotation, die u.a. für jede Transformation eine Neuberechnung der Atomkoordinaten erfordert, die Problematik der multiplen lokalen Oberflächentopologien und zudem die Kombination beider Probleme mit der Anforderung an eine ausreichend hohe Detailauflösung der Oberflächenrepräsentationen, erzeugen hochkomplexe Rechenprobleme, welche entsprechend erweiterten Rechenbedarf mit sich bringen.

Daher gibt es auch einige Ansätze, derartig aufwendige Untersuchungen bereits im Vorfeld in ihrem Umfang zu begrenzen. Es wird z.B. versucht, entsprechend der Abfolge molekularer Erkennungsvorgänge eine derartige Orientierungssuche sukzessiv und gerichtet in einer selektierenden, verfeinernden Weise durchzuführen, d.h von einer groben zu einer detaillierten Bestimmung [Vakser, 1995 und 1996].

Auch ist es von Nutzen, wenn man bereits Aussagen über die voraussichtlichen Interaktionsabschnitte auf den Oberflächen der Interaktionspartner und man somit versucht, die Berührungspunkte bereits tendenziell vorherzusagen und die feineren Untersuchungen nur für den entsprechend eingegrenzten Konformationsraum durchzuführen [Jones & Thornton, 1997].



## 2 Motivation und Lösungsstrategie der Arbeit

### 2.1 Ideen und Ziele der Arbeit

Logische Folgen der hohen Anforderungen oben beschriebener Methoden und Probleme sowie der steigende Anzahl strukturgelöster Proteine, welche als Dockingpartner zu berücksichtigen sind, sind somit das Streben nach Verbesserung der Oberflächenkorrelationsmethoden und die Suche nach qualifizierten Vorfiltern, welche mit ausreichend geringem Zeit- und Rechenaufwand die Dauer derartiger Hauptuntersuchungen und die Summe der gemeinsamen Laufzeiten vermindert. Als Fernziel mag hier z.B. eine Datenbankanalyse stehen, bei der eine neuartige Proteinstruktur gegen eine mehr oder weniger eingeschränkte Gruppe strukturbekannter, potentieller Dockingpartner auf ihre Interaktionsfähigkeit und Bindungsqualität ausgetestet wird.

Ziel dieser Arbeit ist es, einen Vorfilter zu entwickeln, der den Konformationsraum einschränkt, welcher in einer nachfolgenden, zeitaufwendigen Oberflächenkorrelationsanalyse zur Vorhersage eines Protein-Dockings untersucht werden muss.

Hierfür wird zugrunde gelegt, dass spezialisierte, intermolekulare Wasserstoffbrückenbindungen im Laufe der Evolution in den Kontaktflächen spezifischer produktiver Molekülkomplexe bevorzugt wurden. Die Geometrie und die Häufigkeit dieser Wasserstoffbrückenbindungen sollte somit auffällig besser bzw. höher sein als zwischen nichtspezialisierten Moleküloberflächen, die in räumliche Nähe zueinander gebracht werden. Es wird somit die Molekülorientierung gesucht, bei der überproportional viele und gute Bindungen ausgebildet werden können.

Die verwendete Methode basiert auf der transformations-invarianten Parametrisierung geometrischer Eigenschaften einer Gruppe von je zwei oder drei bindungsfähigen Atomen innerhalb einer Oberfläche. Eine vollständige Oberflächenbeschreibung wird für zwei Moleküle erstellt, indem alle Paare oder Triplets von Atomen zusammengestellt und unter Vorselektion parametrisiert werden. Für jedes Atompaar oder -triplett werden dann geeignete Gruppen von Bindungspartnern auf dem entgegengesetzten Molekül gesucht. Das positive Suchergebnis ist gekennzeichnet durch eine Übereinstimmung bzw. Komplementarität der transformations-invarianten Eigenschaften. Bei der Suche werden vereinfachte Methoden des *Geometric Hashing* verwendet, bei dem eine derartige Liste nach Eigenschaften sortiert und gezielt angesprochen und durchsucht werden kann.

Damit auch hier das Kombinatorikproblem bei der Evaluierung der möglichen Gruppenpaarungen nicht unvorteilhaft groß wird, muss die Anzahl der verwendeten, lokal

bindenden Wechselwirkungen ausreichend gering sein, weshalb die Wahl auf Wasserstoffbrückenbindungen fiel. Es werden somit die Ideen der Oberflächenrepräsentationen von Connolly [Connolly, 1986] und der Arbeitsgruppe um Ruth Nussinov [Lin *et al.*, 1994] [Norel *et al.*, 1994 und 1995] fortgeführt durch eine weitere Verringerung der Anzahl repräsentierender Oberflächenpunkte in Ergänzung um die Betrachtung chemischer Bindungseigenschaften.

Das resultierende, in der Sprache C geschriebene Programm soll ohne weitere Manipulationen die standardisierten Dateien der Protein-Datenbank (*Protein Data Bank*, PDB, Brookhaven National Laboratory) nutzen. Es soll trotz problemloser Parametermanipulation automatisiert ablaufen können und eine einfache Schnittstelle zum Oberflächenkorrelations-Programm besitzen oder eventuell integriert und von diesem selbst aufgerufen werden.

Ein weiteres eng mit der eigenen Vorhersage verknüpft Ziel muss eine Analyse der vorhandenen aus den Kristallstrukturen gelösten nativen Proteinkomplexe sein, um für die verwendeten Evaluierungsparameter eine sinnvolle Wahl der Grenzwerte treffen zu können. Falsch negative Ergebnisse einer derartig selektierenden Voruntersuchung können nachfolgend in der Hauptuntersuchung nicht mehr gefunden werden.

## 2.2 Komponenten des vorgestellten Docking-Verfahrens

Abbildung I-2 enthält eine schematische Übersicht der vielfältigen Arbeitsschritte, die den Umgang mit den einzelnen Molekülen und das eigentliche Moleküldocking betreffen. In der folgenden Kurzbeschreibung finden sich erklärende Ergänzungen sowie zusätzliche Aufgabenstellungen, die im Rahmen dieser Arbeit bearbeitet werden und zur Vor- und Nachbereitung der Simulationen dienen.

### **Datensätze und Parameter:**

- Zunächst einmal erfolgt die Zusammenstellung einer repräsentativen, nichtredundanten Auswahl nativer strukturaufgeklärter Nichtimmunoglobulin-Proteinkomplexe.
- Für diese Komplexe können alle existenten intermolekularen Wasserstoffbrückenbindungen bestimmt werden. Hieraus werden mehrere Ensembles von realitätsbezogenen Selektionsgrenzwerten der transformations-invarianten geometrischen Eigenschaften für die monomolekularen Paaren bzw. Triplets der Wasserstoffbrückenbindungs-Stellen sowie deren bimolekularen Kombinationen festgelegt.

- Durch Anwendung dieser Selektionsgrenzen auf die Kombinationen existenter Wasserstoffbrückenbindungsatome werden die Referenzergebnisse identifiziert, die nachfolgend in den eigenen Interaktionsvorhersagen als richtig positive Ergebnisse wiedergefunden werden müssen.

#### **Aufbereitung der Einzelmoleküle:**

- Die vorhandenen 3D-Informationen der einzelnen Moleküle erfordern sodann eine Aufbereitung, bei der bindungsfähige Atome identifiziert und ihre Bindungsgeometrien, d.h. die Positionen der Wasserstoffatome und die Bindungsvektoren festgelegt werden.
- Bei der Erstellung der Molekülrepräsentationen werden aus den bindungsfähigen Atomen eines Proteins alle möglichen Paare bzw. Triplets zusammengestellt und anhand ihrer monomolekularen transformations-invarianten Eigenschaften vorselektiert. Die verbleibenden Atompaaire bzw. -triplets stellen zusammen mit ihren Bindungsgeometrien eine eigenständige Repräsentation des jeweiligen Moleküls für die hier betrachtete Dockingmethode dar. Sie sind bei einem gewählten Satz von Selektionsgrenzwerten nur einmalig für jedes Molekül anzulegen und könne daher permanent abgelegt und für jedes Docking aufgerufen werden.

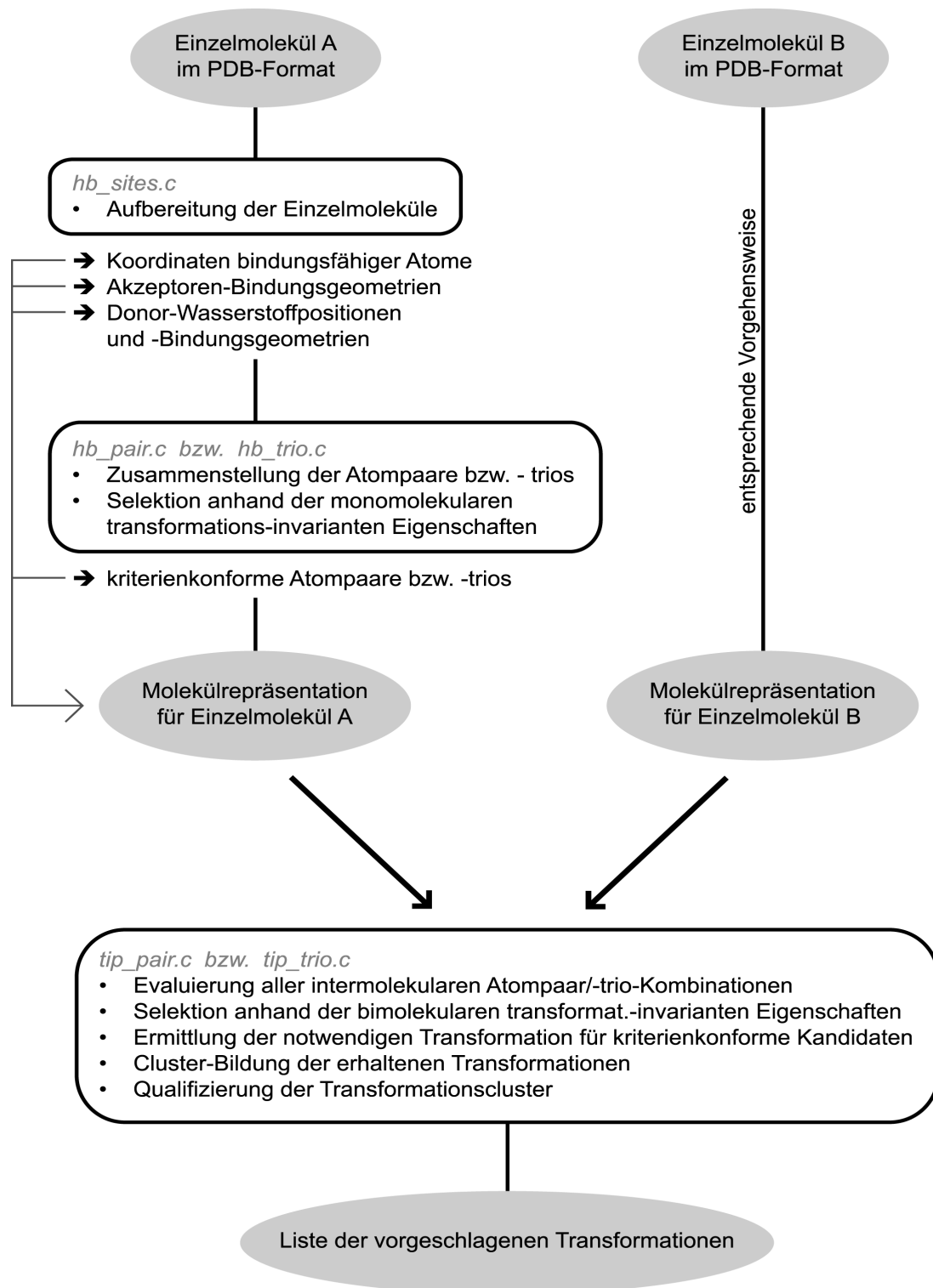
#### **Docking zweier Moleküle:**

- Für das eigentliche Docking zweier Moleküle werden die entsprechenden Datensätze aufgerufen und alle möglichen Kombinationen der Atompaaire bzw. -triplets anhand der entstehenden bimolekularen transformations-invarianten Eigenschaften selektiert. Die kriterienkonformen Atomkombinationen, die sogenannten Kandidaten, spezifizieren somit die Berührungspunkte der beiden Moleküle, bei denen in richtiger Molekülorientierung eine zeitgleiche Ausbildung von zwei bzw. drei intermolekularen Wasserstoffbrückenbindungen möglich scheint.
- Ein sehr bedeutender Arbeitsschritt ist nun die Festlegung der notwendigen, eindeutigen Rotationen und Translationen, die diese erfolgversprechende Orientierung herstellt.
- Gleichartige Transformationen können in gemeinsamen Gruppen zusammengestellt werden, da ja zugrunde gelegt wird, dass die einzige richtige Ausrichtung, welche zum nativen Komplex führt, überproportional häufig über die mögliche Ausbildung multipler, intermolekularer Wasserstoffbrückenbindungen bestimmt werden kann.
- Diese separaten Orientierungen werden abschließend einer Qualifizierung anhand ihrer Belegungsfrequenz mit kriterienkonformen Kandidaten unterzogen. Die erfolgversprechendsten Ergebnisse sollen dabei isoliert und eventuell nachfolgenden

Dockingmethoden zur Verfügung gestellt werden. Im einfachsten Fall ist dieses eine absteigend sortierte Auflistung nach den Belegungszahlen oder ein festgelegter Grenzwert für die Minimalbelegung.

**Qualitative Bewertung einzelner Methoden:**

- Eine Qualitätsbewertung der verwendeten Methoden für die gewählten Referenzkomplexe erfolgt an verschiedenen Punkten im Arbeitsablauf:
  1. Analyse der Evaluierungs- und Wiederfindungsquoten richtig positiver Ergebnisse
  2. Analyse der Quote richtig bestimmter Orientierungen für diese Kandidaten
  3. Analyse der Fähigkeit, diese richtigen Orientierungen in einer Gruppe zusammenzufassen und gegenüber den Gruppen falsch positiver Ergebnisse herauszustellen, d.h. als “potentiell erfolgsversprechend“ zu qualifizieren
  4. Analyse der benötigten Laufzeit des eigentlichen Dockingvorganges



**Abbildung I-2:** Eine Übersicht der generellen Vorgehensweise mit Aufbereitung der Einzelmoleküle zur Erstellung einer methodenspezifischen Molekülrepräsentation für das eigentliche Docking zweier Proteinmoleküle.



## II Theoretische und methodische Grundlagen

In diesem Kapitel sollen theoretische und methodische Grundlagen vorgestellt werden, die im Rahmen dieser Arbeit verwendet werden, jedoch an ihren generellen Prinzipien keine grundlegenden Änderungen oder Weiterentwicklungen erfahren haben.

### 3 Wasserstoffbrückenbindungen

#### **Beschreibung:**

In Wasserstoffbrückenbindungen wird ein Wasserstoffatom gleichzeitig von zwei Atomen gebunden, was dem Intermediatzustand des Protonentransfers einer Brönstedt-Säure zu einer Brönstedt-Base gleichgestellt werden könnte. Genauer gesagt lässt sich eine Wasserstoffbrückenbindung beschreiben als eine schwache elektrostatische Anziehungskraft zwischen dem Proton eines Wasserstoffatoms und der Elektronenwolke eines benachbarten elektronegativen Atoms. Das kovalent bindende Atom nennt man den Proton-Donator, das andere den Proton-Akzeptor.

#### **Atomtypen und -bindungsfähigkeiten:**

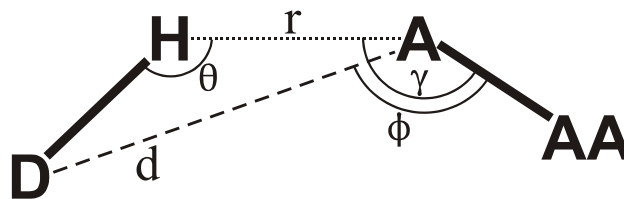
Atome haben entweder keine Wasserstoffbrückenbindungs-Eigenschaften (Kohlenstoffatome), können eindeutig nur Akzeptor- oder nur Donator-Eigenschaften besitzen oder in Abhängigkeit ihrer Umgebung entweder als Akzeptor oder als Donator in Erscheinung treten. In biochemischen Systemen können Stickstoff-, Sauerstoff- und Schwefelatome der Proteinhauptkette und der Aminosäureseitenketten Bindungseigenschaften besitzen, wobei Sulfhydrylgruppen und aromatische Akzeptoren nur sehr schwache Interaktionen ermöglichen und daher in der Regel vernachlässigt werden.

Aufgrund der Auflösungseigenschaften experimenteller Methoden zur Strukturaufklärung ist es häufig nicht möglich, für atomare Substrukturen zuverlässig die richtige Orientierung zu bestimmen. Hierzu zählen insbesondere die Amid-Endgruppen der Aminosäuren Asparagin und Glutamin und die gesamte Ringstruktur des Histidins, wodurch es notwendig wird, die jeweils alternativen Atomanordnungen zu berücksichtigen, so dass den Atomen gegebenenfalls andersartige bzw. doppelte Bindungseigenschaften zugeteilt werden müssen. Eine Tabelle in Anhang A2 führt die bindungsfähigen Atome der Aminosäuren mit Nomenklatur, ihren van-der-Waals-Radien und Bindungseigenschaften auf [McDonald, Thornton, 1994].

**Eigenschaften:**

Wasserstoffbrückenbindungen sind in ihren Eigenschaften gut untersucht und charakterisiert worden [Baker and Hubbard, 1984] [Ippolito, Alexander, Christianson, 1990] [Jeffrey & Saenger, 1991]. Ihre Stärke ist deutlich abhängig von der Ausrichtung der Bindungspartner zueinander und somit am stärksten, wenn eine kollineare Anordnung der Donator-Proton-Bindung zum Akzeptoratom gegeben ist. So weist z.B. das grundlegende und sehr stabile Proteinstrukturmotiv der  $\alpha$ -Helix derartige Arrangements in großer Regelmäßigkeit und Menge auf. Entgegen diesem  $180^\circ$ -Winkel werden Bindungswinkel von  $160^\circ \pm 20^\circ$  als normale bis schwache Bindungen eingestuft, wobei die Kombination mit dem Bindungsabstand zwischen Proton und Akzeptor zu berücksichtigen ist. Hier werden  $1,2 \text{ \AA}$  bis  $1,3 \text{ \AA}$  für starke und  $2,0 \text{ \AA}$  für normale Bindungskräfte angenommen.

Innerhalb gewisser Grenzen, die durch die geometrischen Kriterien von Baker und Hubbard beschrieben und in Abbildung II-1 dargestellt werden, sind Wasserstoffbrückenbindungen flexibel aber funktional.



**Abbildung II-1:** Eine Wasserstoffbrückenbindung und die Kriterien von Baker und Hubbard.

A	Akzeptoratom	
AA	das kovalent-vorhergehende Atom eines Akzeptoratoms ( <i>antecedent atom</i> )	
D	Donatoratom	
H	Proton	
d	Abstand zwischen Donator und Akzeptor	$\leq 3,9 \text{ \AA}$
r	Abstand zwischen Akzeptor und Proton	$\leq 2,5 \text{ \AA}$
$\theta$	Winkel der D-H...A-Verbindungen	$\geq 90^\circ$
$\phi$	Winkel der D-A-AA-Verbindungen	$\geq 90^\circ$
$\gamma$	Winkel der H...A-AA-Verbindungen	$\geq 90^\circ$



Die Abbildung vermittelt den deutlich vektoriellen und gerichteten Charakter dieser Bindungen, der mittels Abstands- und Winkelangaben gut beschrieben und somit mathematisch leicht getestet werden kann. Dieses bildet die Grundlage für die verwendete und später erläuterten Methoden der transformations-invarianten Parametrisierung und der Selektion.

Die Bindungsenergien von Wasserstoffbrückenbindungen liegen in einer Größenordnung von 3 kcal/mol bis 7 kcal/mol. Genauso wie ihre, von den Elektronegativitäten der beteiligten Atome abhängigen Bindungslängen, sind die Energien zwischen denen kovalenter (C-C: 1,54 Å, 83 kcal/mol) und einzelnen van-der-Waals-Bindungen (3 Å bis 4 Å, 1 kcal/mol) einordnet, wie die folgende Auflistung zeigt [Stryer, 1975].

Wasserstoffbrückenbindung	Bindungslänge
O—H ..... O	2,70 Å
O—H ..... O <sup>-</sup>	2,63 Å
O—H ..... N	2,88 Å
N—H ..... O	3,04 Å
N <sup>+</sup> —H ..... O	2,93 Å
N—H ..... N	3,10 Å

Wasserstoffbrückenbindungen sind reversibel. Schwächere Bindungen können jederzeit gelöst werden, wenn sich z.B. durch Konformationsänderungen ein anderer Bindungspartner mit besser geeigneten Eigenschaften anbietet.

### **Protonenpositionen:**

Wasserstoffpositionen können mit der Methode der Röntgenkristallographie als häufigstes Instrument der Strukturaufklärung, nicht bestimmt werden. Für die Protonposition oder -positionen eines Donatoratoms gibt es jedoch einfache Modelle, die auf den geometrischen Eigenschaften des Atoms und der kovalent angebondenen Atome seiner näheren Umgebung beruhen. Diese Modelle können anhand der Kernresonanzspektroskopie, welche Wasserstoffpositionen bestimmen kann, bestätigt werden. Da die Bindungslänge vom Donator zum Proton im Regelfall mit 1,00 Å und die van-der-Waals-Radien von Stickstoff- und Sauerstoffatomen im Regelfall mit 1,50 Å bis 1,70 Å angenommen werden, bedeutet die Existenz eines gebundenen Protons keine berücksichtigenswerte Änderung der räumlichen Ausdehnung eines Donators, sondern zumeist sogar eine Penetration der van-der-Waals-Radien von Donator und Akzeptor aufgrund der anziehenden Wirkung [Xu *et al.*, 1997].

Dementsprechend können die Protonpositionen für die Donatoren gegebener Atomensembles, wie sie die Aminosäuren darstellen, bestimmt werden, ohne dass die räumliche Umgebung innerhalb einer Proteinstruktur zu beachten ist. Im Anhang A3 befindet sich die Berechnungsmethode der Positionen aus den jeweils vorliegenden Geometrie-Daten und eine Auflistung der Positionierungsvorgaben für die klassifizierten Donatortypen. Unterschieden werden "Donatoren im Ringsystem und Stickstoffe der Peptidbindung", "Donatoren direkt außerhalb eines Ringsystems", "zwei- und drei-protonige Aminogruppen" sowie "primäre Hydroxyl- und Sulfhydrylgruppen". Die Ausarbeitung entstammt allgemein verwendeter Praxis, wie sie auch in dem weitverbreiteten Programm HBPLUS [McDonald, Thornton, 1993, 1994] ausgeführt und zur Bestimmung von Wasserstoffbrückenbindungen in Proteinstrukturen angewendet wird.

#### **Akzeptorgeometrien:**

Einem Akzeptor ist es häufig möglich, mehrere Protonen mit seinen freien Elektronenpaaren zugleich zu binden. Beide Bindungen haben dabei den geometrischen Zwängen zu gehorchen. Die stärksten physiologischen Akzeptoren sind ein einfach negativ geladenes Sauerstoffatom und das Sauerstoffatom einer Carbonylgruppe. Bei der symmetrischen Anordnung der Bindungen und freien Elektronenpaare wird angenommen, dass die stärksten Bindungsfähigkeiten entsprechend der Tetraeder- bzw. planaren Dreiecksform anzufinden sind. Statistische Analysen strukturgelöster Proteinkontaktflächen bestätigen durchschnittliche H...A-AA-Bindungswinkel von etwa 130° [Xu *et al.*, 1997].

#### **Brückenbildende Wassermoleküle:**

Wassermoleküle (H<sub>2</sub>O) besitzen die Möglichkeit, sowohl als doppelte Protonen-Donatoren wie auch als doppelte Protonen-Akzeptoren sowie in entsprechend gekreuzter Funktionsweise aufzutreten. Diese Eigenschaften werden bei der Ausbildung der Eisstruktur besonders deutlich. Damit können Wassermoleküle in einer brückenbildenden Funktion tätig werden, indem sie in beliebiger Weise zwischen zwei Bindungsstellen eingespannt werden, die ansonsten entweder ein wenig zu weit voneinander entfernt oder auch zudem für eine direkte Interaktion inkompatibel sind. Es können ganze, hochgeordnete Netzwerke von Wassermolekülbrücken in Proteinstrukturen ausgebildet werden. Die Positionen einzelner Wassermoleküle können derart gut von umgebenden oder benachbarten Proteinfunktionen koordiniert sein, dass ihre Position in hochauflösenden Röntgenstrukturanalysen genau festgelegt werden kann. Der Beitrag dieser brückenbildenden Wirkung für die Vermittlung [Creighton, 1993] [Bhat *et al.*, 1994] und Stabilisierung [Bhat *et al.*, 1994] [Helms & Wade, 1995] ist bekannt.

**Intermolekulare Wasserstoffbrückenbindung:**

Grundlegend für diese Arbeit ist die vielfach gestützte *a priori* Annahme [Janin & Chothia, 1990] [Stickle *et al.*, 1992] [McDonald, Thornton, 1993, 1994], dass an produktiven Proteininteraktionen zumeist zwei oder mehr qualifizierte, intermolekulare Wasserstoffbrückenbindungen mit stabilisierender und spezifizierender Funktion beteiligt sind. Diese Annahme sowie die Realitätstreue obiger Kriterien für Wasserstoffbrückenbindungen werden in umfangreichen Analysen nichtredundanter, strukturgelöster Proteinkontaktflächen bestätigt [Janin *et al.*, 1988 und 1990] [Xu *et al.*, 1997]. Einige der getroffenen Aussagen dieser statistischen Untersuchungen sollen hier stichwortartig zusammengefasst werden, da sie z.T. selber beobachtet werden konnten, so dass im späteren Verlauf der Ausführungen hierauf verwiesen wird.

- durchschnittlich 10,7 Wasserstoffbrückenbindungen pro Kontaktfläche, mit einer Standardabweichung von 12,4, was auf eine enorme Varianz der Kontaktflächenbeschaffenheit hinweist
- starke Korrelation der Bindungsanzahl mit einem Korrelationskoeffizient von 0,89 zur Größe der Kontaktfläche die zu einer linearen Abhängigkeit führt:  
Bindungsanzahl =  $5,34 \cdot asa \cdot 10e-3 / \text{Å}^2$   
(*asa* = *accessible surface area*, hier als *total buried access surface*, d.h. die vom Lösungsmittel durch den Komplexpartner völlig abgeschirmte Oberfläche)
- Dichte intermolekularer Wasserstoffbrückenbindungen abnehmend mit Zunahme der Kontaktfläche
- etwa 50 von 319 Kontaktflächen mit zwei und weniger intermolekularen Wasserstoffbrückenbindungen und etwa 17,4% ungesättigte Bindungsstellen inmitten der Kontaktfläche
- etwa ein brückenbildendes Wassermolekül pro 3,8 Wasserstoffbrückenbindungen (mit großer Varianz) zwischen inkompatiblen Bindungsstellen
- durchschnittliche Bindungsgeometrien ( $2,9 \pm 0,2$  Å D-A- und  $2,0 \pm 0,2$  Å H-A-Abstand und  $150,8 \pm 17,1$ ° DH-A-Winkel), durchschnittliche H···A-AA-Bindungswinkel von etwa 132° und somit größtenteils normale bis schwache Bindungsstärken [Jeffrey & Saenger, 1991]
- intermolekulare Wasserstoffbrückenbindungen insgesamt schwächer und variabler als intramolekulare Wasserstoffbrückenbindungen, da nur eingeschränkt kleine Umlagerungen an der Oberfläche geschehen können und somit ein globales Energieminimum nicht zu erreichen ist - daher aber auch reversibele Bindungen

- Haupt-Hauptketten-Bindungen (22,6% intermolekularer Bindungen zwischen  $\beta$ -Strängen, 64,8% der intramolekularen Bindungen in  $\alpha$ -Helices und  $\beta$ -Faltblättern) tendenziell besser als Haupt-Seiten- oder Seiten-Seitenketten-Bindungen

**Verwendung intermolekularer WBB:**

Intermolekulare Wasserstoffbrückenbindungen besitzen eine große Universalität innerhalb von Proteinstrukturen und -komplexen. Die günstige Anzahl dieser lokal bindenden, spezifischen Wechselwirkungen und ihre sehr gut abstrahierbare, mathematisch erfassbare und selektionierbare Form sind der Ausgangspunkt für ihre Verwendung in dem hier vorgestellten Projekt, der Entwicklung eines Vorfilters zur Identifikation potentiell produktiver Molekülkomplexkonformationen. Bereits angesprochen und dokumentiert wurde ihre Verwendung in der qualitativen Bewertung von Komplexkonformationen sowie in vergleichbaren Aufgaben wie der Sekundärstrukturbestimmung und Faltungsvorhersage.

## 4 Methoden der transformations-invarianten Parametrisierung und des *Geometric Hashing*

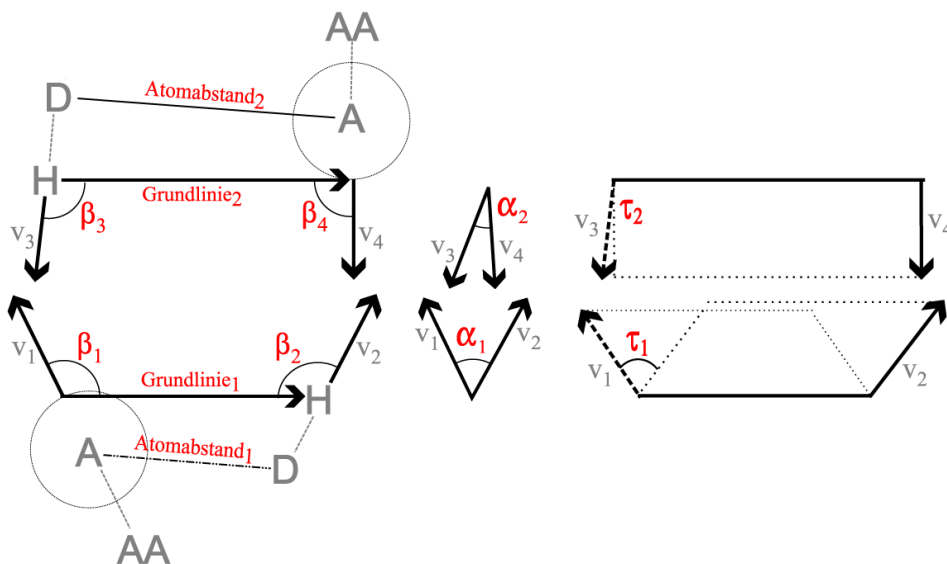
Die einfache Lokalisierung potentieller Wasserstoffbrückenbindungs-Partner auf der Oberfläche von Molekülen erbringt kaum relevante Information zur angestrebten Identifikation der Komplexkonformation, da sich generell jeder Donator mit jedem Akzeptor paaren ließe, es sei denn, die direkt umgebende Oberflächentopographie verhindere eine Anordnung, die den Kriterien von Baker und Hubbard genügt. Dementsprechend ineffektiv wäre die Methode, jede mögliche intermolekulare Akzeptor-Donator-Paarung zu evaluieren, indem sämtliche Atome eines Moleküls der notwendigen Transformation unterzogen würden, um abschließend eine Bewertung der Anordnung durchzuführen.

Transformations-invariante Eigenschaften sind unabhängig von jeglicher Lageänderung durch Rotation oder Translation im dreidimensionalen Raum. Zu diesen Eigenschaften zählen bezüglich der betrachteten Anwendung Bindungseigenschaften, Punktabstände von Koordinatenpunkten eines rigiden Körpers und Winkel zwischen Vektoren eines rigiden Vektorkonstrukts.

Ein Paar oder Triplet von Wasserstoffbrückenbindungs-Stellen auf Protein "A" besitzt derartige, monomolekulare transformations-invariante Eigenschaften. Diese müssen eine qualitative Entsprechung ihrer relativen, räumlichen Position und ihrer gerichteten Bindungscharakteristika in einem Paar oder Triplet auf Protein "B" finden, damit zeitgleich zwei bzw. drei kriterienkonforme produktive Wasserstoffbrückenbindungen ausgebildet werden können. Die qualitative Entsprechung ist hierbei gekennzeichnet durch bimolekulare transformations-invariante Eigenschaften, welche nun durch die Kombination der Eigenschaften der Paare bzw. Triplets zustande kommen. Es handelt sich um Komplementarität der Bindungseigenschaften und Differenzen oder Summen von Abständen oder Winkeln. Werden die Kriterien nicht erfüllt, so können keine ausreichend starken Wasserstoffbrückenbindungen an diesen beiden Stellen zugleich ausgebildet werden. Die korrespondierende Molekülanordnung erscheint damit mangels Bindungsspezifität als nicht erfolgversprechend. Scheint eine gleichzeitige Ausbildung der bindenden Kräfte möglich, muss die korrespondierende Molekülanordnung an die nachfolgende Oberflächenkorrelationsanalyse übergeben werden.

Abbildung II-2 zeigt das Schema derartiger geometrischer transformations-invarianter Eigenschaften, wie sie für zwei Paare von Wasserstoffbrückenbindungs-Stellen aufgestellt werden können. Die Abbildung soll an dieser Stelle jedoch lediglich veranschaulichen, welcher Art transformations-invariante Eigenschaften sind. Eine ausführliche Beschreibung der verwendeten mono- und bimolekularen Eigenschaften, ihrer geometrischen Grundlagen und der Grenzwerte, die zur Selektion ihrer Größen verwendet werden, erfolgt zu einem späteren Zeitpunkt. Es ist ersichtlich, dass nur geeignete Zustände für die einzelnen Parameter bzw. für die Kombination korrespondierender Parameter zu einem derartigen Arrangement führen können. Das Ideal, bei dem sich alle Atome sehr nahe kommen können und ihre Bindungsfähigkeiten sehr gut zueinander ausgerichtet sind, ist durch parallele Grundlinien mit geringer Längendifferenz und antiparallele Repräsentationsvektoren beschrieben. Daraufhin muss die eigentliche Selektion erfolgen. Grundsätzlich ist zu beachten, dass selbstverständlich nur ein Donator mit einem Akzeptor bindend interagieren kann, wobei bestimmte Atome beide Rollen einnehmen können.

Die Berechnungen der einzelnen Größen sind triviale Geometrie und Vektorrechnung, die Selektionsschritte sind einfache Vergleiche mit oberen und unteren Grenzwerten. Für Triplet-Paarungen werden alle drei Dreiecksseiten einer Prüfung als Paar-Paarung unterzogen.



**Abbildung II-2:** Eine schematische Übersicht transformations-invarianter Eigenschaften. Weitere Erläuterungen im Text.

Die transformations-invarianten Parametrisierung [Nussinov & Wolfson, 1991] wurde im Protein-Docking bereits zuvor verwendet bei der Suche nach zwischenmolekularer Komplementarität von vier [Connolly, 1986] und zwei oder drei auffälligen Oberflächenpunkten [Lin *et al.*, 1994] [Fischer *et al.*, 1995] [Norel *et al.*, 1995].

Bei den Evaluierungen aller möglichen, zwischenmolekularen Kombinationen von Paaren oder Triplets im eigentlichen Dockingschritt werden vereinfachte Prinzipien des *Geometric Hashing* verwendet. Diese entstammen zusammen mit der Methode transformations-invarianter Parametrisierung der Form-Charakterisierung und Objekt-Erkennung in der Robotersteuerung (*Computer Vision*) [Schwartz & Sharir, 1987] [Lamdan *et al.*, 1988 und 1990]. Die Gruppe von Nussinov und Wolfson [Nussinov & Wolfson, 1991] [Fischer *et al.*, 1992 und 1995] verwendete diesen Algorithmus für die Suche nach sequenzunabhängigen dreidimensionalen Strukturmotiven in Proteinen, wobei die Lage der C $\alpha$ -Atome der Aminosäuren als Koordinatenpunkte verwendet wurden. Eine spezialisiertere Beschreibung molekularer Oberflächen (*sparse critical points*) [Lin *et al.*, 1994] [Norel *et al.*, 1995] führte zu einem erfolgreichen Docking-Algorithmus [Fischer *et al.*, 1995] [Lin *et al.*, 1996].

Der *Geometric Hashing*-Algorithmus basiert im Original auf dem Vergleich der Anordnung von indizierten Punkten im dreidimensionalen Raum. und unterteilt sich in zwei Arbeitsgänge. Im ersten Schritt (*preprocessing*) wird für eine Struktur "A" für jede Gruppe (*reference set*) von drei Koordinatenpunkten (*interest points* also Atompositionen oder spezielle Oberflächenpunkte) ein relatives Koordinatensystem aufgespannt, in das die Koordinaten aller übrigen Punkte der Struktur überführt werden. Diese transformierte Strukturbeschreibung wird als Eigenschaft (*signature*) dieser Punktgruppe in einer Speicherstruktur abgelegt. Die Speicherstruktur wird dabei über die relativen Koordinaten der Punktgruppe indiziert (*hash table*). Im zweiten Schritt (*recognition*) wird anhand der relativen Koordinaten einer jeden Gruppe von Koordinatenpunkten der Vergleichsstruktur "B", die Speicherstruktur abgefragt. Finden sich an dem adressierten Suchpunkt ein oder mehrere Einträge der Struktur "A", so können die jeweiligen Signaturen der beiden Strukturen verglichen werden. Hierbei lässt sich bewerten, wie viele simultane Ähnlichkeiten es an anderen Stellen des Koordinatensystems noch gibt.

Die Eigenschaften, die sich für eine Gruppe von Koordinatenpunkten ergibt und in der Speicherstruktur abgelegt wird, muss nicht durch die kartesischen Koordinaten der übrigen Punkte beschrieben werden. Generell kann jedes algorithmische Konzept zur Lösung des

Problems eingesetzt werden, das nur irgendwie die in dieser Untersuchung bedeutende Eigenschaft der Punktgruppe beschreibt.

Ebenso kann das Prinzip variiert werden, indem die Adressierung der Speicherstruktur nicht über die Koordinaten selber, sondern über eine oder mehrere der transformations-invarianten Eigenschaften erfolgt. Hierdurch kann man einerseits eine einfach sortierte Tabelle erstellen und gezielt in den Bereichen abfragen, die der einzelnen Eigenschaft der Suchanfrage entsprechen, andererseits lassen sich hochdimensionale Tabellen erstellen, bei denen mit der ersten Adressierung direkt Kombinationen von Eigenschaften evaluiert werden.

Die hier implementierte Umsetzung orientiert sich an der sehr vereinfachten Variante. Es werden die monomolekularen Eigenschaften der Paare oder Triplets eines Dockingpartners einmalig ermittelt und selektiert, bevor die sie nach dem Atomabstand geordnet für die Programmlaufzeit in eine indizierte Tabelle im Arbeitsspeicher des Rechners abgelegt werden. Somit kann anhand des Indizierungsschlüssels direkt auf die Paare bzw. Triplets zugegriffen werden, die für ein Vergleichsensemble mit seinen spezifischen Eigenschaften überhaupt als Interaktionspartner in Frage kommen.



## 5 Räumliche Transformationen und Clusteranalysen von Transformationsangaben

### 5.1 Eindeutige Transformationsbeschreibung im dreidimensionalen Raum

Für die native Anordnung der Bindungspartner im produktiven Molekülkomplex werden multiple Interaktionen zugrunde gelegt, die unabhängig voneinander in den Evaluierungen der jeweiligen Paar- oder Triplett-Paarungen zu identifizieren sind. Daher ergibt sich die Schlussfolgerung, dass die notwendige Transformation eines Dockingpartners für jedes dieser richtig positiven Einzelergebnisse identisch zu erfolgen hat. Gleichartige Transformationen können somit gefunden und in Gruppen zusammengefasst werden, wenn sie eindeutig für den dreidimensionalen Raum beschrieben und somit auch verglichen werden können.

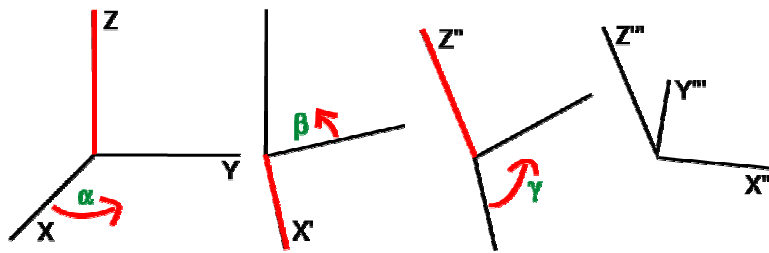
Da die Aufgabe nicht darin besteht, zwei Punkte deckungsgleich anzuordnen, sondern Position und Orientierung eines Vektorsystems festzulegen, bei dem zwei gekoppelte Vektoren komplementär zu einem korrespondierenden Paar gekoppelter Vektoren anzuordnen sind, erfordert die Transformation in festgelegter Abfolge Rotation und Translation. Für diese Arbeit wird vereinbart, dass für eine Transformationsbeschreibung zuerst die Rotation und dann die Translation erfolgt.

Eine Rotationsbeschreibung zur Positionierung eines Vektors mit vorgegebener Endausrichtung ist eindeutig, wenn drei Rotationsachsen zusammen mit der Reihenfolge und dem Drehsinn der Einzelrotationen festgelegt werden [Bronstein, 2000]. Generell lässt sich zwar jegliche Position und Orientierung auch mit nur zwei Rotationen beschreiben, hierbei ist jedoch die erste Rotationsachse - außer bei identischen Transformationen - stets verschieden und muss der Beschreibung als dritte Angabe zugerechnet werden. Der notwendige Rotationswinkelraum, um alle Positionen und Orientierungen realisieren zu können umfasst  $360^\circ * 180^\circ * 360^\circ$ . Die Rotationsbeschreibung zur Positionierung eines einfachen Punktes benötigt stets eine Rotationsangabe weniger.

Die Translationsbeschreibung erfolgt trivial für einen kartesischen Koordinatenraum durch die Aufteilung in Einzeltranslationen entlang der drei global festgelegten Hauptachsen, wobei ihre Reihenfolge bei der Bestimmung oder Ausführung ohne Einfluss auf den Endzustand ist.

## 5.2 Eulerwinkel mit X-Konvention

Bei den drei Rotationsachsen kann es sich um die drei Hauptachsen eines globalen, kartesischen Koordinatensystems handeln, es können jedoch auch Achsen veränderlicher Lage verwendet werden, solange die erste Rotationsachse global und stets identisch festgelegt ist und sich die nachfolgenden Rotationsachsen in Abhängigkeit von der ersten und zweiten Rotation ergeben, zueinander jedoch stets in einer festen Orientierung stehen. Derartiges verwendet die Rotationsbeschreibung über die Eulerwinkel mit X-Konvention. Sie verwendet für die erste Rotation mit dem Winkel  $\alpha$  die globale Z-Achse des Koordinatensystems, wobei zugleich das orthogonale Achsensystem um denselben Winkel rotiert wird. Die zweite Rotation mit dem Winkel  $\beta$  erfolgt um die neue X'-Achse und die letzte Rotation mit dem Winkel  $\gamma$  um die nun zweifach rotierte Z''-Achse, wie Abbildung II-3 illustriert.



**Abbildung II-3:** Eulerwinkel mit X-Konvention - die sukzessiven Einzelrotationen erfolgen um die Z-, X'- und Z''-Achsen. Weitere Erläuterungen im Text.

Eine beachtenswerte Eigenschaft der Eulerwinkel mit X-Konvention ist es, dass eine  $0^\circ, 0^\circ, 0^\circ$ -Rotation nicht eindeutig beschrieben werden kann. Dieses gilt sogar generell, wenn die X'-Rotation  $\beta=0^\circ$  bzw. überhaupt besonders klein ist. Hierbei werden die Rotationsachsen der ersten und letzten Rotation um Z ( $\alpha$ ) und Z'' ( $\gamma$ ) nur sehr wenig voneinander abgeändert, weshalb  $\alpha$  und  $\gamma$  in jeglicher Kombination nur die gewünschte Summe der Z-Rotationen ergeben müssen. Ist also  $\beta=0^\circ$  und die Summe von  $\alpha$  und  $\gamma$   $360^\circ$ , so wird hierdurch das Objekt nur einmal um die Z-Achse rotiert und das Ganze beschreibt *de facto* eine  $0^\circ, 0^\circ, 0^\circ$ -Rotation. Für die Transformationsbestimmung ist somit darauf zu achten, dass die Anfangspositionen der Objekte bezüglich der Rotation nicht der nativen Endposition entsprechen dürfen, da eine eindeutig gleichartige Beschreibung der multiplen, richtig positiven Ergebnisse nicht gegeben ist.

### 5.3 Gruppierung und Clusterung gleichartiger Transformationen

Da Rotation und Translation getrennt erfolgen, ist es auch möglich, gleichartige Transformationen zu identifizieren indem zunächst Ergebnisse gleichartiger Rotationen gruppiert werden und anschließend innerhalb dieser Untergruppen weiter nach gleichartigen Transformationen gesucht wird. Hierbei könne jeweils verschiedene Methoden zur Gruppierung verwendet werden.

#### **Gruppierung mit regelmäßigen Repräsentanten:**

Ein Wertebereich kann, wenn seine Abmessungen bekannt sind, in regelmäßige Abschnitte jeweils gleicher Größe unterteilt werden. Jeder Bereich erhält dann einen repräsentativen Wert, der seinen Mittelpunkt markiert und dem alle Wertobservationen seines Einzugsbereiches zugeordnet werden. Problematisch an dieser Einsortierungsmethode kann die Existenz fixierter Bereichsgrenzen sein, wenn diese eine natürlich zusammengehörige Gruppe gleichartiger Wertobservationen zertrennt und benachbarten Repräsentanten zuordnet. Ein Wertpunkt muss mit den Repräsentanten verglichen und seinem Ähnlichsten zugerechnet werden. Die Methode ist sinnvoll, wenn im gesamten Wertebereich Ergebnisse zu erwarten sind.

#### **Hierarchische Cluster-Analysen:**

Unter Clustern versteht man eher natürliche Ansammlungen und Zusammenstellung von gleichartigen Wertobservationen. Diese zeichnen sich durch eine erhöhte Auftrittsdichte aus und grenzen sich harmonisch von Ansammlungen andersartiger Wertobservationen ab. Derartige Cluster besitzen keine vordefinierten Bestimmungsgrenzen im Werteraum und erhalten ihren repräsentativen Wert letztlich unter der Berücksichtigung aller zugehörigen Observationen.

Generell entsprechen zu Beginn hierarchischer Clusteranalysen alle Wertobservationen je einem einzelnen Cluster, der einen berechenbaren Abstand zu seinen Nachbarn hat. Sukzessive werden nun die jeweils am engsten benachbarten Cluster vereinigt, solange ihre Abstände unterhalb einem festgelegten Abstandskriteriums liegen. Liegt kein ausreichend geringer Abstand zwischen zwei Clustern mehr vor, so findet keine Vereinigung mehr statt und alle Cluster sind ausgebildet. Gibt es kein maximales Abstandskriterium oder ist es zu groß gewählt für die jeweilige Datenmenge, so vereinigen sich zuletzt alle Observationen in einer einzigen Gruppe. Die Clustergröße kann auch begrenzt werden durch einen Maximalabstand der beinhalteten Observationen zum derzeitigen Clusterrepräsentanten, damit die Cluster nicht zu undefiniert groß werden und eine inakzeptable Varianz aufweisen.

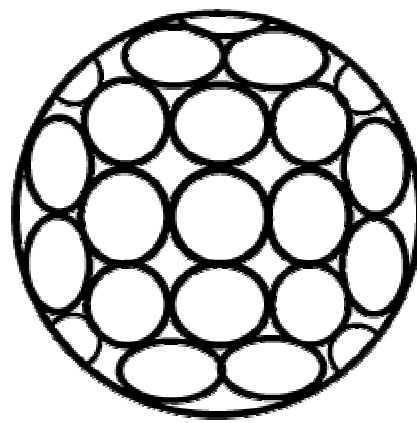
Verschiedene Algorithmen unterscheiden sich in der Berechnung der Abstände zwischen den Clustern: *Average Linkage*-Clusteranalysen verwenden den Abstand zum momentanen Clusterrepräsentanten, *Complete Linkage*-Clusteranalysen den Abstand zur am weitesten entfernten Observationen und *Single Linkage*-Clusteranalysen den Abstand zum nächsten, benachbarten Wertpunkt.

#### 5.4 Regelmäßige Unterteilung des Rotationsraums

Verwendet man die drei Hauptkoordinatenachsen als Rotationsachsen und unterteilt die jeweiligen Rotationsräume in regelmäßige Abschnitte wie es bei Längen- und Breitengradeinteilungen eine Kugel der Fall ist, so entsteht eine unregelmäßige Aufteilung des gesamten Raumes. Einzelnen Rotationsrepräsentanten sind ungleichgroße Winkelräume zugeordnet. Abbildung II-4a verdeutlicht dieses anhand eines derartigen Längen- und Breitengradnetzes für zwei Rotationswinkel. Eaton Lattman hat eine Methode vorgestellt [Lattman, 1972], welche regelmäßig verteilte Rotationsrepräsentanten erzeugt, denen ein stets identischer Winkelraum zugeschrieben wird. Abbildung II-4b versucht, dieses grob schematisch für zwei Rotationswinkel darzustellen, wobei selbstverständlich keine nichtzugeordneten Zwischenräume verbleiben dürfen sondern überlappende Bereiche entstehen. Für eine Darstellung aller drei Rotationswinkel ist das Kugelmodell nicht mehr ausreichend, weshalb auf einen Darstellungsversuch verzichtet wurde.



**Abbildung II-4a:**  
ungleich große Winkelräume



**Abbildung II-4b:**  
gleich große Winkelräume

Verwendet werden die Eulerwinkel  $\theta_1$ ,  $\theta_2$ ,  $\theta_3$ , wobei  $\theta_2$  in fester, wählbarer Schrittweite inkrementiert wird und die beiden anderen Winkel regelmäßig dazu erzeugt werden. Für eine  $\theta_2$ -Inkrementierung in  $30^\circ$ -Schritten ergeben sich 564 Rotationsrepräsentanten, für jede Halbierung der Schrittweite eine Verachtfachung der Clusteranzahl (etwa 4.400 bei  $15^\circ$ , etwa 119.000 bei  $5^\circ$ ). Eine Implementierung [Meyer, 1996], bzw. Portierung aus der Programmiersprache C++ [Zimmermann, 1999] der Methode wurden auf die Programmiersprache C übertragen und verwendet. Sie enthält eine Umwandlung der Rotationswinkel zu Eulerwinkeln mit X-Konvention, welche somit zur Rotationsgruppierung und zur anschließenden Transformationsbeschreibung verwendet werden.



### **III Material und Methoden**

Dieses Kapitel befasst sich nach einer Auflistung verwendeter Hilfsmittel hauptsächlich mit den Methoden, die innerhalb dieser Arbeit Verwendung gefunden haben. Es handelt sich einerseits um konkrete Anwendungen grundlegender Algorithmen, die hier ihre spezifische Anpassung erhalten, andererseits um Lösungen, die aufgabenspezifisch neu entwickelt und gelöst wurden. Um die Zusammenhänge der vielfältigen Problemstellungen und -lösungen zu verdeutlichen, kommt es in aufeinanderfolgenden Abschnitten zu einer chronologischen Darstellung der Arbeitsschritte in der Aufbereitung einzelner Proteinmoleküle und in dem eigentlichen Dockingvorgang zweier Moleküle. Daran schließt sich eine kurze Erläuterung der Auswahl verwendeter Test-Datensätze an. Diese wurden herangezogen, um geeignete, realitätsgetreue Grenzwerte für die bis dahin angeführten Selektionsmöglichkeiten der Methoden abzuleiten, was im letzten Abschnitt erläutert wird. Konkrete Parametervariationen und endgültige Werte für die verschiedenen Größen werden allerdings erst im Ergebnisteil dieser Arbeit präsentiert, da ihre Zusammenstellung die Effizienz des Dockings wesentlich beeinflussen kann und somit einen Großteil der Arbeitsleistung ausmacht.

## **6 Hilfsmittel**

### **6.1 Programmiersprachen, Rechensysteme, Betriebssysteme und Compiler**

Alle selbstverfassten Programme wurden in der Programmiersprache C nach dem Ansi-Standard und ohne Verwendung zusätzlicher Bibliotheken verfasst [Kernighan & Ritchie, 1988]. Ein Quelltextfragment von M.L. Connolly liegt in der Programmiersprache FORTRAN vor. Es wurde zur Integration modifiziert, mit den MIPSpro Compilern übersetzt und eingebunden. Kleinere Hilfsprogramme zur Koordinatentransformation o.ä. wurden teilweise in der Interpretersprache PERL implementiert und sind systemunabhängig unter den Versionen 5.005 und 5.6.x lauffähig. Die verwendeten Rechner, Betriebssysteme und Compiler sind in Tabelle III-1 zusammengefasst worden.

Hersteller	Rechner	Prozessor	Betriebssystem	C-Compiler
Silicon Graphics, Inc.	Challenge S	R5000	IRIX 6.5.5f	MIPSpro Compilers v.7.2.1
Silicon Graphics, Inc.	Indigo 2 Impact	R4400	IRIX 6.5.5f	MIPSpro Compilers v.7.2.1
Silicon Graphics, Inc.	Octane	R10000	IRIX 6.5.5f	MIPSpro Compilers v.7.2.1
Silicon Graphics, Inc.	O2	R10000	IRIX 6.5.5f	MIPSpro Compilers v.7.2.1
Silicon Graphics, Inc.	Challenge XL	R12000	IRIX 6.5.8f	MIPSpro Compilers v.7.3.0
SUN Microsystems	Enterprise 10000	UltraSPARC	Solaris 5.5.1	WorkShop Compilers v.5.0
diverse Hersteller	diverse IBM-kompatible PCs		diverse Linux-Distributionen und Compilerversionen	

**Tabelle III-1:** Übersicht der verwendeten Computersysteme, Betriebssysteme und Compiler

## 6.2 Quelltexte und Computerprogramme

Zur Identifikation oberflächennaher Atome und zur Erlangung von Oberflächennormalenvektoren wurde ein FORTRAN-Quellcode aus der Arbeitsgruppe von Michael L. Connolly [Connolly, 1986] verwendet, bei dem es sich nicht um die ursprüngliche Variante des MS-Programmpakets handelt. Dieses Programm (MDS, *Molecular Dot Surface*), welches wiederum eine Umsetzung des *Rolling Ball*-Algorithmus von Lee und Richards [Lee & Richards, 1971] ist, wurde zur Einbindung in die C-Programme in der Ein- und Ausgabe modifiziert.

Größere Felder wurden nach einer eigenen Implementierung des QUICKSORT-Algorithmus sortiert [Press *et al.*, 1992]. Eine Implementierung [Meyer, 1996] bzw. C++-Portierung [Zimmermann, 1999] der Methode zur Erstellung regelmäßiger Rotations-Repräsentanten nach Lattman [Lattman, 1972] wurden weiter auf C portiert und eingebunden. Quelltexte von Joachim Reichelt und Uta Lessel wurden zur Einbindung von Datenstrukturen und Unterrountinen des Programms BRAGI verwendet.

### **BRAGI (Braunschweig GBF Interactive Protein Modeling), Version 5.0**

[Schomburg und Reichelt, 1988]

Von Schomburg und Reichelt an der GBF (Gesellschaft für Biotechnologische Forschung mbH, Braunschweig) entwickeltes Programm zur Darstellung, Dokumentation, Interpretation und Modellierung von Proteinstrukturen.



**HBPLUS**, Version 3.06

[McDonald *et al.*, 1993] [McDonald & Thornton, 1994]

Erzeugt polare Wasserstoffatome für gegebene Atome eines Proteins und ist grundlegend für die eigene Protonenpositionsbestimmung. Außerdem berechnet HBPLUS Wasserstoffbrückenbindungen entsprechend den Kriterien von Baker und Hubbard und dient damit der Festlegung richtig positiver Referenzergebnisse und der Überprüfung eigener Vorhersagen. Das Programm ist für den akademischen Gebrauch frei verfügbar.

**RASMOL**, Version 2.5 - 2.7.2.1 und 2.6-ucb

[Sayle, 1994 und 1995; Bernstein, 2001; Molinaro, 1996]

Ein Programm zur Darstellung, Dokumentation und Interpretation von Proteinmolekülen, welches auf PC und Workstations eingesetzt werden kann.

**6.3 Proteinstrukturdaten**

Die verwendeten Proteinstrukturdaten entstammen ausnahmslos der Protein-Datenbank, PDB, ursprünglich *Brookhaven National Laboratory* [Bernstein *et al.*, 1997], mittlerweile *Research Collaboratory for Structural Bioinformatics* [Berman *et al.*, 2000]. Einzelne Einträge dieser Datenbank können eindeutig über einen vierstelligen Zahlen- und Buchstabencode identifiziert werden. Beinhaltet ein Datensatz mehr als eine Proteinkette, so sind die einzelnen Moleküle in der Regel durch einen einstelligen Zahlen- oder Buchstabenidentifikator gekennzeichnet. Gegebenenfalls wurden einzelne Proteinketten in getrennten Dateien verwaltet und einer Positions-Transformation unterzogen, jedoch keinerlei Residueneinträge untersuchter Ketten anderweitig manipuliert.

## 7 Erstellung der Dockingrepräsentation einzelner Proteinmoleküle

Als notwendige Datengrundlage für ein einzelnes Proteinmolekül, das der behandelten Dockingsimulation unterzogen werden soll, bedarf es eines Strukturdatensatzes entsprechend dem zuvor erläuterten Informationsgehalt der Protein-Datenbank. Die hierin enthaltene Information wird aufgearbeitet, um eine geeignete, effiziente Molekülrepräsentation für die Methode zu erhalten. Geht man von einem festgesetzten Ensemble an Evaluierungsgrenzwerten aus, so muss diese Aufbereitung nur einmalig für jedes Einzelmolekül erfolgen. Sie ist somit nicht zeitkritisch im Gegensatz zum eigentlichen Dockingvorgang, welcher als Datengrundlage ausschließlich diese Molekülrepräsentation mit ihren hochspezialisierten Informationen nutzt. In der Aufbereitung können auch zeitintensive Arbeitsschritte vorgenommen werden, wenn die erbrachte Informationsqualität dieses rechtfertigt und die Abwägung des hierfür permanent benötigten Speicherplatzbedarfs einer Molekülrepräsentation gegen den Aufwand der Neuberechnung zu jedem Docking positiv ausfällt.

Die Vorbereitung der Moleküle wird in zwei Schritte unterteilt. Zunächst eine Qualifizierung und Charakterisierung der bindungsfähigen Atome und sodann eine Zusammenstellung aller Paar- bzw. Triplettkombinationen dieser Atome unter Vorselektion ihrer transformationsinvarianten Eigenschaften.

Große Bedeutung hat die kombinatorische Natur dieser Aufgabenstellung, die letzten Endes in der benötigten Laufzeit des eigentlichen Dockingexperiments ihre Wirkung zeigt. Daher gilt es, von vornherein möglichst wenige Atome und hieraus möglichst wenige repräsentierende Atumpaare und -tripletts zusammen zu stellen, um die Anzahl notwendiger Kombinationsevaluierungen möglichst gering zu halten. Hierbei darf nur auf Informationen zurückgegriffen werden, die dem Informationsgehalt des Einzelmoleküls zu entnehmen sind, da der Komplexpartner und seine räumliche Position im Rahmen der realitätsgetreuen Aufgabenstellung unbekannt sind. Die Prozesse zur Bewertung und Entscheidung sind dabei mit Vorsicht zu behandeln, da der fehlerhafte Ausschluss eines essenziell wichtigen Atoms oder seiner essenziell wichtigen Paar- oder Triplettszusammenstellung mit ein oder zwei Nachbarn in der Kontaktfläche direkt dazu führt, dass die hiervon abhängigen, richtig positiven Ergebnisse nicht evaluiert werden und somit nicht gefunden werden können, was bei einem eventuellen Mangel derartig relevanter Ergebnisse fatale Folgen haben muss.

## 7.1 Qualifizierung und Charakterisierung bindungsfähiger Atome

### 7.1.1 Qualifizierung bindungsfähiger Atome

Um intermolekulare Wasserstoffbrückenbindung ausbilden zu können, müssen die beteiligten Atome geeignete Bindungseigenschaften aufweisen. Eine Position direkt an der Oberfläche des Proteinmoleküls oder zumindest sehr nahe darunter und eine gerichtete Bindungsfähigkeit zum Lösungsmittel bzw. zum Bindungspartner hin sind Grundvoraussetzungen.

- Die Bindungseigenschaften der verschiedenen Atomtypen bezüglich Wasserstoffbrückenbindungen wurden bereits zuvor in einem Abschnitt erläutert und können der Tabelle im Anhang A2 entnommen werden. Schwefelatome werden nicht als Donatoren qualifiziert, da Sulfhydrylgruppen zu bindungsschwach sind. Elektronenwolken aromatischer Ringsysteme werden nicht als Akzeptoren qualifiziert, da ihre Bindungseigenschaften ebenfalls zu schwach ausgeprägt und darüber hinaus zu wenig gerichtet sind. Atome ohne Akzeptor- bzw. Donatoreigenschaften bedürfen keiner weiteren Berücksichtigung.
- Als Moleküloberfläche wird die lösungsmittelzugängliche Oberfläche verwendet, wie sie über den *Rolling Ball*-Algorithmus [Lee & Richards, 1971] mittels einer eingebunden Implementierung von Connolly [Connolly, 1986] erzeugt wird. Der Sondenradius wird entsprechend dem Radius eines Wassermoleküls mit 1,4 Å angenommen und die diskreten Oberflächenpunkte werden mit einer Dichte von fünf Punkten pro Quadratangstrom erzeugt. Atome, die zu weit von der Moleküloberfläche entfernt liegen, werden als ungeeignet klassifiziert und nicht weiter im Rahmen der hier entwickelten Methode berücksichtigt.
- Um zu bestimmen, ob ein Atom eine geeignet gerichtete Bindungsfähigkeit besitzt, wird zunächst der Bindungsvektor bestimmt, wie im nachfolgenden Abschnitt ausgeführt wird. Der Vektor wird mit den von der Connolly-Unterroutine ebenfalls ausgegebenen Normalenvektoren der Oberflächenpunkte des Atoms bzw. aller Punkte innerhalb eines 3-Å-Abstands verglichen. Je geringer die Winkelabweichung des Interaktionsvektors zu einem der Oberflächennormalenvektoren, desto gradliniger und stabiler kann eine intermolekulare Wasserstoffbrückenbindung ausgebildet werden. Erweist sich die Richtcharakteristik als unzureichend, so wird das Atom als ungeeignet klassifiziert und nicht weiter im Rahmen der hier entwickelten Methode berücksichtigt.

Die verwendeten Identifikationskriterien für Oberflächennähe und geeignete Richtcharakteristik der Bindungsfähigkeit werden im Ergebnisteil empirisch aus realen Werten richtig positiver Untersuchungsergebnisse der nativen Protein-Protein-Komplexen hergeleitet.

### 7.1.2 Charakterisierung bindungsfähiger Atome

Es wurde bereits dargelegt, dass Wasserstoffbrückenbindungen umso stabiler sind, je besser eine kollineare Anordnung der Donator-Proton-Bindung und dem Akzeptoratom gegeben ist. Für die Bindung vom Proton zum Akzeptor und weiter zu dessen kovalent angebundenem Atom sind an sich Bindungswinkel entsprechend den Aufenthaltswahrscheinlichkeiten freier Elektronenpaare zu erwarten. Für  $sp^3$ - bzw.  $sp^2$ -hybridisierte Atome entspräche dieses Tetraeder- bzw. planaren Dreiecks-Winkeln. Im Rahmen dieser Arbeit wird jedoch zunächst darauf verzichtet, multiple Akzeptorrepräsentationen zu implementieren, um die vollständige Verdopplung jeder akzeptorbezogenen Rechnung zu umgehen. Es wird also eine Kollinearität aller vier Atome ( $D-H\cdots A-AA$ ) angenommen. Diese entspricht dann auch der Kollinearität zweier Vektoren, die durch die kovalenten Bindungen von  $D-H$  und  $A-AA$  gegeben sind. Daher liegt es nahe, für die bindungsfähigen Atome je einen Interaktionsvektor festzulegen, der fortan die Richtung der stärkstmöglichen Bindungsfähigkeit und -kraft wiedergibt. Dieser Vektor ist nicht abhängig von der Lage des zu dem Zeitpunkt unbekanntes Bindungspartners innerhalb der Wasserstoffbrückenbindung, sondern von Atomtyp und räumlicher Anordnung der nächsten, kovalent angebundenen Atome im eigenen Molekül.

- Für Donatoren entspricht dieser Richtungsvektor mit Einheitslänge der Richtung der Donator-Proton-Bindung, wobei die Protonposition je nach Donatortyp entsprechend den grundlegenden Ausführungen zu Wasserstoffbrückenbindungen und der Tabelle im Anhang A3 bestimmt ist. Besitzt ein Donator zeitgleich mehrere Protonen, oder müssen mehrere Positionen exemplarisch für einen Protonaufenthalt angenommen werden, so sind ihm auch entsprechend multiple Interaktionsvektoren zugeordnet.
- Für Akzeptoren entspricht der jeweils einzige Interaktionsvektor der Verlängerung der beschriebenen  $AA-A$ -Bindung bzw. bei Lokalisierung des Akzeptors in einem Ringsystem oder allgemein innerhalb einer Atomkette, einer Kombination der beiden kovalenten Anbindungen von  $AA_1$  und  $AA_2$  zum Akzeptor.
- Atome mit Bindungseigenschaften als Akzeptor und Donator erhalten alle entsprechenden Charakterisierungen.

Da es bei der Akzeptor-Donator-Interaktion in der Regel zu einer Berührung oder gar Penetration der Atom-van-der-Waals-Schalen kommt, werden den Atomen explizite

Interaktionspunkte zugeordnet, welche zugleich den Ortsvektor zum Ursprung der Bindungsvektoren darstellen. Sie entsprechen den Protonpositionen bzw. dem kollinearen Punkt auf der van-der-Waals-Schale des Atomzentrums entlang der Richtung des Interaktionsvektors. In Abbildung III-1 lässt sich das beschriebene, repräsentierende Vektormodell von Akzeptoren und Donatoren nachvollziehen.

### *7.1.3 Disqualifizierung bindungsfähiger Atome aufgrund intramolekularer*

#### *Wasserstoffbrückenbindungen*

Akzeptoren für Wasserstoffbrückenbindungen sind in unterschiedlichem Maße in der Lage, zeitgleich mehr als eine Bindung einzugehen. Insbesondere Akzeptoren wie einfach negative geladene Sauerstoffatome mit Radialsymmetrie und sehr starker Bindungsfähigkeit aufgrund der hohen Elektronegativität bilden häufig zwei Interaktionen aus, wenn die Lage beider Bindungspartner die Erfüllung der Kriterien von Baker und Hubbard zulassen [McDonald & Thornton, 1994]. Donatoren hingegen können zu jedem Zeitpunkt nur eine Wasserstoffbrückenbindung eingehen. Für alle Atome wäre eventuell vorstellbar, dass sie sich in andauerndem Wechselspiel mit mehr als einem Interaktionspartner befinden und alternierend Bindungen öffnen und schließen.

Atome, die entsprechend ihrer Geometriebewertungen an sich zur weiteren Berücksichtigung qualifiziert sind, könnten bei geeigneter Umgebung auch Bindungspartner innerhalb des eigenen Moleküls finden und intramolekulare Wasserstoffbrückenbindungen ausbilden. Damit würde sich ihre Eignung zum Aufbau intermolekularer Wasserstoffbrückenbindungen erheblich einschränken. Für Donatoren gilt dieses in höherem Maße als für Akzeptoren, da sie eine eventuell vorhandene Bindung auf jeden Fall zugunsten des intermolekularen Kontakts auflösen müssten, was jedoch bei dem Übergang von einer schwachen intramolekularen zu einer starken intermolekularen Interaktion plausibel ist.

Die mögliche Existenz intramolekularer Wasserstoffbrückenbindungen lässt sich für jedes Einzelmolekül mittels HBPLUS durch einen Test auf geeignete Nachbaratome postulieren. Hierdurch kann man stets eine Liste identifizierbarer Atome erstellen, die eventuell ganz oder teilweise in ihrer Bindungsfähigkeit abgesättigt sind und dem Docking mittels Wasserstoffbrückenbindungen gegebenenfalls nicht mehr zur Verfügung stehen. Atome mit besonders guten Eigenschaften für intermolekulare Bindungen sollten besonders ungeeignet für andersartige Kontakte sein und somit diesen Listen nicht angehören. Hier liegt der

Angriffspunkt dieses Bewertungsschrittes, der die Spezifität der Gruppe qualifizierter Atome erhöhen soll.

In der Implementierung dieser Überlegungen kann prinzipiell unter drei Varianten gewählt werden:

- Bei der **Verwendung aller Atome** wird vollständig auf diese Selektionsmöglichkeit verzichtet. Jedes Atom mit bindungsfähiger Interaktionsgeometrie wird im weiteren Verlauf berücksichtigt.
- Bei der **Verwendung aller ungesättigten Atome** werden nur Bindungsstellen weiterverwendet, die zumindest noch eine freie Bindungskapazität besitzen. Somit sind hier ausschließlich bindungsfreie Donatoren und maximal einfachgebundene Akzeptoren zugelassen.
- Bei der **Verwendung aller ungebundenen Atome** werden auch einfach gebundene Akzeptoren vom weiteren Verlauf ausgeschlossen, d.h. berücksichtigte Atome dürfen keinerlei intramolekularen Bindungspartner vorgeschlagen bekommen.

Problematisch an dem Vorgehen ist die Tatsache, dass HBPLUS lediglich auf Übereinstimmung mit den Bindungskriterien achtet und somit z.B. auch multiple Bindungen für einen Donator ausgeben kann und darüber hinaus keine weitere Bewertung der Bindungsqualitäten vornimmt. Somit kann es zu fehlerhaften Ausschlüssen essenzieller Atome kommen, ohne dass es gelungen wäre, derartige Fehlentscheidungen im direkten Vergleich ausmachen zu können. Daher werden alle drei Optionen im weiteren Verlauf parallel durchgeführt und bewertet.

#### *7.1.4 Zusammenfassung der Repräsentation bindungsfähiger Atome*

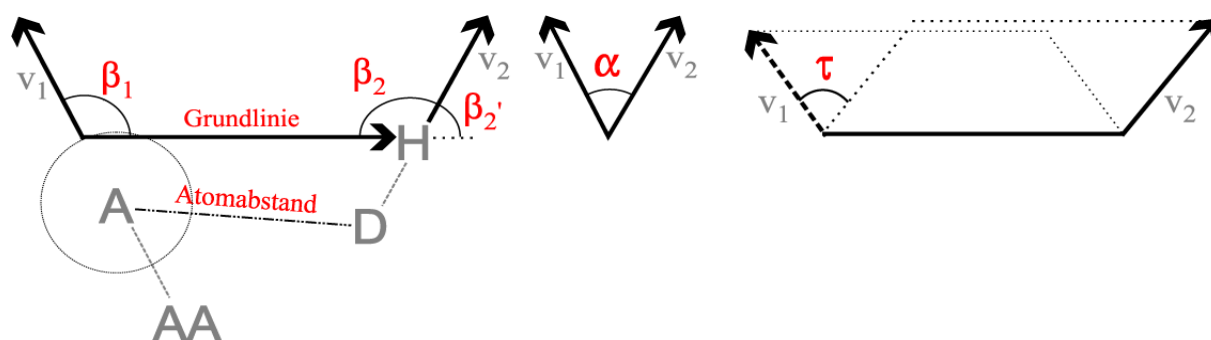
Atome, die als bindungsfähig für intermolekulare Wasserstoffbrückenbindungen eingestuft werden, liegen nah unter der lösungsmittelzugänglichen Oberfläche und weisen eine ausreichend gute Bindungsorientierung zur potentiellen Molekülkontaktfläche hin auf. Des Weiteren können Atome, die diesen Anforderungen genügen, optional von der Berücksichtigung im Docking ausgenommen werden, wenn die mögliche Existenz einer oder mehrerer intramolekularer Wasserstoffbrückenbindungen aufgezeigt werden kann. Eine weitere Option besteht, Akzeptoren nur dann auszuschließen, wenn zumindest zwei intramolekulare Bindungen postuliert werden.

Eine generelle Darstellung von Akzeptoren und Donatoren für Wasserstoffbrückenbindungen erfolgt jeweils durch den Bindungstyp (Akzeptor, Donator oder beides), durch die kartesischen Raumkoordinaten des Atomzentrums und durch einen Richtungsvektor, der entsprechend der

stärksten Bindungsfähigkeit ausgerichtet ist. Jeder Interaktionsvektor hat seinen Ursprung in einem weiteren Koordinatenpunkt, der das Atom charakterisiert. Es handelt sich um den Interaktionspunkt, d.h. um das Zentrum eines Protons für Donatoren und um den Schnittpunkt der AA—A-Bindungsachse mit der van-der-Waals-Schale für Akzeptoren. Multiple Protonenpositionen oder Bindungseigenschaften erfordern die Charakterisierung aller möglichen Erscheinungsarten des entsprechenden Atoms.

## 7.2 Monomolekulare transformations-invariante Eigenschaften eines Paares bindungsfähiger Atome

Bindungsfähige Atome werden mit je zwei Raumkoordinaten und einem Interaktionsvektor beschrieben. Mehrprotonige Donatoren sowie Atome mit Donator- und Akzeptorbindungseigenschaften bekommen entsprechend mehrfache Beschreibungen. Aus diesen Beschreibungen ergibt sich für eine Paarzusammenstellung zweier Atome innerhalb derselben Moleküloberfläche ein repräsentierendes Modell wie es das Vektorkonstrukt in Abbildung III-1 zusammen mit den notwendigen Bezeichnern wiedergibt.



**Abbildung III-1:** Das monomolekulare Vektorkonstrukt eines Paares bindungsfähiger Atome.

Gezeigt sind ein Akzeptoratom (A) mit einem einzelnen, kovalent angebondenen Atom (AA) sowie ein Donator (D) und sein einziges Proton (H). Die gerichtete Grundlinie verbindet die beiden Interaktionspunkte der Atome, welche jeweils Ursprung der entsprechenden Interaktionsvektoren ( $v_1$ ,  $v_2$ ) für die intermolekulare Bindung sind. Eingetragen sind die  $\beta$ -Innenwinkel (bzw.  $\beta_2'$  als Außenwinkel), der  $\alpha$ -Innenwinkel und der gerichtete  $\tau$ -Torsionswinkel, die im Text weiter beschrieben werden.

Im Einzelnen können hieraus folgende transformations-invariante geometrische Eigenschaften abgeleitet werden. Bedeutung und Eigenschaften einzelner Größen, welche richtungsweisend für die Festlegung der Kriteriengrenzwerte sind, sollen hier erläutert werden.

- **Atomabstand**  
euklidischer Abstand zwischen den beiden Atomzentren
- **Grundlinienvektor**  
Verbindungsvektor zwischen den Interaktionspunkten mit eindeutiger Richtung vom Atom niedrigerer zum Atom höherer Identifikationsnummer
- **Grundlinienlänge**  
euklidischer Abstand zwischen den Interaktionspunkten

Atomabstand und Grundlinienlänge sind eng miteinander gekoppelt. Der Atomabstand darf einen Mindestabstand nicht unterschreiten, der durch die Summe der van-der-Waals-Radien gegeben ist. Eine Atom-Penetration ist, wie von einem Teil des Lennard-Jones-Potentials beschrieben, energetisch derart unvorteilhaft, dass von einem Artefakt oder einem Fehler der Struktur bzw. der Datengrundlage ausgegangen werden muss. Genauso muss es eine natürliche Obergrenze geben, da es um die Zusammenstellungen von Atompaaaren innerhalb einzelner Molekülkontaktflächen geht und nicht um Paare, deren Atome in völlig unabhängigen Regionen der Oberfläche lokalisiert sind. Die angestrebte Art des Proteindockings muss versuchen, Ergebnisse aufgrund kleiner Kontaktflächen zu bekommen, da moleküldynamische Effekte wie ein *Induced Fit* hier zunächst außer Acht gelassen werden. Die Einschränkung der Abstände muss es ermöglichen, eine ausreichende Anzahl von Paaren an Wasserstoffbrückenbindungsstellen zu finden. Sie dürfen jedoch nicht allzu groß gewählt sein, da dann keine Kopplung der Bindungsgeometrien aufgrund räumlicher Nähe gegeben ist und die übrigen Parameter uneingeschränkt alle Werte annehmen können. Auch kann ein zu großer Abstand innerhalb dieser Vektorrepräsentation unvorhersehbare topografische Oberflächenmerkmale kaschieren, die im Komplex eine Interaktion durch Molekülpenetration verhindern würden.

- zwei **Innenwinkel  $\beta_1$  und  $\beta_2$**  (bzw. **Außenwinkel  $\beta_2'$** )  
jeweils zwischen einem Interaktions- und dem Grundlinienvektor  
Wertebereich:  $0^\circ \leq \beta \leq 180^\circ$ ,  $\beta_2 = 180^\circ - \beta_2'$

Die  $\beta$ -Winkel alleine erbringen keine Selektionsmöglichkeit, da alle Werte in unabhängiger Kombination angenommen und je nach Oberflächentopographie dennoch zwei ideal parallele Wasserstoffbrückenbindungen aufgebaut werden können. Geht man jedoch von der



Idealvorstellung planarer und paralleler Interaktionsflächen sowie äquidistanter Atompositionen aus, so gibt es eine Präferenz für den Wertebereich um  $90^\circ$ .

- **Differenz der  $\beta$ -Winkel**

torsionsunabhängig, daher ungleich dem Innenwinkel  $\alpha$

Wertebereich:  $0^\circ \leq |\beta_1 - \beta_2| \leq 180^\circ$  bzw.  $0^\circ \leq |\beta_1 - \beta_2| \leq (\beta_{\max} - \beta_{\min})$

Diese Eigenschaft beschreibt die Trapezform des Vektorkonstrukts, wenn man die drei Vektoren in eine gemeinsame Ebene projiziert. Ohne Berechnung des Betrags erhielte man eine Aussage über Auf- oder Zusprenzung des Trapezes und somit über die Parallelität der Interaktionsvektoren. Allerdings ist diese Aussage torsionsunabhängig, weshalb die angenommenen Werte auch kleiner oder gleich dem jeweiligen  $\alpha$ -Winkel sein müssen. Die Differenz kann maximal den Wert  $\beta$ -Maximum abzüglich  $\beta$ -Minimum annehmen.

- **Torsionswinkel  $\tau$  mit Torsionsrichtung**

zwischen den beiden Normalenvektoren der Ebenen, welche von den Interaktionsvektoren mit dem Grundlinienvektor aufgespannt werden; dieses entspricht dem Winkel zwischen den Kreuzvektoren: Grundlinienvektor  $\times$  Interaktionsvektor<sub>1 bzw. 2</sub>

Wertebereich:  $-180^\circ \leq \tau \leq 180^\circ$

Der  $\tau$ -Winkel beschreibt ausschließlich die Torsion, d.h. die Spreizung der Interaktionsvektoren in die Raumbreite. Er beachtet nicht die Trapezform wie es in Abbildung III-1 den Anschein haben mag. Er ist an sich stets positiv und kleiner  $180^\circ$ , wobei man die Torsionsrichtung dennoch bestimmen und für die Spezifikation verwenden kann. Diese wird später zur Beschreibung der bimolekularen Differenz der Torsionswinkel ausgeführt.  $\tau$ -Winkel von  $180^\circ$  sind möglich, beschreiben jedoch zwei Interaktionsvektoren, die gegenläufige Ausrichtung besitzen. Diese sollten somit nur schwerlich zwei intermolekulare Wasserstoffbrückenbindungen entsprechend der Idealvorstellung ausbilden können, wenn sie nicht z.B. auf zwei entgegengesetzten Seiten eines Hügels liegen, der in ein entsprechendes Tal der interagierenden Moleküloberfläche hineinragt.

- **Innenwinkel  $\alpha$**

zwischen den Interaktionsvektoren

torsionsabhängig, daher ungleich  $\beta$ ; trapezformabhängig, daher ungleich  $\tau$

Wertebereich:  $0^\circ \leq \alpha \leq 180^\circ$

Der  $\alpha$ -Winkel trifft eine Aussage über das Maß der Parallelität der beiden Interaktionsvektoren eines Atompaars. Er kann jedoch ihre Lage bezüglich der Grundlinie nicht berücksichtigen und

kann keine Aussage über die Trapezform des Vektorkonstrukts treffen. Der Winkel ist stets positiv, kann keine geringeren Werte als der  $\tau$ -Winkel ( $\alpha=\tau$  für  $|\beta_1-\beta_2|=0^\circ$  bzw. entsprechend  $\beta_1+\beta_2=180^\circ$ ) annehmen und keine größeren als die Summe der  $\beta$ -Winkel ( $\alpha=\beta_1+\beta_2$  für  $\tau=0$ ).

### **7.3 Monomolekulare transformations-invariante Eigenschaften eines Triplets bindungsfähiger Atome zuzüglich Triplett-Vektordivergenz und Triplett- Bindungsseite**

Da ein Dreieck aus drei bindungsfähigen Atomen eines Moleküls zunächst einfach wie die Zusammenstellung der drei Dreiecksseiten, d.h. den Abschnitten zwischen je zwei dieser Atome betrachtet wird, werden alle transformations-invarianten Eigenschaften der Paare bindungsfähiger Atome unverändert zur obigen Beschreibung übernommen. Hierbei entsteht eine entsprechend höhere Selektionskraft, da die Kriterien für alle Kanten zeitgleich erfüllt werden müssen. Für Atomtripletts ergeben sich jedoch ein zusätzliches, selektionierbares Kriterium und eine zusätzliche, später verwendbare Eigenschaft.

Das optimale Atomdreieck, welches planar unter der Moleküloberfläche liegt, hat drei Interaktionsvektoren, die alle parallel zueinander und senkrecht zur Dreiecksebene stehen. Somit haben auch alle Vektoren einen  $0^\circ$ -Winkel bzw.  $180^\circ$ -Winkel zum Normalenvektor der Dreiecksebene, wenn dieser auf derselben bzw. der entgegengesetzten Ebenenseite aufgesetzt ist. (Siehe hierzu die ersten beiden Dreiecke in Abbildung III-2.)

Steht das Dreieck senkrecht zur Moleküloberfläche und behalten die Interaktionsvektoren weiter ihre senkrechte Ausrichtung zur Oberfläche bei, so haben diese alle einen identischen  $90^\circ$ -Winkel zur Dreiecksnormalen. Die Normale selber kommt dabei parallel zur Oberfläche zu liegen. (Siehe das dritte Dreieck in Abbildung III-2.)

Zeigen die Vektoren hingegen auf unterschiedliche Seiten der Dreiecksebene, so weisen sie eine zunehmende, maximal  $180^\circ$ -große Differenz dieser Winkel zur Normalen auf. Das soll mit dem vierten Dreieck aus Abbildung III-2 angedeutet werden, wobei zu beachten ist, dass die drei Interaktionsvektoren selber noch eine ausreichend gute Orientierung zur Moleküloberfläche haben.

Indem für alle Interaktionsvektoren die Winkel zur Dreiecksnormalen bestimmt und anhand der drei Winkeldifferenzen verglichen und ausreichend selektiert werden, kann man sicherstellen, dass die Vektoren eines Triplets alle in dieselbe Richtung weisen oder zumindest in ausreichend symmetrischer Form divergieren. Letztere Bewertung kann nicht getroffen werden, wenn direkt die drei Winkel zwischen den Interaktionsvektoren bewertet werden.

Winkel der lv's zum Nv:	alle 0°	alle 180°	alle 90°	0° bis 180°
Differenzen dieser Winkel:	alle 0°	alle 0°	alle 0°	0° bis 180°
Atomanordnung:	Uhrzeigersinn	Gegenuhrzeigersinn	unbestimmt	unbestimmt
Bindungsseite:	zum Nv	entgegen dem Nv	unbestimmt	unbestimmt

**Abbildung III-2:** Triplett-Vektordivergenz und Triplett-Bindungsseite

Die Moleküloberfläche sei in dieser Abbildung stets oberhalb der Triplets, parallel zu den ersten beiden Dreiecksflächen anzunehmen. Die Atome sind mit ansteigenden Identifikatoren nummeriert und besitzen kleine blaue Interaktionsvektoren (lv). Sie seien ausreichend nah an der Oberfläche positioniert mit einer ausreichend guten Orientierung der Interaktionsvektoren. Die großen roten Dreiecknormalenvektoren (Nv) ergeben sich jeweils aus dem Kreuzprodukt der Atomverbindungen  $(1 \rightarrow 2) \times (2 \rightarrow 3)$  und stehen daher einmal über und einmal unter der Dreiecksfläche. Sie weisen hier einmal zur Moleküloberfläche hin, einmal zum Molekülinneren hin und zweimal oberflächenparallel. Unter den Dreiecken finden sich Angaben zu den Winkeln zwischen Interaktions- und Normalenvektoren, zu den Differenzen dieser Winkel untereinander und zu dem Rückschluss, der aus einer eindeutigen Winkelverteilung auf die Atomanordnung bzw. die Dreiecksbindungsseite geschlossen werden kann.

Zudem ermöglicht der Vergleich der Winkeldifferenzen zur Dreiecksnormalen eine Festlegung des Drehsinns der Anordnung aller Atomidentifikationsnummern bei Aufsicht auf die Normale. Der Dreiecksnormalenvektor ist eindeutig bestimmt als Kreuzvektor zweier Atomverbindungsvektoren  $\text{Atom1} \rightarrow \text{Atom2}$  und  $\text{Atom2} \rightarrow \text{Atom3}$ , wobei Atom1 bis Atom3 streng aufsteigende Atomidentifikationsnummern besitzen. Die Beschreibung des Drehsinns lässt sich zur Beschreibung der Bindungsseite des Dreiecks verwenden. Diese kann postuliert werden, wenn alle Winkel der Interaktionsvektoren zur Normalen ausreichend nahe an 0° bzw. 180° liegen. Dann weisen alle Vektoren derart eindeutig zu einer der Dreiecksseiten, dass nur diese für die Interaktion mit einem anderen Triplett in Betracht gezogen werden muss. Die rückwärtige Anordnung würde nur abweisende Interaktionsvektoren darbieten. Die beiden Zustände werden einfach stellvertretend für "Uhrzeiger-" und "Gegenuhrzeigersinn" als "1" oder "2" gekennzeichnet und dem Atomtriplett als Eigenschaft zugeschrieben.

Sind alle Winkel zur Dreiecksnormalen ausreichend ähnlich, d.h. divergieren die Interaktionsvektoren nicht zu sehr, weisen jedoch nicht eindeutig genug zu einer Dreiecksseite,

so liegen sie eher flach zur Dreiecksebene. Diese steht dann voraussichtlich tendenziell senkrecht zur Moleküloberfläche und muss daher von beiden Seiten zur Interaktion evaluiert werden. Diese Eigenschaft eines Atomtripletts wird mit "0" gekennzeichnet.

Die vollständigen Ausführungen zu den möglichen Anordnungen einer Triplet-Triplett-Kombination finden sich bei der umfassenden Beschreibung des gesamten Dockingablaufs.

Atomtripletts besitzen somit im Vergleich zu Atompaaren zwei zusätzliche transformations-invariante Eigenschaften, die als Triplet-Vektordivergenz und Triplet-Bindungsseite bezeichnet werden sollen.

- **Triplet-Vektordivergenz:**

Anhand der Divergenz kann zusätzlich zu den bisherigen monomolekularen transformations-invarianten Eigenschaften selektiert werden, ob ein Atomensemble günstige, d.h. ausreichend gleichartige Bindungseigenschaften vorweist und in die Molekülrepräsentation aufgenommen wird.

- **Triplet-Bindungsseite:**

Die Festlegung der Bindungsseite ist – wenn sie festgelegt werden kann – eine Eigenschaft, die ihre Verwendung später im eigentlichen Dockingablauf unmittelbar vor dem Evaluierungsaufwurf der Kombination zweier Triplets finden wird.

Weitere transformations-invariante Eigenschaften, wie z.B. die Winkel zwischen den Grundlinienvektoren eines Atomdreiecks finden nach kurzfristiger Betrachtung vorerst keine Verwendung, da sie zu stark abhängig sind von den betrachteten Grundlinienlängen und zudem einer sehr großen Varianz unterliegen.

#### **7.4 Erstellung der Molekülrepräsentationen unter Selektion monomolekularer transformations-invarianter Eigenschaften**

Die vollständige Repräsentation eines Proteinmoleküls für das Docking mittels Paaren oder Triplets intermolekularer Wasserstoffbrückenbindungen wird einmalig erstellt, indem zunächst einmal für jedes Einzelmolekül die bindungsfähigen Atome identifiziert, qualifiziert und charakterisiert werden. Die Liste dieser oberflächennahen, eindeutig identifizierbaren Atome wird zusammen mit den eindeutig zuordbaren Atomrepräsentationen permanent und speicherplatzsparend in binärem Format abgelegt, da sie eine gemeinsame, universelle Grundlage für das Paar- und das Triplet-Docking bildet.

Aus diesen Atomlisten werden in systematischer Vorgehensweise alle mögliche Paare bzw. Triplets nichtredundant zusammengestellt und direkt einer Selektion der monomolekularen transformations-invarianten Eigenschaften unterzogen. Die vollständige Molekülrepräsentation ergänzt sich somit durch die Ablage der Listen kriterienkonformer Atompaare bzw. -triplets anhand ihrer eindeutigen Atomidentifikationsnummern. Solange die Zusammensetzung der hier verwendeten Selektionsgrenzwerte unverändert bleibt, sind die Listen für jeden folgenden Dockingvorgang gültig. Zusatzinformation eines jeden Listeneintrages ist der Atomabstand des Paares bzw. der ersten Dreieckskante. Anhand dieser Werte werden die Listen zum Abschluss der Zusammenstellung der Molekülrepräsentation aufsteigend sortiert, um später den Dockingvorgang zu beschleunigen. Atomtriplets wird zusätzlich die Festlegung der Bindungsseite zugeschrieben. Auch diese Datenablage erfolgt speicherplatzsparend in binärem Format.

Jedem Paar oder Triplet könnten auch die jeweiligen Werte aller weiteren monomolekularen transformations-invarianten Eigenschaften gespeichert werden, damit diese nicht bei jeder Verwendung im eigentlichen Dockingvorgang neu zu berechnen sind. Dieses verbietet jedoch der Bedarf an permanentem Speicherplatz aufgrund des relativ geringen Rechenaufwandes zur Neuerzeugung, insbesondere unter Berücksichtigung der Existenz mehrprotoniger Donatoren und ambivalent bindungsfähiger Atome mit ihren vielfältigsten Paar- und Tripletzusammenstellungen. Wie eingangs erwähnt, mag es aber von Interesse sein, hier andersartige Informationen für eine spätere Verwendung zur Verfügung zu stellen. Dieses können zum Beispiel Informationen zur Oberflächenbeschreibung für alternative Methoden der Molekülorientierung sein. Die Erstellung derartiger Informationen darf in dieser zeitlich unkritischen Aufbereitungsphase durchaus rechenintensiv sein.

## 8 Docking zweier Proteinmoleküle

Im Gegensatz zur Aufbereitung und Repräsentation der einzelnen Proteinmoleküle, ist der eigentliche Dockingvorgang für zwei Moleküle ein äußerst zeitkritischer Arbeitsgang. Seine Effizienz hat somit allerhöchste Priorität, wenn es das Ziel ist, eine schnelle Vorauswahl potentiell produktiver Molekülkonstellationen für eine nachfolgende, genauere aber auch zeitaufwendigere Dockingmethode zu präsentieren. Aus diesem Grunde werden als Datengrundlage die zuvor erstellten Molekülrepräsentationen verwendet, die eine vollständige Neuberechnung ihrer methodenspezifischen Informationsgehalte unnötig machen oder auf ein sinnvoll vertretbares Minimum beschränken.

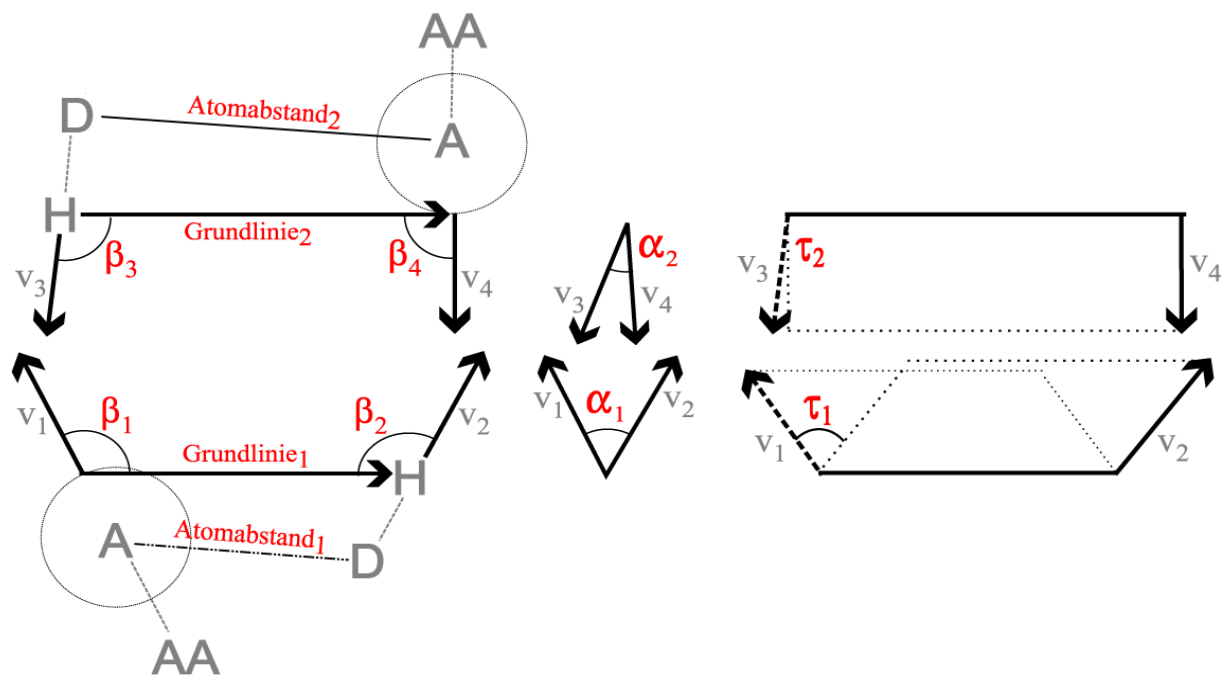
Nach dem Einlesen der Charakterisierungen bindungsfähiger Atome und der Listen kriterienkonformer Atompaaire bzw. -tripletts kommt es zur Erstellung aller möglichen intermolekularen Zusammenstellungen dieser Paare oder Tripletts zur Bewertung ihrer kombinationsgegebenen, bimolekularen transformations-invarianten Eigenschaften anhand der verwendeten Kriteriengrenzwerte. Kriterienkonforme Zusammenstellungen werden durch diese Evaluierung als sogenannte Kandidaten identifiziert, was bedeutet, dass für ihr Atomensemble die mögliche zeitgleiche Ausbildung von zwei bzw. drei intermolekularen Wasserstoffbrückenbindungen postuliert wird.

Die Lokalisierung der beiden beteiligten Atompaaire bzw. -tripletts beschreibt für jedes Molekül einen Berührungspunkt mit seinem Komplexpartner. Somit steht jeder Kandidat für das Auftreten einer Orientierung, bei der die beiden Moleküle an ihren Oberflächen in produktiven Kontakt treten könnten. Für jede Orientierung bedarf es der Herleitung einer geeigneten Transformationsvorschrift für eines der Moleküle, welche die Komplexkonstellation aus der ursprünglich unabhängigen Lage der Einzelmoleküle herbeiführt. Gleichartige Transformationen können sodann in Ergebnisgruppen zusammengefasst werden. Diese Cluster dienen der Identifikation und Qualifikation der häufigsten Transformationsbeschreibungen, unter denen die einzig richtig positive Molekülkonstellation des nativen Molekülkomplexes aufgrund der Erwartung überproportional vieler und qualitativ guter Bindungen vermutet wird.

Nachfolgend sollen nun zunächst getrennt die in sich geschlossenen Themengebiete der bimolekularen transformations-invarianten Eigenschaften intermolekularer Paar- und Triplettkombinationen, der Molekülorientierung, der Transformation und Transformationsclusterung sowie der Clusterqualifizierung erläutert werden, bevor der Gesamtverlauf eines Dockings umfassend dargelegt wird.

## 8.1 Bimolekulare transformations-invariante Eigenschaften der Kombination zweier Paare bindungsfähiger Atome

Das Vektormodell, welches die Paarzusammenstellung zweier bindungsfähiger Atome innerhalb einer Moleküloberfläche beschreibt, wurde erläutert. Im eigentlichen Docking werden nun zwei derartige Repräsentationen aus den unterschiedlichen Molekülen zu einem gemeinsamen Vektorkonstrukt kombiniert. Hier liegen sich je zwei Atome zur eventuellen Ausbildung einer intermolekularen Wasserstoffbrückenbindung gegenüber. Das resultierende Modell einer derartigen Paar-Paarkombination zeigt Abbildung III-3 zusammen mit den notwendigen Bezeichnungen.



**Abbildung III-3:** Das vollständige, bimolekulare Vektorkonstrukt zweier Paare bindungsfähiger Atome. Die Bezeichnungen entsprechen der Darstellung des monomolekularen Vektorkonstrukts. Die Richtungen der Grundlinienvektoren ist aufeinander abgestimmt. Die Paare gegenüberliegender  $\beta$ -Winkel werden stets mit  $(\beta_1+\beta_3)$  sowie  $(\beta_2+\beta_4)$  bezeichnet.

Ziel ist es, aus der Kombination aller transformations-invarianten Eigenschaften die Einschätzung abzuleiten, ob eine zeitgleiche Ausbildung zweier funktioneller Wasserstoffbrückenbindungen, welche den Kriterien von Baker und Hubbard gehorchen, für die vier beteiligten Atome möglich ist. Für oberflächenparallele Kontaktflächen entspricht die

Idealvorstellung zwei gleichartigen und kooperativen Bindungen mit besonderem Augenmerk auf eine möglichst antiparallele Ausrichtung der jeweiligen Interaktionsvektoren verbunden mit einer möglichst parallelen Anordnung der Grundlinienvektoren.

Im Einzelnen können hieraus folgende bimolekulare transformations-invariante Eigenschaften abgeleitet werden:

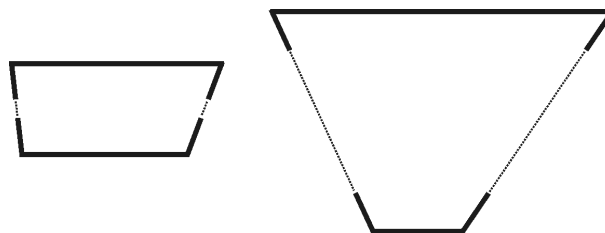
- **Komplementarität der Bindungseigenschaften**

Ein Akzeptor kann nur mit einem Donator und *vice versa* interagieren. Gleichartige Eigenschaften in derart räumlicher Nähe wirken sich kontraproduktiv aus und können im Normalfall keinen evolutiven Hintergrund besitzen.

- **Differenz der Atomabstände und Differenz der Grundlinienlängen**

Optimum:  $0 \text{ \AA}$ , Wertebereich:  $0 \text{ \AA} \leq |\text{Differenz}| \leq |\text{Abstand}_{\text{max}} - \text{Abstand}_{\text{min}}|$

Diese Größen beschreiben am besten das Prinzip, aus den vorliegenden geometrischen Eigenschaften auf Erfüllbarkeit der Bindungskriterien von Baker und Hubbard zu schließen, ohne die absolute räumliche Anordnung des Atomquartetts und damit die spezifischen Kriterienvoraussetzungen dieser Einschätzungshilfe zu kennen. Abbildung III-4 zeigt, wie eine direkte Aussage über die resultierende Bindungslänge der Wasserstoffbrücke getroffen werden kann. Wenn man von antiparallelen Interaktionsvektoren ausgeht, können diese eventuell auch für eine Kombination von Atompaares mit großem Unterschied der Grundlinienlängen gegeben sein. Die dafür notwendigen Bindungslängen lägen jedoch weit außerhalb der effektiv möglichen Bindungskräfte einer Wasserstoffbrückenbindung.



**Abbildung III-4:** Beispiel für die unterschiedlichen Auswirkungen einer kleinen (links) oder großen (rechts) Differenz der Grundlinienlängen auf die notwendigen Bindungslängen (gestrichelte Linien). Die Endpunkte der Trapezformen entsprechen den Interaktionspunkten der beteiligten Atome.

- **Summe gegenüberliegender  $\beta$ -Innenwinkel**

Optimum:  $180^\circ$ , Wertebereich:  $2 \cdot \beta_{\text{min}} \leq (\beta_1 + \beta_3)$  bzw.  $(\beta_2 + \beta_4) \leq 2 \cdot \beta_{\text{max}}$



Diese Eigenschaft stellt sicher, dass die Grundlinien mit Abweichungen parallel zueinander zu liegen kommen, wenn die Interaktionsvektoren aufeinander ausgerichtet sind, damit auch beide Wasserstoffbrückenbindungen zeitgleich reale Längen aufweisen können. Im Idealfall ist die Summe gegenüberliegender  $\beta$ -Winkel jeweils  $180^\circ$ . Hierbei ist jedoch zu beachten, dass die  $\beta$ -Winkel torsionsabhängig sind und dieses Ideal auch eine  $\tau$ -Differenz von  $0^\circ$  voraussetzt.

- **Summe aller  $\beta$ -Innenwinkel**

Optimum:  $360^\circ$

Wertebereich:  $4 \cdot \beta_{\min} \leq (\beta_1 + \beta_2 + \beta_3 + \beta_4) \leq 4 \cdot \beta_{\max}$  und auch

Wertebereich:  $2 \cdot (\beta_1 + \beta_3) \leq (\beta_1 + \beta_2 + \beta_3 + \beta_4) \leq 2 \cdot (\beta_1 + \beta_3)$ , bzw. entspr. mit  $(\beta_2 + \beta_4)$

Dieses Kriterium wäre an sich hinfällig, wenn die Vorgaben der  $\beta$ -Winkel und der Summe gegenüberliegender  $\beta$ -Winkel erfüllt werden. Es hat sich jedoch herausgestellt, dass derart übergeordnete und zusammenfassende Selektionsmöglichkeiten sehr gut geeignet sind, ausgleichende Eigenschaften in Kombination zu bevorzugen. Hiermit ist gemeint, dass z.B. eine Summe gegenüberliegender  $\beta$ -Winkel eng an seinem Kriteriumminimum liegen mag, die zweite Summe der Paarzusammenstellung dann aber ausgleichend optimal oder größer, nicht jedoch ebenfalls an der unteren Grenze gegeben sein muss. Die Grenzwerte der Gesamtsumme sind hierzu enger zu wählen als die angeführten mathematischen Minima und Maxima.

- **Differenz der  $\tau$ -Torsionswinkel unter Berücksichtigung der Torsionsrichtung**

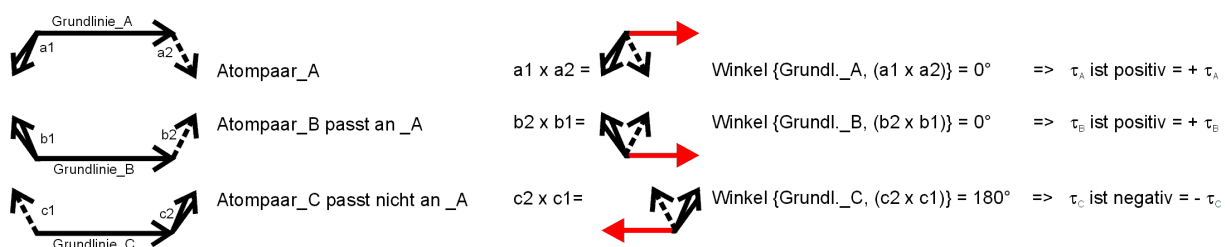
Optimum:  $0^\circ$ , Wertebereich:  $0^\circ \leq |\tau_1 - \tau_2| \leq |2 \cdot \tau_{\max}|$

Dieses Kriterium testet, ob die vier Interaktionsvektoren bei einer parallelen Anordnung der Grundlinienvektoren überhaupt in ihrer Raumbtiefausrichtung ausreichend übereinstimmen können.

Hierzu muss erläutert werden, wie die Spreizrichtung in Abhängigkeit vom betrachteten Molekül bestimmt wird, wobei Bezug genommen wird auf die erläuternde Abbildung III-5.

Der Betrag des  $\tau$ -Winkels wird zwischen den beiden Normalenvektoren der von Grundlinienvektor und Interaktionsvektoren aufgespannten Ebenen ermittelt, wodurch die seitliche Trapezöffnung des Vektorkonstrukts ignoriert wird. Bestimmt man nun für das Atompaar des ersten Moleküls den Kreuzvektor der beiden Normalen mit  $(n_{A1} \times n_{A2})$  und vergleicht ihn durch Winkelmessung mit dem Grundlinienvektor, so stellt man Parallelität oder Antiparallelität fest und kann per Konvention vereinbaren, dass der Winkel  $\tau_A$  nun positiven oder negativen Drehsinn besitzt. Für das Atompaar des zweiten Moleküls geht man entsprechend vor, bestimmt den Kreuzvektor der Normalen jedoch invertiert mit  $(n_{B2} \times n_{B1})$  und beachtet, dass die Richtung des Grundlinienvektors für diesen Vergleich entsprechend der

Ausrichtung beim ersten Molekül gezogen wird. Nun kann die Differenz stets mit  $\tau_A - \tau_B$  gebildet werden und ist somit in der Lage, zwischen den beiden dargestellten Fällen einer ausreichen kleinen Differenz von  $|(+\tau_A) - (+\tau_B)|$  und einer eventuell zu großen Differenz von  $|(+\tau_A) - (-\tau_B)|$  –und entsprechenden Fällen mit negativem  $\tau_A$ – zu unterscheiden. Ist eine Orientierung zwar gegenseinnig aber in ausreichend geringem Maße, so ist eine produktive Doppelinteraktion problemlos vorstellbar, da z.B.  $|(+1^\circ) - (-1^\circ)| = (+2^\circ)$  ja eine fast optimale Übereinstimmung beschreiben.



**Abbildung III-5:** Schematische Darstellung zur Bestimmung der Torsionsrichtung des  $\tau$ -Ebenenwinkels. Dargestellt sind die Grundlininvektoren, die Normalenvektoren (a1 bis c2) der nichteingetragenen Ebenen und die hieraus per angegebenen Kreuzprodukt ermittelten Vergleichsvektoren in Rot. Genauere Details im Text.

- nicht verwendet: **Differenz der  $\alpha$ -Innenwinkel**

An sich müsste es sich mit dieser Eigenschaft analog zur Differenz der  $\tau$ -Torsionswinkel verhalten. Da der  $\alpha$ -Innenwinkel jedoch nicht nur auf die seitliche Spreizkomponente der Trapezform reduziert wird, sondern auch die Torsionskomponente enthält, lässt sich keine analoge Unterscheidung einer “nach Innen-“ oder “nach Außen-“ gerichteten Spreizung treffen. Hierdurch ist auch die Differenzbildung nicht eindeutig und das Kriterium findet keine weitere Anwendung innerhalb dieser Arbeit.

## 8.2 Bimolekulare transformations-invariante Eigenschaften der Kombination zweier Triplets bindungsfähiger Atome

Analog zu vorherig getroffenen Aussagen für die monomolekularen Eigenschaften, werden die bisher vorgestellten Kriterien einer intermolekularen Paar-Paar-Kombination einfach auf die drei Dreieckskante-Dreieckskante-Kombinationen übertragen. Wie bei der Erstellung der Atomtriplets erhöht sich bei Verwendung identischer Grenzwerte sich auch hier der

Selektionsdruck gegenüber den Paarbewertungen, da alle Kriterien zeitgleich für alle Kanten erfüllt werden müssen.

Da die Triplett-Vektordivergenz lediglich zur Qualifizierung eines Atomtripletts verwendet wird, jedoch nicht als transformations-invariante Eigenschaft weitere Verwendung findet, ergibt sich hier keine zusätzliche bimolekulare Vergleichs- und Bewertungsmöglichkeit.

Die zweite tripleteigene Eigenschaft des Drehsinns der Atomanordnung, bzw. der daraus abgeleiteten Dreiecks-Bindungsseite ist keine derartige qualitative Einschätzung. Die Kombination zweier ausgewiesener Bindungsseiten ist jedoch eine bimolekulare Eigenschaft und findet hier seine Verwendung vor der eigentlichen Evaluierung der übrigen geometrischen Eigenschaften. Sie legt fest, in welcher Anordnung zwei Triplets zueinander evaluiert werden, was wiederum determiniert, welche Dreieckskanten kombinieren. Die möglichen Orientierungen werden erst bei der Gesamtdarstellung eines Dockingablaufs im Anschluss an die Vorstellung der Molekülorientierungsmethode genauer und mit einer Abbildungen dargestellt. Hier soll jedoch angemerkt werden, dass zwei Triplets mit gegenläufigem Drehsinn und zwei Triplets mit gleichläufigem Drehsinn jeweils nur in einer Dreiecksanordnung mit den festgelegten Bindungsseiten zueinander evaluiert werden müssen. Ist für eines der Dreiecke die Bindungsseite nicht eindeutig bestimmbar, so sind beide Anordnungen zu berücksichtigen und mit annähernd doppeltem Rechenaufwand zu untersuchen.

### **8.3 Molekülorientierung und Erstellung der Transformationscluster**

In der Aufarbeitung der Moleküle wird ein Molekül bereits in eine abstrakte Form übertragen, indem bindungsfähige Atome durch Vektoren und Bindungseigenschaften beschrieben und zu Paaren bzw. Triplets kombiniert werden, welche den monomolekularen Selektionsanforderungen der Methode genügen. Im Docking wird dann fast nur noch mit ortsgebundenen Vektoren und den transformations-unabhängigen Eigenschaften gearbeitet, obwohl es doch das Ziel ist, eine Molekültransformation zu erhalten, welche die Konstellation eines produktiven Protein-Protein-Komplexes herbeiführen mag.

Ergebnis einer erfolgreichen Evaluierung im Docking ist zunächst einmal ein kriterienkonformer Kandidat. Nun muss für jeden einzelnen Kandidaten eine eindeutige Transformationsvorschrift bestimmt werden, welche die Moleküle entsprechend zueinander positioniert. Hierfür bedarf es eines Orientierungsmodells als Bindeglied zwischen den beteiligten Atomen und der notwendigen Transformation. Der Kandidat kann von da an einzig durch diese

Transformationsvorschrift charakterisiert werden, denn die Identität seiner Atome ist nicht mehr von Bedeutung.

Die Methode beruht darauf, dass überproportional viele intermolekulare Bindungen in der Kontaktfläche ausgebildet werden. Diese können dann als Kandidaten identifiziert und somit mit gleichartigen Transformationen beschrieben werden. Die gleichartigen Transformationen werden in sogenannten Clustern zusammengefasst, wobei das Einzelergebnis nochmals seine singuläre Bedeutung verliert, indem es nur noch zur Transformationsbeschreibung des Ergebnisclusters und zu dessen Belegungszahl beiträgt.

### *8.3.1 Orientierungsmodell der Molekülanordnung*

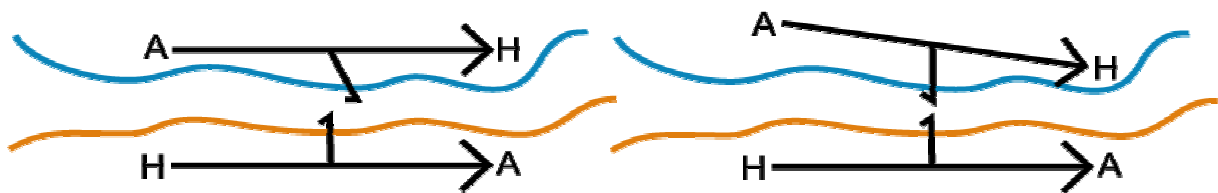
Aufgrund der Gleichstellung eines Atomtriplets mit drei gekoppelten Atompaares erfolgt auch die Orientierung im Triplett-Docking zunächst getrennt für jede Dreiecksseite.

Entsprechend den bisherigen Ausführungen sind mit den verwendeten Methoden zum Zeitpunkt der Identifikation eines Kandidaten als Informationen gegeben: Die absoluten Koordinaten der beteiligten Atome in kartesischen Koordinaten und die relativen Vektorkonstrukte der beteiligten Atompaares, aus denen sich z.B. die absoluten Koordinaten der Interaktionspunkte herleiten lassen. Die geometrischen Grundlagen des Orientierungsmodells sind also sehr eingeschränkt. Sie beinhalten vor allem keinerlei weitere Informationen über die Beschaffenheit, z.B. über die Planarität geschweige denn Komplementarität der beiden Moleküloberflächen in näherer Umgebung der beteiligten Atome.

Das verwendete Orientierungsmodell ist daher recht simpel. Es entstammt der Idealvorstellung einer Paar-Paarung, die ein Rechteck beschreibt. Bei diesem besitzen die Verbindungsvektoren der Atompaares gleiche Länge, die Atome sind in gleichem Abstand zur komplementären, idealerweise planaren Moleküloberfläche positioniert und ihre Interaktionsvektoren stehen senkrecht auf den Verbindungsvektoren. In diesem Fall kämen die Verbindungsvektoren zwischen den Atomen, wie auch zwischen den jeweiligen Interaktionspunkten (die Grundlinie) parallel zueinander zu liegen, während die Interaktionsvektoren antiparallel zum jeweiligen Bindungspartner ausgerichtet sind. Reduziert man die beiden Interaktionsvektoren eines jeden Atompaares durch Mittelung zu einem einzelnen Vektor, der im Mittelpunkt der Verbindungsvektoren seinen Ursprung hat, so sind auch diese beiden Mittelvektoren zueinander antiparallel. Diese genaue Antiparallelität bleibt selbst dann erhalten, wenn das gesamte bimolekulare Vektorkonstrukt kein orthogonales Rechteck sondern ein Parallelogramm oder ein Trapez beschreibt. Da dieser Idealfall und seine Varianten im Regelfall nicht anzutreffen ist,

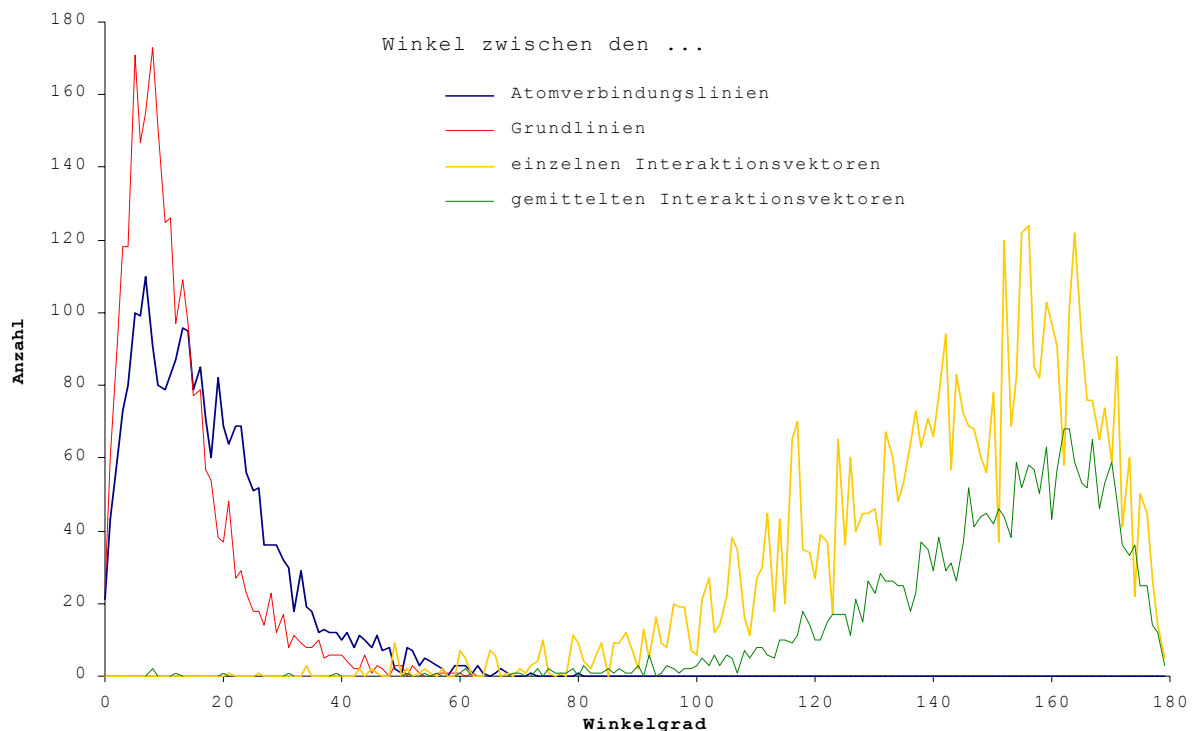
lassen sich hieraus zwei Orientierungsmodelle ableiten, die in Abbildung III-6 wiedergegeben sind:

- parallele Verbindungsvektoren und quasiantiparallele Mittelvektoren, die gegenläufig in einer gemeinsamen Ebene aller Vektoren angeordnet sind
- antiparallele Mittelvektoren und quasiparallele Verbindungsvektoren, die gleichläufig in einer gemeinsamen Ebene aller Vektoren angeordnet sind



**Abbildung III-6:** Schematische Darstellung der beiden im Text genauer vorgestellten Orientierungsmodelle mit den Grundlinienvektoren zwischen den Interaktionspunkten und den Mittelvektoren. Diese werden aus den nichtingezeichneten Interaktionsvektoren aller bindungsfähigen Positionen erzeugt.

Um die Entscheidung für eine dieser Orientierungsmodelle zu fällen, werden für alle Referenzergebnisse des Paar-Dockings die Winkelverteilungen zwischen den entsprechenden Vektoren aus den nativen Atompositionen der Molekülkomplexstrukturen betrachtet. Das Diagramm findet sich in Abbildung III-7.



**Abbildung III-7:** Die Häufigkeitsverteilungen unterschiedlicher Winkel innerhalb des Vektorkonstruktes, das zur Molekülorientierung verwendet wird. Verwendet wurden die 2360 Referenzergebnisse der fakultativen und obligatorischen Proteinkomplexe beim Paar-Docking mit allen Atomen unter Verwendung des später vorgestellten weiten Grenzwertensembles. Weitere Details im Text.

Im Vergleich der Parallelitäten zwischen den jeweiligen Atomverbindungsvektoren und zwischen den jeweiligen Grundlinienvektoren zeigen die Grundlinienvektoren die schmalere Verteilung. Mit überwiegender Mehrheit beträgt hier die Winkelabweichung weniger als  $20^\circ$  und somit etwa  $15^\circ$  bis  $20^\circ$  weniger als bei der vergleichbaren Menge für den Vektor zwischen den Atomzentren. Dieses ist an sich problemlos nachvollziehbar, da die Bindungsdistanz von Proton zur Akzeptor-van-der-Waals-Schale stets sehr gering sein muss, unabhängig davon, wie unsymmetrisch das zugrunde liegende Vektorkonstrukt mit seinen Atompositionen ist.

Die Interaktionsvektoren der jeweiligen Bindungspartner stehen bei weitem nicht so gut antiparallel zueinander wie das Idealmodell vermuten lassen möchte. Maximale Verteilungswerte werden etwa bei  $150^\circ$  bis  $165^\circ$  erreicht, was zu einem großen Teil auch an der Annahme kollinearere Proton-Akzeptor-kovalent-vorhergehendes-Atom-Bindungen liegen wird, wie bereits bei der Vorstellung der Akzeptor-Repräsentierung erläutert wurde. Dementsprechend zeigen sich auch die Ergebnisse für die Messung der Antiparallelität hieraus erzeugter

Mittelvektoren. Die Kurve verläuft wesentlich flacher und weiter vom 180°-Optimum entfernt als die vorgestellten analogen Verteilungen der Parallelitäten.

Es ist somit abzuleiten, dass eine deutlich bessere Parallelität der Grundlinienvektoren von stets kleiner 60°, in der Regel kleiner 30° gegeben ist als eine Antiparallelität der Mittelvektoren von zumeist größer 100°. Daher wird das Orientierungsmodell paralleler Grundlinien und quasiantiparalleler Mittelvektoren verwendet.

Die einfache Konstruktion der Orientierungsmethode hat zur Folge, dass die Transformationsbeschreibung eines jeden Kandidaten, der nativ nicht optimal parallele Grundlinien besitzt, fehlerbehaftet ist, weil er zu dieser parallelen Anordnung gezwungen wird. Der Fehler wird verringert im Docking der Atomtripletts. Hier werden vor der Translationsbestimmung zunächst die Rotationen der Dreieckskanten unabhängig bestimmt und dann zur Ergebnisbeschreibung gemittelt, so dass die Mehrfachbestimmung Fehler ausgleichen kann. Unter Vorwegnahme einzelner Ergebnisse seien hier als Richtwerte angegeben, dass im Triplet-Docking nur für maximal 70% obligatorischer und 60% fakultativer Proteinkomplexe die richtige Rotation erhalten wird. Im Paar-Docking sinken die Quoten auf maximal 50% und 30%.

### 8.3.2 Erstellung der Transformationsvorschriften

Die verwendeten, eindeutigen Transformationsvorschriften beinhalten, dass das im Programmaufruf erst-aufgerufene Molekül stets in seiner Lage festgehalten wird, wogegen das zweit-aufgerufene Molekül zur Verlagerung freigegeben ist. Es wird zunächst entsprechend der notwendigen Kombination der Eulerwinkel rotiert, um das Orientierungsmodell zu erfüllen und anschließend der Translation entsprechend den gerade erzeugten, absoluten Koordinaten unterzogen.

#### 8.3.2.1 Ermittlung der Rotationswinkel

Um die Vektorpositionierungen des Orientierungsmodells herbeizuführen, setzt man zunächst die beiden Systeme aus zwei Vektoren in ein globales Koordinatensystem. Es folgen die Bestimmung der Rotationen in festgelegter Reihenfolge, die die beiden Systeme einander gegenüber setzen. Hierbei kommen die Grundlinienvektoren nun mit einer gemeinsamen Richtung auf einer Hauptachse zu liegen und die Mittelvektoren in gegenläufiger Orientierung in einer der Hauptebenen. Die notwendigen Einzelrotationen können nun verrechnet werden und aus der Gesamtrrotationsmatrix lassen sich die drei Eulerwinkel der verwendeten Rotationsvereinbarung isolieren.

Ebenso wäre es möglich, durch Kreuzvektorberechnung aus den beiden Systemen zweier Vektoren zwei Systeme mit je drei gekoppelten Vektoren zu erstellen. Hiermit bleibt das Gleichungssystem einer Rotationsberechnung nicht mehr unterbestimmt und die Gesamtrationsmatrix kann abgeleitet werden.

Für Atomtriplets werden die Rotationswinkel für jede Dreiecksseite getrennt bestimmt. Dieses stellt sicher, dass erfüllbare Orientierungen bestimmt werden, die nicht aus der Evaluierung falscher Dreiecksbindungsseiten stammen können. Hierbei wird bei der Mittelung der drei Rotationswinkelsätze bewertet, wie groß die Summe der Abweichungen für die einzelnen Winkel der Winkelsätze zum entsprechenden Winkel des gemittelten Endergebnisses ist.

#### 8.3.2.2 Ermittlung des geeigneten Rotationsrepräsentanten

Bevor die eigentliche Rotation der Atome des zweit-aufgerufenen Moleküls durchgeführt wird, wird an dieser Stelle bereits der belegte Rotationscluster für diesen Kandidaten identifiziert. Es handelt sich also wie bereits erläutert nicht um natürlich begrenzte Cluster gleichartiger Rotationen, sondern um die Zuordnung zu dem nächstgelegenen Rotationsrepräsentanten.

Die verwendete regelmäßige Verteilung von Rotationsrepräsentanten nach E. Lattman wurde im Kapitel methodischer Grundlagen vorgestellt [Lattman, 1972]. Die Liste ist aufsteigend sortiert nach den regelmäßig inkrementierten  $\beta$ -Werten und innerhalb dieser Gruppen aufsteigend nach den  $\alpha$ -Werten.

Für den ermittelten Rotationswinkelsatz wird diese zweidimensional angelegte und direkt indizierte Liste der Rotationsrepräsentanten gezielt nur in geeigneten Abschnitten unter Rücksichtnahme auf die  $360^\circ$ -Periodizität der Rotationswinkel abgefragt. Er wird hierbei nach dem Minimum der Summe der Abweichungen aller drei Rotationswinkel gesucht und somit die stellvertretende Winkelkombination identifiziert, welche die erforderliche Rotation am besten nähert.

#### 8.3.2.3 Rotation und Ermittlung der Translation

Bevor für einen Kandidaten die notwendige Translation festgelegt werden kann, sind seine hierfür notwendigen Koordinatenpunkte der zuvor bestimmten Rotation zu unterziehen. Die Rotation erfolgt anhand der Winkelkombination des entsprechend identifizierten Rotationsrepräsentanten, um Fehler der Orientierung und der Rotations-Abstrahierung hier zusammenzufassen und dennoch nachfolgend gleichartige Translationen für zusammengehörige Kandidaten zu erhalten. Dieses Vorgehen hat sich empirisch als geeigneter herausgestellt, um



einheitliche Ergebnisse zu liefern als die Translationsermittlung mit Koordinatenpunkten, welche mit den Winkeln der Orientierungsmethode rotiert wurden.

Nach der Rotation liegen alle Orientierungsvektoren in zwei Parallelebenen vor. Es hat nun noch eine Verschiebung stattzufinden, die alle Vektoren in eine einzige Ebene legt und zudem die notwendige räumliche Nähe der interagierenden Atome erzeugt.

Die Verschiebung erfolgt so, dass der Mittelpunkt der Interaktionspunkte beteiligter Atome eines Paares oder Triplets des zweit-aufgerufenen Moleküls deckungsgleich auf den entsprechenden Mittelpunkt des erst-aufgerufenen, fixierten Moleküls zu liegen kommt.

Diese Translationsbeschreibung des Kandidaten kennzeichnet das positive Evaluierungsergebnis, indem es ohne weitere Angabe der beteiligten Atome dem entsprechenden Rotationsrepräsentanten zugeschrieben wird.

### 8.3.3 Bildung der finalen Transformationscluster

Aufgrund des methodischen Vorgehens, für die Festlegung der Translation zunächst den Rotationsrepräsentanten zu identifizieren und diesem auch anschließend die Translation als Beschreibung des kriterienkonformen Kandidaten zuzuordnen, ist bereits die Hälfte der Arbeit zur Bestimmung der finalen Transformationscluster erledigt. Die Gruppierung der Kandidaten entsprechend ihren Rotationsvorschriften ist erfolgt. Innerhalb dieser Untergruppen werden noch die natürlichen Cluster von Kandidaten erstellt, welche aufgrund gleichartiger Translationen zusammengehören. Hierbei können die Rotationsgruppen unabhängig und nacheinander abgearbeitet werden.

Zunächst werden die Transformationstupel eliminiert, die nicht eine ausreichende Menge gleichartiger Transformationsnachbarn besitzen, was die Gesamtanzahl der Kandidaten und den folgenden Rechenaufwand der Gruppenbildung reduziert. Für die verbleibenden Translationstupel werden alle *Single-Linkage*-Cluster mit einem geeigneten Abstandskriterium für die nahe benachbarten Punkte erstellt. Jeder Cluster hat eine Mindestbelegung aufzuweisen. Für den größten Cluster werden die enthaltenen Transformationstupel gemittelt und die resultierende Translation zusammen mit der Anzahl der Kandidaten dieses Clusters als Qualifizierungsmerkmal des Rotationsrepräsentanten eingesetzt. Dieses stellt einen finalen Transformationscluster dar.

Die verwendeten Anzahls- und Abstandskriterien wurden empirisch aus den Analysen der Referenzkandidaten ermittelt und werden mit der entsprechenden Dokumentation im Ergebnisteil ausgeführt.

#### **8.4 Qualifizierung und Sortierung der Transformationscluster**

Bei der Verwendung einer regelmäßigen 30°-Aufteilung des Rotationsraums ergeben sich 546 Repräsentanten. Diese werden zum Ende des Evaluierungsvorganges jeweils charakterisiert durch ihren größten Translationscluster und dessen Kandidatenbelegung, die die vorgegebene Mindestbelegung erreichen muss. Anhand dieser Beschreibung können diese finalen Transformationscluster nun verglichen, qualifiziert und entsprechend dieser Qualität sortiert werden.

Unter dem Begriff der Cluster-Qualifizierung verstehe ich Identifikationsmethoden die, den Rotationscluster mit den richtig positiven Ergebnissen gegenüber den übrigen 563 Rotationsclustern als “den Richtigen“ heraus stellen oder zumindest in eine Auswahl der “wohlmöglich Richtigen“ einordnen.

Der Gedankengang hierzu lautet:

- Intermolekulare Wasserstoffbrückenbindungen finden sich überproportional häufig in den spezialisierten Kontaktflächen spezifischer Molekülkomplexe.
- Alle richtig positiven Ergebnisse erfordern die gleiche Orientierung.
- Sie fallen somit in einen der Rotationscluster, welcher damit eventuell bereits eine signifikant höhere Belegungszahl aufweisen kann als Rotationscluster mit ausschließlich falsch positiven Ergebnissen.
- Innerhalb dieses Rotationsclusters bilden die richtig positiven Ergebnisse einen geschlossenen Translationscluster, welcher
  - a) der größte Translationscluster in diesem Rotationscluster und
  - b) sogar der größte Translationscluster in allen Rotationsclustern sein sollte.
- Somit sollte der richtige Rotationscluster über die Belegungszahl seines größten Translationsclusters identifizierbar sein.

Als Bewertungskriterien zeichnen sich hier entweder die Belegungen des Rotations- und/oder Translationsclusters gemessen an einem absoluten Schwellenwert oder seine Position in einer absteigenden Sortierung nach Belegungszahlen ab.

Zusätzlich zu derartig fixierten Qualifizierungskriterien wurden noch Möglichkeiten untersucht, eine dynamische Cluster-Qualifizierung durchzuführen, deren Kriterium mit Eigenschaften der behandelten Moleküle korreliert. Derartige Eigenschaften waren z.B. die Größe bzw. Atomanzahl der Dockingpartner oder ein wenig spezifischer die Anzahl oder Dichte der bindungsfähigen Oberflächenatome. Statische und dynamische Bewertungskriterien können problemlos kombiniert werden. Derartige Kombinationen wurden ansatzweise getestet, konnten jedoch noch keine eindeutige und unfehlbare Präferenzen hervorbringen, so dass zunächst nur die Sortierungspositionen erörtert werden.

## **8.5 Gesamtablauf eines Dockings unter Erläuterung aller möglichen Anordnungen einer Paar-Paar- bzw. Triplett-Triplett-Kombination**

### *8.5.1 Übersicht über die Arbeitsschritte*

1. Programmaufruf mit Molekül- und Kettenidentifikatoren beider Moleküle "A" und "B".
2. Einlesen der Liste aller diskreten Rotationsrepräsentanten aus der regelmäßigen Einteilung des gesamten Rotationsraumes nach Eaton Lattman für einen gewählten Repräsentantenabstand. Diese könnten auch zur Laufzeit generiert werden, müssten dann jedoch jedesmal noch nach den  $\beta$ - und  $\alpha$ -Winkeln aufsteigend sortiert werden.
3. Gegebenenfalls das Einlesen der Liste von Referenzergebnissen, die für diese Molekülpaarung als richtig positive Ergebnisse wiederzufinden sind.
4. Einlesen des gewählten Ensembles an Kriteriengrenzwerten. Es ist darauf zu achten, dass bei der einmaligen Erstellung der Molekülrepräsentationen und dem hier durchgeführten Docking sowie gegebenenfalls der Bestimmung der Referenzergebnisse dieselben Kriteriengrenzwerte zugrunde gelegt werden.
5. Einlesen der Liste bindungsfähiger Atome mit ihren Raumkoordinaten und den Charakterisierungen ihrer Bindungsfähigkeiten (Akzeptor, Donator oder beides) und -eigenschaften (Interaktionsvektoren) für beide Moleküle mit der direkten Zugriffsmöglichkeit auf diese zusammengefassten Informationen über einen atomidentifikator-indizierten Zeigerstapel.
6. für Paare:  
Einlesen der Liste repräsentierender Atompaare für das erste Molekül "A", wobei zu jedem Paar aus den bereits erfassten Atomspezifikationen die monomolekularen transformations-invarianten Eigenschaften einmalig für dieses Docking berechnet und in einer gemeinsamen Datenstruktur abgelegt werden. Die Paare wurden bereits zuvor bei ihrer Generierung der Selektion auf Kriterienkonformität unterzogen und anschließend aufsteigend nach ihrem Atomabstand sortiert.

für Triplets:

Einlesen der Listen repräsentierender Atomtripletts für das erste Molekül "A". Die Dreieckskanten werden hierbei unabhängig als Atompaaire behandelt, die über ihre Atomidentifikatoren-Kombination referenziert werden können und analog zu den Paaren behandelt werden. Daher müssen ihre Eigenschaften auch beim Auftreten in mehreren Dreiecken nur einmalig berechnet und im Arbeitsspeicher abgelegt werden. Des Weiteren bedarf es an dieser Stelle der Berücksichtigung, dass ein Dreieck aus den Atomen "a1-a2-a3" später auch evaluiert werden muss in den gedrehten Anordnungen als "a2-a3-a1" und "a3-a1-a2", wie direkt im Anschluss erläutert und in Abbildung III-8 dargestellt.

Auch hierfür erfolgt keine Neuberechnung der Eigenschaften, da die Dreieckskanten in gleichbleibender Orientierung jedoch andersartiger Anordnung verwendet werden. Jedoch erfolgt die aufsteigende Sortierung der Dreiecke entsprechend dem Atomabstand ihrer ersten Dreieckskante mittels einer Implementierung des QUICKSORT-Algorithmus [Press *et al.*, 1992] erst im Anschluss an die Erzeugung der zwei gedrehten, jetzt als eigenständige Dreiecke geführten Auftrittsformen.

7. Nachdem alle Atompaaire bzw. -tripletts von Molekül "A" geordnet erfasst wurden, können Indizierungspunkte entsprechend dem sortierten Merkmal gesetzt werden. Diese werden bei der späteren Evaluierung aller intermolekularen Paar-Paar- bzw. Triplett-Triplett-Kombinationen genutzt, um spezifisch nur Partner mit geeignetem Sortierungsmerkmal auszutesten und für Partner, deren Eigenschaft außerhalb der kriterienkonformen Abweichung liegt, die Evaluierung gar nicht erst zu starten. Die Sortierung erfolgte entsprechend dem Atomabstand der Paare bzw. der ersten Dreieckskante und die Zugriffsmarken werden zunächst einfach im Abstand von 1,0 Å auf den jeweils ersten Vertreter dieser Eigenschaft gesetzt. Dieses Prinzip kann entsprechend den Vorstellungen der *Geometric Hashing*-Methode verfeinert werden durch multiple, mehrdimensionale Merkmalsortierungen und -indizierungen.
8. **"Äußere Schleife":**

Schrittweises Einlesen und Abarbeiten der Listen repräsentierender Atompaaire bzw. -tripletts des zweiten Moleküls "B". Für jeden Eintrag müssen folgende Arbeitsschritte angegangen werden:

  - 8.1. Berechnung der monomolekularen transformations-invarianten Eigenschaften des Atompaares bzw. aller Dreieckskanten.
  - 8.2. Identifikation des unteren Indizierungspunktes des Atomabstandmerkmals, für das alle Paare bzw. Triplets des ersten Moleküls "A" sortiert wurden, entsprechend dem angewendeten Selektionsgrenzwert.

(Beispiel: Atomabstand für Paar von Molekül "B" = 8,5 Å, Atomdifferenz-Grenzwert  $\leq 2,0$  Å, somit kriterienkonformer Atomabstand für Paar von Molekül "A" = 6,5 Å bis 10,5 Å und daher unterer Indizierungspunkt für das erste Paar mit  $\geq 6,0$  Å bei einer Indizierung im Abstand von je einem Ångström)

### 8.3. "Innere Schleife":

Sequenzieller Abruf aller Paare bzw. Triplets von Molekül "A", die in ihrer Auflistung hinter dem unteren Indizierungspunkt folgen. Sobald der erste Vertreter evaluiert wird, dessen Atomabstand die obere Selektionsgrenze (Atomabstand von Molekül "B" zuzüglich maximaler Atomabstanddifferenz) überschreitet, kann die innere Schleife abgebrochen werden, da nur weitere Vertreter mit noch größeren Atomabständen folgen.

- Für jeden Evaluierungsaufwurf dieser inneren Schleife müssen gegebenenfalls zwei mögliche Anordnungen des Paares bzw. Triplets von Molekül "B" berücksichtigt werden (entsprechend 8.3.1. und 8.3.2.), wie die Ausführungen im Anschluss an diese Auflistung der Arbeitsschritte erläutern.
- Für jede dieser Anordnungen, welche prinzipiell eigenständige Paar-Paar- bzw. Triplet-Triplett-Kombinationen darstellen kann hier ein Test erfolgen, ob es sich um ein Referenzergebnis handelt, bevor folgende Arbeitsschritte auszuführen sind:

für Paare:

- Überprüfung auf Kriterienkonformität bimolekularer transformations-invarianter Eigenschaften, die sich für die jeweilige Anordnung der aktuellen Paar-Paar-Kombination ergeben. Diese Eigenschaften müssen jeweils aus den vorbereiteten monomolekularen Eigenschaften neu berechnet werden. Die Evaluierung hat alle Bindungsfähigkeiten als Akzeptor und/oder Donator eines jeden Atoms und alle Bindungseigenschaften bei mehrprotonigen Donatoren in jeglichen resultierenden Kombinationsmöglichkeiten zu berücksichtigen. Sie kann abgebrochen bzw. für eine alternative Kombination gestartet werden, sobald ein Selektionskriterium nicht erfüllt wird. Eine kriterienkonforme Paar-Kombination wird als Kandidat bezeichnet.
- Festlegung und Sicherung der erforderlichen Transformationen zur Molekülorientierung des Kandidaten.

für Triplets:

- Analoge Überprüfung der Kriterienkonformität bimolekularer transformations-invarianter Eigenschaften, getrennt für alle Dreiecksseiten. Eine vollständig kriterienkonforme Triplet-Kombination wird als Kandidat bezeichnet.

- Festlegung und Sicherung der erforderlichen Transformationen zur Molekülorientierung des Kandidaten.
9. Da die Translationen der Kandidaten zuvor bei der Orientierung als eigenständiges positives Ergebnis dem entsprechenden Rotationsrepräsentanten zugeordnet worden sind, müssen nun noch die Translationscluster für jeden diskreten Rotationscluster ermittelt werden.
  10. In der Ergebnisbildung werden die entstandenen Transformationscluster qualifiziert, d.h. entsprechend ihren Kandidaten-Belegungszahlen bewertet.
  11. Ergebnisausgabe und Programmabschluss. An dieser Stelle erfolgt gegebenenfalls die Auswertung zur Referenzwiederfindung, wenn zuvor vermerkt wurde, ob ein Kandidat einer Eintragung der Referenzenliste entspricht.

#### 8.5.2 Ausführungen zu den möglichen Anordnungen einer Paar-Paar- bzw. Triplet-Triplett-Kombination

Alle Paare bzw. Triplets wurden in der Molekülaufbereitung nichtredundant mit streng aufsteigenden Atom-Identifikationsnummern erzeugt. Dieses muss in der Dockingsimulation berücksichtigt werden, da *in natura* nicht immer die Atome niedrigerer bzw. höherer Identifikatoren miteinander eine Wasserstoffbrückenbindung eingehen.

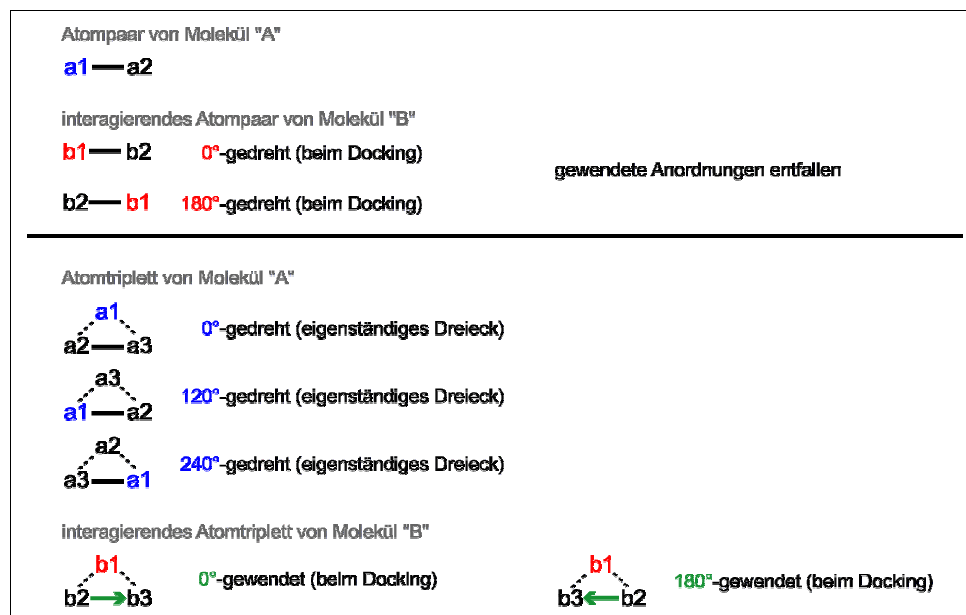
für Paare:

Betrachtet man zwei Atompaare verschiedener Moleküle "a1-a2" und "b1-b2", so müssen zwei mögliche Anordnungen der Atominteraktionen in ihrer Kombination berücksichtigt werden, wie der Abbildung III-8 zu entnehmen ist. Dieses entspricht der 180°-Rotation des Vektorkonstruktes um den Mittelpunkt des "b1-b2"-Verbindungsvektors. Die Werte der einzelnen transformations-invarianten Eigenschaften bleiben dabei unverändert, abgesehen von der Vertauschung der Bezeichner  $\beta_1$  und  $\beta_2$ , ihren entsprechenden Außenwinkeln sowie der Richtung des Grundlinienvektors. Obwohl einige der Evaluierungsberechnungen der beiden Anordnungen identische Ergebnisse erbringen, wurde keine gemeinsame Bewertung durchgeführt, um Übersichtlichkeit und Wiederverwertung universeller Programmroutinen zu gewährleisten.

für Triplets:

Analog obiger 180°-Rotation der Paare gibt es für die Triplets zusätzlich zur Ursprungsanordnung zwei 120°-Drehungen um den Mittelpunkt des "b1-b2-b3"-Dreiecks. Diese werden jedoch anders als bei den Paaren bereits beim Einlesen der Triplettlisten (allerdings für Molekül "A", was aber der identischen Konsequenz entspricht) als eigenständige Dreiecke erfasst und hier nun abgearbeitet.

An dieser Stelle kommt jedoch eine um  $180^\circ$  gedrehte Anordnung bezüglich der Dreiecksebene hinzu, wenn nicht in der Molekülaufbereitung eindeutig entschieden werden konnte, welche Dreiecksseite in der Moleküloberfläche zum Moleküläußeren weist. Diese Anordnung entfällt für die Paare aufgrund der gegenseitigen Rotationssymmetrie um die parallelen Grundlinien. So ergeben sich also für die Atomdreiecke "a1-a2-a3" und "b1-b2-b3" insgesamt sechs alternative, kombinatorische Anordnungsmöglichkeiten, wie sie ebenfalls in Abbildung III-8 nachvollzogen werden können. Auch hierbei bedarf es nur der richtigen Umbenennung der einmalig berechneten monomolekularen transformations-invarianten Eigenschaften. Die entstehenden Paar-Evaluierungen der Dreieckskanten sind jedoch je einmal  $180^\circ$  gedreht und in den anderen beiden Fällen völlig neuartig und müssen daher stets vollständig evaluiert werden.



**Abbildung III-8:** Schematische Darstellung aller möglichen Anordnungen einer Paar-Paar- bzw. Triplet-Triplett-Kombination, die zu berücksichtigen sind, damit alle potentiellen Atomkontakte evaluiert werden. Weitere Details im Text.

## 9 Auswahl der Test-Datensätze

Datengrundlage für alle Untersuchungen bilden ausgewählte Dateien der Protein-Datenbank. In ihr sind zum 24. April 2001 für 11750 Proteineinträge die Raumkoordinaten und Bezeichnungen aller Atome enthalten, deren Positionen aus den Strukturlösungen von Röntgenstruktur- und NMR-Daten bestimmt werden konnten. Eine noch unpublizierte Untersuchung [Vakser & Sali, 2001] hat hieraus 473 nichtredundante Protein-Protein-Komplexe isoliert, welche als Vorauswahl übernommen wurde.

Verwendete Proteinkomplexe müssen folgenden Kriterien genügen:

- Die Struktur wurde durch Röntgenanalyse bestimmt, um Einheitlichkeit der Untersuchungsobjekte und Testbedingungen sicherzustellen.
- Zwei oder mehr sequenzverschiedene Ketten sind in der asymmetrischen Kristalleinheit enthalten, um Kontaktflächen zwischen Heterooligomeren zu untersuchen und Kristallkontakte auszuschließen.
- Eine Polypeptidkette wird von 30 oder mehr Aminosäuren gebildet, um sicherzustellen, faltungsfähige Proteinmoleküle zu bearbeiten.
- Die Datensätze besitzen eine Auflösung von 2,5 Å oder besser, da die Qualität der Positionsbestimmung die Genauigkeit geometrischer Größen direkt beeinflusst.
- Die Ketten sind nicht durch Disulfidbrücken miteinander verbunden, um Proteinkomplexe zu bearbeiten, bei denen die Faltung der getrennten Moleküle in Lösung möglichst ähnlich der Form im Komplex ist und bei denen die Molekülanordnung ausschließlich durch nichtkovalente Wechselwirkungen zustande kommt. Dieses Kriterium besteht zwar auch für die obligatorischen Protein-Protein-Komplexe, auf die Besonderheiten ihrer Faltung und der eventuellen Instabilität der Einzelketten in Lösung wurde jedoch bereits hingewiesen.

Diese Auswahl ist möglichst nichtredundant. Bei mehrfachen Strukturlösungen identischer oder in offensichtlich hohem Maße homologer Komplexe wird der Datensatz mit der besseren Auflösung, bzw. der Datensatz mit der Wildtyp-Variante der Proteine aufgenommen.

Soweit erfüllen etwa 100 Protein-Protein-Komplexe diese Bedingungen. Eine weitere, manuelle Selektion eliminiert hieraus jegliche Komplexe, an denen ein Immunoglobulin beteiligt ist oder welche trotz ausreichend differierender Kettenlänge dennoch als Homodimere in der Kurzbeschreibung und im Sequenzvergleich gekennzeichnet sind.



Ebenfalls anhand der Kurzbeschreibung, zudem jedoch durch eine visuelle Kontrolle können die verbleibenden Datensätze in zwei Gruppen eingeteilt werden: Eigentlichere Untersuchungsgegenstände sind Komplexe, deren Einzelmoleküle sich voraussichtlich unabhängig voneinander falten können. Sie können somit stabil in einer Plasmaumgebung vorliegen, wo auch deren Zusammentreffen und richtige Anordnung zur produktiven Interaktion ohne weitere Hilfsmaßnahme stattfindet. Hiervon strikt zu differenzieren sind Komplexe, die entweder aus einer posttranslationalen, proteolytischen Spaltung hervorgehen oder bei denen die Ketten zwar getrennt synthetisiert, diese jedoch direkt bei der gekoppelt organisierten Translation, gegebenenfalls unter Hilfestellung von faltungshelfenden Chaperonen miteinander zu einem permanenten Komplex vereint werden. Derartige Komplexe sind im Fall der Propeptidkomplexe nur temporäre Zustände in der Reifung des beteiligten Proteins oder im anderen Fall zumeist obligatorische Komplexe, deren Einzelkomponenten ansonsten instabil und funktionsuntüchtig sind. Solche Datensätze wurden identifiziert und in einer eigenständigen Untersuchungsgruppe identisch zu den restlichen Komplexen behandelt. Ebenfalls mit in diese Gruppe aufgenommen wurden fünf Komplexe aus vorhergehenden Arbeiten [Krämer 1998], die dieser Gruppe zuzuordnen sind, wenn sie auch nicht in der Auflistung von Vakser und Sali [Vakser & Sali, 2001] aufgeführt werden.

Somit werden zwei Testgruppen kokristallisierter Nichtimmunoglobulin-Komplexe definiert als Gegenstand der folgenden Untersuchungen:

- 62 fakultative Protein-Protein-Komplexe
- 16 obligatorische Protein-Protein-Komplexe

Die Vertreter dieser Untersuchungsgruppen sind in der Tabelle des Anhangs A4 zusammen mit den betrachteten Polypeptidkettenidentifikatoren, den erzielten Auflösungen der Röntgenstruktur und den Kurzbeschreibungen aus dem PDB-Eintrag aufgeführt.

Eine weitere Reduktion der Datensatz-Anzahl erfolgt nach Bewertung der Komplexe in Bezug auf ihre Eignung für die behandelte Methode anhand der Betrachtung ihrer Anzahl intermolekularer Wasserstoffbrückenbindungen. Ebenso wird für die Eignung betrachtet, ob und wie viele kriterienkonforme Paar- und Triplettkombinationen sich aus diesen Bindungen unter Anwendung der verschiedenen Ensembles von Selektionsgrenzwerten erstellen lassen. Zusammen mit einer Charakterisierung der Komplexe wird dieses nach der Vorstellung der verwendeten Selektionskriterien im Ergebnissteil dargelegt. Unter Vorwegnahme dortiger

Ausführungen soll hier nur erwähnt werden, dass unter den leichtesten Selektionsbedingungen für das Paar-Docking 58 und für das Triplett-Docking 37 der anfangs 62 fakultativen Proteinkomplexe und beide Male alle 16 obligatorische Proteinkomplexe sinnvoll lauffähig sind.

## **10 Festlegung der Ensembles von Kriteriengrenzwerten und Identifikation der Referenzergebnisse**

Die Zusammenstellung der Ensembles von Kriteriengrenzwerten, die zur Evaluierung der monomolekularen und bimolekularen transformations-invarianten Eigenschaften verwendet werden, erfolgt aus der Betrachtung real existenter Werte in den entsprechenden Vektorkonstrukten aus intermolekularen Wasserstoffbrückenbindungen strukturgelöster Proteinkomplexe. Die Anwendung der Ensemble auf diese Komplexe erzeugt die kriterienkonformen Referenzergebnisse, welche in der eigenen Dockingvorhersage dann als richtig positive Ergebnisse wiederzufinden und aufzuzeigen sind.

### **10.1 Real existente Werte für die transformations-invarianten Eigenschaften**

Die Untersuchung nativer, strukturgelöster Proteinkomplexe im Hinblick auf die geometrischen Eigenschaften ihrer intermolekularer Wasserstoffbrückenbindungen und deren Paar- und Tripletzzusammenstellungen dient der Herleitung entsprechender, realitätsbezogener Grenzwerte, die in den eigenen Dockingexperimenten verwendet werden können. Unter Anwendung des Programms HBPLUS werden für die Test-Datensätze jegliche intermolekularen Wasserstoffbrückenbindungen, die den Kriterien von Baker und Hubbard gehorchen, zusammen mit den hierfür erforderlichen Protonpositionen erzeugt. Aus diesen Einzelbindungen lassen sich nichtredundant alle möglichen Paar- und Triplet-Kombinationen zusammenstellen. Die hier untersuchten transformations-invarianten Eigenschaften lassen sich für diese Kombinationen anhand der nativen Positionen aller beteiligten Atome, welche über Kettenbezeichner, Residuenummer, Aminosäure- und Atombezeichner identifiziert und zudem als Akzeptor oder Donator charakterisiert werden, sowie der vorgegebenen Protonpositionen und eigenen Akzeptorrepräsentationen berechnen, gegebenenfalls selektieren und in Listenform ausgeben.

### **10.2 Zusammenstellung der verwendeten Ensembles von Kriteriengrenzwerten**

Die Häufigkeitsverteilungen des Auftretens realer Eigenschaften in beliebig bestimmbarer Wertebereichen jeglicher untersuchter geometrischen Eigenschaften aus den zuvor erstellten Listen lassen sich einfach in Histogrammdarstellung visualisieren und auswerten. Ziel ist es, sinnvolle und effektive Selektionsgrenzwerte für die Wertebereiche der einzelnen Eigenschaften zusammenzustellen. Hierbei können folgende Typen von Grenzwerten unterschieden werden:

- Ein "direkter, expliziter Grenzwert" beschreibt die Wertebegrenzung eines Parameters durch einen explizit festgelegten Minimum- und/oder Maximumschwellewert, der die

natürliche Verteilungskurve für diese Eigenschaft z.T. recht abrupt begrenzt. Als Beispiel sei der Atomabstand aufgeführt, der in einzelnen Komplexen bis zu 70 Å erreichen kann, dann jedoch nicht mehr ein Paar von Bindungen innerhalb einer einzelnen, durchgehenden, lokalen Kontaktfläche beschreibt, sondern völlig voneinander unabhängige Kontaktpunkte, deren Interaktionsvektor-Geometrien keinerlei gekoppelten Abhängigkeiten unterliegen. Die selektierende Wirkung derartiger Grenzwerte kann enorm sein, soll jedoch eine universelle Verwendbarkeit der Methode gewährleisten, die auf der Existenz eingeschränkter Wertebereiche durch sterische Zwänge beruht.

- Ein "natürlicher Grenzwert" liegt vor, wenn die Verteilungskurve eines Parameters einen eigenen, natürlichen Minimum- und/oder Maximumwert besitzen und dazwischen einen normalen Verlauf zeigt. Dieser Verlauf kann indirekt beeinflusst sein durch Manipulationen an Grenzwerten korrelierender Eigenschaften. An sich hat eine derartige Grenze keine selektierenden Auswirkungen bei der Anwendung auf die Daten, aus der die Verteilungskurve selber erstellt wurde. Bei der Anwendung auf andere Datensätze und bei der Selektion "unnatürlicher" Geometrien besitzen sie dennoch ihre berechtigte Wirkung und werden daher implementiert.
- Ein "indirekter, impliziter Grenzwert" beschreibt die Wertebegrenzung eines Parameters, die durch die direkte oder natürliche Begrenzung eines stark korrelierten Parameters aufgrund der zugrunde liegenden Geometrien erzwungen wird. So kann z.B. die Differenz der Atomabstände mathematisch nicht größer sein als die Differenz von zugelassenen Maximal- und Minimalwerten der Abstände selbst. Diese Grenzwerte können defensiv oder aggressiv festgelegt werden. Defensiv entsprechen sie im Wesentlichen natürlichen Grenzwerten, welche eventuell den Einflüssen direkter Grenzen unterworfen sind. Sie besitzen dann gegebenenfalls keine eigene Selektionswirkung oder filtern nur vereinzelte Werteausreißer. Aggressiv festgelegt beschneiden sie jedoch zunehmend den gültigen Bereich theoretisch möglicher Werte. Dieses ist z.B. sinnvoll bei der Kombination zweier maximaler oder zweier minimaler Abstände, die jeweils eine geringe Differenz erzeugen, wohingegen das Zusammentreffen eines minimalen und eines maximalen Abstandes mit einer resultierenden großen Differenz zwar kombinatorisch in der Evaluierung möglich, allerdings im Sinne der Selektion zweier zeitgleich existenter Wasserstoffbrückenbindungen unsinnig ist.

Das Vorgehen bei der Zusammenstellung verwendeter Kriteriengrenzwerte wurde in vorhergehenden Arbeiten [Krämer, 1998] bereits erläutert. Es orientiert sich zudem eng an Grenzwertkombinationen vergleichbarer Arbeiten über die Verwendung transformations-

invarianter Eigenschaften im Geometrischen Docking [Wolfson & Nussinov, 1998], der grundlegenden Publikation der hier behandelten Methode [Meyer, Wilson, Schomburg, 1996] und der vorhergehenden Diplomarbeit [Krämer, 1998]. Es handelt sich im Wesentlichen um einen sequenziellen aber rückgekoppelten Prozess aus Beobachtung der natürlichen Werteverteilungen, der expliziten Eingrenzung elementarer Eigenschaften, die der Beschreibung des Suchziels der zugrunde liegenden Methode dienen und der folgenden, natürlichen und indirekten Anpassung übriger Eigenschaften, um die Selektionskraft der verwendeten Parameter auszunutzen. Der Prozess unterliegt dabei einer empirischen Kontrolle anhand der Überprüfung von Anzahl und Qualität der Ergebnisse für die nativen Proteinkomplexe, die als Referenzergebnisse identifiziert werden und in der eigenen Dockingvorhersage wiederzufinden sein müssen. Entsprechend diesen empirischen, rückgekoppelten Prozessen wurden auch die Grenzwerte ergänzender Eigenschaften bestimmt, die nur Atomtripletts betreffen.

Das Suchziel der Methode für die Zusammenstellung der Kriteriengrenzwerte soll durch folgende Argumentationsfolge beschrieben sein:

- Gesucht werden eine mengenmäßig für den Rechenbedarf sinnvoll begrenzte Anzahl von Paaren bzw. Tripletts intermolekularer Wasserstoffbrückenbindungen, die lokal eingeschränkte, d.h auch ausreichend kleine Kontaktflächen erfassen und dennoch durch eine ausreichend große Anzahl von positiven Suchergebnissen beschreiben können, so dass gegebenenfalls eine universelle Anwendbarkeit der Methode gegeben ist. Diese wird primär durch eine gezielte, direkte Festlegung der gültigen Atomabstände und Grundlinienlängen angestrebt.
- Gesucht werden zudem möglichst gleichlange, starke Wasserstoffbrückenbindungen, die dargestellt werden durch möglichst antiparallel angeordnete Interaktionsvektoren bei parallelen Grundlinien. Die Parameter der Summe gegenüberliegender  $\beta$ -Winkel und der  $\tau$ -Winkeldifferenz können dieses am besten beschreiben, indem eine  $\tau$ -Differenz-Grenze sicherstellt, dass die Wirkvektoren in der räumlichen Tiefe nicht zu weit divergieren, und die  $\beta$ -Summe die verbleibende Spreizrichtung testet und komplementäre Eigenschaftskombinationen identifiziert.
- Der  $\alpha$ -Winkel ist hingegen ein Maß für die Parallelität der Wirkvektoren eines Paares von Wasserstoffbrückenbindungsstellen. Das Kriterium bestimmt, wie viel Wert auf die Existenz gleichgerichteter Wasserstoffbrückenbindungen zur kooperativen Bindung gelegt wird.

Die bedeutendsten Kriterien sind somit die monomolekularen Atomabstände und Grundlinienlängen sowie deren jeweilige bimolekulare Differenzen, die  $\alpha$ -Winkel, Summen gegenüberliegender  $\beta$ -Winkel und die  $\tau$ -Winkel mit ihren Differenzen.

### 10.3 Erstellung und Wiederfindung der Referenzergebnisse

Indem man für jeden Test-Datensatz die real existierenden bzw. von HBPLUS postulierten, intermolekularen Wasserstoffbrückenbindungen identifiziert, aus ihnen die möglichen Paare und Triplets von Bindungen erstellt und die resultierenden, mono- und bimolekularen transformations-invarianten Eigenschaften der Selektion anhand der zusammengestellten Ensembles festgelegter Kriteriengrenzwerte unterwirft, erhält man für jeden speziellen Proteinkomplex eine Liste kriterienkonformer Interaktionskombinationen. Diese Referenzergebnisse gilt es als positive Ergebnisse in der eigenen Dockingsimulation der entsprechenden Proteinmoleküle mit den identischen Grenzwerten wiederzufinden. Sie sind dann weiter mit den richtigen Transformationsvorschriften zu charakterisieren und in ihrer Bedeutung als die einzigen richtig positiven Ergebnisse bzw. die einzig richtige Molekülorientierung herauszustellen.

Die Listen der Referenzergebnisse eines Molekülkomplexes sind für ein Ensemble an Kriteriengrenzwerten einmalig anzulegen und mit Hilfe der eindeutigen Atomidentifikatoren abzuspeichern. Sie können beim Docking dieser Moleküle eingelesen werden, um zu beliebigen Zeitpunkten der Prozesse von Selektion, Charakterisierung, Ergebniserstellung und -ausgabe die Wiederfindung, Erhaltung und qualitative Einschätzung der Endergebnisse zu ermöglichen.

### 10.4 Bedeutungen und Auswirkungen zu differenzierender Typen von Selektionskriterien

Es gilt zu beachten, dass zu der Identifikation richtig positiver Referenzergebnisse lediglich die mono- und bimolekularen Selektionskriterien herangezogen werden. Die übrigen zuvor beschriebenen Selektionsschritte zur Identifikation bindungsfähiger Atome, zur Bewertung der Vektordivergenz bei Atomtriplets, zur Einschätzung der Orientierungsfähigkeit kriterienkonformer Kandidaten und zur weiteren, später beschriebenen Ergebnisverarbeitung in der Ergebnisclusterung betreffen ausschließlich die Evaluierung vor und in der eigenen Dockingsimulation. Sie wirken präselektiv in der Zusammenstellung der Molekülrepräsentation und legen postselektiv fest, ob ein kriterienkonformes Atomensemble ausreichende Orientierungsfähigkeit aufweist und in der weiteren Ergebnisfindung erhalten bleibt.

Es muss daher im späteren Verlauf unterschieden werden nach den Auswirkungen einzelner Selektionskriterien auf die Evaluierungsquote, die Wiederfindungsquote und die Erhaltungsquote der Referenzergebnisse:

- Die **Evaluierungsquote** beschreibt, ob ein Atomensemble überhaupt in die Molekülrepräsentation aufgenommen, im Docking mit den richtigen Partnern und mit den richtigen Atom-Atom-Paarungen kombiniert und dann der Selektion bimolekularer transformations-invarianter Eigenschaften zugeführt wird.
- Die **Wiederfindungsquote** beschreibt, wie viele der richtig positiven Zusammenstellungen von Atomensembles diese Selektion bimolekularer transformations-invarianter Eigenschaften erfolgreich bestehen.
- Die **Erhaltungsquote** beschreibt, wie viele der richtig positiven Ergebnisse im Verlauf der weiterführenden Ergebnisermittlung erhalten bleiben und dann in der endgültigen Ergebnisausgabe des Dockingprogramms mit der richtigen Transformationsvorschrift und einer ausreichenden Qualifizierung versehen werden. Erhaltungsquoten können an mehreren Zwischenschritten ermittelt werden und ermöglichen dann eine Einschätzung der Rotations- und Translationsbestimmung, verschiedener Stufen der Ergebnisclusterung und der abschließenden Ergebnisqualifizierung.





## IV Ergebnisse und Diskussion

An dieser Stelle beginnen die Vorstellung und Diskussion der Ergebnisse dieser Arbeit. Die Begründung für die ungewöhnliche Folge der einzelnen Abschnitte und weshalb Präsentation und Diskussion zusammen ausgeführt werden, ist identisch. Die Wahl einiger Parametergrößen insbesondere in den Methodenteilen außerhalb der eigentlichen Selektion mono- und bimolekularer transformations-invarianter Eigenschaften, stellt sich als überaus abhängig von der Qualität einzelner Arbeitsgänge heraus. Zudem erfordert die Einzelpräsentation für Teilabschnitte eine abstrakte Form anhand relativer Entwicklungen und Qualitätsbewertungen. Daher führt es zu großen Irritationen, wenn man der Chronologie der Arbeitsschritte und ihrer Teilergebnisse nicht folgen kann. Hierzu gehört das Verständnis der Begründung getroffener Entscheidungen, welches nur durch die verknüpfte Diskussion erzeugt werden kann.

Die Abfolge der Ergebnispräsentation erfolgt ähnlich rückgekoppelt wie die Methodenentwicklung:

- Zunächst sollen die gewählten Grenzwerte dokumentiert werden. Hierbei wird chronologisch mit der Molekülaufparametrisierung und dem eigentlichen Docking inklusive der Arbeitsschritte zur Orientierungsbestimmung und Ergebnisqualifizierung vorgegangen.
- Erst anschließend erfolgt die Vorstellung der Test-Datensätze, da sich aus der Anwendung der Grenzwerte mono- und bimolekularer transformations-invarianter Eigenschaften hierfür die Anzahlen erzielbarer Referenzergebnisse ergeben. Diese wiederum ermöglichen eine Betrachtung der Eignung unterschiedlicher Komplexe für die vorgestellten Methoden und Parametergrößen.
- Ausführungen zu den Problematiken der Kombinatorik vorgestellter Methoden und der Spezifität von Suchergebnissen erfolgen hier anhand der konkreten Ergebnisse für die vorgestellten Test-Datensätze. Sie dienen damit dem Verständnis der Diskrepanz zwischen grundlegender Selektionsanforderungen und realen Untersuchungsbedingungen. Hierdurch begründen sich notwendige Einbußen der Ergebnisfindung und

-weiterverarbeitung in Zusammenhang mit der Qualität und den Anforderungen der umfassenden Endergebnisse.

- Die mitentscheidende Qualität einzelner Arbeitsabschnitte zeigt kritische Methoden und Implementierungen auf, die mögliche Ansatzpunkte zur Verbesserung der vorgestellten Ergebnisse liefern.
- Die Präsentation der Endergebnisse, wie sie aus allen Arbeitsschritten heraus dem Anwender ausgegeben werden, ermöglicht eine Einschätzung der Gesamtleistung aller vorgestellten Methoden. Sie verhilft zu einer Bewertung der Qualität und lässt so Rückschlüsse auf die generelle Anwendbarkeit zu, wenn unabhängig voneinander strukturgelöste Einzelmoleküle diesen Untersuchungen unterzogen werden sollen, ohne dass das richtige Ergebnis bekannt ist.
- Hiervon abgekoppelt erfolgt die Betrachtung der benötigten Laufzeiten der Dockingvorgänge unter Berücksichtigung der Vielzahl von Methoden- und Parameterkombinationen. Auch sie ist bedeutend für die individuelle Anwendbarkeit der Methodener als eventuell zeitkritischer Vorfilter einer eigentlichen Hauptuntersuchung.

## 11 Festlegung selektierender Grenzwerte

Aus der Analyse strukturgelöster Protein-Protein-Komplexe bezüglich real existenter Werte für die betrachteten Kriterien werden Grenzwerte für die einzelnen Selektionsschritte folgender drei Wirkungsbereiche festgelegt, die separat diskutiert werden:

- Identifikation bindungsfähiger Atome  
(“präselektiv“, beeinflusst die Evaluierungsquote der Referenzergebnisse)
- Identifikation repräsentierender Atumpaare und -tripletts anhand monomolekularer transformations-invarianter Eigenschaften  
(“präselektiv“, beeinflusst die Evaluierungsquote der Referenzergebnisse)

Identifikation potentiell produktiver, intermolekularer Kombinationen von Atumpaaren und -tripletts anhand bimolekularer transformations-invarianter Eigenschaften  
(beeinflusst die Wiederfindungsquote der Referenzergebnisse)

- Orientierungsabhängige Einschätzung und Qualifizierung kriterienkonformer Kandidaten  
(“postselektiv“, beeinflusst die Erhaltungsquote der Referenzergebnisse)

Soweit nicht anders ausgeführt, liegen den Diagrammen und Tabellen dieser Dokumentation stets die Datensätze der obligatorischen Protein-Protein-Komplexe zugrunde. Diese werden der Aufbereitung und den Dockingexperimenten mit Atomtripletts unter Verwendung aller bindungsfähigen Atome anhand des weiten Grenzwertensembles unterzogen.

Angaben zu den resultierenden Anzahlen bindungsfähiger Atome, repräsentierender Atumpaare und -tripletts sowie sich daraus ergebender Evaluierungsaufrufe finden sich bei der Beschreibung der Testdatensätze und Komplextypen.

Generell kann hier bereits die Bedeutung der Auswahlsschritte nachvollzogen werden: Je mehr Atome als bindungsfähig eingestuft werden, desto mehr repräsentierende Atumpaare und -tripletts lassen sich hieraus für ein Molekül zusammenstellen und desto mehr intermolekulare Kombinationen dieser Atomensembles müssen evaluiert werden. Direkt proportional zur Anzahl der Evaluierungen steigen zum Einen die Anzahl falsch positiver kriterienkonformer Ergebnisse, was Identifikation und Qualifizierung der richtig positiven Ergebnisse erschwert und zum

Anderen direkt gekoppelt die benötigte Programm-Laufzeit. Demzufolge wurde auch zusätzlich eine Nachselektion kriterienkonformer Ergebnisse anhand der Qualität ihrer Orientierungsbestimmung eingeführt. Ziel muss es sein, zu jedem Zeitpunkt im Gesamtablauf der Methoden die Spezifität der Ergebnisse nicht unnötig herabzusetzen durch die vermeidbare Qualifizierung ungeeigneter Interaktionspartner. Hierbei muss –den Ergebnissen vorweggreifend– in Kauf genommen werden, dass makellose Evaluierungs-, Wiederfindungs- und Erhaltungsquoten von 100% nicht erreicht werden können. Nur durch die Herabsetzung der selektiven Kraft eines jeden Kriteriums bis zur Unwirksamkeit könnte in einer unüberschaubaren Menge an Ergebnissen sicher auch jedes richtig positive Ergebnis erhalten werden. In allen Suchproblemen, die ein gewisses Unschärfemoment enthalten müssen, ist dieses Problem hoher Wiederfindungsquoten gegenüber erzielbarer Ergebnisspezifität (*recall versus specificity*) enthalten.

## 11.1 Identifikation bindungsfähiger Atome

### 11.1.1 Wasserstoffbrückenbindungsfähige Atomtypen

Generell muss ein Atom Wasserstoffbrückenbindungseigenschaften aufweisen können, was eine Abfrage entsprechend der standardisierten Tabelle im Anhang A2 sicherstellt. Die Tabelle enthält die Atomfunktionen, welche auch HBPLUS zugrunde legt, wobei Sulphydrylgruppen und aromatische Akzeptoren in der Erstellung der Referenzergebnisse und in der Molekülaufbereitung aus der Berücksichtigung ausgeschlossen werden, so dass an dieser Stelle keine Verluste durch Fehlcharakterisierung auftreten.

### 11.1.2 Nähe zur lösungsmittelzugänglichen Moleküloberfläche

Atome gelten als ausreichend oberflächennah, wenn sie selber Connolly-Oberflächenpunkte besitzen und somit Teil der Oberfläche sind. Um darzustellen, dass keine Atome essenzieller Bindungen bei diesem Vorgehen übergangen werden, weil sie ohne eigene Oberflächenpunkte zwischen zwei Nachbaratomen versenkt in der Oberfläche liegen, werden in Tabelle IV-1 die Auswirkungen verschiedener Zulassungen präsentiert. Qualifiziert werden hier testweise auch alle Atome, die im Abstand von 2,0 Å oder 3,0 Å irgendeinen Oberflächenpunkt aufweisen können.

	eigener Oberflächenpunkt	Oberflächenpunkt ≤2,0 Å entfernt	Oberflächenpunkt ≤3,0 Å entfernt
∅# qualifizierte Atome	244	249	251
∅# repräsentierende Atomtripletts	7.714	8.190	8.455
∅# Evaluierungsaufrufe	44.935.057	49.635.864	52.509.426
# evaluierte Referenzergebnisse (von 923)	902	902	902
Evaluierungsquote	97,7 %	97,7 %	97,7 %

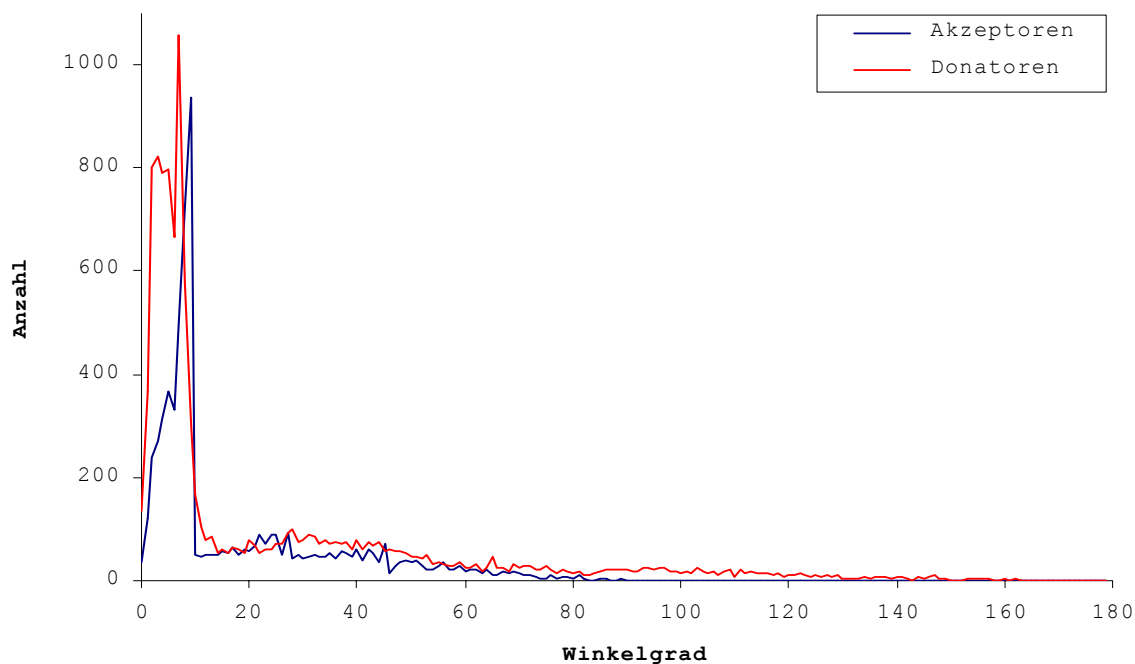
**Tabelle IV-1:** Parametervariation der notwendigen Nähe bindungsfähiger Atome zu Punkten der lösungsmittelzugänglichen Moleküloberfläche. Weitere Details im Text.  
(∅# - durchschnittliche Anzahl, # - Gesamtanzahl)

Es können hier keinerlei Auswirkung auf die Evaluierungsquote aufgezeigt werden, da stets 97,7% der Referenzergebnisse evaluiert werden. Die Anzahlen gebildeter Atomtripletts und ihrer kombinatorischen Evaluierungsaufrufe steigt jedoch wirkungslos und daher ohne Notwendigkeit an.

### 11.1.3 Ausrichtung der Bindungscharakteristik

Die geeignet gerichtete Bindungsfähigkeit eines Atoms wird postuliert, wenn sein charakterisierender Interaktionsvektor weniger als 15° abweicht von dem Normalenvektor eines Oberflächenpunktes innerhalb eines 3,0 Å-Abstands vom Atomzentrum. Bei erfolgloser Suche wird angenommen, dass das Atom zwar ausreichend nahe der Oberfläche liegt jedoch aufgrund der bevorzugten Interaktionsrichtung unfähig zur Ausbildung einer spezifischen und stabilen intermolekularen Wasserstoffbrückenbindung ist und daher nicht weiter im Rahmen der hier entwickelten Methode berücksichtigt werden muss.

Das Histogramm in Abbildung IV-1 zeigt für die bislang qualifizierten Donatoren und Akzeptoren die kleinsten, erreichbaren Winkel. Diese weisen eine überdeutliche Präferenz für den Wertebereich unter 10° auf. Tabelle IV-2 führt dann die Vergleichswerte der Evaluierungsquoten für 10°, 15° und 20° an und verdeutlicht, dass eine Selektionseinschränkung auf diesen bevorzugten Bereich jedoch zu großen Abstrichen in der Evaluierung essenzieller Triplettkombinationen führt. Die Grenzerweiterung auf Winkel bis zu 20° bewirkt gegenüber 15° eigentlich nur einen 25%-Anstieg der Evaluierungsaufrufe ohne Quotengewinn.



**Abbildung IV-1:** Ausrichtung der Bindungscharakteristik: Die Verteilung real existenter minimaler Winkel der Interaktionsvektoren zu Oberflächennormalen benachbarter Punkte der lösungsmittelzugänglichen Moleküloberfläche.

	Winkel $\leq 10^\circ$	Winkel $\leq 15^\circ$	Winkel $\leq 20^\circ$
∅# qualifizierte Atome	235	244	252
∅# repräsentierende Atomtripletts	6.725	7.714	8.623
∅# Evaluierungsaufrufe	33.947.629	44.935.057	56.522.150
# evaluierte Referenzergebnisse (von 923)	874	902	903
Evaluierungsquote	94,7 %	97,7 %	97,8 %

**Tabelle IV-2:** Parametervariation der Interaktionsvektor-Ausrichtung bindungsfähiger Atome zu Normalenvektoren benachbarter Punkte der lösungsmittelzugänglichen Moleküloberfläche. Weitere Details im Text.  
(∅# - durchschnittliche Anzahl, # - Gesamtanzahl)

#### 11.1.4 Disqualifikation aufgrund intramolekularer Wasserstoffbrückenbindungen

Selbst derart qualifizierte Atome können optional wieder disqualifiziert und aus der weiteren Berücksichtigung ausgenommen werden, wenn für sie die Existenz intramolekularer Wasserstoffbrückenbindungen postuliert werden kann.

Tabelle IV-3 soll mit den Durchschnittswerten der obligatorischen Proteinkomplexe einen Einblick in die mengenmäßige Auswirkung dieses Vorgehens geben. Reduktionen repräsentierender Triplets um bis zu 50,7% und repräsentierender Paare um bis zu 36,2% können bewirkt werden. Dieses begründet im Weiteren noch deutlichere Abnahmen der kombinatorischen Evaluierungsaufrufe im eigentlichen Docking um bis zu 75,9% bzw. um bis zu 59,3%. Diese enormen Reduktionen lassen bereits Rückschlüsse auf die benötigten Laufzeiten zu. Spätere Erläuterungen werden die weitere positive Wirkung einer erheblichen Verbesserung des Verhältnisses richtig positiver zu falsch positiver Endergebnisse, darlegen.

Da es hierbei jedoch, wie ebenfalls später erörtert, auch zunehmend zum fehlerhaften Ausschluss essenzieller Atome mit Abnahmen der Evaluierungsquote auf bis zu 80,7% kommt, ist durchgehend der Einsatz aller drei Varianten zu diskutieren. Die bedeutend geringere Abnahme der Evaluierungsquoten gegenüber den Evaluierungsaufrufen und den hervorgehenden kriterienkonformen Ergebnissen ist jedoch eine beachtenswerte Demonstration, wie Ergebnisspezifität hier verbessert werden kann.

	alle Atome	ungesättigte Atome	ungebundene Atome
∅# qualifizierte Atome	244	224	199
∅# repräsentierende Atomtripletts	7.714	5.205 (-32,5%)	3.806 (-50,7%)
∅# Evaluierungsaufrufe	44.935.057	20.429.602 (-54,5%)	10.822.723 (-75,9%)
Evaluierungsquote (von 923 Ref.)	97,7 %	89,3 %	84,7 %
∅# repräsentierende Atompaare	2.445	1.994 (-20,9%)	1.561 (-36,2%)
∅# Evaluierungsaufrufe	7.016.425	4.470.037 (-36,3%)	2.858.964 (-59,3%)
Evaluierungsquote (von 1253 Ref.)	98,3 %	89,3 %	80,7 %

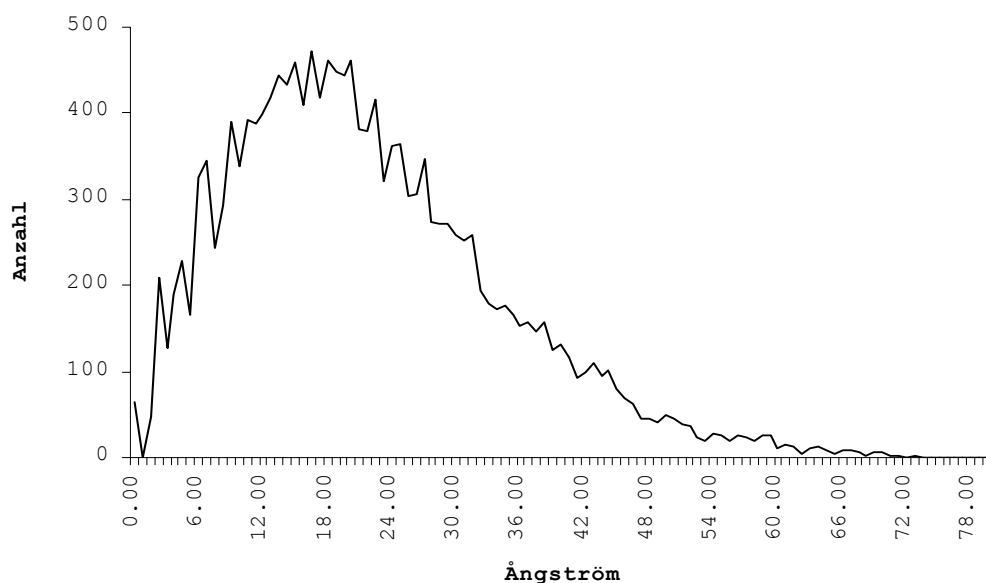
**Tabelle IV-3:** Parametervariation: Auswirkungen der optionalen Disqualifikation bindungsfähiger Atome aufgrund intramolekularer Wasserstoffbrückenbindungen. Weitere Details im Text.  
(∅# - durchschnittliche Anzahl)

## 11.2 Auswahl mono- und bimolekularer transformations-invarianter Eigenschaften

Ausgehend von dem Ensemble, das bereits bei vorherigen Arbeiten [Krämer, 1998] zusammengestellt und verwendet wurde, haben sich vorerst drei diskrete Zusammenstellungen von Selektionsgrenzwerten herauskristallisiert. Diese besitzen tendenziell abnehmende Selektionskraft, indem sie durch zunehmend erweiterte Gültigkeitsbereiche beschrieben werden und somit im weiteren Verlauf auch zunehmende Anzahlen kriterienkonformer Referenzergebnisse erzeugen.

### 11.2.1 Verteilung unselektiert auftretender Atomabstände

Betrachtet man zunächst einmal für alle berücksichtigten Komplexe die Verteilung der Atomabstände aller Paare, die sich aus den an intermolekularen Wasserstoffbrückenbindungen beteiligten Atomen eines Moleküls nichtredundant zusammenstellen lassen, so ergibt sich die Verteilungskurve aus Abbildung IV-2.



**Abbildung IV-2:** Verteilung der unselektiert auftretenden Entfernungen zwischen Atomen nachgewiesener intermolekularer Wasserstoffbrückenbindungen innerhalb der Oberflächen von Einzelmolekülen obligatorischer Proteinkomplexe.

Es finden sich somit Abstände von 2 Å bis 75 Å mit einer deutlichen Betonung des Bereichs von 2 Å bis 35 Å und einer maximalen Auftretshäufigkeit zwischen 10 Å und 23 Å.

Die Ergebnisse, welche hier mit einem Atomabstand von 0 Å erfasst werden, sind keinesfalls Artefakte. Sie beschreiben den Umstand, dass entweder ein Akzeptoratom zeitgleich zwei

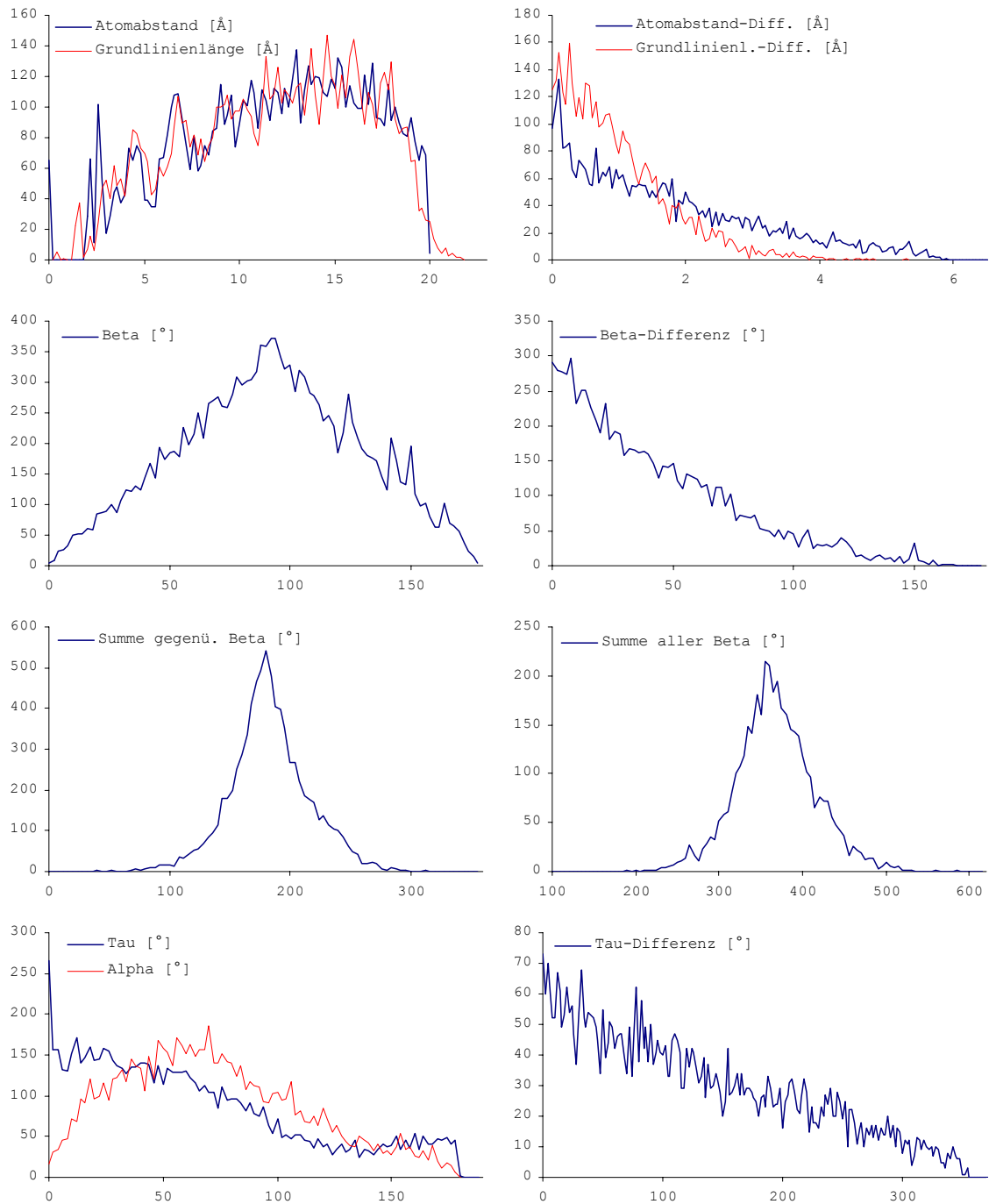


intermolekulare Wasserstoffbrückenbindungen zu unterschiedlichen Donatoren unterhält oder dass HBPLUS für ein Atom zwei Interaktionspartner mit akzeptablen Bindungskonditionen ausfindig gemacht hat und anzeigt. Da in beiden Fällen die Methode zur Bestimmung der Molekülorientierung nicht anwendbar ist, weil kein genügend definierter Grundlinienvektor aufzuspannen ist, werden diese Ergebnisse ignoriert und später durch eine entsprechende Untergrenze ausgeschlossen. Bei Atomabständen zwischen  $0 \text{ \AA}$  und  $2 \text{ \AA}$  muss es sich um fehlerhafte Proteinstrukturen handeln, da hiermit eine Penetration der van-der-Waals-Hüllen zweier Nachbaratome beschrieben würde. Dieses war jedoch bei keinem der hier berücksichtigten Test-Datensätze der Fall.

Obwohl eine initiale Beschränkung des Gültigkeitsbereichs auf z.B.  $10 \text{ \AA}$  bis  $23 \text{ \AA}$  die größte Anzahl an Referenzergebnissen für einen derartigen  $13 \text{ \AA}$ -Bereich zur Verfügung stellen würde, ist ein derartiger Schritt nicht Erfolg versprechend im Hinblick auf die universelle Anwendbarkeit und die Strategie der Selektion.  $13 \text{ \AA}$  entsprechen in etwa der Ausdehnung von drei gestreckten Peptidbindungen, wie sie in  $\beta$ -Strängen vorliegen würden. Zum Einen beschreiben  $10 \text{ \AA}$  bis  $23 \text{ \AA}$  eine kreisrunde, planare Fläche von  $314 \text{ \AA}^2$  bis  $1662 \text{ \AA}^2$ , die bereits einige fakultative Komplexe nicht vorweisen können (siehe Charakterisierung der Komplexe in Tabelle IV-11). Hierbei muss berücksichtigt werden, dass für einen derartigen Kreisradius insbesondere die Angaben molekularer Oberflächen aufgrund der halbkugelförmigen Atomoberflächen deutlich größer sind als für eine ebene Fläche. Zum Anderen ist der Abstand zu groß, als dass die Bindungsorientierungen der Atome einer ausreichend starken Kopplung unterliegen. Hierdurch werden die Wertebereiche, die für die übrigen geometrischen Größen eingenommen werden können, nur unzureichend eingegrenzt wie auch die nachfolgende Untersuchung verdeutlicht.

### 11.2.2 Grundlegende Tendenzen der Werteverteilungen unter Vorselektion der Atomabstände von $2 \text{ \AA}$ bis $20 \text{ \AA}$

Um eine Übersicht über die natürlichen Grenzen der Werteverteilungen aller Größen zu erhalten, erfolgt die Betrachtung bei einer ausschließlichen Vorselektion der Atomabstände von  $0 \text{ \AA}$  bis  $20 \text{ \AA}$ . Die Diagramme der Abbildung IV-3 verdeutlichen die Ergebnisse. Die darauf folgende Auflistung enthält eine kurze Charakterisierung der einzelnen Größen, wobei mit angegebenen Optimalwerten Bezug genommen wird auf die Idealvorstellung eines Vektorkonstrukts in Rechteckform.



**Abbildung IV-3:** Werteverteilung der mono- und bimolekularen transformations-invarianten Eigenschaften aller 7896 Paare aus Atomen intermolekularer Wasserstoffbrückenbindungen der obligatorischen Proteinkomplexe bei ausschließlicher **Vorselektion der Atomabstände von 0 Å bis 20 Å**. Als Funktionswerte ist auf der Ordinate stets die Anzahl der Wertebereiche in einem kleinen Wertintervall aufgetragen. Die Einheiten der Abszisse sind mit [Å] für Ångström und [°] für Winkelgrade jeweils im Histogramm angegeben.

Die Definitionen und Beschreibungen der mono- und bimolekularen Eigenschaften finden sich in den Kapiteln 7.2 und 8.1 bei der Vorstellung der verwendeten Repräsentationen und Vektorkonstrukte.

- Atomabstand kompletter Wertebereich 2 Å bis 20 Å,  
siehe vorherigen Abschnitt zu 0 Å-Vertretern
- Differenz der Atomabstände bevorzugt Optimum < 0,5 Å,  
linearer Abfall zum natürlichen 6,0 Å-Maximum
- Grundlinienlänge vergleichbar mit Atomabständen,  
maximal 2\*van-der-Waals-Radius über 20 Å hinaus
- Differenz der Grundlinienlängen vergleichbar mit der Differenz der Atomabstände,  
deutlichere Betonung < 2.0 Å entsprechend der Trapezform  
des Vektormodells mit geringerer Differenz an den  
Interaktionspunkten, nur noch vereinzelt > 4.0 Å
- Innenwinkel  $\beta_1, \beta_2, \beta_3, \beta_4$  Maximum beim 90°-Optimum,  
Abfall über den kompletten Wertebereich nach 0° und 180°
- Differenzen der Winkel  $|\beta_1 - \beta_2|$  und  $|\beta_3 - \beta_4|$  Maximum beim 0°-Optimum,  
Abfall über den fast kompletten Wertebereich bis etwa 160°
- Summen gegenüberliegender Innenwinkel  $\beta_1 + \beta_3$  und  $\beta_2 + \beta_4$  deutliches Maximum beim 180°-Optimum,  
auffallend schneller Abfall nach 100° und 260°,  
verflachender Auslauf bei 65° und 300°,  
der maximal mögliche Wertebereich wird NICHT genutzt
- Summe aller  $\beta$ -Innenwinkel deutliches Maximum beim 360°-Optimum,  
auffallend schneller Abfall nach 275° und 470°,  
verflachender Auslauf bei 200° und 520°,  
der maximal mögliche Wertebereich wird NICHT genutzt
- Torsionswinkel  $|\tau|$  kompletter Wertebereich 0° bis 180°,  
leichte Bevorzugung des 0°-Optimums
- Differenz der Torsionswinkel  $|\tau_1 - \tau_2|$  linearer Abfall über den kompletten  
Wertebereich 0° bis 360°
- Innenwinkel  $\alpha$  kompletter Wertebereich 0° bis 180°,  
bevorzugt 20° bis 110° und NICHT das 0°-Optimum

Drei wichtige Schlussfolgerungen können aus dieser Aufstellung gezogen werden, um sie in der anschließenden Festlegung der Grenzwertzusammenstellungen umzusetzen:

- Generell bestätigt sich die Vorstellung des idealen Vektorkonstrukts mit seinen antiparallel ausgerichteten Interaktionsvektoren durch die Bevorzugung der optimalen Wertebereiche der Längendifferenzen und aller Eigenschaften, die an die  $\beta$ -Innenwinkel geknüpft sind. Hinzu kommt durch den Vergleich der Differenzen der Atomabstände und der Grundlinienlängen, die Festigung der vorgeschlagenen Trapezform mit geringerer Differenz an den Interaktionspunkten.
- Übergreifende Kriterien können natürlich enger beschränkte Wertebereiche einnehmen, als die Werteverteilung der zugrunde liegenden Größen rechnerisch erlauben würde. So sind Minimum- und Maximumwerte der Summen gegenüberliegender  $\beta$ -Winkel deutlich um jeweils etwa  $60^\circ$  höher bzw. niedriger als sie aufgrund der von den einzelnen  $\beta$ -Winkeln angenommenen Werte des kompletten  $0^\circ$ -bis- $180^\circ$ -Wertebereiches sein müssten. Ebenso verhält es sich mit der Summe aller  $\beta$ -Winkel, die sogar noch enger bestimmt ist als sie sich aus den Summen gegenüberliegender  $\beta$ -Winkel ableiten ließe. Noch deutlicher fällt es in beiden Fällen bei der Betrachtung der abermals enger definierten bevorzugten Bereiche auf.

Es lässt sich ableiten, dass in diesen Fällen eine Bindungsstelle mit vergleichsweise schlechten Bindungsgeometrien kompensiert werden muss durch die Existenz einer gekoppelten Bindungsstelle mit umso besseren Eigenschaften. Dieses wird auch in vergleichbaren Arbeiten postuliert [Nussinov & Wolfson, 1991]. In der angestrebten Selektion kann dieses umgesetzt werden, indem die übergreifenden Kriterien entsprechend strikter selektiert werden als es die mathematischen Grenzen vorschlagen würden.

- Die Wertebereiche, die von den Größen  $\beta$ ,  $\tau$  und  $\alpha$  angenommen werden, umfassen uneingeschränkt die, von dem Modell her zur Verfügung stehenden Maximalbereiche. Besonders für den Torsionswinkel  $\tau$  mag dieses auf Anhieb überraschen. Es ist aber nachzuvollziehen, wenn man sich einen intermolekular in ein  $\beta$ -Faltblatt eingefügten  $\beta$ -Strang vorstellt. Die Existenz einer derartigen Interaktion wurde bereits mit Abbildung I-1 angedeutet. Hierbei können sich die beiden Wasserstoffbrückenbindungen eines Atompaares auf verschiedenen Seiten des Stranges ausbilden und somit einen exakten  $180^\circ$  Winkel zueinander einnehmen.

Eine Reduzierung dieser annehmbaren Wertebereiche kann nur bewirkt werden, wenn eine erhöhte Kopplung der Bindungsgeometrien beider Atome vorausgesetzt wird. Dieses erfolgt durch die Evaluierung ausschließlich näher benachbarter Bindungsstellen. Völlig unvorhersagbare geometrische Oberflächenbeschaffenheiten, die willkürliche Geometrien

des Vektormodells ermöglichen, müssen eliminiert werden, indem primär die Atomabstände und Grundlinienlängen sowie deren Differenzen deutlich enger als die hier betrachteten 0 Å bis 20 Å gefasst werden.

Diese Beobachtungen lassen sich in Einklang bringen mit den bereits zuvor geäußerten Argumentationen zur Wahl der Atomabstände. Unglücklicherweise schließt eine Längenbeschränkung den sehr ergebnisreichen Abschnitt von 10 Å bis 23 Å aus Abbildung IV-2 der Verteilung unselektierter Atomabstände im Folgenden aus.

### 11.2.3 Die drei Ensembles der Selektionsgrenzwerte

Vorangegangene Betrachtungen zeigen die Tendenzen auf, in denen grundlegende Grenzen gesetzt werden. Man verfolgt daraufhin die Auswirkungen auf die übrigen Größen und legt dann hierfür natürliche und indirekte Grenzwerte fest. Welche Selektionswirkung die in Tabelle IV-4 vorgestellten Grenzwertensemble auf die Generierung gültiger Referenzergebnisse für die untersuchten Protein-Protein-Komplexe haben, wird bei der Vorstellung der beiden Testgruppen dargelegt.

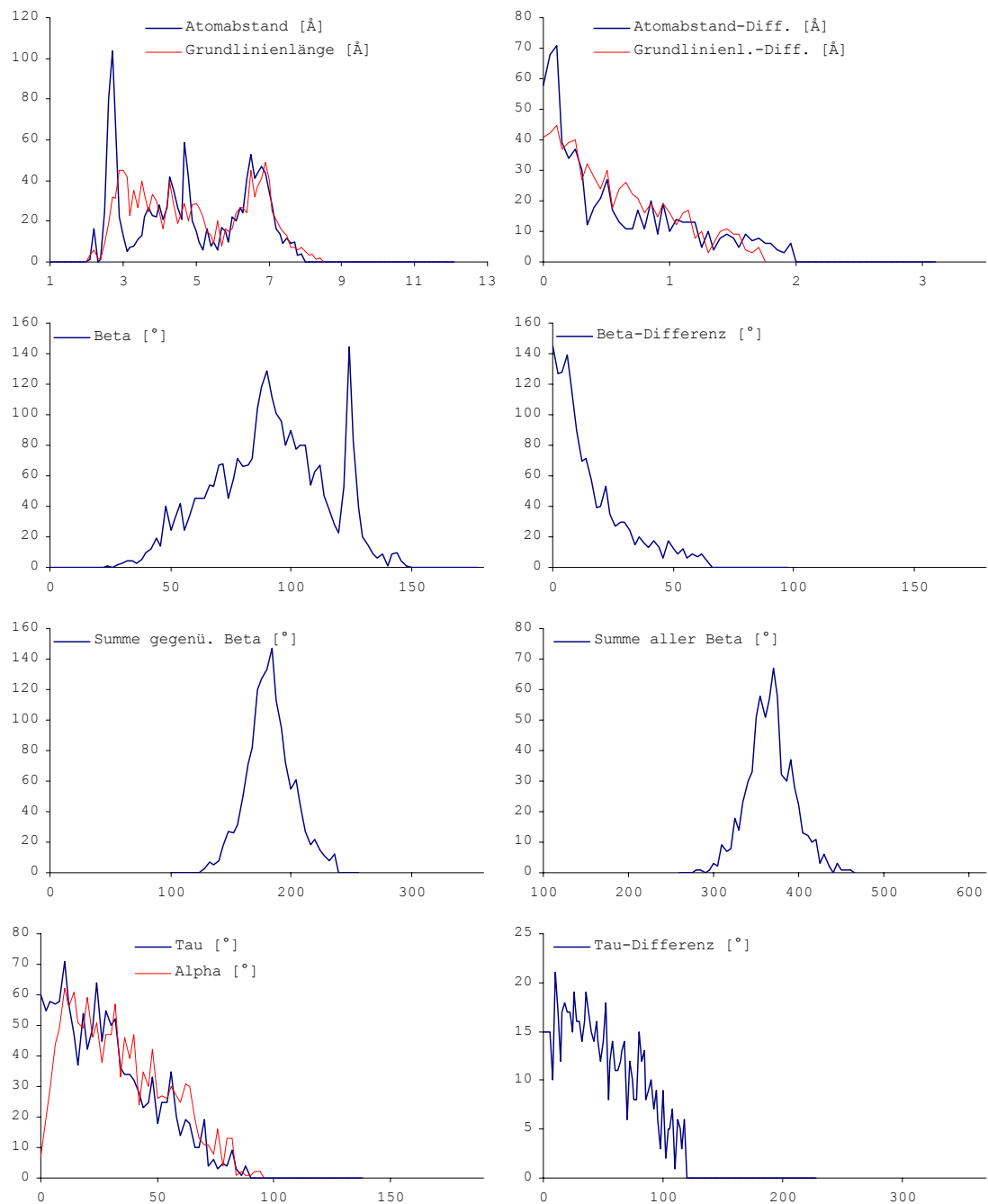
- Aus Orientierungs- und Vergleichsgründen wird das, in vorherigen Arbeiten [Krämer, 1998] entwickelte Grenzwertensemble übernommen und um nichtlimitierende, natürliche Grenzwerte für hinzugekommene Selektionsgrößen ergänzt. Es orientiert sich an Atomabständen von 2,0 Å bis 8,0 Å bei einer maximalen Abstandsdifferenz von 2,0 Å. Insgesamt besitzt es die engsten Grenzen und das größte Selektionsvermögen, weshalb es im Folgenden als **enges Grenzwertensemble** bezeichnet wird. Die Verteilungskurven für die geometrischen Größen aller verbleibenden Referenzergebnisse finden sich in Abbildung IV-4a.
- Das **mittlere Grenzwertensemble** wurde komplett neu bestimmt und orientiert sich an Atomabständen von 2,0 Å bis 10,0 Å bei einer maximalen Abstandsdifferenz von ebenfalls 2,0 Å. Einzig der Grenzwert für den Innenwinkel  $\alpha$  fällt 5° enger aus als beim engen Grenzwertensemble. Die Verteilungskurven für die geometrischen Größen aller verbleibenden Referenzergebnisse finden sich in Abbildung IV-4b.
- Das **weite Grenzwertensemble** wurde ebenfalls komplett neu bestimmt, da es selbst bei der Verwendung des mittleren Ensembles noch einige Testkomplexe gibt, die keinerlei Referenzergebnisse aufweisen können. Es orientiert sich an Atomabständen von 2,0 Å bis 12,0 Å bei einer maximalen Abstandsdifferenz von 3,0 Å. Es besitzt somit die weitesten

Grenzen und geringste Selektionskraft. Die Verteilungskurven für die geometrischen Größen aller verbleibenden Referenzergebnisse finden sich in Abbildung IV-4c.

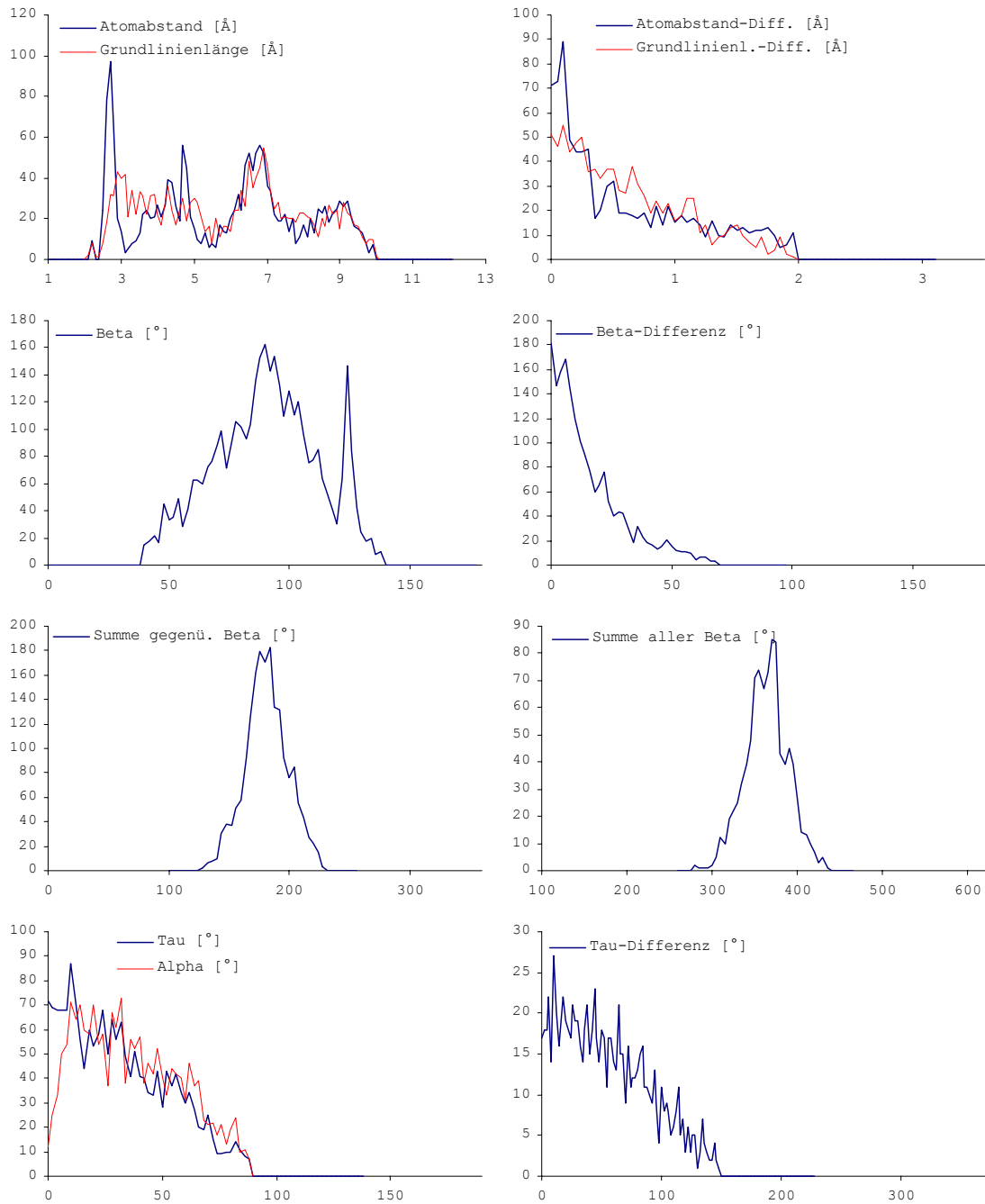
	enges Grenzwert- ensemble	mittleres Grenzwert- ensemble	weites Grenzwert- ensemble
Atomabstand	2,0 - 8,0 Å	2,0 - 10,0 Å	2,0 - 12,0 Å
Differenz der Atomabstände	0,0 - 2,0 Å	0,0 - 2,0 Å	0,0 - 3,0 Å
Grundlinienlänge	2,0 - 8,5 Å	2,0 - 8,0 Å	2,0 - 12,0 Å
Differenz der Grundlinienlängen	0,0 - 1,75 Å	0,0 - 2,0 Å	0,0 - 3,0 Å
Innenwinkel $\beta_1, \beta_2, \beta_3, \beta_4$	25 - 150°	40 - 140°	20 - 160°
Differenzen der Innenwinkel $ \beta_1 - \beta_2 $ und $ \beta_3 - \beta_4 $	0 - 65°	0 - 70°	0 - 90°
Summen gegenüberliegender Innenwinkel $\beta_1 + \beta_3$ und $\beta_2 + \beta_4$	129 - 240°	130 - 230°	110 - 250°
Summe aller $\beta$ -Innenwinkel	(258 - 480°)*	280 - 440°	270 - 460°
Torsionswinkel $ \tau $	0 - 90°	0 - 90°	0 - 135°
Differenz der Torsionswinkel $ \tau_1 - \tau_2 $	0 - 120°	0 - 150°	0 - 220°
Innenwinkel $\alpha$	0 - 95°	0 - 90°	0 - 135°

**Tabelle IV-4:** Die Kombinationen der einzelnen Selektionsgrenzwerte in den drei Grenzwertensembles mono- und bimolekularer transformations-invarianter Eigenschaften.

\* - Diese natürliche Grenzen wurden nachträglich eingefügt und entsprechen den mathematisch möglichen Minimum- und Maximumwerten aus den Summen gegenüberliegender  $\beta$ -Innenwinkel.

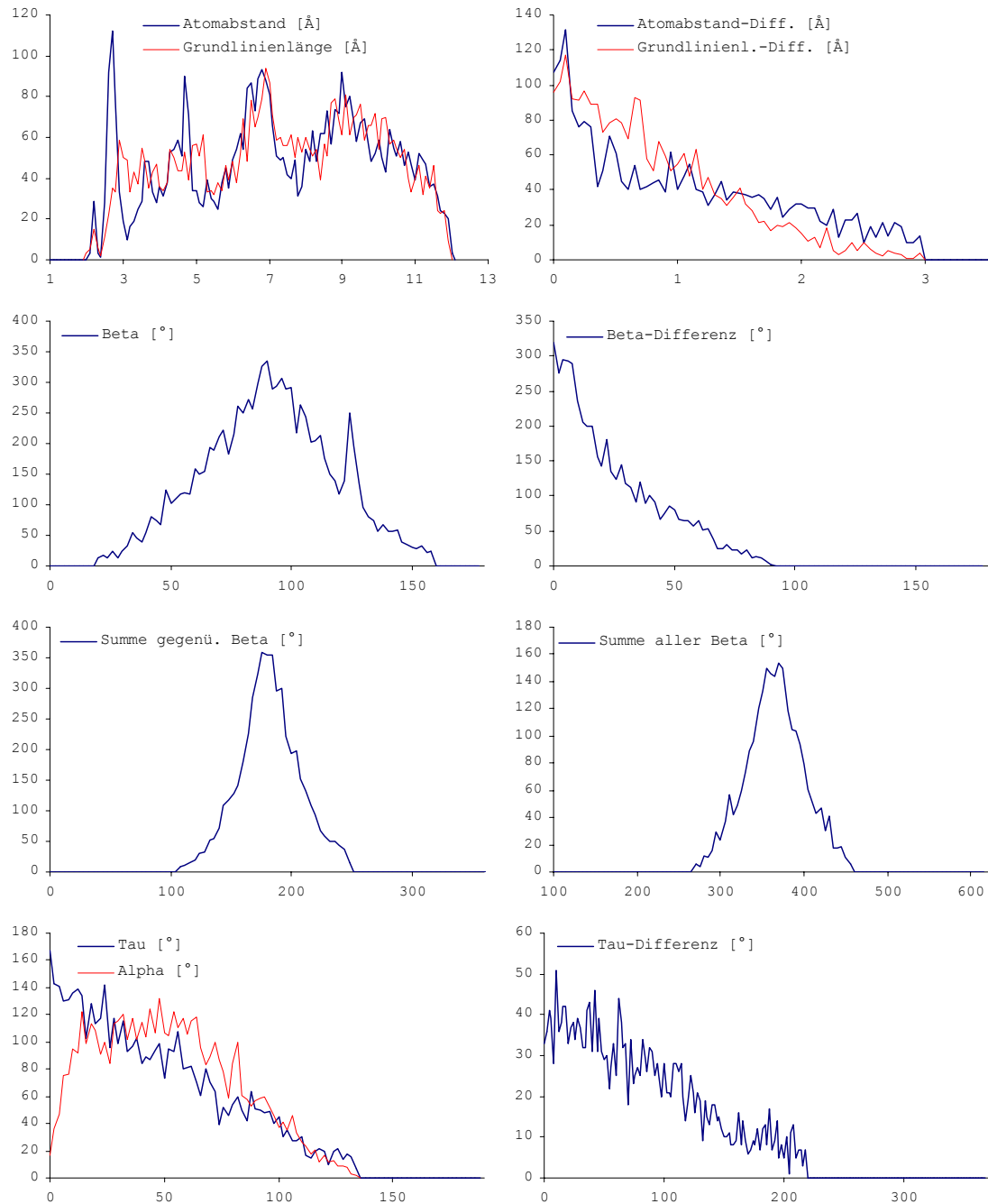


**Abbildung IV-4a:** Werteverteilung der mono- und bimolekularen transformations-invarianten Eigenschaften aller 1408 Paare aus Atomen intermolekularer Wasserstoffbrückenbindung der obligatorischen Proteinkomplexe bei vollständiger Selektion mit dem **engen Grenzwertensemble**. Als Funktionswerte ist auf der Ordinate stets die Anzahl der Wertebereiche in einem kleinen Wertintervall aufgetragen. Die Einheiten der Abszisse sind mit [Å] für Ångström und [°] für Winkelgrade jeweils im Histogramm angegeben.



**Abbildung IV-4b:** Werteverteilung der mono- und bimolekularen transformations-invarianten Eigenschaften aller 1838 Paare aus Atomen intermolekularer Wasserstoffbrückenbindung der obligatorischen Proteinkomplexe bei vollständiger Selektion mit dem **mittleren Grenzwertensemble**. Als Funktionswerte ist auf der Ordinate stets die Anzahl der Wertobservations in einem kleinen Wertintervall aufgetragen. Die Einheiten der Abszisse sind mit [Å] für Angström und [°] für Winkelgrade jeweils im Histogramm angegeben.





**Abbildung IV-4c:** Werteverteilung der mono- und bimolekularen transformations-invarianten Eigenschaften aller 4894 Paare aus Atomen intermolekularer Wasserstoffbrückenbindung der obligatorischen Proteinkomplexe bei vollständiger Selektion mit dem **weiten Grenzwertensemble**. Als Funktionswerte ist auf der Ordinate stets die Anzahl der Wertobservations in einem kleinen Wertintervall aufgetragen. Die Einheiten der Abszisse sind mit [Å] für Ångström und [°] für Winkelgrade jeweils im Histogramm angegeben.

### 11.2.4 Triplett-Vektordivergenz und Triplett-Bindungsseite

Einem Atomtriplett wird eine ausreichende Bindungskoordination zugestanden, wenn seine drei Interaktionsvektoren nicht zu weit voneinander abweichen, sondern möglichst kooperativ in dieselbe Richtung weisen.

Der Grenzwert für die Triplettvektordivergenz wird mit  $105^\circ$  festgelegt, da bei niedrigeren Werten der Verlust essenzieller Atomtripletts zu groß ist. Bei höheren Werten bleibt die Evaluierungsquote annähernd konstant. Es steigt dann hingegen dann die Anzahl falsch positiver Triplettts und dementsprechend unnötiger Evaluierungen funktionslos an, wie Tabelle IV-5 belegt.

	∅# Triplettts	∅# Evaluierungsaufrufe	# evaluierte Referenzen (von 923)	Evaluierungsquote
$\leq 65^\circ$	5.560	22.196.452	689	74,6 %
$\leq 75^\circ$	6.361	29.754.702	771	83,5 %
$\leq 85^\circ$	6.987	36.371.103	830	89,9 %
$\leq 95^\circ$	7.433	41.511.179	887	96,1 %
$\leq 105^\circ$	7.714	44.935.057	902	97,7 %
$\leq 115^\circ$	7.871	46.947.022	908	98,4 %
$\leq 125^\circ$	7.950	47.926.476	910	98,6 %
$\leq 135^\circ$	7.972	48.191.305	911	98,7 %
$\leq 145^\circ$	7.973	48.198.531	911	98,7 %

**Tabelle IV-5:** Parametervariation der maximal zugelassenen Triplett-Vektordivergenz von  $\leq 65^\circ$  bis  $\leq 145^\circ$ . Weitere Details im Text.  
(∅# - durchschnittliche Anzahl, # - Gesamtsumme)

Die Festlegung der Dreiecksbindungsseite hat keinen Einfluß mehr auf die Anzahl repräsentierender Atomtripletts. Wenn eine eindeutige Bindungsseite postuliert werden kann, zu der alle Interaktionsvektoren zeigen, legt sie jedoch fest, in welcher Orientierung zwei Dreiecke evaluiert werden.

Die Grenzwerte werden mit  $\leq 75^\circ$  zum Normalenvektor hin und mit  $\geq 105^\circ$  entgegen dem Normalenvektor festgelegt, da bei Werten näher gegen  $90^\circ$  die Evaluierungsquote aufgrund ausbleibender, essenzieller Evaluierungsaufrufe von Dreiecken mit nahezu planarer Anordnung der Interaktionsvektoren sinkt. Bei entfernteren Werten bleibt die Evaluierungsquote annähernd konstant, wohingegen dann die Anzahl funktionsloser beidseitiger Evaluierungen ansteigen, wie Tabelle IV-6 belegt.

	∅# Evaluierungsaufrufe	# evaluierte Referenzen (von 923)	Evaluierungsquote
$\leq 45^\circ$ & $\geq 135^\circ$	46.622.043	908	98,4 %
$\leq 55^\circ$ & $\geq 125^\circ$	46.572.522	908	98,4 %
$\leq 65^\circ$ & $\geq 115^\circ$	46.231.704	906	98,2 %
$\leq 75^\circ$ & $\geq 105^\circ$	44.935.057	902	97,7 %
$\leq 85^\circ$ & $\geq 95^\circ$	42.245.440	893	96,7 %
$\leq 90^\circ$ & $\geq 90^\circ$	40.438.083	882	95,6 %

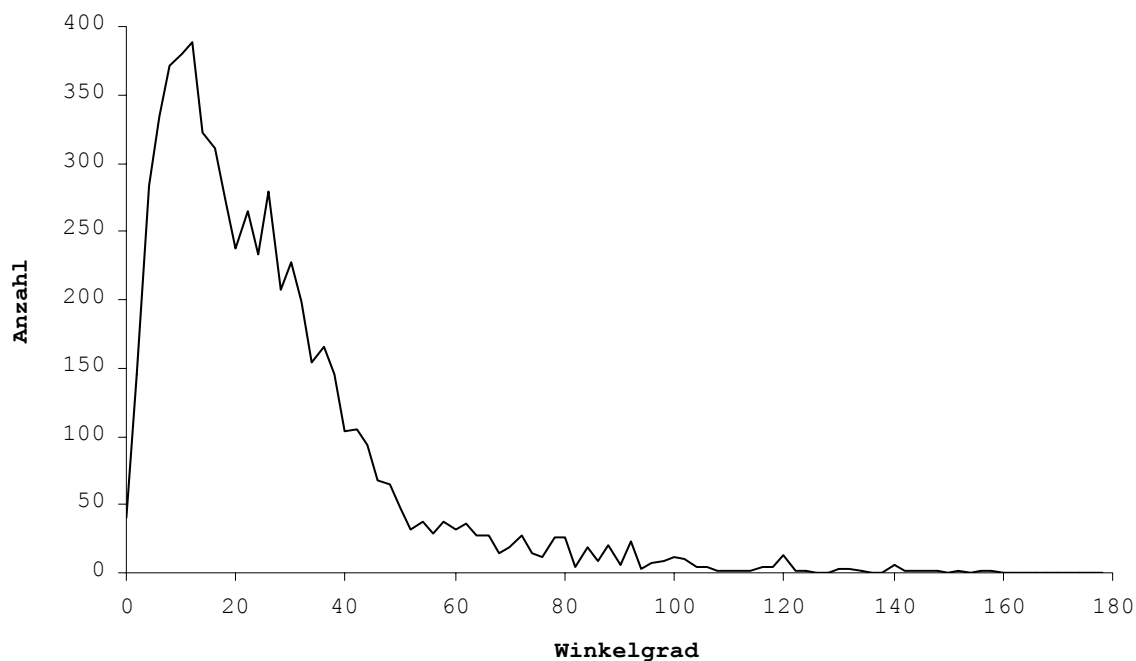
**Tabelle IV-6:** Parametervariation für die Festlegungsgrenzen der Triplet-Bindungsseite mit Vergleichsgrenzwerten von  $90^\circ \pm 45^\circ$  bis  $90^\circ \pm 0^\circ$ . Weitere Details im Text. (∅# - durchschnittliche Anzahl, # - Gesamtsumme)

### 11.3 Orientierungsabhängige Nachbewertung und Qualifizierung kriterienkonformer Kandidaten

#### 11.3.1 Nachbewertung multipler Rotationsbestimmungen eines Atomensembles

Sind multiple Rotationsbestimmungen möglich bzw. notwendig, so wird abschließend ein gemittelter Rotationswinkelsatz erzeugt. Indem die beteiligten Einzelrotationen mit diesem verglichen werden, soll bewertet werden wie plausibel diese gemittelte Rotation eigentlich ist oder ob die drei Seiten eines Dreiecks völlig kontraproduktive Orientierungen erzeugen, die nicht zeitgleich erfüllt werden können. Die Ähnlichkeit der multiplen Bestimmungen wird bewertet über die Summe der Abweichungen aller einzelnen Rotationswinkel eines jeden Winkelsatzes zum entsprechenden Rotationswinkel des gemittelten Winkelsatzes.

Abbildung IV-5 gibt die Werte für die Tripletreferenzen im Histogramm aus und weist eine deutliche Bevorzugung der Differenzwerte kleiner  $50^\circ$  auf. Die Kurve endet jedoch hier nicht abrupt sondern verflacht nur und deutet somit auf eine nicht unbeträchtliche Anzahl weniger optimaler Referenzergebnisse hin. Daher wurde nochmals empirisch überprüft, welche Auswirkungen systematische Grenzwertänderungen ( $50^\circ$  bis  $90^\circ$  und unselektiert) auf die Wiederfindungsquote hervorrufen. Die Evaluierungsquote und die Anzahl der initiierten Evaluierungen sind von dieser postselektiven Bewertung nicht beeinflusst. Es ist hingegen von Interesse wie die Auswirkungen speziell für die Referenzen sind, deren Orientierungsbestimmung letztendlich erfolgreich durchgeführt werden kann und die somit dem richtig positiven Rotationsrepräsentanten zugeteilt wurden. Tabelle IV-7 führt die entsprechenden Werte an.



**Abbildung IV-5:** Nachselektion multipler Rotationsbestimmungen eines Atomensembles. Aufgetragen ist die Summe der Abweichungen der einzelnen Rotationswinkel jedes Rotationswinkelsatzes zum entsprechenden Rotationswinkel des gemittelten Winkelsatzes. Weitere Details im Text.

	# gefundene Referenzen (von 923)	Wiederfindungs- quote	# Kandidaten mit richtig positiver Rotationsbestimmung	
$\leq 50^\circ$	782	84,7 %	522	56,6 %
$\leq 60^\circ$	809	87,7 %	535	58,0 %
$\leq 70^\circ$	831	90,0 %	541	58,6 %
$\leq 80^\circ$	841	91,1 %	544	58,9 %
$\leq 90^\circ$	850	92,1 %	545	59,0 %
unselektiert	862	93,4 %	545	59,0 %

**Tabelle IV-7:** Parametervariation der maximal zulässigen Abweichung in der Rotationsmittelung mit Grenzwerten von  $\leq 50^\circ$  bis  $\leq 90^\circ$  und ohne Begrenzung zur Nachselektion multipler Rotationsbestimmungen eines Atomensembles. Weitere Details im Text.  
Für alle Untersuchungen gilt: 932 Referenzen, davon 902 (97,7%) evaluiert mit durchschnittlich  $44,9 \cdot 10^6$  Evaluierungsaufrufen.  
(# - Gesamtanzahl)

Es zeigt sich, dass die Auswirkungen verschiedener Grenzwerte geringer sind für die Referenzen, deren Orientierungsbestimmung richtig erfolgt und dem erforderlichen Ergebnis entspricht. Die Wiederfindungsquote für diese Untergruppe aller Referenzergebnisse sinkt bei der Verwendung des 70°-Grenzwerts von den maximal erreichbaren 59,0% nur auf 58,6%, wohingegen der Verlust für die gesamte Wiederfindung von maximal 93,4% auf 90,0% erfolgt.

Um die eigentliche postselektive Wirkung dieses Bewertungskriteriums beurteilen zu können und die Kosten-Nutzen-Abwägung einer Grenzwertfestlegung abzuschätzen, muss zum ersten Mal mehr als nur das Schicksal der richtig positiven Ergebnisse verfolgt werden. Die Spezifität einer Selektion zu erhöhen bedeutet, falsch positive Ergebnisse signifikant stärker zu eliminieren als richtig positive. In Tabelle IV-8 soll daher beispielhaft für die Paar- und Triplettevaluierungen innerhalb der Testgruppe obligatorischer Proteinkomplexe das Docking des Komplexes 1kvd\_Ba innerhalb einer Grenzwertpermutation dokumentiert werden.

Triplet-Docking von 1kvd_Ba	≤50°	≤70°	≤90°	unselektiert
# evaluierte Referenzen (von 53)	52	52	52	52
# gefundene Referenzen (von 53)	52	52	52	52
# Kandidaten	106.087 (-86,2 %)	187.186 (-44,0 %)	250.129 (-25,0 %)	334.039
CPU-Sekunden	121 s (-92,6 %)	304 s (-81,3 %)	583 s (-64,2 %)	1630 s

Paar-Docking von 1kvd_Ba	≤50°	≤70°	≤90°	unselektiert
# evaluierte Referenzen (von 63)	63	63	63	63
# gefundene Referenzen (von 63)	63	63	63	63
# Kandidaten	320.860	330.238	334.737	337.902
CPU-Sekunden	386 s	425 s	441 s	454 s

**Tabelle IV-8:** Parametervariation der maximal zulässigen Abweichung in der Rotationsmittelung mit Grenzwerten von ≤50° bis ≤90° und ohne Begrenzung zur Nachselektion multipler Rotationsbestimmungen eines Atomensembles. Ergebnisszahlen für ein Triplet- und ein Paar-Docking mit dem Komplex 1kvd\_Ba. Weitere Details im Text.  
(# - Gesamtanzahl)

Wenn in diesem Beispiel auch die Wiederfindungsquoten unverändert bleiben, so sind die Kandidaten- und Laufzeiteinsparungen, die mit dieser Nachselektion erzielt werden können, beachtlich. Abnahmen gegenüber der unselektierten Vorgehensweise um 25,0% bis 68,2% für die Kandidatenzahlen und sogar um 64,2% bis 92,6% für die Laufzeit bei gleichbleibend 52

gefundenen Referenzergebnissen verdeutlichen die Bedeutung dieses Selektionskriteriums, welches zunächst nicht offensichtlich ist und willkürlich anmutet.

Die Ergebnisse des entsprechenden Paar-Dockings mit 1kvd\_Ba wurden aufgeführt, um zu zeigen, dass die Einsparungseffekte hier deutlich geringer ausfallen. Da es keine Entsprechung zu der multiplen Orientierungsbestimmung der drei Triplettkanten gibt, finden hier die Rotationsmittelung nur bei multiplen Protonenpositionen in der Evaluierung transformationsinvarianter Eigenschaften statt.

Der Grenzwert wird entsprechend den dargestellten Entwicklungen mit  $70^\circ$  festgelegt. Niedrige Werte setzen die Wiederfindung deutlich herab. Höhere Grenzwerte führen zu enormen Zunahmen falsch positiver Endkandidaten und der Programm-Laufzeit, bewirken jedoch vergleichsweise geringe Gewinne in der Wiederfindung richtig positiver Endergebnisse.

### 11.3.2 *Translationsabhängige Erhaltung in der Ergebnisclusterung*

In der Tabelle IV-9 werden die Auswirkungen verschiedener Entfernungskriterien in der Ergebnisqualifizierung des Triplet- bzw. Paar-Dockings für die Testgruppen obligatorischer und fakultativer Proteinkomplexe bei der Anwendung der verschiedenen Grenzwertensembles verglichen. Es handelt sich zum Einen um eine Ergebnisausdünnung anhand der Translationsangabe, welche ein kriterienkonformes Evaluierungsergebnis innerhalb der zuvor bestimmten Rotationsgruppe beschreibt. Zum Anderen ist die anschließende *Single-Linkage*-Clusterbildung der Ergebnisse durchzuführen, die auf der Ähnlichkeit derselben Translationseigenschaft innerhalb homogener Gruppen beruht.

Die Ergebnisausdünnung entspricht einer Nachbarsuche, bei der jedes Translationsergebnis einer intermolekularen Paarkombination zumindest zwei und jedes Translationsergebnis einer Triplettkombination zumindest einen Nachbarn innerhalb des Entfernungskriteriums besitzen muss, um weiter berücksichtigt zu werden. Diese Mindestbelegung von drei bzw. zwei gleichartigen Ergebnissen kommt aus der Betrachtung tatsächlich erhaltener Referenzzahlen, die bei der Beschreibung der Test-Datensätze dargelegt wird. Sie wird auch für die Mindestbelegung finaler Ergebniscluster verwendet.

Triplett-Docking	# Referenzen mit richtiger Rotation nach der Ergebnisausdünnung				# Referenzen mit richtiger Rotation im finalen Ergebnisscluster			
	obligatorisch (von 932)		fakultativ (von 530)		obligatorisch (von 932)		fakultativ (von 530)	
≤0,3 Å	406	44,0%	128	24,2%	289	31,3%	55	10,4%
≤0,5 Å	500	54,2%	171	32,3%	457	49,5%	136	25,7%
≤1,0 Å	528	57,2%	207	39,1%	514	55,7%	195	36,8%
≤1,5 Å	532	57,6%	218	41,1%	516	55,9%	215	40,6%
≤2,0 Å	533	57,7%	218	41,1%	516	55,9%	215	40,6%
≤2,5 Å	534	57,9%	218	41,1%	525	56,9%	215	40,6%
≤3,0 Å	535	58,0%	219	41,3%	525	56,9%	216	40,8%
≤3,5 Å	535	58,0%	219	41,3%	525	56,9%	216	40,8%
≤4,0 Å	535	58,0%	219	41,3%	525	56,9%	216	40,8%

Paar-Docking	obligatorisch (von 1253)				fakultativ (von 1107)			
	obligatorisch (von 1253)		fakultativ (von 1107)		obligatorisch (von 1253)		fakultativ (von 1107)	
≤0,3 Å	186	14,8%	37	3,3%	131	10,8%	33	3,0%
≤0,5 Å	391	31,2%	99	8,9%	361	28,8%	90	8,1%
≤1,0 Å	513	40,9%	216	19,5%	513	40,9%	216	19,5%
≤1,5 Å	540	43,1%	280	25,3%	537	42,9%	277	25,0%
≤2,0 Å	547	43,7%	295	26,6%	547	43,7%	295	26,6%
≤2,5 Å	547	43,7%	301	27,2%	547	43,7%	301	27,2%
≤3,0 Å	547	43,7%	305	27,6%	547	43,7%	305	27,6%
≤3,5 Å	547	43,7%	305	27,6%	547	43,7%	305	27,6%
≤4,0 Å	547	43,7%	305	27,6%	547	43,7%	305	27,6%

**Tabelle IV-9:** Parametervariation des Entfernungskriteriums für Ergebnisobservationen zur translationsabhängigen Erhaltung in der Ergebnisausdünnung und -clustering für das Triplett- und das Paar-Docking. Prozentuale Angaben beziehen sich auf die jeweilige Gesamtanzahl erzielbarer Referenzergebnisse.

Hellgrau-schattierte Bereiche markieren konstante Höchstwerte, die auch mit höheren Grenzwerten bis zu 8,0 Å nicht gesteigert werden können.

obligatorisch / Triplets: je 932 Referenzen, davon 902 (97,7%) evaluiert, 831 (90,0%) gefunden und 541 (58,6%) mit der richtigen Rotation bestimmt

obligatorisch / Paare: je 1253 Referenzen, davon 1232 (98,3%) evaluiert, 1163 (92,8%) gefunden und 547 (43,7%) mit der richtigen Rotation bestimmt

fakultativ / Triplets: je 530 Referenzen, davon 485 (91,5%) evaluiert, 446 (84,2%) gefunden und 233 (44,0%) mit der richtigen Rotation bestimmt

fakultativ / Paare: je 1107 Referenzen, davon 1069 (96,6%) evaluiert, 986 (89,1%) gefunden und 331 (29,9%) mit der richtigen Rotation bestimmt

(# - Gesamtanzahl)

Zusammenfassend und der späteren Diskussion vorweggreifend zeigt sich, dass die Rotations- und Translationsbestimmungen im Triplett-Docking zuverlässiger gelingen als im Paar-

Docking. Zudem lässt sich die Transformation für obligatorische Proteinkomplexe besser bestimmen als für fakultative Komplexe.

Im Triplet-Docking werden 58,6% der Referenzen obligatorischer Komplexe mit der richtigen Rotation beschrieben. Diese bleiben bei einem Abstandskriterium von zumindest 3,0 Å mit Maximalwerten von 58,0% fast vollzählig in der Ausdünnung erhalten. In der Ergebnisclustering sind zuletzt zehn Orientierungen nicht an der Bildung des größten Translationsclusters beteiligt sondern bilden eigenständige Gruppen. Daher sinkt die finale Erhaltungsquote auf 56,9%. Diese kann sich natürlich verändern, wenn auch falsch positive Kandidaten in diese Clusterbildung eingreifen. Für fakultative Komplexe erfolgt die Rotationsbestimmung nur bei 44,0% entsprechend den richtigen Erwartungen. Von diesen Referenzen bleiben zunächst 41,3% und zuletzt 40,8% bei Maximalabständen von zumindest 3,0 Å erhalten.

Das universelle Translations-Entfernungskriterium im Triplet-Docking wird mit 3,0 Å festgesetzt.

Im Paar-Docking werden 43,7% der Referenzen obligatorischer Komplexe mit der richtigen Rotation beschrieben und bleiben bereits mit einem sehr geringen Grenzwert von 2,0 Å vollständig bis zum Ende erhalten. Für fakultative Komplexe sinkt die Erfolgsquote der Rotationsbestimmung auf 29,9% aller Referenzen. Hiervon bleiben bis zuletzt nur 27,6% bei einem Entfernungsgrenzwert von 3,0 Å erhalten. Die nichtverbliebenen Kandidaten fallen bereits bei der Nachbarsuche aus der Evaluierung, da sie nur eine Zweiergruppe im richtigen Rotationscluster bilden können und die erforderliche Minimalbelegung von zumindest drei Vertretern nicht erbringen.

Auch hier wird ein gemeinsamer Grenzwert von 3,0 Å festgesetzt.

Höhere Grenzwerte bis zu 8,0 Å ermöglichen keine Quotensteigerung.

Für stärker selektionierende Grenzwertensembles könnten die vorgestellten Grenzwerte unter geringsten Referenzverlusten um tendenziell 0,5 Å bis 1,0 Å gesenkt werden (Daten nicht gezeigt). Hierauf wird jedoch zunächst verzichtet, da einerseits die momentane Rotationsbestimmung abschließende Arbeiten mit allen Referenzergebnissen gar nicht ermöglicht und andererseits weder Ergebnisspezifität noch Programm-Laufzeit dieses Vorgehen für das enge oder mittlere Grenzwertensemble erfordern, wie spätere Ergebnisse belegen werden.



### 11.4 Tabellarische Zusammenstellung aller verwendeten Selektionswerte

In der folgenden Tabelle IV-10 sind alle verwendeten Parameter zusammen gestellt, die zu verschiedenen Zeitpunkten ihre spezifische Selektionswirkung ausüben. Ihre genaue Natur, der Zusammenhang und die Herleitung der gewählten Grenzwerte wurden erläutert. Ebenso die Existenz dreier Grenzwertensembles der eigentlichen Evaluierung transformations-invarianter Eigenschaften mono- und bimolekularer Kombinationen bindungsfähiger Atome. Hinzugefügt wurden noch die verschiedenen Bewertungsquoten, welche den Verbleib der richtig positiven Referenzergebnisse zu unterschiedlichen Zeitpunkten verdeutlichen.

Atomabstand zur lösungsmittelzugänglichen Moleküloberfläche		nur Atome mit eigenen Oberflächenpunkten			Evaluierungsquote
geeignete Orientierung der Interaktionsvektoren		$\leq 15^\circ$			
Triplet-Vektordivergenz		$\leq 105^\circ$			
Festlegung der Triplet-Bindungsseite		eindeutig zur Normalen: $\leq 75^\circ$ eindeutig entgegen d. Normalen: $\geq 105^\circ$			
Atomabstand	monomolekular	2,0 - 8,0 Å	2,0 - 10,0 Å	2,0 - 12,0 Å	Evaluierungsquote
Grundlinienlänge		2,0 - 8,5 Å	2,0 - 8,0 Å	2,0 - 12,0 Å	
Innenwinkel $\beta_1, \beta_2, \beta_3, \beta_4$		25 - 150°	40 - 140°	20 - 160°	
$\beta$ -Winkel-Differenz $ \beta_1 - \beta_2 $ und $ \beta_3 - \beta_4 $		$\leq 65^\circ$	$\leq 70^\circ$	$\leq 90^\circ$	
Torsionswinkel $ \tau $		$\leq 90^\circ$	$\leq 90^\circ$	$\leq 135^\circ$	
Innenwinkel $\alpha$		$\leq 95^\circ$	$\leq 90^\circ$	$\leq 135^\circ$	
Differenz der Atomabstände	bimolekular	$\leq 2,0$ Å	$\leq 2,0$ Å	$\leq 3,0$ Å	Wiederfindungsquote
Differenz der Grundlinienlängen		$\leq 1,75$ Å	$\leq 2,0$ Å	$\leq 3,0$ Å	
Summen gegenüberliegender Innenwinkel $\beta_1 + \beta_3$ und $\beta_2 + \beta_4$		129 - 240°	130 - 230°	110 - 250°	
Summe aller $\beta$ -Innenwinkel		258 - 480°	280 - 440°	270 - 460°	
Torsionswinkel-Differenz $ \tau_1 - \tau_2 $		$\leq 120^\circ$	$\leq 150^\circ$	$\leq 220^\circ$	
Fehlersumme in der Rotationsmittelung		$\leq 70^\circ$			
Nachbarnabstand vor der Translationsclustering		für Paare und Triplets: $\leq 3,0$ Å			Erhaltungsquote
Nachbarnanzahl vor der Translationsclustering		für Paare: $\geq 3$ für Triplets: $\geq 2$			
Entfernungskriterium der <i>Single-Linkage</i> -Translationsclustering		für Paare und Triplets: $\leq 3,0$ Å			

**Tabelle IV-10:** Zusammenstellung aller verwendeten Kriteriengrenzwerte. Die drei optional zu wählenden Grenzwertensembles der mono- und bimolekularen transformations-invarianten Eigenschaften sind grau unterlegt.

## 12 Beschreibung und Eignungsbestimmung der Test-Datensätze für die vorgestellten Dockingmethoden und Selektionsgrenzwerte

Die Tabellen im Anhang A4 enthalten eine Aufstellung der fakultativen und obligatorischen Protein-Protein-Komplexe aus der Vorselektion heterooligomerer Test-Datensätze [Vakser, Sali, 2001]. Sie sind ergänzt um einige Informationen, welche Vakser und Sali zur Beschreibung der Komplexe ihrer Zusammenstellung liefern. Zudem sind Angaben zur jeweiligen Anzahl intermolekularer Wasserstoffbrückenbindungen hinzugefügt, wie sie von HBPLUS zwischen den entsprechenden Peptidketten aus den Strukturen postuliert werden. Des Weiteren wird angegeben, wie viele richtig positive Referenzergebnisse aus diesen Wasserstoffbrückenbindungen erzeugt werden können, wenn auf alle hieraus möglichen Paar- bzw. Tripletzzusammenstellungen das enge, mittlere oder weite Grenzwert-Ensemble selektionierend angewendet wird.

Es muss beachtet werden, dass alle dort aufgeführten Eigenschaften nur bestimmt werden können, wenn die Struktur des Komplexes vorliegt und analysiert werden kann. Es ist nicht möglich, diese konkreten Zahlen aus den Einzelmolekülen abzuleiten.

### 12.1 Ausdehnung der Kontaktflächen

Die Varianz innerhalb der 62 vorselektierten fakultativen Komplexe ist auffallend hoch. So besitzen etwa 80% der Komplexe eine durch den Kontaktpartner abgeschirmte molekulare Kontaktfläche (*buried molecular surface area, msa*) von  $300 \text{ \AA}^2$  bis  $1400 \text{ \AA}^2$  mit je 12 bis 88 Aminosäureresiduen, die innerhalb dieser Fläche lokalisiert und somit mitverantwortlich sind für genau die spezifische Kontaktvermittlung, welche innerhalb dieser Arbeiten ausgenutzt werden sollen. Etwa 15% der Komplexe besitzen Kontaktflächen von  $1400 \text{ \AA}^2$  bis  $2300 \text{ \AA}^2$  mit je 91 bis 191 Residuen. Die verbleibenden 5% sind mit  $2500 \text{ \AA}^2$  bis  $3300 \text{ \AA}^2$  und 174 bis 220 Residuen recht großflächig.

Die Gruppe der 16 obligatorischen Komplexe verhält sich insgesamt gar nicht auffällig homogener als die zuvor beschriebene. Etwa 50% der Komplexe weisen eine molekulare Kontaktfläche von  $650 \text{ \AA}^2$  bis  $1400 \text{ \AA}^2$  mit je 61 bis 125 Kontaktresiduen auf, wohingegen die übrigen 42% Kontaktflächen von  $1700 \text{ \AA}^2$  bis  $2500 \text{ \AA}^2$  mit je 152 bis 210 Residuen besitzen. Ein einzelner Komplex ist mit etwa  $3400 \text{ \AA}^2$  bei 235 Residuen wieder am oberen Ende des Spektrums vertreten.

fakultative Komplexe			
<i>msa</i>	300 Å <sup>2</sup> - 1400 Å <sup>2</sup>	1400 Å <sup>2</sup> - 2300 Å <sup>2</sup>	2500 Å <sup>2</sup> - 3300 Å <sup>2</sup>
Anteil der Testgruppe	80%	15%	5%
# AS in Kontaktfläche	12 - 88 (einmalig 105)	91 - 191	174 - 220
# intermolek. WBB	1 - 24	13 - 44	41 - 55

obligatorische Komplexe			
<i>msa</i>	650 Å <sup>2</sup> - 1400 Å <sup>2</sup>	1700 Å <sup>2</sup> - 2500 Å <sup>2</sup>	3400 Å <sup>2</sup>
Anteil der Testgruppe	50%	42%	8%
# AS in Kontaktfläche	61 - 125	152 - 210	235
# intermolek. WBB	10 - 80	19 - 46	42

**Tabelle IV-11:** Übersicht einiger Komplexeigenschaften, weitere Details im Text.

*msa*: Molekulare Kontaktfläche, die durch den Komplexpartner abgeschirmt wird (*buried molecular surface area*)

#AS: Anzahl der Aminosäuren in der abgeschirmten Kontaktfläche

#intermolek. WBB: Anzahl der intermolekularen Wasserstoffbrückenbindungen

Aus Anhang fällt bei Betrachtung der zusammenfassenden Tabelle IV-11 auf, dass ein höherer Anteil der obligatorischen Komplexe Kontaktflächen von 1700 Å<sup>2</sup> und mehr besitzt, was bei der geringen Messgrundlage jedoch nicht zu hoch zu bewerten ist. Aussagekräftiger ist die bedeutend höhere Anzahl der involvierten Residuen bei vergleichbarer Kontaktflächen der kleineren Komplexe.

Bereits hieraus lässt sich ein Hinweis ableiten, dass Interaktionsflächen fakultativer und obligatorischer Proteinkomplexe signifikante Unterschiede aufweisen. Innerhalb der Kontaktflächen obligatorischer Komplexe scheint eine auffallend höhere Packungsdichte der einzelnen Reste zu herrschen, was auch ein höheres Maß an Ordnung voraussetzt. Es stellt sich die Frage, ob die Gestaltung der Kontaktflächen z.B. in Hinsicht auf die Orientierung der Aminosäureseitenketten identisch sein kann. Es gilt, mehr Reste auf vergleichbarem Raum unterzubringen. Dieses kann vielleicht durch eine zur Interaktionsfläche hin gerichtete Präsentation der Hauptkette im Gegensatz zur Präsentation der voluminöseren Seitenketten erreicht werden.

Tabelle IV-12 enthält eine Aufstellung der statistischen Mittelwerte und Standardabweichungen einiger der betrachteten Eigenschaften und bestätigt die größeren Kontaktflächen obligatorischer Komplexe mit  $1658,3 \pm 804,6$  Å gegenüber den  $936,4 \pm 630,4$  Å fakultativer Komplexe. Es wird bei Betrachtung der hohen Standardabweichungen nochmals deutlich, wie inhomogen beide

Gruppen sind. In Relation zu den Beträgen der Mittelwerte sind die Abweichungen der fakultativen Test-Datensätze jedoch nochmals deutlich größer als für obligatorische Komplexe.

	$msa$ [ $\text{\AA}^2$ ]	$asa$ [ $\text{\AA}^2$ ]	#wbb	$5.34*asa*10e-3/\text{\AA}^2$
62 fakultative Komplexe	936,4 $\pm 630,4$	3085,5 $\pm 2063,0$	15,2 $\pm 12,0$	16,5 $\pm 11,0$
12 der 16 obligatorischen Komplexe	1658,3 $\pm 804,6$	5472,0 $\pm 2662,2$	44,5 $\pm 21,0$	29,8 $\pm 18,0$

**Tabelle IV-12:** Mittelwerte und Standardabweichungen einiger Kontaktflächeneigenschaften für die Testgruppen, weitere Details im Text.

$msa$ : molekulare Kontaktfläche, die durch den Komplexpartner abgeschirmt wird (*buried molecular surface area*)

$asa$ : lösungsmittelzugängliche Kontaktfläche, die durch den Komplexpartner abgeschirmt wird (*buried solvent accessible surface area*)

#wbb: Anzahl der von HBPLUS beschriebenen intermolekularen Wasserstoffbrückenbindungen

## 12.2 Anzahl intermolekularer Wasserstoffbrückenbindungen

Untersuchungen von Janin und Chothia an fakultativen Protein-Protein-Komplexen haben mit einem Korrelationskoeffizienten von 0,89 eine starke Abhängigkeit der Anzahl intermolekularer Wasserstoffbrückenbindungen zur lösungsmittelzugänglichen Kontaktfläche, die durch den Komplexpartner abgeschirmt wird (*buried solvent accessible surface area, asa*) bestimmt [Janin & Chothia, 1990]. Diese lineare Abhängigkeit lässt sich mit dem Term  $5,34*asa*10e-3/\text{\AA}^2$  ausdrücken.

Für die zusammengestellten Test-Datensätze der fakultativen Komplexe kann dieses Ergebnis auch in etwa bestätigt werden. So werden entsprechend den Angaben in Tabelle IV-12 bei  $asa$ -Kontaktflächen von  $3085,5 \pm 2063,0 \text{\AA}^2$  rechnerisch  $16,5 \pm 11,0$  intermolekulare Bindungen erwartet und von HBPLUS  $15,2 \pm 12,0$  Bindungen postuliert. Für die obligatorischen Komplexe hingegen liegen die Verhältnisse doch signifikant anders. Kontaktflächen von  $5472,0 \pm 2662,2 \text{\AA}^2$  ließen  $29,8 \pm 18,0$  Bindungen erwarten, wohingegen jedoch  $44,5 \pm 21,0$  Bindungen aus der Struktur bestimmt werden können.

Hier liegt also für obligatorische Proteinkomplexe neben den größeren Kontaktflächen zudem eine auffallend höhere Dichte intermolekularer Wasserstoffbrückenbindungen vor. Diese kann insbesondere für die Komplexe mit kleineren Kontaktflächen auch bereits aus Tabelle IV-11 abgeleitet werden.

Ebenfalls aus dieser Tabelle IV-11 bzw. den detaillierten Tabellen im Anhang wird eine weitere Varianz innerhalb der Testgruppe ersichtlich. Es gibt obligatorische Komplexe mit  $697 \text{ \AA}^2$  molekularer Kontaktfläche und 75 intermolekularen Wasserstoffbrückenbindungen (2hnt\_Fc), jedoch auch mit  $1704 \text{ \AA}^2$  und nur 28 Bindungen (1aks\_Ab). Ebenso auffällig ist der Komplex 1mhl\_Ca, der mit  $3403 \text{ \AA}^2$  zwar den großflächigsten Kontakt aufbaut, jedoch "nur" 42 intermolekulare Bindungen vorweisen kann.

### 12.3 Anzahl erzielbarer Referenzergebnisse

Tabelle IV-13 soll unter Verwendung der verschiedenen Methoden und Grenzwertensembles einen Überblick für die Anzahl der kriterienkonformen Referenzergebnisse geben. Die Menge dieser Ergebnisse spielt die bedeutende Rolle, wenn es später darum geht, das einzige richtig positive Ergebnis anhand überproportional häufiger Observationen zu identifizieren.

Paar-Docking	16 obligatorische Komplexe		58 fakultative Komplexe	
	$\varnothing$ # Paar-Referenzen	#Komplexe $\geq 2$ Ref.	$\varnothing$ # Paar-Referenzen	#Komplexe $\geq 2$ Ref.
eng	25,0 $\pm$ 23,3 (4 - 102)	16	4,6 $\pm$ 7,5 (0 - 38)	33
mittel	32,4 $\pm$ 29,7 (8 - 133)	16	6,1 $\pm$ 9,3 (0 - 47)	40
weit	78,3 $\pm$ 61,4 (10 - 253)	16	17,9 $\pm$ 21,6 (0 - 119)	55

Triplett-Docking	16 obligatorische Komplexe		58 fakultative Komplexe	
	$\varnothing$ # Triplett-Referenzen	#Komplexe $\geq 2$ Ref.	$\varnothing$ # Triplett-Referenzen	#Komplexe $\geq 2$ Ref.
eng	7,5 $\pm$ 9,7 (0 - 40)	10	0,7 $\pm$ 2,1 (0 - 10)	6
mittel	14,9 $\pm$ 20,7 (1 - 89)	13	1,5 $\pm$ 4,3 (0 - 21)	9
weit	57,7 $\pm$ 72,3 (6 - 315)	16	8,5 $\pm$ 15,5 (0 - 95)	37

**Tabelle IV-13:** Durchschnittliche Anzahl ( $\varnothing$ #) erzielbarer Referenzergebnisse und Anzahl (#) geeigneter obligatorischer und fakultativer Proteinkomplexe mit jeweils zumindest zwei kriterienkonformen Referenzergebnissen für die Verwendung im Paar- bzw. Triplett-Docking mit den unterschiedlichen Selektionsgrenzwert-Ensembles. Für die Referenzergebnisse angeführt sind Mittelwerte mit Standardabweichungen sowie eingeklammert die auftretenden Minima und Maxima.

Zunächst einmal offenbart sich hier ein genereller Unterschied zwischen dem Docking mit Paaren oder mit Triplets intermolekularer Wasserstoffbrückenbindungen. Die erhöhte Selektionsanforderungen der Kriterienerfüllung an zugleich allen drei Dreieckskanten bewirkt, dass es je nach Vergleichsgrundlage die eineinhalb bis fünffache Anzahl an Referenzpaaren als an Referenztriplets gibt.

Der Unterschied zwischen obligatorischen und fakultativen Proteinkomplexen lässt sich direkt aus den unterschiedlichen Anzahlen intermolekularer Wasserstoffbrückenbindungen aus Tabelle IV-12 ableiten. Obligatorische Komplexe besitzen etwa die dreifache Anzahl intermolekularer Bindungen und können hieraus etwa die fünffache Anzahl Referenzpaare und sogar die etwa zehnfache Anzahl Referenztripletts bilden. Diese Fähigkeiten sind aufgrund der dahinter liegenden Kombinatorikalgorithmien nachvollziehbar, die in einem eigenen Kapitel ausführlich betrachtet werden sollen.

Die Unterschiede der Referenzzahlen bei Verwendung der verschiedenen Grenzwertensembles sind für obligatorische und fakultative Komplexe wieder vergleichbar. Gegenüber dem engen Selektionsgrenzen erzeugten die mittleren Grenzen etwa die eineinhalb- und zweifache Anzahl Referenzpaare und -tripletts. Das weite Grenzwertenssemble sorgt für die drei- bis vierfache Anzahl an Referenzpaaren und sogar sieben- bis zwölffache Anzahl an Referenztripletts.

## **12.4 Eignung der Test-Datensätze für unterschiedliche Methoden und**

### **Grenzwertensembles**

Ebenfalls Bezug nehmend auf Tabelle IV-13 bzw. die entsprechenden Tabellen im Anhang soll hier dargelegt werden, inwieweit eine Protein-Protein-Komplex überhaupt für die vorgestellten Methoden und Parameter geeignet ist. Problem einer derartigen Vorselektion, die für eine nachfolgende Untersuchung den Suchraum eingrenzen soll, ist die Gefahr, den richtig positiven Suchabschnitt zu verwerfen. Hierdurch kann das richtig positive Endergebnis in der Hauptuntersuchung nicht mehr gefunden werden, da es gar nicht weiter getestet wird.

In dieser Arbeit wird generell nach einem überproportional häufigen Auftreten von Paaren bzw. Tripletts intermolekularer Wasserstoffbrückenbindungen innerhalb der spezialisierten Kontaktflächen gesucht, um dann anhand der multiplen, gleichartigen Transformationsbeschreibungen die richtige Molekülorientierung vorherzusagen. Somit sind Komplexe ungeeignet, die zu wenige intramolekulare Wasserstoffbrückenbindungen aufweisen können, um hieraus überhaupt Paare oder Tripletts zusammen zu stellen. Ebenso ungeeignet sind Komplexe, deren Interaktionen zu weit verteilt oder zu unspezifisch für eine kooperative Bindung sind, um entsprechend den verwendeten Selektionsgrenzwerten eine ausreichende Anzahl kriterienkonformer Suchergebnisse zusammen zu stellen. Das grundlegende Eignungskriterium ausreichender Referenzzahlen muss also vom Molekülkomplex bzw. den beiden Einzelmolekülen, die der Dockingsimulation unterzogen werden sollen, erfüllt werden. Hierbei stellt sich erstmals die Frage, ob es entsprechend der späteren Aufgabenstellung des

Dockings möglich ist, bereits einem einzelnen Proteinmolekül im Vorhinein anzusehen, ob es diese Eignungskriterien erfüllen kann.

Für obligatorische Komplexe zeigt sich, dass alle 16 Vertreter im Minimalfall vier Referenzpaare aufweisen können. Für Referenztripletts muss bereits differenziert werden nach den verwendeten Grenzwertensembles. So stellen die weiten Grenzen ebenfalls allen Komplexen zwischen sechs und 315 Referenzen zur Verfügung. Die mittlere Selektion zeigt schon drei Komplexe mit lediglich einem Referenzergebnis, was eine Bewertung aufgrund multipler Observationen schon vereitelt. Die Verwendung des engen Grenzwertensembles disqualifiziert sogar drei Komplexe mit nur einem und weitere drei Komplexe völlig ohne ein Referenztriplett. Die große Varianz kann ebenso wie die jeweiligen Maximalergebnisse der Tabelle IV-13 entnommen werden. Erwähnt werden soll lediglich, dass es selbst unter Verwendung der strengsten Selektionsvorschriften Komplexe mit 102 Referenzpaaren und 40 Referenztripletts gibt. Andererseits kann der Komplex 1mhl\_Ca mit der größten Kontaktfläche ( $3400 \text{ \AA}^2$  *msa*,  $11205 \text{ \AA}^2$  *asa*) nur ein bzw. kein kriterienkonformes Tripletts bei Anwendung des mittleren bzw. engen Grenzwertensembles erzeugen. Seine 42 intermolekularen Wasserstoffbrückenbindungen liegen einfach über eine zu große Fläche verteilt.

Für fakultative Komplexe muss sich die Situation mit der deutlich geringeren Anzahl intermolekularer Interaktionen zwangsweise noch verschärft präsentieren. Beim Docking mit Atompaaaren weisen mit abnehmender Selektionsstärke nur 33, 40 bzw. 55 aller 62 Komplexe überhaupt zumindest zwei Referenzergebnisse auf. Für Referenztripletts liegen die entsprechenden Zahlen bei lediglich sechs, neun und 37 Komplexen. Von den fakultativen Proteinkomplexen sind somit nur 56,5%, 64,5% bzw. 88,7% im Paar- und 9,7%, 14,5% bzw. 59,7% im Tripletts-Docking überhaupt ansatzweise für diese Methoden und die angewendeten Grenzwertzusammenstellungen geeignet. Weshalb hierfür nicht einfach ein noch selektionsschwächeres Grenzwertensemble zusammen gestellt wird, findet seine Begründung in einer unvertretbar großen Zunahme resultierender Kandidatenzahlen und der zeitlichen Dauer eines Dockings. Die einhergehende Herabsetzung der Ergebnisspezifität macht es in den überwiegenden Fällen unmöglich, richtig positive Ergebnisse in ihrer Bedeutung herauszustellen.

Um im weiteren Verlauf der Ergebnisdarstellung und -diskussion nicht stets auf fakultative Proteinkomplexe verweisen zu müssen, die überhaupt nicht in der Lage sind, die notwendigen

Suchergebnisse vorzuweisen, sollen hier die jeweiligen Untersuchungsgrundlagen neu festgelegt werden. Für obligatorische Komplexe erfolgt ein Hinweis in der entsprechenden Diskussion.

- Für die nachfolgenden Qualitätsbewertungen der Methoden, die sich in den Evaluierungs-, Wiederfindungs- und Erhaltungsquoten widerspiegeln, werden stets alle Komplexe verwendet, die überhaupt jemals Referenzergebnisse aufweisen können. Das disqualifiziert lediglich die vier Komplexe 1bun\_Ab, 1tco\_Ab, 1ytf\_Db und 2pcc\_Ab und hinterlässt eine Testgruppe mit 58 fakultativen Komplexen.
- Für die übrigen Ergebnisbestimmungen wie die Rangfestlegung des richtigen Transformationsclusters innerhalb aller richtig und falsch positiven Kandidaten, werden alle fakultativen Komplexe ignoriert, die abhängig vom verwendeten Grenzwertensemble weniger als zwei Referenzergebnisse erbringen können. Der Einfachheit halber erfolgt diese Neubestimmung der Testgruppen nach folgendem Schema, wobei die Tabelle im Anhang A6 die jeweiligen Vertreter identifiziert.

fakultative Komplexe mit zwei und mehr Referenzergebnissen	... im <b>Paar- Docking</b>	... im <b>Triplet- Docking</b>
... beim <b>engen und mittleren</b> Grenzwertensemble	33	6
... beim <b>weiten</b> Grenzwertensemble	55	37

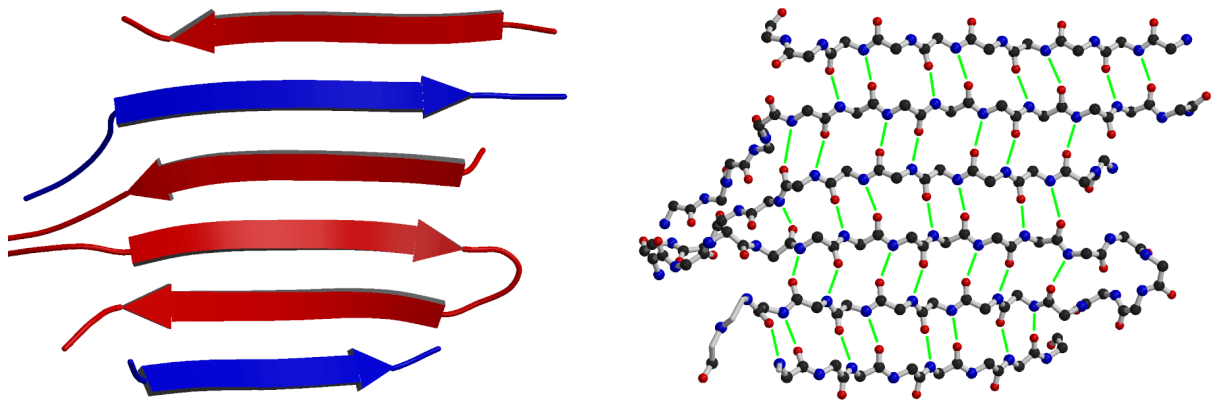
## 12.5 Strukturelle Grundlagen der Unterschiede obligatorischer und fakultativer Proteinkomplexe

Da die Unterschiede obligatorischer und fakultativer Proteinkomplexe an dieser Stelle sehr deutlich werden, soll nun die strukturelle Ursache erläutert werden.

Abbildung IV-6a zeigt einen Ausschnitt der Moleküloberfläche des obligatorischen Protein-Protein-Komplexes 2ltn\_Aa, der bereits zu Beginn der Ausführungen in Abbildung I-1 vorgestellt wurde. Dargestellt sind die antiparallelen  $\beta$ -Stränge eines typischen, sehr stabilen  $\beta$ -Faltblattes. Sekundärstrukturelemente wie  $\alpha$ -Helices und  $\beta$ -Faltblätter sind essenziell für Proteinstrukturen und werden ausgebildet bzw. stabilisiert durch eine Vielzahl regelmäßiger Wasserstoffbrückenbindungen zwischen den Stickstoff- und Carbonylsauerstoffatomen der beteiligten Hauptkettenabschnitte. Diese Wasserstoffbrückenbindungen besitzen in der Regel Geometrien, die den Idealvorstellungen sehr nahe kommen. Sie bewirken somit durch ihr gehäuftes Auftreten sehr starke, kooperative Bindungseffekte.



Die unterschiedlichen Einfärbungen der  $\beta$ -Stränge in Abbildung IV-6a sollen verdeutlichen, dass sie von unterschiedlichen Bindungspartnern des Komplexes stammen und gemeinsam in dieser verschränkten Anordnung ein intermolekulares  $\beta$ -Faltblatt ausbilden. Derartig intermolekular gebildete oder fortgesetzte  $\beta$ -Faltblätter sind mehrfach nachgewiesen und als Merkmal hochspezifischer Kontaktflächen zwischen Proteinmolekülen identifiziert worden [Xu *et al.*, 1997].



**Abb. IV-6a und IV-6b:** Identische Ausschnitte aus der Kontaktfläche des obligatorischen Proteinkomplexes 2ltu\_Ab. Links eine schematische Darstellung, welche den Aufbau eines intermolekular zusammengesetzten  $\beta$ -Faltblatts durch die unterschiedliche, molekülabhängige Einfärbung der einzelnen Stränge verdeutlicht. Rechts eine Darstellung der strukturbestimmten Atompositionen mit grün eingezeichneten Wasserstoffbrückenbindungen zwischen den Stickstoff- und Sauerstoffatomen der Hauptkettenabschnitte.

Abbildung IV-6b ignoriert nun die unterschiedliche Herkunft der einzelnen Hauptkettenabschnitte und zeigt die vorliegenden Atompositionen sowie die regelmäßigen Muster von Wasserstoffbrückenbindungen, welche ausgebildet werden. Bald die Hälfte der eingezeichneten Interaktionen sind im vorliegenden Fall intermolekulare Wasserstoffbrückenbindungen mit besten Eigenschaften und enormer Auftrettsdichte. Auch die überdurchschnittlich gute Qualität dieser Wasserstoffbrückenbindungen im Vergleich zu intermolekularen Bindungen außerhalb von Sekundärstrukturelementen und unter Beteiligung von Akzeptoren oder Donatoren, die in Seitenketten lokalisiert sind, wurde bereits manifestiert [Xu *et al.*, 1997].

Untersucht man für derartige Interaktionen einmal die Stellung der Vektoren, die die einzelnen Bindungsstellen in diesen Methoden repräsentieren, da die Positionen der eigentlichen Bindungspartner ja unbekannt sind, so ist auch hier eine überproportional gute Qualität hinsichtlich Abstand und Ausrichtung zur lösungsmittelzugänglichen Moleküloberfläche zu attestieren. Die Interaktionsvektoren stehen zumeist fast optimal senkrecht unter der Oberfläche und weisen gradlinig zum potentiellen Partner.

Die übrigen obligatorischen Proteinkomplexe besitzen diese intermolekularen Sekundärstrukturelemente in unterschiedlichem Maße, bezüglich sowohl Häufigkeit als auch Ausdehnung. Fakultative Komplexe können entsprechende Strukturen besitzen, wobei Menge und Qualität nicht vergleichbar sind mit den Entsprechungen in obligatorischen Komplexen.

Angesicht der ausgedehnten, intermolekular parallelen Anordnung der Proteinhauptketten in diesem Beispiel wird ersichtlich, wieso trotz vergleichbarer Ausdehnung der Kontaktflächen sowohl die Anzahl beteiligter Aminosäuren als auch die Anzahl ausgebildeter, intermolekularer Wasserstoffbrückenbindungen bei obligatorischen Proteinkomplexen derart höher ist als bei fakultativen Komplexen. Ebenso kann anhand der regelmäßigen Bindungsanordnungen nachvollzogen werden, weshalb auch die Menge kriterienkonformer Paar- und Triplettszusammenstellungen, insbesondere bei den stärker selektierenden Grenzwertensembles, entsprechend höher ausfällt.

## **12.6 Überlegungen zu einem präselektiven Eignungstest einzelner Proteinmoleküle für die vorgestellten Methoden und Grenzwertensembles**

In diesem Ergebniszusammenhang taucht erstmals die Fragestellung auf, ob ein Proteinkomplex eigentlich für diese Methode des Dockings mittels intermolekularen Wasserstoffbrückenbindungen geeignet ist. Unter Berücksichtigung der tatsächlich zugrunde liegenden Problemstellung muss die Problematik sogar weiter geführt werden. Diese Eignung muss eigentlich aus Eigenschaften der zwei unabhängig von einander kristallisierten und in ihrer Struktur aufgeklärten Einzelmoleküle, welche miteinander im Docking ausgetestet werden sollen, abgeleitet werden.

Es gilt zudem, in Abhängigkeit vom anzuwendenden Grenzwertensemble die Fähigkeit einzuschätzen, eine ausreichende Anzahl intermolekularer Wasserstoffbrückenbindungen in notwendiger Dichte mit guter Qualität ausbilden zu können. Hierbei ist zu beachten, dass sich die Moleküloberflächen einer Proteinkontaktfläche im gelösten monomolekularen Zustand und

im komplexierten bimolekularen Zustand in ungewissem Maße unterscheiden, wie die nachgewiesene Existenz der geringfügigen Konformationsänderungen zur Bindung (*induced fits*) belegt. Insbesondere in den vorliegenden, explizit auf die einzelnen Atompositionen bezogenen Methoden, ziehen bereits kleine Positionsänderungen große Effekte nach sich.

Die in dieser Arbeit vorgestellte Voruntersuchung soll eine Untergruppe möglicher potentiell produktiver Molekülorientierungen für eine nachfolgende, zeitlich weitaus aufwendigere Hauptuntersuchung erzeugen. Werden ihr Einzelmoleküle unterzogen, die überhaupt nicht in der Lage sind, die notwendigen Eigenschaften zu erbringen, so wird die einzige richtig positive Orientierung eventuell verworfen. Sie kommt daher in der Hauptuntersuchung nicht zur Evaluierung und kann somit nie aufgedeckt werden, selbst wenn ein vollständiges Geometrisches Docking als alleinstehende Untersuchungsmethode dazu eventuell in der Lage wäre.

Die hier aufgeworfene Fragestellung eines nochmal vorgelagerten, präselektiven Eignungstests ist eindeutig nicht Gegenstand dieser Arbeit. Ihre Bedeutung wird an dieser Stelle nur erstmals ganz offensichtlich, da die parameterabhängige Eignung vieler bislang unauffälliger Untersuchungsobjekte widerlegt werden konnte. Zur Verfügung stehende Eigenschaften einzelner Proteinmoleküle in diesem Zusammenhang sind z.B. die Ausmaße der Moleküloberfläche und die Anzahl der oberflächennahen und bindungsfähigen Atome. Die hieraus zusammenstellbare Dichte bindungsfähiger Atome ließe sich gegebenenfalls verfeinern durch Identifikation von separaten Molekülbereichen mit unterschiedlicher Eigenschaftsausprägung. Rudimentäre Untersuchungen in der Methodenentwicklung haben hier keine eindeutigen Korrelationen und offensichtliche Entscheidungshilfen aufgezeigt.

Sollten sich entsprechende Einschätzungshinweise ausfindig machen lassen und ließe sich ein präselektiver Eignungstest implementieren, so muss man sich des weiteren Risikos bewusst sein, das man hier geschaffen hat. Bedeutet der Eignungstest, dass zwei Einzelmoleküle für eine der vorgestellten Methoden und eines der Grenzwertensembles geeignet sind, so verlässt man sich fortan darauf, dass das Docking mit Wasserstoffbrückenbindungen die richtig positive Molekülorientierung bestimmen und auch weiterqualifizieren kann. Die Gefahr der präselektiven Fehleinschätzung hat sich somit vorerst nur verlagert bzw. weitere Angriffspunkte gefunden. Disqualifiziert der Eignungstest für die Voruntersuchung, so wird allenfalls ohne Gefahr des frühzeitigen Verlustes die eigentliche Hauptuntersuchung verstärkt gefordert, unter

unspezifischeren Molekulantorientierungen und unlimitiertem Rechenaufwand das richtige Ergebnis selbstständig zu finden.

## 13 Algorithmen der Kombinatorik und die Spezifität erzielbarer Suchergebnisse

Nachdem im Methodenteil die Auftrennung der Arbeitsschritte mit ihren multiplen Mechanismen zur Vorselektion der Elemente einer Molekülrepräsentation ausgeführt wurden und zudem eine Beschreibung der Test-Datensätze erfolgt ist, soll hier das Problem der Kombinatorik erörtert werden. Diese Problematik taucht immer auf, wenn zwei Punktmengen miteinander verglichen und auf partielle Übereinstimmungen überprüft werden sollen. Es ist grundlegend für die Dauer derartiger Evaluierungen und gibt Hinweise auf die Spezifität erzielbarer Resultate. Diese drückt sich in der Relation der Menge richtig positiver zur Menge falsch positiver Suchergebnisse aus. Die folgenden Ausführungen der beiden Themenkreise bilden eine Grundlage der Ursachen und Begründungen aller nachfolgenden Ergebnisblöcke.

### 13.1 Kombinatorik der Suchmethode

Es wurde bislang darauf verzichtet für die Test-Datensätze auf die Zahl ihrer bindungsfähigen Atome sowie auf die Anzahl der daraus zusammenstellbaren Paare bzw. Triplets und der wiederum hieraus intermolekular kombinierbaren Paar-Paar- und Triplet-Triplett-Evaluierungen einzugehen.

Zunächst einmal soll Tabelle IV-14 eine Übersicht geben, wie viele Paare und Triplets sich überhaupt rechnerisch aus gegebenen Punktmengen bilden lassen. Die Berechnungsformeln für diese nichtredundanten Kombinationen aus einer Menge von  $n$  Punkten bzw. Atomen finden sich im Abbildungskasten IV-7.

Des Weiteren werden in der Tabelle Abschätzungen angegeben, wie viele Kombinationen aus den Paaren oder Triplets nichtredundant erzeugt werden können. Hierbei sind zudem die bereits vorgestellten multiplen Anordnungen zu berücksichtigen, um alle eventuellen Atom-Atom-Interaktionen zu beachten. Für Paare ist das doppelte Produkt der Atompaare ausreichend, da jede Kombination stets in zwei Anordnungen evaluiert werden muss. Für Triplets entspräche dieses einer Verdreifachung des Produkts. Hinzu kommt jedoch noch die gewendete Anordnung, bei der die Dreiecksseiten anders zusammengelegt werden, woraus eine Versechsfachung des Produkts resultieren würde. Durch den Ansatz, die bindungsfähige Seite eines Atombereichs zu identifizieren sobald dieses eindeutig möglich scheint, kann gegebenenfalls eine Dreiecksseite hier ausgenommen werden. Es wird also versucht, die Evaluierungsaufrufe so weit wie möglich

einzudämmen, weshalb der angegebene Berechnungsterm in Abbildung IV-7 sowie die Werte der Tabellenspalte also nur die optimal geringste Anzahl notwendiger Aufrufe darstellt.

# Atompaaare:	$\sum_{m=1}^{m=n-1} n - m$	entspricht	$\frac{n * (n - 1)}{2}$	für $n > 1$
# Atomtriplets:	$\sum_{o=1, m=2}^{o=n-2, m=n-1} o * (n - m)$	entspricht	$\frac{n * (n - 1) * (n - 2)}{6}$	für $n > 2$
# Atompaar-Kombinationen: #Atompaaare_A * #Atompaaare_B * 2				
# Atomtriplett-Kombinationen: #Atomtriplets_A * #Atomtriplets_B * 3				

**Abbildung IV-7:** Berechnungsformeln für die Zusammenstellung nichtredundanter Atompaaare bzw. -triplets und ihrer Kombinationen aus einer gegebenen Menge von  $n$  Atomen. (# - Gesamtanzahl)

# Atome	# Atompaaare	# Atompaar-Kombinationen	# Atomtriplets	# Atomtriplett-Kombinationen
2	1	2	-	-
3	3	18	1	3
4	6	72	4	48
5	10	200	10	300
6	15	450	20	1.200
8	28	1.568	56	9.408
10	45	4.050	120	43.200
25	300	180.000	2.300	15,9 e+06
50	1.225	3,0 e+06	19.600	1.200,0 e+06
100	4.950	49,0 e+06	161.700	78.400,0 e+06
250	31.125	1.900,0 e+06	2,6 e+06	19,8 e+12
500	124.750	31.100,0 e+06	20,7 e+06	1.300,0 e+12
1.000	499.500	499.000,0 e+06	166,2 e+06	82.800,0 e+12

**Tabelle IV-14:** Kombinatorische Entwicklung in der Zusammenstellung von Atompaaaren bzw. -triplets und ihrer Kombinationen. Um die Lesbarkeit zu erhöhen, wurde die wissenschaftliche Exponentialschreibweise für Werte über einer Millionen (1,0e+06) und einer Billionen (1,0e+12) abgewandelt.

Überaus auffällig ist der Zuwachs der Funktionswerte der Atompaar- bzw. -triplett-Zunahme mit fast der zweiten bzw. dritten Potenz. Diese wird nochmals potenziert in der jeweiligen Anzahl intermolekularer Evaluierungs-Kombinationen. So wären rein rechnerisch bei nur 100 bindungsfähigen Atomen bereits 49 Millionen Paar-Paar- bzw. 78 Milliarden Triplet-Triplett-Evaluierungen von Nöten. Reale Test-Datensätze weisen durchschnittlich 200 bis 250 bindungsfähige Atome auf. Selbstverständlich reduziert sich diese Anzahl erheblich durch die Vorselektion der Atompaare und -tripletts mit den Grenzwerten der monomolekularen transformations-invarianten Eigenschaften.

Tabelle IV-15 soll nun für die Testgruppe der obligatorischen Proteinkomplexe die realen Mittelwerte und Standardabweichungen bindungsfähiger Atome und kriterienkonformer Atomzusammenstellungen wiedergeben. Hierbei werden alle Einzelmoleküle zusammengefasst und nicht nach Rezeptor und Ligand unterschieden. Es muss hingegen differenziert werden nach der optionalen Stufe der Atomdisqualifikation aufgrund intramolekularer Wasserstoffbrückenbindungen und nach den verwendeten Grenzwertensembles. Um die große Varianz innerhalb der Testgruppe zu verdeutlichen, werden zudem die jeweiligen minimalen und maximalen Observationswerte in Klammern angegeben.

Fakultative Komplex unterscheiden sich nicht wesentlich für die hier betrachteten Eigenschaften.

	alle Atome	ungesättigte Atome	ungebundene Atome
# Atome	244 ±159 (80 - 819)	221 ±139 (75 - 729)	199 ±121 (71 - 644)
# Paare			
eng	741 ±489 (202 - 2477)	591 ±367 (167 - 1918)	489 ±287 (147 - 1555)
mittel	1019 ±682 (264 - 3431)	807 ±506 (221 - 2647)	668 ±395 (193 - 2141)
weit	2446 ±1801 (576 - 8746)	1934 ±1348 (492 - 6739)	1561 ±1026 (434 - 5247)
# Triplets			
eng	588 ±380 (125 - 1876)	411 ±249 (91 - 1307)	324 ±185 (76 - 994)
mittel	1133 ±724 (228 - 3557)	781 ±459 (175 - 2406)	618 ±340 (139 - 1823)
weit	7714 ±5861 (1370 - 27114)	5205 ±3746 (1028 - 18130)	3806 ±2546 (838 - 12770)

**Tabelle IV-15:** Reale Mittelwerte, Standardabweichungen sowie Minimum- und Maximumwerte (in Klammern) für die reale Anzahl bindungsfähiger Atome und hieraus gebildeter kriterienkonformer Atompaare und -tripletts. Alle Angaben für die Testgruppe obligatorischer Proteinkomplexe unter Verwendung der engen, mittleren oder weiten Grenzwerte und den optionalen Stufen der Atomdisqualifikation.

Grob zusammenfassen zeigt sich, dass durchschnittlich etwa 200 bis 250 bindungsfähige Atome qualifiziert werden.

Diese erzeugen je nach Grenzwertensemble 500 bis 2500 Atompaare. Hierbei ist von engem nach mittlerem Ensemble eine Zunahme um etwa 30% und zwischen mittlerem und weitem Ensemble um etwa 150% zu verzeichnen. Die Unterschiede zwischen jeweilig beobachtetem Minimum und Maximum betragen zunehmend mit Atomanzahl und Abnahme der Selektionskraft etwas mehr als einen Zehnerpotenz.

Ebenso werden etwa 300 bis 8000 Atomtripletts erzeugt mit analogen Zuwachsraten von fast 100% bzw. 600% bis 700%. Diese deutlich höheren Zuwachsraten für die Triplets wurde bereits mit den Kombinatorikalgorithmien begründet. Auch die Zunahme der Varianz mit der Atomanzahl und der Selektionskraft-Abnahme ist entsprechend abzuleiten. Sind es bei den Atompaaren noch Unterschiede zwischen 576 und 8746 Paaren für alle Atome und das weite Grenzwertensemble, so werden hier bereits zwischen 1370 und 27114 Triplets kriterienkonform zusammengestellt.



Abschließend soll Tabelle IV-16 mit analogen Angaben die realen Zahlen der Evaluierungsaufrufe bimolekularer transformations-invarianter Eigenschaften im Moleküldocking der obligatorischen Proteinkomplexe zusammenfassen.

Diese entsprechen nicht einfach den Funktionswerten oben angeführter Formeln, wenn sie auf die gerade präsentierten Mittelwerte realer Observationen von Atompaaren und -tripletts angewendet werden. Entgegen den Annahmen der exemplarischen Wertentwicklung interagieren im Docking von Heterooligomeren unterschiedlich große Moleküle. Die Determinierung der Dreiecks-Bindungsseite gelingt nicht immer, weshalb gegebenenfalls doppelseitige Evaluierungen durchzuführen sind. In der vorliegenden Implementierung werden zudem nicht alle Kombinationen zur Evaluierung aufgerufen, sondern nur derartige Partner aus der indizierten Liste weiter untersucht, deren Atomabstand innerhalb der kriterienkonformen Grenzen liegt.

Alle Angaben der Tabelle erfolgen in Millionen ( $1,0e+06$ ), um die Lesbarkeit zu erhöhen.

	alle Atome	ungesättigte Atome	ungebundene Atome
<b>Paar-Docking</b>			
eng	1,1 ±1,0 (0,2 - 3,7)	0,7 ±0,6 (0,1 - 2,3)	0,5 ±0,4 (0,1 - 1,4)
mittel	2,0 ±1,8 (0,4 - 6,6)	1,3 ±1,1 (0,3 - 4,1)	0,9 ±0,7 (0,2 - 2,6)
weit	11,8 ±11,1 (1,8 - 38,2)	7,5 ±7,0 (1,1 - 23,3)	4,8 ±4,1 (0,8 - 12,4)
<b>Triplet-Docking</b>			
eng	1,0 ±0,8 (0,2 - 3,2)	0,5 ±0,4 (0,1 - 1,5)	0,3 ±0,2 (0,05 - 0,8)
mittel	3,6 ±2,8 (0,7 - 10,2)	1,8 ±1,4 (0,4 - 5,0)	1,1 ±0,8 (0,2 - 3,1)
weit	170,5 ±163,0 (20,1 - 547,3)	80,4 ±79,4 (9,6 - 286,7)	42,7 ±38,9 (5,8 - 145,3)

**Tabelle IV-16:** Reale Mittelwerte, Standardabweichungen sowie Minimum- und Maximumwerte (in Klammern) für die Anzahlen von Evaluierungsaufrufen im Moleküldocking. Alle Angaben für die Testgruppe obligatorischer Proteinkomplexe unter Verwendung der engen, mittleren oder weiten Grenzwerte und den optionalen Stufen der Atomdisqualifikation erfolgen in Millionen ( $1,0e+06$ ).

Bereits im allergeringsten Falle unter maximaler Selektion und größtmöglicher präselektiver Atomdisqualifikation beträgt die reale Menge der Aufrufe durchschnittliche 500.000 Paar- und 310.000 Triplettevaluierungen. Diese Zahlen steigern sich Grenzwert- und

Atomdisqualifikationsabhängig auf durchschnittliche 11,8 Millionen Paar- und 170,5 Millionen Triplett-Evaluierungsaufrufe. Innerhalb der jeweiligen Disqualifikationsstufe entspricht das einer Zunahme um den Faktor 10 bei Paaren und den Faktor 140 bis 170 bei Triplets. Die jeweiligen Maximalobservationen müssen gar mit 38,2 Millionen Paar- bzw. kombinatorisch enorm vervielfältigten 547,3 Millionen Triplett-Aufrufen registriert werden. Hierbei ist zudem noch zu berücksichtigen, dass eine einzige Triplett-evaluierung drei vollständigen Paarevaluierungen inklusive dreimaliger Rotationsbestimmung entspricht, wenn sie für einen kriterienkonformen Kandidaten bis zum Ende durchgeführt wird. Die Unterschiede der jeweiligen minimalen und maximalen Observationen betragen zwischen 1430% im stark selektierten Paar-Docking mit ungebundenen Atomen und 2630% im schwach selektierten Triplett-Docking mit allen Atomen.

Trotz der enormen Summen, die hier vorgestellt werden, sind diese Zahlen wie bereits angemerkt bei Weitem nicht so groß wie es die vorgeführten mathematischen Kombinatorik-Grundlagen erwarten lassen. An dieser Stelle soll damit vorerst nur aufgezeigt werden, welche Größenordnungen von Evaluierungsaufrufen und damit verbundenem Rechenaufwand auftreten und wie groß die Unterschiede nicht nur in Abhängigkeit von Methoden- und Grenzwertwahl, sondern auch innerhalb der Testgruppen sind. Zudem wird nochmals verdeutlicht wie groß die Bedeutung aller Selektionsschritte von der Atomqualifizierung über die Paar- und Triplettzusammenstellung mit jeder verwendbaren Vorselektionsmöglichkeit bis zum eigentlichen Evaluierungsaufruf ist. Die optionale Atomdisqualifikation aufgrund intramolekularer Wasserstoffbrückenbindungen zeigt außergewöhnliche Spareffekte, darf jedoch nicht zu jeder Gelegenheit verwendet werden, wie sich herausstellen wird.

### 13.2 Spezifität von Ergebnissen

Suchmethoden besitzen einen grundlegenden Zielkonflikt. Einerseits gilt es, möglichst viele richtig positive Ergebnisse zu ermöglichen und auch wiederzufinden. Andererseits müssen die richtig positiven Ergebnisse auch qualifiziert werden, um sie zuletzt in ihrer Bedeutung gegenüber den falsch positiven Ergebnissen kenntlich zu machen. Hierzu wird ein möglichst hoher Anteil richtig positiver Ergebnisse an allen Evaluierungsergebnissen benötigt. Richtige Ergebnisse dürfen in einer derartigen Voruntersuchung weder verloren gehen noch unzureichend qualifiziert werden, da die hiervon abhängige Hauptuntersuchung das einzig korrekte Ergebnis dann nicht evaluiert und somit nicht finden kann.

Als zusätzliche Verschärfung der Problemstellung im vorliegenden Fall ist neben der präselektiven besonders die zeitkritische Natur des Vorfilters zu werten. Die Summe der eigenen

Dauer und der eingegrenzten Dauer der Hauptuntersuchung soll geringere Werte erzielen als die Laufzeit einer vollständigen Hauptuntersuchung.

Weite Grenzwerte einer Selektion ermöglichen mehr richtig positive Ergebnisse, wie die Vorstellung der Test-Datensätze gezeigt hat. Hier konnte eine drei- bis vierfache Zunahme der Referenzpaare und eine sieben- bis zwölfache Zunahme der Referenztripletts durch die Verwendung weiter statt enger Selektionsgrenzwerte belegt werden. Diese erzeugt jedoch zugleich eine Zunahme der Evaluierungsaufrufe um das Zehnfache im Paar-Docking und um einen Faktor zwischen 140 und 170 im Triplet-Docking.

Hieraus können noch keine direkten Schlußfolgerungen auf die Zunahmequoten kriterienkonformer Evaluierungsergebnisse gezogen werden. Die unmittelbaren Abhängigkeiten von Molekülgröße, Evaluierungsaufrufen, Anzahl aller –insbesondere falsch positiver– Kandidaten und der Laufzeit können aber bereits bestätigt werden. Die Teilergebnisse hierzu werden in den letzten beiden Ergebnisabschnitten vollends zusammengeführt.

Die hier vorgeführten Teilergebnisse ermöglichen ebenso keinen Einblick in die Qualität einzelner Methoden im Arbeitsablauf. Sie werden jedoch einerseits benötigt um nachzuvollziehen, weshalb bei fast allen selektionierend wirkenden Evaluierungsstufen nie 100%-Erfolgsquoten für Referenzergebnisse erzielt werden.

Dieses ist prinzipiell möglich – aber die Kosten, einzelne Referenzergebnisse mit ihren suboptimalen Werten durch entsprechend geweitete Selektionsgrenzen zu erhalten, potenzieren sich geradezu selbst in der Anzahl resultierender Evaluierungsaufrufe, in der benötigten Laufzeit eines Dockings und in einer fatalen Herabsetzung der Ergebnisspezifität. Die Notwendigkeit, dennoch derart die Selektionskraft zu senken wie es bereits bei der Verwendung aller Atome mit dem weiten Grenzwertensemble praktiziert wird, hat sich schon bei der Betrachtung der Referenzzahlen und generellen Eignungen der Testgruppen herausgestellt. Insbesondere fakultative Proteinkomplexe benötigen im Prinzip jedes Referenzergebnis, das für sie erzeugt werden kann.

Somit zeigt sich bereits zu diesem Zeitpunkt der bereits häufig angeführte Spezifitätsverlust bezüglich der Auffindung richtig positiver Ergebnisse. Ihr Anteil an der Menge aller zuletzt verbleibender Kandidaten nimmt stetig ab, ist für den präselektiven Filter aber von grundlegender Bedeutung. Der Zielkonflikt hat also allergrößten Einfluss auf den globalen

Anspruch der Methode, der zwischen universeller Gültigkeit, genereller Zuverlässigkeit und sinnvoller Anwendbarkeit zu optimieren ist.

Die Bewertung, wie groß der Spezifitätsverlust für den vorliegenden Fall ist, kann erst nach der Betrachtung der ebenfalls einflussreichen Qualitäten aller weiteren Arbeitsschritte präsentiert werden, da diese zuvor belegen muss, weshalb jede Selektionsgrenze ihren Grenzwert erhalten hat.

## 14 Evaluierungs- und Wiederfindungsquoten

Die Tabellen im Anhang A7 enthalten für das Paar- und das Triplet-Docking obligatorischer und fakultativer Proteinkomplexe unter Anwendung der drei verschiedenen Grenzwertensembles jeweils die Summen und prozentualen Anteile erzielbarer, hiervon auch evaluierter und als kriterienkonforme Kandidaten zuletzt auch gefundener Referenzergebnisse. Es wurde darauf verzichtet, die  $18 \cdot (16+58) = 1332$  Einzelergebnisse aufzuführen.

In der Tabelle IV-17 sind hieraus einige Informationen zusammengetragen. Unter Verzicht auf die Wiedergabe aller absoluten Summen, enthält sie die prozentualen Angaben der jeweiligen Evaluierungs- und Wiederfindungsquoten.

Triplet-Docking	#ref	alle Atome			ungesättigte Atome			ungebundene Atome		
		%eval	%found	(%found)	%eval	%found	(%found)	%eval	%found	(%found)
16 obligatorische K.										
eng	120	99,2	98,3	(99,2)	99,2	98,3	(99,2)	98,3	97,5	(99,2)
mittel	239	92,1	91,6	(99,5)	91,2	90,8	(99,5)	90,0	89,5	(99,5)
weit	923	97,7	90,0	(92,1)	89,3	84,2	(94,3)	84,7	80,9	(95,5)
58 fakultative K.										
eng	42	100,0	97,6	(97,6)	97,6	95,2	(97,6)	83,3	83,3	(100,0)
mittel	93	93,5	93,5	(100,0)	80,6	80,6	(100,0)	72,0	72,0	(100,0)
weit	530	91,5	84,2	(92,0)	69,1	64,5	(93,4)	60,6	56,8	(93,8)
Paar-Docking	#ref	alle Atome			ungesättigte Atome			ungebundene Atome		
		%eval	%found	(%found)	%eval	%found	(%found)	%eval	%found	(%found)
16 obligatorische K.										
eng	399	99,5	97,0	(97,5)	96,7	94,5	(97,7)	92,7	90,7	(97,8)
mittel	519	97,3	96,5	(99,2)	94,0	93,3	(99,2)	90,0	89,2	(99,1)
weit	1253	98,3	95,6	(97,2)	87,2	84,7	(97,1)	80,7	79,2	(98,1)
58 fakultative K.										
eng	286	98,3	94,1	(95,7)	83,6	80,4	(96,2)	78,0	75,2	(96,4)
mittel	376	96,8	94,9	(98,1)	78,5	76,9	(98,0)	73,1	71,5	(97,8)
weit	1107	96,6	92,3	(95,6)	76,0	73,4	(96,7)	67,4	65,9	(97,9)

**Tabelle IV-17:** Übersicht der Evaluierungs- und Wiederfindungsquoten für die möglichen Bedingungskombinationen. Details im Text.

eng, mittel, weit: verwendetes Grenzwertenset

#ref: Summe erzielbarer Referenzergebnisse

%eval: Evaluierungsquote, bzgl. Summe erzielbarer Referenzergebnisse

%found: Wiederfindungsquote, bzgl. Summe erzielbarer Referenzergebnisse

(%found): Wiederfindungsquote, bzgl. Summe evaluierter Referenzergebnisse

Da die nachfolgenden Betrachtungen gewissermaßen “mehrdimensionalen“ Entwicklungstendenzen und multiplen, teilweise überlappenden Begründungen unterliegen, erfolgt diese Generalisierung ohne auf Einzelkomplexe mit ihrer nachgewiesenen Varianz einzugehen. Die einzelnen Tendenzen sollen beschrieben und begründet werden, um Einflüsse der zwei unterschiedlichen Methoden, der unterschiedlichen Selektionsgrenzwerte und der unterschiedlichen Typen von Proteinkomplexen zu verdeutlichen.

#### **14.1 Erzielte Evaluierungsquoten**

Anhand der Evaluierungsquote soll im Wesentlichen bewertet werden, ob die Methoden der Molekülaufbereitung, also der Identifikation bindungsfähiger Atome und ihrer kriterienkonformen monomolekularen Paar- bzw. Triplettzusammenstellungen, fehlerfrei und in vertretbarem Maße verlustfrei funktionieren. Der fehlerhafte Ausschluss essenzieller Atome oder Atompaare bzw. -tripletts vom eigentlichen Docking verhindert, dass die hiervon abhängigen, richtig positiven Ergebnisse überhaupt gefunden werden können. Zusätzlich zu den anderweitig vorgestellten, hauptsächlich geometrischen Kriterien treten hier zum Einen die Atomdisqualifikationen aufgrund der eventuellen Existenz intramolekularer Wasserstoffbrückenbindungen, zum Anderen die Festlegung der Bindungsseite für Atomdreiecke und der damit verknüpften Triplett-Triplett-Anordnungen in Effekt.

Ein Referenzergebnis gilt somit als evaluiert, wenn es mit den durchgehend richtigen Atom-Atom-Zuordnungen den Bewertungen der Akzeptor-Donator-Komplementarität und der bimolekularen transformations-invarianten Eigenschaften zugeführt wird.

Unabhängig vom betrachteten Komplextyp und von der Dockingmethode zeigen die Evaluierungsquoten drei generelle, sich kombinatorisch überlagernde Tendenzen. Um diese Entwicklungen, vor allem aber auch ihre unterschiedlichen Ausprägungen zu verdeutlichen, zeigt Tabelle IV-18 eine nochmalige Reduktion der vorhandenen Daten. Wiedergegeben werden die jeweiligen vier Kombinationsextrema der Zustandsmöglichkeiten “enges oder weites Grenzwertensensemble“ bei “Verwendung aller oder nur ungebundener Atome“.

obligatorisch / Triplet				obligatorisch / Paar		
99,2%	↘	98,3%	enge Kriteriengrenzwerte	99,5%	↘	92,7%
↘		↘		↘		↘
97,7%	↘	84,7%	weite Kriteriengrenzwerte	98,3%	↘	80,7%
alle Atome		ungebundene Atome		alle Atome		ungebundene Atome
100,0%	↘	83,3%	enge Kriteriengrenzwerte	98,3%	↘	78,0%
↘		↘		↘		↘
91,5%	↘	60,6%	weite Kriteriengrenzwerte	96,6%	↘	67,4%
fakultativ / Triplet				fakultativ / Paar		

**Tabelle IV-18:** Übersicht der generellen Tendenzen und Vergleiche für die Evaluierungsquoten. Details im Text.

- Zunächst einmal die Feststellung, dass bei Verwendung strenger Selektionskriterien und aller Atome Evaluierungsquoten zwischen 98,3% und 100% erreicht werden können. Es werden somit z.B. keine Fehler bei der Zuordnung der Bindungsfähigkeiten gemacht.
- Die Quoten sind durchgehend abnehmend in Abhängigkeit von der nachlassenden Selektionsstärke der engen, mittleren und weiten Grenzwertensembles bezüglich der monomolekularen transformations-invarianten Eigenschaften.
- Ebenso sind die Quoten durchgehend abnehmend in Abhängigkeit von der sinkenden Zahl der zur Verfügung stehenden bindungsfähigen Atome, wenn man alle, nur bindungsungesättigte oder nur völlig ungebundene Atome zulässt.
- Zuletzt sind die jeweils vergleichbaren Werte zwischen den unterschiedlichen Komplextypen stets niedriger für die fakultativen Proteinkomplexe.

Der Schlüssel zur Begründung aller Tendenzen ist die Anforderung, die an die Qualität der einzelnen intermolekularen Wasserstoffbrückenbindung gestellt wird. Da die Einzelbindung selbst nie direkt bewertet werden kann, weil sie in der realen Aufgabenstellung ohne vorher bekannten Proteinkomplex nicht strukturgelöst ist, betrifft diese Aussage besonders die Bewertung der Ausrichtung repräsentierender Interaktionsvektoren. Sie stehen im besten Fall wie ein Normalenvektor senkrecht zur Moleküloberfläche und somit in geradliniger Orientierung zur wahrscheinlichen Position des Bindungspartners.

- Die Verwendung des engen Grenzwertensembles qualifiziert zwei bzw. drei Bindungen nur dann, wenn sie alle sehr gut geeignet und aufeinander abgestimmt sind. Die beteiligten Atome besitzen gültige Orientierungen der Interaktionsvektoren zur Moleküloberfläche

hin, welche im Falle des Triplet-Dockings zudem sehr gleichartig sind und die Anforderungen der Triplet-Vektordivergenz unterschreiten. Die Verwendung weniger strenger Selektionsgrenzen erlaubt hingegen Wasserstoffbrückenbindungen, bei denen die einzelnen Interaktionsvektoren weniger optimal geartet sein müssen. Diese Vektoren bestehen den Orientierungstest dabei eventuell nicht, weshalb betroffene Atome als bindungsunfähig eingestuft werden. Zudem werden Triplets von Wasserstoffbrückenbindungen als Referenzen qualifiziert, bei denen die beteiligten Atomdreiecke zu stark divergierende Interaktionsvektoren aufweisen, wodurch das entsprechende Atomtriplett nicht zur Molekülrepräsentation verwendet wird. In beiden Fällen könnten die entsprechenden Selektionsgrenzwerte geweitet werden, wobei jedoch bereits auf die Auswirkungen bezüglich der Zunahme der Evaluierungsaufrufe und der Programmlaufzeiten hingewiesen wurde.

- Mit der zunehmenden Anzahl von Atomen, die aufgrund der möglichen Existenz intramolekularer Wasserstoffbrückenbindungen als bindungsunfähig disqualifiziert werden, steigt die Zahl fehlerhaft ausgeschlossener Atome, wodurch die Evaluierungsquote sinkt. HBPLUS bewertet lediglich die Konformität der Positionsgegebenheiten mit den Kriterien von Baker und Hubbard und vermerkt dann die mögliche Bindung. Hierbei können multiple Interaktionen selbst für Donatoren ausgegeben werden. Es ist bislang nicht möglich, qualitative Unterschiede zwischen den postulierten intra- und intermolekularen Wasserstoffbrückenbindungen festzulegen. Das Ziel hierbei wäre gewesen, eventuell über die minderwertige Güte der intramolekularen Bindung zu argumentieren, dass diese Bindung voraussichtlich zu schwach ist und daher entweder *in vivo* gar nicht stabil existiert oder zugunsten einer optimaleren intermolekularen Bindung geöffnet wird, sobald sich ein geeigneter Interaktionspartner anbietet.

Die Häufigkeit fehlerhafter Atom-Disqualifikationen steigt, wenn die Anforderungen an die Qualität der intermolekularen Wasserstoffbrückenbindung durch die Verwendung weiterer Grenzwertensembles sinken. Wegen der schlechteren Ausrichtung zur Moleküloberfläche, können die Interaktionsvektoren nun Orientierungen einnehmen, die flacher zur Oberfläche und somit besser zu intramolekularen Partnern stehen.

- Beide Entwicklungstendenzen sind stärker bei fakultativen als bei obligatorischen Proteinkomplexen ausgebildet, da gerade hier die generell unterschiedlichen Qualitäten intermolekularer Wasserstoffbrückenbindungen anzutreffen sind. Die strukturellen Unterschiede bezüglich den Bindungen und besonders bezüglich der hohen Qualität und Anzahl einzelner Bindungen innerhalb intermolekular ausgebildeter Sekundärstrukturelemente bei den obligatorischen Komplexen wurde zuvor erläutert. Die



einzelnen Bindungspartner sind viel optimaler und senkrechter zur Moleküloberfläche orientiert. Sie werden daher seltener als bindungsunfähig eingestuft, geraten seltener in Verdacht intramolekulare Interaktionen eingehen zu können und bilden seltener übermäßig divergierende Triplet-Zusammenstellungen.

Alle drei Tendenzen kombinieren im Triplet-Docking zudem mit der zusätzlichen Bewertung der Vektordivergenz. Beim Docking mit weiten Selektionsgrenzwerten unter Verwendung nur ungebundener Atome kommt es zu den größten Verlusten der Evaluierungsquoten. Erzielt werden hier 84,7% für obligatorische und lediglich 60,6% für fakultative Proteinkomplexe.

#### **14.2 Erzielte Wiederfindungsquoten**

In die Wiederfindungsquote geht hauptsächlich die richtige Funktionsweise der Evaluierung bimolekularer transformations-invarianter Eigenschaften ein, die sowohl die geometrischen Eigenschaften aber auch die richtige Festlegung der Bindungseigenschaften zur Gewährleistung der Akzeptor-Donator-Komplementarität umfasst. Zusätzlich kommen speziell für das Triplet-Docking Effekte der Rotationsbestimmung zum Tragen. Es wird zwar noch nicht bewertet, ob die richtige Orientierung vorgeschlagen wird, jedoch müssen die individuellen Bestimmungen für jede einzelne Dreiecksseite ausreichend gleichartige Rotationsvorschriften erbringen.

Ein Referenzergebnis gilt somit als gefunden, wenn hierfür erfolgreich eine Vorschrift zur Molekülorientierung bestimmt werden konnte und es somit der nachfolgenden Auswertung aller anonymisierter Ergebnisse in der Transformationsclusterung und Clusterqualifizierung zugeführt wird.

Die Wiederfindungsquoten zeigen selbstverständlich die selben Tendenzen wie die Evaluierungsquoten, da Ergebnisse nur gefunden werden können wenn sie überhaupt evaluiert worden sind. Demzufolge ist von Interesse, ob die beschriebenen und begründeten Entwicklungen durch die zusätzlich ausgeführten Arbeits- und Selektionsschritte eventuell noch verstärkt oder von anderen Tendenzen überlagert werden. Dieses lässt sich einfacher verfolgen, wenn man zunächst einmal die bisherigen Verluste ignoriert, indem die jeweiligen Anzahlen gefundener Referenzen nicht mit der Menge erzielbarer Referenzen sondern mit der Menge wirklich evaluierter Referenzen verglichen werden. Die entsprechenden prozentualen Angaben zur Bewertung des zwischenzeitlichen Verlustes essenzieller Ergebnisse sind in Tabelle IV-17 jeweils eingeklammert.

Generell können bei der Evaluierung mono- und bimolekularer transformations-invarianter Eigenschaften Verluste auftreten, weil die Referenzen in Einzelfällen mit minimal differierenden Protonenpositionen bestimmt werden. Damit kann es vorkommen, dass in der Referenzidentifikation ein Kandidat Selektionsgrenzwerte gerade noch erfüllt, wohingegen er in dem eigenen Dockingsexperiment gerade außerhalb der Grenzen bestimmt wird.

Im Paar-Docking bestehen unabhängig vom Komplexotyp, vom Grenzwertensemble und den verwendeten Atomen 95,7% bis 99,2% der evaluierten Referenzen auch die Selektion der bimolekularen transformations-invarianten Eigenschaften und die Festlegung der Rotationswinkel. Paar-Paar-Kombinationen unterliegen der Rotationsmittelung und der damit verbundenen Abweichungsbewertung ebenfalls, wenn multiple Protonenpositionen vorhanden sind.

Im Triplet-Docking werden unabhängig vom Komplexotyp und den Atomen sogar Werte von 99,2% bis 100% erreicht, solange das enge oder mittlere Grenzwertensemble zur Selektion verwendet wird. Bei Anwendung der weiten Grenzwerte sinken die Evaluierungsquoten auf 92,0% bis 95,5% und die tatsächlichen Wiederfindungsquoten entsprechend noch ein wenig tiefer.

Dieses findet seine Begründung wieder in der geringeren Anforderung an die Qualität intermolekularer Wasserstoffbrückenbindungen, da die Interaktionsvektoren auch zur Bestimmung der notwendigen Molekülrotation verwendet werden. Strengere Grenzwerte erfordern Atomdreiecke mit sehr gleichartigen Ausrichtungen dieser Vektoren, weshalb sich auch die drei unabhängigen Rotationsbestimmungen der Dreieckskanten nur wenig unterscheiden und gut zur ebenfalls gleichartigen, endgültigen Rotationsvorschrift mitteln lassen. Weniger strenge Grenzwerte erlauben innerhalb der Grenzen für die Triplet-Vektordivergenz abweichende Orientierungen der Interaktionsvektoren. Die resultierenden Rotationsbeschreibungen können so unterschiedlicher ausfallen. Sie werden gemittelt und einzeln mit der Endrotation verglichen, wobei die Abweichung häufiger als zu groß angesehen werden kann. Diese nachträgliche Disqualifikation ansonsten kriterienkonformer Kandidaten aufgrund unzureichender Eindeutigkeit der Orientierungsbeschreibung trifft in diesem Maße nur Triplets und nicht die einfach rotationsbestimmten Paare.

### 14.3 Diskussion der Evaluierungs- und Wiederfindungsquoten

In der Vorstellung der Test-Datenätze wurde bereits gezeigt, dass das enge und mittlere Grenzwertensemble für fakultative Proteinkomplexe aufgrund der geringen Anzahl und Qualität

intermolekularer Wasserstoffbrückenbindungen bis auf wenige Ausnahmen ungeeignet ist, da es nicht ausreichend Referenzergebnisse zulässt.

Generell steigen mit abnehmender Selektionsstärke der Grenzwertensembles natürlich die Zahlen erzielbarer Referenzergebnisse. Dementgegen wurde hier dargelegt, dass die Gefahr der fehlerhaften Atomdisqualifikation bei der Verwendung des weiten Ensembles signifikant steigt. Ebenso gewinnen methodenbedingte Effekte – wie die Triplet-Disqualifikation aufgrund zu großer Vektordivergenz und nachträgliche Kandidaten-Disqualifikation aufgrund uneinheitlicher Rotationsbestimmung – an Einfluss und senken die Evaluierungs- und Wiederfindungsquoten. Wie bereits angedeutet und später belegt, verbietet sich aber die Verwendung weiter Selektionsgrenzen unter Berücksichtigung aller Atome wegen der immensen und stark laufzeitbestimmenden Anzahl notwendiger Evaluierungsaufrufe und der mangelnden Spezifität der Endergebnisse. Diese mangelnde Spezifität macht es in der Regel unmöglich die richtig positiven Ergebnisse überhaupt auszusondern und in ihrer Bedeutung gegenüber der Masse falsch positiver Ergebnisse kenntlich zu machen. Bemühungen, dieses schlechte Verhältnis von richtig positiven Ergebnissen zu Evaluierungsaufrufen bzw. falsch positiven Ergebnissen zu verbessern, wurden bereits angesprochen. So kann im Triplet-Docking durch die Verwendung nur ungebundener Atome die Anzahl der Evaluierungsaufrufe um 75,9% gesenkt werden bei einem bedeutend geringerem Rückgang der Evaluierungsquoten von 97,7% auf 84,7%.

Für diese neuen Tendenzen gilt es also ebenfalls Risiko und Nutzen abzuwägen. Eine Minimierung der Verlustrisiken führt jedoch zwangsläufig zu unzureichend spezifischen Ergebnissen.

Somit verschärft sich nochmals die angesprochene Aufgabenstellung die Eignung eines Molekülkomplexes bzw. der beiden Einzelmoleküle einer Dockingsimulation im Vorhinein einschätzen zu können. Es muss bewertet werden, ob zum Einen die Verwendung im Paar- oder vielleicht auch ein Triplet-Docking möglich ist und ob zum Anderen jeweils genügend Referenzergebnisse bei der Verwendung des engen, mittleren oder weiten Grenzwertensembles gegeben sind. In letztem Fall ist zu differenzieren, ob es immer noch genügend Referenzen sind, damit ein Docking mit nur ungebundenen Atomen nicht in Gefahr läuft zu viele der essenziellen Evaluierungen zu verhindern und ob die Anzahl der sich ergebenden Evaluierungsaufrufe noch eine vertretbare Programmlaufzeit erwarten lässt.

## 15 Orientierungsabhängige Erhaltungsquoten

Analog zum vorhergehenden Abschnitt soll Tabelle IV-19 bezüglich der Orientierungsmethoden die spezifische Zusammenfassung der Tabellen im Anhang A7 sein. Unter Verzicht auf die absoluten Summen enthält sie die prozentualen Angaben, welcher Anteil der erzielbaren Referenzergebnisse nach erfolgreicher Passage der Evaluierung bimolekularer transformationsinvarianter Eigenschaften auch dem richtigen Rotationsrepräsentanten zugeordnet werden. Da ausschließlich die Moleküle aufgetrennter, nativer Protein-Protein-Komplexe wieder miteinander kombiniert werden, nachdem eines der Moleküle einer festgelegten Transformation unterzogen wurde, ist nicht nur die Rotation sondern auch die notwendige Translation bekannt.

Triplett-Docking		#ref	alle Atome			ungesättigte Atome			ungebundene Atome		
	%rot		(%rot)	#rot	%rot	(%rot)	#rot	%rot	(%rot)	#rot	
16 obligatorische K.											
eng		120	69,3	(69,5)	82	68,3	(69,5)	83	67,5	(69,2)	81
mittel		239	63,6	(69,4)	152	63,2	(69,6)	151	62,8	(70,1)	150
weit		923	58,6	(65,1)	541	55,4	(65,8)	511	53,5	(66,1)	494
58 fakultative K.											
eng		42	59,5	(61,0)	25	59,5	(62,5)	25	57,1	(68,6)	24
mittel		93	55,9	(59,8)	52	52,7	(65,3)	49	49,5	(68,7)	46
weit		530	44,0	(52,2)	233	34,5	(53,5)	183	30,9	(54,5)	164

Paar-Docking		#ref	alle Atome			ungesättigte Atome			ungebundene Atome		
	%rot		(%rot)	#rot	%rot	(%rot)	#rot	%rot	(%rot)	#rot	
16 obligatorische K.											
eng		399	50,1	(51,7)	200	49,4	(52,3)	197	47,9	(52,8)	191
mittel		519	52,2	(54,1)	271	50,7	(54,3)	263	49,8	(55,3)	256
weit		1253	44,5	(46,6)	558	41,3	(48,7)	517	39,7	(50,2)	498
58 fakultative K.											
eng		286	31,5	(33,5)	90	29,0	(36,1)	83	26,9	(35,8)	77
mittel		376	32,7	(34,5)	123	27,9	(36,3)	105	26,1	(36,4)	98
weit		1107	30,5	(33,1)	338	24,9	(33,9)	276	22,9	(34,8)	254

**Tabelle IV-19:** Übersicht der Erhaltungsquoten nach der Rotationsbestimmung für die möglichen Bedingungskombinationen. Details im Text.  
eng, mittel, weit: verwendetes Grenzwertenssemble  
#ref: Summe erzielbarer Referenzergebnisse  
%rot: Erhaltungsquote, gemessen an der Summe erzielbarer Referenzergebnisse  
(%rot): Erhaltungsquote, gemessen an der Summe gefundener Referenzergebnisse  
#rot: Summe der Referenzergebnisse mit richtig bestimmter Rotation

### 15.1 Einflüsse und Qualität der Rotationsbestimmung

Um auch hier zunächst einmal die eigenständig tendenzbehafteten Effekte der bisherigen Molekülaufbereitung und der eigentlichen Evaluierung vernachlässigen zu können, finden sich

jeweils eingeklammert in der Tabelle 19 die prozentualen Angaben, die sich nicht auf die Anzahl erzielbarer sondern auf die Anzahl gefundener Referenzen bezieht. Bezüglich dieser Erfolgsquote der Rotationsbestimmung können vier generelle Tendenzen beobachtet werden:

- Atomtriplets lassen sich immer bedeutend zuverlässiger orientieren als Atompaaire der gleichen Komplexe.
- Obligatorische Komplexe lassen sich immer bedeutend zuverlässiger orientieren als fakultative Komplexe bei der Verwendung gleicher Methoden.
- Bei Verwendung der engen oder mittleren Grenzwertensembles gelingt die Orientierung zuverlässiger als bei der Verwendung der weiten Selektionsgrenzen.
- Die Erfolgsquote bei zunehmender Atomdisqualifikation aufgrund der möglichen Existenz intramolekularer Wasserstoffbrückenbindungen steigt, wenn auch zumeist in nur geringem Maße.

Für diese Tendenzen können zwei Begründungen angeführt werden:

- Die Rotationsbestimmung im Triplett-Docking erfolgt redundant, so dass Fehlbestimmungen an einer Dreiecksseite durch die Mittelung der Endrotation ausgeglichen werden können.
- Obligatorische Komplexe besitzen qualitativ hochwertigere, intermolekulare Wasserstoffbrückenbindungen, resultierend in optimaler antiparallelen Interaktions- und Orientierungsvektoren. Diese Interaktionsvektoren werden positiv selektioniert bei der Verwendung strengerer Selektionsgrenzwerte. Ebenso steigt ihr Anteil beim Ausschluss von Atomen, deren Interaktionsvektoren derart ungünstig zur Molekloberfläche orientiert sind, dass sie in den Verdacht geraten, anderweitig ihre Bindungsfähigkeiten abzusättigen. Je höher der Anteil eindeutig und optimal ausgerichteter Interaktionsvektoren ist, desto höher fällt die Quote richtiger Orientierungsbestimmungen aus.

Da sich die beschriebenen Effekte überlagern, lassen sich die Zustandskombinationen zum Auftreten der minimal und maximal erzielten Erfolgsquoten einfach nachvollziehen:

- Paar-Docking fakultativer Komplexe mit allen Atomen unter Verwendung des weiten Grenzwertensembles kennzeichnet das Minimum, bei dem lediglich 33% aller bestimmten Referenzen die richtige Rotation zugeschrieben wird. Für obligatorische Komplexe liegt diese Erfolgsquote immerhin bei 46,6%.
- Die maximalen Werte im Paar-Docking finden sich mit ausschließlich ungebundenen Atomen unter Verwendung des mittleren Grenzwertensembles für fakultative Komplexe mit 34,4% und für obligatorische Komplexe mit 55,3%.

Die Minima und Maxima sind unter identischen Bedingungen auch im Triplett-Docking zu finden, liegen aber für fakultative Komplexe mit 52,2% bis 68,7% und für obligatorische Komplexe mit 65,1% bis 70,1% jeweils 20% bis 30% höher.

Alle diese Ergebnisse müssen zudem vor dem Hintergrund gesehen werden, dass die hierbei verwendeten Rotationsrepräsentanten in regelmäßigen 30°-Abständen zueinander angeordnet sind. Diese Rastergröße ist unbefriedigend groß und erfordert, dass eine nachfolgende Dockingmethode einen immer noch recht großen Konformationsraum genauer und zeitaufwendiger untersuchen muss.

Unvermeidliche Schlußfolgerung dieser Erfolgsquoten für die richtige Rotationsbestimmung muss sein, dass die angewendete Orientierungsmethode qualitativ unzureichend ist. Die Abhängigkeit von den einzelnen Atombindungen, aus denen die verwendeten Orientierungsvektoren hergeleitet werden, erweist sich als fatal. Die zwanghaft parallele Ausrichtung der Grundlinien ist artifiziell, wie bereits die Analyse der realen Werte für die Referenzergebnisse der Proteinkomplexe gezeigt hat. Kleine Abweichungen in einzelnen Atompositionen oder in den bestimmten Protonpositionen resultieren aufgrund des einfachen Orientierungsmodells sofort in direkt proportional andersartigen Rotationsvorschriften. Ebenso innerhalb dieser Arbeit ansonsten gänzlich unberücksichtigt ist der Umstand, dass sich die Kontaktflächen eines ungebundenen Moleküls in Lösung, in kristalliner Form und in ihrem Erscheinungsbild im spezifischen Molekülkomplex unterscheiden wird. Daher muss sich auch hier die atombezogene Orientierungsmethode als ungeeignet erweisen.

Weitere Ausführungen zu dieser kritischen und weitreichenden Erkenntnis finden sich im letzten Abschnitt dieses Kapitels.

Unabhängig von den schlechten Resultaten der Molekülorientierung an sich, müssen zu den oben erläuterten Quoten und Tendenzen nun auch noch die Verluste aus der Molekülaufbereitung und den Evaluierungsvorgängen hinzugezogen werden. Deren Tendenzen verlaufen teils mit teils gegen die hier vorgestellten Entwicklungen, so dass sich neue Minima und Maxima unter andersartigen Methoden- und Parameterkombinationen ergeben.

- Im Paar-Docking verbleiben zuletzt die meisten der erzielbaren Referenzergebnisse mit richtig bestimmter Rotation bei der Verwendung aller Atome mit dem engen oder mittleren Grenzwertensemble. Die wenigsten werden beim Docking mit nur ungebundenen Atomen

und den weiten Selektionsgrenzen richtig orientiert. Für fakultative Komplexe liegen diese Quoten zwischen 22,9% und 32,7%. Obligatorische Komplexe können höhere Quoten mit 39,7% bis 52,2% aufweisen.

- Im Triplet-Docking sind die Ergebnisse ein wenig besser mit 30,9% bis 59,5% für fakultative und 53,5% bis 69,3% für obligatorische Komplexe.

Generell bedeuten diese weiten Spannen von 20% bis 50% verbleibender Referenzergebnisse, dass statistisch im schlechtesten Fall alleine fünf und im besten Fall noch zwei richtig positive Ergebnisse im Paar-Docking gegeben sein müssen, damit für eine Referenz der richtige Rotationscluster ermittelt wird. Für multiple Bestimmungen der richtigen Orientierung bedarf es eines entsprechenden Vielfachen, um ansatzweise die Möglichkeit zu erhalten, eine signifikante Überbelegung der richtig positiven Transformationsbeschreibung detektieren zu können. Für Triplets sind zwar die Quoten höher aber, wie bereits ausgeführt die Referenzzahlen um bis zu 90% geringer! Mit sinkender Selektionskraft der Grenzwerte steigen diese Referenzzahlen, es sinkt jedoch die Qualität der verwendeten Wasserstoffbrückenbindungen mit direkten Auswirkungen auf die Qualität der Transformationsbestimmung und überproportionalen Effekten auf die Anzahl falsch positiver Ergebnisse, wie noch gezeigt werden wird.

Tabelle IV-19 enthält die entsprechenden absoluten Angaben, wie viele Referenzergebnisse für die gesamte jeweilige Testgruppe der 58 fakultativen und 16 obligatorischen Proteinkomplexen zuletzt im richtigen Rotationscluster zu finden sind. Ohne hier durchgehend die internen Inhomogenitäten der Testgruppen auszuführen oder auf die Optionen der Atomdisqualifikation einzugehen, lassen sich Rückschlüsse bereits aus den Gesamtsummen ziehen:

- So erweisen sich die fakultativen Komplexe als völlig ungeeignet für ein Triplet-Docking mit engen oder mittleren Kriteriengrenzwerten, da sie selbst im allerbesten Fall nur 52 richtig orientierte Referenztriplets für alle 58 Komplexe vorweisen können. Die Eignung für das Paar-Docking mit engen oder mittleren Selektionsgrenzen wird bei maximal 123 Referenzen vom Einzelfall abhängen. Ausreichend überproportionale Observationen richtiger Ergebnisse dürfen für Vertreter dieses Komplextyps generell nicht erwarten werden. Einzig das Docking mit weiten Grenzwerten kann ernsthaft für fakultative Komplexe in Erwägung gezogen werden mit maximal 233 Referenztriplets und 338 Referenzpaaren.
- Obligatorische Protein-Protein-Komplexe erweisen sich dementsgegen auch nach diesem weiteren verlustreichen Arbeitsschritt als augenscheinlich geeignet für jegliche Methode und jedes Grenzwertenssemble. Es werden Referenzzahlen zwischen 81 und 558 für jeweils

16 Testobjekte nachgewiesen. Allerdings darf auch hier die Varianz innerhalb der Gruppe nicht vollends in Vergessenheit geraten, da im Paar-Docking unter den besten Voraussetzungen (alle Atome und weite Selektionsgrenzen) der ungeeignetste Komplex gerade noch drei Referenzpaare im richtigen Rotationscluster positionieren kann. Unter den schlechtesten (nur ungebundene Atome und enge Selektionsgrenzen) können bereits fünf Komplexe nur maximal zwei Referenzpaare mit richtiger Rotation ausweisen. Im Triplet-Docking sind es im besten Fall bereits zwei und im schlechtesten Fall sogar acht Komplexe mit maximal zwei richtig orientierten Referenzen.

Der Grund, weshalb hier zunächst nur die eingeschränkte Eignung der gesamten Testgruppen kommentiert wurde, soll nur in dem wiederholten Hinweis stecken, wie bedeutend es sein wird, einem einzelnen Proteinmolekül seine Eignung bereits vor der eigentlichen Untersuchung zuschreiben zu können. Und selbst wenn es gelingt, z.B. durch visuelle Untersuchung den voraussichtlichen Komplextyp bestimmen zu können, so kann wohlmöglich keine Garantie für eine ausreichende Eignung gegeben werden, wie die Menge der unzureichend geeigneten obligatorischen Komplexe verdeutlichen soll.

## 15.2 Qualität der Translationsbestimmung

Tabelle IV-20 enthält eine Zusammenfassung der ausführlicheren Tabellen in Anhang A8 für die Translationsbestimmung obligatorischer Proteinkomplexe im Paar- und Triplet-Docking mit den verschiedenen Grenzwertensembles. Bewertet wird jeweils der euklidische Abstand zwischen zwei kartesischen Koordinatenpunkten, was in diesem Fall gleichbedeutend ist mit der Abweichung zweier Translationsbeschreibungen. Es handelt sich hierbei einerseits um die bekannte Ursprungstranslation, der das jeweils zweit-aufgerufene Protein vor der Molekülaufbereitung unterzogen wurde. Andererseits wird die gemittelte Translationsvorschrift aller Referenzergebnisse, die dem richtigen Rotationsrepräsentanten zugeordnet wurden, verwendet. In Tabelle IV-20 wurde nun der Übersicht halber über die Einzelergebnisse der gesamte Testgruppe gemittelt und die Standardabweichung berechnet.



Triplett-Docking	alle Atome		ungesättigte Atome		ungebundene Atome	
	Mittelwert	Std.abw.	Mittelwert	Std.abw.	Mittelwert	Std.abw.
eng	1,35	± 0,24	1,35	± 0,24	1,33	± 0,25
mittel	1,41	± 0,24	1,40	± 0,23	1,39	± 0,25
weit	1,38	± 0,22	1,37	± 0,19	1,38	± 0,20

Paar-Docking	alle Atome		ungesättigte Atome		ungebundene Atome	
	Mittelwert	Std.abw.	Mittelwert	Std.abw.	Mittelwert	Std.abw.
eng	1,40	± 0,22	1,41	± 0,21	1,44	± 0,23
mittel	1,41	± 0,18	1,41	± 0,17	1,43	± 0,22
weit	1,35	± 0,17	1,36	± 0,18	1,37	± 0,18

**Tabelle IV-20:** Die gemittelten Abweichungen und Standardabweichungen der Transformationsvorschriften aller richtig rotationsorientierten Referenzergebnisse obligatorischer Proteinkomplexe zur bekannten Ursprungstransformation. "Eng, mittel und weit" kennzeichnen das verwendete Grenzwertensemble. Alle Angaben erfolgen in Ångström.

An dieser Stelle kann erfreulicherweise berichtet werden, dass keinerlei auffällige Tendenzen zu verzeichnen sind. Alle Translationsabweichungen bewegen sich im Rahmen von 1,33 Å bis 1,44 Å mit moderaten Standardabweichungen von etwa 0,21 Å. Da mit der erfolgten Zuordnung zum richtigen Rotationsrepräsentanten auch die exakt richtige Rotation vor der Translationsbestimmung veranlasst wird, entsprechen diese Werte zugleich den *RMSD*-Werten der Endposition des transformierten Moleküls zu seiner Originalposition im nativen Komplex. Die *root mean square deviation* beschreibt die mittlere Lageabweichung aller Atompositionen zu ihren Originalpositionen in der nativen Konformation.

Diese Konformität lässt die Vermutung zu, dass das einfache Orientierungsmodell, bei dem als Translationsbestimmung einfach die Mittelpunkte zwischen den jeweiligen Interaktionspunkten der Atompaaire bzw. -tripletts nach der Rotation übereinander geschoben werden, einen auffälligen jedoch stets gleichartigen Fehler verursacht. Voraussichtlich müssen diese Mittelpunkte noch ein wenig von den Grundlinien entlang der Mittelvektoren aus dem Orientierungsmodell verlegt werden.

Der Fehler in der Verschiebung ist einheitlich und hat somit keinerlei Einfluss auf die nachfolgenden Arbeitsschritte der Ergebnisclustering. Diese achtet nur auf relative Abstände der Kandidaten untereinander, weshalb auf eine genauere Fehlerbetrachtung und Methodenoptimierung zu diesem Zeitpunkt verzichtet wurde.

### 15.3 Diskussion der Rotationsbestimmung und Überlegungen zu alternativen

#### Orientierungsmethoden

Die Bedeutungen des Orientierungsmodells und der richtigen Bestimmung der Transformationsvorschrift wurden im Methodenteil hergeleitet. Sie liegen darin, dass ein kriterienkonformer Kandidat nach der erfolgreichen Passage der Selektion bimolekularer transformations-invarianter Eigenschaften nur noch in Form seiner Transformationsbeschreibung existiert, indem er mit den Translationswerten einem bestimmten Rotationsrepräsentanten zugeordnet wird. In der weiteren Ergebnisclustering wird er verworfen oder in eine Gruppe mit gleichartigen Transformationsvorschriften aufgenommen. Die größten Gruppen gleichartiger Kandidaten beschreiben zuletzt Molekülorientierungen, für die offensichtlich besonders viele Paare oder Triplets von intermolekularen Wasserstoffbrückenbindungen zeitgleich ausgebildet werden können. Spezialisierte Kontaktflächen in Proteinkomplexen sollten überproportional viele Interaktionen ausbilden können und die notwendige Transformation somit auf den vorderen Plätzen einer Sortierung nach diesen Clusterbelegungszahlen zu finden sein.

In diesem Ergebnisabschnitt zeigt sich jedoch, dass die momentane Methode zur Rotationsbestimmung enorme Verluste in der Erhaltung der essenziellen Referenzergebnisse verursacht. Nur für etwa 35% bis 70% der gefundenen Referenzen wird der richtige Rotationsrepräsentant identifiziert, wodurch die Spezifität, d.h. der Referenzgehalt der finalen Transformationscluster, eklatant herabgesetzt wird. Die Schrittweite der regelmäßig über den gesamten Rotationsraum verteilten Rotationsrepräsentanten ist zudem mit derzeit  $30^\circ$  noch unangemessen groß, wenn man zur Kenntnis nimmt, dass die momentan verwendete Hauptuntersuchung eines Geometrischen Dockings ohne diese Konformations-Vorselektion ein initiales Suchraster von  $20^\circ$  verwendet.

Die Methode zur Rotationsbestimmung stellt sich hiermit vorzeitig als deutlich schwächstes Glied der Abfolge aller implementierten Methoden heraus und hat größten Anteil an einer Minderung der Gesamtqualität endgültiger Ergebnisse.

Der primäre Fehler muss in der Erstellung der Orientierungsvektoren liegen, welche explizit von den Positionen der beteiligten Atome abhängig sind. Einerseits ist eine vollständige Parallelität der Grundlinien aus den verwendeten Vektorkonstrukten nicht gegeben, wie bereits die Betrachtung der Werteverteilung für reale Positionen der Atome von Referenzergebnissen aus den nativen Komplexen aufgezeigt hat. Andererseits sind die verwendeten Interaktionsvektoren

durch eine –bei einzelnen Atomtypen auch durch maximal zwei– kovalente Bindungen benachbarter Atome definiert und unterliegen somit primär Zwängen der Protein-Hauptketten-Geometrien.

Entsprechend dieser Schlußfolgerungen und der Bedeutung des Arbeitsganges wurden auch einige Anstrengungen unternommen, alternative Orientierungsmethoden zu finden, um diesen Übergang von der Identifikation der Berührungspunkte zweier Moleküle zur Ermittlung der notwendigen Transformation fehlerfreier zu gestalten. Das generelle Lösungsbestreben sollte dabei dahin gehen, dass die Beschaffenheit der Oberflächenumgebung repräsentierender Atompaaire und -tripletts mehr Berücksichtigung finden. Es sind keinerlei derartige Informationen über die Oberflächentopographie in den Interaktionsvektoren der Atome enthalten.

Generell denkbar sind andersartige Vektoren, die entweder ein einzelnes Atom oder ein einzelnes Paar bzw. Triplet in der Orientierungsbestimmung vertreten, indem sie anstatt der Mittelvektoren aufeinander ausgerichtet werden. Die Erzeugung dieser neuen Orientierungsvektoren kann durchaus zeitintensiv sein, wenn sie in der Molekülaufbereitung erfolgt und der Vektor dann als Eigenschaft dem Atom, Paar oder Triplet zugeschrieben und permanent abgelegt wird.

- Erste Versuche wurden mit Oberflächennormalenvektoren der Connolly-Oberflächenpunkte unternommen. Hierbei wurden entweder die Normalen der Punkte beteiligter Atome oder Normalen der Punkte in einem gewissen Umkreises um die Paar- und Triplettszusammenstellungen herum verwendet.
- Ebenso wurde mit verschiedenen Varianten von Oberflächentriangulierungen experimentiert. Hierzu wurden alle Atomen, nur oberflächennahe oder nur bindungsfähige Atomen herangezogen. Sogenannte “*α-shapes*“ [Mucke, Edelsbrunner, 1992] entsprechen quasi den lösungsmittelzugänglichen Oberflächen der Connolly-Methoden, werden aber ebenso als Triangulierung der Atomzentren ausgeführt. Allen diesen Dreiecksflächen lassen sich Normalenvektoren aufsetzen. Den Atomen, welche die Eckpunkte bilden, können so Vektoren aus den direkt oder indirekt angrenzenden Flächen zusammengestellt werden. Hierbei lässt sich variieren, wie groß der Einzugsbereich gewählt werden kann.
- Für das Triplet-Docking wurde auch versucht, einfach über die Atomdreiecksnormalen zu orientieren, wenn es gelang, ihre Ausrichtungen “zum Proteininneren“ oder “zur

Moleküloberfläche“ eindeutig zu bestimmen. Diese kann jedoch z.B. nicht geschehen, wenn ein stumpfes Dreieck senkrecht unter der Moleküloberfläche steht. Derartiges ist häufig der Fall bei intermolekular fortgesetzten Sekundärstrukturen aufgrund des Zickzack-Verlaufs der Proteinhauptkette.

Bis zum jetzigen Zeitpunkt ist es jedoch nicht gelungen, eine Methode zu finden und zu implementieren, welche nicht nur in der Betrachtung von Einzelergebnissen sondern in einer universellen Gesamtaufrechnung für alle Referenzen eines Proteins bessere Resultate hätte liefern können.

- Überlegt, jedoch nicht implementiert wurde, ob man für ein Atompaar eine gemeinsame Achse festlegen könnte, um die dann die Moleküle gegeneinander verdreht werden unter Minimierung ihrer van-der-Waals-Überlappung. Derartige Methoden verlangen jedoch ein ausreichend gute Molekülabstraktionen und erfordern eine beträchtlich größere Anzahl an Koordinatentransformationen, selbst wenn jeweils nur lokale Bereiche betrachtet werden. Der bedeutendste Nachteil eines derartigen Vorgehens ist, dass es nicht zur Molekülaufbereitung sondern beim Docking für jeden kriterienkonformen Kandidaten auszuführen wäre.

Zwei Methoden, die voraussichtlich die weiteste Verbreitung bei der Lösung derartiger Problemstellungen besitzen, wurden bislang nicht implementiert. Sie sollen hier aber als Anregung zur Problemlösung erwähnt werden:

- Der Ferro-Hermans-Algorithmus beschreibt eine Methode, mit der zweimal drei oder mehr Punkte miteinander überlagert werden, wobei die Summe der Abstände aller korrelierten Partner minimiert wird (*least square fit*) [Ferro, Hermans, 1977]. Dieses entspricht der Problematik, die bei einer Überlagerung (*fitting*) homologer Körper wie z.B. zweier ähnlicher Proteinstrukturen gelöst wird. Da es zumindest zweimal drei Punkte verlangt, weil die Rotation um eine Achse von vier Punkten nicht bestimmt werden kann, ließe sich die Methode zunächst nur für das Triplet-Docking verwenden. Hierfür sollte zuvor die Ähnlichkeit der Dreiecke der Referenzergebnisse untersucht werden. Für das Paar-Docking gälte es, einen jeweiligen dritten Koordinatenpunkt zu generieren. Dieser könnte z.B. die Spitze des bisherigen Mittelvektors bzw. seiner Inversen sein. Hiermit kommt aber wieder die Frage auf, ob damit etwas gewonnen wird, da die Abhängigkeit vom

Vektorkonstrukt aus den Atompositionen nicht gelöst wird. Somit muss ein anderer, für die Oberfläche repräsentativer Punkt bestimmt werden.

- Die Gruppe um Ruth Nussinov verwendet eine vergleichbare Strategie [Wolfson & Nussinov, 1998], wobei jedoch nicht mit Punkten gearbeitet wird, sondern zur Rotationsbestimmung mehrere Vektoren mit möglichst geringem Gesamtfehler zueinander antiparallel ausgerichtet werden [Schwartz & Sharir, 1987]. Dieses funktioniert somit bereits mit minimal zweimal zwei Vektoren, wobei je ein Vektorenpaar, wie in den hier vorgestellten Vektorkonstrukten, miteinander gekoppelt ist. Erste Versuche könnten bereits mit den bestehenden Interaktionsvektoren unternommen werden. Allerdings wird genau dieses Prinzip mit der bisherigen Methode über die Mittelvektoren ansatzweise imitiert. Um jedoch von den Atompositionen unabhängig zu werden, wären auch hierfür andersartig bestimmte, von der Oberflächentopographie abhängige und den Atomen zugeschriebene Vektoren notwendig.

## 16 Ergebnisqualität vollständiger Dockingexperimente

Nachdem alle verwendeten Kriteriengrenzwerte und die untersuchten Datensätze vorgestellt wurden, konnte erstmals bewertet werden ob und wie Proteinkomplexe für diese Methoden des Dockings oder für einzelne Grenzwertensembles geeignet sein können. Da multiple Observationen derselben Transformationsvorschrift benötigt werden, ist eine ausreichende Anzahl von Referenzergebnissen für die generelle Eignung ausschlaggebend. Ganz allgemein lässt sich eine größere Anzahl kriterienkonformer Suchergebnisse durch die Verwendung selektionsschwächere Grenzwerte erzeugen.

Die Algorithmen, welche die Problematik der stufenweise aufbauenden Kombinatorik dieser Aufgabenstellung beschreiben, wurden ebenso bereits vorgestellt. Sie begründen, weshalb die Verwendung weiterer Selektionsgrenzwerte eine vielfach höhere Zunahme der durchzuführenden Evaluierungsaufrufe bewirkt als eine Zunahme der Referenzergebnisse. Schließt man von der Menge initiiertes Evaluierungen auf die Anzahl erfolgreicher Passagen aller Selektionen, so bedeutet dieses eine deutlichere Verschlechterung des wichtigen Verhältnisses richtig positiver zu falsch positiven Suchergebnissen. Dieser Spezifitätsverlust der richtig positiven Kandidaten erschwert die weitere Aufgabenstellung, die Bedeutung dieser Referenzen gegenüber der Masse falsch positiver Ergebnisse in der Ergebnisqualifizierung kenntlich zu machen.

Die detaillierte Beschreibung der erzielten Evaluierungs-, Wiederfindungs- und orientierungsabhängigen Erhaltungsquoten haben eine Bewertung einzelner Arbeitsschritte hinsichtlich dem Verbleib der Referenzergebnisse ermöglicht. Hierbei wurde deutlich, dass in Abhängigkeit von den verwendeten Grenzwertensembles und von den optionalen Disqualifikationen von Atomen mit eventuellen intramolekularen Wasserstoffbrückenbindungen Verluste zwischen 0% und 40% in der Wiederfindung auftreten. Diese müssen hingenommen werden, da weitere Grenzwerte zwar mehr Referenzen zulassen, jedoch zugleich das Risiko fehlerhafter Ergebnisverluste erhöhen, wenn man eine ungehinderte Zunahme falsch positiver Ergebnisse unterbinden möchte. Den bislang mit Abstand bedeutendsten Kritikpunkt aller vorgestellten Einzelmethoden stellt die Orientierungsbestimmung kriterienkonformer Kandidaten dar. Für 30% bis 65% der Kandidaten wird die falsche Rotationsvorschrift und ein verkehrter Rotationsrepräsentant bestimmt. Dieses verschärft das Problem der Ergebnisspezifität, da die Gruppe der richtig orientierten positiven Ergebnisse entsprechend an ihrer Bedeutung einer geschlossenen Einheit verliert.

Dieser Ergebnisabschnitt betrachtet nun erstmals den Gesamtablauf des Dockings unter Berücksichtigung aller richtig und falsch positiven Kandidaten. Unverändert interessiert insbesondere der Verbleib der Referenzergebnisse bzw. das Abschneiden der einzig richtigen Molekülorientierung. Diese ist bekannt, so dass man für den entsprechenden Rotationsrepräsentanten stets bestimmen kann, wie viele dieser Kandidaten auch Referenzergebnisse sind und welchen Rang innerhalb einer absteigenden Sortierung nach Belegungszahlen er einnimmt.

Es werden drei Zeitpunkte zur Bewertung gewählt:

- Die Situation nach der Rotationsbestimmung,
- nach der Ergebnisausdünnung mittels einer Nachbarsuche der Translationsvorschriften
- und in den finalen Transformationsclustern. Diese werden vom jeweils größten *Single-Linkage-Cluster* von Kandidaten mit einer gleichartigen Translation innerhalb jeder der 564 Rotationsgruppen gebildet.

In diesem Kapitel soll erläutert werden, ob und bei welchen Methoden- und Parameterkombinationen es möglich ist, die einzig richtige Molekülorientierung in einer ausreichend guten Sortierungsposition zu finden. Es wird auch betrachtet, wie gut dann die Bestimmung der Translation gemessen an der bekannten Ursprungsverschiebung erfolgt. Ebenso soll versucht werden einen Rückschluss zu ziehen, wie viele Referenzergebnisse ein Komplex von Proteinmolekülen voraussichtlich aufweisen muss, damit die Spezifität dieser Ergebnisse für eine erfolgreiche Bestimmung ausreicht.

### 16.1 Docking obligatorischer Proteinkomplexe

Die Betrachtung erfolgt zunächst einmal für die obligatorischen Komplexe, da für fast alle Methoden und Grenzwertensembles die Rechnungen durchgeführt werden können. Mit den Ausnahmen von drei Komplexen im Triplett-Docking mit mittleren Grenzwerten und sechs Komplexen im Triplett-Docking mit engen Grenzwerten, kann jederzeit mehr als ein Referenzergebnis vorgewiesen werden, wodurch eine multiple Bestimmung der richtigen Orientierung an sich möglich ist. Somit können alle Tendenzen hier an einer weitestgehend identischen Datengrundlage vorgestellt werden.

Für das Paar- und das Triplett-Docking mit weiten Selektionsgrenzen ergeben sich Lücken in den folgenden Ergebnistabellen wenn ein Docking nicht vollständig durchgeführt werden kann. Dieses betrifft die größten Komplexe, bei denen aufgrund der Kombinatorik entweder der Bedarf an Arbeitsspeicher und/oder die Rechenzeit zu groß wird. Einerseits muss ein Limit bei

6000 Zuordnungen für einen Rotationsrepräsentanten gesetzt werden, da ansonsten keine Bearbeitung der translationsabhängigen Vorgänge stattfinden kann. Andererseits werden hierfür Rechenzeiten von mehr als zwölf Stunden auf einer SGI®-Origin mit R10000-Prozessor benötigt, worauf im letzten Kapitel genauer eingegangen werden soll.

### 16.1.1 Ergebnistabellen

Die folgenden Tabellen IV-21a bis IV-23c beschreiben das Paar-Docking der 16 obligatorischen Proteinkomplexe unter Verwendung der drei Grenzwertensembles und jeweils mit den Optionen der Verwendung aller, nur ungesättigter und nur ungebundener Atome. Die Tabellen IV-24a bis IV-26c entsprechen diesem Schema, enthalten jedoch die Angaben für das Triplet-Docking.

Zu jedem Komplex wird nach seinem Namen die Anzahl erzielbarer Referenzergebnisse angegeben. Danach folgen identische Angaben für die drei Bewertungsabschnitte. Aufgeführt wird jeweils die Anzahl der Kandidaten in der richtigen Rotationsgruppe, die Anzahl der darin enthaltenen richtig positiven Ergebnisse und der Rang, den dieser Rotationsrepräsentant mit seinen Kandidaten in einer absteigenden Sortierung aller 564 Rotationsgruppen einnimmt. Für diese Sortierungsposition wird die Anzahl der Rotationen mit gleicher oder höherer Belegungszahl ausgegeben. Ebenfalls in der Tabelle wird die Abweichung der gemittelten Translationsbeschreibung aller Kandidaten des größten richtig rotierten Transformationsclusters von der bekannten Ursprungsverschiebung aufgeführt.

Legende für die Tabellen IV-21a bis IV-26c:

Jede Zeile enthält nach dem Namen des einzelnen Komplexes folgende Spalteneinträge, die sich immer auf die Ergebnisgruppe des Rotationsrepräsentanten beziehen, der die richtige Orientierung beschreibt:

<i>#Referenzen</i>	Anzahl der erzielbaren Referenzergebnisse des Test-Datensatzes
<i># hits</i>	Anzahl der Kandidaten zum Betrachtungszeitpunkt
<i># ref</i>	Anzahl der Referenzen zum Betrachtungszeitpunkt
<i>rank</i>	Positionierung des Clusters in einer absteigenden Sortierung
<i>trans</i>	Abweichung der gemittelten Translationsbeschreibung aller Kandidaten im finalen Transformationsclusters von der bekannten Ursprungstransformation.

Es sind drei Betrachtungszeitpunkte aufgeführt: nach der Rotationsbestimmung (*Rotationscluster*), nach der Ergebnisausdünnung (*Ergebnisausdünnung*) und nach der finalen Clusterbildung (*Transformationscluster*).

Die jeweils letzte Zeile einer Tabelle enthält die Summen der Kandidaten- und Referenzahlen. Grauschattierungen in einzelnen Zeilenabschnitten markieren Ergebnisse bzw. Teilabschnitte, bei denen weniger als zwei Referenzergebnisse in der Ergebnismenge enthalten sind.



	# Referenzen	Rotationscluster			Ergebnisausdünnung			Transformationscluster			trans [Å]
		# hits	# ref	rank	# hits	# ref	rank	# hits	# ref	rank	
1ab9_Bc	29	206	14	4	34	14	1	15	14	1	1,25
1aks_Ab	21	217	8	77	27	8	2	10	8	1	1,62
1ept_Ba	16	78	5	32	10	5	9	5	5	10	1,31
1ept_Ca	6	55	2	247	0	0	564	0	0	564	20,89
1ept_Bc	15	144	6	60	23	6	1	8	6	1	1,41
1huc_Ba	6	85	2	333	5	2	66	5	2	9	1,05
1kvd_Ba	30	112	19	2	26	19	1	22	19	1	1,12
1lya_Ba	50	278	29	53	45	29	1	32	29	1	1,23
1mhl_Ca	20	575	11	67	36	10	61	14	10	1	1,28
1pnk_Ba	27	762	13	372	49	13	11	20	13	1	1,03
1spb_Sp	6	151	2	294	6	2	159	3	2	348	1,01
1wht_Ab	18	386	10	3	48	10	1	11	10	1	1,27
2hnt_Fc	4	67	1	115	0	0	564	0	0	564	20,89
2kin_Ab	36	327	26	44	39	26	4	34	26	1	1,48
2ltn_Ab	102	426	50	1	197	50	1	65	50	1	1,98
2pcd_Nb	13	465	2	297	27	2	165	3	0	539	31,78
	399	4.334	200		572	196		247	194		

Tabelle IV-21a: Paar-Docking mit dem engen Grenzwertensensemble und allen Atomen.

	# Referenzen	Rotationscluster			Ergebnisausdünnung			Transformationscluster			trans [Å]
		# hits	# ref	rank	# hits	# ref	rank	# hits	# ref	rank	
1ab9_Bc	29	121	14	1	24	14	1	15	14	1	1,25
1aks_Ab	21	156	8	28	18	8	2	10	8	1	1,62
1ept_Ba	16	71	5	3	10	5	5	5	5	7	1,31
1ept_Ca	6	32	2	409	0	0	564	0	0	564	20,89
1ept_Bc	15	100	5	59	19	5	1	7	5	1	1,51
1huc_Ba	6	46	2	461	3	2	111	3	2	111	1,43
1kvd_Ba	30	98	19	1	26	19	1	22	19	1	1,12
1lya_Ba	50	192	28	17	38	28	1	31	28	1	1,23
1mhl_Ca	20	393	10	60	16	9	101	13	9	1	1,28
1pnk_Ba	27	489	13	234	30	13	1	14	13	1	1,53
1spb_Sp	6	82	2	316	6	2	20	3	2	162	1,01
1wht_Ab	18	230	10	16	23	10	1	11	10	1	1,27
2hnt_Fc	4	45	1	249	0	0	564	0	0	564	20,89
2kin_Ab	36	192	26	59	36	26	1	31	26	1	1,44
2ltn_Ab	102	365	50	1	182	50	1	55	50	1	1,51
2pcd_Nb	13	298	2	238	6	2	345	3	0	468	32,59
	399	2.910	197		437	193		223	191		

Tabelle IV-21b: Paar-Docking mit dem engen Grenzwertensensemble und ungesättigten Atomen.

	# Referenzen	Rotationscluster			Ergebnisausdünnung			Transformationscluster			trans [Å]
		# hits	# ref	rank	# hits	# ref	rank	# hits	# ref	rank	
1ab9_Bc	29	99	13	1	23	13	1	14	13	1	1,22
1aks_Ab	21	107	8	54	18	8	1	10	8	1	1,62
1ept_Ba	16	63	5	2	10	5	3	5	5	7	1,31
1ept_Ca	6	23	2	425	0	0	564	0	0	564	20,89
1ept_Bc	15	81	4	13	18	4	1	6	4	4	1,57
1huc_Ba	6	32	2	491	3	2	75	3	2	75	1,43
1kvd_Ba	30	83	19	1	26	19	1	22	19	1	1,12
1lya_Ba	50	160	28	13	38	28	1	31	28	1	1,23
1mhl_Ca	20	238	7	236	8	6	113	8	6	1	1,71
1pnk_Ba	27	306	13	340	21	13	1	14	13	1	1,53
1spb_Sp	6	58	2	306	3	0	112	3	0	112	21,34
1wht_Ab	18	143	9	128	18	9	1	10	9	1	1,23
2hnt_Fc	4	41	1	157	0	0	564	0	0	564	20,89
2kin_Ab	36	140	26	19	36	26	1	31	26	1	1,44
2ltn_Ab	102	339	50	1	181	50	1	55	50	1	1,51
2pcd_Nb	13	206	2	148	3	2	331	3	2	331	2,06
	399	2.119	191		406	185		215	185		

Tabelle IV-21c: Paar-Docking mit dem engen Grenzwertensensemble und ungebundenen Atomen.

	# Referenzen	Rotationscluster			Ergebnisausdünnung			Transformationscluster			trans [Å]
		# hits	# ref	rank	# hits	# ref	rank	# hits	# ref	rank	
1ab9_Bc	41	294	24	11	56	24	2	27	24	1	1,52
1aks_Ab	30	382	16	2	82	16	1	22	16	1	1,67
1ept_Ba	18	121	4	6	16	4	8	5	0	59	6,53
1ept_Ca	11	76	7	415	7	7	113	7	7	4	1,12
1ept_Bc	21	233	8	14	24	8	66	11	8	2	1,20
1huc_Ba	8	139	1	228	6	1	317	3	0	457	28,09
1kvd_Ba	41	140	26	2	41	26	1	30	26	1	1,24
1lya_Ba	59	446	37	6	97	37	1	45	37	1	1,15
1mhl_Ca	33	880	16	171	124	15	44	25	15	1	1,42
1pnl_Ba	33	1.162	13	388	104	13	183	23	13	1	1,03
1spb_Sp	8	244	2	167	17	2	147	4	0	303	32,75
1wht_Ab	22	579	11	50	79	11	38	15	11	4	1,40
2hnt_Fc	8	106	3	130	8	3	53	5	3	20	1,12
2kin_Ab	40	487	30	132	84	30	5	40	30	1	1,46
2ltn_Ab	133	588	69	1	318	69	1	86	69	1	1,48
2pcd_Nb	13	741	4	264	83	4	259	13	4	6	1,98
	519	6.618	271		1.146	270		361	263		

Tabelle IV-22a: Paar-Docking mit dem mittleren Grenzwertensemble und allen Atomen.

	# Referenzen	Rotationscluster			Ergebnisausdünnung			Transformationscluster			trans [Å]
		# hits	# ref	rank	# hits	# ref	rank	# hits	# ref	rank	
1ab9_Bc	41	160	22	2	35	22	1	24	22	1	1,41
1aks_Ab	30	248	14	6	48	14	1	18	14	1	1,51
1ept_Ba	18	100	4	2	13	4	8	5	0	38	6,53
1ept_Ca	11	43	7	491	7	7	24	7	7	2	1,12
1ept_Bc	21	155	7	27	20	7	14	10	7	1	1,25
1huc_Ba	8	80	1	378	0	0	564	0	0	564	20,89
1kvd_Ba	41	121	26	1	38	26	1	30	26	1	1,24
1lya_Ba	59	300	37	1	56	37	1	43	37	1	1,16
1mhl_Ca	33	601	13	74	64	12	13	22	12	1	1,52
1pnl_Ba	33	719	13	271	46	13	93	16	13	2	1,58
1spb_Sp	8	135	2	195	16	2	10	4	2	128	0,97
1wht_Ab	22	358	11	37	45	11	3	15	11	1	1,40
2hnt_Fc	8	82	3	131	8	3	22	5	3	8	1,12
2kin_Ab	40	294	30	59	48	30	2	37	30	1	1,45
2ltn_Ab	133	506	69	1	281	69	1	81	69	1	1,47
2pcd_Nb	13	476	4	216	33	4	179	10	4	4	1,74
	519	4.378	263		758	261		327	257		

Tabelle IV-22b: Paar-Docking mit dem mittleren Grenzwertensemble und ungesättigten Atomen.

	# Referenzen	Rotationscluster			Ergebnisausdünnung			Transformationscluster			trans [Å]
		# hits	# ref	rank	# hits	# ref	rank	# hits	# ref	rank	
1ab9_Bc	41	124	21	6	33	21	1	23	21	1	1,40
1aks_Ab	30	174	14	7	37	14	1	17	14	1	1,58
1ept_Ba	18	89	4	2	13	4	3	5	0	29	6,53
1ept_Ca	11	32	6	466	6	6	19	6	6	4	1,09
1ept_Bc	21	119	6	10	18	6	5	8	6	1	1,47
1huc_Ba	8	58	1	389	0	0	564	0	0	564	20,89
1kvd_Ba	41	99	26	1	38	26	1	30	26	1	1,24
1lya_Ba	59	249	37	1	49	37	1	43	37	1	1,16
1mhl_Ca	33	376	11	189	38	10	8	19	10	1	1,77
1pnl_Ba	33	480	13	195	34	13	5	16	13	1	1,58
1spb_Sp	8	99	2	117	15	2	5	3	2	302	1,63
1wht_Ab	22	224	10	120	28	10	1	13	10	1	1,21
2hnt_Fc	8	71	3	71	6	3	44	3	3	208	1,31
2kin_Ab	40	215	30	8	43	30	1	37	30	1	1,45
2ltn_Ab	133	456	69	1	268	69	1	81	69	1	1,47
2pcd_Nb	13	319	3	178	21	3	86	9	3	4	1,77
	519	3.184	256		647	254		313	250		

Tabelle IV-22c: Paar-Docking mit dem mittleren Grenzwertensemble und ungebundenen Atomen.

	# Referenzen	Rotationscluster			Ergebnisausdünnung			Transformationscluster			trans [Å]
		# hits	# ref	rank	# hits	# ref	rank	# hits	# ref	rank	
1ab9_Bc	78	1998	40	177	1270	40	175	781	40	121	14,28
1aks_Ab	72	2758	35	54	2044	34	85	1465	34	156	14,87
1ept_Ba	41	607	13	82	229	13	168	25	0	220	17,35
1ept_Ca	16	604	7	109	248	7	188	35	0	297	18,51
1ept_Bc	45	1619	22	36	994	22	24	218	0	299	16,10
1huc_Ba	34	1029	9	337	446	8	269	33	0	387	26,90
1kvd_Ba	63	797	34	4	426	34	2	71	0	15	17,88
1lya_Ba	155	3553	79	14	2618	79	7	1826	79	15	19,06
1mhl_Ca	153										
1pnk_Ba	118										
1spb_Sp	28	1958	10	150	1079	10	171	239	0	11	34,53
1wht_Ab	58	5299	27		3706	27		2045	27		8,70
2hnt_Fc	10	855	3	58	397	3	29	64	0	90	14,25
2kin_Ab	73	4097	47		2890	47		2106	47		16,96
2ltn_Ab	253	2746	147	1	2044	140	1	1395	140	1	4,39
2pcd_Nb	56										
	1.253	27.920	473		18.391	464		10.303	367		

Tabelle IV-23a: Paar-Docking mit dem weiten Grenzwertenssemble und allen Atomen.

	# Referenzen	Rotationscluster			Ergebnisausdünnung			Transformationscluster			trans [Å]
		# hits	# ref	rank	# hits	# ref	rank	# hits	# ref	rank	
1ab9_Bc	78	996	39	177	423	39	47	88	39	13	4,98
1aks_Ab	72	1761	32	25	1054	32	56	498	32	68	12,04
1ept_Ba	41	460	13	53	152	13	86	18	13	158	1,61
1ept_Ca	16	386	7	66	111	7	117	10	7	362	1,51
1ept_Bc	45	1023	19	20	503	19	8	31	19	313	0,98
1huc_Ba	34	633	3	334	174	3	361	10	3	503	1,51
1kvd_Ba	63	594	33	5	290	33	1	54	33	7	1,54
1lya_Ba	155	2268	75	2	1316	75	1	348	75	1	6,28
1mhl_Ca	153	5277	29		3247	29		1501	29		12,08
1pnk_Ba	118										
1spb_Sp	28	1080	10	112	416	10	83	82	0	9	35,15
1wht_Ab	58	3083	22	70	1616	22	99	247	22	176	9,06
2hnt_Fc	10	659	3	62	252	3	47	53	0	35	13,33
2kin_Ab	73	2383	47	78	1267	47	90	227	47	156	7,25
2ltn_Ab	253	2212	144	1	1575	137	1	1030	137	1	4,59
2pcd_Nb	56	4635	9		3137	9		1914	9		17,95
	1.253	27.450	485		15.533	478		6.111	465		

Tabelle IV-23a: Paar-Docking mit dem weiten Grenzwertenssemble und ungesättigten Atomen.

	# Referenzen	Rotationscluster			Ergebnisausdünnung			Transformationscluster			trans [Å]
		# hits	# ref	rank	# hits	# ref	rank	# hits	# ref	rank	
1ab9_Bc	78	730	35	154	214	35	147	45	35	8	1,82
1aks_Ab	72	1222	31	5	596	31	18	218	31	5	8,63
1ept_Ba	41	398	13	19	137	13	8	18	13	69	1,61
1ept_Ca	16	283	7	49	72	7	32	10	7	133	1,51
1ept_Bc	45	727	15	4	298	15	4	26	0	66	13,29
1huc_Ba	34	471	3	325	116	3	266	9	3	430	1,47
1kvd_Ba	63	441	33	2	178	33	1	47	33	1	1,17
1lya_Ba	155	1786	75	4	882	75	2	121	75	7	0,86
1mhl_Ca	153	3335	22	172	1573	22	115	214	22	79	4,25
1pnk_Ba	118	4023	32	29	1375	31	68	177	31	1	3,54
1spb_Sp	28	712	10	79	247	10	36	66	0	6	35,28
1wht_Ab	58	1895	20	239	766	20	202	51	20	260	1,57
2hnt_Fc	10	526	3	91	153	3	134	10	0	409	14,21
2kin_Ab	73	1634	47	6	713	47	6	108	47	12	1,07
2ltn_Ab	253	1913	144	1	1330	137	1	825	137	1	3,20
2pcd_Nb	56	3070	8	153	1613	8	178	478	8	32	13,47
	1.253	23.166	498		10.263	490		2.423	462		

Tabelle IV-23a: Paar-Docking mit dem weiten Grenzwertenssemble und ungebundenen Atomen.

	# Referenzen	Rotationscluster			Ergebnisausdünnung			Transformationscluster			trans [Å]
		# hits	# ref	rank	# hits	# ref	rank	# hits	# ref	rank	
1ab9_Bc	10	25	6	4	15	6	1	5	5	3	1,83
1aks_Ab	5	20	0	45	6	0	16	2	0	196	8,64
1ept_Ba	6	10	6	16	6	6	3	6	6	1	1,06
1ept_Ca	0	5	0	62	0	0	564	0	0	564	20,89
1ept_Bc	4	17	1	7	9	0	2	3	0	27	8,6
1huc_Ba	1	5	0	131	0	0	564	0	0	564	20,89
1kvd_Ba	12	27	8	1	15	8	1	6	4	1	2,14
1lya_Ba	13	24	7	40	9	7	4	5	5	4	2,82
1mhl_Ca	0	32	0	16	11	0	3	3	0	45	20,62
1pnk_Ba	7	47	5	31	10	4	23	4	4	13	2,37
1spb_Sp	1	10	1	89	2	0	103	2	0	103	47,02
1wht_Ab	5	33	4	3	12	3	2	5	3	2	2,89
2hnt_Fc	0	3	0	143	0	0	564	0	0	564	20,89
2kin_Ab	15	40	12	1	24	12	1	14	12	1	1,98
2ltn_Ab	40	182	31	1	146	31	1	50	31	1	1,88
2pcd_Nb	1	22	1	138	2	0	252	2	0	252	44,14
	120	502	82		267	77		107	70		

Tabelle IV-24a: Triplet-Docking mit dem engen Grenzwertenset und allen Atomen.

	# Referenzen	Rotationscluster			Ergebnisausdünnung			Transformationscluster			trans [Å]
		# hits	# ref	rank	# hits	# ref	rank	# hits	# ref	rank	
1ab9_Bc	10	15	6	3	11	6	1	5	5	2	1,83
1aks_Ab	5	15	0	21	6	0	6	2	0	124	8,64
1ept_Ba	6	9	6	9	6	6	2	6	6	1	1,06
1ept_Ca	0	3	0	121	0	0	564	0	0	564	20,89
1ept_Bc	4	16	1	3	9	0	1	3	0	25	8,6
1huc_Ba	1	2	0	262	0	0	564	0	0	564	20,89
1kvd_Ba	12	26	8	1	15	8	1	6	4	1	2,14
1lya_Ba	13	18	7	16	9	7	1	5	5	3	2,82
1mhl_Ca	0	22	0	8	6	0	7	2	0	150	28,73
1pnk_Ba	7	31	5	10	8	4	5	4	4	5	2,37
1spb_Sp	1	5	1	85	2	0	43	2	0	43	47,02
1wht_Ab	5	15	4	5	8	3	1	5	3	1	2,89
2hnt_Fc	0	2	0	183	0	0	564	0	0	564	20,89
2kin_Ab	15	32	12	1	19	12	1	13	12	1	1,87
2ltn_Ab	40	157	31	1	133	31	1	48	31	1	1,85
2pcd_Nb	1	8	1	251	2	0	117	2	0	117	44,14
	120	376	82		234	77		103	70		

Tabelle IV-24b: Triplet-Docking mit dem engen Grenzwertenset und ungesättigten Atomen.

	# Referenzen	Rotationscluster			Ergebnisausdünnung			Transformationscluster			trans [Å]
		# hits	# ref	rank	# hits	# ref	rank	# hits	# ref	rank	
1ab9_Bc	10	15	6	2	11	6	1	5	5	2	1,83
1aks_Ab	5	13	0	5	4	0	11	2	0	71	8,64
1ept_Ba	6	9	6	5	6	6	2	6	6	1	1,06
1ept_Ca	0	3	0	71	0	0	564	0	0	564	20,89
1ept_Bc	4	16	1	1	9	0	1	3	0	19	8,6
1huc_Ba	1	2	0	189	0	0	564	0	0	564	20,89
1kvd_Ba	12	25	8	1	15	8	1	6	4	1	2,14
1lya_Ba	13	14	7	18	7	7	3	5	5	3	2,82
1mhl_Ca	0	9	0	91	0	0	564	0	0	564	20,89
1pnk_Ba	7	25	5	3	8	4	1	4	4	4	2,37
1spb_Sp	1	3	1	141	2	0	27	2	0	27	47,02
1wht_Ab	5	13	3	2	8	3	1	5	3	1	2,89
2hnt_Fc	0	1	0	301	0	0	564	0	0	564	20,89
2kin_Ab	15	25	12	1	16	12	1	12	12	1	1,72
2ltn_Ab	40	154	31	1	130	31	1	48	31	1	1,85
2pcd_Nb	1	8	1	100	2	0	79	2	0	79	44,14
	120	335	81		218	77		100	70		

Tabelle IV-24c: Triplet-Docking mit dem engen Grenzwertenset und ungebundenen Atomen.

	# Referenzen	Rotationscluster			Ergebnisausdünnung			Transformationscluster			trans [Å]
		# hits	# ref	rank	# hits	# ref	rank	# hits	# ref	rank	
1ab9_Bc	22	40	12	5	23	12	2	9	9	1	1,75
1aks_Ab	13	38	4	39	15	3	17	3	3	131	3,83
1ept_Ba	8	11	5	112	5	5	42	3	3	55	1,68
1ept_Ca	6	12	3	24	5	3	14	3	3	24	1,75
1ept_Bc	9	26	2	22	9	0	34	4	0	26	8,65
1huc_Ba	1	18	0	9	9	0	1	3	0	36	26,62
1kvd_Ba	25	34	15	1	24	15	1	14	10	1	3,31
1lya_Ba	24	59	16	2	29	16	1	19	16	1	1,47
1mhl_Ca	1	66	1	41	27	1	1	5	1	24	2,72
1pnk_Ba	11	92	6	28	32	6	3	5	5	13	2,36
1spb_Sp	2	18	2	132	6	2	67	2	2	276	1,83
1wht_Ab	5	74	5	1	24	5	3	6	5	14	2,17
2hnt_Fc	5	9	2	32	7	2	2	3	2	20	1,73
2kin_Ab	17	57	12	4	31	12	2	14	12	1	1,40
2ltn_Ab	89	381	66	1	342	66	1	118	66	1	2,46
2pcd_Nb	1	44	1	109	10	0	110	2	0	424	44,14
	239	979	152		598	148		213	137		

Tabelle IV-25a: Triplet-Docking mit dem mittleren Grenzwertensemble und allen Atomen.

	# Referenzen	Rotationscluster			Ergebnisausdünnung			Transformationscluster			trans [Å]
		# hits	# ref	rank	# hits	# ref	rank	# hits	# ref	rank	
1ab9_Bc	22	24	12	1	18	12	1	9	9	1	1,75
1aks_Ab	13	28	3	5	14	2	1	3	0	70	8,80
1ept_Ba	8	9	5	90	5	5	24	3	3	42	1,68
1ept_Ca	6	10	3	10	3	3	22	3	3	14	1,75
1ept_Bc	9	21	2	7	9	0	12	4	0	17	8,65
1huc_Ba	1	12	0	11	6	0	3	2	0	109	20,68
1kvd_Ba	25	33	15	1	24	15	1	14	10	1	3,31
1lya_Ba	24	39	16	3	22	16	1	18	16	1	1,63
1mhl_Ca	1	35	1	63	14	1	5	4	0	26	51,42
1pnk_Ba	11	57	6	10	25	6	2	5	5	2	2,36
1spb_Sp	2	9	2	115	4	2	31	2	0	144	47,02
1wht_Ab	5	27	5	17	8	5	14	6	5	1	2,17
2hnt_Fc	5	8	2	21	7	2	2	3	2	18	1,73
2kin_Ab	17	34	12	5	22	12	3	12	12	1	1,74
2ltn_Ab	89	338	66	1	305	66	1	110	66	1	2,45
2pcd_Nb	1	17	1	220	2	0	281	2	0	281	44,14
	239	701	151		488	147		200	131		

Tabelle IV-25b: Triplet-Docking mit dem mittleren Grenzwertensemble und ungesättigten Atomen.

	# Referenzen	Rotationscluster			Ergebnisausdünnung			Transformationscluster			trans [Å]
		# hits	# ref	rank	# hits	# ref	rank	# hits	# ref	rank	
1ab9_Bc	22	23	12	1	18	12	1	9	9	1	1,75
1aks_Ab	13	21	3	3	10	2	1	2	0	176	8,64
1ept_Ba	8	9	5	53	5	5	16	3	3	32	1,68
1ept_Ca	6	8	2	17	2	2	75	2	2	75	1,81
1ept_Bc	9	20	2	5	9	0	10	4	0	15	8,65
1huc_Ba	1	7	0	30	4	0	7	2	0	68	27,47
1kvd_Ba	25	32	15	1	24	15	1	14	10	1	3,31
1lya_Ba	24	34	16	2	21	16	1	17	16	1	1,63
1mhl_Ca	1	16	1	171	5	1	52	3	1	49	3,16
1pnk_Ba	11	41	6	2	20	6	1	5	5	1	2,36
1spb_Sp	2	8	2	60	4	2	18	2	0	99	47,02
1wht_Ab	5	18	5	9	8	5	4	6	5	1	2,17
2hnt_Fc	5	6	2	34	6	2	4	2	2	67	1,68
2kin_Ab	17	27	12	4	18	12	4	12	12	1	1,74
2ltn_Ab	89	323	66	1	297	66	1	109	66	1	2,41
2pcd_Nb	1	13	1	136	2	0	189	2	0	189	44,14
	239	606	150		453	146		194	131		

Tabelle IV-25c: Triplet-Docking mit dem mittleren Grenzwertensemble und ungebundenen Atomen.

	# Referenzen	Rotationscluster			Ergebnisausdünnung			Transformationscluster			trans [Å]
		# hits	# ref	rank	# hits	# ref	rank	# hits	# ref	rank	
1ab9_Bc	65	1989	31	102	1841	31	99	948	31	135	16,08
1aks_Ab	56	2801	24	100	2663	24	99	1716	24	147	16,71
1ept_Ba	35	484	14	73	407	14	65	39	0	172	15,74
1ept_Ca	14	406	8	116	311	8	131	17	0	280	17,06
1ept_Bc	27	1317	14	115	1180	14	112	136	0	227	22,93
1huc_Ba	6	671	1	174	511	1	181	28	0	256	13,27
1kvd_Ba	53	806	34	14	717	34	14	156	0	26	14,32
1lya_Ba	109	3134	65	82	2927	65	82	2149	65	56	25,03
1mhl_Ca	81										
1pnk_Ba	57										
1spb_Sp	13	1676	7	142	1467	7	154	495	0	39	31,79
1wht_Ab	30										
2hnt_Fc	5	562	2	95	452	2	100	31	0	187	24,05
2kin_Ab	50										
2ltn_Ab	315	3620	223	1	3457	223	1	2262	223	1	4,79
2pcd_Nb	7										
	923	17.466	423		15.933	423		7.977	343		

Tabelle IV-26a: Triplet-Docking mit dem weiten Grenzwertenset und allen Atomen.

	# Referenzen	Rotationscluster			Ergebnisausdünnung			Transformationscluster			trans [Å]
		# hits	# ref	rank	# hits	# ref	rank	# hits	# ref	rank	
1ab9_Bc	65	585	30	113	446	30	111	77	30	18	4,22
1aks_Ab	56	1384	17	84	1227	17	80	262	0	145	21,50
1ept_Ba	35	315	14	63	245	14	55	21	0	167	29,81
1ept_Ca	14	220	7	51	147	7	65	11	0	165	25,84
1ept_Bc	27	683	12	73	544	12	83	59	12	95	4,66
1huc_Ba	6	341	1	172	232	1	180	17	0	246	36,67
1kvd_Ba	53	583	32	6	507	32	6	121	32	1	3,66
1lya_Ba	109	1470	60	87	1232	60	82	130	0	80	31,30
1mhl_Ca	81										
1pnk_Ba	57										
1spb_Sp	13	819	7	56	677	7	45	326	0	2	31,43
1wht_Ab	30	2059	19	78	1733	19	79	148	19	143	2,12
2hnt_Fc	5	387	2	86	295	2	80	20	0	184	17,97
2kin_Ab	50	1584	34	81	1356	34	78	215	34	83	1,99
2ltn_Ab	315	2763	220	1	2611	220	1	1583	220	1	4,06
2pcd_Nb	7										
	923	13.193	455		11.252	455		2.990	347		

Tabelle IV-26a: Triplet-Docking mit dem weiten Grenzwertenset und ungesättigten Atomen.

	# Referenzen	Rotationscluster			Ergebnisausdünnung			Transformationscluster			trans [Å]
		# hits	# ref	rank	# hits	# ref	rank	# hits	# ref	rank	
1ab9_Bc	65	382	29	106	263	29	108	53	29	5	3,21
1aks_Ab	56	768	17	76	635	17	73	82	17	65	3,58
1ept_Ba	35	231	14	57	171	14	49	20	14	92	1,68
1ept_Ca	14	127	4	87	62	4	149	5	4	281	1,18
1ept_Bc	27	407	11	70	291	11	77	34	11	31	4,23
1huc_Ba	6	241	1	136	150	1	154	12	0	215	26,30
1kvd_Ba	53	402	32	4	342	32	2	106	32	1	3,52
1lya_Ba	109	1014	60	88	804	60	89	91	60	14	1,11
1mhl_Ca	81	1849	17	132	1443	17	131	128	17	53	3,09
1pnk_Ba	57	1943	28	97	1408	28	85	147	28	3	1,63
1spb_Sp	13	531	7	31	434	7	27	199	0	7	30,98
1wht_Ab	30	1046	16	100	785	16	109	110	16	15	2,12
2hnt_Fc	5	293	2	69	208	2	65	18	0	111	32,25
2kin_Ab	50	889	34	60	701	34	54	111	34	23	2,88
2ltn_Ab	315	2422	219	1	2291	219	1	1472	219	1	3,81
2pcd_Nb	7	1436	3	150	1109	3	139	39	3	174	3,60
	923	13.981	494		11.097	494		2.627	484		

Tabelle IV-26a: Triplet-Docking mit dem weiten Grenzwertenset und ungebundenen Atomen.

### 16.1.2 Spezifität richtiger Ergebnisse

Die Tabellen IV-27 und IV-28 enthalten für das Paar- und Triplet-Docking abgeleitete Informationen aus den ausführlicheren Ergebnistabellen der einzelnen Dockingexperimente. Diese beziehen sich auf die Summen der Kandidaten und der Referenzergebnisse der gesamten Testgruppe für die einzelnen Bewertungsschritte. Einerseits wird angegeben, um welchen prozentualen Anteil die Menge aller Kandidaten von der Rotationsbestimmung zur Ergebnisausdünnung und von dort zur finalen Clusterbildung abnimmt. Andererseits wird der jeweilige prozentuale Anteil richtig positiver Ergebnisse an der Menge aller Kandidaten aufgeführt, solange noch zumindest ein Referenzergebnis in dieser Gruppe enthalten ist.

#### 16.1.2.1 Spezifität im Paar-Docking obligatorischer Proteinkomplexe

	Abnahme der Kandidaten [%]		Anteil richtig positiver Ergebnisse an allen Kandidaten [%]		
	Rot.clust. → Ausdünnung	Ausdünnung → Trans.cluster	Rot.clust.	Ausd.	Trans.clust.
<b>enge Grenzwerte</b>					
alle Atome	-86,8	-56,8	4,6	34,3	79,5
ungesättigte A.	-85,0	-49,0	6,8	44,2	86,8
ungebundene A.	-80,8	-47,0	9,0	45,6	87,3
<b>mittlere Grenzwerte</b>					
alle Atome	-82,7	-68,5	4,1	23,6	75,4
ungesättigte A.	-82,7	-56,9	6,0	34,4	79,8
ungebundene A.	-79,7	-51,6	8,0	39,3	80,2
<b>weite Grenzwerte</b>					
alle Atome	-41,6	-59,9	1,7	2,5	3,8
ungesättigte A.	-47,8	-70,5	1,8	3,1	7,8
ungebundene A.	-55,7	-76,4	2,2	4,8	19,9

Tabelle IV-27: Spezifitätsentwicklung im Paar-Docking. Prozentuale Abnahme der Kandidatenzahlen von der Rotationsbestimmung (Rot.clust.) zur Ergebnisausdünnung (Ausd.) und weiter zur Bildung der Translationscluster (Trans.clust.). Anteil richtig positiver Ergebnisse an der Menge aller Kandidaten. Die zugrunde liegenden Summenwerte der Kandidaten- und Referenzzahlen finden sich in den Tabellen IV-21a bis IV-23c.

#### **Spezifität: Paar-Docking mit engem und mittlerem Grenzwertensemble:**

Zunächst sollen einmalig anhand des Paar-Dockings mit allen Atomen unter Verwendung des engen Grenzwertensembles die Bedeutung verwendeter Angaben verdeutlicht werden:

- Im richtigen Rotationscluster befinden sich für die gesamte Testgruppe nach der Evaluierung bimolekularer transformations-invarianter Eigenschaften 4334 Kandidaten

inklusive 200 Referenzergebnisse. Dieses entspricht einem Anteil richtig positiver Ergebnisse von lediglich 4,6%.

- Anhand der translationsabhängigen Nachbarsuche werden diese Ergebnisse ausgedünnt auf 572 Kandidaten, was einer Abnahme um 86,8% entspricht. Hier sind immer noch 196 Referenzen enthalten, die somit einen gesteigerten Anteil von 34,3% ausmachen.
- Die Bildung der finalen Translationscluster verringert die Kandidatenzahlen um 56,8% auf insgesamt 247 Ergebnisse. Für einen Komplex enthält der richtige Rotationscluster zwar drei Kandidaten, jedoch kein Referenzergebnis. Von den verbleibenden 244 Kandidaten der Testgruppe sind 194 richtig positive Ergebnisse. Dieses entspricht einem endgültigen Referenzanteil von 79,5%.

Die Spezifität der Referenzergebnisse in diesem Beispiel kann also kontinuierlich gesteigert werden, indem in der Nachbearbeitung kriterienkonformer Kandidaten deutlich mehr falsch positive als richtig positive Ergebnisse eliminiert werden. Die Grundlage der Eliminierung ist die hier bestätigte Annahme, dass unspezifische, falsch positive Ergebnisse nicht mit gleichartigen Translationsbeschreibungen in ausreichend hoher Dichte auftreten.

Die Verwendung ungesättigter oder ungebundener Atome senkt primär die Anzahl der Atome, die für die Molekülrepräsentation verwendet werden. Daher ist die Anzahl aller Kandidaten nach der Evaluierung und Rotationsbestimmung bereits um bis zu 50% geringer als bei der Verwendung aller Atome (siehe Tabellen IV-21a bis IV-23c). Da bei der Atomvorselektion bereits der relative Anteil essenziell richtiger Referenzatome gesteigert wird, beträgt der Referenzanteil zu diesem Zeitpunkt auch bis zu 9,0%. Da die Ausdünnungseffekte etwas geringere Wirkung zeigen können, weil bereits mit einer spezifischeren Atomauswahl begonnen wird, beträgt der maximale Anteil der Referenzergebnisse in den finalen Transformationsclustern unter Verwendung nur ungebundener Atome 87,3% mit 185 von 212 Ergebnissen.

Die Verwendung des engen Grenzwertensembles hat auch zur Folge, dass drei der 16 obligatorischen Komplexe nach der Rotationsbestimmung nur ein bzw. zwei richtig positive Ergebnisse mit der richtigen Rotation bestimmt haben. Dieses betrifft zwei Komplexe mit den kleinsten Bindungspartnern, die überhaupt von Anfang an nur vier bzw. sechs Referenzpaare aufweisen können, aber auch einen Komplex mit anfangs 13 erzielbaren Referenzen. Die für Atompaaire besonders schlechte Orientierungsbestimmung verteilt die übrigen Referenzen voneinander unabhängig auf andere Rotationsrepräsentanten und lässt sie damit verloren gehen. Da im Paar-Docking in der Ergebnisausdünnung zumindest Dreiergruppen translationsähnlicher



Ergebnisse erwartet werden, bleiben bei den beiden kleinen Komplexen im richtigen Rotationscluster zuletzt überhaupt keine Kandidaten erhalten. Auch kein falsch positives Ergebnis liegt nahe genug an den richtig positiven Ergebnissen, um mit ihnen einen Translationscluster zu bilden. Der größere Komplex kann zunächst 27 Ergebnisse inklusive den beiden Referenzen erhalten. Er verliert diese jedoch bei der Bildung eines finalen Translationsclusters, der aus drei falsch positiven Ergebnisse zusammengesetzt wird.

Zuletzt stellen sich also drei obligatorische Komplexe, die anfangs vier, sechs bzw. 13 Referenzergebnisse vorweisen können, für die Verwendung der engen Selektionsgrenzwerte im Paar-Docking als ungeeignet heraus. Sie können keine finalen Transformationscluster mit richtig positiven Ergebnissen erzeugen.

Bei der Verwendung nur ungebundener Atome ändert sich das Bild ein wenig. Dieselben zwei kleinen Komplexe erbringen natürlich auch hier keinen richtig rotierten Transformationscluster, der dritte Test-Datensatz enthält seine zwei richtig positiven Ergebnisse in einer finalen Dreiergruppe und ein weiterer ebenfalls recht kleiner Komplex produziert ein Endergebnis ohne Referenzen.

Im Paar-Docking mit den mittleren Grenzwerten zeigen sich sehr vergleichbare Werte in den Entwicklungen. Die Ausgangszahlen liegen jedoch für die Kandidaten etwa 50% und für die Referenzen etwa 35% höher als beim Docking mit den engen Selektionsgrenzen, was bereits einen geringere Spezifität direkt zu Beginn beschreibt.

Die Abnahme der Kandidatenzahlen und damit gekoppelt die Zunahme des Referenzanteils in der Ergebnisausdünnung und der finalen Clusterbildung fallen ein wenig geringer aus, da die größeren Atomabstände ein etwas dichteres und homogeneres Netz von Translationsnachbarn erzeugen. Zuvor isolierte Ergebnisse können so an größere Gruppen angebunden werden. Die erreichbaren Referenzanteile in den finalen Ergebnissen liegen zwischen 75,4% und 80,2%.

Bei der Verwendung der mittleren Grenzwerte passiert es je einmal bei den optionalen Atomdisqualifikationen, dass kein finales Ergebnis mit richtiger Rotation erzeugt wird. Drei Komplexe im Docking mit allen Atomen und ansonsten je ein Komplex erzeugen richtig rotierte Endergebnisse mit Translationsbeschreibungen ohne Mitwirkung von Referenzergebnissen.

Zusammenfassend für das Paar-Docking mit engen oder mittleren Grenzwertensembles kann man feststellen, dass die Kandidatenzahlen durch die Nachbearbeitung kriterienkonformer Evaluierungsergebnisse um 80% und nochmals 50% gesenkt werden. Hierdurch steigt der Anteil

richtig positiver Ergebnisse an den finalen Ergebnisgruppen je nach optionaler Atomvorauswahl von 4% bis 9% auf 75% bis fast 90%.

### **Spezifität: Paar-Docking mit weitem Grenzwertensemble:**

Die Verwendung der weiten Selektionsgrenzen hat einen enormen Einfluss auf die Ergebniszahlen. Da die Testgruppe nur beim Docking mit völlig ungebundenen Atomen vollständig abgearbeitet werden kann, lassen sich nur hier richtige Vergleichswerte zu den anderen Grenzwertensembles erstellen. Insgesamt werden 23166 Kandidaten mit der richtigen Rotation beschrieben. Bei den vorhergehenden analogen Dockingexperimenten waren dieses 2119 bzw. 3184 Kandidaten. Da die Referenzsumme in diesem Vergleich nur von 191 bzw. 256 auf 498 zunimmt, bestätigt sich die geringere Spezifität mit einem Anteil richtig positiver Ergebnisse von anfänglich nur 2,2%.

Da durch die große Menge von Atompaaren das Netzwerk der Translationsbeschreibungen falsch positiver Ergebnisse gegenüber der Verwendung mittlerer Grenzwerte nochmals deutlich gleichmäßiger gestreut wird, sinkt die Quote der Kandidatenabnahme auf 55,7% in der Ergebnisausdünnung. Die Verlustquote in der Bildung finaler Transformationscluster ist mit 76,4% dagegen höher als oben beschrieben, weil hier alle Ergebnisse, die nicht dem einen größten Cluster angehören, pauschal wegfallen.

Dennoch ist am Ende im allergünstigsten Fall nur ein Referenzanteil von 19,1% zu verzeichnen entgegen den zuvor erreichten 87,3% und 80,2%.

Für die Dockingansätze mit allen bzw. mit ungesättigten Atomen fallen diese finalen Referenzanteile mit 3,8% und 7,8% sogar noch geringer aus. Diese Ergebnisse sind jedoch nicht vollwertig vergleichbar, da nicht alle Einzelexperimente durchgeführt werden können. Da diese Moleküle allerdings nochmals bedeutend höhere Zahlen von Atompaaren und Evaluierungsaufrufen vorweisen können, kann die Spezifität nur weiter abnehmen bei der dazu relativ geringeren Zunahme von Referenzergebnissen, die sie besteuern können.

Dieses ist auch daran ersichtlich, dass beim Docking mit allen Atomen sieben der elf Komplexe Endergebnisse ohne Referenzbeteiligung hervorbringen. Bei den übrigen Endergebnissen beträgt der Referenzanteil im schlechtesten Fall 34 von 1465 Kandidaten und im besten Fall auch nur 140 von 1395 Kandidaten.

Im Docking mit nur ungebundenen Atomen besitzen zwar alle Experimente ein richtig rotiertes Endergebnis, aber drei Mal ist daran keinerlei Referenz beteiligt. In sechs Dockingexperimenten liegt der Referenzanteil mit etwa 40% bis 80% immer noch tiefer als beim Paar-Docking mit den stärker selektiven Grenzwerten.

## 16.1.2.2 Spezifität im Triplett-Docking obligatorischer Proteinkomplexe

Von den tendenziellen Entwicklungen her lassen sich alle Beobachtungen, die für das Paar-Docking obligatorischer Proteinkomplexe erläutert wurden auf das Triplett-Docking übertragen.

	Abnahme der Kandidaten [%]		Anteil richtig positiver Ergebnisse an allen Kandidaten [%]		
	Rot.clust. → Ausdünnung	Ausdünnung → Trans.cluster	Rot.clust.	Ausd.	Trans.clust.
<b>enge Grenzwerte</b>					
alle Atome	-46,8	-59,9	18,8	32,5	73,7
ungesättigte A.	-37,8	-56,0	24,7	36,8	76,1
ungebundene A.	-34,9	-54,1	26,4	38,3	76,9
<b>mittlere Grenzwerte</b>					
alle Atome	-38,9	-64,4	15,8	26,0	67,2
ungesättigte A.	-30,4	-59,0	21,9	31,2	71,6
ungebundene A.	-25,2	-57,2	25,0	33,3	72,0
<b>weite Grenzwerte</b>					
alle Atome	-8,8	-49,9	2,4	2,7	4,9
ungesättigte A.	-14,7	-73,4	3,4	4,0	15,8
ungebundene A.	-20,6	-76,3	3,5	4,5	20,0

Tabelle IV-28: Spezifitätsentwicklung im Triplett-Docking. Prozentuale Abnahme der Kandidatenzahlen von der Rotationsbestimmung (Rot.clust.) zur Ergebnisausdünnung (Ausd.) und weiter zur Bildung der Translationscluster (Trans.clust.). Anteil richtig positiver Ergebnisse an der Menge aller Kandidaten. Die zugrunde liegenden Summenwerte der Kandidaten- und Referenzzahlen finden sich in den Tabellen IV-24a bis IV-26c.

**Spezifität: Triplett-Docking mit engem und mittlerem Grenzwertensemble:**

Es ist zunächst auffällig, dass bei der Verwendung der beiden selektionsstärkeren Grenzwertensembles die Wirkung der Ergebnisausdünnung mit Abnahmen der Kandidatenzahlen um etwa 35% insgesamt schwächer ausfallen als im Paar-Docking mit Reduktionen um etwa 80%. Die Triplett-Dockings produzieren allerdings von Anfang an nur etwa ein Achtel der Ergebniszahlen und aufgrund der etwas erfolgreicheren Rotationsbestimmung beträgt der Referenzanteil für den richtigen Rotationsrepräsentanten anfangs bereits 15,8% bis 26,4%. Daher fällt ein Großteil der Kandidaten weg, die im Paar-Docking aufgrund völlig fehlerhafter Rotationsvorschriften auch abnorme Translationen zugeschrieben bekommen, daher in der Ausdünnung keine Nachbarn finden und eliminiert werden können. Außerdem wird im Triplett-

Docking wegen den geringeren Referenzzahlen nur auf zwei und nicht drei gleichartige Translationsbeschreibungen vorselektiert.

Der endgültige Referenzanteil fällt mit 67,2% bis 76,9% etwa 10% niedriger aus als beim Paar-Docking. Die Einzelergebnisse verraten jedoch, dass zum Teil die Hälfte aller Komplexe finale Transformationscluster vorweisen kann, die ausschließlich aus richtig positiven Kandidaten zusammengesetzt sind. Ein großer Anteil zollt andererseits dem simultanen Selektionsdruck auf alle Dreiecksseiten Tribut und kann überhaupt keine Referenzen vorweisen oder keine bzw. zu wenige Referenzen richtig orientieren. Besitzen diese Komplexe dennoch ausschließlich falsch positiv besetzte finale Transformationscluster, so senken diese den relativen Referenzanteil.

### **Spezifität: Triplet-Docking mit weitem Grenzwertensemble:**

Die geringere Ergebnisspezifizierung durch Ergebnisausdünnung und Clusterbildung treffen auch im Triplet-Docking unter Verwendung des weiten Grenzwertensembles zu. Die finalen Referenzquoten sind allerdings auch hier nur für das Docking mit nur ungebundenen Atomen vollständig und vergleichbar mit vorherigen Ergebnissen. Nur drei Komplexe weisen Endergebnisse ohne Referenzen aus, da sie anfänglich zu wenige richtig positive Ergebnisse in das Docking einbringen. Bei den übrigen Test-Datensätzen schwanken die Anteile zwischen etwa 14% und 70%.

Unter Verwendung von mehr Atomen, dem hiermit verbundenen Anstieg des Kombinatorikproblems und der abnehmenden Wirkung der Nachselektion ist zu beobachten, dass Referenzen nur noch in den größten gerechneten Dockingexperimenten im Endergebnis enthalten sind. Hier ist ihr Anteil an den finalen Transformationsclustern, die 948 bis 2262 Kandidaten umfassen, jedoch mit einem einzigen Spitzenwert von 10% und einem Durchschnitt von 4,9% viel zu gering.

#### *16.1.3 Qualifizierung richtig rotierter Transformationscluster*

Die Tabellen IV-29 und IV-30 enthalten für das Paar- und das Triplet-Docking die Übersicht, wie viele Komplexe der Testgruppe obligatorischer Proteinkomplexe zu den drei Betrachtungszeitpunkten zumindest zwei Referenzergebnisse vorweisen können und wie viele dieser Komplexe eine Sortierungsposition innerhalb der ersten 100 aller 564 Rotationscluster einnehmen.

## 16.1.3.1 Sortierungsposition im Paar-Docking obligatorischer Proteinkomplexe

		# Komplexe mit ≥2 Referenzergebnissen			hiervon: # Komplexe mit Sortierungsposition ≤100		
		Rot.- cluster	Ausd.	Trans.- cluster	Rot.- cluster	Ausd.	Trans.- cluster
<b>enge Grenzwerte</b>							
alle Atome	(von 16)	15	14	13	10	12	12
ungesättigte A.	(von 16)	15	14	13	10	11	11
ungebundene A.	(von 16)	15	13	13	8	11	12
<b>mittlere Grenzwerte</b>							
alle Atome	(von 16)	15	15	13	8	11	13
ungesättigte A.	(von 16)	15	15	14	10	14	13
ungebundene A.	(von 16)	15	15	14	9	15	12
<b>weite Grenzwerte</b>							
alle Atome	(von 11)	11	11	4	7	6	2
ungesättigte A.	(von 13)	13	13	11	10	11	5
ungebundene A.	(von 16)	16	16	13	11	10	10

Tabelle IV-29: Obligatorische Komplexe mit mindestens zwei Referenzpaaren und mit Sortierungspositionen von Rang 100 und besser.  
Absolute Anzahl der Komplexe nach der Rotationsbestimmung (Rot.cluster), nach der Ergebnisausdünnung (Ausd.) im finalen Transformationscluster (Trans.cluster).

**Qualifizierung: Paar-Docking mit engem und mittlerem Grenzwertensemble:**

Bei Verwendung der engen und mittleren Grenzwertensembles im Paar-Docking besitzt ein Komplex durchgehend nur ein Referenzergebnis mit richtiger Rotationsbestimmung. Dieses geht zudem stets in der Ergebnisausdünnung verloren, weshalb kein Endergebnis mit richtig positiven Kandidaten erbracht werden kann.

Generell sind zwei richtig orientierte Referenzen an sich zu wenig, da sie in diesen Fällen stets die Hilfe eines falsch positiven Ergebnisses mit korrekter Transformationsvorschrift benötigen. Es bestätigt sich nochmals die Gefahr, beim Besitz weniger richtig positiver Ergebnisse essentielle Kandidaten durch fehlerhafte Atomdisqualifikation zu verlieren. Allerdings nicht nur durch den Ausschluss eines richtig positiven Atoms, was bei den selektionsstarken Grenzwerten vergleichsweise seltener vorkommt als beim weiten Grenzwertensemble, sondern nun auch durch den Verlust wichtiger Translationsnachbarn. Zuletzt besitzen drei bzw. zwei Komplexe keine Referenzen mehr im Endergebnis.

Recht eindeutige Resultate können beschreiben, wie viele Kandidaten im finalen Transformationscluster enthalten sein müssen, damit eine Einsortierung unter den ersten 100

Positionen erreicht wird. Bei Verwendung des engen Grenzwertensembles ist mit mehr als drei Kandidaten und beim mittleren Ensemble mit mehr als vier Kandidaten immer ein derartiger Rang zu erlangen. In 71 von 76 Fällen ist dieser Rang innerhalb der ersten zehn Positionen im Regelfall sogar Platz Eins.

Die Anzahl guter Platzierungen ist zu den beiden früheren Betrachtungszeitpunkten zumeist geringer, da die Ergebnisspezifität noch nicht so gut ausgeprägt ist. Daher gibt es falsch positive Ergebnisgruppen, die noch größer sind.

#### **Qualifizierung: Paar-Docking mit weitem Grenzwertensemble:**

Ebenso die geringere Ergebnisspezifität kommt bei der Verwendung der weiten Selektionsgrenzen zum Tragen. Nach der Rotationsbestimmung und der Ergebnisausdünnung gibt es keinen einzigen Komplex, der nicht richtig orientierte Referenzergebnisse vorweisen könnte, die Sortierungspositionen sind jedoch zu unzuverlässig und offensichtlich nur vom Zufall geprägt. Bei der Bildung finaler Cluster qualifizieren sich beim Docking mit allen Atomen sieben von elf Komplexen ohne ein Referzergebnis. Die hierbei erzielten Sortierungspositionen schwanken unspezifisch zwischen Rang elf und Rang 387. Auch bei den Komplexen, die bis zuletzt richtig positive Kandidaten mit sich führen, machen diese keinen ausschlaggebenden Anteil an allen enthaltenen Kandidaten aus. Zum Teil sind bei 1465 Kandidaten nur 34 Referenzpaare enthalten. Dieses Verhältnis ist unsignifikant und die Belegung dieses Endergebnisses ist nicht von der Referenzbelegung beeinflusst, weshalb die Sortierungsposition ebenfalls als referenzunabhängig bewertet werden muss.

Ein wenig bessert sich diese Situation bei der Vorauswahl der verwendeten Atome. Beim Docking mit nur ungebundenen Atomen ist die Spezifität am Ende hoch genug, dass zumindest 13 von 16 Komplexen verbleibende Referenzpaare im richtig rotierten Endergebnis enthalten. In zehn Fällen können auch Positionen unter den ersten 100 Plätzen erzielt werden. Allerdings gibt es auch zwei Experimente, bei denen sieben Referenzen in einer finalen Zehnergruppe und drei Referenzen in einer finalen Neunergruppe unzureichend sind für eine gute Positionierung, obwohl diese Referenzbelegungen deutlich überproportionale Referenzanteile gegenüber dem Testgruppendurchschnitt von 19,1% darstellen. Die Clusterbelegung ist zuletzt einfach viel zu gering um gegen zufällige Cluster aus den Massen falsch positiver Ergebnisse herauszuragen. Ebenso gibt es zwei Experimente, bei denen finale Ergebniscluster völlig ohne Referenzbeteiligung sehr gut einsortiert werden. Zum Teil sind 51 Kandidaten inklusive 20 Referenzpaaren nicht ausreichend, um eine bessere Positionierung als Rang 260 zu erzielen.

Insgesamt muss die Schlussfolgerung gezogen werden, dass man aus den Belegungszahlen bei der Verwendung der weiten Selektionsgrenzen nicht auf die Qualität der Sortierung schließen kann. Dieses ist noch nicht einmal eindeutig möglich, wenn man den Referenzanteil im finalen Transformationscluster kennt, was in der realen Aufgabenstellung ohne Kenntnis der Komplexstruktur nicht gegeben ist, da sowohl gute als auch schlechte Positionen mit guten Verhältnissen nachgewiesen werden können.

### 16.1.3.2 Sortierungsposition im Triplet-Docking obligatorischer Proteinkomplexe

		# Komplexe mit $\geq 2$ Referenzergebnissen			hiervon: # Komplexe mit Sortierungsposition $\leq 100$		
		Rot.-cluster	Ausd.	Trans.-cluster	Rot.-cluster	Ausd.	Trans.-cluster
<b>enge Grenzwerte</b>							
alle Atome	(von 16)	8	8	8	8	8	8
ungesättigte A.	(von 16)	8	8	8	8	8	8
ungebundene A.	(von 16)	8	8	8	8	8	8
<b>mittlere Grenzwerte</b>							
alle Atome	(von 16)	13	12	12	11	12	10
ungesättigte A.	(von 16)	13	12	10	12	12	10
ungebundene A.	(von 16)	13	12	10	13	12	10
<b>weite Grenzwerte</b>							
alle Atome	(von 11)	10	10	4	5	6	2
ungesättigte A.	(von 13)	12	12	9	11	11	5
ungebundene A.	(von 16)	15	15	13	11	10	11

Tabelle IV-30: Obligatorischen Komplexe mit mindestens zwei Referenztripletts und mit Sortierungspositionen von Rang 100 und besser.  
Absolute Anzahl der Komplexe nach der Rotationsbestimmung (Rot.cluster), nach der Ergebnisausdünnung (Ausd.) im finalen Transformationscluster (Trans.cluster).

#### **Qualifizierung: Triplet-Docking mit engem und mittlerem Grenzwertenssemble:**

Beim Triplet-Docking obligatorischer Proteinkomplexe mit den engen und mittleren Grenzwertenssembles ist sogar noch eine eindeutigere Aussage möglich als beim Paar-Docking. Komplexe, deren finaler Translationscluster mehr als zwei Kandidaten aufweist, erhalten mit einer Ausnahme eine Sortierungsposition unter den ersten 100 Plätzen.

Aufgrund der simultanen Selektion aller Dreiecksseiten erbringt jedoch die Hälfte bzw. ein Viertel der Komplexe nicht diese Grundvoraussetzung. Mit den engen Selektionsgrenzen

erbringt die Hälfte der Komplexe ohne Referenzergebnisse auch überhaupt keinen richtig orientierten Transformationscluster.

#### **Qualifizierung: Triplet-Docking mit weitem Grenzwertensemble:**

Mit weiten Selektionsgrenzwerten können alle Komplexe Referenzen nach der Ergebnisausdünnung aufweisen. Bei 2,7% bis 4,5% Anteil an allen verbleibenden kriterienkonformen Kandidaten können diese jedoch im Regelfall wie auch im Paar-Docking keinerlei spezifische Auswirkung auf die Rangfestlegung ausüben. Diese ist somit eher dem Zufall überlassen und sortiert sogar beim Docking mit nur ungebundenen Rotationsclustern mit 1443 Kandidaten inklusive 17 Referenztripletts zum Teil nur auf Rang 131 nach der Ergebnisausdünnung ein.

Mangels Spezifität wird in sieben von elf Experimenten im Docking mit allen Atomen und in immer noch drei von 16 Fällen im Docking mit ungebundenen Atomen das Endergebnis ohne Referenzbeteiligung festgelegt. Die Sortierungspositionen können dabei nur im letzten Fall, bei dem von den spezifischsten Atomen ausgegangen wird, ansatzweise mit der Einwirkung der Referenzen in Zusammenhang gebracht werden. Zumindest erzielt jeder finale Transformationscluster mit mehr als elf Referenzen, die allerdings nur zwischen 14% und 70% der enthaltenen Kandidaten ausmachen, eine Position besser als Rang 100.

Dennoch kann auch hier keine Mindestbelegung mit Kandidaten, geschweige denn Referenztripletts, bestimmt werden, bei der eine gute Positionierung gewährleistet wird. Gerade mal zu Rang 174 reichen zum Teil 39 Kandidaten, wogegen anderswo 34 Kandidaten für Rang 31 qualifizieren.

#### *16.1.4 Qualität der Translationsbestimmung*

Alle Tabellen für das Docking mit engen und mittleren Grenzwerten zeigen, dass die Translationsbestimmung finaler Transformationscluster mit richtiger Rotation deutlich bestimmt wird von den enthaltenen Referenzergebnissen. Translationsabweichungen von 1,0 Å bis 2,1 Å im Paar-Docking und bis 3,8 Å in einem vereinzelt Extremfall des Triplet-Dockings werden immer erreicht wenn auch nur ein richtig positiver Kandidat an der Bestimmung beteiligt ist. Ein grundlegender Fehler von 1,3 Å bis 1,4 Å, der aus dem Orientierungsmodell entsteht, wurde bereits aufgezeigt und vorerst hingenommen. Tendenziell steigen die Abweichung von der bekannten Translation mit abnehmendem Referenzanteil am Endergebnis. Ist kein richtig



positiver Ergebnis an der Translationsfindung beteiligt, so weichen die vorgeschlagenen Verschiebungen um 6,5 Å und mehr vom richtigen Ergebnis ab.

Für alle Dockingvorgänge mit den weiten Selektionsgrenzen ist die Qualität der Translationsbestimmung schon deutlich weniger vom Referenzanteil abhängig, da die Ergebnisse tendenziell schlechter sind. Im Docking mit nur ungebundenen Atomen kommt es bei Referenzbeteiligungen bereits zu Abweichungen von 1,1 Å bis 4,7 Å im Paar-Docking und 0,9 Å bis 13,3 Å im Triplet-Docking. Dieses ist nicht verwunderlich, wenn nur acht von 478 Kandidaten richtig positive Ergebnisse mit der richtigen Translationsvorgabe sind. Tendenzuell nimmt der Fehler auch hier mit abnehmendem Referenzanteil zu, wenn er auch im Regelfall unter 5,0 Å liegt. Finale Transformationscluster ohne Referenzbeteiligung erzeugen willkürliche Translationsvorschriften mit Abweichungen von 13,3 Å und mehr.

## 16.2 Docking fakultativer Proteinkomplexe

Das Docking fakultativer Protein-Protein-Komplexe weist enorme Lücken in den Ergebnistabellen auf, die im Anhang A9 zusammengestellt sind. Hierfür liegen bedeutende Gründe vor:

- Die generelle und die methodenbedingte Eignung eines Test-Datensatzes für ein Dockingexperiment.
- Die sinnvolle Durchführbarkeit eines Dockingexperimentes.

### 16.2.1 Eignung fakultativer Proteinkomplexe

Bereits nach der Präsentation aller selektierenden Grenzwerte konnte in der Vorstellung der Test-Datensätze eine deutlich herabgesetzte Eignung fakultativer Komplexe für die vorgestellten Dockingmethoden festgestellt werden. Gesucht werden zuletzt multiple Observationen identischer Transformationsbeschreibungen. Eine sehr große Anzahl fakultativer Komplexe ist jedoch je nach Methoden- und Parameterwahl nicht in der Lage, eine entsprechende Anzahl richtig orientierter Evaluierungsergebnisse bereitzustellen.

Hauptursache hierfür ist die, im Vergleich zu den obligatorischen Proteinkomplexen, geringere Qualität intermolekularer Wasserstoffbrückenbindungen in den teilweise kleineren Kontaktflächen und hiermit verknüpft die schlechteren Geometrien der Interaktionsvektoren, die bindungsfähige Atome beschreiben.

- Sie erhöht das Risiko fehlerhafter Atomdisqualifikationen.

- Sie vermindert die Zahl essenziell wichtiger repräsentierender Atompaae und -tripletts, da Interaktionsvektoren weniger gleichartig beschaffen sind und somit die geforderten monomolekularen transformations-invarianten Eigenschaften weniger gut erfüllen.
- Sie vermindern die Zahl essentiell wichtiger kriterienkonformer Evaluierungsergebnisse bimolekularer transformations-invarianter Eigenschaften, da Interaktionsvektoren auch weniger optimal zum Bindungspartner ausgerichtet sind.
- Ihrer Rotationsbestimmung kann weniger zuverlässig erfolgen.

Von anfänglich 58 fakultativen, heterooligomeren Protein-Protein-Komplexen können bereits vier Test-Datensätze nie ein einziges Referenzergebnis in ein Docking einbringen.

Abhängig von den verwendeten Methoden und Grenzwertensembles wurden die Test-Datensätze ausgesondert, die weniger als zwei Referenzen in ein Docking einbringen können, um überhaupt multiple Observationen richtig positiver Ergebnisse prinzipiell ermöglichen zu können. Nur 33 Paar- und sechs Triplet-Dockingexperimente bei der Verwendung enger und mittlerer Selektionsgrenzwerte sowie 55 Paar- und 37 Triplet-Dockingexperimente bei der Verwendung weiter Selektionsgrenzwerte sind hierzu in der Lage und werden in Anhang A6 identifiziert.

Diese Untergrenze von zwei Referenzen betrachtet die Ausgangssituation eines Dockings. Der Verlust durch fehlerhafte Rotationsbestimmungen ist hier nicht berücksichtigt. Zudem werden im Paar-Docking drei ähnliche Translationsbeschreibungen in der Ergebnisausdünnung verlangt werden, um zuletzt eine ausreichende Ergebnisspezifität ermöglichen zu können. Hierdurch sind Komplexe mit nur zwei richtig positiven Ergebnissen an dieser Stelle zudem auf die Existenz eines falsch positiven Kandidaten mit gleichartiger Translationsbeschreibung angewiesen oder verlieren die Referenzen.

Die Verlustquoten einzelner Arbeitsvorgänge wurden in vorhergehenden Ergebnisabschnitten diskutiert. Es wurde dargestellt, dass Verluste an Referenzergebnissen mit den vorhandenen Methoden in Kauf genommen werden müssen, um im Endergebnis eine ausreichende Ergebnisspezifität für geeignete Test-Datensätze herzustellen damit richtig positive Ergebnisse in ihrer Bedeutung gegenüber falsch positiven Resultaten gekennzeichnet werden können.

Es verwundert daher nicht, dass zuletzt nur ein Bruchteil der Dockingexperimente fakultativer Proteinkomplexe noch richtig positive Kandidaten mit korrekter Rotationsbestimmung bis in die finalen Transformationscluster einbringen kann. Tabelle IV-31 enthält entsprechende Angaben.

Es wurden vorbereitende Teilrechnungen durchgeführt, die nur die Ergebnisse mit richtigen Rotationsbestimmungen verarbeiten. Aufgrund der benötigten Rechenzeiten sind für das Docking unter Verwendung weiter Selektionsgrenzen mit allen und ungesättigten Atomen vollständige Rechnungen nur für die Komplexe durchgeführt worden, die zumindest ein Referenzergebnis in dem korrekten Transformationscluster vorweisen.

### 16.2.2 Sinnvolle Durchführbarkeit von Dockingexperimenten

Ein weiterer, großer Teil der Rechnungen unter Verwendung des weiten Grenzwertensembles konnte für die fakultative Komplexe nicht zu Ende gerechnet werden bzw. wurde gar nicht erst gestartet. Betroffen sind je nach Atomvorselektion vier bis 16 Experimente im Paar- und drei bis 37 Experimente im Triplett-Docking, wie auch das größte Dockingproblem mit allen Atomen und dem mittleren Grenzwertensemble.

Die Spezifitätsuntersuchungen an den obligatorischen Proteinkomplexen, die in dieser Hinsicht durchgehend bessere Eigenschaften vorweisen, haben gezeigt, dass es ab einer gewissen Menge von kriterienkonformen Kandidaten einfach nicht mehr möglich ist, die richtig positiven Ergebnisse herauszustellen. Ein Dockingexperiment wird mit den vorgestellten Methoden kein ausreichend gutes Ergebnis erbringen können, wenn nicht außergewöhnliche viele Referenzen für einen Komplex zu bestimmen sind, was für die fakultativen Test-Datensätze nie der Fall ist. Das letzte Ergebniskapitel wird zeigen, dass es in gewissem Maße möglich ist, diese Kandidatenzahlen und die Dauer des Dockingexperimentes bereits zuvor aus den Molekülrepräsentationen der Einzelmoleküle abzuschätzen.

Die entsprechenden Rückschlüsse für eine Eignung fakultativer Proteinkomplexe im Rahmen der Aufgabenstellung und für eine sinnvolle Durchführbarkeit mit vertretbaren Rechenzeiten wurden für die einzelnen Dockingexperimente dieser Parameterwahl bereits gezogen. Daher enthält Tabelle IV-31 eine dementsprechende Spalte mit der Anzahl nichtausgeführter Dockingexperimente.

Auch im Docking obligatorischer Komplexe können unter Berücksichtigung aller oder ungesättigter Atome nicht alle Dockingexperimente durchgeführt werden. Aufgrund der geringen Größe verbleibender Testgruppen soll nicht weiter über den jeweiligen relativen Anteil dieser Fälle spekuliert werden. Die später vorgestellte Abschätzung der Größe und Lösbarkeit der Dockingprobleme lassen an sich auch keine Auffälligkeiten fakultativer Komplexe an den Tag treten. Anhand der deutlich geringeren Anzahl von Referenzergebnissen bei gleichen

Rechenanforderungen und Kandidatenzahlen lässt sich jedoch – ohne konkrete Beweisführung – vermuten, dass auch der Anteil der Kontaktfläche an der gesamten Moleküloberfläche für fakultative Komplexe geringer sein könnte als für obligatorische Komplexe. Diese wäre eine ergänzende Aussage zu der Vorstellung der strukturellen Unterschiede in der Architektur der unterschiedlichen Kontaktflächen und zu den Beobachtungen unterschiedlicher Dichten interagierender Aminosäuren in den Kontaktflächen.

### 16.2.3 *Spezifität, Qualifizierung und Translationsbestimmung*

Tabelle IV-31 gibt die Übersicht, wie viele fakultative Komplexe je nach Methoden- und Parameterwahl zuletzt weniger als zwei Referenzen in die finale Clusterbildung einbringen können und wie viele Dockingexperimente als nichtlauffähig eingestuft bzw. in Erwartung eindeutig unzureichender Ergebnisse nicht durchgeführt wurden.

Weiter wird aufgeführt, wie viele gestartete Dockingexperimente zuletzt zwei richtig orientierte Referenzen vorweisen und wie viele dieser Komplexe zudem auch Sortierungspositionen von zumindest Rang 100 erreichen. Zuletzt ist für die Test-Datensätze mit Referenzbeteiligung im finalen Transformationscluster noch der Mittelwert des entsprechenden Referenzanteils eingetragen.

Die detaillierten Ergebnistabellen durchgeführter Dockingexperimente finden sich im Anhang A9.

			# <2 Ref.	# nicht- lauffähig	# ≥2 Ref.	# ≤Rang 100	Referenz- anteil [%]
<b>Paar-Docking</b>							
enge Grenzwerte							
alle Atome	(von 33)		24		9	7	80,0
ungesättigte A.	(von 33)		24		9	7	85,9
ungebundene A.	(von 33)		27		6	6	90,9
mittlere Grenzwerte							
alle Atome	(von 33)		23	1	9	7	65,6
ungesättigte A.	(von 33)		23		10	8	62,4
ungebundene A.	(von 33)		23		10	8	64,2
weite Grenzwerte							
alle Atome	(von 55)		29	19	7	3	4,5
ungesättigte A.	(von 55)		43	8	4	3	5,9
ungebundene A.	(von 55)		43	4	8	5	21,5
<b>Triplett-Docking</b>							
enge Grenzwerte							
alle Atome	(von 6)		2		4	3	70,6
ungesättigte A.	(von 6)		2		4	4	100,0
ungebundene A.	(von 6)		2		4	4	100,0
mittlere Grenzwerte							
alle Atome	(von 6)		1		5	5	77,3
ungesättigte A.	(von 6)		1		5	4	81,3
ungebundene A.	(von 6)		0		6	6	84,2
weite Grenzwerte							
alle Atome	(von 37)		20	15	2	1	10,0
ungesättigte A.	(von 37)		29	4	4	3	33,4
ungebundene A.	(von 37)		30	3	4	3	50,3

Tabelle IV-31: Übersicht der Ergebnisse im Docking fakultativer Proteinkomplexe.  
Details im Text.

# <2 Ref.: Komplexe mit weniger als zwei Referenzen im finalen Transformationscluster.

# nichtlauffähig: nicht durchgeführte Dockingexperimente (siehe Text)

# ≥2 Ref.: Komplexe mit zumindest zwei Referenzen im finalen Transformationscluster.

# ≤Rang 100: hiervon: Komplexe mit Sortierungspositionen von Rang 100 oder besser.

Referenzanteil: Anteil der Referenzen an allen Kandidaten im finalen Transformationscluster.

Die Spezifität der Endergebnisse ist für das Paar-Docking der Test-Datensätze, die zuletzt noch zumindest zwei richtig rotierte Referenzen besitzen, präsentiert sich völlig vergleichbar mit den Ergebnissen obligatorischer Proteinkomplexe. Bei Verwendung der engen und mittleren Selektionsgrenzen werden Referenzanteile von 62,4% bis 90,9% erzielt. In den meisten der Fälle, die richtig positive Ergebnisse bis zum Ende mit sich führen, können auch

Sortierungspositionen besser als Rang 100 erreicht werden. Die hier erreichte Quote ist allerdings nicht hoch wie bei den obligatorischen Komplexen, welche zudem als Mindestbelegung hierfür weniger Kandidaten benötigen.

Mit dem weiten Grenzwertenssemble sind je nach Atomdisqualifikation auch nur 4,5% bis maximal 21% Referenzanteil nach allen Spezifizierungsbemühungen möglich. Es kann genauso wenig eine Mindestbelegung für gute Sortierungspositionen festgelegt werden. Zum Teil sind 852 Kandidaten in einem finalen Transformationscluster nicht ausreichend, eine bessere Positionierung als Rang 116 zu erreichen.

Für das Triplet-Docking überraschen zunächst die guten Referenzanteile in den wichtigen Endergebnissen. Sie liegen sogar bei der Verwendung weiter Grenzwerte zwischen 10% und 50%. Dieses liegt jedoch zunächst einmal auf eine sehr geringe Anzahl durchgeführter Dockingexperimente zurückzuführen. Nur vier bis sechs Experimente für die selektionsstarken und sogar nur zwei bzw. vier Experimente für die selektionsschwachen Grenzwertenssembles setzen die Bedeutung dieser Quoten stark herab. Ansonsten kann auch hier beobachtet werden, dass Komplexe mit zumindest zwei Referenztripletts in fast allen Fällen auch beste Sortierungspositionen bekommen.

Auch die translationsbestimmung folgt dem Docking obligatorischer Komplexe. Wenn genügend Referenzergebnisse im finalen Transformationscluster vorliegen, sind die Abweichungen im Regelfall geringer als 3,0 Å, zunehmend mit abnehmendem Referenzanteil und willkürlich bei Endergebnissen ohne Referenzbeteiligung.

Diese Teilergebnisse, die derart präsentiert doch nur wenige Unterschiede zwischen fakultativen und obligatorischen Protein-Protein-Komplexe aufzeigen, müssen differenziert betrachtet werden:

- Sie zeigen, dass die implementierten Methoden keinen generellen Unterschied machen zwischen den Molekültypen, wenn der Molekülkomplex geeignet ist für diese Behandlung.
- Diese Eignung fakultativer Komplexe ist jedoch in keiner Weise vergleichbar mit der Eignung obligatorischer Komplexe. Die geringere Dichte bindungsfähiger Atome in den spezialisierten Kontaktflächen und besonders auch die vergleichsweise schlechteren Geometrien der Interaktionsvektoren vermindern zu verschiedenen Zeitpunkten der Molekülaufbereitung und der Dockingsimulation die Erfolgsquote, mit der richtig positive

Ergebnisse erzielt und mit den korrekten Transformationsbeschreibungen versehen werden. Fakultative Komplexe benötigen eine deutlich höhere Anzahl erzielbarer Referenzergebnisse, damit diese zuletzt mit richtigen Transformationsbeschreibungen in ihrer Bedeutung herausgestellt werden können.

Nur vier bis zehn von anfänglich 58 fakultativen Komplexen für das Paar-Docking und zwei bis sechs Komplexe für das Triplett-Docking sind soweit geeignet, dass sie überhaupt zumindest zwei Referenzen in die abschließende Ergebnisfindung einbringen können.

### 16.3 Diskussion der Ergebnisqualität

#### **Spezifität:**

Der Anteil richtig positiver Ergebnisse an dem größten Transformationscluster mit der richtigen Rotationsbeschreibung fasst alle Gewinne und Verluste der einzelnen Methoden zusammen. Ebenso gehen hier alle Tendenzen ein, die durch die Verwendung unterschiedlicher Grenzwertensembles oder optionaler Atomdisqualifikationen initiiert werden.

In Übereinstimmung mit allen vorherigen Ausführungen dieser Ergebnispräsentation und Diskussion lassen sich folgende Grundsätze zusammenfassen:

- Die Vorselektion von Atomen, für die keine intramolekularen Wasserstoffbrückenbindungen vorhergesagt werden können, senkt signifikant die anfängliche Anzahl der Interaktionspunkte zur Beschreibung einer Moleküloberfläche. Damit erhöht sich der Anteil von Atomen, die deutlich zu einem potentiellen intermolekularen Bindungspartner ausgerichtet sind. Verringert wird dadurch die Zahl der Ergebnisse mit falscher Rotationsbestimmung. So steigt der relative Anteil von Referenzergebnissen mit richtiger Rotationsbestimmung im Laufe der Ergebnisqualifizierung, wodurch die Spezifität zunimmt und die Identifikation der richtig positiven Kandidaten wahrscheinlicher macht.
- Enge Grenzwertensembles erzeugen eine höhere Ergebnisspezifität, da sie höheren Wert auf die Qualität möglicher intermolekularer Wasserstoffbrückenbindungen legen. Sie wirken an verschiedenen Angriffspunkten und eliminieren besonders die unspezifischen, falsch positiven Ergebnisse mit den nichtoptimierten Interaktionsgeometrien:
  - Sie bevorzugen gleichartig gerichtete Interaktionsvektoren, für die seltener potentielle intramolekulare Bindungspartner gefunden werden können, die zu fehlerhafter Atomdisqualifikation führen.
  - Sie bevorzugen Vektorzusammenstellungen, die eher die Anforderungen der geeigneten Orientierung zur Moleküloberfläche und zueinander erfüllen, wie es die Bewertung der Triplett-Divergenz prüft.
  - Sie erhöhen die Quote richtiger Rotationsbestimmungen speziell für den relativen Anteil der Referenzergebnisse in ihrem Transformationscluster.
  - Sie erhalten eine spezifische Charakteristik spezialisierter Kontaktflächen, indem unterschiedliche Konzentrationen bindungsfähiger Atomensembles in der translationsabhängigen Ergebnisverarbeitung erhalten bleiben. Die besser lokalisierten und abgegrenzten Translationscluster können in der Ausdünnung und



der finalen Clusterbildung einfacher gegenüber den verstreuten Einzelergebnissen zufälliger, falsch positiver Kandidaten isoliert werden, weil sie deutlich überproportional belegt sind.

- Da im Triplet-Docking zwar die Quote richtiger Rotationsbestimmungen höher, die Gesamtanzahl der Referenzergebnisse jedoch geringer ist, fällt der Referenzanteil zuletzt etwas geringer aus als im Paar-Docking.

Insgesamt finden sich somit bei der Überlagerung dieser Abhängigkeiten und Entwicklungen die höchsten Anteile richtig positiver Ergebnisse an den finalen Transformationsclustern mit richtig erfolgter Rotationsbestimmung im Paar- bzw. Triplet-Docking mit den ungebundenen Atomen unter Verwendung des engen Grenzwertensembles. Erreicht werden dabei 87,3% bzw. 76,9%. Die schlechtesten Quoten bei der Verwendung aller Atome mit den mittleren Selektionsgrenzen sind mit immerhin 75,4% bzw. 67,2% zu verzeichnen.

Unter Verwendung weiter Grenzwerte sinkt die Spezifität der Endergebnisse drastisch. Berücksichtigt man nur finale Transformationscluster, in denen überhaupt richtig positive Ergebnisse enthalten sind, so werden beim Docking mit ungebundenen Atomen gerade einmal Referenzanteile von 20% erzielt.

### **Ergebnisqualifizierung und Translationsbeschreibung:**

Finale Transformationscluster mit hohem oder ausschließlichem Referenzanteil können ab einer Mindestgröße sehr gut in ihrer Bedeutung identifiziert werden. Im Regelfall gelingt bei Verwendung enger und mittlerer Selektionsgrenzen eine Qualifizierung des einzig richtig orientierten Suchergebnisses innerhalb der ersten zehn aller 564 Positionen einer absteigenden Sortierung nach Belegungszahlen finaler Transformationscluster.

Bei hohem Referenzanteilen erfolgt die Translationsbestimmungen sehr zuverlässig mit Abweichungen von der bekannten richtigen Verschiebung von zumeist weniger als 2,0 Å inklusive einem methodenbedingten Fehler von etwa 1,3 Å.

Mit zunehmender Ergebniszahl durch Verwendung des weiten Grenzwertensembles oder unselektierter Atome sinkt die Ergebnisspezifität und es kann nicht mehr sichergestellt werden, dass Referenzergebnisse einen hohen Anteil der Endergebnisse ausmachen oder überhaupt in ihnen enthalten sind. Die Sortierungsposition und die Qualität der Translationsbestimmung sind zunehmend willkürlich und dem Zufall überlassen.

**Grenzwertensembles:**

Stark selektionierende Grenzwerte können bewirken, dass Komplexe überhaupt keine oder zu wenig Referenzen bereitstellen können. Diese können dann bei der Rotationsbestimmung, der Ergebnisausdünnung oder der finalen Clusterbildung verloren gehen. Die Gefahr ist größer für die dreifach selektierten Triplets. Sie steigt weiter, wenn bereits die verwendeten Atome vorselektiert werden anhand der möglichen Existenz intramolekularer Wasserstoffbrückenbindungen, bei der es zu fehlerhaften Disqualifikationen kommen kann.

Ansonsten bewirkt diese Vorselektion durchgehend eine Erhöhung der Spezifität der Endergebnisse, die sich in einer zuverlässigeren Identifikation richtig positiver Ergebnisse und besserer Translationsbestimmung bemerkbar macht.

Weite Selektionsgrenzen erhöhen zunächst in allen Fällen die Anzahl der Referenzergebnisse und senken damit an sich das Risiko, richtig positive Ergebnisse in der Ergebnisausdünnung und der finalen Clusterbildung durch fehlende Translationsnachbarn zu verlieren.

Diese Gewinne kommen jedoch bereits nach der Rotationsbestimmung nicht mehr voll zum Tragen. Einerseits steigt das Risiko der fehlerhaften Atomdisqualifikation bei der größeren Anzahl von weniger optimalen Bindungsorientierungen, die mit der schwächeren Selektion zugelassen werden. Weiter ist die Quote richtiger Rotationsbestimmungen mit diesen Interaktionsvektoren geringer. Zudem erhöht sich die Anzahl der falsch positiven Ergebnisse um ein Vielfaches gegenüber der Referenzanzahl. Durch die homogenere Streuung der Translationsbeschreibungen verliert die Ergebnisausdünnung ihre Wirksamkeit und die schon anfangs geringe Ergebnisspezifität kann nur wenig gesteigert werden. Daher ist auch nicht mehr sicher zu stellen, dass der finale Transformationscluster von den richtig positiven Ergebnissen abhängig ist. Die völlige Abwesenheit dieser Ergebnisse bewirkt unspezifische Translationsbestimmungen, die im Regelfall weit von der richtigen Beschreibung entfernt sind. Die Qualität der Translationsbestimmungen unter Referenzbeteiligung sind abhängig vom jeweiligen Referenzanteil.

Eine Einflussnahme des Referenzanteils an den Endergebnissen kann eigentlich nur beim Docking obligatorischer Komplexe mit nur ungebundenen Atomen festgestellt werden. Und selbst in dieser Konstellation sind die Resultate in großem Maße von den Eigenschaften der beteiligten Moleküle abhängig, für welche eine sehr große Varianz bereits nachgewiesen wurde.

**Fakultative Komplexe:**

Fakultative Komplexe haben zum Teil deutlich weniger optimale intermolekulare Wasserstoffbrückenbindungen. Sie können daher von Anfang an weniger Referenzergebnisse

zur Verfügung stellen. Ihre Bindungsstellen werden häufiger fehlerhaft vom Docking ausgeschlossen, da sie mit ihrer Bindungsorientierungen in Verdacht geraten, intramolekulare Bindungen eingehen zu können. Ebenso gelingt die Rotationsbestimmung weniger zuverlässig als mit den regelmäßigen Interaktionsvektoren obligatorischer Komplexe. Somit verstärken sich alle oben aufgeführten Gefahrenquellen für einen Verlust der richtig positiven Ergebnisse und es schwächen sich alle Effekte der Spezifitätssteigerung in der Nachbearbeitung kriterienkonformer Evaluierungsergebnisse mangels guter Orientierungsbestimmung ab.

Fakultative Komplexe sind nur im Einzelfall für ein Docking mit den selektionsstärkeren Grenzwertensembles geeignet. Sie scheitern daran, dass ihre zwar spezialisierten aber weniger gut optimierten und zum Teil kleinen Kontaktflächen, die mit der Qualität intramolekularer Hauptkettenkontakte aus Proteinfaltungsvorgängen nicht mithalten können, nicht genügend kriterienkonforme Ergebnisse zur Verfügung stellen. Andererseits benötigen sie struktur- und methodenbedingt bald die doppelte Anzahl erzielbarer Referenzergebnisse. Nur in den seltensten Fällen besteht dann kein Unterschied zu obligatorischen Komplexen. Deutlich häufiger bieten Komplexe kein oder nur ein Referenzergebnis, wodurch eine Erfüllung des Suchziels multipler Observationen identischer Transformationsbeschreibungen vereitelt wird bzw. auf die Existenz falsch positiver Ergebnisse angewiesen ist.

Die Verwendung weiter Selektionsgrenzen erzeugt wie auch bei den obligatorischen Komplexen eine beträchtliche –vielleicht sogar anteilmäßig größere– Anzahl von Dockingproblemen die zu anspruchsvoll sind, um sie in vertretbarer Rechenzeit zu lösen. Es wurde anhand der betrachteten Ergebnisse gezeigt, dass unter der dann auftretenden Ergebnismenge die richtig positiven Kandidaten in zunehmendem Maße auch nicht zuverlässig identifiziert und in ihrer Bedeutung herausgestellt werden können.

#### **Voraussetzungen für ein erfolgreiches Docking:**

Tabelle IV-32 versucht eine abschließende Übersicht zu geben, wie viele Kandidaten in einem finalen Transformationscluster vorliegen müssen, damit bei den entsprechenden Kombinationen der Methoden und Parameter eine signifikante und gute Positionierung z.B. unter den ersten 100 der 564 Ränge innerhalb einer absteigenden Sortierung nach den Belegungszahlen des größten finalen Transformationsclusters eines Rotationsrepräsentanten erzielt wird.

	Paar-Docking		Triplett-Docking	
	alle Atome	ungebundene A.	alle Atome	ungebundene A.
<b>obligatorische Komplexe</b>				
enge Grenzwerte	≥ 4	≥ 4	≥ 3	≥ 2
mittlere Grenzwerte	≥ 5	≥ 4	≥ 4	≥ 3
weite Grenzwerte	(218)	(51)	(1716)	(39)
<b>fakultative Komplexe</b>				
enge Grenzwerte	≥ 6	≥ 5	≥ 3	≥ 3
mittlere Grenzwerte	≥ 9	≥ 5	≥ 4	≥ 3
weite Grenzwerte	(852)	(727)	(530)	(162)

Tabelle IV-32: Kandidatenbelegung finaler Transformationscluster mit richtiger Rotation für eine Sortierungsposition unter den ersten 100 von 564 Rängen.  
Für das Docking mit dem weiten Grenzwertensemble ist es aufgrund der großen Varianz der Testgruppen und aufgrund der unzureichenden Spezifität der Endergebnisse nicht möglich, derartige Grenzwerte universell festzulegen. Eingeklammerte Werte kennzeichnen Belegungszahlen einzelner Dockingexperimente, die trotz ihrer Größe NICHT für eine Positionierung besser als Rang 100 ausreichen. Die jeweils aufwendigsten Dockingexperimente sind hierbei für das Docking mit allen und mit ungesättigten Atomen nicht enthalten.

Es zeigt sich, dass für obligatorische Protein-Protein-Komplexe bei engen und mittleren Selektionsgrenzen zumindest drei bis fünf kriterienkonforme Kandidaten im Regelfall ausreichend sind, um Spitzenpositionierungen und sehr gute Translationsvorschriften zu erreichen. Bei einer vorzeitigen Erhöhung der Spezifität durch Verwendung nur ungebundener Atome können diese minimalen Belegungszahlen sogar zumeist noch um eine Observation gesenkt werden.

Um auf die Eignung eines Komplexes für dieses Docking schließen zu können, muss noch die unzureichende Orientierungsmethode berücksichtigt werden. Ein Komplex muss somit an sich die zwei- bis dreifache Menge erzielbarer Referenzergebnisse zur Verfügung stellen, damit man den Endergebnissen dieser präselektiven Untersuchung vertrauen kann, ohne die einzig richtige Orientierung vor der eigentlichen Hauptuntersuchung einer genaueren Dockinganalyse zu verlieren.

Für fakultative Komplexe verdoppelt sich zum Teil die Anzahl notwendiger Referenzergebnisse, damit eine Eignung und eine gute Sortierungsposition sichergestellt werden können. Dieses liegt an der geringeren Qualität der Interaktionsvektoren, die eine höhere Fehlerquote in der Atomdisqualifikation und vor allem eine geringere Erfolgsquote in der Rotationsbestimmung bewirkt.

Für die Verwendung des weiten Grenzwertensembles kann zu keiner Methoden- und Parameterkombination und für keinen der Komplextypen eine zuverlässige Mindestbelegung der finalen Transformationscluster bestimmt werden. Gute Positionierungen können zwar erreicht werden, sind jedoch nicht gezwungenermaßen von richtig positiven Ergebnissen abhängig und besitzen keine einheitlichen Belegungszahlen. Selbst für die erfolgreichen Dockingexperimente mit nur ungebundenen Atomen können keine generellen Grenzen bestimmt werden. Zu groß sind die Auswirkungen der individuellen Eigenschaftskombinationen, die sich aus der jeweiligen Molekülkombination in den sehr variablen Testgruppen ergeben. Der Spezifitätsverlust ist zu groß und der Rang in der Sortierung ist zu den größten Teilen willkürlich für Endergebnisse ob mit oder ohne Referenzen.

Das folgende Kapitel wird unter anderem die Frage adressieren, ob für ein Dockingexperiment im Vorhinein die Möglichkeit eingeschätzt werden kann, eine ausreichende Ergebnisspezifität erreichen zu können.

#### **16.4 Enger positionierte Rotationsrepräsentanten und alternative statische und dynamische Qualifizierungsmethoden**

Drei Rotationswinkel sind notwendig für die eindeutige Rotationsvorschrift eines Vektors. Aus der regelmäßigen Unterteilung des  $360^\circ \cdot 180^\circ \cdot 360^\circ$ -Rotationsraums in  $30^\circ$ -Winkelschritten [Lattman, 1972] folgt die Existenz von 564 diskreten Rotationsrepräsentanten. Eine Halbierung der Verteilungsabstände führt zu einer Verachtfachung der Repräsentanten. Alle bisherigen Ergebnisse beziehen sich auf diese  $30^\circ$ -Unterteilung, da die Methode zur Rotationsbestimmung nur sehr unzuverlässig funktioniert. Im Vergleich zu einem Geometrischen Docking, das für eine vollständige Analyse ohne eine Vorselektion der Molekülorientierungen z.B. etwa 2600 anfängliche Rotationsvorschriften auswertet und dann sukzessive verfeinert, sind diese 564 Repräsentanten eine zu ungenaue Bestimmung.

Die Verwendung von  $20^\circ$ -Winkelschritten verteilt die identische Anzahl von kriterienkonformen Kandidaten auf 1867 Rotationsrepräsentanten. Hierdurch könnten die translationsabhängige Ergebnisausdünnung und die Bildung finaler Transformationscluster eventuell unnötig, sicher aber vereinfacht werden. Mit der implementierten Rotationsbestimmung kann momentan nicht sichergestellt werden, dass richtig positive Kandidaten dem korrekten Rotationsrepräsentanten zugeteilt werden. Erst recht kann nicht davon ausgegangen werden, dass ausreichend richtig positive Ergebnisse korrekt orientiert werden, dass sie auch die Ergebnisausdünnung mit einem oder sogar zwei Nachbarn mit ähnlicher Translationsbeschreibung überdauern.

Als Qualifizierungsmethode der als Endergebnisse ausgegebenen Transformationsvorschriften wurde eine absteigende Sortierung aller Rotationsrepräsentanten anhand der Belegungszahlen ihrer größten finalen Transformationscluster kriterienkonformer Kandidaten verwendet. Bislang wurde bewertet, wie zuverlässig die einzig richtige Orientierungsbeschreibung sich auf einem der ersten 100 Ränge positionieren konnte. Dieses gelingt je nach Komplextyp, Methoden- und Parameterwahl mit wechselndem Erfolg.

Die Vereinbarung, als Qualifikationskriterium eine Sortierungsposition von Rang 100 zu verwenden, beschreibt eine Qualifizierungsmethode mit einem fixierten, statischen Grenzwert. Es wird dabei immer die gleiche Anzahl von 100 Transformationsbeschreibungen zur Weiterbearbeitung übergeben.

Ebenfalls einer statischen Qualifizierungsmethode entspricht die Vereinbarung, jede Transformationsvorschrift weiter zu geben, die von einer festzulegenden diskreten Anzahl an Kandidaten beschrieben wird. Hierbei schwankt die Anzahl qualifizierter Endergebnisse nach Methoden-, Parameterwahl und den Eigenschaften der Molekülkombination. Ist die Kandidatenzahl zu groß für den gewählten Grenzwert, so ist es möglich, dass keinerlei Auswahl in der Qualifizierung getroffen werden kann. Die Qualifizierung kann zu verschiedenen Zeitpunkten, direkt nach der Rotationsbestimmung aller Kandidaten, nach der Ergebnisausdünnung oder anhand der finalen Transformationscluster erfolgen.

Um auf die Varianzen der untersuchten Molekülkomplexe zu reagieren, kann versucht werden, nicht einen statischen sondern einen dynamischen Qualifikationsgrenzwert für jedes Dockingexperiment zu erzeugen. Als Abhängigkeiten zu den kombinierten Moleküleigenschaften sind z.B. Korrelationen zu den präselektiven  $Ab_2/Ab_3$ -Werten oder zur den postselektiven Anzahl kriterienkonformer Kandidaten vorstellbar. Einfache dynamische Grenzwerte sind die durchschnittliche Anzahl kriterienkonformer Kandidaten bezogen auf die Anzahl aller Rotationsrepräsentanten oder bezogen auf die Anzahl der Repräsentanten, die von zumindest einem Kandidaten belegt sind.

Statische und dynamische Qualifizierungen sind kombinierbar, wobei die statische Grenze als Tiefpass wirken kann, wenn die dynamische Qualifizierung unzureichend ist. Ebenso lassen sich mehrere Qualifizierungen zu unterschiedlichen Zeitpunkten realisieren und als additive Einzelqualifizierungen oder als kombinatorische Multiqualifizierung, die in allen Abschnitten passiert werden muss, einsetzen.

Bei der Verwendung der 564 30°-Rotationsrepräsentanten verteilen sich zu viele Kandidaten unspezifisch auf zu wenige Ergebnisgruppen, als dass eine Qualifizierung direkt nach der Rotationsbestimmung mit niedrigen statischen Belegungswerten erfolgen könnte. Da die Anzahl richtig positiver Ergebnisse bekannt und relativ niedrig ist, ist es ohne Sinn, einfach hohe Qualifikationsgrenzen zu verwenden.

Aus diesem Grund sollen hier provisorische Ergebnisse für ein Docking mit 1867 20°-Rotationsrepräsentanten präsentiert werden, die einen Eindruck vermitteln sollen, ob und welche Einsparungen mit einfachen statischen und dynamischen Qualifizierungsmethoden erzielt werden können. Es soll dabei unabhängig von den Referenzergebnissen und den Translationsvorschriften nur der Verbleib der richtigen Rotationsvorschrift und die Anzahl qualifizierter Endergebnisse betrachtet werden. Es soll hierbei auch nicht nochmals auf die Varianz innerhalb der Testgruppe eingegangen werden, weshalb nur gemittelte Werte für die gesamte Gruppe wiedergegeben werden, um auf mögliche Einsparungen aufwendigerer Folgeuntersuchungen abschätzen zu können.

Tabelle IV-33 beschreibt das Paar- und Triplet-Docking obligatorischer Proteinkomplexe mit allen optionalen Stufen der Atomvorselektion unter Verwendung der engen und mittleren Grenzwertensembles. Ausgeführt wird eine Qualifikation direkt nach der Rotationsbestimmung.

- Einerseits wird angegeben wie viele Rotationsrepräsentanten überhaupt mit mindestens einem Kandidaten belegt sind, wodurch der Verlust essentieller Ergebnisse in der postselektiven Phase eliminiert würde, wenn die Kandidatenzahlen gering genug sind und sich auf eine ausreichende Anzahl Repräsentanten verteilen.
- Andererseits wird eine individuelle und dennoch statische Qualifizierung gezeigt, wenn die Kandidatenzahlen zu hoch sind. Als “maximaler Grenzwert“ wird die minimale Belegungszahl der korrekten Rotationsrepräsentanten innerhalb der Testgruppe, höchstens aber eine Untergrenze von 15 Kandidaten, verwendet. Somit wird mit maximaler Selektionswirkung und ohne Verlust der richtigen Orientierung für alle Komplexe der Testgruppe gearbeitet. Diese Zahl ist in der realen Aufgabenstellung selbstverständlich nicht bekannt und soll hier nur verdeutlichen, welche Einsparungen mit einem “optimalen Qualifizierungsgrenzwert“ möglich wären. Die maximale Untergrenze von 15 Kandidaten beschreibt bereits den Umstand, dass eine völlig unzureichende Ergebnisspezifität zum Bewertungszeitpunkt vorliegt.
- Zusätzlich wird betrachtet, wie viele der 16 obligatorischen Proteinkomplexe nach der Ergebnisausdünnung bereits keinen Kandidaten mit richtiger Rotationsbeschreibung mehr besitzen und das essentielle Ergebnis damit verloren haben. Hier ist anzumerken, dass im

Paar-Docking vorerst unverändert auf drei Translationsnachbarn selektiert wurde, was bei der geringeren Kandidatendichte in den kleineren Rotationsclustern eigentlich nicht mehr angebracht ist.

- Die letzte Spalte verwendet einen einfachen dynamischen Qualifikationsgrenzwert. Zugelassen werden alle Rotationsgruppen, die direkt nach der Rotationsbestimmung mehr Kandidaten aufweisen können als die mittlere Anzahl aller Kandidaten pro Rotationsrepräsentant. (Summe aller Kandidaten dividiert durch 1867, die Anzahl der Repräsentanten).

	Rotationsbestimmung		Ausdünnung		≥ $\bar{\phi}$ Kandidaten pro Rot.cluster
	≥1	≥max.Grenzwert	≥1		
<b>Paar-Docking</b>					
<b>enges Grenzwertensemble</b>					
alle Atome	$\bar{\phi}$ 1866	(≥19) ≥15	$\bar{\phi}$ 1733	5*Null	4*Null
ungesättigte Atome	$\bar{\phi}$ 1865	≥13	$\bar{\phi}$ 1622	6*Null	3*Null
ungebundene Atome	$\bar{\phi}$ 1864	≥6	$\bar{\phi}$ 1837	6*Null	2*Null
<b>mittleres Grenzwertens.</b>					
alle Atome	$\bar{\phi}$ 1867	(≥42) ≥15	$\bar{\phi}$ 1838	4*Null	3*Null
ungesättigte Atome	$\bar{\phi}$ 1843	(≥23) ≥15	$\bar{\phi}$ 1649	4*Null	2*Null
ungebundene Atome	$\bar{\phi}$ 1866	≥13	$\bar{\phi}$ 1697	4*Null	1*Null
<b>Triplett-Docking</b>					
<b>enges Grenzwertensemble</b>					
alle Atome	$\bar{\phi}$ 1340	≥1	$\bar{\phi}$ 1340	6*Null	$\bar{\phi}$ 791
ungesättigte Atome	1*Null			6*Null	2*Null
ungebundene Atome	1*Null			7*Null	1*Null
<b>mittleres Grenzwertens.</b>					
alle Atome	$\bar{\phi}$ 1565	≥4	$\bar{\phi}$ 942	4*Null	$\bar{\phi}$ 833
ungesättigte Atome	$\bar{\phi}$ 1411	≥3	$\bar{\phi}$ 806	5*Null	$\bar{\phi}$ 826
ungebundene Atome	$\bar{\phi}$ 1274	≥2	$\bar{\phi}$ 864	3*Null	$\bar{\phi}$ 780

Tabelle IV-33: Statische und dynamische Qualifizierung bei 1867 20°-Rotationsclustern  
Angabe der durchschnittlich nach der Rotationsbestimmung oder Ergebnisausdünnung anhand ihrer Kandidatenzahlen qualifizierten Rotationsrepräsentanten. Maximale Grenzwerte (maximal 15, Mindestbelegung für die Testgruppe in Klammern) ermöglichen die optimale Qualifikation für alle Komplexe. "Null" kennzeichnet, dass die richtige Rotation von keinem Kandidaten beschrieben wird. Weitere Details im Text.

Zunächst einmal kann pauschal beobachtet werden, dass selbst bei Verwendung der selektionsstarken Grenzwertenssembles im Paar-Docking die Kandidatenzahlen stets zu hoch sind als dass auch mit den höchsten statischen Qualifikationsgrenzen nennenswerte Einsparungen erzielt werden können. Andererseits sind nach der translationsabhängigen



Ergebnisausdünnung vier bis sechs Komplexe ohne ein Ergebnis mit korrekter Rotationsvorschrift und eine Qualifizierung wäre zu spät. Es muss hierbei beachtet werden, dass trotz verminderter Kandidatenzahl pro Rotationsrepräsentant unverändert nach drei Translationsnachbarn gesucht wurde. Der einfach gewählte dynamische Qualifikationsgrenzwert liegt für das Paar-Docking durchgehend zu hoch, so dass zwischen zwei und vier Komplexe die richtige Orientierung nicht qualifizieren können.

Im Triplet-Docking mit dem engen Grenzwertensemble können nur unter Verwendung aller Atome auch alle Test-Datensätze zumindest ein richtig orientiertes Ergebnis vorweisen. Hierbei qualifizieren sich durchschnittlich 1340 Rotationsrepräsentanten. Ebenso qualifiziert sich die richtige Orientierung sicher für alle Komplexe bei der Bewertung mit dem dynamischen Grenzwert. Durchschnittlich 791 Rotationsrepräsentanten und damit etwa 50% aller möglichen Resultate werden an die nachfolgende Untersuchung weitergeleitet. Die optionalen Vorselektionen bindungsfähiger Atome verhindern eine vollständig verlustfreie Qualifizierung ebenso wie die Bewertung nach der Ergebnisausdünnung, da im besten Fall ein und im schlechtesten Fall bereits sieben Komplexe das richtige Ergebnis verlieren.

Mit den mittleren Selektionsgrenzwerten gelingt eine vollständig verlustfreie Qualifizierung nach der Rotationsbestimmung mit niedrigen statischen und dynamischen Grenzwerten. Zwischen 1274 und 1565 aller 1867 Rotationen weisen zumindest einen Kandidaten auf. Die Einsparungen können gesteigert werden, wenn die jeweiligen optimalen Qualifizierungsgrenzwerte verwendet werden und sich durchschnittlich 864 bis 942 Orientierungen auszeichnen. Die Ergebnisausdünnung erzeugt noch drei bis sieben Komplexe ohne richtig orientiertes Ergebnis. Ebenfalls eine etwa 50prozentige Selektion aller möglichen Rotationsvorschriften erzeugt die Verwendung des dynamischen Grenzwertes, der ohne Verlust 833 bis 780 Resultate weiterleitet.

Tabelle IV-34 enthält die entsprechenden Angaben für das Docking mit weiten Selektionsgrenzwerten unter Verwendung nur ungebundener Atome. Da die Anzahl kriterienkonformer Kandidaten auch für die 20°-Rotationsrepräsentanten die bisherige statische Maximalgrenze von 15 Kandidaten immer ohne Selektionseffekt übersteigt, werden mit den jeweilig optimalen Grenzwerten die Auswirkungen auch für die folgenden Bewertungszeitpunkte nach der Ergebnisausdünnung und nach der finalen Clusterbildung betrachtet. Die Wirkung der dynamischen Qualifizierungsmethode wird unverändert für die Ergebnisse nach der Rotationsbestimmung beschrieben.

	Rot.bestimmung ≥max.Grenzwert	Ausdünnung ≥max.Grenzwert	Trans.clust. ≥max.Grenzwert	≥ $\emptyset$ Kandidaten pro Rot.cluster
Paar-Docking	(≥98) ≥15 $\emptyset$ 1861	≥4 $\emptyset$ 1624	≥1 $\emptyset$ 1680	$\emptyset$ 1107
Triplett-Docking	(≥47) ≥15 $\emptyset$ 1513	≥10 $\emptyset$ 1236	≥4 $\emptyset$ 1003	$\emptyset$ 840

Tabelle IV-34: Statische und dynamische Qualifizierung bei 1867 20°-Rotationsclustern (Docking mit dem weiten Grenzwertensemble und ungebundenen Atomen)  
Angabe der durchschnittlich nach der Rotationsbestimmung oder Ergebnisausdünnung anhand ihrer Kandidatenzahlen qualifizierten Rotationsrepräsentanten. Maximale Grenzwerte (maximal 15, Mindestbelegung für die Testgruppe in Klammern) ermöglichen die optimale Qualifikation für alle Komplexe. "Null" kennzeichnet, dass die richtige Rotation von keinem Kandidaten beschrieben wird. Weitere Details im Text.

Im Docking mit den weiten Selektionsgrenzwerten liegen die minimalen Belegungszahlen innerhalb der Testgruppe direkt nach der Rotationsbestimmung bei 98 kriterienkonformen Paar- bzw. 47 Triplett-Kombinationen. Die Spezifität ist selbst bei der Verteilung auf die 1867 Rotationsrepräsentanten derart gering, dass es sinnlos ist, derart hohe Qualifizierungsgrenzen anzusetzen, wenn doch nur einige wenige Referenzen wirklich Bedeutung besitzen. Dafür weisen jedoch erstmals alle Komplexe auch Ergebnisse mit richtiger Rotationsvorschrift nach der Ergebnisausdünnung und sogar in den finalen Transformationsclustern auf, wenn hier auch nicht immer Referenzergebnisse enthalten sind (Daten nicht gezeigt). Die jeweils optimalen Qualifikationsgrenzen führen allerdings immer noch zu durchschnittlich 1680 Qualifizierungen im besten Fall des Paar- und zu 1003 Qualifizierungen im besten Fall des Triplett-Dockings. Die Varianz dieser Daten ist erwartungsgemäß sehr groß, da die kombinatorischen Effekte voll ausgespielt werden (Daten nicht gezeigt). Ein wenig konstanter und dabei auch effektiver ist die Wirkung des dynamischen Qualifizierungsgrenzwerts, der durchschnittlich 1107 bzw. 840 Rotationen im Paar- bzw. Triplett-Docking an eine nachfolgende Untersuchung weiterleitet, die somit bis zu 50% weniger anfängliche Orientierungen zu untersuchen hat.

Zusammenfassend zeigt sich, dass mit geeigneten statischen und dynamischen Qualifikationsgrenzwerten und dem Umstieg auf eine engmaschigere Unterteilung des Rotationsraums schnell durchschnittliche Vorselektionen von etwa 50% aller Rotationen zu erreichen sind. Eine zuverlässige Rotationsbeschreibung wird die Anzahl richtig rotierter Ergebnisse direkt um das Zwei- bis Dreifache erhöhen, so dass eventuell sogar mit noch enger gesetzten Rotationsrepräsentanten auch noch eindeutigere Qualifizierungen mit geringerem Risiko und höheren Einsparungsraten möglich sind.

Die Wahl des Grenzwertes ist hierbei wieder sehr kritisch, da auch hier die fehlerhafte Disqualifikation der einzig richtigen Transformationsvorschrift droht, wenn er zu hoch angesetzt wird. Für die Kandidatenzahlen im Triplet-Docking mit mittleren Evaluierungsgrenzwerten sind die beispielhaft gewählten Schwellenwerte von zumindest einem Kandidaten nach der Rotationsbestimmung und von der durchschnittlichen Kandidatenzahl pro Rotationsrepräsentant gerade gut geeignet. Engere Selektionsgrenzwerte bewirken, dass zu viele Komplexe keine richtig rotierten Kandidaten mehr besitzen. Zu weite Grenzen erzeugen zu viele unspezifische Evaluierungsergebnisse als dass selbst nach der riskanten Spezifitätssteigerung durch Ergebnisausdünnung und Bildung der Transformationscluster ein guter Selektionseffekt selbst mit den optimalen Grenzwerten sicher erzielt werden könnte.

Die Einteilung des Rotationsraums in  $20^\circ$ -Abschnitte führt nochmals eindrucksvoll das Problem der unzuverlässigen Rotationsbestimmung vor. Nur bei der unspezifischen Evaluierung mit dem weiten Grenzwertensemble bestehen alle Komplexe auch die Ergebnisausdünnung mit richtig rotierten Kandidaten. Besonders beim Paar-Docking mit den engen und mittleren Selektionsgrenzwerten wird die Problematik deutlich, dass zunächst zu viele Ergebnisse für eine effektvolle Qualifizierung erzielt werden, nach der Ergebnisausdünnung aber die korrekte Orientierung bei bis zu sechs Test-Datensätzen verloren gegangen ist.

## 17 Programmlaufzeiten in Abhängigkeit von den kombinatorischen Effekten

Ein letzter Faktor in der Bewertung der vorgestellten Methoden ist die Dauer der einzelnen Untersuchungen, wenn eine zeitkritische Verwendung erwogen werden soll. Hier werden die Laufzeiten für obligatorische Protein-Protein-Komplexe angeführt, da für sie die Verwendbarkeit aller Methoden und Parameterkombinationen besser gegeben ist. Für fakultative Test-Datensätze gibt es jedoch keine generellen Unterschiede in der zugrunde liegenden Kombinatorik und dem Laufzeitverhalten.

Wie bereits erläutert wurde der Gesamttablauf in drei große Arbeitsgänge aufgeteilt, von denen zwei zusammen gehören zur Erstellung der Molekülrepräsentation. Die Arbeitsabschnitte bauen aufeinander auf, nutzen jeweils permanent abgelegte Ergebnisdaten der vorhergehenden Bearbeitungsstufen und sparen somit eine wiederholte Neuberechnung aller Informationen.

### 17.1 Erstellung der Molekülrepräsentation

Die Molekülaufbereitung mit der Qualifizierung und Charakterisierung bindungsfähiger Atome beinhaltet die einmalige Berechnung der lösungsmittelzugänglichen Moleküloberfläche. Diese wird momentan von der eingebundenen Unterfunktion einer frei verfügbaren FORTRAN-Implementierung der *Rolling Ball*-Algorithmen nach Lee und Richards [Lee & Richards, 1971] aus der Arbeitsgruppe um M. Connolly [Connolly, 1986] vorgenommen. Sie ist nach Angaben im Quelltext nur wenig laufzeitoptimiert. Dennoch werden die Einzelmoleküle in durchschnittlich etwa  $4,44 \pm 6,21$  Sekunden (minimal 0,49 Sekunden, maximal 30,29 Sekunden) reiner Prozessornutzung einer SGI<sup>®</sup>-Octane mit R10000-Prozessor verarbeitet. Zu dieser Molekülaufbereitung muss noch eine Bearbeitung mit HBPLUS und die Erstellung der Listen intramolekularer Wasserstoffbrückenbindungen oberflächennaher Atome hinzugerechnet werden. Auch diese einmaligen Arbeiten werden in  $1,48 \pm 1,07$  Sekunden (maximal 4,80 Sekunden) erledigt.

Da diese gesamte Molekülaufbereitung nur ein einziges Mal ausgeführt werden muss, wurden hierbei keine weiteren Bemühungen zur Optimierung angestellt. Es sind aber Implementierungen der Oberflächenberechnung mit weitaus geringerem Zeitbedarf bekannt.

Die Zusammenstellung der Atompaaire und -tripletts eines einzelnen Proteinmoleküls unter Selektion monomolekularer transformations-invarianter Eigenschaften ist abhängig von der

Anzahl der verwendeten Atome und von den verwendeten Selektionsgrenzwerten, da mit abnehmender Selektionsstärke mehr Evaluierungen bis zum Ende durchgeführt werden müssen. Dieser Arbeitsgang ist bei Verwendung unterschiedlicher Optionen zur Atomdisqualifikation aufgrund intramolekularer Wasserstoffbrückenbindungen und unterschiedlicher Grenzwertensembles vor jedem Docking auszuführen. Alternativ können die Paar- und Tripletlisten für die verschiedenen Parameterkombinationen an unterschiedlichen Orten oder mit differenzierten Namen hinterlegt und gezielt abgefragt werden.

Insgesamt stellt dieser Arbeitsgang jedoch keine Ansprüche, da er selbst mit allen Atomen und den weitesten Kriteriengrenzwerten auf einer SGI<sup>®</sup>-Octane lediglich  $0,36 \pm 0,32$  (maximal 1,54 Sekunden) für die Erstellung der Paarlisten und  $0,19 \pm 0,22$  Sekunden (maximal 1,10 Sekunden) für die Tripletlisten in Anspruch nimmt.

## 17.2 Docking zweier Moleküle

Kritisch ist der eigentliche Dockingvorgang. Es wurde bereits dargelegt, welche kombinatorischen Effekte aus der Atomqualifikation zur Paar- und Tripletztzusammenstellung und von dort aus zur intermolekularen Kombinations-Evaluierung zu tragen kommen.

- Somit ist die Laufzeit in großem Maße abhängig von der Paar- oder Tripletanzahl beider Moleküle und daher auch direkt von der Atomanzahl, optionaler Disqualifikation und der verwendeten Selektionsstärke monomolekularer transformations-invarianter Eigenschaften. Insbesondere für Atomtripletts wurde bezüglich ihrer Anzahl versucht, zusätzliche Bewertungen wie die Vektordivergenz anzuführen und unnötige Evaluierungen zu verhindern, wenn es gelingt, sicher die Bindungsseite eines Dreiecks zu identifizieren.
- Zweiter Einflussfaktor ist die Selektionsstärke für bimolekulare transformations-invariante Eigenschaften. Bei den sehr großen Mengen von Evaluierungen, die jeweils gestartet werden müssen, ergeben sich bereits erhebliche Laufzeitänderungen durch die Dauer der einzelnen Evaluierungen bzw. dadurch wie früh sie gegebenenfalls als nicht-kriterienkonform abgebrochen werden können.
- Als letzter Faktor kommt die weitere Ergebnisverarbeitung hinzu. Diese beginnt mit der notwendigen Transformationsbestimmung eines kriterienkonformen Kandidaten inklusive der Suche nach dem ähnlichsten Rotationsrepräsentanten. Hier ergibt sich bereits noch eine weitere Zeitabhängigkeit, wenn an der Methode festgehalten wird. Bei der Verwendung einer höheren Dichte der Repräsentanten, die angestrebt werden muss, verachtfacht sich die Anzahl der Repräsentanten bei jeder Halbierung der Winkelschritte. Ebenfalls in den abschließenden Arbeitsschritten enthalten ist die Erstellung der Translationscluster und die Sortierung aller finalen Transformationscluster anhand ihrer Belegungszahlen. Die

Clusterung geschieht momentan mit der hierarchischen Methode zur Erstellung von *Single-Linkage*-Clustern. Diese weist bei Zunahme der Datenpunkte zunehmend ungünstiges Laufzeitverhalten auf. Der Faktor würde allerdings an Bedeutung verlieren, wenn die Rotationsbestimmung zuverlässiger gelöst werden kann und sich dann die Rotationsrepräsentanten enger setzen lassen. Dadurch verteilt sich die unveränderte Gesamtmenge kriterienkonformer Kandidaten auf ein Vielfaches an Rotationsclustern, die somit bedeutend geringere Belegungen besitzen. Als Folge würde entweder eine intermediäre Ergebnisausdünnung unnötig und die Clusterung selber würde dennoch beschleunigt oder man könnte eventuell vollständig auf die Clusterung verzichten. Letzteres wäre der Fall, wenn eine Ergebnisausdünnung bereits alle richtig positiven Kandidaten isoliert und so in ihrer Bedeutung anhand der signifikant besseren Belegungszahlen mit ausreichender Spezifität herausstellt.

#### 17.2.1 *Programmlaufzeiten*

Die Tabellen in Anhang A10 enthalten für das Paar- und Triplett-Docking mit den drei Grenzwertensembles und für die drei optionalen Stufen der Atomdisqualifikation für jedes Dockingexperiment die benötigten Programmlaufzeiten. Alle Dockingrechnungen wurden auf einer SGI®-Origin mit R10000-Prozessor ausgeführt. Angegeben ist die reine Nutzungsdauer eines Prozessors.

Die folgende Tabelle IV-33 enthält eine Zusammenfassung dieser detaillierten Angaben. Sie führt jeweils die Mittelwerte und Standardabweichungen für die gesamte Testgruppe obligatorischer Proteinkomplexe auf. Da die Standardabweichungen die Mittelwerte in vielen Fällen übertreffen, enthält die Tabelle zudem eine Auflistung der Minimum- und Maximumwerte, um diese enorme Varianz vorzuführen. Für die Werte der Dockingexperimente mit den weiten Grenzwerten ist zu beachten, dass nur unter Verwendung ungebundener Atome die gesamte Testgruppe abgearbeitet wurde. Beim Docking mit ungesättigten Atomen sind die Laufzeiten der drei größten Komplexe und beim Docking mit allen Atomen die Werte für die fünf größten Komplexe nicht enthalten, da die Rechnungen aufgrund ihrer Dauer nicht zu Ende geführt wurden.

	alle Atome	ungesättigte Atome	ungebundene Atome
<b>Paar-Docking</b>			
enge Grenzwerte	34,8 ±41,4	18,1 ±19,2	11,2 ±10,0
mittlere Grenzw.	71,2 ±91,4	33,3 ±37,8	19,3 ±19,0
weite Grenzw.	5257,7 ±8162,6 (11) 87,6 ± 136,0 min	1809,5 ±2650,0 (13) 30,2 ± 44,2 min	1904,8 ±2981,8 (16) 31,8 ± 49,7 min
<b>Triplett-Docking</b>			
enge Grenzwerte	2,1 ±6,0	1,2 ±0,8	0,8 ±0,6
mittlere Grenzw.	4,5 ±3,3	2,3 ±1,8	1,6 ±1,3
weite Grenzw.	5139,6 ±6615,0 (11) 85,7 ± 110,3 min	1012,7 ±1211,6 (13) 16,9 ± 20,2 min	738,7 ±1087,3 (16) 12,3 ± 18,1 min
<b>Paar-Docking</b>			
enge Grenzwerte	4,5 - 166,4	3,0 - 75,4	2,2 - 39,5
mittlere Grenzw.	7,4 - 358,5	4,6 - 143,1	3,3 - 73,5
weite Grenzw.	165,0 - (27528) (~7,7 h)	59,7 - (9585)	32,5 - 815,3
<b>Triplett-Docking</b>			
enge Grenzwerte	0,6 - 6,0	0,3 - 2,9	0,2 - 2,5
mittlere Grenzw.	0,8 - 12,5	0,5 - 6,4	0,4 - 5,3
weite Grenzw.	133,5 - (19589) (~5,4 h)	39,5 - (4023)	22,1 - 3952,5

Tabelle IV-33: Mittelwerte, Standardabweichungen sowie minimale und maximale Programmlaufzeiten im Docking obligatorischer Komplexe. Alle Angaben sind in Sekunden gemacht außer bei den gekennzeichneten Minutenwerten (min). Details im Text.

### Docking mit engem und mittlerem Grenzwertensemble:

Die absoluten Dockingzeiten betragen je nach Vorselektion der Atome im Paar-Docking durchschnittlich 11,2 bis 24,8 Sekunden (maximal 166,4 s) mit den engen und 19,3 bis 71,2 Sekunden (maximal 358,5 s) mit den mittleren Selektionsgrenzen. Das entsprechende Triplett-Docking wird in 0,8 bis 2,1 Sekunden (maximal 6,0 s) bzw. 1,6 bis 4,5 Sekunden (maximal 12,5 s) vollendet.

Für die Laufzeitentwicklungen bei der Verwendung enger und mittlerer Grenzwertensembles lassen sich sehr gut relative Verhältnisse bestimmen:

- Ein Docking mit den mittleren Grenzwerten dauert etwa doppelt so lang wie ein Docking mit den engen Grenzwerten.
- Ein Docking mit allen Atomen dauert etwa dreimal so lange wie ein Docking mit nur ungebundenen Atomen. Bei der Verwendung ungesättigter Atome wird etwa die Hälfte der Zeit eines Dockings mit allen Atomen benötigt.

- Ein Paar-Docking benötigt etwa die 16fache Zeit eines entsprechenden Triplet-Dockings. Diese Differenz nimmt ein wenig ab bei der Vorselektion der Atome und differiert dann um einen Faktor von 13.
- Die Varianz innerhalb der Testgruppe kommt mit zunehmendem Evaluierungsaufwand stärker zum Tragen:
  - Im Paar-Docking mit ungebundenen Atomen beträgt der Laufzeitunterschied des kürzesten und längsten Dockings bei beiden Grenzwertensembles etwa einen Faktor von 20.
  - Beim Docking mit allen Atomen wächst diese Differenz bereits auf einen Faktor von 40 bzw. 50 bei der Verwendung der engen bzw. mittleren Grenzwerte an.
  - Im Triplet-Docking sind die Unterschiede der Evaluierungsaufufe und damit die Grundlage der Varianzen aufgrund der dreifachen Selektion deutlich geringer. Die Laufzeitunterschiede liegen daher stets zwischen den Faktoren 10 bis 15.

Insgesamt kann bei der Verwendung dieser Selektionsgrenzen noch von einer größtenteils linearen Abhängigkeit der Evaluierungsaufufe zur Laufzeit ausgegangen werden. Erst größere Komplexe im Paar-Docking zeigen stärker ansteigenden Zeitbedarf. Diese Entwicklung wird bei der Betrachtung des Dockings mit weiten Grenzwerten noch besser zu sehen sein und soll dann bei der Darstellung der Abhängigkeiten der Dockingdauer erläutert werden.

Nimmt man ganz naiv und pauschal an, dass ein Geometrisches Docking ohne Vorselektion der Molekülanordnungen 60 Minuten pro Molekülpaarung benötigt, und dass es möglich ist, die hierbei zu untersuchenden Molekülorientierungen um 80% zu senken (entsprechend etwa auf 100 von 564 Rotationen mit zusätzlicher Translationsbeschreibung), so würde der Zeitbedarf der Hauptuntersuchung auf 12 Minuten gesenkt. Im allerschlechtesten Fall benötigt dann diese Vorselektion anhand eines Dockings mit Wasserstoffbrückenbindungen selber weitere 6 Minuten. Im Regelfall würde die Voruntersuchung jedoch mit durchschnittlich je nach Methoden- und Parameterwahl 0,8 bis 71 Sekunden kaum ins Gewicht fallen, so dass nur die Laufzeit des Geometrischen Dockings zeitbestimmend ist.

#### **Docking mit weitem Grenzwertensemble:**

Die Zahlen für das Docking mit dem weiten Grenzwertensemble sind nicht derart durchgängig zu analysieren. Bei der Verwendung ungesättigter Atome sind bereits die drei größten, bei der Verwendung aller Atome sogar die fünf größten Proteinkomplexe nicht berechnet worden. Ihr Zeitbedarf liegt weit über dem, was für ihre Anwendung vertretbar ist und es wurde bereits



nachgewiesen, dass die enormen Ergebnismengen eine richtige Ergebnisfindung mangels Spezifität der richtig positiven Kandidaten vereitelt.

Das Paar-Docking mit ungebundenen Atomen benötigt durchschnittlich mit 31,8 Minuten bereits die 100fache Dauer des Dockings mit mittleren Grenzwerten. Für das Triplet-Docking beträgt der Unterschied sogar das 460fache.

Die Unterschiede zwischen Paar- und Triplet-Docking haben sich angeglichen und das Paar-Docking benötigt nur noch die zweieinhalbfache Zeit.

Die Varianz innerhalb der Testgruppe beträgt bereits beim Paar-Docking mit ungebundenen Atomen mit minimal 32,5 und maximal 8152,3 Sekunden einen Faktor von 250. Im Triplet-Docking ist der Unterschied zwischen 22,1 und 3952,5 Sekunden mit einem Faktor von 180 noch etwas geringer. Beide Male werden die Varianzen aus der Verwendung stärker selektionierender Grenzwerte um ein Vielfaches übertroffen.

Für die Verhältnisse bei Verwendung ungesättigter oder aller Atome lassen sich allenfalls Vermutungen anstellen. Maximale Rechenzeiten für den elftgrößten von 16 Komplexen liegen aber bereits bei etwa 7,7 Stunden im Paar- und 5,4 Stunden im Triplet-Docking. Hiermit zeichnet sich bereits eine Nichtanwendbarkeit dieser Grenzwerte für derart große Dockingprobleme ab.

Andererseits bewegen sich die Rechenzeiten kleiner Komplexe je nach Vorselektion der Atome mit minimalen 32,5 bis 165,0 Sekunden im Paar- und 2,1 bis 133,5 Sekunden im Triplet-Docking durchaus in Bereichen, die für die Konformationsvorselektion eines Geometrischen Dockings durchaus sinnvoll sind, wenn ausreichend spezifische Ergebnisse erzielt werden können.

Hiermit stellen sich die Fragen, was die Ursache für den übermäßigen Anstiegs des Zeitbedarfs bei Verwendung der weiten Selektionsgrenzen ist und ob man die benötigte Rechendauer bereits vor dem eigentlichen Docking abschätzen kann.

### 17.2.2 *Abhängigkeiten der Dockingdauer*

Zu Beginn des Kapitels wurde angeführt, dass der Zeitbedarf eines Dockings abhängig sein muss von

- der zugrunde liegenden Atomanzahl also der Molekülgröße,

- von den verwendeten Selektionsgrenzwerten, die Einfluss haben auf den Umfang der Molekülrepräsentation und Anzahl sowie Dauer der Evaluierungen
- und zuletzt von der Anzahl kriterienkonformer Kandidaten, die in der Ergebnisverarbeitung zu verwerten sind.

Im Laufe dieser Arbeit wurden bereits Größen vorgestellt und verwendet, die diese drei Effektoren sehr gut kennzeichnen können:

- Das errechenbare Produkt der repräsentierenden Paare bzw. Triplets beteiligter Proteinmoleküle nach der Selektion monomolekularer transformations-invarianter Eigenschaften. Diese Produkt wird erweitert um einen Faktor Zwei bzw. Drei, welcher der Anzahl notwendiger Orientierungen entspricht, die zu betrachten sind, damit jede mögliche Atom-Atom-Interaktion berücksichtigt wird. Diese Werte sollen im Folgenden einfach als "Ab2-" bzw. "Ab3-Werte" bezeichnet werden und entsprechen der Anzahl theoretischer Evaluierungsaufrufe.
- Die gezählte Anzahl der Evaluierungsaufrufe bimolekularer transformations-invarianter Eigenschaften.
- Die gezählte Anzahl kriterienkonformer Kandidaten nach der Evaluierung bimolekularer transformations-invarianter Eigenschaften und der Rotationsbestimmung.

Der besonders interessante Aspekt an der berechenbaren Anzahl der Ab2/Ab3-Werte ist die Möglichkeit, sie vor dem eigentlichen Docking bestimmen zu können. Hierzu müssen nur die sehr schnell anzufertigenden Molekülrepräsentationen der Dockingpartner erstellt und die entsprechenden Paar- oder Triplettzahlen entsprechend kombiniert werden.

Somit ist zu betrachten, welche Abhängigkeit zwischen den drei beschriebenen Größen und der für die Test-Datensätze bekannten Dockingdauer besteht.

Die Histogramme der Abbildung IV-8 tragen für alle einzelnen Dockingexperimente obligatorischer Proteinkomplexe den berechneten Ab2- bzw. Ab3-Wert gegen die Anzahl der erfolgten Evaluierungsaufrufe und gegen die Anzahl der resultierenden kriterienkonformen Kandidaten vor der Ergebnisausdünnung auf.

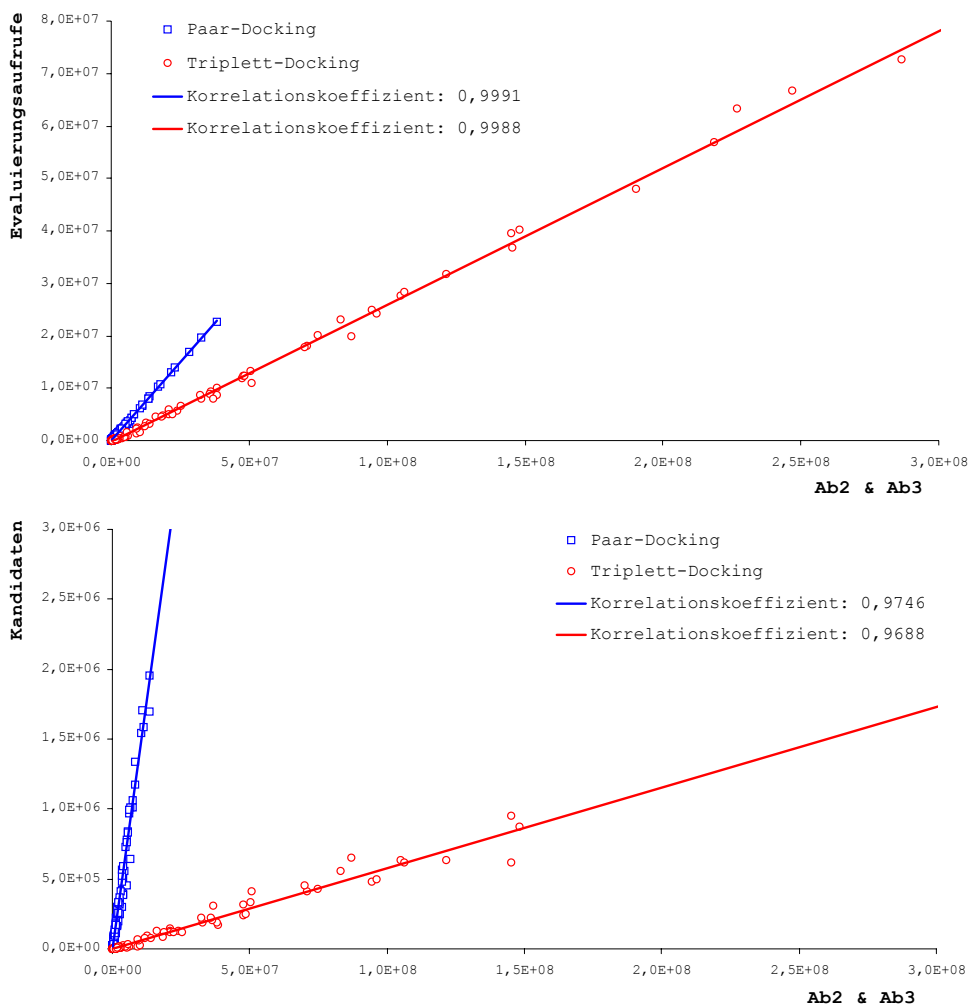


Abbildung IV-8: Aufgetragen sind die Ab2- und Ab3-Werte gegen die Anzahl der Evaluierungsaufrufe (oben) bzw. der kriterienkonformen Kandidaten (unten) für alle durchgeführten Dockingexperimente obligatorischer Proteinkomplexe.

Beide Male sind sehr gute lineare Abhängigkeiten mit Korrelationskoeffizienten von wenigstens 0,97 sowohl für die Werte der Paar- als auch der Triplett-Dockingexperimente deutlich ersichtlich. Diese werden mit den entsprechenden Trendlinien illustriert. Die Histogramme sind auf Ab2/Ab3-Werte bis 300 Millionen begrenzt, obwohl höhere Werte für einige Test-Datensätze vorliegen. Hier sind jedoch keine Funktionswerte bekannt, da die Dockingexperimente nicht vollständig zu Ende gerechnet wurden

Unter Berücksichtigung dieser linearen Abhängigkeiten muss nun nur noch die Korrelation der Ab2/Ab3-Werte zur Dockingdauer betrachtet werden.

### Docking mit engem und mittlerem Grenzwertensensemble:

Die Histogramme der Abbildung IV-9 stellen diese Abhängigkeit getrennt nach Paar- und Triplet-Docking für alle Observationen des Dockings mit engen und mittleren Selektionsgrenzwerten dar.

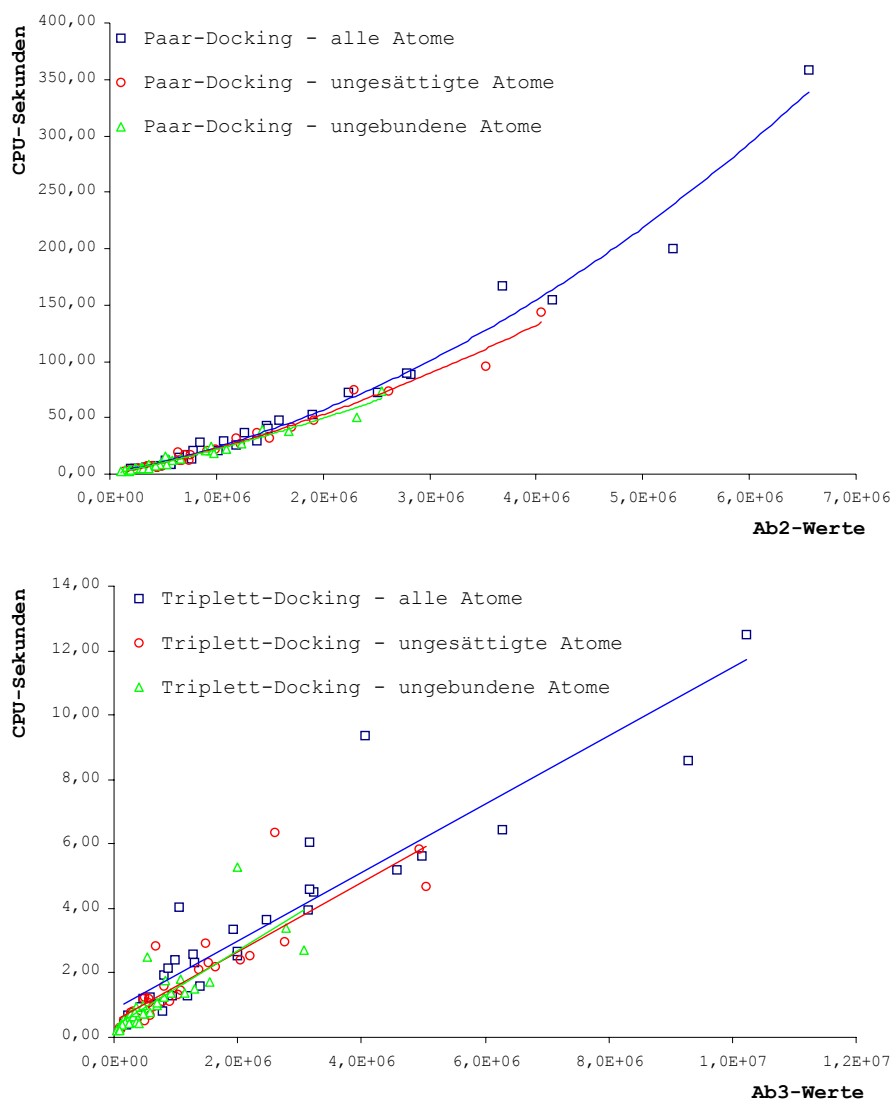


Abbildung IV-9: Aufgetragen sind für die Verwendung des engen und mittleren Grenzwertensamples die Ab2-Werte des Paar-Dockings (oben) und die Ab3-Werte des Triplet-Dockings (unten) gegen die Dauer der einzelnen Dockingexperimente.

In die Histogramme sind Trendlinien eingezeichnet, die für die betrachteten Bereiche jeweils recht deckungsgleich verlaufen. Diese Tendenzen sind beim Triplet-Docking sehr linearer Natur für

die gesamte Testgruppe. Einzelne Komplexe weisen für verschiedene Atomdisqualifikationen für sich deutliche Korrelationsgeraden auf, die individuell flacher oder steiler verlaufen. So setzen sich in Abbildung IV-9 zwei Komplexe mit auffällig steiler steigenden Korrelationsgeraden von der übrigen Testgruppe ab. Für die Testgruppe betrachtet erscheint die Korrelation im Triplett-Docking nicht außerordentlich gut zu sein. Hierbei ist allerdings zu berücksichtigen, dass die Werte der Dockingergebnisse mit engen und mittleren Grenzwerten zusammen dargestellt sind und in die Trendlinie einfließen. Daher sollen hier auch weiter keine Korrelationskoeffizienten angegeben werden.

Im Paar-Docking deuten sich bei Ab2-Werten über 3 Millionen eine deutlich nichtlinear zunehmende Entwicklung an. Die Trendlinien fächern zunehmend auf, wobei größere Atomzahlen größere Kurvensteigungen bewirken.

Diese im Weiteren noch zunehmende Auffächerung lässt sich wie folgt begründen. Wenn die Anzahl der Ab2- bzw. Ab3-Werte und somit auch der kriterienkonformen Kandidaten identisch ist, jedoch einmal mit allen und einmal mit vorselektierten Atomen gearbeitet wird, so muss im ersten Fall ein kleinerer und im zweiten Fall ein größerer Molekülkomplex gedockt werden. Somit liegen im ersten Fall die Translationsbeschreibungen aufgrund der höheren Dichte verwendeter Atome auch näher beieinander. Die Ergebnisausdünnung kann daher im Vergleich nicht so effektiv arbeiten, weshalb die finale Transformationsclusterung mit mehr Punkten zu arbeiten hat und entsprechend mehr Zeit benötigt.

Die angesprochene Laufzeitabhängigkeit von der Ergebnisverarbeitung nach der erfolgreichen Kandidatenidentifikation ist somit begründet.

Insgesamt ist die Korrelation für das Paar-Docking jedoch in diesen Bereichen recht gut, so dass es durchaus möglich ist, über die Ab2-Werte die Dockingdauer ab zu schätzen. Für das Triplett-Docking ergibt sich in den Ab3-Wertebereichen bis zu 12 Millionen kaum die Notwendigkeit die Programmlaufzeit abschätzen zu müssen, da die Laufzeiten hier bei maximal 12,5 Sekunden liegen.

#### **Docking mit weitem Grenzwertensemble:**

Die entsprechenden Histogramme für das Triplett-Docking mit weiten Selektionsgrenzwerten sind in Abbildung IV-10 wiedergegeben. Die verzeichneten Ab2-Werte sind bei 14 Millionen, die Ab3-Werte bei 160 Millionen begrenzt, da das Docking größerer Komplexe unter Verwendung aller und ungesättigter Atome nicht zu Ende gerechnet wurde.

Deutlich ist in beiden Fällen die nichtlineare Entwicklung und die Auffächerung der polynomischen Trendlinien für die unterschiedlichen Atomdisqualifikationen zu sehen. Die großen Laufzeitunterschiede bei gleichbleibenden Ab2/Ab3-Werten sind entsprechend obiger Begründung auf den gestiegenen Rechenbedarf in der Ergebnisverarbeitung und hier besonders in der finalen *Single-Linkage*-Clusterbildung zurückzuführen. Andere Clustermethoden, die effizienter mit den deutlich größeren Mengen von Datenpunkten umgehen können, werden diese Diskrepanz senken.

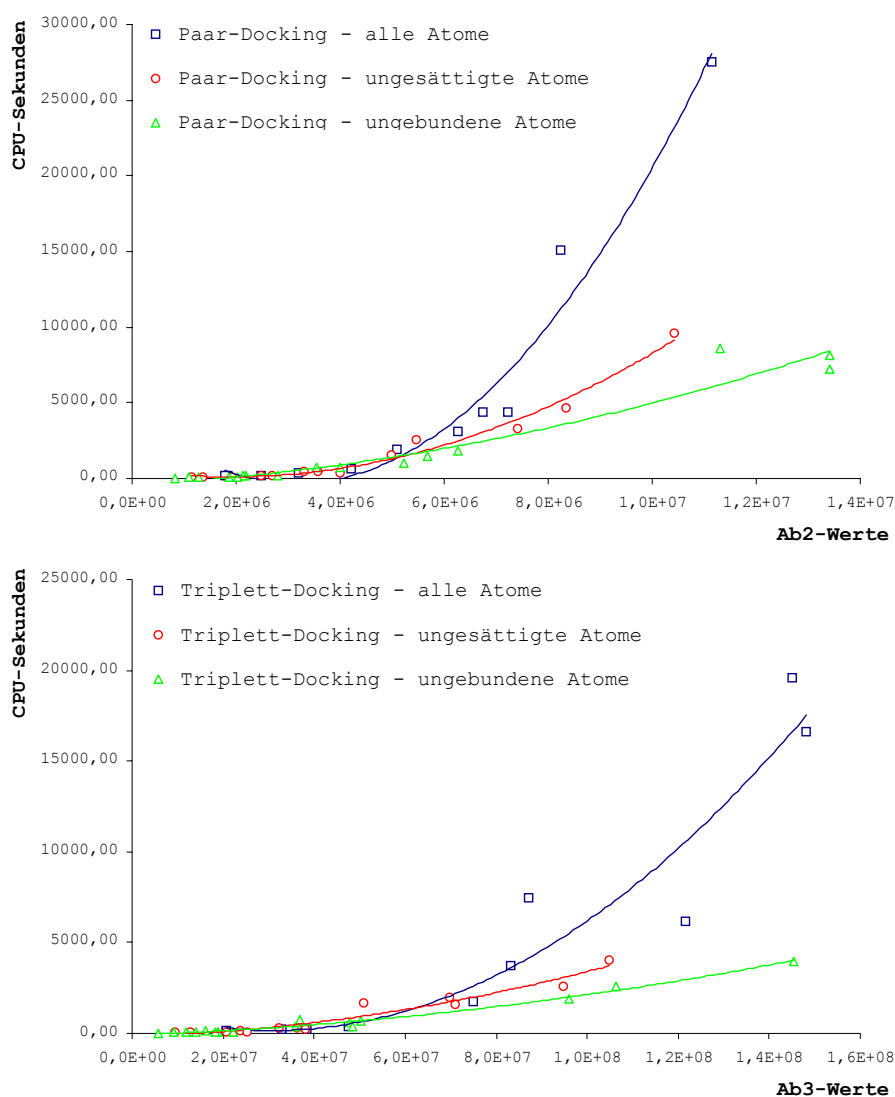


Abbildung IV-10: Aufgetragen sind für die Verwendung des weiten Grenzwertensembles die Ab2-Werte des Paar-Dockings (oben) und die Ab3-Werte des Triplett-Dockings (unten) gegen die Dauer der einzelnen Dockingexperimente.

Anhand der Kurven wird deutlich, weshalb die größten Komplexe unter Verwendung ungesättigter und aller Atome nicht zu Ende gerechnet wurden:

- Mit ungesättigten Atomen stehen Rechnungen mit Ab2-Werten von 18 bis 23 Millionen und Ab3-Werten von 190 bis 286 Millionen aus. Diese Werte betragen fast das Doppelte der entsprechenden Werte für die bisher längsten Rechnung, welche am Ende eines Kurvenverlaufs stehen, der die lineare Phase deutlich verlassen hat.
- Mit allen Atomen stehen Rechnungen mit Ab2-Werten von 17 bis 38 Millionen und Ab3-Werten von 226 bis 548 Millionen aus. Die dazugehörigen Kurven wurden bei Maximalwerten von 11 Millionen und über 7,5 Stunden Rechenzeiten im Paar-Docking und 145 Millionen und über 5,5 Stunden Rechenzeit im Triplett-Docking vorerst beendet.
- Die fakultativen Komplexe, deren Rechnungen zum vorhergehenden Ergebniskapitel nicht abgeschlossen bzw. gestartet wurden, weisen alle Ab2-Werte von 13 bis 50 Millionen und/oder Ab3-Werte von 40 bis 436 Millionen auf.

### 17.3 Diskussion der Programmlaufzeiten

Die Erstellung der Molekülrepräsentation mit Bestimmung und Charakterisierung der bindungsfähigen Atome ist wenig zeitkritisch, da sie nur einmalig für jedes einzelne Proteinmolekül durchzuführen ist.

Die Zusammenstellung der kriterienkonformen Atompaaire und -tripletts muss jedesmal in Abhängigkeit der verwendeten monomolekularen Selektionsgrenzwerte und der gewählten Atomdisqualifikation erfolgen. Sie benötigt jedoch bislang niemals mehr als durchschnittlich 0,36 Sekunden Rechenzeit mit einem absoluten Spitzenwert von 1,54 Sekunden. Die Ergebnisse könnten zudem ebenfalls permanent abgelegt und gezielt abgefragt werden.

Die Dauer des Proteindockings ist von der Methoden- und Parameterwahl und in sehr großem Maße von der Molekülgröße bzw. dem Umfang der Molekülrepräsentationen beteiligter Dockingpartner abhängig. Es konnte eine sehr gute, lineare Korrelation der Ab2- bzw. Ab3-Werte zu den im Docking zählbaren Evaluierungsaufrufen und zu den hieraus hervorgehenden kriterienkonformen Kandidaten aufgezeigt werden. Diese Ab2/Ab3-Werte lassen sich nach der Erstellung der Molekülrepräsentationen und somit vor dem Moleküldocking berechnen und enthalten indirekte Angaben über die Molekülgröße und Dichte bindungsfähiger Atome.

Für die Verwendung der selektionsstarken Grenzwertensembles kann eine relativ gute Abhängigkeit der Dockingdauer zu diesen Werten beobachtet werden, so dass vor der Durchführung des Dockings der Zeitbedarf abgeschätzt werden kann.

Das Triplet-Docking mit engen und mittleren Selektionsgrenzen erfolgt sehr schnell mit einem maximalen Durchschnittswert von nur 4,5 Sekunden und einem absoluten Spitzenwert innerhalb der Testgruppe obligatorischer Komplexe von 12,5 Sekunden. Das entsprechende Paar-Docking weist bereits einen Durchschnittswert von 71,2 Sekunden mit einer Standardabweichung von 91,4 Sekunden auf. Nur drei der 16 Dockingexperimente benötigen lange Rechenzeiten von über 90 Sekunden bis maximal 358,5 Sekunden.

Mit den weiten Selektionsgrenzwerten kommt es zu einem deutlich stärkeren Anstieg der Programmlaufzeiten mit maximalen Durchschnittswerten unter Verwendung aller Atome von 5258 Sekunden im Paar- und 5140 Sekunden im Triplet-Docking. Hierin sind die Rechnungen der fünf größten Komplexe nicht enthalten.

Die Trendlinien der Histogramme zeigen eine erwartete Abhängigkeit von der Molekülgröße bzw. den Ab<sub>2</sub>/Ab<sub>3</sub>-Werten zu den Programmlaufzeiten. Diese zeigt jedoch nicht mehr linearen sondern zunehmend exponentiellen Kurvencharakter. Zusätzlich lässt sich eine bedeutend stärkere Zunahme der Laufzeiten in Abhängigkeit von der Atomvorselektion und somit von der Dichte bindungsfähiger Atome in der Moleküloberfläche verzeichnen. Hieraus kann geschlossen werden, dass die Ergebnisausdünnung, vor allem aber die Bildung finaler *Single-Linkage*-Cluster der Transformationsbeschreibungen, nicht nur unzureichende Wirkung sondern auch sehr schlechte Bearbeitungszeiten für die deutlich zunehmende Menge der Datenpunkte, die zu verarbeiten sind, besitzen.

Die Korrelation beobachteter Dockingzeiten zu den hieraus erstellten Trendlinien im Docking mit weiten Selektionsgrenzen erscheint insbesondere bei der Verwendung aller Atome nicht sehr gut zu sein. Dieses liegt primär an der größtmöglichen Varianz, die diese Kombination bietet. Zudem sind die Trendlinien jeweils aus nur elf Datenpunkten erstellt.

Dennoch können diese Trendlinien durchaus ihren Sinn erfüllen, wenn sie auch für die hohen Ab<sub>2</sub>/Ab<sub>3</sub>-Werte zunehmende Abweichungen aufweisen. Hierzu müssen die Untersuchungsergebnisse der Laufzeitvorhersage mit den Ergebnissen der Untersuchung zur Entwicklung der Ergebnisspezifität und der Ergebnisqualifizierung richtig positiver Ergebnisse kombiniert werden. Grob vereinfacht zeigt sich, dass beim Docking mit dem weiten Grenzwertensemble eine ausreichende Spezifität nur erzielt werden kann, wenn einerseits genügend Referenzen vorgewiesen werden können, andererseits aber auch ein Paar-Docking



nicht länger als etwa 200 Sekunden und ein Triplet-Docking nicht länger als etwa 100 Sekunden dauert. Diese eventuellen Richtwerte sind hier aber nur ganz grobe Abschätzungen und bei der vorgestellten Varianz innerhalb der Testgruppen sind abweichende Ergebnisse in positive und negative Richtungen zu beobachten.

Hiermit werden jedoch wieder zwei Frage angeschnitten, die Ergebnisse, nicht jedoch der Gegenstand dieser Arbeit sind:

- Ist es möglich, aus einer Bewertung der Evaluierungskombinatorik einen Rückschluss zu ziehen, ob das Dockingexperiment überhaupt in der Lage sein wird, eine voraussichtlich ausreichende Ergebnisspezifität zu erzeugen?
- Ist es möglich, für eine Molekülpaarung im Vorhinein abzuschätzen, ob es für diese Art und die gewählten Parameter des Dockings geeignet ist, indem genügend Referenzergebnisse zur Verfügung gestellt werden können?

## **V Abschlussdiskussion, Ausblick und Zusammenfassungen**

### **18 Zusammenfassende Diskussion und Ausblick**

In den vorangegangenen Ergebnisabschnitten wurden alle Methoden, ihre Implementierungen, die verwendeten Parameter, die eigentlichen Untersuchungsgegenstände und die zuletzt erzielten Ergebnisse ausführlich dargestellt und auch diskutiert. In einigen Fällen wurden auch bereits Alternativen zu den verwendeten Umsetzungen sowie weitere, unterschiedlich weit ausgeführte Denkanstöße für Weiterentwicklungen und Verbesserungen geliefert. Um diese detaillierten Diskussionen nicht zu wiederholen, sollen hier die Kernaussagen zusammengefasst und mit weiterführenden Arbeitsansätzen in einen Kontext gerückt werden. Ansonsten soll an dieser Stelle einmalig auf die Ausführungen der speziellen Abschnitte verwiesen werden.

Der gesamte Arbeitsablauf wird unterteilt in eine Aufbereitung der einzelnen Proteinmoleküle und eine Zusammenstellung der Molekülrepräsentationen, die im den abschließenden Dockingexperimenten verwendet werden.. Ein ganz bedeutender Anteil des eigentlichen Dockings ist nach der Selektion geeigneter Substrukturen, die miteinander interagieren können, die Erstellung der notwendigen Transformationsvorschriften und die Aufbereitung aller Observationen, um richtig positive Ergebnisse zu isolieren und in ihrer Bedeutung gegenüber den zufälligen, falsch positiven Ergebnissen herauszustellen.

Das eigentliche Docking sucht zunächst auf den Moleküloberflächen beider Dockingpartner nach Paaren oder Triplets von Atomen, deren kombinierte Bindungseigenschaften die zeitgleiche Ausbildung von zwei bzw. drei hochwertigen intermolekularen Wasserstoffbrückenbindungen bei geeigneter Molekülorientierung möglich erscheinen lassen. Zueinander passende Paare bzw. Triplets werden gesucht und identifiziert anhand transformations-invarianter Eigenschaften der relativ zueinander statisch positionierten und orientierten Bindungscharakteristika ihrer Atome.

Die Festlegung aller selektionierenden Grenzwerte, die an den verschiedensten Stellen im gesamten Arbeitsablauf zur Wirkung kommen, erfolgte stets in einer rückgekoppelten Betrachtung des Verbleibs der bekannten Referenzergebnisse aus den Test-Datensätzen. Zur Zeit existieren drei Grenzwertensembles mit starker, mittlerer und schwacher Selektionskraft. Entsprechend der Abnahme der Selektionsstärke steigt die Anzahl richtig positiver Ergebnisse, die in einem Dockingexperiment gefunden werden können. Im Triplet-Docking werden die gleichen Selektionsgrenzwerte wie im Paar-Docking verwendet. Der erzeugte Selektionsdruck ist jedoch bedeutend größere, da alle drei Kanten eines Atomdreiecks, zeitgleich die Evaluierung erfolgreich bestehen müssen.

Während der Vorauswahl der Test-Datensätze wurden zunächst einfach qualifizierte heterooligomere Protein-Protein-Komplexe von Nichtimmunoglobulinen identifiziert. Hierbei fiel auf, dass eine weitere grundlegende Unterscheidung getroffen werden muss, die hauptsächlich den Entstehungsprozess der Komplexe betrifft und anhand einer optischen Kontrolle der Testkandidaten durchzuführen ist.

Fakultative Proteinkomplexe setzen sich aus zwei Einzelmolekülen zusammen, die unabhängig voneinander in der Proteinbiosynthese translatiert und gefaltet werden. Sie weisen im Regelfall globuläre Form auf und können beide getrennt und stabil in Lösung vorliegen.

Obligatorische Proteinkomplexe zeigen dagegen häufig eine ineinander verschlungene Anordnung. Zumindest einer der Dockingpartner besitzt eine Struktur, die frei in Lösung nicht faltungsfähig und stabil ist. Derartige Komplexe könnten bei einer koordinierten Translation unter Mitwirkung faltungshelfender Chaperone zusammen gelagert werden und würden voraussichtlich Struktur und Funktion verlieren, wenn der Komplex getrennt wird. Oder es handelt sich um Proteine und ihre posttranslational proteolytisch abgetrennten jedoch noch komplexierten Propeptide, deren Abtrennung die Funktion des Einzelmoleküls aktiviert. Bei vergleichbarer Ausdehnung der spezialisierten intermolekularen Kontaktflächen finden sich bei obligatorischen Proteinkomplexen in der Regel deutlich mehr intermolekulare

Wasserstoffbrückenbindungen. Diese weisen zudem auffällig hochwertigere Bindungsqualitäten auf, die sich in besseren geometrischen Eigenschaften ausdrücken. Diese werden von der Molekülrepräsentation beschrieben und in den Dockingexperimenten anhand der Selektionsgrenzwerte bewertet. Die besseren Bindungen stammen in großen Teilen aus intramolekular zusammengesetzten Sekundärstrukturelementen, die in der Faltung der Peptidkette bzw. der beiden Peptidketten ausgebildet werden und hochspezifische Muster von Wasserstoffbrückenbindungen erzeugen.

Die Anwendung der Grenzwertensembles auf die beiden Testgruppen hat gezeigt, dass zwischen den Komplextypen aber auch innerhalb der Gruppen Unterschiede und Varianzen zu entdecken sind.

Die Unterschiede drücken sich primär in der Anzahl kriterienkonformer Referenzergebnisse aus, die in den Komplexen zu finden sind. So zeigt sich eine große Anzahl der fakultativen Komplexe als ungeeignet für die stark selektionierenden Methoden des Triplet-Dockings und/oder für das enge und mittlere Grenzwertenset. Nur drei der 16 obligatorischen Komplexe vermögen im Triplet-Docking mit engen Selektionsgrenzwerten überhaupt kein Referenzergebnis in die Evaluierung einzubringen, wogegen hier nur acht von 58 fakultativen Komplexen überhaupt zumindest ein einziges Referenzergebnis präsentieren können.

Die Varianzen innerhalb der Testgruppen basiert zunächst einmal einfach auf den unterschiedlichen Größen der einzelnen Moleküle. Hieraus resultieren unterschiedliche Mengen repräsentierender Atompaaire und -tripletts und daraus je nach Molekülkombination auch sehr unterschiedliche Zahlen notwendiger Evaluierungsaufrufe.

Werden die grundlegenden Algorithmen betrachtet, die einerseits die nichtredundante Kombinatorik der Paar- und Triplettszusammenstellungen und andererseits die Kombinatorik vollständiger Paar-Paar- und Triplet-Triplett-Evaluierungen beschreiben, so zeigt sich, dass aus gegebenen Punktmengen ungeheuer große Mengen von Bewertungsaufrufen entstehen. Die Zahlen steigen dabei für Tripletts um ein Vielfaches stärker an als für Paare. Diese wird weiter verstärkt durch die Notwendigkeit, alle Anordnungen zweier Paare oder Tripletts zueinander auszutesten, um jede mögliche Atom-Atom-Interaktion zu betrachten. Dementgegen wirkt die größere Selektionskraft im Triplet-Docking.

Insgesamt ist zu verzeichnen, dass durch die Verwendung selektionsschwächerer Grenzwertensembles die Zunahme der Evaluierungsaufrufe die Zunahme der Referenzergebnisse bedeutend übertrifft. Zunächst bedeutet die Verwendung weiter Selektionsgrenzwerte eine Zunahme der Anzahl von Referenzergebnissen, die einer Vielzahl der

Test-Datensätze überhaupt erst eine grundlegende Eignung zur Anwendung der Methoden vermittelt. Da starke lineare Korrelationen der Anzahl von Evaluierungsaufrufen zur Anzahl selektierter und kriterienkonformer Kandidaten aufgezeigt werden können, sinkt aber zugleich auch der relative Anteil richtig positiver Ergebnisse an der Gesamtzahl aller Kandidaten.

Somit steigt die Bedeutung postselektiver Mechanismen mit der Aufgabe, falsch positive Evaluierungsergebnisse anhand ihrer Transformationsvorschriften zu identifizieren und zu eliminieren. Zugleich muss die Gruppe richtig positiver Ergebnisse mit gleichartigen Transformationsvorschriften versehen und in ihrer Einheit als überproportional große Gruppe zusammengehöriger Observationen qualifiziert werden. Um diese elementare Spezifität der Referenzergebnisse zu jedem Zeitpunkt der Untersuchungen zu erhöhen, wurden auch zusätzliche präselektive Bewertungen implementiert, um Atome, welche eventuelle durch intramolekulare Wasserstoffbrückenbindungen bereits in ihrer Bindungsfähigkeit abgesättigt sind, frühzeitig aus der Molekülrepräsentation zu entfernen. Für Atomtripletts wurden zudem Möglichkeiten geschaffen zur zusätzlichen Vorselektion geeigneter Interaktionscharakteristika und zur Bestimmung einer Bindungsseite, was die Anzahl notwendiger Evaluierungsaufrufe um die Hälfte senken kann, wenn sie eindeutig ist.

Der unzuverlässigste und verlustreichste Abschnitt im gesamten Arbeitsablauf ist eindeutig die Methode der Rotationsbestimmung im Anschluss an die Identifikation eines kriterienkonformen Evaluierungskandidaten. Zwischen 40% und 70% der richtig positiven Ergebnisse bekommen hier eine falsche Orientierungsvorschrift und sind somit für die nachfolgende Ergebnisaufbereitung verloren. Generell erfolgt die Rotationsbestimmung:

- für obligatorische Komplexe zuverlässiger als für fakultative Komplexe, da optimalere intermolekulare Wasserstoffbrückenbindungen zugrunde liegen,
- für das Triplet-Docking zuverlässiger als für das Paar-Docking, da redundante Bestimmungen einzelne Fehler eliminieren können,
- für selektionsstarke Grenzwertensembles zuverlässiger als für das weite Grenzwertensemble, da optimalere Bindungscharakteristika bevorzugt werden.

Für die Eignung der Moleküle bzw. der Komplexe ergeben sich somit entsprechend höhere Grundanforderungen, die von den fakultativen Komplexen eigentlich gerade einmal 75% der Gruppe beim Paar-Docking und etwa 40% beim Triplet-Docking mit weiten Selektionsgrenzen leisten können.

Um die unzureichende Qualität der Rotationsbestimmung zunächst ein wenig auszugleichen und überhaupt mit richtig positiven Ergebnissen weiter arbeiten zu können, wird vorerst mit einer Unterteilung des Rotationsraums in 564 Winkelräume mit äquidistant in 30°-Abständen

zueinander positionierten Rotationsrepräsentanten gearbeitet. Diese Auflösung ist generell noch zu ungenau als dass sie zur Erstellung qualitativ überzeugender Ergebnisse ausreichen würde.

Die Dauer der Molekülaufbereitungen ist wenig zeitkritisch aber auch unbedeutend kurz. Die Laufzeiten der Dockingexperimente sind ganz eindeutig von mehreren Faktoren abhängig. Hierzu zählen besonders die Größen der Molekülrepräsentationen der Dockingpartner, die einerseits von den Molekülgrößen selbst aber auch von den verwendeten Grenzwertensembles abhängig sind. Erhebliche Gewinne können erzielt werden, wenn frühzeitig ungeeignete Atome vom Docking ausgeschlossen werden, was aber bei einer kritischen Anzahl von Referenzergbnisse große Risiken birgt, da diese Entscheidungen nicht völlig fehlerfrei getroffen werden können.

Während alle Dockingexperimente mit dem engen oder mittleren Grenzwertensemble in vertretbaren Laufzeiten vervollständigt werden können, verursacht die Verwendung der weiten Selektionsgrenzwerte Rechenzeiten, die der Aufgabenstellung einer schnellen und präselektiven Untersuchung nicht gerecht werden können. Grundsätzlich ist der Evaluierungsaufwand, der aus der Kombinatorik der Einzelmoleküle resultiert, hierfür die Ursache. Es konnte gezeigt werden, dass die Methodenwahl und die Implementierung der Bildung von *Single-Linkage*-Clustern aus allen Kandidaten gleichartiger Rotation anhand ihrer Translationsvorschriften für die anfallenden Observationszahlen ihr ungünstiges Laufzeitverhalten wirksam präsentiert.

Betrachtet man die, mit den vorliegenden Methoden und Implementierungen erzielbaren, Endergebnisse der vollständigen Dockingexperimente unter Verwendung der 30°-Rotationsrepräsentanten, so lassen sich zwei grundlegende Extrema eines funktionierenden und eines nichtfunktionierenden Dockings formulieren, die auch genau so in den eigenen Dockingergebnissen anzutreffen sind:

- Die Komplexe können für das Docking mit den engen und mittleren Grenzwertensembles ausreichend Referenzergbnisse bereitstellen und auch mit den richtigen Rotationsvorschriften qualifizieren.
  - Der Referenzanteil an allen Endergebnissen mit richtiger Rotationsbeschreibung beträgt für fakultative und obligatorische Komplexe 62,4% bis 87,3%.
  - Die Qualifizierung dieser richtigen Ergebnisse gelingt im Regelfall sehr zuverlässig mit einer Sortierungspositionierung innerhalb der ersten zehn aller 564 Ränge.
  - Die Qualität der Translationsbeschreibung ist sehr hoch und entspricht zum Teil nur einem einfachen methodischen Fehler.

- Die durchschnittlichen Rechenzeiten sind sehr gering und liegen im Paar-Docking je nach Atomvorselektion und Grenzwertensemble zwischen 11,2 Sekunden und 71,2 Sekunden sowie im Triplett-Docking sogar nur zwischen 0,8 Sekunden und 4,5 Sekunden.
- Mit den wenigsten Ausnahmen können fast alle Komplexe ausreichend Referenzen für das Docking mit den weiten Selektionsgrenzwerten bereit stellen. Hierbei steigt jedoch die Anzahl falsch positiver Ergebnisse überproportional zur Referenzanzahl.
  - Der Referenzanteil kann für obligatorische Proteinkomplexe trotz aller prä- und postselektiven Anstrengungen nur auf maximal 20% angehoben werden.
  - Die Qualifizierung der richtig positiven Ergebnisse ist unzuverlässig. Sie sind nur selten in besonders umfangreichen finalen Transformationsclustern mit sehr geringer Gewichtung enthalten. Ansonsten qualifizieren sich unspezifische Ergebniscluster, deren Belegungszahlen zufallsbestimmt größer sind.
  - Die Qualität der Translationsbeschreibung ist abhängig vom Referenzanteil und erfolgt somit zunehmend unzutreffend und in vielen Fällen völlig willkürlich ohne den Beitrag eines richtig positiven Ergebnisses.
  - Die größten Dockingexperimente können nicht zu Ende geführt werden, da sie die Anforderungen einer derartigen Untersuchung an Rechenzeit und/oder Arbeitsspeicher überschreiten. Zugleich konnte auch gezeigt werden, dass derart große Problemstellungen mangels Ergebnisspezifität mit den vorliegenden Methoden niemals qualitativ ausreichende Ergebnisse hervorbringen können.
  - Die durchschnittlichen Rechenzeiten der durchgeführten, d.h. der elf kleinsten und schnellsten Experimente, liegen je nach Methode und Atomvorselektion zwischen 12 Minuten und 1,5 Stunden bei einem Spitzenwert von 7,7 Stunden.

Die Varianzen innerhalb aller Test-Datensätze ermöglichen fließende Übergänge zwischen den beschriebenen Extrema. Ebenso ist der Anstieg der Selektionsstärke der drei Grenzwertensembles nicht in irgendeiner Weise linear, so dass weitere Abstufungen eventuell bessere Ergebnisse bezüglich der gesamten Testgruppen ermöglichen.

Mit der zunehmenden Selektionsstärke der Selektionsgrenzwerte und mit allen prä- und postselektiven Mechanismen, welche die Ergebnisspezifität steigern sollen, steigt das Risiko, die Anzahl richtig positiver Ergebnisse zu sehr zu beschränken oder die richtige Transformationsvorschrift fehlerhaft zu verlieren.

Eine letzte Tabelle IV-34 soll nochmals zusammenfassen, wie viele Dockingexperimente für welche Methoden- und Parameterkombinationen erfolgreich und vollständig abgeschlossen

werden konnten und wie viele richtig positionierte Kandidaten hierfür in etwa im finalen Transformationscluster vorliegen müssen. Als Erfolgskriterium wird eine Sortierungsplatzierung innerhalb der ersten 100 Ränge einer absteigenden Sortierung nach den Belegungszahlen finaler Transformationscluster angewendet.

		erfolgreiche Qualifikationen	erforderliche Ergebniszahl im finalen Transformationscluster
<b>Obligatorische Proteinkomplexe</b>			
Paar-D	- enges/mittleres Gwe.	11 bis 13	von 16 $\geq 4$ bis $\geq 5$
Triplett-D	- enges/mittleres Gwe.	8 bis 10	von 16 $\geq 2$ bis $\geq 4$
Paar-D	- weites Grenzwertens.	2, 5, 10	von 11,13,16 nicht bestimmbar
Triplett-D	- weites Grenzwertens.	2, 5, 11	von 11,13,16 nicht bestimmbar
<b>Fakultative Proteinkomplexe</b>			
Paar-D	- enges/mittleres Gwe.	6 bis 8	von 58 $\geq 5$ bis $\geq 9$
Triplett-D	- enges/mittleres Gwe.	3 bis 6	von 58 $\geq 3$ bis $\geq 4$
Paar-D	- weites Grenzwertens.	3, 3, 5	von 39,50,54 nicht bestimmbar
Triplett-D	- weites Grenzwertens.	1, 3, 3	von 43,54,55 nicht bestimmbar

Tabelle IV-34: Übersicht erfolgreicher Dockingexperimente und der hierfür notwendigen Belegungszahlen im finalen Transformationscluster mit der richtigen Rotationsbeschreibung. Die Variationen innerhalb der Methoden-Parameter-Kombinationen stammen aus den optionalen Atomvorselektionen.

Die Tabelle IV-34 folgt im Wesentlichen der bereits detailliert dargelegten Argumentation, dass die Qualität der betrachteten, bewerteten und auch zur Transformationsbeschreibung verwendeten intermolekularen Wasserstoffbrückenbindungen elementar bedeutend ist.

Obligatorische Komplexe besitzen derartig hochwertige Bindungen in großem Maße. Das Triplett-Docking, die Verwendung selektionsstarker Grenzwerte und die präselektiven Atomdisqualifikationen aufgrund der eventuellen Existenz intramolekularer Bindungspartner verstärken einzeln und in Kombination die Anforderung, die an zwei bzw. drei gleichartige Interaktionspunkte in einer spezialisierten Kontaktfläche gestellt werden. Hoher Selektionsdruck erzeugt zunächst weniger falsch positive Evaluierungsergebnisse und kann sie auch im Nachhinein besser eliminieren, wodurch die Spezifität der Referenzergebnis ausreichend gesteigert werden kann, um zuverlässige Qualifizierungen und qualitativ hochwertige Transformationsvorschriften zu erzeugen. Hoher Selektionsdruck erfordert aber auch Proteinkomplexe mit besonders großer Anzahl von Interaktionen. Auch ist die Vorselektion bindungsfähiger Atome nicht fehlerfrei und so steigt das Risiko die wichtigen Referenzen schon vor der Selektion zu verlieren. Nach der Selektion besteht dieselbe Gefahr, wenn zu wenige

richtig positive Ergebnisse gleichartig beschrieben werden, um eine erfolgreiche Passage der postselektiven Bewertungen zu bestehen.

Zu geringer Selektionsdruck erzeugt genügend Referenzergebnisse. Diese können aber in der viel stärker angewachsenen Menge falsch positiver Ergebnisse nicht mehr identifiziert und qualifiziert werden. Zudem werden Programmlaufzeiten generiert, die für eine derartige Voruntersuchung nicht vertretbar sind.

Es ist ersichtlich, dass die Methoden und Implementierungen an sich funktionieren, wenn auch einzelne elementare Mängel bestehen, die jedoch sehr gut zu definieren sind. Es ist genauso ersichtlich, dass das Untersuchungsobjekt für die Anwendung der Methoden und Parameter geeignet sein muss.

Diese Eignung ergibt sich für den vorliegenden Fall primär aus den Eigenschaften des Proteinkomplexes bzw. der Einzelmoleküle. Es müssen genügend Referenzergebnisse zur Verfügung gestellt werden können und gleichzeitig darf die Ergebnisspezifität nicht durch zu viele falsch positive Evaluierungsergebnisse so sehr herabgesetzt werden, dass eine Qualifizierung und ein verhältnismäßiges Laufzeitverhalten nicht mehr gewährleistet werden können. Die Unterschiede und Merkmale müssen genauer aus differenzierten Vergleichen obligatorischer und fakultativer Protein-Protein-Komplexe sowie aus den Beobachtungen funktionierender und nichtfunktionierender Dockingexperimente abgeleitet werden. Es konnte bereits gezeigt werden, dass aus den sogenannten Ab2- und Ab3-Werten Rückschlüsse auf eine generelle Eignung bereits vor dem Start des eigentlichen Dockingexperiments gezogen werden können.

Die sich ergebende Notwendigkeit einer derartigen präselektiven Einschätzung der Eignung noch vor dieser Untersuchung, die selbst einen Vorfilter einer aufwendigeren Hauptuntersuchung darstellt, und auch die sich hieraus ergebenden Gefahren wurden bereits erläutert.

Einige weitere Ausblicke sollen Möglichkeiten zur Weiterentwicklung der vorgestellten Arbeit kurz darstellen und auch in einen gewissen Prioritätenrahmen stellen.

Das bedeutendste Problem, das gelöst werden muss, ist die unzureichende Methode der Rotationsbestimmung. Eine erfolgreiche Evaluierung transformations-invarianter Eigenschaften identifiziert mit den beiden beteiligten Atompaaaren oder -tripletts die Berührungspunkte, an denen die Moleküle miteinander in Kontakt treten könnten. Die Erstellung der genauen



Rotations- und Translationsvorschrift, um diesen Kontakt mit komplementär verlaufenden Oberflächenprofilen und ohne van-der-Waals-Überlappung herbeizuführen, ist neu zu implementieren. Mit einer gesicherten Bestimmung würde die Anzahl richtig positiver Ergebnisse mit korrekter Rotationsvorschrift in etwa verdoppelt, was in den meisten Fällen mit einer Mindestzahl von 4 bis 9 Referenzen für ein erfolgreiches Docking ausreichen würde. Ebenfalls könnten dann die regelmäßigen Rotationsrepräsentanten deutlich enger gesetzt werden. Die gleichbleibende Anzahl kriterienkonformer Kandidaten verteilt sich so auf ein Vielfaches von Repräsentanten, so dass die Grundanforderungen für die Eignung der Proteinmoleküle weiter sinkt und die richtige Orientierung eventuell sogar ohne Ergebnisausdünnung und/oder finale Clusterbildung ausreichend überproportional gut belegt sind und in ihrer Bedeutung identifiziert werden können.

Es wäre wünschenswert, von der Verwendung diskreter Rotationsrepräsentanten Abstand zu gewinnen. Befindet sich die Gruppe richtig positiver Ergebnisse genau auf der Zuordnungsgrenze benachbarter Repräsentanten, kann es zu einer zwanghaften Teilung der Gruppe kommen, welche die richtige Ergebnisfindung vereitelt. Die Erstellung natürlich zusammengehöriger Cluster, ähnlich der *Single-Linkage-Cluster* der Translationsbeschreibungen innerhalb der zuvor eingeteilten Rotationsgruppen, ist anzustreben. Hierbei sind Methoden zu wählen, die mit den enormen Mengen aller kriterienkonformen Kandidaten umgehen können, wenn nicht eine überlappende Voreinteilung stattfinden soll.

Alternative Qualifizierungsmethoden wurden bereits vorgestellt und diskutiert. Mit geeigneten Qualifizierungsgrenzwerten kann es schnell gelingen, eine generelle Halbierung aller möglichen Endergebnisse ohne Verlust des richtigen Ergebnisses herbeizuführen. Noch höhere und vor allem auch noch zuverlässigere Selektionswirkungen können mit einer besseren Rotationsbestimmung erzielt werden. Große Bedeutung ist der Möglichkeit dynamischer Qualifizierungsgrenzwerte beizumessen, die Abhängigkeiten zu den kombinierten Moleküleigenschaften der Dockingpartner herstellen, um auf die großen Varianzen innerhalb der Komplextypen reagieren zu können.

Die Selektionsmöglichkeiten in der Evaluierung mono- und bimolekularer transformationsinvarianter Eigenschaften sind noch nicht vollständig ausgereizt. So werden noch keine direkten Korrelationen und dynamische Evaluierungsgrenzwerte für geometrische Größen bestimmt, die strukturbedingte Abhängigkeiten aufweisen. In einigen Fällen wird durch übergreifende

Selektionsgrößen sichergestellt, dass eine nichtoptimale Interaktion nur gekoppelt mit einer hochwertigen Interaktion ein kriterienkonformes Ergebnis bilden können.

Die grundlegende Methodik der transformations-invarianten Parametrisierung und auch viele Abschnitte der vorgelegten Implementierung können für Adaptionen verwendet werden. Vor allem die Verwendung elektrostatischer Wechselwirkungen bietet sich an. Diesen fehlen jedoch bereits in großem Maße der gerichtete Vektorcharakter der Wasserstoffbrückenbindungen.

Unter Berücksichtigung der Herkunft obligatorischer Protein-Protein-Komplexe und der sehr guten Performance, die insbesondere für die regelmäßigen Muster von Wasserstoffbrückenbindungen innerhalb der Sekundärstrukturelemente sogar mit den bestehenden Mängeln aufgezeigt wurde, ist vielleicht eine Anwendung der Methode in der Faltungserkennung bzw. -simulation von Polypeptiden zu erwägen. Stabile Proteinfaltungen beruhen nicht nur auf Zusammenlagerungen hydrophober Bereiche, die sich so von polarem Lösungsmittel abschirmen, sondern auch auf der Ausbildung vieler stabiler und optimierter intramolekularen Bindungen. Eine schnelle Bewertung und alternativer Konformationen durch die Quantifizierung und Qualifizierung kooperativer Bindungen ist hier notwendig.

Zwei weitere Überlegungen haben diese Arbeit von Anfang an begleitet und wurden auch bereits erwähnt: Brückenbildende Wasseratome, die als Bindeglied zwischen zwei bindungsfähigen Atomen eingespannt sind und eine molekulare Dynamik bzw. Flexibilität der Aminosäureresiduen an der Oberfläche der Proteinmoleküle.

Brückenbildende Wassermoleküle haben sich bei Untersuchungen obligatorischer Proteinkomplexe mit einer Häufigkeit von etwa eine Observation pro 3,8 interkompatiblen Wasserstoffbrückenbindungen gefunden [Janin *et al.*, 1988 und 1990] [Xu *et al.*, 1997]. Dieser Wert weist erwartungsgemäß eine große Varianz auf. Ein Wassermolekül kann im Extremfall entsprechend seinem tetraedrischen Aufbau simultan bis zu vier Koordinationen eingehen. Dieses entspricht dem Grundaufbau der Eisstruktur mit je zwei Wasserstoffbrückenbindungen in Donator- und Akzeptorfunktion. Cytosolische Proteine lagern Wasser so an, dass eiskristallähnliche Gitter entstehen [Will, 1993]. Somit können Wassermoleküle kompatible Bindungsstellen verknüpfen, die zu weit auseinander liegen, um in direkten Kontakt zu treten, und auch entsprechend voneinander entfernte inkompatible Bindungsstellen. Teilweise werden

ganze Netzwerke von Wassermolekülen aufgespannt, die an ihren Rändern von Proteinfunktionen fixiert sind.

Diese Erscheinung darf also an sich nicht vernachlässigt werden, wenn es um intermolekulare Wasserstoffbrückenbindungen geht. Angesichts der bereits bestehenden Probleme qualitativer und quantitativer, kombinatorischer Natur werden jedoch keinerlei Möglichkeiten gesehen, hier ansatzweise Erfolge erzielen zu können. Die Einbeziehung brückenbildender Wassermoleküle führt fast jeder Bewertungsmöglichkeit neue Freiheitsgrade hinzu. Die bisherige Vorgabe, dass Akzeptoren nur mit Donatoren Bindungen eingehen können wird aufgehoben, wenn nur der direkte Atom-Atom-Kontakt nicht notwendig ist. Die sterischen Freiheiten, der recht gut definierten Abstands- und Winkelvorgaben werden drastisch erweitert, da Wassermoleküle keine kovalenten Anbindungen besitzen und außer in Koordination frei translatiert und rotiert werden können. Jegliche geometrischen Kriterien können in allen Bindungen vollständig ausgereizt werden, wodurch auch indirekt kaum mehr ein Rückschluss gezogen geschweige denn eine effektive Selektionsmöglichkeit für die geometrische Anordnung der Bindungsstellen zueinander eingesetzt werden können.

Es ist als sicher anzunehmen, dass sich die Oberflächen von ungebundenen und in Komplexen gebundenen Molekülen in den Kontaktflächenbereichen in unterschiedlichem Maße unterscheiden. Eine aktuelle Publikation weist darauf hin, dass es eventuell "nur" 20% der Aminosäureresiduen in den Kontaktflächen sind, deren Seitenketten im Rahmen einer bindungsinduzierten Anpassung (*induced fit*) Lageveränderungen erfahren. Hinzu kommen Effekte der Strukturaufklärung, dieses sind insbesondere Effekte in der "unnatürlichen" Kristallpackung, und Unsicherheiten bzw. Ungenauigkeiten in der Strukturinterpretation. Aus diesen Gründen werden grundsätzlich das Starre Docking (*rigid body docking*), bei dem die Einzelmoleküle verwendet werden, wie sie in ihren Einzel- oder Komplexstrukturen vorliegen, und das Flexible Docking (*flexible docking*) unterschieden, bei dem den Oberflächen gewisse Bewegungsmöglichkeiten eingeräumt werden. Eine weitere Unterscheidung wird gemacht mit dem Docking zweier Moleküle, deren Strukturen aus einer gemeinsamen Komplexstrukturlösung extrahiert wurden und die wie in dieser Arbeit wieder zusammengefügt werden sollen (*bound case*) und dem Docking zweier Moleküle, deren Strukturen unabhängig voneinander gelöst wurden (*unbound case*). Im Fall eines Flexiblen Dockings könnten Konformationsänderungen der Aminosäureseitenketten mit Rotameren-Bibliotheken diskret simuliert oder auch mit einer Dynamiksimulation und/oder Energieminimierung mit einem neuen Endzustand versehen werden, der dann jedoch auch nicht zugleich dem gebundenen und dem ungebundenen Zustand entsprechen kann.

Auch hinsichtlich dieser Problematik wird die vorliegende Methode zum derzeitigen Entwicklungsstand ungeeignet sein, entsprechende Adaptionen sinnvoll einbinden zu können. Versuche, eine Unschärfe in die Lokalisation einzelner Atome oder Residuen einzubringen oder ihre eventuell möglichen Aufenthaltsorte umfassen zu beschreiben, führen zum direkten Verlust aller konkreten Orts- und Richtungscharakteristika, die für die transformations-invariante Parametrisierung unerlässlich sind. Die Verwendung diskreter Alternativkonformationen können angesichts des bereits bestehenden Kombinatorikproblems nicht ernsthaft in Betracht genommen werden, da bereits eine einzelne Atomverlagerung nicht nur andersartige sondern auch gegebenenfalls völlig neuartige Atompaare und -tripletts entstehen lässt, welche in einer nichtoptimierten Suche prinzipiell gegen alle Atomensembles des Dockingpartners ausgetestet werden müssen. Die zwangsläufige Entwicklung, wenn mehr als ein Atom verlagert wird, soll hier ohne weitere Ausführung nur ins Bewusstsein gerufen sein. Unabhängig von dem kombinatorischen Problem wird sich auch hier die Methode der Rotationsbestimmung als ungeeignet erweisen, da auch sie völlig auf die Richtung einzelner Atombindungen angewiesen ist obwohl sich die generelle Anordnung der Dockingpartner nicht ändern wird.

Die einzigen Möglichkeiten neben der oben beschriebenen Erstellung einer einzigen alternativen Oberflächenbeschaffenheit, die einen energiearmen Zustand beschreibt, könnte eventuell sein, beste Geometriezustände herzustellen. Hierbei nehmen die Atome für die Paar-Paar- bzw. Triplet-Triplett-Evaluierungen jeweils die optimalsten Positionen zueinander ein, welche mit den zur Verfügung stehenden Bewegungsfreiräumen der beteiligten Aminosäuren realisiert werden können. Die verwendeten Kriteriengrenzwerte sollten entsprechend stärker selektierend ausfallen können als sie das bisherige enge Grenzwertensemble zusammenstellt. Ob und wie das Problem noch transformations-invariant zu lösen ist, bleibt kommenden Studien zu ergründen.

## 19 Kurzzusammenfassung

Diese Arbeit beschreibt und diskutiert ein Methode, die räumliche Anordnung zweier spezifisch interagierender Proteinmoleküle *in silico* aus den PDB-Datensätzen der Einzelmoleküle zu bestimmen, indem ein physiko-chemischer Vorfilter lokale, kooperative Bindungen zwischen den Molekülen sucht und als Ergebnis eine sortierte Liste von Transformationsvorschriften ausgibt.

Als Suchkriterium potentiell produktiver Anordnungen werden Paare bzw. Triplets bindungsfähiger, oberflächennaher Atome verwendet, die zeitgleich zwei bzw. drei intermolekulare Wasserstoffbrückenbindungen ausbilden können. Hierzu werden die Protonen-Akzeptoren und -Donatoren in eine Vektorrepräsentation überführt und die Atomensembles durch eine auf Abständen und Winkeln basierende transformations-invariante Parametrisierung beschrieben und bewertet. Zur Anwendung kommen vereinfachte Methoden des Geometric Hashing und verschiedene Ensembles von Evaluierungsgrenzwerten, die aus einer Analyse der verwendeten geometrischen Eigenschaften bei nativen Molekülkomplexe bekannter Struktur zusammengestellt wurden. Für funktionell und sterisch akzeptable Gruppen von Interaktionsstellen werden die notwendigen Transformationsvorschriften hergeleitet. Eine höhere Anzahl gleichartiger Vorschriften, die in einer Folge von clusterbildenden Arbeitsschritten identifiziert werden, soll zuletzt die einzig richtige Molekülkonformation beschreiben und als positives Ergebnis qualifizieren. Zwei Testgruppen von verschiedener Molekülkomplextypen wurden erstellt. Sie zeigen große strukturelle Unterschiede, indem einerseits zwei globuläre Moleküle aufeinander treffen und andererseits zwei Moleküle sehr verschlungene Anordnungen vorweisen und dabei extensive, intermolekular zusammengesetzte Sekundärstrukturelemente mit hochspezifischen Mustern hochwertiger Wasserstoffbrückenbindungen besitzen.

Es können unterschiedliche Eignungen für die verschiedenen Methoden und Evaluierungsgrenzwerte festgestellt werden: Nur optimale Komplexe besitzen derart gute Bindungen, dass auch mit selektionsstarken Grenzwerten genügend kriterienkonforme Paar- und Triplettensembles mit richtiger Transformationsvorschrift identifizieren können, um vorderste Positionen in der sortierten Liste von Endergebnissen einzunehmen. Die benötigten Rechenzeiten unter Verwendung der selektionsstarken Grenzwertensembles liegen pro Molekülpaarung durchschnittlich zwischen 11,2 und 71,2 Sekunden im Docking von Bindungspaaren und zwischen 0,8 und 4,5 Sekunden im Docking von Bindungstriplets. Die Verwendung selektionsschwacher Evaluierungsgrenzwerten erhöht meist derart die Zahl falsch positiver Kandidaten, dass es nicht mehr gelingt, die richtige Transformation in ihrer Bedeutung zu kennzeichnen und als potentiell korrektes Ergebnis zu qualifizieren. Eng gekoppelt sind Laufzeiten der Dockingexperimente, die den Ansprüchen einer derartigen Voruntersuchung nicht genügen.

## 20 Abstract

### Studies, Characterisation and Efficient Handling of Surface-Parameters for the Simulation of Molecular Protein-Protein-Interactions

This thesis describes and discusses a method to determine *in silico* the orientation of two specifically interacting protein molecules from the PDB-datasets of the single molecules. A physico-chemical prefilter looking for local cooperative interactions has been developed and provides a ranked list of molecule transformations.

As the search criterion for probably productive orientations we use pairs and triplets of atoms which are capable of the simultaneous formation of two or three intermolecular interactions. For this, hydrogen acceptors and donors are handled by vectorized representations and the groups of atoms can be described and evaluated using a transformation invariant parameterization based upon distances and angles. Geometric hashing techniques and different sets of qualification thresholds derived from analysis of the used geometric features in native complexes are applied. For functionally congruent and sterically acceptable ensembles of binding sites, the necessary transformations are computed. A significantly higher number of similar observations identified by clustering steps should describe the only true positive molecule conformation and mark its importance with a high-ranking position.

Two subsets of protein-protein complexes have been created. They have important structural differences, as on the one hand two globular molecules interact while on the other hand two molecules show a very entangled arrangement. This can hardly come about from separated monomers and exhibits an extensive amount of intermolecular composed secondary structure elements with highly specific patterns of hydrogen bonds.

Depending on the subsets dissimilar qualifications can be determined for the different methods and thresholds: When using the highly restrictive evaluation thresholds, only the more optimal complexes offer enough suitable interactions to provide enough criteria conform pair- and triplet-ensembles to identify the true positive transformation with high accuracy. Using these thresholds and different options of preselecting binding sites, the running time of a single molecule docking averages between 11.2 and 71.2 seconds looking for pairs of hydrogen bonds and between 0.8 and 4.5 seconds looking for triplets.

Usage of the less restrictive ensemble of evaluation thresholds provides sufficient true positive results in most although not all experiments. At the same time the number of false positive candidates often rises over a critical ratio where it becomes impossible to identify the only true orientation in its significance and securely qualify it as a potentially correct result. Insufficient specificity of the results is closely connected to running times of experiments that cannot suffice the needs of the intended prefiltering analysis.

# Anhang

## A1 Abkürzungen, Symbole und Begriffe

Neben einzelnen Abkürzungen stehen hier einige Begriffe im Mittelpunkt, die zum Teil mit ganz speziellen Bedeutungen in dieser Arbeit genutzt werden. In diesen Fällen ist neben einer sehr kurzen Beschreibung vor allem in eckigen Klammern ein Verweis auf das Kapitel mit der Einführung, Herleitung und Bedeutungsklärung des Begriffes innerhalb dieser schriftlichen Ausführungen angeführt.

$\alpha, \beta, \gamma, \theta, \tau, \phi$	Angabe von Winkeln in Winkelgrad; für die transformations-invariante Eigenschaften [7.2]
A, Akz., Akzeptor	Protonenakzeptor [3]
Å	Ångström, $10^{-10}$ Meter
AA	<i>antecedent atom</i> , kovalent an einen Akzeptor gebundenes Atom [3]
alle, ungesättigte, ungebundene Atome	beziehen sich auf eine Vorselektion bindungsfähiger Atome [7.1.3]
asa	<i>accessible surface area</i> , lösungsmittelzugängliche Moleküloberfläche [12.1]
bimolekulare Eigenschaften	Eigenschaften, die sich aus der Kombination monomolekularer Eigenschaften ergeben [8.1]
bindungsfähiges Atom	Protonenakzeptor oder -donator [3]
D, Don., Donator	Protonendonator [3]
Grundlinie, Grundlinienvektor	Verbindungsline zweier Interaktionspunkte [7.2]
h	Stunde
H	Wasserstoffatom bzw. Proton [3]
Interaktionspunkt, -vektor	beschreiben die Charakteristik eines bindungsfähigen Atoms [7.1.2]
min	Minute
monomolekulare Eigenschaften	Eigenschaften, die von einem einzelnen Molekül abhängig sind [7.2]
msa	<i>molecular surface area</i> , molekulare van-der-Waals-Oberfläche [12.1]
N	Stickstoffatom [3]
O	Sauerstoffatom [3]
PDB	<i>Protein Data Base</i> , Protein Datenbank [6.3]
RMS	<i>root mean square</i> , mittlere quadratische Fehler
sek	Sekunde
Triplett-Bindungsseite	Eigenschaft eines Atomtriplets [7.3]
Triplett-Vektordivergenz	Eigenschaft der Interaktionsvektoren eines Atomtriplets [7.3]
ungebundene Atome	siehe "alle Atome" [7.1.3]
ungesättigte Atome	siehe "alle Atome" [7.1.3]
Wasserstoffbrückenbindungs-Stelle	bindungsfähiges Atom [3]

## A2 Wasserstoffbrückenbindungsatome, van-der-Waals Radien

Atome haben entweder keine Wasserstoffbrückenbindungseigenschaften, können eindeutig nur Akzeptor-(A)- oder nur Donator-(D)-Eigenschaften besitzen oder in Abhängigkeit ihrer Umgebung als Akzeptor oder Donator (B, beides) in Erscheinung treten.

Für Atome, die nachfolgend nicht explizit aufgeführt werden, gilt:

Stickstoffatome der Peptidbindungen (N): 1,70 Å; Donatoren  
 Sauerstoffatome der Peptidbindungen (O): 1,40 Å; Akzeptoren  
 N-terminale Stickstoffatome (N): 1,70 Å; Donatoren  
 C-terminale Sauerstoffatome (OXT): 1,40 Å; Akzeptoren

α-Kohlenstoffatome der Peptidbindungen (CA): 2,00 Å

Kohlenstoffe der Peptidbindungs-Carboxylgruppen (C): 1,70 Å

weitere Kohlenstoffatome (div.): 1,70 Å

Kohlenstoffatome besitzen keine Wasserstoffbrückenbindungseigenschaften

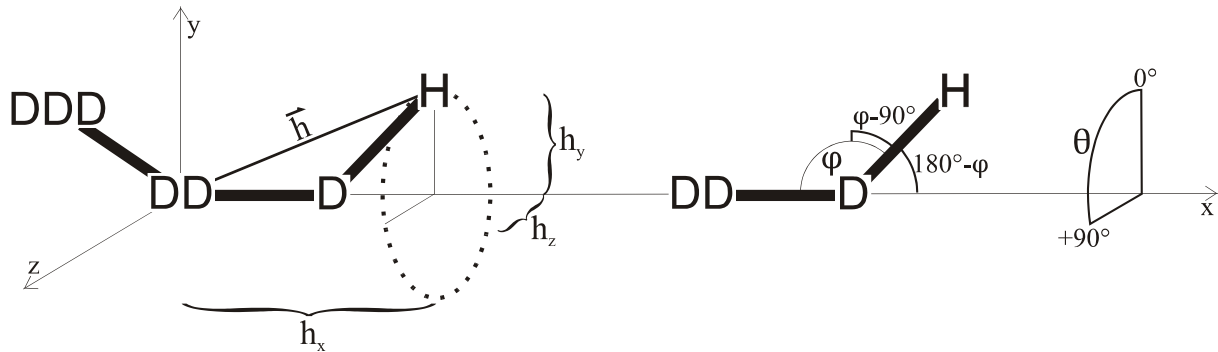
ARGININE (ARG)	OE2 GLU 1.50 A
NE ARG 1.70 D	
NH1 ARG 1.70 D	GLUTAMINE (GLN)
NH2 ARG 1.70 D	OE1 GLN 1.40 A
	NE2 GLN 1.70 B
ASPARAGINE (ASN)	
OD1 ASN 1.40 B	GLUTAMIC ACID / GLUTAMINE
AD1 ASN 1.40 B	AMBIGUOUS (GLX)
ND2 ASN 1.70 B	AE1 GLX 1.40 B
AD2 ASN 1.70 B	(AVERAGE OF OE2 AND NE2 OF GLU AND GLN)
ASPARTIC ACID (ASP)	AE2 GLX 1.55 B
OD1 ASP 1.50 A	
OD2 ASP 1.50 A	HISTIDINE (HIS)
	ND1 HIS 1.60 B
ASPARTIC ACID / ASPARAGINE	AD1 HIS 1.60 B
AMBIGUOUS (ASX)	CD2 HIS 2.00 B
AD1 ASX 1.40 B	AD2 HIS 2.00 B
(AVERAGE OF OD2 AND ND2 OF ASP AND ASN)	CE1 HIS 2.00 B
AD2 ASX 1.55 B	AE1 HIS 2.00 B
	NE2 HIS 1.60 B
	AE2 HIS 1.60 B
CYSTEINE / CYSTIN (CYS)	
SG CYS 1.80 -	LYSINE (LYS)
	NZ LYS 2.00 D
GLUTAMIC ACID (GLU)	
OE1 GLU 1.50 A	METHIONINE (MET)



SD MET 1.80 -	CZ UNK 1.85 D
	(AVERAGE OF PHE/ARG CZ)
SERINE (SER)	CZ2 UNK 1.85 D
OG SER 1.60 B	(AVERAGE OF PHE/ARG CZ)
THREONINE (THR)	CZ3 UNK 1.85 D
OG1 THR 1.60 B	(AVERAGE OF PHE/ARG CZ)
TRYPTOPHAN (TRP)	N UNK 1.70 B
NE1 TRP 1.70 D	ND1 UNK 1.60 B
	ND2 UNK 1.70 B
	NE UNK 1.70 B
	NE1 UNK 1.70 B
	NE2 UNK 1.70 B
	NH1 UNK 1.70 B
	NH2 UNK 1.70 B
	NZ UNK 2.00 B
	O UNK 1.40 B
	OD1 UNK 1.50 B
	OD2 UNK 1.50 B
	OE1 UNK 1.50 B
	OE2 UNK 1.50 B
	OG UNK 1.60 B
	OG1 UNK 1.50 B
	(AVERAGE OF -OH AND =O)
	OH UNK 1.60 B
	OXT UNK 1.40 A
	SD UNK 1.80 -
	SG UNK 1.80 -
	UNK UNK 1.70 -
	(AVERAGE DEFAULT VALUE)
UNKNOWN (UNK)	
C UNK 1.70 -	
CA UNK 2.00 -	
CB UNK 2.00 -	
CD UNK 2.00 -	
CD1 UNK 2.00 -	
CD2 UNK 2.00 -	
CE UNK 2.00 -	
CE1 UNK 2.00 -	
CE2 UNK 2.00 -	
CE3 UNK 2.00 -	
CG UNK 2.00 -	
CG1 UNK 2.00 -	
CG2 UNK 2.00 -	
CH2 UNK 2.00 -	

### A3 Übersicht über die Positionsbestimmung der Protonen

Für vier der fünf verschiedenen Grundtypen der Positionsbestimmung von Protonen kann diese Methode mit wechselnden Werten für Winkel und Bindungslängen genutzt werden. Angaben zu den einzelnen Grundtypen finden sich auf der nächsten Seite.



#### globale Atompositionen

D Donoratom

DD kovalent vorhergehendes Atom eines Donoratoms

DDD kovalent vorhergehendes Atom eines DD (Lage in X-Z-Ebene, hinter X-Y-Ebene)

H Proton

#### lokales Koordinatensystem (Ursprung in DD)

$$\vec{x}_0 = \overrightarrow{DD > D}_0$$

$$\vec{y}_0 = \vec{x}_0 \times \overrightarrow{DD > DDD}_0$$

$$\vec{z}_0 = \vec{x}_0 \times \vec{y}_0$$

#### weitere Angaben

$\vec{h}$  lokaler Ortsvektor zum Proton

$|h|$  Bindungslänge D>H

$\varphi$  Winkel der DD-D-H-Verbindungen

$\theta = [0^\circ; 360^\circ]$ , wobei  $0^\circ = \vec{y}_0$ -parallel und  $90^\circ = \vec{z}_0$ -parallel = vor X-Y-Ebene

$r$  Kreisradius der Protonposition um die x-Achse

$$r = |h| \cdot \cos(\vec{h}, \vec{y}_0) = |h| \cdot \cos(\varphi - 90^\circ)$$

#### lokale Richtungsvektoren der Protonposition (mit richtiger Länge)

$$\vec{h}_x = \vec{x}_0 \cdot |h| \cdot \cos(\vec{h}, \vec{x}_0) = \vec{x}_0 \cdot |h| \cdot \cos(180^\circ - \varphi)$$

$$\vec{h}_y = \vec{y}_0 \cdot r \cdot \cos \theta$$

$$\vec{h}_z = \vec{z}_0 \cdot r \cdot \sin \theta$$

Rotation und Translation führen zu den Koordinaten im globalen Koordinatensystem

Grundtyp und Atome	Protonenpositionen	Bindungslänge
<b>Donator im Ringsystem und Stickstoff der Peptidbindung</b>  HIS ND1/AD1, CD2/AD2, CE1/AE1, NE2/AE2; ARG NE; TRP NE1; Peptidbindung N	eine fixierte Protonposition  in einer Ebene mit D, DD <sub>1</sub> und DD <sub>2</sub>  identische Winkel DD <sub>1</sub> -D-H und DD <sub>2</sub> -D-H	1,00 Å
<b>Donator direkt außerhalb eines Ringsystems</b>  TYR OH	zwei mögliche, fixierte Protonpositionen  in einer Ebene mit D, DD, DDD <sub>1</sub> , DDD <sub>2</sub>  $\phi = 110^\circ$ $\theta = 90^\circ$ oder $270^\circ$	1,00 Å
<b>zwei-protonige Aminogruppen</b>  ASN/ASX OD1/AD1, ND2/AD2; GLN/GLX OE1/AE1, NE2/AE2; ARG NH1, NH2	zwei fixierte Protonenpositionen  in einer Ebene mit D, DD, DDD <sub>1</sub> , DDD <sub>2</sub>  $\phi = 120^\circ$ $\theta = 90^\circ$ und $270^\circ$	1,00 Å
<b>drei-protonige Aminogruppen</b>  LYS NZ; N-terminales N	drei fixierte Protonenpositionen  $\phi = 110^\circ$ $\theta = 90^\circ, 210^\circ, 330^\circ$	1,01 Å
<b>primäre Hydroxyl- und Sulfhydrylgruppen</b>  SER OG; THR OG1; CYS SG	frei rotationsfähig  $\phi = 110^\circ$ (SER, THR), $96^\circ$ (CYS) $\theta = [0^\circ; 360^\circ]$  es werden exemplarisch drei fixierte Positionen bestimmt: $\theta = 90^\circ, 210^\circ, 330^\circ$	1,00 Å (SER, THR)  1,33 Å (CYS)

## A4 Verwendete Test-Datensätze

Die Test-Datensätze der beiden zusammengestellten Gruppen fakultativer und obligatorischer Protein-Protein-Komplexe werden hier mit ihren üblichen 4-Buchstabencodes aufgeführt. Diesen folgen die jeweils ausgetestete Peptidketten-Kombination mit den, in dem Datensatz verwendeten Kettenbezeichnern, die Angabe der erzielten Auflösung aus der Röntgenstrukturbestimmung in Ångström und zwei Spalten Kurzbeschreibung aus dem entsprechenden PDB-Eintrag.

### Obligatorische Protein-Protein-Komplexe:

1ab9	BC	1.60	Serine Protease/Peptide	
1aks	AB	1.80	Serine Protease	alpha-Trypsin; Chain: A, B; Ec: 3.4.21.4
1ept	BA	1.80	Hydrolase (Serine Protease)	Porcine E-Trypsin (E.C. 3.4.21.4)
1ept	CA	1.80	Hydrolase (Serine Protease)	Porcine E-Trypsin (E.C. 3.4.21.4)
1ept	CB	1.80	Hydrolase (Serine Protease)	Porcine E-Trypsin (E.C. 3.4.21.4)
1huc	BA	2.10	Thiol Protease	Cathepsin B (E.C. 3.4.22.1)
1kvd	AB	1.80	Toxin	
1lya	BA	2.50	Lysosomal Aspartic Protease	Cathepsin D (E.C. 3.4.23.5)
1mhl	CA	2.25	Myeloperoxidase	Myeloperoxidase; Chain: A, C, B, D; Ec: 1.11.1.7; Other_Details: Isoform C
1pnk	AB	1.90	Antibiotic Resistance	
1spb	SP	2.00	propeptid Complex (Serine Proteinase/Prosegment)	Subtilisin Bpn' Prosegment; Chain: P; Subtilisin Bpn'; Chain: S;
1wht	AB	2.00	Serine Carboxypeptidase	Serine Carboxypeptidase II (E.C. 3.4.16.1) Complexed With L-Benzylsuccinate
2hnt	FC	2.50	Serine Protease	gamma-Thrombin
2kin	AB	1.90	Motor Protein	
2ltn	AB	1.70	Lectin	Pea Lectin
2pcd	NB	2.15	Dioxygenase	Protocatechuate 3,4-Dioxygenase (E.C. 1.13.11.3)

### Fakultative Protein-Protein-Komplexe:

1abr	BA	2.14	Complex (Glycosidase/Carbohydrate)	Abrin-A Complexed With Two Sugar Chains
1acb	EI	2.00	Hydrolase (Serine Protease)	alpha-Chymotrypsin (E.C. 3.4.21.1) Complex With Eglin C
1apy	AB	2.00	Hydrolase	Aspartylglucosaminidase; Chain: A, B, C, D
1bpl	BA	2.20	Glycosyltransferase	alpha-1,4-Glucan-4-Glucanohydrolase; Chain: A, B; Synonym: alpha-Amylase (Bla)
1brs	CF	2.00	Endonuclease	Barnase (G Specific Endonuclease) Complexed With Barstar Mutant With Cys40 Replaced By Ala and Cys82 Replaced by Ala
1bun	AB	2.45	Toxin	beta2-Bungarotoxin; Chain: A, B; Synonym: Phosphatide Acyl-Hydrolase, Phosphatidylcholine2-Acylhydrolase
1cgj	EI	2.30	Serine Protease/Inhibitor Complex	alpha-Chymotrypsinogen Complex With Human Pancreatic Secretory Trypsin Inhibitor Variant 4
1cho	EI	1.80	Complex (Serine Proteinase-Inhibitor)	alpha-Chymotrypsin (E.C. 3.4.21.1) Complex With Turkey Ovomucoid Third Domain (OMTKY3)
1cpc	BA	1.66	Light Harvesting Protein	C-Phycocyanin
1dfj	IE	2.50	Complex (Endonuclease/Inhibitor)	Ribonuclease A; Chain: E; Synonym: Rnase A; Molecule: Ribonuclease Inhibitor; Chain: I; Synonym: Rnase Inhibitor
1fcb	AB	2.40	Oxidoreductase (Ch-OH(D)-Cytochrome(A))	Flavocytochrome b2 (E.C. 1.1.2.3)
1fin	AB	2.30	Complex (Transferase/Cyclin)	Cyclin-Dependent Kinase 2; Chain: A, C; Synonym: Cdk2 Cyclin A; Chain: B, D;
1got	BA	2.00	Complex (GTP-Binding/Transducer)	Gt-alpha/Gi-alpha Chimera; Chain: A Gt-beta; Chain: B beta1, Transducin beta Subunit Gt-Gamma; Chain: G;
1got	BG	2.00	Complex (GTP-Binding/Transducer)	Gt-alpha/Gi-alpha Chimera; Chain: A Gt-beta; Chain: B beta1, Transducin beta Subunit Gt-Gamma; Chain: G;
1gua	AB	2.00	Complex (GTP-Binding/ATP-Binding)	Rap1A; Chain: A C-Raf1; Chain: B
1hbh	DC	2.20	Oxygen Carrier	Hemoglobin (Deoxy)
1hcg	AB	2.20	Coagulation Factor	Blood Coagulation Factor Xa
1htr	BP	1.62	Aspartyl Protease	Progastricsin (Pepsinogen C) (E.C. 3.4.23.3)
1lpb	BA	2.46	Hydrolase (Carboxylic Esterase)	Lipase (E.C. 3.1.1.3) Complexed With Colipase and Inhibited By Undecane Phosphonate Methyl ester
1lta	EC	2.20	Enterotoxin	Heat-Labile Enterotoxin (Lt) Complex With Galactose
1lts	AC	1.95	Toxin	Heat-Labile Enterotoxin (Lt)

1mda	HL	2.50	Electron Transport	Methylamine Dehydrogenase (E.C. 1.4.99.3) Complex With Amicyanin
1mee	AI	2.00	Complex(Serine Proteinase-Inhibitor)	Mesentericopeptidase (E.C. 3.4.21.14) Peptidyl Peptide Hydrolase Complex With Eglin-C
1mmo	CH	2.20	Oxidoreductase (Monooxygenase)	Methane Monooxygenase Hydrolase (E.C. 1.14.13.25) (Methane Hydroxylase)
1mmo	EC	2.20	Oxidoreductase (Monooxygenase)	Methane Monooxygenase Hydrolase (E.C. 1.14.13.25) (Methane Hydroxylase)
1mmo	EH	2.20	Oxidoreductase (Monooxygenase)	Methane Monooxygenase Hydrolase (E.C. 1.14.13.25) (Methane Hydroxylase)
1ppf	EI	1.80	Hydrolase(Serine Proteinase)	Human Leukocyte Elastase (Hle) (Neutrophil Elastase (Hne)) Complex With The Third Domain Of Turkey Ovomucoid Inhibitor
1prc	CL	2.30	Photosynthetic Reaction Center	Photosynthetic Reaction Center
1prc	CM	2.30	Photosynthetic Reaction Center	Photosynthetic Reaction Center
1prc	ML	2.30	Photosynthetic Reaction Center	Photosynthetic Reaction Center
1pvc	12	2.40	Virus	Poliovirus Type 3, Sabin Strain; Chain: 0, 1, 2, 3, 4
1pya	DC	2.50	Carboxy-Lyase	Pyruvoyl-Dependent Histidine Decarboxylase (L-Histidine Carboxylase) (E.C. 4.1.1.22)
1pyt	CA	2.35	pro? Ternary Complex (Zymogen)	Procarboxypeptidase A; Chain: A, B; Proproteinase E; Chain: C Chymotrypsinogen C; Chain: D;
1req	CD	2.00	Isomerase	Methylmalonyl-Coa Mutase; Chain: A, B, C, D Details: Chains A and C Include Coenzyme B12, Desulpho-CoA, and A Glycerol
1scu	ED	2.50	Ligase (ATP-Binding)	Succinyl-Coa Synthetase (Succinate-Coa Ligase) (Adp-Forming) (E.C. 6.2.1.5)
1sgp	EI	1.40	Complex (Serine Protease/Inhibitor)	Streptomyces Griseus Proteinase B; Chain: E; Turkey Ovomucoid Inhibitor; Chain: I; Domain: Third Domain
1slu	BA	1.80	Complex (Serine Protease/Inhibitor)	Ecotin; Chain: A; Synonym: Trypsin Inhibitor; Anionic Trypsin; Chain: B; Ec: 3.4.21.4;
1smp	AI	2.30	Complex (Metalloprotease/Inhibitor)	Serratia Metallo Proteinase; Chain: A; Erwinia Chrysanthemi Inhibitor; Chain: I;
1stf	EI	2.37	Hydrolase(Sulphydryl Proteinase)	Papain (Cys 25 Carboxymethylated) (E.C. 3.4.22.2) Complexed With The Inhibitor Stefin B (Cystatin B)
1tab	EI	2.30	Hydrolase (Serine Proteinase)	Trypsin (E.C. 3.4.21.4) Complex With Bowman-Birk Inhibitor (AB-I)
1taf	BA	2.00	Complex (Two Transcription Factors)	Tfiid TBP Associated Factor 42; Chain: A; Fragment: Residues 19 - 86; Tfiid TBP Associated Factor 62; Chain B
1tco	AB	2.50	Complex (Hydrolase/Isomerase)	Serine/Threonine Phosphatase B2; Chain: A, B; Fragment: Chain A Catalytic Subunit, Res 18-392. Chain B Regulatiory Subunit
1tco	AC	2.50	Complex (Hydrolase/Isomerase)	Serine/Threonine Phosphatase B2; Chain: A, B; Fragment: Chain A Catalytic Subunit, Res 18-392. Chain B Regulatiory Subunit
1tii	AC	2.25	Enterotoxin	Heat Labile Enterotoxin Type Iib; Chain: D, E, F, G, H, A, C; Synonym: Lt-Iib
1ubs	BA	1.90	Complex (Lyase/Peptide)	Tryptophan Synthase; Chain: A, B; L-Serine; Chain: C; Bound To B Subunit
1vrt	AB	2.20	Nucleotidyltransferase	Hiv-1 Reverse Transcriptase; Chain: A, B; Synonym: Hiv-1 Rt;
1wdc	CA	2.00	Muscle Protein	Scallop Myosin; Chain: A, B, C; Fragment: Proteolytic Fragment, Regulatory Domain;
1yrn	BA	2.50	Complex (Two DNA-Binding Proteins/DNA)	Mating-Type Protein A-1; Chain: A; Domain: Homeodomain; Synonym: Mat A1; Mating-Type Protein alpha-2; Chain B; Domain: Homeodomain
1ytf	DB	2.50	Complex (Transcription Regulation/DNA)	Yeast Tata-Box Binding Protein; Chain: A; Fragment: C-Terminal 180 Amino Acids;
2bbk	HL	1.75	Electron Transport	Methylamine Dehydrogenase (Madh) (E.C. 1.4.99.3)
2kai	BI	2.50	Complex (Proteinase-Inhibitor)	Kallikrein A (E.C. 3.4.21.8) Complex With Bovine Pancreatic Trypsin Inhibitor
2kau	CA	2.00	Hydrolase (Urea Amido)	Klebsiella Aerogenes Urease; Ec: 3.5.1.5; Synonyms: Urea Amidohydrolase, Urease;
2kau	CB	2.00	Hydrolase (Urea Amido)	Klebsiella Aerogenes Urease; Ec: 3.5.1.5; Synonyms: Urea Amidohydrolase, Urease;
2pcc	AB	2.30	Oxidoreductase/Electron Transport	Yeast Cytochrome C Peroxidase (Ccp) Complex With Yeast Iso-1-Cytochrome C
2pka	BA	2.05	Serine Proteinase	Kallikrein A (E.C. 3.4.21.8)
2ptc	EI	1.90	Complex (Proteinase/Inhibitor)	beta-Trypsin (E.C. 3.4.21.4) Complex With Pancreatic Trypsin Inhibitor
3sic	EI	1.80	Complex (Proteinase/Inhibitor)	Subtilisin Bpn' Complex With A Modified Ssi (Streptomyces Subtilisin Inhibitor)
4aah	CD	2.40	Oxidoreductase (Pqq(A)-Choh(D))	Methanol Dehydrogenase; Chain: A, B, C, D;
4htc	HI	2.30	Hydrolase(Serine Protease)	alpha-Thrombin (E.C. 3.4.21.5) Complex With Recombinant Hirudin (Variant 2, Lys 47)
4sgb	EI	2.10	Complex(Serine Proteinase-Inhibitor)	Serine Proteinase B Complex With The Potato Inhibitor PCI-1
8atc	AB	2.50	Transferase (Carbamoyl-P, Aspartate)	Aspartate Carbamoyltransferase (Aspartate Transcarbamylase) (R State) Complex With N-Phosphonacetyl-L-Aspartate (PALA)
8ruc	EK	1.60	Lyase (Carbon-Carbon)	Ribulose-1,5-Bisphosphate Carboxylase/Oxygenase; Chain: A, C, E, G, I, J, K, L; Synonym: Rubisco

## A5 Beschreibungen und Eignungsbestimmung der Test-Datensätze

Die Tabellen enthalten für die fakultativen und obligatorischen Protein-Protein-Komplexe die 4-Buchstabencodes und einige beschreibende Angaben zu den jeweiligen Kontaktflächen sowie die Anzahl der erzielbaren Referenzergebnisse bei Verwendung unterschiedlicher Selektionsgrenzwerte. Für die Komplexe mit teilweise eingeklammerten Werten sind einzelne Werte gegebenenfalls nicht gegeben, da sie nicht der Proteinkomplex-Liste von Vakser & Sali [Vakser & Sali, 2001] entstammen.

Legende:

- msa: Molekulare Kontaktfläche, die durch den Komplexpartner abgeschirmt wird (*molecular surface area*)
- asa: Lösungsmittelzugängliche Kontaktfläche, die durch den Komplexpartner abgeschirmt wird (*accessible surface area*)
- #res: Anzahl der Aminosäuren in der Kontaktfläche
- #wbb: Anzahl der intermolekularen Wasserstoffbrückenbindungen
- #asa\*5,34\*10e-3 / Å<sup>2</sup>: lineare Abhängigkeit der Anzahl intermolekularer Wasserstoffbrückenbindungen zur Lösungsmittelzugängliche Kontaktfläche, die durch den Komplexpartner abgeschirmt wird [Janin & Chothia, 1990]
- #Ref.paare, -tripletts: Anzahl der Paar- bzw. Triplet-Referenzergebnisse, die aus den intermolekularen Wasserstoffbrückenbindungen gebildet werden können und den Selektionsanforderungen des entsprechenden Grenzwertensembles gehorchen.
- eng, mittel, weit: Kennzeichnung des entsprechenden Grenzwertensembles, das zur Selektion angewendet wird.

Mit den Schattierungen wird gekennzeichnet, wenn ein Komplex nur zwei (hellgrau) ein (mittelgrau) oder gar kein (dunkelgrau) Referenzergebnis für die jeweilige Selektion aufweisen kann.

### Obligatorische Protein-Protein-Komplexe:

	msa [Å <sup>2</sup> ]	asa [Å <sup>2</sup> ]	#res	#wbb	asa*5.34 *10e-3/Å <sup>2</sup>	#Ref.paare			#Ref.tripletts		
						eng	mittel	weit	end	mittel	weit
1ab9 B C				(34)		29	41	78	10	22	65
1aks A B	1704	5604	170	28	29,9	21	30	72	5	13	56
1ept B A	1137	3735	86	29	19,9	16	18	41	6	8	35
1ept C A	678	2211	61	59	11,8	6	11	16	0	6	14
1ept C B	1352	4453	111	80	23,8	15	21	45	4	9	27
1huc B A	1388	4559	125	78	24,3	6	8	34	1	1	6
1kvd A B				(32)		30	41	63	12	25	53
1lya B A	1942	6452	187	19	34,5	50	59	155	13	24	109
1mhl C A	3403	11205	235	42	59,8	20	33	153	0	1	81
1pnk A B				(22)		27	33	118	7	11	57
1spb S P	669	2228	68	10	11,9	6	8	28	1	2	13
1wht A B	2512	8277	187	40	44,2	18	22	58	5	5	30
2hnt F C	697	2269	63	75	12,1	4	8	10	0	5	5
2kin A B				(10)		36	40	73	15	17	50
2ltn A B	2260	7537	210	29	40,2	102	133	253	40	89	315
2pcd N B	2158	7134	152	46	38,1	13	13	56	1	1	7
Summe	19900	65664	1655	633	351	399	519	1253	120	239	923
Mittelwert	1658,3	5472,0	138,0	44,6 (39,6)	29,2	25,0	32,4	78,3	7,5	14,9	57,7
Std.abw.	804,7	2662,2	58,2	24,2 (22,0)	14,2	23,3	29,7	61,4	9,7	20,7	72,3

## Fakultative Protein-Protein-Komplexe:

	msa [Å <sup>2</sup> ]	asa [Å <sup>2</sup> ]	#res	#wbb	asa*5.34 *10e-3/Å <sup>2</sup>	Ref.paare			Ref.tripletts		
						eng	mittel	weit	end	mittel	weit
1abr BA	1052	3499	77	14	18,7	1	2	7	0	0	1
1acb EI	472	1547	41	9	8,3	5	5	14	0	1	7
1apy AB	1685	5520	169	37	29,5	30	40	68	10	20	44
1bpl BA	1788	5961	114	40	31,8	27	27	74	5	6	48
1brs CF	453	1529	26	12	8,2	1	2	11	0	0	0
1bun AB	315	1045	13	3	5,6	0	0	0	0	0	0
1cgg EI	607	1993	51	9	10,6	4	5	14	0	0	5
1cho EI	448	1466	43	10	7,8	5	6	20	0	1	11
1cpc BA	941	3048	57	8	16,3	0	0	1	0	0	0
1dfj IE	761	2584	32	6	13,8	0	0	1	0	0	0
1fcb AB	1635	5333	104	13	28,5	5	4	7	0	0	2
1fin AB	1023	3403	64	17	18,2	3	4	11	0	0	3
1got BA	750	2494	49	17	13,3	5	9	22	0	1	16
1got BG	1461	4746	97	16	25,3	1	1	6	0	0	1
1gua AB	379	1293	18	12	6,9	7	7	23	1	1	14
1hbh DC	486	1584	39	9	8,5	2	2	10	0	0	3
1hcg AB	528	1777	34	10	9,5	1	2	11	0	0	4
1htr BP	1413	4668	91	25	24,9	16	26	41	4	16	36
1lpb BA	485	1622	27	11	8,7	1	3	14	0	0	7
1lta EC	426	1409	25	4	7,5	0	0	2	0	0	0
1lts AC	737	2436	37	9	13,0	0	0	2	0	0	0
1mda HL	837	2763	59	7	14,8	0	1	2	0	0	0
1mee AI	536	1769	56	10	9,4	4	4	14	0	0	3
1mmo CH	839	2774	49	12	14,8	1	0	6	0	0	0
1mmo EC	3137	10300	213	55	55,0	2	7	50	0	1	16
1mmo EH	1205	4021	59	22	21,5	3	11	23	0	2	8
1ppf EI	408	1323	42	6	7,1	2	2	8	0	0	2
1prc CL	1028	3424	72	19	18,3	1	3	16	0	0	5
1prc CM	1605	5286	114	28	28,2	6	8	27	0	1	7
1prc ML	2626	8586	174	41	45,8	3	6	38	0	0	8
1pvc I2	1391	4598	105	27	24,6	4	8	28	0	0	7
1pya DC	1909	6289	191	44	33,6	38	47	119	10	21	95
1pyt CA	630	2106	43	11	11,2	1	2	7	0	0	0
1req CD	3211	10502	220	55	56,1	10	18	62	0	2	18
1scu ED	1072	3489	78	9	18,6	0	1	3	0	0	1
1sgp EI	320	1061	39	9	5,7	6	7	15	0	1	9
1slu BA	580	1906	37	5	10,2	1	1	6	0	0	1
1smp AI	541	1820	40	9	9,7	1	1	7	0	0	1
1stf EI	545	1797	51	5	9,6	1	1	3	0	0	0
1tab EI	404	1364	35	7	7,3	4	5	10	0	0	5
1taf BA	1035	3335	78	12	17,8	1	2	6	0	0	1
1tco AB	1200	3819	79	8	20,4	0	0	0	0	0	0
1tco AC	350	1156	20	2	6,2	0	0	1	0	0	0
1tii AC	624	2039	41	5	10,9	0	1	2	0	0	0
1ubs BA	788	2621	49	16	14,0	3	2	14	0	0	5
1vrt AB	1407	4641	92	16	24,8	1	0	6	0	0	0
1wdc CA	1144	3709	59	14	19,8	1	1	6	0	0	1
1yrn BA	389	1281	26	6	6,8	1	1	4	0	0	0
1ytf DB	647	2068	45	6	11,0	0	0	0	0	0	0
2bbk HL	957	3175	59	19	17,0	3	7	21	0	0	8
2kai BI	333	1094	32	8	5,8	3	5	11	0	0	2
2kau CA	570	1872	37	13	10,0	3	2	13	0	0	3
2kau CB	1052	3468	88	22	18,5	21	24	48	5	9	31
2pcc AB	340	1138	12	1	6,1	0	0	0	0	0	0
2pka BA	1689	5517	154	33	29,5	22	23	64	6	8	39
2ptc EI	431	1430	38	12	7,6	6	8	18	0	0	8
3sic EI	485	1603	53	11	8,6	4	4	20	0	1	10
4aah CD	1126	3749	82	24	20,0	4	1	22	1	0	9
4htc HI	880	2957	48	16	15,8	5	9	20	0	1	15
4sgb EI	382	1249	48	6	6,7	3	4	10	0	0	4
8atc AB	455	1536	34	9	8,2	1	1	6	0	0	1
8ruc EK	1104	3706	63	14	19,8	2	3	12	0	0	5
Summe	58057	191298	4122	945	1021,5	286	376	1107	42	93	530
Mittelwert	936,4	3085,5	66,5	15,2	16,5	4,6	6,1	17,9	0,7	1,5	8,5
Std.abw.	630,4	2063,0	46,7	12,0	11,0	7,5	9,3	21,6	2,1	4,3	15,5





## A6 Differenzierte Zulassung der fakultativer Protein-Protein-Komplexe

Abgeleitet aus der vorhergehenden Tabelle werden hier die fakultativen Komplexe mit einem Kreuz gekennzeichnet, die bei Verwendung des angegebenen Grenzwertensembles (eng und mittel oder weit) und der jeweiligen Methode jeweils zwei oder mehr Referenzergebnissen aufweisen können.

	Paar-Docking		Triplett-Docking			Paar-Docking		Triplett-Docking	
	eng und mittel	weit	eng und mittel	weit		eng und mittel	weit	eng und mittel	weit
1abr BA		×			1pya DC	×	×	×	×
1acb EI	×	×		×	1pyt CA		×		
1apy AB	×	×	×	×	1req CD	×	×		×
1bpl BA	×	×	×	×	1scu ED		×		
1brs CF		×			1sgp EI	×	×		×
1bun AB					1slu BA		×		
1cgj EI	×	×		×	1smp AI		×		
1cho EI	×	×		×	1stf EI		×		
1cpc BA					1tab EI	×	×		×
1dfj IE					1taf BA		×		
1fcb AB	×	×		×	1tco AB				
1fin AB	×	×		×	1tco AC				
1got BA	×	×		×	1tii AC		×		
1got BG		×			1ubs BA	×	×		×
1gua AB	×	×		×	1vrt AB		×		
1hbh DC	×	×		×	1wdc CA		×		
1heg AB		×		×	1yrn BA		×		
1htr BP	×	×	×	×	1ytf DB				
1lpb BA		×		×	2bbk HL	×	×		×
1lta EC		×			2kai BI	×	×		×
1lts AC		×			2kau CA	×	×		×
1mda HL		×			2kau CB	×	×	×	×
1mee AI	×	×		×	2pcc AB				
1mmo CH		×			2pka BA	×	×	×	×
1mmo EC	×	×		×	2ptc EI	×	×		×
1mmo EH	×	×		×	3sic EI	×	×		×
1ppf EI	×	×		×	4aah CD		×		×
1prc CL		×		×	4htc HI	×	×		×
1prc CM	×	×		×	4sgb EI	×	×		×
1prc ML	×	×		×	8atc AB		×		
1pvc 12	×	×		×	8ruc EK	×	×		×
<b>jeweilige Anzahl der zugelassenen Komplexe:</b>					<b>33</b>	<b>55</b>	<b>6</b>	<b>37</b>	

## A7 Evaluierungs-, Wiederfindungs und Erhaltungsquoten

Legende:      ref: erzielbare Referenzen  
                   eval: evaluierte Referenzen  
                   found: gefundene Referenzen  
                   rot: Referenzen mit richtiger Rotationsbestimmung  
                   #sum: Gesamtsummen der Referenzergebnisse in der jeweiligen Spalte  
                   %ref: prozentualer Anteil bzgl. #sum-ref  
                   %eval: prozentualer Anteil bzgl. #sum-eval  
                   %found: prozentualer Anteil bzgl. #sum-found

### Im Triplet-Docking

16 obligatorische Komplexe					58 fakultative Komplexe					
Verwendung des engen Grenzwertensembles										
	ref	eval	found	rot		ref	eval	found	rot	
#sum	120	119	118	82	alle	#sum	42	42	41	25
%ref		99,2	98,3	68,3	Atome	%ref		100,0	97,6	59,5
%eval			99,2	68,9		%eval			97,6	59,5
%found				69,5		%found				61,0
#sum	120	119	118	82	ungesättigte	#sum	42	41	40	25
%ref		99,2	98,3	68,3	Atome	%ref		97,6	95,2	59,5
%eval			99,2	68,9		%eval			97,6	61,0
%found				69,5		%found				62,5
#sum	120	118	117	81	ungebundene	#sum	42	35	35	24
%ref		98,3	97,5	67,5	Atome	%ref		83,3	83,3	57,1
%eval			99,2	68,6		%eval			100,0	68,6
%found				69,2		%found				68,6
Verwendung des mittleren Grenzwertensembles										
	ref	eval	found	rot		ref	eval	found	rot	
#sum	239	220	219	152	alle	#sum	93	87	87	52
%ref		92,1	91,6	63,6	Atome	%ref		93,5	93,5	55,9
%eval			99,5	69,1		%eval			100,0	59,8
%found				69,4		%found				59,8
#sum	239	218	217	151	ungesättigte	#sum	93	75	75	49
%ref		91,2	90,8	63,2	Atome	%ref		80,6	80,6	52,7
%eval			99,5	69,3		%eval			100,0	65,3
%found				69,6		%found				65,3
#sum	239	215	214	150	ungebundene	#sum	93	67	67	46
%ref		90,0	89,5	62,8	Atome	%ref		72,0	72,0	49,5
%eval			99,5	69,8		%eval			100,0	68,7
%found				70,1		%found				68,7
Verwendung des weiten Grenzwertensembles										
	ref	eval	found	rot		ref	eval	found	rot	
#sum	923	902	831	541	alle	#sum	530	485	446	233
%ref		97,7	90,0	58,6	Atome	%ref		91,5	84,2	44,0
%eval			92,1	60,0		%eval			92,0	48,0
%found				65,1		%found				52,2
#sum	923	824	777	511	ungesättigte	#sum	530	366	342	183
%ref		89,3	84,2	55,4	Atome	%ref		69,1	64,5	34,5
%eval			94,3	62,0		%eval			93,4	50,0
%found				65,8		%found				53,5
#sum	923	782	747	494	ungebundene	#sum	530	321	301	164
%ref		84,7	80,9	53,5	Atome	%ref		60,6	56,8	30,9
%eval			95,5	63,2		%eval			93,8	51,1
%found				66,1		%found				54,5

## Im Paar-Docking

16 obligatorische Komplexe					58 fakultative Komplexe					
Verwendung des engen Grenzwertensembles										
	ref	eval	found	rot		ref	eval	found	rot	
#sum	399	397	387	200	alle Atome	#sum	286	281	269	90
%ref		99,5	97,0	50,1		%ref		98,3	94,1	31,5
%eval			97,5	50,4		%eval			95,7	32,0
%found				51,7		%found				33,5
#sum	399	386	377	197	ungesättigte Atome	#sum	286	239	230	83
%ref		96,7	94,5	49,4		%ref		83,6	80,4	29,0
%eval			97,7	51,0		%eval			96,2	34,7
%found				52,3		%found				36,1
#sum	399	370	362	191	ungebundene Atome	#sum	286	223	215	77
%ref		92,7	90,7	47,9		%ref		78,0	75,2	26,9
%eval			97,8	51,6		%eval			96,4	34,5
%found				52,8		%found				35,8
Verwendung des mittleren Grenzwertensembles										
	ref	eval	found	rot		ref	eval	found	rot	
#sum	519	505	501	271	alle Atome	#sum	376	364	357	123
%ref		97,3	96,5	52,2		%ref		96,8	94,9	32,7
%eval			99,2	53,7		%eval			98,1	33,8
%found				54,1		%found				34,5
#sum	519	488	484	263	ungesättigte Atome	#sum	376	295	289	105
%ref		94,0	93,3	50,7		%ref		78,5	76,9	27,9
%eval			99,2	53,9		%eval			98,0	35,6
%found				54,3		%found				36,3
#sum	519	467	463	256	ungebundene Atome	#sum	376	275	269	98
%ref		90,0	89,2	49,3		%ref		73,1	71,5	26,1
%eval			99,1	54,8		%eval			97,8	35,6
%found				55,3		%found				36,4
Verwendung des weiten Grenzwertensembles										
	ref	eval	found	rot		ref	eval	found	rot	
#sum	1.253	1.232	1.198	558	alle Atome	#sum	1.107	1.069	1.022	338
%ref		98,3	95,6	44,5		%ref		96,6	92,3	30,5
%eval			97,2	45,3		%eval			95,6	31,6
%found				46,6		%found				33,1
#sum	1.253	1.093	1.061	517	ungesättigte Atome	#sum	1.107	841	813	276
%ref		87,2	84,7	41,3		%ref		76,0	73,4	24,9
%eval			97,1	47,3		%eval			96,7	32,8
%found				48,7		%found				33,9
#sum	1.253	1.011	992	498	ungebundene Atome	#sum	1.107	746	730	254
%ref		80,7	79,2	39,7		%ref		67,4	65,9	22,9
%eval			98,1	49,3		%eval			97,9	34,0
%found				50,2		%found				34,8

## A8 Die Translationsbestimmung obligatorischer Komplexe

In der Tabelle sind die Abweichungen der gemittelten Translationsvorschriften der richtig rotationsbestimmten Referenzergebnisse obligatorischer Proteinkomplexe zu der bekannten Ursprungstranslation angegeben. Unter den Spalten finden sich Mittelwerte und Standardabweichungen der entsprechenden Einträge. Lücken in der Tabelle kennzeichnen Untersuchungen, bei denen für weniger als zwei Referenzen die richtige Rotation bestimmt wurde.

Alle Angaben bezeichnen die Abweichung als euklidischer Abstand in Angström.

“alle“, “<2“ und “<1“ kennzeichnen die Verwendung aller, ungesättigter oder nur ungebundener Atome.

Paar-Docking	enge Grenzwerte			mittlere Grenzwerte			weite Grenzwerte		
	alle	< 2	< 1	alle	< 2	< 1	alle	< 2	< 1
1ab9_B_c	1,26	1,26	1,24	1,47	1,36	1,35	1,47	1,43	1,45
1aks_A_b	1,49	1,49	1,49	1,64	1,52	1,52	1,50	1,45	1,43
1ept_B_a	1,31	1,31	1,31	1,31	1,31	1,31	1,19	1,19	1,19
1ept_C_a	1,09	1,09	1,09	1,12	1,12	1,09	1,07	1,07	1,07
1ept_B_c	1,20	1,30	1,33	1,18	1,25	1,27	1,15	1,16	1,25
1huc_B_a	1,58	1,58	1,58				1,64	1,70	1,70
1kvd_B_a	1,07	1,07	1,07	1,10	1,10	1,10	1,06	1,06	1,06
1lya_B_a	1,36	1,36	1,36	1,32	1,32	1,32	1,26	1,24	1,24
1mhl_C_a	1,25	1,25	1,76	1,35	1,53	1,79	1,32	1,39	1,64
1pnk_B_a	1,48	1,48	1,48	1,53	1,53	1,53	1,49	1,48	1,48
1spb_S_p	1,49	1,49	1,49	1,49	1,49	1,49	1,28	1,28	1,28
1wht_A_b	1,44	1,44	1,42	1,44	1,44	1,41	1,49	1,55	1,51
2hnt_F_c				1,31	1,31	1,31	1,31	1,31	1,31
2kin_A_b	1,49	1,49	1,49	1,52	1,52	1,52	1,48	1,48	1,48
2ltn_A_b	1,57	1,57	1,57	1,57	1,57	1,57	1,44	1,45	1,45
2pcd N b	1,97	1,97	1,97	1,76	1,76	1,92	1,52	1,54	1,37
Mittelwert	1,40	1,41	1,44	1,41	1,41	1,43	1,35	1,36	1,37
Standardabw.	0,22	0,21	0,23	0,18	0,17	0,22	0,17	0,18	0,18
Triplett-Docking									
1ab9_B_c	1,14	1,14	1,14	1,27	1,27	1,27	1,52	1,50	1,49
1aks_A_b				1,81	1,66	1,66	1,69	1,50	1,50
1ept_B_a	1,02	1,02	1,02	1,00	1,00	1,00	1,03	1,03	1,03
1ept_C_a				1,19	1,19	1,12	1,05	1,11	1,07
1ept_B_c				1,44	1,44	1,44	1,25	1,32	1,34
1huc_B_a									
1kvd_B_a	1,03	1,03	1,03	1,14	1,14	1,14	1,09	1,08	1,08
1lya_B_a	1,54	1,54	1,54	1,48	1,48	1,48	1,40	1,39	1,39
1mhl_C_a							1,23	1,28	1,54
1pnk_B_a	1,63	1,63	1,63	1,82	1,82	1,82	1,59	1,59	1,65
1spb_S_p				1,57	1,57	1,57	1,42	1,42	1,42
1wht_A_b	1,34	1,34	1,16	1,26	1,26	1,26	1,47	1,44	1,36
2hnt_F_c				1,23	1,23	1,23	1,23	1,23	1,23
2kin_A_b	1,51	1,51	1,51	1,49	1,49	1,49	1,44	1,44	1,44
2ltn_A_b	1,60	1,60	1,60	1,61	1,61	1,61	1,48	1,48	1,48
2pcd N b							1,75	1,75	1,75
Mittelwert	1,35	1,35	1,33	1,41	1,40	1,39	1,38	1,37	1,38
Standardabw.	0,24	0,24	0,25	0,24	0,23	0,23	0,22	0,19	0,20

## A9 Ergebnistabellen des Dockings fakultativer Komplexe

Aufgeführt sind hier nur fakultative Komplexe, die für die jeweilige Methoden- und Parameterkombination mindestens zwei Referenzergebnisse aufweisen können.

Jede Zeile enthält nach dem Namen des einzelnen Komplexes folgende Spalteneinträge, die sich immer auf die Ergebnisgruppe des Rotationsrepräsentanten beziehen, der die richtige Orientierung beschreibt:

- #Referenzen* Anzahl der erzielbaren Referenzergebnisse des Test-Datensatzes
- # hits* Anzahl der Kandidaten zum Betrachtungszeitpunkt
- # ref* Anzahl der Referenzen zum Betrachtungszeitpunkt
- rank* Positionierung des Clusters in einer absteigenden Sortierung
- trans* Abweichung der gemittelten Translationsbeschreibung aller Kandidaten im finalen Transformationsclustervon der bekannten Ursprungstransformation.

Es sind drei Betrachtungszeitpunkte aufgeführt: nach der Rotationsbestimmung (Rotationscluster), nach der Ergebnisausdünnung (Ergebnisausdünnung) und nach der finalen Clusterbildung (Transformationscluster).

Grauschattierungen in einzelnen Zeilenabschnitten markieren Ergebnisse bzw. Teilabschnitte, bei denen weniger als zwei Referenzergebnisse in der Ergebnismenge enthalten sind.

### A9.1 Docking mit engem Grenzwertensemble

Paar-Docking mit dem engen Grenzwertensemble und allen Atomen											
	# Referenzen	Rotationscluster			Ergebnisausdünnung			Transformationscluster			trans [Å]
		# hits	# ref	rank	# hits	# ref	rank	# hits	# ref	rank	
1acb_Ei	5	161	2	381	0	0	564	0	0	564	20,89
1cgj_Ei	4	135	2	370	9	0	86	3	0	402	28,33
1cho_Ei	5	133	2	331	0	0	564	0	0	564	20,89
1fcb_Ab	5	1180	0	337	84	0	193	5	0	444	44,79
1fin_Ab	3	434	0	237	17	0	197	4	0	291	41,51
1got_Ab	5	929	0	80	86	0	25	9	0	33	45,78
1hbh_Dc	2	97	1	186	3	0	195	3	0	195	40,10
1mee_Ai	4	146	1	279	3	0	380	3	0	380	41,35
1mmo_Ec	2	719	2	352	21	0	257	3	0	538	32,10
1mmo_Eh	3	439	1	183	15	0	181	3	0	531	28,48
1ppf_Ei	2	129	0	268	9	0	74	3	0	377	33,33
1prc_Cm	6	613	0	97	15	0	324	3	0	537	26,34
1prc_Ml	3	306	1	443	12	0	132	3	0	492	54,69
1pvc_Ab	4	920	1	328	61	0	150	4	0	504	76,86
1tab_Ei	4	100	1	244	1	0	403	0	0	564	20,89
1ubs_Ba	3	341	0	456	12	0	314	3	0	531	61,58
2bbk_Hl	3	398	2	53	15	2	328	4	0	372	32,64
2kai_Bi	3	77	3	327	3	0	265	3	0	265	31,28
2kau_Ca	3	235	1	333	0	0	564	0	0	564	20,89
2ptc_Ei	6	137	2	276	6	2	265	3	0	448	17,80
3sic_Ei	4	161	1	416	12	0	106	4	0	188	15,71
4htc_Hi	5	193	1	53	3	0	427	3	0	427	16,86

4sgb_Ei	3	71	0	407	0	0	564	0	0	564	20,89
8ruc_Ek	2	456	0	349	12	0	446	3	0	539	69,72
1apy_Ab	30	240	9	26	45	9	2	9	9	6	1,69
1bpl_Ba	27	566	9	262	51	9	34	10	9	8	1,13
1gua_Ab	7	129	2	423	4	2	298	4	2	136	0,44
1htr_Bp	16	126	7	283	7	7	52	7	7	1	1,94
1pya_Dc	38	174	16	153	30	16	1	21	16	1	1,25
1req_Cd	10	1756	5	144	146	5	79	10	5	50	1,88
1sgp_Ei	6	89	2	430	4	2	198	4	2	112	1,95
2kau_Cb	21	402	5	208	22	5	176	6	5	71	1,47
2pka_Ba	22	169	9	53	22	9	3	9	9	2	1,36

### Paar-Docking mit dem engen Grenzwertensemble und ungesättigten Atomen

	# Referenzen	Rotationscluster			Ergebnisausdünnung			Transformationscluster			trans [Å]
		# hits	# ref	rank	# hits	# ref	rank	# hits	# ref	rank	
1acb_Ei	5	82	2	424	0	0	564	0	0	564	40,10
1cgj_Ei	4	75	2	308	6	0	42	3	0	180	45,60
1cho_Ei	5	75	2	312	0	0	564	0	0	564	41,61
1fcb_Ab	5	674	0	310	20	0	307	4	0	367	31,28
1fin_Ab	3	246	0	244	3	0	390	3	0	390	43,03
1got_Ab	5	501	0	23	29	0	23	5	0	95	27,95
1hbh_Dc	2	57	1	74	3	0	64	3	0	64	25,24
1mee_Ai	4	80	1	389	0	0	564	0	0	564	16,86
1mmo_Ec	2	368	1	351	0	0	564	0	0	564	47,51
1mmo_Eh	3	258	1	176	6	0	195	3	0	379	45,12
1ppf_Ei	2	86	0	101	6	0	58	3	0	209	54,92
1prc_Cm	6	347	0	203	3	0	424	3	0	424	38,30
1prc_Ml	3	178	1	436	1	0	345	0	0	564	61,58
1pvc_Ab	4	565	1	368	22	0	143	4	0	320	69,53
1sgp_Ei	6	45	1	490	3	1	196	3	1	196	1,96
1tab_Ei	4	73	1	228	0	0	564	0	0	564	32,10
1ubs_Ba	3	205	0	381	6	0	215	3	0	413	20,89
2bbk_Hl	3	265	2	28	7	2	234	4	0	162	20,89
2kai_Bi	3	53	2	316	3	0	154	3	0	154	20,89
2kau_Ca	3	127	1	252	0	0	564	0	0	564	20,89
3sic_Ei	4	101	1	317	4	0	134	4	0	73	20,89
4htc_Hi	5	132	1	160	3	0	296	3	0	296	20,89
4sgb_Ei	3	50	0	415	0	0	564	0	0	564	20,89
8ruc_Ek	2	291	0	334	6	0	375	3	0	485	20,89
1apy_Ab	30	162	9	4	38	9	1	9	9	3	1,69
1bpl_Ba	27	322	9	217	34	9	1	10	9	1	1,13
1gua_Ab	7	91	2	303	4	2	139	4	2	71	0,44
1htr_Bp	16	82	7	285	7	7	14	7	7	1	1,94
1pya_Dc	38	114	16	118	25	16	1	19	16	1	1,21
1req_Cd	10	993	2	172	44	2	53	5	2	155	2,00
2kau_Cb	21	224	4	201	9	4	108	5	4	27	1,51
2pka_Ba	22	113	9	53	15	9	1	9	9	1	1,36
2ptc_Ei	6	79	2	204	3	2	225	3	2	225	2,55

### Paar-Docking mit dem engen Grenzwertensemble und ungebundenen Atomen

	# Referenzen	Rotationscluster			Ergebnisausdünnung			Transformationscluster			trans [Å]
		# hits	# ref	rank	# hits	# ref	rank	# hits	# ref	rank	
1acb_Ei	5	65	2	292	0	0	564	0	0	564	20,89
1cgj_Ei	4	47	2	378	6	0	12	3	0	113	27,95
1cho_Ei	5	53	2	103	0	0	564	0	0	564	20,89
1fcb_Ab	5	470	0	296	9	0	345	3	0	508	46,30
1fin_Ab	3	165	0	260	1	0	341	0	0	564	20,89
1got_Ab	5	318	0	38	13	0	56	4	0	154	41,68
1hbh_Dc	2	33	0	148	3	0	45	3	0	45	40,10
1mee_Ai	4	57	1	425	0	0	564	0	0	564	20,89
1mmo_Ec	2	219	1	367	0	0	564	0	0	564	20,89
1mmo_Eh	3	169	1	150	3	0	245	3	0	245	54,92

1ppf_Ei	2	53	0	200	3	0	116	3	0	116	44,52
1prc_Cm	6	238	0	186	3	0	292	3	0	292	69,53
1prc_Ml	3	115	1	341	0	0	564	0	0	564	20,89
1pvc_Ab	4	379	1	369	9	0	265	3	0	499	56,45
1req_Cd	10	657	2	152	28	0	9	4	0	207	46,27
1sgp_Ei	6	27	1	482	3	1	107	3	1	107	1,96
1tab_Ei	4	53	1	298	0	0	564	0	0	564	20,89
1ubs_Ba	3	120	0	336	6	0	53	3	0	215	61,58
2bbk_Hl	3	214	2	8	7	2	113	4	0	101	43,03
2kai_Bi	3	42	2	348	3	0	125	3	0	125	31,28
2kau_Ca	3	79	1	225	0	0	564	0	0	564	20,89
2kau_Cb	21	145	2	196	0	0	564	0	0	564	20,89
2ptc_Ei	6	43	2	255	0	0	564	0	0	564	20,89
3sic_Ei	4	80	1	193	4	0	85	4	0	53	45,60
4htc_Hi	5	95	1	100	3	0	182	3	0	182	16,86
4sgb_Ei	3	32	0	393	0	0	564	0	0	564	20,89
8ruc_Ek	2	221	0	227	6	0	265	3	0	427	43,17
1apy_Ab	30	130	9	3	34	9	1	9	9	2	1,69
1bpl_Ba	27	230	9	161	17	9	5	10	9	1	1,13
1gua_Ab	7	68	2	221	4	2	70	4	2	43	0,44
1htr_Bp	16	65	7	128	7	7	5	7	7	1	1,94
1pya_Dc	38	90	13	127	20	13	1	16	13	1	1,21
2pka_Ba	22	91	9	58	12	9	3	9	9	1	1,36

### Triplett-Docking mit dem engen Grenzwertensemble und allen Atomen

	# Referenzen	Rotationscluster			Ergebnisausdünnung			Transformationscluster			trans [Å]
		# hits	# ref	rank	# hits	# ref	rank	# hits	# ref	rank	
1apy_Ab	10	40	8	1	22	8	1	6	0	1	7,50
2kau_Cb	5	32	3	11	9	2	23	3	0	60	74,60
1bpl_Ba	5	53	3	2	9	3	41	3	3	102	2,88
1htr_Bp	4	14	3	10	5	3	2	3	3	9	3,25
1pya_Dc	10	28	4	2	17	3	1	8	3	2	4,34
2pka_Ba	6	16	4	12	5	4	18	3	3	31	3,07

### Triplett-Docking mit dem engen Grenzwertensemble und ungesättigten Atomen

	# Referenzen	Rotationscluster			Ergebnisausdünnung			Transformationscluster			trans [Å]
		# hits	# ref	rank	# hits	# ref	rank	# hits	# ref	rank	
1pya_Dc	10	22	4	1	12	3	1	5	0	2	7,66
2kau_Cb	5	18	3	2	9	2	1	3	0	20	74,60
1apy_Ab	10	31	8	1	17	8	1	4	4	5	1,81
1bpl_Ba	5	23	3	5	3	3	49	3	3	19	2,88
1htr_Bp	4	11	3	2	3	3	13	3	3	8	3,25
2pka_Ba	6	14	4	1	5	4	2	3	3	5	3,07

### Triplett-Docking mit dem engen Grenzwertensemble und ungebundenen Atomen

	# Referenzen	Rotationscluster			Ergebnisausdünnung			Transformationscluster			trans [Å]
		# hits	# ref	rank	# hits	# ref	rank	# hits	# ref	rank	
1pya_Dc	10	15	3	2	7	2	3	3	0	11	7,77
2kau_Cb	5	13	3	3	9	2	1	3	0	15	74,60
1apy_Ab	10	29	8	1	17	8	1	4	4	4	1,81
1bpl_Ba	5	17	3	4	3	3	18	3	3	8	2,88
1htr_Bp	4	9	3	5	3	3	5	3	3	4	3,25
2pka_Ba	6	11	4	1	5	4	2	3	3	4	3,07

## A9.2 Docking mit mittlerem Grenzwertensemble

Paar-Docking mit dem mittleren Grenzwertensemble und allen Atomen											
	# Referenzen	Rotationscluster			Ergebnisausdünnung			Transformationscluster			trans [Å]
		# hits	# ref	rank	# hits	# ref	rank	# hits	# ref	rank	
1acb_Ei	5	254	2	269	13	0	388	4	0	401	40,68
1cgj_Ei	5	204	2	418	9	0	451	3	0	530	38,66
1cho_Ei	6	216	3	194	34	3	21	7	0	56	35,10
1fcb_Ab	4	2010	0	273	321	0	161	15	0	130	34,42
1fin_Ab	4	778	0	140	86	0	135	8	0	154	56,17
1got_Ab	9	1524	1	100	275	0	54	12	0	225	44,40
1hbh_Dc	2	151	1	233	7	0	155	4	0	131	39,48
1mee_Ai	4	251	1	137	18	0	264	8	0	20	32,93
1mmo_Eh	11	706	3	374	47	3	338	4	0	527	8,01
1ppf_Ei	2	186	0	444	16	0	277	4	0	370	26,79
1prc_Cm	8	953	0	367	36	0	519	5	0	449	88,69
1prc_Ml	6	500	2	335	34	2	156	6	0	98	46,89
1pvc_Ab	8	1520	3	320	228	3	207	7	0	484	47,87
1sgp_Ei	7	152	1	318	18	0	147	3	0	526	44,26
1ubs_Ba	2	631	0	356	46	0	401	6	0	269	37,30
2bbk_Hl	7	616	3	163	89	3	120	6	0	439	53,13
2kai_Bi	5	135	2	155	5	0	387	5	0	99	31,54
2kau_Ca	2	382	1	338	17	0	393	4	0	437	46,06
2ptc_Ei	8	243	3	94	24	0	216	7	0	87	36,06
3sic_Ei	4	255	1	395	39	0	65	6	0	171	11,19
4htc_Hi	9	300	2	44	31	0	75	6	0	95	23,84
4sgb_Ei	4	128	0	309	12	0	169	3	0	504	10,50
8ruc_Ek	3	768	2	313	111	0	78	6	0	443	43,36
1apy_Ab	40	373	11	23	86	11	1	23	11	1	2,44
1bpl_Ba	27	885	10	185	122	10	112	13	10	52	0,97
1gua_Ab	7	222	3	438	21	3	313	9	3	14	0,78
1htr_Bp	26	189	10	65	24	10	2	10	10	1	2,00
1mmo_Ec	7	1234	2	351	70	2	452	6	2	300	1,53
1pya_Dc	47	284	23	42	73	23	1	36	23	1	1,09
1tab_Ei	5	152	2	209	14	2	159	5	2	159	2,10
2kau_Cb	24	625	8	186	67	8	167	11	8	25	1,08
2pka_Ba	23	271	8	15	32	8	42	8	8	28	1,27

nicht gerechnet: 1req Cd

Paar-Docking mit dem mittleren Grenzwertensemble und ungesättigten Atomen											
	# Referenzen	Rotationscluster			Ergebnisausdünnung			Transformationscluster			trans [Å]
		# hits	# ref	rank	# hits	# ref	rank	# hits	# ref	rank	
1acb_Ei	5	122	2	408	3	0	399	3	0	399	24,44
1cgj_Ei	5	98	2	492	6	0	214	3	0	409	27,95
1cho_Ei	6	103	3	456	7	3	220	4	0	192	40,35
1fcb_Ab	4	1147	0	251	87	0	249	9	0	73	33,65
1fin_Ab	4	430	0	142	31	0	82	5	0	179	56,04
1got_Ab	9	776	1	135	69	0	142	6	0	274	46,63
1hbh_Dc	2	82	1	249	3	0	195	3	0	195	18,50
1mee_Ai	4	158	0	64	6	0	299	3	0	449	33,46
1mmo_Ec	7	641	0	228	9	0	510	3	0	540	3,76
1mmo_Eh	11	405	3	356	18	3	213	4	0	323	8,01
1ppf_Ei	2	112	0	426	7	0	263	4	0	215	26,79
1prc_Cm	8	508	0	463	7	0	508	4	0	374	77,29
1pvc_Ab	8	914	3	405	66	3	373	6	0	302	47,47
1sgp_Ei	7	80	1	402	9	0	126	3	0	426	44,26
1ubs_Ba	2	368	0	284	12	0	427	6	0	62	37,30
2bbk_Hl	7	395	3	165	41	3	70	5	0	295	48,64
2kai_Bi	5	100	2	101	5	0	242	5	0	59	31,54
2kau_Ca	2	224	1	161	6	0	379	3	0	495	54,58
2ptc_Ei	8	134	3	140	6	0	298	6	0	28	35,84
3sic_Ei	4	154	1	342	16	0	74	6	0	55	11,19
4htc_Hi	9	209	1	149	16	0	110	4	0	262	30,62



4sgb_Ei	4	87	0	302	6	0	172	3	0	362	23,83
8ruc_Ek	3	478	0	344	61	0	20	5	0	347	31,62
1apy_Ab	40	233	9	8	51	9	2	21	9	1	2,75
1bpl_Ba	27	471	10	199	54	10	8	13	10	3	0,97
1gua_Ab	7	155	3	322	15	3	165	9	3	9	0,78
1htr_Bp	26	132	10	50	16	10	2	10	10	1	2,00
1prc_Ml	6	291	2	286	7	2	315	4	2	214	0,81
1pya_Dc	47	187	23	42	52	23	1	32	23	1	0,96
1req_Cd	18	1725	2	260	123	2	417	11	2	49	2,05
1tab_Ei	5	106	2	231	9	2	125	3	2	424	1,61
2kau_Cb	24	347	4	141	26	4	53	6	4	64	1,26
2pka_Ba	23	188	8	6	31	8	1	8	8	6	1,27

### Paar-Docking mit dem mittleren Grenzwertensemble und ungebundenen Atomen

	# Referenzen	Rotationscluster			Ergebnisausdünnung			Transformationscluster			trans [Å]
		# hits	# ref	rank	# hits	# ref	rank	# hits	# ref	rank	
1acb_Ei	5	97	2	163	0	0	564	0	0	564	20,89
1cgj_Ei	5	66	2	428	6	0	77	3	0	229	27,95
1fcb_Ab	4	790	0	213	38	0	292	7	0	61	33,49
1fin_Ab	4	266	0	340	10	0	258	4	0	243	43,98
1got_Ab	9	160	0	144	0	0	564	0	0	564	20,89
1hbh_Dc	2	51	0	239	3	0	134	3	0	134	18,50
1mee_Ai	4	122	0	50	6	0	191	3	0	386	33,46
1mmo_Ec	7	206	0	424	0	0	564	0	0	564	20,89
1mmo_Eh	11	164	0	228	0	0	564	0	0	564	20,89
1ppf_Ei	2	70	0	431	0	0	564	0	0	564	20,89
1prc_Cm	8	251	0	401	0	0	564	0	0	564	20,89
1pvc_Ab	8	101	1	397	0	0	564	0	0	564	20,89
1sgp_Ei	7	52	1	265	6	0	85	3	0	249	44,26
1ubs_Ba:	2	206	0	363	3	0	479	3	0	479	45,02
2bbk_Hl	7	303	3	74	31	3	37	4	0	374	53,24
2kai_Bi	5	87	2	74	5	0	179	5	0	43	31,54
2kau_Ca	2	133	1	192	0	0	564	0	0	564	20,89
2kau_Cb	24	219	2	160	6	0	352	3	0	485	39,11
2ptc_Ei	8	74	3	159	1	0	337	0	0	564	20,89
3sic_Ei	4	122	1	180	16	0	24	6	0	35	11,19
4htc_Hi	9	140	1	219	13	0	40	4	0	144	30,62
4sgb_Ei	4	51	0	439	0	0	564	0	0	564	20,89
8ruc_Ek	3	150	0	271	3	0	275	3	0	275	41,98
1apy_Ab	40	184	9	2	48	9	1	21	9	1	2,75
1bpl_Ba	27	315	10	274	32	10	10	13	10	2	0,97
1cho_Ei	6	53	3	491	3	3	229	3	3	229	2,26
1gua_Ab	7	109	3	292	9	3	166	9	3	6	0,78
1htr_Bp	26	99	10	17	13	10	1	10	10	1	2,00
1prc_Ml	6	166	2	307	7	2	99	4	2	89	0,81
1pya_Dc	47	150	19	19	42	19	1	28	19	1	0,86
1req_Cd	18	1102	2	203	76	2	63	7	2	74	2,00
1tab_Ei	5	82	2	211	9	2	51	3	2	343	1,61
2pka_Ba	23	144	8	14	22	8	2	8	8	2	1,27

**Triplett-Docking mit dem mittleren Grenzwertensemble und allen Atomen**

	# Referenzen	Rotationscluster			Ergebnisausdünnung			Transformationscluster			trans [Å]
		# hits	# ref	rank	# hits	# ref	rank	# hits	# ref	rank	
1bpl_Ba	6	81	3	31	29	3	30	4	0	165	50,96
1apy_Ab	20	66	10	1	40	10	1	9	6	1	1,56
1htr_Bp	16	27	11	1	13	11	1	11	11	1	3,30
1pya_Dc	21	55	11	1	43	11	1	14	7	1	3,57
2kau_Cb	9	48	5	39	15	4	44	4	4	67	1,72
2pka_Ba	8	25	7	37	10	7	15	6	6	4	2,74

**Triplett-Docking mit dem mittleren Grenzwertensemble und ungesättigten Atomen**

	# Referenzen	Rotationscluster			Ergebnisausdünnung			Transformationscluster			trans [Å]
		# hits	# ref	rank	# hits	# ref	rank	# hits	# ref	rank	
1apy_Ab	20	38	7	1	28	7	1	4	0	9	6,75
1bpl_Ba	6	38	3	8	11	3	9	3	3	101	2,88
1htr_Bp	16	23	11	1	13	11	1	11	11	1	3,30
1pya_Dc	21	46	11	1	34	11	1	10	4	1	2,06
2kau_Cb	9	20	4	85	7	3	28	3	3	60	2,09
2pka_Ba	8	14	6	1	7	6	1	5	5	1	3,02

**Triplett-Docking mit dem mittleren Grenzwertensemble und ungebundenen Atomen**

	# Referenzen	Rotationscluster			Ergebnisausdünnung			Transformationscluster			trans [Å]
		# hits	# ref	rank	# hits	# ref	rank	# hits	# ref	rank	
1apy_Ab	20	42	9	1	30	9	1	6	5	3	1,73
1bpl_Ba	6	25	3	16	9	3	6	3	3	53	2,88
1htr_Bp	16	18	11	1	13	11	1	11	11	1	3,30
1pya_Dc	21	33	8	1	23	8	1	9	4	1	2,01
2kau_Cb	9	14	4	52	5	3	27	3	3	42	2,09
2pka_Ba	8	18	7	1	8	7	1	6	6	1	2,74

## A9.3 Docking mit weitem Grenzwertensemble

Paar-Docking mit dem weiten Grenzwertensemble und allen Atomen											
	# Referenzen	Rotationscluster			Ergebnisausdünnung			Transformationscluster			trans [Å]
		# hits	# ref	rank	# hits	# ref	rank	# hits	# ref	rank	
1acb_Ei	14	2183	4		1460	4		360	0		23,40
1brs_Cf	11	921	1		437	0		78	0		35,04
1cgj_Ei	14	1944	6		1166	6		329	0		23,09
1cho_Ei	20	1859	7		1157	6		242	0		36,72
1got_Bg	6	1426	1		446	1		44	0		16,16
1gua_Ab	23	2235	6		1524	6		565	0		24,22
1hbh_Dc	10	1544	0		701	0		31	0		42,51
1hcg_Ab	11	1986	4		1195	4		508	0		35,45
1lta_Ec	2	675	0		265	0		38	0		34,37
1lts_Ac	2	1076	0		456	0		41	0		31,08
1mda_Hl	2	5149	1		3866	1		2672	0		27,69
1mmo_Ch	6	4558	1		2426	0		268	0		28,93
1ppf_Ei	8	1898	2		1194	2		325	0		25,80
1prc_Ml	38	4576	8		2485	8		228	0		15,74
1scu_Ed	3	5516	0		3250	0		1288	0		29,01
1sgp_Ei	15	1303	4		741	4		109	0		34,72
1slu_Ba	6	4843	0		3690	0		2655	0		36,83
1smp_Ai	7	5655	0		4082	0		2363	0		22,96
1stf_Ei	3	2757	2		1794	2		356	0		28,63
1tab_Ei	10	1254	2		721	2		168	0		22,10
1taf_Ba	6	303	4		82	4		14	0		21,32
1tii_Ac	2	563	1		204	0		24	0		21,13
1wdc_Ca	6	573	0		151	0		14	0		52,01
1yrn_Ba	4	225	1		69	0		16	0		32,05
2kai_Bi	11	1087	2		579	2		73	0		23,03
2kau_Ca	13	3864	4		2152	4		307	0		42,02
2pka_Ba	64	2074	24		1301	24		403	0		20,44
3sic_Ei	20	2425	4		1458	4		492	0		39,50
4sgb_Ei	10	1089	2		571	2		86	0		24,42
1apy_Ab	68	2796	18	118	1785	18	136	852	18	116	20,56
1htr_Bp	41	1355	17	40	491	17	69	30	17	220	2,03
1mee_Ai	14	2044	4	78	1274	4	27	188	4	157	9,83
1pya_Dc	119	2109	52	74	1127	51	39	128	51	97	2,27
1pyt_Ca	7	2732	3	179	1726	2	70	459	2	85	17,72
2ptc_Ei	18	1804	5	269	1115	5	283	184	5	271	12,16
4htc_Hi	20	2375	9	87	1452	9	78	512	9	52	12,64

nicht gerechnet: 1abr\_Ba, 1bpl\_Ba, 1fcb\_Ab, 1fin\_Ab, 1got\_Ab, 1lpb\_Ba, 1mmo\_Ec, 1mmo\_Eh, 1prc\_Cl, 1prc\_Cm, 1pvc\_Ab, 1req\_Cd, 1ubs\_Ba, 1vrt\_Ab, 2bbk\_Hl, 2kau\_Cb, 4aah\_Cd, 8atc\_Ab, 8ruc\_Ek

Paar-Docking mit dem weiten Grenzwertensemble und ungesättigten Atomen											
	# Referenzen	Rotationscluster			Ergebnisausdünnung			Transformationscluster			trans [Å]
		# hits	# ref	rank	# hits	# ref	rank	# hits	# ref	rank	
1abr_Ba	7	5144	2		3135	2		1284	0		56,33
1acb_Ei	14	1108	3		451	3		19	0		23,88
1apy_Ab	68	1534	14		688	14		115	0		12,20
1bpl_Ba	74	4394	23		2584	23		618	0		23,97
1brs_Cf	11	562	1		162	0		15	0		39,89
1cgj_Ei	14	996	6		368	6		30	0		27,93
1cho_Ei	20	980	7		413	6		121	0		36,95
1fin_Ab	11	3952	0		2072	0		226	0		49,26
1got_Bg	6	838	1		192	1		13	0		16,42
1gua_Ab	23	1392	6		717	6		89	0		34,40
1hbh_Dc	10	802	0		235	0		26	0		42,44
1hcg_Ab	11	1309	2		617	0		62	0		38,99
1lpb_Ba	14	2630	2		1299	2		66	0		22,98
1lta_Ec	2	413	0		109	0		23	0		35,17
1lts_Ac	2	697	0		222	0		22	0		36,31
1mda_Hl	2	3458	1		2072	1		969	0		28,16

1mee_Ai	14	1183	2	538	1	26	0	47,34	
1mmo_Ch	6	2636	1	975	0	35	0	20,17	
1ppf_Ei	8	1173	2	509	2	41	0	30,51	
1prc_Cl	16	4386	3	2185	3	272	0	55,70	
1prc_Cm	27	4878	5	2127	4	343	0	58,56	
1prc_Ml	38	2572	6	993	6	40	0	46,61	
1pyt_Ca	7	1552	2	679	0	53	0	31,80	
1scu_Ed	3	3004	0	1143	0	182	0	24,83	
1sgp_Ei	15	681	3	233	3	55	0	46,55	
1slu_Ba	6	2850	0	1692	0	465	0	30,24	
1smp_Ai	7	3049	0	1500	0	220	0	15,39	
1stf_Ei	3	1685	0	767	0	109	0	36,24	
1tab_Ei	10	831	2	360	2	23	0	27,29	
1taf_Ba	6	200	4	57	4	7	0	21,50	
1tii_Ac	2	269	0	58	0	9	0	19,84	
1ubs_Ba	14	3730	3	2093	3	514	0	33,21	
1wdc_Ca	6	409	0	87	0	11	0	51,27	
1yrn_Ba	4	151	1	33	0	8	0	28,73	
2kai_Bi	11	773	2	312	0	29	0	35,27	
2kau_Ca	13	2000	2	676	2	17	0	54,72	
2kau_Cb	48	3152	14	1588	14	97	0	60,97	
2ptc_Ei	18	984	4	415	4	29	0	32,99	
3sic_Ei	20	1367	4	583	4	72	0	52,90	
4aah_Cd	22	2428	3	908	3	70	0	36,58	
4sgb_Ei	10	745	1	306	0	40	0	25,43	
8atc_Ab	6	4081	0	2545	0	927	0	45,96	
8ruc_Ek	12	4677	2	2833	1	1390	0	34,75	
1htr_Bp	41	919	17	232	17	25	17	40	1,88
1mmo_Eh	23	4133	6	1876	6	599	6	1	10,98
1pya_Dc	119	1314	48	542	48	87	48	2	1,05
2pka_Ba	64	1372	24	653	24	61	24	184	3,26

nicht gerechnet: 1fcb Ab, 1got Ab, 1mmo Ec, 1pvc Ab, 1req Cd, 1vrt Ab, 2bbk Hl, 4htc Hi

### Paar-Docking mit dem weiten Grenzwertensemble und ungebundenen Atomen

	# Referenzen	Rotationscluster			Ergebnisausdünnung			Transformationscluster			trans [Å]
		# hits	# ref	rank	# hits	# ref	rank	# hits	# ref	rank	
1abr_Ba	7	3992	0	36	2126	0	75	566	0	7	63,66
1acb_Ei	14	724	3	144	222	3	98	15	0	214	34,93
1bpl_Ba	74	2983	23	191	1375	23	181	85	0	259	16,95
1brs_Cf	11	314	1	105	70	0	154	10	0	162	32,08
1cgj_Ei	14	656	6	207	161	6	325	17	0	168	28,94
1cho_Ei	20	570	7	142	168	6	104	23	0	37	42,22
1fin_Ab	11	2569	0	100	1001	0	113	113	0	11	46,46
1got_Ab	22	4357	1	181	2240	1	109	278	0	287	37,10
1got_Bg	6	581	0	178	127	0	95	9	0	382	26,14
1hbh_Dc	10	440	0	178	91	0	286	10	0	310	17,34
1hcg_Ab	11	913	2	251	334	0	323	39	0	192	31,82
1lta_Ec	2	262	0	63	65	0	46	10	0	136	34,16
1lts_Ac	2	473	0	121	110	0	113	12	0	141	50,04
1mda_Hl	2	2725	1	128	1455	1	140	356	0	181	31,83
1mee_A	14	845	2	65	312	1	48	19	0	227	47,07
1mmo_Ch	6	1670	1	200	408	0	93	20	0	77	20,39
1mmo_Ec	50	3405	8	214	830	7	348	20	0	350	15,28
1ppf_Ei	8	718	2	297	228	2	383	16	0	424	46,71
1prc_Cl	16	2555	3	23	842	3	21	57	0	16	34,93
1prc_Cm	27	2920	5	254	801	4	349	48	0	70	58,87
1prc_Ml	38	1424	6	72	351	6	29	13	0	295	49,06
1pyt_Ca	7	1077	2	142	346	0	191	36	0	31	42,34
1scu_Ed	3	1982	0	122	602	0	167	34	0	204	17,40
1sgp_Ei	15	387	3	171	113	3	164	30	0	14	46,66
1slu_Ba	6	2117	0	90	1077	0	58	196	0	23	30,92
1smp_Ai	7	1862	0	190	628	0	189	33	0	164	45,17
1stf_Ei	3	1069	0	217	360	0	201	22	0	218	44,37
1tab_Ei	10	629	2	150	224	2	121	15	0	374	26,66
1taf_Ba	6	137	4	37	36	4	3	5	0	181	24,17

1tii_Ac	2	125	0	50	16	0	96	4	0	271	20,82
1ubs_Ba	14	2226	3	45	905	3	17	59	0	180	30,34
1wdc_Ca	6	216	0	246	38	0	79	4	0	478	42,42
1yrn_Ba	4	81	1	207	0	0	564	0	0	564	20,89
2kai_Bi	11	609	2	72	227	0	58	21	0	170	35,99
2kau_Ca	13	1130	2	147	259	2	122	9	0	428	44,10
2kau_Cb	48	1740	9	259	604	9	71	30	0	221	29,09
2ptc_Ei	18	496	4	250	135	4	201	11	0	287	33,69
3sic_Ei	20	939	4	49	331	4	80	33	0	163	49,22
4aah_Cd	22	1665	1	152	472	1	265	30	0	216	38,29
4htc_Hi	20	1134	6	134	450	6	65	32	0	152	18,49
4sgb_Ei	10	482	1	241	157	0	134	11	0	327	25,15
8atc_Ab	6	2816	0	94	1427	0	80	111	0	192	65,54
8ruc_Ek	12	3258	2	172	1729	1	171	727	0	125	37,91
1apy_Ab	68	998	14	67	356	14	49	41	14	38	1,61
1gua_Ab	23	965	6	61	432	6	23	71	6	21	3,28
1htr_Bp	41	644	17	11	140	17	20	22	17	4	1,87
1lpb_Ba	14	1536	2	112	600	2	41	35	2	161	6,58
1mmo_Eh	23	2573	6	75	821	6	118	125	6	2	13,72
1pya_Dc	119	976	44	83	351	44	15	82	44	1	1,06
2bbk_Hl	21	2483	8	66	1244	8	92	155	8	298	11,73
2pka_Ba	64	1016	23	119	360	23	152	27	23	139	1,32

nicht gerechnet: 1fcb\_Ab, 1pvc\_Ab, 1req\_Cd, 1vrt\_Ab

### Triplett-Docking mit dem weiten Grenzwertensemble und allen Atomen

	# Referenzen	Rotationscluster			Ergebnisausdünnung			Transformationscluster			trans [Å]
		# hits	# ref	rank	# hits	# ref	rank	# hits	# ref	rank	
1acb_Ei	7	2002	2		1844	2		418	0		26,66
1cgj_Ei	5	1580	2		1421	2		508	0		24,80
1cho_Ei	11	2020	3		1881	3		890	0		35,58
1gua_Ab	14	2145	4		1983	4		711	0		35,75
1hbh_Dc	3	675	1		488	1		36	0		42,84
1hcg_Ab	4	1621	1		1477	1		743	0		36,04
1htr_Bp	36	747	18		531	18		30	0		13,21
1mee_Ai	3	1888	0		1713	0		226	0		33,89
1ppf_Ei	2	1683	1		1539	1		467	0		36,34
1prc_Ml	8	2975	2		2555	2		245	0		46,10
1sgp_Ei	9	1514	3		1401	3		723	0		35,03
1tab_Ei	5	1295	1		1190	1		366	0		17,26
1ubs_Ba	5	4137	1		3718	1		1065	0		34,26
2kai_Bi	2	825	0		699	0		121	0		18,98
2kau_Ca	3	2410	3		2010	3		144	0		43,28
2pka_Ba	39	1674	25		1504	25		235	0		29,92
2ptc_Ei	8	1774	2		1628	2		379	0		26,89
3sic_Ei	10	2370	2		2146	2		1021	0		38,93
4aah_Cd	9	2650	2		2231	2		132	0		20,24
4sgb_Ei	4	1191	0		1087	0		315	0		26,53
1apy_Ab	44	2411	25	108	2203	25	106	530	25	174	8,32
1pya_Dc	95	1443	50	83	1222	50	78	219	50	13	3,70

nicht gerechnet: 1bpl\_Ba, 1fcb\_Ab, 1fin\_Ab, 1got\_Ab, 1lpb\_Ba, 1mmo\_Ec, 1mmo\_Eh, 1prc\_Cl, 1prc\_Cm, 1pvc\_Ab, 1req\_Cd, 2bbk\_Hl, 2kau\_Cb, 4htc\_Hi, 8ruc\_Ek

**Triplett-Docking mit dem weiten Grenzwertensemble und ungesättigten Atomen**

	# Referenzen	Rotationscluster			Ergebnisausdünnung			Transformationscluster			trans [Å]
		# hits	# ref	rank	# hits	# ref	rank	# hits	# ref	rank	
1acb_Ei	7	729	2		583	2		44	0		36,88
1bpl_Ba	48	3266	19		2861	19		390	0		42,27
1cgj_Ei	5	603	2		434	2		57	0		32,18
1cho_Ei	11	839	3		719	3		309	0		37,94
1fin_Ab	3	2366	0		1978	0		361	0		47,89
1gua_Ab	14	1162	4		1009	4		376	0		35,49
1hbh_Dc	3	220	1		123	0		10	0		44,39
1hcg_Ab	4	842	0		692	0		93	0		43,19
1lpb_Ba	7	1897	0		1620	0		121	0		23,57
1mee_Ai	3	818	0		660	0		59	0		35,32
1mmo_Ec	16	2885	2		2099	2		44	0		40,56
1mmo_Eh	8	2297	4		1812	4		98	0		22,68
1ppf_Ei	2	930	1		801	1		213	0		39,12
1prc_Cl	5	2462	2		2004	2		93	0		55,36
1prc_Cm	7	2408	0		1799	0		53	0		97,21
1prc_Ml	8	1377	1		1050	1		66	0		47,43
1sgp_Ei	9	661	3		581	3		151	0		44,60
1tab_Ei	5	718	0		594	0		81	0		19,89
1ubs_Ba	5	1750	1		1390	1		99	0		20,26
2bbk_Hl	8	2674	1	143	2359	1	139	699	1	140	18,45
2kai_Bi	2	459	0		364	0		89	0		31,06
2kau_Ca	3	906	1		605	1		50	0		52,41
2kau_Cb	31	1937	9		1604	9		152	0		51,35
2ptc_Ei	8	731	2		595	2		58	0		35,06
3sic_Ei	10	1061	2		895	1		245	0		34,77
4aah_Cd	9	1270	0		927	0		55	0		16,62
4htc_Hi	15	1353	5		1188	5		183	0		17,50
4sgb_Ei	4	747	0		648	0		180	0		27,07
8ruc_Ek	5	3495	0		3031	0		1323	0		39,16
1apy_Ab	44	917	20	113	746	20	100	137	20	36	5,38
1htr_Bp	36	350	18	112	206	18	126	20	18	53	3,02
1pya_Dc	95	720	44	96	553	44	77	91	44	5	2,32
2pka_Ba	39	875	25	124	710	25	126	75	25	119	2,21

nicht gerechnet: 1fcb Ab, 1got Ab, 1pvc Ab, 1req Cd

**Triplett-Docking mit dem weiten Grenzwertensemble und ungebundenen Atomen**

	# Referenzen	Rotationscluster			Ergebnisausdünnung			Transformationscluster			trans [Å]
		# hits	# ref	rank	# hits	# ref	rank	# hits	# ref	rank	
1acb_Ei	7	413	2	97	308	2	88	30	0	64	39,78
1apy_Ab	44	562	19	63	438	19	50	37	0	86	20,74
1bpl_Ba	48	1728	19	148	1374	19	144	94	0	163	66,54
1cgj_Ei	5	305	2	183	198	2	184	18	0	171	27,51
1cho_Ei	11	332	3	121	248	3	105	30	0	107	37,09
1fin_Ab	3	1217	0	105	898	0	102	73	0	33	46,63
1gua_Ab	16	2212	2	152	1731	2	158	138	0	85	39,66
1hcg_Ab	4	536	0	160	400	0	176	29	0	219	41,11
1lpb_Ba	7	1042	0	41	832	0	31	84	0	52	24,13
1mee_Ai	3	556	0	103	427	0	104	53	0	83	35,47
1mmo_Ec	16	1226	0	134	744	0	136	20	0	194	67,38
1mmo_Eh	8	1216	4	72	856	4	70	29	0	141	6,83
1ppf_Ei	2	442	1	188	346	1	187	40	0	213	44,35
1prc_Cl	5	1143	2	96	822	2	81	64	0	19	35,23
1prc_Cm	7	1135	0	137	721	0	130	19	0	161	103,53
1prc_Ml	8	648	1	73	440	1	64	40	0	16	48,31
1sgp_Ei	9	303	3	87	254	3	60	79	0	16	45,74
1tab_Ei	5	446	0	111	338	0	122	63	0	106	19,60
1ubs_Ba	5	838	1	88	593	1	79	68	0	29	20,15
2bbk_Hl	8	1630	1	155	1355	1	150	162	0	206	17,35

2kai_Bi	2	357	0	113	273	0	105	67	0	49	30,37
2kau_Ca	3	420	1	117	258	1	95	13	0	151	47,38
2kau_Cb	31	917	6	87	673	6	86	56	0	98	48,83
2ptc_Ei	8	285	2	140	210	2	126	10	0	293	17,50
3sic_Ei	10	632	2	99	502	1	89	152	0	54	32,79
4aah_Cd	9	843	0	128	579	0	131	21	0	227	17,64
4sgb_Ei	4	373	0	75	300	0	58	39	0	24	27,72
8ruc_Ek	5	2300	0	89	1952	0	92	863	0	76	39,56
1htr_Bp	36	232	18	47	131	18	39	20	18	5	3,02
1pya_Dc	95	510	40	93	374	40	75	77	40	4	2,27
2pka_Ba	39	587	22	90	447	22	88	46	22	54	1,05
4htc_Hi	15	759	5	82	604	5	74	26	5	233	5,51
nicht gerechnet:		1fcb Ab, 1pvcA b, 1req Cd									

## A10 Ab2/Ab3-Werte, Evaluierungsaufrufe, Kandidatenzahlen und Dockinglaufzeiten

Alle Angaben der folgenden Tabelle beziehen sich auf die angegebenen obligatorischen Proteinkomplexe. "Ab2" und "Ab3" bezeichnet die Ab2- bzw. Ab3-Werte, die sich aus den kriterienkonformen Paar- bzw. Tripletzzusammenstellungen bindungsfähiger Atome zur Molekülrepräsentation berechnen lassen. "Aufrufe" und "Kandidaten" entsprechen den im Docking zählbaren Evaluierungsaufrufen bimolekularer transformations-invarianter Eigenschaften und der Anzahl hieraus hervorgehender, kriterienkonformer Kandidaten. "CPU" bezeichnet die jeweils benötigte Rechenzeit, die auf einem R10000-Prozessor einer SGI<sup>®</sup>-Origin für den Dockingvorgang benötigt wurde.

Grauschattierungen weisen auf die Komplexe hin, deren Rechnungen nicht durchgeführt wurden, da ihre Ab2/Ab3-Werte und Evaluierungsaufrufe derart große Laufzeiten erwarten lassen, dass sie im Rahmen dieser Untersuchungen nicht weiter von Interesse sind.

		Paar-Docking				Triplet-Docking			
		Ab2	Aufrufe	Kand.	CPU [s]	Ab3	Aufrufe	Kand.	CPU [s]
enge Grenzen	lab9_Bc	722.176	439.298	90.156	17,08	803.049	232.392	4.579	1,93
	laks_Ab	783.636	481.606	105.165	20,74	875.790	260.636	5.061	2,14
alle Atome	lept_Ba	236.964	144.230	34.023	5,28	224.112	70.863	1.917	0,70
	lept_Ca	200.074	122.988	29.243	4,46	157.458	49.302	1.153	0,45
	lept_Bc	516.516	313.624	67.370	11,70	464.928	139.196	3.224	1,22
	lhuc_Ba	414.100	253.104	47.335	7,68	297.375	82.501	1.604	0,63
	lkvd_Ba	336.960	204.156	39.316	6,23	430.950	105.562	2.826	0,93
	llya_Ba	1.062.870	661.928	135.129	29,35	990.333	291.858	6.653	2,42
	lmhl_Ca	2.826.310	1.705.586	286.645	88,72	2.463.330	620.901	8.082	3,64
	lpnk_Ba	3.685.776	2.288.018	422.069	166,36	3.168.564	937.309	13.297	6,04
	lspb_Sp	655.140	408.164	81.948	14,95	578.088	179.357	2.869	1,24
	lwht_Ab	1.475.472	892.644	177.488	42,51	1.290.396	332.174	5.900	2,33
	2hnt_Fc	289.920	176.536	32.124	4,94	194.670	51.492	960	0,38
	2kin_Ab	1.262.210	787.458	155.454	36,33	1.267.500	358.675	6.326	2,57
2ltn_Ab	849.000	512.574	124.193	28,64	1.045.440	285.519	12.813	4,03	
2pcd_Nb	2.238.486	1.379.036	252.158	72,21	1.926.624	522.331	7.866	3,33	
enge Grenzen	lab9_Ba	404.460	245.160	47.582	7,82	312.276	85.858	2.183	0,83
	laks_Ab	535.410	328.864	69.881	12,35	472.068	140.580	3.116	1,21
unge-sättigte Atome	lept_Ba	178.794	109.020	26.595	4,04	160.701	50.034	1.514	0,53
	lept_Ca	136.080	83.408	20.145	3,01	90.420	26.973	825	0,29
	lept_Bc	350.966	214.118	45.718	7,38	266.271	76.614	2.422	0,79
	lhuc_Ba	259.852	158.890	29.665	4,51	136.500	39.585	944	0,35
	lkvd_Ba	269.940	163.610	31.699	4,92	295.932	72.173	2.243	0,69
	llya_Ba	702.180	434.596	87.569	16,83	493.119	131.120	4.063	1,26
	lmhl_Ca	1.918.280	1.169.776	193.474	48,23	1.368.900	347.411	4.878	2,11
	lpnk_Ba	2.293.928	1.441.662	259.670	75,40	1.478.217	419.291	7.391	2,92
	lspb_Sp	368.628	228.722	45.254	7,28	223.896	70.168	1.327	0,55
	lwht_Ab	913.662	553.798	106.804	21,16	564.792	141.446	2.720	1,06
	2hnt_Fc	216.544	131.172	23.872	3,58	117.096	29.767	672	0,26
	2kin_Ab	754.220	462.440	91.760	17,43	560.748	147.943	3.323	1,22
2ltn_Ab	633.600	382.832	93.126	19,02	668.028	175.025	9.748	2,83	
2pcd_Nb	1.384.056	858.136	158.308	36,55	812.565	222.993	3.906	1,57	
enge Grenzen	lab9_Bc	299.040	180.984	36.996	5,96	222.870	59.653	1.705	0,61
	laks_Ab	358.308	219.268	48.681	8,04	265.230	76.617	1.969	0,75
ungebundene Atome	lept_Ba	147.350	89.750	21.954	3,30	117.990	35.979	1.216	0,40
	lept_Ca	98.000	59.856	14.777	2,19	53.475	15.231	596	0,20
	lept_Bc	239.970	145.286	32.776	5,07	163.134	43.791	1.904	0,56
	lhuc_Ba	186.690	113.792	22.538	3,37	88.008	25.500	710	0,26
	lkvd_Ba	190.440	114.242	22.976	3,48	178.716	42.465	1.553	0,46
	llya_Ba	550.080	335.054	71.246	12,92	352.512	91.146	3.133	0,97
	lmhl_Ca	1.239.758	752.836	127.058	26,74	797.202	200.302	2.834	1,26
	lpnk_Ba	1.427.490	886.778	168.638	39,48	820.050	238.074	4.636	1,76
	lspb_Sp	245.032	151.592	31.998	4,96	137.295	45.369	887	0,37
	lwht_Ab	589.612	353.822	71.146	12,52	314.826	76.968	1.666	0,64
	2hnt_Fc	180.224	108.232	19.872	2,97	89.856	22.405	562	0,21
	2kin_Ab	488.920	298.626	61.765	10,85	338.688	90.592	2.288	0,79
2ltn_Ab	519.672	311.668	78.900	15,93	538.785	143.603	8.713	2,51	
2pcd_Nb	903.312	553.690	105.026	20,71	467.166	129.079	2.475	0,97	
mittlere Grenzen	lab9_Bc	1.379.044	681.346	134.227	29,61	3.150.252	479.264	8.728	3,96
	laks_Ab	1.488.940	738.830	168.520	40,74	3.238.386	498.217	10.369	4,52
alle Atome	lept_Ba	453.604	217.332	48.648	7,96	949.440	143.412	3.455	1,27
	lept_Ca	398.036	194.242	45.693	7,42	734.784	115.494	2.571	1,05
	lept_Bc	1.010.846	485.644	105.196	20,88	2.003.760	296.018	6.275	2,55
	lhuc_Ba	768.240	373.858	73.052	12,95	1.179.216	158.788	3.503	1,30
	lkvd_Ba	569.050	273.094	51.745	8,76	1.391.940	180.695	4.629	1,59
	llya_Ba	1.901.036	953.402	202.936	53,25	3.155.106	492.017	12.313	4,59
	lmhl_Ca	5.287.300	2.627.718	458.519	199,53	9.282.000	1.280.507	19.744	8,60
	lpnk_Ba	6.553.210	3.275.280	642.796	358,49	10.233.489	1.596.420	26.870	12,50
	lspb_Sp	1.177.820	599.378	125.279	26,21	2.001.474	329.718	6.332	2,65
	lwht_Ab	2.783.388	1.375.486	285.802	90,00	4.583.040	643.086	13.752	5,22
	2hnt_Fc	577.818	292.794	52.194	8,58	782.694	110.031	1.902	0,80
	2kin_Ab	2.506.212	1.271.542	248.887	72,93	4.969.221	752.775	13.356	5,63
2ltn_Ab	1.581.020	739.694	175.720	48,05	4.055.163	541.376	25.675	9,35	
2pcd_Nb	4.146.912	2.050.136	395.069	154,56	6.284.850	886.313	15.678	6,44	
mittlere Grenzen	lab9_Bc	736.064	359.344	67.645	12,07	1.028.355	141.355	3.519	1,32
	laks_Ab	984.600	485.426	107.053	21,47	1.637.964	242.247	5.572	2,21
unge-sättigte Atome	lept_Ba	342.432	164.764	37.684	5,93	628.650	95.724	2.572	0,93
	lept_Ca	263.610	128.088	29.872	4,57	382.800	58.317	1.652	0,61
	lept_Bc	676.992	326.608	69.705	12,31	1.078.992	153.205	4.109	1,45
	lhuc_Ba	495.040	238.724	46.690	7,54	576.450	73.255	2.015	0,69
	lkvd_Ba	445.480	215.114	40.409	6,54	907.920	117.029	3.449	1,13
	llya_Ba	1.221.890	613.374	127.105	27,20	1.524.915	224.754	7.175	2,31
	lmhl_Ca	3.525.472	1.740.390	299.686	95,84	5.039.400	680.741	11.052	4,68
	lpnk_Ba	4.049.910	2.033.892	384.769	143,13	4.937.112	731.677	14.638	5,84
	lspb_Sp	654.728	332.290	69.504	12,16	774.090	125.754	2.873	1,10
	lwht_Ab	1.699.204	835.372	172.884	41,84	2.050.698	288.230	6.473	2,41
	2hnt_Fc	437.250	221.182	39.778	6,28	490.500	65.277	1.387	0,53
	2kin_Ab	1.491.682	748.510	143.297	31,87	2.197.800	316.364	6.745	2,55
2ltn_Ab	1.178.716	551.920	131.186	31,58	2.611.476	330.482	19.124	6,35	
2pcd_Nb	2.610.528	1.289.752	250.612	73,13	2.753.520	377.802	7.892	2,98	



		Paar-Docking				Triplett-Docking			
		Ab2	Aufrufe	Kand.	CPU	Ab3	Aufrufe	Kand.	CPU
mittlere Grenzen	lab9_Bc	534.352	259.354	51.938	8,91	702.204	94.659	2.746	0,98
	laks_Ab	658.878	324.448	73.474	13,27	914.688	131.825	3.560	1,37
	lept_Ba	285.768	137.440	31.746	4,89	484.272	72.005	2.199	0,74
	lept_Ca	191.970	93.346	21.665	3,25	242.136	36.135	1.279	0,42
	lept_Bc	469.224	224.032	49.396	8,23	703.836	95.718	3.240	1,06
	lhuc_Ba	352.418	168.708	34.421	5,30	352.782	44.712	1.337	0,46
	lkvd_Ba	317.356	151.458	29.146	4,53	579.828	73.174	2.406	0,75
	llya_Ba	958.820	480.288	102.037	20,25	1.081.350	160.113	5.540	1,80
	lmhl_Ca	2.314.368	1.136.676	195.272	50,46	3.081.933	415.240	6.392	2,70
	lpnk_Ba	2.552.072	1.280.208	251.622	73,47	2.783.721	426.004	8.789	3,40
	lspb_Sp	430.680	217.078	48.277	7,86	474.777	80.049	1.930	0,74
	lwht_Ab	1.093.820	533.504	112.714	22,65	1.140.360	161.097	3.677	1,36
	2hnt_Fc	362.408	181.602	32.082	4,91	392.805	50.666	1.102	0,43
	2kin_Ab	977.392	484.322	93.437	18,40	1.290.348	187.313	3.979	1,51
2ltn_Ab	955.110	445.796	109.108	24,80	1.999.596	260.636	16.234	5,27	
2pcd_Nb	1.684.096	815.326	165.309	38,27	1.555.587	208.059	4.819	1,73	
weite Grenzen	lab9_Bc	7.236.992	4.285.786	1.013.379	4.386,94	121.625.196	31.870.940	630.234	6.206,83
	laks_Ab	8.244.384	4.926.212	1.336.745	15.062,97	148.143.060	40.189.745	873.686	16.614,14
	lept_Ba	1.845.504	1.056.968	299.487	164,98	20.825.649	5.499.850	143.566	147,59
	lept_Ca	1.774.080	1.032.336	299.019	185,02	20.817.147	5.884.162	131.798	133,50
	lept_Bc	5.081.544	2.988.928	776.736	1.893,65	74.851.821	20.147.003	427.905	1.751,11
	lhuc_Ba	4.208.574	2.469.672	559.917	599,74	47.626.680	11.883.907	239.772	388,76
	lkvd_Ba	2.486.496	1.441.962	330.238	191,97	32.823.180	8.089.476	187.186	250,58
	llya_Ba	11.144.056	6.679.148	1.708.187	27.529,69	144.986.526	39.524.705	949.780	19.588,97
	lmhl_Ca	32.835.192	19.762.116			547.276.644	140.443.104		
	lpnk_Ba	38.185.036	22.711.630			474.223.860	126.875.960		
	lspb_Sp	6.271.020	3.764.802	967.211	3.145,26	83.048.472	23.035.683	558.231	3.749,72
	lwht_Ab	17.133.754	10.264.988			246.940.110	66.682.488		
	2hnt_Fc	3.196.232	1.908.590	412.331	323,53	38.340.273	9.959.003	174.253	207,22
	2kin_Ab	13.868.208	8.349.666			226.853.136	63.374.164		
2ltn_Ab	6.756.420	3.777.388	1.013.460	4.351,10	87.197.715	19.912.023	651.772	7.496,47	
2pcd_Nb	28.113.522	16.842.614			412.034.760	105.588.698			
weite Grenzen	lab9_Bc	3.991.776	2.344.952	505.112	410,66	38.197.836	8.773.359	189.926	255,82
	laks_Ab	5.471.928	3.249.584	836.421	2.588,16	70.973.454	18.100.558	409.419	1.587,75
	lept_Ba	1.349.064	775.702	220.849	84,43	12.687.345	3.333.332	91.519	73,77
	lept_Ca	1.151.280	668.252	186.100	59,71	9.574.803	2.594.894	60.123	39,50
	lept_Bc	3.317.820	1.951.302	479.688	425,82	36.246.210	9.265.354	204.359	287,18
	lhuc_Ba	2.701.152	1.583.502	345.583	188,16	23.907.168	5.611.313	125.031	118,31
	lkvd_Ba	1.919.300	1.111.012	245.708	96,28	20.830.260	4.946.814	122.463	110,33
	llya_Ba	7.400.556	4.427.108	1.058.816	3.253,47	69.917.688	17.889.892	450.372	1.950,87
	lmhl_Ca	21.816.312	13.118.672			286.648.593	72.611.213		
	lpnk_Ba	23.276.506	13.886.852			218.375.850	56.969.911		
	lspb_Sp	3.563.586	2.139.190	524.092	443,24	32.403.270	8.733.293	221.519	324,93
	lwht_Ab	10.423.192	6.244.950	1.541.491	9.585,36	104.749.296	27.646.387	630.900	4.023,29
	2hnt_Fc	2.473.700	1.489.518	317.017	156,37	25.307.520	6.552.334	116.441	102,78
	2kin_Ab	8.345.554	5.009.944	1.176.617	4.641,41	94.676.673	24.872.071	479.994	2.592,86
2ltn_Ab	4.986.846	2.788.938	732.732	1.590,86	50.779.680	11.050.226	410.157	1.697,84	
2pcd_Nb	17.826.660	10.731.114			190.438.950	47.922.680			
weite Grenzen	lab9_Bc	2.797.100	1.637.188	367.779	182,57	22.220.019	4.997.197	121.645	107,67
	laks_Ab	3.556.440	2.116.878	564.055	687,81	35.818.020	8.971.664	225.117	357,12
	lept_Ba	1.103.392	636.892	182.909	52,88	9.008.034	2.349.726	66.991	44,91
	lept_Ca	837.056	486.532	133.816	32,49	5.805.891	1.541.773	37.723	22,08
	lept_Bc	2.187.760	1.285.232	325.018	151,71	18.877.062	4.700.143	118.500	98,12
	lhuc_Ba	1.872.276	1.098.094	257.495	89,85	13.884.822	3.291.583	80.241	52,98
	lkvd_Ba	1.290.814	742.720	172.029	51,09	12.009.108	2.825.496	75.195	46,01
	llya_Ba	5.678.720	3.383.866	832.775	1.443,90	47.713.674	12.315.527	315.055	629,71
	lmhl_Ca	13.413.600	8.033.580	1.700.762	8.152,29	145.251.270	36.764.446	614.787	3.952,48
	lpnk_Ba	13.406.508	7.997.398	1.955.864	7.206,87	106.233.630	28.403.203	614.879	2.590,87
	lspb_Sp	2.142.962	1.270.408	337.658	153,12	16.281.036	4.461.163	128.716	114,90
	lwht_Ab	6.278.026	3.736.542	990.961	1.799,97	50.346.300	13.335.038	332.884	675,63
	2hnt_Fc	2.021.880	1.215.078	257.655	94,34	18.398.505	4.652.799	84.737	60,65
	2kin_Ab	5.239.872	3.115.422	761.928	984,27	48.332.940	12.431.116	251.309	406,01
2ltn_Ab	3.990.168	2.218.096	593.088	765,38	36.874.026	7.939.245	308.738	788,85	
2pcd_Nb	11.304.832	6.769.494	1.582.737	8.628,80	96.094.026	24.183.446	495.659	1.870,19	



## Literaturverzeichnis

- [Ackermann *et al.*, 1995] Ackermann F, Herrmann G, Kummert F, Posch S, Sagerer G, Schomburg D, "Protein docking combining symbolic descriptions of molecular surfaces and grid-based scoring functions", *Ismb* (1995), **3**, 3-11.
- [Baker und Hubbard, 1984] Baker EN and Hubbard RE, "Hydrogen bonding in globular proteins", *Prog. Biophys. Mol. Biol.* (1984) **44**, 97-179.
- [Berman *et al.*, 2000] Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE, "The Protein Data Bank", *Nucleic Acids Research* (2000) **28**, 235-242 <http://www.rcsb.org/pdb/>.
- [Bernstein *et al.*, 1977] Bernstein FC, Koetzle TF, Williams GJB, Meyer EF, Brice MD, Rodgers JR, Kennard O, Shimanouchi T and Tasumi M, "The Protein Data Bank: a computer based archival file for macromolecular structures", *J. Mol. Biol.* (1977) **112**, 535-542.
- [Bhat *et al.*, 1994] Bhat TN, Bentley GA, Boulot G, Greene MI, Tello D, Dall'Aqua W, Souchon H, Schwarz FP, Mariuzza RA, Poljak RJ "Bound water molecules and conformational stabilization help mediate an antigen-antibody association", *Proc. Natl. Acad. Sci. USA* **91** (1994) 1089-1093.
- [Bronstein *et al.*, 2000] Bronstein IN, Semendjajew KA, Musiol G, Mühlig H, "Das Taschenbuch der Mathematik" 5., überarbeitete, erweiterte Ausgabe, Verlag Harri Deutsch.
- [CCP4, 1994] Collaborative Computational Project, Number 4, "The CCP4 Suite: Programs for Protein Crystallography", *Acta Cryst. D* (1994) **50**, 760-763.
- [Chau, Dean, 1994] Chau P-L, Dean PM, "Electrostatic complementarity between proteins and ligands. 1. Charge disposition, dielectric and interface effects", *J. Compu.-Aided Mol. Design* (1994) **8**, 513-525.

- [Chothia & Janin 1975] Chothia C, Janin J, "Principles of Protein-Protein Recognition", *Nature* (1975) **256**, 705-708.
- [Connolly, 1983] Connolly ML, "Analytical molecular surface calculation", *J. Appl. Crystallogr.* (1983) **16**, 548-558.
- [Connolly, 1983] Connolly ML, "Solvent-accessible Surfaces of Proteins and Nucleic Acids", *Science* (1983) **221**, 709-713.
- [Connolly, 1986] Connolly ML, "Shape complementarity at the hemoglobin  $\alpha 1\beta 1$  subunit interface", *Biopolymers* (1986) **25**, 1229-1247.
- [Connolly, 1986] Connolly ML, "Plotting protein surfaces", *J. Mol. Graphics* (1986) **4**, 93-96.
- [Connolly, 1986] Connolly ML, "Molecular Dot Surface", Quelltext mds.f, FORTRAN-Portierung.
- [Creighton, 1993] Creighton TE, "Proteins" 2nd edition, Freeman, San Francisco.
- [Ferro, Hermans, 1977] Ferro DR, Hermans J, "A Different Best Rigid-Body Molecular Fit Routine", *Acta Cryst. A* (1977) **33**, 345-347.
- [Fersht, 1984] Fersht AR, "Basis of biological specificity", *Trends Biochem. Sci.* (1984) **9**, 145-147.
- [Finkelstein *et al.*, 1995] Finkelstein AV, Gutin AM, Badretdinov AY, "Why do protein architectures have Boltzmann-like statistics", *Proteins: Struct. Funct. Genet.* (1995) **23**, 151-162.
- [Fischer *et al.*, 1992] Fischer D, Bachar O, Nussinov R, Wolfson HJ, "An Efficient Automated Computer Vision Based Technique for Detection of Three Dimensional Structural Motifs in Proteins", *J. Biomol. Struct. and Dyn.* (1992), **9:4**, 769-789.
- [Fischer *et al.*, 1995] Fischer D, Lin SL, Wolfson HL, Nussinov R, "A geometry-based suite of molecular docking processes", *J. Mol. Biol.* (1995) **248:2**, 459-477.

- [Fischer *et al.*, 1995] Fischer D, Nussinov R, Wolfson HJ, "A 3D Sequence-Independent Representation of the Protein Data Bank" *Protein Eng.* (1995) **8:10**, 981-997.
- [Gabb *et al.*, 1997] Gabb HA, Jackson RM, Sternberg MJE, "Modelling Protein Docking using Shape Complementarity, Electrostatics and Biochemical Information" *J. Mol. Biol.* (1997) **272**, 106-120.
- [Helms & Wade, 1995] Helms V, Wade RC, "Thermodynamics of Water Mediating Protein-Ligand Interactions in Cytochrome P450cam: A Molecular Dynamics Study", *Biophys. J.* (1995) **69**, 810-824.
- [Helmer-Citterich, Tramontano, 1994] Helmer-Citterich M, Tramontano A, "PUZZLE: a new method for automated protein docking based on surface shape complementarity", *J. Mol. Biol.* (1994) **235:3**, 1021-1031.
- [Honig & Yang, 1995] Honig B, Yang AS, "Free Energy Balance in Protein Folding", *Adv. Protein Chem.* (1995) **46**, 27-59.
- [Hubbard, 1992] Hubbard S, "ACCESS", (1992) EMBL.
- [Ippolito, Alexander, Christianson, 1990] Ippolito JA, Alexander RS, Christianson DW, "Hydrogen bond stereochemistry in protein structure and function", *J. Mol. Biol.* (1990) **264**, 457-471.
- [Janin *et al.*, 1988] Janin J, Miller S, Chothia C, "Surface, subunit interfaces and interior of oligomeric proteins", *J. Mol. Biol.* (1988) **204**, 155-164.
- [Janin & Chothia, 1990] Janin J, Chothia C, "The structure of protein-protein recognition sites", *J. Biol. Chem.* (1990) **265**, 16027-16030.
- [Jeffrey & Saenger, 1991] Jeffrey GA, Saenger W, "Hydrogen Bonding in Biological Structure", Springer, Berlin (1991)
- [Jiang & Kim, 1991] Jiang F, Kim SH, "Soft docking matching of molecular surface cubes", *J. Mol. Biol.* (1991) **219**, 79-102

- [Jones & Thornton, 1996] Jones SD, Thornton JM, "Principles of Protein-Protein Interactions Derived From Structural Studies", *Proc. Natl Acad. Sci.* (1996) **93**, 13-20.
- [Jones & Thornton, 1997] Jones SD, Thornton JM, "Analysis of Protein-Protein Interaction Sites using Surface Patches", *J. Mol. Biol.* (1997) **272:1**, 121-132.
- [Kabsch & Sander, 1983] Kabsch W, Sander C, "Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features", *Biopolymers* (1983) **22**, 2577-2637.
- [Katchalski-Katzir *et al.*, 1992] Katchalski-Katzir E, Shariv I, Eisenstein M, Friesem A, Aflalo C, Vakser I, "Molecular surface recognition: determination of geometric fit between protein and their ligands by correlation techniques", *Proc. Natl. Acad. Sci. USA* (1992) **89**, 2195-2199.
- [Kernighan & Ritchie, 1990] Kernighan BW, Ritchie DM, "The C programming language, second edition, Ansi C", Prentice-Hall International (1990).
- [Krämer, 1998] Krämer P, "Diplomarbeit: Kriterien zur Vorhersage von Protein-Docking - Die Verwendung von Wasserstoffbrückenbindungen" (1998)
- [Kuntz *et al.*, 1982] Kuntz ID, Blaney JM, Oatley SJ, Langridge R, Ferrin TE, "A Geometric Approach to Macromolecule-Ligand Interactions" *J. Mol. Biol.* (1982) **161**, 269-288.
- [Lamdan *et al.*, 1988] Lamdan Y, Wolfson HJ, "Geometric Hashing: a general and efficient model-based recognition scheme", *Proc. IEEE Int. Conf. On Computer Vision*, Tampa FL, (1988) 238-249.
- [Lamdan *et al.*, 1990] Lamdan Y, Schwartz JT, Wolfson HJ, "Affine Invariant Model-Based Object Recognition", *IEEE Transactions on Robotics and Automation* (1990) **6:5**, 578-589.
- [Larsen, Goodsell, 1998] Larsen TA, Olson AJ, Goodsell DS, "Morphology of Protein-Protein Interfaces", *Structure* (1998) **6:4**, 427-427.

- [Laskowski, 1995] Laskowski RA, "SURFNET: A program for visualizing molecular surfaces, cavities and intermolecular interactions", *J. Mol. Graph.* (1995) **13**, 323-330.
- [Lattman, 1972] Lattman E, "Optimal sampling of the rotation function", *The molecular replacement method* Rossmann Ed. (1972).
- [Lee & Richards, 1971] Lee B and Richards F, "The interpretation of protein structures: Estimation of static accessibility", *J. Mol. Biol.* (1971) **55**, 379-400.
- [Lengauer *et al.*, 1996] Lengauer T, Rarey M, "Computational methods for biomolecular docking", *Curr Opin Struct Biol.* (1996) **6:3**, 402-406.
- [Lin *et al.*, 1994] Lin SL, Nussinov R, Fischer D, Wolfson HJ, "Molecular surface representations by sparse critical points", *Proteins* (1994) **18**, 94-101.
- [Lin *et al.*, 1995] Lin SL, Tsai CJ, Nussinov R, "A study of four-helix bundles: Investigating protein folding via similar architectural motifs in protein cores and in subunit interfaces", *J. Mol. Biol.* (1995) **248**, 151-161.
- [Lin *et al.*, 1996] Lin SL, Nussinov R, "Molecular Recognition Via Face Center Representation of Molecular Surface", *J. Mol. Graphics* (1996) **14**, 78-90.
- [McDonald *et al.*, 1993] McDonald IK, Naylor D, Jones D, Thornton J, "HBPLUS Hydrogen Bond Calculator v 2.25", University College London, London.
- [McDonald & Thornton, 1994] McDonald IK and Thornton J, "Satisfying hydrogen bonding potentials in proteins", *J. Mol. Biol.* (1994) **238**, 777-793.
- [McCoy *et al.*, 1997] McCoy AJ, Epa VC, Colman PM, "Electrostatic Complementarity at Protein/Protein Interfaces." *J. Mol. Biol.* (1997) **268**, 570-584.
- [Meyer, 1996] Meyer M, Implementierung des Rotations-Samplings nach Lattman [Lattman, 1972], persönliche Mitteilung an O. Zimmermann.

- [Meyer, Wilson, Schomburg, 1996] Meyer M, Wilson P, Schomburg D, "Hydrogen bonding and molecular surface shape complementarity as a basis for protein docking", *J. Mol. Biol.* (1996) **264**, 199-210.
- [Mucke, Edelsbrunner, 1992] Mucke E, Edelsbrunner H, "Three-dimensional Alpha Shapes", *Volume Visualisation Workshop Proceedings, Boston, MA* (1992).
- [Norel *et al.*, 1994] Norel R, Lin SL, Wolfson HJ, Nussinov R, "Shape complementarity at protein-protein interfaces", *Biopolymers* (1994) **34**, 933-940.
- [Norel *et al.*, 1995] Norel R, Lin SL, Wolfson HJ, Nussinov R, "Molecular surface complementarity at protein-protein interfaces: the critical role played by surface normals at well placed, sparse points in docking", *J. Mol. Biol.* (1995) **353:2**, 263-273.
- [Norel *et al.* 1997] Norel R, Lin SL, Xu D, Wolfson HJ, Nussinov R, "Molecular surface variability and induced conformational changes upon protein-protein associations. In: Structure, Motion, Interaction and Expression of Biological Macromolecules", *The Proceedings of the 10th Conversation, Adenine Press, Albany, NY* (1997).
- [Norel *et al.*, 1999] Norel R, Petrey D, Wolfson HJ, Nussinov R, "Examination of shape complementarity in docking of unbound proteins.", *Proteins* (1999) **36:3**, 307-317.
- [Nussinov & Wolfson, 1991] Nussinov R, Wolfson HJ, "Efficient Detection of three-dimensional structural motifs in biological macromolecules by computer vision techniques", *Proc. Natl. Acad. Sci. USA* (1991) **88:23**, 10495-10499.
- [Ritchie, Kemp, 1999] Ritchie DW, Kemp GJL, "Fast Computation, Rotation and Comparison of Low Resolution Spherical Harmonic Molecular Surfaces", *J. Comp. Chem.* (1999) **20**, 383-395.
- [Ritchie, Kemp, 1999] Ritchie DW, Kemp GJL, "Protein Docking Using Spherical Polar Fourier Correlations", *Proteins: Struct. Func. Genet.* (2000) **39**, 178-194.
- [Press *et al.*, 1992] Press WH, Flanery BP, Teukolsky SA, Vetterling WT, "Numerical Recipes in Fortran", *Cambridge University Press, Cambridge*.



- [Roth, Neal, Lenhoff, 1996] Roth CM, Neal BL, Lenhoff AM, "Van der Waals interactions involving proteins", *Biophys. J.* (1996) **70**, 977-987.
- [Sayle and Milner-White, 1995] Sayle RA, Milner-White EJ, "RasMol: Biomolecular graphics for all", *Trends in Biochemical Sciences* (1995) **20**, 374-376.
- [Sayle, 1994; Bernstein, 2001; Molinaro, 1996] Sayle R, 1984; Bernstein HJ, 1998-2001; Molinaro M, 1996, "RasMol" Sayle: Biomolecular Structures Group, Glaxo Wellcome Research & Development, Stevenage, Hertfordshire, UK; Bernstein: Bernstein + Sons, Bellport, NY, USA; Molinaro: UC Regents/ModularCHEM Consortium.
- [Schomburg und Reichelt, 1988] Schomburg D und Reichelt J "BRAGI: a comprehensive protein modeling program system", *J. Mol. Graphics* (1988) **6**, 161-165.
- [Schwartz & Sharir, 1987] Schwartz JT, Sharir M, "Identification of Partially Obscured Objects in Two and Three Dimensions by Matching Noisy Characteristic Curves", *Int.J.Robotics Res.* (1987) **6:2**, 29-44.
- [Stickle *et al.*, 1992] Stickle DF, Presta LG, Dill KA and Rose GD, "Hydrogen Bonding in Globular Proteins.", *J. Mol. Biol.* (1992) **226**, 1143-1159.
- [Shoichet *et al.*, 1992] Shoichet BK, Bodian DI, Kuntz ID, "Molecular Docking using Shape Descriptors" *J. Comput. Chem.* (1992) **13**, 380-397.
- [Shoichet *et al.*, 1993] Shoichet BK, Kuntz ID, "Matching Chemistry and Shape in Molecular Docking", *Protein Eng.* (1993) **6:7**, 723-732.
- [Stryer, 1975] Stryer L, "Biochemistry", W.H. Freeman & Company, New York 1975.
- [Tsai, Lin, Wolfson, Nussinov, 1997] Tsai CJ, Lin SL Wolfson H, Nussinov R, "Studies of protein-protein interfaces: a statistical analysis of the hydrophobic effect", *Protein Sci.* (1997) **6:1**, 53-64.
- [Tsai *et al.*, 1998] Tsai CJ, Xu D, Nussinov R. "Protein folding *via* binding, and *vice versa*", *Folding & Design* (1998) **3**, R71-R80.

- [Vakser, Aflalo, 1994] Vakser IA, Aflalo C, "Hydrophobic Docking: a proposed enhancement to molecular recognition techniques", *Proteins* (1994) **20:4**, 320-329.
- [Vakser, 1995] Vakser IA, "Protein docking for low-resolution structures", *Protein Eng.* (1995) **8**, 371-377.
- [Vakser, 1996] Vakser IA, "Low-resolution docking: Prediction of complexes for underdetermined structures", *Biopolymers* (1996) **39**, 455-464.
- [Vakser & Sali, 2001] Vakser IA, Sali A, "COMBASE, a database of protein-protein complexes", *noch nicht publiziert* <http://guitar.rockefeller.edu/sub-pages/combase.html>
- [Will, 1993] Will RD, "Geheimnis Wasser", Knauer-Verlag, München 1993.
- [Wilson & Stanfield, 1993] Wilson IA and Stanfield RL, "Antibody-antigen interaction", *Curr. Opinion Struct. Biol.* (1993) **3**, 113-118.
- [Wolfson & Nussinov, 1998] Wolfson HJ, Nussinov R, "From computer vision to protein structure ans association", *Computational Methods in Molecular Biology*, Elsevier Sience B.V. (1998) 313-334.
- [Xu, *et al.*, 1997] Xu D, Lin SL, Nussinov R, " Protein binding versus protein folding: The role of hydrophilic bridges in protein associations", *J. Mol. Biol.* (1997) **265**, 68-84.
- [Xu, Tsai, Nussinov, 1997] Xu D, Tsai C-J, Nussinov R, "Hydrogen bonds and salt bridges across protein-protein interfaces", *Prot. Engineering* (1997) **10:9**, 999-1012.
- [Zimmermann, 1997] Zimmermann O, persönliche Mitteilung.
- [Zimmermann, 1999] Zimmermann O, C++-Portierung der Implementierung [Meyer, 1996] des Rotations-Samplings nach Lattman [Lattman, 1972], persönliche Mitteilung.