

**Entwicklung von Methoden zur Strukturvorhersage von
Proteinkomplexen und zur Untersuchung von
Proteinzuständen mit
Hochdruckkernspinresonanzspektroskopie**



Universität Regensburg

DISSERTATION ZUR ERLANGUNG DES
DOKTORGRADES DER NATURWISSENSCHAFTEN (DR. RER. NAT.)
DER FAKULTÄT FÜR BIOLOGIE UND VORKLINISCHE MEDIZIN
DER UNIVERSITÄT REGENSBURG

vorgelegt von

Bärbel Kieninger

aus Regensburg

im Jahr 2015

Promotionsgesuch eingereicht am: 24. Juni 2015

Die Arbeit wurde angeleitet von: Prof. Dr. Dr. Hans Robert Kalbitzer

Unterschrift:

Inhaltsverzeichnis

1	Einleitung	
1.1	Wechselwirkungen von Proteinen	1
1.1.1	Strukturbestimmung von Proteinkomplexen	2
1.1.2	Evolutionäre Konservierung der Bindestellen	4
1.1.3	Strukturelle Beweglichkeit und Mehrzustandsmodell	5
1.1.4	Korrespondenz zwischen freien und gebundenen Zuständen	6
1.1.5	Hochdruck-NMR-Spektroskopie zur Untersuchung der Proteinzustände	7
1.1.6	Ziele dieser Arbeit	8
1.2	Untersuchte Proteine	9
1.2.1	G-Protein gekoppelte Signaltransduktion: Ras und Byr2	10
1.2.2	A β und Alzheimersche Krankheit	15
1.2.3	IE1 und Persistenz des Cytomegalievirus	17
2	Material und Methoden	
2.1	Untersuchte Proteine und Datensätze	19
2.1.1	Byr2-RBD	19
2.1.2	IE1-CTD	21
2.1.3	Alzheimer- β -Amyloidpeptid (A β)	21
2.1.4	Ras-Protein	22
2.2	Theorie	23
2.2.1	Molekulardynamik	23
2.2.2	Homologie-Modellierung mit PERMOL	28
2.2.3	Thermodynamik	32
2.2.4	Fit-Verfahren	37
2.3	Neuimplementierung von PERMOL	40
2.3.1	Erweiterung der Methode	40
2.3.2	Programmiersprache und Programmaufbau	40
2.3.3	Aufbereitung der Daten	41
2.3.4	Erfassung der Eingabeparameter	42
2.3.5	Alignment	43
2.3.6	Informationsübertragung	46
2.3.7	Berechnung der Restraints und Gewichtung verschiedener PDBs	47

2.3.8	Modifikationen an dem Programm CNS	50
2.4	Implementierung von MultiStates	54
2.4.1	Anwendung des Levenberg-Marquardt-Verfahrens und einer alternativen Lösungsmöglichkeit	54
2.4.2	Besonderheiten des Parameterraums	61
2.4.3	Erzeugung und Einsatz von Testdatensätzen	65
2.4.4	Bestimmung eines Lösungsbereichs im Parameterraum	67
2.4.5	Implementierung der entwickelten Verfahren mit MATLAB	68
3	Ergebnisse	
3.1	Strukturvorhersagen von Byr2-RBD mit PERMOL	72
3.1.1	Rekonstruktion von Byr2-RBD aus PDB-1K8R	72
3.1.2	Verbesserung der NMR-Struktur von Byr2-RBD	77
3.1.3	Modellierung des Komplexes Byr2-RBD mit Ras aus homologen Strukturen	82
3.2	Erstellung eines Homologiemodells des Proteins IE1(476-491) im Komplex mit Chromatin	94
3.2.1	Erstellung von Alignments zu LANA(4-17)	94
3.2.2	Modellrechnungen mithilfe von PERMOL und CNS	96
3.2.3	Beurteilung der Modelle	97
3.2.4	NMR-spektroskopische Untersuchungen von IE1(1-491) und IE1(1-475)	104
3.3	Optimierung der Anfangsbedingungen von Algorithmus 1 durch Untersuchungen des Parameterraums	109
3.3.1	Die Form der Funktion Norm über dem Parameterraum	109
3.3.2	Geeignete Wahl der Startpunkte für Algorithmus 1 und Funktionstest	120
3.4	Untersuchung des Einflusses statistischer Fehler	123
3.4.1	Einfluss auf die Funktionsweise von Algorithmus 1 und die Größe der Norm	123
3.4.2	Wirkung auf die Gibbs-Energiedifferenzen und die Besetzungszahlen	126
3.4.3	Beeinflussung der chemischen Verschiebungen der Zustände	132
3.5	Bestimmung eines Lösungsbereichs um das mathematische Minimum	137
3.5.1	Untersuchung der Symmetrie der Fehlerbereiche	137
3.5.2	Definition des Lösungsbereichs über geeignete Punkte des Parameterraums	140
3.5.3	Stabilitätstest der entwickelten Methode bei Variation der Anzahl der verwendeten Daten	142
3.6	Untersuchung des Druckverhaltens des Proteins Byr2-RBD	145

3.6.1	Vorbereitung der Daten	145
3.6.2	Versuch einer Auswertung der chemischen Verschiebungen mit Algorithmus 1	146
3.6.3	Untersuchung der Signalvolumina	149
3.6.4	Multimodale Datenanalyse	150
3.7	Anwendung von Algorithmus 1 auf weitere experimentell gewonnene Datensätze	156
3.7.1	Untersuchung des Druckverhaltens der chemischen Verschiebungen von Ras	156
3.7.2	Auswertung der chemischen Verschiebungen des Alzheimer- β -Amyloidpeptids	161
4	Diskussion	
4.1	Beurteilung der Methode und Ergebnisse von PERMOL	165
4.1.1	Rekonstruktion von Byr2-RBD aus 1K8R	165
4.1.2	Verbesserung der Struktur von Byr2 aus NMR-spektroskopischen Daten	166
4.1.3	Homologiemodelle des Komplexes Byr2-RBD mit Ras	167
4.1.4	Homologiemodelle von IE1-CTD im Komplex mit menschlichem Chromatin	171
4.1.5	Zusammenfassende Beurteilung von PERMOL	172
4.2	Beurteilung der Methoden und Ergebnisse von MultiStates	174
4.2.1	Anwendung des Levenberg-Marquardt-Algorithmus auf das vorliegende Problem	174
4.2.2	Untersuchung des Parameterraums, Optimierung der Anfangsparameter und Funktionstest	175
4.2.3	Untersuchungen der Auswirkungen der statistischen Fehler	177
4.2.4	Angabe eines Lösungsbereichs	178
4.2.5	Beurteilung der Ergebnisse von Byr2-RBD	179
4.2.6	Beurteilung der Ergebnisse für das Protein Ras	182
4.2.7	Beurteilung der Ergebnisse für das Alzheimer- β -Amyloidpeptid	184
4.2.8	Bewertung der Ergebnisse von MultiStates	185
5	Zusammenfassung	187
6	Literaturverzeichnis	189
7	Publikationsverzeichnis	197
8	Abkürzungsverzeichnis	198

1 Einleitung

1.1 Wechselwirkungen von Proteinen

Bei der Entwicklung des Lebens auf der Erde entstanden mehrere Klassen komplexerer Moleküle, die heute noch millionenfach in jeder Zelle vorkommen, darunter die Aminosäuren und die Nukleotide. Beide haben die Eigenschaft, sich zu Ketten zusammenschließen zu können. Während insgesamt fünf verschiedene Arten von Nukleotiden ausreichen, um die Makromoleküle DNS und RNS aufzubauen, gibt es 22 proteinogene Aminosäuren. Diese bilden die Bausteine der Proteine. Beim Aufbau zu Polypeptidketten, der über Peptidbindungen erfolgt, ist somit eine große Vielzahl an Variationen möglich. Diese spiegelt sich in dem breiten Spektrum an Funktionen, die Proteine übernehmen können, wider: als Grundbausteine beim Aufbau von Strukturen, als Transporter durch den Körper, als Enzyme und vieles andere mehr. Dabei ist die dreidimensionale Faltung der Ketten elementar.

Bei allen biologischen Prozessen, an denen Proteine beteiligt sind, wechselwirken diese mit anderen Molekülen. Diese Wechselwirkungen sind sehr spezifisch, jedes Proteinmolekül bindet mit nur einem oder einigen wenigen der vielen verschiedenen Molekültypen, mit denen es zusammentrifft. Diese Spezifität hat ihre Ursache in der Variabilität der Proteine bezüglich ihrer dreidimensionalen Struktur und ihrer physikochemischen Eigenschaften. Die genaue Kenntnis der Interaktionen ist elementar für das Verständnis der physiologischen Prozesse, insbesondere dann, wenn Wirkstoffe entwickelt werden sollen, die in diese Prozesse eingreifen, oder wenn pathogene Mutationen vorliegen.

Das Aufeinandertreffen der an einer Interaktion beteiligten Moleküle geschieht durch Diffusion. Zur Bindung kommt es, wenn die Eigenschaften der Oberflächen der stoßenden Teilchen komplementär zueinander sind, sich also in Form und Ladung ergänzen. Die Lebensdauer des Komplexes hängt dabei von der Stärke der Kräfte zwischen den beiden Partnern ab. Aber auch stabile Bindungen zwischen biologischen Makromolekülen werden im Allgemeinen nicht durch kovalente Wechselwirkungen stabilisiert, sondern die Protein-Interaktionen beruhen auf den gleichen elementaren Wechselwirkungen, die auch die Proteine selbst - ebenso wie Nukleinsäuren - zu ihrer Faltung antreiben und festigen: van der Waals-Kräfte, Wasserstoffbrückenbindungen, elektrostatische Interaktionen und Energieanteile, die als hydrophobe Wechselwirkungen bezeichnet werden (Alberts et al., 2011).

1.1.1 Strukturbestimmung von Proteinkomplexen

Um die Struktur von Proteinkomplexen zu untersuchen, sind im Prinzip die gleichen Methoden anwendbar, die man zur Strukturbestimmung von einzelnen Proteinen oder deren Domänen verwendet, allen voran die Röntgenkristallographie (X-RAY) und die Kernspinresonanzspektroskopie (NMR). Beiden Techniken ist jedoch gemein, dass diese Methoden bei Proteinkomplexen deutlich schwieriger zu realisieren sind als bei einzelnen Proteinen.

Röntgenstrukturen werden erzeugt, indem aus dem Beugungsmuster, das bei der Röntgenbeugung an den Atomen regelmäßig aufgebauter Proteinkristalle entsteht, auf die dreidimensionale Struktur rückgerechnet wird. Dazu müssen die Makromoleküle durch hohe Proteinkonzentration und Zugabe geeigneter Präzipitationsmittel kristallisiert werden und starre Strukturen bilden. Eine zusätzliche Schwierigkeit bei der Kristallisation von Komplexen im Vergleich zu der von einzelnen Proteinen liegt unter anderem darin, dass die Mengenverhältnisse der beiden Partner exakt stimmen müssen (Pingoud et al., 2002).

Bei der Strukturbestimmung mittels Kernspinresonanzspektroskopie galt eine Größe der Molekülkomplexe von etwa 300 Aminosäuren lange Zeit als obere Grenze, da die Linienbreite in den Spektren unter anderem von der Beweglichkeit des Moleküls in Lösung abhängt. Deswegen werden die Signale, aus deren Frequenz und Volumen auf die Konfiguration des Proteins rückgeschlossen wird, mit zunehmender Größe des untersuchten Komplexes immer breiter und sind schlechter vom Untergrundrauschen zu unterscheiden. Durch die Entwicklung der TROSY-Technik ist inzwischen die Strukturbestimmung auch bei größeren Komplexen möglich (Wüthrich, 2003), jedoch nach wie vor sehr aufwendig und zeitraubend. Verschiedene Möglichkeiten, die durch die größere Anzahl resonanzfähiger Atome bedingte Überlagerung der Signale in den Spektren zu vermeiden, sind die selektive Isotopenmarkierung, bei der nur spezielle Aminosäuren - zum Beispiel alle Valine, Leucine, Isoleucine, Phenylalanine und Tyrosine - isotopenmarkiert werden (Schedlbauer et al., 2008), oder die Inteinmethode (David et al., 2004).

Daraus erklärt sich die Notwendigkeit zur Entwicklung von Computerprogrammen, sogenannten *Docking*-Programmen, die aus den Strukturen ihrer Untereinheiten auf den Proteinkomplex schließen. Die meisten dieser Programme beruhen auf der Annahme, dass die Bindungspartner komplementär zueinander sind und einfach solange gegeneinander gedreht

werden müssen, bis eine optimale Orientierung gefunden ist, in der diese Komplementarität gegeben ist. Die Grundidee dieses Schlüssel-Schloss-Prinzips geht auf den Chemiker Emil Fischer zurück (Fischer 1894). Entscheidend jedoch für die erfolgreiche Entwicklung von *Docking*-Programmen war ein von Ephraim Katchalski-Katzir entwickelter Algorithmus auf der Basis von Fourier-Korrelationen, der eine Beurteilung der verschiedenen Orientierungen der beiden Bindungspartner zueinander erlaubt. Dazu werden die beiden zu dockenden Proteinstrukturen zuerst in zwei getrennte dreidimensionale Gitter projiziert. Als nächstes wird jedem Gitterpunkt ein Wert zugeordnet, je nachdem, ob er sich außerhalb, innerhalb oder auf dem Rand des Moleküls befindet. Dann werden die beiden Gitter aufeinander verschoben und verdreht und für jede Position ein Score berechnet, der ein Maß für die Passform ist (Katchalski-Katzir et al., 1992). Diese Methode, die ursprünglich nur die geometrische Passform überprüfte, wurde von verschiedenen Arbeitsgruppen aufgegriffen und um die Möglichkeit erweitert, parallel dazu die physikochemische Komplementarität zu beurteilen (Ritchie, 2008).

Die eigentliche Schwierigkeit liegt jedoch darin, dass viele komplexbildende Proteine beim Kontakt mit dem Bindungspartner ihre Form ändern und sich aneinander anpassen. Einige *Docking*-Programme versuchen dieses Problem zu umgehen, indem sie einzelne Aminosäurereste der *Interfaces*, der Bindestellen der beiden Partner, oder auch Teile des Proteinrückgrats beweglich machen, um diese Anpassungen zulassen zu können (Bonvin 2006). Ein anderer Ansatz ist der, dass durch geeignete Filter, die vor der Fourierkorrelationsberechnung auf die Gitter angewendet werden (*Soft-Docking*), die Oberfläche der Proteine „weicher“ gemacht wird (Yoshimori und Del Carpio M., 2001).

Im Allgemeinen beschränken sich die bei Komplexbildung entstehenden strukturellen Verformungen nicht nur auf die Bindestelle, sondern verändern die gesamte Geometrie des Moleküls mehr oder weniger stark. Es konnte gezeigt werden, dass ein Zusammenhang zwischen der intrinsischen Flexibilität von Proteinen und der Größe der tatsächlichen strukturellen Änderungen besteht, die sie beim *Docking* mit ihrem Partner machen (Marsh et al., 2012). Nachweisen kann man solche lokalen Strukturänderungen an Aminosäuren außerhalb der Bindestelle zum Beispiel durch NMR-Experimente. Für das in dieser Arbeit verwendete Protein Byr2-RBD zeigten sich bei Komplexbildung mit Ras nicht nur Änderungen der chemischen Verschiebungen der im *Interface* liegenden Aminosäuren, sondern auch weiter von der Bindestelle entfernt liegende Reste wichen etwas von den

Werten des ungebundenen Zustands ab, was nur durch Veränderungen im lokalen Magnetfeld und damit durch Umformung der Struktur in deren Umgebung erklärbar ist (Gronwald et al., 2001).

Diese kleineren oder größeren Konformationsänderungen bei der Komplexbildung sind der Schlüssel dafür, wie die Signalübertragung oder die Wirkung von Proteinen als Enzyme erklärt werden kann (Alberts et al., 2011).

1.1.2 Evolutionäre Konservierung der Bindestellen

Will man die Struktur eines Proteins ohne die aufwendige Durchführung von Experimenten bestimmen, durchsucht man die Datenbanken nach sequenzähnlichen Proteinen und zieht aus deren Struktur Rückschlüsse auf die gesuchte dreidimensionale Konformation. Diese Technik nennt man Homologiemodellierung. Sie beruht auf der Tatsache, dass während der Evolution zwar Mutationen die Sequenzen verändert haben, strukturell wichtige Aminosäuren aber erhalten oder durch ähnliche ersetzt wurden. Diese Ähnlichkeit kann sich auf ihre physikalisch-chemischen Eigenschaften - wie Polarität oder Ladung - oder auch auf ihre Größe beziehen. Dadurch wird das für eine Familie typische Faltungsmuster erhalten. Von einer Strukturerhaltung kann man ausgehen, wenn eine Sequenzähnlichkeit von mindestens 20% vorliegt, wobei dieser Wert allerdings von der Proteinfamilie selbst und von der Größe des Proteins abhängig ist (Chotia und Lesk, 1986, 1987; Wood und Pearson, 1999; Sander und Schneider, 1991).

Eine stärkere Konservierung findet man aber nicht nur bei strukturell, sondern auch bei funktionell wichtigen Resten, da bereits einzelne Mutationen im Bereich einer Bindestelle die Bindungseigenschaften stark modifizieren können. Ein Beispiel dafür ist das Krankheitsbild der Sichelzellenanämie, bei der die Veränderung einer einzelnen Aminosäure ausreicht, um die Bindungsspezifität der Hämoglobinmoleküle so zu erweitern, dass sie selbst miteinander interagieren und sich Fäden aus ihnen bilden (Ingram, 1956).

Die gezielte Suche nach konservierten Bereichen auf der Moleküloberfläche durch Sequenzvergleich bei homologen Proteinen kann auch verwendet werden, um an der Interaktionsfläche beteiligte Aminosäuren ausfindig zu machen (Lichtarge et al., 1996). Daneben sind auch korrelierte Mutationen bekannt, bei denen die Mutation eines

Bindungspartners durch eine zweite Mutation beim anderen Bindungspartner kompensiert wird, um die Bindungsfähigkeit zu erhalten (Pazos et al., 1997).

Diese Erkenntnisse legen die Idee nahe, die Technik der Homologiemodellierung auf Proteinkomplexe zu übertragen. Dieser Ansatz wurde auch bei der Entwicklung des Programms HOMCOS (Fukuhara und Kawabata, 2008) gewählt.

1.1.3 Strukturelle Beweglichkeit und Mehrzustandsmodell

Die Möglichkeit, sich bei Komplexbindung an den Bindungspartner anzupassen (*induced fit*), liegt in der Natur der Proteine. Sie sind keine starren Einheiten, sondern besitzen viele Freiheitsgrade. Die kleinsten Bewegungen laufen im Femto- und Pikosekundenbereich ab und sind atomare Fluktuationen wie Bindungslängenänderungen kovalent gebundener Atome. Etwas länger dauern Bewegungen der Aminosäureseitenketten oder strukturelle Veränderungen größerer Bereiche, weil dabei mehr Masse bewegt wird. Die Proteinfaltung vom ungefalteten zum gefalteten Zustand dagegen kann sogar mehrere Sekunden in Anspruch nehmen. Während atomare Fluktuationen nur lokal ablaufen, sind kollektive Bewegungen globale Änderungen der Struktur und können als Übergang zwischen zwei Zuständen beschrieben werden. Jedoch lassen sich thermische Bewegungen sehr schwer von echten Zustandsänderungen unterscheiden (Huber, 1979).

Anfinsen definierte vor mehr als 40 Jahren den nativen Zustand eines Proteins in seinem normalen physiologischen Milieu als den dreidimensionalen Zustand mit der geringsten Gibbs-Energie des Systems (Anfinsen, 1973). Diese liegt im Allgemeinen 20-60 kJ/mol niedriger als der denaturierte entfaltete Zustand. Inzwischen hat sich die Vorstellung eines Faltungstrichters in der Energielandschaft des Proteins durchgesetzt, wobei als Energielandschaft die Funktion der Gibbs-Energie in Abhängigkeit von den Freiheitsgraden des Moleküls bezeichnet wird (Abbildung 1-1). Die ungleichmäßige, zerfurchte Oberfläche mit mehreren Minima symbolisiert die Möglichkeit mehrerer energetisch günstiger Faltungszustände (Dill und Chan, 1997).

Diese Sichtweise erlaubt auch eine Unterscheidung zwischen thermischer Bewegung, also einer Bewegung innerhalb eines Tales in der Energielandschaft, und einer Zustandsänderung, bei der die Konfiguration sich kollektiv soweit ändert, dass ein nächstes Tal erreicht wird. Ein Zustand wird daher immer durch ein Ensemble aus Konfigurationen repräsentiert.

In welchem Besetzungsverhältnis nun die einzelnen Zustände vorliegen, hängt von der Tiefe und Form ihrer Minima ab. Zudem wird die Form der Energiefläche von den äußeren Bedingungen wie Temperatur, Druck, Lösungsmittel oder pH beeinflusst.

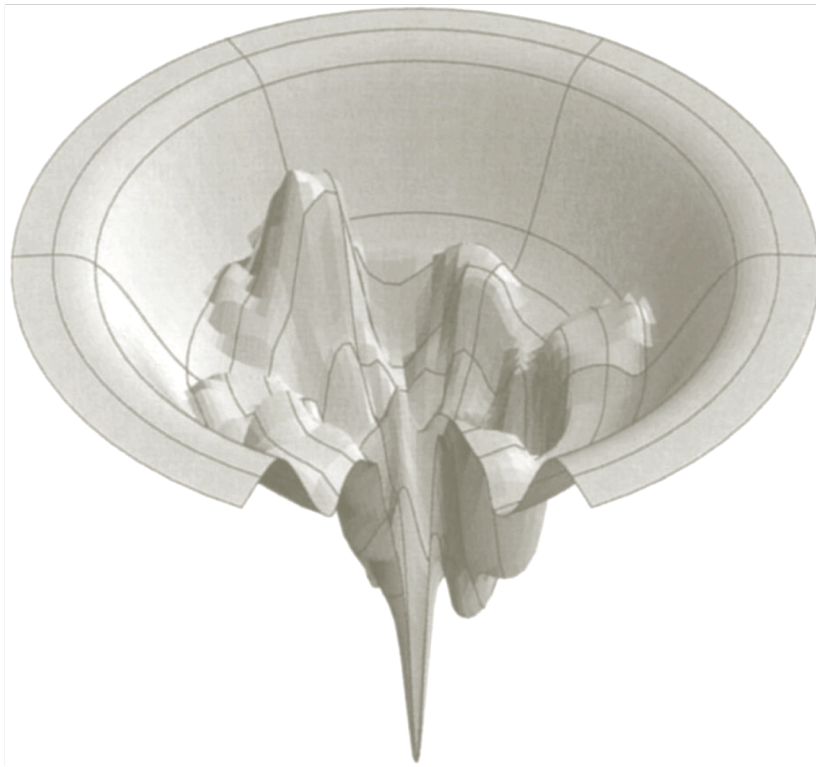


Abbildung 1-1: Energielandschaft eines Proteins

Um die Energielandschaft eines Proteins im Dreidimensionalen darzustellen, wählt man zwei der N Freiheitsgrade des Proteins aus und trägt auf der vertikalen Achse dazu die Abhängigkeit der reduzierten Gibbs-Energie auf.

(Abbildung nach: Dill und Chan, 1997)

1.1.4 Korrespondenz zwischen freien und gebundenen Zuständen

Überträgt man die Schlussfolgerung von Marsh et al., dass eine größere Flexibilität im ungebundenen Zustand mit größeren strukturellen Änderungen bei der Bindung verknüpft ist (Marsh et al., 2012), auf das Mehrzustandsmodell, könnte das bedeuten, dass bei größerer Beweglichkeit mehrere energetisch ähnliche Zustände in der Energielandschaft des ungebundenen Proteins vorhanden sind und einer dieser Zustände dem gebundenen bereits ähnlich sein muss, damit die Bindung überhaupt stattfinden kann. Dieses Konzept der Koexistenz verschiedener Zustände wurde bereits vor mehr als 40 Jahren vorgeschlagen und

ist als Monod-Wyman-Changeux-Modell oder konformationelle Selektion als Ergänzung zum Induced-Fit-Modell bekannt (Abbildung 1-2) (Changeux und Edelstein, 2005).

Von Tobi und Bahar wurde durch theoretische Berechnung für mehrere Proteinkomplexe ein diesem Modell entsprechender Zusammenhang zwischen den signifikanten Strukturänderungen bei Komplexbildung und der intrinsischen Bewegung der Proteine, das über ein Federmodell mathematisch realisiert wurde, im ungebundenen Zustand gezeigt (Tobi und Bahar, 2005).

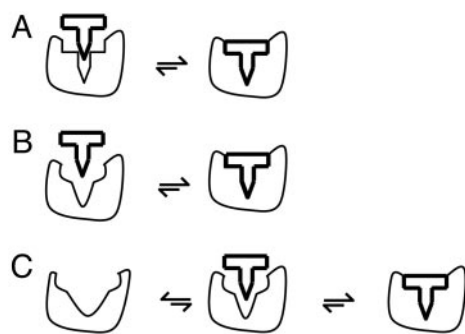


Abbildung 1-2: Modelle für Protein-Interaktionen

(A) Schlüssel-Schloss-Prinzip
(B) Induced-Fit-Modell
(C) Monod-Wyman-Changeux-Modell
(Konformationelle Selektion)

(Abbildung entnommen aus: Tobi und Bahar, 2005)

1.1.5 Hochdruck-NMR-Spektroskopie zur Untersuchung der Proteinzustände

NMR-Experimente werden seit mehreren Jahren verwendet, um verschiedene Zustände eines Proteins nachzuweisen, da bei dieser Technik die Moleküle in Lösung vermessen werden können und ihre natürliche Bewegung nicht eingeschränkt wird. Die grundlegende Idee ist davon auszugehen, dass im Gleichgewicht alle funktionellen Zustände eines Proteins vorliegen, aber meist eine der Konformationen dominiert. (Akasaka, 2006)

Dieses Gleichgewicht kann man nun durch Hochdruck verschieben, wodurch sich die Besetzungsverhältnisse verändern. Die Hoffnung dabei ist, unter Normaldruck weniger besetzte Zustände stabilisieren und "sichtbar" machen zu können. Es wird angenommen, dass die Verschiebung zugunsten von Konfigurationen verläuft, deren molares Volumen kleiner ist als das des bei Normaldruck höchstbesetzten. Bei sehr hohen Drücken würde man damit ein entfaltetes Protein erhalten. Dass es tatsächlich möglich ist, Populationsverhältnisse zwischen funktionellen Konfigurationen zu beeinflussen, wurde erstmals für das Protein Ras anhand von Phosphor-NMR-spektroskopischen Experimenten gezeigt: Während bei Normaldruck der Effektoren-bindende Zustand dominiert, konnte durch Druck das Gleichgewicht zugunsten des Guanosintriphosphat-Austauschfaktoren-bindenden Zustands verschoben werden. Die

Differenz der Gibbs-Energie dieser zwei Zustände wurde bei Normaldruck zu $1,5 \text{ kJ mol}^{-1}$ bestimmt, die Differenz des spezifischen Volumens zu $17,2 \text{ mL mol}^{-1}$ (Kalbitzer et al., 2009).

Das Ziel bei diesen Untersuchungen ist es, durch die Verschiebungen des Gleichgewichts Aminosäuren zu identifizieren, die maßgeblich am Übergang zu einem speziellen Zustand beteiligt sind. Dazu muss eine Anpassung der Daten an ein thermodynamisches Modell erfolgen (siehe Kapitel 2.2.3).

Diese funktionellen Aminosäuren, die bei Strukturänderungen hin zu einem Zustand beteiligt sind, der eine hohe Affinität zu einem bestimmten Bindungspartner aufweist, könnten nun als Ansatzpunkt für Wirkstoffe dienen. Dabei sind insbesondere außerhalb von Bindestellen liegende Reste interessant, da sie allosterisch wirken und nicht die Kontaktfläche blockieren, sondern die Wahrscheinlichkeit einer Interaktion mit dem betrachteten Bindungspartner durch eine Verschiebung des Gleichgewichts verändern können. Dies könnte gezielt beim Drug Design ausgenutzt werden (Kalbitzer et al., 2013).

1.1.6 Ziele dieser Arbeit

Das zentrale Thema der Arbeitsgruppen am Lehrstuhl für Biophysik der Universität Regensburg ist die strukturbasierte Analyse von Funktion und Interaktion biologischer Makromoleküle in Lösung. Dazu stehen mehrere NMR-Spektrometer, die mit einer Drucksprunganlage zur Durchführung von Kernspinexperimenten unter Hochdruck ausgestattet sind, und verschiedene computerbasierte Techniken zur Verfügung. Das dort in Zusammenarbeit mit der Bruker BioSpin GmbH entwickelte Softwarepaket AUREMOL (Gronwald et al., 2004; Gronwald und Kalbitzer, 2004) dient zur Visualisierung und Auswertung der NMR-Spektren und zur Proteinstrukturbestimmung.

Ein Ziel dieser Arbeit ist es, ein Homologiemodell des C-terminalen Endes des Proteins IE1 des Cytomegalievirus in Komplex mit menschlichem Chromatin zu erstellen, um weitere Erkenntnisse über die lebenslange Persistenz dieses Virus im menschlichen Körper gewinnen zu können und den Mechanismus der Latenz zu verstehen. Daraus ergibt sich die Notwendigkeit, das in AUREMOL integrierte Modul PERMOL (Möglich et al., 2005), das der Erstellung von Homologiemodellen einer einzelnen Domäne und der Strukturverfeinerung dient, zu verbessern und neu zu implementieren und dabei um die Möglichkeit zu erweitern, Proteinkomplexe aus homologen gebundenen Strukturen - die aus zwei, oder wie im

vorliegenden Fall, aus noch mehr Sequenzstücken bestehen - vorhersagen zu können. Dieser Ansatz scheint erfolgversprechend, da die Bindebereiche besonders stark konserviert sind und die Informationen, die man aus der homologen Struktur für den gesuchten Komplex ziehen kann, bereits von zwei aneinander angepassten Partnern stammen. Insbesondere muss nicht mehr nach der richtigen Orientierung zueinander gesucht werden, und außerdem wird die Bindung zwischen den Partnern durch die gleichen Kräfte getrieben, die auch die Faltung eines einzelnen Proteins stabilisieren. Das für PERMOL entwickelte Grundkonzept (siehe Kapitel 2.2.2) kann dabei direkt übernommen werden. Bei der Neuimplementierung sollen zudem die Limitierungen der alten Version behoben werden, wobei vor allem die Fehler bei der Berechnung des Alignments, der Auswahl der Aminosäuren und der Benennung der Atome beseitigt und eine benutzerfreundlichere Oberfläche geschaffen werden sollen.

Ein zweites Ziel dieser Arbeit ist es, die mit NMR-spektroskopischen Methoden erstellte Struktur der Ras-bindenden Domäne des Proteins Byr2 von *Schizosaccharomyces Pombe* zu verbessern und aus einer Analyse der Hochdruckdaten auf dessen funktionelle Konformationen zu schließen. Dazu soll eine Methode - MultiStates - entwickelt werden, mit der aus dem Verhalten der chemischen Verschiebungen - unter Einbeziehung aller verfügbarer Aminosäuren - bei Druckvariation eines Proteins die Besetzungsverhältnisse seiner Zustände abgeschätzt werden können, um so wiederum auf funktionell wichtige Aminosäuren zu schließen. Dabei wird an die Arbeit von Köhler et al. (2012) angeknüpft, nachdem sich die Änderungen der chemischen Verschiebungen aus einem atom- und aminosäureabhängigen Anteil und einem strukturell bedingten Anteil zusammensetzen, und für den zweitgenannten ein neues mathematisches Verfahren zu dessen Auswertung entwickelt.

1.2 Untersuchte Proteine

Im nun folgenden Kapitel wird kurz der Zusammenhang zwischen den Proteinen Byr2 von *Schizosaccharomyces Pombe* und der monomeren GTPase Rat sarcoma (Ras) des Menschen erläutert. Des Weiteren wird auf das Protein IE1 des Cytomegalievirus und das mit der Alzheimerschen Krankheit verknüpfte Protein β -Amyloidpeptid ($A\beta$) eingegangen. An den Daten zu diesen Makromolekülen - NMR-spektroskopischen Messwerten und Strukturen - wurden die in dieser Arbeit entwickelten Methoden hauptsächlich getestet.

1.2.1 G-Protein gekoppelte Signaltransduktion: Ras und Byr2

Ras - Signalknotenpunkt bei der Regulation von Wachstums- und Differenzierungsprozessen

Die monomere GTPase Ras ist zentral an der Regulation verschiedener Wachstums- und Differenzierungsprozesse beteiligt, indem sie Signale von Zelloberflächenrezeptoren weitergibt. Beim Menschen unterscheidet man vor allem zwischen H-Ras, K-Ras und N-Ras, die zwar leicht unterschiedliche Funktionen besitzen, aber von denen vermutet wird, dass sie auf sehr ähnliche Weise als molekulare Schalter dienen: Ras kann zwischen zwei verschiedenen Konformationszuständen wechseln, dem aktiven - GTP gebundenen - und dem inaktiven - GDP gebundenen - Zustand. Die Lebensdauer der beiden Formen wird durch zwei Klassen von regulatorischen Proteinen bestimmt, den Guanin-Nukleotid-Austauschfaktoren (GEFs), die sich an die GTPase anlagern und dadurch den Austausch des gebundenen GDP zu GTP bewirken, und den GTPase-aktivierenden Proteinen (GAPs), die den Vorgang der GTP-Hydrolyse beschleunigen (Wittinghofer und Vetter, 2011; Tamanoi, 2013).

Ras wird seit mehreren Jahrzehnten intensiv erforscht, da Punktmutationen an diesem Protein mit 30% aller bekannten Krebserkrankungen assoziiert werden. Durch die Veränderungen in der Primärsequenz entstehen Moleküle, die immer in der aktiven Form vorliegen und damit zu unkontrollierter Proliferation der Krebszellen beitragen (Rajalingam et al., 2007; Schubbert et al., 2007; Karboub und Weinberg, 2008).

Proteine, die wie Ras an der Signaltransduktion beteiligt sind, liegen typischerweise in mehr als einem Konformationszustand vor. Im Ras-Zyklus wechselwirkt dieses Protein mit mindestens acht verschiedenen Bindungspartnern mit zum Teil unterschiedlichen Bindungsstellen, woraus acht funktionelle Zustände resultieren (Kalbitzer et al., 2013). Zwei der konformationellen Hauptzustände im aktiven Ras, der Zustand bei der Bindung an GEFs und der bei der Bindung an Effektoren, konnten direkt an Ras•Mg²⁺•GppNHp mit ³¹P-NMR-Spektroskopie unterschieden werden (Geyer et al., 1996; Spörner et al., 2001, Spörner et al., 2005; Kalbitzer et al., 2009). Ras im Komplex mit GTP wurde von Spörner et al. bereits untersucht (Spörner et al., 2010).

In einer Arbeit von Kalbitzer et al. konnten die vier Zustände, die im aktivierten Ras erwartet werden, durch Hochdruck-NMR-spektroskopische Experimente nachgewiesen werden (Kalbitzer et al., 2013). Dazu wurden sowohl die chemischen Verschiebungen als auch die

Volumina der Signale ausgewertet und die interessierenden Parameter durch Anpassung an thermodynamische Modelle bestimmt (Abbildung 1-3). Die Besetzungszahlen der Zustände ließen sich aus den Differenzen der molaren Gibbs-Energien ΔG_{ij}^0 und der partiellen molaren Volumina ΔV_{ij}^0 zwischen den Zuständen i und j berechnen (Tabelle 1-1). Die zu diesen Untersuchungen gewonnenen Datensätze wurden im Rahmen dieser Arbeit zum Testen der neu entwickelten Methode verwendet.

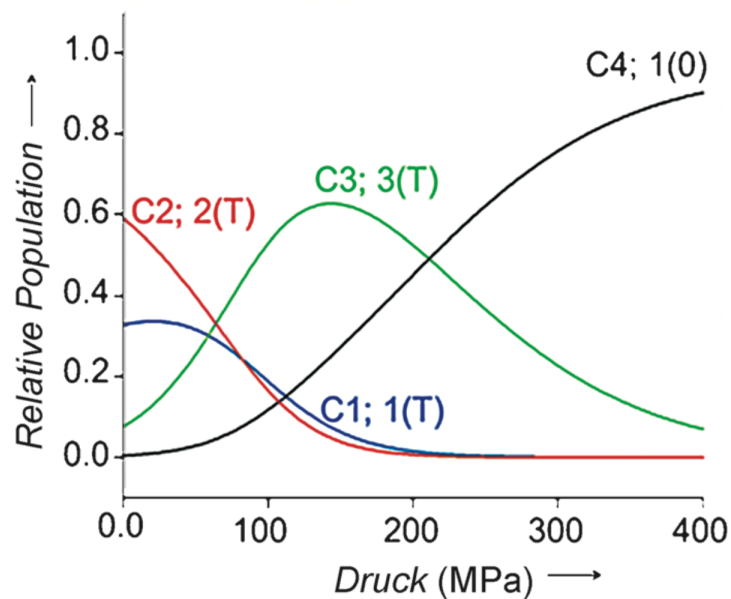


Abbildung 1-3: Populationsverhältnisse beim aktiven Ras in Abhängigkeit vom Druck

Die relative Populationen der verschiedenen Konformationszustände von aktivem Ras im Komplex mit dem nicht-hydrolysierenden GTP-Analog GppNHp:

- C1: Zustand, an den GEFs binden
- C2: Zustand, an den Effektoren binden
- C3: Zustand, an den GAPs binden
- C4: Zustand, bei dem vermutlich das Nukleotid freigesetzt wird

(aus: Kalbitzer et al., 2013)

Tabelle 1-1: Gibbsenergien der Konformationsübergänge bei Ras

Konformationsübergänge und zugehörige Differenzen der molaren Gibbs-Energie und der partiellen molaren Volumina von Ras im Komplex mit dem nicht-hydrolysierenden GTP-Analog GppNHp bei 303 K

(Werte aus: Kalbitzer et al., 2013)

Übergang	ΔG_{ij}^0 [kJ mol ⁻¹]	ΔV_{ij}^0 [kJ mol ⁻¹]
C2→ C1	1,5 ± 0,2	-18 ± 1
C2→ C3	5,2 ± 0,3	-81 ± 3
C2→ C4	12,4 ± 0,4	-115 ± 2

Byr2 - Vermittlung der Signalweiterleitung

Das Protein Byr2 ist eine Serin/Threonin spezifische Kinase des Hefepilzes *Schizosaccharomyces Pombe*. Dieser Hefepilz wurde erstmals 1893 aus ostafrikanischem Hirsebier isoliert und dient seitdem als Modellorganismus in der Zell- und Molekularbiologie.

Untersucht wird in der vorliegenden Arbeit die zweite der insgesamt vier Domänen von Byr2, die mit Ras1 (Byr2-RBD, Ras-bindende Domäne) interagiert. Bei den Hefepilzen übernimmt Byr2 die Rolle von Raf bei Säugetieren im MAP-Kinase Zyklus. Es fungiert somit als Effektorprotein und vermittelt die Signalweiterleitung. Wie andere Ras-bindende Proteine könnte es als Ansatzpunkt dienen, die Aktivität von Ras zu beeinflussen (Masuda et al., 1995; Tu et al., 1997).

Die Abfolge seiner Sekundärstrukturelemente ist $\beta\beta\alpha\beta\beta\alpha$ und entspricht dem Faltungsmuster der Familie von Ubiquitin. In der PDB-Datenbank (<http://www.pdb.org>) (Berman et al., 2000) findet man zwei Einträge:

Die unter der Nummer PDB-1I35 abgelegte Struktur ist ein Bündel aus zehn Modellen (jeweils Aminosäuren 71-165, die jedoch mit 1-95 nummeriert wurden). Die Strukturbestimmung erfolgte mithilfe von Kernspinresonanzexperimenten in Lösung am Konstrukt Byr2(65-180). Eine Zuordnung der chemischen Verschiebungen liegt für alle Aminosäuren außer P153, S154, R160 und P180 vor. Für die Region P153 bis E163 waren nur sehr schwache Signale zu finden, weswegen auch nur ein rudimentäres Seitenkettenassignment möglich war. Die Proteinstrukturbestimmung wurde erschwert durch die Neigung des Proteins zu schwachen unspezifischen Aggregationen, die sich in Resonanzlinienverbreiterung äußerten, und durch vermutlich verschiedene Konformationen

im C-terminalen Bereich. Zusätzlich zeigten manche Signale eine Aufspaltung in mehrere Resonanzpeaks, woraus mehrere Zustände im langsamen Austausch miteinander geschlossen wurden. Nur ein Teil dieser Aufspaltungen kann durch die Cis-Trans-Isomerie benachbarter Proline erklärt werden (Huber et al., 2000; Gronwald et al., 2001).

Zur Erstellung der Struktur PDB-1K8R wurde eine Röntgenstrukturanalyse des Komplexes aus GppNHp gebundenem H-Ras(1-166) mit Byr2(71-180) durchgeführt. Die erzielte Auflösung beträgt 0,3 nm (Scheffzek et al., 2001).

Von Wohlgemuth et al. wurden die Ras-bindenden Domänen verschiedener Proteine verglichen (Wohlgemuth et al., 2005). Dabei zeigte sich, dass die Primärsequenz von Byr2 nur eine Sequenzähnlichkeit von nur 14-16% zu anderen bekannten Ras-bindenden Proteinen wie AF6, RalGEF oder C-Raf aufweist. Trotzdem besitzt das Makromolekül das gleiche Faltungsmuster wie die Vergleichsproteine.

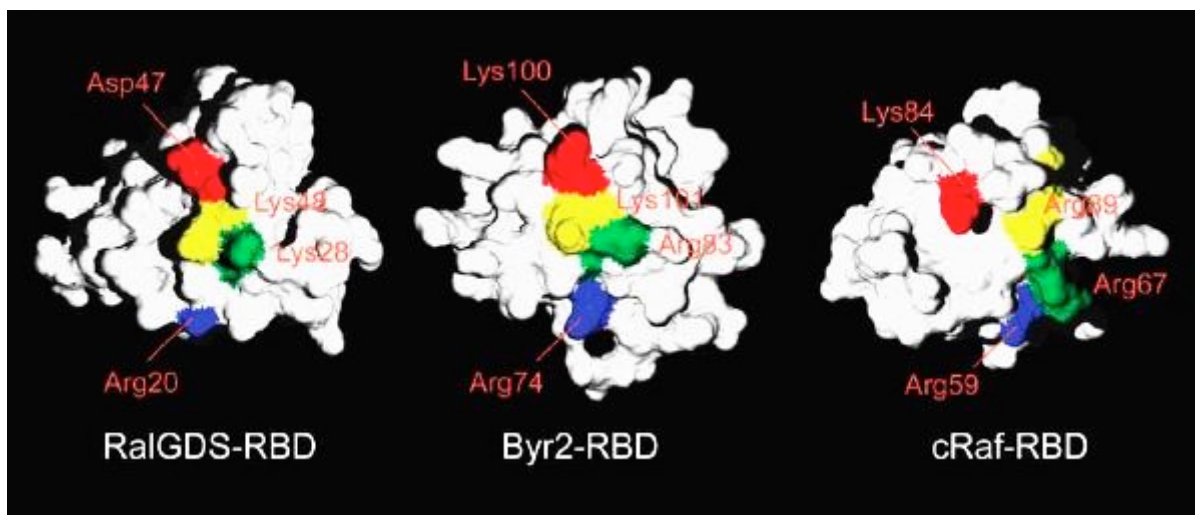


Abbildung 1-4: Strukturen verschiedener Ras-bindenden Proteine

Strukturen der Ras-bindenden Proteine RalGDS-RBD, Byr2-RBD und C-Raf-RBD im Vergleich. Farblich dargestellt die positiv geladenen Aminosäuren, die als besonders wichtig für die Bindung an Ras eingeschätzt werden und die im Alignment für das erstellte Homologiemodell einander zugeordnet sind.

(aus Wohlgemuth et al., 2005)

Des Weiteren war ein Sequenzalignment zwischen den betrachteten Proteinen und Byr2-RBD erstellt worden, was aber nur unter Einbeziehung der Sekundärstruktur möglich war. Eine Untersuchung der Bindestelle und der Art der Wechselwirkung der Ras-bindenden Proteine ergab, dass die Bindung an Ras jeweils durch die Wasserstoffbrücken zweier intermolekularer

β -Faltblätter vermittelt wird. Zusätzlich wird sie durch die Polarität der Seitenketten der wechselwirkenden Aminosäuren verstärkt, wobei aufseiten der Ras-bindenden Domänen jeweils positiv geladene Reste - Arginine und Lysine - der ersten beiden β -Faltblätter und der ersten Helix kennzeichnend für die Bindung sind (Abbildung 1-4) (Wohlgemuth et al., 2005).

Während für Byr2-RBD bisher keine Ergebnisse aus Untersuchungen mit Hochdruckexperimenten zur Identifizierung verschiedener Zustände publiziert wurden, ist das strukturhomologe Protein Ubiquitin eingehend mit dieser und anderen Techniken untersucht worden. Dabei stellte sich heraus, dass bei Normaldruck zwei gefaltete Zustände koexistieren (Kitahara et al., 2001), von denen der bei Normaldruck weniger besetzte als Intermediat der Faltung identifiziert werden konnte, wobei sich die Strukturänderungen im Vergleich zum Grundzustand vor allem im C-Terminus lokalisieren lassen (Kitahara und Akasaka, 2003; Jackson, 2006). Weitere Untersuchungen zeigten, dass sich dieser Zustand durch Mutation einer einzigen Aminosäure (Q41N) stabilisieren lässt und die resultierenden Strukturänderungen genau den Änderungen entsprechen, die das Protein bei einer Bindung mit dem aktivierenden Enzym E1 ausführt. Der Erkennungsvorgang von Ubiquitin und E1 ist deshalb ein sehr schönes Beispiel für eine konformationelle Selektion, bei welcher der bei der Bindung eingenommene Zustand bereits ohne Vorliegen des Bindungspartners als schwächer populierter Zustand nachgewiesen werden kann (Kitazawa et al., 2014). Außerdem konnte belegt werden, dass die erste Helix der Struktur Schwingungen ausführt, bei denen sich die Atome um bis zu 0,3 nm hin- und herbewegen (Kitahara 2005).

Auch das zu Byr2-RBD strukturhomologe Protein RalGDS-RBD wurde mit Hochdruckexperimenten untersucht, und strukturelle Veränderungen, die die Bindung an Rap1A begünstigen, konnten nachgewiesen werden (Inoue et al., 2000; Inoue et al., 2001).

Autoinhibierung als allgemeines Prinzip zur Regulierung der Kinaseaktivität

Autoinhibierung ist ein allgemeines Prinzip, das bei vielen Proteinen bekannt ist. Dabei legt sich ein Teil des Proteins selbst über die Bindestelle, so dass die Signalweitergabe blockiert wird. Zum Beispiel ist die Proteinkinase C in Abwesenheit von aktivierenden Cofaktoren einer Autoinhibierung durch die regulatorische Domäne unterworfen (Orr und Newton, 1994).

Auch für die GEFs RasGRP1 und SOS ist bekannt, dass sie durch dieses Prinzip die Interaktion mit Ras steuern können. In beiden Fällen wird die Ras-Bindestelle ebenfalls durch einen interdomainen Linker blockiert (Iwig et al., 2013; Sondermann et al., 2004).

1.2.2 A β und Alzheimersche Krankheit

Die Alzheimer Krankheit ist eine irreversible neurodegenerative Erkrankung und häufigste Form der primären Demenz. Der erste Fall der Erkrankung wurde bereits 1901 von Alois Alzheimer beschrieben. Klinisch präsentieren sich die Patienten mit zunächst progredienten Limitationen des Kurzzeitgedächtnisses, während im Langzeitgedächtnis gewohnte Tätigkeitsabläufe und emotionale Erlebnisse meist noch gut präsent sind. Im weiteren Verlauf manifestieren sich ein reduziertes Sprachvermögen mit gehemmtem Sprachfluss und reduziertem Vokabular sowie Verhaltensauffälligkeiten. Auch fallen zunehmend Störungen der Feinmotorik auf, so dass die Betroffenen bereits früh bei komplexeren Tätigkeiten auf Unterstützung angewiesen sind. Im fortgeschrittenen Stadium kommt es zu einem zunehmenden kognitiven Verfall der Patienten, die dann selbst nahestehende Angehörige nicht mehr erkennen und die Affektkontrolle verlieren. Die Betroffenen sind mit einfachsten Anforderungen des Alltags, wie beispielsweise das morgendliche Aufstehen aus dem Bett, überfordert. Ein typisches Symptom ist die Fortbewegung durch unsichere Trippelschritte. Auch stellt sich meist eine Harn- und Stuhlinkontinenz ein. Häufig werden die Patienten im Endstadium der Erkrankung vollkommen immobil und sind gänzlich auf fremde Hilfe angewiesen. Die verbleibende Lebenserwartung nach Diagnosestellung beträgt etwa sieben bis zehn Jahre. Von der Alzheimer Krankheit sind fast ausschließlich Menschen höheren Alters betroffen.

Pathognomonisches histomorphologisches Korrelat im Gehirn der Betroffenen sind Amyloid-Ablagerungen, deren Hauptbestandteil das β -Amyloid-Peptid (A β) ist. Gemäß der allgemein akzeptierten Amyloidhypothese ist A β die primäre Ursache der Alzheimer Krankheit (Hardy und Selkoe, 2002). A β entsteht aus einem Vorläuferprotein, dem Amyloid-Precursor-Protein (APP), bei dem es sich um ein integrales Membranprotein handelt. Durch die Proteinasen β - und μ -Sekretase können aus APP die beiden Hauptprodukte A β (1-40) und A β (1-42) freigesetzt werden, die aus 40 beziehungsweise 42 Aminosäuren bestehen. Es wird vermutet, dass das A β als Teil des APP in der Membran eine α -helikale Konformation einnimmt, und auch in Lösung zeigen sich teilweise gefaltete, α -helikale Strukturen (Temussi et al., 2003), wie auch mit NMR-Spektroskopie nachgewiesen werden konnte (Vivekanandan et al., 2011).

Das pathogenetische Grundproblem der Alzheimer Krankheit scheint die Aggregation der entstehenden Moleküle zu Oligomeren zu sein (Haass und Selkoe, 2007): Es bilden sich die typischen Fibrillen der Amyloid-Plaques, in denen das A β eine sich überkreuzende β -Faltstrangstruktur („cross beta“ structure) einnimmt (Petkova et al., 2006).

Dazu stellt sich die Frage, wie dieser Übergang von teilweise gefalteten, α -helikalen Strukturen zu einer Quartärstruktur mit β -Faltblattcharakter zustande kommen kann, insbesondere, ob dabei stabile Intermediate auftreten, die sich mit dem in Kapitel 1.1.5 vorgestelltem Ansatz über die Beobachtung der chemischen Verschiebungen des Peptids bei verschiedenen Drücken nachweisen lassen. Diese Idee wurde bereits 2013 verfolgt und aus den durchgeführten Experimenten das Vorliegen von drei deutlich unterscheidbaren Zuständen bei Normaldruck geschlossen (Munte et al., 2013). Abbildung 1-5 zeigt die relativen Populationen der drei Konformationszustände. Der vorliegende Datensatz soll nun als Testdatensatz für die in dieser Arbeit entwickelten Methode zum Anpassen der Parameter der Gibbs-Energiedifferenzen dienen.

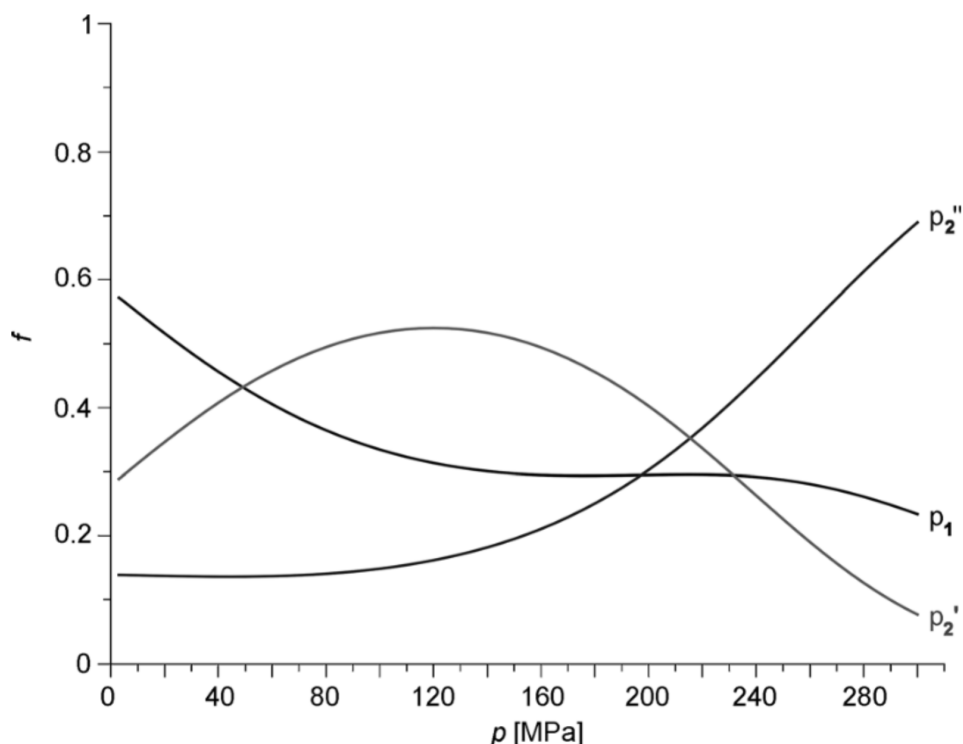


Abbildung 1-5: Populationsverhältnisse von A β in Abhängigkeit vom Druck

Relative Populationen der drei Konformationszustände von A β bei einer Temperatur von 277 K

(aus Munte et al., 2013)

1.2.3 IE1 und Persistenz des Cytomegalievirus

Das Cytomegalievirus (CMV) ist ein Vertreter der Herpesvirusfamilie (Humanes Herpesvirus 5, HHV-5), das von mehr als 50% der Weltbevölkerung in sich getragen wird. Die Übertragung erfolgt entweder intrauterin, über die Muttermilch oder durch engen Kontakt. Die Infektion bei einem gesunden Erwachsenen erfolgt meist asymptomatisch, selten treten Splenomegalie, Fieber oder Hepatitis auf. Klinisch bedeutsam wird das Virus bei intrauteriner Infektion - was zu neurologischen Schäden führen kann - oder bei Reaktivierung oder Neuinfektion unter Immunsuppression. Außerdem ist das Virus für Komplikationen bei Transplantationen verantwortlich (Paulus und Nevels, 2009).

Das Virus besitzt eine Eigenschaft, die allen Herpesviren gemein ist, nämlich die lebenslange Persistenz im Körper. Jedoch bauen diese Viren ihr Genom nicht in die menschliche DNS ein. Für ein anderes Mitglied der Herpesfamilie, HHV-8, dem Karposi-Sarkom assoziierten Herpesvirus, konnte gezeigt werden, dass das virale Genom an die menschliche DNS angeheftet wird. Dadurch geht während der latenten Phase, in der kaum neue Virus-DNS produziert wird, das virale Genom nicht verloren und wird an die Tochterzellen weitergegeben. Dieses Anheften des Viruserbguts an die menschlichen Histone erfolgt mithilfe des Proteins LANA. Die Röntgenkristallstruktur PDB-1ZLA enthält ein Modell von LANA(4-17) im Komplex mit Fragmenten menschlicher Histone und DNS (Barbera et al., 2006; Feeney und Parish, 2009).

Das Protein IE1 des Cytomegalievirus wird besonders in der ersten Phase der Proteinexpression (Immediate Early Phase, deswegen der Name IE1) produziert und wird ebenfalls mit Persistenz im menschlichen Körper verknüpft (Paulus und Nevels, 2009). Experimente der Arbeitsgruppe von Michael Nevels und Christina Paulus des Instituts für Mikrobiologie und Hygiene in Regensburg konnten zeigen, dass IE1 an die Virus-DNS bindet. Außerdem wurde vermutet, dass es genauso wie LANA mit den menschlichen Histonen eine Bindung eingehen kann und dass diese Bindung C-terminal geschieht. Verdrängungsexperimente mit LANA ließen schließen, dass IE1 an dieselbe Stelle des Chromatins - der sogenannten sauren Tasche - wie dieses bindet.

Diese saure Tasche, eine Furche, gebildet durch die Histone H2A und H2B an der Oberfläche des Chromatins, fungiert auch als Bindestelle für verschiedene weitere Proteine (Kalashnikova et al., 2013). Außer für den eben genannten Komplex mit LANA gibt es

weitere Kristallstrukturen dieser sauren Tasche: im Komplex mit dem Protein RCC1 (PDB-3MVD), einem GEF des Ran-Proteins, einer weiteren GTPase der Ras-Superfamilie, und im Komplex mit SIR3 (Silent Information Regulator) (PDB-3TU4) von *Saccharomyces cerevisiae* (Makde et al., 2010; Armache et al., 2011).

Im Rahmen dieser Arbeit stellte sich die Frage, in welcher Weise diese Bindung von IE1 an die menschlichen Histone geschehen könnte. Die kürzlich veröffentlichte Kristallstruktur (Abbildung 1-6) von rhIE1 (PDB-4WID), der IE1-Variante des bei Rhesusaffen vorkommenden Cytomegalievirus, dessen Primärsequenz dem IE1 des für den Menschen relevanten Virus sehr ähnlich ist, kann darauf keine Antwort geben, da der C-terminale Bereich, der als verantwortlich für die Wechselwirkung identifiziert wurde, in der Struktur fehlt (Scherer et al., 2014). Informationen können aber aus den obengenannten Strukturen PDB-1ZLA, PDB-3MVD und PDB-3TU4 gewonnen werden, um ein Homologiemodell des Komplexes des C-Terminus von IE1 und dem menschlichen Chromatin zu erstellen. Dieser Ansatz wurde in der vorliegenden Arbeit gewählt.

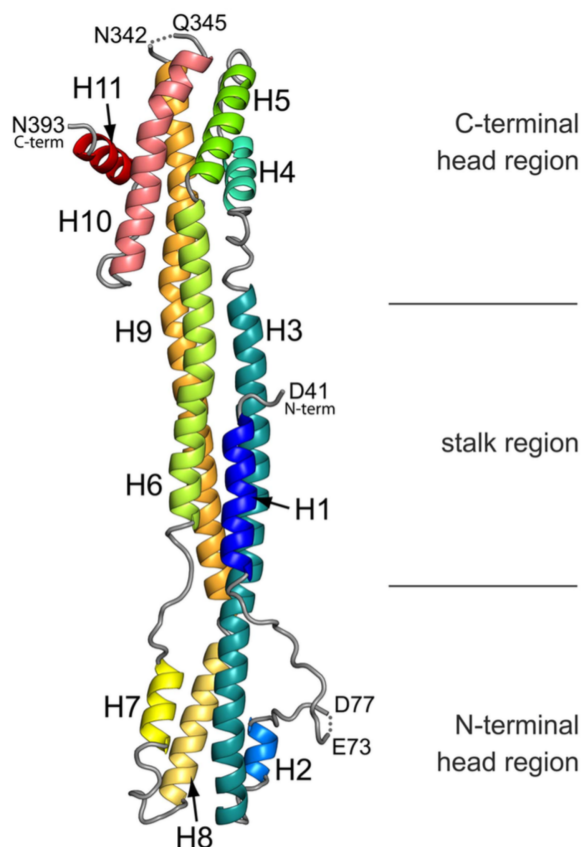


Abbildung 1-6: Struktur von rhIE1(41-393), PDB-4WID

Die mithilfe von Röntgenkristallographie gewonnene Struktur besteht aus 11 Helices. Für die C- und die N-terminalen Aminosäuren konnte aufgrund nur schlecht definierter Elektronendichte keine Aussage gemacht werden.

(aus Scherer et al., 2014)

2. Materialien und Methoden

2.1 Untersuchte Proteine und Datensätze

2.1.1 Byr2-RBD

Als Testprotein für das Programm PERMOL wurde in der vorliegenden Arbeit die Ras bindende Domäne von Byr2 verwendet. Dazu dienten die beiden in Kapitel 1.2.1 genannten Strukturen PDB-1I35 (Gronwald et al., 2001) und PDB-1K8R (Scheffzek et al., 2001) und ein Set von experimentellen Restraints, das von Wolfram Gronwald zur Verfügung gestellt und zur Strukturberechnung von 1I35 verwendet worden war. Es umfasst 824 Abstandsrestraints, 29 Wasserstoffbrückenrestraints und 28 residuale Restkopplungsrestraints. Die dihedralen Winkelrestraints wurden dagegen nicht übernommen.

Des Weiteren waren ein ^1H - ^1H -NOESY-Spektrum des Proteins und eine Liste chemischer Verschiebungen verfügbar (Huber et al., 2000), die unter der Nummer 4463 der Biological Magnetic Resonance Data Bank (BMRB) (<http://bmrb.wisc.edu>) (Ulrich et al., 2008) zu finden sind. Außerdem dienten die beiden Strukturen PDB-4G0N (Fetics et al., 2015) und PDB-1LFD (Huang et al., 1998) der PDB-Datenbank als Modellierungsgrundlage zur Erstellung von Homologiemodellen des Komplexes Byr2-RBD mit Ras mithilfe von PERMOL.

Zum Nachweis der Proteinzustände mit Hochdruck standen TROSY-HSQC-Spektren von Byr2-RBD zur Verfügung, die am selben Konstrukt wie oben beschrieben aufgenommen worden waren. Die Aufnahmen entstanden 2005 in Japan in Zusammenarbeit von Werner Kremer und Ryo Kitahara. Als Druckpunkte wurden 3 MPa, 25 MPa, 50 MPa, 75 MPa, 100 MPa, 125 MPa, 150 MPa, 175 MPa, 200 MPa und 225 MPa gewählt. Allerdings musste das als letztes aufgenommene Spektrum (175 MPa) zur Auswertung verworfen werden, da das Protein nicht so lange stabil war. Die Messungen fanden bei einer Temperatur von 298 K und einem pH von 6,9 statt. Die Spektren wurden an einem DRX-800 Spektrometer (Bruker, Billerica, USA) aufgenommen, die ^1H Frequenz betrug dabei 800,13 MHz, die ^{15}N Frequenz 81,08 MHz. Die Druckexperimente fanden in druckresistenten Quarzzellen statt.

In der direkten Dimension wurden 2048, in der indirekten jedoch nur 64 Datenpunkte als komplexe Datenmatrix aufgenommen.

Die Proben enthielten 25 mM DTE, 200 mM deuteriertes Glycin, 20 mM Phosphatpuffer,

0,5 mM EDTA, 0,5 mM NaN_3 und 0,1 mM DSS zur Referenzierung.

Das Vorgehen zur Gewinnung von Byr2(65-180) wurde in der Dissertation von Fritz Huber ausführlich beschrieben (Huber, 2001). Byr2 wurde in *Escherichia coli* Stämmen BL21 DE3 unter Verwendung des Plasmids pGEX4T3 (GE Healthcare Life Sciences, Chalfont St Giles, UK) exprimiert. Das erhaltene Fusionsprotein GST-Byr2 wurde auf eine Ni-NTA-Säule aufgetragen. Nach der Elution unspezifisch bindender Proteine wurde das Fusionsprotein durch Inkubation bei 4°C unter Anwesenheit von 10 U/ml Thrombin (SERVA, Heidelberg) über Nacht auf der Säule gespalten. Danach wurde Byr2 von der Säule eluiert und auf 20 mg/ml konzentriert. Dazu wurden Zentrifugalkonzentratoren (Sigma-Aldrich, St. Louis, USA) verwendet. Schließlich wurde Byr2 dann mittels Größenausschluss-Chromatographie weiter gereinigt und in 20 mM Kaliumphosphat, 0,5 mM EDTA, 0,5 mM NaN_3 and 3 mM DTE bei pH 6,9 überführt. Nur der monomere Anteil (über 95%; abgeschätzt mithilfe von SDS Polyacrylamid Gelelektrophorese) wurde für die weiteren Experimente verwendet. Das Protein wurde schließlich wieder auf 20 mg/ml konzentriert, in flüssigem Stickstoff schockgefroren und bei -80°C bis unmittelbar vor den jeweiligen NMR-Messungen aufbewahrt.

Für die Isotopenmarkierung (^{15}N und ^{13}C) wurde Minimalmedium anstelle des Standardmediums I für das Wachsen der Zellen verwendet. 1 l Medium enthielt dabei 1 g $^{15}\text{NH}_4\text{Cl}$ (> 98%, Cambridge Isotope Laboratorien), 10 g Glucose, 7,5 g $\text{Na}_2\text{HPO}_4 \cdot 2\text{H}_2\text{O}$, 3 g KH_2PO_4 , 0,5 g NaCl , 0,25 g $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$, 14 mg $\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$, 50mg Ampicillin, 100 μg $\text{ZnSO}_4 \cdot 7\text{H}_2\text{O}$, 30 μg MnCl_2 , 300 μg H_3BO_3 , 200 μg $\text{CoCl}_2 \cdot 6\text{H}_2\text{O}$, 10 μg $\text{CuCl}_2 \cdot 2\text{H}_2\text{O}$, 20 μg $\text{NiCl}_2 \cdot 6\text{H}_2\text{O}$, and 30 μg $\text{Na}_2\text{MoO}_4 \cdot 2\text{H}_2\text{O}$. Am Ende wurden noch 5 mg EDTA und 2 mg $\text{FeSO}_4 \cdot 2\text{H}_2\text{O}$ zugefügt. Ampicillin und Schwermetallkomplex Lösungen wurden steril gefiltert, bevor sie zum autoklavierten Medium gegeben wurden. Für nur ^{15}N Markierung wurde die Inkubation der Zellen wie oben beschrieben durchgeführt. Für das doppelt markierte Protein wurde die Glucose durch 2 g/l $^{13}\text{C}_6$ -Glucose (>99%, CIL) ersetzt, und die Zellen wurden nach dem Animpfen für 6 Stunden bei 35°C inkubiert.

Die Strukturbestimmung erfolgte mithilfe von NMR-Experimenten bei einer Temperatur von 298 K und einem pH von 6,9 in den Spektrometern DRX 600 und DRX 800 (Bruker, Billerica, USA) (Huber et al., 2000; Gronwald et al., 2001).

2.1.2 IE1-CTD

Als Vorlage zur Berechnung des Homologiemodells wurde die Struktur PDB-1ZLA der PDB-Datenbank benutzt. Die experimentelle Auflösung beträgt 0,29 nm.

Außerdem standen von Hans Robert Kalbitzer aufgenommene NMR-Spektren des Proteins zur Verfügung. Die Aufnahmen fanden an zwei verschiedenen Konstrukten des Proteins statt, zum einen an rekombinantem IE1 in voller Länge von 491 Aminosäuren (IE1-Full, 1-491), zum zweiten an einer C-terminal um 16 Aminosäuren verkürzten Variante (Aminosäuren 1-475), die beide von der Arbeitsgruppe von Michael Nevels und Christina Paulus zur Verfügung gestellt wurden. Die vermessenen Proben enthielten 95 μM (IE1-Full) bzw. 83 μM (IE1-1-475) Protein in PBS mit 1 mM DTT in $^1\text{H}_2\text{O}/^2\text{H}_2\text{O}$ (92%/8%) beziehungsweise $^2\text{H}_2\text{O}$ bei einem pH von 6,5. Zur Referenzierung der chemischen Verschiebungen der Wasserstoffatome war als interner Standard Natrium-2,2-dimethyl-2-silapentan-2-Sulfonsäure (DSS) beigemischt.

Die NMR-Experimente wurden an einem Bruker Avance 600 Spektrometer, das mit einem 5mm TXI Kryoprobenkopf ausgestattet ist, bei einer ^1H -Frequenz von 600,03 MHz bei einer Temperatur von 298 K aufgenommen. Für die 1D ^1H -NOESY-Experimente mit Vorsättigung während der Relaxationszeit wurde eine Mischzeit von 10 ms gewählt. Unter Verwendung einer MLEV-17 Entkoppelungssequenz von 70 ms wurden ^1H - ^1H 2D NOESY Spektren mit einer Mischzeit von 300 ms und ^1H - ^1H 2D TOCSY Spektren zur Zuordnung der chemischen Verschiebungen der Atome der mobilen Regionen aufgenommen.

2.1.3 Alzheimer- β -Amyloidpeptid (A β)

Zur Datenakquisition am Alzheimer- β -Amyloidpeptids wurde eine Reihe von ^1H - ^{15}N -HSQC-Spektren bei 11 verschiedenen Druckpunkten (3 MPa, 20 MPa, 40 MPa, 60 MPa, 80 MPa, 100 MPa, 120 MPa, 140 MPa, 160 MPa, 180 MPa, 200 MPa) bei einer Temperatur von 277 K aufgenommen. Die Messungen erfolgten an einem Bruker Avance 600 Spektrometer mit einer ^1H -Frequenz von 600,13 MHz. Die Probe enthielt 474 μM ^{15}N -angereichertes humanes A β (1-40), das den Aminosäuren 672 – 711 von βAPP770 der SwissProt-Datenbank (<http://www.ebi.ac.uk/uniprot>) (Bairoch, Boeckmann et al. 2004), Eintrag P05067, entspricht, in 50 μM $[\text{D}_{11}]$ Tris(hydroxymethyl) aminomethan, 90 mM NaCl, 50 μM 4,4-Dimethyl-4-silapentan-1-Sulfonsäure, 0,1 mM Dioxan, 1 mM NaN_3 , 0,5 mM $[\text{D}_{16}]$ Ethylen-diamintetraessigsäure und 8% $^2\text{H}_2\text{O}$ bei einem pH von 7,0 (Munte et al., 2013).

Für 31 der 40 Aminosäuren waren die chemischen Verschiebungen der Stickstoffatome und der zugehörigen Wasserstoffatome des Proteinrückgrats für alle Druckpunkte vollständig und wurden zu dem hier untersuchten Datensatz zusammengefasst.

Die Daten wurden von Claudia Munte zur Verfügung gestellt. Das Vorgehen zur Präparation der Proben wurde von Munte et al. ausführlich beschrieben (Munte et al., 2013). Um einheitliches, nicht-aggregiertes A β (1-40) Peptid zu bekommen erfolgte eine Vorbehandlung der Proben mit Hexafluoro-2-propanol (HFIP, Sigma-Aldrich, St. Louis, USA). 1 mg des lyophilisierten Peptids wurde in 2 ml HFIP verdünnt und für 15 min in ein Ultraschallbad gegeben. Nach einer 60 minütigen Äquilibrierung der resultierenden klaren Lösung wurde HFIP wieder verdampft, der resultierende Peptid-Film in flüssigem Stickstoff schockgefroren und über Nacht unter Vakuum (0,025 MPa) getrocknet. Das vorbehandelte Peptid wurde dann in 0,2 ml einer 100 mM NaOH Lösung resuspendiert und der pH unter Zugabe geringer Mengen an HCl auf 7,0 adjustiert. Das gewünschte Volumen wurde durch Zugabe einer Pufferlösung mit exakt dem gleichen pH erzielt, die geeignete Mengen an Tris-d₁₁, NaCl, EDTA-d₁₆, DSS und entweder ¹H₂O oder ²H₂O enthielt. Die fertigen NMR-Proben enthielten typischerweise 0,5, 0,75 oder 1 mM A β (1-4) in 50 mM Tris-d₁₁, pH 7.00, 90 mM NaCl, 0.5 mM EDTA-d₁₆, 0.05 oder 0.1 mM DSS in entweder 90% ¹H₂O/10% ²H₂O oder 99.5% ²H₂O. Zur Entfernung größerer Aggregate wurden manche Proben bei 50,000 g und 277 K für 3 Stunden zentrifugiert und nur der Überstand wurde hier für die NMR-spektroskopischen Experimente verwendet.

2.1.4 Ras-Protein

Wild-Typ und Mutanten des H-Ras(1-166) wurden unter Verwendung von ptac-Vektoren in *Escherichia coli* Kulturen (Stamm BL21) exprimiert und gemäß der Methode von Tucker et al. isoliert und gereinigt (Tucker et al., 1986). Der Nukleotidaustausch zu GppNHp erfolgte in Analogie zu dem von John et al. beschriebenen Procedere (John et al., 1990). Eine einheitliche Anreicherung mit ¹⁵N Isotopen wurde unter Verwendung des „new minimal medium“ (Budisa et al., 1995) durchgeführt, das 1 g/l ¹⁵NH₄Cl enthält. Freie Nukleotide und Phosphate wurden mittels Gel-Filtration entfernt. Der resultierende Reinheitsgrad des Proteins war > 95% (gemäß SDS-PAGE).

Zur Untersuchung der verschiedenen Konformationszustände des Ras-Proteins fanden ¹H-¹⁵N-HSQC-Experimente an Ras(1-166) im Komplex mit dem nicht-hydrolysierenden GTP-Analog GppNHp statt. Die Probe enthielt 1,84 mM ¹⁵N-isotopen-markiertes

Ras(wt)·Mg²⁺· GppNHp in 40 mM Tris/HCl, 10 mM MgCl₂, 2 mM DTE, 0,1 mM DSS und 12% D₂O. Gemessen wurde bei einem pH-Wert von 7,4 und einer Temperatur von 303 K. Die experimentellen Auflösungen betrugen 9,390 Hz (¹H) bzw. 11,087 Hz (¹⁵N). Die Messungen wurden an einem Bruker Avance-800 NMR-Spektrometer durchgeführt mit den Resonanzfrequenzen von 800,12 MHz für ¹H und 81,08 MHz für ¹⁵N. Die Spektrenaufnahme erfolgte bei 3 MPa, 20 MPa, 40 MPa, 60 MPa, 81 MPa, 101 MPa, 120 MPa, 140 MPa, 160 MPa und 180 MPa. (Kalbitzer et al., 2013)

Bei diesem Datensatz war die zur Auswertung notwendige Druckkorrektur mit Random-Coil-Daten nach Köhler et al. bereits durchgeführt worden (Köhler et al., 2012). Die chemischen ¹H- und ¹⁵N-Verschiebungen konnten für 116 Aminosäuren für alle 10 Druckpunkte registriert und zu dem hier untersuchten Datensatz zusammengefasst werden.

Die Spektren waren von Ina Rosnizeck aufgenommen und zur Verfügung gestellt worden.

2.2 Theorie

2.2.1 Molekulardynamik

Molekulardynamik bezeichnet Computersimulationen in der molekularen Modellierung, bei denen Wechselwirkungen zwischen Atomen und Molekülen und resultierende räumliche Bewegungen iterativ berechnet und visualisiert werden. Zur Berechnung von Proteinstrukturen gibt es in diesem Kontext im Wesentlichen zwei verschiedene Ansätze. Zum einen werden *ab initio* Berechnungen durchgeführt, bei denen außer der Primärsequenz keine weiteren Informationen über die Struktur bekannt sein müssen. Das Makromolekül wird dabei zusammen mit den Molekülen des Lösungsmittels simuliert und der Faltungsprozess dadurch nachvollzogen. Diese Methode ist jedoch nach wie vor aufgrund der benötigten Rechenkapazitäten nur für kleinere Makromoleküle praktikabel. Beim zweiten Ansatz wird die virtuelle Strukturfaltung beschleunigt, indem bereits vorhandene Informationen über die Struktur des Makromoleküls aus entweder experimentellen Untersuchungen (NMR- oder Röntgenspektroskopie) oder aus strukturhomologen Proteinen gewonnene Informationen (Restraints) a priori in die Berechnungen mit einbezogen werden.

Strukturrechnungen mit CNS

Die in der vorliegenden Arbeit durchgeführten molekulardynamischen Simulationen erfolgten mit dem Programm CNS (Brünger et al., 1998). Das System wird nach der Force-Field-

Methode durch (Pseudo-)Teilchen beschrieben, die an den Orten der eigentlichen Atomkerne sitzen und denen je nach Atomart und Bindungspartner eine bestimmte Masse, Ladung und Größe zugeordnet wird. Für jedes dieser Teilchen wird eine Kraftfeldfunktion aufgestellt, die die Wechselwirkungen der Teilchen untereinander repräsentiert. Von bestimmten Startbedingungen - definiert durch Ort und Geschwindigkeit - ausgehend wird mithilfe der Newtonschen Bewegungsgleichungen ihre Geschwindigkeitsänderung in festgelegten Zeitabständen berechnet. Dann erfolgt die Bewegung, also eine Änderung der Koordinaten des Teilchens, die sich aus seiner momentanen Geschwindigkeit und der Länge des Zeitintervalls ergibt, auf die dann wieder eine Angleichung der Geschwindigkeit folgt. Dies wird so lange fortgesetzt, bis die vorgegebene Anzahl der Iterationsschritte erreicht ist. Die Kraftfeldfunktion setzt sich aus verschiedenen Komponenten zusammen, zum einen aus Potentialen, die das *a priori*-Wissen über Proteine wie Bindungslängen und favorisierte Winkel, elektrostatische oder Van-der-Waals Wechselwirkungen in die Strukturrechnung einfließen lassen und für alle Proteine gültig sind - im Folgenden physikalische Potentiale genannt - , zum anderen aus Potentialen, die aus den zusätzlich eingelesenen Restraints - experimentell erstellten oder wie hier aus homologen Proteinen gewonnenen - aufgebaut werden. Zur Beurteilung der Strukturen dient die Summe der Energien aus den Kraftfeldfunktionen der einzelnen Atome, die umso kleiner ist, je besser die Restraints und die *a priori*-Bedingungen erfüllt sind.

Die Strukturrechnung selbst erfolgt in mehreren Schritten und wird durch Aufruf von verschiedenen Skripten gesteuert, die unter struktur.usc.edu/cns/ heruntergeladen werden können:

Im ersten Schritt (Skript *generate.inp*) werden aus der eingelesenen Primärsequenz Listen erstellt, die alle Atome des Proteins sowie deren Bindungen et cetera enthalten, die bei der Berechnung der Energiefunktion später abgearbeitet werden müssen (Strukturfiles). Im zweiten Schritt (Skript *generate_extended.inp*) muss eine Startstruktur für die Simulation erzeugt werden, ein sogenannter "ausgestreckter Strang" des Proteins ohne Faltung, bei dem die Aminosäuren aneinandergehängt, die Bindungslängen angepasst und Überlappungen der Atome unterbunden werden. Dann folgt der Prozess der Faltung, in der sich die Atome bewegen dürfen. Dazu wird ein *Simulated Annealing* (Skript *anneal.inp*) durchgeführt, das das Protein durch gerichtete Atombewegung in eine energetisch günstige Konformation führt und das in mehrere Phasen unterteilt ist: eine Phase bei hoher Temperatur (also hohen

Geschwindigkeiten der Pseudoatome), dann zwei Abkühlphasen und einer abschließenden Energieminimierung, bei der die Gewichtung der einzelnen Potentiale zueinander noch einmal leicht verändert wird. Durch Variation der Anfangsbedingungen müssen sehr viele (zum Beispiel 1000) Strukturen generiert werden in der Hoffnung, das absolute Minimum der Energiehyperebene zu finden. Den Abschluss bildet ein *Waterrefinement* (Linge et al., 2003), das nur mit den besten Strukturen, also denjenigen, die die geringste Energie aufweisen, durchgeführt wird und bei dem das Protein zusammen mit Wassermolekülen simuliert wird, um die Wechselwirkungen zum Lösungsmittel zu berücksichtigen.

Die Änderungen der Geschwindigkeiten werden dabei mithilfe der Newtonschen Bewegungsgleichungen aus den Potentialen abgeleitet.

Evaluierung der Strukturen

Für die Evaluierung der Strukturen selbst und der Bindestelle bei Komplexen können folgende Methoden angewendet werden:

Berechnung des RMSD

Als Maß dafür, wie gut der zur Strukturrechnung verwendete Restraintsatz das daraus berechnete Strukturbündel definiert, kann der RMSD-Wert, der als Wurzel aus der Summe aller Abweichungen der Atomkoordinaten der Struktur berechnet wird, angewandt werden, wobei man hierzu - nach optimaler Überlagerung - paarweise alle Strukturen eines Bündels miteinander vergleicht. Für zwei Strukturen, deren N Atompositionen durch die Vektoren \mathbf{v} und \mathbf{w} gegeben sind, errechnet sich der RMSD zu

$$\begin{aligned} RMSD(\mathbf{v}, \mathbf{w}) &= \sqrt{\frac{1}{N} \sum_{i=1}^N \|\mathbf{v}_i - \mathbf{w}_i\|^2} = \\ &= \sqrt{\frac{1}{N} \sum_{i=1}^N (v_{ix} - w_{ix})^2 + (v_{iy} - w_{iy})^2 + (v_{iz} - w_{iz})^2} \quad (\text{Gleichung 1}) \end{aligned}$$

Diese Berechnung kann zum Beispiel mit den Programmen MOLMOL (Korari et al., 1996) oder PyMOL (DeLano, 2002) erfolgen.

Überprüfung der Winkel und Bindungslängen

Eine Überprüfung der Winkel und Bindungslängen der erhaltenen Strukturen kann mit dem Programm PROCHECK-NMR (Laskowski et al., 1998) durchgeführt werden. Die dihedralen Winkel werden nach Ramachandran (Ramachandran et al., 1963) beurteilt, außerdem werden Bindungslängen, Bindungswinkel und die Planarität der Ringsysteme der Seitenketten mit aus Datenbanken gewonnenen Werten verglichen. Dabei ist allerdings zu beachten, dass Abweichungen von diesen Mittelwerten nicht unbedingt als Mängel zu bewerten sind (Schrauber et al., 1993).

Überprüfung der dipolaren Restkopplungen

Wurden bei der Strukturrechnung auch dipolare Restkopplungsrestraints (Clore et al., 1998; Lipsitz und Tjandra, 2004) verwendet, muss mit dem Programm Pales überprüft werden, inwieweit die Strukturen mit den experimentellen Daten kompatibel sind (Zweckstetter, 2008).

Überprüfung des NMR-R-Faktors

Der NMR-R-Faktor (Gronwald et al., 2000) vergleicht die Volumina von Peaks eines echten NOESY-Experiments mit denen eines anhand einer Zuordnungsliste und des Strukturbündels simulierten Spektrums, dessen Volumina der Crosspeaks aus den Atomabständen der Strukturen berechnet wurden. Er kann zum Vergleich der Güte zweier berechneter Strukturen eingesetzt werden. In dieser Arbeit wurde der R-Faktor der langreichweitigen Signale verwendet, da dieser am besten die globale Orientierung der Sekundärstrukturelemente bewerten kann.

Beurteilung der Bindestelle bei Komplexen

Von Schumann et al. wurde eine Methode vorgeschlagen, die anhand von zwei Kriterien entscheidet, ob eine Aminosäure eines berechneten Proteinkomplexes der Bindestelle zugeordnet wird (Schumann et al., 2007):

- Die lösungsmittelzugängliche Oberfläche der betrachteten Aminosäure nimmt bei Komplexbildung um mindestens 5% ab. Dieses Kriterium kann mit dem Programm MOLMOL überprüft werden.
- Mindestens ein Atom der Aminosäure ist näher als 0,5 nm an einem Atom, das dem Bindungspartner angehört. Dieses Kriterium kann mit PERMOL (siehe Kapitel 2.2.2) überprüft werden, indem für den zu untersuchenden Komplex Abstandsrestraints

zwischen den Bindungspartnern erstellt werden. Die PDB-Struktur des zu untersuchenden Komplexes wird dazu als Sequenz- und als Strukturfile verwendet. Da man bewusst nur nach intermolekularen Abständen sucht, muss die Differenz der Nummer der Aminosäuren entsprechend groß gewählt werden, ab der überhaupt nach passenden Atomen gesucht werden soll. Dazu empfiehlt es sich, die Zahl, mit der die Nummerierung der Aminosäuren des zweiten Proteins begonnen werden soll, beim Einlesen der Sequenzen sehr hoch zu setzen. Bei der Auswahl der für die Suche verwendeten Atome müssen für dieses Kriterium alle Atomsorten markiert werden.

Die Bindestellen selbst der beiden Partner müssen auf Komplementarität in Bezug auf Form - beurteilbar durch graphische Darstellung - und Ladung überprüft werden. Außerdem lässt sich feststellen, ob ein Zusammenhalt durch die Ausbildung von Wasserstoffbrücken stabilisiert sein könnte (Meyer et al., 1996; McCoy et al., 1997).

Ein weiterer Hinweis, wo die Interaktionsfläche eines Proteins auf dessen Oberfläche zu finden ist, kann eine Analyse der aus NMR-Experimenten gewonnenen chemischen Verschiebungen zeigen (vergleiche Kapitel 2.2.3). Dazu müssen bei verschiedenen Drücken jeweils die chemischen Verschiebungen der Atome gewonnen und nach einer aminosäureabhängigen Druckkorrektur die verbleibenden Änderungen untersucht werden. Eine Anpassung an die Funktion

$$\delta^{korr} = \delta_0^{korr} + B_1^*(p - p_0) + \frac{1}{2}B_2^*(p - p_0)^2 \quad (\text{Gleichung 2})$$

liefert die Koeffizienten B_1^* und B_2^* . Wie von Beck-Erlach et al. beschrieben, können über das Verhältnis B_2^*/B_1^* Aminosäuren identifiziert werden, für die eine höhere Wahrscheinlichkeit besteht, der Bindestelle anzugehören (Beck-Erlach et al., 2014). Zu in dieser Art auffälligen Residuen gehören diejenigen, deren Verhältnis positiv ist oder mehr als eine Standardabweichung von Null abweicht. Für thermodynamische Zweizustandsmodelle kann gezeigt werden, dass dieses Verhältnis dem Quotienten aus den Differenzen der partiellen molaren Kompressibilitätsfaktors $\Delta\beta_{ij}^{o'}$ und der partiellen molaren Volumina ΔV_{ij}^o der Zustände i und j entspricht:

$$\frac{B_2^*}{B_1^*} = -\frac{\Delta\beta_{ij}^{o'}}{\Delta V_{ij}^o} \quad (\text{Gleichung 3})$$

Informationen aus experimentellen Untersuchungen

Hinweise darauf, ob eine vorhergesagte Struktur mit der Wirklichkeit übereinstimmt oder nicht, können eine Vielzahl an biochemischen Experimenten liefern. Als Beispiel seien hier Bindungsstudien in Kombinationen mit Mutagenese-Experimenten genannt, die über die Beteiligung einzelner Aminosäuren an Komplexbildungen Aufschluss geben können. Auch NMR-Experimente liefern Erkenntnisse über Bindestellen - durch einen Vergleich der chemischen Verschiebungen mit und ohne den bindenden Partner - oder über bewegliche Bereiche, da diese im Vergleich zu stärker fixierten Regionen eine schmalere Linienbreite aufweisen.

Sekundärstrukturvorhersage

Zur Validierung der Ergebnisse aus der Modellierung kann ein Vergleich zwischen den Bereichen, in denen die berechneten Modelle bestimmte Sekundärstrukturelemente aufweisen, und den Bereichen, für die sich anhand der Abfolge der Aminosäuren α -helikale Bereiche oder Faltblattstruktur mit hoher Wahrscheinlichkeit vorhersagen lassen, angestellt werden. Zur Sekundärstrukturvorhersage aus der Primärsequenz eignen sich die Programme PredictProtein (<http://www.predictprotein.org>) oder JPred3 (<http://www.compbio.dundee.ac.uk/www-jpred>).

Graphische Darstellung

Zur graphischen Darstellung der Proteinstrukturen kann das Programm PYMOL verwendet werden. Insbesondere bietet es auch die Möglichkeit, die Bindestelle durch Animation - Drehungen und Zoomen - gut darzustellen. Mit der frei erhältlichen Software VirtualDub lassen sich daraus Filmsequenzen (Format AVI) erstellen, die mit dem Programm Movie Maker (Microsoft, Redmond, USA) weiter aufbereitet werden können, zum Beispiel um Beschriftungen oder Formatänderungen vorzunehmen.

2.2.2 Homologie-Modellierung mit PERMOL

Grundlagen der Methode

Die Methode der Homologiemodellierung beruht auf der Beobachtung, dass Proteine mit ähnlicher Primärsequenz meist gleiche Faltungsmuster besitzen (Chotia und Lesk, 1986). Informationen aus bekannten Raumstrukturen können deswegen auf sequenzähnliche Proteine unbekannter Struktur übertragen werden. Vor allem bei großer Sequenzähnlichkeit lassen sich mit diesem Verfahren gute Modelle erzeugen. Bei weniger starker Ähnlichkeit liefert diese

Methode Konformationen, die als Startstrukturen für ab initio Rechnungen oder als Hilfe beim Assignment für NMR-spektroskopische Aufnahmen eingesetzt werden können, um die Vorhersage der gesuchten Struktur zu beschleunigen.

Die Vorgehensweise bei der Homologiemodellierung erfolgt im Wesentlichen in zwei Schritten: Zuerst muss die Protein-Data-Bank (PDB) (Berman et al., 2000), in der 3D-Strukturdaten von inzwischen über 100000 Proteinen und Nukleinsäuren hinterlegt sind, mit einem der bekannten Algorithmen FASTA, BLAST oder PSI-BLAST nach sequenzhomologen Proteinen bekannter Struktur durchsucht werden. Die passende Auswahl kann über die höchste sequentielle Übereinstimmung, nachgewiesene evolutionäre Verwandtschaft, Qualität der Struktur oder auch über Informationen über die Funktion - die selben Bindungspartner oder die Katalyse der gleichen Reaktion können hier als Kriterium dienen - getroffen werden.

Im zweiten Schritt erfolgt die Gewinnung von Informationen aus den gewählten Modellen und deren Übertragung über ein Alignment auf die unbekannte Struktur zur Modellierung.

Erste Version von PERMOL

Für das 2004 am Institut für Biophysik und physikalischer Biochemie unter der Leitung von Hans Robert Kalbitzer entwickelte Programm PERMOL (Möglich et al., 2005) wurde folgender Ansatz gewählt: Die gesuchten Informationen aus den strukturhomologen Proteinen werden in Form von Restraints extrahiert und auf die zu untersuchende Sequenz übertragen, die dann als Grundlage für eine Modellierung für eine molekulardynamische Simulation - ursprünglich für die Programme DYANA oder CNS gedacht - verwendet werden können. Insbesondere der (damals) neue Einsatz von dihedralen Winkeln, die sehr gut zur Rekonstruktion der Sekundärstruktur geeignet sind und die sich zusätzlich zu Informationen über Abstände und Wasserstoffbrückenbindungen verarbeiten lassen, brachte bereits bei dieser ersten Version zufriedenstellende Ergebnisse. Für das Alignment wurde das Programm CLUSTAL-X verwendet und dessen Ausgabefile eingelesen.

Diese erste Fassung von PERMOL wurde in Perl/Tk erstellt und auf mehreren Betriebssystemen getestet. Es besitzt eine graphische Benutzeroberfläche, über die mehrere Optionen zur Erstellung der Restraints angewählt werden können.

Zweite Version von PERMOL

Die erste in AUREMOL integrierte Version wurde von Josef Scheiber implementiert, indem der vorliegende PERL-Code in ANSI-C umgeschrieben und angepasst wurde. Dabei hatte man die Möglichkeit geschaffen, das Alignment auch in AUREMOL durchführen zu können. Die Implementierung des Alignments erfolgte nach Needleman und Wunsch, wobei jedoch die von diesen vorgeschlagenen Identitätswerte durch die der BLOSUM 62-Matrix ersetzt wurden (Needleman und Wunsch, 1970). Als weitere Verbesserung wurde die Option implementiert, auch Restraints für Seitenkettenatome zu erstellen, außerdem wurde bereits hier eine Gewichtung der Informationen aus verschiedenen Strukturfiles nach lokaler Sequenzähnlichkeit vorgeschlagen, aber nicht in adäquater Weise verwirklicht und getestet (Scheiber, 2003).

Konrad Brunner führte die Arbeit an dieser in AUREMOL integrierten Version weiter (Brunner, 2006). Auf ihn geht auch die Möglichkeit zurück, die dihedralen Winkelrestraints nach dem Schema, wie die Regionen in Ramachandran-Plots unterschieden werden (*favoured*, *additionally allowed*, *generally allowed*, *disallowed regions*), auszuwählen und ausgeben zu lassen. Darüber hinaus entwickelte er zusammen mit Carolina Cano eine auf PERMOL basierende Methode zur Strukturverbesserung von experimentell mithilfe von Kernspinresonanzspektroskopie erstellten Proteinstrukturen (Cano et al., 2009). Dazu werden die Strukturen in das Programm eingelesen, ein Alignment mit der eigenen Proteinsequenz erstellt und Restraints daraus berechnet. Führt man mit diesen neuen Zwangsbedingungen Strukturrechnungen durch, erhält man insbesondere dann ein verbessertes Strukturbündel, wenn durch die experimentelle Bestimmung nicht ausreichend viele oder sich widersprechende Restraints gewonnen worden waren. Auch eine Kombination der experimentellen und der von PERMOL erzeugten Restraints kann in Einzelfällen sinnvoll sein.

Verwendete Restraints

Die Informationen, die helfen, eine unbekannte Struktur zu bestimmen, stecken in den Raumkoordinaten der zur Verfügung stehenden bekannten strukturhomologen Proteine. Es werden drei Sorten von Restraints - Abstandsrestraints, dihedrale Winkelrestraints und Wasserstoffbrückenrestraints - generiert. Diese dienen dann bei der Strukturrechnung als Zwangsbedingungen, die idealerweise erfüllt werden sollen. Aus diesen Bedingungen werden Potentiale erstellt, die so konzipiert sind, dass sie keinen Energiebeitrag liefern, wenn die

Bedingung exakt erfüllt ist. Außerdem werden aus ihnen nach Newton während der Strukturrechnung durch Ableitung Kräfte auf die beteiligten Atome berechnet, die so wirken, dass das System der im Restraint definierten Bedingung näherkommt. Bei der Verwendung von CNS (Brünger et al., 1998) als Molekulardynamikprogramm müssen diese Restraints folgende Form haben (<http://nmr.cit.nih.gov/xplor-nih/doc/current/xplor/>; Schwieters, 2006):

Abstandsrestraints

assign (Atom 1) (Atom 2) real_1 real_2 real_3

Die Atome 1 und 2 entsprechen den beiden Atomen, zwischen denen ein Abstand definiert werden soll. In dieser Arbeit wurde die derzeit aktuelle Version 1.2 von CNS und die von den Entwicklern des Programmes zur Verfügung gestellten Steuerdateien eingesetzt, die im Internet unter structure.usc.edu/cns/ zu finden sind. Wie in den Voreinstellungen der Steuerfiles von CNS festgelegt, wurde für die Strukturrechnungen für die Abstandsrestraints ein biharmonisches Potential verwendet. Die drei Variablen real_1, real_2 und real_3 definieren die Form des Potentials: Das Minimum liegt bei dem Wert real_1. Die Größen real_2 und real_3 bestimmen die Steigung für kleinere beziehungsweise größere Abstände als real_1.

Wasserstoffbrückenrestraints

Wasserstoffbrücken werden in der gleichen Weise erfasst wie die Abstandsrestraints. Analog dem bei Strukturbestimmung mit NMR üblichen Vorgehen werden für eine Wasserstoffbrücke die Abstände zwischen dem Donator und dem Akzeptor und zwischen dem die Bindung vermittelnden Wasserstoffatom und dem Akzeptor angegeben.

Dihedrale Winkelrestraints

assign (Atom 1) (Atom 2) (Atom 3) (Atom 4) real_1 real_2 real_3 integer_1

Die Atome 1 bis 4 entsprechen den Atomen, die den dihedralen Winkel definieren. Das Potential selbst hat die Form $E_{CDIH} = S \sum real_1 * a^{integer_1}$, wobei wie in (<http://nmr.cit.nih.gov/xplor-nih/doc/current/xplor/>) beschrieben bei Strukturrechnungen für real_1 der Wert 1 und für integer_1 der Wert 2 gesetzt wird. Der Skalierungsfaktor S wird

nicht über den Restraint selbst festgelegt (siehe Kapitel 3.2.5). Die Variablen `real_2` und `real_3` legen den Bereich fest, `real_2 ± real_3`, innerhalb dessen sich der Winkel befinden soll. Befindet sich der Winkel in diesem Bereich, wird die Variable `a` Null und der Restraint liefert keinen Energiebeitrag; andernfalls entspricht `a` der Abweichung des Winkels von `real_2`.

2.2.3 Thermodynamik

Chemische Prozesse wie Proteinfaltungen laufen nur dann spontan ab, wenn es sich um exergone Reaktionen handelt und dabei die Änderung der Gibbs-Energie negativ ist. Betrachtet man ein System aus vielen gleich aufgebauten Proteinen, die miteinander über das Lösungsmittel und Stöße in Kontakt stehen, dann stellt sich ein Gleichgewicht zwischen den möglichen Zuständen ein. Die Besetzung der einzelnen Zustände ist von den Differenzen ihrer Gibbs-Energien G_m abhängig und ist im thermischen Gleichgewicht boltzmannverteilt. Durch Änderung der äußeren Bedingungen wie Druck, Temperatur, pH oder Variation des Lösungsmittels können die Gleichgewichte zwischen den Zuständen verschoben werden. Dies ist der Grundgedanke der Beobachtung der Signale unter Variation von Druck.

Bei konstanter Temperatur T_0 gilt für die Besetzungszahl n_m eines Zustands m

$$n_m = \frac{e^{-\frac{G_m}{RT_0}}}{\sum_{m=1}^M e^{-\frac{G_m}{RT_0}}} \quad (\text{Gleichung 4}),$$

wobei R die allgemeine Gaskonstante mit einem Wert von $8,314 \text{ J mol}^{-1} \text{ K}^{-1}$ ist. Daher gilt für das Verhältnis zweier Zustände l und m :

$$\frac{n_l}{n_m} = e^{-\frac{G_l - G_m}{RT_0}} = e^{-\frac{\Delta G_{lm}}{RT_0}} \quad (\text{Gleichung 5})$$

Dieses Verhältnis verändert sich bei Variation des Drucks, woraus folgt, dass sich auch die Gibbs-Energiedifferenz ΔG_{lm} zwischen den Zuständen l und m mit dem Druck ändern muss. Deren Verhalten $\Delta G_{lm}(T_0, p)$ ist aber per se nicht bekannt. Nach Heremans und Smeller kann man eine Taylorentwicklung

$$\Delta G_{lm}(T_0, p) = \Delta G_{lm}^0(T_0, p_0) + \Delta V_{lm}^0(p - p_0) - \frac{\Delta \beta_{lm}^{0'}}{2}(p - p_0)^2 \quad (\text{Gleichung 6})$$

um den Druck p_0 ansetzen (Heremans und Smeller, 1998). Die Größen ΔV_{lm}^0 und $\Delta \beta_{lm}^{0'}$ lassen sich physikalisch als Differenzen des partiellen molaren Volumens und des partiellen molaren Kompressibilitätsfaktors bei der Temperatur T_0 und dem Druck p_0 interpretieren.

Zustände im langsamen Austausch

NMR-Experimente werden seit mehreren Jahren verwendet, um verschiedene Zustände eines Proteins zu untersuchen. Befinden sich 2 Zustände l und m im langsamen Austausch miteinander ($|\Delta\omega_{lm}| \gg 1/\tau_{ex}$, wobei $\Delta\omega_{lm}$ die Differenz der Resonanzfrequenzen des betrachteten Kernspins und τ_{ex} die Austauschrate der Zustände ist), kann aminosäureweise für jeden Zustand ein eigenes Signal beobachtet werden. Deswegen lässt sich über einen Vergleich der Signalvolumina direkt auf die Besetzungszahlen schließen, die zu den Signalvolumina direkt proportional sind:

$$\frac{V_l}{V_m} = \frac{n_l}{n_m} \quad (\text{Gleichung 7})$$

Aus den Volumenverhältnissen der M Zustände kann man mit Gleichung 5 für alle Druckpunkte die Werte von $\Delta G_{lm}(T_0, p)$ berechnen. Daraus lassen sich - zum Beispiel durch Minimierung der Summe der Fehlerquadrate - die Größen $\Delta G_{lm}^0(T_0, p_0)$, ΔV_{lm}^0 und $\Delta\beta_{lm}^{0'}$ einfach bestimmen.

Im Prinzip kann man so aus den Signalen einer einzigen Aminosäure die die Differenzen der Gibbs-Energiedifferenzen beschreibenden Parameter abschätzen, solange die chemischen Verschiebungen ihrer Einzelzustände eine so große Differenz haben, dass die Signale getrennt nachweisbar sind und vor allem als zu einer einzigen Aminosäure gehörig erkannt werden. Geht man wie bei dem Datensatz von A β von einer Anzahl von $P = 10$ verschiedenen Drücken aus, bei denen das Protein beobachtet wird, und für die Anzahl der Zustände $M = 3$ aus, werden über 30 Gleichungen 6 Variable bestimmt.

Besser ist es allerdings, die Messdaten mehrerer Aminosäuren gleichzeitig in der Ausgleichsrechnung zu berücksichtigen, was durch vorherige Mittelwertbildung der $\Delta G_{lm}(T_0, p)$ für alle Druckpunkte unaufwendig geschehen kann.

Bei der Auswertung ist zu beachten, dass es verschiedene Effekte gibt, die dazu führen, dass die Volumina mancher Signale in den Spektren kleiner als erwartet sind, was die Auswertung erschwert. Dazu gehört insbesondere der Signalverlust, der auf den Austausch der Protonen mit dem Lösungsmittel zurückzuführen ist, oder Relaxationsprozesse, die den INEPT-Transfer weniger effektiv machen. Zudem kann sich durch Polymerisation die Konzentration des zu beobachtenden Proteins verändern (Kalbitzer et al., 2013).

Zustände im schnellen Austausch

Befinden sich die Zustände im schnellen Austausch miteinander ($|\Delta\omega_{lm}| \ll 1/\tau_{ex}$), dann misst man statt der chemischen Verschiebungen der einzelnen Zustände ein einziges Signal, dessen chemische Verschiebung dem besetzungszahlgewichteten Mittelwert der Einzelzustände entspricht. Aus diesem Signal lässt sich aber nicht direkt auf die Besetzungszahlen schließen und somit auch nicht auf $\Delta G_{lm}^0(T_0, p_0)$, ΔV_{lm}^0 und $\Delta\beta_{lm}^{0'}$.

Um dennoch Informationen aus diesen Daten zu gewinnen, können folgende Überlegungen angestellt werden. Man betrachte ein Ensemble aus N gleichen Proteinen, die M verschiedene Zustände einnehmen können (Nummerierung der Zustände: $m = 1, \dots, M$). Diese Zustände befinden sich untereinander im schnellen Austausch. Da sich ihnen verschiedene Energieniveaus zuordnen lassen, sind sie nach Boltzmann nicht gleich besetzt: N_m sei die Anzahl der Proteine im m -ten Zustand, die Summe $\sum_m N_m$ muss der Anzahl N der Proteine entsprechen. Die Besetzungszahlen der Zustände seien n_m mit $n_m = \frac{N_m}{N}$ und $\sum_m n_m = 1$.

Nun wird dieses System bei P verschiedenen Drücken beobachtet (Nummerierung der Drücke: $p = 1, \dots, P$). Durch Druckänderung verschieben sich die Energieniveaus, was eine Änderung der Besetzungszahlen bewirkt, also n_{mp} .

Sichtbar werden diese Verschiebungen der Energieniveaus gegeneinander durch die Beobachtung der chemischen Verschiebungen von L Atomen des Proteins (Nummerierung der Atome: $l = 1, \dots, L$). Stammen die zur Verfügung stehenden Informationen aus HSQC-Spektren, können das die chemischen Verschiebungen der Stickstoffatome des Proteinrückgrats oder die der zugehörigen Amidprotonen sein. Die beobachtete chemische Verschiebung des l -ten Atoms δ_{pl} beim Druck p ergibt sich aus der gewichteten Mittelung der chemischen Verschiebungen der einzelnen Zustände, die im Folgenden mit a_{lmp} bezeichnet werden:

$$\delta_{lp} = \sum_m a_{lmp} * n_{mp} \quad (\text{Gleichung 8})$$

Da ein Zustand einem Tal im Phasenraum entspricht und nicht einer bestimmten festen Konfiguration, sind die Werte a_{lmp} selbst auch wieder Mittelwerte aus den chemischen Verschiebungen, die zu einem bestimmten Atom beim Durchlaufen der verschiedenen möglichen dreidimensionalen Konformationen eines bestimmten Zustands gehören. Ebenso ist die einem Zustand zugeordnete Energie ein Mittelwert.

Des Weiteren geht man von der Annahme aus, dass sich die druckinduzierten Änderungen der chemischen Verschiebungen aus zwei Anteilen zusammensetzen, die sich zu dem beobachteten Effekt aufaddieren. Der erste Anteil ist unabhängig von der Struktur des beobachteten Proteins und spiegelt das Verhalten der Aminosäuren wider, verschieden auf Druckänderung zu reagieren, was sich durch ihre unterschiedlichen Größen und ihren Aufbau begründen lässt. Eine Variation des Drucks führt daher zu aminosäurespezifischen Änderungen der chemischen Verschiebung. Diese wurden in (Köhler et al., 2012) bestimmt und eine Methode zu ihrer Korrektur beschrieben:

$$\delta_{lp}^{korr} = \delta_{lp} - B_{l1}(p - p_0) - B_{l2}(p - p_0)^2 \quad (\text{Gleichung 9})$$

Dabei ist der Druck p_0 der Normaldruck. Die Koeffizienten B_{l1} und B_{l2} sind aminosäure- und atomspezifisch und entsprechen den von Köhler et al. publizierten Zahlenwerten (Köhler et al., 2012). Sie wurden aus einer Taylorentwicklung des Verhaltens der chemischen Verschiebungen gewonnen, die an Random-Coil-Peptiden, die keine geordneten Strukturen aufweisen und deswegen auch keinen strukturspezifischen Anteil besitzen sollten, gemessen.

Um diese Korrektur an den experimentellen Daten durchzuführen, wurde in dieser Arbeit ein von Markus Beck-Erlach erarbeitetes und zur Verfügung gestelltes Skript für das Datenanalyseprogramm R (<http://www.r-statistik.de>) verwendet.

Diese aminosäureabhängige Druckabhängigkeit muss ebenso für die chemischen Verschiebungen der einzelnen Zustände gelten, die durch Subtraktion eines aminosäure-, atom- und druckabhängigen Wertes $korr_{lp}$ in die druckunabhängigen chemischen Verschiebungen a_{lm}^{korr} überführt werden können:

$$a_{lm}^{korr} = a_{lmp} - korr_{lp} \quad (\text{Gleichung 10})$$

Daraus folgt:

$$\begin{aligned} \sum_m a_{lmp} * n_{mpd_{lp}} &= \\ &= \sum_m (a_{lm}^{korr} + korr_{lp}) * n_{mp} = \\ &= \sum_m a_{lm}^{korr} * n_{mp} + korr_{lp} * \sum_m n_{mp} = \\ &= \sum_m a_{lm}^{korr} * n_{mp} + korr_{lp} \quad (\text{Gleichung 11}) \end{aligned}$$

Führt man also die Korrektur an den gemittelten chemischen Verschiebungen durch, wird dadurch aus Gleichung 8:

$$\delta_{lp}^{korr} = \delta_{lp} - korr_{lp} = \sum_m a_{lm}^{korr} * n_{mp} \quad (\text{Gleichung 12})$$

Nach dieser Korrektur kann davon ausgegangen werden, dass die verbleibende Änderung der chemischen Verschiebung die Effekte des Drucks auf die einzelnen Zustände widerspiegelt. Da die einzelnen Zustände verschiedenen gemittelten Konformationen entsprechen, die ungleich stark gepackt sein können und deren Oberflächen mehr oder weniger gut für das Lösungsmittel zugänglich sind, ist zu erwarten, dass sie auf Hochdruck unterschiedlich reagieren - energetisch günstiger oder ungünstiger werden. Dadurch verändern sich die Gibbs-Energien der Zustände und damit das Verhältnis ihrer Besetzungen. Dieser verbleibende Anteil kann nun weiter ausgewertet werden.

Betrachtet man die Daten eines einzelnen Atoms l , entspricht Gleichung 12 in Kombination mit den Gleichungen 4 und 6 P Gleichungssystemen mit $M + 3(M - 1)$ zu bestimmenden Unbekannten: Den M korrigierten chemischen Verschiebungen der einzelnen Zustände und den Differenzen $\Delta G_{1m}^0(T_0, p_0)$, ΔV_{1m}^0 und $\Delta \beta_{1m}^{0'}$ mit $m = 2, \dots, M$. Nimmt man beispielsweise wie bei dem Datensatz von Aß für die Anzahl der Druckpunkte $P = 10$ und für die Anzahl der Zustände $M = 3$ an, dann müssten über 10 Gleichungssysteme 9 Variablen angepasst werden. Hier zeigt sich die ungleich schwierigere Bestimmung der Parameter im Vergleich zu dem Fall des langsamen Austauschs und die Notwendigkeit, die gesuchten Größen aus den Daten mehrerer Aminosäuren zu ermitteln. Bei gleichzeitiger Betrachtung von L Atomen erhöht sich die Anzahl der Gleichungen auf $L * P$, und die Anzahl der zu bestimmenden Variablen berechnet sich zu $L * M + 3(M - 1)$, da die korrigierten chemischen Verschiebungen für jedes Atom unterschiedlich sind. Nimmt man wie eben $P = 10$ und $M = 3$ an und kombiniert dies mit $L = 31$, ergibt dies 310 Gleichungen zur Bestimmung von 99 unbekannten Variablen. Zur Lösung dieses Problems müssen Verfahren zur Bestimmung nichtlinearer Ausgleichsprobleme angewandt werden.

Erschwert wird die Lösung dadurch, dass die Anzahl M der anzunehmenden Zustände nicht - wie bei der Auswertung der Volumina - bekannt ist.

2.2.4 Fit-Verfahren

Multimodales Vorgehen

Von Kremer et al., Kalbitzer et al. und Munte et al. wurden Ergebnisse von Hochdruck-NMR-Spektren für die Prionproteine des Menschen und des Syrischen Hamsters, für das Protein Ras und das Alzheimer- β -Amyloid vorgestellt (Kremer et al., 2007; Kalbitzer et al., 2013; Munte et al. 2013). Hier wurde eine interaktive, vielschichtige Auswertung der Daten beschrieben, bei der insbesondere berücksichtigt wurde, dass für die untersuchten Proteine Zustände im langsamen und schnellen Austausch koexistieren und auch Übergangsraten zu finden sind, die in der Grenzregion zwischen den beiden Fällen liegen und daher einer Auswertung schwer zugänglich sind. Deswegen werden neben den chemischen Verschiebungen auch die Volumina der Signale zum Informationsgewinn herangezogen.

Zur Bestimmung der gesuchten thermodynamischen Größen aus den chemischen Verschiebungen wurde - nach Durchführung der beschriebenen aminosäureabhängigen Druckkorrektur - speziell nach Aminosäuren gesucht, die für einen bestimmten Übergang zu einem Zustand hin charakteristisch sind, indem deren chemische Verschiebungen der weiteren Zustände nahezu identisch sind. Diese lassen sich anhand ihrer sigmoidalen Form identifizieren. Durch Abschätzen der Größen ΔG_{1i}^0 , ΔV_{1i}^0 und $\Delta \beta_{1i}^{0'}$ für jede Aminosäure wurde eine Verteilung erstellt, durch deren Optimierung sich iterativ an die optimalen Werte herangetastet wurde.

Lösungsverfahren für nichtlineare Ausgleichsprobleme

Während die Parameter a_{lm}^{korr} , die die korrigierten chemischen Verschiebungen der einzelnen Zustände repräsentieren und deswegen für jedes betrachtete Atom getrennt optimiert werden können, sind die Größen ΔG_{1i}^0 , ΔV_{1i}^0 und $\Delta \beta_{1i}^{0'}$ für alle Atome gleich. Daher liegt die Idee nahe, sie durch gleichzeitige Optimierung aller Funktionen δ_{lp}^{korr} zu bestimmen.

Eine Möglichkeit besteht darin, mit der Methode der kleinsten Quadrate eine Funktion aufzustellen, deren Minimum die optimalen Parameter repräsentiert. Da ΔG_{1i}^0 , ΔV_{1i}^0 und $\Delta \beta_{1i}^{0'}$ über e-Funktionen mit den anzupassenden druckkorrigierten chemischen Verschiebungen verknüpft sind, liegt demnach bei diesem Vorgehen ein nichtlineares Optimierungsproblem vor.

Die Lösung nichtlinearer Ausgleichsprobleme kann über das Verfahren nach Gauß-Newton erfolgen, das iterativ in jedem Schritt die zu minimierende Funktion durch eine quadratische Näherung ersetzt, die explizit berechnet werden kann. Dadurch wird das zu lösende nichtlineare Gleichungssystem in eine Folge linearer Ausgleichsprobleme überführt.

Eine Weiterentwicklung dieses Verfahrens ist der Levenberg-Marquardt-Algorithmus, der das Gauß-Newton-Verfahren mit einer Regularisierungstechnik verknüpft, die absteigende Funktionswerte erzwingt. Dieses Verfahren zeichnet sich durch bessere Konvergenz aus. Sowohl beim Gauß-Newton- als auch beim Levenberg-Marquardt-Verfahren sind die gewählten Anfangsbedingungen entscheidend für eine erfolgreiche Anwendung. Sinnvoll ist es, von mehreren Kombinationen von Parametern aus den Algorithmus zu starten. Um geeignete Kombinationen zu finden, müssen die Besonderheiten des jeweiligen Parameterraums berücksichtigt werden (Dahmen und Reusken, 2008).

Kombinationen chemischer Verschiebungen verschiedener Elemente

Um die benötigten chemischen Verschiebungen zu gewinnen, eignen sich ^1H - ^{15}N -HSQC-Spektren, da man damit die druckinduzierten Änderungen gut analysieren kann. Dazu werden die verwendeten Datensätze zuerst mit TopSpin (<http://www.bruker.com>) prozessiert und mithilfe des Hauptresonanzpeaks der als interner Standard verwendeten Substanz DSS geeicht. Mit dem Programm AUREMOL (Gronwald und Kalbitzer, 2004) können mithilfe von vorliegenden Listen der chemischen Verschiebungen die Peaks des bei Normaldruck aufgenommenen Spektrums den zugehörigen Atomen des Proteins zugeordnet werden. Diese Zuordnung wird dann sukzessive auf die bei höheren Drücken aufgenommenen Spektren übertragen. Durch Bestimmung der Maxima der Peaks können mit AUREMOL und TopSpin die Werte der chemischen Verschiebungen erfasst werden.

Die Wahrscheinlichkeit, die gesuchten Parameter richtig zu bestimmen, kann dadurch erhöht werden, dass man die chemischen Verschiebungen verschiedener Elemente - namentlich H_N , N , C_α und C' - miteinbezieht, da diese eine unterschiedliche Sensitivität auf bestimmte Arten von Strukturänderungen zeigen. So zeichnen sich die C' -Atome des Backbones dadurch aus, dass sie besonders empfindlich Variationen in der Sekundärstruktur detektieren können.

Aufgrund der unterschiedlichen Frequenzbereiche der möglichen Elemente muss allerdings eine Gewichtung stattfinden. Diese kann über Gewichtungsfaktoren geschehen, die aus den

Standardabweichungen der Atome, die in der BMRB-Datendank zu finden sind, berechnet werden. Wie von Schumann et al. beschrieben, bewährt sich diese Vorgehensweise bei der Identifizierung von Protein-Protein-Interaktionen (Schumann et al., 2007) und soll auch hier zur Anwendung kommen.

Kombinierte Betrachtung der chemischen Verschiebungen und Volumina der Signale

Bei der Auswertung experimenteller Datensätze und der Frage nach der Anzahl der Zustände muss immer berücksichtigt werden, dass die Zustandsänderungen häufig nicht auf einer Zeitskala ablaufen, die die Annahme eines schnellen Austauschs rechtfertigen. Bei der Erfassung der chemischen Verschiebungen muss deswegen immer auch eine Volumenbetrachtung der Signale miteinbezogen werden, die darüber Auskunft gibt, ob in dem beobachteten Peak die Information aller Zustände steckt. Eine Möglichkeit, die gemittelte chemische Verschiebung eines Peaks zu eruieren, wenn dieser in Unterpeaks aufgespalten ist, die sich im Spektrum sehr nah sind und von denen angenommen wird, dass sie zu einer einzigen Aminosäure gehören, besteht darin, das in AUREMOL implementierte Verfahren zum Peakpicking zu verwenden mit der Option, die Koordinaten des Schwerpunkts des Signalsystems zu benutzen.

2.3 Neuimplementierung von PERMOL

2.3.1 Erweiterung der Methode

Ziel dieser Arbeit war es, die in Kapitel 2.2.2 beschriebenen Ideen aufzugreifen und das Programm neu zu implementieren. Dabei wurde PERMOL um die Möglichkeit erweitert, auch Restraints für Homologiemodelle zu generieren, die aus zwei oder mehr Sequenzstücken bestehen, so dass insbesondere Komplexe vorhergesagt werden können, bei denen im Vergleich zur Struktur, aus der die Informationen gewonnen werden, ein Bindungspartner durch ein sequenzähnliches anderes Protein ersetzt wird. Zwei große Vorteile dieser Methode gegenüber Dockingprogrammen, die aus Einzelstrukturen den Komplex vorhersagen, liegen auf der Hand: Zum einen ist die Bindestelle bereits vorgegeben und die Bindungspartner müssen nicht mehr gegeneinander rotiert werden. Zum anderen sind die Bindungspartner des Modells bereits aneinander angepasst, so dass besonders bei der Bindung nötige Winkeländerungen des Proteinrückgrats gegenüber den ungebundenen Partnern gleich als Informationen in die Konfiguration des neu zu modellierenden Komplexes miteinfließen können.

Dabei konnte das Grundkonzept von PERMOL - die Erstellung von künstlichen, also nicht aus Experimenten gewonnenen Restraints, auf deren Grundlage dann eine Strukturrechnung durchgeführt werden kann - direkt übernommen werden. Für die beschriebene Erweiterung war deswegen vor allem eine Umstrukturierung des Programms nötig, um intern mehr als eine Proteinsequenz verarbeiten zu können. Hierbei ist insbesondere das Alignment betroffen, da hier beim Sequenzvergleich mehrere Sequenzstücke verglichen werden müssen.

Zusätzlich musste das Programm CNS - einzelne Dateien des Quellcodes und der Steuerfiles - so abgeändert werden, dass die molekulardynamischen Simulationen der Komplexstrukturen durchgeführt werden können.

2.3.2 Programmiersprache und Programmaufbau

Da eine sukzessive Umstellung des Programmes AUREMOL, in das das Modul PERMOL integriert ist, von ANSI-C auf die neuere Programmiersprache C++ in Kombination mit den Klassenbibliotheken von Qt vorangetrieben wird, wurde auch zur Neuimplementierung und Erweiterung von PERMOL diese Kombination verwendet.

Vereinfacht ist das Programm nach folgendem Schema aufgebaut (Abbildung 2-1):

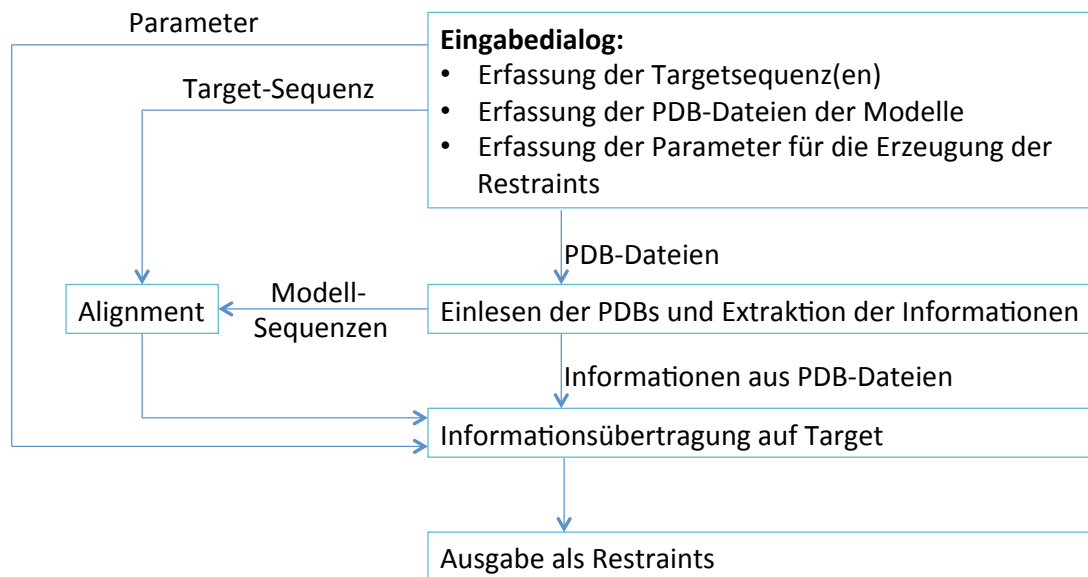


Abbildung 2-1: Schematischer Aufbau des Programms PERMOL

Die Informationen, die den Modellen entnommen werden, werden über das Alignment auf die Sequenz der unbekannten Proteinstruktur übertragen und als Restraints ausgegeben.

2.3.3 Aufbereitung der Daten

Die Informationen, die aus den sequenzhomologen Proteinen gewonnen werden, stammen aus den in PERMOL eingelesenen PDB-Dateien. Diese können jedoch nur bei kompatibler Aufbereitung optimal verarbeitet werden. Deswegen müssen vor Verwendung folgende zwei Schritte durchgeführt werden:

Hinzufügen von Wasserstoffatomen

Wie von Cano et al. gezeigt, erzielt man mit der PERMOL zugrundeliegenden Methode die besten Ergebnisse, wenn man zur Strukturrechnung Abstände zwischen Wasserstoffatomen verwendet (Cano et al., 2009). Deren Koordinaten lassen sich durch die Orte der schwereren Atome der Umgebung berechnen. Dazu kann das entsprechende Modul in AUREMOL eingesetzt werden.

IUPAC-Nomenklatur

Entscheidend für gute Ergebnisse ist auch, dass die Namen der Atome der IUPAC-Nomenklatur (Markley et al., 1998) entsprechen, da Auremol und damit auch PERMOL mit dieser Konvention arbeitet und nur so eine Rückkonvertierung in die in CNS gebräuchliche Nomenklatur bei der Ausgabe der Restraints fehlerfrei funktionieren kann. Eine Konvertierung aus den gängigsten Namenskonventionen kann ebenfalls mit AUREMOL durchgeführt werden. Um stereospezifische Vertauschungen auszuschließen, muss anschließend noch eine Überprüfung mit dem von Philipp Schneider entwickelten Programm Iupacify gemacht werden. Mit diesem Programm werden die Koordinaten aller über Winkel definierten Atomnamen auf Konsistenz mit den IUPAC-Konventionen hin abgeglichen und notfalls eine Vertauschung der Atomnamen vorgenommen (Schneider, 2014).

2.3.4 Erfassung der Eingabeparameter

Der Dialog zur Erfassung der Eingabeparameter beinhaltet die Möglichkeit, eine oder mehrere Zielsequenzen, die Pfade für die als Modelle zu verwendenden PDB-Dateien und den Pfad für die Ausgabedateien aufzunehmen. Die Zielsequenzen können entweder direkt im 1-Letter-Code oder alternativ als AUREMOL-Sequenzfile eingelesen oder aber auch aus einer PDB-Datei extrahiert werden. Für jede Sequenz ist die Nummer der ersten Aminosäure festzulegen, die bei der Ausgabe der Restraints verwendet werden soll. Der Dialog sieht folgendermaßen aus (Abbildung 2-2):

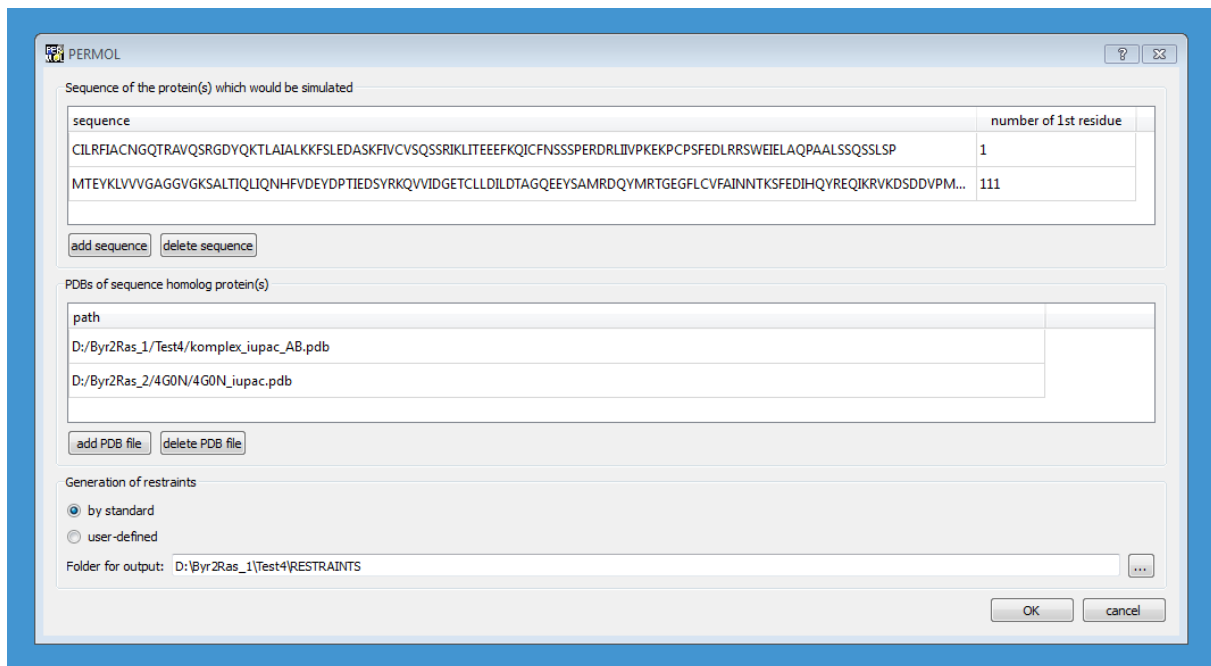


Abbildung 2-2: Startdialog des Programms PERMOL

Der Startdialog nimmt die Sequenzen der gesuchten Struktur und die Pfade der PDB-Dateien der Modellierungsgrundlagen und der Ausgabedateien auf. Außerdem kann entschieden werden, ob die Restraints nach benutzerdefinierten Parametern oder *by standard* generiert werden sollen.

2.3.5 Alignment

Ziel eines Alignments ist es zu entscheiden, welche Aminosäure des Modells am besten durch welche Aminosäure der gesuchten Struktur ersetzt werden muss, damit das Faltungsmuster rekonstruiert werden kann. Dazu wird ein Sequenzvergleich im 1-Letter-Code durchgeführt, wobei zusätzlich Leerstellen - auch Löcher oder Gaps genannt - sowohl in der Zielsequenz als auch in der Sequenz des vorhandenen Modells eingefügt werden dürfen. Die verschiedenen Möglichkeiten der Zuordnung werden durch Scores beurteilt, wobei Löcher und Nichtübereinstimmungen bestraft und Übereinstimmungen belohnt werden. Ein optimales globales Alignment wird mit dem Needleman-Wunsch-Algorithmus erreicht. Beim Scoring kann dabei die bewährte BLOSUM 62-Matrix verwendet werden (Henikoff und Henikoff, 1992), die bei Nichtübereinstimmung zweier Aminosäuren deren Ähnlichkeit bezüglich ihrer physikalischen und chemischen Eigenschaften berücksichtigt, was für den Strukturerhalt von Relevanz ist, und die bei Übereinstimmung evolutionär seltene Konservierungen stärker belohnt.

Stehen dem Benutzer jedoch weitere Informationen wie zum Beispiel das Wissen um die Sekundärstrukturelemente der gesuchten Struktur zur Verfügung oder wurde bereits ein anderes Programm wie ClustalW2 (<http://www.ebi.ac.uk/tools/msa/clustalw2>), EMBOSS Stretcher (http://www.ebi.ac.uk/tools/psa/emboss_stretcher) oder LALIGN (<http://www.ebi.ac.uk/tools/psa/lalign>) für das Alignment verwendet, kann das von PERMOL generierte Alignment als Grundlage dienen und dann entsprechend editiert werden.

Es wurde der Algorithmus nach Needleman-Wunsch mit folgenden Besonderheiten implementiert:

- Die Werte für Gap Penalty, Mismatch Penalty und Match Award des Algorithmus, die in der alten Version eingeführt worden waren, um eine Möglichkeit zu haben, das Alignment zu beeinflussen, sich aber im Laufe dieser Arbeit als nicht zielführend herausgestellt haben, wurden jeweils auf Null gesetzt, als Bewertungsfunktion wird einzig die BLOSUM 62-Matrix verwendet.
- Lücken in der Sequenz von Target oder Modell werden mit "-" gekennzeichnet.
- Unbekannte Aminosäuren werden mit "X" gekennzeichnet und dadurch beim Berechnen des Scores weniger bestraft als Lücken.
- Nicht zusammenhängende Sequenzstücke werden mit einem "+" abgetrennt. Dies ermöglicht die Verwendung von mehr als einer Sequenz und erlaubt damit die Bearbeitung von Proteinkomplexen.
- Besteht die gesuchte Struktur aus mehreren Sequenzstücken, müssen diese in der Reihenfolge, wie sie mit den Sequenzstücken des Modells aligniert werden sollen, eingelesen werden.
- Sequenzstücke der gesuchten Struktur werden jeweils nur mit einem anderen Sequenzstück des Modells aligniert.
- Lücken am Anfang oder Ende einer Sequenz werden nicht wie in der BLOSUM 62-Matrix vorgesehen mit -6 bestraft, sondern gehen nicht in den Score ein. Dadurch wird erreicht, dass einzelne Aminosäuren am Anfang oder Ende eines Sequenzstücks nicht durch Lücken abgetrennt werden, was für die Informationsübertragung nicht sinnvoll wäre.

Bei der Neuimplementierung des Alignments konnte bereits auf eine von Michael Ebel in C++ erstellte Version zurückgegriffen und weiter ausgebaut werden.

Bereits die erste in Perl implementierte Fassung von PERMOL besaß eine graphische Benutzeroberfläche, bei der das Alignment dargestellt wurde, wobei die Farben den Grad der Sequenzkonservierung symbolisierten. Aminosäuren, für die Restraints erstellt werden sollten, konnten durch Anklicken oder anhand dieses Farbcodes ausgewählt werden. Diese Möglichkeit ging bei der zweiten Implementierung verloren, da dies in ANSI-C nicht in adäquater Weise zu verwirklichen war. Bei der Erstellung der neuen Version wurde die Idee der Darstellung nun wieder aufgegriffen und ausgebaut: Durch Anklicken können die Aminosäuren oder Bereiche markiert werden, für die Restraints berechnet oder nicht berechnet werden sollen. Außerdem kann das Alignment durch Einfügen oder Löschen von Lücken - dargestellt durch "-" - direkt editiert werden. Die so neu definierte Zuordnung wird dann auf Konsistenz mit Target- und Modellsequenz überprüft, der zugehörige Score berechnet und als Alignment für die Informationsübertragung verwendet oder - bei Bedarf - wieder auf das von PERMOL berechnete Alignment zurückgesetzt (Abbildung 2-3).

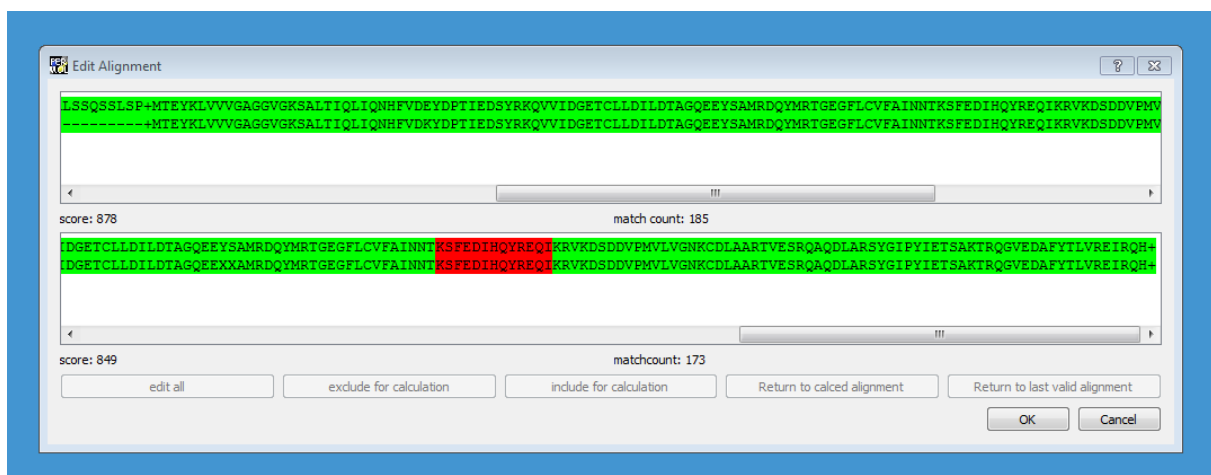


Abbildung 2-3: Graphische Darstellung des Alignments

Die zu untersuchenden Sequenzen werden mit jeder eingegeben PDB-Datei getrennt align und dargestellt, wobei der Dialog dynamisch der Anzahl der PDB-Dateien angepasst wird. Die Alignments sind direkt editierbar und die Lücken ("-") dürfen beliebig verschoben werden. Dabei ist darauf zu achten, dass die "+"-Zeichen, die die Sequenzstücke abteilen, in der Darstellung am Ende des Editiervorgangs wieder übereinander zu liegen müssen. Für die so neu entstandene Zuordnung wird automatisch der neue Score berechnet. Außerdem kann über diesen Dialog eine Auswahl getroffen werden, für welche Aminosäuren Restraints berechnet werden sollen (grün markiert). Mit der Maus können einzelne Aminosäuren oder Bereiche markiert werden, um sie auszuschließen (rot markiert) oder wieder hinzuzunehmen.

2.3.6 Informationsübertragung

Für die Informationsübertragung wurden folgende Überlegungen angestellt und genauso implementiert:

Über das Alignment wird jeder Aminosäure der gesuchten Struktur eine oder keine Aminosäure des Modells zugeordnet. Sind sich die beiden miteinander alignnten Aminosäuren nicht sehr ähnlich, muss überlegt werden, inwieweit Informationen übertragen werden können. Informationen aus dem Proteinerückgrat, also Abstände zwischen Atomen der Hauptkette, die charakteristischen die Sekundärstruktur stabilisierenden Wasserstoffbrücken und auch die dihedralen Winkel Φ , Ψ und Ω können direkt übernommen werden, solange weder die Aminosäure der Zielstruktur noch die des Modells eine der Aminosäuren Prolin oder Glycin ist, deren Aufbau und Benennung sich von den anderen Aminosäuren unterscheiden.

Für die Seitenketten können folgende Regeln angewandt werden:

- Ein Abstandsrestraint ist durch zwei Atome und deren Abstand definiert. Der aus dem Modell berechnete Restraint wird dann übernommen, wenn die beiden Atomnamen in den entsprechenden Aminosäuren der Zielstruktur wieder vorkommen. Dieser Ansatz beruht auf der Tatsache, dass in der IUPAC-Nomenklatur nur Atome gleich benannt sind, die an gleicher Stelle beim Aufbau stehen und sich insbesondere auch stereochemisch entsprechen.
- Wasserstoffbrücken werden ebenfalls als Abstandsrestraints codiert. Für die beiden beteiligten Atome muss überprüft werden, ob die entsprechenden Atome der ersetzten Aminosäuren ebenfalls als Donator beziehungsweise Akzeptor geeignet sind, um den Restraint übernehmen zu können.
- Die Wasserstoffatome der Methylgruppen werden nicht zum Informationsgewinn genutzt, da bei ihnen aufgrund ihrer geringen Größe und ihrer Lage am Ende der Seitenkettenäste eine Drehung der gesamten Gruppe viel weniger behindert wird als bei schwereren Atomen. Auch stellen sie keine Donatoren oder Akzeptoren dar, die stabile Wasserstoffbrücken ausbilden könnten und damit innerhalb der Proteinstruktur fixiert wären.
- Die X-Winkel der Seitenketten beschreiben eine von mehreren Konformationsmöglichkeiten - Rotamere genannt (Dunbrack, 2002) -, die auf die Rotationsmöglichkeiten um die Atombindungen zurückzuführen sind und die aber gleichzeitig durch die Größe der Atome, die eine freie Drehung verhindern,

eingeschränkt sind. Eine Übertragung dieser Winkelinformationen ist dann sinnvoll, wenn sich die Seitenketten der beiden über das Alignment einander zugeordneten Aminosäuren ähneln. Das ist dann der Fall, wenn nur ein Unterschied der Länge vorliegt, also nur eine CH₂-Gruppe eingefügt oder weggelassen wurde, oder das Ende der Seitenkette nur leicht modifiziert wurde wie bei der Ersetzung von Glutamin durch Glutamat. Dann lässt sich zu dem vorhandenen Rotamer des Modells ein entsprechendes Rotamer der unbekannten Struktur finden, das durch die X-Winkel teilweise oder komplett beschrieben werden kann.

2.3.7 Berechnung der Restraints und Gewichtung verschiedener PDBs

Wie von Cano et al. beschrieben, arbeitet PERMOL am besten, wenn folgende Auswahl zur Erstellung der Restraints getroffen wird (Cano et al., 2009):

- Abstandsrestraints zwischen allen H α -Atomen des Modells, die in der Sequenz einen Mindestabstand von zwei Aminosäuren besitzen und deren Abstand im Raum zwischen 0,18 nm und 1,5 nm liegt
- Abstandsrestraints zwischen allen HN-Atomen und Seitenkettenwasserstoffatomen des Modells, die nicht zur selben Aminosäure gehören und deren Abstand im Raum zwischen 0,18 nm und 0,6 nm liegt
- Alle Φ , Ψ - und X-Winkel
- Alle Wasserstoffbrücken

Bei der Neuimplementierung von PERMOL wurden zwei Möglichkeiten realisiert, eine benutzerdefinierte Auswahl (Abbildung 2-4) oder *by standard*. Wird diese zweite Möglichkeit gewählt, werden die Restraints, wie eben beschrieben, erstellt mit der Ausnahme, dass nur dihedrale Winkel ausgegeben werden, die in *most favoured* Bereichen im Ramachandran-Plot liegen.

Wenn diese automatische Auswahl nicht genutzt wird, müssen die Parameter zur Erzeugung über einen weiteren Dialog ausgewählt werden. Diese Option ist dann hilfreich, wenn keine Wasserstoffatome verwendet werden sollen oder das Programm eingesetzt wird, Abstände in Proteinstrukturen in einem festgelegten Bereich zu bestimmen, um diese Informationen beispielsweise bei einer Fehlersuche bei einem Satz experimenteller Restraints - man denke an eine falsche Zuordnung bei NMR-Untersuchungen - zu nutzen, die Bindestelle bei

Komplexen, wie in Kapitel 2.2.1, beschrieben zu untersuchen oder Wasserstoffbrücken in den errechneten Strukturen zu finden.

Die generierten Restraints werden als Textdateien ausgegeben und können direkt in CNS eingelesen werden.

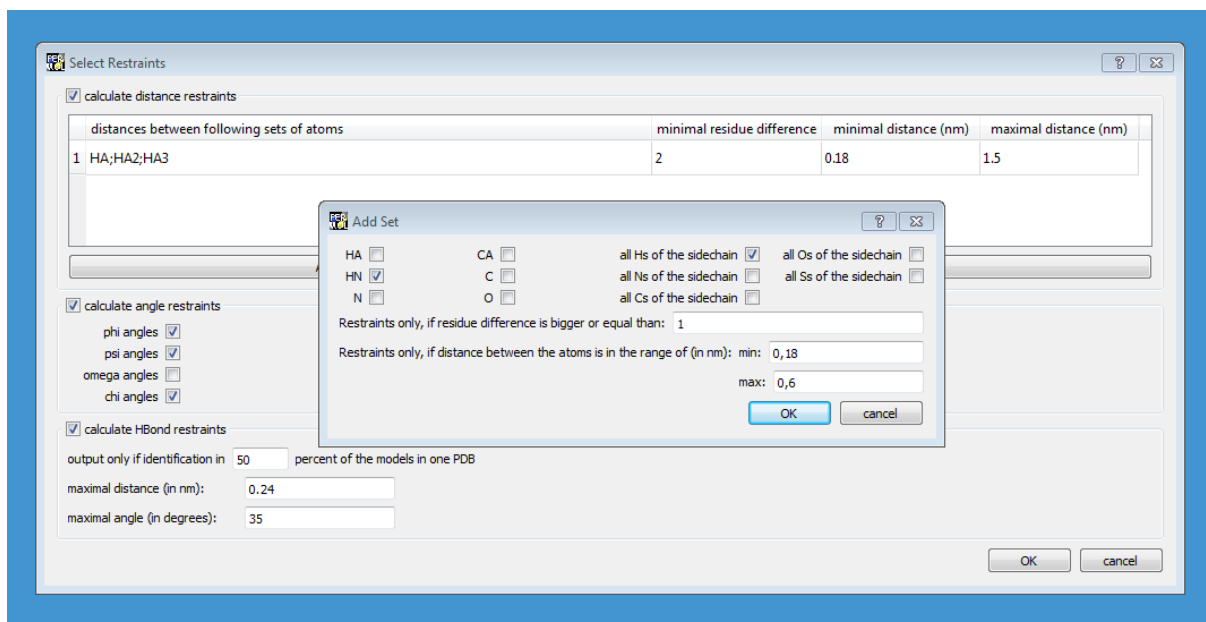


Abbildung 2-4: Dialog des Programms PERMOL zur Auswahl der Parameter bei der Erzeugung der Restraints

Zusätzliche Abstandsrestraints können über den Dialog "Add Set" angefordert werden, wobei außer der Atomsorte auch der Mindestabstand in der Sequenz zwischen den Aminosäuren, denen diese Atome angehören, und ein maximaler und minimaler Abstand, zwischen dem sich der auszugebende Restraint bewegen darf, angegeben werden müssen. Für die dihedralen Winkelrestraints muss die Art des Winkels und der Bereich im Ramachandran-Plot angegeben werden, in dem sie liegen sollen. Die Suche nach Wasserstoffbrücken erfolgt über deren typische Geometrie, die aber auch nach Wunsch abgeändert werden kann.

Bei der konkreten Berechnung der Restraints muss unterschieden werden, ob eine oder mehrere PDB-Dateien als Modellierungsgrundlage zur Verfügung stehen. Wird nur eine PDB-Datei verwendet und handelt es sich dabei um eine mithilfe von NMR-spektroskopischen Experimenten erhaltene Struktur, besitzt diese im Allgemeinen mehrere Modelle. Für jedes dieser Modelle kann nun der Abstand zwischen zwei Atomen X und Y berechnet werden. Wie von Brunner beschrieben, liefern Bildung des Mittelwertes und Berechnung der Standardabweichung über alle Modelle die gewünschten Informationen für den Restraint (Brunner, 2006): Es wird ein Konfidenzintervall berechnet unter der Voraussetzung, dass die Strukturen als statistische Stichproben angesehen werden müssen

und die generierten Abstände normalverteilt sind. Bei der Implementierung wurde ein Konfidenzniveau von 95% angenommen, so dass als obere und untere Grenze jeweils die doppelte Standardabweichung verwendet wird. Für die dihedralen Winkelrestraints kann dasselbe Vorgehen gewählt werden.

Wasserstoffbrücken werden ebenfalls für jedes Modell unabhängig voneinander bestimmt, wobei zur Identifizierung die typische Geometrie dieser Wechselwirkungen, ein maximaler Abstand von 0,24 nm zwischen Donator und Akzeptor und einem Wasserstoffbrückenwinkel von $180^\circ \pm 35^\circ$, genutzt wird. Ob eine Wasserstoffbrücke als Restraint ausgegeben wird, hängt von einer benutzerdefinierten in Prozent angegebenen Schwelle ab, in wie vielen Modellen diese mindestens gefunden werden muss. Mittelwert und Fehlergrenzen werden dabei analog zu den Abstandsrestraints berechnet. Als mögliche Donatoren können alle stark elektronegativen Atome - Stickstoffe oder Sauerstoffe -, an die ein Wasserstoffatom kovalent gebunden ist, dienen, als Akzeptoren kommen Atome mit einem freien Elektronenpaar in Frage. Die Restraints werden so erstellt, dass sich aus jeder Wasserstoffbrücke zwei Zwangsbedingungen - der Abstand zwischen Donator und Akzeptor und der Abstand zwischen dem die Bindung vermittelnden Proton und dem Akzeptor - ergeben.

Ist die strukturhomologe Konformation dagegen mit Röntgenspektroskopie bestimmt worden, liegt in der PDB im Allgemeinen nur ein Modell vor. Wie von Brunner bereits für Vorgängerversion von PERMOL realisiert (Brunner, 2006), wird dann die Berechnung der Standardabweichung ersetzt durch eine Abschätzung der Abweichung, die sich aus der Auflösung R ergibt und nach Holton und Alber einem Drittel von R entspricht (Holton und Alber, 2004), und der Abstands- beziehungsweise Winkeländerung aufgrund der thermischen Bewegung. In diese geht der Temperaturfaktor B_i (zu finden in den Spalten 61-66 des entsprechenden Atoms in der PDB-Datei) des Atoms ein - bei Wasserstoffatomen wird ersatzweise der des kovalent gebundenen schweren Atoms verwendet -, der aus der PDB-Datei mitausgelesen werden kann. Der daraus resultierende Anteil für jedes Atom i ist nach Cruickshank gegeben durch $\sqrt{\frac{B_i}{8 * \pi^2}}$ (Cruickshank, 1999). Daraus folgt für das Quadrat der Standardabweichung $\sigma^2(d_{ij})$ des Abstands d_{ij} zweier Atome i und j in Näherung (Brunner, 2006):

$$\sigma^2(d_{ij}) = \frac{B_i + B_j}{8 * \pi^2} + \frac{R}{9}$$

Wird zur Erstellung der Restraints mehr als eine PDB-Datei verwendet, erstellt man für jede von ihnen ein separates Alignment mit der Zielsequenz und berechnet die gesuchten Informationen. Dadurch werden Restraints, die die gleichen Atome miteinander verknüpfen, je nach Anzahl der verwendeten PDB-Dateien mehrfach generiert. Diese werden nun vor der Ausgabe so miteinander kombiniert, dass die Information einer PDB-Datei, deren Sequenz in unmittelbarer Umgebung der den betrachteten Restraint definierenden Atome der Targetsequenz am ähnlichsten ist, ein größeres Gewicht hat als die der anderen verwendeten PDB-Dateien. Für jede Position k innerhalb eines Alignments einer betrachteten PDB-Datei l wird deshalb ein lokaler Score L_k^l definiert, der sich aus dem Eintrag der Blosumtabelle s_k des betrachteten Atoms, seines Vorgängers und seines Nachfolgers zusammensetzt. Da dieser lokale Score positiv sein soll - er soll als Gewichtung dienen -, muss eine konstante Zahl dazuaddiert werden. Die kleinste vorkommende Zahl der BLOSUM 62-Matrix ist -6, so dass 18 als konstanter Summand gewählt werden kann:

$$L_k^l = \sum_{j=-1}^1 s_{k+j} + 18$$

Für einen speziellen Restraint, in den die Koordinaten von N Atomen einfließen, wird für jede PDB ein lokaler mittlerer Score über die Aminosäuren, denen diese N Atome angehören, eruiert. Der Mittelwert und die Standardabweichung zur Ausgabe des gesuchten Restraints werden dann durch diese nach den lokalen Scores gewichtete Mittelwertbildung berechnet.

Sollen Restraints für mehrere Proteine wie bei der Simulation eines Komplexes erstellt werden, gibt es prinzipiell zwei Möglichkeiten, mit denen CNS später arbeiten kann. Die eine Möglichkeit besteht darin, die Definition eines Atoms, das üblicherweise durch die Aminosäurenummer und seinen Namen festgelegt ist, um einen sequenzabschnittsspezifischen Namen - in den PDB-Dateien als *chain identifier* bezeichnet - zu erweitern. Die zweite - für PERMOL gewählte - besteht darin, darauf zu achten, dass die Aminosäurenummern eines Proteins nicht mit denen der anderen Proteine überlappen.

2.3.8 Modifikationen an dem Programm CNS

Verarbeitung der von PERMOL erzeugten Restraints

Die dafür vorgesehenen Protokolle von CNS und der Quellcode müssen beim Einsatz von den hier generierten Restraints etwas modifiziert werden, da sie für experimentelle Restraints, gewonnen durch NMR-Experimente, optimiert worden sind. Insbesondere muss berücksichtigt werden, dass bei dieser Art der Restrainterstellung deutlich mehr Bedingungen

pro Aminosäure entstehen, die zu verarbeiten sind - im Mittel etwa 30 Restraints pro Aminosäure im Vergleich zu etwa 16 bei Strukturbestimmung mit NMR. Die erzeugten Abstandsrestraints ersetzen dabei die durch NOESY-Spektren erhaltenen Abstände, die über den Nuklear Overhauser Effekt bestimmt werden. Die erzeugten dihedralen Winkel werden statt der Winkel verwendet, die aus den chemischen Verschiebungen vorhergesagt werden können, was zum Beispiel mit dem Programm TALOS+ (Shen et al., 2009) geschehen kann. Die generierten Wasserstoffbrückenrestraints ersetzen die experimentell gemessenen. Folgende Anpassungen müssen vorgenommen werden:

- Es muss genügend Speicherplatz für die Restraints bereitgestellt werden, da sonst nicht alle Bedingungen eingelesen werden können. Dazu müssen die entsprechenden Dateien des Quellcodes (NMR-Modul *readdata*) abgeändert werden. Für das Waterrefinement ist das Skript entsprechend anzupassen (Parameter *nrestraints*).
- Um die Gewichtung zwischen den Energiebeiträgen aus den physikalischen Potentialen und den durch die Restraints erzeugten ähnlich der Gewichtung zu gestalten, wie sie beim Einsatz experimenteller Daten vorliegt und für welche die die Strukturberechnung steuernden Skripten optimiert wurden, können die Skalierungsparameter entsprechend kleiner gemacht werden, damit ein einzelner Restraint nicht so viel Gewicht bekommt. Dazu müssen die Parameter *md.hot.noe*, *md.hot.cdih*, *md.cool.noe*, *md.cool.cdih*, *md.cart.noe*, *md.cart.cdih*, *md.pow.noe* und *md.pow.cdih* im Skript *anneal.inp* verändert werden. Im Protokoll des Waterrefinements geschieht dies durch Veränderung der Parameter *scale* in den jeweiligen die entsprechenden Restraints betreffenden Abschnitten. Eine sinnvolle Wahl erkennt man daran, dass sie die Energiebeiträge der physikalischen Potentiale auf ein Minimum reduzieren. Dieses Vorgehen ist insbesondere bei geringer Sequenzähnlichkeit sinnvoll, da die vorgeschlagenen Restraints die Konfiguration zwar formen sollen, wenn die übertragenen Informationen gut passen, aber eben auch keinen zu großen Einfluss haben sollen, wenn durch die geänderte Sequenz ein Restraint nicht mehr realisierbar und damit falsch ist. Um dabei die Wasserstoffbrücken, die in CNS intern genauso wie Abstandsrestraints gehandhabt werden, nicht zu niedrig zu gewichten, kann man sie als Ausgleich mehrfach einlesen. Indem der Skalierungsparameter kleiner gewählt wird, werden die biharmonischen Potentiale der Abstandsrestraints flacher, wodurch die Kräfte bei Nichterfüllung kleiner werden und somit der Einfluss der entsprechenden Bedingung verringert wird.

- Die Zeitintervalle (Parameter *md.hot.ss*, *md.cool.ss*, *md.cart.ss* und *md.pow.ss*), nach denen während der Simulation die Bewegung unterbrochen werden soll, um neue Kräfte und damit eine Geschwindigkeitskorrektur zu berechnen, müssen deutlich kleiner gewählt werden, weil durch die große Anzahl von Restraints die Summe der Kraft auf ein einzelnes Atom beziehungsweise seine Geschwindigkeit sehr groß werden kann und dieses Atom sich damit aus dem programmtechnisch vorgegebenen Bereich, in dem es sich aufhalten darf, hinausbewegen würde, was einen Abbruch der Simulation bewirkt. Entsprechend muss dafür die Anzahl der Iterationsschritte erhöht werden (Parameter *md.hot.step*, *md.cool.step*, *md.cart.step* und *md.pow.step*).

Modellierung von Komplexstrukturen

Zusätzlich zu den eben erwähnten Änderungen müssen für die Simulation eines Komplexes für drei der vier in Kapitel 2.2.1 erläuterten Schritte - Erzeugung der Strukturfiles, Generierung einer geeigneten Startstruktur, *Simulated Annealing* und *Waterrefinement* - Modifikationen an den Steuerfiles vorgenommen werden:

- **Erzeugung der Strukturfiles**
Für jede Sequenz muss ein eigenes Strukturfile erzeugt werden. Dabei ist darauf zu achten, dass die Nummer der ersten Aminosäure mit der in PERMOL gewählten übereinstimmen muss.
- **Generierung einer geeigneten Startstruktur**
Zur Generierung einer geeigneten Startstruktur müssen alle Strukturfiles gleichzeitig eingelesen und verarbeitet werden. Dazu sind im Steuerfile weitere Variable zu definieren (Abbildung 2-5 oben) und einzulesen (Abbildung 2-5 unten). Es entsteht eine PDB-Datei, die einen ausgestreckten Strang enthält, der aus mehreren Sequenzstücken besteht.
- **Waterrefinement**
Während es im Steuerfile für das *Simulated Annealing* bereits vorgesehen ist, mehrere Proteinsequenzen verwenden zu können, so dass sich mehrere Strukturfiles direkt abarbeiten lassen, muss im Steuerfile des *Waterrefinements* wie bei der Generierung der Startstruktur die Möglichkeit geschaffen werden, ein oder mehrere weitere Proteine simultan modellieren zu können.

```

{===== molecular structure =====}

{* structure file(s) *}
{==>} structure_file_1="PSF/byr2.mtf";
{==>} structure_file_2="PSF/ras.mtf";
{==>} structure_file_3="";

...

structure
  if (&structure_file_1 # " ") then
    @&structure_file_1
  end if
  if (&structure_file_2 # " ") then
    @&structure_file_2
  end if
  if (&structure_file_3 # " ") then
    @&structure_file_3
  end if
end

```

Abbildung 2-5: Das modifizierte Steuerfile generate_extended.inp

Ausschnitte aus dem Steuerfile generate_extended.inp, das an den Stellen, wo der Pfad und Name der Strukturfiles definiert (oben) und eingelesen wird, entsprechend abgeändert wurde, dass zwei oder mehr Proteine verarbeitet werden können.

2.4 Implementierung von MultiStates

Das in dieser Arbeit entwickelte Programm MultiStates dient dazu, aus den chemischen Verschiebungen der Atome eines Proteins, die für verschiedene Drücke bestimmt wurden, Rückschlüsse auf die Populationsverhältnisse der vermuteten Zustände in Abhängigkeit vom Druck zu bestimmen unter der Voraussetzung, dass ein schneller Austausch zwischen den Zuständen vorliegt. Insbesondere soll die Möglichkeit geschaffen werden, Aminosäuren zu identifizieren, die, wie in Kapitel 1.1.5 beschrieben, maßgeblich am Übergang zu einem speziellen Zustand beteiligt sind.

Im Folgenden wird die Anwendung des Algorithmus nach Levenberg-Marquardt auf das Problem beschrieben.

2.4.1 Anwendung des Levenberg-Marquardt-Verfahrens und einer alternativen Lösungsmöglichkeit

Formulierung des Minimierungsproblems

Ausgangspunkt ist die Gleichung 12

$$\delta_{lp}^{korr} = \delta_{lp} - korr_{lp} = \sum_m a_{lm}^{korr} * n_{mp},$$

die für alle Atome, die in die Auswertung miteinbezogen werden sollen, gelten muss.

Stellt man nun die gemessenen gemittelten chemischen Verschiebungen für alle betrachteten Atome und Drücke als $L \times P$ -Matrix dar, $D = (\delta_{lp}^{korr})$, und definiert eine Matrix $A = (a_{lm}^{korr})$ als $L \times M$ -Matrix aus den druckkorrigierten chemischen Verschiebungen der einzelnen Zustände für die betrachteten Atome und eine $M \times P$ -Matrix $N = (n_{mp})$, die die Besetzungszahlen enthält, kann diese Gleichung zu $D = A * N$ umformuliert werden.

Da mögliche Fehler bei den Messwerten δ_{lp}^{korr} nicht zu Widersprüchen bei der Bestimmung der Unbekannten führen dürfen und zugleich nach einer Lösung gesucht wird, die möglichst nahe an den Messwerten liegt, kann eine Fehlermatrix R eingeführt werden, die die Abweichungen r_{lp} enthält:

$$D + R = A * N$$

Daraus folgt:

$$R = A * N - D \quad (\text{Gleichung 13})$$

Es wird angenommen, dass so viele Messwerte vorliegen, damit dieses Gleichungssystem

überbestimmt ist und deswegen keine exakte Lösung gefunden werden kann. Zum Auffinden der am besten zu den Messwerten passenden Lösung wird R nun nach der Methode der kleinsten quadratischen Abweichung minimiert. Das ist gleichbedeutend mit der Minimierung der Frobeniusnorm der Matrix R , die als Wurzel aus der Summe aller Betragsquadrate aller Matricelemente definiert ist:

$$\|R\|_2 = \|A * N - D\|_2 = \sqrt{\sum_l \sum_p r_{lp}^2} = \sqrt{\sum_l \sum_p [(\sum_m a_{lm}^{korr} * n_{mp}) - \delta_{lp}^{korr}]^2} \stackrel{!}{=} \min$$

Dies wiederum entspricht der Minimierung des Betragsquadrats eines Vektors F mit folgender Definition:

$$\|F\|_2^2 = \left\| \begin{pmatrix} r_{11} \\ \vdots \\ r_{1p} \\ r_{12} \\ \vdots \\ r_{2p} \\ \vdots \\ r_{L1} \\ \vdots \\ r_{LP} \end{pmatrix} \right\|_2^2 = \left\| \begin{pmatrix} \sum_m a_{1m}^{korr} n_{m1} - \delta_{11}^{korr} \\ \vdots \\ \sum_m a_{1m}^{korr} n_{mp} - \delta_{1p}^{korr} \\ \sum_m a_{2m}^{korr} n_{m1} - \delta_{21}^{korr} \\ \vdots \\ \sum_m a_{2m}^{korr} n_{mp} - \delta_{2p}^{korr} \\ \vdots \\ \sum_m a_{Lm}^{korr} n_{m1} - \delta_{L1}^{korr} \\ \vdots \\ \sum_m a_{Lm}^{korr} n_{mp} - \delta_{LP}^{korr} \end{pmatrix} \right\|_2^2 \stackrel{!}{=} \min \quad (Definition 1)$$

Dieser zu minimierende Ausdruck wird im Folgenden "Norm" genannt. Gegenüber R hat sie den Vorteil, dass sie nicht matrix-, sondern vektorwertig und außerdem durch das Quadrat leichter zu handhaben ist.

Zur Lösung dieses Minimierungsproblems nimmt man für die Elemente der Matrix N folgende Einschränkungen vor:

1. Für jeden Druck p muss die Summe aller Besetzungszahlen 1 ergeben. Das bedeutet für die Elemente jeder Spalte von N : $\sum_m n_{mp} = 1$
2. Für die Elemente jeder Zeile, also für die Besetzungszahlen eines Zustandes m , erlaubt man nur stetige Änderungen mit dem Druck. Über Gleichung 5 und Gleichung 6 werden sie nach Boltzmann durch das Verhalten der Größen $\Delta G_{lm}(p)$, die die Differenzen der Gibbs-Energien zwischen den Zuständen l und m darstellen, mit dem Druck verknüpft. Ohne Einschränkung des Systems kann Zustand 1 als Referenz gewählt werden, bezüglich der alle Differenzen angegeben werden, so dass

$$\Delta G_m(p) := \frac{\Delta G_{1m}(p)}{RT} \quad (\text{Definition 2})$$

für alle $m = 2, \dots, M$ definiert sei. Näherungsweise kann das Verhalten mit dem Druck mit einer Taylorentwicklung an der Stelle p_0 für alle p beschrieben werden,

$$\Delta G_m(p) = \Delta G_m^0 + \Delta V_m^0(p - p_0) - \frac{\Delta b_m^{0'}}{2}(p - p_0)^2 \text{ für alle } m = 2, \dots, M,$$

wobei ΔG_m^0 der Differenz der Gibbs-Energie der Zustände 1 und m an der Stelle p_0 geteilt durch die universelle Gaskonstante R und die Temperatur T entspricht. ΔV_m^0 und $\Delta b_m^{0'}$ können physikalisch als partielle molare Volumendifferenz beziehungsweise Differenz der partiellen molaren Kompressibilitätsfaktoren, jeweils ebenso durch $R * T$ geteilt, gedeutet werden. Dies entspricht einer Modifizierung von Gleichung 6 mit Definition 2. Die Größen ΔG_m^0 , ΔV_m^0 und $\Delta b_m^{0'}$ werden im Folgenden modifizierte Differenzen der Gibbs-Energie, des partiellen molaren Volumens und des partiellen molaren Kompressibilitätsfaktors genannt.

Mit diesen Einschränkungen ergibt sich ein Fit-Funktionssystem mit $L * M + 3(M - 1)$ Parametern, die durch gleichzeitige Optimierung von $L * P$ Gleichungen so bestimmt werden sollen, dass die Norm minimal wird.

Da aber experimentelle Daten immer mit Messfehlern behaftet sind, kann das mathematische Minimum, das durch die Optimierung gefunden wird, nur als Näherung für den eigentlich das System erzeugenden Parametersatz gesehen werden. Deshalb muss zusätzlich untersucht werden, wie groß der Bereich um das mathematische Minimum ist, in dem der echte Parametersatz mit ziemlicher Sicherheit liegt. Die Größe dieses Gebiets wiederum ist von mehreren Faktoren abhängig - von der Anzahl der Atome, für die chemische Verschiebungen zur Verfügung stehen, vom mittleren Fehler pro Messpunkt, von der Anzahl der Druckmesspunkte und von der Lage des gefundenen Minimums im Parameterraum.

Im Folgenden werden nun Methoden zur Bestimmung des mathematischen Minimums vorgestellt und im Anschluss daran eine Vorgehensweise vorgeschlagen, die zusätzlich dazu das Gebiet rund um das gefundene Extremum untersucht und einen Bereich eingrenzt, der den echten Parametersatz mit hoher Wahrscheinlichkeit enthält.

Algorithmus 1 - Lösung nach Levenberg-Marquardt

Als Lösungsweg bietet sich der Algorithmus nach Levenberg-Marquardt (Dahmen und Reusken, 2008) an, der eine Variation des Gauß-Newton-Verfahrens ist mit dem Vorteil, dass bei korrekter Handhabung Minima nicht übersprungen werden. Bei diesem Lösungsverfahren werden alle Parameter gleichzeitig optimiert, so dass ein nichtlineares Problem entsteht. Dessen Lösung wird iterativ über eine Reihe geeigneter linearer Probleme angenähert. Dazu fasst man die Matrix F als Funktion aller Parameter auf und ersetzt F durch eine lineare Approximation mittels Taylorentwicklung, bei der nur die nullte und erste Ordnung berücksichtigt wird. Dadurch ergibt sich folgendes Minimierungsproblem:

$$\|F(x_0) + D_F(x_0)\Delta x\|_2^2 \stackrel{!}{=} \min \quad (\text{Gleichung 14})$$

In den Spaltenvektoren x_0 und Δx mit x_0 und $\Delta x \in \mathbb{R}^{LM+3(M-1)}$ sind die Parameter zusammengefasst. D_F ist die Jakobimatrix von F , die durch Bildung der partiellen Ableitungen der Matrixelemente r_{lp} nach den verschiedenen Parametern entsteht, die nun explizit berechnet werden:

Für r_{lp} gilt:

$$r_{lp} = \left(\sum_m a_{lm}^{korr} * n_{mp} \right) - \delta_{lp}^{korr}$$

Damit gilt für $\frac{\partial r_{lp}}{\partial a_{ij}^{korr}}$ mit $i = 1, \dots, L$ und $j = 1, \dots, M$:

$$\frac{\partial r_{lp}}{\partial a_{ij}^{korr}} = n_{jp} \text{ falls } i = l \quad (\text{Gleichung 15})$$

$$\frac{\partial r_{lp}}{\partial a_{ij}^{korr}} = 0 \text{ falls } i \neq l \quad (\text{Gleichung 16})$$

Die partiellen Ableitungen nach ΔG_m^0 , ΔV_m^0 und Δb_m^0 sind:

$$\frac{\partial r_{lp}}{\partial \Delta G_i^0} = \sum_m a_{lm}^{korr} \frac{\partial n_{mp}}{\partial \Delta G_m^0} = n_{ip} \left(\sum_m a_{lm}^{korr} n_{mp} - a_{li}^{korr} \right) \quad (\text{Gleichung 17})$$

$$\frac{\partial r_{lp}}{\partial \Delta V_i^0} = \sum_m a_{lm}^{korr} \frac{\partial n_{mp}}{\partial \Delta V_i^0} = n_{ip} (p_p - p_0) \left(\sum_m a_{lm}^{korr} n_{mp} - a_{li}^{korr} \right) \quad (\text{Gleichung 18})$$

$$\frac{\partial r_{lp}}{\partial \Delta b_i^0} = \sum_m a_{lm}^{korr} \frac{\partial n_{mp}}{\partial \Delta b_i^0} = -n_{ip} \frac{(p_p - p_0)^2}{2} \left(\sum_m a_{lm}^{korr} n_{mp} - a_{li}^{korr} \right) \quad (\text{Gleichung 19})$$

Die daraus resultierende Form der Matrix D_F ist in Abbildung 2-6 dargestellt.

Wäre das Problem linear, so wäre $D_F(x_0)$ eine für alle x_0 konstante Matrix. Sie wäre also von

der Wahl der Parameter unabhängig. In diesem Fall müsste man das Gleichungssystem

$$D_F^T D_F \Delta x = -D_F^T F(x_0) \quad (\text{Gleichung 20})$$

für ein beliebiges x_0 nur einmal lösen und könnte den richtigen Parametersatz direkt durch $x_0 + \Delta x$ erhalten. Wenn sich aber wie hier D_F mit x_0 verändert und somit ein nichtlineares Ausgleichsproblem vorliegt, muss man diesen Schritt immer wieder wiederholen. Man erhält eine Folge von Vektoren $x_0^{k+1} = x_0^k + \Delta x^k$ in der Hoffnung, mit jedem Iterationsschritt k näher an das Minimum heranzukommen (Gauß-Newton-Verfahren), was bei diesem Verfahren nicht gewährleistet ist.

Der Levenberg-Marquardt-Algorithmus jedoch erzwingt absteigende Funktionswerte durch Einführung eines Regulierungsparameters μ , der in jedem Iterationsschritt angepasst wird.

Das Minimierungsproblem wird dann folgendermaßen formuliert:

$$\|F(x_0) + D_F(x_0)\Delta x\|_2^2 + \mu^2 \|\Delta x\|^2 \stackrel{!}{=} \min \quad (\text{Gleichung 21})$$

Dies ist gleichbedeutend mit:

$$\left\| \begin{pmatrix} F(x_0) \\ 0_{LM+3(M-1)} \end{pmatrix} + \begin{pmatrix} D_F(x_0) \\ \mu E_{LM+3(M-1)} \end{pmatrix} \Delta x \right\|_2^2 \stackrel{!}{=} \min \quad (\text{Gleichung 22})$$

Setzt man $\begin{pmatrix} F(x_0) \\ 0_{LM+3(M-1)} \end{pmatrix} = F^{mod}(x_0)$ und $\begin{pmatrix} D_F(x_0) \\ \mu E_{LM+3(M-1)} \end{pmatrix} = D_F^{mod}(x_0)$, erhält man:

$$\|F^{mod}(x_0) + D_F^{mod}(x_0)\Delta x\|_2^2 \stackrel{!}{=} \min \quad (\text{Gleichung 23})$$

Jetzt arbeitet man genauso wie beim Gauß-Newton-Verfahren, das heißt, man löst das Gleichungssystem

$$(D_F^{mod}(x_0))^T D_F^{mod}(x_0) \Delta x = -(D_F^{mod}(x_0))^T F^{mod}(x_0) \quad (\text{Gleichung 24})$$

solange, bis man ein lokales Minimum erreicht hat.

μ wird bei jeder Iteration so gewählt, dass man bei jedem Schritt zu einem kleineren $\|F\|_2^2$ kommt: Bewirkt die berechnete Korrektur Δx eine Vergrößerung der Norm, wird sie verworfen und eine neue Korrektur mit einem vergrößerten μ berechnet.

Das hat den Vorteil, dass – bei korrekter Handhabung des Algorithmus – das Minimum quasi nicht „übersprungen“ werden kann. Umgekehrt wird, wenn die Änderung pro Iterationsschritt zu stagnieren droht, der Regulierungsparameter μ verkleinert.

Sowohl die Methoden nach Gauss-Newton als auch die nach Levenberg-Marquardt haben den Nachteil, dass man keine Garantie hat, dass das gefundene Minimum das absolute Minimum ist. Deswegen ist es sinnvoll, von mehreren Startpunkten des Parameterraumes die Suche zu beginnen. Insbesondere können auch mehrere Minima mit der gleichen Norm von F gefunden werden.

Der Parameterraum hat $L * M + 3(M - 1)$ Dimensionen, $L * M$ Werte für die korrigierten chemischen Verschiebungen aller L Aminosäuren und aller M Zustände und die Werte für ΔG_m^0 , ΔV_m^0 und $\Delta b_m^{0'}$ aller $m = 2, \dots, M$, also die Parameter, die zur Beschreibung der Druckabhängigkeit der freien Enthalpie benötigt werden. Für diese Parameter müssen sinnvolle Anfangswerte für den Algorithmus gewählt werden.

Als Startwerte für die druckkorrigierten chemischen Verschiebungen (Matrix A) eignen sich Werte nahe den gemessenen korrigierten chemischen Verschiebungen. Deswegen ist es naheliegend, den ersten Zustand mit den gemessenen korrigierten chemischen Verschiebungen bei dem Druckwert p_0 , den letzte Zustand mit denen des höchsten Drucks und alle weiteren mit Werten dazwischen zu initialisieren. Dahinter steckt die Überlegung, dass Zustand 1 der Zustand sein soll, der bei p_0 am höchsten populiert ist, während der letzte Zustand am Ende der Druckreihe dominieren soll.

Diese Initialisierung der Matrix A kann nun mit verschiedenen Kombinationen aus ΔG_m^0 , ΔV_m^0 und $\Delta b_m^{0'}$ kombiniert werden, um die Wahrscheinlichkeit zu erhöhen, dass der Algorithmus das richtige Minimum findet.

Algorithmus 2 - Feine Rasterung des Parameterraums der Größen ΔG_m^0 , ΔV_m^0 und $\Delta b_m^{0'}$

Als alternativer Lösungsweg kann der durch die Größen ΔG_m^0 , ΔV_m^0 und $\Delta b_m^{0'}$ mit $m = 2, \dots, M$ aufgespannte Parameterraum gerastert werden. Durch die Koordinaten der Rasterpunkte werden die Funktionen der Gibbs Energiedifferenzen festgelegt und damit die Matrix N bestimmt.

Der mathematische Ansatz ist genauso wie bei dem Vorgehen nach Levenberg-Marquardt. Da jedoch die Parameter ΔG_m^0 , ΔV_m^0 und $\Delta b_m^{0'}$ fixiert sind, besteht die Jakobimatrix von F nur aus der rechten Hälfte der in Abbildung 2-6 für Algorithmus 1 gezeigten Matrix. Dadurch wird sie für alle Δx konstant und Gleichung 12 zu einem überbestimmten linearen Gleichungssystem, bei dem exakt in nur einem Schritt mit linearer Regression die Matrix A der hypothetischen korrigierten chemischen Verschiebungen so berechnet werden kann, dass die Norm minimal wird. Deren Wert wird dann für jeden Punkt des Raumes gespeichert.

Bei der Anwendung dieses Verfahrens muss bedacht werden, dass die Anzahl der zu untersuchenden Rasterpunkte und die dadurch entstehenden Rechenzeiten und Datenmengen exponentiell mit der Zahl M ansteigen. In dieser Arbeit wird es vor allem zur graphischen Darstellung des Parameterraumes von Zweizustandssystemen verwendet.

2.4.2 Besonderheiten des Parameterraums

Anzahl der Zustände

Wendet man die hier vorgestellten Algorithmen auf einen experimentellen Datensatz an, so ist im Allgemeinen nicht klar, wie viele relevante Zustände das zu untersuchende Makromolekül besitzt. Interessant sind nur diejenigen, die eine so geringe Energie haben, dass ihre Populationsgröße einen experimentellen Nachweis erlaubt. Von Proteinen ist bekannt, dass Konfiguration und Funktion eng miteinander verknüpft sind. Geht man davon aus, dass die Entstehung einzelner Zustände durch Anpassung an die jeweiligen Bindungspartner erfolgte, kann man sich bei der Wahl der Zahl M an deren Anzahl orientieren. Dabei ist eventuell die Anzahl der Zustände des Proteins, die im langsamen Austausch miteinander stehen, abzuziehen.

Ein sinnvolles Vorgehen ist es, die Anzahl der angenommenen Zustände sukzessive zu erhöhen, bis das Modell die vorhandenen Daten erklärt. Als Anhaltspunkt kann die zu erwartende Norm verwendet werden, die der Wurzel der Summe der Quadrate aller Messfehler entspricht. Deren Größenordnung lässt sich aus der erzielten Auflösung des

Spektrums abschätzen. Diese ist nicht nur von den Akquisitionsparametern des Spektrums abhängig, sondern zusätzlich vor allem von der Relaxationszeit und der Struktur des untersuchten Proteins.

Liegen ideale oder nahezu ideale, also mit nur sehr kleinen Messfehlern behaftete Daten vor, können die aus dem Fit erhaltenen Besetzungszahlen als Kriterium dafür dienen, ob die angenommene Anzahl der Zustände korrekt ist oder nicht. Hat man die korrekte Zahl der relevanten Zustände bereits erreicht und nimmt einen weiteren dazu, können bei Messwerten ohne Fehler folgende Fälle auftreten:

1. Die Besetzungszahlen der relevanten Zustände ändern sich kaum und der neu hinzugenommene Zustand ist über den gesamten Messbereich nur sehr schwach populiert.
2. Einer der relevanten Zustände wird quasi aufgespalten in zwei Zustände. Deren Besetzungszahlen zeigen über den gesamten Messbereich gleiches Verhalten.
3. Einer der relevanten Zustände wird quasi aufgespalten und die beiden resultierenden Zustände zeigen spiegelsymmetrisches Verhalten um den eigentlichen Zustand.

Daraus folgt, dass bei Annahme von mehr als den eigentlich relevanten Zuständen im Prinzip unendlich viele Minima im Parameterraum vorliegen können.

Bei experimentellen Daten mit Messfehlern allerdings werden vor allem diejenigen Parametersätze durch eine besonders niedrige Norm begünstigt, die einen Teil der Fehler beim Fit ausgleichen. Deswegen kann dieses Kriterium nur bedingt als Maßstab herangezogen werden.

Eine Ausnahme bilden Zustände, die über den gesamten Druckbereich eine annähernd gleiche Besetzung aufweisen. Da die beschriebenen Methoden auf der Registrierung von Änderungen der chemischen Verschiebung mit dem Druck beruhen und konstante Besetzungen keine Änderung hervorrufen, können sie nicht nachgewiesen werden. Daraus folgt, dass eine weitere Anzahl Zustände existieren kann, deren Nachweis mit dieser Methode unmöglich ist und deren Besetzungszahlen trotzdem in der Größenordnung der für die Algorithmen sichtbaren Zustände liegen könnten.

Symmetrien im Parameterraum

Bei den hier vorgestellten Algorithmen werden nicht die absoluten Werte der Variablen der Taylorentwicklung, sondern die Differenzen ΔG_m^0 , ΔV_m^0 und $\Delta b_m^{0'}$ betrachtet, die laut Definition 1 als Differenzen zwischen den Größen des ersten und des m-ten Zustands für alle $m = 2, \dots, M$ festgelegt sind. Zur Differenzbildung müssen die Zustände nummeriert werden, wofür es bei M Zuständen $M!$ verschiedene Möglichkeiten gibt. Diese Permutationen repräsentieren alle dieselbe Lösung und bilden mit Matrizen A' eine minimale Norm, die aus Vertauschung der Spalten aus A hervorgehen.

Bei einem Zweizustandssystem heißt das, dass zwei gleichwertige Minima ($\Delta G_{m,min}^0$, $\Delta V_{m,min}^0$, $\Delta b_{m,min}^{0'}$) und $(-\Delta G_{m,min}^0, -\Delta V_{m,min}^0, -\Delta b_{m,min}^{0'})$ existieren und der durch ΔG_2^0 , ΔV_2^0 und $\Delta b_2^{0'}$ aufgespannte Raum bei zu jedem Punkt optimierten korrigierten chemischen Verschiebungen eine Punktsymmetrie bezüglich der Norm aufweist.

Bei Dreizustandssystemen resultieren 6 gleichwertige Minima (Abbildung 2-7), bei Vierzustandssystemen sind es bereits 24. Bei der Implementierung der Algorithmen muss also einerseits darauf geachtet werden, dass gleichwertige Lösungen auch als solche erkannt werden, andererseits kann durch die Ausnutzung der dadurch entstehenden Symmetrien der Raum, der nach dem Minimum abgesucht wird, eingeschränkt werden.

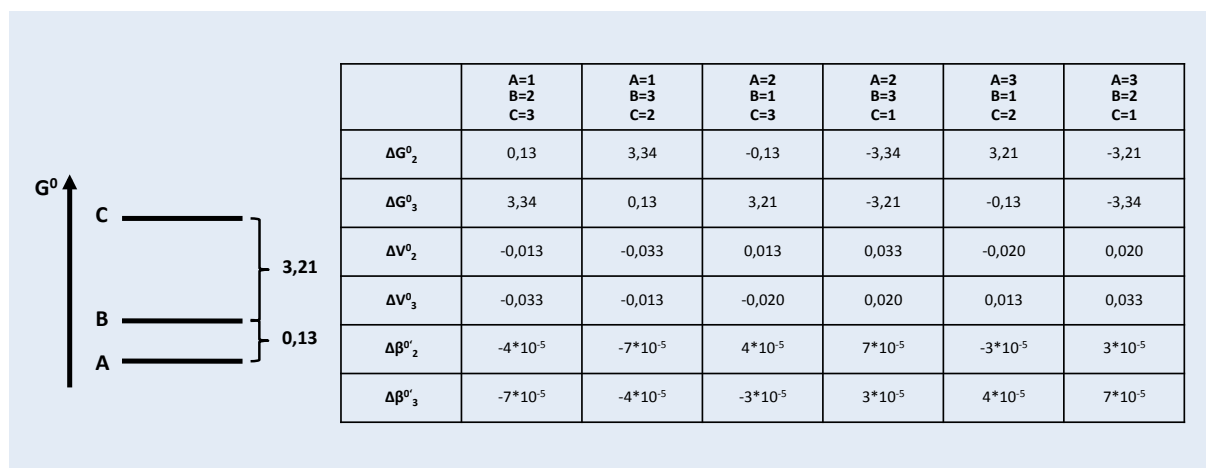


Abbildung 2-7: Bei Dreizustandssystemen resultieren 6 gleichwertige Minima

Zahlenbeispiel: Durch Permutation bei der Benennung der Zustände erhält man bei einem Dreizustandssystem sechs Minima im Parameterraum, die alle dieselbe Lösung repräsentieren.

Größenordnungen der Parameter ΔG_m^0 , ΔV_m^0 und Δb_m^0

Für eine erfolgreiche Bestimmung des Minimums muss der Parameterraum in dem Bereich untersucht werden, in dem das Extremum zu erwarten ist. Orientieren kann man sich dazu an Werten von Proteinen, für die diese Parameter bereits bestimmt wurden. In Tabelle 2-1 sind einige Beispiele zusammengestellt.

Aus Gleichung 5 und Definition 1 folgt $-\ln \frac{n_1(p_0)}{n_m(p_0)} = \Delta G_m^0$. Wählt man für ΔG_m^0 Werte zwischen 0 und 4,5, erfasst man Besetzungsverhältnisse zwischen 1:1 und 1:99. Damit lässt sich der bei Normaldruck eigentlich relevante Bereich abdecken und das Verhalten der interessierenden Zustände beschreiben. Dieser Bereich ist allerdings nicht ausreichend, da mit Druckänderung auch die Populationsanteile von Zuständen, die bei Normaldruck keine nachweisbare Besetzung haben, soweit zunehmen können, dass sie nicht vernachlässigt werden dürfen.

Als höchste zu testende Werte für ΔG_m^0 können die Werte angesehen werden, die beim Übergang zur Entfaltung auftreten, die bei höheren Drücken einsetzt, was im Bild der Zustände eine Populationszunahme des denaturierten Zustands bedeutet. Für die Ras-bindende Domäne von RalGDS wurde beispielsweise gezeigt, dass bei 200 MPa bereits 30% des Proteins denaturiert vorliegen (Inoue et al., 2000). Die Gibbs-Energiedifferenzen bei Normaldruck zwischen den bei Normaldruck dominierenden und denaturierten Zuständen sind vor allem abhängig von Proteingröße und Polymerisationszustand und liegen zwischen 20 und 60 kJ/mol. Das entspricht Werten von ΔG_m^0 von etwa 8 bis 24. Allerdings laufen solche Entfaltungen normalerweise auf einer langsameren Zeitskala ab und führen deswegen nicht zu einer Änderung der chemischen Verschiebung, sondern zu einer Volumenabnahme der Peaks. Daher können sie im Allgemeinen vernachlässigt werden.

Der Parameter ΔV_m^0 kann positive und negative Werte annehmen, je nachdem, ob das Verhältnis der Besetzungszahlen sich mit steigendem Druck vergrößert oder verkleinert. Nimmt man als größte Änderung an, dass das Populationsverhältnis eines Zustands während einer Druckänderung von 200 MPa von einer nicht nachweisbaren Besetzung auf 30% ansteigt, folgt daraus $\Delta V_m^0 \approx \frac{-10-0,4}{200 \text{ MPa}} \approx -0,05 \text{ MPa}^{-1}$ beziehungsweise $+0,05 \text{ MPa}^{-1}$.

Der in der Arbeit von Kalbitzer et al. angewandte und hier übernommene Ansatz, für den

Verlauf für $\Delta G_m(p)$ eine Taylornäherung anzusetzen, beruht darauf, dass man einen annähernd linearen Verlauf mit dem Druck erwartet (Kalbitzer et al., 2013). Die Werte des Parameters $\Delta b_m^{0'}$, der das quadratische Glied moduliert, sind deshalb um mindestens zwei Größenordnungen kleiner als die von ΔV_m^0 anzunehmen.

Tabelle 2-1: Experimentell bestimmte Parameter $\Delta G_m^0 (= \frac{\Delta G_0}{RT})$ und $\Delta V_m^0 (= \frac{\Delta V_0}{RT})$ für einige Proteine

Protein	$\Delta G_0/RT$	$\Delta V_0/RT$ [in MPa ⁻¹]
Ras(wt)*	0,60	-0,007
	2,06	-0,032
	4,92	-0,046
ShaPrP-(90-230)**	1,3	-0,019
	4,9	nicht bestimmt
	9,2	nicht bestimmt
huPrP-(121-230)**	1,2	-0,034
	4,3	-0,026
	7,4	-0,050
A β ***	0,67	-0,017
	1,31	-0,005

* Kalbitzer et al., 2013

** Kremer et al., 2007

*** Munte et al., 2013, 277 Kelvin

2.4.3 Erzeugung und Einsatz von Testdatensätzen

Erzeugung

Um die vorgestellten Algorithmen testen und optimieren und außerdem die Abhängigkeiten des Lösungsbereichs von verschiedenen Parametern des betrachteten Gesamtsystems untersuchen zu können, müssen Testdatensätze aus Sets von chemischen Verschiebungen erzeugt werden, die der Theorie gehorchen.

Zur Erzeugung der Matrix A wählt man zuerst die Anzahl der betrachteten Atome L und die Anzahl M der Zustände. Für die in dieser Arbeit untersuchten Testdaten wurde M gleich 2, 3 oder 4 und L gleich 80 verwendet. Die $L * M$ Elemente, die die Matrix A füllen und die korrigierten chemischen Verschiebungen der Zustände darstellen, werden zufällig aus einem Bereich bestimmt, der den experimentellen Daten entspricht, mit denen die Ergebnisse zu vergleichen sind. Da Systeme erzeugt werden sollten, die Datenreihen, bei denen

Amidprotonen gemessen wurden, ähneln, wurden diese Werte a_{lm} aus einem Bereich zwischen 8,0 ppm und 9,2 ppm gewählt, wodurch bei zwei Zuständen eine mittlere absolute Differenz von 0,3 ppm resultiert. Als angenommene Messpunkte wurden die Druckpunkte 3 MPa, 25 MPa, 50 MPa, 75 MPa, 100 MPa, 125 MPa, 150 MPa, 175 MPa, 200 MPa und 225 MPa für den Vektor p verwendet.

Die Elemente der Matrix N werden durch ΔG_m^0 , ΔV_m^0 und $\Delta b_m^{0'}$ bestimmt, welche zufällig oder systematisch gewählt werden können. Ein idealer Testdatensatz lässt sich mit $D_{ideal} = A * N$ generieren.

Statistische Fehler

Den Einfluss von statistischen Messfehlern kann man dadurch abschätzen, indem bei diesen idealen Datensätzen zusätzlich Fehler verschiedener Größenordnung angelegt werden. Dazu werden Fehlermatrizen $R_{erzeugend}$ verschiedener Größenordnungen erzeugt, die dann mit $D_{test} = D_{ideal} + R_{erzeugend}$ die zu testenden Sets von chemischen Verschiebungen D_{test} bilden. Die Matrixelemente von $R_{erzeugend}$ werden so gewählt, dass sie eine Normalverteilung um Null bilden. Die Breite dieser Verteilung ist abhängig von dem mittleren statistischen Fehler f_{mittel} , der für die Daten angenommen werden soll.

Bei der Testdatenerzeugung wurden Fehlermatrizen R in drei Größenordnungen verwendet. Als mittlerer Fehler pro Messpunkt und Aminosäure wurde entweder $1,8 \cdot 10^{-5}$ ppm (bezeichnet mit s), $1,8 \cdot 10^{-4}$ ppm (m) oder $1,8 \cdot 10^{-3}$ ppm (l) angenommen, was bei dem angenommenen Bereich für die chemischen Verschiebungsdifferenzen einem relativen Messfehler $\frac{\Delta\Delta\delta}{\Delta a}$ von 0,00006 (s), 0,0006 (m) beziehungsweise 0,006 (l) entspricht. Bei 80 Atomen und 10 Messpunkten entspricht das in etwa den Normen 0,0005 (s), 0,005 (m) und 0,05 (l). Ideale Datensätze wurden mit Fehler 0 gekennzeichnet. Aus diesen Größen wurde dann für jeden Datensatz einer Datenreihe die Matrix D mit $D = A * N + R$ berechnet.

Wendet man die Algorithmen auf einen mit Fehlern erzeugten Testdatensatz an, wird anhand der durchgeführten Ausgleichsrechnung das mathematische Minimum des Systems gefunden. Dabei kann es passieren, dass dieses Extremum nicht exakt den Parametern entspricht, die zur Erzeugung des Datensatzes dienten, die gefundene Norm also kleiner ist als die am gesuchten Ort im Parameterraum. Im Falle der in dieser Arbeit getesteten experimentellen Datensätze ist

dies zu erwarten, weil zwar die Elemente von $R_{erzeugend}$ - also die Messfehler - als statistisch normalverteilt angenommen werden können, aber die Anzahl der gemessenen Werte einer Druckkurve eines Atoms, die bei den vorliegenden experimentellen Datensätzen entweder 10 oder 11 beträgt, so klein ist, dass die idealen Parameter zum Fitten beeinflusst werden, zumal auch eine Anpassung der Elemente der Matrix A erlaubt ist.

Will man das Verhalten der Testverfahren an Daten mit Fehlern testen, ist auch hier das mathematische Minimum zunächst nicht bekannt. Ein Test muss deswegen in zwei Stufen erfolgen: Zuerst wird das mathematische Minimum gesucht, indem man das beschriebene Levenberg-Marquardt-Verfahren verwendet und dabei als Startpunkt die Koordinaten des erzeugenden Parametersatzes wählt. Von diesem ist bekannt, dass er sich in der Nähe des mathematischen Minimums befindet und so mit großer Sicherheit gefunden werden kann.

In einem zweiten Schritt wird der zu testende Algorithmus, der den gesamten Parameterraum absucht, auf den Datensatz angewandt und dann verglichen, ob das richtige Extremum gefunden werden kann.

Einfluss statistischer Messfehler

Erzeugt man zu einem idealen Datensatz D_{ideal} nicht nur eine, sondern statistisch viele zufällige Fehlermatrizen $R_{erzeugend}$ zu einem festgelegten mittleren statistischen Fehler f_{mittel} , wird ein Bereich im durch ΔG_m^0 , ΔV_m^0 und $\Delta b_m^{0'}$ aufgespannten Parameterraum um den tatsächlichen Parametersatz gefunden, in dem die zugehörigen mathematischen Minima liegen. Durch Testen von verschiedenen Werten für f_{mittel} kann man Abhängigkeiten zwischen f_{mittel} und der Größe dieses gefundenen Bereichs feststellen. Ebenfalls untersucht werden kann die mittlere Differenz zwischen der Norm an der Stelle der erzeugenden Parameter im Orbit und den minimal gefundenen Normen.

Eine Abhängigkeit dieser Größen von der Lage des tatsächlichen Datensatzes im Parameterraum lässt sich durch Variation von D_{ideal} bestimmen.

2.4.4 Bestimmung eines Lösungsbereichs im Parameterraum

Als ersten Schritt zur Lösung des Problems muss das mathematische Minimum wie oben beschrieben nach der Methode der kleinsten Quadrate gefunden werden. Aus der Norm am Minimum $Norm_{math}$ kann dann eine Norm festgelegt werden, die als obere Grenze verwendet wird, bis zu der umliegende Orte um das Extremum des Parameterraums dem Lösungsbereich

zugerechnet werden. Im zweiten Schritt wird die Umgebung des mathematischen Minimums systematisch gerastert oder zufällig Punkte ausgewählt, für die jeweils die minimale Norm durch Optimierung der Matrix A bestimmt wird und die dann der Lösung zugerechnet werden, wenn die Norm unter der festgelegten Schwelle liegt.

Zur Abschätzung eines passenden Schwellenwerts der Norm kann nach Sachs aus der Norm am mathematischen Minimum der durchschnittliche Messfehler σ berechnet werden,

$$\sigma = \frac{Norm_{math}}{\sqrt{L * P - K}} \quad (Gleichung 25),$$

wobei K die Anzahl der Parameter, $K = L * M + 3 * (M - 1)$, ist (Sachs, 2004). Damit kann man die Norm der erzeugenden Fehlermatrix $R_{erzeugend}$ zu

$$Norm_{erzeugend} = Norm_{math} * \sqrt{\frac{L * P}{L * P - K}} \quad (Gleichung 26)$$

bestimmen. Allerdings wird bei der Suche nach dem Minimum nicht $Norm_{min}$ mit der Norm der Fehlermatrix verglichen, da bei der Optimierung auch die Matrix A angepasst wird und somit ein kleinerer Wert am Ort des erzeugenden Parametersatzes gefunden wird, das mit

$$Norm_{opt} = Norm_{math} * \sqrt{\frac{L * P - L * M}{L * P - K}} \quad (Gleichung 27)$$

abgeschätzt werden kann. Laut Sachs bilden alle Punkte des Parameterraums, deren Norm kleiner oder gleich dieser Schwelle $Norm_{opt}$ ist, ein Konfidenzgebiet mit einem Konfidenzniveau von 68% (Sachs, 2004).

Diese Abschätzung des Lösungsbereichs ist allerdings nur korrekt, wenn die Messfehler des zu untersuchenden Systems statistisch gaußverteilt sind. Strenggenommen darf eine Datenkombination - wie in Kapitel 2.2.4 vorgeschlagen - nicht angewendet werden, da durch diese Transformation der Daten die Verteilungsdichte der Unsicherheiten ebenfalls verändert wird und der Lösungsbereich nicht mehr bestimmt werden kann.

2.4.5 Implementierung der entwickelten Verfahren mit MATLAB

Übersicht

Zur Erstellung des Programms MultiStates wurde MATLAB verwendet (Versionen R2013b und R2014b), da diese Software speziell für das Rechnen mit großen Matrizen viele Funktionen bereithält, die den Umgang damit komfortabel machen.

Die Algorithmen benötigen folgende Eingabeparameter:

- Die Zahl M (die Anzahl der angenommenen Zustände)
- den Zeilenvektor p (der die Drücke in aufsteigender Reihenfolge in der Einheit Megapascal enthält)
- die $L \times P$ -Matrix D (die die korrigierten chemischen Verschiebungen aller Atome - L ist Anzahl der Atome - an allen gemessenen Druckpunkten - P ist Anzahl der Druckpunkte - enthält)

Die erstellten Funktionen sind so konzipiert, dass das erste Element von p , der Normaldruck, automatisch als das p_0 des Algorithmus gewählt wird, um das die Taylorentwicklung von $G_{lm}(p)$ stattfindet.

Rückgabewerte sind die Parameter des mathematischen Minimums, also die Parameter ΔG_m^0 , ΔV_m^0 und Δb_m^0 , die die Gibbs Energiedifferenzen an der Stelle p_0 , die Differenzen der partiellen molaren Volumina und die Differenzen der partiellen molaren Kompressibilitätsfaktoren, jeweils geteilt durch $R \cdot T$, darstellen, die zugehörige Matrix A , die die korrigierten chemischen Verschiebungen der Zustände der Atome enthält, und die berechnete Norm.

Implementierung von Algorithmus 1

Kern des Programms ist die Lösung des linearen Gleichungssystems

$$(D_F^{mod}(x))^T D_F^{mod}(x) \Delta x = -(D_F^{mod}(x))^T F^{mod}(x) \quad (\text{Gleichung 28})$$

mit der Methode von Levenberg-Marquardt. Dazu müssen in jedem Iterationsschritt die Matrizen $F(x)$ und $D_F(x)$, die zu minimierende Funktion Norm und deren Jakobimatrix, berechnet werden.

Die modifizierten Matrizen müssen dabei nicht explizit erzeugt werden, da durch einfache Rechnung gezeigt werden kann, dass

$$(D_F^{mod}(x))^T D_F^{mod}(x) = (D_F(x))^T D_F(x) + \mu^2 E(L \cdot M + 3(M - 1))$$

und

$$(D_F^{mod}(x))^T F^{mod}(x) = (D_F(x))^T F(x)$$

gilt. Deswegen wurde alternativ die einfachere Gleichung

$$(D_F(x))^T D_F(x) + \mu^2 E(L \cdot M + 3(M - 1)) \Delta x = -(D_F(x))^T F(x) \quad (\text{Gleichung 29})$$

implementiert. Dafür stellt MATLAB die Funktion \ (Links-Division) zur Verfügung.

Um sicherzugehen, dass die Iteration auch ein Ende findet, falls der Algorithmus in Bereichen landet, in denen kein Minimum gefunden werden kann, wurden die Variablen *Stpcnt* und *Epsilon* eingeführt. *Epsilon* dient als obere Schranke für die Ableitung der Zielfunktion, wobei im Idealfall so lange iteriert wird, bis diese Schranke unterschritten ist. *Stpcnt* begrenzt die Anzahl der Iterationsschritte von einem Rasterpunkt aus. Dadurch werden zu lange Rechenzeiten von ungünstigen Rasterpunkten aus vermieden. Zur Untersuchung der Testdaten wurden *Epsilon* auf $1 \cdot 10^{-15}$ und *Stpcnt* auf 10000 gesetzt, wobei letztere Größe bei der Untersuchung der experimentellen Daten auf 40000 erhöht wurde, um eine höhere Sicherheit zu bekommen, das gesuchte Minimum zu finden.

Ein Überbau startet diesen Algorithmus von verschiedenen Punkten des Parameterraumes aus, um das mathematische Minimum zu finden. Eine geeignete Wahl der Startpunkte ergibt sich aus der Form der Funktion Norm über dem Parameterraum (siehe Kapitel 3.3.1).

Um für Testdaten mit Fehlern das mathematische Minimum zu finden und dann mit dem Minimum, das der eigentliche Algorithmus liefert, vergleichen zu können, kann an die Methode von Levenberg-Marquardt auch der das System erzeugende Parametersatz direkt übergeben werden (siehe Kapitel 2.4.3). Diese Methode wird im Folgenden Algorithmus 1M genannt.

Implementierung von Algorithmus 2

Bei Algorithmus 2 wird im Gegensatz zum Levenberg-Marquardt-Verfahren keine Iteration durchgeführt, um das Minimum zu finden, sondern stattdessen zu jedem Rasterpunkt des Parameterraums, aufgespannt aus ΔG_m^0 , ΔV_m^0 und $\Delta \beta_m^{0'}$ mit $m = 2, \dots, M$, die Matrix N berechnet und dann mithilfe linearer Regression, die in MATLAB wieder über Links-Division verwirklicht wurde, die Matrix A so bestimmt, dass die Norm zu dem gegebenen Punkt minimal wird. Die daraus resultierenden Normwerte werden für jeden Rasterpunkt gespeichert und können graphisch dargestellt werden.

Zusatzfunktionen

Außerdem wurden in MATLAB mehrere Funktionen und Skripten implementiert:

- Funktionen zur graphischen Darstellung der druckabhängigen Besetzungszahlen

- Funktionen zur Darstellung der gemessenen druckkorrigierten chemischen Verschiebungen gegen den Druck und im Vergleich dazu die berechnete optimale Kurve
- Skripten zur Erstellung der Fehlermatrizen und Testdaten
- Skript, das zu einem gegebenen Parametersatz eine bestimmte Anzahl Testdatensätze mit einem wählbaren statistischen Fehler erzeugt zur Ermittlung des Lösungsbereichs
- Variationen der Algorithmen, bei denen ein oder mehrere Parameter festgehalten werden können
- Funktionen zur Untersuchung der Matrix A

3 Ergebnisse

3.1 Strukturvorhersagen von Byr2-RBD mit PERMOL

3.1.1 Rekonstruktion von Byr2-RBD aus PDB-1K8R

Modellierungsgrundlage

Um zu testen, ob die neu implementierte Version von PERMOL einen konsistenten Satz Restraints liefert, wurde die Struktur 1K8R der PDB-Datenbank verwendet. Für die darin enthaltene Ras-bindende Domäne von Byr2 wurde ein Homologiemodell berechnet, wobei zur Erstellung der Restraints ein Alignment mit der eigenen Primärsequenz - der denkbar einfachste Fall - erfolgte. Dadurch kann man überprüfen, ob die Methode geeignet ist, die Sekundär- und Tertiärstruktur adäquat zu rekonstruieren.

Die Modellierungsgrundlage PDB-1K8R, welche mit aus Röntgenkristallographie gewonnenen Daten erstellt wurde, besitzt eine Auflösung von 0,3 nm und enthält neben der hier untersuchten Ras-Bindedomäne GppNHp-gebundenes Ras(1-166) und einige kleinere Liganden. Die Strukturbestimmung war an dem Konstrukt Byr2(71-180) durchgeführt worden, wobei allerdings für die Aminosäuren E127 bis E139 und L166 bis P180 aufgrund schlecht definierter Elektronendichte keine Strukturbestimmung möglich war und sie deshalb in PDB-1K8R fehlen. Der erste dieser Bereiche zeigte jedoch α -helikale Eigenschaften (Scheffzek et al., 2001).

Alignment und Strukturrechnung

Nach Hinzufügen der Wasserstoffatome mit AUREMOL und einer Überprüfung der IUPAC-Nomenklatur wurde die PDB-Datei in PERMOL eingelesen und als Sequenz die vollständige Primärstruktur von Byr2(71-180) verwendet. Für den Bereich, der in der X-Ray-Struktur fehlt, konnten keine Restraints generiert werden. Beim Einlesen der PDB-Struktur erkennt PERMOL die Anzahl der fehlenden Aminosäuren anhand der Nummerierung und ersetzt diese im Alignment durch ein "X" (Abbildung 3-1). Dadurch wird der Algorithmus zur Berechnung des besten Alignments stabiler. In dem hier vorliegenden Fall wird dadurch verhindert, dass die Sequenzstücke ungünstig abgetrennt werden und nicht E126 der Aminosäure E139 zugeordnet wird.

```

spByr2  71  +CILRFIACNGQTRAVQSRGDYQKTLAIALKKFSLEDASKFIVCVSQSSRI
spByr2  71  +CILRFIACNGQTRAVQSRGDYQKTLAIALKKFSLEDASKFIVCVSQSSRI

spByr2 121  KLITEEEFKQICFNSSSPERDRLIIVPKEKPCPSFEDLRRSWEIELAQPA
spByr2 121  KLITEEXXXXXXXXXXXXXRDLIIVPKEKPCPSFEDLRRSWEIE-----

spByr2 171  ALSSQSSLSP+
spByr2 171  -----+

```

Abbildung 3-1: PERMOL - Einlesen der PDB-Struktur

Anhand der Nummerierung der Aminosäuren im PDB-File erkennt der Algorithmus, dass für die Atome einiger Residuen keine Koordinaten vorliegen. Bei der Berechnung des Scores werden diese durch ein "X" (für unbekannte Aminosäure) ersetzt und wirken sich weniger negativ als Lücken bei der Berechnung des Alignments aus. Außerdem wird dadurch verhindert, dass die Sequenzstücke ungünstig unterteilt werden, weil "X" nicht wie eine Lücke innerhalb der Sequenz verschiebbar ist.

Die Parameter zur Erzeugung der Restraints wurden wie in Kapitel 2.3.7 beschrieben ("by standard") gewählt. Damit wurden 7431 Abstandsrestraints, 328 dihedrale Winkelrestraints und 45 Wasserstoffbrückenrestraints erstellt und mit diesen die Strukturrechnungen mit CNS durchgeführt.

Dazu wurden im Protokoll, dass das *Simulated Annealing* steuert, folgende Parameter eingesetzt:

Die Anzahl der Iterationen bei der Hochtemperatursimulation betrug 15000 in Kombination mit einem Zeitschritt von 0,001 units¹, die der beiden Abkühlungsphasen jeweils 3000 bei einem Zeitintervall von 0,005 units, und für die abschließende Energieminimierung wurden 200 Iterationen durchgeführt. Die Wahl der Skalierungsfaktoren erfolgte wie bei der Verwendung experimenteller Restraints.

Es wurden 1000 Strukturen erzeugt, aus denen dann die 10 energetisch günstigsten ausgewählt und einem *Waterrefinement* unterzogen wurden. Vor und nach dieser Verfeinerung setzen sich die Gesamtenergien der energetisch günstigsten Struktur wie in Tabelle 3-1 dargestellt zusammen.

¹ CNS verwendet das AKMA-Einheitensystem, 1 unit entspricht hier also $4,88882 \cdot 10^{-14}$ s.

Tabelle 3-1: Energiebeiträge zur Gesamtenergie für die energetisch günstigste Struktur vor und nach dem *Waterrefinement*

Die beiden Komponenten, die durch die von PERMOL erstellten Restraints zur Kraftfeldfunktion beitragen, E_{noe} (Abstands- und Wasserstoffbrückenrestraints) und E_{cdih} (dihedrale Winkelrestraints), haben nur einen marginalen Anteil an der Gesamtenergie und wurden somit konsistent erstellt. Beim Vergleich ist zu beachten, dass E_{cdih} im Gegensatz zu E_{noe} im letzten Schritt des Simulated Annealings doppelt so stark gewichtet ist wie beim *Waterrefinement*.

Energiebeiträge (kcal/mol)	vor dem <i>Waterrefinement</i>	nach dem <i>Waterrefinement</i>
E_{bond}	0,396	44,391
E_{angle}	0,396	154,487
E_{dihe}	541,332	568,742
E_{impr}	1,148	57,4341
E_{vdW}	1,499	-468,321
E_{elec}	0	-4477,810
E_{noe}	0,0443	0,740
E_{cdih}	0,346	7,302
E_{overall}	574,847	-4113,040

Daraus ist ersichtlich, dass der Beitrag aus den erstellten Restraints marginal ist. Eine Überprüfung aller Zwangsbedingungen ergab, dass alle Abstandsrestraints und dihedralen Winkelrestraints erfüllt sind, wobei für die Abstandsrestraints eine Grenze von $\pm 0,5$ nm und für die Winkelrestraints eine Grenze von $\pm 5^\circ$ angenommen wurde, innerhalb derer die Bedingung als erfüllt angesehen wird. Das zeigt, dass das Restraintsset von PERMOL korrekt erstellt wurde und in sich konsistent ist. Abbildung 3-2 zeigt das berechnete Strukturbündel nach dem *Waterrefinement*.

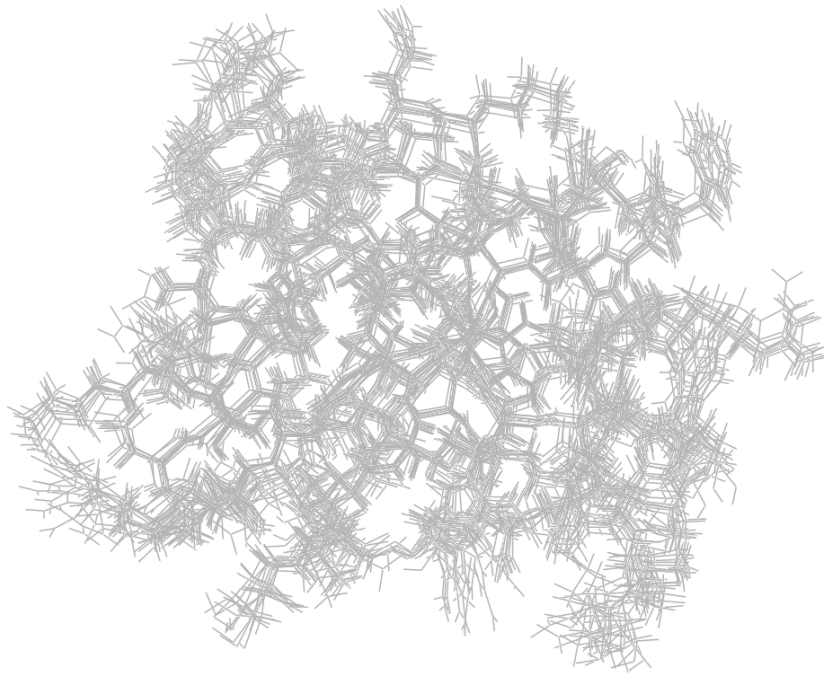


Abbildung 3-2: Strukturbündel nach dem *Waterrefinement*

Berechnetes Strukturbündel von Byr2-RBD nach dem *Waterrefinement* ohne die Aminosäuren E127 bis E139 und L166 bis P180, für die keine Restraints generiert werden konnten, weil sie in der als Modellierungsgrundlage verwendeten Struktur PDB-1K8R fehlen.

Vergleich mit PDB-1K8R

In den Bereichen des Proteins, die in der X-RAY-Struktur in Form von Atomkoordinaten vorhanden waren, sind auch die berechneten Strukturen gut definiert und zeichnen sich durch ein enges Strukturbündel mit einem RMSD der schweren Atome des Proteinrückgrats von $0,052 \text{ nm} \pm 0,007 \text{ nm}$ aus. Eine Analyse der Sekundärstrukturelemente mit dem Programm PyMOL zeigte übereinstimmend für die Struktur 1K8R und die mithilfe von PERMOL generierten Strukturen α -helikale Bereiche für die Aminosäuren Q92 bis F102 und E156 bis P160 und Faltblattstruktur für F110 bis S115, I120 bis L122 und L143 bis P147. Minimale Abweichungen existieren um den Turn zwischen dem ersten und zweiten β -Strang. Diese Sekundärstrukturelemente reichen in der X-RAY-Struktur von Aminosäure I72 bis I76 und von Aminosäure T82 bis Q86. In allen Strukturen des berechneten Bündels jedoch beinhalten sie jeweils noch eine Aminosäure mehr - von I72 bis A77 und von Q81 bis Q86 (Abbildung 3-3). Außerdem fällt eine leichte Verdrehung gegenüber der Röntgenstruktur auf.

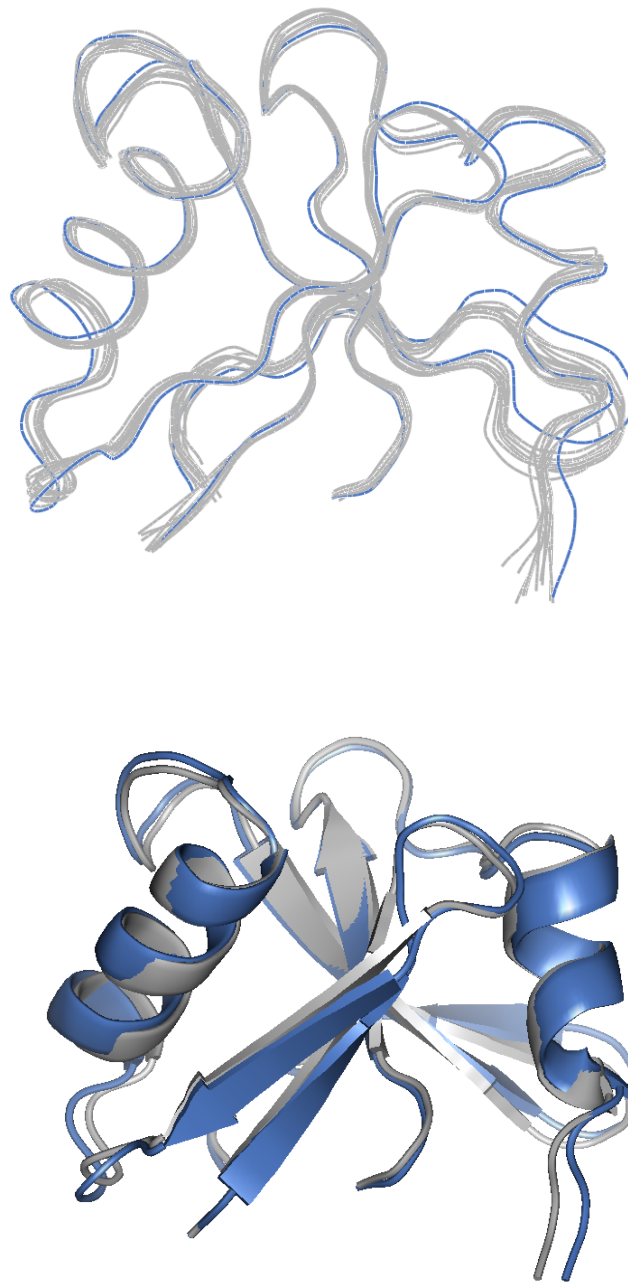


Abbildung 3-3: Vergleich zwischen berechnetem Strukturbündel und Modellierungsgrundlage PDB-1K8R

Berechnetes Strukturbündel (grau) im Vergleich zur Modellierungsgrundlage PDB-1K8R (blau): Das Proteinrückgrat zeigt nahezu denselben Verlauf (oben) und die Sekundärstrukturelemente entsprechen sich bis auf den Turn zwischen dem ersten und zweiten Faltblatt (unten vorne mittig zu erkennen). Geringfügige Abweichungen finden sich im Bereich der Bindestelle zu Ras, das in der Struktur der Modellierungsgrundlage vorhanden war, aber bei der Berechnung des neuen Strukturbündels nicht miteinbezogen wurde.

Weitere geringfügige Abweichungen zwischen dem berechneten Bündel und der Röntgenstruktur findet man im Bereich nach der letzten Helix und einigen wenigen der

angrenzenden Aminosäuren. Diese Bereiche entsprechen genau den Interaktionsstellen mit Ras, das in der verwendeten Modellierungsgrundlage vorhanden war. Da dies bei der Strukturrechnung nicht berücksichtigt wurde, sind kleine Änderungen der Konformation an der eigentlichen Bindestelle ein Zeichen dafür, dass die erstellten Restraints nicht nur die Struktur rekonstruieren können, sondern auch Anpassung an veränderte Bedingungen erlauben.

Der RMSD der schweren Atome des Proteinrückgrats zwischen der Struktur PDB-1K8R und dem Bündel beträgt paarweise berechnet im Mittel $0,059 \text{ nm} \pm 0,006 \text{ nm}$.

3.1.2 Verbesserung der NMR-Struktur von Byr2-RBD

Vorgehen

Für die Simulation der unter der Nummer PDB-1I35 hinterlegten Struktur der Ras-bindenden Domäne von Byr2 wurden 824 Abstandsrestraints, je 44 dihedrale Phi- und Psi-Winkel, 28 dipolare Restkopplungsrestraints und 29 Wasserstoffbrückenrestraints verwendet. Damit entstand ein Strukturbündel, das mit einem RMSD der schweren Atome des Proteinrückgrats mit einem Wert von $0,134 \text{ nm}$ nicht gut definiert ist und dessen Verteilung der Winkelkombinationen im Ramachandran-Plot nur einen Anteil von 67,8% im *most-favoured*-Bereich aufweist (Gronwald et al., 2001). Zur Strukturanalyse wurden jeweils nur die Aminosäuren C71 bis L158 herangezogen, deren Koordinaten auch in der Struktur PDB-1I35 zu finden sind.

Zur Verbesserung dieser Struktur wurden zuerst eine erneute Strukturrechnung und ein *Waterrefinement* mit den experimentellen Restraints durchgeführt, auf die eine Strukturverfeinerung, wie von Cano et al. vorgeschlagen (Cano et al., 2009), folgte.

Strukturrechnung mit experimentellen Restraints

Für die erneute Strukturrechnung wurden dieselben Abstands-, Wasserstoffbrücken- und dipolaren Restkopplungsrestraints verwendet, die auch zur Erstellung des Strukturbündels PDB-1I35 zur Verfügung standen. Lediglich die dihedralen Winkelrestraints wurden mit dem Programm TALOS+ aus den unter der Nummer 4463 in der Biological Magnetic Resonance Bank hinterlegten chemischen Verschiebungen neu generiert.

Eine Analyse der Abstandsrestraints, bei der die experimentell gewonnenen Restraints nach Aminosäuren sortiert und gezählt werden, deren Atome an dem jeweiligen Restraint beteiligt sind, zeigt, dass für die C-terminale Helix kaum Informationen vorliegen (Abbildung 3-4).

Für die Verwendung der dipolaren Restkopplungsrestraints wurden die dafür benötigten Tensorkomponenten D_a und R unabhängig voneinander variiert - D_a in einem Bereich von 8 bis 17 und R von 0,2 bis 0,7 -, um die am besten zu den übrigen Restraints passende Kombination zu ermitteln, die sich durch die geringste Gesamtenergie bei der Strukturrechnung und die kleinsten von dem Programm Pales berechneten Q-Values auszeichnet.

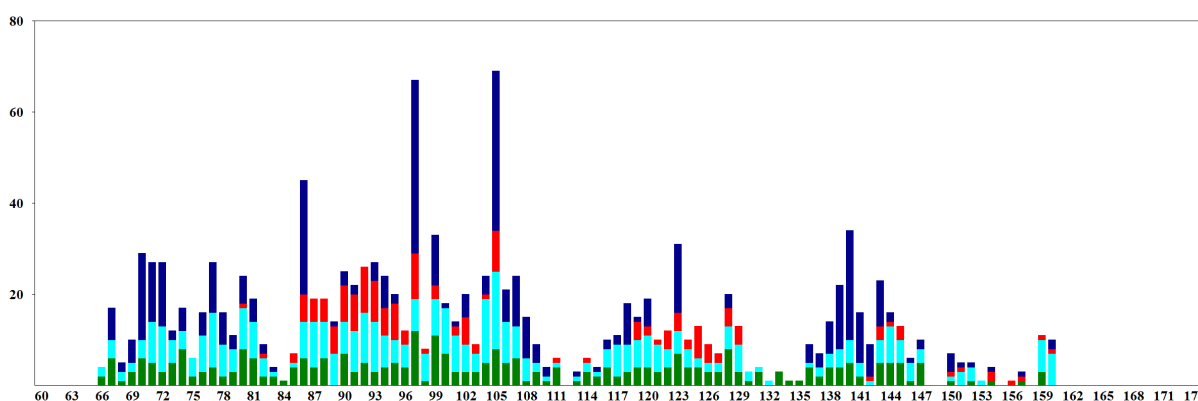


Abbildung 3-4: Analyse der experimentell gewonnenen Abstandsrestraints von Byr2-RBD

Die Anzahl der Abstandsrestraints (nach oben aufgetragen) in Abhängigkeit von den Aminosäuren: Von den 824 Restraints sind 207 langreichweitig (dunkelblau, Abstand der Aminosäuren in der Sequenz größer als 4), 84 mittelreichweitig (rot), 217 sequentiell (cyan) und 316 intraresidual (grün). Aminosäuren, die Sekundärstrukturelementen angehören, zeichnen sich durch eine höhere Anzahl an Restraints aus, wobei für helikale Bereiche mittelreichweitige und für Faltblattstrukturen langreichweitige Restraints typisch sind. Für den C-Terminus sind kaum Restraints vorhanden.

Mit CNS wurden 1000 Strukturen erzeugt, die 10 energieärmsten Strukturen ausgewählt und einem Waterrefinement unterzogen. Dazu wurde das von Jens Linge (Linge et al., 2003) erstellte Skript für CNS zur Durchführung um den Term E_{sani} , der den Energiebeitrag aus den dipolaren Restkopplungen erfasst, erweitert. Diese 10 Strukturen bilden Strukturbündel 1.

Für diese Strukturrechnung wurde eine Sequenz von 116 Aminosäuren, Byr2(65-180), simuliert, da an diesem Konstrukt die NMR-Messungen durchgeführt worden waren.

Strukturrechnung mit den mit PERMOL generierten Restraints

Im nächsten Schritt wurde das so erhaltene Strukturbündel in PERMOL mit sich selbst align und neue Restraints ("by standard", siehe Kapitel 2.3.7) erstellt. Generiert wurden 4646 Abstandsrestraints, 43 Wasserstoffbrückenrestraints und 409 dihedrale Winkelrestraints. Mit diesem Set wurden zwei neue Strukturbündel berechnet, einmal in Kombination (Strukturbündel 3) und einmal ohne (Strukturbündel 2) die experimentell gemessenen dipolaren Restkopplungsrestraints, und anschließend einem *Waterrefinement* unterzogen.

Vergleich der erhaltenen Strukturbündel

Um die drei erhaltenen Strukturbündel untereinander und mit der Struktur 1I35 vergleichen zu können, wurde eine Analyse durchgeführt, bei der nur die Aminosäuren 71-158 miteinbezogen wurden. Zur Beurteilung wurden verschiedene Methoden angewandt - die Erstellung von Ramachandran-Plots, Berechnung der RMSD-Werte unter Einbeziehung aller schweren Atome beziehungsweise nur der schweren Atome des Proteinrückgrats und eine Berechnung des NMR-R-Faktors. Außerdem wurde eine Überprüfung der dipolaren Restkopplungsrestraints mithilfe des Q-Values vorgenommen, den das Programm Pales ausgibt. Die Ergebnisse sind in Tabelle 3-2 dargestellt.

Tabelle 3-2: Vergleich der drei berechneten Strukturbündel bezüglich verschiedener Bewertungskriterien

Methode	Strukturbündel 1 (berechnet aus experimentellen Restraints)	Strukturbündel 2 (berechnet aus PERMOL-Restraints ohne RDCs)	Strukturbündel 3 (berechnet aus PERMOL-Restraints mit RDCs)
Ramachandran (70-158)*			
Residuen in			
<i>most favoured regions</i>	87,15 %	87,65 %	86,14 %
<i>additional allowed regions</i>	8,66 %	10,14 %	10,90 %
<i>generously allowed regions</i>	2,14 %	0,65 %	1,41 %
<i>disallowed regions</i>	1,15 %	1,65 %	1,65 %
RMSD (70-158)*			
schwere Atome des Rückgrats	0,185 nm ± 0,034 nm	0,069 nm ± 0,010 nm	0,073 nm ± 0,007 nm
alle schweren Atome	0,287 nm ± 0,038 nm	0,155 nm ± 0,016 nm	0,166 nm ± 0,014 nm
R-Faktor			
langreichweitige Signale	0,22	0,21	0,21
Q-Value			
	0,0075	0,5423	0,0046

***in Klammern die betrachteten Aminosäuren**

Auffällig ist, dass bereits Strukturbündel 1 eine deutlich bessere Qualität aufweist als die Struktur PDB-1I35, da die Anzahl der Aminosäuren, die im Ramachandran-Plot im *most-favoured*-Bereich liegt, um 20 % größer ist und auch der RMSD deutlich kleiner wurde.

Abbildung 3-5 zeigt das Strukturbündel 3. Die Sekundärstrukturelemente sind, bis auf die dritte Helix, sehr gut definiert. Diese dritte Helix ist in allen drei Bündeln wegen des Mangels an Restraints der inhomogenste Bereich. Bei manchen der Strukturen ist sie bei Aminosäure L158 unterbrochen oder besitzt hier einen Knick.

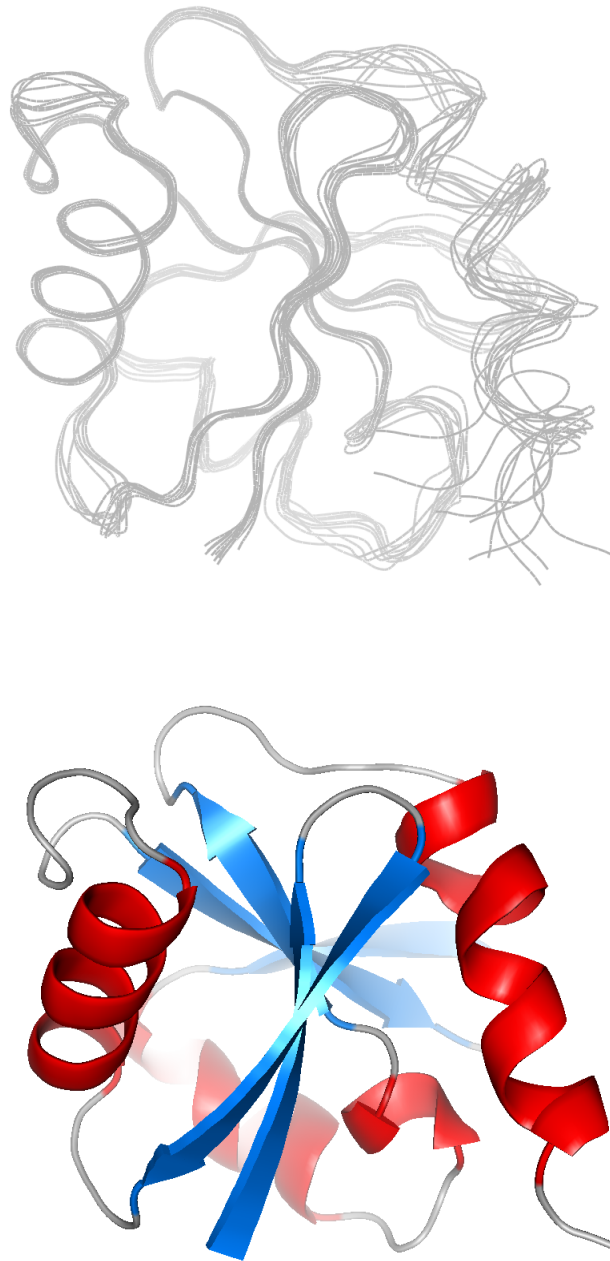


Abbildung 3-5: Strukturbündel 3

Strukturbündel 3, das auf der Grundlage von mit PERMOL erzeugten Restraints in Kombination mit den experimentell bestimmten dipolaren Restkopplungsrestraints generiert wurde. Oben die besten 10 einem *Waterrefinement* unterzogenen Strukturen, die ein enges Bündel bilden, unten eine Darstellung der Sekundärelemente. Neben den drei Haupthelices ist rechts unten die im Vergleich zu 1I35 zusätzlich erzeugte helikale Schleife zu sehen.

Der erste β -Strang reicht von Aminosäure I72 bis A77, der zweite von G80 bis C86, der dritte von F110 bis Q116, der vierte von R119 bis I123 und der fünfte von R142 bis P147. Damit

sind der zweite, der dritte und der fünfte β -Strang jeweils um zwei Aminosäuren länger als bei Struktur PDB-1I35, der vierte um eine nach vorne verschoben. Die α -Helices 1, 2 und 3 reichen von Aminosäure Y91 bis F102, E125 bis N134 und von S154 bis A167, womit die zweite Helix sogar um 3 Aminosäuren länger ist als in PDB-1I35 (E127 bis F133). Die dritte Helix reicht in der Vergleichsstruktur sogar nur bis Aminosäure L158. Zusätzlich zu den bereits von Gronwald et al. ermittelten Sekundärstrukturelementen (Gronwald et al., 2001) wurde in allen neu berechneten Konfigurationen eine helikale Schleife aus vier Aminosäuren vor dem fünften β -Strang - Residuen P138 bis D141 - gefunden.

3.1.3 Modellierung des Komplexes Byr2-RBD mit Ras aus homologen Strukturen

Auswahl der Modellierungsgrundlagen

Zur Überprüfung der Funktion und zur Untersuchung der Möglichkeiten, die die neu implementierte Version von PERMOL bietet, wurden verschiedene Modellrechnungen des Komplexes Byr2-RBD mit Ras aus homologen Strukturen durchgeführt und mit der Röntgenstruktur dieses Komplexes, PDB-1K8R, verglichen.

Als Modellierungsgrundlagen eignen sich in diesem Fall Strukturen von Komplexen, bei denen einer der beiden Bindungspartner gleich bleibt. Für das Protein Ras sind verschiedene Bindungspartner bekannt und eingehend analysiert worden. Fast sämtliche der von Gronwald et al. und Wohlgemuth et al. untersuchten Bindungspartner von Ras wie C-Raf oder RalGEF stammen aber von Säugetieren (Gronwald et al., 2001; Wohlgemuth et al., 2005). Wie in Kapitel 1.2.1 beschrieben, beträgt die Sequenzähnlichkeit zu dem von einem Hefepilz stammenden Byr2 zu den Vergleichsproteinen nur zwischen 14% und 16%. Trotz dieser geringen Sequenzähnlichkeit liegt jedoch große strukturelle Ähnlichkeit vor. Als Modellierungsgrundlage für den Komplex der Ras-bindenden Domäne von Byr2 mit Ras wurden die Kristallstrukturen PDB-1LFD (Huang et al., 1998) - Ras im Komplex mit RalGDS - und PDB-4G0N (Fetics et al., 2015) - Ras im Komplex mit Raf - ausgewählt.

Alignment unter Einbeziehung der Sekundärstruktur

Liegt - wie hier - nur eine so geringe Sequenzähnlichkeit vor, ist es nicht sinnvoll, ein globales Alignment durchzuführen, das allein auf der Suche nach dem besten Score beruht (Merkl und Waack, 2009). Wenn die Sekundärstrukturelemente der Proteine bekannt sind, dürfen diese bei der Zuordnung nicht außer Acht gelassen werden. Von Wohlgemuth et al.

wurde ein Sequenzalignment zwischen den Ras-bindenden Domänen von Byr2 und RalGDS beziehungsweise C-Raf vorgeschlagen, bei dessen Erstellung die Faltung dieser Proteine miteinbezogen worden war (Wohlgemuth et al., 2005; Abbildung 3-6). Dieses wurde für die Informationsübertragung zur Erstellung der Modelle verwendet. Dazu konnte die Möglichkeit des Editierens eines von PERMOL vorgeschlagenen Alignments eingesetzt und getestet werden.

Eine Analyse der beiden ausgewählten Strukturen zeigt, dass für die dritte Helix von Byr2-RBD keine Informationen extrahiert werden können, da diese weder in PDB-1LFD noch in PDB-4G0N enthalten ist. Deshalb wurden von Byr2 jeweils nur die Aminosäuren C71 bis E146 simuliert, von Ras die Aminosäuren M201 bis H366. Die in den Modellen enthaltenen und an Ras gebundenen Ionen und das Nukleotid wurden bei den Rechnungen nicht miteinbezogen. Stattdessen wurden die entsprechenden angrenzenden Bereiche durch den Einsatz einiger weniger weiterer Abstandsrestraints stabilisiert.

```

spByr2  71  CILRFIAC---NGQTRAVQSRG--DYQKTLAIALKKFSLE--DASKFIVC
cRaf    56  NTIRVFLP---NKQRTVVNVN-GMSLHDCLMKALKVRGLQ--PECCAVF
mRalGDS 16  CIIRVSLDVDNGNMYKSILVTSQDKAPTIVIRKAMDKHNLDEDEPEDYELL

spByr2 121  VSQS--SRIKLITEEEFKQICFNSSSPERDRLIIVPKEKPCPSFEDL
cRaf    106  RLLHEHKGK KARLDWN--TDAASLI----GEELQVDFL
mRalGDS 66  QIIS--EDHKLKIPENANVFYAMNSAAN--YDFILKKR

```

Abbildung 3-6: Sequenzalignment zwischen den drei Ras-bindenden Proteinen Byr2, C-Raf und mRalGDS

Sequenzalignment zwischen den Ras-bindenden Proteinen Byr2, C-Raf und mRalGDS, das auf einem Vergleich der Sekundärstrukturelemente beruht: Die wenigen identischen Aminosäuren wurden rot eingefärbt.

(nach Wohlgemuth, Kiel et al. 2005)

PDB-1LFD als Modellierungsgrundlage

Mit PERMOL wurden 22427 Abstandsrestraints, 877 dihedrale Winkelrestraints und 156 Wasserstoffbrückenrestraints ("by standard") erstellt. Dabei wurden sowohl intramolekulare als auch intermolekulare Informationen über Abstände und Wasserstoffbrücken aus der

Modellierungsvorlage extrahiert und an die neue Sequenz angepasst. Die intermolekularen Restraints verknüpfen die beiden Bindungspartner miteinander.

Damit wurden 1000 Strukturen simuliert, die besten 10 ausgewählt und einem *Waterrefinement* unterzogen. Die Strukturrechnung erfolgte mit folgenden Gewichtungen der Restraints in den einzelnen Phasen des *Simulated Annealings*:

- Hochtemperatursimulation: 26 (Skalierungsfaktor für den Abstandsrestraint-Energieterm) bzw. 36 (Skalierungsfaktor für den dihedralen Winkelrestraint-Energieterm)
- 1. Abkühlungsphase: 26 bzw. 71
- 2. Abkühlungsphase: 26 bzw. 71
- Abschließende Minimierung: 13 bzw. 143

Während des *Waterrefinements* betrug der Gewichtungsfaktor für die Abstandsrestraints 13, der für die dihedralen Winkelrestraints 71. Um die Relation zwischen Wasserstoffbrückenrestraints und den Abstandsrestraints wiederherzustellen, wurden erstgenannte 4-fach eingelesen und ihnen damit ein höheres Gewicht gegeben. Die Anzahl der Iterationen bei der Hochtemperatursimulation betrug 15000 in Kombination mit einem Zeitschritt von 0,001 units, die der beiden Abkühlungsphasen jeweils 3000 bei einem Zeitintervall von 0,005 units, und für die abschließende Energieminimierung wurden 200 Iterationen getätigt. Dieselben Parameter wurden auch für die noch folgenden Strukturrechnungen mit PDB-4G0N beziehungsweise mit PDB-1LFD und PDB-4G0N als Modellierungsgrundlage verwendet.

Das erhaltene Strukturbündel besitzt einen RMSD der schweren Atome des Proteinrückgrats von $0,029 \text{ nm} \pm 0,004 \text{ nm}$ (Abbildung 3-7). Die Gesamtenergie der besten Struktur betrug vor dem *Waterrefinement* 3110,7 kcal/mol, nach der Verfeinerung -10086,5 kcal/mol. 38 der erzeugten Abstands- und 22 der Winkelrestraints, die von PERMOL vorgeschlagen wurden, konnten nicht erfüllt werden. Bei ihnen wichen die in den Strukturen gemessenen Abstände beziehungsweise Winkel um mehr als 0,5 nm beziehungsweise 5° von den in den Restraints definierten Werten ab.

Für die Simulationen wurden Chain A (RalGDS) und Chain B (Ras) von PDB-1LFD verwendet. Ein Test, bei dem Chain C und Chain D als Modellierungsgrundlage dienten, ergab deutlich schlechtere Strukturen.

Für einen Vergleich mit der experimentell bestimmten Struktur dieses Komplexes PDB-1K8R, konnten nur die Aminosäuren I71 bis R126 und F140 bis E146 betrachtet werden, da die dazwischenliegenden Aminosäuren in der Röntgenstruktur nicht vorhanden waren. Abbildung 3-8 zeigt Byr2-RBD aus PDB-1K8R und aus einer der Strukturen des Bündels, die ideal aufeinander gelegt wurden.

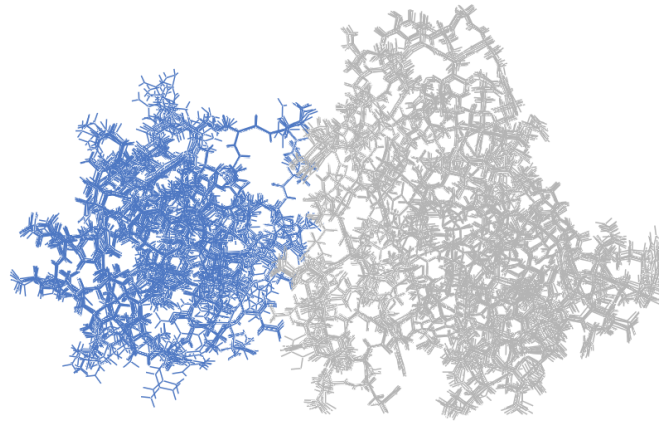


Abbildung 3-7: Strukturbündel von Byr2-RBD im Komplex mit Ras

Das erzeugte Strukturbündel von Byr2-RBD (blau) im Komplex mit Ras (grau), wobei die Struktur PDB-1LFD als Modellierungsgrundlage diente.

Die Sekundärstrukturelemente von Byr2-RBD der berechneten Strukturen stimmen von der Position innerhalb der Sequenz und auch vom Neigungswinkel zueinander gut mit der Röntgenstruktur überein, sind jedoch tendenziell etwas kürzer.

Die Struktur PDB-1LFD enthält nicht den Wildtyp von Ras, sondern die Mutante RasE31K. Das Protein Ras musste also bei der Simulation in fast identischer Weise wie in der Modellstruktur rekonstruiert werden. Es stimmt in Aufbau und Geometrie mit PDB-1K8R gut überein. Nur im Bereich der Bindungsstelle von GppNHp zeigen sich leichte Abweichungen des Proteinrückgrats, da dieses Molekül bei der Erstellung des Homologiemodells vernachlässigt wurde.

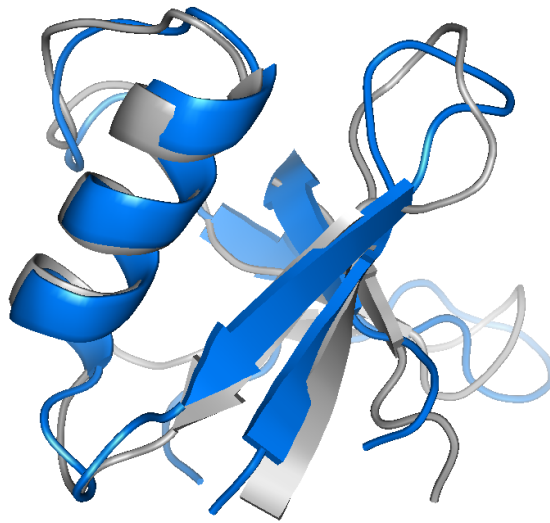


Abbildung 3-8: Byr2-RBD aus 1K8R im Vergleich zu Byr2-RBD und einer Struktur des Bündels

Byr2-RBD aus 1K8R (blau) im Vergleich zu Byr2-RBD und einer Struktur des Bündels: Gut erkennbar ist die Übereinstimmung der Sekundärstrukturelemente in Lage innerhalb der Sequenz und Orientierung zueinander.

Betrachtet man jedoch den Komplex als Ganzes, stellt man fest, dass die beiden Strukturen von Byr2-RBD, die neu berechnete und die der Röntgenstruktur, bei festgehaltenem Ras gegeneinander um etwa 30° verdreht (Abbildung 3-9) und damit nicht optimal modelliert sind. Eine Erklärung dafür liefert ein Vergleich der Sequenzen von Byr2-RBD und dem Homolog RalGDS, dessen Struktur als Modellierungsgrundlage diente. Dieses besitzt zwischen den beiden ersten Helices, einem elementaren Bereich der Bindestelle, drei Aminosäuren mehr (Abbildung 3-6).

Der mittlere paarweise berechnete RMSD der schweren Atome des Proteinrückgrats für den gesamten Komplex zwischen den generierten Konfigurationen und der Vergleichsstruktur betrug $0,061 \text{ nm} \pm 0,003 \text{ nm}$.

Bei der Analyse von PDB-1K8R unter den in Kapitel 2.2.1 dargelegten Kriterien - Abstand und Zugänglichkeit des Lösungsmittels - wurden die Aminosäuren der dritten Helix bei Byr2 außer Acht gelassen, da sie in den gerechneten Modellen fehlen. Dabei wurden die Aminosäuren I72, R74, N79 bis A84, Q86, K101 und S103 als Bindestelle aufseiten von Byr2 identifiziert. Diese stimmt mit denen der erzeugten Strukturen komplett überein. Zusätzlich jedoch erfüllt Aminosäure K100 die genannten Bedingungen.

Für Ras können bei PDB-1K8R die Residuen Q225, D233, I236 bis R241 und Y264 dem Interaktionsbereich zugerechnet werden, die in den Modellen ebenfalls der Bindestelle angehören, wobei auch hier weitere Aminosäuren, I221, I224, E231 und M267, die angewandten Kriterien erfüllen.

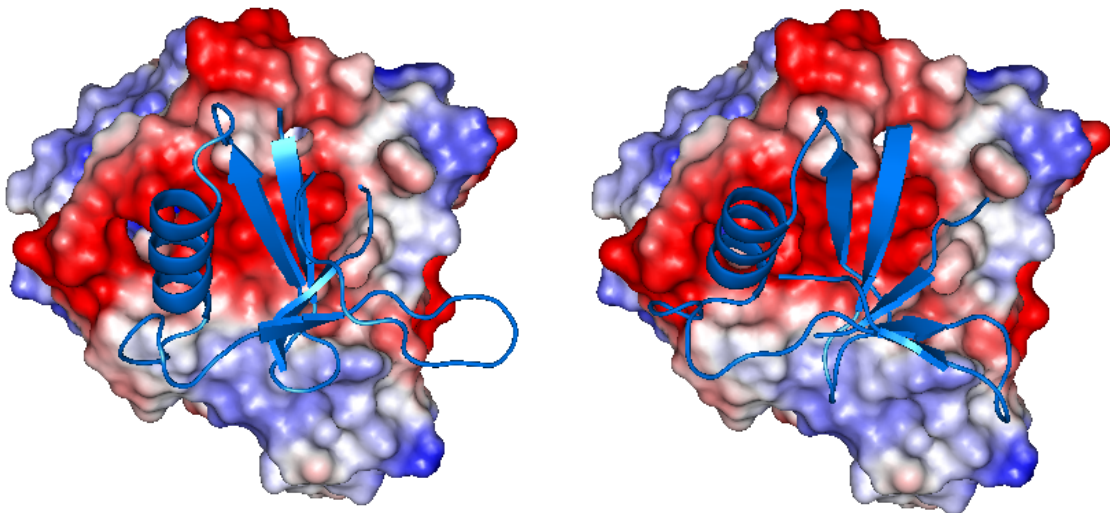


Abbildung 3-9: Gegenüberstellung der Struktur 1K8R (links) und einer der berechneten Modellstrukturen, welche mit 1LFD (RalGDS im Komplex mit Ras) als Modellierungsgrundlage entstanden ist (rechts)

Zu erkennen ist die leicht andere Orientierung der beiden Komplexpartner zueinander. Die Bindestelle selbst jedoch stimmt sowohl aufseiten von Ras als auch aufseiten von Byr2 gut überein, und auch die den Komplex stabilisierenden Wasserstoffbrücken werden - bis auf eine im Bereich des Turns zwischen dem ersten und zweiten β -Faltblatt von Byr2 liegende - durch die Simulation richtig wiedergegeben.

Der Zusammenhalt selbst, der vor allem durch Komplementarität der Ladungen geschaffen wird, ist in PDB-1K8R durch intermolekulare Wasserstoffbrücken zwischen dem zweiten β -Strang von Byr2 und dem zweiten β -Strang von Ras der betrachteten Domänen stabilisiert. Zusätzlich findet man Wasserstoffbrücken zwischen den Aminosäuren R74 beziehungsweise K101 bei Byr2 und E237 beziehungsweise D238 bei Ras.

Diese Bindungen sind bis auf die Wasserstoffbrücke zu Q81 von Byr2, die in keinem der Strukturen des Bündels gefunden werden kann, in den berechneten Homologiemodellen vorhanden.

PDB-4G0N als Modellierungsgrundlage

Analog zu PDB-1LFD erfolgte eine Strukturrechnung mit PDB-4G0N als Modellierungsgrundlage. Diese Struktur enthält nicht RalGDS sondern C-Raf im Komplex mit Ras. Mit PERMOL wurden 19856 Abstands-, 704 dihedrale Winkel- und 137 Wasserstoffbrückenrestraints erzeugt und damit Strukturrechnungen mit anschließendem *Waterrefinement* vorgenommen. Obwohl die allgemeine Geometrie der Tertiärstruktur der erhaltenen Konfigurationen die typische Form aufweist, wurden die Sekundärstrukturelemente unbefriedigend modelliert (Tabelle 3-3). Im Vergleich zu PDB-1K8R sind sie entweder innerhalb der Sequenz verschoben ($\beta 1$, $\beta 2$, $\alpha 1$ und $\beta 5$) oder werden aufgrund der schlechten Geometrie von den darstellenden Programmen gar nicht als solche erkannt ($\beta 3$, $\beta 4$).

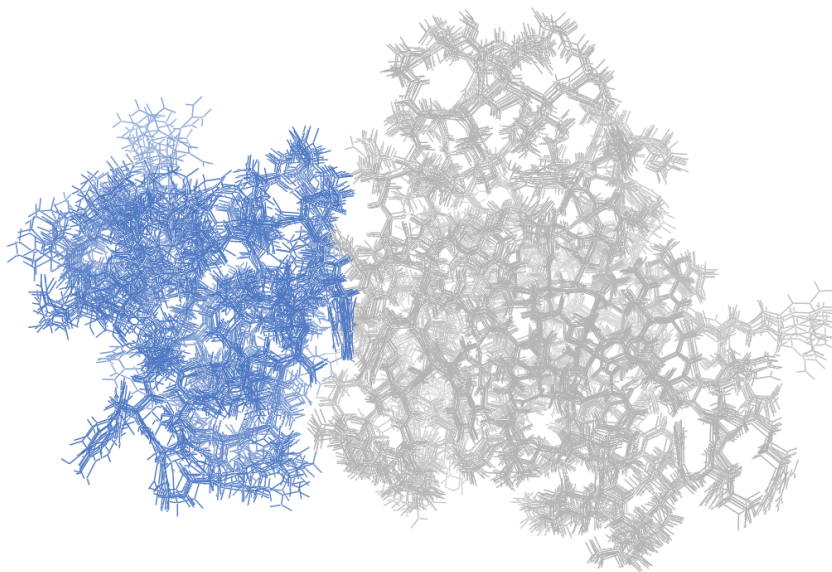


Abbildung 3-10: Strukturbündel von Byr2-RBD im Komplex mit Ras

Das erzeugte Strukturbündel von Byr2-RBD (blau) im Komplex mit Ras (grau), wobei die Struktur PDB-4G0N als Modellierungsgrundlage diente.

Das erhaltene Strukturbündel besitzt einen RMSD der schweren Atome des Proteinrückgrats von $0,093 \text{ nm} \pm 0,024 \text{ nm}$ (Abbildung 3-10). Die Gesamtenergie der besten Struktur betrug jedoch vor dem Waterrefinement $1596,8 \text{ kcal/mol}$, nach der Verfeinerung $-10882,8 \text{ kcal/mol}$ und ist damit geringer als die des Strukturbündels mit PDB-1LFD als

Modellierungsgrundlage. Von allen erzeugten Restraints konnten nur 6 Abstandsrestraints und 14 Winkelrestraints nicht erfüllt werden.

Eine Analyse der Bindestelle ergab, dass zwar der Bereich, den Byr2 auf der Oberfläche von PDB-1K8R abdeckt, auch bei diesem Strukturbündel als Bindestelle fungiert, aber zusätzlich die Aminosäuren der ersten Helix und die der beiden Loops davor und danach von Byr2-RBD ebenfalls engen Kontakt zu Ras haben.

Eine Überprüfung der Wasserstoffbrücken zeigte, dass sowohl diejenigen, die die beiden Faltblätter von Byr2 und Ras verknüpfen, als auch die beiden zusätzlichen zu den Aminosäuren 37 und 38 von Ras in den Strukturen vorhanden sind.

Obwohl dieses Bündel auf den ersten Blick aufgrund des größeren RMSDs und der schlechter definierten Sekundärstrukturelemente weniger gut gelungen scheint im Vergleich zu dem mit PDB-1LFD als Modellierungsgrundlage erzeugten, zeigen eine Überprüfung der Energien, der die Komplexpartner verknüpfenden Wasserstoffbrücken und auch die Tatsache, dass bei PDB-4G0N zum Informationsgewinn für die Strukturrechnung deutlich weniger der erzeugten Restraints nicht erfüllt werden können, dass die Geometrie der erhaltenen Strukturen besser an die Sequenz angepasst ist als bei PDB-1LFD als Grundlage.

PDB-1LFD und PDB-4G0N als Modellierungsgrundlage

Als dritte Möglichkeit wurden PDB-1LFD und PDB-4G0N als Modellierungsgrundlagen automatisch, wie in Kapitel 2.3.7 beschrieben, anhand eines lokalen Scores kombiniert und 22021 Abstandsrestraints, 867 dihedrale Winkelrestraints und 193 Wasserstoffbrückenrestraints erzeugt. Die Strukturrechnung und das anschließende Waterrefinement ergaben ein Strukturbündel, dessen RMSD der schweren Atome des Proteinerückgrats mit $0,069 \text{ nm} \pm 0,03 \text{ nm}$ in der Größenordnung des mit PDB-1LFD als Modellierungsgrundlage berechneten Strukturbündels liegt und dessen Sekundärstrukturelemente ebenfalls eine diesem Bündel vergleichbare Güte aufweisen (Tabelle 3-3).

Tabelle 3-3: Sekundärstrukturelemente der berechneten Bündel im Vergleich zur Röntgenstruktur PDB-1K8R

Die Zahlen geben die Nummern der entsprechenden Aminosäuren an. Während bei PDB-4G0N als Modellierungsgrundlage die Faltblätter und die Helix entweder auf der Sequenz verschoben oder überhaupt nicht geometrisch befriedigend rekonstruiert werden konnten, stimmen sie bei PDB-1LFD als Modell oder bei einer Kombination aus 4G0N und 1LFD in etwa mit PDB-1K8R überein. Die beiden weiteren Helices $\alpha 2$ und $\alpha 3$ können nicht verglichen werden, da $\alpha 2$ in 1K8R fehlt und $\alpha 3$ in den verwendeten Modellierungsgrundlagen nicht mehr vorhanden ist.

Sekundär- struktur- element	Vergleichs- struktur PDB-1K8R	Bündel mit Modellierungsgrundlage		
		PDB-1LFD	PDB-4G0N	1LFD und 4G0N
$\beta 1$	72 - 76	71 - 76	74 - 78	71 - 76
$\beta 2$	82 - 86	81 - 86	80 - 83	83 - 87
$\alpha 1$	90 - 102	91 - 102	92 - 103	92 - 101
$\beta 3$	110 - 115	113 - 115	-	113 - 115
$\beta 4$	120 - 122	121 - 123	-	121 - 123
$\beta 5$	141 - 144	143 - 145	143 - 146	142 - 145

Von den berechneten Restraints werden 58 Abstandsrestraints nicht erfüllt und zeigen eine Abweichung von größer 0,5 nm, bei den dihedralen Winkelrestraints sind es 39, die mehr als 5° von dem vorgeschlagenen Winkel abweichen. Die Gesamtenergie vor und nach dem *Waterrefinement* (Tabelle 3-4) lag jeweils deutlich über der der beiden anderen Modelle, woran auch eine Bereinigung um die Anteile E_{noe} und E_{cdih} der generierten Restraints nichts änderte.

Bei der Berechnung aller drei Modelle wurden zusätzlich Versuche mit anderen Gewichtungen der von PERMOL generierten Restraints in Bezug auf die über die physikalischen Potentiale definierten Zwangsbedingungen durch Variation der Skalierungsfaktoren unternommen. Dabei stellte sich heraus, dass bei höherer Gewichtung zwar erreicht werden kann, dass Strukturen gefunden werden können, bei denen wesentlich weniger Restraints nicht erfüllt werden. Allerdings fällt bei diesen Konfigurationen auf, dass die Energiebeiträge aus den physikalischen Potentialen, zum Beispiel aus den Bindungslängen oder dem elektrostatischen Anteil, deutlich höher liegen als bei der schlussendlich gewählten Gewichtung. Außerdem zeigte sich, dass bei Erhöhung der Skalierungsfaktoren der erzeugten Restraints ein erheblich größerer Anteil der insgesamt erzeugten Strukturen keine adäquate Faltung aufweist.

Tabelle 3-4: Energiebeiträge nach dem *Waterrefinement*

Verglichen werden die Energiebeiträge der drei berechneten Strukturbündel, für die PDB-1LFD, PDB-4G0N und eine Kombination aus 1LFD und 4G0N als Modellierungsgrundlage dienten.

Energiebeiträge (kcal/mol)	PDB-1LFD als Modell	PDB-4G0N als Modell	1LFD und 4G0N als Modelle
E_{bond}	163,58	136,81	172,89
E_{angle}	507,83	445,79	536,12
E_{impr}	202,08	172,58	221,19
E_{vdW}	-1376,86	-1355,85	-1315,67
E_{elec}	-11432,30	-11802,40	-11597,40
E_{dihe}	1343,88	1345,70	1448,02
E_{noe}	446,22	132,14	731,27
E_{cdih}	50,11	42,37	77,48
E_{overall}	-10086,50	-10882,80	-9726,09
$E_{\text{overall-korr}}$	-10582,83	-11057,31	-10534,84

Eine Analyse der die Interaktionsflächen definierenden Aminosäuren ergab, dass diese aufseiten von Byr2 vollkommen mit den Residuen von PDB-1K8R übereinstimmen, aufseiten von Ras aber zusätzlich die Aminosäuren S217 und M267 die beiden Kriterien Abstand und Zugänglichkeit durch ein Lösungsmittel erfüllen. Eine genauere Untersuchung ergab jedoch, dass nur geringfügige Unterschiede in der Konfiguration an diesen Stellen vorliegen. Im Gegensatz zu dem Strukturbündel, in dem nur PDB-1LFD als Modellierungsgrundlage verwendet wurde, ist die Orientierung von Byr2-RBD auf der Oberfläche von Ras richtig gedreht (Abbildung 3-11). Wie ein Vergleich der Aminosäuren, die aufseiten von Byr2 der Bindestelle angehören, zeigt, nehmen sie auf der Oberfläche von Ras die gleichen Stellen ein wie in der experimentell gefundenen Vergleichsstruktur PDB-1K8R. Allerdings ist bei manchen der Residuen erkennbar, dass die Stellung der Seitenketten eine etwas andere ist (Abbildung 3-12).

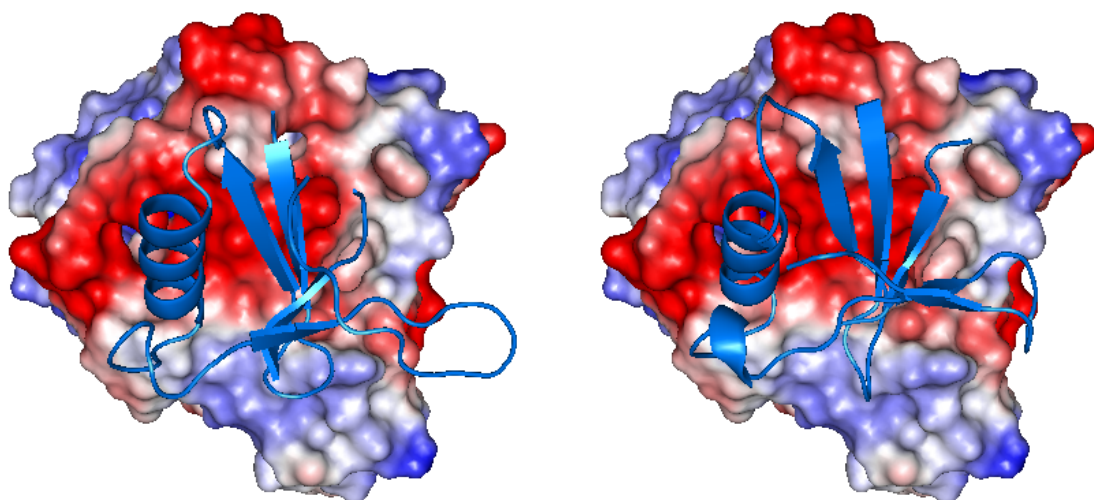


Abbildung 3-11: Gegenüberstellung der Struktur PDB-1K8R (links) und eines der berechneten Modelle (rechts) auf Grundlage des kombinierten Datensatzes

Zur Modellierung wurden Informationen aus den beiden Strukturen PDB-1LFD und PDB-4G0N kombiniert. Im Gegensatz zu einer Modellierung nur mit 1LFD als Modellierungsgrundlage sind die beiden Komplexpartner genauso zueinander orientiert wie in der Röntgenstruktur PDB-1K8R des Komplexes.

Die Wasserstoffbrücken zwischen den beiden Komplexpartnern werden wie in den beiden anderen Homologiemodellen durch die Simulation gut wiedergegeben. Sie stabilisieren das intermolekulare Faltblatt und verknüpfen zusätzlich die Aminosäuren R74 beziehungsweise K101 von Byr2 mit den Aminosäuren E237 und D238 von Ras genauso wie in der Röntgenstruktur 1K8R.

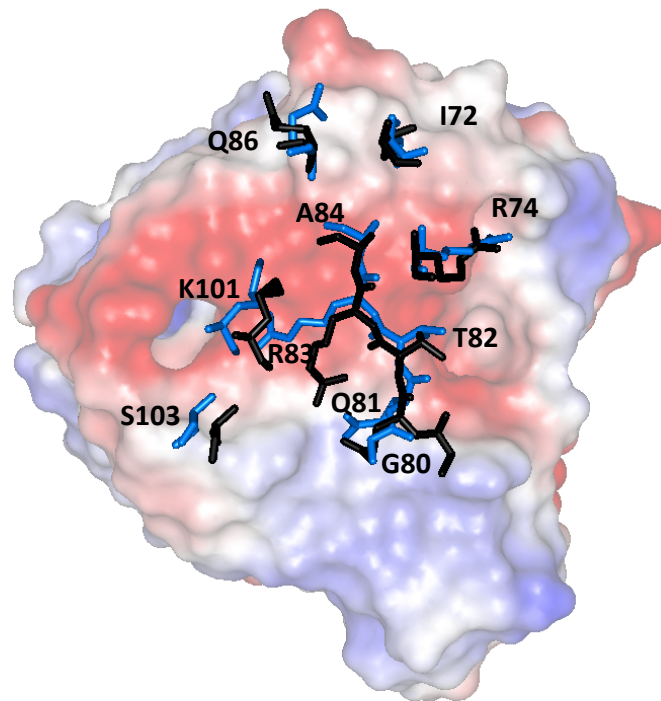


Abbildung 3-12: Oberfläche des Proteins Ras mit Darstellung der Aminosäuren von Byr2 der Bindestellen

Im Vergleich in Schwarz die Konfigurationen wie in PDB-1K8R, in Blau eines der berechneten Modelle. Die Lage auf der Oberfläche stimmt beinahe perfekt überein, jedoch sind manche Seitenketten anders orientiert.

3.2 Erstellung eines Homologiemodells des Proteins IE1(476-491) im Komplex mit Chromatin

3.2.1 Erstellung von Alignments zu LANA(4-17)

Modellstruktur

Als Grundlage für die Modellierung diente die unter der Identifikationsnummer 1ZLA in der PDB-Datenbank hinterlegte Struktur des N-Terminus des Proteins LANA des Kaposi-Sarkom assoziierten Herpesvirus (HHV8) in Komplex mit einem Nukleosom, um das DNA geschlungen ist. Sie wurde mithilfe von Röntgenspektroskopie erstellt, und die Auflösung beträgt 2,9 Ångström. Die Oberfläche des Nukleosoms enthält eine sogenannte saure Tasche, gebildet durch die Histone H2A und H2B, eine Rille mit mehreren negativ geladenen Aminosäuren, in die das LANA-Peptid - untersucht wurden die Aminosäuren 1 bis 23 - bindet und eine Haarnadelstruktur formt. Da LANA auch an die Virus-DNA andocken kann, könnte die gleichzeitige Bindung an das menschliche Chromatin der Schlüssel für die Latenz dieser Viren sein (Barbera et al., 2006; Feeney und Parish, 2009).

Durch Mutageneseexperimente konnte gezeigt werden, dass der C-Terminus von IE1 des Cytomegalievirus, das ebenfalls den Herpesviren angehört, auch in diese saure Tasche bindet und die Bindung aufseiten des Nukleosoms durch nahezu identische Aminosäuren geschieht wie bei LANA. Daher wurde die Vermutung angestellt, dass die Bindung in gleicher oder sehr ähnlicher Weise geschieht, obwohl nur eine geringe Sequenzähnlichkeit zwischen dem C-Terminus von IE1 und dem N-Terminus von LANA besteht. Zur Erstellung des Modells wurden deswegen nicht nur das Peptid selbst, sondern auch die Bereiche von H2A und H2B, die die Bindestelle bilden, modelliert. Die Ergebnisse dieser Strukturvorhersagen wurden von Mücke et al. veröffentlicht und mit in einer kurzen Filmsequenz veranschaulicht, die ebenfalls im Rahmen dieser Arbeit entstand (Mücke et al., 2014).

Wie in Kapitel 1.2.3 beschrieben, wird die saure Tasche auf der Oberfläche des Chromatins unter anderem auch von den Proteinen RCC1 und SIR3 genutzt, von denen ebenfalls Komplexstrukturen in der PDB-Datenbank abgelegt sind. Diese Proteine wurden als mögliche Modellierungsgrundlagen verworfen, da bei SIR3 eine wesentlich größere Anzahl an Aminosäuren in die Interaktion einbezogen ist und die bindenden Reste von RCC1 eine helikale Struktur bilden, was für den C-Terminus von IE1 aufgrund des mittig gelegenen Prolins eher ungünstig scheint. Ein weiterer Grund, LANA zu bevorzugen, ist die Tatsache, dass es wie IE1 das Protein eines Herpesvirus ist und eine ähnliche Funktion vermutet wird.

Alignment

Da wie im vorliegenden Fall die Sequenzähnlichkeit sehr gering und deswegen eine Zuordnung nicht eindeutig ist, wurden zur Erstellung des Alignments verschiedene Programme getestet und deren Ergebnisse verglichen. Zwei der vorgeschlagenen Möglichkeiten wurden ausgewählt: zum einen das Alignment, das der in PERMOL integrierte Algorithmus, aber auch einige andere Programme wie ClustalW2, als Lösung ausgeben (Alignment 1), zum zweiten eine Lösung des Programms LALIGN (Alignment 2). Weitere Alignments wurden verworfen, da das Bindungsmotiv von LANA (GMRLRSG) nicht auf dem interessierenden Bereich von IE1 zu liegen kam.

Beiden Alignments ist gemein, dass drei Aminosäuren identisch sind. In der Sequenz von IE1 findet man die Aufeinanderfolge RS (R486 bis S487) einmal, in LANA ist sie zweimal vorhanden, so dass dieses Sequenzstück bei Alignment 1 mit den Aminosäuren R9 bis R10, bei Alignment 2 mit R12 bis R13 alignt wird (Abbildung 3-13).

```
LANA    4    PGMRLR--SGRSTGAP    17
IE1     476  GGKSTHPMVTRSKADQ    491
              *      :      **..

LANA    4    -----PGMRLRSGRSTGAP    17
IE1     476  GGKSTHPMVTRSKADQ---    491
              *    **    .
```

Abbildung 3-13: Alignments zwischen dem N-Terminus des Proteins LANA (HHV 8) und dem C-Terminus des Proteins IE1 (HHV 5)

Zwei Möglichkeiten eines Alignments zwischen dem N-Terminus des Proteins LANA (HHV 8) und dem C-Terminus des Proteins IE1 (HHV 5), die beide in die gleiche saure Tasche, die von den Histonen H2A und H2B gebildet wird, binden.

Oben: Alignment 1

Unten: Alignment 2

(publiziert in Mücke et al., 2014)

3.2.2 Modellrechnungen mithilfe von PERMOL und CNS

Die Strukturrechnung erfolgte mit dem Programm CNS. Simuliert wurden die Aminosäuren G476 bis Q491 von IE1, A14 bis V107 des Histons H2A und R30 bis K122 des Histons H2B. Für diese drei Proteinstücke wurden drei einzelne Strukturfiles erstellt. Als zusätzliche Bedingungen für die Modellierung wurden für beide Alignments Abstandsrestraints (18336 für Alignment 1, 13085 für Alignment 2), dihedrale Winkelrestraints (je 782 für beide Alignments) und Wasserstoffbrückenrestraints (151 für Alignment 1, 152 für Alignment 2) mit PERMOL erstellt. Die dafür gewählten Parameter entsprechen den in Kapitel 2.3.7 beschriebenen ("by standard"). Auf Grundlage dieser beiden Sets von Restraints wurden je 1000 Strukturen für jedes Alignment berechnet. Jeweils die 10 Strukturen, deren Gesamtenergien am geringsten waren, wurden ausgewählt und einem *Waterrefinement* unterzogen. Die erhaltenen Konformationen wurden zu den Strukturbündeln Modell 1 und Modell 2 zusammengefasst (Abbildung 3-14).

Die beiden berechneten Modelle bilden jeweils wie die Modellierungsgrundlage ein β -Haarnadelmotiv - zwei antiparallele β -Stränge, verbunden durch einen Turn.

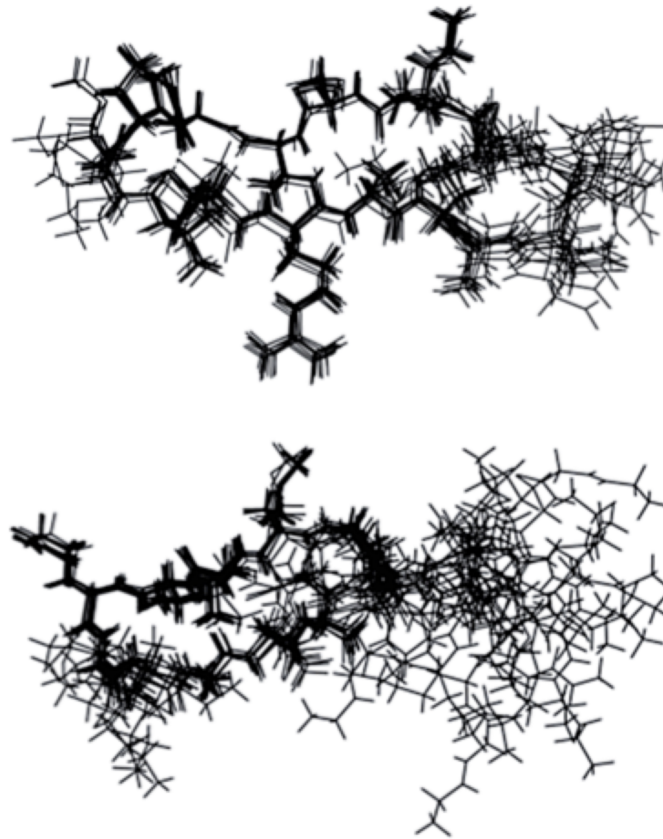


Abbildung 3-14: Strukturbündel Modell 1 und Modell 2

Darstellung der beiden Strukturbündel (jeweils nur der C-Terminus von IE1), wobei links der Turn des Motivs zu erkennen ist.

Oben: Modell 1

Unten: Modell 2

(publiziert in Mücke et al., 2014)

3.2.3 Beurteilung der Modelle

Alignment

Berücksichtigt man bei der Erstellung des Alignments zusätzlich zur Sequenz auch die Sekundärstruktur von LANA, die eine Haarnadelstruktur bildet und deren β -Turn durch die Aminosäuren L8, R9, S10 und G11 gebildet wird, stellt man fest, dass bei Alignment 1 dieser Turn deutlich günstiger zu liegen kommt, da dadurch das Prolin P482 im Turn liegt und das Faltblatt nicht behindert.

RMSD

Die Atome der simulierten Aminosäuren der Histone nehmen durch die verwendeten Restraints im Vergleich zu der Struktur PDB-1ZLA nahezu dieselben Plätze ein. Der RMSD der schweren Atome des Proteinrückgrats wurde daher nur für das simulierte Sequenzstück

von IE1 berechnet. Dabei ist ein deutlicher Unterschied zwischen beiden Modellen zu finden: Während er für Modell 1 0,038 nm ist, also ein gut definiertes Strukturbündel vorliegt, ist er für Modell 2 0,200 nm groß. Diese deutlich schlechtere Strukturbestimmung hat ihre Ursache darin, dass für die Aminosäuren 476 bis 480 keine entsprechenden Informationen von LANA übertragen werden konnten (vergleiche Abbildung 3-13, Alignment).

Wasserstoffbrücken

Eine Untersuchung der berechneten Strukturen mit den Programmen PERMOL, PyMOL und MOLMOL, die eine Suche nach Wasserstoffbrücken erlauben, ergab, dass bei Modell 1 die β -Haarnadelstruktur vom C-Terminus von IE1 durch vier Wasserstoffbrücken zwischen Hauptkettenatomen stabilisiert wird (Abbildung 3-15). Die Bindung an die durch die Histone H2A und H2B geformte saure Tasche erfolgt durch 10 zwischenmolekulare Wasserstoffbrücken (Abbildung 3-16). Wie ein Vergleich mit den in PDB-1ZLA vorhandenen Wasserstoffbrücken zeigt, konnten durch die Simulation zusätzlich zu bereits bei der Modellierungsvorlage ausgebildeten Brücken sowohl zwei inter- als auch zwei intramolekulare Wasserstoffbrücken gefunden werden, die bei LANA nicht vorhanden sind (Abbildung 3-17, rot gekennzeichnet).

Eine Untersuchung von Modell 2 ergab dagegen nur drei intra- und fünf intermolekulare Wasserstoffbrücken.

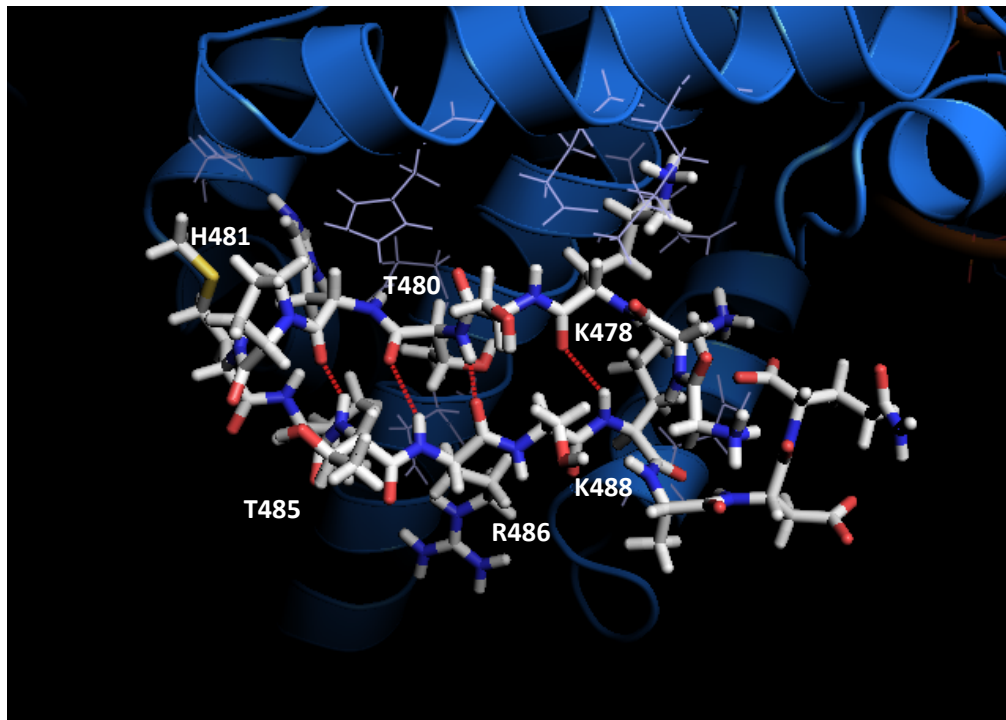


Abbildung 3-15: Die das Faltblatt stabilisierende intramolekularen Wasserstoffbrücken von IE1

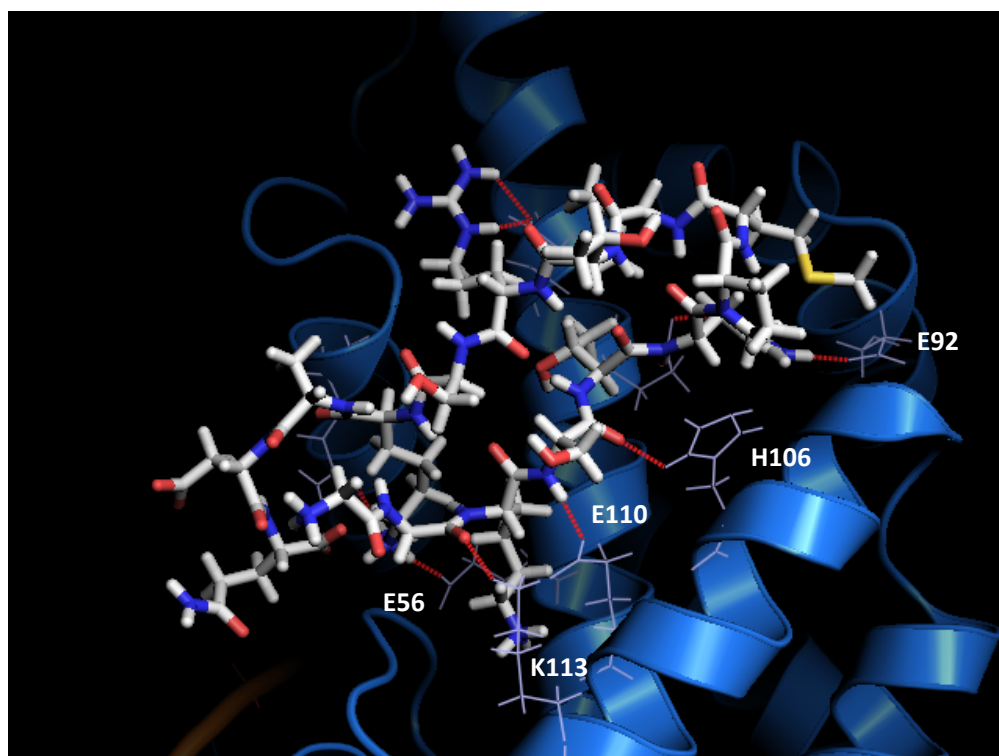
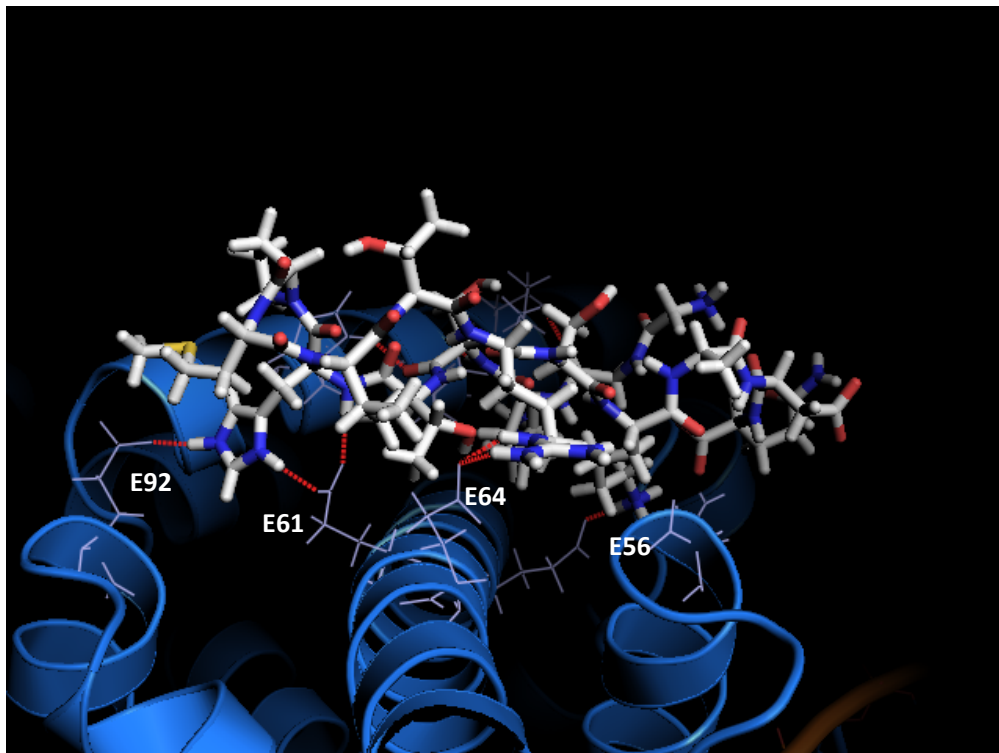


Abbildung 3-16: Intermolekulare Wasserstoffbrücken von Modell 1

Die Sekundärstrukturelemente (Cartoon-Optik) der intermolekularen Wasserstoffbrücken von Modell 1 zwischen IE1 und den Histonen sind blau dargestellt, schwach graublau die Seitenketten.

Oben: Wasserstoffbrücken zu H2A (E92, E61, E64 und E56)

Unten: Wasserstoffbrücken zu H2B (K113, E110 und H106), wobei allerdings diejenige zu Q44 nicht zu sehen ist.

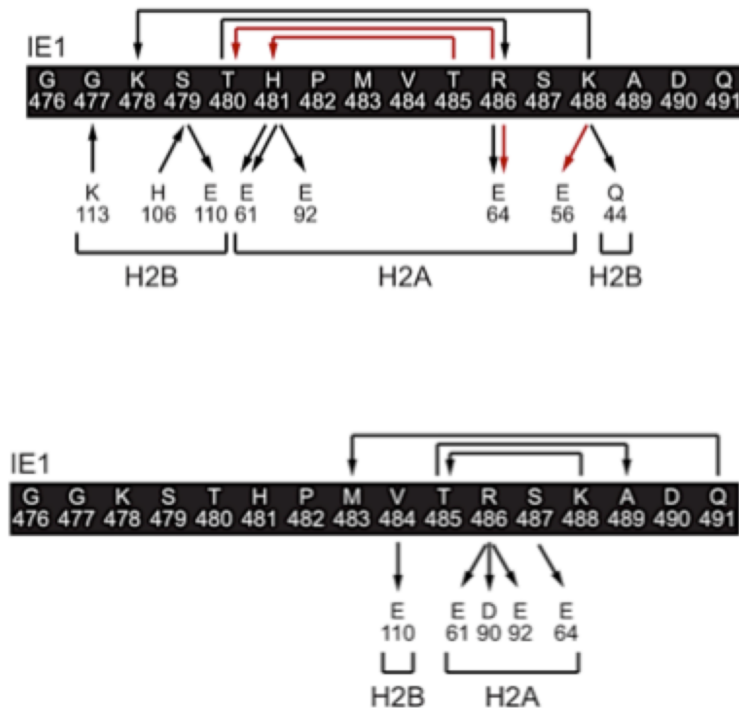


Abbildung 3-17: Inter- und intramolekulare (zu H2A und H2B) Wasserstoffbrücken der beiden berechneten Modelle für den C-Terminus von IE1

Der Pfeil zeigt vom Donator zum Akzeptor. Rot eingezeichnet sind die Wasserstoffbrücken, die bei der Struktur PDB-1ZLA, die als Modellierungsgrundlage diente, nicht vorhanden sind.

Oben: Modell 1, unten: Modell 2

(publiziert in Mücke et al., 2014)

Vergleich mit experimentellen Ergebnissen

Parallel zu diesen Strukturrechnungen wurden am Institut für Medizinische Mikrobiologie und Hygiene der Universität Regensburg verschiedene *in vivo*, *in vitro* und *in silico* Experimente durchgeführt. Außerdem wurden Bindungsstudien zwischen den Histonen, die die saure Tasche bilden, und dem C-Terminus von IE1 erstellt. Dabei erfolgte ein Vergleich mit dem Wildtyp von H2A und mehreren Mutanten, bei denen einfach oder mehrfach Aminosäuren durch Alanin ersetzt waren. Sowohl die Ergebnisse aus den *in silico* als auch aus den Mutageneseexperimenten lassen Modell 1 wahrscheinlicher erscheinen als Modell 2.

Komplementarität in Form und Ladung

Da Modell 1 aufgrund des strukturell besser passenderen Alignments und den Übereinstimmungen mit den experimentellen Ergebnissen deutlich dem zweiten Modell vorzuziehen ist, wurde das erste Modell weiter untersucht. Dabei ließ sich eine sehr gute Komplementarität in Form und Ladung zwischen dem C-Terminus von IE1 und der sauren

Tasche feststellen, die in Abbildung 3-18 deutlich als roter Bereich auf der Oberfläche des Nukleosoms erkennbar ist.

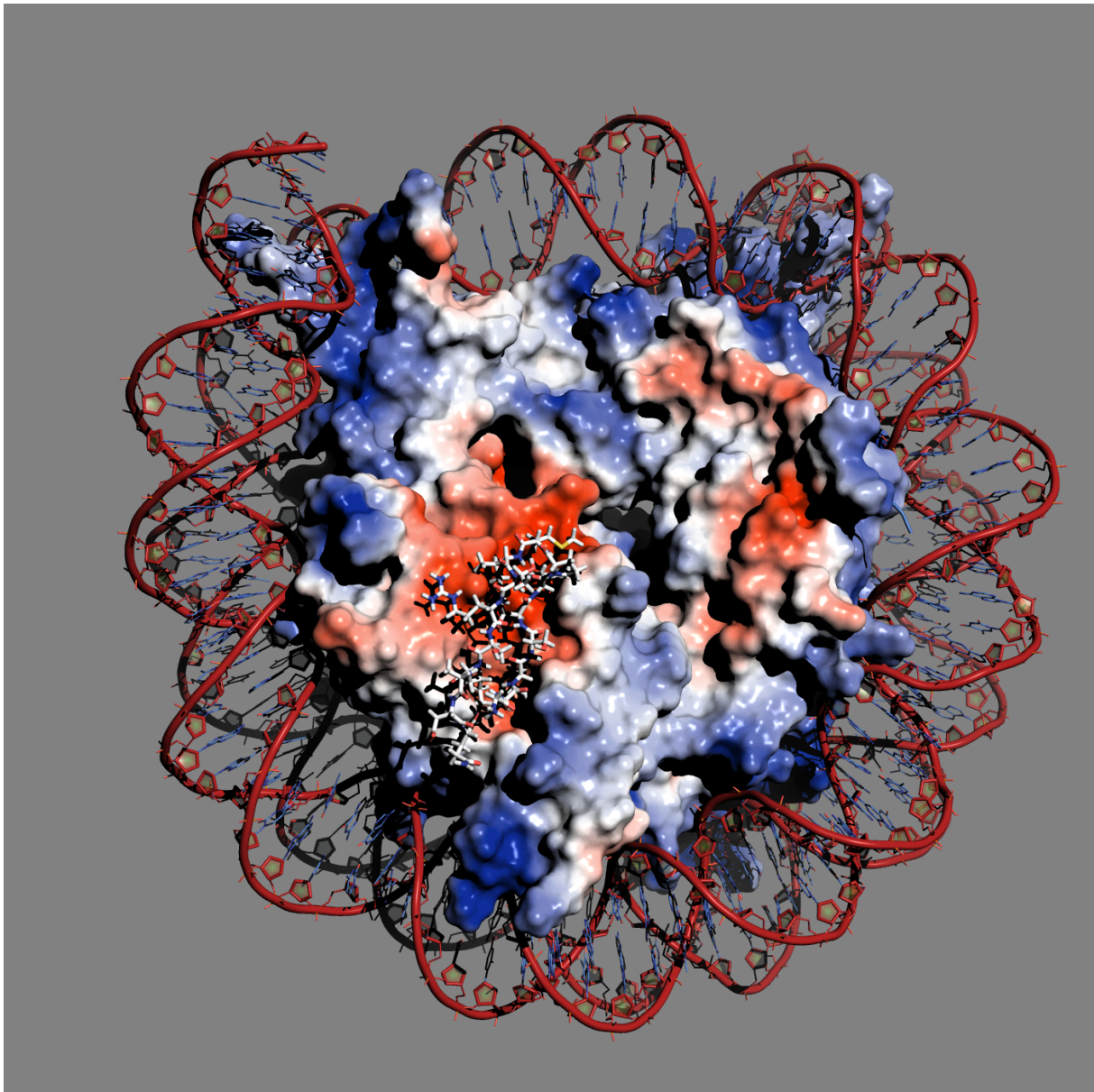


Abbildung 3-18: Nukleosom

Nukleosom, umgeben von menschlicher DNA, dargestellt als Oberfläche, bei der Aminosäuren mit negativ geladenen Seitenketten rot und positiv geladenen Seitenketten blau gefärbt sind. Deutlich zu erkennen ist die sogenannte saure Tasche, in die der C-Terminus von IE1 bindet, der selbst im Bindebereich positiv geladen ist. IE1₄₇₆₋₄₉₁ ist als Sticks abgebildet.

(publiziert in Journal of Virology, Januar 2014, Titelbild)

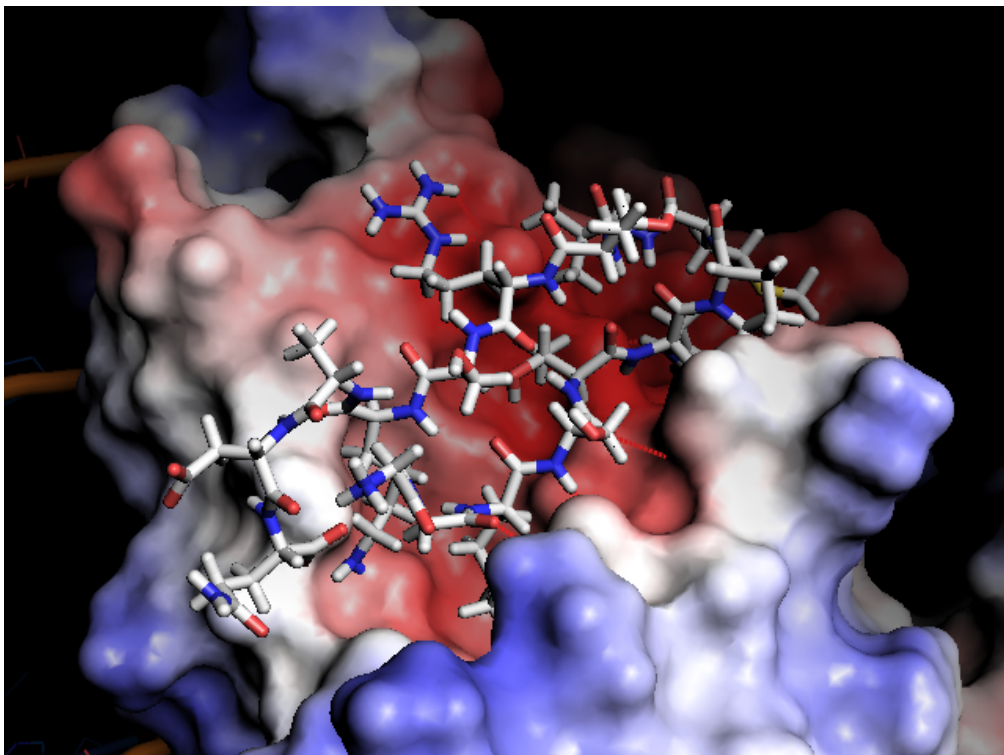
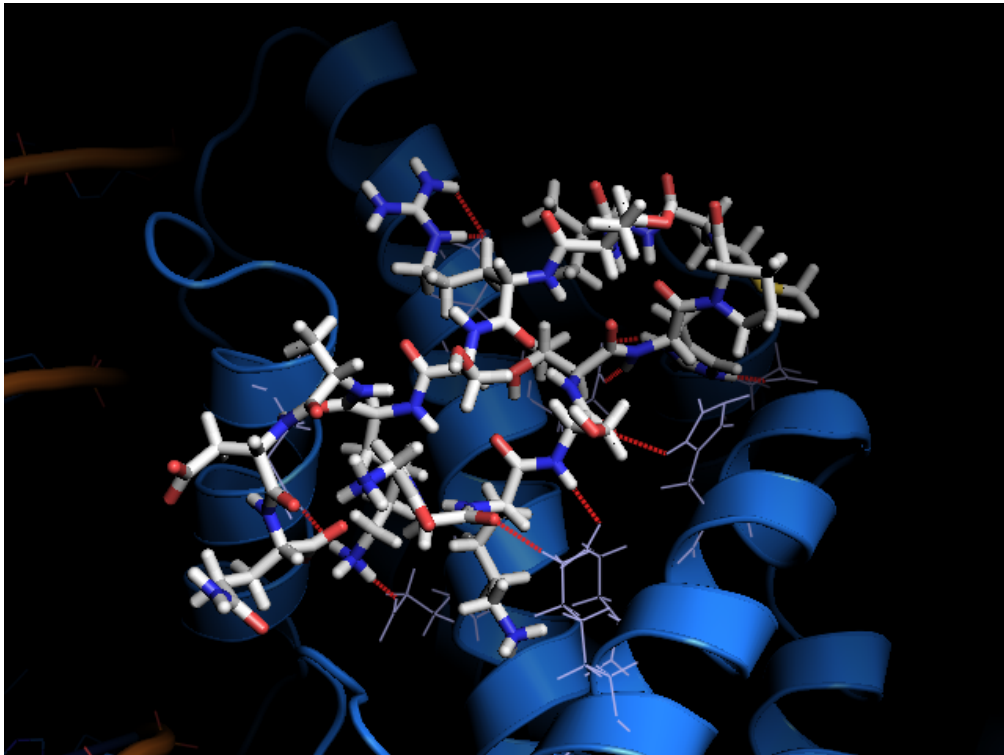


Abbildung 3-19: Modell 1 - Komplementarität der Oberfläche des Nukleosoms und der Struktur von IE1₄₇₆₋₄₉₁

Bei dem berechneten Modell 1 liegt auch eine gute Komplementarität der Oberfläche des Nukleosoms und der Struktur von IE1₄₇₆₋₄₉₁ vor. Dieses C-terminale Ende passt sich gut in die gegebenen Kavitäten ein. In der oberen Darstellung sind nur die Sekundärstrukturelemente der Histone und deren bindingsrelevante Seitenketten zu sehen. Unten ist aus dem gleichen Blickwinkel die Oberfläche des Nukleosoms abgebildet.

Abbildung 3-18 zeigt Modell 1, wobei die Oberfläche des Chromatins, umgeben von der DNA, so gestaltet ist, dass Aminosäuren mit positiv geladenen Seitenketten blau und Aminosäuren mit negativ geladenen Seitenketten rot dargestellt sind. Der Bindebereich, die saure Tasche ist deutlich rot und wird dominiert von Asparaginsäuren und Glutaminsäuren. Der C-Terminus von IE1 ist in Sticks-Darstellung zu sehen und besitzt im Bindebereich durch zwei Lysine, ein Arginin und ein Histidin positive Ladung. Die Komplementarität der Form zeigt Abbildung 3-19.

3.2.4 NMR-spektroskopische Untersuchungen von IE1(1-491) und IE1(1-475)

Zusätzlich wurden NMR-spektroskopische Untersuchungen an nicht isotopenmarkiertem Protein durchgeführt. Für das CMV Protein IE1 ist aufgrund seiner Masse von 72 kDa eine sehr große Linienbreite für die Signale der Atome der gut gefalteten Bereiche des Proteins zu erwarten, so dass sie im Spektrum nicht vom Hintergrundrauschen unterschieden werden können. Es stellte sich die Frage, ob der C-Terminus in Lösung ohne die Nähe von Chromatin ungefaltet ist und sich dann bei Bindung strukturell entsprechend anpasst. Für diesen Bereich müssten dann aufgrund seiner Beweglichkeit im Gegensatz zum restlichen Protein scharfe Signale messbar sein. Zur Identifizierung wurden deswegen Experimente am ganzen Protein und an einer verkürzten Variante, der die letzten 16 Aminosäuren - der für die Bindung relevante C-Terminus - fehlten, unternommen. Die erhaltenen Spektren wurden mit TOPSPIN und AUREMOL prozessiert und anhand des Eichpeaks von DSS geeicht.

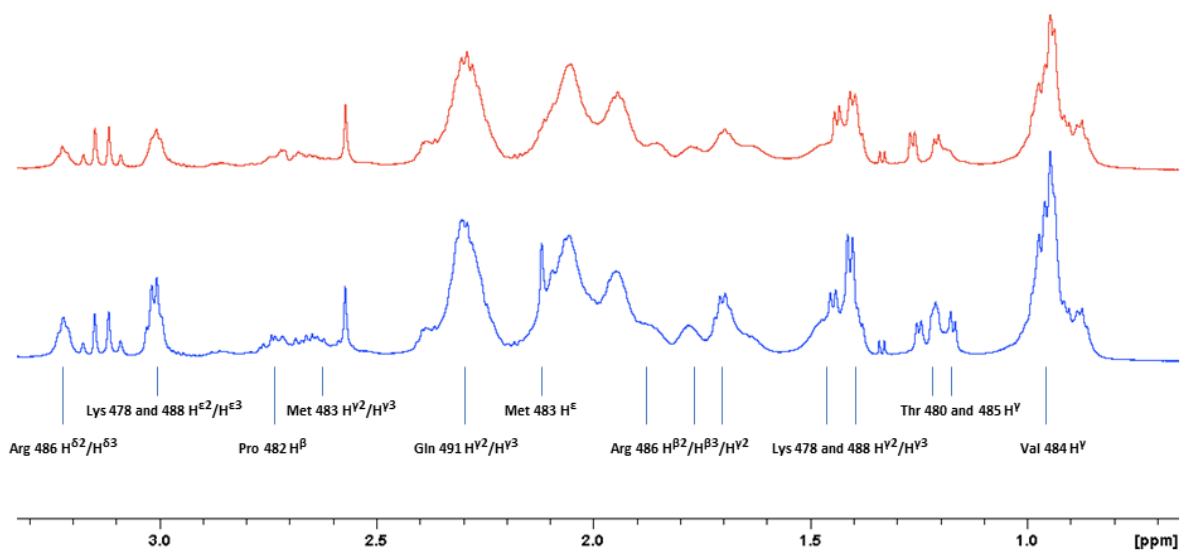


Abbildung 3-20: Vergleich der 1D-NOESY-Spektren des ganzen Proteins (blau) und der um den C-Terminus verkürzten Variante (rot), gemessen in D₂O

Bei den Spektren des ganzen Proteins konnten im Vergleich zur verkürzten Variante zusätzliche Peaks gefunden werden, die sich anhand der Signale der aufgenommenen 2D-NOESY- und 2D-TOCSY-Experimente dem C-Terminus zuordnen ließen (Abbildung 3-20 und Tabelle 3-5).

Tabelle 3-5: Chemische Verschiebungen der Wasserstoffatome des C-terminalen Endes von IE1

Aminosäure	Atomname	Chemische Verschiebung (in ppm)
Threonin 473	H ^N	8.32
	H ^α	4.36
	H ^β	4.18
	H ^γ	1.21
Glycin 476	H ^N	8.38
	H ^α	4.03/4.06
Glycin 477	H ^N	8.29
	H ^α	3.99/4.00
Lysin 478	H ^N	8.24
	H ^α	4.40
	H ^β	1.87/1.76
	H ^γ	1.49/1.43
	H ^δ	1.68
	H ^ε	2.99/3.02
Serin 479	H ^N	8.43
	H ^α	4.39
	H ^β	3.96/3.99
Threonin 480	H ^N	8.22
	H ^α	4.33
	H ^β	4.29
	H ^γ	1.25
Histidin 481	H ^α	4.98
	H ^β	3.22/3.12
	H ^{δ2}	7.24
	H ^{ε1}	8.35
Prolin 482	H ^α	4.45
	H ^{β2}	2.30
	H ^{β3}	1.94
	H ^γ	2.06
Methionin 483	H ^N	8.66
	H ^α	4.53
	H ^β	2.08/2.05
	H ^γ	2.58/2.64
	H ^ε	2.12
Valin 484	H ^N	8.25
	H ^α	4.21
	H ^β	2.12
	H ^γ	0.96
Threonin 485	H ^N	8.20
	H ^α	4.21
	H ^β	4.13
	H ^γ	1.18
Arginin 486	H ^N	8.37
	H ^α	4.43
	H ^β	1.89/1.78
	H ^γ	1.67/1.63
	H ^δ	3.22
	H ^ε	7.15
Lysin 488	H ^N	8.44
	H ^α	4.32
	H ^β	1.86/1.72
	H ^γ	1.48/1.43
	H ^δ	1.65
	H ^ε	3.05/2.99
Alanin 489	H ^N	8.34
	H ^α	4.36
	H ^β	1.44
Glutamin 491	H ^N	7.80
	H ^α	4.17
	H ^β	2.13/1.96
	H ^γ	2.30
	H ^{ε21}	7.55
	H ^{ε2}	6.84

Verifiziert werden konnte diese Zuordnung durch fünf sequenzspezifische NOE-Kontakte:

- H^α von G477 zu H^N von K478
- H^α von K478 zu H^N von S479
- H^α von P482 zu H^N von M483
- H^α von M483 zu H^N von V484
- H^α von K488 zu H^N von A489

Dagegen wurden keine NOE-Kontakte gefunden, die bereits eine β -Haarnadel-Struktur ähnlich Modell 1 oder 2 im ungebundenen Zustand bestätigen würden.

Für den C-Terminus lässt sich außerdem feststellen, dass er bereits bei T473 flexibel sein muss, da nicht nur für zwei sondern für drei Threonine Signale identifiziert werden können. Sie lassen sich durch ihre Linienbreite - T473 mit 6,7 Hz, T480 mit 5,6 Hz und T485 mit 4,1 Hz - und folglich durch ihre Beweglichkeit unterscheiden und können auf diese Weise zugeordnet werden.

Signale, die von der Aminosäure T473 stammen, sind auch in den Spektren der verkürzten Variante zu finden, wo sie sich an drittletzter C-terminaler Position befindet. Allerdings ist der Peak aufgrund der Änderung der Umgebung deutlich verschoben und wegen der größeren Beweglichkeit deutlich schärfer.

Zusätzliche weitere Datenanalysen zeigten, dass mindestens eine weitere Region des Proteins beweglich ist, da einige weitere Peaks in den Spektren beider Konstrukte zu finden sind. Insbesondere lassen sich die typischen Muster eines Histidins, eines Tyrosins und eines Phenylalanins feststellen, die alle drei in der Sequenz von IE1 sehr selten sind. Das in der Sequenz sehr häufig vertretene Methionin, das sich normalerweise leicht über sein typisches Muster im TOCSY-Spektrum identifizieren lässt, kann nicht ermittelt werden. Damit ist der N-Terminus als weitere bewegliche Region auszuschließen. Ebenso lässt sich auch kein Tryptophan nachweisen. Gefunden werden dagegen mindestens zwei Alanine, ein Arginin, des weiteren vier Muster in den TOCSY-Spektren, die von Asparaginsäuren oder Glutaminen stammen könnten, zwei oder mehr Isoleucine, ein Leucin, zwei oder mehr Lysine und ein Threonin. Ein Vergleich mit der Sequenz legt den Schluss nahe, dass der bewegliche Bereich die Aminosäuren A248 bis D269 umfasst (Abbildung 3-21). Zwei durchgeführte

Sekundärstrukturanalysen mit den Programmen PredictProtein und JPred3 sagten außerdem übereinstimmend für diesen Sequenzbereich einen Loopbereich vorher.

```
MESSAKRKMDPDNPDEGPSSKVPRPETPVTKATTFLLQTMLRKEVNSQLSLGDPLFP
ELAEESLKT FERVTEDCNENPEKDVLAELVKQIKVRVDMVRHRIKEHMLKKYTQTEEK
FTGA FNMMGGCLQNALDILDKVHEPFEEMKCIGLT MQSMYENYIVPEDKRE MW
MACIKEL HDVSKGAANKLGALQAKARAKKDELRRKMMYMCYRNIEFFTKNSAFP
KTTNGCSQAMAALQNLQPQCSPDEIMAYAQKIFKILDEERDKVLTHIDHIFMDILTTC
VETMCNEYKVTSDACMMTMYGGISLLSEFCRVLCCYVLEETSVMLAKRPLITKPEVIS
VMKRRIEEICMKVFAQYILGADPLRVCSVDDDLRAIAEESDEEEAIVAYTLATAGASS
SDSLVSPPEPVPATIPLSSVIVAENS DQEESEQSDEEEEGAQEEREDTVSVKSEPVS
EIEEVAPEEEEDGAEEPTASGGKSTHPMVTRSKADQ
```

Abbildung 3-21: Primärsequenz des Proteins IE1

Primärsequenz des Proteins IE1, die 491 Aminosäuren umfasst. Um den Bereich zu identifizieren, der außer dem C-Terminus noch beweglich ist, kann dieser selbst ausgeschlossen werden (durchgestrichen). Rot gekennzeichnet sind alle Methionine, deren typisches Muster nicht gefunden wurde. Grün markiert sind die Aminosäuren, die in der Sequenz seltener zu finden sind, aber (vermutlich) je einmal in dem gesuchten Sequenzstück vorhanden sein müssen. Grau unterlegt ist der resultierende Bereich, der auch mit der weiteren gefundenen Aminosäurezusammensetzung kompatibel ist.

3.3 Optimierung der Anfangsbedingungen von Algorithmus 1 durch Untersuchungen des Parameterraums

3.3.1 Die Form der Funktion Norm über dem Parameterraum

Wie in Kapitel 2.2.4 erwähnt, muss bei der Anwendung des Levenberg-Marquardt-Algorithmus auf sinnvolle Anfangsparameter geachtet werden, wobei idealerweise mehrere Punkte im Parameterraum als Startpunkte getestet werden sollten. Diese lassen sich aus der Form der Funktion Norm über dem Parameterraum ableiten, die anhand von Testdatensätzen untersucht wurde.

Zweizustandssysteme:

Systematische Variation von ΔG_2^0 und ΔV_2^0 bei $\Delta b_2^{0'}$ gleich Null

Um eine Vorstellung von der Form der Funktion Norm über dem Parameterraum bei Zweizustandssystemen zu bekommen, wurde eine Datenreihe mit Algorithmus 2 untersucht, für deren Erzeugung ΔG_2^0 systematisch von 0 bis 5 und ΔV_2^0 von $-0,04 \text{ MPa}^{-1}$ bis $0,02 \text{ MPa}^{-1}$ variiert und mit $\Delta b_2^{0'}$ gleich 0 kombiniert wurde, so dass 49 verschiedene Datensätze entstanden. Wie in Kapitel 2.4.1 definiert, ist ΔG_2^0 die durch Teilen durch $R \cdot T$ modifizierte Gibbs-Energiedifferenz der Zustände bei Normaldruck, und ΔV_2^0 und $\Delta b_2^{0'}$ sind die ebenfalls durch $R \cdot T$ geteilten Differenzen des partiellen molaren Volumens und des Kompressibilitätsfaktors.

Abbildung 3-22 zeigt die Anteile der beiden Zustände in Abhängigkeit vom Druck für diese 49 Datensätze. Dabei wird ersichtlich, dass bei der hier verwendeten systematischen Variation der Parameter ΔG_2^0 und ΔV_2^0 auch Systeme entstehen, bei denen einer der beiden Zustände über den gesamten Messbereich viel zu schwach populierte ist, um mit den beschriebenen Methoden nachweisbar zu sein.

Bei der Anwendung von Algorithmus 2 wurden in einem feinen Raster, in dem ΔG_2^0 und ΔV_2^0 variiert wurden, für jeden Punkt die nach A optimierte Norm berechnet und gespeichert. Abbildung 3-23 zeigt die Ergebnisse der Datenreihe im Überblick. Dargestellt ist nur die Ebene, in der $\Delta b_2^{0'}$ gleich Null gehalten wurde. Nach rechts sind jeweils die Werte für ΔV_2^0 variiert, nach oben die für ΔG_2^0 . Die Daten sind als Höhenplots angelegt, wobei blaue Farbe eine kleinere Norm anzeigt als gelbe. Das Minimum wurde mit einem roten Kreis gekennzeichnet, das dem zur Erzeugung der Daten verwendeten Parameterset entspricht.

Jeder der Plots weist eine Punktsymmetrie um den Nullpunkt auf, was sich aus der Möglichkeit der Vertauschung der Zustände ergibt.

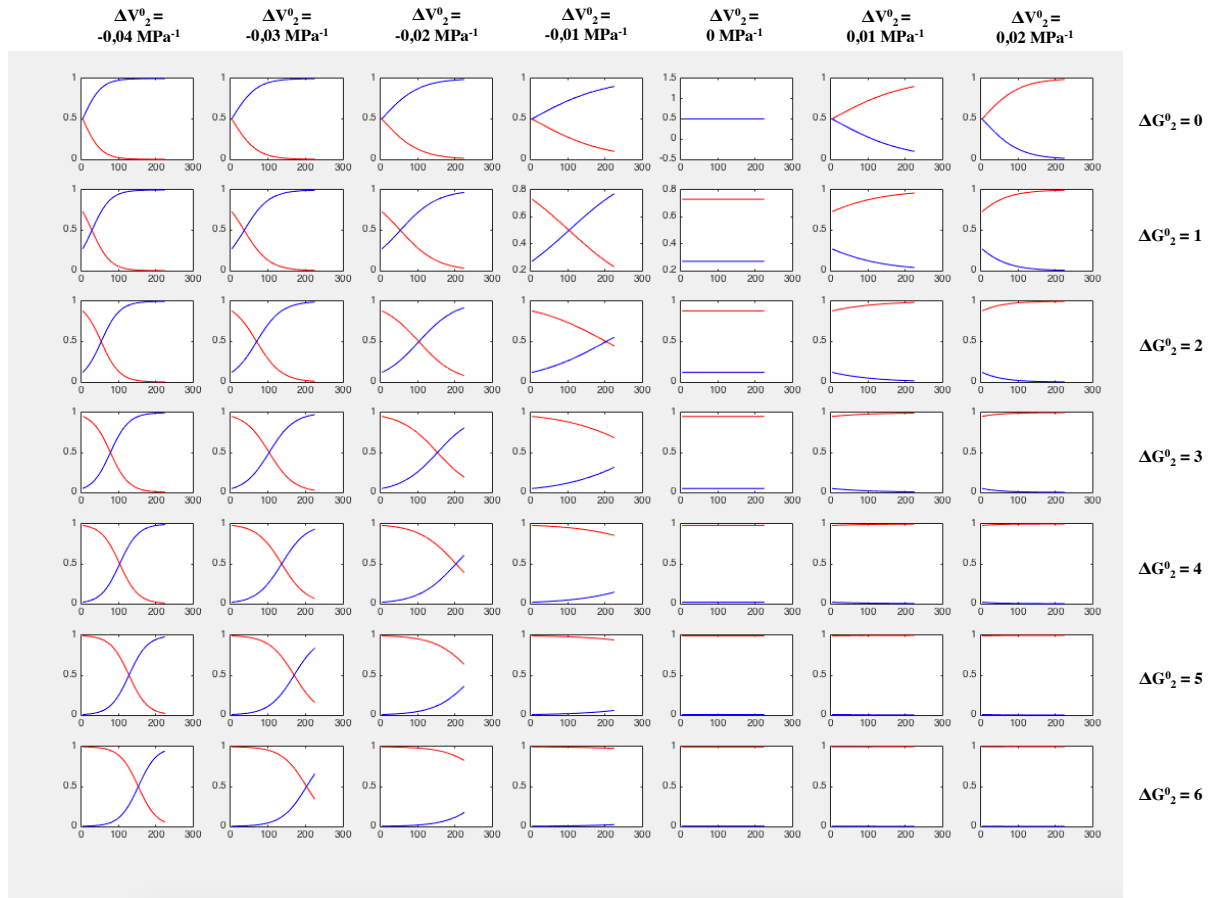


Abbildung 3-22: Anteile der Besetzungszahlen der beiden Zustände in Abhängigkeit vom Druck

Anteile der Besetzungszahlen (nach oben) der beiden Zustände in Abhängigkeit vom Druck (nach rechts, aufgetragen in MPa) für verschiedene Parametersätze im Überblick: Vertikal wurde der Parameter ΔG_2^0 variiert, horizontal der Parameter ΔV_2^0 [MPa⁻¹]. Die für einen Datensatz verwendeten Werte sind jeweils am oberen beziehungsweise rechten Rand ersichtlich.

Außerdem ist als Linie die Punkteschar ΔV_2^0 gleich Null zu erkennen, die als Wand durch jede der Graphiken läuft (siehe auch Abbildungen 3-24 bis 3-27). Die fünfte senkrechte Reihe der Plots, die mit dem Parameter ΔV_2^0 gleich Null erzeugt wurde, zeigt keine Struktur. Bei dieser Parameterwahl bleiben die Populationsverhältnisse gleich (siehe Abbildung 3-22), so dass die in der Matrix D gespeicherten korrigierten chemischen Verschiebungen für jede Aminosäure einen festen Wert über den gesamten Messbereich haben. Um das Gleichungssystem für die Punkte des Parameterraums, für die ΔV_2^0 ungleich Null ist, stimmig zu lösen, bleibt nur die Möglichkeit, die korrigierten chemischen Verschiebungen für Zustand 1 und 2 gleichzusetzen, die dann vom Algorithmus als Minimum gefunden wird. Daraus ergibt sich

für jeden Punkt des Parameterraums eine einheitliche minimale sehr kleine Norm und damit eine einheitliche Färbung.

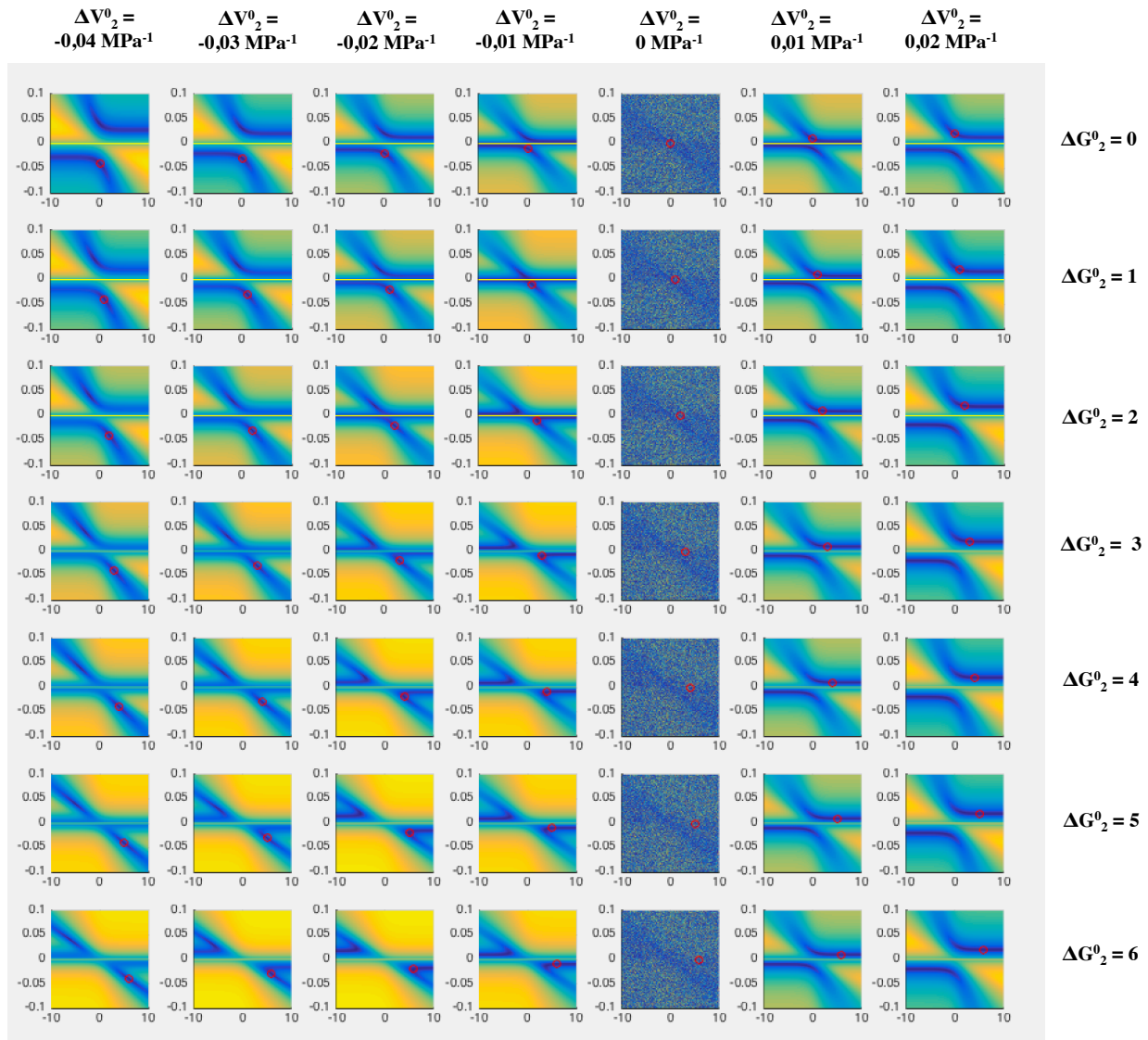


Abbildung 3-23: Überblick der Norm in Abhängigkeit ΔG_2^0 und ΔV_2^0 für die untersuchten Datensätze der Datenreihe

Die untersuchte Datenreihe im Überblick: Für jeden Datensatz wird die Norm in Abhängigkeit von ΔG_2^0 (nach rechts) und ΔV_2^0 (nach oben in MPa^{-1} aufgetragen) als Höhenplot (wobei blau kleinen Normen und gelb großen Normen entspricht, siehe auch Abbildung 3-24) bei optimalen Werten für A dargestellt. Es ist jeweils die Ebene $\Delta \mathbb{b}_2^{0'} = 0$ gezeigt, die das Minimum enthält. Der rote Kreis markiert die Lage des Minimums.

Für die verschiedenen Datensätze wurde vertikal der Parameter ΔG_2^0 variiert, horizontal der Parameter ΔV_2^0 [MPa^{-1}]. Die für einen Datensatz verwendeten Werte sind jeweils am oberen bzw. rechten Rand ersichtlich.

Die $\Delta G_2^0 - \Delta V_2^0$ - Ebene weist zwei Gräben auf, in denen die Norm besonders klein ist. Die Form dieser Gräben ist von den Parametern abhängig. Aus den ersten vier senkrechten Datenreihen in Abbildung 3-23 ist ersichtlich, dass der Winkel, der entsteht - wenn man sich die Gräben als zwei durchgezogene sich schneidende Geraden vorstellt - sowohl mit größer werdendem ΔG_2^0 als auch mit steigendem Betrag von ΔV_2^0 kleiner wird. Daraus folgt, dass die Grabenform von den dem System zu Grunde liegenden Parametern abhängt.

Lässt man zusätzlich auch eine Variation von $\Delta b_2^{0'}$ zu (bei den hier gezeigten Beispielen wurde diese Größe nicht variiert), kann man Parametersätze finden, deren Gräben einen praktisch identischen Verlauf zeigen und bei denen nur das Minimum in diesen Gräben verschoben ist. Auffällig bei solchen Systemen ist, dass auch deren Verlauf der Besetzungszahlen mit dem Druck fast identisch ist. Ein Beispiel dafür, wo dies, obwohl $\Delta b_2^{0'}$ gleich Null gehalten wurde, im Ansatz zu erkennen ist, sind die Datensätze mit ΔG_2^0 gleich 4 mit ΔV_2^0 gleich $-0,04 \text{ MPa}^{-1}$, ΔG_2^0 gleich 3 mit ΔV_2^0 gleich $-0,03 \text{ MPa}^{-1}$ und ΔG_2^0 gleich 2 mit ΔV_2^0 gleich $-0,02 \text{ MPa}^{-1}$, deren Schnittpunkte der Zustände jeweils bei 100 MPa liegen und deren Gräben das beschriebene Verhalten aufweisen.

Vergleicht man nun die Lage der Minima, gekennzeichnet durch den roten Kreis, bei den einzelnen Datensätzen einer Datenreihe (Abbildung 3-23), so stellt man fest, dass es zwei Gruppen von Datensätzen gibt: Bei der einen Gruppe liegt das Extremum in dem Graben, der parallel der Linie mit ΔV_2^0 gleich Null verläuft, bei der zweiten Gruppe im anderen Graben. Wie ein Vergleich mit den Besetzungszahlen (Abbildung 3-22) der verschiedenen Datensätze zeigt, entspricht erstgenannte Gruppe Datensätzen, bei denen der am Anfang der Druckreihe höher populierte Zustand am Ende der Druckreihe der schwächer besetzte ist. Bei der zweiten Gruppe bleibt der stärker besetzte Zustand über den gesamten Messbereich der dominanter.

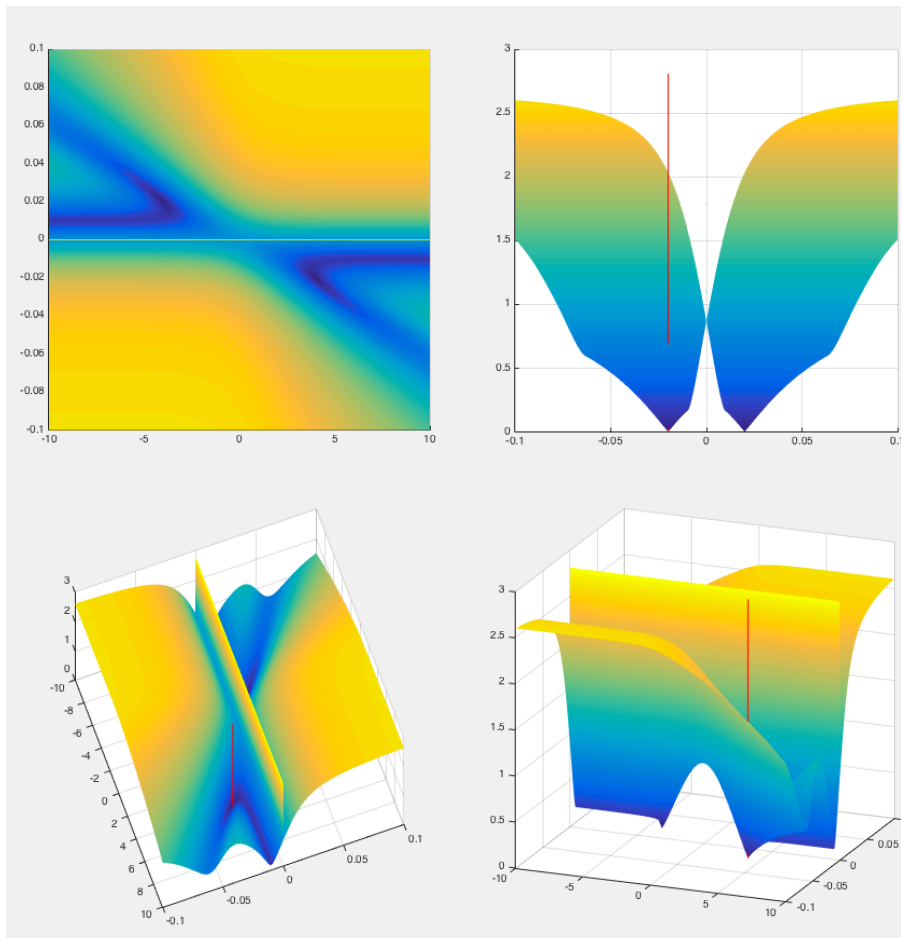


Abbildung 3-24: Die Funktion Norm als Höhenplot in verschiedenen Ansichten

Die Darstellung erfolgte über der Ebene $\Delta G_2^0 - \Delta V_2^0$, ΔG_2^0 variiert von -10 bis 10, ΔV_2^0 von -0,1 bis 0,1. $\Delta b_2^{0'}$ ist in der ganzen Ebene 0, die Matrix A wurde an jedem Punkt der Ebene so gewählt, dass die Norm minimal ist.

Zur Erzeugung des untersuchten Datensatzes wurden folgende Parameter verwendet:

$$\begin{aligned}\Delta G_2^0 &= 4 \\ \Delta V_2^0 &= -0,02 \text{ MPa}^{-1} \\ \Delta b_2^{0'} &= 0\end{aligned}$$

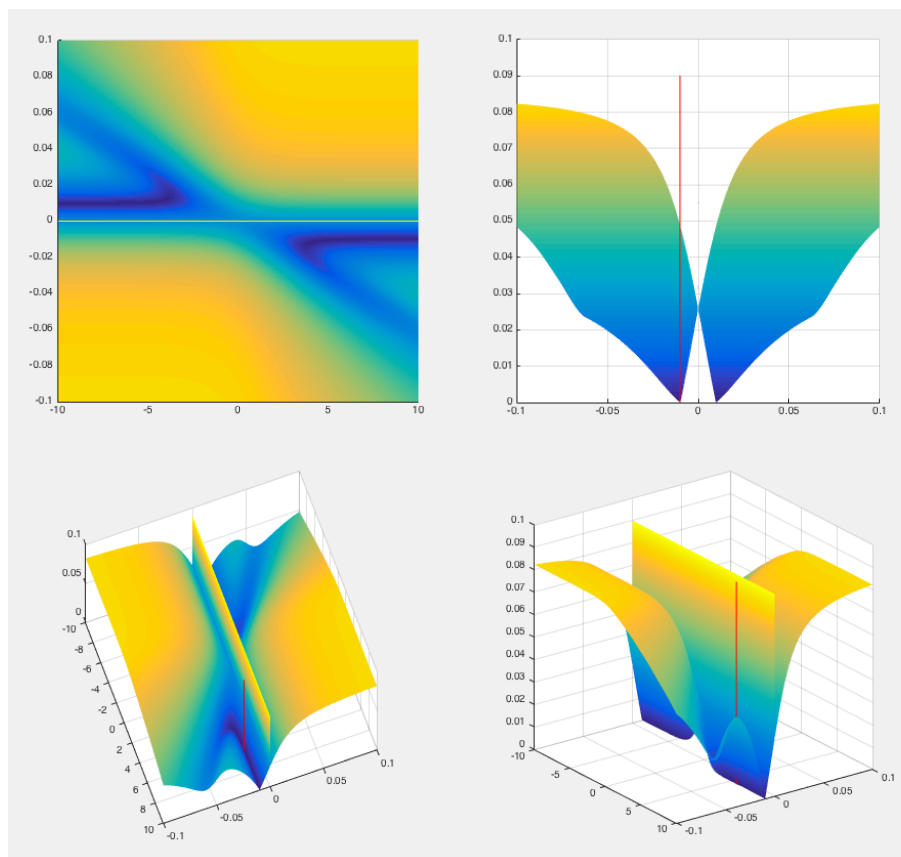


Abbildung 3-25: Die Funktion Norm als Höhenplot in verschiedenen Ansichten

Die Darstellung erfolgte analog Abbildung 3-24.

Zur Erzeugung wurden folgende Parameter verwendet:

$$\begin{aligned}\Delta G_2^0 &= 6 \\ \Delta V_2^0 &= -0,01 \text{ MPa}^{-1} \\ \Delta b_2^{0'} &= 0\end{aligned}$$

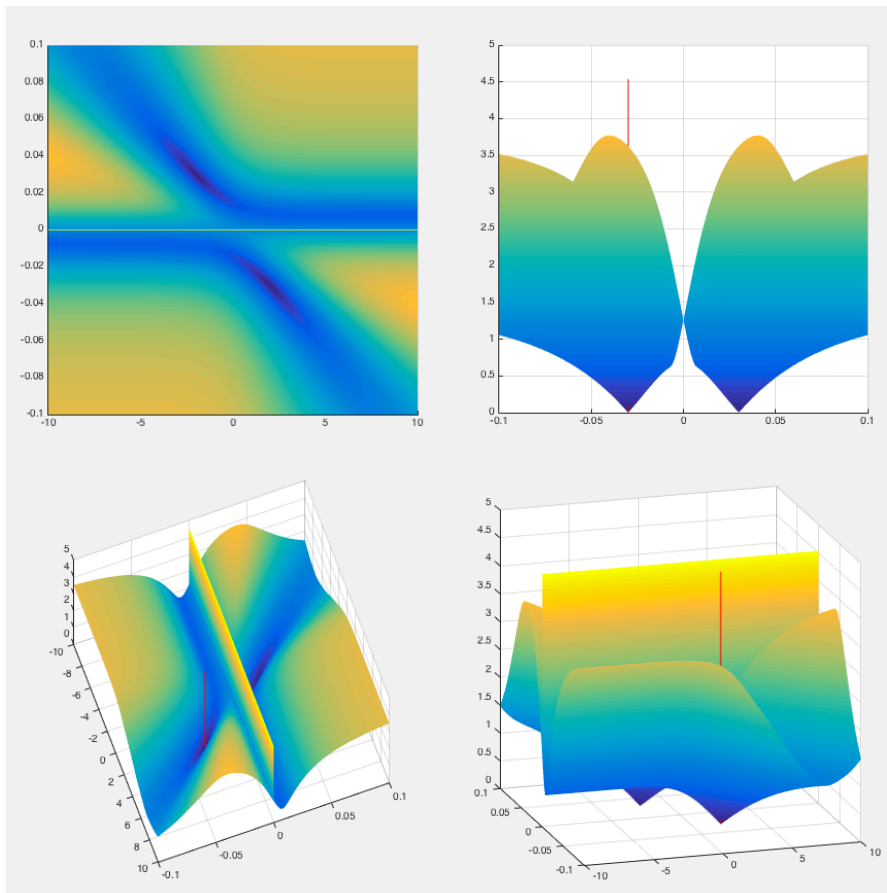


Abbildung 3-26: Die Funktion Norm als Höhenplot in verschiedenen Ansichten

Die Darstellung erfolgte analog Abbildung 3-24.

Zur Erzeugung wurden folgende Parameter verwendet:

$$\begin{aligned}\Delta G_2^0 &= 2 \\ \Delta V_2^0 &= -0,03 \text{ MPa}^{-1} \\ \Delta b_2^{0'} &= 0\end{aligned}$$

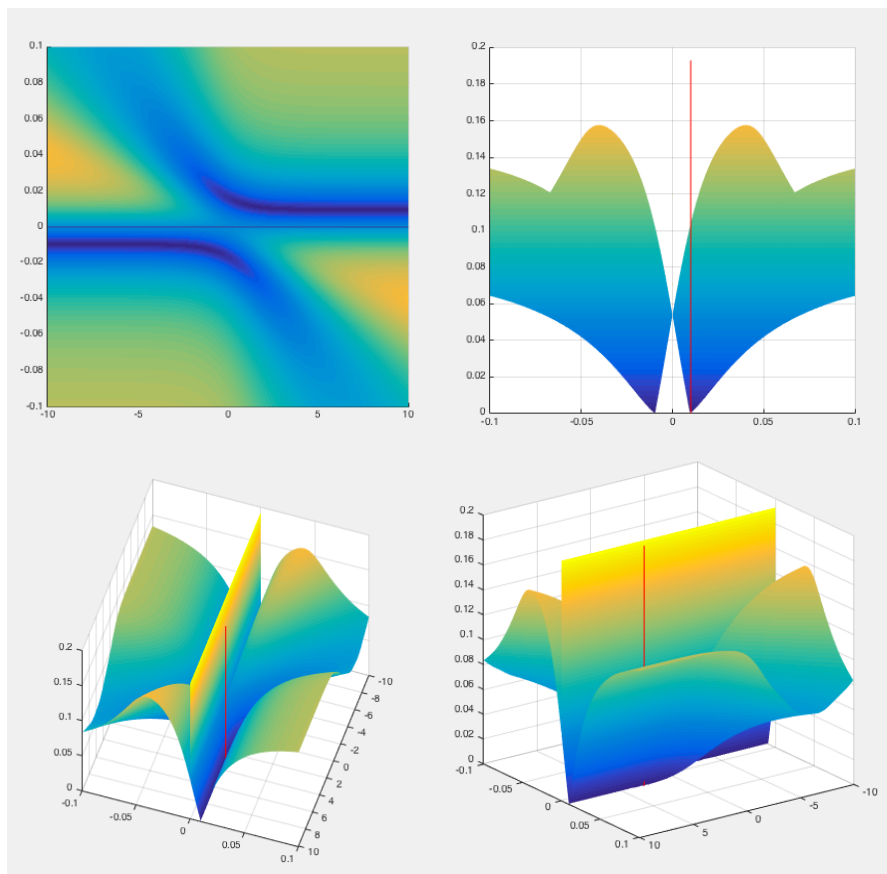


Abbildung 3-27: Die Funktion Norm als Höhenplot in verschiedenen Ansichten

Die Darstellung erfolgte analog Abbildung 3-24.

Zur Erzeugung wurden folgende Parameter verwendet:

$$\begin{aligned}\Delta G_2^0 &= 3 \\ \Delta V_2^0 &= 0,01 \text{ MPa}^{-1} \\ \Delta b_2^{0'} &= 0\end{aligned}$$

Die Abbildungen 3-24 bis 3-27 zeigen für verschiedene Datensätze im Vergleich die soeben als Höhenplots dargestellten Ergebnisse als Relief aus mehreren Perspektiven. In Abbildung 3-23 kann man erkennen, dass bei den Graphiken im oberen linken Bereich die Winkel α wesentlich größer sind als bei denen im unteren linken Bereich. Bei den kleineren Winkeln hängen das untere und das rechte Grabenstück zusammen und bilden ein spitzes V. An den Abbildungen 3-24 und 3-25 lässt sich sehen, wie sich die Fläche verhält, die die Norm über dem Parameterraum bildet, wenn sich das Minimum in diesem Graben verschiebt. Während bei den Graphiken in Abbildung 3-24 das Extremum deutlich ausgebildet ist, ist es bei Abbildung 3-25 über die Ecke auf die andere Grabenhälfte gewandert, die parallel zu ΔV_2^0 gleich Null verläuft. Deutlich festzustellen ist, dass das Minimum dort an einer sehr flachen Stelle des Grabens ist. Dies konnte als typisches Merkmal für Systeme beobachtet werden, deren Populationsverhältnisse sich kaum ändern.

Untersucht wurde auch, wie die Wahl der Elemente der Matrix von A sich auf die Form des Parameterraums auswirkt. Dabei zeigte sich, dass der Verlauf des Grabens in den oben dargestellten Reliefs nur durch die Parameter ΔG_2^0 , ΔV_2^0 und $\Delta b_2^{0'}$ bestimmt ist. Die Höhendifferenzen sind jedoch abhängig von der über alle betrachteten Atome gemittelten Differenz zwischen den korrigierten chemischen Verschiebungen der beiden Zustände. Sie werden umso größer, je größer diese Differenz ist. Im Grenzfall, bei dem die beiden korrigierten chemischen Verschiebungen für jede Aminosäure jeweils gleich sind - die Matrix A also zeilenweise gleiche Einträge besitzt -, wird aus dem Relief eine ebene Fläche, aus der keine Rückschlüsse mehr auf die erzeugenden Parameter gezogen werden können.

Die Norm im ΔG_2^0 - ΔV_2^0 - $\Delta b_2^{0'}$ -Raum für ideale Datensätze

Zur Berechnung der eben gezeigten Beispiele war $\Delta b_2^{0'}$ gleich Null gewählt und auch nur die entsprechende das Minimum enthaltene Ebene im Parameterraum betrachtet worden. Um nun die Abhängigkeit der Norm vom quadratischen Term der Taylorentwicklung von $\Delta G_2(p)$ für Systeme mit zwei Zuständen zu untersuchen, wurde für jede ΔV_2^0 - $\Delta b_2^{0'}$ -Ebene das Minimum der Norm gesucht und die dadurch entstehenden Kurven im ΔG_2^0 - ΔV_2^0 - $\Delta b_2^{0'}$ -Raum dargestellt. Die Kurven ändern sich stetig mit ΔG_2^0 , ΔV_2^0 und $\Delta b_2^{0'}$ des erzeugenden Parametersatzes. Abbildung 3-28 zeigt drei Beispiele, bei denen für ΔG_2^0 die Werte 1, -1 und -3 angenommen wurden, ΔV_2^0 und $\Delta b_2^{0'}$ hatten jeweils die Werte $-0,02 \text{ MPa}^{-1}$ beziehungsweise $5 \cdot 10^{-5} \text{ MPa}^{-2}$. Daraus ergeben sich die in Abbildung 3-28 gezeigten Besetzungen der Zustände.

Der ΔG_2^0 - ΔV_2^0 - $\Delta b_2^{0'}$ -Raum weist eine Punktsymmetrie auf, so dass zu einem Parametersatz immer zwei gleichfarbige gebogene Kurven gehören (Abbildung 3-29). In die Richtung von $\Delta b_2^{0'}$ sind die Kurven begrenzt, da die Norm ein Plateau erreicht und das Minimum nicht mehr auf einen Punkt in der ΔV_2^0 - $\Delta b_2^{0'}$ -Ebene begrenzt werden kann. In die andere Richtung verläuft die Kurve parallel zur ΔG_2^0 -Achse unbegrenzt weiter.

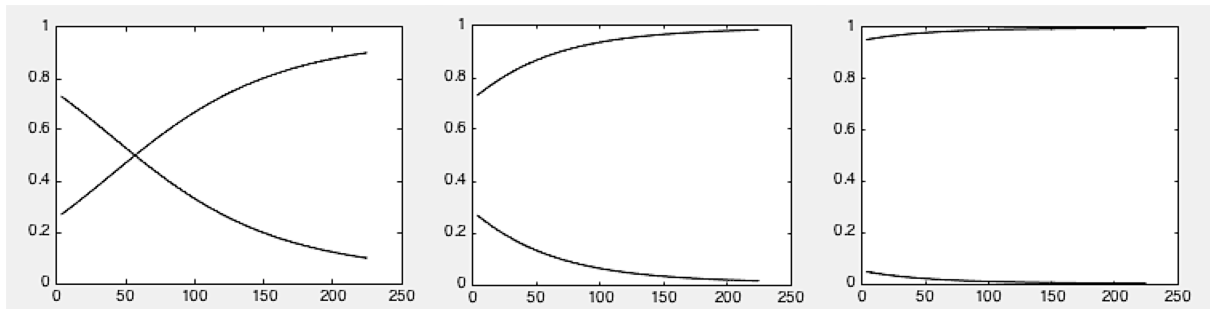


Abbildung 3-28: Die Anteile der beiden Zustände in Abhängigkeit vom Druck

Bei den gezeigten Beispielen wurden für ΔG_2^0 die Werte 1, -1 und -3 angenommen und die Anteile der beiden Zustände in Abhängigkeit vom Druck (in MPa) aufgetragen:

Links: $\Delta G_2^0 = 1$, $\Delta V_2^0 = -0,02 \text{ MPa}^{-1}$ und $\Delta b_2^{0'} = -5 \cdot 10^{-5} \text{ MPa}^{-2}$ (entspricht der blauen Kurve in Abbildung 3-29)

Mitte: $\Delta G_2^0 = -1$, $\Delta V_2^0 = -0,02 \text{ MPa}^{-1}$ und $\Delta b_2^{0'} = -5 \cdot 10^{-5} \text{ MPa}^{-2}$ (entspricht der grünen Kurve)

Rechts: $\Delta G_2^0 = -3$, $\Delta V_2^0 = -0,02 \text{ MPa}^{-1}$ und $\Delta b_2^{0'} = -5 \cdot 10^{-5} \text{ MPa}^{-2}$ (entspricht der roten Kurve)

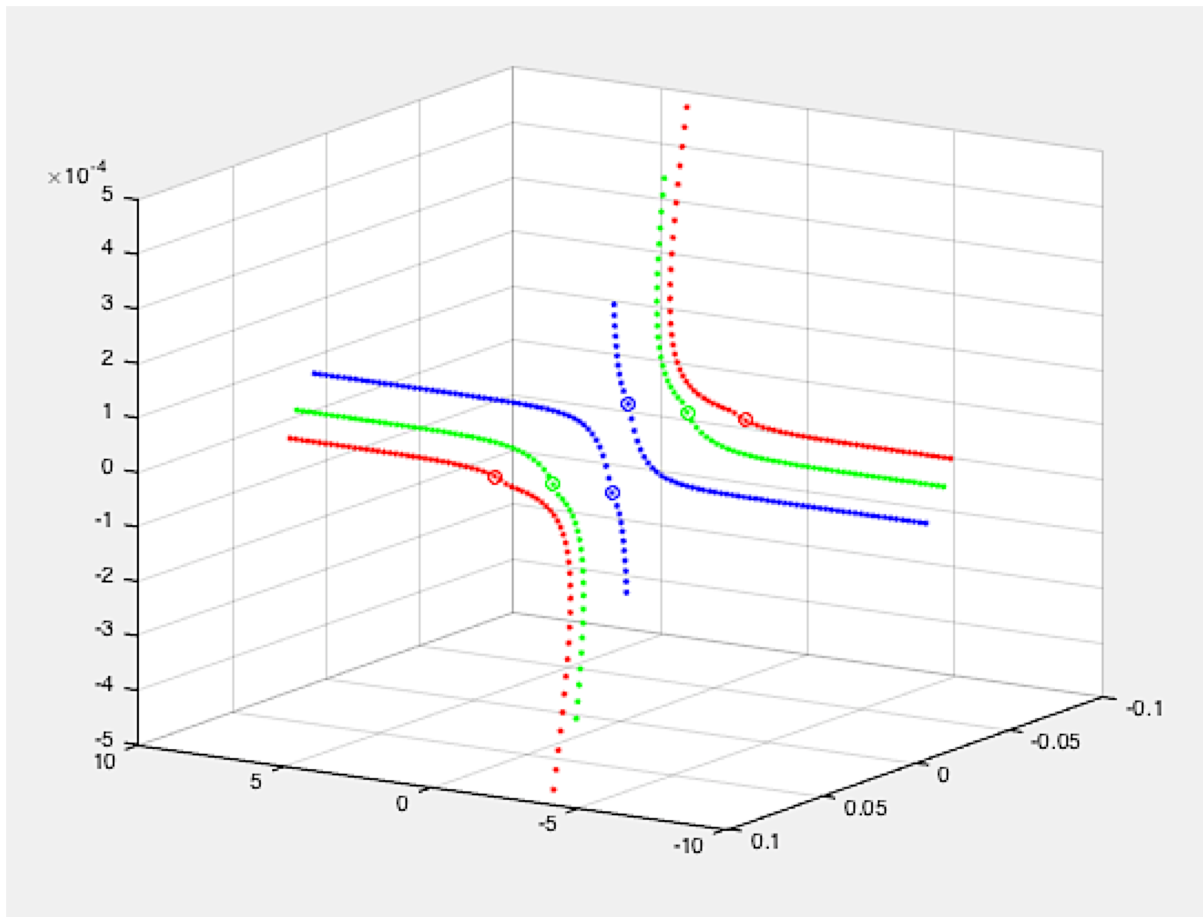


Abbildung 3-29: Minimale Norm durch den Parameterraum

Die minimale Norm durch den Parameterraum, nach rechts ΔG_2^0 , nach hinten ΔV_2^0 in MPa^{-1} , nach oben $\Delta b_2^{0'}$ in MPa^{-2} aufgetragen. Als Kreise sind jeweils die Punkte im Parameterraum gekennzeichnet, die den erzeugenden Systemparametern entsprechen.

Wie an den Anteilen der Zustände zu sehen ist, erzeugt der Parametersatz, der der linken Graphik in Abbildung 3-28 zugrunde liegt, ein System, dessen Besetzungszahlen sich mit dem Druck deutlich verändern. Ein Vergleich mit Abbildung 3-29 zeigt, dass sich der Kreis bei der zugehörigen Kurve durch den Parameterraum (blau), der den erzeugenden Parametern entspricht, auf dem Ast der Kurve, der in Richtung der $\Delta b_2^{0'}$ -Achse verläuft, befindet, während der eingezeichnete Kreis bei den beiden anderen dargestellten Systemen auf dem anderen Ast liegt.

Die Norm längs der Kurve durch den Raum

Untersucht wurde, wie sich die Norm entlang dieser in der Abbildung 3-29 dargestellten Kurven verhält (Abbildung 3-30). Betrachtet wurde der Verlauf der hinteren rechten Äste der

Graphik, für die vorderen dazu spiegelsymmetrischen Kurven erhält man einen am Nullpunkt der ΔG_2^0 -Achse gespiegelten Verlauf.

Wie eine genaue Analyse der Daten ergab, besitzen die drei untersuchten Datensätze die Minima der Norm an den jeweils erwarteten Stellen. Dabei fällt auf, dass ein Zusammenhang besteht zwischen der Größe der Änderungen in den Besetzungszahlen mit dem Druck und der Ausprägung der Extrema, wobei wenig Änderung nur schwache Minima hervorbringt. Der prinzipielle Verlauf entlang der ΔG_2^0 - Achse ist bei allen drei Kurven derselbe, zuerst ein steiler Abfall zum Minimum hin, dann ein Anstieg und nach Erreichen einer lokal höchsten Stelle ein kontinuierlicher Abfall, bei dem sich die Norm asymptotisch einem datensatzabhängigen Wert nähert. Dieser Verlauf konnte bei allen Datensätzen, deren Gräben kleinster Norm durch den Raum, wie in Abbildung 3-26 gezeigt, ein flaches V bilden, festgestellt werden. Datensätze, die wie der in Abbildung 3-24 ein spitzes V bilden, besitzen zusätzlich ein Nebenminimum, das jedoch nicht so tief wie das eigentliche Extremum ist.

Ebenfalls durchgeführte Untersuchungen zeigten, dass bei einer weiteren Steigerung des Absolutwertes von ΔG_2^0 das Minimum ab einem gewissen Punkt numerisch nicht mehr nachweisbar ist. Das entspricht Systemen, bei denen der eine Zustand viel stärker besetzt ist als der andere. Gleiches Verhalten findet sich auch bei einer Verkleinerung des Absolutwertes von ΔV_2^0 , was bewirkt, dass die Änderungen der Besetzungszahlen und damit der korrigierten chemischen Verschiebungen mit dem Druck sehr gering werden.

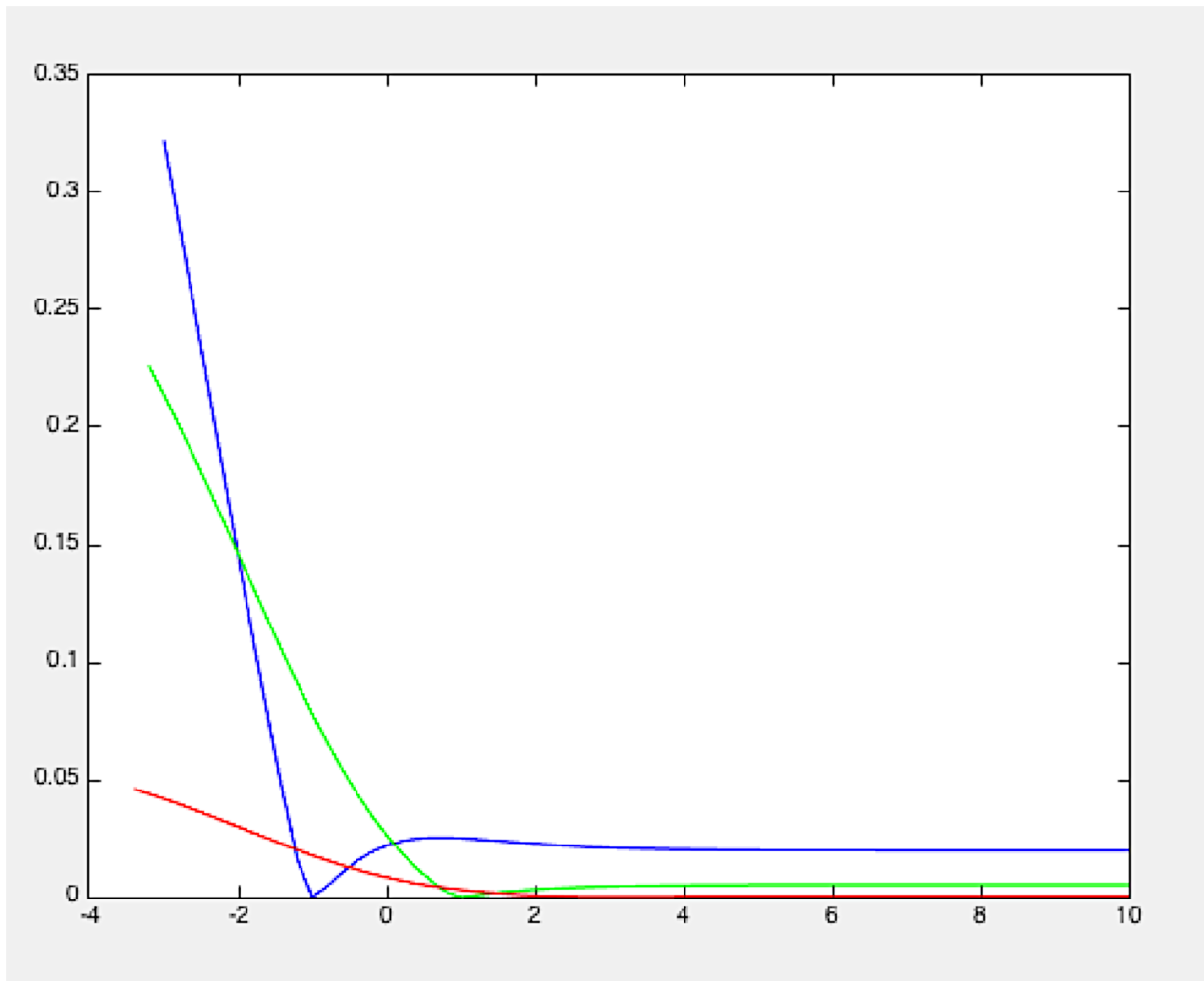


Abbildung 3-30: Norm längs der Kurve durch den Raum

Die Norm, dargestellt in Abhängigkeit von ΔG_2^0 entlang der in Abbildung 3-29 gezeigten Kurven, wobei die in der Graphik rechts hinten liegenden Äste betrachtet wurden. Der prinzipielle Verlauf ist bei allen drei Datensätzen gleich entlang der ΔG_2^0 -Achse, zuerst ein steiler Abfall der Norm zum Tal hin, dann ein Anstieg und schließlich ein erneutes leichtes Abfallen, das sich allerdings einem konstanten Niveau annähert, das bei den drei Datensätzen unterschiedlich ist. Je weniger sich die Besetzungszahlen mit dem Druck ändern, desto weniger stark ist das Minimum ausgeprägt.

Untersuchung des Parameterraums bei Drei- und Vierzustandssystemen

Die Untersuchung des Parameterraums bei Drei- und Vierzustandssystemen gestaltete sich insgesamt schwieriger, da durch die größere Anzahl an Dimensionen und Symmetrien die bei nur zwei Zuständen durchgeführte Rasterung und bildliche Darstellung nicht möglich ist. Dass auch hier der Parameterraum von Linien, an denen die Norm besonders kleine Werte hat, durchzogen wird, wird in Kapitel 3.4.2 indirekt gezeigt.

Um dem Ziel, geeigneten Anfangsbedingungen für das implementierte Levenberg-Marquardt-Verfahren näherzukommen, wurde die Verteilung der Punkte im Raum untersucht, die sich

jeweils durch Permutation der Benennungen der Zustände für ein betrachtetes System ergibt. Sie zeigt, dass die Raumdiagonalen mit gleichen Werten für alle ΔG_m^0 im $\Delta G_2^0 - \Delta G_3^0$ - Raum bei Dreizustandssystemen beziehungsweise im $\Delta G_2^0 - \Delta G_3^0 - \Delta G_4^0$ - Raum bei Vierzustandssystemen diejenigen Achsen durch den Raum sind, denen die gesuchten Minima am nächsten kommen (Abbildung 3-31) und deswegen als Startpunkte für den Algorithmus ideal sind. Besonders nah an dieser Geraden sind die gesuchten Punkte im negativen Bereich.

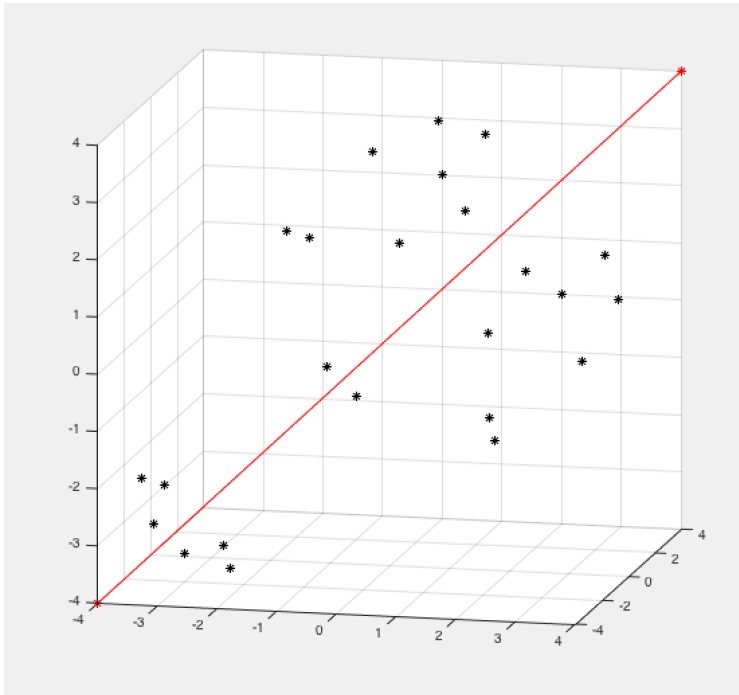


Abbildung 3-31: Minima im Vierzustandssystem

Der $\Delta G_2^0 - \Delta G_3^0 - \Delta G_4^0$ - Raum bei einem Vierzustandssystem: Die rote Diagonale ist die Linie durch den Raum, bei denen alle drei betrachteten Variablen den gleichen Wert annehmen. Im Vergleich zu anderen Raumdiagonalen kommen dieser Linie die durch Vertauschung der Benennungen erzeugten Raumpunkte, die alle die gleiche Besetzung der Zustände repräsentieren, am nächsten, wodurch sich auf ihr liegende Punkte als Startpunkte für den Algorithmus nach Levenberg-Marquardt anbieten.

3.3.2 Geeignete Wahl der Startpunkte für Algorithmus 1 und Funktionstest

Wahl der Startpunkte

Aus den Ergebnissen des letzten Kapitels lassen sich mehrere Folgerungen ableiten:

- Als Startpunkte für die ΔG_m^0 , $m = 2, \dots, M$, eignen sich Kombinationen, die auf der in Abbildung 3-31 eingezeichneten roten Achse liegen.
- Da im negativen Bereich dieser Achse die gesuchten Minima im Raum näher kommen, muss nur dieser Bereich abgesucht werden (siehe Abbildung 3-31).
- Da der Algorithmus vom Startpunkt aus einen Weg auf eine Linie kleiner Norm durch den Raum suchen wird, ist es wichtig, durch eine Vielzahl von verschiedenen Anfangspunkten den Ort, an dem er auf der Linie landet, zu variieren, da sich gezeigt hat, dass es entlang der Linie Bereiche gibt, von denen aus die Norm weg vom eigentlich gesuchten Parametersatz kleiner wird (siehe Abbildung 3-30).

Verschiedene weitere Versuche an Testdatensätzen bestätigten diese Folgerungen und ließen folgende Wahl der Anfangsparameter favorisieren:

- Variation der Parameter ΔG_m^0 für $m = 2, \dots, M$ von -11 bis 0 mit einer Schrittweite von 0,5, wobei jedoch bei der Untersuchung von mehr als zwei Zuständen zusätzlich $\Delta G_m^0 = \Delta G_n^0$ für alle $n, m = 2, \dots, M$ gilt
- Zu jedem Wert ΔG_2^0 wird ΔV_2^0 mit 0,005 initialisiert, bei mehr als zwei Zuständen werden die weiteren ΔV_n^0 mit $\Delta V_n^0 = \Delta V_{n-1}^0 + 0,005$ gewählt
- Δb_m^0 wird jeweils mit Null initialisiert
- Die Initialisierung der Matrix A , die die chemischen Verschiebungen der einzelnen Zustände für alle Aminosäuren enthält, erfolgt wie in Kapitel 2.4.1 beschrieben

Auf der Suche nach geeigneten Startpunkten im Parameterraum für das Levenberg-Marquardt-Verfahren wurden alternativ dazu auch verschiedene systematische Rasterungen in allen Dimensionen getestet, deren Gitterpunkte als Anfangsbedingungen verwendet wurden. Außerdem wurde versucht, den zu überprüfenden Orbit zuerst grob zu rastern und zu den Punkten direkt mit Algorithmus 2 die Norm zu berechnen, um dann in mehreren Stufen - es wurde ein Versuch mit 20 Stufen unternommen - immer den Raum um die gefundene kleinste Norm weiter zu analysieren. Im Anschluss daran wurde vom so gefundenen idealsten Rasterpunkt aus Algorithmus 1 aus gestartet. Diese Methode brachte jedoch nur dann gute Ergebnisse, wenn man bereits im ersten Schritt eine sehr kleine Rastergröße gewählt hatte, wodurch sie für mehr als zwei Zustände aufgrund der großen Anzahl an Dimensionen unbrauchbar wurde.

Testergebnisse

Um die Funktionsweise von Algorithmus 1 mit den eben beschriebenen Anfangsbedingungen für die Parameter der Gibbs-Energiedifferenzen zu überprüfen, wurden Testdatensätze generiert, wobei die Werte für ΔG_2^0 zufällig aus einem Bereich von 0 bis 10, die für ΔV_2^0 aus einem Bereich von -0,05 MPa⁻¹ bis +0,05 MPa⁻¹ und die Werte für Δb_2^0 aus einem Bereich von -0,0005 MPa⁻² bis +0,0005 MPa⁻¹ gewählt wurden.

Da bei Populationsverteilungen der Zustände, deren Änderungen mit dem Druck sehr klein sind und deswegen auch nur minimale Änderungen der chemischen Verschiebung bei Druckvariation resultieren, wurde zunächst nahe einer Schwelle gesucht, ab der die Änderungen zu klein sind, um das Minimum zu identifizieren und der Algorithmus an seine

numerische Grenze stößt. Dabei wurde festgestellt, dass bei Populationsverteilungen, bei denen ein Zustand eine Änderung von weniger als 2% erfährt, nur noch für weniger als 10% der Datensätze der Algorithmus erfolgreich eingesetzt werden kann. Daher wurden Datensätze mit solchen Populationsverteilungen vor dem Funktionstest aussortiert.

Für Zwei-, Drei- und Vierzustandssysteme wurden jeweils 100 Datensätze ohne statistische Fehler mit Algorithmus 1 untersucht. Die Testergebnisse sind in Tabelle 3-6 zusammengestellt. Wie zu erwarten, nimmt der Anteil der korrekt identifizierten Minima mit der Anzahl der Zustände ab, da die Anzahl der zu optimierenden Parameter ansteigt.

Die erhaltenen Ergebnisse erscheinen auf den ersten Blick ernüchternd, da in vielen Fällen das Minimum nicht korrekt identifiziert werden konnte. Eine stichprobenartige Analyse der Datensätze, bei denen der Algorithmus nicht die richtige Lösung lieferte, zeigte aber, dass in diesen Fällen Parameterkombinationen auftraten, bei denen einer oder mehrere Zustände sich kaum ändernde Besetzungszahlen oder sehr schwache Populationen über den gesamten betrachteten Druckbereich aufwiesen.

Tabelle 3-6: Testergebnisse mit Algorithmus 1 an idealen Systemen ohne statistische Fehler

Getestet wurden zufällig erzeugte Zwei-, Drei- und Vierzustandssysteme, bei denen die Änderung der korrigierten chemischen Verschiebung im untersuchten Druckbereich mindestens 2% betrug.

Anzahl der Zustände	Anteil der korrekt identifizierten Minima
2	88%
3	74%
4	64%

3.4 Untersuchung des Einflusses statistischer Fehler

Im Folgenden wurde untersucht, welchen Einfluss die statistischen Fehler der korrigierten chemischen Verschiebungen, die experimentell gewonnenen Daten zwangsläufig anhaften, auf den Parameterraum und vor allem auf die aus dem Algorithmus erhaltenen gesuchten Besetzungszahlen der Zustände, auf die Gibbs-Energien und die in der Matrix A zusammengefassten chemischen Verschiebungen haben.

3.4.1 Einfluss auf die Funktionsweise von Algorithmus 1 und die Größe der Norm

Untersuchte Datenreihen

Analog zu der in Kapitel 3.2.2 beschriebenen idealen Datenreihe wurden für Zwei-, Drei- und Vierzustandssysteme jeweils drei verschiedene Datenreihen erzeugt, die sich durch die Größe ihrer statistischen Fehler (s , m und l) unterschieden. Jede dieser neun Datenreihen enthielt 50 Testdatensätze, von denen schon während der Erzeugung die aussortiert worden waren, bei denen einer der beiden Zustände über den gesamten Messbereich eine Populationsgröße von weniger als 2% hatte.

Suche nach dem mathematischen Minimum

Bei Datensätzen mit statistischen Fehlern muss die Überprüfung von Algorithmus 1 in zwei Schritten geschehen. Da angenommen werden muss, dass durch die Fehler auch das Minimum im Parameterraum verschoben wird, muss erst bekannt sein, wo sich der neue gesuchte Raumpunkt befindet. Um diesen zu ermitteln, wurde für alle erzeugten Datensätze Algorithmus 1M benutzt, der als Startpunkte für die Suche mit dem Levenberg-Marquardt-Algorithmus den erzeugenden Parametersatz des jeweiligen Systems verwendet.

Funktionstest

Da nun bekannt war, wo sich die mathematischen Minima (vermutlich) befinden, konnte Algorithmus 1 auf die Datensätze angewandt werden. Dabei zeigte sich jedoch, dass bei mehreren Datensätzen bei Dreizustandssystemen mit Fehlergröße l und Vierzustandssystemen mit Fehlergröße m oder l mit Algorithmus 1 Minima mit einer kleineren Norm gefunden wurden als mit Algorithmus 1M. Somit ist für diese Datenreihen der durchgeführte Funktionstest nicht aussagekräftig. Die übrigen Ergebnisse sind in Tabelle 3-7 zusammengestellt.

Ein Vergleich mit Tabelle 3-6, die die Anteile der korrekt identifizierten Minima der ohne statistische Fehler generierten Daten enthält, zeigt die Tendenz, dass sich statistische Fehler positiv auf die Fähigkeit des Algorithmus 1 auswirken könnten, das Minimum zu finden.

Tabelle 3-7: Vergleich der Anteile korrekt identifizierter Minima von Algorithmus 1

Die Datenreihen unterscheiden sich durch die Anzahl ihrer Zustände beziehungsweise durch unterschiedlich große statistische Fehler. Für die mit * gekennzeichneten Datenreihen konnten keine Aussagen gemacht werden.

Anzahl der Zustände	Trefferquote bei		
	Fehler s	Fehler m	Fehler l
2	94%	92%	86%
3	78%	76%	_*
4	62%	_*	_*

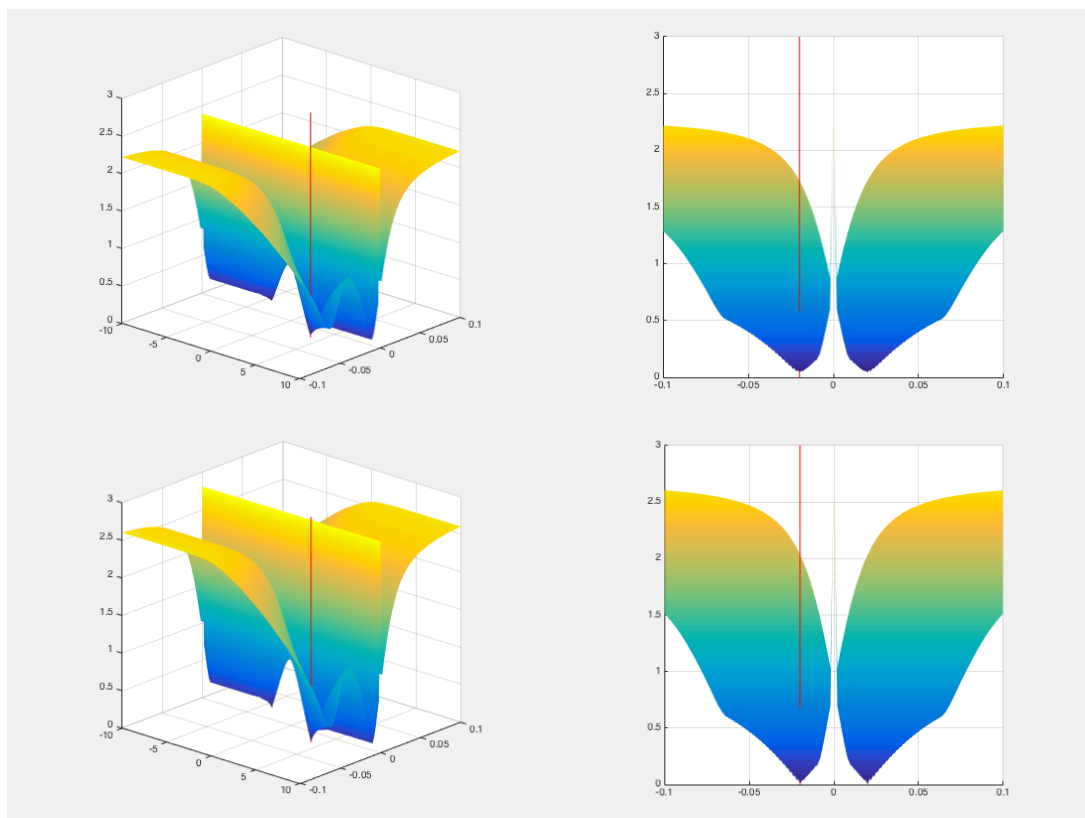


Abbildung 3-32: Untersuchung von Zweizustandssystemen mit Algorithmus 2

Dargestellt wurde zweimal das identische System, einmal mit statistischen Fehlern der Größe l (oben) und einmal ohne Fehler (unten). Durch die Fehler werden die Gräben kleiner Norm breiter und weniger tief. (Erzeugung der Darstellung analog Abbildung 3-24)

Eine Erklärung dafür lieferten Untersuchungen von Zweizustandssystemen mit Algorithmus 2, bei denen der Parameterraum fein gerastert wurde. Die erhaltenen Bilder erwiesen sich im Gegensatz zu denen der idealen Datensätze als weniger scharf, weil die Höhendifferenzen weniger stark ausgebildet waren, und zugleich die Gräben durch den Parameterraum breiter waren, wobei Letzteres für die Minimumssuche eher günstig erscheint (Abbildung 3-32).

Die Norm am mathematischen Minimum

An den generierten Daten wurde auch getestet, ob die in Kapitel 2.4.4 aufgestellten Formeln für die Norm am mathematischen Minimum $\text{Norm}_{\text{math}}$ und die Norm Norm_{erz} , die sich aus der systemerzeugenden Fehlermatrix R berechnet, für die untersuchten Systeme korrekt sind. Zur Überprüfung wurde das Verhältnis zwischen der mathematischen und der erzeugenden Norm berechnet (Tabelle 3-8). Dabei zeigte sich, dass, wie theoretisch vorhergesagt, diese Größe umso kleiner wird, je mehr Zustände die betrachteten Systeme haben und dass außerdem keine Abhängigkeit von der Größe der statistischen Fehler vorliegt.

Tabelle 3-7: Mittelwert und Standardabweichung des Verhältnisses der Normen $\text{Norm}_{\text{math}}$, die an den Orten der mathematischen Minima gefunden werden, zu den Normen Norm_{erz} der Fehlermatrizen R

Anzahl der Zustände	$\text{Norm}_{\text{math}}/\text{Norm}_{\text{erz}}$
2	$0,89 \pm 0,02$
3	$0,83 \pm 0,02$
4	$0,76 \pm 0,02$

Die Ergebnisse stimmen gut mit den Werten überein, die sich anhand Gleichung 25 und Gleichung 26 für die hier betrachteten Systeme aus den chemischen Verschiebungen von 80 Atomen an 10 Messpunkten berechnen lassen und die konkret die theoretischen Zahlenwerte 0,8923, 0,8322 und 0,7673 für Zwei- Drei- und Vierzustandssysteme liefern.

Bei der Minimumssuche vergleicht Algorithmus 1 allerdings nicht zwischen der Norm am mathematischen Minimum $\text{Norm}_{\text{math}}$ und der durch die Fehlermatrizen berechneten Norm Norm_{erz} , da diese nicht bekannt ist. Vielmehr werden mit dem Verfahren auch gleichzeitig die Elemente der Matrix A , die korrigierten chemischen Verschiebungen der Zustände aller Aminosäuren, mitoptimiert, so dass durch deren Anpassung eine Norm am Raumpunkt der erzeugenden Parameter entsteht, die kleiner ist als die eben betrachtete und aus der Fehlermatrix berechnete. Auch die Größe dieser Norm Norm_{opt} lag, wie theoretisch erwartet,

bei allen untersuchten Systemen zwischen denen der beiden oben erfassten Normen und nur geringfügig über $\text{Norm}_{\text{math}}$.

3.4.2 Wirkung auf die Gibbs-Energiedifferenzen und die Besetzungszahlen

Zweizustandssysteme

Um zu untersuchen, wo die mathematischen Minima liegen, die man statt des eigentlichen erzeugenden Parametersatzes im ΔG_2^0 - ΔV_2^0 - $\Delta b_2^{0'}$ -Raum findet, wurden zu verschiedenen Parametersets jeweils 200 verschiedene Fehlermatrizen erzeugt, auf die ideale Matrix D_{ideal} aufaddiert und mit Algorithmus 1M die Suche durchgeführt.

Dabei zeigte sich, dass die Entfernung zwischen den eigentlichen Parametern und denen des mathematischen Minimums mit der Größe der statistischen Fehler steigt. Ein überraschendes Ergebnis dabei war, dass die Minima nicht gleichmäßig um den Punkt im Raum verteilt liegen, der die eigentlichen Parameter repräsentiert, sondern sich entlang der Linie kleinster Norm durch den Parameterraum sammeln. Abbildung 3-33 zeigt für jeweils dasselbe erzeugende System (in den Graphiken mit einem roten Pfeil gekennzeichnet) die mathematischen Minima für 200 zufällig erzeugte Fehlermatrizen. Für das abgebildete Beispiel wurden die Parameter ΔG_2^0 gleich 1, ΔV_2^0 gleich $-0,02 \text{ MPa}^{-1}$ und $\Delta b_2^{0'}$ gleich $3 \cdot 10^{-5} \text{ MPa}^{-2}$ gewählt. Die drei Graphiken unterscheiden sich durch die Größe des mittleren statistischen Fehlers.

Stichprobenartige Untersuchungen mit verschiedenen Systemen zeigten, dass die Breite der Bereiche, in denen die Parameter ΔG_2^0 , ΔV_2^0 und $\Delta b_2^{0'}$ durch die Fehler um die eigentlichen systemerzeugenden Werte variieren, erheblich von der Lage des erzeugenden Systems im Parameterraum abhängt. Insbesondere lieferten Systeme mit Zuständen, deren Besetzungszahlen sich stark ändern, deutlich kleinere Schwankungen als Systeme mit weniger Variationen in den Populationsverhältnissen. Dies stimmt mit der Erwartung überein, da Erstere im Vergleich zu den anderen mit größeren nach der aminosäureabhängigen Druckkorrektur verbleibenden Änderungen der chemischen Verschiebungen einhergehen, und deswegen das Verhältnis zwischen statistischen Fehlern und den interessierenden Verschiebungsänderungen größer ist.

Abbildung 3-33: Mathematische Minima im Raum

Darstellung von jeweils 200 mathematischen Minima im Raum. Die Größe ΔG_2^0 ist nach rechts hinten aufgetragen, nach links hinten ΔV_2^0 und nach oben $\Delta \beta_2^{0'}$. Der rote Pfeil kennzeichnet den Ort des erzeugenden Parametersets:

$$\Delta G_2^0 = 1$$

$$\Delta V_2^0 = -0,02 \text{ MPa}^{-1}$$

$$\Delta \beta_2^{0'} = 3 \cdot 10^{-5} \text{ MPa}^{-2}$$

Für die drei Graphiken wurde die Größe des statistischen Fehlers unterschiedlich gewählt:

Oben: Fehler s

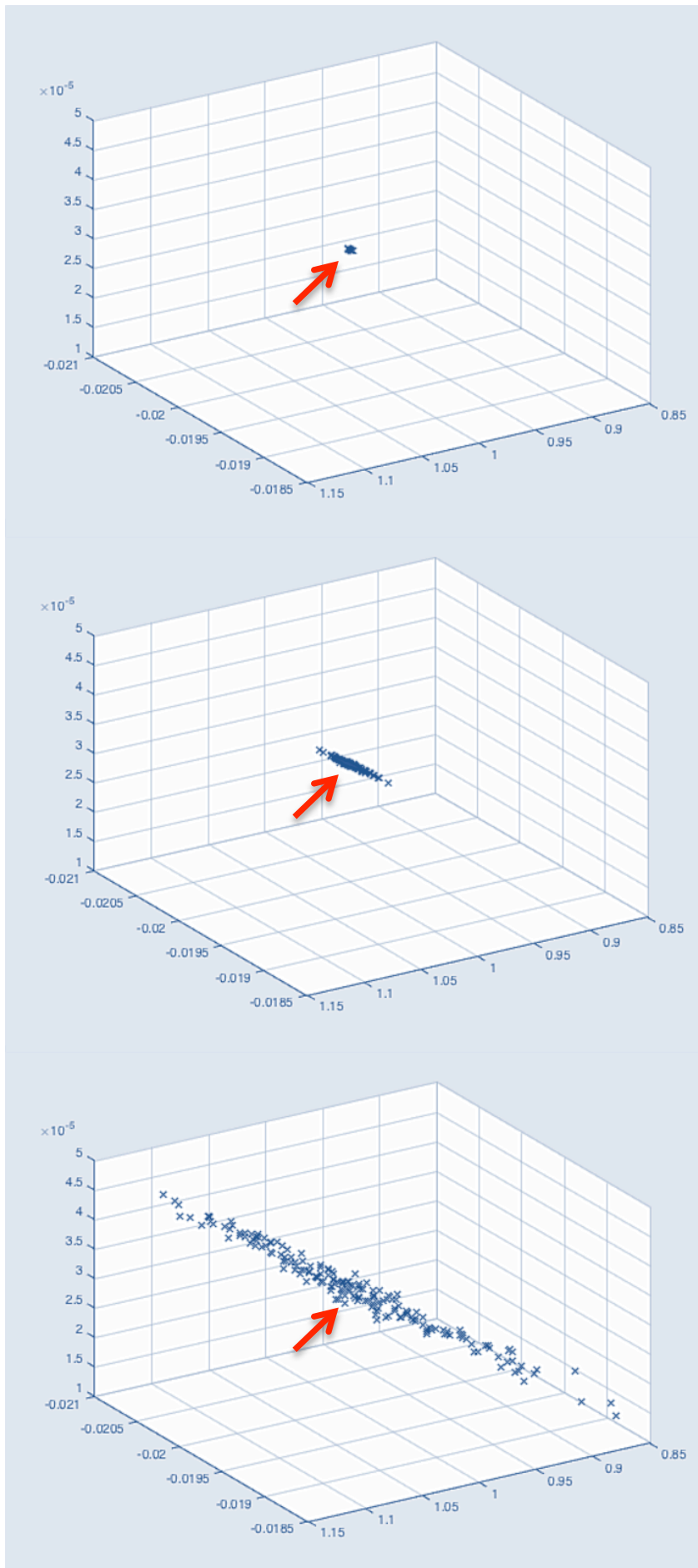
Mitte: Fehler m

Unten: Fehler l

Die mathematischen Minima sind nicht gleichmäßig um den Punkt im Raum verteilt, der das erzeugende System repräsentiert, sondern sie sammeln sich entlang der Kurve der minimalsten Norm durch den Raum.

Achsen:

ΔG_2^0 variiert von 0,85 bis 1,15 (nach rechts), ΔV_2^0 variiert von -0,021 MPa⁻¹ bis -0,0185 MPa⁻¹ (nach vorne), $\Delta \beta_2^{0'}$ von $1 \cdot 10^{-5}$ MPa⁻² bis $5 \cdot 10^{-5}$ MPa⁻² (nach oben)



Aus diesem Ergebnis folgt die Frage, inwieweit die Fehler die Besetzungszahlen der Zustände, die sich aus den verschobenen Parametern ergeben, ändern. Deswegen wurden für verschiedene ideale Datensätze ihre Besetzungszahlanteile gegen den Druck aufgetragen (in den Abbildungen 3-34 und 3-35 rot eingezeichnet). Dann wurden jeweils zehn Fehlermatrizen R einer bestimmten Größenordnung zufällig erzeugt, die mathematischen Minima der resultierenden Systeme mit Algorithmus 1M gesucht und die resultierenden Anteile an den Besetzungen mit denen des idealen Systems verglichen (in den Abbildungen schwarz eingezeichnet). Zwar sind die Kurven etwas gegenüber der idealen Kurve verschoben, doch sie verlaufen ansonsten annähernd parallel. Das bedeutet, dass die prozentualen Anteile zueinander leicht verschoben werden. Die Streuung erfolgt in beide Richtungen und die ideale Kurve wird durch die aus den Fehlern resultierenden eingeschlossen.

Weitergehende Berechnungen an Datensätzen, bei denen der Einfluss der Größe der Fehler systematisch untersucht wurde, lieferten zwei Ergebnisse: Zum einen entscheidet die Größe der Fehler darüber, wie weit die Kurven um den idealen Verlauf streuen, das heißt, werden größere statistische Fehler zur Erzeugung der Matrizen R angenommen, wird der Bereich um den idealen Verlauf breiter als bei kleineren Fehlern (Abbildung 3-34). Zum anderen ist aber dieser Bereich deutlich abhängig von der Lage des untersuchten Systems im Parameterraum. Insbesondere findet man wieder große Unterschiede zwischen Datensätzen, bei denen ein Zustand über den gesamten Druckbereich der höher populierte bleibt, und solchen Datensätzen, auf die dies nicht zutrifft. Bei erstgenannten streuen die Kurven der Besetzungszahlen viel breiter (Abbildung 3-34). Aufgrund dieser starken Abhängigkeit von der Lage des Systems ist es nicht möglich, allgemeingültigen Fehlergrenzen für die Parameter zu ermitteln.

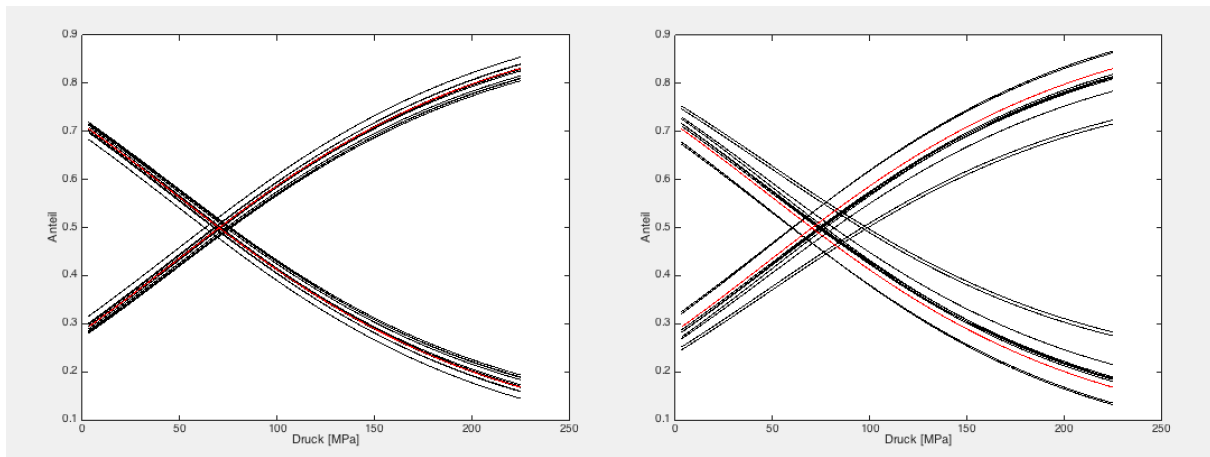


Abbildung 3-34: Vergleich der Besetzungszahlanteile in Abhängigkeit vom Druck

Rot aufgetragen die Anteile, die sich aus den das System erzeugenden Parametern ergeben ($\Delta G_2^0 = 0,8775$, $\Delta V_2^0 = -0,0139 \text{ MPa}^{-1}$ und $\Delta b_2^{0'} = -2,4678 \cdot 10^{-5} \text{ MPa}^{-2}$). Schwarz dazu die Besetzungszahlen, die sich aus den mathematischen Minima ergeben, die durch die Fehlermatrizen erzeugt werden. Links wurde für die Norm der Fehlermatrix ein Wert von 0,02 angesetzt, rechts 0,05. Es wurden jeweils 10 Fehlermatrizen zufällig erzeugt. Die entstehenden Kurven laufen parallel zu der idealen Besetzung und zeigen damit keinen wesentlich anderen Verlauf. Nur der Anteil der Zustände zueinander verändert sich leicht. Der entstehende Bereich schließt die rote Kurve ein und wird mit größer werdendem Fehler immer breiter.

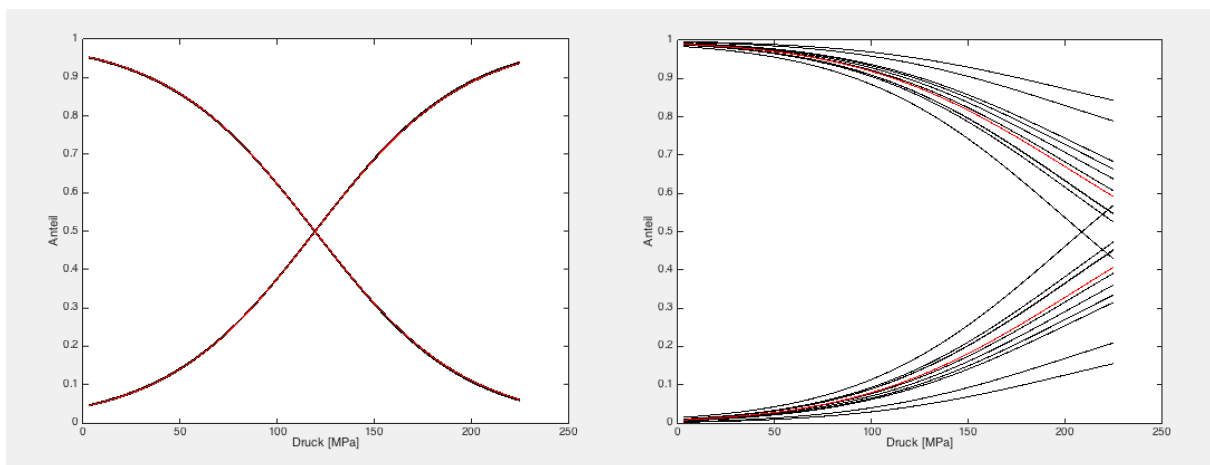


Abbildung 3-35: Einfluss der Fehler auf die Besetzungszahlanteile

Der Einfluss der Fehler auf die Besetzungszahlanteile ist stark von der Lage des erzeugenden Parametersets im Parameterraum abhängig. Für die beiden gezeigten Beispiele wurden jeweils 10 Fehlermatrizen zufällig erzeugt. Deren Norm war jeweils 0,05. Dargestellt sind jeweils die Anteile der Besetzungszahlen des idealen Systems (rot), schwarz dazu die Anteile, die durch die Veränderung aufgrund der Fehler entstehen.

Links: System mit $\Delta G_2^0 = 3,0255$, $\Delta V_2^0 = -0,0261 \text{ MPa}^{-1}$ und $\Delta b_2^{0'} = -1,4840 \cdot 10^{-6} \text{ MPa}^{-2}$. Die entstehenden Werte für ΔG_2^0 streuen zwischen 3,0732 und 3,0055.

Rechts: System mit $\Delta G_2^0 = 4,6838$, $\Delta V_2^0 = -0,0261 \text{ MPa}^{-1}$ und $\Delta b_2^{0'} = -6,0614 \cdot 10^{-5} \text{ MPa}^{-2}$.

Hier streuen die entstehenden Werte für ΔG_2^0 zwischen 5,7915 und 4,1377. Daran wird deutlich, dass sich Fehler gleicher Größe je nach Lage des erzeugenden Systems im Parameterraum sehr unterschiedlich stark bemerkbar machen.

Drei- und Vierzustandssysteme

Wie bei den Zweizustandssystemen wurden die Auswirkungen der statistischen Fehler auf die Besetzungszahlen bei Drei- und Vierzustandssystemen untersucht, indem datensatzweise jeweils mehrere Fehlermatrizen generiert und idealen Datensätzen aufaddiert wurden. Dazu wurden die mathematischen Minima bestimmt, daraus die Besetzungszahlen in Abhängigkeit vom Druck berechnet und mit denen des idealen Systems verglichen. Die Abbildungen 3-36 und 3-37 zeigen exemplarisch zwei Beispiele für Drei- beziehungsweise Vierzustandssysteme. In Rot sind die eigentlichen Besetzungszahlkurven der Zustände dargestellt, die grauen entsprechen denen durch die Fehler verschoben. Es zeigt sich das gleiche Verhalten wie bei Zweizustandssystemen.

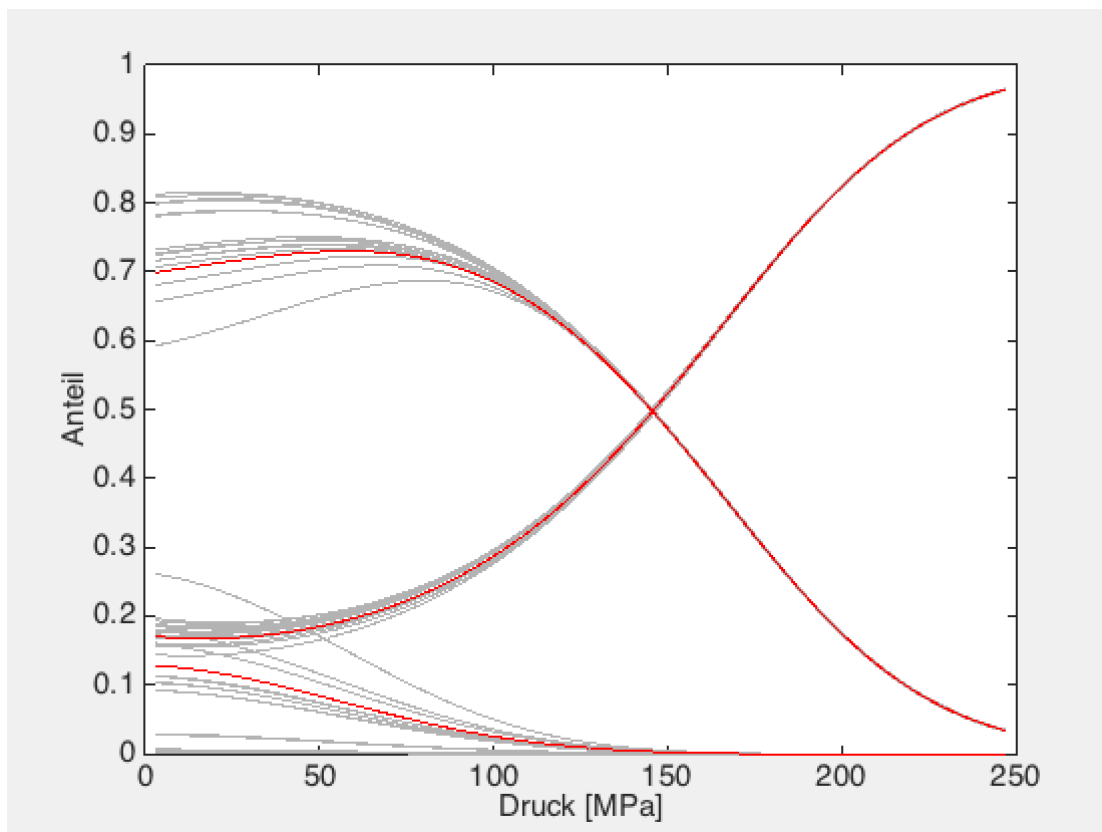


Abbildung 3-36: Vergleich der Besetzungszahlanteile in Abhängigkeit vom Druck bei einem Dreizustandssystem

Rot aufgetragen die Anteile, die sich aus den das System erzeugenden Parametern $\Delta G_2^0 = 1,4092$, $\Delta V_2^0 = 0,0036 \text{ MPa}^{-1}$, $\Delta b_2^{0'} = 1,8879 \cdot 10^{-4} \text{ MPa}^{-2}$, $\Delta G_3^0 = 1,6919$, $\Delta V_3^0 = 0,0038 \text{ MPa}^{-1}$ und $\Delta b_3^{0'} = -2,5826 \cdot 10^{-4} \text{ MPa}^{-2}$ ergeben. Grau dazu die Besetzungszahlen der Systeme, die durch die Fehlermatrizen (Fehlergröße 1) erzeugt werden.

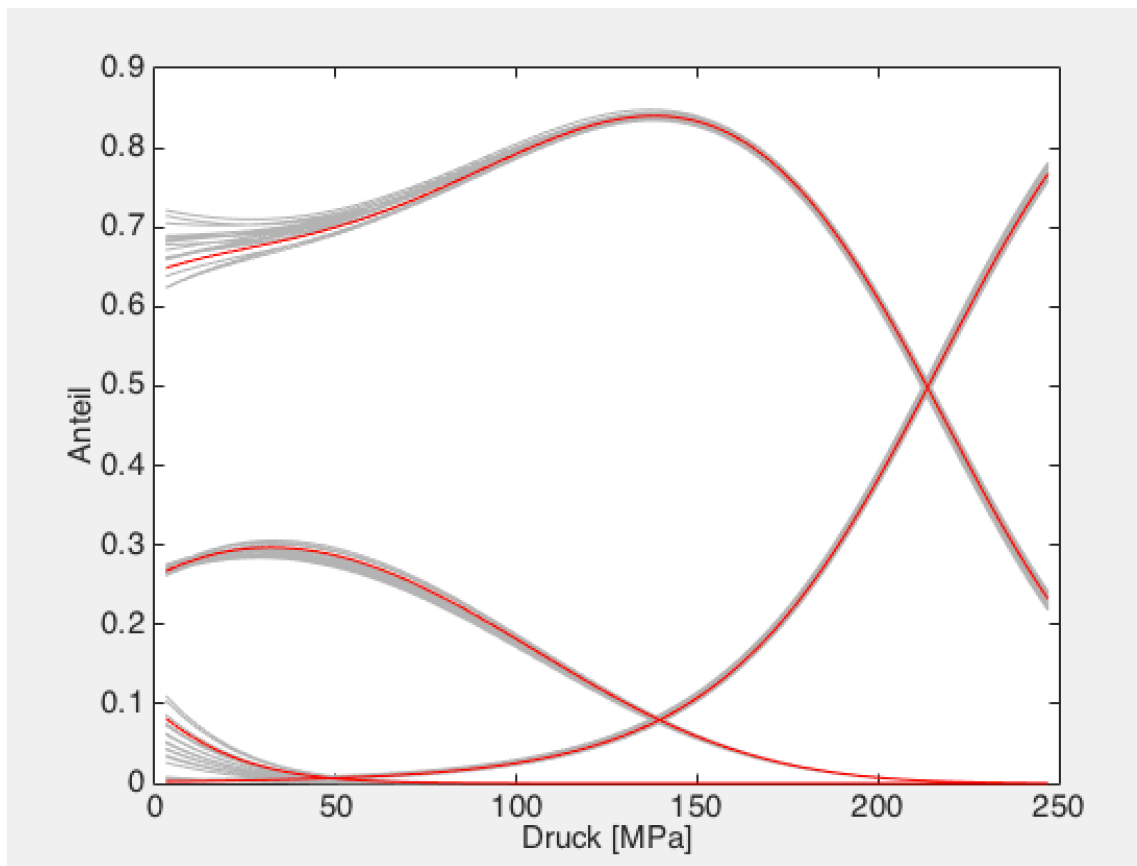


Abbildung 3-37: Vergleich der Besetzungszahlanteile in Abhängigkeit vom Druck bei einem Vierzustandssystem

Rot aufgetragen die Anteile, die sich aus den das System erzeugenden Parametern ergeben, in Grau dazu Besetzungszahlen, die sich aus den mathematischen Minima ergeben, die durch die Fehlermatrizen (Fehlergröße m) erzeugt werden. Es wurden folgende Parameter verwendet:

$\Delta G_2^0 = 0,8869$, $\Delta V_2^0 = -0,0055 \text{ MPa}^{-1}$ und $\Delta b_2^{0'} = -2,3760 \cdot 10^{-4} \text{ MPa}^{-2}$, $\Delta G_3^0 = 2,0702$, $\Delta V_3^0 = 0,0458 \text{ MPa}^{-1}$ und $\Delta b_3^{0'} = -4,6417 \cdot 10^{-4} \text{ MPa}^{-2}$ sowie $\Delta G_4^0 = 5,5801$, $\Delta V_4^0 = -0,0184 \text{ MPa}^{-1}$ und $\Delta b_4^{0'} = 7,6775 \cdot 10^{-5} \text{ MPa}^{-2}$

Die mathematischen Abhängigkeiten der Parameter in der Funktion Norm legen nahe, dass auch bei Drei- und Vierzustandssystemen diese als Linien durch den Raum sichtbar werden könnten, indem man die Koordinaten der durch die Fehlermatrizen gefundenen erzeugten verschobenen Minima gegeneinander graphisch darstellt. Abbildung 3-36 zeigt ein Beispiel eines Dreizustandssystems, bei dem die Parameter genauso wie bei dem in Abbildung 3-38 gezeigten System gewählt wurden. Der Übersichtlichkeit halber wurde hier eine kleinere Fehlergröße (m) gewählt, mit der 50 Fehlermatrizen erzeugt wurden. Daraus geht hervor, dass die Größen ΔG_2^0 , ΔV_2^0 und $\Delta b_2^{0'}$ sehr empfindlich auf die statistischen Fehler reagieren und über weite Bereiche des Parameterraums streuen, aber aufgrund der Abhängigkeit voneinander trotzdem Besetzungszahlen in Abhängigkeit vom Druck resultieren, die den eigentlichen Populationsverhältnissen nahekommen.

Durch Projektion in verschiedene Ebenen des Parameterraums konnten indirekt die Linien kleiner Norm sichtbar gemacht werden, die sich durch Rasterungen der vieldimensionalen Räume nur erahnen ließen.

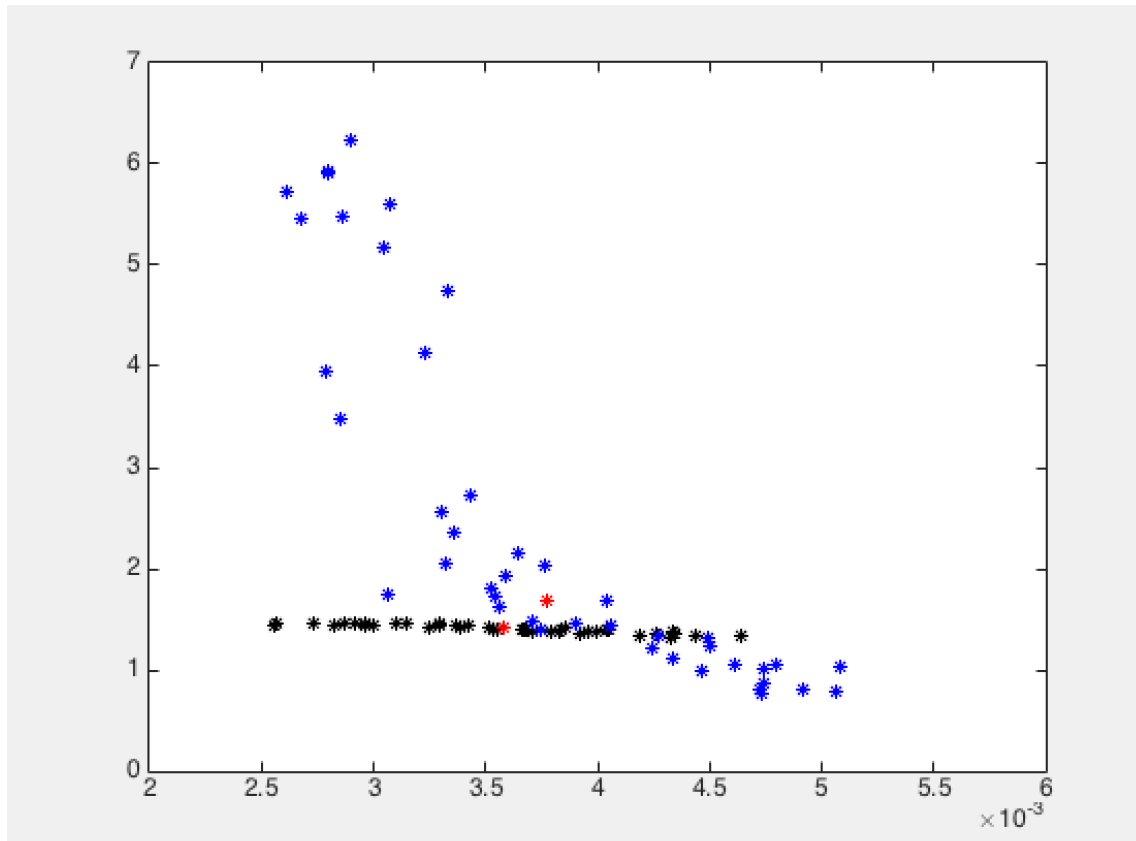


Abbildung 3-38: Abhängigkeit der Parameter bei einem Dreizustandssystem

Zu dem eigentlichen System wurden 50 Fehlermatrizen der Größe m erzeugt und die Minima dieser Systeme bestimmt. Dargestellt als Punkte sind in Schwarz ΔG_2^0 gegen ΔV_2^0 und in Blau ΔG_3^0 gegen ΔV_3^0 der Fehlersysteme. In Rot sind die Punkte des eigentlichen Systems eingezeichnet. Dessen Parameter wurden genauso gewählt wie bei dem in Abbildung 3-36 gezeigten Beispiel.

Zur Erzeugung dieser Graphik wurden zuerst jeweils die Projektionen der erhaltenen Raumpunkte in den ΔG_2^0 - ΔV_2^0 - Unterraum und den ΔG_3^0 - ΔV_3^0 - Unterraum vorgenommen und diese beiden Projektionen dann aufeinandergelegt.

3.4.3 Beeinflussung der chemischen Verschiebungen der Zustände

Überprüfung der Matrix A bei einem idealen System

Als Vorüberlegung wurde zunächst ein einfacher Test durchgeführt: Wie in Kapitel 3.3.1 wurde ein Zweizustandssystem mit $\Delta b_2^{0'}$ gleich Null betrachtet und die Norm mit Algorithmus 2 in der Ebene des Parameterraums, in der $\Delta b_2^{0'}$ gleich Null ist, in Abhängigkeit von den Parametern ΔG_2^0 und ΔV_2^0 berechnet. Bei diesem Verfahren wird für jeden betrachteten Punkt der Ebene jeweils die Matrix A so bestimmt, dass die Norm minimal ist.

Dabei wird der Bereich, in denen sich die Werte der Matrixelemente bewegen dürfen, nicht beschränkt.

Wie beschrieben wurden bei der Generierung der Testdaten zur Erstellung der Matrix A Werte für die chemischen Verschiebungen der einzelnen Zustände aller Atome zufällig aus einem Bereich zwischen 8,00 ppm und 9,2 ppm gewählt. Um nun zu testen, in welchen Gebieten der Ebene die Matrixelemente sich noch in der Nähe dieses Bereichs befinden, wurde eine obere Grenze von 10,2 ppm und eine untere Grenze von 7,00 ppm definiert, innerhalb derer sich die chemischen Verschiebungen bewegen dürfen. Abbildung 3-39 zeigt die untersuchte Ebene für einen Datensatz mit den Parametern ΔG_2^0 gleich 3, ΔV_2^0 gleich $0,01 \text{ MPa}^{-1}$ und $\Delta b_2^{0'} = 0 \text{ MPa}^{-2}$. Gebiete, in denen die Elemente der optimierten Matrix A außerhalb des erlaubten Bereiches liegen, wurden dunkelblau eingefärbt. Untersuchungen anderer Parametersätze mit und ohne Fehler zeigten ein vergleichbares Verhalten. Insbesondere die nahezu konstant bleibende Norm im Graben bei immer größer werdenden Werten für ΔG_2^0 wird dadurch realisiert, dass die Elemente von A verbotene Werte annehmen.

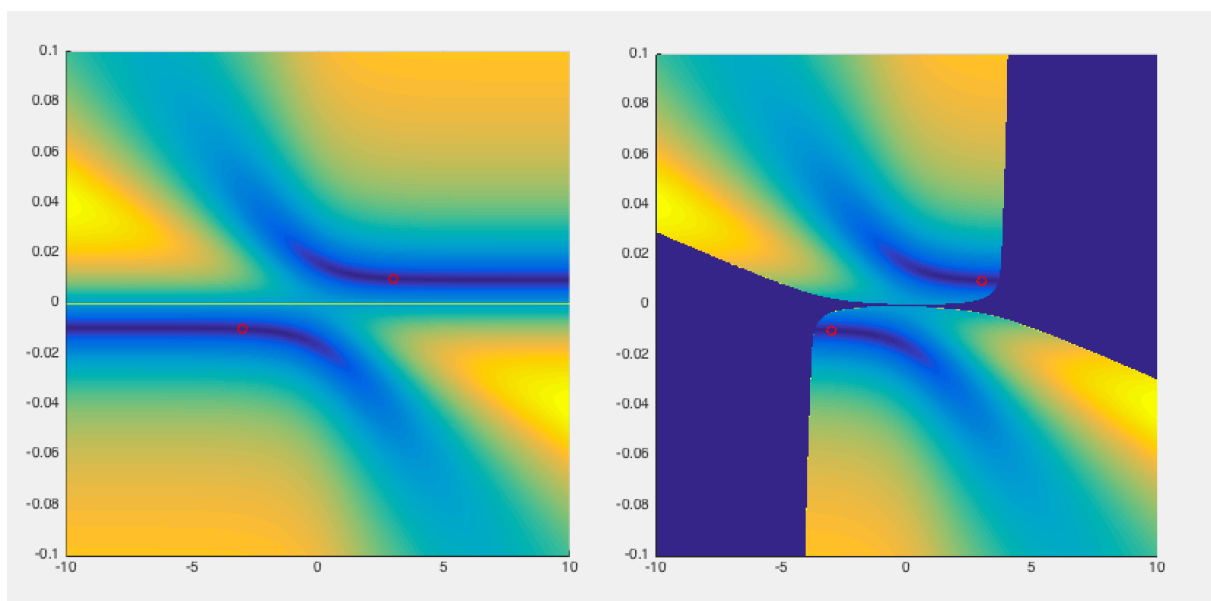


Abbildung 3-39: Untersuchung der Matrixelemente von A für einen mit $\Delta G_2^0 = 3$, $\Delta V_2^0 = 0,01 \text{ MPa}^{-1}$ und $\Delta b_2^{0'} = 0$ erzeugten Datensatz

Die linke Graphik zeigt die Norm als Höhenplot über der $\Delta b_2^{0'} = 0$ - Ebene, rot gekennzeichnet ist das Minimum, das den erzeugenden Parametern entspricht. Im Vergleich dazu wurde in der rechten Graphik das Gebiet, bei dem die Matrixelemente von A bei optimierter Norm im verbotenen Bereich liegen, dunkelblau gefärbt. Nach rechts ist ΔG_2^0 (dimensionslos), nach oben ΔV_2^0 (in MPa^{-1}) aufgetragen.

Einfluss der statistischen Fehler

Des Weiteren wurde stichprobenartig für verschiedene Datensätze mithilfe verschiedener Fehlermatrizen getestet, wie sich die Werte der Elemente der Matrix A verhalten. Dabei zeigten vor allem die korrigierten chemischen Verschiebungen der Atome besonders große Abweichungen von den eigentlichen Werten, die für alle Zustände sehr ähnliche Werte aufwiesen.

Außerdem ergab sich, dass die korrigierten chemischen Verschiebungen für Zustände mit kleineren Besetzungen oder wenig Änderungen über den gesamten Druckbereich deutlich weniger stabil gegenüber statistischen Fehlern waren als diejenigen mit großen Besetzungszahländerungen oder Besetzungszahlen. Die Abbildungen 3-40 und 3-41 zeigen zwei Beispiele verschiedener Dreizustandssysteme, deren Populationsverhältnisse mit dem Druck - jeweils im unteren Teil der Graphik dargestellt - erheblich voneinander abweichen. Die korrigierten chemischen Verschiebungen des untersuchten Atoms waren bei beiden Systemen gleich gewählt worden und hatten die Werte 8,89 ppm (schwarz), 8,61 ppm (rot) und 8,26 ppm (blau). Deutlich zu sehen ist, dass in Abbildung 3-41 die Werte der schwächer besetzten Zustände über einen sehr großen Bereich streuen. Die Größe der angenommenen mittleren Fehler für jeden Druckpunkt entsprach in etwa der aus HSQC-Spektren gewonnenen chemischen Verschiebungen für Amidprotonen.

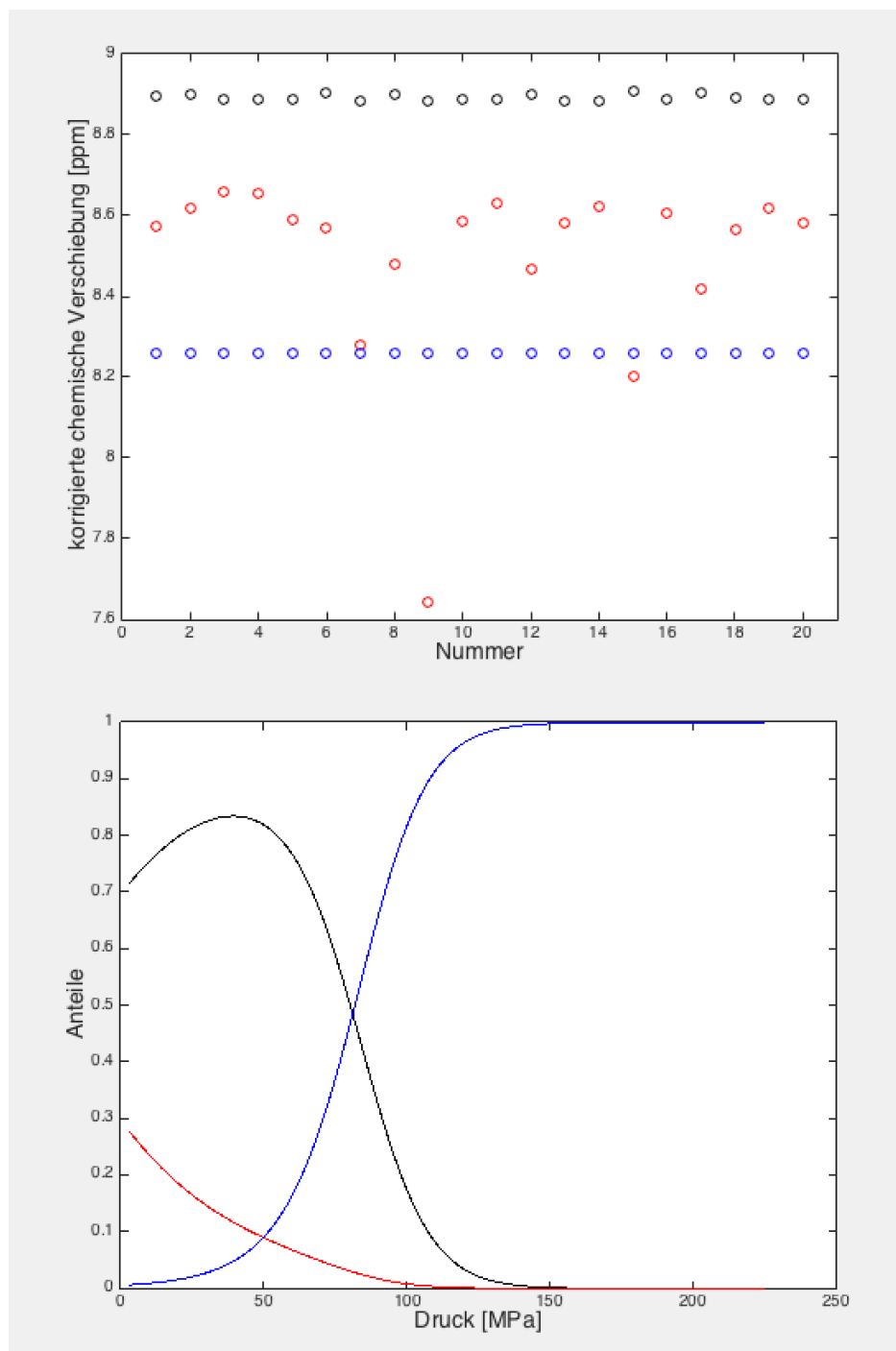


Abbildung 3-40: Einfluss statistischer Fehler auf die optimierten korrigierten chemischen Verschiebungen

Oben: Die Variation der korrigierten chemischen Verschiebungen (nach oben aufgetragen) eines einzelnen Atoms im Dreizustandssystem, die aus Systemen, erzeugt durch 20 verschiedene Fehlermatrizen, durch Optimierung der Matrix A gewonnen wurden (nach rechts die Nummer der Fehlermatrix).

Unten: Populationsanteile des Systems

Am weitesten streuen die Werte des am schwächsten besetzten Zustands.

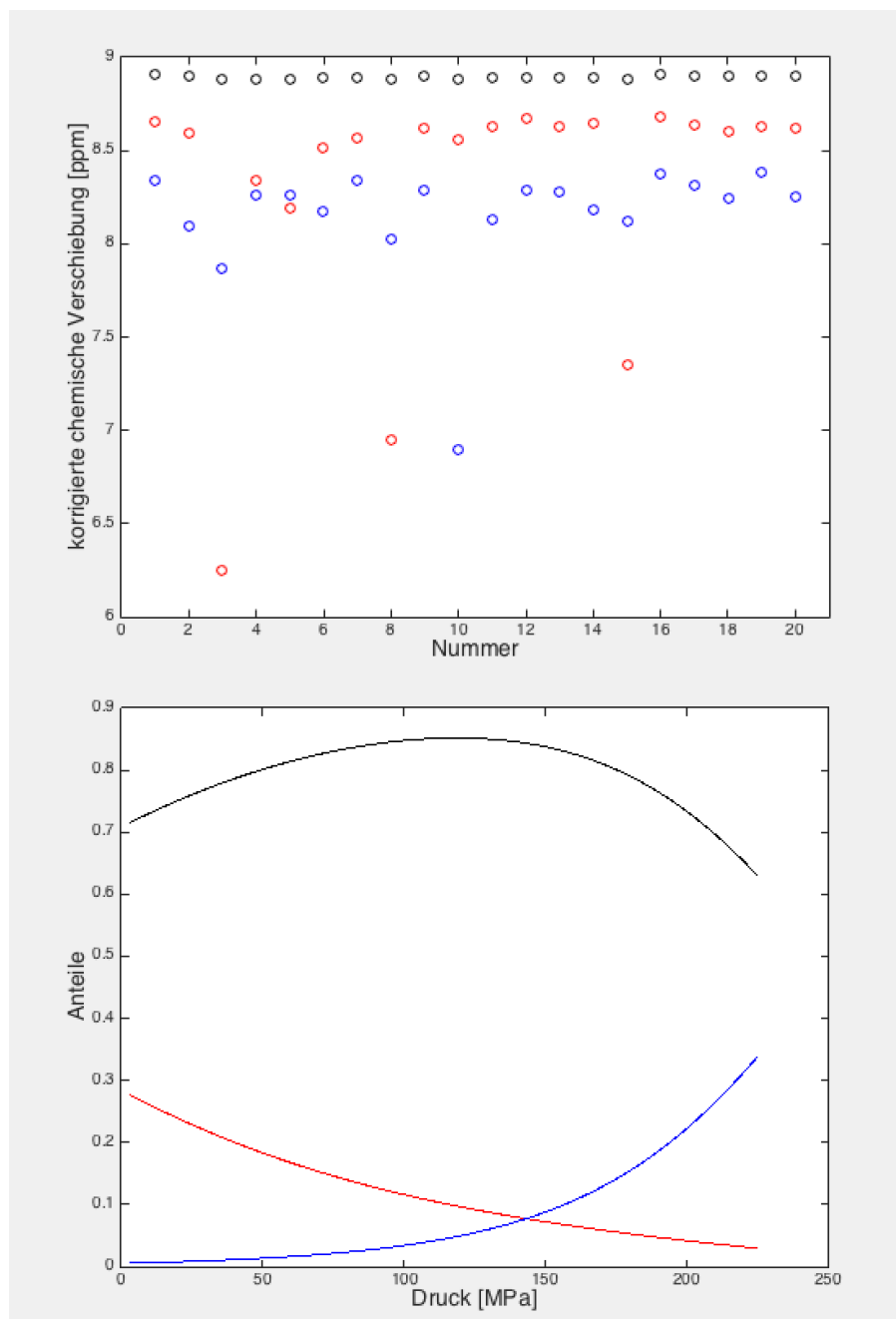


Abbildung 3-41: Einfluss statistischer Fehler auf die optimierten korrigierten chemischen Verschiebungen

Im Gegensatz zu Abbildung 3-40 wurden hier Populationsverhältnisse gewählt, bei denen zwei Zustände im gesamten Druckbereich deutlich schwächer besetzt sind als der dritte. Daraus ergibt sich eine deutlich größere Beeinflussung der Fehler auf die korrigierten chemischen Verschiebungen.

3.5 Bestimmung eines Lösungsbereichs um das mathematische Minimum

Die bis jetzt gezeigten Ergebnisse legen nahe, dass man mit Algorithmus 1 zwar die Möglichkeit hat, das mathematische Minimum des Systems zu finden, aber keine einfache Abschätzung aus der Norm und der Anzahl der Zustände gefunden werden konnte, den Bereich um dieses gefundene Minimum im Parameterraum einzugrenzen, in dem die erzeugenden Parameter des Systems mit großer Sicherheit liegen. Vielmehr wurde festgestellt, dass der Einfluss der Fehler stark von der Lage im Parameterraum abhängt und deswegen für jedes zu untersuchende System eine gesonderte Abschätzung gemacht werden muss.

3.5.1 Untersuchung der Symmetrie der Fehlerbereiche

Es ist daher zu überprüfen, inwieweit sich der Parameterraum bezüglich der Fehlerbereiche symmetrisch verhält, die man um den erzeugenden Parametersatz und um den Parametersatz des mathematischen Minimums legen kann, wenn man Fehler gleicher Größe annimmt. Insbesondere stellt sich die Frage, ob ein so generierter Bereich um das mathematische Minimum den erzeugenden Parametersatz enthält. Um dies zu untersuchen, wurde folgendes Vorgehen gewählt und für Zwei-, Drei- und Vierzustandssysteme mehrfach stichprobenartig getestet:

- Zufällige Wahl der Parameter $\Delta\mathbb{G}_m^0, \Delta\mathbb{V}_m^0$ und $\Delta\mathbb{b}_m^{0'}$ und einer Matrix A , woraus die ideale Matrix D_{ideal} generiert wird
- Wahl der Größe des Fehlers, mit der j Matrizen R_j zufällig erzeugt werden
- Generierung von j Systemen mit $D_j = D_{ideal} + R_j$, für die die mathematischen Minima gesucht werden
- Auswahl zweier mathematischer Minima, deren Parameter maximal weit von den Parametern des idealen Systems - $\Delta\mathbb{G}_m^0, \Delta\mathbb{V}_m^0$ und $\Delta\mathbb{b}_m^{0'}$ - entfernt sind. Diese beiden Minima sind charakterisiert durch ihre Parameter $\Delta\mathbb{G}_m^{01}, \Delta\mathbb{V}_m^{01}$ und $\Delta\mathbb{b}_m^{0'1}$ und eine Matrix A_1 beziehungsweise $\Delta\mathbb{G}_m^{02}, \Delta\mathbb{V}_m^{02}$ und $\Delta\mathbb{b}_m^{0'2}$ und eine Matrix A_2 , aus denen sich dann die Matrizen D^1 und D^2 berechnen lassen
- Erzeugung von jeweils k neuen Systemen durch Erstellunen von k Fehlermatrizen der gleichen Fehlergröße wie oben gewählt und Addieren zu D^1 beziehungsweise D^2 ; dann Bestimmen der mathematischen Minima dieser neuen Systeme
- Überprüfen, ob die durch diese Minima definierten Bereiche im Parameterraum die Parameter des idealen Systems beinhalten

Dabei stellte sich heraus, dass in allen untersuchten Fällen die Parameter des idealen Systems in dem Bereich zu finden waren, der sich durch die Lage der Punkte definieren lässt, die den mathematischen Minima der durch Fehlermatrizen erzeugten Systeme entsprechen, solange deren Anzahl k statistisch groß genug gewählt wurde. Dabei reichten für Zweizustandssysteme 50 Matrizen aus, für Drei- und Vierzustände musste diese Anzahl auf 200 erhöht werden.

Diese Ergebnisse legen dar, dass die Funktion der Norm über dem Parameterraum - also insbesondere die Linien kleiner Norm durch den Raum -, die durch die erzeugenden Parameter ohne die statistischen Fehler entstehen würde, bei idealen Systemen durch diese Fehler nur geringfügig verändert wird.

Im Folgenden sei exemplarisch ein Beispiel eines Zweizustandssystems gezeigt. Das ideale System wurde mit den $\Delta G_2^0 = 5,1253$, $\Delta V_2^0 = -0,0171 \text{ MPa}^{-1}$ und $\Delta b_2^{0'} = -3,1910 \text{ MPa}^{-2}$ und einer zufällig erzeugten Matrix A generiert. Die Norm der Fehlermatrizen betrug 0,02. Abbildung 3-42 zeigt die aus den Fehlersystemen resultierenden Anteile der Besetzungszahlen und die Verteilung im Parameterraum $\Delta G_2^0 - \Delta V_2^0$. In Rot sind jeweils die Anteile beziehungsweise die Parameter des idealen Systems eingezeichnet.

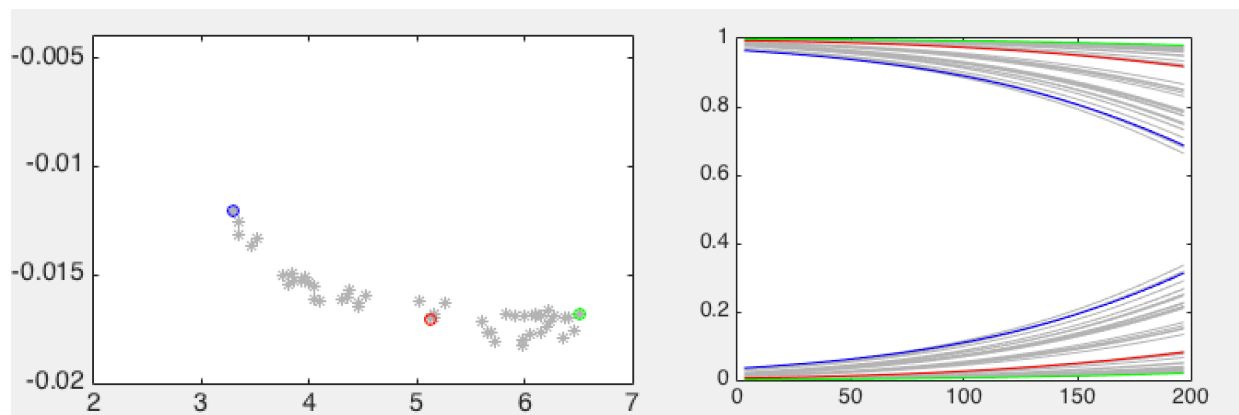


Abbildung 3-42: Aus den Fehlersystemen resultierende Anteile der Besetzungszahlen und Verteilung im Parameterraum $\Delta G_2^0 - \Delta V_2^0$

Auf einen idealen Datensatz, in der linken Graphik als roter Punkt im $\Delta G_2^0 - \Delta V_2^0$ - Parameterraum eingezeichnet (ΔG_2^0 nach rechts und ΔV_2^0 nach oben in MPa^{-1}) und rechts die daraus resultierenden Anteile der Besetzungszahlen in Abhängigkeit vom Druck (in MPa), wurden 50 verschiedene Fehlermatrizen der Norm 0,02 aufaddiert. Die dadurch verschobenen mathematischen Minima beziehungsweise Besetzungszahlen wurden in Grau eingezeichnet. Dann wurden die beiden (linke Graphik) jeweils am Rand liegenden Systeme (blau und grün) für weitere Untersuchungen ausgewählt.

Im nächsten Schritt wurden im Parameterraum die beiden Randpunkte der Verteilung ausgewählt - in den Graphiken mit Blau und Grün gekennzeichnet. Die Parameter des linken Randpunktes (blau) waren $3,2993$, $-0,0120 \text{ MPa}^{-1}$ und $-9,5819 \cdot 10^{-6} \text{ MPa}^{-2}$, die des rechten (grün) $6,5091$, $-0,0168 \text{ MPa}^{-1}$ und $3,42777 \cdot 10^{-5} \text{ MPa}^{-2}$. Die entsprechenden Normen waren $0,0177$ beziehungsweise $0,0180$. Mit den Parametern wurden die beiden Matrizen N_{blau} und $N_{\text{grün}}$ berechnet und mit den entsprechenden ebenfalls vom Algorithmus zurückgegebenen Matrizen A_{blau} und $A_{\text{grün}}$ zu den beiden neuen Systemen D_{blau} und $D_{\text{grün}}$ kombiniert. Für diese beiden Systeme wurden nun wieder jeweils 50 zufällige Fehlermatrizen mit einer Norm von $0,02$ erzeugt und analog zum obigen Vorgehen mathematische Minima gefunden. Deren Verteilungen im Parameterraum und die resultierenden Anteile der Besetzungszahlen zeigen, dass in dem gefundenen Bereich die ursprünglichen Parameter des idealen Systems enthalten sind (Abbildungen 3-43 und 3-44).

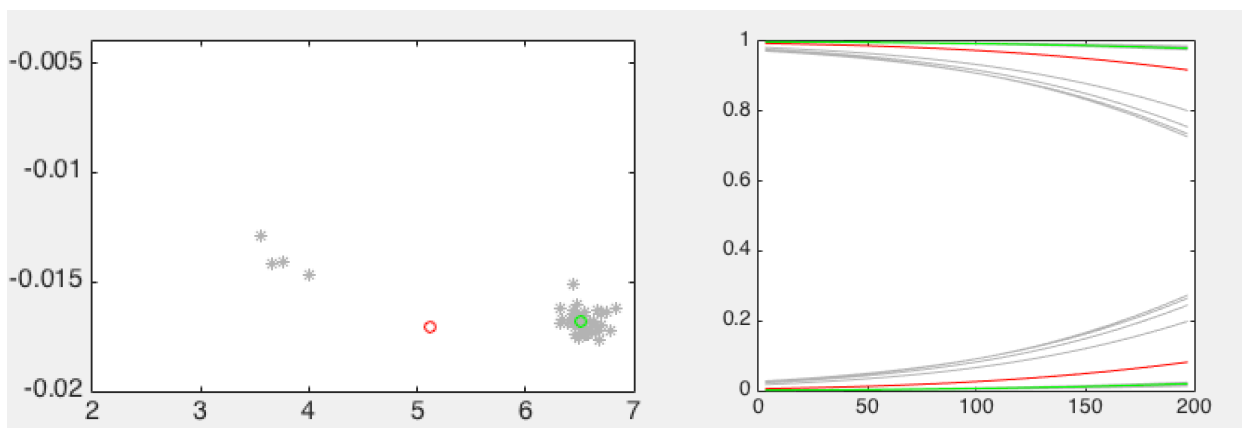


Abbildung 3-43: Wirkung der Fehlermatrizen auf die resultierenden Anteile der Besetzungszahlen und die Verteilung im Parameterraum $\Delta G_2^0 - \Delta V_2^0$ von einem Randpunkt des eigentlichen Systems aus

Für einen der beiden ausgewählten Parametersätze (grün) wurden ebenfalls 50 Fehlermatrizen mit einer Norm von $0,02$ erzeugt und für die resultierenden Datensätze die Minima bestimmt (grau). Die meisten der gefundenen Raumpunkte liegen dicht gedrängt um den grünen Punkt. Der durch die gefundenen Minima definierte Bereich im Parameterraum schließt die Parameter des ursprünglichen Systems (rot) ein.

In der linken Graphik ist nach rechts ΔG_2^0 und nach oben ΔV_2^0 in MPa^{-1} aufgetragen, in der rechten Graphik die Anteile der Besetzungszahlen gegen den Druck (in MPa).

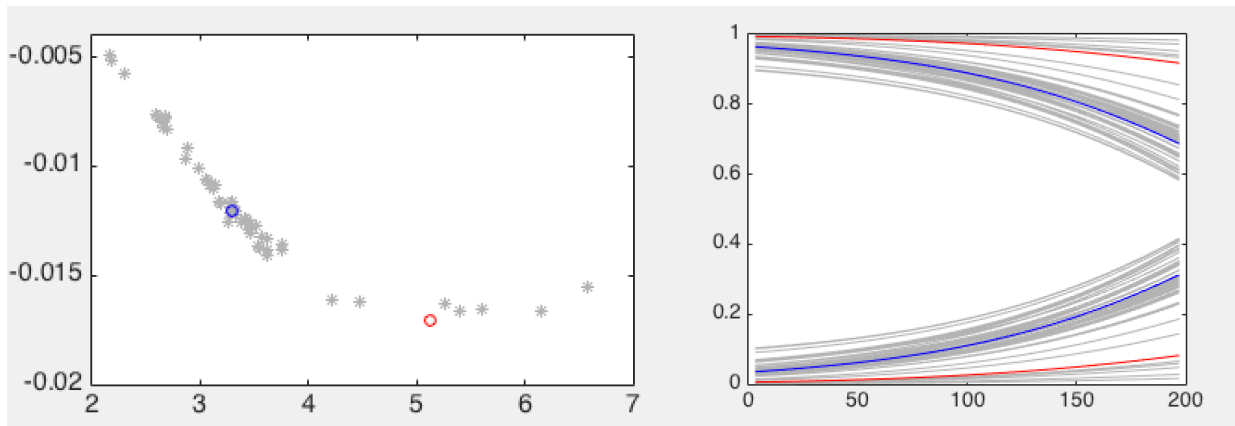


Abbildung 3-44: Wirkung der Fehlermatrizen auf die resultierenden Anteile der Besetzungszahlen und die Verteilung im Parameterraum $\Delta G_2^0 - \Delta V_2^0$ von einem Randpunkt des eigentlichen Systems aus

Analog zu Abbildung 3-43 sind hier die Ergebnisse des zweiten ausgewählten Parametersatzes (blau) dargestellt. Der dadurch definierte Bereich im Parameterraum schließt ebenfalls die Parameter des ursprünglichen Systems (rot) ein.

3.5.2 Definition des Lösungsbereichs über geeignete Punkte des Parameterraums

Beschreibung der Methode

Bei dem vorliegenden mathematischen Problem kann der Lösungsbereich nur schwer über Fehlergrenzen oder Vertrauensintervalle der Parameter, die das Druckverhalten der Gibbs-Energiedifferenzen beschreiben, definiert werden, da dies der Abhängigkeit der Parameter untereinander nicht Rechnung trägt. Vielmehr ist es sinnvoll, den Lösungsbereich über spezielle Raumpunkte zu definieren, die sich aufgrund der Norm am mathematischen Minimum diesem Bereich zuordnen lassen.

Vorausgesetzt, dass das mathematische Minimum zur Matrix $D_{original}$ gefunden wurde, liefert Algorithmus 1 dem Benutzer zwei verwendbare Informationen über das System, an erster Stelle die Parameter ΔG_m^0 , ΔV_m^0 und $\Delta b_m^{0'}$. Wie gezeigt, wird durch diese die Form der Funktion der Norm über dem Parameterraum bestimmt, und es kann davon ausgegangen werden, dass diese Funktion der eigentlichen Funktion des Systems sehr ähnlich ist und sich nur durch die Lage des Minimums auf den Linien kleinster Norm durch den Raum unterscheidet. Die zweite wichtige Information ist die Größe der Norm am mathematischen Minimum. Über sie lässt sich anhand der in Kapitel 2.4.4 beschriebenen Formeln sowohl die Norm $Norm_{erz}$ der Fehlermatrix R als auch die durch die Optimierung der Matrix A kleinere $Norm_{opt}$ abschätzen, die als Schwellenwert dient. Als Lösungsbereich mit einem Konfidenzniveau von 68% kommen alle Raumpunkte infrage, deren Norm kleiner als diese geschätzte Größe ist.

Um Punkte im Parameterraum zu finden, auf die diese Bedingung zutrifft, wurde zuerst versucht, den Raum um das mathematische Minimum systematisch zu rastern. Dabei stellte sich heraus, dass dies aufgrund der Anzahl der Dimensionen des Parameterraumes - bei einer sehr groben Rasterung von nur 10 Rasterpunkten in jeder Dimension sind das bei Vierzustandssystemen bereits eine Milliarde Rasterpunkte - nicht realisierbar ist. Ebenso war es nicht möglich, den Raum um das Minimum durch zufällig gewählte Punkte in adäquater Weise abzubilden, da durch die Abhängigkeit der Parameter das Gebiet durch den Raum sehr flach ist und deswegen schlecht getroffen werden kann.

Daher wurde ein anderer Weg gewählt, gezielt die gesuchten Raumpunkte, die sich entlang der Linien kleinster Norm sammeln, zu finden: Ähnlich wie in Kapitel 3.5.1 beschrieben, werden zum System des gefundenen mathematischen Minimums Fehlermatrizen generiert, deren Norm zwischen Null und Norm_{erz} liegt, und deren Minima bestimmt. Durch Überprüfung der Norm der so gefundenen Raumpunkte in Bezug auf die eigentliche - durch die korrigierten chemischen Verschiebungen definierte - Matrix D_{original} kann dann festgestellt werden, ob sie unter dem Schwellenwert der Norm liegen und somit dem Lösungsbereich angehören. Auf diese Weise lässt sich der interessierende Bereich eingrenzen.

Testergebnisse

Zur Überprüfung, ob diese Methode in der Lage ist, einen Lösungsbereich zu definieren, der den eigentlichen Parametersatz tatsächlich enthält, wurden mehrere Testdatensätze untersucht. Dabei stellte sich heraus, dass sich auf diese Weise tatsächlich Raumpunkte finden lassen, die die obengenannte Bedingung bezüglich der Norm erfüllen. Insbesondere zeigte sich, dass bei allen untersuchten Systemen, bei denen das mathematische Minimum korrekt identifiziert worden war, auch der gesuchte Parametersatz innerhalb des so definierten Lösungsbereichs lag. Voraussetzung dafür war eine ausreichend große Anzahl an Fehlermatrizen, um den Bereich im Parameterraum zu markieren. Dies wurde anhand von Projektionen der Punkte in verschiedene Ebenen überprüft.

Bei manchen Systemen musste die Norm der Fehlermatrizen leicht nach oben korrigiert werden, um bei der Überprüfung der erhaltenen Raumpunkte mit D_{original} Normen über und unter dem Schwellenwert zu erzeugen, um eine Abgrenzung des Lösungsbereichs innerhalb des Parameterraums zu ermöglichen.

3.5.3 Stabilitätstest der entwickelten Methode bei Variation der Anzahl der verwendeten Daten

Für die Erstellung der Testdaten wurden die korrigierten chemischen Verschiebungen der einzelnen Zustände, die die Matrix A füllen, bewusst aus einem Bereich gewählt, wie man ihn für experimentelle Daten bei Amidprotonen findet, und auch deren typische mittlere Änderung zwischen zwei Zuständen wurde in der entsprechenden Größenordnung angenommen. Allerdings wurde bei der zufälligen Erzeugung der Matrix A nicht berücksichtigt, dass es in experimentellen Datensätzen immer Aminosäuren gibt, die einen speziellen Übergang anzeigen, und außerdem ein großer Anteil der Residuen nach der aminosäureabhängigen Druckkorrektur kaum mehr Änderungen ihrer Verschiebungen aufweisen.

Vor der Auswertung der experimentellen Daten erfolgte ein Stabilitätstest, bei dem untersucht wurde, wie sich das Hinzufügen beziehungsweise Weglassen der Daten von Atomen bemerkbar macht, deren korrigierte chemische Verschiebungen sich mit dem Druck kaum ändern. Insbesondere stellte sich die Frage, ob deren Mitberücksichtigung bei der Bestimmung der Parameter nicht sogar hinderlich ist.

Deswegen wurde ein typisches System aus 80 Atomen nachgestellt, wie man es stark vereinfacht bei experimentellen Daten finden könnte, wenn nur Zustände im schnellen Austausch miteinander vorlägen:

- 10 Atome, bei denen die korrigierten chemischen Verschiebungen der ersten beiden Zustände gleich sind und sich vom dritten um 0,5 ppm unterscheiden und deswegen für den Übergang zum dritten Zustand charakteristisch sind
- 10 Atome, bei denen die korrigierten chemischen Verschiebungen des ersten und dritten Zustands gleich sind und sich vom zweiten um 0,5 ppm unterscheiden
- 10 Atome, bei denen die korrigierten chemischen Verschiebungen des zweiten und dritten Zustands gleich sind und sich vom ersten um 0,5 ppm unterscheiden
- 10 "unspezifische" Atome mit drei verschiedenen korrigierten chemischen Verschiebungen mit Abstand 0,2 ppm, unterschiedlich angeordnet
- 40 Atome - A_{unspez} genannt - mit drei genau gleichen chemischen Verschiebungen für die Zustände

Mit den so gewählten chemischen Verschiebungen ließ sich die Matrix A bilden. Dazu wurde ein zufällig gewähltes Parameterset aus $\Delta G_2^0, \Delta V_2^0, \Delta b_2^{0'}, \Delta G_3^0, \Delta V_3^0$ und $\Delta b_3^{0'}$ bestimmt und die Matrix D_{ideal} berechnet. Zu diesem System wurden nun 50 verschiedene Fehlermatrizen der Größe m aufaddiert, so dass 50 Systeme mit Fehlern entstanden. Die chemischen Verschiebungen der letzten 40 Aminosäuren enthielten auf diese Weise nur die statistischen Schwankungen ohne jegliche Information für das System.

Als Norm für die Fehlermatrix wurde ein Wert von 0,005 verwendet. Das entspricht einem mittleren Fehler von 0,00018 ppm je Messpunkt. Dieser ist damit um 3 Größenordnungen kleiner als die Änderungen der chemischen Verschiebungen, die bei den für die Übergänge zwischen den Zuständen anzeigenden Atomen angenommen wurden und deren Werte 0,5 ppm betrugen.

Nun erfolgten drei Testrechnungen: Bei der ersten wurden für diese 50 Systeme jeweils das mathematische Minimum bestimmt, die daraus folgenden Besetzungszahlen berechnet und zum Vergleich gegen den Druck aufgetragen. Bei den beiden anderen Rechnungen wurden nur 60 Atome beziehungsweise nur 40 Atome verwendet, wobei dazu die Informationen der A_{unspec} ganz oder teilweise weggelassen wurden, und die sonst identischen Berechnungen durchgeführt.

Dabei zeigte sich, dass durch das Hinzufügen oder Weglassen der informationslosen Atome, die bei experimentellen Daten Atomen entsprechen, die nach der aminosäureabhängigen Druckkorrektur keine Änderungen mehr zeigen, der Algorithmus nicht beeinträchtigt wird und die Streuung durch die statistischen Fehler in der gleichen Größenordnung bleibt. Abbildung 3-45 zeigt die Besetzungszahlen der Fehlermatrizen für die Berechnungen mit allen Atomen beziehungsweise mit 40 Atomen. Die Folgerung daraus ist, dass keine Vorselektion der Daten bezüglich der Größe der Änderung der Druckabhängigkeit der chemischen Verschiebungen der Atome vor der Auswertung erfolgen muss. Allerdings zeigt sich an der erhaltenen Streuung der Verläufe der Besetzungszahlen mit dem Druck der enorme Einfluss selbst sehr kleiner statistischer Fehler und damit die Empfindlichkeit der mathematischen Funktion Norm an sich.

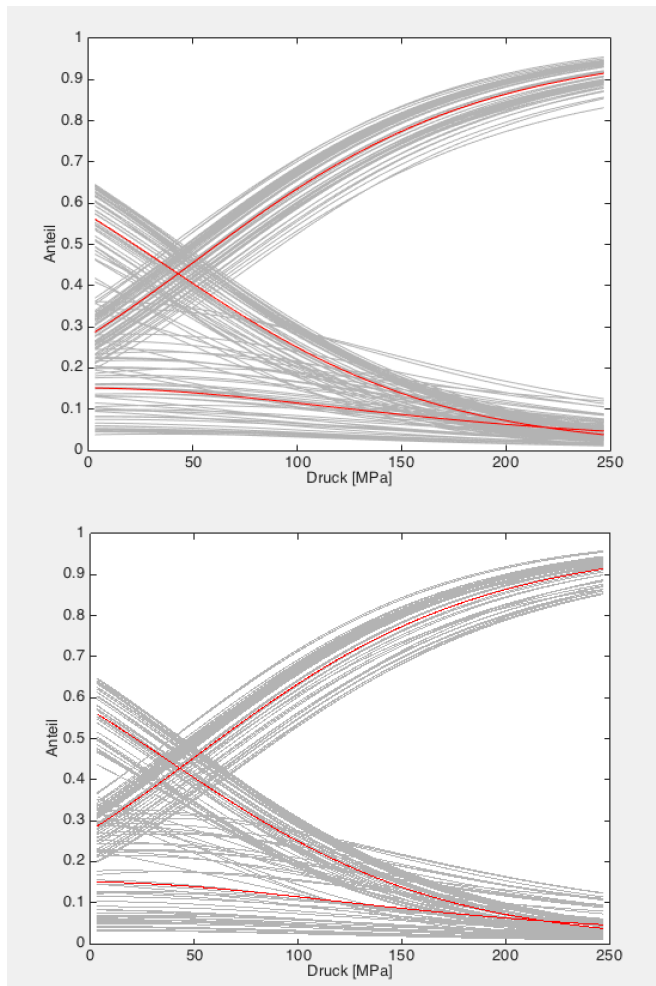


Abbildung 3-45: Besetzungszahlen in Abhängigkeit vom Druck (in MPa), die durch die Fehler verschoben werden

Oben: Komplettes System, bei dem 40 Atome ohne Information bei Algorithmus 1 mitberücksichtigt wurden.

Unten: Vor Anwenden von Algorithmus 1 wurden die informationslosen Atome aussortiert.

Dabei zeigt sich, dass durch die Atome ohne Änderungen der chemischen Verschiebungen die Fehler des Systems nicht vergrößert werden.

3.6 Untersuchung des Druckverhaltens des Proteins Byr2-RBD

3.6.1 Vorbereitung der Daten

Im ersten Schritt wurden die zu untersuchenden TROSY-HSQC-Spektren (Abbildung 3-46) in beiden Dimensionen mit jeweils 4096 Datenpunkten referenziert und prozessiert, wobei in ^{15}N -Richtung eine *Linear Prediction* durchgeführt wurde. Danach wurden die Signale anhand einer vorliegenden Liste chemischer Verschiebungen zugeordnet, wobei als erstes das bei Normaldruck aufgenommene Spektrum bearbeitet und das Ergebnis dann sukzessive auf die weiteren Spektren übertragen wurde.

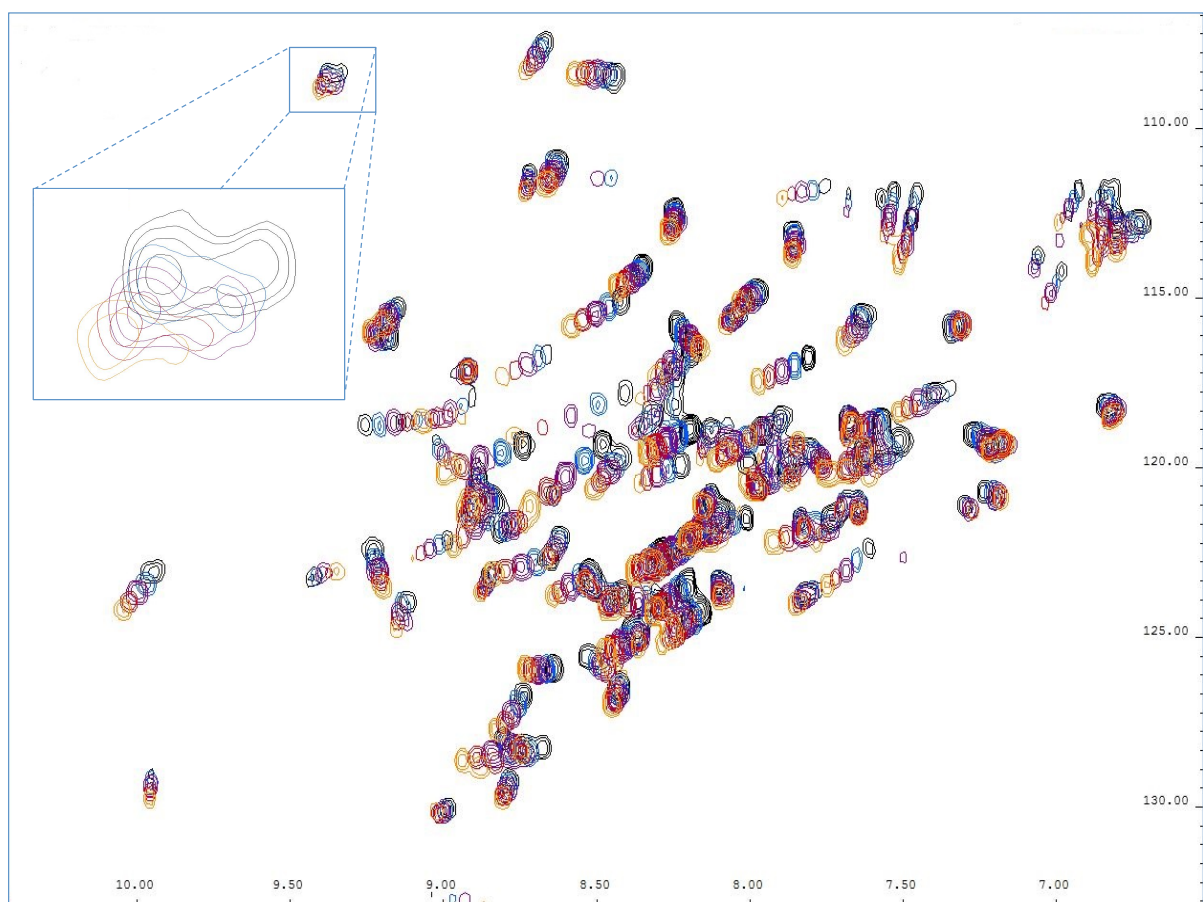


Abbildung 3-46: TROSY-HSQC-Spektren von Byr2-RBD, die bei verschiedenen Drücken aufgenommen wurden

3 MPa (schwarz), 50 MPa (blau), 100 MPa (violett), 150 MPa (rot) und 200 MPa (gelb). Der Übersichtlichkeit halber fehlen in der Abbildung die ebenfalls in die Auswertungen eingegangenen Spektren für 25 MPa, 75 MPa, 125 MPa und 225 MPa.

Oben links ist ein vergrößerter Ausschnitt der Aminosäure 80G zu sehen, bei der die für einige Residuen typische Aufspaltung der Signale deutlich zu erkennen ist.

Eine erste optische Analyse ergab folgende Ergebnisse:

- Die beobachteten druckinduzierten chemischen Verschiebungen der Signale sind reversibel.
- Die Spektren enthalten etliche kleine, nicht zuordenbare Signale.
- Einige Signale zeigen eine deutliche Aufspaltung (Abbildung 3-46, vergrößerter Ausschnitt) in zwei bei Normaldruck fast gleich große Signale. Dies ließ sich nicht in allen Fällen durch Cis-Trans-Isomerie benachbarter Proline erklären. In manchen Fällen ist eine Aufspaltung in drei Signale zu erkennen, wobei das dritte wesentlich kleiner ist als die beiden anderen.
- Die Volumina der Signale der einzelnen Aminosäuren differieren stark.
- Die Gesamtvolumina der Signale in den Spektren variieren ebenfalls stark.

3.6.2 Versuch einer Auswertung der chemischen Verschiebungen mit Algorithmus 1

Um die chemischen Verschiebungen zu erhalten, erfolgte ein *Peakpicking* mit dem Programm AUREMOL. Die Auswertung wurde erschwert durch viele aufgespaltene Signale. Diese wurden - so weit wie möglich - zu einem Signal zusammengefasst und der Wert des Schwerpunkts des Systems verwendet, um die bei Algorithmus 1 vorausgesetzte besetzungszahlgewichtete Mittelung zu erreichen. Zusätzlich nahm bei einigen das Volumen mit Zunahme des Drucks deutlich ab, so dass davon ausgegangen werden muss, dass bei dem untersuchten Protein die Annahme - alle Zustände befinden sich im schnellen Austausch miteinander - nicht gegeben ist und daher zusätzlichen zu den statistischen Fehlern, die bei der Datenauswertung aus der Begrenzung der Auflösung und kleineren Artefakten resultieren, mit weiteren Fehlern bei der Bestimmung der gemittelten chemischen Verschiebungen zu rechnen ist, deren Ursache in unzureichender Erfassung des Gesamtsignals liegt.

Außerdem zeigte sich, dass nur die ^1H chemischen Verschiebungen der Amidprotonen des Proteinrückgrats in die weitere Auswertung einfließen konnten, da die Auflösung in Richtung der Stickstofffrequenzen zu gering war und für die chemischen Verschiebungen der Stickstoffatome keine auswertbaren Druckkurven resultierten. Bei insgesamt 88 Aminosäuren ließen sich ^1H chemische Verschiebungen an allen Druckpunkten finden, für die dann eine aminosäureabhängige Druckkorrektur erfolgte (Köhler et al., 2012). Die resultierenden korrigierten Werte der einzelnen Aminosäuren sind in Abhängigkeit vom Druck in Abbildung 3-47 zu sehen. In diesen ist zu erkennen, dass auch nach dieser Korrektur noch eine deutliche Veränderung mit dem Druck resultiert.

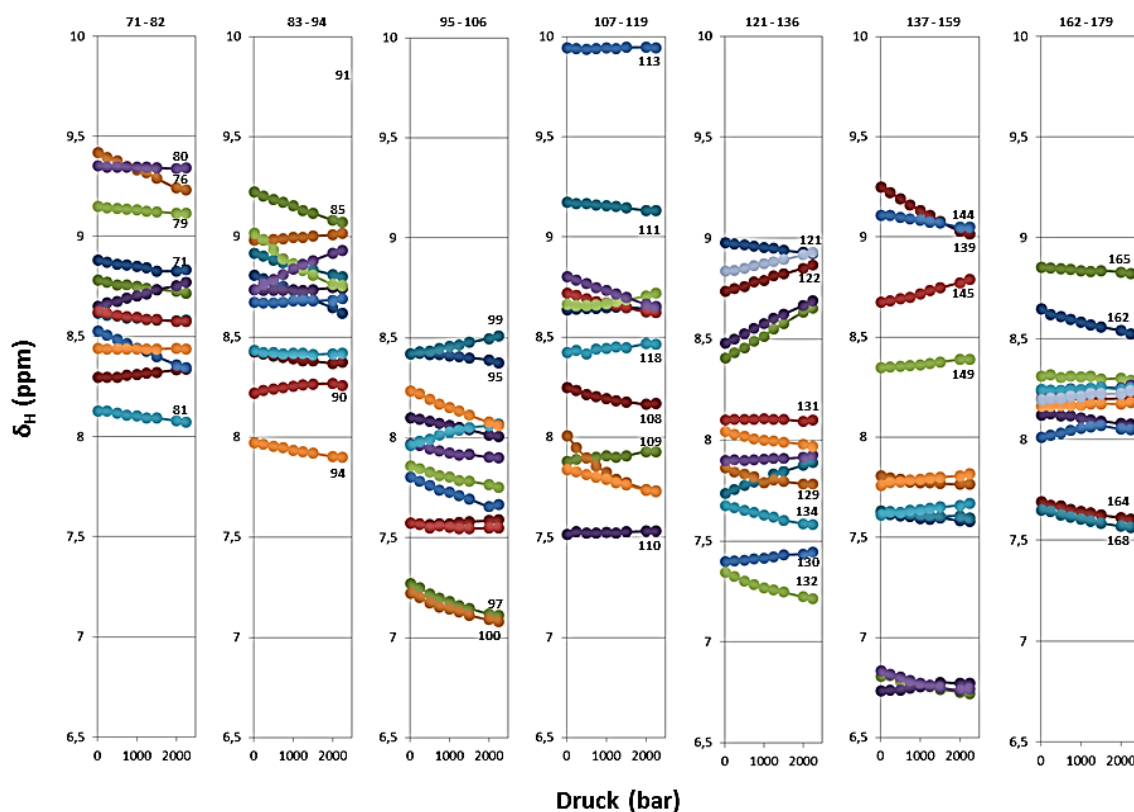


Abbildung 3-47: Chemische Verschiebungen der Amidprotonen des Proteinrückgrats in Abhängigkeit des Drucks nach der aminosäureabhängigen Druckkorrektur

Oben am Rand jeder Spalte sind die Nummern der dargestellten Aminosäuren ablesbar.

Anwendung von Algorithmus 1

Aufgrund der Tatsache, dass einige der Signale in drei Untersignale aufgespalten sind, wurde bei der durchgeführten Auswertung der chemischen Verschiebungen der Amidprotonen des Proteinrückgrats mit Algorithmus 1 angenommen, dass 3 Zustände vorliegen. Unter Einbeziehung aller Aminosäuren wurde das mathematische Minimum mit den Parametern $\Delta G_2^0 = 0,3120$, $\Delta V_2^0 = 0,0011 \text{ MPa}^{-1}$ und $\Delta b_2^0 = 1,21 \cdot 10^{-5} \text{ MPa}^{-2}$ für den Übergang zwischen Zustand 1 und Zustand 2 und $\Delta G_3^0 = 2,4994$, $\Delta V_3^0 = -0,0156 \text{ MPa}^{-1}$ und $\Delta b_3^0 = 3,14 \cdot 10^{-5} \text{ MPa}^{-2}$ zwischen Zustand 1 und Zustand 3 gefunden. Die zugehörige mathematische Norm ist mit 0,0676 sehr hoch, was einem mittleren statistischen Fehler von $3 \cdot 10^{-3} \text{ ppm}$ pro Messpunkt entspräche. In Abbildung 3-48 ist die resultierende Verteilung für die Parameter des mathematischen Minimums gezeigt.

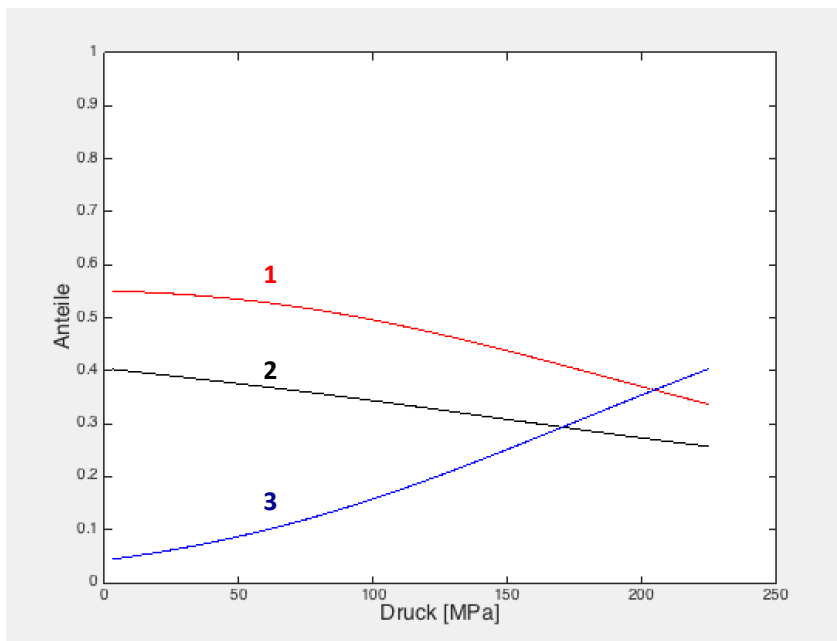


Abbildung 3-48:
Populationsverhältnisse
von Byr2-RBD

Die Populationsverhältnisse der Zustände des Proteins Byr2-RBD, die das mathematische Minimum des Systems nahelegt.

Aufgrund dieser sehr großen mittleren Fehler ergab die Fehleranalyse, die in dem Gebiet um das gefundene Minimum im Parameterraum durchgeführt wurde, einen sehr großen Bereich, der dem Lösungsraum zugeordnet werden muss. Weder die in Kapitel 3.5.2 vorgeschlagene Methode zur Bestimmung der Fehlergrenzen noch eine systematische Rasterung beziehungsweise eine Untersuchung zufällig erzeugter Parameterkombinationen brachte wegen der starken Abhängigkeit der Parameter voneinander und der zugleich vorliegenden Breite des Fehlerbereichs eine befriedigende Lösung zur Fehlerabschätzung.

Außerdem zeigte sich bei der Auswertung dieses Datensatzes, dass er nicht stabil gegenüber dem Weglassen der Daten einzelner Aminosäuren bei der Anwendung von Algorithmus 1 reagierte, indem sich dadurch die Besetzungszahlen und insbesondere deren Verlauf mit dem Druck stark änderten.

Zusätzlich wurden verschiedenste weitere Versuche unternommen, durch Variation der Anzahl der Zustände - es wurden zwei, vier und fünf Zustände angenommen - ein zufriedenstellenderes Modell für den zu untersuchenden Datensatz zu finden, was aber nicht gelang.

Daher muss angenommen werden, dass bei den gewonnenen Daten aufgrund des langsamen Austauschs keine ausreichende besetzungszahlgewichtete Mittelung über alle relevanten Zustände vorliegt und Algorithmus 1 somit nicht angewandt werden kann.

3.6.3 Untersuchung der Signalvolumina

Die Signalaufspaltungen in den Spektren und die unterschiedlichen Größen der Peakvolumina legen nahe, dass die im untersuchten Druckbereich dominierenden Zustände des Proteins im langsamen Austausch miteinander stehen. Deswegen wurde parallel zur Auswertung der chemischen Verschiebungen auch eine Untersuchung der Signalvolumina vorgenommen. Dabei fiel auf, dass die Gesamtvolumina der einzelnen Spektren, gewonnen durch Integration über alle Signale, stark schwankten. Als Grund dafür wurde eine sich ändernde Monomerkonzentration von Byr2-RBD innerhalb der Messreihe vermutet. Daher wurde zunächst eine Korrektur durchgeführt, bei der das Volumen bei Normaldruck des HSQC-Spektrums und zusätzlich des korrespondierenden 1D-Spektrums als Referenz für eine Normierung mit $V_{nor}(p) = a_p * V_{mes}(p)$ verwendet wurde, wobei sich die Korrekturfaktoren a_p direkt aus Verhältnissen der in Abbildung 3-49 dargestellten Gesamtvolumina der Spektren herleiten.

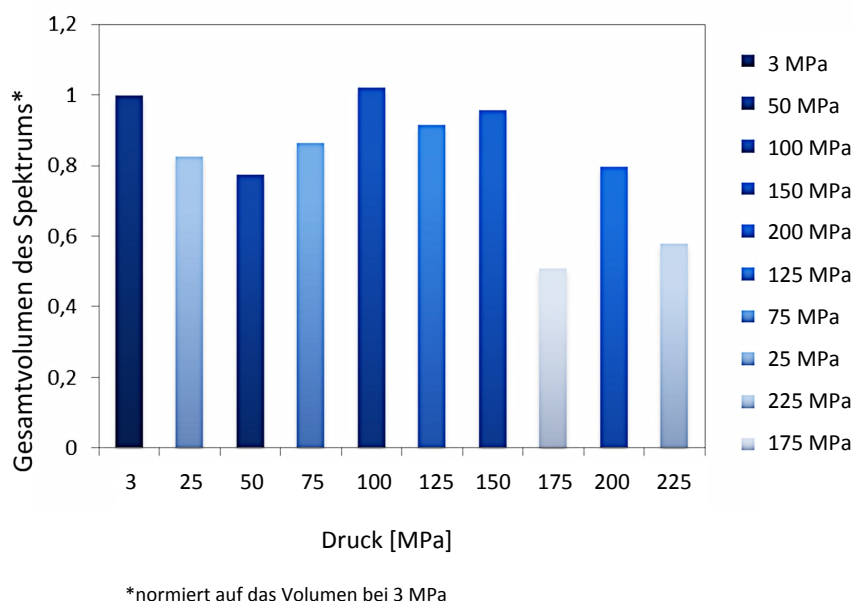


Abbildung 3-49: Gesamtvolumina der Signale der einzelnen HSQC-Spektren

Die Gesamtvolumina der Signale (gewonnen durch Integration des gesamten Spektrums) der einzelnen HSQC-Spektren, normiert auf den Wert bei 3 MPa, die zur Korrektur verwendet wurden, um die Volumina vergleichen zu können. Die farbliche Kodierung entspricht der Reihenfolge der Aufnahme der Spektren (begonnen wurde bei 3 MPa).

Nach dieser Korrektur wurden die Volumina der einzelnen Aminosäuren miteinander verglichen. Dabei zeigte sich ein sehr inhomogenes Verhalten, das sich hauptsächlich durch die der schlechten Auflösung geschuldeten Fehler erklären lässt, die geschätzt in einer Größenordnung von 20% liegen.

Die größten Volumenänderungen mit dem Druck ergeben sich bei den Aminosäuren des C-Terminus. Die zweit- und die drittletzte Aminosäure des Konstrukts - L178 und S179 - besitzen die größten Volumina des gesamten Spektrums. Bei ihnen kann davon ausgegangen werden, dass sie bei allen Zuständen sehr frei beweglich und ihre Signale nicht aufgespalten sind. Sie weisen zudem Werte für die chemischen Verschiebungen auf, die typisch für Random-Coil-Bereiche sind. Die Signale der diesen Aminosäuren vorangehenden Residuen - S173, S176 und S177 - zeigen bei Normaldruck ein Volumen von etwa $48\% \pm 20\%$, das dann bis zum Ende der Druckreihe auf etwa $22\% \pm 20\%$ abfällt (Abbildung 3-50). Dieses Verhalten zeigen keine weiteren Signale in den untersuchten Spektren.

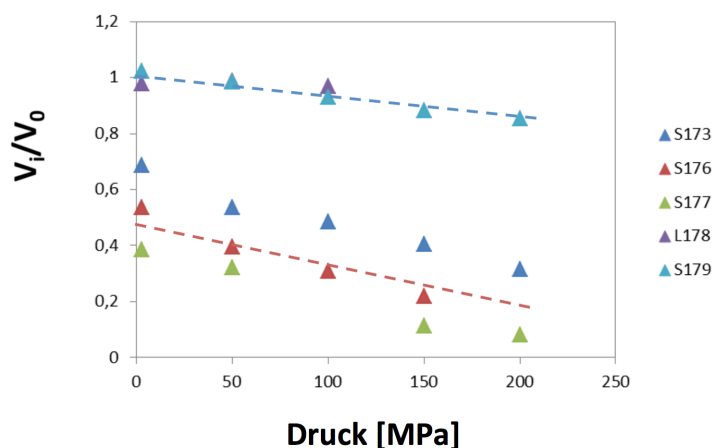


Abbildung 3-50: Signalvolumenvergleich

Ein Volumenvergleich der Aminosäuren des C-Terminus zeigt für die zweit- und drittletzte Aminosäure des Konstrukts kaum Veränderungen mit dem Druck. Sie dienen hier zur Normierung (V_0). Die Signale dagegen, die den unmittelbar davor liegenden Aminosäuren zugeordnet werden konnten, weisen bereits bei Normaldruck nur ein Volumen von etwa 48% auf, das dann bei 200 MPa auf nur noch etwa 22% abgefallen ist.

3.6.4 Multimodale Datenanalyse

Seitenkettenanalyse

Wie eine Analyse der Seitenketten ergab, ist die Struktur von Byr2-RBD so aufgebaut ist, dass die Aminosäuren nach der letzten Helix nicht mehr zum hydrophoben Kern gehören (Abbildung 3-51).

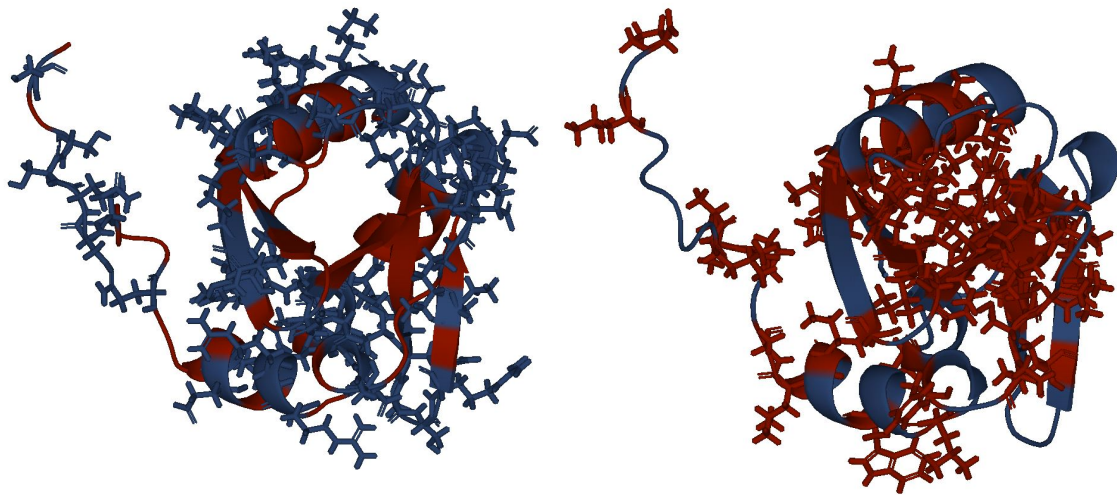


Abbildung 3-51: Analyse der Seitenketten von Byr2-RBD

Die Abbildungen links und rechts unterscheiden sich dadurch, dass rechts alle nicht-polaren Seitenketten (rot: Glycine, Alanine, Valine, Leucine, Isoleucine, Proline, Phenylalanine, Methionine, Tryptophane und Cysteine) zu sehen sind, links dagegen die übrigen (blau). Deutlich erkennbar ist der in sich abgeschlossene hydrophobe Kern, dem der C-Terminus nicht mehr angehört.

Ergebnisse aus molekulardynamischen Rechnungen

Da auch bei den von Wolfram Gronwald und Fritz Huber durchgeführten NMR-Experimenten für die Aminosäuren S161 bis P180 keine Abstandsrestraints gefunden werden konnten (siehe Abbildung 3-4), war dieser Proteinbereich bei den in Kapitel 3.1.2 beschriebenen Strukturrechnungen nicht scharf definiert (Abbildung 3-52 rechts). Unter den 20 energieärmsten Konformationen finden sich jedoch vier, bei denen sich wie bei der hier abgebildeten (Abbildung 3-52 links) der C-Terminus an der Oberfläche anlagert und sogar Wasserstoffbrücken zum restlichen Protein bilden könnte. Bei den Strukturrechnungen wurden keine diese strukturelle Eigenheit begünstigenden Restraints verwendet. Eine Analyse der erhaltenen Konformationen legt folgende Verknüpfungen nahe:

- L178 O - K101 H_{ζ2}
- S174 O_γ - R83 H_{η12}
- A170 O - R74 H_{η12}
- L166 H_N - E163 O
- A167 H_N - I164 O

In Abbildung 3-52 sind die betreffenden Aminosäuren des C-Terminus in Cyan dargestellt, diejenigen der Ras-Bindestelle, über die sich das Ende legt, sind rot eingefärbt.

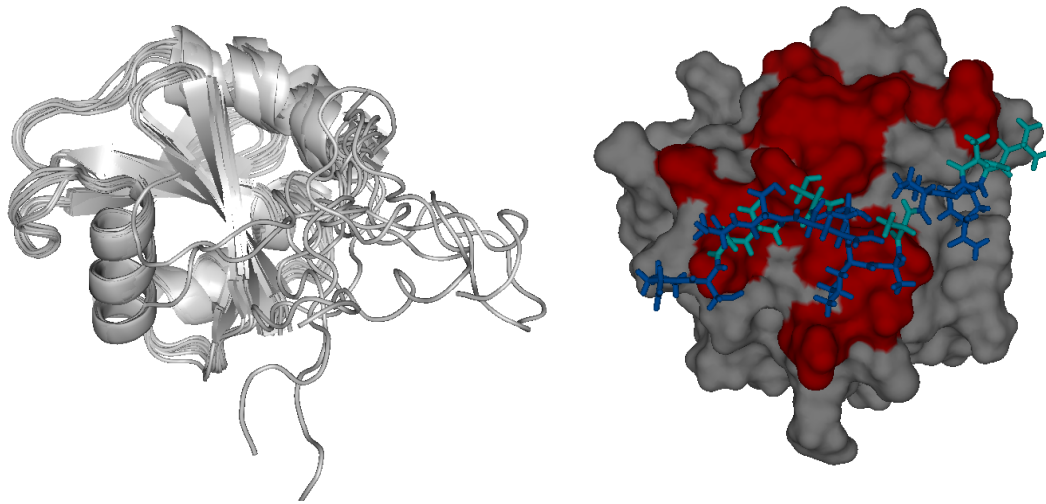


Abbildung 3-52: Molekulardynamische Berechnungen Byr2-RBD

Links einige der energieärmsten berechneten Strukturen. Da für den C-Terminus keine Abstandsrestraints vorhanden waren, wird hier keine einheitliche Konformation angenommen. Allerdings sind unter den 20 energieärmsten 4 zu finden, die sich eng an das restliche Molekül (grau/rot als Oberfläche, blau/cyan in Stickdarstellung der C-Terminus) anschmiegen (rechts abgebildet eine der Konformationen, wobei der bewegliche Teil ab Aminosäure E165 angenommen wurde) und sich über die Ras-Bindestelle (rot) legen. Eine Analyse der abgebildeten Struktur zeigte, dass Wasserstoffbrücken zwischen den in cyan dargestellten Aminosäuren und dem restlichen Molekül vorhanden sein könnten.

Analyse der Signalaufspaltungen

Einen Hinweis darauf, dass diese Konformationen, die sich bei den Strukturrechnungen trotz fehlender fixierender Restraints als energetisch günstig erwiesen, tatsächlich in dieser oder ähnlicher Weise bei Byr2-RBD realisiert werden könnten, liefern auch die Signalaufspaltungen. Diese betreffen insbesondere die Aminosäuren I76, A77, G80 (Abbildung 3-46, Vergrößerung), L143 und K150, die alle in Nachbarschaft des C-Terminus - in der in Abbildung 3-52 rechts gezeigten Konformation - sind. Diese Aminosäuren gehören auch teilweise der Ras-Bindestelle an. Bei jeder dieser Aminosäuren liegt eine Aufspaltung in zwei oder drei Signale vor. Ein Vergleich der Volumina der beiden größeren zeigt, dass der stärker besetzte bei Normaldruck $60\% \pm 10\%$ des aus diesen beiden Peaks gebildeten Gesamtvolumens ausmacht, während der manchmal erkennbare dritte Anteil im Vergleich zu denen der beiden anderen deutlich kleiner ist. Wie in Abbildung 3-46 (Vergrößerung) für Aminosäure G80 zu erkennen, verschiebt sich das Volumen bei höher werdendem Druck

zugunsten des bereits bei Normaldruck größeren Signals, das bei 200 MPa $85\% \pm 10\%$ des Volumens der beiden zu vergleichenden Signale auf sich vereint. Diese Ergebnisse deuten ebenso wie die Analyse der Signalvolumina darauf hin, dass der C-Terminus beweglich ist und zwischen einem freien und einem an die Oberfläche des Proteins angeschmiegenen Zustand wechseln könnte.

Vorhersage einer Interaktionsfläche

Des Weiteren wurde eine Auswertung der chemischen Verschiebungen nach Beck-Erlach et al. durchgeführt (Beck-Erlach et al., 2014). Aufgrund der schlechten Auflösung der Spektren wurden zur Analyse nur Aminosäuren herangezogen, deren Änderung der aminosäureabhängig druckkorrigierten chemischen Verschiebungen im betrachteten Druckbereich mindestens 0,1 ppm betrug. Wie in Kapitel 2.2.1 dargelegt, werden die Koeffizienten B_1^* und B_2^* bestimmt und deren Verhältnis berechnet (Tabelle 3-9).

Die Standardabweichung σ dieser Verteilung beträgt $2,2205 \text{ GPa}^{-1}$. Die Abbildungen 3-53 und 3-54 stellen das Verhältnis B_2^*/B_1^* für die untersuchten Aminosäuren und ihre Lage im Protein Byr2-RBD dar. Wie von Beck-Erlach et al. beschrieben, deuten positive Werte für B_2^*/B_1^* oder kleinere als $-\sigma$ auf Interaktionsflächen hin (Beck-Erlach et al., 2014), was ebenfalls das in Abbildung 3-49 gezeigte Modell unterstützt, da die meisten der auffälligen Aminosäuren (R74, I97, K100, K101, S103, E105, W162) in den vermuteten Bindungsbereich des C-Terminus fallen.

Tabelle 3-9: Die Koeffizienten B_1 und B_2 und deren Verhältnisse für Byr2-RBD

Ausgewertet wurden die nach Köhler et al. druckkorrigierten chemischen Verschiebungen der Amidprotonen des Proteinrückgrats

Aminosäure	B_1 [GPa ⁻¹]	B_2 [GPa ⁻²]	B_2/B_1 [GPa ⁻¹]
R74	0,5153	0,0550	0,1067
I76	-0,8860	0,2935	-0,3313
A77	-0,8748	0,4278	-0,4891
R83	-0,7691	-0,6390	0,8309
V85	-0,7642	0,7277	-0,9523
S87	-0,6656	1,3743	-2,0647
Y91	-1,9926	6,8997	-3,4627
Q92	1,1799	-2,5807	-2,1872
I97	-1,1491	3,7500	-3,2635
K100	-0,9759	3,1237	-3,2010
K101	-0,9720	2,7791	-2,8592
S103	-0,6618	1,8044	-2,7265
E105	0,8280	-3,0872	-3,7287
D106	-0,9220	1,1957	-1,2969
V112	-2,3770	10,3359	-4,3483
Q116	0,7861	0,9591	-1,2199
R119	-0,5544	0,4411	-0,7956
L122	0,6141	-0,2493	-0,4059
T124	1,2188	-0,6628	-0,5438
E126	1,0935	-1,2973	-1,1864
E127	0,7497	-0,4850	-0,6469
C132	-0,9227	3,0818	-3,3400
E139	-1,3494	2,4239	-1,7963
I145	0,4301	0,8384	1,9495
W162	-0,7287	1,8843	-2,5859

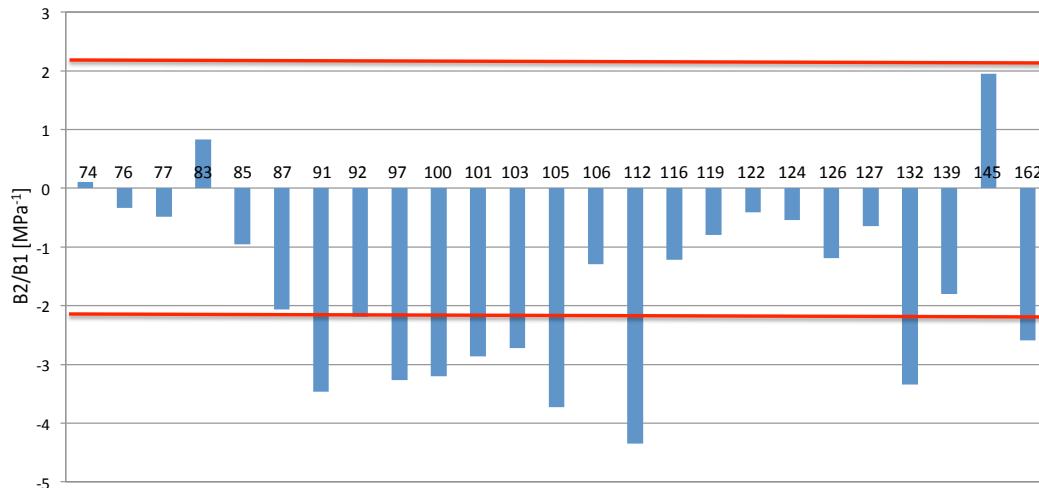


Abbildung 3-53: Verhältnis B_2/B_1 , dargestellt für die untersuchten Aminosäuren im Vergleich (vergleiche Tabelle 3-9)

Ein Hinweis darauf, dass die betrachtete Aminosäure zu einem Interaktionsbereich gehört, liegt dann vor, wenn das berechnete Verhältnis positiv oder kleiner als $-\sigma$ (in Rot eingezeichnet) ist.

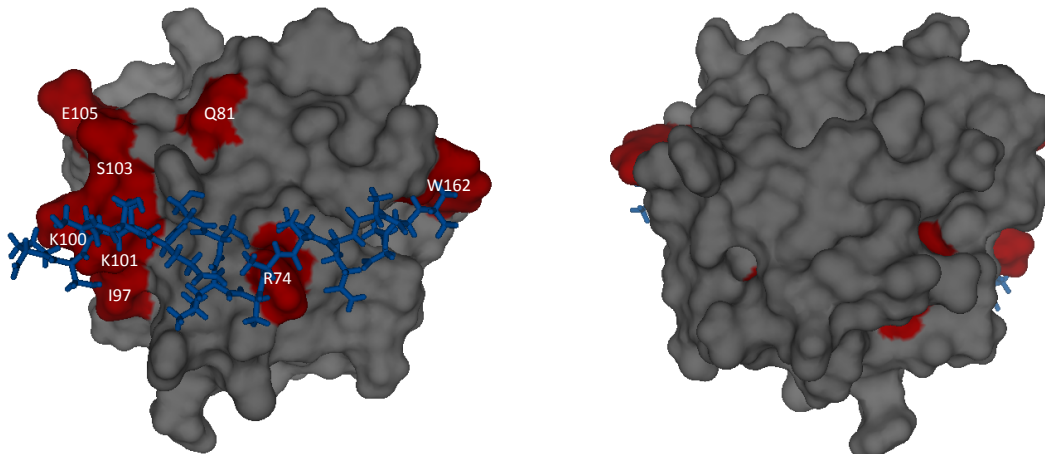


Abbildung 3-54: Aminosäuren, die vermutlich einem Interaktionsbereich angehören

Die Darstellung der Aminosäuren 70-165 erfolgte als Oberfläche, der bewegliche C-Terminus (166-180) als Stick. Rot gefärbt sind die als Indikator für eine Interaktionsstelle identifizierte Residuen. Links zu sehen ist der vermutete Interaktionsbereich zwischen dem C-Terminus und dem übrigen Molekül, rechts die Rückseite von Byr2-RBD (Drehung um 180 Grad).

Eine Ausnahme bildet Aminosäure Q81, die in dem berechneten Modell etwas vom C-Terminus entfernt liegt und ebenfalls das typische auf eine Interaktionsfläche hindeutende Verhalten zeigt. Für sie ist jedoch aus einer Analyse von PDB-1K8R bekannt, dass sie zur

Bindestelle mit Ras gehört. Alle übrigen auffälligen Aminosäuren liegen im Inneren des Proteins, was für eine globale Zustandsänderung bei Bindung des C-terminalen Endes an die Oberfläche spricht. Keine der mit dieser Methode identifizierten Aminosäuren liegt gut zugänglich auf der von dieser Interaktion nicht betroffenen Seite von Byr2-RBD (Abbildung 3-54, rechte Graphik).

3.7 Anwendung von Algorithmus 1 auf weitere experimentell gewonnene Datensätze

3.7.1 Untersuchung des Druckverhaltens der chemischen Verschiebungen von Ras

Im Gegensatz zu den Spektren des Proteins Byr2-RBD, in denen sehr viele Signale deutlich sichtbar aufgespalten sind, zeigen die Daten von Ras, die bei einer Temperatur von 303 K aufgenommen worden waren, dieses Verhalten nicht. Trotzdem unterliegen einige Signale Volumenänderungen bei Druckerhöhung, und korrespondierend dazu finden sich neue, aber nicht zuordenbare Signale in den Hochdruckspektren. Dies deutet darauf hin, dass mindestens einer der Zustände zu den anderen in langsamen Austausch steht. Deswegen muss auch bei diesem Datensatz hinterfragt werden, ob die beobachteten Signale besetzungszahlgewichtete Mittelungen der chemischen Verschiebungen aller bei Normaldruck relevanten Zustände sind, die in schnellem Austausch miteinander stehen, oder ob für manche Atome zusätzliche Anteile aus weiteren Zuständen im langsamen Austausch zu diesen die Messwerte beeinflussen.

Freie Minimumsuche

Analog zu Kalbitzer et al. wurden zur Auswertung die korrigierten chemischen Verschiebungen der Stickstoffatome des Proteinrückgrats verwendet (Kalbitzer et al., 2013), jedoch im Gegensatz zu der dort publizierten Vorgehensweise keinerlei Informationen aus Volumenuntersuchungen der Signale, aus den Spektren niedrigerer Temperatur oder den Untersuchungen mit ^{31}P -NMR-Spektroskopie. In einem ersten Schritt wurde unter der Annahme, dass vier Zustände vorliegen, mit Algorithmus 1 eine Suche nach dem mathematischen Minimum des Systems durchgeführt. Tabelle 3-10 enthält die Parameter Differenz der molaren Gibbs-Energien ΔG^0 , der partiellen molaren Volumina ΔV^0 und der molaren Kompressibilitätskoeffizienten $\Delta \beta^{0'}$. Im Gegensatz zu dem für das Protein Byr2 vorliegenden Datensatz zeigten sich die hier untersuchten Daten wesentlich stabiler beim Weglassen der korrigierten chemischen Verschiebungen einzelner Atome. Abbildung 3-55 zeigt die Populationsverhältnisse der Zustände. Dabei wurde, um einen besseren Vergleich

mit Abbildung 1-3 (nach Kalbitzer et al., 2009) anstellen zu können, die aus den gefundenen Parametern berechneten Besetzungszahlen bis zu einem Druck von 400 MPa dargestellt, obwohl nur Daten bis 180 MPa vorlagen.

Tabelle 3-10: Die Werte für die Differenzen der molaren freien Energie, der partiellen molaren Volumina und die molaren Kompressibilitätskoeffizienten für das Protein Ras des mathematischen Minimums des Systems.

Die Fehleranalyse zeigt jedoch, dass diese Werte aufgrund der Breite der Konfidenzintervalle das System nur unzureichend charakterisieren. Bei einem Konfidenzniveau von 68% erstreckt sich das Konfidenzintervall beispielsweise für ΔG_{21}^0 von 0,8 kJ mol⁻¹ bis 2,3 kJ mol⁻¹, für die übrigen Differenzen der freien Energien sind die Intervalle noch um vieles breiter.

Übergang	ΔG^0 [kJ mol ⁻¹]	ΔV^0 [mL mol ⁻¹]	$\Delta \beta^{0'}$ [mL MPa ⁻¹ mol ⁻¹]
2 → 1	1,93	-2,59	0,060
2 → 3	6,57	-28,40	-0,024
2 → 4	7,37	19,05	0,243

Eine Fehleranalyse, wie in Kapitel 3.5.2 beschrieben, lieferte die Vertrauensintervalle der Besetzungsanteile für ein Konfidenzniveau von 68% (Tabelle 3-11). Aufgrund der großen Menge zu optimierender Parameter bei vier Zuständen sind die Vertrauensintervalle sehr groß. Die Folgerung daraus ist, dass die in Tabelle 3-10 zusammengefassten Werte aufgrund der Breite der Konfidenzbereiche das System unzureichend charakterisieren: Als Beispiel sei das Konfidenzintervall für ΔG_{21}^0 genannt, das sich von 0,8 kJ mol⁻¹ bis 2,3 kJ mol⁻¹ erstreckt. Dieses Verhalten überrascht nicht und entspricht demjenigen der untersuchten und in Kapitel 3.4 beschriebenen Testdatensätze. Auch fällt auf, dass der Wert für $\Delta \beta_{24}^{0'}$ in einem Bereich liegt, der viel größer als der aus der Physik heraus erwartete liegt.

Ein Vergleich mit den von Kalbitzer et al. publizierten Daten (Kalbitzer et al., 2013) zeigt, dass mit Algorithmus 1 für die vier angenommenen Zustände im Wesentlichen dasselbe Verhalten der Besetzungszahlen der Zustände mit dem Druck bestimmt werden konnte und der Lösungsbereich die publizierten Druckabhängigkeiten der Zustände enthält (Vergleich Abbildung 1-3 mit Tabelle 3-11 und Abbildung 3-55).

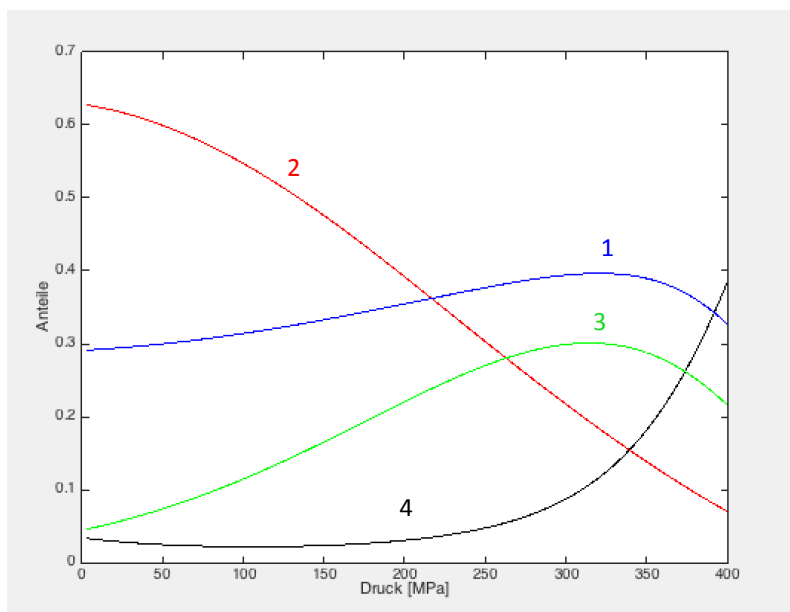


Abbildung 3-55: Populationsverhältnisse von Ras

Die Populationsverhältnisse der Zustände des Proteins Ras, die das mathematische Minimum des Systems nahelegt, wenn alle anzupassenden Parameter frei wählbar sind.

Tabelle 3-11: Die Populationsanteile der vier angenommenen Zustände für das Protein Ras des mathematischen Minimums und in Klammern darunter jeweils das Vertrauensintervall des jeweiligen Wertes bei einem Konfidenzniveau von 68%, das allerdings nur unter der Voraussetzung gilt, dass das mathematische Minimum des Systems korrekt bestimmt wurde

Zustand	Populationsanteil bei		
	bei 3 MPa	100 MPa	bei 200 MPa
2	62,8% (48,2% - 72,2%)	54,8% (41,6% - 65,4%)	39,2% (24,9% - 59,4%)
1	29,2% (19,3% - 40,3%)	31,5% (23,7% - 40,4%)	35,5% (29,4% - 40,5%)
3	4,6% (0,4% - 11,8%)	11,5% (1,8% - 21,8%)	22,1% (4,9% - 34,4%)
4	3,4% (0,0% - 15,8%)	2,2% (0,6% - 13,3%)	3,1% (0,9% - 14,8%)

Abbildung 3-56 zeigt die gemessenen und dann aminosäureabhängig druckkorrigierten chemischen Verschiebungen der Stickstoffatome im Vergleich zu dem berechneten Verlauf, der sich aus den optimierten Parametern des gefundenen Minimums ergibt, exemplarisch für die Aminosäuren Y4, V7, K42 und E49.

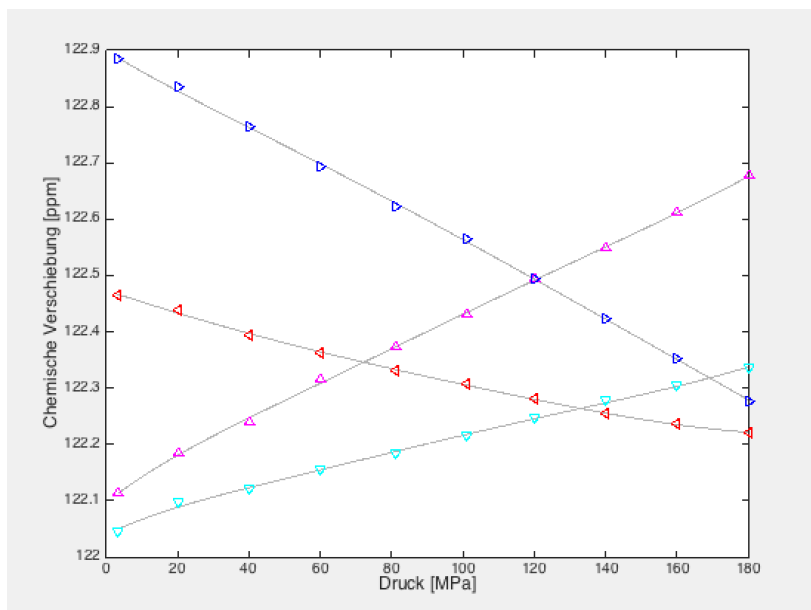


Abbildung 3-56: Vergleich zwischen gemessenen Werten und berechneten Parametern

Gemessene druckkorrigierte chemische Verschiebungen der Stickstoffatome des Proteinrückgrats im Vergleich zu dem aus den optimierten Parametern berechneten Verlauf exemplarisch für vier Aminosäuren:

Cyan: Y4
Magenta: V7
Blau: K42
Rot: E49

Fixierung der Größe ΔG_2^0 bei der Minimumsuche

Um den Parameterraum weiter einzuschränken, wurde in einem zweiten Schritt eine etwas modifizierte Form von Algorithmus 1 verwendet, der bei der Minimumsuche den Parameter ΔG_2^0 festhält, das heißt den Wert der modifizierten freien Gibbs-Energiedifferenz bei Normaldruck zwischen dem an GEFs bindenden und dem an Effektoren bindenden Zustand. Diese Größe ist aus ^{31}P -NMR-spektroskopischen Experimenten bekannt (Kalbitzer et al., 2009), wurde allerdings für eine Temperatur von 278 K bestimmt. Trotzdem wurde näherungsweise eine Kombination mit den zu untersuchenden chemischen Verschiebungen vorgenommen, die bei 303 K aufgenommen worden waren, um die Einsatzmöglichkeiten des Algorithmus zu testen. Abbildung 3-57 zeigt die so bestimmten Besetzungszahlen der Zustände des damit neu bestimmten mathematischen Minimums.

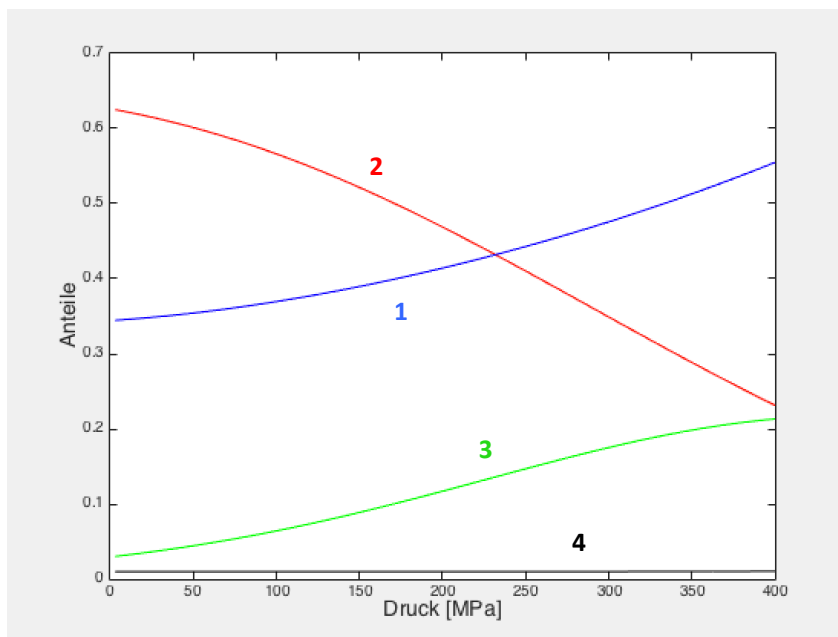


Abbildung 3-57: Populationsverhältnisse von Ras bei Fixierung des Verhältnisses von Zustand 1 und 2 bei Normaldruck

Die Populationsverhältnisse der Zustände des Proteins Ras, die das mathematische Minimum des Systems nahelegt.

In der anschließend durchgeführten Fehlerrechnung zeigte sich, dass auch hier die Stabilisierung einer der drei Gibbs-Energiedifferenzen deutlich bemerkbar ist: Der Lösungsbereich wird stark eingeschränkt. Abbildung 3-58 zeigt exemplarisch den Effekt auf die Anteile der Besetzungszahlen für 20 Parameterkombinationen, die alle zum Konfidenzintervall von 68% gehören.

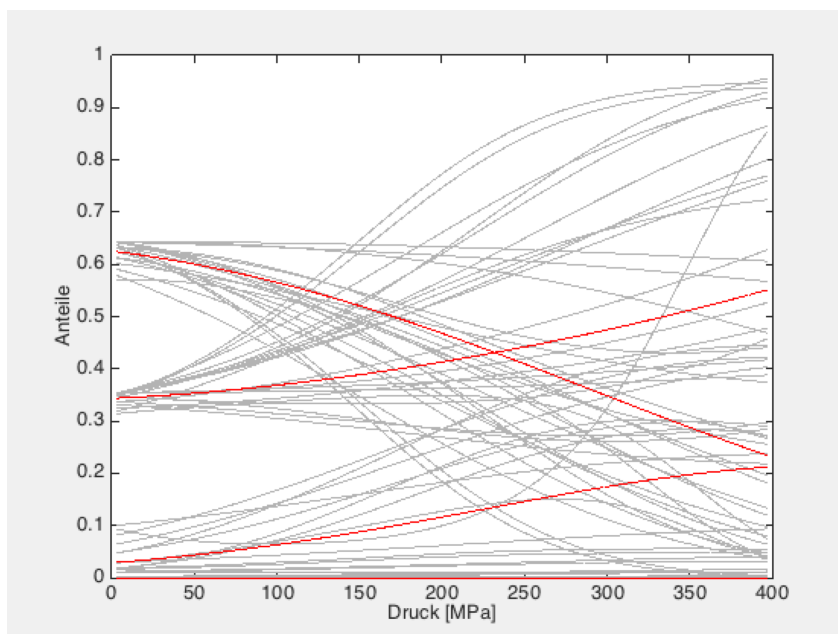


Abbildung 3-58: Fehleranalyse

In Grau 20 Beispiele für Verteilungen der Besetzungszahlen der vier Zustände aus dem Lösungsbereich mit einem Konfidenzniveau von 68% für das Protein Ras in der aktiven Form unter der Annahme, dass ΔG_2^0 den Wert $1,5 \text{ kJ mol}^{-1}$ hat

In einem nächsten Schritt wurden aus dem Lösungsbereich diejenigen Datenkombinationen herausgefiltert, bei denen der Schnittpunkt der beiden höchstbesetzten Zustände bei etwa 125 MPa liegt, da bei diesem Druck mittels ^{31}P -NMR-spektroskopischen Experimenten eine

Gleichbesetzung des GEF-bindenden und des Effektor-bindenden Zustands beobachtet worden war (Kalbitzer et al., 2009). Auf diese Weise konnte der Lösungsbereich weiter eingeschränkt werden: Aus den der Lösung zugehörigen Punkten im Parameterraum wurden diejenigen ausgewählt, bei denen die beiden betrachteten Zustände ein Gleichgewicht zwischen 100 MPa und 150 MPa einnehmen. Die Bereiche, in denen sich die interessierenden Parameter dann bewegen, sind in Tabelle 3-12 zusammengefasst. Auf diese Weise konnte der anfänglich sehr große Lösungsbereich deutlich eingeschränkt werden, ist aber immer nicht um ein Vielfaches größer als der von Kalbitzer et al. publizierte, bei dem weitere zusätzliche Informationen aus den Signalvolumina verwendet wurden.

Tabelle 3-12: Die Werte für die Differenzen der molaren freien Energie, die der partiellen molaren Volumina und die der molaren Kompressibilitätsfaktoren für das Protein Ras, die sich bei Auswertung unter der Annahme ergeben, dass die vier Zustände im schnellen Austausch miteinander stehen

Übergang	ΔG^0 [kJ mol ⁻¹]	ΔV^0 [mL mol ⁻¹]	$\Delta \beta^{0'}$ [mL MPa ⁻¹ mol ⁻¹]
2 → 1	1,5	-4,69 bis -12,97	0,031 bis 0,134
2 → 3	4,39 bis 20,69	-12,8 bis -23,3	-0,003 bis 0,124
2 → 4	25,9 bis 29,22	9,8 bis 23,8	0,191 bis 0,321

3.7.2 Auswertung der chemischen Verschiebungen des Alzheimer- β -Amyloidpeptids

Der vorliegende Datensatz des Proteins A β zeichnete sich durch eine besonders sorgfältige Datenakquisition bei der Erstellung der Listen chemischer Verschiebungen aus, was sich positiv auf die Größe der statistischen Fehler auswirkte. Wie erste Untersuchungen ergaben, waren die Änderungen der chemischen Verschiebungen mit dem Druck nach der aminosäureabhängigen Druckkorrektur sowohl für die Stickstoffatome des Proteinrückgrats als auch für die ihrer Protonen nur noch sehr klein. Außerdem zeigte sich, dass einige Atome kein stetig steigendes oder fallendes Druckverhalten aufweisen, was die Annahme von mehr als zwei Zuständen rechtfertigt, wie dies auch von Munte et al. bei der Auswertung dieses Datensatzes angenommen worden war (Munte et al., 2013). Ebenso wie bei den anderen beiden experimentellen Datensätzen existieren neben den chemischen Verschiebungsänderungen mit dem Druck auch Variationen in den Signalvolumina. Diese wurden bei der nun folgenden Auswertung vernachlässigt.

Datenkombination nach Schumann et al.

Da das Protein A β mit insgesamt nur 40 Aminosäuren eher klein ist und von diesen nur für 31 Reste chemische Verschiebungen vorlagen, wurden die korrigierten chemischen Verschiebungen der Stickstoffatome des Proteintrückgrats und die ihrer Protonen, wie in Kapitel 2.2.4 nach Schumann et al. (Schumann et al., 2007) beschrieben, kombiniert. Das mathematische Minimum des Systems hat eine Norm von 0,0143, dessen Parameter in Tabelle 3-13 zusammengestellt sind, und Abbildung 3-59 zeigt die daraus resultierende Verteilung der Zustände.

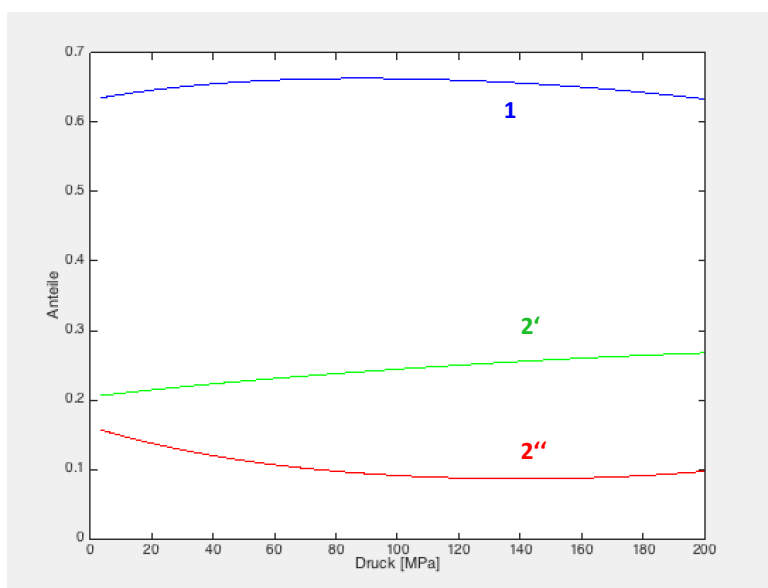


Abbildung 3-59: Populationsverhältnisse für A β , wenn die Daten der Stickstoffatome und die ihrer Protonen nach Schumann et al. kombiniert werden

Die Besetzungszahlen wurden aus den Parametern des mathematischen Minimums des Systems berechnet.

Tabelle 3-13: Die Werte für die Differenzen der Gibbs-Energie, die der partiellen molaren Volumina und die der molaren Kompressibilitätsfaktoren für das Protein Alzheimer- β -Amyloidpeptid des mathematischen Minimums des Systems

Übergang	ΔG^0 [kJ mol ⁻¹]	ΔV^0 [mL mol ⁻¹]	$\Delta \beta^0$ [mL MPa ⁻¹ mol ⁻¹]
1 \rightarrow 2'	1,79	37,4	0,2900
1 \rightarrow 2''	6,82	-7,00	0,0098

Die durchgeführte Fehleranalyse ergab, dass die Funktion Norm über dem Bereich im Parameterraum, in dem das Minimum liegt, in einem sehr flachen Tal liegt. Damit wird der Lösungsbereich sehr groß. Eine Untersuchung der resultierenden Anteile der Besetzungen möglicher Parameterkombinationen, die dem Lösungsbereich angehören, zeigen einen annähernd parallelen Verlauf zu den in Abbildung 3-59 dargestellten Verläufen der Besetzungszahlen mit dem Druck, die dem mathematischen Minimum entsprechen, und sind nur nach oben oder unten verschoben. Allerdings streuen die Besetzungszahlen so weit, dass

sich die Bereiche der Zustände stark überlappen. So kann aus dieser Analyse nur die Aussage gewonnen werden, dass sich die Besetzungszahlen mit dem Druck kaum verändern. Insbesondere ist eine Bestimmung der prozentualen Verteilung der Zustände an den gemessenen Druckpunkten nicht möglich.

Datenkombination über das gyromagnetische Verhältnis

Zusätzlich wurde eine zweite Möglichkeit der Datenkombination getestet, indem die korrigierten chemischen Verschiebungen der Stickstoffatome und ihrer Protonen über das gyromagnetische Verhältnis gewichtet zu einer Matrix D mit einer Zeilenlänge von 62 zusammengefasst wurden.

Die Anwendung von Algorithmus 1 auf die so kombinierten Daten lieferte ein Minimum mit einer sehr kleinen Norm von nur 0,008. Die diesem Extremum entsprechenden Parameter (Tabelle 3-14) des Systems ergab die in Abbildung 3-60 gezeigte Verteilung der Zustände. Auch diese ergibt Besetzungszahlen, die sich unter Hochdruck kaum verändern.

Tabelle 3-14: Die Werte für die Differenzen der molare Gibbs-Energien, die der partiellen molaren Volumina und die der molaren Kompressibilitätsfaktoren für das Protein A β

Übergang	ΔG^0 [kJ mol ⁻¹]	ΔV^0 [mL mol ⁻¹]	$\Delta\beta^{0'}$ [mL MPa ⁻¹ mol ⁻¹]
1 \rightarrow 2'	2,77	1,53	0,0005
1 \rightarrow 2''	5,92	17,67	0,1266

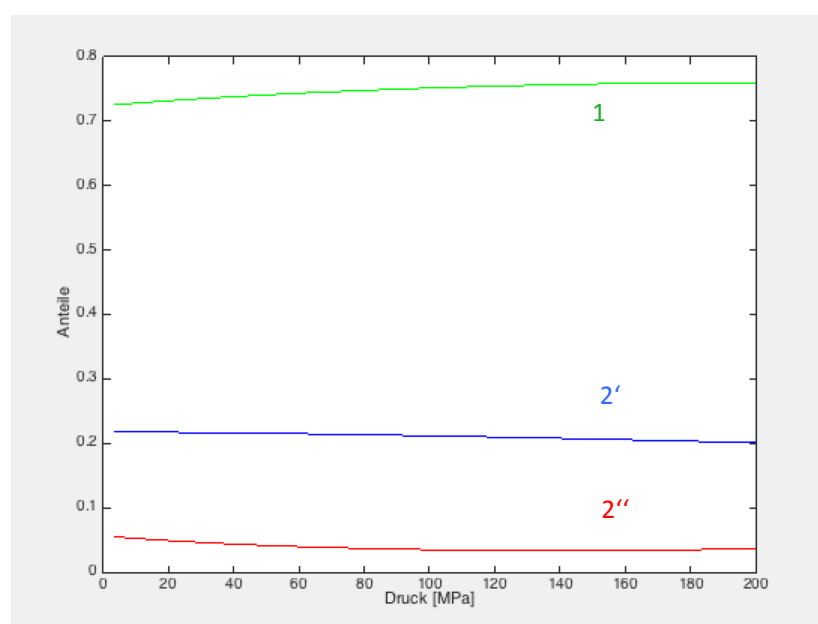


Abbildung 3-60: Besetzungszahlen in Abhängigkeit vom Druck

Die Populationsverhältnisse der Zustände des Proteins A β , die das mathematische Minimum des Systems nahelegt, wenn die verwendeten Daten der Stickstoffatome und ihrer Protonen nach ihrem gyromagnetischen Verhältnis gewichtet in die Auswertung einfließen.

Die durchgeführte Fehleranalyse ergab, dass das Minimum im Gegensatz zu der Gewichtung nach Schumann etwas stabiler ist und hier der Lösungsbereich im Parameterraum besser eingegrenzt werden kann: Im Vertrauensintervall (Konfidenzniveau 68%) weist Zustand 1 immer einen Besetzungszahlanteil von über 50% auf. Alle im Lösungsbereich befindlichen Parameterkombinationen zeigen einen fast linearen Verlauf mit dem Druck. Die beiden unterschiedlichen Gewichtungsmethoden liefern damit im Wesentlichen die gleichen Ergebnisse.

Ein Vergleich mit dem publizierten Verlauf der Besetzungszahlen (Abbildung 1-5) (Munte et. al., 2013) zeigt deutliche Unterschiede und kann mit den im Rahmen dieser Arbeit durchgeführten Rechnungen - die nur auf der Auswertung der chemischen Verschiebungen beruhen - nicht nachvollzogen werden. Für die dort dargestellten Ergebnisse war derselbe experimentelle Datensatz untersucht worden, weswegen zum Vergleich die Norm an dem Punkt im Parameterraum berechnet wurde, der den publizierten Parametern entsprach. Diese liegt mit 0,0137 deutlich höher als die Norm der mit dem Verfahren nach Levenberg-Marquardt identifizierten Werte.

Mit der gleichen Gewichtung wurde des Weiteren probenhalber eine Testrechnung unter der Annahme durchgeführt, dass nur zwei relevante Zustände in dem beobachteten Druckbereich vorliegen. Das dabei gefundene Minimum hatte eine Norm von 0,0418 und liegt damit etwa um das fünffache höher als die kleinste Norm, die man erhält, wenn man drei Zustände annimmt. Dies deutet darauf hin, dass, obwohl sich die Besetzungszahlen des dritten Zustands kaum mit dem Druck verändern und sehr klein sind, es sich aber um eine echte weitere relevante Konfiguration handelt, deren chemische Verschiebungen von denen der anderen Zustände zumindest für einige Aminosäuren deutlich abweicht.

Zusammenfassend muss daher festgestellt werden, dass das gewählte Modell, dass drei Zustände im schnellen Austausch miteinander stehen, ein deutlich anderes Ergebnis liefern als ein Modell, das langsamen Austausch mitberücksichtigt, wie es von Munte et al. zur Datenauswertung gewählt wurde.

4 Diskussion

4.1 Beurteilung der Methode und Ergebnisse von PERMOL

Bei der Neuimplementierung des in AUREMOL integrierten Programms PERMOL stand im Vordergrund, die guten Leistungen der Vorgängerversion zur Erstellung von Homologiemodellen und zur Strukturverbesserung auch auf die Vorhersage von Komplexstrukturen zu übertragen, da gerade hier durch die große Anzahl an Aminosäuren die klassischen Methoden der experimentellen Strukturbestimmung - Kernspinresonanzspektroskopie und Röntgenspektroskopie - an ihre Grenzen stoßen. Im Folgenden werden nun die in Kapitel 3.1 und 3.2 beschriebenen Ergebnisse der auf der Methode von PERMOL durchgeführten Strukturrechnungen diskutiert.

4.1.1 Rekonstruktion von Byr2-RBD aus PDB-1K8R

Der erste beschriebene Test von PERMOL (Kapitel 3.1.1) stellt einen einfachen Funktionstest dar, der zeigen soll, ob das Programm in der Lage ist, einen konsistenten Satz Restraints so zu erzeugen, dass dieser das Testprotein zusammen mit dem *a priori* Wissen über Proteine - die typischen Bindungslängen und Winkel, Ladungen, die Größen der Pseudoteilchen und vieles andere mehr - , das bei den Simulationen mit CNS über die physikalischen Potentiale in die Strukturrechnung miteinfließt, wieder rekonstruieren kann.

Als Teststruktur wurde die mithilfe von Röntgenkristallographie gewonnene Konfiguration von Byr2-RBD aus PDB-1K8R (Scheffzek et al., 2001) verwendet. Die Erzeugung von Abstandsrestraints, dihedralen Winkelrestraints und Wasserstoffbrückenrestraints aus den vorhandenen Atomkoordinaten und die damit durchgeführten Strukturrechnungen mit anschließendem Waterrefinement ergaben ein enges Strukturbündel, das die ursprüngliche, zu rekonstruierende Struktur in weiten Teilen der Sequenz einschließt. Die geringfügigen Abweichungen im Bereich des ersten und zweiten β -Strangs und des C-Terminus finden ihre Erklärung vermutlich darin, dass in der Struktur PDB-1K8R, das als Modell diente, Byr2 eigentlich im Komplex mit Ras vorhanden ist und die Bindestelle genau den genannten Bereich abdeckt. Eine leichte Drehung des Turns, wie sie hier beobachtet wurde (Abbildung 3-3), beruht wohl darauf, dass diese Stellung ohne den Bindungspartner Ras energieärmer und damit wahrscheinlicher ist. Sie darf deswegen nicht als Nachteil der Methode gedeutet werden, sondern vielmehr als deren Vorteil, sich Veränderungen der Umgebung- zumindest bis zu einem gewissen Grad - anpassen zu können.

Die Rechnungen zeigen auch, dass die Methode in der Lage ist, durch die getroffene Auswahl der Parameter insbesondere die Sekundärstrukturelemente gut rekonstruieren zu können, also aus dem Modell in adäquater Weise zu abstrahieren und als Restraints für die Strukturrechnung wieder zugänglich zu machen.

An den Energiebeiträgen ist ersichtlich, dass die erstellten Restraints in sich konsistent und auch mit den physikalischen Potentialen von CNS kompatibel sind. Insbesondere wurden auch keine Zwangsbedingungen erzeugt, die nicht erfüllbar wären.

4.1.2 Verbesserung der Struktur von Byr2 aus NMR-spektroskopischen Daten

Die in Kapitel 3.1.2 dargestellten Ergebnisse zur Verbesserung der NMR-Struktur von Byr2-RBD - in der PDB-Datenbank unter der Nummer 1I35 zu finden - zeigen, wie sich die von PERMOL erzeugten Restraints als Ersatzrestraints anstatt der experimentell gefundenen verwenden lassen. Dieses Verfahren kann, wie von Cano et al. beschrieben, als Imputation gesehen werden, bei der die fehlenden oder falschen Informationen eines eher schlechten experimentellen Datensatzes durch ein Set von durch Schätzwerte gewonnenen Restraints ersetzt werden kann (Cano et al., 2009). Diese Schätzwerte stammen aus Strukturen, die anhand des fehlerhaften Datensatzes berechnet werden.

Im vorliegenden Fall wurde ein Teil der Daten, nämlich die dihedralen Winkel, durch neu generierte ersetzt, die mit der aktuellen Version des Programms Talos+ erzeugt worden waren, und zusätzlich der Einsatz der dipolaren Restkopplungsrestraints verbessert. Die mit diesem neuen Datensatz erzeugten Strukturen - Strukturbündel 1, das außerdem einem *Waterrefinement* unterzogen wurde,- zeigten bereits eine sichtliche Verbesserung im Vergleich zu dem ursprünglich berechneten Strukturbündel PDB-1I35: Im Ramachandran-Plot stellt man eine deutliche Verschiebung hin zu den erlaubten Bereichen fest, was die Strukturen wahrscheinlicher macht: Während bei PDB-1I35 nur 67,8% der Winkelkombinationen der betrachteten Aminosäuren im *most favoured* Bereich liegen, entfallen bei Strukturbündel 1 87,15 % in diese Kategorie. Ebenso ist der RMSD sowohl der schweren Atome des Proteinrückgrats als auch aller schweren Atome deutlich geringer geworden.

Zusätzlich zu dieser bereits deutlich verbesserten Qualität des Strukturbündels wurde im darauffolgenden Schritt, dem Einsatz von mit PERMOL erzielten Ersatzrestraints, eine

weitere Verfeinerung der erhaltenen Konfigurationen erzielt. Wie die Tabelle 3-1 belegt, verringert sich bei den Rechnungen mit den neu generierten Restraints vor allem der RMSD-Wert noch mehr.

Zum Vergleich wurden zwei Strukturbündel gerechnet, deren Unterschied in der Erzeugung darin lag, dass bei einer der beiden Rechnungen zusätzlich die experimentellen dipolaren Restkopplungsrestraints dazugenommen wurden. Das so erzeugte Bündel - Strukturbündel 3 - zeigt im Vergleich zu Strukturbündel 1 eine leichte Verbesserung des R-Faktors, das heißt, es passt besser zu dem zugehörigen NOESY-Spektrum, aus dem die ursprünglichen Abstandsrestraints extrahiert worden waren: Dabei wurde der Faktor der langreichweitigen Signale betrachtet, der sich von 0,22 auf 0,21 verbesserte. Die Informationen über dipolare Restkopplungen ergänzen offensichtlich die eher lokalen Informationen über Abstände und Winkel gut, da sie die Orientierung verschiedener kovalenter Bindungen, im vorliegenden Fall zwischen den N- und HN-Atomen des Backbones, miteinander verknüpfen und somit globale Informationen in die Strukturrechnung mit einbringen. Dadurch wird die Stellung der Sekundärstrukturelemente zueinander optimiert, was auch an dem berechneten Q-Value ersichtlich ist, der mit einer Größe von nur 0,0046 einen sehr guten Wert aufweist.

Dieser Wert ist bei dem vergleichenden Strukturbündel 2, das ohne den expliziten Einsatz der residualen Restkopplungsrestraints berechnet worden ist, jedoch um gleich zwei Größenordnungen höher. Es sind also die Informationen, die in Strukturbündel 1 über die residualen Restkopplungen bereits enthalten waren, durch die Erstellung der Ersatzrestraints offensichtlich wieder teilweise verlorengegangen. Das deutet darauf hin, dass die von PERMOL extrahierten Informationen leichte Schwachpunkte aufweisen, wenn die Parameter - wie in der Arbeit von Cano et al. als ideal publiziert - gewählt werden (Cano et al., 2009). Eine Möglichkeit, dem System globale Informationen hinzuzufügen, wäre, die Abstandsrestraints, dihedralen Winkelrestraints und die Informationen über Wasserstoffbrücken durch Restraints in der Art der residualen Restkopplungen zu ergänzen, um die neu implementierten Version leicht erweitert werden könnte.

4.1.3 Homologiemodelle des Komplexes Byr2-RBD mit Ras

Um das Verfahren der Homologiemodellierung von Komplexen zu testen, wurde das bereits mehrfach in dieser Arbeit untersuchte Protein Byr2-RBD ausgewählt und Modelle zusammen

mit seinem Interaktionspartner Ras berechnet. Die Ergebnisse konnten mit der durch Röntgenkristallographie gewonnenen Struktur PDB-1K8R verglichen werden.

Die erste Schwierigkeit dabei lag darin, geeignete Modellierungsmodelle zu finden, die der Informationsgewinnung dienen konnten. In der Literatur sind einige weitere Bindungspartner des Proteins Ras eingehend beschrieben worden, die eine Strukturhomologie zu Byr2 zeigen. Experimentell bestimmte Strukturen lagen allerdings nur für RalGDS und C-Raf im Komplex mit Ras vor, die, wie auch die anderen bekannten Bindungspartner von Ras, nur eine eher geringe Sequenzähnlichkeit zu der Ras bindenden Domäne von Byr2 von etwa 15% aufweisen. Für das Alignment wurde daher nicht auf die von PERMOL berechnete - mit dem auf dem Algorithmus nach Needleman und Wunsch beruhenden - Zuordnung der Aminosäuren zurückgegriffen, sondern eine von Wohlgemuth et al. vorgeschlagene verwendet, die auf einem Sekundärstrukturvergleich der Ras-bindenden Proteine beruht (Wohlgemuth et al., 2005). Diese Zuordnung ist sinnvoll und daran erkennbar, dass die Sekundärstrukturelemente bei den berechneten Modellen - die einmal auf der Grundlage von PDB-1LFD (RalGDS im Komplex mit Ras), einmal von PDB-4G0N (C-Raf im Komplex mit Ras) und als Drittes auf einer Kombination von diesen beiden Strukturen berechnet wurden - im Vergleich zu der durch Röntgenkristallographie gewonnenen Konfiguration PDB-1K8R dieses Komplexes an den richtigen Stellen wiedergegeben werden (Tabelle 3-3).

Außer dem Auffinden eines passenden Alignments lag bei der getesteten Methode eine weitere Schwierigkeit in der Handhabung des Programms CNS, dessen Einsatz zur Strukturrechnung für Restraints, die mit experimentellen Methoden gewonnen worden waren, optimiert ist. Deren Anzahl ist im Allgemeinen viel kleiner als die Menge der mit PERMOL generierten Restraints. Neben kleinerer Anpassung im Quellcode selbst, um die so entstehenden Restraints einlesen und verarbeiten zu können, mussten die den Programmablauf steuernden Skripten so erweitert werden, dass die gleichzeitige Optimierung mehrerer Sequenzstücke möglich ist. Des Weiteren mussten, wie in Kapitel 2.3.8 ausführlich beschrieben, die Skalierungsparameter, die Anzahl der Iterationsschritte und die Zeitintervalle zwischen den Energie- und Kraftberechnungen beim Simulated Annealing angepasst werden. Als Maßstab dafür, wann die gewählten Parameter optimale Werte aufweisen, zählte das Kriterium, ein Minimum der Energiebeiträge aus den physikalischen Potentialen anzustreben. Dies ist deswegen als sinnvoll anzusehen, da nur so gewährleistet ist, dass "entspannte" Strukturen resultieren, also Strukturen, deren Bindungslängen und Winkel optimal sind und bei denen keine Atome Überlappungen aufweisen. Insbesondere bei der Erstellung von

Homologiemodellen, bei denen die in den Restraints zusammengefassten Informationen von strukturhomologen, aber eben nicht sequenzgleichen Proteinen stammen, dürfen die aus den Restraints resultierenden Kräfte zwar die Strukturen formen. Einzelne Bedingungen aber, die aufgrund einer veränderten Umgebung - wie zum Beispiel beim Ersetzen einer Aminosäure durch eine größere - nicht mehr realisierbar sind, dürfen die Strukturen nicht in mit den physikalischen Potentialen unvereinbare Konfigurationen verschieben. Vielmehr müssen die Restraints als mögliche wahrscheinliche Vorschläge gesehen werden, die bei Nichterfüllung im Gegensatz zu experimentellen Restraints nicht auf einen Mangel in der gefundenen Konfiguration hinweisen, sondern Folge der Anpassung an die veränderte Sequenz sind. Insofern muss bei der Bewertung der Strukturen auch die restraintkorrigierte Gesamtenergie, also abzüglich der Beiträge aus den restraintbasierten Energien, als Maßstab dienen.

Bei der Beurteilung der drei berechneten Strukturbündel aus den verschiedenen Modellierungsgrundlagen fällt zuallererst auf, dass die Methode trotz der geringen Sequenzähnlichkeit in der Lage ist, die Bindestelle und insbesondere auch die intersequentiellen Wasserstoffbrücken korrekt vorherzusagen, was als Stärke dieses Verfahrens hervorzuheben ist.

Bei den ersten beiden vorgestellten Modellierungen, bei denen die Ras-Bindedomänen von RalGDS beziehungsweise von C-Raf, jeweils im Komplex mit Ras, als Informationsquelle für mögliche Abstände, Winkel und Wasserstoffbrücken dienten, scheint auf den ersten Blick die Struktur, die PDB-1LFD (RalGDS) als Grundlage hatte, die überlegeneren Ergebnisse zu liefern, da die Sekundärstrukturelemente wesentlich besser geformt sind und auch deren Stellung zueinander sehr gut modelliert wurde (Abbildungen 3-7 und 3-8). Dagegen weisen insbesondere die β -Stränge des Strukturbündels, bei dem PDB-4G0N als Grundlage diente, eine eher schlechte Geometrie auf. Eine mögliche Ursache liegt darin, dass die Sequenz von C-Raf im Vergleich zu RalGDS nochmals weniger Übereinstimmung zeigt und so insgesamt fast 3000 Restraints weniger erzeugt werden konnten, was sich auch in dem schlechteren RMSD-Wert der schweren Atome des Proteinerückgrats von etwa 0,093 nm widerspiegelt. Trotzdem werden auch hier die Wasserstoffbrücken fast korrekt modelliert. Bei beiden erzeugten Strukturbündeln sind allerdings mehr Aminosäuren von Byr2 in engem Kontakt mit Ras als bei der Röntgenstruktur des untersuchten Komplexes, PDB-1K8R.

Das Bild, PDB-1LFD sei eindeutig die bessere Modellierungsgrundlage, wird jedoch relativiert durch die Tatsache, dass die erhaltenen Strukturen trotz passender Bindestelle und Fixierung durch Wasserstoffbrücken auf der Oberfläche von Ras im Vergleich zu PDB-1K8R um etwa 30° verdreht erscheinen (Abbildung 3-9). Wie ein Vergleich der Energien (Tabelle 3-4) zeigt, entstehen durch PDB-4G0N als Modell Strukturen, deren Energiebeiträge aus den physikalischen Potentialen geringer sind als bei PDB-1LFD als Modellierungsgrundlage. Dies deutet darauf hin, dass die Geometrie der mit PDB-4G0N erzeugten Konfigurationen besser zur Sequenz von Byr2-RBD passt.

Als Drittes wurde ein Strukturbündel mit PDB-1LFD und PDB-4G0N als Informationsquellen berechnet. Das Ergebnis sind Konfigurationen, deren Sekundärstrukturelemente (Tabelle 3-3) und RMSD vergleichbar sind mit denen, die unter diesen Gesichtspunkten zuerst deutlich besser abschnitten und als Modellierungsgrundlage die Struktur PDB-1LFD hatten. Im Gegensatz zu diesen zeichnen sie sich jedoch durch eine deutlich bessere Orientierung von Byr2 auf der Oberfläche von Ras aus (Abbildung 3-11). Ein Vergleich der Positionen der der Bindestelle angehörenden Aminosäuren ergibt eine sehr gute Übereinstimmung der Lage zu PDB-1K8R. Lediglich die Ausrichtung einiger Seitenketten (Abbildung 3-12) deutet auf die Notwendigkeit hin, ein weiteres Refinement der Strukturen durch eine abschließende Energieoptimierung ohne die von PERMOL erzeugten Restraints durchzuführen.

Abbildung 1-4 zeigt die Lage der beim Alignment einander zugeordneten positiv geladenen Aminosäuren Arginin und Lysin der Strukturen der Ras-Bindedomänen von RalGDS, C-Raf und Byr2. Es ist zu erkennen, dass deutliche strukturelle Unterschiede zwischen diesen drei Ras-bindenden Proteinen vorliegen, so dass das Ergebnis des Homologiemodells, das insbesondere die Bindestelle sehr gut wiedergibt, positiv zu bewerten ist. Somit ist die Methode in der Lage, die Strukturänderungen und nötigen Anpassungen, die aus dem Austausch der Aminosäuren resultieren, adäquat bei der Formung der neuen Modelle zu berücksichtigen, und auch die Idee, die Informationen aus mehreren PDBs zu kombinieren und über die Berechnung von lokalen Scores je nach Grad der Übereinstimmung mit der Sequenz des zu modellierenden Moleküls zu gewichten, wurde sinnvoll und gewinnbringend umgesetzt, wie dies auch in ähnlicher Art und Weise von Merkl und Waack (2009) vorgeschlagen wird.

4.1.4 Homologiemodelle von IE1-CTD im Komplex mit menschlichem Chromatin

Im Gegensatz zu den eben diskutierten Ergebnissen der Komplexstruktur von Byr2 mit Ras lag für die Modellierung des C-terminalen Endes des Proteins IE1 des Cytomegalievirus, das den Herpesviren angehört, im Verbund mit menschlichem Chromatin keine aus experimentellen Daten gewonnene Struktur vor, die zur Validierung hätte dienen können. Vielmehr sollte geprüft werden, ob ein ähnlicher Bindungsmechanismus wie er für das Protein LANA des Karposisarkom-assoziierten Herpesvirus bekannt vorliegen könnte, und das mit Röntgenkristallographie untersucht worden war (PDB-1ZLA). Hierbei zeigte sich, dass die Bindung bei LANA im Gegensatz zu IE1 N-terminal geschieht. Die beteiligten Aminosäuren formen dabei eine β -Haarnadelstruktur aus.

Für die Aminosäuren des C-Terminus von IE1 war bekannt, dass sie die Bindung in die "saure Tasche" - den einzigen größeren zugänglichen negativ geladenen Bereich auf der Oberfläche der Histone, die durch die Proteine H2A und H2B gebildet wird - ausüben, in die auch LANA bindet. Verschiedene Programme zum Alignen der Sequenzen lieferten zwei Zuordnungen, für die Modelle gerechnet wurden.

Diese beiden Modelle wurden unter verschiedenen Gesichtspunkten untersucht. Dabei wurde festgestellt, dass das in Kapitel 3.2.2 als Modell 1 bezeichnete Strukturbündel dem anderen Modell deutlich überlegen ist, da in Modell 2 ein vorhandenes Prolin die Ausbildung der Faltblattstruktur stört, wohingegen diese Aminosäure im anderen berechneten Modell im Turn des Haarnadelmotivs zu liegen kommt. Des Weiteren zeichnet sich das bevorzugte Strukturbündel aufgrund von vier das Faltblatt stabilisierenden und zehn intersequenziellen Wasserstoffbrücken zu H2A oder H2B deutlich vor den anderen berechneten Konfigurationen aus, die nur insgesamt acht entsprechende Nebervalenzbindungen vorweisen können. Insbesondere sagt Modell 1 auch vier Wasserstoffbrücken vorher - zwei intra- und zwei intersequentielle - die in der Struktur PDB-1ZLA, also in der Modellierungsvorlage LANA mit den Histonen, nicht vorhanden sind. Auch die Mutageneseexperimente, die parallel zu diesen Berechnungen von Mitarbeitern des Institutes für Medizinische Mikrobiologie und Hygiene der Universität durchgeführt wurden, passen ebenfalls besser zu Modell 1. Eine sehr gute Ladungs- und Formkomplementarität sprechen ebenso für dieses Strukturbündel.

Unter dem Gesichtspunkt möglicher Zustandsmodelle stellte sich die Frage, in welcher Form der C-Terminus von IE1 ohne die räumliche Nähe von menschlichem Chromatin vorliegt. Wie die dazu durchgeführten NMR-spektroskopischen Untersuchungen zeigten, ist dieses

Sequenzstück sehr beweglich. Die Suche nach Anhaltspunkten dafür, dass bereits vor der Bindung in die saure Tasche eine Haarnadelstruktur vorliegen könnte, verlief negativ, da keine NOE-Kontakte zwischen den beiden β -Strängen gefunden werden konnten, die dies bestätigen würden. Dagegen konnte ein zweiter beweglicher Bereich des Proteins IE1, die Aminosäuren A248 bis D269, identifiziert werden.

Bei der kürzlich veröffentlichten Struktur PDB-4WID (Scherer et al., 2014) handelt es sich um die Konfiguration von rhIE1 des Rhesusaffen, dessen Primärsequenz dem hier untersuchten Protein IE1 sehr ähnlich ist. Anhand eines Sequenzvergleichs (Mücke et al., 2005) lässt sich die Lage des zweiten als beweglich identifizierten Bereiches auf ein Gebiet projizieren, dessen Mitte in den Loop zwischen Helix 8 und Helix 9 (siehe Abbildung 1-6) fällt. Daher kann vermutet werden, dass diese Stelle als Scharnier zwischen diesen beiden Helices dienen könnte und eine funktionelle Bedeutung hat.

Die experimentell durchgeführten Analysen sind gut damit vereinbar, dass die Bindung von IE1 an das menschliche Chromatin in gleicher oder sehr ähnlicher Weise, wie das berechnete Modell vorschlägt, stattfinden könnte. Einen Beweis für die Richtigkeit der berechneten Konfiguration kann jedoch nur eine experimentelle Strukturbestimmung, wie für LANA durchgeführt, bringen.

4.1.5 Zusammenfassende Beurteilung von PERMOL

Die im Rahmen dieser Arbeit durchgeführten Strukturrechnungen zeigen, dass die Neuimplementierung und Umstrukturierung des in AUREMOL integrierten Moduls PERMOL wie die Vorgängerversion effektiv zur Erstellung von Homologiemodellen und zur Strukturverbesserung eingesetzt werden kann. Mit der neu implementierten Version besteht nun auch die Möglichkeit, Restraints für zwei oder mehr Sequenzstücke zu generieren, wodurch sich eine neue Einsatzmöglichkeit bei der Berechnung von Protein-Protein-Komplexen eröffnet.

Wie die vorgestellten Beispiele darlegen, können selbst bei geringer Sequenzähnlichkeit des zu ersetzenden Interaktionspartners Modelle erzeugt werden, bei denen die Bindestelle und die Orientierung gut aneinander angepasst sind. Die von PERMOL vorgeschlagenen Restraints formen die beiden wechselwirkenden Moleküle und bringen sie offensichtlich durch die daraus bei der Strukturrechnung resultierenden Potentiale so zueinander, dass der

Abstand zum Partner passend ist. Folglich kann dann das Verfahren des *Simulated Annealings* und des abschließenden *Waterrefinements* durch kleine Änderungen der Winkel des Proteinrückgrats und der Ausrichtung der Seitenketten energetisch günstige Konfigurationen finden. Dadurch wird die geometrische Komplementarität optimiert. Allerdings sind dabei durch die Restraints Grenzen gesetzt, so dass für optimale Ergebnisse eine weitere Strukturverfeinerung ohne den Einsatz von Restraints folgen müsste.

Im Gegensatz dazu liegt bei dem Programm HOMCOS (Fukuhara und Kawabata, 2008), das ebenfalls zur Vorhersage von Proteinkomplexen entwickelt wurde, der Schwerpunkt auf dem Auffinden homologer Strukturkomplexe. Die Erzeugung des Modells jedoch erfolgt nur durch einfaches Ersetzen der Atome durch diejenigen der zueinander alignnten Aminosäuren. Detailliertere Anpassungen, insbesondere bei so geringer Sequenzähnlichkeit wie bei den im Rahmen dieser Arbeit untersuchten Bindungspartnern von Ras - Byr, Raf und RalGDS - oder Veränderungen beim Alignment sind hierbei nicht vorgesehen. Außerdem steht nur die Eingabemöglichkeit für zwei Sequenzstücke zur Verfügung. Damit wäre die Vorhersage so komplexer Modelle wie das des C-Terminus von IE1 mit der sauren Bindetasche, die durch Teile der Proteine H2A und H2B gebildet wird, mit dieser Software genauso wenig möglich gewesen wie mit dem ebenfalls zur Berechnung von Homologiemodellen erstellten Programm SWISS-MODEL (Arnold et al., 2005).

Mit der Neuimplementierung von PERMOL konnte damit eine Lücke in Bereich der Modellierung von Komplexen geschlossen und ein vielfältig einsetzbares Werkzeug geschaffen werden, welches durch die GUI-Programmierung mit einer optisch ansprechenden und benutzerfreundlichen Oberfläche ausgestattet wurde, wie sie auch in den anderen Modulen in AUREMOL angestrebt wird. Die Sequenz des zu untersuchenden Makromoleküls kann direkt im 1-Letter-Code eingegeben oder die Möglichkeit genutzt werden, dass diese unmittelbar aus einer PDB-Datei extrahiert wird, was vor allem für den Einsatz zur Strukturverbesserung anwenderfreundlich ist. Das berechnete Alignment ist nun direkt editierbar, und es können dadurch sehr einfach und schnell mehrere Möglichkeiten der Aminosäurezuordnungen getestet werden, was insbesondere bei der Simulation von Komplexen hilft, die am besten passende Zuordnung zu finden. Zusätzlich ist die Auswahl der Aminosäuren, für die die Restraints erstellt werden sollen, durch einfaches Anwählen mit der Maus möglich.

4.2 Beurteilung der Methoden und Ergebnisse von MultiStates

Die Idee, aus dem Verhalten von Proteinen unter Hochdruckeinfluss auf die Anzahl der relevanten konformationellen Zustände und ihre Besetzungsverhältnisse zueinander zu schließen, ist seit mehreren Jahren Gegenstand der Forschung (Akasaka, 2006). Ausgangspunkt dabei ist die Annahme, dass die Zustände einer Boltzmannverteilung gehorchen. Durch Variation des Drucks können ihre Gibbs-Energien beeinflusst werden, was Populationsverschiebungen nach sich zieht und über die Veränderung der chemischen Verschiebungen verfolgt werden kann. Die Untersuchungen der letzten Jahre ergaben, dass die Aminosäuren je nach Größe und Aufbau ein leicht verschiedenes, jedoch spezifisches Verhalten an den Tag legen. Erst nach einer Korrektur der gemessenen chemischen Verschiebungen mit druckabhängigen Random-Coil-Werten (Köhler et al., 2012) kann davon ausgegangen werden, dass die verbleibenden Änderungen den strukturspezifischen Veränderungen des Proteins mit dem Druck zuzuordnen sind und damit - im Bild der Zustände - eine Populationsverschiebung zwischen den Zuständen des Proteins anzeigen.

Bei der bisher angewandten Methode zur Untersuchung dieses strukturabhängigen Anteils (Kremer et al., 2007; Kalbitzer et al., 2013; Munte et al., 2013) wurden die chemischen Verschiebungen aminosäureweise betrachtet und eine Abschätzung der interessierenden Größen vorgenommen. Über eine Analyse ihrer Verteilungen schloss man dann auf die Parameter der Funktion der Gibbs-Energiedifferenzen zwischen den Zuständen. Zur Vereinfachung der Analysen wurde bei einigen untersuchten Datensätzen die Taylornäherung dieser Größe nach dem linearen Glied abgebrochen und damit der Einfluss der molaren partiellen Kompressibilitätsfaktoren vernachlässigt.

Ziel dieses Projekts war es nun, eine schnellere und robustere Methode zu entwickeln, bei der die Informationen aus den chemischen Verschiebungen aller Aminosäuren gleichzeitig verarbeitet werden können. Der dazu entwickelte Algorithmus und die Ergebnisse dazu sollen nun im Folgenden diskutiert werden.

4.2.1 Anwendung des Levenberg-Marquardt-Algorithmus auf das vorliegende Problem

Gegeben waren die bereits druckkorrigierten chemischen Verschiebungen der betrachteten Atome an den gemessenen Druckpunkten. Um nun die gesuchten Besetzungszahlen der einzelnen Zustände und deren Verhalten mit dem Druck zu finden, musste zunächst eine geeignete Funktion entwickelt werden, deren Minimum auf die Gibbs-Energiedifferenzen und

deren Druckabhängigkeit Rückschlüsse zulässt und die mit den Messwerten am besten verträglich ist.

Dabei wurde folgendes Vorgehen gewählt: Um die Messwerte alle gleichzeitig betrachten zu können, wurden die gegebenen Größen zu Matrizen zusammengefasst, wobei die den Daten anhaftenden statistischen Fehler durch Einführung einer Fehlermatrix R berücksichtigt wurden. Ziel war es nun, für die Elemente von R Werte zu ermitteln, die das entstehende Gleichungssystem so lösen, dass die Frobeniusnorm dieser Fehlermatrix minimal wird. Durch Umformulierung in Vektorschreibweise konnte eine Form der zu minimierenden Funktion - Norm genannt - gefunden werden, auf die die üblichen Verfahren zur Lösung nichtlinearer Ausgleichsprobleme angewandt werden konnten. Ein solches liegt vor, weil die Populationsverteilung keine lineare Funktion der Gibbs-Energie ist und daher die gesamte Funktion $A * N$ keine lineare Funktion dieser Größe ist.

Zur Lösung des Ausgleichsproblems wurde das für derartige Fragestellungen bewährte Verfahren nach Levenberg-Marquardt gewählt, da dies im Gegensatz zu dem einfacheren Verfahren nach Gauss-Newton besser konvergiert. Im Laufe der Arbeit zeigte sich aber, dass dieses Verfahren nicht immer zu einem zufriedenstellenden Ergebnis kommt und das gesuchte Minimum im Raum nicht findet. Der Grund liegt in der Besonderheit des Parameterraums, von Linien, auf denen die Norm sehr klein ist, durchzogen zu sein und die zusätzlich die Eigenschaft haben, sich sehr langsam fallend einem kleinen minimalen Normwert immer mehr zu nähern. Landet der Algorithmus also auf seinem Weg durch den Parameterraum an einer ungünstigen Stelle auf einer dieser Linien, läuft er vom eigentlich gesuchten Raumpunkt weg. Dieses Problem würde sich jedoch auch mit einem anderen Algorithmus, zum Beispiel mit einem Monte-Carlo-Verfahren, nicht zufriedenstellender lösen lassen.

4.2.2 Untersuchung des Parameterraums, Optimierung der Anfangsparameter und Funktionstest

Wie schon anhand der nichtlinearen Abhängigkeiten zu vermuten war, ist die Funktion der Norm über dem Parameterraum komplexer aufgebaut und nicht mit einem einfachen Trichter zu vergleichen, in den der Algorithmus von jedem Punkt des Parameterraums findet. Eine geeignete Wahl von Anfangsbedingungen ist deswegen elementar wichtig, um gute Ergebnisse mit dem Verfahren nach Levenberg-Marquardt zu erzielen. Daher wurde ein

Schwerpunkt der Arbeit darauf gelegt, die Besonderheiten des Parameterraums zu erfassen, um die daraus gewonnenen Erkenntnisse bei der Minimumsuche nutzen zu können.

Dabei wurden zwei Merkmale des Parameterraums erkannt, die beide bei der Wahl der Startpunkte Beachtung finden müssen:

Wie in Kapitel 2.4.2 erläutert, entstehen im Parameterraum Symmetrien, die aus der Möglichkeit der Permutation bei der Benennung der einzelnen Zustände resultieren. Eine Untersuchung deren Lage im $\Delta\mathbb{G}_m^0$ -Unterraum des Parameterraums zeigt (siehe Abbildung 3-30), dass die Raumdiagonale, die durch Gleichsetzen aller $\Delta\mathbb{G}_m^0$ - die normierten Gibbs-Energiedifferenzen bei Normaldruck - entsteht, als Symmetrieachse fungiert, um die die Minima angeordnet sind. Ihr kommen sie im Mittel viel näher als allen anderen gedachten Raumachsen. Unabhängig von der Wahl der Parameter finden sich die kürzesten Abstände zwischen dieser Diagonale und den Extrema bei negativen Werten von $\Delta\mathbb{G}_m^0$. Deswegen bietet es sich an, Startpunkte des Algorithmus auf dieser Linie zu wählen.

Neben den beschriebenen Symmetrien wird - wie in Kapitel 3.3.1 ausführlich dargelegt - die Funktion der Norm über dem Parameterraum von Linien geprägt, die sich durch eine besonders kleine Norm auszeichnen und auf denen die Minima liegen. Die untersuchten Zweizustandssysteme, bei denen nur die Ebene, in der $\Delta\mathbb{b}_2^{0'}$ gleich Null analysiert wird, liefert ein sehr anschauliches Bild dieser Linien durch den Raum, die als Gräben der Hyperebene Norm über dem Parameterraum betrachtet werden können (siehe zum Beispiel Abbildung 3-24). Diese Gräben sind sehr schmal, werden aber - wie in Kapitel 3.4.1 untersucht - durch die den Daten anhaftenden statistischen Fehler etwas breiter, was bewirkt, dass der Algorithmus zur Minimumsuche leichter den Graben finden kann.

Durch einen Vergleich mehrerer Parameterkombinationen wurde exemplarisch für Zweizustandssysteme gezeigt, wie sich der Verlauf der Gräben mit Variation der Parameter ändert. Dabei ist insbesondere zu beachten, dass Kombinationen gefunden werden können, die einen annähernd gleichen Verlauf der Linien kleiner Norm durch den Raum haben und die sich nur durch die Stelle auf der Linie unterscheiden, wo sich das Minimum befindet. An den betrachteten Beispielen ist auch ersichtlich, dass bei Systemen, die sich durch annähernd gleichen Linienverlauf auszeichnen, auch deren Anteile der Zustände ein ähnliches Verhalten mit dem Druck zeigen (siehe Abbildungen 3-22 und 3-23).

Allerdings wird die Norm auf diesen Linien kleiner Norm durch den Raum nicht zu dem gesuchten Minimum hin immer kleiner, wie für Zweizustandssysteme anschaulich dargestellt wurde (siehe Abbildung 3-30), sondern es existieren Bereiche, von denen aus die Norm sich einem konstanten tiefer liegenden Wert immer weiter annähert und ihr Verlauf entlang der Linie immer flacher wird. Diese Linien biegen bei allen untersuchten Datensätzen auf Geraden ein, die parallel der ΔG_2^0 - Achsen durch den Raum verlaufen. Möglich ist das dadurch, dass bei immer größer werdendem ΔG_2^0 der zweite Zustand immer mehr an Bedeutung verliert. Die korrigierten chemischen Verschiebungen des höherbesetzten Zustands nehmen dann die Mittelwerte der gemessenen druckkorrigierten Werte an, während die des anderen Zustands sehr große oder sehr kleine Werte annehmen, um trotz immer schwächerer Gewichtung die Abweichungen der Messwerte von den Mittelwerten optimal auszugleichen.

Aus diesen Besonderheiten des Parameterraumes wird klar, dass es zum Auffinden des Minimums wichtig ist, an der richtigen Stelle auf der Linie kleinster Norm durch den Parameterraum zu landen, um nicht mit dem Algorithmus, der nach immer kleiner werdenden Werten für die Norm sucht, in die falsche Richtung zu immer größer werdenden Werten für ΔG_m^0 abzurutschen. Aufgrund dieser Ergebnisse lässt sich erkennen, wie wichtig die aufwendigen Untersuchungen des Parameterraums sind, um das Verfahren nach Levenberg-Marquardt anwenden zu können.

Die in Kapitel 3.3.2 und 3.4.1 beschriebenen Funktionstests für Datensätze mit und ohne statistische Fehler zeigen, dass es trotz geschickter Vorauswahl der Startpunkte nicht möglich ist, für alle untersuchten Systeme die minimalste Norm sicher ausfindig zu machen. Dabei muss allerdings bedacht werden, dass zur Untersuchung einer großen Anzahl Datensätze eine zeitsparende Automatisierung zur Auswertung angewandt wurde. Liegt im Gegensatz dazu nur ein experimenteller Datensatz vor, kann durch das Testen weiterer Startpunkte und Interpretation von Indizien aus Zwischenergebnissen die Wahrscheinlichkeit deutlich erhöht werden, die richtigen Parameter zu finden.

4.2.3 Untersuchungen der Auswirkungen der statistischen Fehler

Des Weiteren stand in dieser Arbeit im Vordergrund festzustellen, wie sich Messfehler bei der Datenaufnahme auf die Systeme auswirken. Um den Einfluss statistischer Fehler der Messgrößen abzuschätzen, wurden auf die idealen Datensätze Fehlermatrizen verschiedener

Größenordnungen aufaddiert, für diese neu entstandenen Systeme deren mathematische Minima bestimmt und mit den erzeugenden Parametern verglichen.

Dabei zeigte sich, dass die Extrema durch die Fehler im Parameterraum verschoben werden. Das eigentliche Ziel, aus diesen Analysen je nach Fehlergröße mittlere typische Abweichungen der Parameter bestimmen zu können, konnte jedoch nicht erreicht werden, da sich die Größe der Fehler als stark abhängig von der Lage der erzeugenden Parameter im Parameterraum erwiesen. Insbesondere wurden Datensätze mit großen Besetzungszahländerungen deutlich weniger durch die Störungen beeinflusst als Datensätze mit kleinen Änderungen.

Um nun zu sehen, wohin die Minima durch die Fehler verschoben werden, wurden verschiedenen Datensätzen jeweils eine große Anzahl zufällig erzeugter Fehlermatrizen gleicher Größenordnung aufaddiert. Hierbei war festzustellen, dass sich die Minima im Parameterraum nicht gleichmäßig um den Punkt des eigentlich erzeugenden Parametersatzes sammeln, sondern entlang der Linien kleinster Norm verschoben werden (siehe Abbildung 3-33) und folglich die Parameter ΔG_m^0 , ΔV_m^0 und Δb_m^0 durch die Fehler über einen sehr breiten Bereich des Parameterraums streuen. Damit wurde deutlich, dass die Auswertung der korrigierten chemischen Verschiebungen nur für sehr kleine Fehler und günstige Parameterkombinationen überhaupt in der Lage ist, diese Größen abzuschätzen. Trotzdem resultieren - aufgrund der starken Abhängigkeit der Parameter - Besetzungszahlen, deren Verlauf mit dem Druck sich stark ähneln. Sie zeigen prinzipiell das gleiche Verhalten, nur ihre Anteile variieren je nach Fehlergröße und Lage des Systems im Parameterraum mehr oder weniger stark.

4.2.4 Angabe eines Lösungsbereichs

Untersucht man ein unbekanntes System mit dem vorgeschlagenen Verfahren, ist zunächst das mathematische Minimum zu bestimmen. Als möglicher Lösungsbereich kommen dann alle Punkte in der Nähe dieses Extremums unter einer gewissen Norm in Frage.

Deswegen war das nächste konkrete Ziel dieses Projekts, einen geeigneten Schwellenwert für diese Norm abzuleiten und eine Methode zu entwickeln, passende Punkte im Raum aufzuspüren, die dem Lösungsbereich zuzurechnen sind.

Dazu konnte das übliche bei der Minimierung der Fehlerquadratsumme verwendete Vorgehen (Sachs, 2004) auf die hier vorliegende Funktion angewendet werden (siehe Kapitel 2.4.4), das erlaubt, aus der Norm am mathematischen Minimum sowohl die Größe der Norm der Fehlermatrix als auch einen geeigneten Schwellenwert der Norm zu berechnen, um einen Vertrauensbereich im Parameterraum zu definieren, dem der gesuchte Parametersatz mit einer Wahrscheinlichkeit von 68% angehört.

Nun stellte sich die Frage, wie Raumpunkte zu finden sind, auf die diese Bedingung - eine kleinere Norm zu besitzen als der berechnete Schwellenwert - zutrifft. Dazu wurden im Rahmen dieser Arbeit zunächst verschiedene Versuche unternommen, den Raum systematisch zu rastern oder durch zufällige Parameterwahl Punkte auszuwählen. Allerdings erwiesen sich diese Ansätze aufgrund der multiplen Dimensionen des Parameterraums und der Form des Lösungsbereichs als nicht zielführend. Daher wurde, wie in Kapitel 3.5.2 dargestellt, eine neue und bisher in der Literatur nicht beschriebene Methode entwickelt. Als Besonderheit dieser Methode wird der Raum um das mathematische Minimum gezielt durch Fehlermatrizen abgetastet, da davon ausgegangen werden kann, dass sich die Funktion Norm über dem Parameterraum des Systems des gefundenen Minimums von dem gesuchten erzeugenden System nur durch die Lage des Extremums unterscheidet. Die typischen Linien kleiner Norm durch den Raum aber bleiben stationär und die gewünschten Raumpunkte sind genau auf und um diese Linien zu finden. Die Korrektheit dieser Methode wurde anhand mehrerer zufällig erzeugter Zwei-, Drei- und Vierzustandssysteme überprüft. In allen Fällen enthielt der generierte Lösungsbereich den Raumpunkt des gesuchten Parametersatzes.

Aus den so gewonnenen Raumpunkten können dann die Verteilungen der Besetzungszahlen berechnet und für diese der gewünschte Signifikanzbereich angegeben werden.

4.2.5 Beurteilung der Ergebnisse von Byr2-RBD

Für die Auswertung der bei verschiedenen Drücken aufgenommenen Spektren des Proteins Byr2-RBD mussten zunächst eine Prozessierung der Daten und eine Zuordnung der Signale durchgeführt werden. Dabei stellte sich aber heraus, dass die Grundannahme bei der Auswertung mit Algorithmus 1 - die Signale, deren chemische Verschiebungen verarbeitet werden sollen, müssen im schnellen Austausch miteinander stehen - bei den untersuchten Daten nicht erfüllt ist. Trotzdem wurde eine Auswertung mit dem in dieser Arbeit entwickelten Verfahren versucht.

Dabei konnte ein mathematisches Minimum gefunden werden, für das eine Populationsverteilung folgt, bei der zwei Zustände mit etwa 40% beziehungsweise 55% deutlich höher besetzt sind als ein weiterer dritter Zustand. Allerdings deutet die Tatsache, dass dieses Minimum nicht stabil ist gegenüber dem Weglassen der Daten einzelner Aminosäuren, darauf hin, dass die Fehler nicht statistisch verteilt sind, sondern bei einigen Aminosäuren durch den langsamen Austausch Teile des Signals verloren gegangen sind und somit keine Mittelung über alle Zustände vorliegt. Auch eine Fehleranalyse schlug aufgrund der sehr großen gefundenen Norm und des daraus resultierenden großen Bereichs im Parameterraum fehl. Die Auswertung der chemischen Verschiebungen mit dem entwickelten Algorithmus musste somit als undurchführbar eingestuft werden.

Bei den folgenden weiteren Auswertungen wurde der Schwerpunkt der Untersuchungen auf das C-terminale Ende des untersuchten Konstrukts gelegt. Die Tatsache, dass für diesen Bereich keine mittel- oder langreichweitigen Abstandsrestraints in den NOESY-Spektren gefunden werden konnten (Gronwald et al., 2001), unterstützt die Vermutung über dessen größere Beweglichkeit. Ebenso ergibt dies die durchgeführte Analyse der Seitenketten des Proteins. Die gefundene Verteilung polarer und nicht-polarer Seitenketten (siehe Abbildung 3-51) zeigt deutlich den durch die Helices 1 und 2 und die fünf β -Faltblätter gebildeten hydrophoben Kern, der eine kompakte Einheit - ein *Ubiquitin fold* - bildet, dem die Aminosäuren 166-180 nicht mehr angehören.

Dass es dennoch neben dem Zustand relativ freier Beweglichkeit stabile Konformationen geben könnte, ergibt die Auswertung der Signalvolumina: Nur die allerletzten Aminosäuren des Konstrukts, L178 und S179, weisen ein Volumen auf, das deutlich größer ist als das der vorhergehenden Residuen. Außerdem liegen deren chemische Verschiebungen in Random-Coil-typischen Bereichen. Die zugeordneten Signale der Aminosäuren S173, S176 und S177 dagegen besitzen bereits bei Normaldruck nur ein Volumen von etwa 48% des eigentlich erwarteten, das unter dem Einfluss größer werdenden Drucks auf etwa 22% absinkt. Das bedeutet, die Gesamtsignale könnten durch mehrere Zustände im langsamen Austausch aufgespalten sein und deren Besetzungsverhältnisse durch den Druck gegeneinander verschoben werden, wie dies zum Beispiel auch für das Protein Ras mit ^{31}P -NMR-spektroskopischen Untersuchungen nachgewiesen worden war (Kalbitzer et al., 2009). Für die beobachteten Signale des C-terminalen Bereiches ließen sich im Gegensatz zu denen anderer Aminosäuren, die ebenfalls eine Aufspaltung in mehrere Signale aufweisen und dem Kern des

Proteins angehören, keine weiteren Signale in unmittelbarer Nähe zuordnen, die den anderen Zuständen entsprechen würden, so dass eine stärkere Konformationsänderung in diesem Bereich bei einem Zustandsübergang zu vermuten ist. Das ist auch kompatibel mit der Beobachtung mehrerer kleiner in den Spektren gefundener, aber nicht zugeordneter Signale. Diese Ergebnisse bestätigen das bereits in der Arbeit von Gronwald et al. vermutete Vorliegen mehrerer Konformationen des C-terminalen Bereiches (Gronwald et al., 2001).

Denkbar sind also außer dem Zustand, bei dem sich der C-Terminus frei bewegt, Konformationen, bei denen sich dieser Bereich an den gefalteten Kern anlagert und zu dem beweglichen Zustand im langsamen Austausch steht. Die durchgeführten Struktur-rechnungen lieferten hier Vorschläge, die ohne den Einsatz stabilisierender Restraints zufällig entstanden und energetisch sehr günstig sind. Allen diesen Konformationen, bei denen sich das betrachtete Ende an das übrige Molekül anlagert, ist gemein, dass immer derselbe Bereich um Aminosäure R74 auf der Oberfläche betroffen ist. Wie eine Analyse zeigte, könnte der C-Terminus, der die Aminosäuren L166 bis P180 umfasst, über Wasserstoffbrücken an der vermuteten Interaktionsfläche stabilisiert sein.

Weitere Hinweise darauf, dass diese Konformationen in dieser oder in sehr ähnlicher Weise in vivo realisiert sein könnten, liefern auch die Signalaufspaltungen einiger Signale von Aminosäuren des gefalteten Kerns des Proteins, I76, A77, G80 (siehe Abbildung 3-46, Vergrößerung), L143 und K150. Auch hier ist eine Verschiebung der Volumenverhältnisse mit dem Druck beobachtbar, die sich im Rahmen der Fehlergrenzen analog zu der Abnahme der Signalvolumina der C-terminalen Aminosäuren bewegt. Das ist als starker Hinweis dafür zu werten, dass sie über Konformationsänderungen des C-Terminus miteinander verknüpft sind. Auch die Analyse nach Beck-Erlach et al., die eben diesen Bereich als Interaktionsfläche vorhersagt, unterstützt diese Vermutung (Beck-Erlach et al., 2014).

Die Beobachtung, dass manche der Signale des Proteinkerns eine Aufspaltung in mehr als zwei Unterpeaks aufweisen, könnte ein Hinweis darauf sein, dass es mehrere leicht unterschiedliche Konformationen der letzten Aminosäuren der untersuchten Domäne geben könnte, die energetische ähnlich günstig und an den Kern angelagert sind.

In den Untersuchungen von Huber an demselben Konstrukt von Byr2-RBD, wie dem in dieser Arbeit analysierten, wird ebenfalls von Signalaufspaltungen berichtet, die aber bei der Zugabe

von Ras zu dem Protein verschwanden (Huber, 2001). Dies wäre kompatibel mit der Vorstellung, dass eine Anlagerung des C-Terminus an die Proteinoberfläche nicht mehr möglich ist, wenn die bevorzugte Interaktionsfläche bereits von Ras besetzt ist.

Auch ist es interessant, dass die durch den C-Terminus solchermaßen überdeckten Aminosäuren zur Ras-Bindestelle gehören. Dabei stellt sich die Frage, ob dieses Verhalten physiologisch oder erst durch die Abtrennung der angrenzenden Domänen möglich ist. Im Prinzip wäre es auch denkbar, dass, wie von Iwig et al. für das Ras-bindende Protein RasGRP1 oder von Sondermann et al. für das Ras-bindende Protein SOS beschrieben (Iwig et al., 2013; Sondermann et al., 2004), ein Vorgang der Autoinhibierung vorliegt, über den die Bindung an Ras gesteuert werden könnte.

Zusammenfassend kann man feststellen, dass sich die Hoffnung, für die untersuchte Ras-Bindedomäne von Byr2 zwei strukturell unterschiedliche Zustände und damit ein ähnliches Verhalten wie für das eingehend unter Hochdruck untersuchte strukturhomologe Protein Ubiquitin nachweisen zu können, bestätigte, da hier ebenfalls deutliche Hinweise vorliegen, dass der C-Terminus wie auch bei Ubiquitin für konformationelle Änderungen verantwortlich ist (Kitahara und Akasaka, 2003; Jackson, 2006). In Anbetracht der schlechten Auflösung der Spektren jedoch konnte nur eine eher qualitative Aussage aufgrund mehrerer Indizien gemacht werden. Die Option, einen Ras-bindenden und einen Ras-nichtbindenden Zustand identifizieren zu können, scheiterte an der Tatsache, dass sich der flexible C-Terminus vermutlich genau über die Ras-Bindestelle legt und damit alle strukturellen Umwandlungen, die bei der Anpassung an Ras geschehen müssen, überdeckt.

4.2.6 Beurteilung der Ergebnisse für das Protein Ras

Zur Analyse der chemischen Verschiebungen des Proteins Ras in seiner aktiven Form wurden, wie von Kalbitzer et al. dargestellt, vier Zustände angenommen (Kalbitzer et al., 2013). In einem ersten Schritt wurde mit Algorithmus 1 eine Parameteranpassung vorgenommen, ohne weiteres Wissen über das System zu verwenden. Insbesondere flossen dabei keine Ergebnisse aus der Auswertung der Signalvolumina mit ein. Die Populationsverhältnisse in Abhängigkeit vom Druck, die aus dem so identifizierten mathematischen Minimum folgen, bestätigen die von Kalbitzer et al. publizierten Ergebnisse. Trotzdem weichen die Werte der Differenzen der Gibbs-Energien und der partiellen molaren Volumina sehr weit von den publizierten Werten ab. Eine Fehleranalyse an diesem Datensatz

bestätigte die an den Testdatensätzen gemachte Beobachtung, dass diese Parameter nicht stabil und durch die statistischen Fehler weit von ihren eigentlichen Werten abweichen können, während sich über den Verlauf der Besetzungszahlen mit dem Druck durch die Auswertung der chemischen Verschiebungen durchaus verwertbare Aussagen treffen lassen.

Bei der Beurteilung der Ergebnisse ist zu beachten, dass bei der im Rahmen dieser Arbeit durchgeführten Datenauswertung ein anderes Modell für das Protein Ras angenommen wurde als bei den bereits von Kalbitzer et al. publizierten Ergebnissen: Dort wurde - aufgrund der Volumenabnahme mancher Signale bei hohen Drücken - angenommen, dass der vierte Zustand mit den drei anderen, die sich im schnellen Austausch untereinander befinden, im langsamen Austausch steht und damit für einen Teil der beobachteten Signale zu Volumenabnahmen führt. Wie Abbildung 3-58 und Tabelle 3-12 zeigen, sind die Besetzungszahlen des mit MultiStates angenommenen vierten Zustands jedoch so klein, dass er bei der Datenanpassung nur die Aufgabe übernimmt, Messfehler auszugleichen. Dadurch wird indirekt das Ergebnis von Kalbitzer et al. bestätigt, dass vier Zustände - der vierte befindet sich im langsamen Austausch zu den drei funktionellen Hauptkonformationen - die an aktivem Ras aufgenommenen Spektren erklären.

Die weiteren an den Daten von aktivem Ras durchgeführten Rechnungen zeigten, wie sich Informationen aus anderen Quellen über die Parameter, in diesem Fall der Differenz der Gibbs-Energie zwischen dem GEF-bindenden und dem Effektor-bindenden Zustand, mit der entwickelten Methode zur Auswertung der druckkorrigierten chemischen Verschiebungen sinnvoll kombinieren lassen: Bei der Implementierung des Algorithmus nach Levenberg-Marquardt wurde die Möglichkeit geschaffen, einen oder mehrere Parameter während der Optimierung auf einem festen Wert zu halten. Durch die Hinzunahme einer weiteren Information über das System - die beiden eben genannten Zustände stehen bei etwa 125 MPa im Gleichgewicht miteinander (Kalbitzer et al., 2009) - konnten Parameterkombinationen innerhalb des Lösungsbereichs ausfindig gemacht werden, die auch diese Bedingung erfüllen. Dies zeigt, auf welche Weise der Algorithmus zur Datenauswertung eingesetzt werden kann, führt aber gleichzeitig die generellen Schwierigkeiten vor Augen, die die Auswertung der chemischen Verschiebungen mit sich bringt, insbesondere dann, wenn mehr als drei Zustände angenommen werden: Die Signifikanzbereiche für die Parameter, die den Verlauf der Gibbs-Energiedifferenzen mit dem Druck bestimmen - ΔG_{1m}^0 , ΔV_{1m}^0 und $\Delta \beta_{1m}^{0'}$ -, werden sehr groß, und durch die freie Anpassung ohne Einschränkung für die Größen nach oben oder unten

kann die Differenz der Kompressibilitätsfaktoren $\Delta\beta_{1m}^{0'}$ sehr große Werte annehmen, die nach Lehrmeinung zu groß sind und der Physik nicht entsprechen und weswegen bei der Auswertung der Daten am Protein Ras von Kalbitzer et al. diese Größe für alle Übergänge gleich Null gesetzt wurde (Kalbitzer et al., 2014). Dies erklärt auch die deutlich kleineren Werte für die Standardabweichung der publizierten Ergebnisse im Vergleich zu den mit MultiStates berechneten, da dadurch der Parameterraum weiter eingeschränkt wird.

4.2.7 Beurteilung der Ergebnisse für das Alzheimer- β -Amyloidpeptid

Bei der Auswertung der vorliegenden Daten des Alzheimer- β -Amyloidpeptids wurden unter Einbeziehung von 62 Atomen, der Stickstoffatome des Proteinrückgrats und deren kovalent gebundener Wasserstoffatome, zwei verschiedene Arten der Datenkombination getestet und jeweils ein Extremum ausfindig gemacht. Die aus den dabei gefundenen Minima folgenden Besetzungszahlen der Zustände entsprechen sich im Rahmen der Fehlergrenzen, bestätigen allerdings nicht die Ergebnisse der Arbeit von Munte et al. (Munte et al., 2013). Die Ursache muss vermutlich darin gesucht werden, dass bei den hier untersuchten Spektren Volumenänderungen der Signale beobachtet werden können, die von Munte et al. ebenfalls zur Auswertung herangezogen worden waren und daraus auf die Besetzungszahlen der Zustände geschlossen worden war. Insofern lag bei der Analyse ein anderes Modell als das von MultiStates gewählte - das von besetzungszahlgemittelten chemischen Verschiebungen ausgeht und die Möglichkeit des intermediären oder langsamen Austauschs nicht berücksichtigt - zugrunde.

Zusätzlich muss für die vorhandenen Daten des Alzheimer- β -Amyloidpeptids festgestellt werden, dass es sich dabei um einen extrem schwer auswertbaren Datensatz handelt, da die chemischen Verschiebungen nach der aminosäureabhängigen Druckkorrektur nur mehr sehr kleine Änderungen mit dem Druck aufweisen. Der von Munte et al. gewählte Ansatz zur Deutung, einer der Zustände weise Random-Coil-ähnliche Strukturen auf (Munte et al., 2013), könnte eine Erklärung für dieses Verhalten sein. Der mit Algorithmus 1 aber berechnete, sich mit dem Druck kaum verändernde Verlauf der Besetzungszahlen würde die hier vorliegenden sehr schwachen Änderungen der korrigierten chemischen Verschiebungen ebenfalls begründen, wobei dies Random-Coil-ähnliche Strukturen nicht ausschließt.

Ein Versuch, die Daten auch an ein Modell mit nur zwei Zuständen anzupassen, deutet darauf hin, dass ein weiterer Zustand zur Erklärung der Daten angenommen werden muss. Dies legt

auch der Druckverlauf der korrigierten chemischen Verschiebungen einiger Atome nahe und bestätigt somit die von Munte et al. gemachte Annahme von mindestens drei relevanten koexistierenden Zuständen bei Normaldruck.

4.2.8 Bewertung der Ergebnisse von MultiStates

Für Systeme mit einer so großen Anzahl an Variablen im Verhältnis zur Anzahl der gemessenen Größen eine optimale Lösung zu finden, die die Daten erklärt, ist schwierig. Aufgrund der Komplexität des Systems müssen deswegen alle zur Verfügung stehenden Daten gleichzeitig betrachtet werden.

Die Ergebnisse der vorliegenden Arbeit zeigen, dass ein Algorithmus entwickelt werden konnte, der durch die Berücksichtigung der Besonderheiten des Parameterraums eine solche Lösung liefern kann. Dabei sind die hier vorgeschlagenen Anfangsbedingungen sind dem Wissen über die Form der zu optimierenden Funktion Norm über dem Parameterraum optimal angepasst.

Allerdings stößt das Verfahren an seine Grenzen, wenn die untersuchten Änderungen der chemischen Verschiebungen sehr klein sind und dadurch das Minimum numerisch nicht mehr identifiziert werden kann. Zudem bewirkt das Vorhandensein von statistischen Fehlern, dass das mathematische Minimum vom eigentlich gesuchten Parametersatz wegverschoben wird. Wie weit dies geschieht, hängt von der Fehlergröße und von der Lage der erzeugenden Parameter im Parameterraum ab. Diese Verschiebung durch die Fehler erfolgt aber nicht willkürlich, sondern es entstehen Systeme, deren Verlauf der Besetzungszahlen dem tatsächlichen System sehr ähnlich ist. Da der eigentlich gesuchte Parametersatz durch die Fehler nicht gefunden werden kann, muss deswegen abgeschätzt werden, wie groß der Bereich um das mathematische Minimum ist, der als Lösungsbereich in Frage kommt. Der in der vorliegenden Arbeit neu entwickelte Ansatz, den Lösungsbereich über Fehlermatrizen zu generieren, eröffnet die Möglichkeit, die gesuchten Raumpunkte, die sich entlang der Linien kleinster Norm sammeln, zu finden, ohne aufwendig den Raum um das mathematische Minimum rastern zu müssen.

Während jedoch die die Gibbs-Energiedifferenzen beschreibenden Parameter ΔG_m^0 , ΔV_m^0 und $\Delta \beta_m^{0'}$, die von Algorithmus 1 als optimal ausgegeben werden, durch die möglichen statistischen Fehler der Daten deutlich von ihren tatsächlichen Werten wegverschoben sein

können, zeigt sich der Verlauf der Zustände mit dem Druck deutlich stabiler gegenüber solchen Störungen. Sie müssen deswegen als die eigentlich ein System charakterisierenden Größen gelten.

Die Anwendung des Algorithmus auf experimentelle Datensätze kann jedoch durch mehrere Faktoren erschwert werden: Zum einen können die Änderungen der chemischen Verschiebungen wie bei dem in dieser Arbeit untersuchten Protein Alzheimer- β -Amyloidpeptid so klein sein, dass sie an der Grenze zur Nachweisbarkeit liegen, zum anderen kann eine Auswertung wie auch bei Byr2-RBD durch unzureichende Mittelung erschwert werden, wenn die Zustandsänderungen auf einer langsameren Zeitskala ablaufen. Liegen wie bei dem hier ebenfalls untersuchten Protein Ras (in seiner aktiven Form) vier Zustände vor, dann wird die Anzahl der anzupassenden Parameter sehr groß und damit der Lösungsbereich sehr schwer eingrenzbar. Auch wenn die Auswertung experimenteller Datensätze in einigen Fällen aus den genannten Gründen erschwert oder unmöglich ist, so legen die Größenordnungen der gefundenen statistischen Fehler nahe, dass durch die heute verfügbaren experimentellen Techniken NMR Spektren mit einer hinreichenden Genauigkeit aufgenommen werden können, um den Verlauf der Besetzungszahlen mit dem Druck vorherzusagen. Bei der Anwendung von MultiStates muss allerdings sehr genau überprüft werden, inwieweit die Annahme des schnellen Austauschs für die in die Auswertung einbezogenen Atome gegeben ist, und eine zusätzliche Analyse der Signaltolumina ist unerlässlich.

Zusammengefasst ermöglicht es der Algorithmus also, die Auswertung der Daten zu beschleunigen, weil der aufwendige iterative Prozess, sich über das Anpassen der Parameter an die chemischen Verschiebungen jeder einzelnen Aminosäure und die statistische Auswertung den gesuchten Größen zu nähern, entfällt. Als nächster Schritt bei der Weiterentwicklung des Algorithmus MultiStates muss nun eine Analyse der ebenfalls mit dem Verfahren optimierten chemischen Verschiebungen der einzelnen Zustände erfolgen, um dem von Kalbitzer et al. formulierten Ziel (Kalbitzer et al., 2013) näher zu kommen, allosterisch wirkende Aminosäuren zu identifizieren und gezielt über Populationsverschiebungen in die Interaktionen der untersuchten Proteine eingreifen zu können.

5 Zusammenfassung

In der vorliegenden Arbeit werden zwei bioinformatische Methoden dargestellt, die zur Untersuchung von Proteinen neu- beziehungsweise weiterentwickelt und getestet wurden. Bei beiden stand insbesondere der Aspekt der Anpassung an den Bindungspartner bei der Ausbildung eines Komplexes im Mittelpunkt.

Die erste Methode dient zur Vorhersage von Proteinraumstrukturen und ist eine Weiterentwicklung des am Institut für Biophysik und physikalische Biochemie unter Anleitung von Hans Robert Kalbitzer entwickelten Homologiemodellierungsprogramms PERMOL, das darauf beruht, aus strukturhomologen Proteinen Abstandsrestraints, dihedrale Winkelrestraints und Wasserstoffbrückenrestraints zu extrahieren, die für die gesuchte unbekannte Struktur als Grundlage einer molekulardynamischen Simulation des Proteins verwendet werden können. Neben der völligen Neuimplementierung mit einer benutzerfreundlicheren und optisch ansprechenderen GUI-Programmierung, die mit C++ in Kombination mit der Klassenbibliothek Qt umgesetzt wurde, lag der Schwerpunkt vor allem auf dem neuen Aspekt der Modellierung von Proteinkomplexen. Außerdem wurde die Möglichkeit geschaffen, die Informationen aus mehreren PDB-Strukturen gewichtet nach lokaler Sequenzähnlichkeit zum zu modellierenden Protein miteinander zu kombinieren. Beispielrechnungen an Byr2-RBD im Komplex mit Ras und IE1 im Komplex mit menschlichem Histon zeigten die Einsatzmöglichkeiten und Stärken dieser Methode. Sehr gute Ergebnisse wurden dabei bei der Anpassung der Interaktionspartner aneinander erzielt. PERMOL ist Teil des Softwareprojekts AUREMOL.

Die zweite Methode, MultiStates, zielt auf die Vorhersage des Verhaltens der verschiedenen Zustände eines Proteins unter Hochdruck ab und zeigt druckinduzierte Verschiebungen des Gleichgewichts zwischen ihnen auf. Diese Zustände entsprechen Tälern in der Hyperebene, die die Gibbs-Energie über dem Konfigurationsraum des beobachteten Proteins bildet, und es besteht die Vermutung, dass sie mit den Zuständen korrelieren, die bei der Bindung an Interaktionspartner eingenommen werden. Neben der Anzahl der Zustände selbst interessiert dabei das Verhalten der Differenzen der Gibbs-Energien und den dadurch bewirkten Besetzungszahländerungen mit dem Druck. Für die Gibbs Energie wurde als Näherung eine Taylorentwicklung bis zur zweiten Ordnung gewählt. Der in diesem Projekt neu entwickelte Ansatz zur Bestimmung dieser Taylorkoeffizienten fasst die vorliegenden Daten zu Matrizen zusammen und formuliert das Problem als überbestimmtes nichtlineares Gleichungssystem,

dessen Minimum gefunden werden muss. Die Lösung erfolgt über das Verfahren nach Levenberg-Marquardt. Zur Optimierung der Testverfahren und zur Untersuchung des Einflusses von statistischen Fehlern, wie sie bei experimentell bestimmten Daten immer vorliegen, wurde zusätzlich zu den drei untersuchten experimentellen Datensätzen eine große Anzahl an Testdatensätzen generiert. Neben der Optimierung der Methoden zum Auffinden des mathematischen Minimums und einer Abschätzung der Fehlergrenzen lag der Schwerpunkt vor allem auf der Untersuchung des Parameterraums selbst, in dessen Form die Ursache für die entwickelte Vorgehensweise bei der Suche nach dem Minimum liegt.

Mit den entwickelten Methoden konnte für die Ras-Bindedomäne des Proteins Byr2 eine Verbesserung der NMR-Struktur erreicht werden. Durch eine multimodale Analyse NMR-spektroskopisch generierter Daten unter Hochdruck, weiterer molekulardynamischer Simulationen und bioinformatischen Methoden konnte für dieses Makromolekül außerdem gezeigt werden, dass der C-Terminus zwei Zustände besitzt - entweder frei beweglich ist oder sich über die Ras-Bindestelle legt. Dies deutet auf den Mechanismus einer Autoinhibierung hin, wie sie auch für andere Ras-Bindepartner bekannt ist.

Für den C-Terminus des Proteins IE1 des Cytomegalievirus konnte ein Homologiemodell im Komplex mit menschlichem Chromatin erstellt werden, das mit den experimentellen Ergebnissen aus Mutagenesexperimenten kompatibel ist und die Vermutung unterstützt, dass IE1 als Bindeglied zwischen der Virus-DNA und menschlichem Erbgut fungieren könnte und somit an dem Mechanismus der Persistenz dieses Virus während der latenten Phase beteiligt ist.

6 Literaturverzeichnis

- Akasaka**, Kazuyuki (2006). "Fluctuation of Proteins by Pressure Perturbation." Chem Rev **106**:1814-1835.
- Alberts**, Bruce, **Johnson**, Alexander, **Lewis**, Julian, **Raff**, Martin, **Roberts**, Keith, **Walter**, Peter (2011). Molekularbiologie der Zelle, 5. Auflage. Weinheim, Wiley-VCH.
- Anfinsen**, Christian Boehmer (1973). "Principles that govern the folding of Protein chains." Science **181**(4096):223-230.
- Armache**, KJ, **Garlick**, JD, **Canzio**, D, **Narlikar**, GJ, **Kingston**, RE (2011). "Structural basis of silencing: SIR3 BAH domain in complex with a nucleosome at 3.0 Å resolution." Science **334**:977-982.
- Arnold**, Konstantin, **Bordoli**, Lorenza, **Jürgen**, Kopp, **Schwede**, Torsten (2005). "The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling." Bioinformatics **22**(2): 195-201.
- Bairoch**, Amos, **Boeckmann**, Brigitte, **Ferro**, Serenella, **Gasteiger**, Elisabeth (2004). "Swiss-Prot: juggling between evolution and stability." Brief Bioinform **5**(1): 39-55.
- Barbera**, Andrew J, **Chodaparambil**, Jayanth V, **Kelley-Clarke**, Brenna, **Joukov**, Vladimir, **Walter**, Johannes C, **Luger**, Karolin, **Kaye**, Kenneth M (2006). "The nucleosomal surface as a docking station for Kaposi's sarcoma herpesvirus LANA." Science, **311**(5762):856-61.
- Beck-Erlach**, Markus, **Köhler**, Jörg, **Möser**, Beate, **Horinek**, Dominik, **Kremer**, Werner, **Kalbitzer**, Hans Robert (2014). "Relationship between Nonlinear Pressure-Induces Chemical Shift Changes and Thermodynamic Parameters." J Phys Chem B **118**: 5681-5690.
- Berman**, Helen M., **Westbrook**, John, **Feng**, Z., **Gilliland**, G., **Bhat**, T. N., **Weissig**, H., **Shindyalov**, I. N., **Bourne**, Philip E. (2000). "The Protein Data Bank." Nucleic Acids Res **28**(1): 235-242.
- Bonvin**, Alexandre M.J.J. (2006). "Flexible protein-protein docking." Curr Opin Struct Biol **16**(2):194-200.
- Brünger**, Axel T, **Adams**, Paul D, **Clore**, G Marius, **DeLano**, Warren L, **Gros**, Piet, **Grosse-Kunstleve**, Ralf W, **Jiang**, Jian-Sheng, **Kuszewski**, John, **Nilges**, Michael, **Pannu**, Navraj S, **Read**, Randy J, **Rice**, Luke M, **Simonson**, Thomas, **Warren**, Gregory L (1998). "Crystallography & NMR system: A new software suite for macromolecular structure determination." Acta Crystallogr D Biol Crystallogr **54**:905-921.
- Brunner**, Konrad (2006). "Modellierung, Strukturverbesserung und sequentielle Zuordnung als voll-automatische Module für die automatisierte Proteinstrukturbestimmung im Softwareprojekt AUREMOL." Dissertation, Universität Regensburg.
- Brunner**, Konrad, **Gronwald**, Wolfram, **Trenner**, Jochen M., **Neidig**, Klaus-Peter, **Kalbitzer**, Hans Robert (2006). "A general method for the unbiased improvement of solution NMR structures by the use of related X-ray data, the AUREMOL-ISIC algorithm." BMC Struct Biol **6**: 14.
- Budisa**, N, **Steipe**, B, **Demange**, P, **Eckerskorn**, C, **Kellermann**, J, **Huber**, R (1995). "High-level biosynthetic substitution of methionine in proteins by its analogs 2-aminohexanoic acid, selenomethionine, telluromethionine and ethionine in Escherichia coli." Eur. J. Biochem. **230**(2):788-96.

Cano, Carolina, **Brunner**, Konrad, **Baskaran**, Kumaran, **Elsner**, Ralph, **Munte**, Claudia Elisabeth, **Kalbitzer**, Hans Robert (2009). "Protein structure calculation with data imputation: the use of substitute restraints." J Biomol NMR **45**(4):397-411.

Changeux, Jean-Pierre, **Edelstein**, Stuart J. (2005). "Allosteric mechanisms of signal transduction." Science **308**(5727): 1424-1428.

Chothia, Cyrus, **Lesk**, Arthur M. (1986). "The relation between the divergence of sequence and structure in proteins." EMBO J **5**(4): 823-826.

Chothia, Cyrus, **Lesk**, Arthur M. (1987). "The evolution of protein structures." Cold Spring Harb Symp Quant Biol **52**: 399-405.

Clore, Marius G, **Gronenborn**, Angela M, **Bax**, Ad (1998). "A Robust Methode for Determining the Magnitude of the Fully Asymmetric Alignment Tensor of Orientated Macromolecules in the Absence of Structural Information." J Mag Res **133**: 216-221.

Cruickshank, D W J (1999). "Remarks About Protein Structure Precision." Acta Cryst D **55**:583-601.

David, Ralf, **Richter**, Michael P O, **Beck-Sickinger**, Annette G (2004). "Review article: Expressed protein ligation. Methods and applications." Eur J Biochem **271**:663-677.

DeLano, Warren L (2002). "The PyMOL User's Manual." DeLano Scientific, San Carlos, California, USA.

Dahmen, Wolfgang, **Reusken**, Arnold (2008). Numerik für Ingenieure und Natur-wissenschaftler, 2., korr. Auflage. Springer Lehrbuch.

Dill, Ken. A., **Chan**, Hue Sun (1997). "From Levinthal to pathways to funnels." Nat Struct Biol **4**(1): 10-19.

Dunbrack, Roland L Jr. (2002). "Rotamer libraries in the 21st century." Curr Opin Struct Biol **12**(4):431-440.

Ebel, Michael (2012). "Reimplementierung des Moduls Align im Rahmen des Softwareprojekts AUREMOL.", Bericht zum Praktischen Studiensemester, Universität Regensburg.

Feeney, Katherine. M., **Parish**, Joanna L. (2009). "Targeting mitotic chromosomes: a conserved mechanism to ensure viral genome persistence." Proc Biol Sci **276**(1662): 1535-1544.

Fetics, Susan K, **Guterres**, Hugo, **Kearney**, Bradley M, **Buhrman**, Greg, **Ma**, Buyong, **Nussinov**, Ruth, **Mattos**, Carla (2015). "Allosteric Effects of the Oncogenic RasQ61L Mutant of RasRBD." Structure **23**(3): 505-516.

Fischer, Emil (1894). "Einfluss der Configuration auf die Wirkung der Enzyme." Berichte der deutschen chemischen Gesellschaft **27**(3): 2985-2993.

Fukuhara, N, **Kawabata**, T (2008). "HOMCOS: a server to predict interacting protein pairs and interacting sites by homology modeling of complex structures." Nucleic Acids Res. **36**(Web Sever issue): W185-9.

Geyer, Matthias, **Schweins**, Thomas, **Herrmann**, Christian, **Prisner**, Thomas, **Wittinghofer**, Alfred, **Kalbitzer**, Hans Robert (1996). "Conformational Transitions in P21^{ras} and in Its Complexes with the Effector Protein Raf-RBD and the GTPase Activating Protein GAP." Biochemistry **35**:10308-10320.

Gronwald, Wolfram, **Brunner**, Konrad, **Kirchhöfer**, Renate, **Nasser**, Adel, **Trenner**, Jochen, **Ganslmeier**, Bernhard, **Riepl**, Hubert, **Ried**, Andreas, **Scheiber**, Josef, **Elsner**, Ralph, **Neidig**, Klaus-Peter, **Kalbitzer**, Hans Robert (2004). "AUREMOL, a New Program for the Automated Structure Elucidation of Biological Macromolecules." Bruker Reports **154/155**: 11-14.

Gronwald, Wolfram, **Huber**, Fritz, **Grünwald**, Petra, **Spörner**, Michael, **Wohlgemuth**, Sabine, **Herrmann**, Christian, **Kalbitzer**, Hans Robert (2001). "Solution structure of the Ras binding domain of the protein kinase Byr2 from *Schizosaccharomyces pombe*." Structure **9**(11): 1029-1041.

Gronwald, Wolfram, **Kalbitzer**, Hans Robert (2004). "Automated Structure Determination of Proteins by NMR Spectroscopy." Prog. NMR Spectrosc. **44**: 33-96.

Gronwald, Wolfram, **Kirchhöfer**, Renate, **Görler**, Adrian, **Kremer**, Werner, **Ganslmeier**, Bernhard, **Neidig**, Klaus-Peter, **Kalbitzer**, Hans Robert (2000). "RFAC, a program for automated NMR R-factor estimation." J Biomol NMR **17**: 137-151.

Haass, Christian, **Selkoe**, Dennis J (2007). "Soluble protein oligomers in neurodegeneration: lessons from the Alzheimer's amyloid β -peptid." Nat. Rev. Mol. Cell Biol. **8**:101-112.

Hardy, John, **Selkoe**, Dennis J (2002). "The Amyloid Hypothesis of Alzheimer's Disease: Progress and Problems on the Road to Therapeutics." Science **297**:353-356.

Henikoff, Steven, **Henikoff**, Jorga G (1992). "Amino acid substitution matrices from protein blocks." Proc Natl Acad Sci USA **89**(22):10915-10919.

Huang, Lan, **Hofer**, Franz, **Martin**, Steven G, **Kim**, Sung-Hou (1998). "Structural basis for the interaction of Ras with RalGDS." Nat. Struct. Biol. **5**:422-426.

Huber, Fritz, **Gronwald**, Wolfram, **Wohlgemuth**, Sabine, **Herrmann**, Christian, **Geyer**, M., **Wittinghofer**, Alfred, **Kalbitzer**, Hans Robert (2000). "Sequential NMR assignment of the RAS-binding domain of Byr2." J Biomol NMR **16**(4): 355-356.

Heremans, K, **Smeller**, L (1998). "Protein Structure and Dynamics at High Pressure." Biochim. Biophys. Acta **1386**:353-370.

Holton, James, **Alber**, Tom (2004). "Automated Protein Crystal Structure Determination using ELVES." Proc Natl Acad Sci U.S.A. **101**:1537-1542.

Huber, Fritz (2001). "Strukturaufklärung der Ras-bindenden Domäne von Byr2 mit multidimensionaler NMR-Spektroskopie." Dissertation, Universität Regensburg.

Huber, Robert (1979). "Conformational flexibility in protein molecules." Nature **280**(5723): 538-539.

Ingram, Vernon Martin (1956). "A specific chemical difference between the globins of normal human and sickle-cell anaemia haemoglobin." Nature **178**(4537): 792-794.

Inoue, Kyoko, **Yamada**, Hiroaki, **Akasaka**, Kazuyuki, **Herrmann**, Christian, **Kremer**, Werner, **Maurer**, Till, **Döker**, Rolf, **Kalbitzer**, Hans Robert (2000). "Pressure-induced local unfolding of the Ras binding domain of RalGDS." Nat Struct Biol **7**: 547-550.

Inoue, Kyoto, **Maurer**, Till, **Yamada**, Hiroaki, **Herrmann**, Christian, **Horn**, Gudrun, **Kalbitzer**, Hans, Robert, **Akasaka**, Kazuyuki (2001). "High-pressure NMR study of the complex of a GTPase Rap1A with its effector RalGDS." FEBS Lett **506**(3):180-4.

Iwig, Jeffrey S, **Vercoulen**, Yvonne, **Das**, Rahul, **Barros**, Tiago, **Limnander**, Andre, **Che**, Yan, **Pelton**, Jeffrey G, **Wemmer**, David E, **Roose**, Jeroen P, **Kuriyan**, John (2013). "Structural analysis of autoinhibition in the Ras-specific exchange factor RasGRP1." Dötsch V, ed eLife 2013; 2:e00813.doi:10.7554/eLife.00813.

Jackson, Sophie E (2006). "Ubiquitin: a small protein folding paradigm." Org. Biomol. Chem. **4**:1845-1853.

John, J, **Sohmen**, R, **Feuerstein**, J, **Linke**, R, **Wittinghofer**, A, **Goody**, RS (1990). "Kinetics of interaction of nucleotides with nucleotides-free H-ras p21." Biochemistry **29**(25):6058-65.

Kalbitzer, Hans Robert, **Rosnizeck**, Ina C., **Munte**, Claudia Elisabeth, **Narayanan**, Sunilkumar Puthenpurackal, **Kropf**, Veronika, **Spoerner**, Michael (2013). "Intrinsische allosterische Hemmung von Signalproteinen durch Stabilisierung gering besetzter, durch Hochdruck-NMR-Spektroskopie nachweisbarer Interaktionszustände." Angewandte Chemie **125**: 14492-14496.

Kalbitzer, Hans Robert, **Spörner**, Michael, **Ganser**, Petra, **Hozsa**, Constantin, **Kremer**, Werner (2009). "Fundamental Link between Folding States and Functional States of Proteins." J Am Chem Soc **131**:16714-16719.

Karnoub, Antoine E, **Weinberg**, Robert A (2008). "Ras oncogenes: split personalities." Nat Rev Mol Cell Biol **9**:517-531.

Katchalski-Katzir, Ephraim, **Shariv**, I., **Eisenstein**, M., **Friesem**, Asher A., **Aflalo**, Claude, **Vakser**, Ilya A. (1992). "Molecular surface recognition: determination of geometric fit between proteins and their ligands by correlation techniques." Proc Natl Acad Sci U S A **89**(6): 2195-2199.

Kalashnikova, Anna A, **Porter-Goff**, Mary E, **Muthurajan**, Uma M, **Luger**, Karolin, **Hansen**, Jeffrey C. (2013). "The role of nucleosome acid patch in modulating higher order chromatin structure." J.R.Soc. Interface **10**:20121022.

Kitahara, Ryo, **Akasaka**, Kazuyuki (2003). "Close identity of a pressure-stabilized intermediate with a kinetic intermediate in protein folding." PNAS **100**(6):3167-3172.

Kitahara, Ryo, **Yokoyama**, Shigeyuki, **Akasaka**, Kazuyuki (2005). "NMR Snapshots of a Fluctuating Protein Structure: Ubiquitin at 30 bar - 3 kbar." J. Mol. Biol. **347**:277-285.

Kitazawa, Soichiro, **Kameda**, Tomoshi, **Kumo**, Ayumi, **Yagi-Utsomi**, Maho, **Baxter**, Nicola J, **Kato**, Koichi, **Williamson**, Mike P, **Kitahara**, Ryo (2014). "Close identity of the alternatively folded state N₂ of ubiquitin with the conformation of the protein bound to ubiquitin-activating enzyme." Biochemistry **53**(3):447-9.

Köhler, Jörg, **Beck-Erlach**, Markus, **Crusca Jr.**, Edson, **Kremer**, Werner, **Munte**, Claudia Elisabeth, **Kalbitzer**, Hans Robert (2012). "Pressure Dependence of ¹⁵N Chemical Shifts in Model peptides Ac-Gly-Gly-X-Ala-NH₂." Materials **5**: 1774-1786.

Korari, Reto, **Billeter**, Martin, **Wüthrich**, Kurt (1996). "MOLMOL: a program for display and analysis of macromolecular structures." J Mol Graphics **14**: 51-55.

Kremer, Werner, **Kachel**, Norman, **Kuwata**, Kazuo, **Akasaka**, Kazuyuki, **Kalbitzer**, Hans Robert (2007). "Species-specific differences in the intermediate states of human and Syrian hamster prion protein detected by high pressure NMR spectroscopy." J Biol Chem **282**(31): 22689-22698.

Laskowski, Roman A, **MacArthur**, Malcolm W, **Thornton**, Janet M (1998). "Validation of protein Models derived from experiment." Curr Opin Struct Biol **8**: 631-639.

Lichtarge, Olivier, **Bourne**, Henry, **Cohen**, Fred (1996). "Evolutionarily conserved Galphabetagamma binding surfaces support a model of the G protein-receptor complex." Proc Natl Acad Sci U S A **93**(15): 7507-7511.

Linge, Jens P, **Williams**, Mark A, **Spronk**, Christian AEM, **Bonvin**, Alexandre MJJ, **Nilges**, Michael (2003). "Refinement of Protein Structures in Explicit Solvent." Proteins **50**: 496-506.

Lipsitz, Rebecca S, **Tjandra**, Nico (2004). "Residual Dipolar Couplings in NMR Structure Analysis." Annu Ref Biophys Biomol Struct **33**: 387-413.

Makde, RD, **England**, JR, **Yennawar**, HP, **Tan**, S (2010). "Structure of RCC1 chromatin factor bound to the nucleosome core particle." Nature **467**:562-566.

Masuda, Tadayuki, **Kariya**, Ken-ichi, **Shinkai**, Masayuki, **Okada**, Tomoyo, **Kataoka**, Tohru (1995). "Protein kinase Byr2 is a target of Ras1 in the fission yeast *Schizosaccharomyces pombe*." J Biol Chem **270**(5): 1979-1982.

Markley, John L, **Bax**, Ad, **Arata**, Yoji, **Hilbers**, CW, **Kaptein**, Robert, **Sykes**, Brian D, **Wright**, Peter E, **Wüthrich**, Kurt (1998). "Recommendations for the presentation of NMR structures of protein and nucleic acids." Pure & Appl. Chem. **70**(1):117-142.

Marsh, Joseph A., **Teichmann**, Sarah A., **Forman-Kay**, Julie D. (2012). "Probing the diverse landscape of protein flexibility and binding." Curr Opin Struct Biol **22**: 643-650.

McCoy, Airlie J, **Epa**, V Chandana, **Colman**, Peter M (1997), "Electrostatic Complementarity at Protein/Protein Interfaces." J Mol Biol **268**: 570-584.

Merkel, Rainer, **Waack**, Stephan (2009). Bioinformatik Interaktiv. Grundlagen, Algorithmen, Anwendungen, 2. Auflage. Weinheim, Wiley-VCH.

Meyer, Michael, **Wilson**, Peter, **Schomburg**, Dietmar (1996). "Hydrogen Bonding and Molecular Surface Shape Complementarity as a Basis of Protein Docking." J Mol Biol **264**: 199-210.

Möglich, Andreas, **Weinfurtner**, Daniel, **Gronwald**, Wolfram, **Maurer**, Till, **Kalbitzer**, Hans Robert (2005). "PERMOL: restraint-based protein homology modeling using DYANA or CNS." Bioinformatics **21**(9): 2110-2111.

Mücke, Katrin, **Paulus**, Christina, **Bernhardt**, Katharina, **Gerrer**, Katrin, **Schön**, Kathrin, **Fink**, Alina, **Sauer**, Eva-Maria, **Asbach-Nitzsche**, Alexandra, **Harvardt**, Thomas, **Kieninger**, Bärbel, **Kremer**, Werner, **Kalbitzer**, Hans Robert, **Nevels**, Michael (2014). "Human Cytomegalovirus Major Immediate Early 1 Protein Targets Host Chromosomes by Docking to the Acid Pocket on the Nucleosome Surface". Journal of Virology **88**(2):1228.

Munte, Claudia Elisabeth, **Beck-Erlach**, Markus, **Kremer**, Werner, **Köhler**, Jörg, **Kalbitzer**, Hans Robert (2013). "Getrennte Konformationszustände des Alzheimer- β -Amyloidpeptids - Nachweis mit Hochdruck-NMR-Spektroskopie." Angewandte Chemie **125**: 9111-9116.

Needleman, Saul B, **Wunsch**, Christian D (1970). "A general method applicable to the search for similarities in the amino acid sequence of two proteins." J Mol Biol **48**(3): 443-53.

Orr, Jeffrey W, **Newton**, Alexandra C (1994) "Intra-peptide regulation of protein kinase C." J Biol Chem **269**(11):8383-7.

Paulus, Christina, **Nevels**, Michael (2009). "The human cytomegalovirus major immediate-early proteins as antagonists of intrinsic and innate antiviral host responses." Viruses **1**(3): 760-779.

Pazos, Florencio, **Helmer-Citterich**, Manuela, **Ausiello**, Gabriele, **Valencia**, Alfonso (1997). "Correlated mutations contain information about protein-protein interaction." J Mol Biol **271**(4): 511-523.

Petkova, Aneta T, **Yau**, Wai-Ming, **Tycko**, Robert (2006). "Experimental Constraints on Quaternary Structure in Alzheimer's β -Amyloid Fibrils." Biochemistry **45**:498-512.

Pingoud, Alfred, **Urbanke**, Claus, **Hogget**, Jim, **Jeltsch**, Albert (2002). "Biochemical Methods." , 1. Auflage. Weinheim, Wiley-VCH.

Rajalingam, Krishnaraj, **Schreck**, Ralf, **Rapp**, Ulf R, **Albert**, Štefan (2007). "Ras oncogenes and their downstream targets." Biochim Biophys Acta **1773**:1177-1195.

Ramachandran, GN, **Ramakrishnan**, C, **Sasisekharan**, V (1963): "Stereochemistry of polypeptide chain configurations." J Mol Biol **7**: 95-99.

Ritchie, David W. (2008). "Recent progress and future directions in protein-protein docking." Curr Protein Pept Sci **9**(1): 1-15.

Sachs, Lothar (2004). Angewandte Statistik: Angewandte Statistik, 11. Auflage. Berlin, Heidelberg, New York, Springer.

Rosnizeck, Ina, **Flichtinski**, Daniel, **Lopes**, Rui Pedro, **Kieninger**, Bärbel, **Herrmann**, Christian, **Kalbitzer**, Hans Robert, **Spörner**, Michael (2014). "Elucidating the Mode of Action of a Typical Ras State 1(T) Inhibitor." Biochemistry **53**(24):3867-78.

Sander, Chris, **Schneider**, Reinhard (1991). "Database of homology-derived protein structures and the structural meaning of sequence alignment." Proteins **9**(1): 56-68.

Schauber, Hannelore, **Eisenhaber**, Frank, **Argos**, Patrick (1992). "Rotamers: To be or not to be? An Analysis of Amino Acid Side-chain Conformations in Globular Proteins." J Mol Biol **230**: 592-612.

Schedlbauer, Andreas, **Auer**, Renate, **Ledolter**, Karin, **Tollinger**, Martin, **Kloiber**, Karin, **Lichtenecker**, Roman, **Ruedisser**, Simon, **Hommel**, Ulrich, **Schmid**, Walther, **Konrat**, Robert (2008). "Direct methods and residue type specific isotope labeling in NMR structure determination and model-driven sequential assignment." Journal of Biomolecular NMR **42**(2):111-127.

Scheffzek, Klaus, **Grunewald**, Petra, **Wohlgemuth**, Sabine, **Kabsch**, Wolfgang, **Tu**, Hua, **Wigler**, Mike, **Wittinghofer**, Alfred, **Herrmann**, Christian (2001). "The Ras-Byr2RBD complex: structural basis for Ras effector recognition in yeast." Structure **9**(11): 1043-1050.

Scheiber, Josef Heinrich (2003). "Entwicklung einer auf Moleküldynamik basierende Methode zur Homologie-Modellierung der Tertiärstruktur von Proteinen." Diplomarbeit, Universität Regensburg.

Scherer, Myriam, **Klingl**, Stefan, **Sevvana**, Madhumati, **Otto**, Victoria, **Schilling**, Eva-Maria, **Stump**, Joachim D, **Müller**, Regina, **Reuter**, Nina, **Sticht**, Heinrich, **Muller**, Yver A, **Stamminger**, Thomas (2014). "Crystal Structure of Cytomegalovirus IE1 Protein Reveals Targeting of TRIM Family Member PML via Coiled-Coil Interactions." PLoS Pathog **10**(11): e1004512.

Schneider, Philipp (2014). "NMR-Datenanalyse zur automatisierten stereospezifischen Zuordnung." Diplomarbeit, Universität Regensburg.

Schubbert, Suzanne, **Shannon**, Kevin, **Bollag**, Gideon (2007). Hyperactive Ras in developmental disorders and cancer. " Nat Rev Cancer **7**:295-308.

Schumann, Frank H, **Riepl**, Hubert, **Maurer**, Till, **Gronwald**, Wolfram, **Neidig**, Klaus-Peter, **Kalbitzer**, Hans Robert (2007). "Combined chemical shift changes and amino acid specific chemical shift mapping of protein-protein interactions." J Biomol NMR **39**(4): 275-289.

Schwieters, Charles D, **Kuszewski**, John J, **Clore**, G Marius (2006). "Using Xplor-NIH for NMR molecular structure determination." Prog. NMR Spectrosc. **48**: 47-62.

Shen, Yang, **Delaglio**, Frank, **Cornilescu**, Gabriel, **Bax**, Ad (2009). "TALOS+: A hybrid method for predicting protein backbone torsion angles from NMR chemical shifts." J Biomol NMR **44**: 213-223.

Sondermann, Holger, **Soisson**, Stephen M, **Boykevich**, Sean, **Yang**, Shao-Song, **Bar-Sagi**, Dafna, **Kuriyan**, John (2004). "Structural analysis of autoinhibition in the Ras activator Son of sevenless." Cell **119**(3): 393-405.

Spörner, Michael, **Herrmann**, Christian, **Vetter**, Ingrid R, **Kalbitzer**, Hans Robert, **Wittinghofer**, Alfred (2001). "Dynamic properties of the Ras switch I region and its importance for binding to effectors." Proc Natl Acad Sci USA **98**:4944-4949.

Spörner, Michael, **Nuehs**, Andrea, **Ganser**, Petra, **Herrmann**, Christian, **Wittinghofer**, Alfred, **Kalbitzer**, Hans Robert (2005). "Conformational states of Ras complexed with the GTP analogue GppNHp or GppCH2p: implications for the interaction with effector proteins." Biochemistry **44**:2225-2236.

Spörner, Michael, **Hosza**, Constantin, **Pötl**, Johann A, **Ganser**, Petra, **Geyer**, Matthias, **Kalbitzer**, Hans Robert (2010). "Conformational States of Human Rat Sarcoma (Ras) Protein Complexed with Its Natural Ligand GTP and Their Role for Effector Interaction and GTP Hydrolysis." J Biol Chem **285**(51):39768-39778.

Tamanai, Fukuhiko (2013). "The Enzymes: Inhibitors of the Ras Superfamily G-proteins, Part A." 1. Auflage. Amsterdam, Elsevier.

Temussi, Piero Andrea, **Masino**, Laura, **Pastore**, Annalisa (2003). "From Alzheimer zu Huntington: why ist a structural understanding so difficult?" EMBO J **22**:355-361.

Tobi, Dror, **Bahar**, Ivet (2005). "Structural changes involved in protein binding correlate with intrinsic motions of proteins in the unbound state." Proc Natl Acad Sci U S A **102**(52): 18908-18913.

Tu, Hua, **Barr**, Maureen, **Dong**, Dennis L., **Wigler**, Mike (1997). "Multiple regulatory domains on the Byr2 protein kinase." Mol Cell Biol **17**(10): 5876-5887.

Tucker, J, **Sczakiel**, G, **Feuerstein**, J, **John**, J, **Goody**, RS, **Wittinghofer**, A (1986). "Expression of p21 proteins in Escherichia coli and stereochemistry of the nucleotide-binding site." EMBO J. **5**(6):1351-8.

Ulrich, Eldon L., **Akutsu**, Hideo, **Doreleijers**, Jürgen F., **Harano**, Yoko, **Ioannidis**, Yannis E., **Lin**, Jundong, **Livny**, Miron, **Mading**, Steve, **Maziuk**, Dimitri, **Miller**, Zachary, **Nakatani**, Eiichi, **Schulte**, Christopher F., **Tolmie**, David E., **Wenger**, R. Kent, **Yao**, Hongyang, **Markley**, John L. (2008). "BioMagResBank." Nucleic Acids Res **36**(Database issue): D402-408.

Vivekanandan, Subrananian, **Brender**, Jeffrey R, **Lee**, Shirley Y, **Ramamoorthy**, Ayyalusamy (2011). "A partially folded structure of amyloid-beta(1-40) in an aqueous environment." Biochem. Biophys. Res. Commun. **411**:312-316.

Wittinghofer, Alfred, **Vetter**, Ingrid R (2011). "Structure-Function Relationship of the G Domain, a Canonical Switch Motif." Annu Rev Biochem **80**:943-71.

Wohlgemuth, Sabine, **Kiel**, Christina, **Krämer**, Astrid, **Serrano**, Luis, **Wittinghofer**, Fred, **Herrmann**, Christian (2005). "Recognizing and defining true Ras binding domains I: biochemical analysis." J Mol Biol **348**(3):741-58.

Wood, Todd C., **Pearson**, William R. (1999). "Evolution of protein sequences and structures." J Mol Biol **291**(4): 977-995.

Wüthrich, Kurt (2003). "NMR studies of structures and function of biological macromolecules (Nobel lectures)." Angew. Chem. Int. Ed **42**(29): 3340-3363.

Yoshimori, Atsushi, **Del Carpio M.**, Carlos A. (2001). "A Novel Protein Soft Docking Algorithm Based on Spectral Analysis of Shape Complementarity." Genome Informatics **12**: 354-355.

Zweckstetter, Markus (2008). "NMR: prediction of molecular alignment from structure using the PALES software." Nature Protocols **3**(4): 679-690.

7 Eigene Publikationen

Rosnizeck, Ina, **Flichtinski**, Daniel, **Lopes**, Rui Pedro, **Kieninger**, Bärbel, **Herrmann**, Christian, **Kalbitzer**, Hans Robert, **Spörner**, Michael (2014). "Elucidating the Mode of Action of a Typical Ras State 1(T) Inhibitor." *Biochemistry* **53**(24):3867-78.

Mücke, Katrin, **Paulus**, Christina, **Bernhardt**, Katharina, **Gerrer**, Katrin, **Schön**, Kathrin, **Fink**, Alina, **Sauer**, Eva-Maria, **Asbach-Nitzsche**, Alexandra, **Harvardt**, Thomas, **Kieninger**, Bärbel, **Kremer**, Werner, **Kalbitzer**, Hans Robert, **Nevels**, Michael (2014). "Human Cytomegalovirus Major Immediate Early 1 Protein Targets Host Chromosomes by Docking to the Acic Pocket on the Nucleosom Surface". *Journal of Virology* **88**(2):1228.

Kieninger, Bärbel, **Danilova**, Nina, **Cano**, Carolina, **Brunner**, Konrad, **Schneider**, Philipp, **de Sanctis**, Silvia, **Donaubauer**, Harald, **Harsch**, Tobias, **Kremer**, Werner, **Kalbitzer**, Hans Robert (2013). "ISIC and PERMOL - multifunctional tools for the determination of protein structure." Poster, GDCH 2013.

Danilova, Nina, **Kieninger**, Bärbel, **Schneider**, Philipp, **Cano**, Carolina, **Brunner**, Konrad, **Arnold**, Martin Reinhard, **Kitahara**, Ryo, **Elsner**, Ralph, **Huber**, Fritz, **Akasaka**, Kazuyuki, **Herrmann**, Christian, **Kremer**, Werner, **Kalbitzer**, Hans Robert (2013). "Analysis of high pressure induced conformational changes of TmCSP and Byr2-RBD using AUREMOL tools." Poster, GDCH 2013.

Donaubauer, Harald, **Harsch**, Tobias, **Malloni**, Wilhelm, **de Sanctis**, Silvia, **Kieninger**, Bärbel, **Danilova**, Nina, **Brunner**, Konrad, **Trenner**, Jochen, **Gronwald**, Wolfram, **Neidig**, Klaus-Peter, **Kalbitzer**, Hans Robert (2011). "AUREMOL: New modules for data processing via the AUREMOL-SSA/ICA and protein quality control." Poster, Euromar 2011.

Donaubauer, Harald, **Harsch**, Tobias, **Malloni**, Wilhelm, **de Sanctis**, Silvia, **Kieninger**, Bärbel, **Danilova**, Nina, **Brunner**, Konrad, **Trenner**, Jochen, **Gronwald**, Wolfram, **Neidig**, Klaus-Peter, **Kalbitzer**, Hans Robert (2010). "AUREMOL: Automatic protein structure determination from NMR data." Poster, GDCH 2010.

Donaubauer, Harald, **Harsch**, Tobias, **Malloni**, Wilhelm, **de Sanctis**, Silvia, **Kieninger**, Bärbel, **Danilova**, Nina, **Brunner**, Konrad, **Trenner**, Jochen, **Gronwald**, Wolfram, **Neidig**, Klaus-Peter, **Kalbitzer**, Hans Robert (2009). "AUREMOL: Automatic protein structure determination from NMR data." Poster, Frühjahrstagung DPG 2009.

Donaubauer, Harald, **Harsch**, Tobias, **Malloni**, Wilhelm, **de Sanctis**, Silvia, **Kieninger**, Bärbel, **Danilova**, Nina, **Brunner**, Konrad, **Trenner**, Jochen, **Gronwald**, Wolfram, **Neidig**, Klaus-Peter, **Kalbitzer**, Hans Robert (2008). "AUREMOL: Automatic protein structure determination from NMR data." Poster, GDCH 2008.

8 Abkürzungsverzeichnis

Aminosäuren

A (oder Ala)	Alanin
C (oder Cys)	Cystein
D (oder Asp)	Asparaginsäure
E (oder Glu)	Glutaminsäure
F (oder Phe)	Phenylalanin
G (oder Gly)	Glycin
H (oder His)	Histidin
I (oder Ile)	Isoleucin
K (oder Lys)	Lysin
L (oder Leu)	Leucin
M (oder Met)	Methionin
N (oder Asn)	Asparagin
P (oder Pro)	Prolin
Q (oder Gln)	Glutamin
R (oder Arg)	Arginin
S (oder Ser)	Serin
T (oder Thr)	Threonin
W (oder Trp)	Tryptophan
Y (oder Tyr)	Tyrosin

Übrige Abkürzungen

CMV	Cytomegalievirus
CTD	C-terminale Domäne
DNS	Desoxyribonukleinsäure
GAP	GTPase-aktivierendes Protein
GDP	Guanosindiphosphat
GEF	Guanosintriphosphat -Austauschfaktor (Guanine Nucleotide Exchange Factor)
GppNHp	5'-Guanylyl Imidodiphosphat
GTP	Guanosintriphosphat
HSQC	2D-heteronuklear korreliertes NMR-Experiment (Heteronuclear Single Quantum Coherence)
K	Kelvin
MPa	Megapascal
nm	Nanometer
NMR	Kernspinspektroskopie
ppm	part per million
RBD	Ras-Bindedomäne
RNS	Ribonukleinsäure
TROSY	NMR-technisches Verfahren insbesondere zur Untersuchung großer Proteine (Transverse Relaxation Optimized Spectroscopy)
X-RAY	Röntgenkristallographie

9 Danksagung

Zum Abschluss meiner Arbeit möchte ich noch meine Dankbarkeit gegenüber den folgenden Personen aussprechen:

- Meinem Doktorvater, Herrn Prof. Dr. Dr. Hans Robert Kalbitzer, der mir die Chance gab, an seinem Lehrstuhl diese Arbeit anzufertigen und sie durch viele interessante Diskussionen und Ideen bereicherte.
- PD Dr. Michael Spörner und Dr. Ina Rosnizeck für die fruchtbare Zusammenarbeit am Ras-Protein und bei der tatkräftigen Unterstützung bei der Fertigstellung dieser Arbeit.
- Prof. Dr. Wolfram Gronwald, der mir die Daten für das Protein Byr2 zur Verfügung stellte und der mir insbesondere am Anfang half, die Theorie und Technik der NMR an Proteinen zu verstehen und Strukturrechnungen durchzuführen.
- Prof. Dr. Werner Kremer, der mir die Hochdruckdaten für das Protein Byr2 zur Verfügung stellte und der mich bei der Arbeit an dem Protein IE1 unterstützte.
- Prof. Dr. Claudia E. Munte, die mir geduldig meine Fragen beantwortete, viele Anregungen lieferte und mich mit ihrer Begeisterung für die Wissenschaft ansteckte und mir außerdem die Daten für das Protein Aß zur Verfügung stellte.
- Meinen Arbeitskollegen am Lehrstuhl für das sehr nette Arbeitsklima und insbesondere meinen lieben Kollegen vom AUREMOL-Developer-Team, vor allem Nina Danilova, Tobias Harsch und Harald Donaubaueer.
- Der Arbeitsgruppe Mikrobiologie, insbesondere Dr. Katrin Mücke und Prof. Dr. Michael Nevels, für die interessante und fruchtbare Zusammenarbeit am Protein IE1.
- Michael Ebel, der während seiner Zeit am Lehrstuhl das Alignment des Programms PERMOL neu implementierte, und der mir bei der Strukturierung von PERMOL wertvolle Hinweise gab und außerdem ein zuverlässiger Ansprechpartner für alle mathematischen oder MATLAB-Probleme war.
- Philipp Schneider, der im Rahmen seiner Diplomarbeit das Programm Iupacify schrieb, wodurch er mich bei der Entwicklung von PERMOL unterstützte und mir außerdem die Verwendung von Python nahebrachte.
- Meinen Eltern, meinen Schwiegereltern und ganz besonders Hanne und Gert Althaus, die mich bei der Kinderbetreuung und im Haushalt wunderbar unterstützten

und mir in stressigen Zeiten den Rücken freihielten.

- Und natürlich meiner lieben Familie, meinem Ehemann Martin und unseren Kindern Anna, Florian und Timo, die immer zu mir halten.

DANKE!