

Neue Verfahren zur Beschreibung der
dynamischen Struktur von Biomolekülen
am Beispiel Ribonuklease T_1

Dissertation
zur Erlangung des Doktorgrades
der Naturwissenschaften

vorgelegt beim Fachbereich für Chemische und Pharmazeutische
Wissenschaften der Johann Wolfgang Goethe-Universität in Frankfurt am
Main

von
Dipl.-Chem. Ulrich Schieborr
aus Essen

Frankfurt am Main, 2001
(DF1)

Für Silke, Melina und Julia

Danksagung

Die vorliegende Arbeit entstand unter Leitung von Herrn Prof. Dr. Heinz Rüterjans in der Zeit von Oktober 1996 bis August 2001 am Institut für Biophysikalische Chemie der Johann Wolfgang Goethe-Universität in Frankfurt am Main.

Mein besonderer Dank gilt meinem akademischen Lehrer Prof. Dr. Heinz Rüterjans für die Schaffung hervorragender Arbeitsbedingungen sowie für die Freiheit bei der Auswahl und Bearbeitung des Themas.

Frau Dr. Stefania Pfeiffer und Herrn Dr. Christian Ludwig danke ich für wertvolle Anmerkungen, sowie die ständige Bereitschaft, über meine „rein hypothetischen“ Fragestellungen zu diskutieren.

Frau Dipl.-Biochem. Heide Muckenfuß gilt mein Dank für die Mitarbeit bei der Diederwinkel-Auswertung.

Herrn Dr. Christian Ludwig und Herrn Dr. Jan Engelke danke für die Überlassung der Relaxationsdaten an RNase T_1 .

Außerdem möchte ich allen anderen Mitgliedern des Arbeitskreises für das angenehme Arbeitsklima danken.

Vom Fachbereich Chemische und Pharmazeutische Wissenschaften der Johann
Wolfgang Goethe-Universität als Dissertation angenommen

Dekan: Prof. Dr. Walter E. Müller
1. Gutachter: Prof. Dr. Heinz Rüterjans
2. Gutachter: Prof. Dr. Ernst Bamberg

Tag der Disputation:

Inhaltsverzeichnis

1	Einleitung und Problemstellung	4
2	Theoretische Grundlagen und Methoden	8
2.1	Molekulardynamik-Simulationen	8
2.1.1	Der Integrationsalgorithmus	8
2.1.2	Das Kraftfeld	10
2.1.3	Der LINCS-Algorithmus	11
2.1.4	Simulationsbedingungen	14
2.2	Kollektive Bewegungsmoden	15
2.2.1	Normalmodenanalysen	17
2.2.2	Dynamik aus Strukturensamples	20
2.2.3	Zusammenhang zwischen NMA und EDA	22
2.2.4	Projektion von Proteinbewegung auf kollektive Koordinaten	24
2.2.5	Energiehyperfläche im Raum der kollektiven Koordinaten	25
2.2.6	„Sampling“ Methoden	26
2.2.7	Absolute Entropien	29
2.3	Berechnung von Ordnungsparametern	32
3	Ergebnisse und Diskussion	34
3.1	Molekulardynamik-Simulation an RNase T_1	34
3.1.1	Simulationsbedingungen	34
3.2	Analyse der essentiellen Dynamik	34
3.2.1	Stabilität der Eigenvektoren	35
3.2.2	Die Modenanzahl für den essentiellen Raum	39
3.2.3	Vergleich mit der PCA-Analyse im Distanzraum	45
3.3	Vorurteilsfreie kartesische Moden	53
3.3.1	Annahmen und Berechnungsprozeduren	55
3.3.2	Transformation von Distanzmoden in kartesische Moden .	57

3.3.3	Verhältnis zwischen den Amplituden in Distanzraum und den Amplituden in kartesischen Raum	58
3.3.4	Eliminierung des Rotationsanteils und des Translationsanteils von Bewegungsmoden	59
3.3.5	Berechnung einer neuen Kovarianzmatrix der positionellen Abweichungen	60
3.3.6	Anwendung auf zwei Modellsysteme	61
3.3.7	Anwendung auf ein reales Molekül: RNase T_1	68
3.4	Diederwinkelanalyse	74
3.5	Vergleich mit anderen Methoden	84
3.5.1	Ordnungsparameter	84
3.5.2	NMR Struktur-Cluster	88
4	Zusammenfassung	93
5	Anhang	114
5.1	Herleitung der angepassten Eigenwerte λ_i^{ortho}	114
5.2	Tabellen der Gauß'schen Verteilungen der Diederwinkel	117

Abkürzungsverzeichnis

ACSIP	A verage C umulative S quare I nner P roduct (Mittleres kumulatives quadratisches inneres Produkt)
BPTI	B ovine P ancreatic T rypsin I nhibitor (Pankreatischer Trypsininhibitor aus Rindern)
DG	D istanz G eometrie
EDA	E ssential D ynamics A nalysis (Analyse der essentiellen Dynamik)
JAM	J umping A long M inima model (Minima-Sprungmodell)
LINCS	L INear C onstraint S olver
MC	M onte C arlo
MD	M olekular D ynamik
MSF	M ean S quare F luctuation (mittlere quadratische Fluktuation)
NMA	N ormal M oden A nalyse
NMR	N uclear M agnetic R esonance (Kernmagnetische Resonanz)
PCA	P rincipal C omponent A nalysis (Hauptkomponentenanalyse)
PCR	P rincipal C omponent R estraint (Zwang auf die Hauptkomponenten)
RMSD	R oot M ean S quare D eviation (Wurzel aus den mittleren quadratischen Distanzen)
RMSF	R oot M ean S quare F luctuation (Wurzel aus den mittleren quadratischen Fluktuationen)
RNase	R ibo N uklease
SVD	S ingular V alue D ecomposition (Singularwertzerlegung)
TAM	T otal A mount of M otion (Totale Menge an Bewegung)

1 Einleitung und Problemstellung

In lebenden Organismen nehmen Biomoleküle wie Proteine und Nucleinsäuren die verschiedensten Aufgaben wahr. Aus der DNA ergibt sich ein Bauplan für die Proteine, die ihrerseits als hochwirksame und hochspezifische Katalysatoren auftreten können oder als Botenstoffe oder auch als Transportmittel für andere Moleküle fungieren.

Proteine setzen sich aus einer Abfolge der 20 natürlich vorkommenden Aminosäuren zusammen, die über Amidbindungen miteinander verbunden sind. Die Abfolge der Aminosäurereste, die das Molekül bilden, wird als *Primärstruktur* eines Proteins bezeichnet. Die Primärstruktur bestimmt, welche Konformation das Molekül einnimmt und welche Funktion es im Organismus ausübt. Bisher sind keine Verfahren bekannt, mit deren Hilfe aus der Kenntnis der Aminosäuresequenz die Struktur oder die Funktion des Proteins vorherbestimmt werden kann. Es wurde zwar bereits mit sehr hohem Rechenaufwand versucht, die Faltung eines Proteins zu simulieren [1], dennoch war ein rein rechen technisches Verfahren bisher nicht erfolgreich. Daher bedient man sich physikalischer Methoden, um die Struktur der Biomoleküle experimentell zu ermitteln.

Zwei Verfahren werden verwendet, um zu hochaufgelösten Strukturen von Biomolekülen zu gelangen. Zum einen können Röntgenbeugungs-Experimente an Einkristallen durchgeführt werden. Diese Methode hat den Vorteil, daß die Größe der zu vermessenden Moleküle praktisch unbegrenzt ist [2]. Für diese Art von Strukturaufklärung muß es aber möglich sein, Einkristalle des molekularen Systems herzustellen. Auch kann nicht mit Gewißheit davon ausgegangen werden, daß die Konformation im Einkristall und in wäßriger Lösung - der natürlichen Umgebung der meisten Proteine - identisch ist. Außerdem sind dynamische Parameter, also die Varianzen in der Konformation, mit Beugungsexperimenten nur sehr begrenzt meßbar. Die NMR-Spektroskopie [3-5] ist die zweite physikalische Methode zur Strukturaufklärung. Sie ist zwar auf kleinere Proteine beschränkt, bietet jedoch den Vorteil, daß die Biomoleküle in

Lösung, also in annähernd natürlicher Umgebung, vermessen werden können. Die NMR-Spektroskopie gewährt außerdem Einblick in dynamische Vorgänge auf verschiedenen Zeitskalen von 10^{-12} bis 10^4 s [6, 7].

An die Strukturaufklärung mittels NMR oder Röntgenbeugung schließt sich die Verfeinerung der dynamischen Struktur mit Hilfe von *in silico* Verfahren an. In diesem Zusammenhang ist die Methode der Molekulardynamik-Simulationen (MD) ein sehr wichtiges Verfahren. In MD-Simulationen werden zeitliche Trajektorien der Konformation von Molekülen berechnet. Dabei wird die natürliche Umgebung meist durch explizite Wassermoleküle in der Rechnung berücksichtigt. Die MD-Simulationen dienen neben der Verfeinerung der Struktur insbesondere auch der Identifizierung von korrelierten Bewegungen innerhalb der Moleküle.

Ziel dieser Arbeit ist es, Verfahren zu entwickeln, mit deren Hilfe aus den strukturellen Ensembles von MD-Trajektorien sowohl die dynamische Struktur als auch korrelierte Bewegungen identifiziert und beschrieben werden können.

Als Modellsystem für die Methodenentwicklung wird das Enzym Ribonuklease T_1 (RNase T_1) gewählt. Das Molekül gehört zu der Gruppe der mikrobiellen endolytisch spaltenden Hydrolasen. Es spaltet spezifisch einzelsträngige RNA am 3'-Ende von Guaninnukleotiden, jedoch nicht DNA, doppelsträngige RNA oder organische Phosphordiester [8–11]. RNase T_1 wurde 1957 erstmals aus Takadiastase, einem Extrakt des Schimmelpilzes *Aspergillus oryzae*, isoliert [12]. Die Sequenz wurde 1965 von Takahashi aufgeklärt. RNase T_1 wird mittlerweile in größeren Mengen gentechnisch aus *E. coli* gewonnen [13, 14], wodurch umfangreiche Untersuchungen möglich geworden sind. Mehrere Röntgenstrukturanalysen sowohl von unkomplexierter RNase T_1 [15] als auch im Komplex mit verschiedenen Inhibitoren wurden durchgeführt [16–29]. Auch mit NMR-Methoden wurde eine hochaufgelöste Struktur einschließlich der Hydratation beschrieben [30–32]. Aber auch Relaxationsparameter, die sich besonders für den Vergleich mit *in silico* Experimenten eignen, wurden mit Hilfe der NMR-Spektroskopie bestimmt [33, 34]. RNase T_1 besteht aus 104 Aminosäureresten,

wobei Cys2 und Cys10 sowie Cys6 und Cys103 durch Disulfidbrücken miteinander verbunden sind. Eine reduktive Spaltung der Disulfidbrücken führt zum Verlust der enzymatischen Aktivität [35]. Für einen Vergleich mit Strukturenensembles aus Molekulardynamik-Rechnungen unterscheidet sich die Struktur des gefalteten Proteins in wässriger Lösung nicht wesentlich von der Struktur im Kristall. Aminosäurereste 13 bis 29 bilden eine 4,5 Windungen umfassende α -Helix. Das zentrale fünfsträngige β -Faltblatt ist linksgängig verdrillt und erstreckt sich über die Reste 39-42, 55-62, 76-81, 85-91 und 100-103. Ein kurzes zweisträngiges β -Faltblatt befindet sich im N-terminalen Sequenzabschnitt an den Positionen 4-6 und 9-12.

Im Rahmen dieser Arbeit wird eine MD-Simulation an RNase T_1 durchgeführt und mit Hilfe der entwickelten Methoden ausgewertet. Dabei sind Korrelationen zwischen weit entfernten Atomen von besonderem Interesse, denn solche korrelierten Bewegungen könnten etwa konzertierte Bindungsvorgänge oder andere funktionell relevanten Konformationsübergänge beschreiben [36]. Ichiye und Karplus haben 1991 an Rinderpankreas-Trypsininhibitor (BPTI) Berechnungen durchgeführt [37], aus denen hervorging, daß die Korrelation der Bewegung zwischen zwei Atomen von deren Abstand abhängt. Demnach fällt zunächst die mittlere Korrelation in BPTI von Eins für einen Abstand von Null bis ungefähr zum Radius des Proteins auf einen signifikant negativen Wert. Vom Radius bis zum Durchmesser steigt die mittlere Korrelation dann wieder auf positive Werte an. Dieses Ergebnis legt nahe, daß Atome auf der Oberfläche an gegenüber liegenden Positionen eine gemeinsame Aufgabe für die Funktion des Biomoleküls haben. Die Aussagen der Autoren [37] wurden in der Vergangenheit kontrovers diskutiert [36, 38, 39]. Es ist strittig, ob der beobachtete Effekt einen physikalischen Ursprung hat oder ein Artefakt der Kalkulationen ist. Das Problem bei der Berechnung von Korrelationen bzw. Kovarianzen der Bewegung ist dadurch begründet, daß die Strukturen eines Ensembles in kartesischen Koordinaten vorliegen und vor dem Vergleich überlagert werden müssen, um die globale Bewegung von internen strukturellen Fluktuationen zu trennen. Dazu

wird definiert, welche Gruppe von Atomen die Referenz für die Überlagerung bildet. Diese Definition ist immer in einem gewissen Maße willkürlich, denn innerhalb eines Biomoleküls gibt es keine Atome, die über eine Trajektorie hinweg vollkommen fixiert bleiben. Unterschiedliche Referenzsysteme führen aber auch zu unterschiedlichen Kovarianzen. Eine detaillierte Auswertung der Kovarianzmatrix ist somit mit dem gängigen Verfahren nicht möglich. Die erste Aufgabe dieser Arbeit ist deswegen die Entwicklung eines Verfahrens zur Trennung von interner und globaler Bewegung in Biomolekülen. Das Verfahren soll frei von der Willkür der Definition eines Referenzsystems sein, und es soll zu Kovarianzen führen, die frei von Resten globaler Bewegung sind.

Aus Kovarianzmatrizen lassen sich durch Diagonalisierung Moden korrelierter Bewegung ermitteln, die die dynamische Struktur beschreiben. Die Anzahl der Moden, die zur Beschreibung des strukturellen Ensembles benötigt werden, beträgt etwa ein zehntel der Anzahl der kartesischen Koordinaten [40]. Ein Nachteil der Beschreibung der dynamischen Struktur im Modenraum liegt darin, daß die Amplituden einzelner Moden immer nur eine globale Information der Konformation liefern. Lokale Struktureigenschaften, die auch mit Hilfe der NMR-Spektroskopie meßbar sind, ergeben sich erst aus der Kombination der Amplituden aller relevanten Moden. Zur lokalen Beschreibung der dynamischen Struktur wurden im Rahmen dieser Arbeit Diederwinkel benutzt. Die zweite Aufgabe dieser Arbeit ist die Entwicklung eines Verfahrens, das die Populationen der Diederwinkel in den Strukturen mit einfachen Potentialfunktionen erklären kann. Diese Potentialfunktionen könnten zur weiteren Aufklärung der dynamischen Struktur von Biomolekülen dienen.

2 Theoretische Grundlagen und Methoden

2.1 Molekulardynamik-Simulationen

Molekulardynamik-Simulationen beruhen auf den Gesetzen Newton'scher Mechanik, obwohl allgemein bekannt ist, daß sich Systeme auf atomarer Ebene nach quantenmechanischen Gesetzen verhalten. Ein einfacher Weg, um festzustellen, ob die klassische Näherung realistisch ist, basiert auf der thermischen de Broglie Wellenlänge Λ ,

$$\Lambda = \sqrt{\frac{2 \cdot \pi \cdot \hbar^2}{M \cdot k_B \cdot T}} \quad (1)$$

wobei M die atomare Masse des Systems, T die absolute Temperatur, k_B die Boltzmann-Konstante und \hbar die Planck-Konstante geteilt durch 2π ist. Die klassische Näherung ist erfüllt, wenn $\Lambda \ll a$, wobei a der mittlere Abstand zum nächsten Nachbarn ist. Für niedrige Temperaturen bzw. Massen in der Größenordnung weniger Dalton ist die Bedingung sicher nicht erfüllt. Nimmt man aber eine typische physiologische Temperatur von 300 K und eine Masse von 10 kD an, dann ist $\Lambda \approx 0.01 \text{ \AA}$. Der Abstand eines Atoms zu seinem nächsten Nachbarn beträgt etwa 1 \AA , womit die Bedingung $\Lambda \ll a$ erfüllt ist. Molekulare Dynamik läßt sich also tatsächlich mit Hilfe der Newton'schen Mechanik beschreiben.

2.1.1 Der Integrationsalgorithmus

Der Grundgedanke der Molekulardynamik-Simulation ist die iterative Integration der Newton'schen Bewegungsgleichungen für die Gesamtheit aller Atome eines molekularen Systems. Es existieren verschiedene Algorithmen für die Lösung dieses Problems [41], von denen nur der hier verwendete „leap-frog“-Algorithmus [42, 43] erläutert wird, der eine Alternative zum ebenfalls weit verbreiteten Verlet-Algorithmus [44, 45] darstellt.

Entwickelt man eine Taylorreihe bis zur dritten Ordnung um die Positionen der Atome \vec{x} zu den Zeitpunkten $t + \frac{\Delta t}{2}$ und $t - \frac{\Delta t}{2}$, so erhält man die Gleichungen

2 und 3. Δt ist die Länge eines Integrationsschritts $\overrightarrow{v(t)}$, $\overrightarrow{a(t)}$ und $\overrightarrow{b(t)}$ sind die Geschwindigkeit, die Beschleunigung und die Änderung der Beschleunigung zum Zeitpunkt t .

$$\overrightarrow{v\left(t + \frac{\Delta t}{2}\right)} = \overrightarrow{v(t)} + \overrightarrow{a(t)} \cdot \frac{\Delta t}{2} + \overrightarrow{b(t)} \cdot \left(\frac{\Delta t}{2}\right)^2 + O(\Delta t^3) \quad (2)$$

$$\overrightarrow{v\left(t - \frac{\Delta t}{2}\right)} = \overrightarrow{v(t)} - \overrightarrow{a(t)} \cdot \frac{\Delta t}{2} + \overrightarrow{b(t)} \cdot \left(\frac{\Delta t}{2}\right)^2 + O(\Delta t^3) \quad (3)$$

Subtrahiert man Gleichung 3 von Gleichung 2, so ergibt sich Gleichung 4 mit einem Fehler im dritten Glied der Reihenentwicklung. Gleichung 4 wird zur iterativen Berechnung der Geschwindigkeiten benutzt.

$$\overrightarrow{v\left(t + \frac{\Delta t}{2}\right)} = \overrightarrow{v\left(t - \frac{\Delta t}{2}\right)} + \overrightarrow{a} \cdot \Delta t \quad (4)$$

Unter der Annahme, daß die mittlere Geschwindigkeit innerhalb des Integrationszeitraums gleich der Geschwindigkeit nach der halben Integrationszeit ist, ergibt sich Gleichung 5 für die iterative Berechnung der Ortskoordinaten.

$$\overrightarrow{x(t + \Delta t)} = \overrightarrow{x(t)} + \overrightarrow{v\left(t + \frac{\Delta t}{2}\right)} \cdot \Delta t \quad (5)$$

Abbildung 1 zeigt das Schema der Integrationsmethode. Die Geschwindigkeiten und die Ortskoordinaten werden um einen Zeitraum von $\frac{\Delta t}{2}$ versetzt berechnet. Der Algorithmus wird „leap-frog“ genannt, weil \overrightarrow{x} und \overrightarrow{v} wie Frösche gegenseitig über ihren Rücken springen. Man geht von gegebenen Ortskoordinaten zum Zeitpunkt $t = 0$ und entsprechend der Simulationstemperatur nach eine Maxwell-Verteilungsfunktion „zufällig“ ausgewählten Geschwindigkeiten zum Zeitpunkt $t = -\frac{\Delta t}{2}$ aus. Über Kraftkomponenten $\overrightarrow{f}(\overrightarrow{x})$ können mit Hilfe von Gleichung 4 unter Verwendung des Zusammenhangs zwischen Kraft und Beschleunigung $\overrightarrow{f} = m \cdot \overrightarrow{a}$ neue Geschwindigkeiten zum Zeitpunkt $t = \frac{\Delta t}{2}$ berechnet werden. Aus den berechneten Geschwindigkeiten, den Anfangsgeschwindigkeiten, den Anfangskoordinaten lassen sich mit Hilfe von Gleichung 5

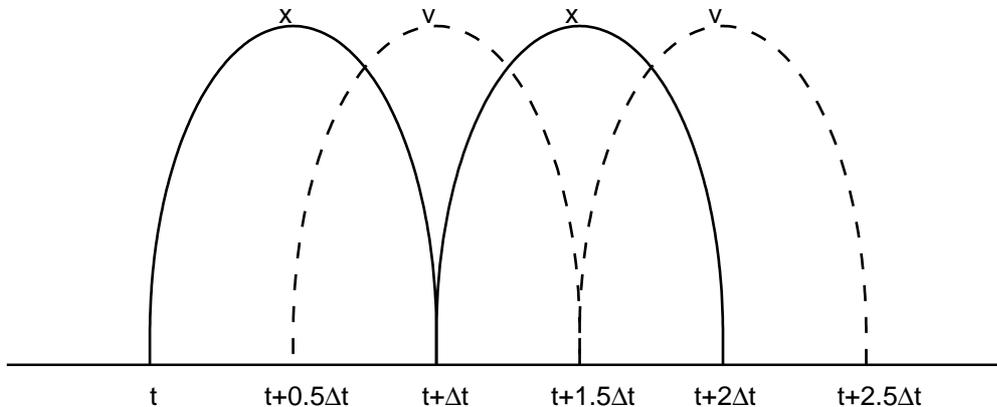


Abbildung 1: Die leap-frog-Integrationsmethode: Der Algorithmus wird „leap-frog“ genannt, weil \vec{x} und \vec{v} wie Frösche gegenseitig über ihren Rücken springen.

die Ortskoordinaten zum Zeitpunkt $t = \Delta t$ berechnen. Die erhaltene Konformation erlaubt nun die erneute Berechnung der Kraftkomponenten.

2.1.2 Das Kraftfeld

Die in jedem Zeitschritt zu berechnenden Beschleunigungen werden durch ein sogenanntes Kraftfeld definiert, das die vollständige Information über die Topologie des des molekularen Systems enthält. Gleichung 6 beschreibt das Potential V des Systems in Abhängigkeit von den Ortskoordinaten \vec{x} . Die Differentiation von V bezüglich der Ortskoordinaten führt zu den Kräften \vec{f} ($\nabla V = \vec{f}$). Es wird zwischen bindenden (Bindungslängen, Bindungswinkel, Diederwinkel, Chiralitätszentren) und nicht-bindenden (van-der-Waals und elektrostatische) Wechselwirkungen unterschieden. Die in Gleichung 6 mit dem Index „0“ gekennzeichneten Größen sind Gleichgewichtswerte, die durch das Kraftfeld definiert werden. Der Potentialterm für kovalente Bindungen z.B. ist proportional zur Differenz zwischen dem Gleichgewichtsabstand der beteiligten Atome b_0 und dem Abstand b , den die beiden Atome in der vorliegenden Struktur aufweisen. Die Kraftkonstante k_{bi} legt fest, wie stark die Rückstellkraft mit der Verletzung des Gleichgewichtsabstandes ansteigen soll. Implizit wird durch k_{bi} auch die

Frequenz der Bindungsschwingung festgelegt. Die Terme für die Bindungswinkel und Chiralitätszentren in Gleichung 6 sind analog der Definition der Bindungslängen. Das Potential als Funktion des Diederwinkel φ weist abhängig von den Substituenten Minima bzw. Maxima für die gestaffelten und überlagerten Konformationen auf. Die Definitionen der Lennard-Jones- und der Coulomb-Wechselwirkungen weisen keine harmonischen Potentiale auf und werden durch die van-der-Waals-Parameter C_{12} bzw. C_6 und die Dielektrizitätskonstante ϵ_r bestimmt.

$$\begin{aligned}
V(\vec{x}) = & \frac{1}{2} \cdot \sum_i k_{b_i} \cdot (b - b_0)^2 && \text{kovalente Bindungen} \\
& + \frac{1}{2} \cdot \sum_i k_{\theta_i} \cdot (\theta - \theta_0)^2 && \text{Bindungswinkel} \\
& + \frac{1}{2} \cdot \sum_i k_{\xi_i} \cdot (\xi - \xi_0)^2 && \text{Chiralitäts - Planaritätszentren} \\
& + \frac{1}{2} \cdot \sum_i k_{\varphi_i} \cdot (1 + \cos(n \cdot \varphi - \delta))^2 && \text{Torsionswinkel} \\
& + \frac{1}{2} \cdot \sum_i \sum_j 2 \cdot \left(\frac{C_{12}}{r_{ij}^{12}} - \frac{C_6}{r_{ij}^6} \right) \cdot S(r) && \text{Lennard - Jones - Wechselwirkung} \\
& + \frac{1}{2} \cdot \sum_i \sum_j 2 \cdot \left(\frac{q_i \cdot q_j}{4 \cdot \pi \cdot \epsilon_0 \cdot \epsilon_r \cdot r_{ij}} \right) \cdot S(r) && \text{Coulomb - Wechselwirkung}
\end{aligned} \tag{6}$$

Die Qualität eines Kraftfeldes hängt entscheidend von dessen Parametrisierung ab. Es ergeben sich für verschiedene Kraftfelder unterschiedliche Ergebnisse bei energetischen Betrachtungen oder Dynamik-Simulationen [46]. Die Kraftkonstanten und Gleichgewichtsparameter werden in einem sogenannten Parametrisierungsverfahren über quantenmechanische Rechnungen oder über den Vergleich mit experimentellen Daten bestimmt.

2.1.3 Der LINCS-Algorithmus

LINCS (**LI**Near **C**onstraint **S**olver) ist ein Algorithmus der die Bindungslängen nach einem Integrationsschritt auf ihre Gleichgewichtswerte korrigiert [47–49]. Der LINCS-Algorithmus wird im Rahmen dieser Arbeit angewendet, weil er stabiler ist und rascher zu Ergebnissen führt als der alternative SHAKE-

Algorithmus [50].

In einem System mit N Partikeln und N Positionsvektoren \vec{x}_i wird in Molekulardynamik-Simulationen Newtons Gesetz $\vec{F}_i = m \cdot \vec{a}_i$ verwandt, wobei mit \vec{F}_i der Kraft-Vektor und mit \vec{a}_i die Beschleunigung auf Atom i dargestellt wird. In dem System treten K zeitunabhängige Zwänge auf, die durch Gleichung 7 beschrieben werden. Hierbei ist $g_j(\vec{x})$ die Differenz aus dem Gleichgewichtsbindungsabstand zweier betrachteter Atome und dem tatsächlichen Abstand in der vorliegenden Struktur $|\vec{x}_{j_1} - \vec{x}_{j_2}|$.

$$g_j(\vec{x}) = |\vec{x}_{j_1} - \vec{x}_{j_2}| - d_j = 0 \quad j = 1, \dots, K \quad (7)$$

Im numerischen Integrationsschema wird LINC6 genau wie SHAKE erst nach einem Integrationsschritt ohne zusätzliche Zwänge angewendet. Der Algorithmus funktioniert in zwei Schritten (s. Abbildung 2). Nachdem die neuen Koordinaten ohne zusätzlichen Zwang berechnet wurden (Abbildung 2 links), wird die Projektion der neuen Bindungsvektoren auf die alten Bindungsvektoren auf Null gesetzt. Der zweite Schritt bezieht sich auf die Bindungslängen-Variationen aufgrund von Rotation (Abbildung 2 rechts). Die Gradientenmatrix der Zwangsgleichungen 7 werden \mathbf{B} genannt. Es ergibt sich Gleichung 8:

$$\vec{B}_{hi} = \frac{\partial g_h}{\partial \vec{x}_i} \quad (i = 1 \dots N, \quad h = 1 \dots K) \quad (8)$$

\mathbf{B} ist eine Matrix der Größe $K \times 3N$ und beschreibt die Richtungen der Zwänge. Gleichung 10 zeigt unter Verwendung der Abkürzung aus Gleichung 9, in welchem Verhältnis die neuen Koordinaten unter Berücksichtigung der Zwänge $\vec{x}_{t+\Delta t}$ zu den freien Koordinaten $\vec{x}_{t+\Delta t}^{unc}$ stehen [49]. \mathbf{M} ist die Massenmatrix und in \vec{d} stehen die Standarddistanzen der Zwänge d_i . $\vec{x}_{t+\Delta t}$ entspricht den Koordinaten in Abbildung 2 Mitte.

$$\mathbf{T} = \mathbf{M}^{-1} \mathbf{B}^T (\mathbf{B} \mathbf{M}^{-1} \mathbf{B}^T)^{-1} \vec{d} \quad (9)$$

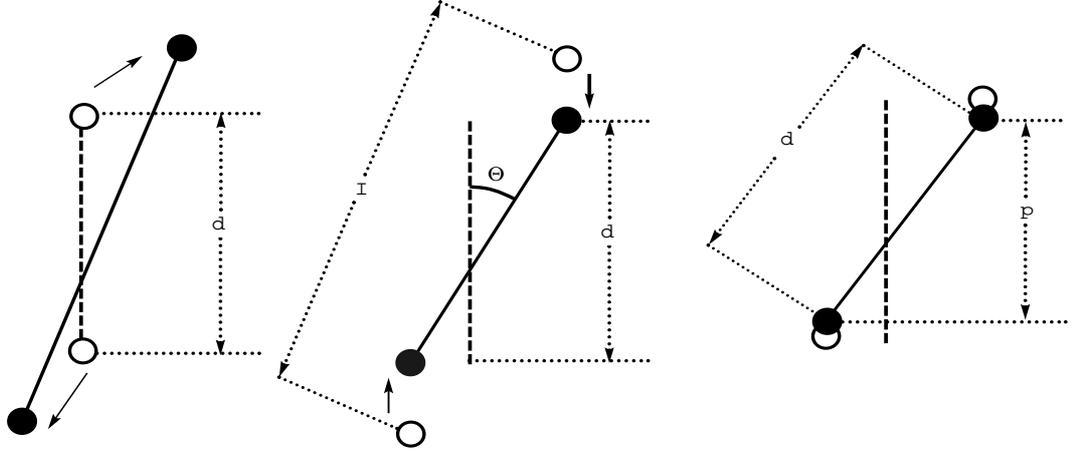


Abbildung 2: Schematische Darstellung des LINCS-Algorithmus. Die Aktualisierung der Ortskoordinaten eines Integrationsschritts teilt sich in drei Schritte auf. Zunächst werden ausgehend von der Ursprungskonformation (links hohle Kugeln) neue Koordinaten ohne die Anwendung eines zusätzlichen Zwangs (links schwarze Kugeln bzw. Mitte hohle Kugeln) berechnet. Die Projektion der neuen Bindungsvektoren auf die alten Bindungsvektoren werden auf Null gesetzt (Mitte schwarze Kugeln bzw. rechts hohle Kugeln). Im letzten Schritt werden die Bindungslängen-Variationen aufgrund von Rotation korrigiert.

$$\begin{aligned}
\overrightarrow{x_{t+\Delta t}} &= (1 - \mathbf{T}_t \cdot \mathbf{B}_t) \cdot \overrightarrow{x_{t+\Delta t}^{unc}} + \mathbf{T}_t \cdot \vec{d} \\
&= \overrightarrow{x_{t+\Delta t}^{unc}} - \mathbf{M}^{-1} \mathbf{B}_t (\mathbf{B}_t \cdot \mathbf{M}^{-1} \cdot \mathbf{B}_t^T)^{-1} (\mathbf{B}_t \cdot \overrightarrow{x_{t+\Delta t}^{unc}} - \vec{d})
\end{aligned} \tag{10}$$

Im zweiten Schritt werden die Bindungslängen-Variationen aufgrund von Rotation korrigiert. Wie aus Abbildung 2 hervorgeht, lassen sich die Längen p_i der Projektionen der korrigierten Bindungsvektoren auf die ursprünglichen Bindungsvektoren mit Hilfe von Gleichung 11 berechnen.

$$p_i = \sqrt{2 \cdot d_i^2 - l_i^2} \quad \text{mit } l_i = d_i \cdot \cos(\Theta) \tag{11}$$

Die endgültigen Koordinaten $\overrightarrow{x_{t+\Delta t}^*}$ werden mit Hilfe von Gleichung 12 bestimmt [49].

$$\overrightarrow{x_{t+\Delta t}^*} = (1 - \mathbf{T}_t \cdot \mathbf{B}_t) \cdot \overrightarrow{x_{t+\Delta t}} + \mathbf{T}_t \cdot \vec{p} \tag{12}$$

2.1.4 Simulationsbedingungen

Im Experiment wie auch *in vivo* findet ein Druck- und Temperatúrausgleich mit der Umgebung statt. In Molekulardynamik-Simulationen werden diese Bedingungen durch Korrekturfaktoren realisiert.

Die Summe aus kinetischer und potentieller Energie eines Simulationssystems würde ohne zusätzliche Temperaturregelung immer konstant bleiben, was einem abgeschlossenen System entsprechen würde. Zur Simulation des Wärmeaustauschs mit der Umgebung wird das System an ein Wärmebad gekoppelt, was durch einen Korrekturfaktor der Geschwindigkeiten implementiert wird [51]. Der Zusammenhang zwischen kinetischer Energie und Temperatur ist über Gleichung 13 definiert [52]. N_{df} bezeichnet die Anzahl der Freiheitsgrade, $E_{kin.}$ ist die kinetische Energie und $v_i(t)$ bezeichnet die Geschwindigkeit des Teilchens i zum Zeitpunkt t .

$$E_{kin.}(t) = \sum_{i=1}^N \frac{1}{2} m_i \cdot v_i^2(t) = \frac{1}{2} N_{df} \cdot k_B \cdot T(t) \quad (13)$$

Der im Rahmen dieser Arbeit verwendete Algorithmus ist eine schwache Kopplung erster Ordnung.

$$\frac{dT}{dt} = \frac{T_0 - T}{\tau} \quad (14)$$

Die Differenz zwischen der Systemtemperatur T und der Umgebungstemperatur T_0 wird nach Gleichung 14 exponentiell mit der Zeitkonstanten τ abgeführt. Die ausgetauschte Wärme wird über einen Skalierungsfaktor λ geregelt, mit dem die Geschwindigkeiten aller Partikel bei jedem Schritt multipliziert werden.

$$\lambda = \sqrt{1 + \frac{\Delta t}{\tau_T} \cdot \left(\frac{T_0}{T(t - \frac{\Delta t}{2})} - 1 \right)} \quad (15)$$

Der Parameter τ_T ist ungefähr aber nicht exakt gleich der Zeitkonstanten τ der Temperaturkopplung aus Gleichung 14.

$$\tau = \frac{2 \cdot C_V \cdot \tau_T}{N_{df} \cdot k_B} \quad (16)$$

C_V ist die Wärmekapazität des Systems. Der Grund, warum $\tau \neq \tau_T$ ist, liegt darin, daß die Änderung der kinetischen Energie durch die Skalierung der Geschwindigkeiten teilweise zwischen kinetischer und potentieller Energie umverteilt wird und dementsprechend die Änderung in der Temperatur kleiner ist, als die Skalierungsenergie [48].

Die Druckkopplung basiert auf dem gleichen Grundgedanken wie die Temperaturkopplung. Das System wird an ein "Druckbad" gekoppelt [51]. Der Druck läßt sich mit Hilfe von Gleichung 17 berechnen, wobei das Virial Ξ Gleichung 18 zu entnehmen ist.

$$P = \frac{2}{3 \cdot V} (E_{kin} - \Xi) \quad (17)$$

$$\Xi = -\frac{1}{2} \sum_{\text{Paare } i,j} (\vec{x}_i - \vec{x}_j) \cdot \vec{F}_{ij} \quad (18)$$

Der Druck wird durch eine Skalierung aller Koordinaten und der Boxgröße mit einem Faktor μ geregelt, welcher eine kinetischen Relaxation erster Ordnung des Drucks P zum Referenzdruck P_0 zur Folge hat.

$$\frac{dP}{dt} = \frac{P - P_0}{\tau_p} \quad (19)$$

Der Skalierungsfaktor ist durch Gleichung 20 gegeben, wobei β die isotherme Kompressibilität beschreibt.

$$\mu = \left[1 + \frac{\Delta t}{\tau_p} \cdot \beta \cdot (P(t) - P_0) \right]^{\frac{1}{3}} \quad (20)$$

2.2 Kollektive Bewegungsmoden

Es wurde in den letzten 15 Jahren gezeigt, daß die Vorbereitung und Durchführung der Berechnungen der „einfache“ Teil eines Simulationsprojektes ist und

daß die Extraktion von biochemisch relevanten Daten aus der Fülle von Informationen den weitaus schwierigeren Teil darstellt [53]. Eine wichtige Analyse-methode ist die Berechnung von Moden korrelierter Bewegung. Grundsätzlich gibt es zwei verschiedene Ansätze, um korrelierte Bewegungen zu ermitteln. Die Moden können aus den atomaren Wechselwirkungen in einer einzelnen experimentellen Struktur oder aus Konformationsensembeln ermittelt werden [54]. Die Analyse von kollektiven Bewegungen kann zur Erforschung der konformationellen Energiehyperfläche benutzt werden, um die Effizienz des Samplings zu verbessern, sowie NMR- und Röntgenstrukturen zu verfeinern [55,56]. Aber auch funktionell relevante Bewegungen in Richtung von wenigen kollektiven Koordinaten, welche dominant zur atomaren Fluktuation beitragen, konnten in einer Reihe von Studien identifiziert werden. Ein Beispiel ist die MD-Simulation an $CRP : (cAMP)_2$ (Komplex des $cAMP$ mit dem $cAMP$ -Rezeptorprotein) [57]. Die Überführung von der geöffneten in die geschlossene Konformation wird allein durch zwei kollektive Koordinaten beschrieben. *Essential dynamics analyses* (EDA) und Normalmodenanalysen (NMA) an $GroEL$ konnten Fluktuationen in der Richtung der konformationellen Änderung zwischen gebundener und freier Form mit Fluktuationen in der Richtung des Bindungsvorgangs korrelieren [58–60]. Funktionell relevante Moden wurden auch für Aspartat- Transcarbamylase [61–63], für α -lytische Protease [64, 65], für das serumretinolbindende Protein [66–68], für HIV-1-Integrase [69] sowie für Neurophysin [70] mit Hilfe von EDA (s. Kapitel 2.2.2) und NMA (s. Kapitel 2.2.1) gefunden. Eine weitere interessante Anwendung von kollektiven Koordinaten sind Untersuchungen von Entfaltungsvorgängen [71–74].

Kollektive Koordinaten können aber auch dazu dienen, einen niedrig dimensionalen Subraum vorherzusagen, in dem signifikante Proteindynamik zu erwarten ist [56]. Die Anzahl der Freiheitsgrade bei der Definition dieses Raumes kann auf etwa 5% der Anzahl der entsprechenden kartesischen Koordinaten reduziert werden. Die atomare Fluktuation innerhalb eines solchen Subraumes macht den Hauptanteil an der gesamten Fluktuation aus [56]. Ein Raum, der durch wenige

niedrig-frequente Normalmoden aufgespannt wird, wurde „wichtiger Subraum“ („important subspace“) genannt [75]. Das Konzept wurde zunächst dazu benutzt, um Kristallstrukturen zu verfeinern [76–78]. Später haben Amadei et al. den Raum als „essentiellen Raum“ (Essential subspace) bezeichnet und dem Verfahren zur Berechnung des Raumes ausgehend von strukturellen Ensembles den Namen „Analyse der essentiellen Dynamik“ („Essential Dynamics Analysis“, EDA) gegeben. Da sich dieser Begriff im allgemeinen Sprachgebrauch durchgesetzt hat, wird er auch im Rahmen dieser Arbeit verwendet.

In manchen Fällen hängt eine kleine Anzahl von Moden mit großer Amplitude von der Länge der zu Grunde liegenden Simulation ab [79]. Außerdem wurde gezeigt, daß sehr langsame Bewegungen vom Kraftfeld abhängen [80]. Dennoch sind die Frequenzverteilungen und Fluktuationen bemerkenswert ähnlich [56]. Außerdem ist der essentielle Subraum in vielen Fällen nahezu unabhängig von der Wahl der beschreibenden Koordinaten bei einer NMA [81, 82], sowie dem Einfluß des Solvens [83, 84] und den Simulationsbedingungen zur Generierung eines strukturellen Ensembles für eine EDA [85, 86].

2.2.1 Normalmodenanalysen

Die Methode der Normalmodenanalyse (NMA) basiert auf der Annahme, daß im Bereich der thermischen Fluktuation die Hyperfläche der Konformationsenergie durch eine quadratische Näherung an einem einzelnen energetischen Minimum charakterisiert werden kann. Anders ausgedrückt: Man schreibt die Potentialfunktion $V(\mathbf{x})$ um eine Referenzstruktur \vec{x}_0 als Taylor’sche Reihenentwicklung. Befindet sich die Referenzstruktur in einem energetischen Minimum, dann ist der Gradient von V an dieser Stelle gleich dem Nullvektor. Wenn man nun vollständige Harmonizität voraussetzt, also die Ableitungen dritter und höherer Ordnung vernachlässigt, dann ist direkt zu zeigen, daß das System durch Richtungsvektoren von Normalmoden \vec{v}_i und Frequenzen ω_i entsprechend Gleichung 21 beschrieben werden kann [53]. Die Matrix \mathbf{M} enthält die atomaren Massen auf ihrer Diagonalen. Die Hess’sche Matrix \mathbf{F} enthält die zweiten Ableitungen

der potentiellen Energie an der Stelle \vec{x}_0 .

$$\mathbf{M}^{-1/2} \cdot \mathbf{F} \cdot \mathbf{M}^{-1/2} \cdot \vec{v}_i = \omega_i^2 \cdot \vec{v}_i \quad (21)$$

Eine direkte Berechnung von Normalmoden eines Moleküls mit N Atomen enthält also die numerische Diagonalisierung einer Matrix $\mathbf{M}^{-1/2} \cdot \mathbf{F} \cdot \mathbf{M}^{-1/2}$ der Größe $3N \times 3N$. Seit etwa 15 Jahren ist es möglich, solche Berechnungen an kleinen Proteinen durchzuführen [87–89].

Um größere Proteine analysieren zu können, wurden auch Diederwinkel als alternative Variablen für die Berechnung der Hess'schen Matrix benutzt, denn die Anzahl der Diederwinkel, die die Konformation eines Proteins beschreibt, beträgt ca. nur ein achtel von der Anzahl der entsprechenden kartesischen Koordinaten [55, 88]. Die Diederwinkel-NMAs wurden auf eine Reihe von Proteinen wie BPTI [88–91], Crambin [89], Lysozym [89, 92], Myoglobin [93, 94], epidermaler Wachstumsfaktor von Mäusen [95] und G-Aktin [96] angewendet. Der Vergleich der Ergebnisse aus dem kartesischen Raum mit den Ergebnissen aus dem Diederwinkelraum hat gezeigt, daß zwar die dominanten Bewegungen mit niedriger Frequenz nahezu vollständig mit der Diederwinkelanalyse erfaßt werden, daß die Frequenzen im Diederwinkelraum aber ca. 10 bis 20 % höher liegen als im kartesischen Raum [81, 82]. Es scheint so, daß das Zulassen von Fluktuationen in den Bindungslängen und den Bindungswinkeln den indirekten Effekt hat, daß die Diederwinkel flexibler werden [55].

Eine andere Methode, um den Rechenaufwand zu reduzieren, liegt darin, nicht die gesamte Hess'sche Matrix für die Diagonalisierung zu verwenden, indem die zweiten Ableitungen zwischen weit entfernten Atomen gleich Null gesetzt werden. Die Reduktion des Rechenaufwands basiert auf einer Unterteilung des Moleküls in drei Teile A , B und C , wobei A und C durch B voneinander räumlich getrennt werden, sodaß ihre Wechselwirkungen vernachlässigt werden können. In diesem Fall berechnet sich die Hess'sche Matrix nach Gleichung 22.

$$\mathbf{H} = \begin{pmatrix} H_{AA} & H_{AB} & 0 \\ H_{BA} & H_{BB} & H_{BC} \\ 0 & H_{CB} & H_{CC} \end{pmatrix} \quad (22)$$

Wechselwirkungen zwischen B und A und zwischen B und C werden durch die statischen Bedingungen in Gleichung 23 angenähert.

$$H_{AA} \cdot \nu_A + H_{AB} \cdot \nu_B = 0 = H_{CC} \cdot \nu_C + H_{BC} \cdot \nu_B \quad (23)$$

Dabei sind ν_A und ν_B die Spaltenmatrizen der atomaren Verschiebungen. Das System läßt sich nach ν_B auflösen. So ist es möglich, die Bewegungsgleichung in Termen der Fluktuation allein für die Atome in *B* aufzuschreiben. Dadurch ist die Diagonalisierung für jeden Teil des Systems separat durchzuführen. Die Verschiebungen der Atome in *A* und *C* werden als Linearkombination ihrer eigenen Normalmoden und der Moden aus der statischen Korrelationsanalyse mit *B* ausgedrückt. Man kann dann die Matrix aus Gleichung 22 in den reduzierten Raum der Bewegungen mit niedrigen Frequenzen transformieren und letztlich die Diagonalisierung durchführen, um angenäherte Normalmoden und ihre Frequenzen zu erhalten [55]. Die Methode führt für lineare Moleküle zu recht exakten Ergebnissen mit Fehlern um 1 bis 2% in den Eigenwerten relativ zur Berechnung der vollen Hess'schen Matrix [97, 98].

Mouawad und Perahia haben vorgeschlagen, Normalmodenanalysen anzunähern, indem Matrizen, die zu diagonalen Blöcken der Hess'schen Matrix gehören, diagonalisiert werden [61]. Dazu wird das Molekül zunächst gedanklich zerteilt und die diagonalen Blöcke der Hess'schen Matrix, die zu den Teilen gehören, werden diagonalisiert. Die Hess'sche Matrix wird daraufhin in dem Subraum, der sich auf die niedrigen Frequenzen der Proteinteile bezieht, erneut diagonalisiert.

Es existieren noch einige andere Verfahren, um schneller oder mit weniger Rechenspeicher Bewegungsmoden aus Hess'schen Matrizen zu berechnen [87, 99–102]. Allen Methoden ist gemeinsam, daß durch die Vereinfachungen

insbesondere Korrelationen zwischen weit entfernten Atomen verschwinden können.

Aber auch wenn die NMA exakt durchgeführt wird, wird eine weitreichende Annahme eingeführt: Man geht davon aus, daß im Bereich der thermischen Fluktuation die Hyperfläche der Konformationsenergie durch eine quadratische Näherung an ein einzelnes energetisches Minimum charakterisiert werden kann. Diese Annahme ist für physiologische Temperaturen falsch, denn es werden zwei wichtige Effekte vernachlässigt [54]. Zum einen wird der Einfluß des Lösungsmittels bei den Rechnungen außer acht gelassen, auch wenn in manchen Fällen einige gebundene Wassermoleküle berücksichtigt werden. Zum anderen wird auch ein einzelnes parabolisches Potential approximiert, wenn die Konformationshyperfläche mehrere Minima aufweist. Dennoch stimmen die mit Hilfe von NMA bestimmten unabhängigen harmonischen Moden von Molekülen relativ gut mit solchen Moden überein, bei denen das Lösungsmittel berücksichtigt wurde und keine Harmonizität vorausgesetzt werden muß [67,103] (s. 2.2.2).

2.2.2 Dynamik aus Strukturensembles

Bereits 1982 haben Studien von Molekulardynamik-Simulationen den Nachweis für kollektive Bewegungen geliefert. Die Analyse von zeitlichen Korrelationsfunktionen von individuellen Atomen in BPTI haben gezeigt, daß atomare Fluktuation als Überlagerung von lokalen Oszillationen und kollektiven Bewegungen auf langen Zeitskalen betrachtet werden können [104]. Eine Kovarianzanalyse der atomaren Fluktuation in BPTI hat außerdem gezeigt, daß sich benachbarte Atome insbesondere innerhalb von Sekundärstrukturelementen kollektiv bewegen [37]. Schon bei den frühen Studien wurde deutlich, daß Methoden zur Ermittlung unkorrelierter generalisierter Koordinaten von großem Wert sind. Die Normalmodenanalyse ist eine Möglichkeit, solche Koordinaten zu berechnen. Sie hat aber den Nachteil, daß für diese Berechnungen die Form der Energiehyperflächen in Abhängigkeit von den Konformationen als rein harmonisch angenommen werden muß. Doch durch die Projektion von Molekulardynamik-

oder Monte Carlo-Trajektorien [105–108] auf Normalmoden ist man in der Lage, die Dynamik ohne eine Annahme von Harmonizität zu analysieren.

Ein kollektiver Koordinatensatz kann als Satz von Eigenvektoren der Kovarianzmatrix aufgefaßt werden. Dieses Verfahren wird im allgemeinen Hauptkomponentenanalyse (PCA = **P**rocedure **C**omponent **A**nalysis) genannt [56,81]. Ein Element der Kovarianzmatrix der die Struktur beschreibenden - üblicherweise kartesischen - Koordinaten berechnet sich nach Gleichung 24

$$c_{jk} = \langle (x_j - \langle x_j \rangle) \cdot (x_k - \langle x_k \rangle) \rangle \quad (24)$$

wobei die Symbole $\langle \dots \rangle$ einen Mittelwert über das strukturelle Ensemble darstellen und x_j, x_k die massengewichteten Koordinaten sind. Gleichung 25 ist die zu Gleichung 24 äquivalente Form in Matrixschreibweise.

$$\mathbf{C} = cov(\mathbf{x}) = \langle (\mathbf{x} - \langle \mathbf{x} \rangle) \cdot (\mathbf{x} - \langle \mathbf{x} \rangle)^T \rangle \quad (25)$$

Die Matrix \mathbf{x} enthält in jeder Spalte die Koordinaten genau einer Struktur und die Werte einer bestimmten Koordinate in einer Zeile. \mathbf{C} ist die Kovarianzmatrix und enthält in der Spalte j und Zeile k die Kovarianz der beiden Koordinaten j und k innerhalb des strukturellen Ensembles. Die Diagonalelemente von \mathbf{C} enthalten die jeweiligen Varianzen der kartesischen Koordinaten.

Die Kovarianzmatrix ist symmetrisch und kann daher immer durch eine orthogonale Koordinatentransformation \mathbf{T} zu einer diagonalen Matrix Λ mit Eigenwerten λ_i diagonalisiert werden [40].

$$\mathbf{C} = \mathbf{T} \cdot \Lambda \cdot \mathbf{T}^T \quad (26)$$

Im transformierten Raum werden die Strukturen in generalisierten Koordinaten q beschrieben. Im Gegensatz zu den ursprünglichen Koordinaten in \mathbf{x} sind die generalisierten Koordinaten in \mathbf{q} unkorreliert.

$$\mathbf{q} = \mathbf{T}^T (\mathbf{x} - \langle \mathbf{x} \rangle) \quad (27)$$

Dementsprechend enthält die Kovarianzmatrix im transformierten Raum Λ ausschließlich Nullen für Elemente außerhalb der Diagonalen.

$$\Lambda = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \lambda_{3N} \end{pmatrix} \quad (28)$$

Die Koordinatentransformation \mathbf{T} besteht aus orthonormalen Spaltenvektoren, die die Bewegungsmoden darstellen. Jeder Mode in \mathbf{T} ist genau eine Varianz λ_i in Λ zugeordnet. λ_i gibt die mittlere quadratische Auslenkung der Mode innerhalb des strukturellen Ensembles an. Im Rahmen einer EDA werden die Spalten in \mathbf{T} so geordnet, daß die Varianzen mit dem Index der Moden abnehmen.

$$\lambda_1 > \lambda_2 > \lambda_3 > \cdots > \lambda_{3N} \quad (29)$$

2.2.3 Zusammenhang zwischen NMA und EDA

Für eine Analyse der essentiellen Dynamik wird die Kovarianzmatrix \mathbf{C} diagonalisiert, während bei Normalmodenanalysen von der Hess'schen Matrix ausgegangen wird. Das Verhältnis zwischen diesen beiden Größen wird aus Gleichung 30 deutlich [56]. k_B ist die Boltzmann-Konstante, T die absolute Temperatur und \mathbf{H} die Hess'sche Matrix.

$$\mathbf{C} = k_B \cdot T \cdot \mathbf{H}^{-1} \quad (30)$$

Die inverse einer Matrix wird üblicherweise durch Auflösung von Gleichung 31 nach \mathbf{H}^{-1} berechnet. Gleichung 31 führt zu $3N$ Gleichungssystemen mit $3N$ Variablen, die genau dann lösbar sind, wenn \mathbf{H} den vollen Rang hat.

$$\mathbf{H} \cdot \mathbf{H}^{-1} = 1 \quad (31)$$

Weil aber bei der NMA keinerlei Rotations- und Translationsfreiheitsgrade berücksichtigt werden, hat \mathbf{H} nur den Rang von $3N - 6$. Deswegen muß die inverse Matrix von \mathbf{H} über eine Singulärwertzerlegung (SVD = Singular Value Decomposition) ermittelt werden.

\mathbf{H} ist symmetrisch und kann daher immer durch eine orthogonale Koordinatentransformation \mathbf{U} zu einer diagonalen Matrix \mathbf{S} mit Eigenwerten s_i diagonalisiert werden.

$$\mathbf{H} = \mathbf{U} \cdot \mathbf{S} \cdot \mathbf{U}^T \quad (32)$$

$$\mathbf{S} = \begin{pmatrix} s_1 & 0 & \cdots & 0 \\ 0 & s_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & s_{3N} \end{pmatrix} \quad (33)$$

Weil \mathbf{H} den Rang $3N - 6$ hat, sind die sechs kleinsten Eigenwerte s_{3N-5} bis s_{3N} im Rahmen der Genauigkeit gleich Null. Deswegen wird in Gleichung 34 eine reduzierte Singulärwertmatrix \mathbf{S}_{red} definiert.

$$\mathbf{S}_{\text{red}} = \begin{pmatrix} s_1 & 0 & \cdots & 0 \\ 0 & s_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & s_{3N-6} \end{pmatrix} \quad (34)$$

In der reduzierten Form der Matrix \mathbf{U} werden dementsprechend gemäß Gleichungen 35 die letzten sechs Spalten vernachlässigt.

$$\mathbf{U}_{\text{red}} = (\vec{u}_1 \vec{u}_2 \dots \vec{u}_{3n-6}) \quad (35)$$

Die inverse Matrix von \mathbf{S}_{red} läßt sich nach Gleichung 36 berechnen, denn alle Eigenwerte s_i in \mathbf{S}_{red} sind ungleich Null.

$$\mathbf{S}_{\text{red}}^{-1} = \begin{pmatrix} s_1^{-1} & 0 & \cdots & 0 \\ 0 & s_2^{-1} & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & s_{3N-6}^{-1} \end{pmatrix} \quad (36)$$

\mathbf{U}_{red} und $\mathbf{U}_{\text{red}}^T$ sind inverse Matrizen, denn \mathbf{U}_{red} besteht aus orthonormalen Spaltenvektoren. Damit ergibt sich Gleichung 37 für die Inverse Hess'sche Matrix \mathbf{H}^{-1} .

$$\begin{aligned} \mathbf{H}^{-1} &= (\mathbf{U}_{\text{red}} \cdot \mathbf{S}_{\text{red}} \cdot \mathbf{U}_{\text{red}}^T)^{-1} \\ &= (\mathbf{U}_{\text{red}}^T)^{-1} \cdot (\mathbf{U}_{\text{red}} \cdot \mathbf{S}_{\text{red}})^{-1} \\ &= (\mathbf{U}_{\text{red}}^T)^{-1} \cdot \mathbf{S}_{\text{red}}^{-1} \cdot (\mathbf{U}_{\text{red}})^{-1} \\ &= \mathbf{U}_{\text{red}} \cdot \mathbf{S}_{\text{red}}^{-1} \cdot \mathbf{U}_{\text{red}}^T \end{aligned} \quad (37)$$

In der Praxis ist sowohl in der EDA als auch in der NMA der Schritt mit dem größten Rechenaufwand die Diagonalisierung der Kovarianzmatrix bzw. Hess'schen Matrix. In der EDA kann der Rechenaufwand dadurch verringert werden, daß nur C_α -Atome berücksichtigt werden. Für die NMA wurden vereinfachte Modelle vorgeschlagen, die mit Hilfe von vereinfachten Kraftfeldern einige Moden mit großer Amplitude voraussagen können [109, 110].

2.2.4 Projektion von Proteinbewegung auf kollektive Koordinaten

Einer der ersten Ansätze, MD- und MC- Simulationen auf kollektive Koordinaten zu projizieren, wurde bereits 1991 an humanem Lysozym durchgeführt [111]. Die Projektion auf den Normalmoden-Raum ist jedoch keine sonderlich anschauliche Methode zur Beschreibung von Zustandsübergängen [112]. Dennoch können Zustandsänderungen beschrieben werden, indem MD-Trajektorien z.B. auf die Ebene, die durch zwei kollektive Koordinaten aufgespannt werden, projiziert werden [81]. Mittlerweile ist die EDA eine allgemein gebräuchliche Methode, um

wichtige Bewegungen in Biomolekülen zu beschreiben [109, 110, 113–119].

Bei EDAs sind Lösungsmittelleffekte auf zweierlei Weise von Bedeutung. Zum einen werden Konformationsübergänge kleiner Amplitude in Wasser beobachtet, die *in vacuo* nicht zu beobachten sind [81, 83]. Diese Übergänge gehen mit Änderungen in der Hydratationsstruktur einher [120], weil eine Feinstruktur der freien Energieoberfläche im Wasser existiert, die im Vakuum nicht berücksichtigt wird [121]. Durch erniedrigte Energiebarrieren für große konformationelle Übergänge werden globale Übergänge häufiger in voll solvatisierten als in nur teilweise solvatisierten Modellen beobachtet [56, 122]. Der zweite Lösungsmittelleffekt besteht in einer überaus starken Dämpfung der zeitlichen Korrelation von Moden mit großer Amplitude [83, 84]. Im Fall einer harmonischen Energieoberfläche kann dieses Problem elegant in NMAs berücksichtigt werden, während die Anwendung auf Langevin-Dynamik-Simulationen noch nicht gelöst ist [56, 123, 124].

2.2.5 Energiehyperfläche im Raum der kollektiven Koordinaten

Nach den Pionierarbeiten an Myoglobin [125] und BPTI [126–130] werden mittlerweile für die meisten Ansätze zur Untersuchung der Energiehyperfläche kollektive Koordinaten verwendet. Ausgehend von nicht-Gaußverteilten Projektionen von MD-Trajektorien auf kollektive Koordinaten konnten nicht-lineare Bewegungen mit großer Amplitude identifiziert werden [40, 114]. Weitere Arbeiten zum Verständnis des anharmonischen Aspektes der Bewegung basieren auf NMAs [112, 131]. In den Arbeiten wurde ein sogenannter “Anharmonizitätsfaktor” eingeführt, der eine klare Trennung zwischen harmonischen und anharmonischen Moden zuläßt. Es hat sich herausgestellt, daß nur ca. 7 bis 12 % aller Moden einem anharmonischen Potential zugeordnet sind [56, 86, 132, 133].

Das Minima-Sprung (JAM = **J**umping **a**long **m**inima) Modell beruht auf einer Aufteilung der Kovarianzmatrix aus Gleichung 24 in zwei Summanden gemäß Gleichung 38.

$$\begin{aligned}
c_{ij} = & \sum_{k=1}^{N_M} f_k \cdot (\langle x_i \rangle_k - \langle x_i \rangle) \cdot (\langle x_j \rangle_k - \langle x_j \rangle) \\
& + \sum_{k=1}^{N_M} f_k \cdot (x_i - \langle x_i \rangle_k) \cdot (x_j - \langle x_j \rangle_k)_k
\end{aligned} \tag{38}$$

N_M ist die Anzahl der konformationellen Zustände. k ist der Zustandsindex. $\langle \dots \rangle_k$ ist der Mittelwert über die Zeitperiode, während der sich das System im k -ten Zustand befunden hat, und f_k ist das Verhältnis der Länge dieser Periode relativ zur gesamten Simulationszeit. Die erste Zeile von Gleichung 38 repräsentiert die atomare Fluktuation hervorgerufen durch einen Wechsel des Zustandes. Die zweite Zeile von Gleichung 38 beschreibt die atomare Fluktuation innerhalb der Zustände. Die Anwendung des Modells hat gezeigt, daß jeder konformationelle Subzustand ein nahezu harmonisches Potential aufweist. Mit Hilfe der Anharmonizitätsfaktoren hat die Analyse der Wahrscheinlichkeitsdichtefunktionen ergeben, daß zwischen drei verschiedenen Typen von Moden unterschieden werden kann: Multihirarchische Moden, hirarchische Moden und harmonische Moden. In Tabelle 1 sind die Eigenschaften der Gruppen von Moden zusammengefaßt. Die Beschreibung der Proteindynamik ist nicht an den nativen Zustand gebunden, denn selbst in einem erweiterten Konformationsraum trägt immer noch eine kleine Anzahl von Moden dominierend zu der gesamten Fluktuation bei [56]. Die Vermutung liegt also nahe, daß es möglich sein müßte, mit Hilfe von Hauptkomponentenanalysen auf Faltungssimulationen die Hypothesen der Faltungsmechanismen zu untersuchen [56].

2.2.6 „Sampling“ Methoden

Ein grundsätzliches Problem von MD und MC Rechnungen ist die Begrenzung der Rechenzeit. Trotz ständig steigender Rechengeschwindigkeit ist die bisher längste MD-Simulation, die berechnet wurde “nur” eine Mikrosekunde lang [1]. Für eine solche Simulation wurde ein halbes Jahr Rechenzeit an verschiedenen Großrechnern benötigt. Weil die Kosten für eine solche Rechnung enorm sind, wird versucht, durch zusätzliche Annahmen das Abtasten des Konformations-

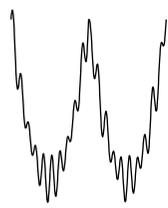
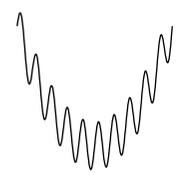
	Multihirarchische Mode	Hirarchische Mode	Harmonische Mode
Energieoberfläche			
Wahrscheinlichkeitsdichte	nicht Gauß-verteilt, mehrere Minima	nahezu Gauß-verteilt, ein Minimum	Gauß-verteilt ein Minimum
Anharmonizitätsfaktor ρ_n	>2	1-2	1
Packungstopologie	signifikante Änderung	keine Änderung	keine Änderung
Fluktuationen in den Diederwinkeln	große Veränderungen in Rückgrat und Seitenketten	große Veränderungen in den Seitenketten	keine Veränderungen
Anteil der Moden relativ zur Anzahl aller Moden	ca. 0.5-1 %	ca. 4-5 %	ca. 95 %
Anteil an der mittleren quadratischen Fluktuation	75-85 %	0-20 %	<5%

Tabelle 1: Eigenschaften von Moden, die mit EDA und JAM ermittelt wurden [56, 133]. Der Anharmonizitätsfaktor ρ_n ist definiert als $\rho_n = \sqrt{\zeta_n / \zeta_n^{har}}$, wobei ζ_n^{har} die erwartete mittlere quadratische Fluktuation aus der NMA ist.

raums in den Simulationen zu beschleunigen.

Es können zusätzliche Potentialterme aufgenommen werden, die auf realen Messwerten an dem System beruhen, wie etwa zeitgemittelte geometrische Randbedingungen aus NMR-Experimenten [134–136]. Auch die Durchführung von mehreren parallelen Simulationen ausgehend von derselben Anfangskonformation ist effizienter als eine lange Simulation zu berechnen [137]. Es existieren noch andere Ansätze zum effektiven Sampling wie etwa multikanonische MC- [138, 139] und MD-Simulationen [140, 141] oder MD mit variablen Zeitschritten [142].

Ein effizientes Abtasten des essentiellen Konformationsraums mit Hilfe von kollektiven Koordinaten ist von besonderem Interesse. Denn die kollektiven Koordinaten mit der größten Amplitude gehen mit keinem bzw. nur mit einem geringen Anstieg der freien Energie einher [40]. So läßt sich eine große Varianz in den Konformeren erzielen, ohne den durch die Energie begrenzten Bereich der thermischen Fluktuation zu verlassen.

Der von Amadei et al. vorgeschlagene Algorithmus basiert auf einem Zusatzpotential, welches das System zwingt, den Abstand in kollektiven Koordinaten zur Referenzstruktur zu vergrößern. Der Gradient an dieses Zusatzpotential zeigt also im transformierten Koordinatenraum von der Referenzstruktur auf die aktuelle Struktur der Trajektorie [143].

Später wurde dieser Algorithmus modifiziert. Nach dem modifizierten Algorithmus wird an Stelle des Zusatzpotentials ein Akzeptanzkriterium eingeführt. Nach Berechnung jedes MD-Schrittes wird die neue Konformation nur dann unverändert akzeptiert, wenn sie im kollektiven Koordinatenraum einen größeren Abstand zur Referenzstruktur hat als die vorherige Struktur. Sollte die Simulation sich wieder in Richtung der Referenzstruktur entwickeln, die neue Struktur also einen geringeren Abstand zur Referenzstruktur aufweisen als ihr Vorgänger, dann wird die Konformation auf die Hyperkugel mit dem vorherigen Radius projiziert. Der kollektive Koordinatenvektor wird also so verlängert, daß er wieder die gleiche Länge aufweist wie in der vorherigen Struktur [144].

Ein eleganter Ansatz zur Regelung der Diversität des strukturellen Ensembles einer Simulation sind die Hauptkomponenten-Zwänge (PCR=**P**roincipal **C**omponent **R**estraint). Nach dieser Methode wird die Simulation eines Ensembles von gleichzeitigen Trajektorien mit Zwängen kombiniert, die auf das Ensemble der Strukturen im Ganzen wirken. Dabei kann die Varianz des Ensembles und/oder die mittlere Position des Ensembles geregelt werden [145].

2.2.7 Absolute Entropien

Eine wichtige Anwendung von kartesischen Kovarianzmatrizen ist die Berechnung von Konformationsentropien [89, 146, 147]. Für ein kanonisches Ensemble legt die thermodynamische Bedeutung der Entropie als Maß für das zugängliche Volumens im Phasenraum den Ansatz nahe, die Entropie aus Mittelwerten der Kovarianzmatrix \mathbf{C} zu berechnen [148]. Karplus und Kushik haben vorgeschlagen, die Entropiedifferenzen zweier Konformationen A und B mittels Gleichung 39 zu berechnen, wobei $\mathbf{C}_{A/B}$ die Kovarianzmatrizen eines Satzes von **internen** Variablen sind.

$$\Delta S = \frac{1}{2} k_B \cdot \ln \frac{\det(\mathbf{C}_B)}{\det(\mathbf{C}_A)} \quad (39)$$

Diese Formulierung ist in der Praxis jedoch nicht sonderlich hilfreich, denn es muß eine Koordinatentransformation durchgeführt werden, wenn interne Koordinaten ausgehend von kartesischen Koordinaten aus Molekulardynamik-Simulationen berechnet werden sollen [148]. Die ursprüngliche Kovarianzmatrix in kartesischen Koordinaten ist singulär, denn der Rang dieser Matrizen wird durch die zuvor eliminierten drei möglichen Rotationsbewegungen und die drei möglichen Translationsbewegungen um insgesamt 6 Dimensionen verringert. Die Determinante von Matrizen mit einem nicht vollem Rang ist Null, wodurch Gleichung 39 nicht anwendbar ist [148]. Schlitter hat eine sehr elegante Methode entwickelt, dieses Problem zu umgehen [148–150]. Die Formel basiert auf einer quantenmechanischen Behandlung von eindimensionalen Freiheitsgraden x mit

Zuständen $|n\rangle$ mit $n \in \{0, 1, 2, \dots\}$. Die Energie eines jeden Zustandes wird E_n genannt, die Masse von x sei M und der Mittelwert $\langle x \rangle$ sei gleich Null. Die kanonische Zustandssumme Z berechnet sich bei einer Temperatur T mit Hilfe von Gleichung 40.

$$Z = \sum_n e^{-\frac{E_n}{k_B \cdot T}} \quad (40)$$

Die Wahrscheinlichkeit p_n , das System im Zustand n anzutreffen, wird durch Gleichung 41 wiedergegeben.

$$p_n = \frac{e^{-\frac{E_n}{k_B \cdot T}}}{Z} \quad (41)$$

Gleichung 42 beschreibt die gesamte Entropie des Systems.

$$S = -k_B \cdot \sum_n p_n \cdot \ln(p_n) \quad (42)$$

Für einen gegebenen Erwartungswert der Varianz $\langle x^2 \rangle = \sum_n p_n \langle n | x^2 | n \rangle$ wird die Entropie durch Variation der Wahrscheinlichkeiten p_n maximiert. Die Varianz $\langle x^2 \rangle$ und die Normierungsbedingung der Wahrscheinlichkeiten $\sum_n p_n = 1$ gehen als Lagrange-Multiplikatoren in die Rechnung ein. Das Ergebnis ist die Bedingung, daß die Energie eines Zustandes proportional zur Varianz des Zustands sein muß. Das System, für das diese Bedingung gilt, ist der einfache harmonische Oszillator (eho). Die Entropie S_{eho} eines eindimensionalen einfachen harmonischen Oszillators wird durch Gleichung 43 ausgedrückt.

$$S \leq S_{eho} = \frac{k_B \cdot \alpha}{e^\alpha - 1} - k_B \cdot \ln(1 - e^{-\alpha}) \quad (43)$$

$$\alpha = \frac{\hbar \cdot \omega}{k_B \cdot T}$$

Wie von Schlitter im Detail hergeleitet wurde [148] ist S_{eho} eine obere Grenze für die tatsächliche Entropie S des Systems. Die Frequenz ω des Oszillators hängt von der quantenmechanisch definierten Varianz $\langle x^2 \rangle$ ab, die durch das Gleichverteilungstheorem aus Gleichung 44 mit der aus klassischen Simulationen

berechenbaren klassischen Varianz in Verbindung steht.

$$m \cdot \omega^2 \cdot \langle x^2 \rangle_c = k_B \cdot T \quad (44)$$

Gleichung 44 gilt nur, wenn $\hbar \cdot \omega \ll k_B \cdot T$ gilt. Diese Näherung ist gut erfüllt, denn die hochfrequenten Bewegungen, für die die Bedingung nicht gilt, tragen nur einen sehr geringen Teil zur gesamten Entropie bei [150].

Für die Verallgemeinerung von Gleichung 43 für mehrere Freiheitsgrade wird in Gleichung 45 eine einfacher zu handhabende Näherung eingeführt. Die Größe S' ist ein rein heuristischer Term, der lediglich noch ein wenig höher liegt als S_{eho} , wodurch S' natürlich auch eine obere Grenze der tatsächlichen Entropie S darstellt.

$$\begin{aligned} S \leq S_{eho} < S' &= \frac{1}{2} k_B \cdot \ln \left(1 + \frac{e^2}{\alpha^2} \right) \\ &= \frac{1}{2} k_B \cdot \ln \left(1 + \frac{k_B \cdot T \cdot e^2}{\hbar^2} \cdot m \cdot \langle x^2 \rangle_c \right) \end{aligned} \quad (45)$$

Die nach den Gleichungen 26 und 27 transformierten Koordinaten q sind unkorreliert. Sie beschreiben also voneinander unabhängige Oszillatoren und die gesamte Entropie des Systems setzt sich daher gemäß Gleichung 46 aus der Summe der einzelnen Entropien der Oszillatoren zusammen. Die Varianz der einzelnen Oszillatoren $\langle q_{ii}^2 \rangle$ ist durch die Eigenwerte λ_i aus den Gleichungen 28 bzw. 26 gegeben.

$$\begin{aligned} S < S' &= \frac{1}{2} k_B \cdot \sum_{i=1}^{3N} \ln \left[1 + \frac{k_B \cdot T \cdot e^2}{\hbar^2} \cdot \langle q_{ii}^2 \rangle \right] \\ &= \frac{1}{2} k_B \cdot \ln \left(\prod_{i=1}^{3N} \left[1 + \frac{k_B \cdot T \cdot e^2}{\hbar^2} \cdot \lambda_i \right] \right) \end{aligned} \quad (46)$$

Das Produkt aus Gleichung 46 kann auch als Berechnung der Determinante einer entsprechenden Diagonalmatrix interpretiert werden. Λ und \mathbf{C} lassen sich gemäß Gleichung 27 über eine orthogonale Koordinatentransformation ineinander überführen. Weil Determinanten invariant mit orthogonalen Koordinatentransformationen sind, folgt Gleichung 47 aus Gleichung 46, wenn für die Berechnung von \mathbf{C} die Massengewichtung berücksichtigt wurde [148].

$$S' = \frac{1}{2} k_B \cdot \ln \left[\det \left(1 + \frac{k_B \cdot T \cdot e^2}{\hbar^2} \cdot \mathbf{C} \right) \right] \quad (47)$$

2.3 Berechnung von Ordnungsparametern

Zur Verifizierung von Molekulardynamik-Simulationen werden im Rahmen dieser Arbeit Ordnungsparameter S^2 aus der Trajektorie bestimmt und mit experimentellen Werten aus NMR-Relaxationsexperimenten verglichen.

Sowohl die Dipol-Dipol-Wechselwirkung zwischen ^{15}N - und ^1H -Kernen als auch die Anisotropie der chemischen Verschiebung der ^{15}N -Kerne haben einen dominanten Einfluß auf die ^{15}N -Spinrelaxation von Proteinen in Lösung [34]. Weil die Winkelabhängigkeit dieser Wechselwirkungen mit sphärischen Tensoren des Ranges Zwei beschrieben werden kann [151], werden die für die NMR-Relaxation relevanten Korrelationsfunktionen $C(t)$ durch Gleichung 48 beschrieben [34].

$$C(t) = \langle P_2(\mu_{t_0} \cdot \mu_{t_0+t}) \rangle = \frac{4\pi}{5} \cdot \sum_{|M| \leq 2} \langle Y_{2M}^*(\Omega_{t_0+t}) \cdot Y_{2M}^*(\Omega_{t_0}) \rangle \quad (48)$$

Die Kugelflächenfunktionen werden durch $Y_{2M}(\Omega)$ symbolisiert, $\Omega = \{\theta, \varphi\}$ ist ein Satz sphärischer Polarwinkel, der die Orientierung des $\text{N} - \text{H}$ - bzw. $\text{C}_\alpha - \text{H}$ - Bindungsvektors beschreibt. $P_2(x)$ ist das zweite Legendre-Polynom und μ ist der Einheitsvektor entlang der Bindungsachse.

Die Überführung von NMR-Relaxationsdaten in mikrodynamische Parameter hängt von der Wahl eines geeigneten Modells ab. Der “modellfreie Ansatz” nach Lipari und Szabo [152, 153] ist weit verbreitet, um eine Analyse durchzuführen, die unabhängig von Details des Bewegungsprozesses ist. Der modellfreie Ansatz beschreibt die Korrelationsfunktion der Bewegung eines Bindungsvektors mit Hilfe eines generalisierten Ordnungsparameters S^2 und einer Korrelationszeit τ . S^2 gibt den Wert der Korrelationsfunktion zum Zeitpunkt $t = \infty$ wieder, während τ ein Maß für die Geschwindigkeit der Bewegung ist.

Im Rahmen dieser Arbeit werden S^2 -Werte aus Trajektorien berechnet. Dazu wird ein modifizierter modellfreier Ansatz verwandt [154, 155]. Um eine schnelle Abnahme von $C(t)$ zu erlauben, wird die Korrelationsfunktion der Bindungsreorientierung durch Gleichung 49 eingeführt.

$$C(t) = S^2 + (C_1 - S^2) \cdot e^{-\frac{t}{\tau}} \quad (49)$$

C_1 bezeichnet den ersten Zeitschritt der Datenaufnahme $C_1 = C(t = t_1)$, wobei $t_1 \ll \tau$ vorausgesetzt wird. Wenn man exakte Werte für $C(t)$ aus Gleichung 48 verwendet, dann lassen sich S^2 und τ mit den Gleichungen 50 und 51 bestimmen [34].

$$S^2 = C(\infty) = \frac{4\pi}{5} \cdot \sum_{|M| \leq 2} |\langle Y_{2M}(\theta, \varphi) \rangle|^2 = \frac{4\pi}{5} \cdot \sum_{|M| \leq 2} \left| \frac{1}{T} \int_0^{\infty} Y_{2M}(\theta, \varphi) dt \right|^2 \quad (50)$$

$$\tau = \frac{1}{C_1 - S^2} \cdot \int_0^{\infty} (C(t) - S^2) dt \quad (51)$$

3 Ergebnisse und Diskussion

3.1 Molekular-dynamik-Simulation an RNase T_1

3.1.1 Simulationsbedingungen

Zur Untersuchung der dynamischen Struktur von RNase T_1 wurde eine ausgedehnte Simulation des Proteins in einer mit 8448 Wassermolekülen gefüllten kubischen Box mit einer Kantenlänge von 65.5 Å durchgeführt.

Zur Minimierung des Rechenaufwands wurde mit dem Ansatz der vereinigten Atome gerechnet [48]. Als Anfangsstruktur diente die Energieminimierte Struktur des NMR-Strukturensembles [31], die zu den experimentellen Randbedingungen die geringsten Verletzungen aufweist. Für die zeitliche Integration wurde der „leap frog“-Algorithmus [42] verwendet. Als Wassermodell wurde SPC2 gewählt [48]. Um die Bindungslängen konstant zu halten wurden der LINCS-Algorithmus [49] gewählt. Dadurch konnte die Integrationszeit für einen Schritt auf 2 fs festgesetzt werden. Das System wurde an ein Wärmebad mit 300 K gekoppelt. Die Simulation wurde mit Hilfe des Programmpaketes GROMACS durchgeführt [48].

Von der gesamten Simulationszeit von 15 ns wurden nur die letzten 13 ns für die Auswertung berücksichtigt, um sicherzustellen, daß sich das System im Gleichgewicht befindet.

3.2 Analyse der essentiellen Dynamik

Ein zentrales Problem von MD-Rechnungen ist die Beschränkung durch die Rechengeschwindigkeit [136]. Für die im Rahmen dieser Arbeit durchgeführte Simulation an RNase T_1 benötigte ein „personal computer“ (PC) der jeweils neuesten Generation ca. 15 Monate. Dennoch wird mit einer gesamten auswertbaren Simulationszeit von 13 ns noch lange nicht die Periodenlänge der langsamsten Bewegungen erreicht. Sehr langsame Bewegungen können durch zusätzliche Kräfte in die Richtungen dieser langsamen Bewegungen erfasst werden.

Als Grundlage zur Erfassung dieser konzentrierten Bewegungsrichtungen dient die durchgeführte MD-Simulation. Die Analyse der essentiellen Dynamik wurde mit 26000 Strukturen in einem Abstand von jeweils 500 fs durchgeführt. Die Strukturen enthielten ausschließlich die C_α -Atome eines jeden Aminosäurerestes, was für die Beschreibung der essentiellen Dynamik ausreicht [40, 156].

Die Strukturen wurden auf ihre Schwerpunkte zentriert und auf ihre mittlere Struktur überlagert. Die Kovarianzmatrix wurde nach Gleichung 25 ermittelt und in ihr Eigenwertsystem mit den Gleichungen 26 und 27 überführt. Die Eigenvektoren mit den höchsten zugeordneten Eigenwerten beschreiben die Bewegungsrichtungen mit den niedrigsten Frequenzen [40, 156].

In Abbildung 3 sind die 40 höchsten der insgesamt 312 Eigenwerte nach abnehmender Größe geordnet dargestellt. Es existieren offensichtlich nur wenige Eigenwerte, die deutlich von Null verschieden sind. Eine Einteilung der Dynamik in einen niederdimensionalen essentiellen und einen hochdimensionalen physikalisch stark beschränkten Konformationsraum ist also prinzipiell möglich. Der relative Anteil der Dynamik aus der Simulation, der mit den ersten N Eigenvektoren dargestellt werden kann, ist Abbildung 4 zu entnehmen. Die relative Fluktuation berechnet sich aus der relativen kumulativen Summe der Eigenwerte. Bereits mit sieben Eigenvektoren läßt sich über 80% der Fluktuation der Trajektorie darstellen.

3.2.1 Stabilität der Eigenvektoren

Die ermittelten Moden müssen innerhalb der Trajektorie stabil sein, um verlässlich zu sein. Die Ergebnisse sollten demnach nur gering von der Auswahl des Teils der Trajektorie abhängen, welchen man als Grundlage der Analyse verwendet.

Die Trajektorie wurde dazu in zwei Hälften aufgeteilt und die EDA getrennt durchgeführt.

Die Abbildung 5 zeigt ein Intensitätsdiagramm der quadratischen inneren Produkte zwischen allen Eigenvektoren der beiden Halbtrajektorien. Es tritt

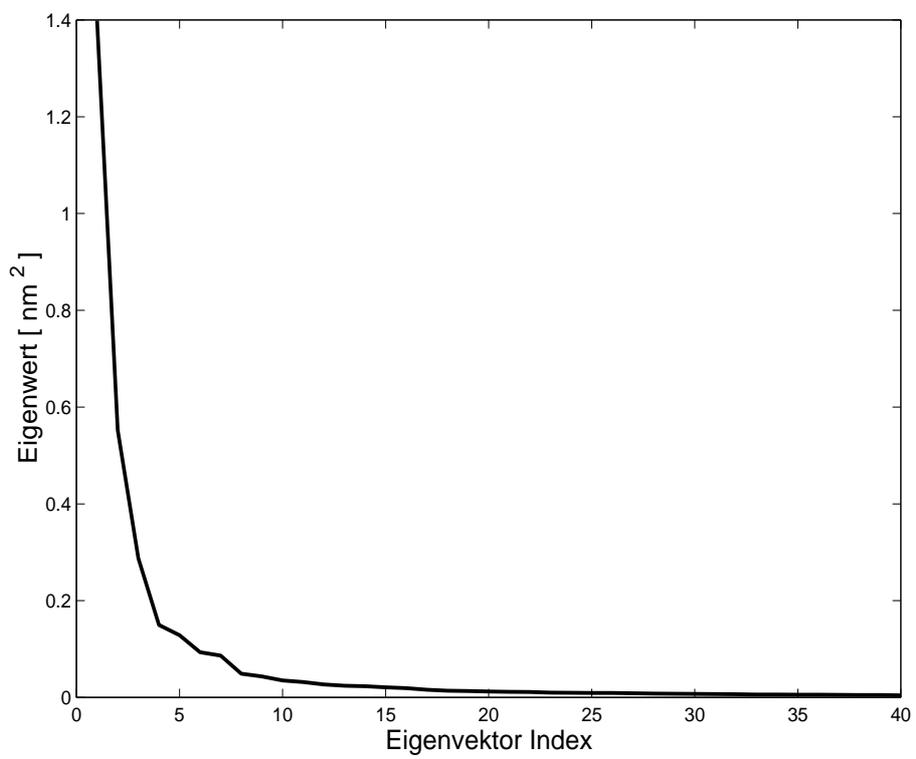


Abbildung 3: Die 40 höchsten der insgesamt 312 Eigenwerte aus der Analyse der essentiellen Dynamik nach abnehmender Größe geordnet.

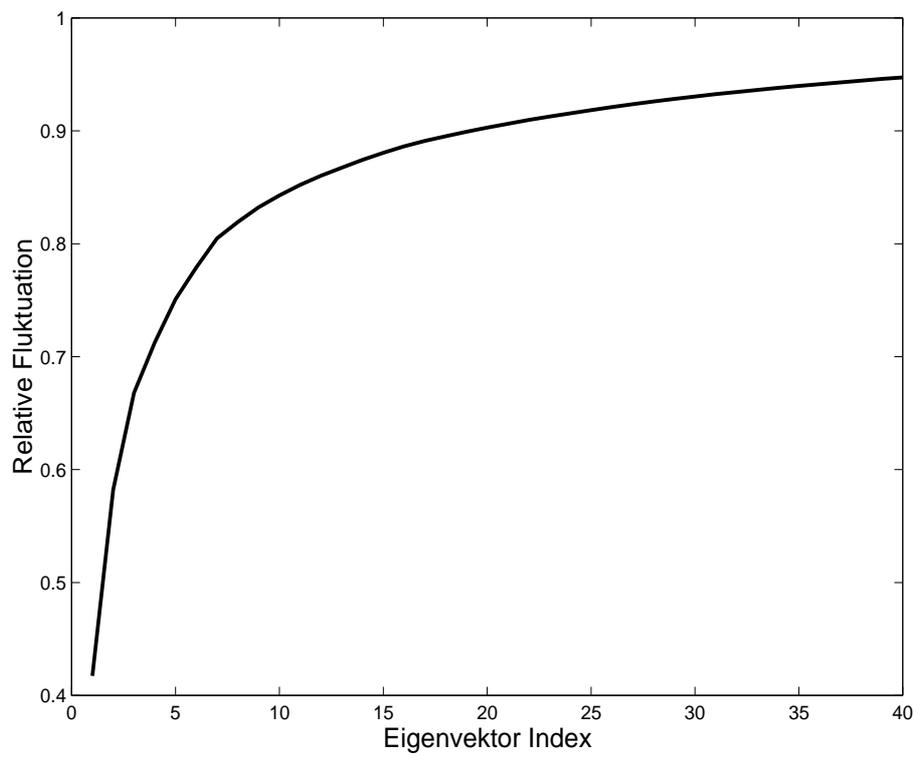


Abbildung 4: Relativer Anteil der Dynamik aus der Simulation, der mit den ersten N Eigenvektoren erfaßt wird.

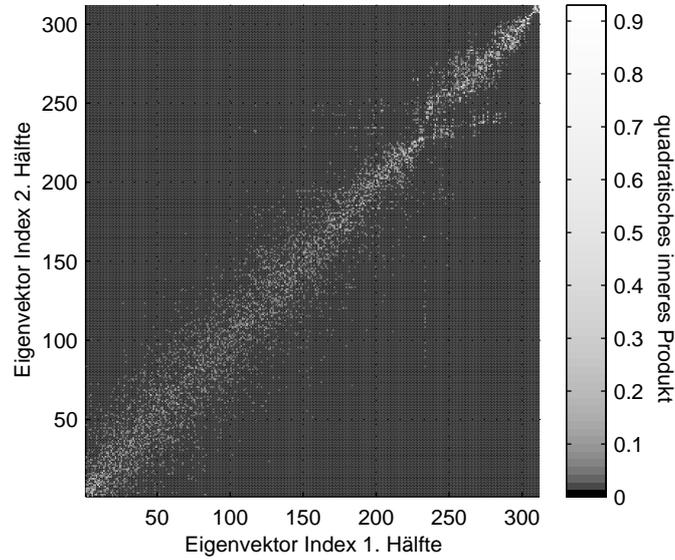


Abbildung 5: Intensitätsdiagramm des quadratischen inneren Produkts zwischen allen Eigenvektoren der beiden Halbtrajektorien

keine signifikante Überlappung zwischen Eigenvektoren mit hohen Eigenwerten (kleiner Index) und Eigenvektoren mit niedrigen Eigenwerten (hoher Index) der jeweils anderen Hälfte der Trajektorie auf. Der größte Teil der Überlappungen liegt in der Nähe der Diagonalen. Die Eigenvektoren beider Halbtrajektorien und die relative Größe ihrer Eigenwerte sind also sehr ähnlich.

Weil die ersten Eigenvektoren den maßgeblichen Anteil der Dynamik beschreiben, bedürfen diese einer genaueren Untersuchung, um festzustellen ob die Ergebnisse der Dynamik in den beiden Halbtrajektorien konsistent sind. Die einzelnen Vektorsätze der Halbtrajektorien lassen sich detailliert mit Hilfe des mittleren kumulativen quadratischen inneren Produktes (ACSIP= „Average Cumulative Square Inner Product“) der Eigenvektoren [85,103] quantitativ vergleichen. Die nach Gleichung 52 berechneten ACSIP-Verläufe geben den Anteil eines Vektors oder Vektorsatzes an, der durch einen Satz von anderen Vektoren beschrieben werden kann [85,157].

$$ACSIP(k) = \frac{1}{D} \sum_{a=1}^D \sum_{b=1}^k \left(\vec{v}_a^A \cdot \vec{v}_b^B \right)^2 \quad (52)$$

$\vec{\nu}_a^A$ und $\vec{\nu}_b^B$ sind die zu vergleichenden Eigenvektoren aus den zwei Datensätzen A und B . D ist die Dimension des Raumes A . k kann zwischen 1 und $3N$ variieren, wenn N die Anzahl der in der Berechnung der Eigenvektoren berücksichtigten Atome ist.

Ein linearer Verlauf zeigt an, daß keine Korrelation zwischen den aufgespannten Räumen auftritt. Einen konvexen Verlauf erhält man, wenn der Vektorsatz B mit den ersten Vektoren von A korreliert. Überlappung von B mit Eigenvektoren niedriger Eigenwerte in A wird durch einen konkaven Verlauf angezeigt [85, 157]. In Abbildung 6 ist der ACSIP der ersten sieben Eigenvektoren der ersten Halbtrajektorie ($D = 1$) dargestellt durch die ersten 100 Moden der zweiten Halbtrajektorie aufgetragen. Auf der Geraden würden ACSIPs liegen, wenn sie unkorreliert wären. Der deutlich konvexe Verlauf bestätigt, daß die Räume, die durch die ersten Eigenvektoren der Halbtrajektorien aufgespannt werden, im wesentlichen identisch sind.

3.2.2 Die Modenanzahl für den essentiellen Raum

Der essentielle Raum sollte genügend Eigenvektoren zur Beschreibung der Dynamik beinhalten, um mit den entsprechenden Koordinaten die Trajektorie beschreiben zu können [143]. Andererseits sollte die Dimension wegen des steigenden Aufwands bei einem eventuellen späteren Abtasten des Raumes nicht zu hoch gewählt werden. In jedem Fall aber dürfen keine Eigenvektoren, die innerhalb der Trajektorie unreproduzierbar sind, berücksichtigt werden. In Abbildung 7 erkennt man die ACSIP-Werte der Räume, die durch die jeweils ersten N Eigenvektoren der Halbtrajektorien aufgespannt werden. Die Kurve steigt bis zu einer Anzahl von etwa sieben bis zehn Eigenvektoren auf einen Wert von 0,61 bis 0.66 an und steigt von da an bis zu Eigenvektor 312 langsam auf Eins.

Mit sieben Eigenvektoren sind über 80% der Dynamik darstellbar (Abbildung 4) und die relative Überlappung der Räume aus den Halbtrajektorien bei dieser Anzahl von Eigenvektoren steigt im Vergleich zu den ersten Eigenvektoren nur noch langsam an. Diese Definition ist eher willkürlich und wird

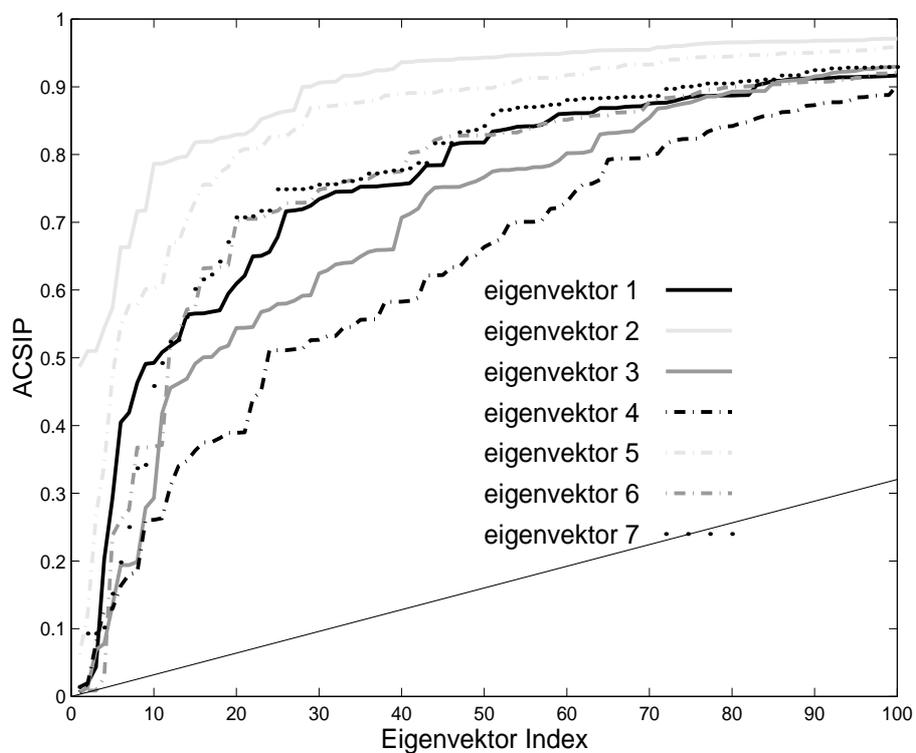


Abbildung 6: Mittlere kumulative quadratische innere Produkte (ACSIP) der ersten sieben Eigenvektoren der ersten Halbtrajektorie dargestellt durch die ersten 100 Moden der zweiten Halbtrajektorie

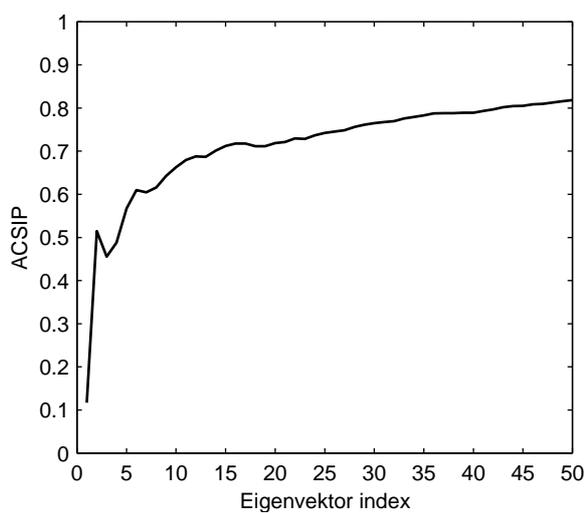
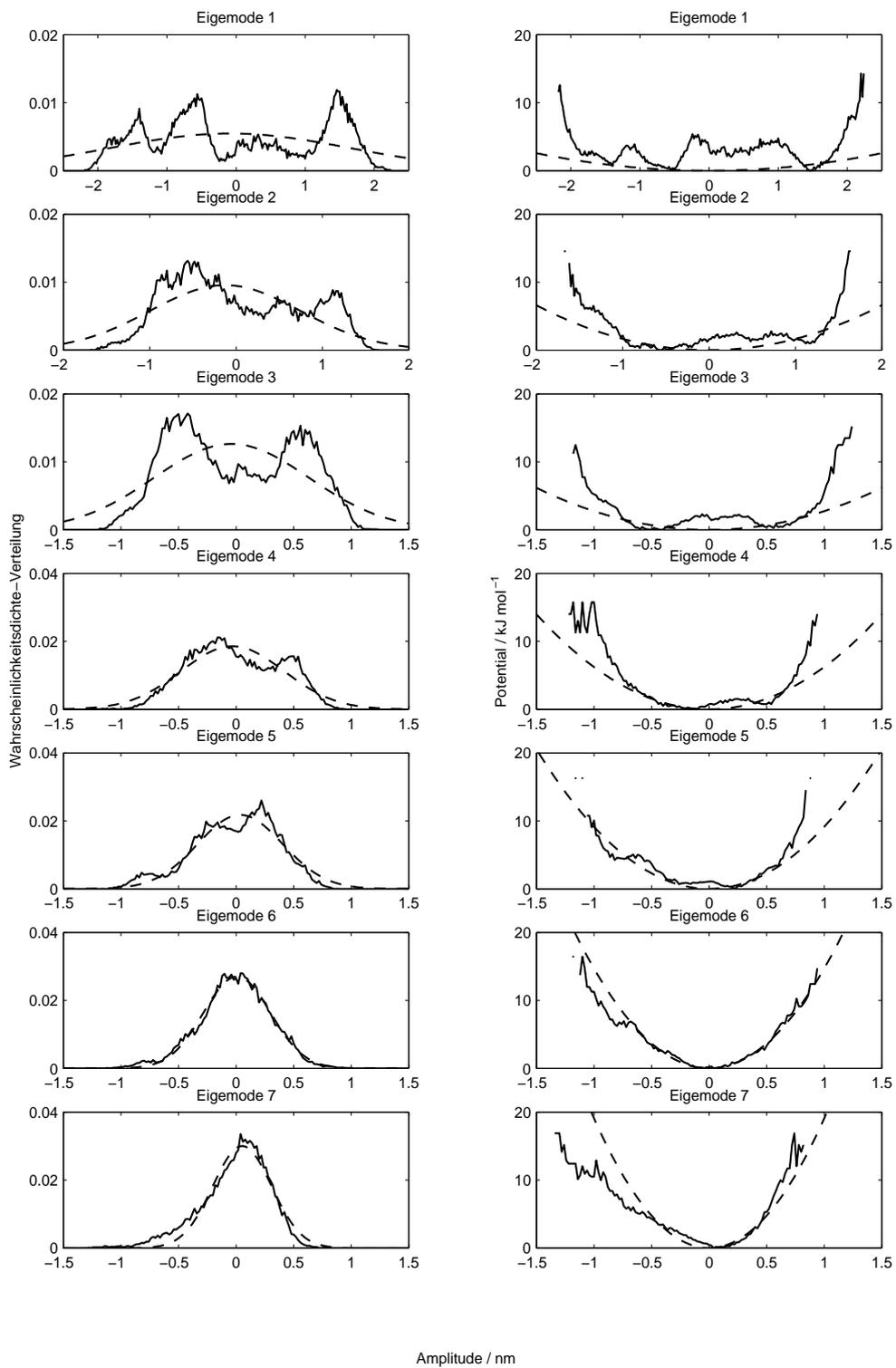
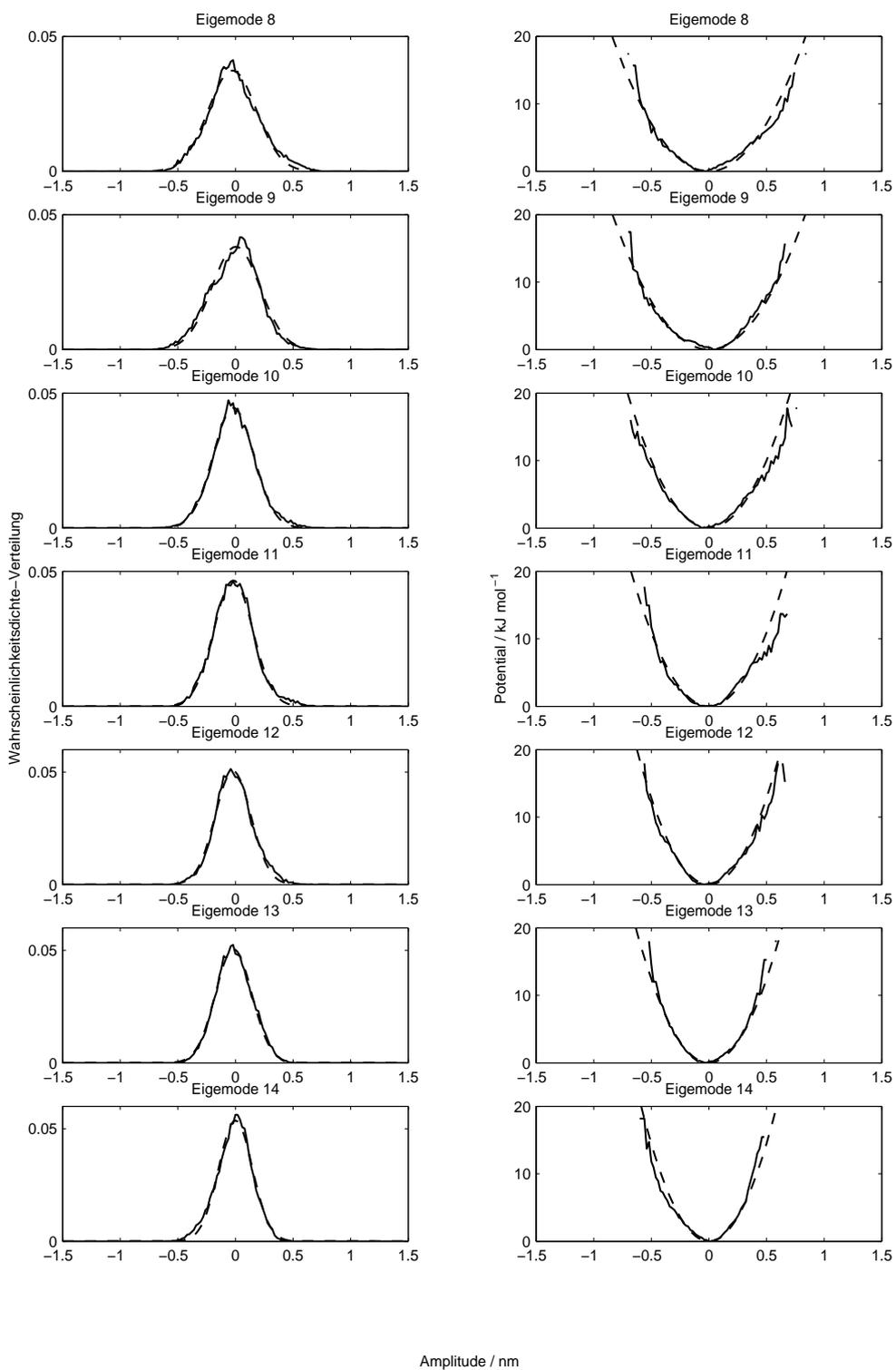


Abbildung 7: Relative Überlappung der Räume, die durch die jeweils ersten N Eigenvektoren der Halbtrajektorien aufgespannt werden.

im folgenden durch eine Betrachtung der Fluktuationen entlang der Moden im transformierten Raum überprüft.

Die relativen Verteilungen der Amplituden in den transformierten Koordinaten werden in Abbildung 8 für die ersten 21 Moden dargestellt. Werden die Eigenwerte als die Varianzen eines harmonischen Oszillators interpretiert, dann ergeben sich die Potentialverläufe, die in Abbildung 8 links gestrichelt dargestellt sind. Auf der rechten Seite von Abbildung 8 erkennt man die Potentialverläufe, die der Wahrscheinlichkeitsdichtefunktion aus der Simulation (durchgezogene Linie) und dem entsprechenden harmonischen Potential (gestrichelt) entsprechen. Nach der in Kapitel 2.2.5 vorgestellten Unterteilung der Moden von Kitao et al. [86] haben die Moden, die den essentiellen Raum bilden, kein harmonisches Potential. Ab der siebten bis achten Mode trifft die Annahme von unabhängigen harmonischen Oszillatoren tatsächlich zu, wodurch bestätigt wird, daß mit den ersten sieben Eigenvektoren die essentielle Dynamik des Proteins beschrieben werden kann.





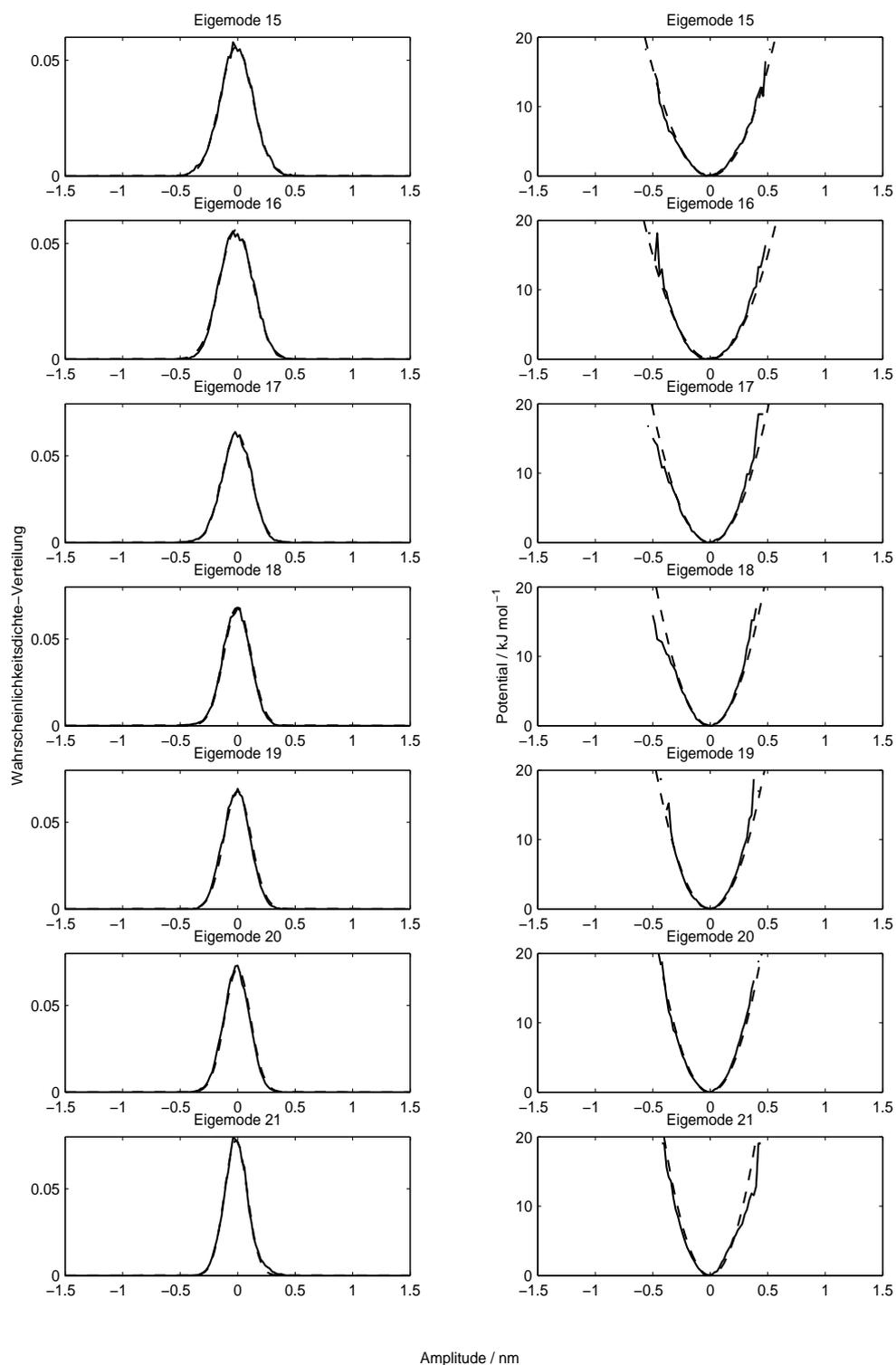


Abbildung 8: Relative Verteilungen der Amplituden in den transformierten Koordinaten für die ersten 21 Moden (links durchgezogen). Werden die Eigenwerte als die Varianzen eines harmonischen Oszillators interpretiert, dann ergeben sich die Potentialverläufe, die links gestrichelt dargestellt sind. Auf der rechten Seite erkennt man die Potentialverläufe, die der Wahrscheinlichkeitsdichtefunktion aus der Simulation (durchgezogene Linie) und dem entsprechenden harmonischen Potential (gestrichelt) entsprechen.

3.2.3 Vergleich mit der PCA-Analyse im Distanzraum

Ein grundsätzliches Problem der Analyse der essentiellen Dynamik aus einer Trajektorie in kartesischen Koordinaten ist bedingt durch die Trennung der Bewegung des gesamten Moleküls von der internen Bewegung. Dieses Problem wird im allgemeinen durch Überlagerung der Strukturen auf eine Referenzstruktur gelöst [36, 158, 159]. Werden interne Strukturkoordinaten wie Distanzen oder Diederwinkel zur Beschreibung der Trajektorie verwendet, dann ist die globale Bewegungen des Moleküls unabhängig von der internen Dynamik. Das von Abseher und Nilges vorgeschlagene Verfahren als Alternative zur EDA besteht aus einer Hauptkomponentenanalyse im Distanzraum (PCA) [36, 86]. Für diese Analyse müssen die Strukturen des Ensembles nicht überlagert werden, da intramolekulare Distanzen unabhängig von Rotation und Translation des Moleküls sind. Durch Vergleich von PCA-Analyse im Distanzraum und EDA lassen sich die Ergebnisse der EDA somit verifizieren.

Auch bei der PCA im Distanzraum werden ausschließlich die Abstände der C_α -Atome zueinander berücksichtigt. Die Ergebnisse der PCA-Analyse sind Eigenwerte mit zugeordneten Distanz-Eigenvektoren der Dimension $N \cdot (N - 1) / 2$, wobei die Anzahl der in der Analyse berücksichtigten Atome mit N bezeichnet wird.

Die bisher benutzte Methode zum Vergleich von EDA und PCA-Eigenvektoren beruht auf einer Reduktion des Informationsgehaltes der beiden Vektorsätze zu vergleichbaren Größen [36]. Im folgenden wird diese Methode an RNase T_1 angewendet.

Die Elemente der PCA-Vektoren können alternativ auch in einer Notation mit zwei Indizes geschrieben werden, wobei jeder Index eine der beteiligten Atome angibt. Für die weitere Analyse werden die Eigenvektoren als symmetrische Matrix mit dem Rang 2 geschrieben. Wie aus Gleichung 53 hervorgeht, haben diese Matrizen eine Dimension von $N \times N$ und enthalten auf ihrer Diagonalen ausschließlich Nullen.

$$\overrightarrow{\nu^{dis}} = \begin{pmatrix} \nu_1^{dis} \\ \nu_2^{dis} \\ \vdots \\ \nu_i^{dis} \\ \vdots \\ \nu_{N/2 \cdot (N-1)}^{dis} \end{pmatrix} \rightarrow \begin{pmatrix} \nu_{2,1}^{dis} \\ \nu_{3,1}^{dis} \\ \vdots \\ \nu_{k,l}^{dis} \\ \vdots \\ \nu_{N,N-1}^{dis} \end{pmatrix} \rightarrow \begin{pmatrix} 0 & \nu_{2,1}^{dis} & \cdots & \nu_{k,l}^{dis} & \cdots & \nu_{N,1}^{dis} \\ \nu_{2,1}^{dis} & 0 & & & & \\ \vdots & & \ddots & & & \vdots \\ \nu_{k,l}^{dis} & & & 0 & & \vdots \\ \vdots & & & & \ddots & \\ \nu_{N,1}^{dis} & \cdots & \cdots & & & 0 \end{pmatrix} \quad (53)$$

Zum in der Literatur beschriebenen Vergleich mit den EDA-Moden [36] werden die Distanzvektoren $\overrightarrow{\nu^{dis}}$ nach Gleichung 54 zu N -dimensionalen Vektoren $\overrightarrow{\nu_{red}^{dis}}$ reduziert. $\nu_{k,red}^{dis}$ ist das k -te Element im Vektor $\overrightarrow{\nu_{red}^{dis}}$.

$$\nu_{k,red}^{dis} = \sum_{l=1}^N (\nu_{k,l}^{dis})^2 \quad (54)$$

Die Eigenvektoren der EDA werden nach Gleichung 55 reduziert. $\overrightarrow{\nu_k^{EDA}}$ ist der 3-dimensionale Vektor aus der EDA-Mode $\overrightarrow{\nu^{EDA}}$, der sich auf das Atom k bezieht. Der sich ergebende reduzierte EDA-Vektor, der sich aus den Elementen $\nu_{k,red}^{EDA}$ zusammensetzt, wird im weiteren mit $\overrightarrow{\nu_{red}^{EDA}}$ bezeichnet.

$$\nu_{k,red}^{EDA} = \left(\overrightarrow{\nu_k^{EDA}} \right)^2 \quad (55)$$

Die Vektoren $\overrightarrow{\nu_{red}^{dis}}$ und $\overrightarrow{\nu_{red}^{EDA}}$ beschreiben laut Literatur [36] die gleichen Eigenschaften, nämlich die Bewegungsamplituden der einzelnen Atome innerhalb einer Mode. Dennoch sind sie nicht nach Gleichung 52 über ACSIP-Werte vergleichbar, da die Vektoren der betrachteten Räume zueinander orthogonal sein müßten. Die Ähnlichkeit zwischen den reduzierten Eigenvektoren im Koordinaten- und im Distanzraum wird mit Hilfe von Korrelationskoeffizienten nach Gleichung 56 quantifiziert [160, 161].

$$r = \frac{\sum_{i=1}^N \left(\nu_{i,red}^{EDA} - \langle \overline{\nu_{red}^{EDA}} \rangle \right) \cdot \left(\nu_{i,red}^{dist} - \langle \overline{\nu_{red}^{dist}} \rangle \right)}{\sqrt{\sum_{i=1}^N \left(\nu_{i,red}^{EDA} - \langle \overline{\nu_{red}^{EDA}} \rangle \right)^2 \cdot \sum_{i=1}^N \left(\nu_{i,red}^{dist} - \langle \overline{\nu_{red}^{dist}} \rangle \right)^2}} \quad (56)$$

In Abbildung 9 ist die Korrelation zwischen dem reduzierten kartesischen Raum und dem reduzierten Distanzraum für die ersten 10 Eigenvektoren für RNase T_1 dargestellt. Der Grauton der Quadrate kodiert den Korrelationskoeffizienten r des jeweiligen Vektorenpaars. Die Zahlen in den Quadraten geben die Wahrscheinlichkeiten t nach Student an [160, 161], mit der die Korrelation der beiden Vektoren zufällig ist. Werte kleiner als 10^{-2} sollen eine signifikante Korrelation anzeigen. Es sind zu jedem reduzierten EDA-Eigenvektor (y-Achse) mehrere PCA-Eigenvektoren (x-Achse) zu erkennen, mit denen eine hohe Korrelation (helle Schattierung) erreicht wird. Bis auf 5 geringe Korrelationen sind alle anderen Korrelationen zwischen den jeweils ersten zehn Eigenvektoren nach dem t -Test signifikant. Es läßt sich demnach nicht jedem EDA-Eigenvektor genau ein PCA-Eigenvektor zuordnen. Im Gegenteil, nahezu jedes Paar von reduzierten Eigenvektoren scheint korreliert zu sein, wenn auch in unterschiedlichem Ausmaß. Die hohen Korrelationen verschiedener Paare ist darin zu begründen, daß entgegen der Meinung in [36] die aus den ursprünglichen Ergebnissen der beiden Analysen unterschiedliche Arten von Informationen entfernt worden sind, um scheinbar vergleichbare Daten zu erhalten. Aus den Ergebnissen der PCA geht die Information verloren, zu welchen Atomen die Abstände eines bestimmten Atoms variieren. Man berechnet die Summe aller Abstandsvarianzen eines bestimmten Atoms. Die EDA-Vektoren dagegen enthalten Abstandsinformationen nur implizit über die Richtungen im 3-dimensionalen Raum und die Referenzstruktur, auf die man sich bezieht.

Aufgrund dieses unbefriedigenden Ergebnisses wurde in dieser Arbeit ein Verfahren entwickelt, wonach man die EDA- und PCA-Räume auch mit der ACSIP-Methode vergleichen kann. Dazu werden die kartesischen EDA-Vektoren in Distanz-, also „quasi-PCA“-Vektoren überführt.

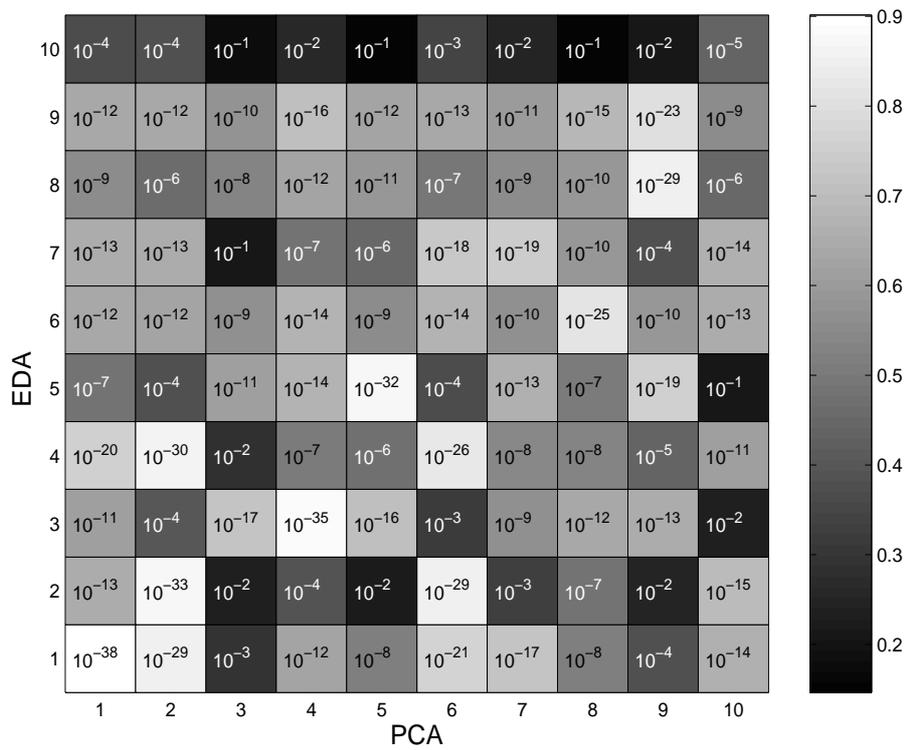


Abbildung 9: Korrelation zwischen dem reduzierten kartesischen und dem reduzierten Distanzraum. Die Zahlen in den Quadraten geben die Wahrscheinlichkeiten t nach Student an [160, 161], mit der die Korrelation der beiden Vektoren zufällig ist.

Die Strukturen, die eine Distanzmode repräsentieren, sind mit Gleichung 57 beschrieben.

$$d_{kl} = q^{dis} \cdot \nu_{kl}^{dis} + \langle d_{kl} \rangle \quad (57)$$

Die Distanz d_{kl} zwischen den Atomen i und j ist eine lineare Funktion der Amplitude q^{dis} der Mode $\overrightarrow{\nu}^{dis}$. Die Distanz bei einer Amplitude von Null ist gleich der Distanz in der Referenzstruktur $\langle d_{kl} \rangle$ - normalerweise die mittlere Distanz im Strukturensamble. Die Ableitung der Distanz nach der Amplitude der Distanzmode ist der Wert ν_{kl}^{dis} im Distanzvektor $\overrightarrow{\nu}^{dis}$.

$$\frac{\partial d_{kl}}{\partial q^{dis}} = \nu_{kl}^{dis} \quad (58)$$

Eine kartesische Mode wird durch einen $3N$ -dimensionalen Vektor $\overrightarrow{\nu}^{cart}$ repräsentiert, der die relativen Geschwindigkeiten aller Atome in den drei Raumrichtungen des Koordinatensystems beinhaltet. Die durch eine kartesische Mode erzeugten Strukturen werden durch Gleichung 59 ausgedrückt.

$$\overrightarrow{x} = \overline{\langle x \rangle} + q \cdot \overrightarrow{\nu}^{cart} \quad (59)$$

\overrightarrow{x} besteht aus den kartesischen Koordinaten einer Konformation, bei der nur eine Mode $\overrightarrow{\nu}^{cart}$ relativ zur Referenzstruktur $\overline{\langle x \rangle}$ ausgelenkt wurde. Alle kartesischen Koordinaten von \overrightarrow{x} hängen linear von der Amplitude q^{cart} der Mode ab. Gleichung 59 kann auch als Satz von 3-dimensionalen Vektorgleichungen für jedes Atom k geschrieben werden.

$$\overline{\langle x_k \rangle} = q^{cart} \cdot \overrightarrow{\nu}_k^{cart} + \overline{\langle x_k \rangle} \quad (60)$$

Die kartesischen Koordinaten $\overline{\langle x_k \rangle}$ des Atoms k werden mit Gleichung 60 berechnet. Die entsprechenden Indizes der kartesischen Moden werden mit $\overrightarrow{\nu}_k^{cart}$ bezeichnet. Die Referenzposition des Atoms k wird $\overline{\langle x_k \rangle}$ genannt.

Die Distanz d_{kl} zwischen zwei Atomen k und l ist die Länge der Differenz

ihrer kartesischen Koordinaten.

$$d_{kl} = |\vec{x}_k - \vec{x}_l| = \left| q^{cart} \cdot \left(\overrightarrow{\nu_k^{cart}} - \overrightarrow{\nu_l^{cart}} \right) + \langle \vec{x}_k \rangle - \langle \vec{x}_l \rangle \right| \quad (61)$$

Die Ableitung von d_{kl} nach der Amplitude der kartesischen Mode q^{cart} zeigt Gleichung 62, die sich zu Gleichung 63 vereinfachen lässt, wenn die Ableitung an der Referenzstruktur berechnet wird, indem $q^{cart} = 0$ gesetzt wird.

$$\frac{\partial d_{kl}}{\partial q^{cart}} = \frac{\left(\overrightarrow{\nu_k^{cart}} - \overrightarrow{\nu_l^{cart}} \right) \cdot \left(\langle \vec{x}_k \rangle - \langle \vec{x}_l \rangle \right) + q^{cart} \cdot \left(\overrightarrow{\nu_k^{cart}} - \overrightarrow{\nu_l^{cart}} \right)}{\left| \langle \vec{x}_k \rangle - \langle \vec{x}_l \rangle \right| + q^{cart} \cdot \left(\overrightarrow{\nu_k^{cart}} - \overrightarrow{\nu_l^{cart}} \right)} \quad (62)$$

$$\frac{\partial d_{kl} (q^{cart} = 0)}{\partial q^{cart}} = \frac{\left(\overrightarrow{\nu_k^{cart}} - \overrightarrow{\nu_l^{cart}} \right) \cdot \left(\langle \vec{x}_k \rangle - \langle \vec{x}_l \rangle \right)}{\left| \langle \vec{x}_k \rangle - \langle \vec{x}_l \rangle \right|} \quad (63)$$

Eine kartesische Mode und die entsprechende Distanzmode müssen den selben Effekt auf den Distanzraum haben. Die Ableitungen der Distanzen nach den jeweiligen Amplituden der Moden q^{dis} und q^{cart} müssen übereinstimmen. Wenn man die Abhängigkeit der Ableitung von q^{cart} in Gleichung 62 vernachlässigt, ergibt sich Gleichung 64.

$$\frac{\partial d_{kl} (q^{cart})}{\partial q^{cart}} \approx \frac{\partial d_{kl} (q^{cart} = 0)}{\partial q^{cart}} = \frac{\partial d_{kl}}{\partial q^{dis}} \quad (64)$$

Die Kombination der Gleichungen 58 und 63 mit Gleichung 64 führt zu Gleichung 65.

$$\nu_{kl}^{dis} = \frac{\left(\overrightarrow{\nu_k^{cart}} - \overrightarrow{\nu_l^{cart}} \right) \cdot \left(\langle \vec{x}_k \rangle - \langle \vec{x}_l \rangle \right)}{\left| \langle \vec{x}_k \rangle - \langle \vec{x}_l \rangle \right|} \quad (65)$$

Gleichung 65 wird unter Verwendung von Gleichung 53 als Vorschrift zur Berechnung der Koordinaten von "pseudo-Distanzvektoren" $\overrightarrow{\nu^{dis}}$ benutzt.

Die durch die Transformation der EDA-Vektoren berechneten pseudo-Distanzvektoren sind fast vollständig orthogonal zueinander. Die Verteilung der Winkel zwischen

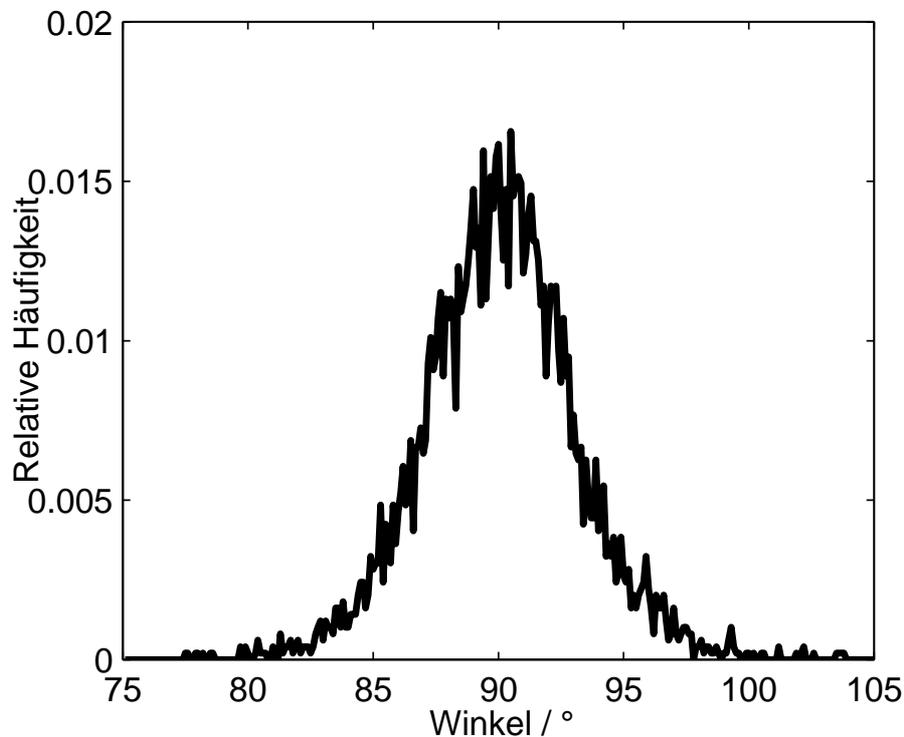


Abbildung 10: Verteilung der Winkel zwischen den ersten 100 pseudo-PCA-Vektoren.

den ersten zehn pseudo-PCA Vektoren ist Abbildung 10 zu entnehmen. Ungeachtet ihrer Winkel zueinander lassen sich die pseudo-Distanzvektoren durch ACSIP mit den PCA-Vektoren vergleichen, denn die PCA-Vektoren sind exakt orthogonal zueinander.

In Abbildung 11 sind die ACSIP-Werte der Darstellung der ersten zehn pseudo-PCA Vektoren durch die ersten zehn PCA-Vektoren gezeigt. Die sigmoidalen Verläufe zeigen eine erhebliche Übereinstimmung zwischen den Vektorsätzen an. Bis auf Vektor 9 und 10 können mit den ersten zehn PCA-Vektoren über 90% der jeweiligen pseudo-Distanzvektoren dargestellt werden. Es existieren also keine Korrelationen zwischen niedrig indizierten EDA-Vektoren und hoch indizierten PCA-Vektoren. Jeder der ersten acht EDA-Vektoren lässt sich eindeutig einem oder zwei PCA-Vektoren zuordnen. Das Ergebnis des entsprechenden in der Literatur beschriebenen Verfahrens [36] (s. Abbildung 9) lässt

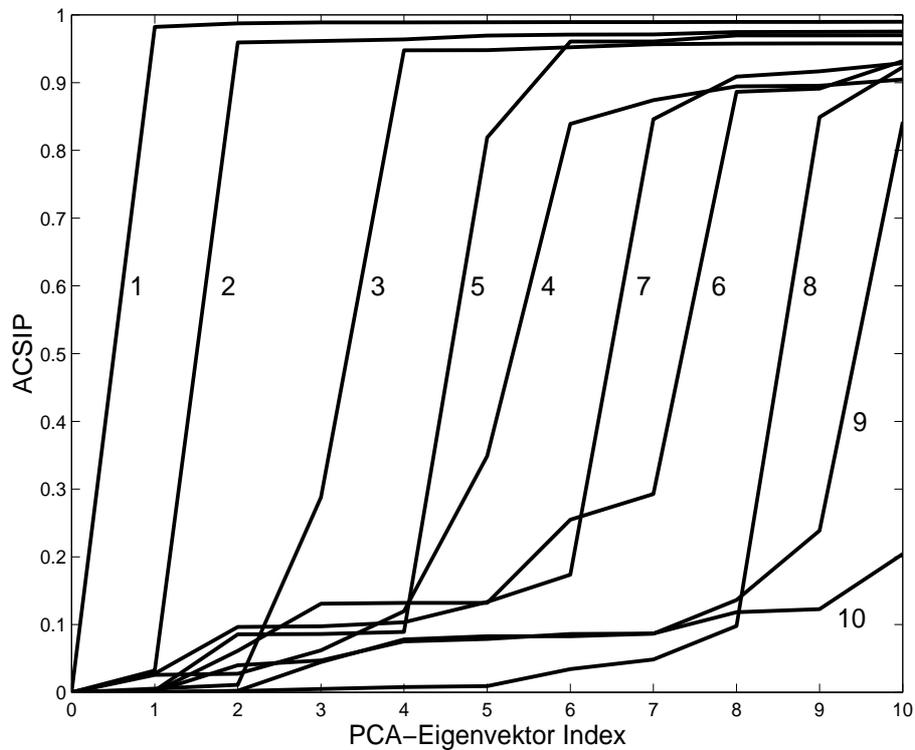


Abbildung 11: ACSIP-Werte bei der Darstellung der ersten zehn pseudo-PCA Vektoren durch die ersten zehn PCA-Vektoren. Die Zahlen neben den Kurven kennzeichnen die Nummer des EDA- Eigenvektors, der dem jeweiligen Verlauf zu Grunde liegt.

eine solche Zuordnung nicht zu. Die neu entwickelte Methode der Überführung der kartesischen Moden in Distanzmoden ist somit dem Standardverfahren vorzuziehen.

Der Überlappungsanteil der ersten 100 transformierten EDA Vektoren mit den ersten 100 PCA Vektoren ist in Abbildung 12 dargestellt. Man erkennt, daß Eigenvektoren mit ähnlichem Index einen hohen Überlappungsanteil aufweisen und Eigenvektoren, deren Index sehr unterschiedlich ist, keine Überlappung aufweisen.

Mit Hilfe des entwickelten Verfahrens konnte nachgewiesen werden, daß EDA und PCA auf die Distanzen sehr ähnliche Auswirkungen haben. Die Frage, ob die EDA-Moden zusätzliche artifizielle Bewegungen beschreiben, wird im weiteren genauer untersucht.

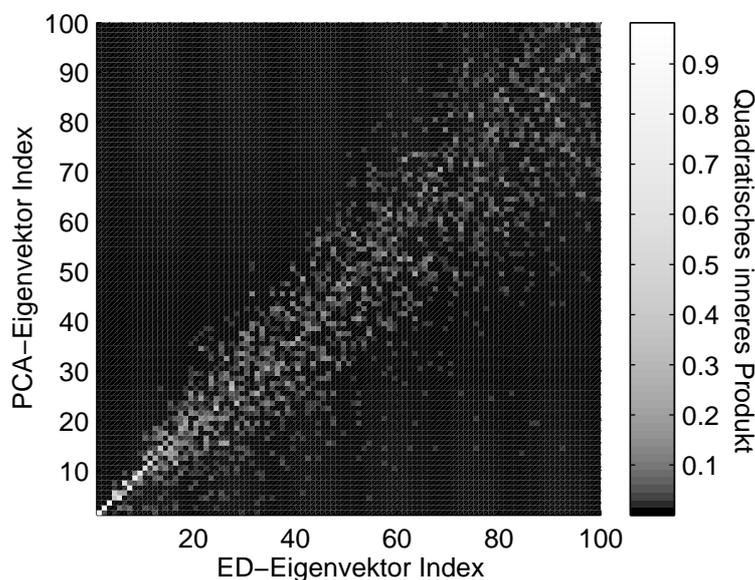


Abbildung 12: Quadratisches inneres Produkt der ersten 100 transformierten EDA Vektoren mit den ersten 100 PCA-Vektoren

3.3 Vorurteilsfreie kartesische Moden

Eine Trennung von interner und globaler Bewegung ist für Untersuchungen der internen Dynamik molekularer Systeme zwingend notwendig [36]. Sowohl die Analyse von MD-Rechnungen, als auch die Interpretation von NMR-Relaxationsdaten hängen von dieser Trennung ab [152]. Fehler führen zu einer voreingenommenen Beschreibung der internen Dynamik [36, 162].

Der übliche Algorithmus zur Entfernung der globalen Bewegung von einer simulierten Trajektorie in kartesischen Koordinaten wird in zwei Schritten durchgeführt. Zunächst werden alle Konformere des strukturellen Ensembles auf ihren jeweiligen Schwerpunkt zentriert. Danach werden alle Strukturen so gedreht, daß die mittlere quadratische Abweichung der Atome (MSD bzw. RMSD) zur mittleren Struktur minimal wird [37, 40, 158, 159].

Korrelationen zwischen weit voneinander entfernten Atomen sind von besonderem Interesse, weil sie im Gegensatz zu Korrelationen von etwa direkt aneinander gebundenen Atomen nicht trivial sind [36]. Die Einbeziehung von

mobilen Atomen für das Referenzsystem kann zu Korrelationen zwischen weit entfernten Atomen auf der Oberfläche des Makromoleküls führen [38]. Um solche Artefakte zu reduzieren, können auch lediglich die Atome, die den rigiden Anteil des Moleküls bilden, als Referenzsystem ausgewählt werden. Ein Nachteil dieses Ansatzes ist, daß die Wahl einer Gruppe von Atomen das Verschwinden von wirklichen Korrelationen bewirken kann, denn alle Korrelationen werden dann relativ zur Bewegung des als rigide angesehenen Anteils berechnet [39]. Zusätzlich kommt die Einbeziehung aller Atome der Null-Drehmoment Bedingung am nächsten [39]. Unter der Null-Drehmomentsbedingung versteht man, daß die Ausrichtung der Hauptträgheitsachsen des Systems unabhängig von den Amplituden der Moden ist.

Die Benutzung interner Variablen, auf die die globale Bewegung keinen Einfluß hat, ist ein möglicher Ansatz, um das Trennungsproblem zu vermeiden. Van Aalten et al. haben EDA im Φ/Ψ -Diederwinkelraum diskutiert, was im wesentlichen zwei Probleme aufwarf. Die Visualisierung von zurückgerechneten Strukturen, die durch die Eigenvektoren beschrieben werden, zeigt physikalisch unsinnige Strukturen, und die essentiellen Eigenvektoren im Φ/Ψ -Raum müssen nicht unbedingt zu großen Fluktuationen in der 3-dimensionalen Struktur führen [157]. Abseher und Nilges haben intramolekulare Distanzen als alternative Koordinaten benutzt [36]. Kovarianzen im Distanzraum haben aber eine andere Bedeutung als Kovarianzen, die aus räumlichen Eigenschaften einzelner Atome berechnet wurden. Die Distanzkovarianzen beschreiben die Korrelation der Bewegung von drei oder vier Atomen. Dementsprechend ist die Korrelation von exakt zwei Atomen nur implizit enthalten und muß in einem nachfolgenden Schritt berechnet werden. Außerdem ist es wünschenswert und manchmal zwingend notwendig, kartesische Moden zu benutzen, wie z.B. zur Identifizierung des essentiellen Konformationsraums [143–145].

Im Rahmen dieser Arbeit wird eine Methode entwickelt, die es ermöglicht kartesische Moden zu berechnen, die frei von der Voreingenommenheit einer Überlagerung und frei von der Wahl einer Untermenge von Atomen als Referenz-

system sind. Mit diesen Moden können korrigierte Kovarianzmatrizen berechnet werden.

3.3.1 Annahmen und Berechnungsprozeduren

Zunächst werden, wie schon beschrieben, Moden aus einer EDA gewonnen. Globale und lokale Bewegung werden dabei durch Überlagerung zur mittleren Struktur unter Berücksichtigung aller Atome als Referenz getrennt. Der Schlüssel zu dem vorgeschlagenen Algorithmus liegt in der Idee, daß die verbliebene Rotation und Translation in den Moden durch die Addition von kompensierenden Moden, die reine Rotation/Translation beschreiben, eliminiert werden kann. Die Gesamtmenge an Bewegung (TAM=Total Amount of Motion) nimmt nur dann ab, wenn Rotation oder Translation durch die addierten Vektoren kompensiert wird.

Der Algorithmus zur Entfernung der Rotation und Translation basiert auf drei Annahmen:

1. Die Menge an Rotation in einer EDA-Mode ist gering relativ zur gesamten durch die Mode beschriebenen Fluktuation.

Hierbei wird die übliche Annahme gelockert, daß EDA-Moden reine interne Bewegung beschreiben. Die EDA-Moden basieren auf einem minimierten RMSD in kartesischen Koordinaten. Diese Moden beinhalten aber immer noch Reste an Rotation und Translation, wie in dieser Arbeit gezeigt wird. Für kleine Änderungen kann die restliche Rotation in einer EDA-Mode durch einen linearen Vektor ausgedrückt werden, weil die Winkelabhängigkeit der Drehung vernachlässigt werden kann. Die Rotation um eine Achse kann durch die relativen Geschwindigkeiten beschrieben werden, die auftreten, wenn die Referenzstruktur um diese Achse gedreht wird. Die Menge an Rotation kann durch Skalierung der relativen Geschwindigkeiten reguliert werden.

2. Die gesamte Menge an Bewegung (TAM = Total Amount of Motion), die durch eine Mode repräsentiert wird, kann durch die Summe der massengewichteten Verschiebungen aller Atome in der

Mode beschrieben werden.

Der TAM-Wert ist ein quantitatives Maß für die Summe von interner und globaler Bewegung in einer Mode. Die Addition von purer Translation/Rotation hat keinen Einfluß auf die interne Bewegung. Die globale Bewegung des Biomoleküls als Ganzes muß also genau dann abnehmen, wenn der TAM sinkt. Es wird angenommen, daß eine Mode ohne globale Bewegung dieselben Konformationen erzeugt, wie eine Mode mit globaler Bewegung. Im letzten Fall enthält die Mode nur zusätzliche Rotation und Translation, die zu einem Anstieg der Verschiebungen der Atome führt. Daher wird die globale Bewegung kompensiert, wenn der TAM minimal ist. Zu diesem Zweck wird der TAM als Zielfunktion benutzt.

3. Wenn globale Bewegung identifiziert werden kann, dann kann diese nicht ein Teil der internen Bewegung sein.

Die interne Bewegung wird also durch Eliminierung einer maximalen Menge an globaler Bewegung erhalten.

Ausgehend von einer Mode, die mittels EDA berechnet wurde, wird der TAM durch Addition dreier Rotationsvektoren (ein Vektor pro Rotation um eine Hauptachse) und dreier Translationsvektoren (ein Vektor pro Translation in Richtung einer Hauptachse) minimiert. Die sich ergebenden Moden werden in einem nachfolgenden Schritt orthonormiert und korrigierte Kovarianzmatrizen werden mit angepassten Eigenwerten zurückgerechnet.

Mit der Absicht, den Ansatz ausgehend von unterschiedlichen Datensätzen zu verifizieren, wurde eine Methode entwickelt, PCA-Moden im Distanzraum in den kartesischen Raum zu überführen, ohne dabei einen expliziten Überlagerungsalgorithmus zu verwenden. Der Vergleich der Ergebnisse ausgehend von EDA und PCA im Distanzraum kann dann zur Überprüfung verwendet werden, ob vorurteilsfreie kartesische Moden wirklich existieren.

Die beiden Datensätze, einer ausgehend von einer PCA im Distanzraum und einer ausgehend von einer EDA im kartesischen Raum, sollten in den selben Moden resultieren, wenn die Voreingenommenheit bezüglich der Wahl des Referenzsystems durch den vorgeschlagenen Algorithmus eliminiert werden kann.

3.3.2 Transformation von Distanzmoden in kartesische Moden

Die Transformation von Distanzmoden in kartesische Moden ist die umgekehrte Richtung des in Kapitel 3.2.3 beschriebenen Verfahrens. Jetzt wird also eine Distanzmode $\overrightarrow{\nu}^{dis}$ gesucht, die die Gleichung 65 erfüllt.

Allerdings beinhalten die Distanzvektoren redundante Informationen, weil bei gegebener Referenzstruktur jede beliebige Komponente eines Eigenvektors auch aus den anderen Komponenten des Vektors berechnet werden kann. Außerdem sind die Komponenten in dem Distanzvektor nicht selbstkonsistent, wenn eine konkrete kartesische Referenzstruktur angenommen wird, weil die Geometrie im 3-dimensionalen Raum nur spezielle Zusammensetzungen von Distanzen zulässt. Es ist also nur in sehr speziellen Fällen möglich, eine Distanzmode zu finden, die Gleichung 65 exakt erfüllt. Daher wird in Gleichung 66 eine Zielfunktion definiert, die minimal wird, wenn der kartesische Vektor der Erfüllung von Gleichung 65 am nächsten kommt. Die Funktion wird durch die mittleren Massen $\frac{m_k+m_l}{2}$ der beteiligten Atome k und l gewichtet. Die Minimierungen wurden mit Hilfe des Levenberg-Marquart Algorithmus [163] durchgeführt, der auf den Gradienten der Zielfunktion (Gleichung 67) zurückgreift. $\nabla_{\nu_l^{cart}}$ steht für den Vektor der partiellen Ableitungen zu den kartesischen Koordinaten des Atoms l . Alle Komponenten der anfänglichen kartesischen Moden wurden auf Null gesetzt.

$$T_{dist2cart}(\overrightarrow{\nu}^{cart}) = \sum_{k=1}^{N-1} \sum_{l=k+1}^N \frac{m_k+m_l}{2} \cdot \left((\overrightarrow{\nu}_k^{cart} - \overrightarrow{\nu}_l^{cart}) \cdot (\overrightarrow{\langle x_k \rangle} - \overrightarrow{\langle x_l \rangle}) - \nu_{kl}^{dis} \cdot |\overrightarrow{\langle x_k \rangle} - \overrightarrow{\langle x_l \rangle}| \right)^2 \quad (66)$$

$$\nabla_{\nu_l^{cart}}(T_{pca2cart}) = \sum_{k=1}^N (m_k + m_l) \cdot (\overrightarrow{\langle x_k \rangle} - \overrightarrow{\langle x_l \rangle}) \cdot \left(\nu_{kl}^{dis} \cdot |\overrightarrow{\langle x_k \rangle} - \overrightarrow{\langle x_l \rangle}| - (\overrightarrow{\nu}_k^{cart} - \overrightarrow{\nu}_l^{cart}) \cdot (\overrightarrow{\langle x_k \rangle} - \overrightarrow{\langle x_l \rangle}) \right) \quad (67)$$

Die den Distanzmoden entsprechenden kartesische Moden können somit mit-

tels der Gleichungen 66 und 67 berechnet werden, wenn eine Referenzstruktur angenommen wird.

3.3.3 Verhältnis zwischen den Amplituden in Distanzraum und den Amplituden in kartesischen Raum

Die kartesischen Vektoren, die durch Transformation vom Distanzraum in den kartesischen Raum erzeugt wurden, werden im weiteren $\overrightarrow{\nu^{cart}}$ genannt. Die entsprechenden normierten kartesischen Vektoren werden $\overrightarrow{\nu^{cart, \hat{n}}}$ genannt.

$$\overrightarrow{\nu^{cart, \hat{n}}} = \frac{\overrightarrow{\nu^{cart}}}{|\overrightarrow{\nu^{cart}}|} \quad (68)$$

Die kartesischen Moden und ihre jeweilige Ableitung können in Anlehnung an die Gleichungen 59 und 63 beschrieben werden.

$$\overrightarrow{x_k} = q^{cart, n} \cdot \overrightarrow{\nu_k^{cart, \hat{n}}} + \langle x \rangle \quad (69)$$

$$\frac{\partial d_{kl} (q^{cart, n} = 0)}{\partial q^{cart, n}} = \frac{(\overrightarrow{\nu_k^{cart, \hat{n}}} - \overrightarrow{\nu_l^{cart, \hat{n}}}) \cdot (\langle \overrightarrow{x_k} \rangle - \langle \overrightarrow{x_l} \rangle)}{|\langle \overrightarrow{x_k} \rangle - \langle \overrightarrow{x_l} \rangle|} \quad (70)$$

Einsetzen von Gleichung 68 führt zu Gleichung 71.

$$\frac{\partial d_{kl} (q^{cart, n} = 0)}{\partial q^{cart, n}} = \frac{(\overrightarrow{\nu_k^{cart}} - \overrightarrow{\nu_l^{cart}}) \cdot (\langle \overrightarrow{x_k} \rangle - \langle \overrightarrow{x_l} \rangle)}{|\overrightarrow{\nu^{cart}}| \cdot |\langle \overrightarrow{x_k} \rangle - \langle \overrightarrow{x_l} \rangle|} \quad (71)$$

Mit den Gleichungen 63 und 64 erhält man Gleichung 72, die das quantitative Verhältnis zwischen den transformierten Variablen im Distanzraum und im kartesischen Raum beschreibt.

$$\frac{\partial d_{kl} (q^{cart, n} = 0)}{\partial q^{cart, n}} = |\overrightarrow{\nu^{cart}}|^{-1} \cdot \frac{\partial d_{kl}}{\partial q^{dis}} \quad (72)$$

3.3.4 Eliminierung des Rotationsanteils und des Translationsanteils von Bewegungsmoden

Wie schon erwähnt wird globale und interne Bewegung eines Moleküls bisher getrennt, indem die Konformer einer Trajektorie entsprechend einem minimalen RMSD-Wert relativ zur mittleren Struktur überlagert werden. Variationen der Koordinaten werden danach als interne Bewegung interpretiert. Diese kartesischen Moden, die im weiteren EDA-Moden oder $\overrightarrow{\nu^{EDA}}$ genannt werden, enthalten immer noch Translations- und Rotationsanteile, wie im Rahmen dieser Arbeit gezeigt wird. Die Subtraktion von Rotation $\overrightarrow{\nu^{rot}}$ und Translation $\overrightarrow{\nu^{trans}}$ müsste dann „vorurteilsfreie“ („biasfree“) interne Bewegung $\overrightarrow{\nu^{bf}}$ beschreiben.

$$\overrightarrow{\nu^{bf}} = \frac{\overrightarrow{\nu^{int}} - \overrightarrow{\nu^{trans}} - \overrightarrow{\nu^{rot}}}{|\overrightarrow{\nu^{int}} - \overrightarrow{\nu^{trans}} - \overrightarrow{\nu^{rot}}|} \quad (73)$$

Der Rotationsvektor $\overrightarrow{\nu^{rot}}$ kann mit Gleichung 74 berechnet werden.

$$\overrightarrow{\nu_k^{rot}} = \sum_{i=1}^3 l_i \cdot (\overrightarrow{d_i} \times \overrightarrow{x_k}) \quad (74)$$

Die Vektoren $\overrightarrow{d_i}$ bezeichnen die normierten Hauptachsen des Systems. Der Ausdruck $\overrightarrow{d_i} \times \overrightarrow{x_k}$ ist ein Vektor, der das Atom k der Referenzstruktur um die Hauptachse $\overrightarrow{d_i}$ dreht. Die skalaren Parameter l_i legen die Menge an Rotation fest. Ein Translationsvektor $\overrightarrow{\nu^{trans}}$ kann mit Gleichung 75 berechnet werden. Die skalaren Parameter t_i legen die Menge an Translation bezüglich der Richtungen der Hauptachsen fest.

$$\overrightarrow{\nu_k^{trans}} = \sum_{i=1}^3 t_i \cdot \overrightarrow{d_i} \quad (75)$$

Durch Anwendung von Annahme 2 (Seite 56) kann die Rotations- und Translationsbewegung durch Minimierung der Zielfunktion in Gleichung 76 in \overrightarrow{t} und \overrightarrow{r} eliminiert werden.

$$T(\vec{l}, \vec{t}) = \sum_{k=1}^N m_k \cdot \left| \frac{\overline{\nu_k^{EDA}} - \sum_{i=1}^3 [l_i \cdot (\vec{d}_i \times \vec{x}_k) + t_i \cdot \vec{d}_i]}{\sqrt{\sum_{m=1}^N \left(\overline{\nu_m^{EDA}} - \sum_{i=1}^3 [l_i \cdot (\vec{d}_i \times \vec{x}_m) + t_i \cdot \vec{d}_i] \right)^2}} \right| \quad (76)$$

$$\vec{l} = \begin{pmatrix} l_1 \\ l_2 \\ l_3 \end{pmatrix}, \quad \vec{t} = \begin{pmatrix} t_1 \\ t_2 \\ t_3 \end{pmatrix}$$

Die Minimierung von T in \vec{l} und \vec{t} führt mit den Gleichungen 74, 75 und 73 zu den vorurteilsfreien Moden.

3.3.5 Berechnung einer neuen Kovarianzmatrix der positionellen Abweichungen

Die Korrelation zwischen atomaren Bewegungen kann durch die Kovarianzmatrix \mathbf{C} der positionellen Abweichungen ausgedrückt werden [40] und berechnet sich nach Gleichung 25 (Seite 21). Die symmetrische Matrix \mathbf{C} kann durch die orthogonale Koordinatentransformation \mathbf{T} zu einer diagonalen Matrix Λ mit Eigenwerten λ_i diagonalisiert werden (Gleichung 26). Die Koordinatentransformation \mathbf{T} besteht aus orthonormalen Spaltenvektoren. Die vorurteilsfreien Moden sind noch nicht orthogonal zueinander. Sie müssen also noch orthogonalisiert werden, bevor eine korrigierte Kovarianzmatrix analog zu Gleichung 25 berechnet wird.

Die Mode mit dem größten Eigenwert $\overline{\nu^{bf}}(\lambda_1)$ bleibt während der Orthogonalisierung unverändert. Von allen anderen Moden werden ihre Projektionen auf die Eigenvektoren mit höheren entsprechenden Eigenwerten subtrahiert.

$$\overrightarrow{\nu^{ortho}(\lambda_i)} = \begin{cases} \frac{\overrightarrow{\nu^{bf}(\lambda_i)} - \sum_{k=1}^{i-1} (\overrightarrow{\nu^{ortho}(\lambda_k)} \cdot \overrightarrow{\nu^{bf}(\lambda_i)}) \cdot \overrightarrow{\nu^{ortho}(\lambda_k)}}{|\overrightarrow{\nu^{bf}(\lambda_i)} - \sum_{k=1}^{i-1} (\overrightarrow{\nu^{ortho}(\lambda_k)} \cdot \overrightarrow{\nu^{bf}(\lambda_i)}) \cdot \overrightarrow{\nu^{ortho}(\lambda_k)}|} & \text{wenn } i \geq 2 \\ \overrightarrow{\nu^{bf}(\lambda_1)} & \text{wenn } i = 1 \end{cases} \quad (77)$$

Die Transformation \mathbf{T}^{ortho} setzt sich aus den orthonormierten vorurteilsfreien Vektoren $\overrightarrow{\nu^{ortho}(\lambda_i)}$ zusammen.

$$\mathbf{T}^{ortho} = \left(\overrightarrow{\nu^{ortho}(\lambda_1)} \quad \overrightarrow{\nu^{ortho}(\lambda_2)} \quad \dots \quad \overrightarrow{\nu^{ortho}(\lambda_{3N})} \right) \quad (78)$$

Die Kovarianzmatrix \mathbf{C}^{bf} kann mit Gleichung 79 berechnet werden. Die Herleitung zur Ermittlung der Diagonalelemente von Λ^{ortho} , den angepassten Eigenwerten λ_i^{ortho} , ist detailliert im Anhang ab Seite 114 beschrieben.

$$\mathbf{C}^{bf} = \mathbf{T}^{ortho} \cdot \Lambda^{ortho} \cdot (\mathbf{T}^{ortho})^T \quad (79)$$

$$\Lambda^{ortho} = \begin{pmatrix} \lambda_1^{ortho} & 0 & \dots & 0 \\ 0 & \lambda_2^{ortho} & \dots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & \lambda_{3N}^{ortho} \end{pmatrix} \quad (80)$$

3.3.6 Anwendung auf zwei Modellsysteme

Zwei Modellsysteme werden konstruiert, um den entwickelten Algorithmus zu testen. Modell 1 zeigt nur eine korrelierte Bewegung, um zu testen, ob in vorurteilsfreien Moden immer noch artifizielle Kovarianzen auftreten. Eine zweite Bewegung, die zur ersten korreliert ist, aber eine wesentlich kleinere Amplitude aufweist, wird in Modell 2 eingeführt, um zu überprüfen, ob Korrelationen auch noch detektierbar sind, wenn sie zwischen weit entfernten Atomen auftreten.

Beide Modelle bestehen aus Pseudomolekülen, die die Form eines Würfels haben. Eine Kante des Würfels wird durch einen Halbkreis (Schleufe A) ersetzt,

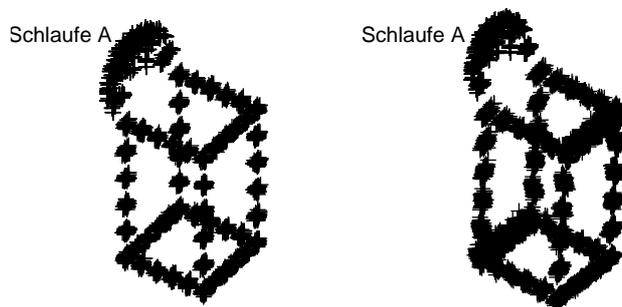


Abbildung 13: Modell 1: Jeweils 1000 Strukturen des Modellsystems sind auf unterschiedliche Weise überlagert worden. Auf der linken Seite ist nur der rigide Teil der pseudo-Moleküle ohne Schlaufe A als Referenzsystem für die Überlagerung benutzt worden. Durch die scharfe Kontur im überlagerten Ensemble wird angezeigt, daß bei der Wahl dieses Referenzsystems die Ausrichtung der Kuben im Raum in allen Konformeren identisch ist. Die rechte Seite zeigt dasselbe strukturelle Ensemble, nachdem die Strukturen mit allen Atomen als Referenz überlagert worden sind. Die Struktur des Ensembles erscheint hier verzerrt, weil die Ausrichtung der Kuben in den einzelnen Strukturen unterschiedlich ist. Durch die Wahl eines ungeeigneten Referenzsystems wurde also durch unterschiedliche Ausrichtung der einzelnen Strukturen scheinbare Bewegung eingeführt.

der um die vorherige Kante des Würfels rotiert. Zusätzlich fluktuieren alle „Atome“ stochastisch um ihre definierten Positionen. Die insgesamt 1000 Strukturen dieses Systems (Modell 1) sind Abbildung 13 zu entnehmen. Auf der linken Seite von Abbildung 13 ist nur der rigide Teil der Moleküle ohne die Schlaufe als Referenzsystem für die Überlagerung benutzt worden. Die einzigen Atome, die an der konzertierten Bewegung beteiligt sind, sind in der Schlaufe lokalisiert. Die rechte Seite von Abbildung 13 zeigt dasselbe strukturelle Ensemble, nachdem die Strukturen mit allen Atomen als Referenzsystem überlagert worden sind. Durch den Überlagerungsalgorithmus ist offensichtlich artifizielle Bewegung in den starren Teil des Systems eingeführt worden. Unter Definition aller Atome als Referenzsystem für die Überlagerung wurde die Kovarianzmatrix mit Gleichung 25 berechnet. Die Beträge der Werte in der resultierenden Matrix sind Abbildung 14 zu entnehmen. Die kartesischen Koordinaten der Strukturen sind nach den Atomen geordnet. Die ersten 52 Atome formen den rigiden Körper des Würfels. Die entsprechenden Nummern der Koordinaten sind 1 bis 156.

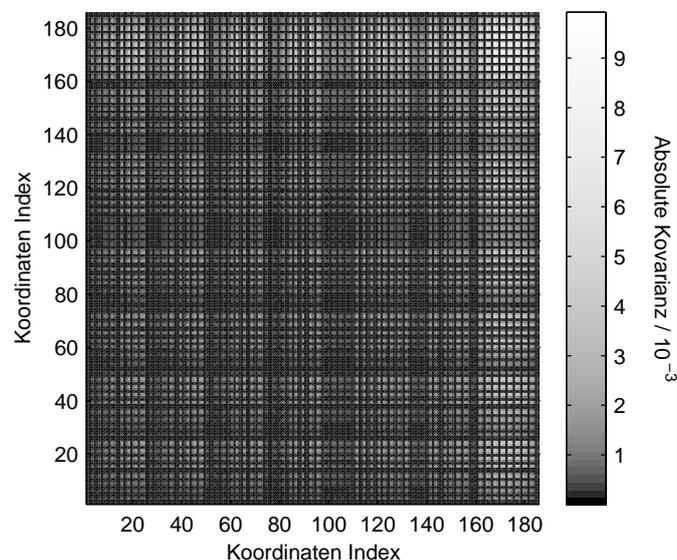


Abbildung 14: Modell 1: Beträge der Kovarianzmatrix bei Definition aller Atome als Referenzsystem für die Überlagerung der einzelnen Strukturen der Trajektorie

Die Koordinaten der Atome 53 bis 62 bilden Schlaufe A und entsprechen den Koordinaten 157 bis 186. Offensichtlich sind durch die Überlagerung artifizielle Kovarianzen eingeführt worden. Die Artefakte treten sowohl zwischen dem rigiden Teil und der Schlaufe auf als auch zwischen Atomen die sich beide im rigiden Teil des Pseudomoleküls befinden.

Für das gleiche System wurde eine PCA-Analyse im Distanzraum durchgeführt, um die Kovarianzen der Bewegung ohne Überlagerungsalgorithmus zu berechnen. Alle Distanzeigenvektoren wurden in kartesische Moden durch Minimierung der Zielfunktion in Gleichung 66 unter Zuhilfenahme der analytischen Gradienten aus Gleichung 67 überführt. Die sich ergebende Kovarianzmatrix zeigt die gleichen Artefakte wie die Matrix in Abbildung 14 (nicht abgebildet). Eine PCA-Analyse im Distanzraum mit anschließender Transformation in kartesische Moden führt also eine ähnliche Voreingenommenheit bezüglich der Unterscheidung zwischen interner und globaler Bewegung ein, wie eine Überlagerung aller Atome. Die Voreingenommenheit wird vermutlich während der Transforma-

tion eingeführt, indem der Nullvektor als anfängliche kartesische Mode gewählt wurde. Der Nullvektor erfüllt sowohl die Null-Translationsbedingung - d.h. die Lage des Schwerpunktes ist unabhängig von der Amplitude des Nullvektors - als auch die Null-Drehmomentsbedingung (Definition: s. S. 54). Diese beiden Eigenschaften kommen der minimalen RMSD-Bedingung sehr nahe. Dementsprechend sind auch die erzeugten kartesischen Moden nahe bei diesen Bedingungen. Dennoch ist keine explizite Überlagerungsmethode angewandt worden. Offensichtlich führt eine PCA von internen Distanzen mit sich anschließender Transformation in den kartesischen Raum nicht direkt zu unvoreingenommenen kartesischen Moden. Im weiteren wird nicht zwischen kartesischen Moden unterschieden, die ausgehend von kartesischen Koordinaten oder ausgehend von Distanzkoordinaten erzeugt wurden, weil die Ergebnisse vom vorgeschlagenen Algorithmus für die Modellsysteme gleich sind.

Die kartesische Mode mit der höchsten entsprechenden Amplitude ist Abbildung 15(a) zu entnehmen. Der Rotations- und Translations-Anteil der Bewegung wurde in (b) eliminiert. Die artifizielle Bewegung konnte vollständig mit Hilfe des Algorithmus' eliminiert werden. Das Ergebnis ist unabhängig von der räumlichen Orientierung der Referenzstruktur. Man erhält identische Vektoren, wenn die 3-dimensionalen Vektoren in (b) auf dieselbe Weise gedreht werden, wie zuvor die Referenzstruktur gedreht wurde. Das Ergebnis ist ebenfalls nahezu unabhängig von der Wahl der Referenzstruktur für die Transformation in den kartesischen Raum. Die Berechnung wurde für alle Strukturen des Ensembles als Referenz durchgeführt und der geringste Überlappungsanteil von zwei beobachteten Vektoren betrug 0.98. Abbildung 16 zeigt exemplarisch die Unterschiede zwischen einigen anderen Moden $\overline{\nu^{EDA}}$ (links) und den entsprechenden orthogonalisierten vorurteilsfreien Vektoren $\overline{\nu^{ortho}}$ (rechts). Nur die ersten beiden Eigenvektoren werden durch den Algorithmus maßgeblich beeinflusst. Dieses Ergebnis ist im Einklang mit der Tatsache, daß die interne Bewegung des Modells 1 in Abbildung 13 mit exakt zwei kartesischen Moden dargestellt werden kann.

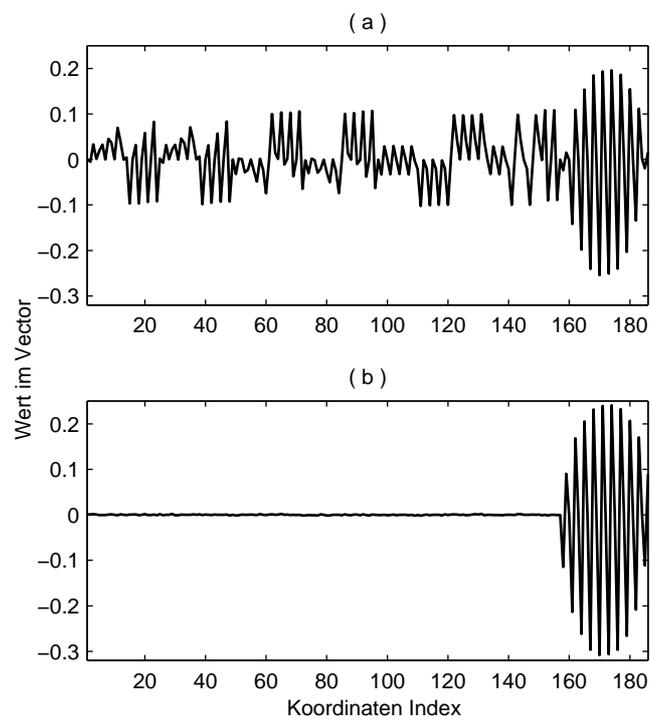


Abbildung 15: Modell1: Die kartesische Mode mit der höchsten entsprechenden Amplitude der EDA ist in (a) abgebildet. Der Rotations- und Translations-Anteil wurde in (b) eliminiert.

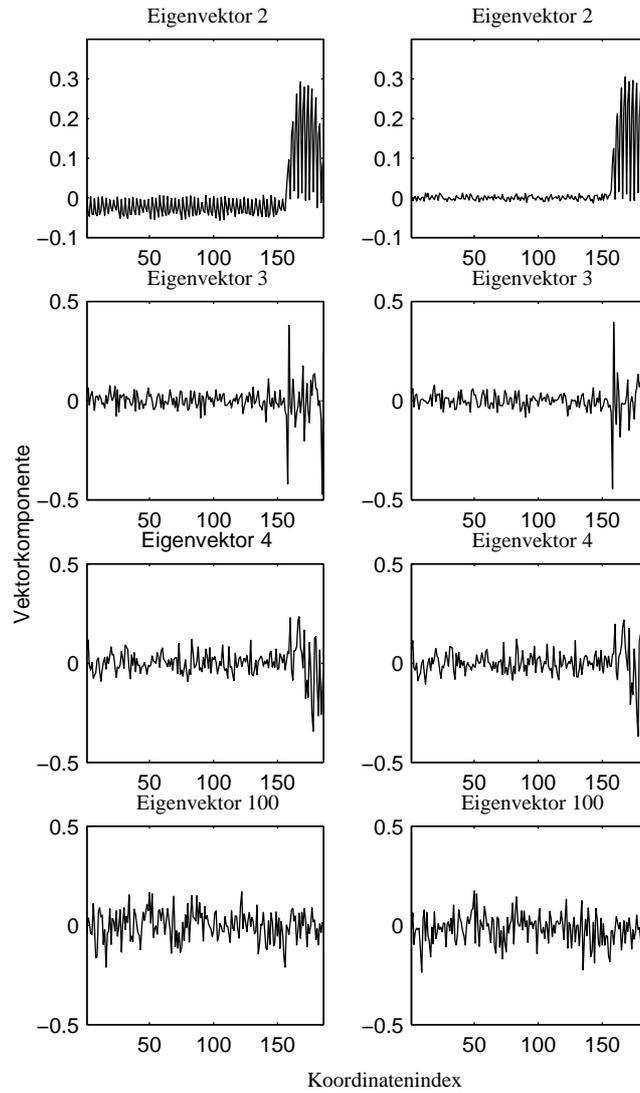


Abbildung 16: Modell 1: Komponenten der Moden 2, 3, 4 und 100. Die linke Spalte zeigt die internen Moden wie in Abbildung 15 (a) für den ersten Eigenvektor. Die rechte Spalte zeigt die entsprechenden vorurteilsfreien kartesischen Moden wie in Abbildung 15 (b) für den ersten Eigenvektor.

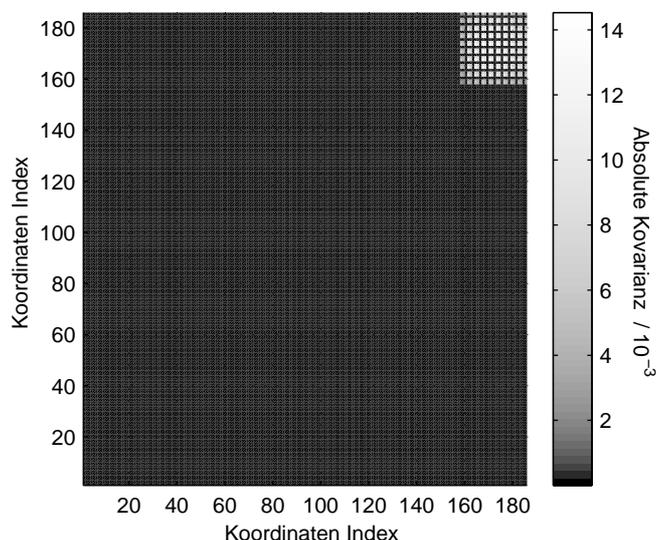


Abbildung 17: Modell 1: Absolute Werte der Kovarianzmatrix, die aus den orthogonalisierten vorurteilsfreien Moden mit Gleichung 79 berechnet wurde.

Die Eigenwerte der orthogonalisierten vorurteilsfreien Vektoren im kartesischen Raum wurden mit Hilfe der Gleichungen 105 und 109 (s. Anhang 5.1) berechnet. Die Kovarianzmatrix, die den orthogonalisierten Moden entspricht, ergibt sich mit Gleichung 79. Abbildung 17 zeigt ein Intensitätsdiagramm der Beträge dieser Kovarianzmatrix. Nahezu alle artifiziellen Kovarianzen, die in Abbildung 14 auftreten, wurden eliminiert. Lediglich die Korrelationen innerhalb der Schlaufe bleibt erhalten. Offensichtlich liefert der entwickelte Algorithmus kartesische Moden, die frei von artifiziellen Korrelationen sind. Ein zweites Modell (Modell 2) wurde entwickelt, um zu überprüfen, ob Korrelationen zwischen Bewegungen von weit entfernten Atomen mit vorurteilsfreien Moden immer noch detektierbar bleiben. Dazu wurde eine konzertierte Bewegung geringer Amplitude in Schlaufe B (Atom 33 bis 36, Koordinaten Index 97-108) auf der gegenüberliegenden Seite von Schlaufe A eingeführt. Abbildung 18 zeigt die Strukturen von der Seite des Würfels aus gesehen. Abbildung 19 zeigt die absoluten Kovarianzen unter Berücksichtigung aller Atome als Referenzsystem (a) und die absoluten Kovarianzen aufgrund der vorurteilsfreien Vektoren (b).

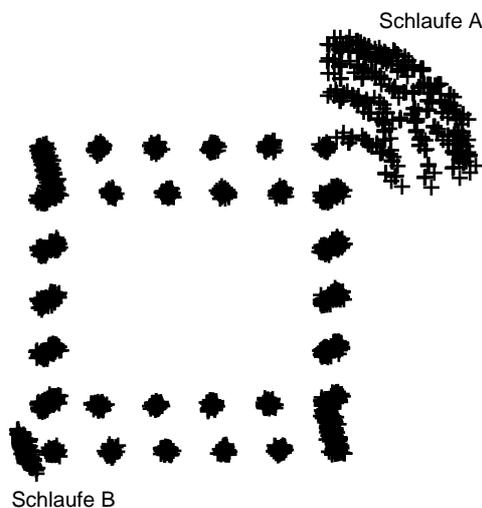


Abbildung 18: Modell 2: Dargestellt sind die 1000 Strukturen des Modells nach Überlagerung des rigiden Teils der Kuben. Es treten zwei zueinander korrelierte Bewegungen unterschiedlicher Amplitude in den Schlaufen A und B auf.

Die vorurteilsfreie Matrix in (b) zeigt sowohl die Kovarianzen in Schlaufe A und Schlaufe B, als auch die Kovarianzen zwischen den beiden Regionen. Folglich eliminiert die Methode der vorurteilsfreien Moden keine Korrelationen über lange Distanzen, sondern verbessert die Detektierbarkeit von kleinen Kovarianzen im Vergleich zu einer EDA.

3.3.7 Anwendung auf ein reales Molekül: RNase T_1

Das beschriebene Verfahren wurde auf die MD-Simulation der RNase T_1 angewendet. Für die Analyse der Moden wurden ausschließlich die 104 C_α -Atome berücksichtigt. Abbildung 20(a) zeigt die mittleren absoluten Kovarianzen der kartesischen Koordinaten aller C_α -Atome wie sie sich mit Hilfe von Gleichung 25 nach Überlagerung auf die mittlere Struktur ergeben. Diese Kovarianzmatrix wurde diagonalisiert und die so berechneten Eigenvektoren wurden in orthonormierte vorurteilsfreie Vektoren überführt. Die entsprechende Kovarianzmatrix ist in Abbildung 20(b) dargestellt. Abbildungen 20(c) und (d) zeigen vorurteilsfreie Kovarianzen ausgehend von der PCA-Analyse im Distanzraum, wobei die Struktur mit dem geringsten (c) und mit dem höchsten (d) RMSD zur mitt-

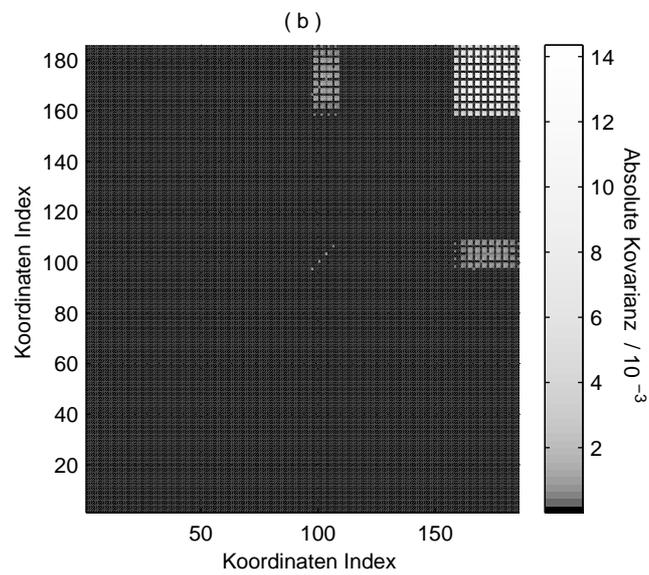
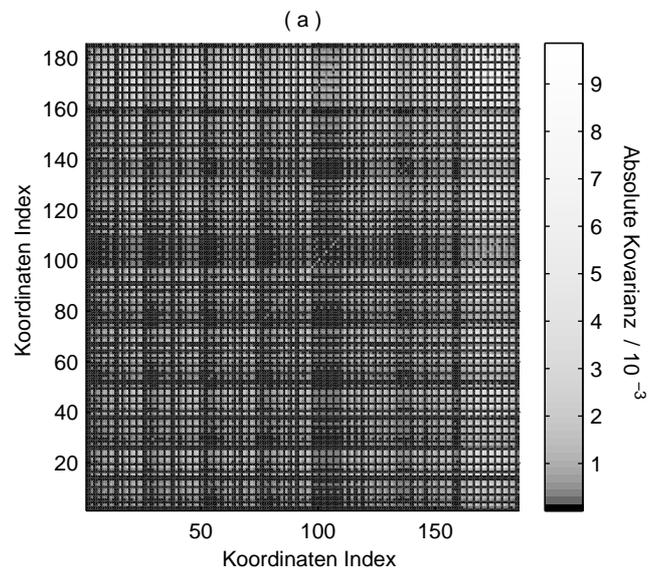


Abbildung 19: Modell 2: Absolute Kovarianzen, die sich auf der Überlagerung aller Atome (a) und den vorurteilsfreien Moden (b) gründen.

leren Konformation als Referenzstruktur verwendet wurde. Nur die ersten 100 Distanzvektoren wurden für Abbildungen (c) und (d) berücksichtigt, da der Rechenaufwand zur Berechnung der Eigenvektoren sehr groß ist (etwa 14 Stunden auf einem Athlon 650 MHz pro Eigenvektor). Die Abbildungen 20(b), (c) und (d) sind fast identisch. Sie zeigen schärfere Konturen als die anfängliche Matrix in (a). Einige Signale verschwinden vollständig. Der Einfluß der Referenzstruktur ist beeindruckend klein. Die entwickelte Methode führt zu sehr ähnlichen angepassten Kovarianzen, unabhängig vom anfänglichen Datensatz. In Abbildung 21 werden die EDA-Moden mit ihren entsprechenden vorurteilsfreien Moden verglichen. Wie schon für die Modellsysteme beobachtet, nimmt die Amplitude der ohnehin hohen Komponenten zu und die Amplituden der kleinen “rauschähnlichen” Signale nehmen ab, wenn die EDA-Moden in vorurteilsfreie Moden überführt werden.

Die Kreuzkorrelation $c(k, l)$ zwischen zwei Atomen k und l kann mit Hilfe von Gleichung 81 beschrieben werden [37].

$$c(k, l) = \frac{\langle (\vec{x}_k - \langle \vec{x}_k \rangle) \cdot (\vec{x}_l - \langle \vec{x}_l \rangle) \rangle}{\sqrt{\langle (\vec{x}_k - \langle \vec{x}_k \rangle)^2 \rangle \cdot \langle (\vec{x}_l - \langle \vec{x}_l \rangle)^2 \rangle}} \quad (81)$$

Abbildung 22 zeigt die atomaren Kreuzkorrelationen und mittleren Kreuzkorrelationen als Funktion der interatomaren Abstände.

Wie Ichiye und Karplus bereits für BPTI beobachtet haben [37], steigt auch für RNase T_1 die Kreuzkorrelationen aus der EDA für große Distanzen in (a) und (c) an. Dieser Effekt wird für vorurteilsfreie Korrelationen nicht beobachtet. Hier fällt die mittlere Korrelation monoton, was im Widerspruch zu der Interpretation von Ichiye und Karplus steht [37]. Die negative Korrelation wird durch Änderungen im Gyrationradius des Proteins hervorgerufen. Wenn sich der Radius des Proteins ändert, scheinen sich die Atome auf unterschiedlichen Seiten der Proteinoberfläche in entgegengesetzter Richtung zu bewegen. Ein korreliertes “Atmen” aller Atome wird jedoch nicht unterschieden von einer “Atmung”,

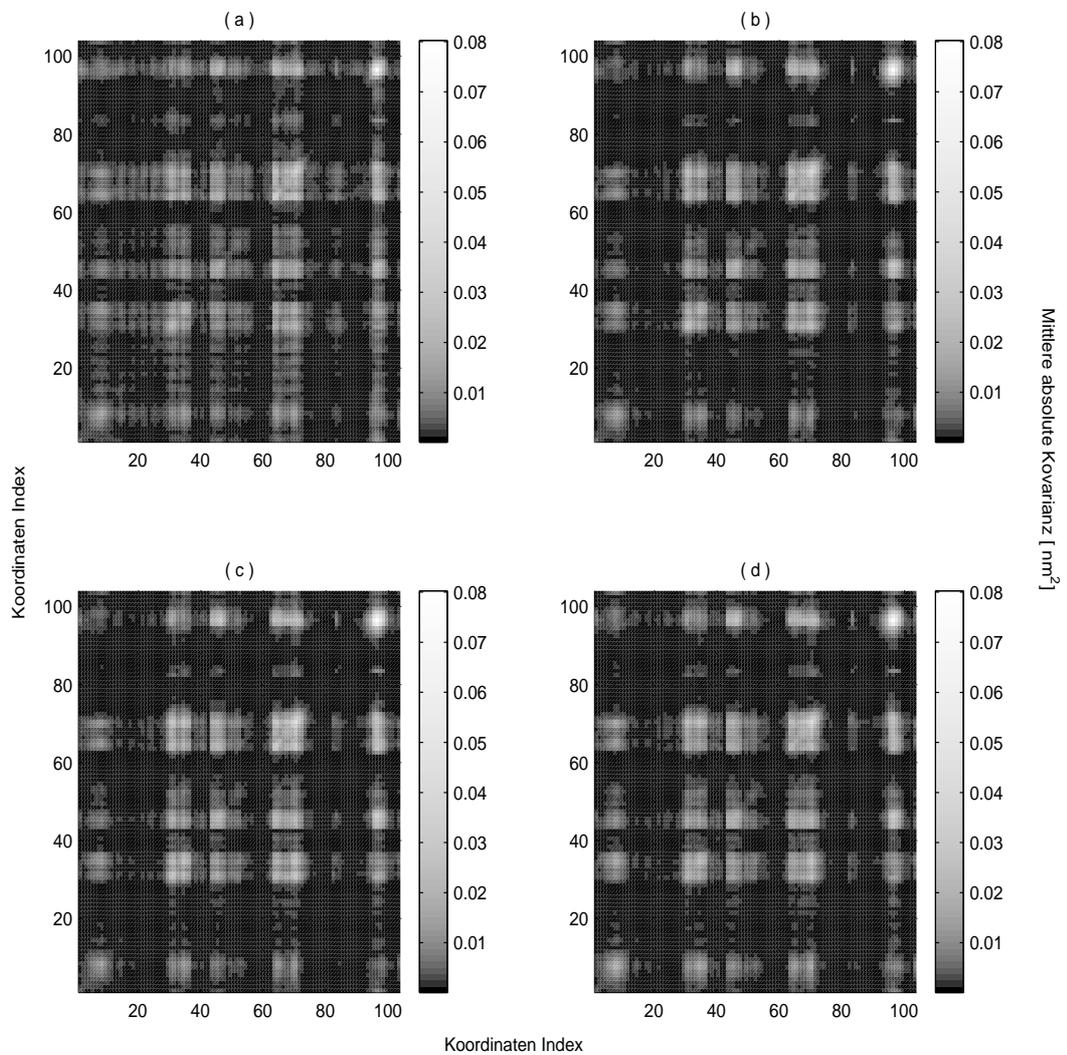


Abbildung 20: Mittlere absolute Kovarianzen der kartesischen Koordinaten aller C_{α} -Atome berechnet aus der 13 ns MD-Simulation an RNase T_1 . Die anfängliche Kovarianzmatrix ist in (a) abgebildet, und die zurückgerechneten vorurteilsfreien Kovarianzen aus der EDA sind (b) zu entnehmen. Die Diagramme (c) und (d) zeigen die vorurteilsfreien Kovarianzen, die man unter Verwendung der PCA-Analyse im Distanzraum erhält. Für (c) wurde die Struktur mit dem geringsten und für (d) die Struktur mit dem höchsten RMSD zur mittleren Konformation als Referenzstruktur für die Transformation in den kartesischen Raum benutzt.

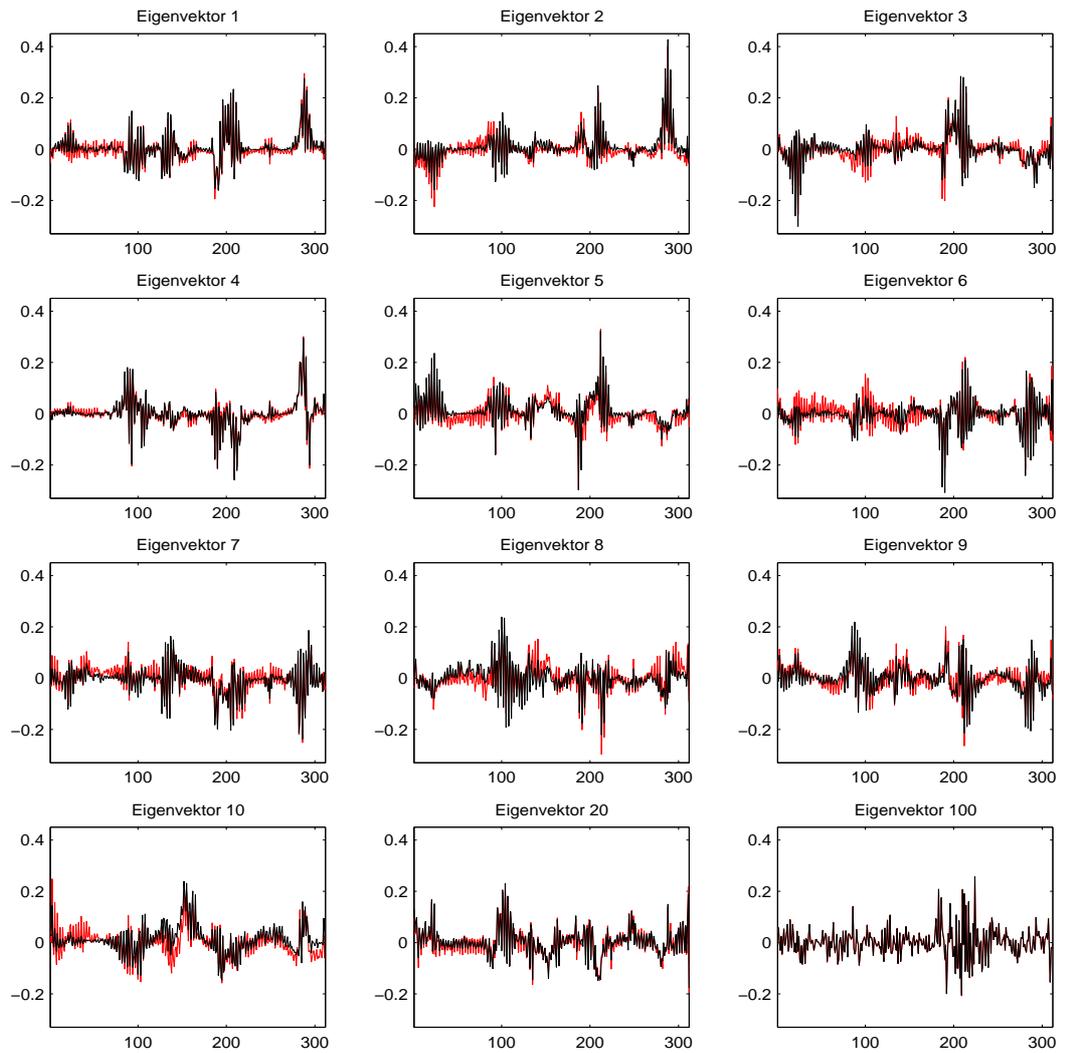


Abbildung 21: RNase T_1 : Vergleich der EDA-Vektoren (rot) und der vorurteilsfreien Vektoren (schwarz). Der größte Effekt der Manipulation ist für die ersten Eigenvektoren mit den dazugehörigen größten Eigenwerten zu erkennen.

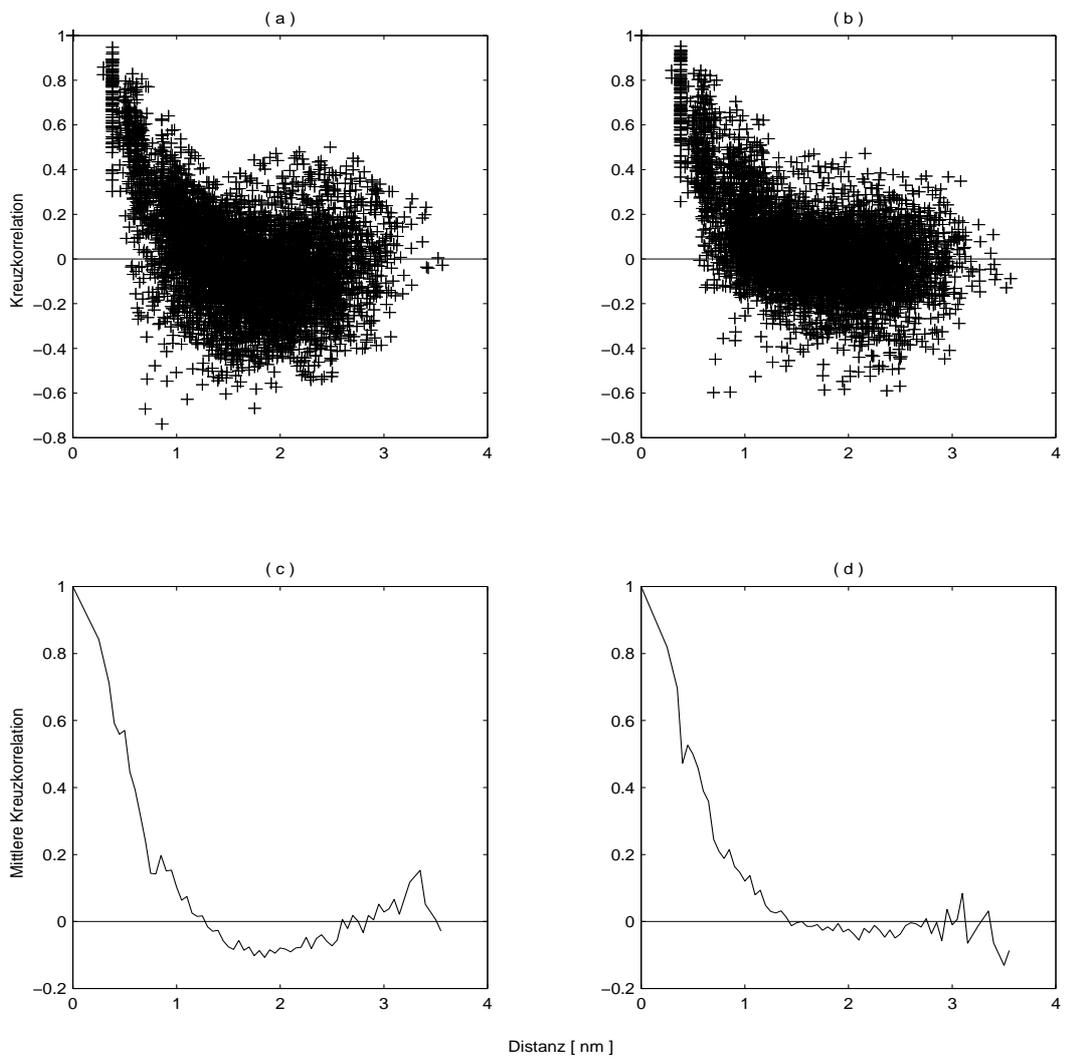


Abbildung 22: Kreuzkorrelation und mittlere Kreuzkorrelation von RNase T_1 als Funktion der interatomaren Distanz. Die Korrelationen unter Berücksichtigung aller Atome für die Überlagerung sind in (a) und (c) dargestellt. (b) und (d) zeigt die vorurteilsfreien Korrelationen.

die nur in einer Hälfte des Proteins stattfindet. Negative Korrelationen zwischen weit entfernten Atomen würden in beiden Fällen berechnet werden. Der Anstieg der EDA-Kreuzkorrelationen in (a) und (c) ist ein Artefakt des Translationsteils des Überlagerungsalgorithmus'. Die Null-Translationsbedingung für den Schwerpunkt führt zu artifiziellen, gleichzeitigen Verschiebungen von Atomen, die die positiven Korrelationen für Atome auf unterschiedlichen Seiten des Moleküls bewirken.

Die PCA-Analyse im Distanzraum für dieselbe Simulation wurde durchgeführt, um zu überprüfen, ob die vorurteilsfreien Vektoren unabhängig von der Referenzstruktur reproduzierbar sind. Die Ergebnisse der korrigierten Kovarianzmatrix sind bereits anhand Abbildung 20 diskutiert worden. Die Verwendung der Struktur mit dem höchsten und der Struktur mit dem niedrigsten RMSD zur mittleren Struktur führte zu nahezu identischen Ergebnissen. In Abbildung 23 werden die vorurteilsfreien Vektoren ausgehend von den Distanzvektoren mit denen ausgehend von der EDA verglichen. Die beiden Datensätze zeigen eine hohe Korrelation. Das bedeutet, daß die vorurteilsfreien Moden unabhängig von der Art der Koordinaten, die für die Matrixdiagonalisierung verwendet werden, reproduzierbar sind.

3.4 Diederwinkelanalyse

Eine alternative Möglichkeit zur Beschreibung der internen Dynamik ist die Verwendung von Diederwinkeln (lokale interne Koordinaten). Durch die Verwendung dieser generalisierten Variablen reduziert sich verglichen mit dem kartesischen Raum die Anzahl der für die Beschreibung einer Konformation notwendigen Variablen. Die Rückgratkonformation eines Aminosäurerestes wird durch die beiden Winkel ϕ und ψ beschrieben, während 12 kartesische Koordinaten für dieselbe Information benötigt werden. Diederwinkel sind aber lokale Strukturinformation, die nicht zwingend mit den globalen Konformationsumwandlungen korrelieren müssen. Wie in dieser Arbeit gezeigt wird, ist es in vielen Fällen dennoch möglich, das Verhalten einzelner Diederwinkel auf ein harmonisches

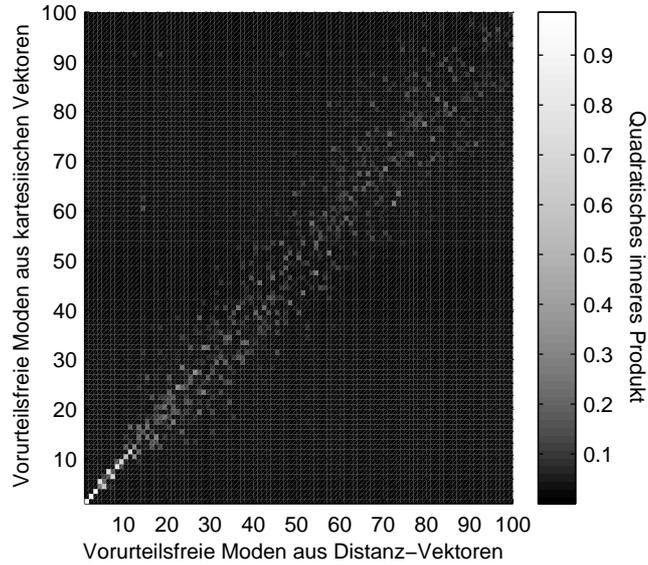


Abbildung 23: RNase T_1 : Quadratisches inneres Produkt der vorurteilsfreien Vektoren aus Distanzmoden unter Verwendung der mittleren Struktur als Referenzstruktur und den vorurteilsfreien Vektoren, die aus einer EDA berechnet wurden.

Potential zurückzuführen.

Im Rahmen dieser Arbeit wurden die Verteilungen aller Diederwinkel des Proteinrückgrates von RNase T_1 sowie die Diederwinkel der Seitenketten χ_1 auf jeweils möglichst einfache harmonische Modelle zurückgeführt.

Das einfachste Modell ist die Darstellung eines Diederwinkels als harmonischer Oszillator. Die entsprechende Definition der freien Energie ΔG als Funktion des Diederwinkels θ ist eine Parabel mit der Kraftkonstante K_θ als Streckfaktor und dem mittleren Winkel $\langle \theta \rangle$ als Symmetrieachse.

$$\Delta G(\theta) = \frac{1}{2} \cdot K_\theta \cdot (\theta - \langle \theta \rangle)^2 \quad (82)$$

Für ein kanonisches Strukturensamble wird die Wahrscheinlichkeitsdichte $\rho(\theta)$ mit Gleichung 83 beschrieben [164–166].

$$\rho(\theta) = \text{const} \cdot e^{-\frac{\Delta G(\theta)}{k \cdot T}} \quad (83)$$

Gleichung 84 ergibt sich aus den Gleichungen 82 und 83.

$$\rho(\theta) = \text{const} \cdot e^{-\frac{K_{\theta} \cdot (\theta - \langle \theta \rangle)^2}{2 \cdot k \cdot T}} \quad (84)$$

Das Integral über die Wahrscheinlichkeitsdichten ρ in allen möglichen Werten für θ muß 1 ergeben.

$$\int \rho(\theta) d\theta = 1 \quad (85)$$

Durch Verwendung der Normierung aus Gleichung 85 ergeben sich aus Gleichung 84 Gauß'sche Glockenkurven (Gleichungen 86 und 87).

$$\rho(\theta) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(\theta - \langle \theta \rangle)^2}{2 \cdot \sigma^2}} \quad (86)$$

$$\sigma = \sqrt{\frac{k \cdot T}{K_{\theta}}} \quad (87)$$

Bei der Betrachtung von Diederwinkeln tritt das Problem auf, daß die Skala der Winkel zyklisch ist. Dieser Eigenschaft wird man gerecht, ohne umständliche mathematische Definitionen aufstellen zu müssen, indem die theoretische Funktion in dem Intervall zwischen -540° und $+540^{\circ}$ berechnet wird. Die Funktionswerte der drei Perioden werden addiert und dem eigentlichen Intervall zwischen -180° und $+180^{\circ}$ zugeordnet. Der durch das Abbrechen der Funktion eingeführte Fehler ist bei den betrachteten Systemen vernachlässigbar klein, wie im weiteren gezeigt wird. Die mathematische Umsetzung der Vereinfachung wird durch Gleichung 88 beschrieben.

$$\rho(\theta) = \sum_{i=-1}^1 \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(\theta + i \cdot 2\pi - \langle \theta \rangle)^2}{2 \cdot \sigma^2}} \quad (88)$$

In dem Fall, daß mehrere Populationen in einem Diederwinkel auftreten oder daß die Häufigkeitsverteilung nicht einer Gauß'schen Verteilungsfunktion entspricht, wird die Näherungsfunktion durch die Summe von n Gauß-Funktionen

mit Gleichung 89 definiert.

$$\rho(\theta) = \sum_{k=1}^n \frac{A_k}{\sqrt{2\pi}\sigma_k} e^{-\frac{(\theta - \langle \theta_k \rangle)^2}{2 \cdot \sigma_k^2}} \quad (89)$$

Die entsprechende vereinfachte Funktion ist Gleichung 90 zu entnehmen.

$$\rho(\theta) = \sum_{k=1}^n \sum_{l=-1}^1 \frac{A_k}{\sqrt{2\pi}\sigma_k} e^{-\frac{(\theta + l \cdot 2\pi - \langle \theta_k \rangle)^2}{2 \cdot \sigma_k^2}} \quad (90)$$

Die Gewichtungsfaktoren A_k der einzelnen Gauß-Funktionen müssen so gewählt sein, daß die Summe aller Wahrscheinlichkeitsdichten 1 ergibt:

$$\sum_{k=1}^n A_k \approx 1 \quad (91)$$

Es treten Abweichungen von Gleichung 91 auf, weil das Integral über die Gauß-Kurven von $-\infty$ bis ∞ Eins ergibt, während in dem zu Grunde liegenden Modell lediglich ein Ausschnitt berücksichtigt wurde. Daher wurde auf die exakte Erfüllung von Gleichung 91 zu Gunsten einer exakteren Funktion verzichtet.

Der Potentialverlauf in Gleichung 92 ergibt sich durch Gleichsetzen der Ausdrücke aus den Gleichungen 83 und 90 nach Vereinfachung.

$$V(\theta) = -k \cdot T \cdot \ln \left[\sum_{k=1}^n \sum_{l=-1}^1 \frac{A_k}{\sqrt{2\pi}\sigma_k} \cdot e^{-\frac{(\theta + l \cdot 2\pi - \langle \theta_k \rangle)^2}{2 \cdot \sigma_k^2}} \right] \quad (92)$$

Die Form des Potentials in Gleichung 92 ist zwar nicht mehr harmonisch, sie geht aber für ausreichend separierte Kurven in harmonische Potentiale über (s. Abbildung 25 Seite 80).

Die Verteilungsfunktionen, die sich aus der Molekulardynamik-Simulation ergeben, werden mit Hilfe genau einer Gauß'schen Verteilungsfunktionen angenähert, wenn die mittlere quadratische Abweichung T_{Gauss} von Näherungsfunktion $\rho_{harmonisch}$ und tatsächlicher Verteilung $\rho_{Simulation}$ unter einen empirisch festgelegten Wert von 10^{-4} liegt. Wenn die Abweichung größer ist, dann wird die Anzahl der Gauß-Kurven sukzessiv erhöht bis der Grenzwert unterschritten

wird.

$$T_{Gauss} = \left\langle (\rho_{Simulation}(\theta) - \rho_{harmonisch}(\theta))^2 \right\rangle \quad (93)$$

Die Abbildung 24 zeigt die Verteilungsfunktionen von vier exemplarischen Diederwinkeln. Die aus der Trajektorie berechnete Verteilung ist in blau dargestellt. Die mit Hilfe von Gleichung 88 bzw. 90 berechneten Wahrscheinlichkeitsdichten sind in rot abgebildet. Die Verteilung von ϕ_{22} (links oben) läßt sich mit Hilfe eines einzigen harmonischen Potentials darstellen. ψ_{64} (rechts oben) zeigt zwei separierte Gauß-verteilte Populationen. Die Einstellungen von ϕ_{65} (links unten) entsprechen auf den ersten Blick zwar auch zwei Populationen, die Verteilung der positiven Winkel läßt sich aber erst durch ein weiteres Potential annähern. Die Wahrscheinlichkeitsdichte des Diederwinkels ψ_{70} ist nicht Gauß-verteilt. Man benötigt fünf harmonische Potentiale, um den Verlauf anzunähern.

Die Potentiale, die gemäß Gleichung 92 den Verläufen aus Abbildung 24 entsprechenden, sind der Abbildung 25 in rot zu entnehmen. Der Verlauf der gestrichelten schwarzen Kurven stellen die harmonischen Potentiale dar, die den einzelnen Populationen entsprechen. Im einfachsten Fall - wie bei ϕ_{22} - verläuft das Potential über den gesamten Winkelbereich parabolisch. Zwei parabolisch geformte "Potentialtöpfe" sind für ψ_{64} zu erkennen. Das Potential der Population um -79° für ϕ_{65} ist ebenfalls parabolisch. Die zweite Population, die durch zwei harmonische Potentiale beschrieben wird, entspricht lediglich für Winkel, die größer als ca. 120° sind, einem harmonischen Potential. Das effektive Potential von ψ_{70} , hat nur sehr geringe harmonische Anteile. Die Ergebnisse der Diederwinkelbetrachtung ist detailliert den Tabellen im Anhang Seite 117 ff. zu entnehmen. Die Winkelverteilung von 102 der insgesamt 206 Diederwinkel des Proteinrückgrats entspricht einem einfachen harmonischen Potential. Weitere 69 Winkel weisen zwei harmonische Potentiale auf, während nur 33 Winkel 3 überlagerte Parabeln benötigen. Die Winkelverteilung von nur 2 Diederwinkeln läßt sich erst mit mehr als drei harmonischen Potentialen nachvollziehen.

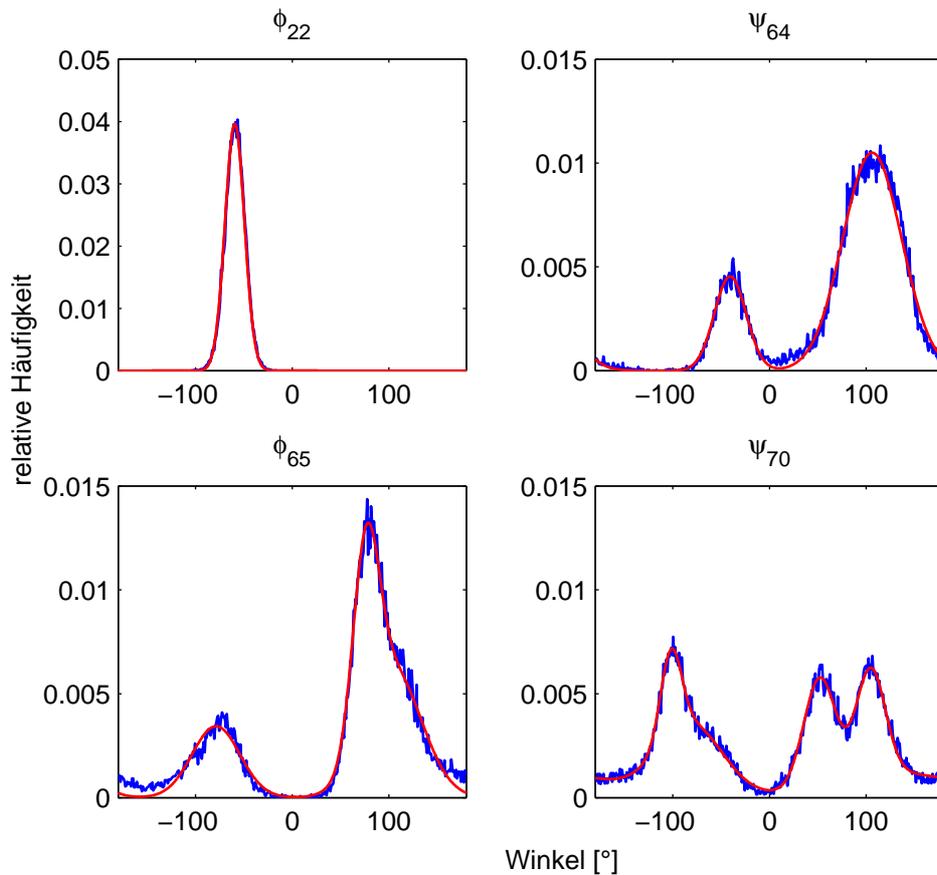


Abbildung 24: Verteilungsfunktionen von Diederwinkeln, die durch 1, 2, 3 bzw. 5 überlagerte harmonische Potentiale beschrieben werden können. Die aus der Trajektorie berechnete Verteilung ist in blau dargestellt. Die mit Hilfe von Gleichung 88 bzw. 90 berechneten Wahrscheinlichkeitsdichten sind in rot abgebildet. Die Diederwinkel sind mit den entsprechenden Aminosäureresten indiziert.

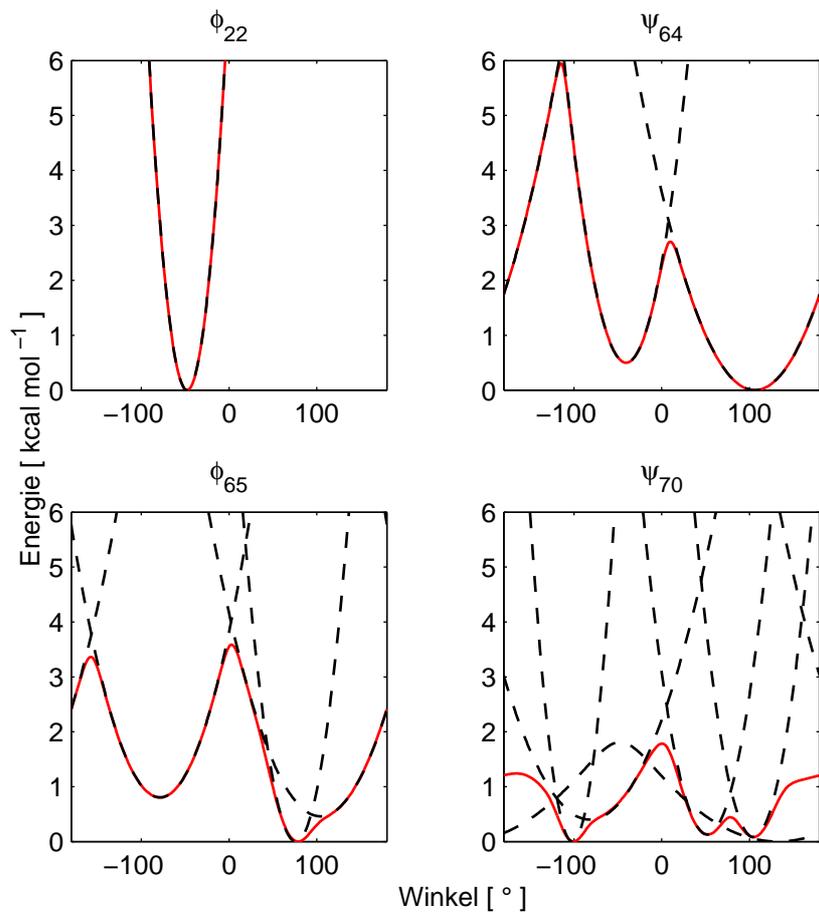


Abbildung 25: Potentiale, die gemäß Gleichung 92 den Verläufen aus Abbildung 24 entsprechen. In rot ist der Potentialverlauf dargestellt, die gestrichelten Linien entsprechen den harmonischen Verläufen, aus denen das Potential zusammengesetzt ist.

Abbildung 26 zeigt das Proteinrückgrat der mittleren Struktur der Simulation von RNase T_1 . Grün eingefärbte Bereiche zeigen Aminosäurereste mit harmonischem Potential in ϕ bzw. ψ . Die Bereiche, für die zwei harmonische Potentiale benötigt werden, sind gelb; für drei und mehr als drei Potentiale sind die Bereiche rot und blau eingefärbt. Die komplette α -Helix zeigt einfache harmonische Potentiale. Auch im restlichen Proteinrückgrat sind die freien Energien erstaunlich gut durch harmonische Potentialverläufe zu beschreiben. Für nur 7 Reste (6,7%) ist weder der Energieverlauf des ϕ noch des ψ -Winkel durch höchstens zwei überlagerte harmonische Potentiale darstellbar. Für 73% aller Reste können beide Rückgratdiederwinkel durch höchstens zwei harmonische Potentiale beschrieben werden.

Abbildung 27 zeigt eine schematische Darstellung von RNase T_1 . Die Sekundärstrukturelemente sind rot (α -Helix) und gelb (β -Faltblatt) gekennzeichnet. Hellblaue Reste kennzeichnen Aminosäurereste, von denen genau ein Rückgratdiederwinkel durch mehr als zwei harmonische Potentiale zu beschreiben ist. Dunkelblaue Reste benötigen mindestens drei harmonische Potentiale in beiden Diederwinkeln. Der Fluktuation der Diederwinkel aller Reste innerhalb von Sekundärstrukturelementen außer TYR42 liegen maximal zwei harmonische Potentiale zu Grunde. Viele Anharmonizitäten treten in unmittelbarer Nachbarschaft zu den Sekundärstrukturelementen auf. Solche Reste erfüllen eine "Scharnier-Funktion" für die sich anschließende Schlaufe, wie in 31, 37, 42, 51-53, 62, 73-74 oder 94. Die Potentiale der Reste 7 und 103 werden vermutlich durch die Disulfidbrücke zwischen 6 und 103 verformt. Es ist also möglich, einen beträchtlichen Anteil der Potentiale der Rückgratdiederwinkel durch sehr einfache differenzierbare Funktionen wiederzugeben. Diese Potentiale beschreiben die dynamische Struktur eines Proteins mit Hilfe von lokalen Strukturinformationen, was die Möglichkeit zur Analyse der Dynamik ganz bestimmter Bereiche im Protein vereinfacht. Die Information über den korrelierten Charakter der Bewegung im Protein geht allerdings verloren, könnte aber durch eine EDA ergänzt werden.

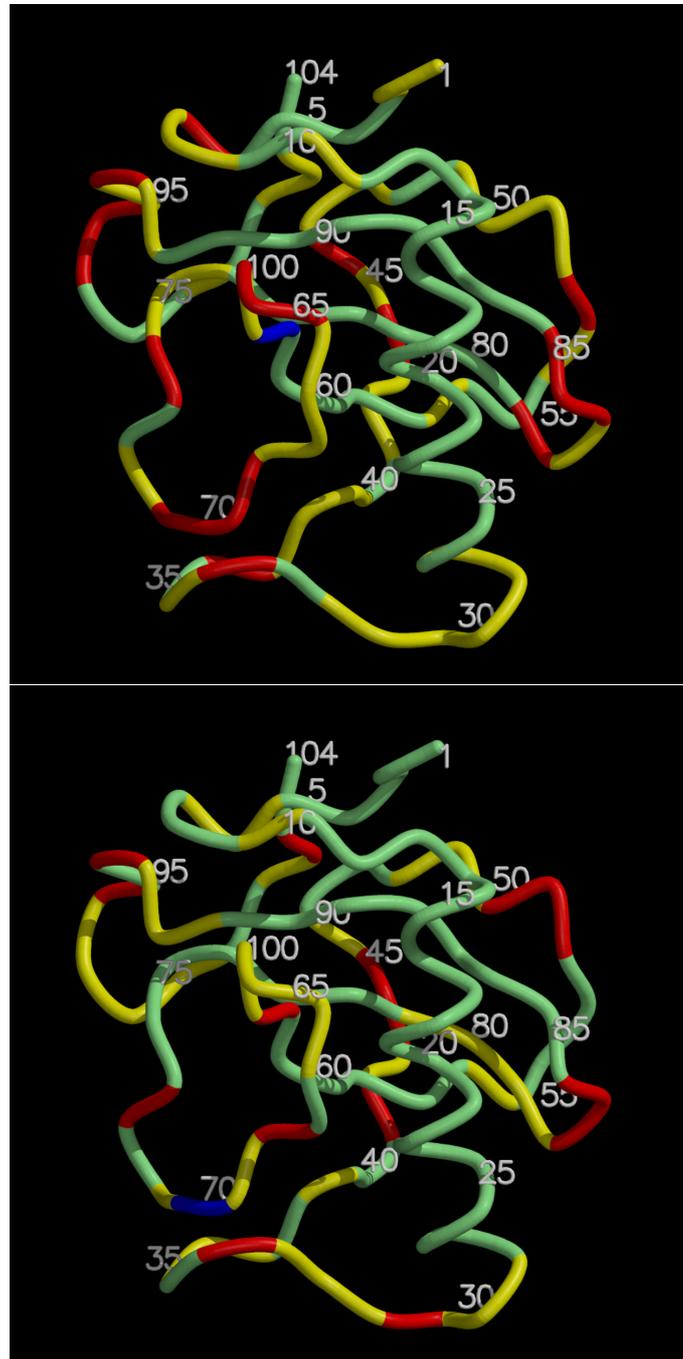


Abbildung 26: Proteinrückgrat von RNase T_1 in der mittleren Konformation der Simulation. Grün eingefärbte Bereiche zeigen Aminosäurereste mit harmonischem Potential in ϕ (oben) bzw. ψ (unten). Die Bereiche, für die zwei harmonische Potentiale benötigt werden, sind gelb; für drei und mehr als drei Potentiale sind die Bereiche rot und blau eingefärbt.

Ein Anwendungsgebiet der Potentialfunktionen ist die MD im Diederwinkelraum. Wenn man die freie Energie der Diederwinkel relativ genau kennt, dann müßte es auch möglich sein mit Hilfe von vereinfachten Kraftfeldern die Potentialhyperfläche der Konformation der Proteine schneller abzutasten, als dies mit konventioneller MD möglich ist. Es wird Bestandteil weiterer Forschung sein, diese Vermutung näher zu untersuchen.

3.5 Vergleich mit anderen Methoden

3.5.1 Ordnungsparameter

Aus dem strukturellen Ensemble der MD-Simulation lassen sich mit den Gleichungen 48 und 50 Ordnungsparameter S^2 berechnen, die auch mit Hilfe der NMR-Spektroskopie zugänglich sind [152, 155]. Die S^2 -Werte wurden für die $N-H$ - und für die $C_\alpha-H_\alpha$ - Bindungsvektoren berechnet und mit den Ergebnissen aus NMR-Relaxationsexperimenten [33, 167] verglichen. Die Abbildungen 28 und 29 zeigen die $1-S^2$ -Werte für die $C_\alpha-H_\alpha$ - bzw. $N-H$ -Bindungsvektoren. Rot unterlegt ist das Ergebnis der NMR-Relaxationsexperimente in seinen Fehlergrenzen. Schwarz sind die Ergebnisse aus der vorurteilsfrei überlagerten MD-Trajektorie dargestellt. Blau und grün sind die $1-S^2$ -Werte, die sich aus zurückgerechneten Trajektorien (Gleichung 59) ausgehend von der EDA ergeben. Die 20 (blau) bzw. 40 (grün) Eigenvektoren mit den höchsten Eigenwerten wurden bei der Berechnung nicht berücksichtigt. Man erkennt innerhalb der grau unterlegten Sekundärstrukturelemente gute Übereinstimmungen zwischen den $1-S^2$ -Werten aus MD und NMR in beiden Abbildungen.

Innerhalb der Schlaufen sind die $1-S^2$ -Werte aus der MD-Trajektorie höher, als die mit NMR-Spektroskopie bestimmten Werte. Für die Aminosäurereste 30-37, 43-49, 68-74, 82-84, 96-99 und 103-104 liegen die $1-S^2$ -Werte des $C_\alpha-H_\alpha$ -Vektors aus der MD-Trajektorie deutlich über den NMR-Werten. Durch Vernachlässigung der langsamen Moden der Trajektorie (hohe Eigenwerte) werden alle $1-S^2$ -Werte außer 62-65, 71, 72 und 96 annähernd in den Bereich

gebracht, der sich auch aus der NMR ergibt. Das ist ein Indiz dafür, daß die weitreichenden Veränderungen in der Struktur der Schlaufe nicht mit Hilfe der NMR- S^2 -Werte erfasst werden können, weil die entsprechenden Periodenlängen höher sind, als sie mit NMR-Relaxation nachgewiesen werden können (ca. 10^2 ns). Innerhalb der Trajektorie waren die Periodenlängen allerdings im oberen Pikosekunden bis unteren Nanosekundenbereich. Erklärbar wäre dies etwa durch zu steile Potentiale im GROMACS-Kraftfeld, was zu höheren Frequenzen und damit zu einer scheinbar beschleunigten Dynamik führt.

Die MD- $1 - S^2$ -Werte der $C_\alpha - H_\alpha$ -Bindungsvektoren stimmen insgesamt besser mit den NMR-Experimenten überein als die $1 - S^2$ -Werte der $N - H$ -Bindungsvektoren. Der erhöhte $1 - S^2$ -Wert von Rest 30 läßt sich weder durch die Betrachtung des $N - H$ -Vektors mittels NMR noch durch die Röntgenstrukturanalyse [15] bestätigen. Lediglich in der NMR-Messung am $C_\alpha - H_\alpha$ -Vektor des Restes 30 ist ebenfalls ein deutlicher Anstieg der $1 - S^2$ -Werte festzustellen. Die hohen $1 - S^2$ -Werte der Reste 63-66 und 84-85 werden nicht durch die NMR aber durch die Kristallstruktur [15] bestätigt. In den übrigen Bereichen stimmt zumindest die Tendenz der $1 - S^2$ -Werte in MD und NMR überein.

In den unteren Diagrammen der Abbildungen 28 und 29 sind jeweils die B-Faktoren der C_α - bzw. N - Atome in der Kristallstruktur [15] aufgetragen. In den Resten 71 und 72 bleiben die MD- $1 - S^2$ -Werte auch bei den zurückgerechneten Trajektorien deutlich höher als die NMR- $1 - S^2$ -Werte. Die B-Faktoren in dem Bereich sind allerdings relativ zu den umgebenen Resten erhöht, was auf auftretende Dynamik hinweist. Der qualitative Verlauf der B-Faktoren korreliert gut mit den MD- $1 - S^2$ -Werten, wenn man von dem Anstieg der B-Faktoren in den Resten 51 bis 53 absieht. In dem Bereich sind bereits in dem NMR-Strukturcluster zwei alternative Konformationen gefunden worden, die sich vermutlich in der Kristallstruktur bei der Interpretation als eine Konformationen in erhöhten B-Faktoren niederschlagen [31]. Da aber die auftretende Dynamik in MD-Simulationen ohnehin wesentlich größer ist als in der Kristallstruktur, schlagen sich die alternativen Rückgratkonformationen nicht in den S^2 -Werten

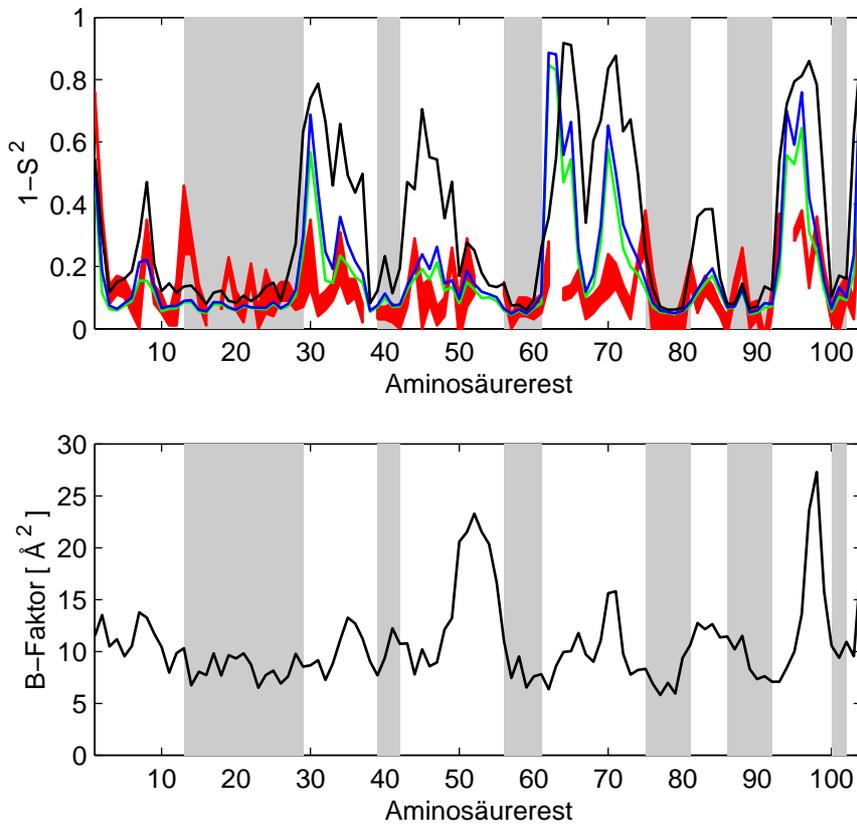


Abbildung 28: Oben: $1 - S^2$ -Werte für die $C_\alpha - H_\alpha$ -Bindungsvektoren. Die $1 - S^2$ -Werte aus der NMR-Spektroskopie sind in ihren Fehlergrenzen rot unterlegt. Schwarz sind die Ergebnisse aus der MD-Trajektorie dargestellt. Blau und grün sind die $1 - S^2$ -Werte, die sich aus zurückgerechneten Trajektorien ausgehend von der EDA ergeben. Die ersten 20 (blau) bzw. 40 (grün) Eigenvektoren wurden bei der Berechnung nicht berücksichtigt. Unten: B-Faktoren der C_α -Atome in der Kristallstruktur [15].

Innerhalb der grau unterlegten Sekundärstrukturelemente stimmen die $1 - S^2$ -Werte aus der Simulation und aus der NMR-Spektroskopie gut überein. Innerhalb der Schleifen sind die $1 - S^2$ -Werte aus der MD-Trajektorie höher als die mit NMR-Spektroskopie bestimmten Werte. Für die Aminosäurereste 30-37, 43-49, 68-74, 82-84, 96-99 und 103-104 liegen die $1 - S^2$ -Werte aus der Trajektorie deutlich über den NMR-Werten. Durch Vernachlässigung der langsamen Moden der Trajektorie (hohe Eigenwerte) werden alle $1 - S^2$ -Werte außer 62-65, 71, 72 und 96 annähernd in den Bereich gebracht, der sich auch aus der NMR ergibt. Der qualitative Verlauf der B-Faktoren korreliert gut mit den MD- $1 - S^2$ -Werten, wenn man von dem Anstieg der B-Faktoren in den Resten 51 bis 53 absieht.

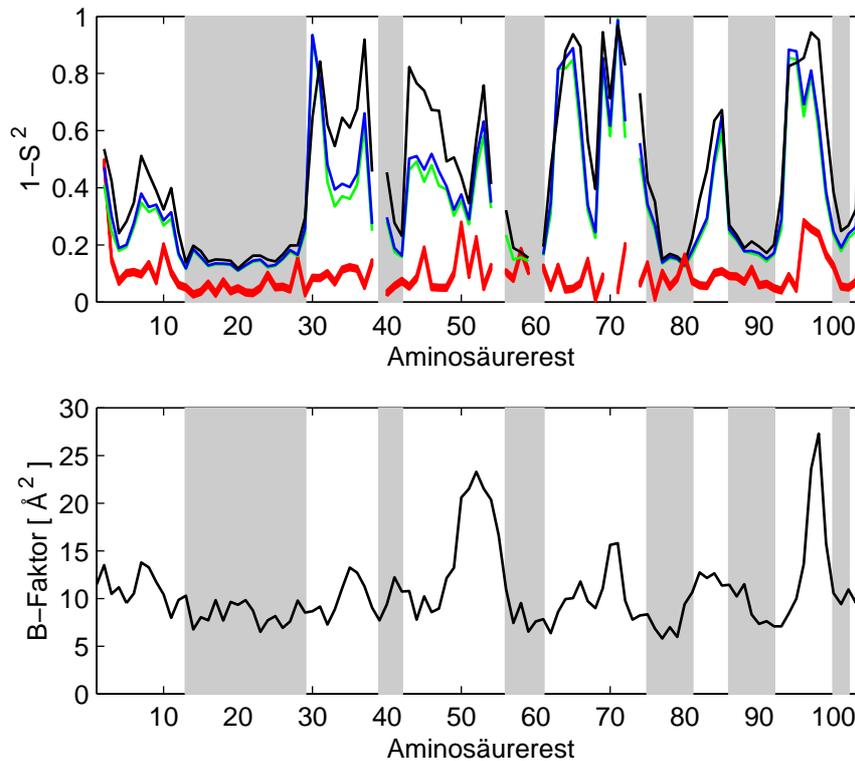


Abbildung 29: Oben: $1 - S^2$ -Werte für die $N - H$ -Bindungsvektoren. Die $1 - S^2$ -Werte aus der NMR-Spektroskopie sind in ihren Fehlergrenzen rot unterlegt. Schwarz sind die Ergebnisse aus der MD-Trajektorie dargestellt. Blau und rot sind die $1 - S^2$ -Werte, die sich aus zurückgerechneten Trajektorien ausgehend von der EDA ergeben. Die ersten 20 (blau) bzw. 40 (rot) Eigenvektoren wurden bei der Berechnung nicht berücksichtigt. Unten: B-Faktoren der Amidstickstoffatome in der Kristallstruktur [15].

Innerhalb der grau unterlegten Sekundärstrukturelemente stimmen die $1 - S^2$ -Werte aus der Simulation und aus der NMR-Spektroskopie gut überein. Innerhalb der Schleifen sind die $1 - S^2$ -Werte aus der MD-Trajektorie höher, als die mit NMR-Spektroskopie bestimmten Werte. Innerhalb der Schleifen sind die $1 - S^2$ -Werte aus der MD-Trajektorie deutlich höher als die mit NMR-Spektroskopie bestimmten Werte. Der qualitative Verlauf der B-Faktoren korreliert gut mit den MD- $1 - S^2$ -Werten, wenn man von dem Anstieg der B-Faktoren in den Resten 51 bis 53 absieht.

der Simulation nieder.

3.5.2 NMR Struktur-Cluster

Neben den Relaxationsparametern gibt auch das NMR-Lösungsstrukturensemble implizit dynamische Parameter wieder, die zum Vergleich mit den Ergebnissen von MD-Rechnungen herangezogen werden können [31].

Abbildung 30 zeigt die lokalen RMSD Werte der C_α -Atome aus dem NMR-Strukturensemble (a) und aus der MD-Rechnung (b). Die vorurteilsfreien Varianzen unterscheiden sich für beide Strukturensembles nur marginal von den Ergebnissen der klassischen Überlagerung. Der qualitative Verlauf der lokalen RMSD Werte stimmt für beide Ensembles recht gut überein. Die Werte aus der Distanzgeometrie liegen etwa eine Größenordnung niedriger als die Werte aus der MD.

Die DG-Rechnung (DG = **D**istanz**G**eometrie) zeigt im Gegensatz zur MD eine erhöhte Flexibilität zu Beginn der α -Helix in den Resten 13 und 14 sowie innerhalb des β -Faltblattes im Rest 88. Rest 29 bildet laut NMR-Lösungsstruktur und Röntgenstruktur das andere Ende der Helix. Die MD legt aber aufgrund des RMSD-Wertes nahe, Rest 29 bereits der sich anschließenden Schlaufenregion zuzuordnen. Die Flexibilität des Schlaufenbereichs der Reste 63 bis 65 ist in der MD-Trajektorie ebenfalls erhöht, was in den lokalen RMSD Werten der DG nicht beobachtet werden kann.

Die kartesische Kovarianzmatrix der MD Rechnung ist in Kapitel 3.3.7 bereits detailliert diskutiert worden. In Abbildung 31 sind die Kovarianzmatrizen des NMR-Strukturensembles mit klassischer (a) und vorurteilsfreier Überlagerung (b) dargestellt. Der Einfluß der vorurteilsfreien Superposition ist für das DG-Cluster geringer als für die MD-Rechnung (vgl. Abbildung 20). Korrelierte Konformationsunterschiede gehen aus dem DG-Cluster nur schlecht hervor. Das im Vergleich zu Molekulardynamik eher stark vereinfachte physikalische Kraftfeld der Distanzgeometrie wird von Abstandsbeschränkungen überlagert, was zu einem verzerrten Korrelationsmuster führt, wenn sich beispielsweise zwei

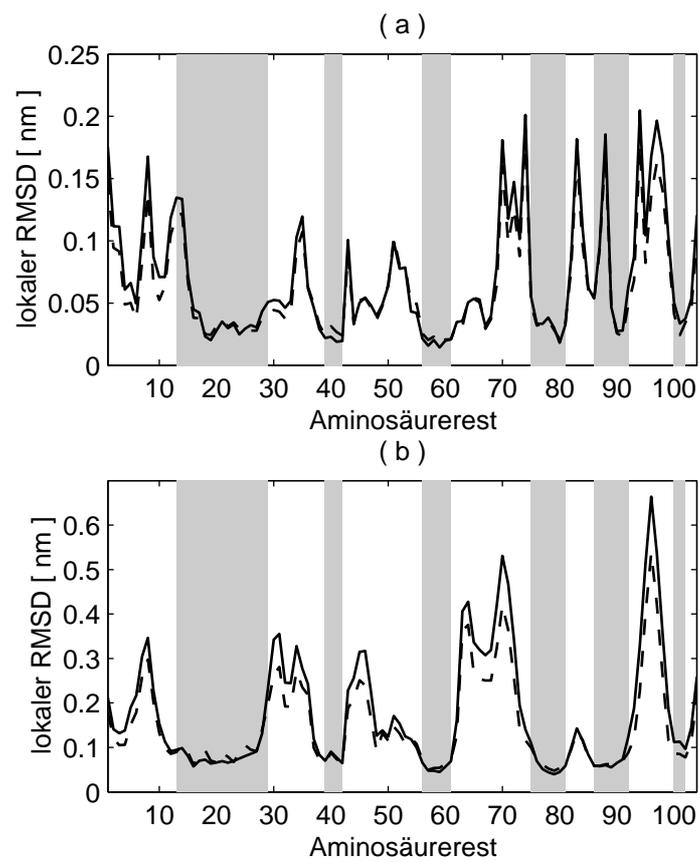


Abbildung 30: Lokale RMSD-Werte der C_{α} -Atome aus dem NMR-Strukturensemble (a) und aus der MD-Rechnung (b). Die gestrichelten Linien beschreiben die klassischen Varianzen, während die durchgezogenen Linien vorurteilsfreien Varianzen entsprechen.

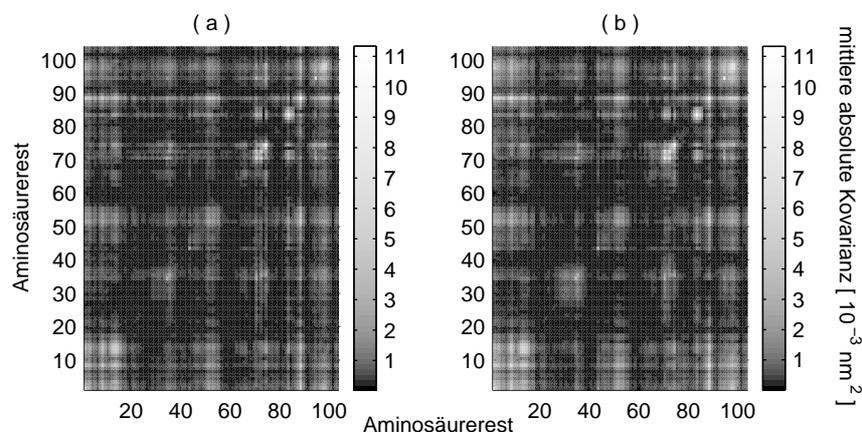


Abbildung 31: Mittlere absolute Kovarianz der kartesischen Koordinaten des NMR-Strukturensembles. Klassische Überlagerungen der Strukturen führen zu (a). Die Kovarianzen der zurückgerechneten vorurteilsfreien Koordinaten sind in (b) abgebildet.

Abstands-Restriktionen im 3-dimensionalen Raum gegenseitig ausschließen.

Die mittlere Verletzung der NOE-Abstandsbeschränkungen pro Aminosäure-rest ist Abbildung 32 zu entnehmen. Insgesamt werden die Abstands-Restriktionen vom MD-Cluster (unterbrochene Linie) stärker verletzt als von den DG-Strukturen (durchgezogene Linie). Das gilt besonders für die Schlaufenregionen, während über weite Bereiche der Sekundärstrukturelemente die Verletzungen etwa gleich groß sind.

Abbildung 33 zeigt, zwischen welchen Aminosäureresten Distanzbeschränkungen existieren, die nur vom den Strukturen einer der beiden Methoden verletzt werden. In Abbildung 33(a) sind die Paare gekennzeichnet, zwischen denen Abstandsbeschränkungen existieren, die von den DG-Strukturen weniger verletzt werden, als von den MD-Strukturen. Abbildung 33(b) zeigt den umgekehrten Fall. Die markierten Positionen in den beiden Diagrammen sind sehr ähnlich. Das ist ein Hinweis dafür, daß nicht selbstkonsistente Abstands-Restriktionen durch das genauere physikalische Kraftfeld der MD gegeneinander ausgetauscht wurden. In den DG-Rechnungen ist ein Teil des Kraftfeldes durch die Abstandsbeschränkungen bestimmt, was dazu führt daß die Beschränkungen weniger verletzt werden als in einer Rechnung, in deren Kraftfeld explizit

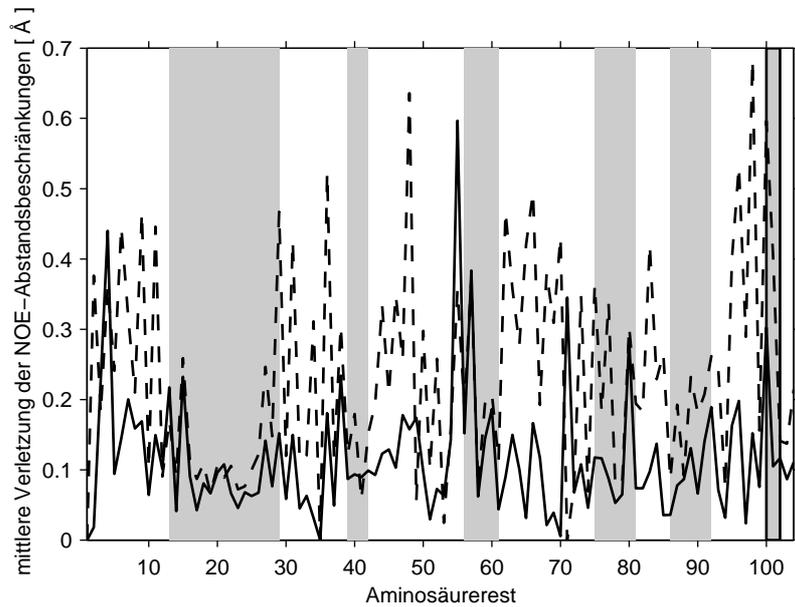


Abbildung 32: Mittlere Verletzung der NOE-Abstandsbeschränkungen pro Aminosäurerest. Die durchgezogene Linie beschreibt die Ergebnisse des Distanzgeometrie-Strukturensembles. Die gestrichelte Linie wurde aus den Daten der MD-Rechnung berechnet.

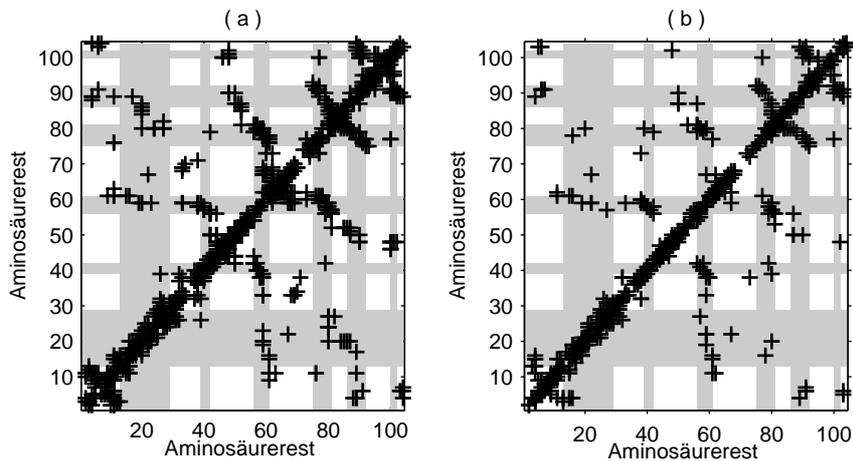


Abbildung 33: Vergleich der Verletzungen der NOE-Abstandsbeschränkungen. (a) zeigt, zwischen welchen Resten Abstandsbeschränkungen durch das MD Ensemble stärker verletzt werden als durch die DG-Strukturen. (b) zeigt den umgekehrten Fall.

keine artifiziiellen Abstandsenergien eingehen (Abbildung 32). Da die Abstands-Restriktionen in der MD-Simulation nicht wesentlich stärker verletzt werden und sich die im Sinne des MD-Kraftfeldes günstigeren Strukturen auch nicht wesentlich von den DG Strukturen unterscheiden (mittlerer RMSD = 3,5 Å) ist davon auszugehen, daß das MD-Strukturensemble auch im Sinne der NMR-Parameter als verfeinerte dynamische Struktur anzusehen ist.

4 Zusammenfassung

Im Rahmen der vorliegenden Arbeit wurde eine Molekulardynamik-Simulation am Modellprotein Ribonuklease T_1 durchgeführt. Der Vergleich simulierter Dynamik von RNase T_1 mit experimentell ermittelten Parametern zur Beschreibung interner Dynamik zeigt eine deutliche Übereinstimmung der Ergebnisse beider Methoden. Die aus der MD-Simulation berechneten Ordnungsparameter S^2 sind innerhalb der Sekundärstrukturelemente im Einklang mit den Ordnungsparametern aus NMR-Relaxationsexperimenten und den B-Faktoren aus der Kristallstruktur. Wie schon in anderen Untersuchungen gezeigt, ist die simulierte Dynamik der Schlaufenregionen größer als die experimentell beschriebene Dynamik. Dennoch sind die Verletzungen der Abstands-Restriktionen aus der NMR-Strukturbestimmung ausgehend von dem NMR-Strukturcluster und dem strukturellen Ensemble der Simulation vergleichbar. Das strukturelle Ensemble der Simulation diente als Modell für die Anwendung von Methoden, die im Rahmen dieser Arbeit entwickelt wurden.

Für die unvoreingenommene Trennung von interner und globaler Bewegung in Biomolekülen wurde ein neuer Algorithmus eingeführt. Das Verfahren besteht aus mehreren Schritten. Zunächst werden die Strukturen "klassisch" überlagert. Die daraus berechnete Kovarianzmatrix der atomaren Fluktuation wird diagonalisiert. Im Modenraum werden die globalen Anteile der Bewegung eliminiert. Schließlich werden die korrigierten Moden zu einer berichtigten Kovarianzmatrix rücktransformiert. Für die Eliminierung wird die "Gesamtmenge an Bewegung" (TAM) durch Addition von kompensierenden Moden, die reine Rotation/Translation beschreiben, minimiert. Es konnte gezeigt werden, daß die globalen Anteile der Bewegung eliminiert werden können. Der Algorithmus ist auf alle kartesische Kovarianzmatrizen anwendbar, unabhängig von der Methode ihrer Generierung. Insbesondere auch Kovarianzmatrizen aus Normalmodenanalysen können mit diesem Algorithmus korrigiert werden. Die vorurteilsfreie Trennung wurde auf die Kovarianzmatrizen aus der MD-Simulation an RNase

T_1 angewendet. Die entsprechenden Kovarianzmatrizen zeigen scharfe Konturen. Eine signifikante Anzahl von Kovarianzen aufgrund von globaler Bewegung konnte durch den Algorithmus eliminiert werden. Mit Hilfe der vorurteilsfreien Methode wurde zudem gezeigt, daß der zuerst 1991 von Ichiye und Karplus gefundene Anstieg der mittleren Korrelation für weit entfernte Atome eine Folge des Überlagerungsverfahrens ist.

Eine andere mögliche Anwendung der vorurteilsfreien Trennung ist die Identifizierung von rigiden Atomen. Dazu wird die anfängliche Mode durch die Differenzen der kartesischen Koordinaten von Strukturpaaren ersetzt. Bei Durchführung einer RMSD-Überlagerung mit sich anschließender Transformation des kartesischen Differenzvektors zu einer vorurteilsfreien Mode führt zur Identifizierung der am wenigsten mobilen Atome.

Anwendungen der vorurteilsfreien Trennung von globaler und interner Dynamik beinhalten auch Abtast-Methoden zur Aufklärung des zugänglichen Konformationsraums. Solche Methoden benötigen in regelmäßigen Abständen Überlagerungen zu einer Referenzstruktur. In diesem Fall sollten die EDA-Moden verwendet werden, denn die "klassische" Überlagerung ist rechentechnisch weniger anspruchsvoll als die vorurteilsfreie Methode. Die restliche Rotation und Translation sollte erst im nachfolgenden Schritt bei der Untersuchung des generierten Ensembles berücksichtigt werden. Die Berechnung von Konfigurationsentropien aus Kovarianzmatrizen hängt ebenfalls von der Trennung von interner und globaler Bewegung ab. Vorurteilsfreie Kovarianzen werden die Präzision der Ergebnisse solcher Rechnungen erhöhen.

Es wurde gezeigt, daß die Beschreibung der dynamischen Struktur mit Hilfe von überlagerten harmonischen Potentialen im Diederwinkelraum, genau wie kollektive Koordinaten, zu einer Reduktion der Anzahl der beschreibenden Parameter führt. Die freien Energien sind gut durch harmonische Potentialverläufe zu beschreiben. Nur 5 von insgesamt 104 Aminosäureresten weisen einen Energieverlauf des ϕ und ψ -Winkels auf, der durch mehr als zwei überlagerte harmonische Potentiale dargestellt wird. Die Energieverläufe von 76 Amino-

säureresten sind sogar in beiden Rückgratdiederwinkeln durch eine oder zwei überlagerte harmonische Potentiale zu beschreiben.

Weitere Untersuchungen werden zeigen, ob die Kombination von lokalen und globalen generalisierten Koordinaten weitere Fortschritte bei der Aufklärung des zugänglichen Konformationsraums mit sich bringen wird. Denkbar wäre ein Abtasten des Konformationsraums mit Zusatzpotentialen für die Diederwinkel unter beschleunigter Bewegung in den Moden mit großer Amplitude.

Literatur

- [1] Duan Y, Kollman PA. Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution. *Science* 1998; 282(5389):740–4.
- [2] Glusker JP, Trueblood KN. *Crystal structure Analysis*. New York: Oxford University Press, 1985.
- [3] Ernst RR, Bodenhausen G, Wokaun A. *Principles of NMR in one and two dimensions*. London: Oxford University Press, 1987.
- [4] Wüthrich K. *NMR of proteins and nucleic acids*. New York: Wiley, 1986.
- [5] Abragam A. *Principles of nuclear magnetism*. Oxford: Clarendon Press, 1961.
- [6] Brooks CL, Karplus M, Pettitt BM. *Proteins: A theoretical perspective of dynamics, structure and thermodynamics*. New York: Wiley, 1988.
- [7] McCammon JA, Harvey SC. *Dynamics of proteins and nucleic acids*. Cambridge: Cambridge University Press, 1987.
- [8] Egami F, Takahashi K, Uchida T. Ribonucleases in taka-diastrase: properties, chemical nature, and applications. *Prog Nucleic Acid Res Mol Biol* 1964;3:59–101.
- [9] Kuriyama Y, Egami F. [preparation and properties of water-insoluble ribonuclease t-1]. *Seikagaku* 1966;38(10):735–40.
- [10] Takahashi K, Uchida T, Egami F. Ribonuclease t1, structure and function. *Adv Biophys* 1970;1:53–98.
- [11] Egami F, Oshima T, Uchida T. Specific interaction of base-specific nucleases with nucleosides and nucleotides. *Mol Biol Biochem Biophys* 1980; 32:250–77.

- [12] K SaFE. *J Biochem Tokyo* 1957;44:735.
- [13] Ikehara M, Ohtsuka E, Tokunaga T, Nishikawa S, Uesugi S, Tanaka T, Aoyama Y, Kikyodani S, Fujimoto K, and KY. Inquiries into the structure-function relationship of ribonuclease t1 using chemically synthesized coding sequences. *Proc Natl Acad Sci U S A* 1986;83(13):4695-9.
- [14] Quaas R, McKeown Y, Stanssens P, Frank R, Blocker H, Hahn U. Expression of the chemically synthesized gene for ribonuclease t1 in escherichia coli using a secretion cloning vector. *Eur J Biochem* 1988;173(3):617-22.
- [15] Martinez-Oyanedel J, Choe HW, Heinemann U, Saenger W. Ribonuclease t1 with free recognition and catalytic site: crystal structure analysis at 1.5 a resolution. *J Mol Biol* 1991;222(2):335-52.
- [16] Heinemann U, Wernitz M, Pahler A, Saenger W, Menke G, Ruterjans H. Crystallization of a complex between ribonuclease t1 and 2'-guanylic acid. *Eur J Biochem* 1980;109(1):109-14.
- [17] Heinemann U, Saenger W. Specific protein-nucleic acid recognition in ribonuclease t1-2'-guanylic acid complex: an x-ray study. *Nature* 1982; 299(5878):27-31.
- [18] Sugio S, Oka K, Ohishi H, Tomita K, Saenger W. Three-dimensional structure of the ribonuclease t1 x 3'-guanylic acid complex at 2.6 a resolution. *FEBS Lett* 1985;183(1):115-8.
- [19] Arni R, Heinemann U, Tokuoka R, Saenger W. Three-dimensional structure of the ribonuclease t1 2'-gmp complex at 1.9-a resolution. *J Biol Chem* 1988;263(30):15358-68.
- [20] Sugio S, Amisaki T, Ohishi H, Tomita K. Refined x-ray structure of the low ph form of ribonuclease t1-2'-guanylic acid complex at 1.9 a resolution. *J Biochem Tokyo* 1988;103(2):354-66.

- [21] Koepke J, Maslowska M, Heinemann U, Saenger W. Three-dimensional structure of ribonuclease t1 complexed with guanylyl-2',5'-guanosine at 1.8 a resolution. *J Mol Biol* 1989;206(3):475–88.
- [22] Kostrewa D, Choe HW, Heinemann U, Saenger W. Crystal structure of guanosine-free ribonuclease t1, complexed with vanadate (v), suggests conformational change upon substrate binding. *Biochemistry* 1989; 28(19):7592–600.
- [23] Lenz A, Heinemann U, Maslowska M, Saenger W. X-ray analysis of cubic crystals of the complex formed between ribonuclease t1 and guanosine-3',5'-bisphosphate. *Acta Crystallogr B* 1991;47 (Pt 4):521–7.
- [24] Ding J, Koellner G, Grunert HP, Saenger W. Crystal structure of ribonuclease t1 complexed with adenosine 2'-monophosphate at 1.8-a resolution. *J Biol Chem* 1991;266(23):15128–34.
- [25] Lenz A, Cordes F, Heinemann U, Saenger W. Evidence for a substrate-binding subsite in ribonuclease t1. crystal structure of the complex with two guanosines, and model building of the complex with the substrate guanylyl-3',5'-guanosine. *J Biol Chem* 1991;266(12):7661–7.
- [26] Arni RK, Pal GP, Ravichandran KG, Tulinsky A, Metcalf P. Three-dimensional structure of gln25-ribonuclease t1 at 1.84-a resolution: structural variations at the base recognition and catalytic sites. *Biochemistry* 1992;31(12):3126–35.
- [27] Heydenreich A, Koellner G, Choe HW, Cordes F, Kisker C, Schindelin H, Adamiak R, Hahn U, Saenger W. The complex between ribonuclease t1 and 3'gmp suggests geometry of enzymic reaction path. an x-ray study. *Eur J Biochem* 1993;218(3):1005–12.
- [28] Lenz A, Choe HW, Granzin J, Heinemann U, Saenger W. Three-dimensional structure of the ternary complex between ribonuclease t1,

- guanosine 3',5'-bisphosphate and inorganic phosphate at 0.19 nm resolution. *Eur J Biochem* 1993;211(1-2):311-6.
- [29] Zegers I, Haikal AF, Palmer R, Wyns L. Crystal structure of rnase t1 with 3'-guanylic acid and guanosine. *J Biol Chem* 1994;269(1):127-33.
- [30] Karimi-Nejad Y, Schmidt JM, Ruterjans H, Schwalbe H, Greisinger C. Conformation of valine side chains in ribonuclease t1 determined by nmr studies of homonuclear and heteronuclear 3j coupling constants. *Biochemistry* 1994;33(18):5481-92.
- [31] Pfeiffer S, Karimi-Nejad Y, Ruterjans H. Limits of nmr structure determination using variable target function calculations: ribonuclease t1, a case study. *J Mol Biol* 1997;266(2):400-23.
- [32] Pfeiffer S, Spitzner N, Lohr F, Ruterjans H. Hydration water molecules of nucleotide-free rnase t1 studied by nmr spectroscopy in solution. *J Biomol NMR* 1998;11(1):1-15.
- [33] Fushman D, Weisemann R, Thüring H, Rüterjans H. Backbone dynamics of ribonuclease t1 and its complex with 2'gmp studied by two-dimensional heteronuclear nmr spectroscopy. *J Bio NMR* 1993;4:61-78.
- [34] Fushman D, Ohlenschlager O, Ruterjans H. Determination of the backbone mobility of ribonuclease t1 and its 2'gmp complex using molecular dynamics simulations and nmr relaxation data. *J Biomol Struct Dyn* 1994; 11(6):1377-402.
- [35] Pace CN, Grimsley GR, Thomson JA, Barnett BJ. Conformational stability and activity of ribonuclease t1 with zero, one, and two intact disulfide bonds. *J Biol Chem* 1988;263(24):11820-5.
- [36] Abseher R, Nilges M. Are there non-trivial dynamic cross-correlations in proteins? *J Mol Biol* 1998;279(4):911-20.

- [37] Ichiye T, Karplus M. Collective motions in proteins: a covariance analysis of atomic fluctuations in molecular dynamics and normal mode simulations. *Proteins* 1991;11(3):205–17.
- [38] Hunenberger PH, Mark AE, van Gunsteren WF. Fluctuation and cross-correlation analysis of protein motions observed in nanosecond molecular dynamics simulations. *J Mol Biol* 1995;252(4):492–503.
- [39] Karplus M, Ichiye T. Comment on a fluctuation and cross correlation analysis of protein motions observed in nanosecond molecular dynamics simulations". *J Mol Biol* 1996;263(2):120–2.
- [40] Amadei A, Linssen AB, Berendsen HJ. Essential dynamics of proteins. *Proteins* 1993;17(4):412–25.
- [41] Berendsen HJ, Van Gunsteren WF, Zwinderman HR, Geurtsen RG. Simulations of proteins in water. *Ann N Y Acad Sci* 1986;482:269–86.
- [42] Freyman RP, Leighton RB, Sands M. The Freyman lecture of physics. Addison-Wesley, 1963, Chapter 9 pages.
- [43] van Gunsteren WF, Berendsen HJC. A leap frog algorithm for stochastic dynamics. *Mol Simulation* 1988;1(3):173–85.
- [44] Verlet L. Computer experiments on classical fluids. i. thermodynamical properties of lennard-jones molecules. *Phys Rev* 1967;159(1):98–103.
- [45] Verlet L. Computer experiments on classical fluids. ii. equilibrium correlation functions. *Phys Rev* 1967;165(1):201–215.
- [46] Schmidt JM, Brüschweiler R, Ernst RR, and D. Joseph RLD, Karplus M. Molecular dynamics simulation of the proline conformational equilibrium and dynamics in antanamide using the charm force field. *J Am Chem Soc* 1993;115:8747–56.
- [47] Berendsen HJC, van der Spoel D, van Drunen R. *Comput Phys Commun* 1995;19.

- [48] van der Spoel D, Berendsen HJC, van Buuren AR, Apol E, Meulenhoff PJ, Sijbers ALTM, van Drunen R. GROMACS user manual. 1995.
- [49] van H. Bekker BH, Berendsen HJC, Fraaije JGEM. Lincs: A linear constraint solver for molecular simulations. *J Comput Chem* 1997;18(12):1463–72.
- [50] Ryckaert JP, Cicotti G, Berendsen HJC. Numerical integration of the cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes. *J Comput Phys* 1977;23:327.
- [51] Berendsen HJC, van Gunsteren WF. Molecular dynamics with coupling to an external bath. *J Chem Phys* 1984;81(8):3684–90.
- [52] van Gunsteren WF, Berendsen HJC. Moleküldynamik-computersimulationen; methodik, anwendungen und perspektiven in der chemie. *Angew Chem* 1990;102:1020–55.
- [53] Case DA. Normal mode analysis of protein dynamics. *Curr Opin Struct Biol* 1994;4:285–90.
- [54] Berendsen HJ, Hayward S. Collective protein dynamics in relation to function. *Curr Opin Struct Biol* 2000;10(2):165–9.
- [55] Hayward S, Go N. Collective variable description of native protein dynamics. *Annu Rev Phys Chem* 1995;46:223–50.
- [56] Kitao A, Go N. Investigating protein dynamics in collective coordinate space. *Curr Opin Struct Biol* 1999;9(2):164–9.
- [57] Garcia AE, Harman JG. Simulations of crp:(camp)₂ in noncrystalline environments show a subunit transition from the open to the closed conformation. *Protein Sci* 1996;5(1):62–71.
- [58] de Groot BL, Vriend G, Berendsen HJ. Conformational changes in the chaperonin groel: new insights into the allosteric mechanism. *J Mol Biol* 1999;286(4):1241–9.

- [59] Ma J, Karplus M. The allosteric mechanism of the chaperonin groel: a dynamic analysis. *Proc Natl Acad Sci U S A* 1998;95(15):8502–7.
- [60] Ma J, Sigler PB, Xu Z, Karplus M. A dynamic model for the allosteric mechanism of groel. *J Mol Biol* 2000;302(2):303–13.
- [61] Mouawad L, Perahia D. Diagonalization in a mixed basis: A method to compute low-frequency normal modes for large macromolecules. *Biopolymers* 1993;33:599–611.
- [62] Hinsen K, Thomas A, Field MJ. Analysis of domain motions in large proteins. *Proteins* 1999;34(3):369–82.
- [63] Thomas A, Hinsen K, Field MJ, Perahia D. Tertiary and quaternary conformational changes in aspartate transcarbamylase: a normal mode study. *Proteins* 1999;34(1):96–112.
- [64] Horstink LM, Abseher R, Nilges M, Hilbers CW. Functionally important correlated motions in the single-stranded dna-binding protein encoded by filamentous phage pf3. *J Mol Biol* 1999;287(3):569–77.
- [65] Miller DW, Agard DA. Enzyme specificity under dynamic control: a normal mode analysis of alpha-lytic protease. *J Mol Biol* 1999;286(1):267–78.
- [66] van Aalten DM, Conn DA, de Groot BL, Berendsen HJ, Findlay JB, Amadei A. Protein dynamics derived from clusters of crystal structures. *Biophys J* 1997;73(6):2891–6.
- [67] de Groot BL, Hayward S, van Aalten DM, Amadei A, Berendsen HJ. Domain motions in bacteriophage t4 lysozyme: a comparison between molecular dynamics and crystallographic data. *Proteins* 1998;31(2):116–27.
- [68] Chau PL, van Aalten DM, Bywater RP, Findlay JB. Functional concerted motions in the bovine serum retinol-binding protein. *J Comput Aided Mol Des* 1999;13(1):11–20.

- [69] Lins RD, Briggs JM, Straatsma TP, Carlson HA, Greenwald J, Choe S, McCammon JA. Molecular dynamics studies on the hiv-1 integrase catalytic domain. *Biophys J* 1999;76(6):2999–3011.
- [70] Kazmierkiewicz R, Czaplewski C, Lammek B, Ciarkowski J. Essential dynamics/factor analysis for the interpretation of molecular dynamics trajectories. *J Comput Aided Mol Des* 1999;13(1):21–33.
- [71] Creveld LD, Amadei A, van Schaik RC, Pepermans HA, de Vlieg J, Berendsen HJ. Identification of functional and unfolding motions of cutinase as obtained from molecular dynamics computer simulations. *Proteins* 1998;33(2):253–64.
- [72] Garcia AE, Hummer G. Conformational dynamics of cytochrome c: correlation to hydrogen exchange. *Proteins* 1999;36(2):175–91.
- [73] Kazmirski SL, Li A, Daggett V. Analysis methods for comparison of multiple molecular dynamics trajectories: applications to protein unfolding pathways and denatured ensembles. *J Mol Biol* 1999;290(1):283–304.
- [74] Roccatano D, Amadei A, Di Nola A, Berendsen HJ. A molecular dynamics study of the 41-56 beta-hairpin from b1 domain of protein g. *Protein Sci* 1999;8(10):2130–43.
- [75] Kidera A, Inaka K, Matsushima M, Go N. Normal mode refinement: crystallographic refinement of protein dynamic structure applied to human lysozyme. *Biopolymers* 1992;32(4):315–9.
- [76] Kidera A, Go N. Refinement of protein dynamic structure: normal mode refinement. *Proc Natl Acad Sci U S A* 1990;87(10):3718–22.
- [77] Kidera A, Go N. Normal mode refinement: crystallographic refinement of protein dynamic structure. i. theory and test by simulated diffraction data. *J Mol Biol* 1992;225(2):457–75.

- [78] Kidera A, Inaka K, Matsushima M, Go N. Normal mode refinement: crystallographic refinement of protein dynamic structure. ii. application to human lysozyme. *J Mol Biol* 1992;225(2):477–86.
- [79] Balsera MA, ans Y. Oono WW, Schulten K. Principal component analysis and long time protein dynamics. *J Phys Chem* 1996;100:2567–72.
- [80] Teeter MM, Case DA. Harmonic and quasiharmonic description of crambin. *J Phys Chem* 1990;94:8091–7.
- [81] Kitao A, Go N. Conformational dynamics of polypeptides and proteins in the dihedral angle space and in the cartesian coordinate space: Normal mode analysis of deca-alanine. *J Comput Chem* 1991;12(3):3.
- [82] Kitao A, Hayward S, Go N. Comparison of normal mode analyses on a small globular protein in dihedral angle space and cartesian coordinate space. *Biophys Chem* 1994;52:107–14.
- [83] Hayward S, Kitao A, Hirata F, Go N. Effect of solvent on collective motions in globular protein. *J Mol Biol* 1993;234(4):1207–17.
- [84] Kitao A, Hirata F, Go N. The effects of solvent on the conformation and the collective motions of proteins: Normal mode analysis and molecular dynamics simulation of melittin in water and in vacuum. *Chem Phys* 1991; 158:447–72.
- [85] de Groot BL, van Aalten DM, Amadei A, Berendsen HJ. The consistency of large concerted motions in proteins in molecular dynamics simulations. *Biophys J* 1996;71(4):1707–13.
- [86] Kitao A, Hayward S, Go N. Energy landscape of a native protein: jumping-among-minima model. *Proteins* 1998;33(4):496–517.
- [87] Brooks B, Karplus M. Harmonic dynamics of proteins: normal modes and fluctuations in bovine pancreatic trypsin inhibitor. *Proc Natl Acad Sci U S A* 1983;80(21):6571–5.

- [88] Go N, Noguti T, Nishikawa T. Dynamics of a small globular protein in terms of low-frequency vibrational modes. *Proc Natl Acad Sci U S A* 1983; 80(12):3696–700.
- [89] Levitt M, Sander C, Stern PS. Protein normal-mode dynamics: trypsin inhibitor, crambin, ribonuclease and lysozyme. *J Mol Biol* 1985;181(3):423–47.
- [90] Levitt M, Sander C, Stern PS. The normal modes of a protein: Native bovine pancreatic trypsin inhibitor. *Int J Quant Chem Quant Biol Sympos* 1983;10:181–99.
- [91] Nishikawa T, Go N. Normal modes of vibration in bovine pancreatic trypsin inhibitor and its mechanical property. *Proteins* 1987;2(4):308–29.
- [92] Gibrat JF, Go N. Normal mode analysis of human lysozyme: study of the relative motion of the two domains and characterization of the harmonic motion. *Proteins* 1990;8(3):258–79.
- [93] Seno Y, Go N. Deoxymyoglobin studied by the conformational normal mode analysis. i. dynamics of globin and the heme-globin interaction. *J Mol Biol* 1990;216(1):95–109.
- [94] Yamato T, Higo J, Seno Y, Go N. Conformational deformation in deoxymyoglobin by hydrostatic pressure. *Proteins* 1993;16(4):327–40.
- [95] Ikura T, Go N. Normal mode analysis of mouse epidermal growth factor: characterization of the harmonic motion. *Proteins* 1993;16(4):423–36.
- [96] Tirion MM, ben Avraham D. Normal mode analysis of g-actin. *J Mol Biol* 1993;230(1):186–95.
- [97] Hao M, Harvey SC. Analyzing the normal mode dynamics of macromolecules by the component synthesis method. *Biopolymers* 1992;32:1393–405.

- [98] Hao M, Scheraga HA. Analyzing the normal mode dynamics of macromolecules by the component sythesis method: Residue clustering and multiple-component approach. *Biopolimers* 1994;34:321–35.
- [99] Durand P, Trinquier G, Sanejouand YH. A new approach for determining low-frequency normal modes in macromolecules. *Biopolimers* 1994;34:759–71.
- [100] Harrison RW. Variational calculation of the normal modes of a large macromolecule: methods and some initial results. *Biopolymers* 1984; 23(12):2943–9.
- [101] Roux B, Karplus M. The normal modes of the gramicidin-a dimer channel. *Biophys J* 1988;53(3):297–309.
- [102] Simonson T, Perahia D. Normal modes of symmetric protein assemblies. application to the tobacco mosaic virus protein disk. *Biophys J* 1992; 61(2):410–27.
- [103] van Aalten DM, Amadei A, Bywater R, Findlay JB, Berendsen HJ, Sander C, Stouten PF. A comparison of structural and dynamic properties of different simulation methods applied to sh3. *Biophys J* 1996;70(2):684–92.
- [104] Swaminathan S, Ichiye T, van Gunsteren W, Karplus M. Time dependence of atomic fluctuations in proteins: analysis of local and collective motions in bovine pancreatic trypsin inhibitor. *Biochemistry* 1982;21(21):5230–41.
- [105] Bahar I, Erman B, Haliloglu T, Jernigan RL. Efficient characterization of collective motions and interresidue correlations in proteins by low-resolution simulations. *Biochemistry* 1997;36(44):13512–23.
- [106] Haliloglu T, Bahar I. Coarse-grained simulations of conformational dynamics of proteins: application to apomyoglobin. *Proteins* 1998;31(3):271–81.

- [107] Kurt N, Haliloglu T. Conformational dynamics of chymotrypsin inhibitor 2 by coarse-grained simulations. *Proteins* 1999;37(3):454–64.
- [108] Kurt N, Haliloglu T. Conformational dynamics of subtilisin-chymotrypsin inhibitor 2 complex by coarse-grained simulations. *J Biomol Struct Dyn* 2001;18(5):713–31.
- [109] Bahar I, Atilgan AR, Erman B. Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. *Fold Des* 1997; 2(3):173–81.
- [110] Tirion MM. Large amplitude elastic motions in proteins from a single-parameter, atomic analysis. *PHYSICAL REVIEW LETTERS* 1996; 77(9):1905–1908.
- [111] Horiuchi T, Go N. Projection of monte carlo and molecular dynamics trajectories onto the normal mode axes: human lysozyme. *Proteins* 1991; 10(2):106–16.
- [112] Hayward S, Kitao A, Go N. Harmonicity and anharmonicity in protein dynamics: a normal mode analysis and principal component analysis. *Proteins* 1995;23(2):177–86.
- [113] Abagyan R, Argos P. Optimal protocol and trajectory visualization for conformational searches of peptides and proteins. *J Mol Biol* 1992; 225(2):519–32.
- [114] Garcia AE. Large-amplitude nonlinear motions in proteins. *PHYSICAL REVIEW LETTERS* 1992;68(17):2696–2699.
- [115] Basu G, Kitao A, Hirata F, Go N. A collective motion description of the 3/10-/alpha-helix transition: implications for a natural reaction coordinate. *J Am Chem Soc* 1994;116:6307–15.

- [116] Garcia AE, Soumpasis DM, Jovin TM. Dynamics and relative stabilities of parallel- and antiparallel-stranded dna duplexes. *Biophys J* 1994; 66(6):1742–55.
- [117] van Aalten DM, Amadei A, Linssen AB, Eijssink VG, Vriend G, Berendsen HJ. The essential dynamics of thermolysin: confirmation of the hinge-bending motion and comparison of simulations in vacuum and water. *Proteins* 1995;22(1):45–54.
- [118] Nadig G, Vishveshwara S. Effect of substrate binding on the dynamics of RNase A: Molecular dynamics simulations of RNA bound and native RNase A. *Biopolymers* 1997;42:505–20.
- [119] Laatikainen R, Saarela J, Tuppurainen K, Hassinen T. Internal motions of native lysozyme are more organized than those of mutants: a principal component analysis of molecular dynamics data. *Biophys Chem* 1998; 73(1-2):1–5.
- [120] Kitao A, Hirata F, Go N. Effect of solvent on the conformation and the collective motions of a protein. 2. structure and hydration in melittin. *J Phys Chem* 1993;97:10223–30.
- [121] Kitao A, Hirata F, Go N. Effect of solvent on the conformation and the collective motions of a protein. 3. free energy analysis by the extended RISM theory. *J Phys Chem* 1993;97:10231–35.
- [122] Andrews BK, Romo T, Clarage JB, and BMP. Characterizing global substates of myoglobin. *Structure* 1998;6(5):587–94.
- [123] Lamm G, Szabo A. Langevin modes of macromolecules. *J Chem Phys* 1986;85:7334–48.
- [124] Kottalam J, Case DA. Langevin modes of macromolecules: applications to crambin and DNA hexamers. *Biopolymers* 1990;29:1409–21.

- [125] Elber R, Karplus M. Multiple conformational states of proteins: a molecular dynamics analysis of myoglobin. *Science* 1987;235(4786):318–21.
- [126] Noguti T, Go N. Structural basis of hierarchical multiple substates of a protein. i: Introduction. *Proteins* 1989;5(2):97–103.
- [127] Noguti T, Go N. Structural basis of hierarchical multiple substates of a protein. ii: Monte carlo simulation of native thermal fluctuations and energy minimization. *Proteins* 1989;5(2):104–12.
- [128] Noguti T, Go N. Structural basis of hierarchical multiple substates of a protein. iii: Side chain and main chain local conformations. *Proteins* 1989; 5(2):113–24.
- [129] Noguti T, Go N. Structural basis of hierarchical multiple substates of a protein. iv: Rearrangements in atom packing and local deformations. *Proteins* 1989;5(2):125–31.
- [130] Noguti T, Go N. Structural basis of hierarchical multiple substates of a protein. v: Nonlocal deformations. *Proteins* 1989;5(2):132–8.
- [131] Hayward S, Kitao A, Go N. Harmonic and anharmonic aspects in the dynamics of bpti: a normal mode analysis and principal component analysis. *Protein Sci* 1994;3(6):936–43.
- [132] Janezic D, Veneble RM, Brooks BR. Harmonic analysis of large systems. iii. comparison with molecular dynamics. *J Comput Chem* 1995;16:1554–66.
- [133] Garcia AE, Blumenfeld R, Hummer G, Krumhasl JA. Multi-basin dynamics of a protein in a crystal environment. *Physica D* 1997;107:225–39.
- [134] Torda AE, Scheek RM, van Gunsteren WF. Time-averaged nuclear overhauser effect distance restraints applied to tendamistat. *J Mol Biol* 1990; 214(1):223–35.

- [135] Torda AE, Brunne RM, Huber T, Kessler H, van Gunsteren WF. Structure refinement using time-averaged j-coupling constant restraints. *J Biomol NMR* 1993;3(1):55–66.
- [136] Marek M. Ermittlung der dynamischen Struktur von Proteinen mit Molekular-Dynamik Rechnungen und NMR-Randbedingungen. 1996.
- [137] Caves LS, Evanseck JD, Karplus M. Locally accessible conformations of proteins: multiple molecular dynamics simulations of crambin. *Protein Sci* 1998;7(3):649–66.
- [138] Hansmann UH, Okamoto Y. New monte carlo algorithms for protein folding. *Curr Opin Struct Biol* 1999;9(2):177–83.
- [139] Hansmann UHE, Okamoto Y. Prediction of peptide conformation by multicanonical algorithm: new approach to the multi-minima problem. *J Comput Chem* 1993;14:1333–8.
- [140] Nakajima N, Nakamura H, Kidera A. Multicanonical ensemble generated by molecular dynamics simulation for enhanced conformational sampling of peptides. *J Phys Chem B* 1997;101(5):817–24.
- [141] Shirai H, Nakajima N, Higo J, Kidera A, Nakamura H. Conformational sampling of cdr-h3 in antibodies by multicanonical molecular dynamics simulation. *J Mol Biol* 1998;278(2):481–96.
- [142] Eastman P, Doniach S. Multiple time step diffusive langevin dynamics for proteins. *Proteins* 1998;30(3):215–27.
- [143] Amadei A, Linssen AB, de Groot BL, van Aalten DM, Berendsen HJ. An efficient method for sampling the essential subspace of proteins. *J Biomol Struct Dyn* 1996;13(4):615–25.
- [144] de Groot BL, Amadei A, van Aalten DM, Berendsen HJ. Toward an exhaustive sampling of the configurational spaces of the two forms of the peptide hormone guanylin. *J Biomol Struct Dyn* 1996;13(5):741–51.

- [145] Abseher R, Nilges M. Efficient sampling in collective coordinate space. *Proteins* 2000;39(1):82–88.
- [146] Karplus M, Ichiye T, Pettitt BM. Configurational entropy of native proteins. *Biophys J* 1987;52(6):1083–5.
- [147] Steinhoff HJ, Schlitter J, Redhardt A, Husmeier D, Zander N. Structural fluctuations and conformational entropy in proteins: entropy balance in an intramolecular reaction in methemoglobin. *Biochim Biophys Acta* 1992; 1121(1-2):189–98.
- [148] Schlitter J. Estimation of absolute and relative entropies of macromolecules using the covariance matrix. *Chem Phys Letters* 1993;215(6):617–21.
- [149] Schafer H, Daura X, Mark AE, van Gunsteren WF. Entropy calculations on a reversibly folding peptide: changes in solute free energy cannot explain folding behavior. *Proteins* 2001;43(1):45–56.
- [150] Schäfer H, Mark AE, van Gunsteren WF. Absolute entropies from molecular dynamics simulation trajectories. *J Chem Phys* 2000;113:7809–7817.
- [151] Abragam A. *Principles of Nuclear Magnetism*. Clarendon Press Oxford, 1961.
- [152] Lipari G, Szabo A. Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. theory and range of validity. *J Am Chem Soc* 1982;104:4546–4559.
- [153] Lipari G, Szabo A. Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 2. analysis of experimental results. *J Am Chem Soc* 1982;104:4559–4570.
- [154] Clore G, Driscoll P, Wingfield P, Gronenborn A. Analysis of the backbone dynamics of interleukin-1 β using two-dimensional inverse detected heteronuclear ^{15}N - ^1H NMR spectroscopy. *Biochemistry* 1990;29:7387–7401.

- [155] Clore G, Szabo A, Bax A, Kay L, Driscoll P, Gronenborn A. Deviations from the simple two-parameter model-free approach to the interpretation of nitrogen-15 nuclear magnetic relaxation of proteins. *J Am Chem Soc* 1990;112:4989–4991.
- [156] de Groot BL, Amadei A, Scheek RM, van Nuland NA, Berendsen HJ. An extended sampling of the configurational space of hpr from *e. coli*. *Proteins* 1996;26(3):314–22.
- [157] van Aalten DMF, de Groot BL, Findlay JBC, Berendsen HJC. A comparison of techniques for calculating protein essential dynamics. *J Comput Chem* 1997;18:169–181.
- [158] Kabsch W, Kabsch H, Eisenberg D. Packing in a new crystalline form of glutamine synthetase from *escherichia coli*. *J Mol Biol* 1976;100(3):283–91.
- [159] McLachlan AD. Gene duplications in the structural evolution of chymotrypsin. *J Mol Biol* 1979;128(1):49–79.
- [160] Press WH, Teukolski SA, Vetterling WT, Flannery BP. *Numerical Recipes in FORTRAN*. Cambridge University Press, 1992, 630-3 pages.
- [161] Sachs L. *Statistische Methoden*. Berlin: Springer Verlag, 1993, 179-82 pages.
- [162] Luginbuhl P, Pervushin KV, Iwai H, Wuthrich K. Anisotropic molecular rotational diffusion in ¹⁵n spin relaxation studies of protein mobility. *Biochemistry* 1997;36(24):7305–12.
- [163] Press WH, Flannery BP, Teukolsky SA, Vetterling WT. *Numerical Recipes in C*. New York: Cambridge University Press, 1988, 542-547 pages.
- [164] McQuarrie DA. *Statistical Mechanics*. New York: Harper Collins Pub., 1976.
- [165] McCammon JA, Wolynes PG, Karplus M. Picosecond dynamics of tyrosine side chains in proteins. *Biochemistry* 1979;18(6):927–42.

- [166] Buck M, Karplus M. Internal and overall peptide group motion in proteins: Molecular dynamics simulations for lysozyme compared with results from x-ray and nmr spectroscopy. *J Am Chem Soc* 1999;121:9645–9658.
- [167] Pfeiffer S, Fushman D, Cowburn D. Simulated and nmr-derived backbone dynamics of a protein with significant flexibility: a comparison of spectral densities for the betaark1 ph domain. *J Am Chem Soc* 2001;123(13):3021–36.

5 Anhang

5.1 Herleitung der angepassten Eigenwerte λ_i^{ortho}

Gleichung 59 für die interne Bewegung bezieht sich auf die Ergebnisse einer EDA. Gleichung 94 zeigt die Definition der Strukturen im essentiellen Raum. Die generalisierte Koordinate von Mode i wird q_i genannt. Die Anzahl der Eigenvektoren, die den essentiellen Raum aufspannen wird mit n bezeichnet.

$$\vec{x} - \langle \vec{x} \rangle = \sum_{i=1}^n q_i \cdot \overrightarrow{\nu^{int}(\lambda_i)} \quad (94)$$

Durch Substitution der internen Mode mit Gleichung 73 ergibt sich Gleichung 96 unter Benutzung der Abkürzung aus Gleichung 95.

$$f_i = \left| \overrightarrow{\nu^{int}(\lambda_i)} - \overrightarrow{\nu^{trans}(\lambda_i)} - \overrightarrow{\nu^{rot}(\lambda_i)} \right| \quad (95)$$

$$\mathbf{x} - \langle \mathbf{x} \rangle = \sum_{i=1}^n q_i \cdot \left(f_i \cdot \overrightarrow{\nu^{bf}(\lambda_i)} + \overrightarrow{\nu^{trans}(\lambda_i)} + \overrightarrow{\nu^{rot}(\lambda_i)} \right) \quad (96)$$

Die Konformationen, die sich bei der Auslenkung einer Mode ergeben ist unabhängig vom Rotationsanteil und vom Translationsanteil der Mode. Auch die Amplitude, die benötigt wird, um dieselbe Konformation darzustellen, unterscheiden sich nicht in Moden mit derselben internen Bewegung und unterschiedlichen Rotations- und Translationsanteilen.

$$k_1 \cdot \overrightarrow{\nu^{bf}} + k_2 \cdot \overrightarrow{\nu^{trans}} + k_3 \cdot \overrightarrow{\nu^{rot}} \equiv k_1 \cdot \overrightarrow{\nu^{bf}} \quad (97)$$

Folglich kann die Fluktuation der vorurteilsfreien Moden mit den selben Werten für die internen Koordinaten q_i ausgedrückt werden.

Die vorurteilsfreien Vektoren $\overrightarrow{\nu^{bf}}$ und ihre orthogonalisierten Moden spannen den gleichen Raum auf. Aber die Werte für die Koordinaten q_i und q_i^{ortho} sind unterschiedlich.

$$(\mathbf{x} - \langle \mathbf{x} \rangle)' = \sum_{i=1}^n q_i \cdot f_i \cdot \overrightarrow{\nu^{bf}(\lambda_i)} = \sum_{i=1}^n q_i^{ortho} \cdot \overrightarrow{\nu^{ortho}(\lambda_i)} \quad (98)$$

Unter Verwendung von Gleichung 77 ergibt sich Gleichung 99. Gleichung 101 ist eine Vereinfachung von Gleichung 99 unter Benutzung der Abkürzung in Gleichung 100.

$$\begin{aligned} \sum_{i=1}^n q_i^{ortho} \cdot \overrightarrow{\nu^{ortho}(\lambda_i)} &= q_1 \cdot f_1 \cdot \overrightarrow{\nu^{ortho}(\lambda_1)} + \\ \sum_{i=2}^n q_i \cdot f_i \cdot \left(\overrightarrow{\nu^{ortho}(\lambda_i)} \cdot \left| \overrightarrow{\nu^{bf}(\lambda_i)} - \sum_{k=1}^{i-1} \left(\overrightarrow{\nu^{ortho}(\lambda_k)} \cdot \overrightarrow{\nu^{bf}(\lambda_i)} \right) \cdot \overrightarrow{\nu^{ortho}(\lambda_k)} \right| + \right. \\ &\quad \left. \sum_{k=1}^{i-1} \left(\overrightarrow{\nu^{ortho}(\lambda_k)} \cdot \overrightarrow{\nu^{bf}(\lambda_i)} \right) \cdot \overrightarrow{\nu^{ortho}(\lambda_k)} \right) \end{aligned} \quad (99)$$

$$g_i = \begin{cases} \left| \overrightarrow{\nu^{bf}(\lambda_i)} - \sum_{k=1}^{i-1} \left(\overrightarrow{\nu^{ortho}(\lambda_k)} \cdot \overrightarrow{\nu^{bf}(\lambda_i)} \right) \cdot \overrightarrow{\nu^{ortho}(\lambda_k)} \right| & \text{for } i \geq 2 \\ 1 & \text{for } i = 1 \end{cases} \quad (100)$$

$$\begin{aligned} \sum_{i=1}^n q_i^{ortho} \cdot \overrightarrow{\nu^{ortho}(\lambda_i)} &= \sum_{i=1}^n q_i \cdot f_i \cdot g_i \cdot \overrightarrow{\nu^{ortho}(\lambda_i)} + \\ \sum_{i=2}^n q_i \cdot f_i \cdot \sum_{k=1}^{i-1} \left(\overrightarrow{\nu^{ortho}(\lambda_k)} \cdot \overrightarrow{\nu^{bf}(\lambda_i)} \right) \cdot \overrightarrow{\nu^{ortho}(\lambda_k)} \end{aligned} \quad (101)$$

Die Summanden von Gleichung 101 sind in den Gleichungen 102 bzw. 103 nach den orthogonalisierten Moden geordnet.

$$\begin{aligned} \sum_{i=1}^n q_i^{ortho} \cdot \overrightarrow{\nu^{ortho}(\lambda_i)} &= \sum_{i=1}^n q_i \cdot f_i \cdot g_i \cdot \overrightarrow{\nu^{ortho}(\lambda_i)} + \\ \sum_{k=1}^{n-1} \sum_{l=k+1}^n q_l \cdot f_l \cdot \left(\overrightarrow{\nu^{ortho}(\lambda_k)} \cdot \overrightarrow{\nu^{bf}(\lambda_l)} \right) \cdot \overrightarrow{\nu^{ortho}(\lambda_k)} \end{aligned} \quad (102)$$

$$\begin{aligned} \sum_{i=1}^n q_i^{ortho} \cdot \overrightarrow{\nu^{ortho}(\lambda_i)} &= q_n \cdot f_n \cdot g_n \cdot \overrightarrow{\nu^{ortho}(\lambda_n)} + \\ \sum_{i=1}^{n-1} \left(q_i \cdot f_i \cdot g_i + \sum_{l=i+1}^n q_l \cdot f_l \cdot \left(\overrightarrow{\nu^{ortho}(\lambda_i)} \cdot \overrightarrow{\nu^{bf}(\lambda_l)} \right) \right) \cdot \overrightarrow{\nu^{ortho}(\lambda_i)} \end{aligned} \quad (103)$$

Die Werte für die Koordinaten der Mode mit dem kleinsten Eigenwert q_n^{ortho}

und der entsprechende neue Eigenwert λ_n^{ortho} kann direkt aus Gleichung 103 abgeleitet werden.

$$q_n^{ortho} = q_n \cdot f_n \cdot g_n \quad (104)$$

$$\lambda_n^{ortho} = \langle (q_n^{ortho})^2 \rangle = \langle (q_n \cdot f_n \cdot g_n)^2 \rangle = f_n^2 \cdot g_n^2 \cdot \lambda_n \quad (105)$$

Die Koordinaten der übrigen Moden und ihre neuen Eigenwerte werden auch aus Gleichung 103 entnommen.

$$q_{i<n}^{ortho} = q_i \cdot f_i \cdot g_i + \sum_{l=i+1}^n q_l \cdot f_l \cdot \left(\overrightarrow{\nu^{ortho}(\lambda_i)} \cdot \overrightarrow{\nu^{bf}(\lambda_l)} \right) \quad (106)$$

$$\lambda_{i<n}^{ortho} = \langle (q_{i<n}^{ortho})^2 \rangle = \left\langle \left(q_i \cdot f_i \cdot g_i + \sum_{l=i+1}^n q_l \cdot f_l \cdot \left(\overrightarrow{\nu^{ortho}(\lambda_i)} \cdot \overrightarrow{\nu^{bf}(\lambda_l)} \right) \right)^2 \right\rangle \quad (107)$$

Die Vereinfachung von Gleichung 107 führt zu Gleichung ??.

$$\begin{aligned} \lambda_{i<n}^{ortho} = & f_i^2 \cdot g_i^2 \cdot \lambda_i + 2 \cdot f_i \cdot g_i \cdot \sum_{k=i+1}^n f_k \cdot \left(\overrightarrow{\nu^{ortho}(\lambda_i)} \cdot \overrightarrow{\nu^{bf}(\lambda_k)} \right) \cdot \langle q_i \cdot q_k \rangle + \\ & \sum_{k=i+1}^n f_k^2 \cdot \left(\overrightarrow{\nu^{ortho}(\lambda_i)} \cdot \overrightarrow{\nu^{bf}(\lambda_k)} \right)^2 \cdot \lambda_k + \\ & 2 \cdot \sum_{k=i+1}^{n-1} \sum_{l=k+1}^n f_k \cdot f_l \cdot \left(\overrightarrow{\nu^{ortho}(\lambda_i)} \cdot \overrightarrow{\nu^{bf}(\lambda_k)} \right) \cdot \left(\overrightarrow{\nu^{ortho}(\lambda_i)} \cdot \overrightarrow{\nu^{bf}(\lambda_l)} \right) \cdot \langle q_k \cdot q_l \rangle \end{aligned} \quad (108)$$

Die Moden $\overrightarrow{\nu^{int}(\lambda_i)}$ sind unkorreliert, weil sie aus der Diagonalisierung einer Kovarianzmatrix resultieren. Folglich müssen alle inneren Produkte der transformierten Koordinaten $\langle q_i \cdot q_k \rangle$ gleich Null sein. Mit dieser Vereinfachung ergibt sich Gleichung 109.

$$\lambda_{i<n}^{ortho} = f_i^2 \cdot g_i^2 \cdot \lambda_i + \sum_{k=i+1}^n f_k^2 \cdot \left(\overrightarrow{\nu^{ortho}(\lambda_i)} \cdot \overrightarrow{\nu^{bf}(\lambda_k)} \right)^2 \cdot \lambda_k \quad (109)$$

5.2 Tabellen der Gauß'schen Verteilungen der Diederwinkel

Aminosäurerest	$\langle \phi_k \rangle$	σ_k	A_K	$T_{Gau}/10^4$
CYS 2	-115.8	14.1	0.47	
CYS 2	-110.6	28.8	0.52	0.3913
ASP 3	-115.5	14.0	0.68	
ASP 3	-90.7	26.7	0.32	0.3638
TYR 4	-107.5	17.0	1.01	0.9477
THR 5	-78.3	11.8	1.00	0.2820
CYS 6	-92.7	20.0	1.01	0.4913
GLY 7	-53.8	13.0	0.64	
GLY 7	-27.4	33.1	0.21	
GLY 7	56.5	15.5	0.15	0.4412
SER 8	-80.0	20.8	0.29	
SER 8	-61.5	12.1	0.69	0.4693
ASN 9	-163.6	12.5	0.42	
ASN 9	-142.7	20.3	0.58	0.7449
CYS 10	-80.1	13.6	0.97	0.7541
TYR 11	-111.2	13.1	0.50	
TYR 11	-81.2	16.4	0.50	0.3147
SER 12	-89.1	15.7	1.00	0.5261
SER 13	-52.7	9.6	0.99	0.5749
SER 14	-60.9	10.9	0.99	0.4545
ASP 15	-58.0	9.8	0.99	0.5062
VAL 16	-60.7	9.8	1.00	0.5388
SER 17	-57.8	10.1	0.99	0.4266
THR 18	-62.8	10.1	0.99	0.4922
ALA 19	-59.2	9.7	0.99	0.4189
GLN 20	-56.1	9.1	1.00	0.4320

Aminosäurerest	$\langle\phi_i\rangle$	σ_i	A_k	$T_{Gau}/10^4$
ALA 21	-57.4	9.7	1.00	0.3669
ALA 22	-59.3	10.0	1.00	0.3586
GLY 23	-56.4	10.8	0.99	0.3596
TYR 24	-51.6	10.8	1.00	0.5134
LYSH 25	-54.4	9.8	0.99	0.3024
LEU 26	-68.3	11.3	0.99	0.4889
HISH 27	-59.8	10.2	0.79	
HISH 27	-78.0	18.0	0.21	0.3412
GLU 28	-75.3	17.0	0.48	
GLU 28	-60.3	11.5	0.52	0.4717
ASP 29	-128.0	23.1	0.51	
ASP 29	-68.7	14.7	0.51	0.8530
GLY 30	140.2	36.7	0.92	
GLY 30	-93.2	22.2	0.08	0.9612
GLU 31	-116.9	25.8	0.80	
GLU 31	-64.1	15.8	0.19	0.3987
THR 32	-90.8	34.3	0.52	
THR 32	-59.5	15.6	0.47	0.4634
VAL 33	-93.7	25.2	1.01	0.6545
GLY 34	-46.9	28.9	0.06	
GLY 34	76.0	24.2	0.65	
GLY 34	70.1	12.0	0.29	0.4038
SER 35	-78.5	28.4	0.42	
SER 35	-58.8	12.0	0.58	0.3974
ASN 36	-95.8	22.8	0.99	0.6706
SER 37	-78.2	16.6	0.47	

Aminosäurerest	$\langle\phi_i\rangle$	σ_i	A_k	$T_{Gau}/10^4$
SER 37	31.2	25.5	0.28	
SER 37	56.0	13.4	0.25	0.4242
TYR 38	-131.7	14.5	0.40	
TYR 38	-100.1	19.1	0.61	0.3939
PRO 39	-82.8	11.2	0.65	
PRO 39	-75.4	6.8	0.36	0.4380
HISH 40	-109.3	14.2	0.63	
HISH 40	-81.6	14.3	0.37	0.3457
LYSH 41	-64.9	12.3	0.99	0.7810
TYR 42	-89.7	22.2	0.44	
TYR 42	-71.3	11.3	0.56	0.4410
ASN 43	-126.1	14.8	0.71	
ASN 43	28.2	18.9	0.26	0.6092
ASN 44	-75.6	10.9	0.14	
ASN 44	-59.0	29.9	0.53	
ASN 44	52.5	17.5	0.33	0.6154
TYR 45	-106.6	22.9	0.89	
TYR 45	-51.7	15.4	0.11	0.3727
GLU 46	-161.4	12.3	0.31	
GLU 46	-134.0	15.0	0.37	
GLU 46	-86.5	29.3	0.26	0.7146
GLY 47	120.0	8.9	0.09	
GLY 47	151.6	38.2	0.92	0.8675
PHE 48	-103.7	16.2	0.50	
PHE 48	-71.2	16.0	0.45	0.6430
ASP 49	-100.7	21.3	1.01	0.4949

Aminosäurerest	$\langle\phi_i\rangle$	σ_i	A_k	$T_{Gau}/10^4$
PHE 50	-115.8	14.7	0.55	
PHE 50	-88.0	22.6	0.45	0.3068
SER 51	-123.1	12.7	0.70	
SER 51	-89.2	16.1	0.29	0.3428
VAL 52	-125.3	14.0	0.77	
VAL 52	-65.0	24.9	0.22	0.6460
SER 53	-142.8	19.0	0.61	
SER 53	-52.4	16.1	0.28	
SER 53	38.0	23.0	0.10	0.5819
SER 54	-74.2	19.5	0.72	
SER 54	-63.2	11.0	0.28	0.3770
PRO 55	-85.6	10.5	1.00	0.4124
TYR 56	-105.0	15.1	1.00	0.7110
TYR 57	-101.8	16.2	0.42	
TYR 57	-118.0	10.8	0.58	0.3529
GLU 58	-96.6	14.1	1.00	0.5663
TRP 59	-135.8	13.0	1.00	0.5547
PRO 60	-64.5	10.7	1.00	0.5100
ILE 61	-94.7	19.5	1.01	0.4984
LEU 62	-147.4	15.5	0.12	
LEU 62	-124.9	11.8	0.30	
LEU 62	-51.4	21.5	0.05	
LEU 62	35.2	19.3	0.24	
LEU 62	56.0	12.0	0.29	
LEU 62	56.0	12.0	0.29	0.4105
SER 63	-100.2	28.1	0.90	

Aminosäurerest	$\langle\phi_i\rangle$	σ_i	A_k	$T_{Gau}/10^4$
SER 63	68.8	9.4	0.11	0.8676
SER 64	-140.6	20.8	0.31	
SER 64	-88.3	29.5	0.52	
SER 64	53.9	15.1	0.17	0.5710
GLY 65	-78.8	24.8	0.21	
GLY 65	76.2	13.5	0.31	
GLY 65	104.3	29.6	0.45	0.7435
ASP 66	-99.0	26.2	0.84	
ASP 66	-71.3	11.6	0.15	0.4611
VAL 67	-116.1	16.1	0.22	
VAL 67	-56.4	13.4	0.78	0.4617
TYR 68	-148.6	17.2	0.25	
TYR 68	-102.7	20.7	0.73	0.5286
SER 69	-111.2	24.7	0.42	
SER 69	-67.8	17.5	0.29	
SER 69	58.5	21.8	0.28	0.5236
GLY 70	-55.3	14.5	0.24	
GLY 70	-13.7	21.1	0.10	
GLY 70	62.5	16.4	0.58	0.9150
GLY 71	-140.4	55.5	0.54	
GLY 71	-65.9	11.1	0.07	
GLY 71	82.1	21.7	0.40	0.9380
SER 72	-115.8	19.1	0.73	
SER 72	-66.9	15.4	0.25	0.5749
PRO 73	-69.7	12.9	0.99	0.6590
GLY 74	-59.0	12.7	0.43	

Aminosäurerest	$\langle\phi_i\rangle$	σ_i	A_k	$T_{Gau}/10^4$
GLY 74	-71.0	31.8	0.35	
GLY 74	67.6	16.7	0.20	0.3879
ALA 75	-61.4	11.7	0.50	
ALA 75	-72.0	23.3	0.50	0.3124
ASP 76	-89.3	24.4	0.30	
ASP 76	-69.4	11.7	0.71	0.3413
ARG 77	-129.1	10.1	1.00	0.3340
VAL 78	-95.0	14.5	1.01	0.3801
VAL 79	-105.9	12.4	1.00	0.6198
PHE 80	-119.7	12.5	1.00	0.4802
ASN 81	-84.4	15.8	1.01	0.7507
GLU 82	-59.5	15.5	0.10	
GLU 82	28.8	24.8	0.25	
GLU 82	58.5	11.1	0.66	0.7078
ASN 83	-104.7	22.3	0.75	
ASN 83	-117.6	12.6	0.25	0.4264
ASN 84	-131.7	15.0	0.77	
ASN 84	-63.3	28.9	0.13	
ASN 84	61.6	17.1	0.10	0.6394
GLN 85	-131.2	25.5	0.12	
GLN 85	-56.3	14.0	0.70	
GLN 85	62.8	11.7	0.15	0.7129
LEU 86	-70.8	13.6	0.99	0.6182
ALA 87	-81.4	15.3	1.00	0.3937
GLY 88	175.2	13.1	0.98	0.4685
VAL 89	-113.7	14.8	1.00	0.8355

Aminosäurerest	$\langle\phi_i\rangle$	σ_i	A_k	$T_{Gau}/10^4$
ILE 90	-107.3	16.1	1.00	0.7130
THR 91	-113.4	17.3	1.00	0.3458
HISH 92	-81.4	11.8	0.99	0.3811
THR 93	-72.7	14.6	0.30	
THR 93	-111.1	16.9	0.70	0.2974
GLY 94	-74.7	24.6	0.22	
GLY 94	68.8	15.0	0.09	
GLY 94	115.5	22.9	0.64	0.9937
ALA 95	-126.2	24.5	0.75	
ALA 95	-69.4	17.4	0.26	0.4664
SER 96	-59.0	17.9	0.48	
SER 96	25.1	25.8	0.24	
SER 96	51.0	15.3	0.25	0.5274
GLY 97	-59.1	17.2	0.13	
GLY 97	78.9	9.4	0.14	
GLY 97	99.6	25.1	0.71	0.4937
ASN 98	-93.7	25.4	1.00	0.7232
ASN 99	-85.1	21.1	0.99	0.8442
PHE 100	-118.2	23.6	0.32	
PHE 100	-80.2	11.1	0.69	0.4076
VAL 101	-112.2	15.9	0.99	0.5612
GLU 102	-62.0	10.7	0.55	
GLU 102	-70.3	17.4	0.45	0.2912
CYS 103	-118.3	15.5	0.21	
CYS 103	-71.2	18.8	0.78	0.4210
THR 104	-122.1	23.7	1.00	0.5563

Aminosäurerest	$\langle\psi_i\rangle$	σ_i	A_k	$T_{Gau}/10^4$
ALA 1	138.5	18.5	0.98	0.6973
CYS 2	105.9	16.3	0.98	0.7520
ASP 3	-54.6	14.8	1.01	0.4080
TYR 4	110.3	12.2	1.00	0.5055
THR 5	79.5	17.0	1.00	0.5089
CYS 6	146.3	14.5	0.84	
CYS 6	88.4	20.0	0.16	0.3198
GLY 7	-52.8	15.9	0.76	
GLY 7	-100.6	14.9	0.22	0.3492
SER 8	-53.8	16.7	0.98	0.5002
ASN 9	145.9	14.6	0.97	0.9924
CYS 10	97.2	12.9	0.39	
CYS 10	124.7	13.9	0.62	0.3923
TYR 11	152.5	13.3	1.00	0.3779
SER 12	162.4	8.7	0.99	0.6156
SER 13	-43.5	10.3	1.00	0.3781
SER 14	-47.3	10.6	1.00	0.4737
ASP 15	-48.0	9.1	1.00	0.3874
VAL 16	-48.5	9.3	1.00	0.2597
SER 17	-47.0	10.0	1.00	0.4679
THR 18	-43.5	10.2	1.00	0.3936
ALA 19	-49.7	9.0	1.00	0.3221
GLN 20	-49.8	8.9	0.99	0.4770
ALA 21	-46.2	9.8	1.00	0.4879
ALA 22	-47.8	9.7	1.00	0.4290
GLY 23	-55.6	10.5	1.00	0.3302

Aminosäurerest	$\langle\psi_i\rangle$	σ_i	A_k	$T_{Gau}/10^4$
TYR 24	-49.5	10.4	1.00	0.7298
LYSH 25	-48.4	10.5	0.99	0.8086
LEU 26	-38.9	12.7	0.98	0.9592
HISH 27	-39.8	11.5	0.99	0.5244
GLU 28	-52.6	15.0	0.98	0.4478
ASP 29	-48.1	15.2	0.53	
ASP 29	107.9	24.1	0.44	0.5729
GLY 30	-177.3	77.0	0.42	
GLY 30	-158.4	19.4	0.57	0.5477
GLU 31	138.9	24.1	0.25	
GLU 31	153.7	11.9	0.38	
GLU 31	-54.4	13.8	0.36	0.3491
THR 32	112.5	31.7	0.74	
THR 32	148.4	14.2	0.26	0.3768
VAL 33	111.6	26.1	0.53	
VAL 33	129.4	12.7	0.48	0.3594
GLY 34	175.4	13.4	0.29	
GLY 34	-168.2	13.4	0.34	
GLY 34	-106.9	18.8	0.35	0.3855
SER 35	-44.9	16.1	0.97	0.7770
ASN 36	12.3	28.1	0.54	
ASN 36	127.2	23.9	0.47	0.4752
SER 37	68.7	25.8	0.37	
SER 37	85.2	15.0	0.62	0.3865
TYR 38	130.1	8.1	0.99	0.6120
PRO 39	117.8	15.5	0.29	

Aminosäurerest	$\langle\psi_i\rangle$	σ_i	A_k	$T_{Gau}/10^4$
PRO 39	134.2	12.0	0.71	0.5675
HISH 40	148.0	10.6	0.99	0.5861
LYSH 41	122.0	14.0	0.99	0.9218
TYR 42	-50.4	12.1	0.25	
TYR 42	69.4	16.1	0.13	
TYR 42	127.2	12.3	0.59	0.7999
ASN 43	60.0	15.4	0.20	
ASN 43	127.3	22.5	0.77	0.5579
ASN 44	-52.4	13.0	0.12	
ASN 44	69.5	26.6	0.33	
ASN 44	116.0	14.0	0.55	0.4067
TYR 45	-58.2	17.1	0.69	
TYR 45	-180.0	79.2	0.14	
TYR 45	-57.4	9.1	0.17	0.4952
GLU 46	38.4	56.9	0.24	
GLU 46	86.2	21.4	0.76	0.8808
GLY 47	-133.5	31.2	0.97	0.7981
PHE 48	134.7	14.6	0.99	0.5141
ASP 49	153.8	13.2	0.20	
ASP 49	125.3	25.0	0.80	0.3945
PHE 50	139.1	14.0	0.81	
PHE 50	116.2	30.6	0.18	0.3594
SER 51	58.0	21.3	0.40	
SER 51	-31.6	19.2	0.16	
SER 51	75.4	12.0	0.43	0.5061
VAL 52	111.2	14.6	0.08	

Aminosäurerest	$\langle\psi_i\rangle$	σ_i	A_k	$T_{Gau}/10^4$
VAL 52	167.3	19.6	0.31	
VAL 52	-88.0	14.8	0.60	0.5039
SER 53	107.9	18.7	1.00	0.5832
SER 54	136.9	8.5	1.00	0.2151
PRO 55	152.3	16.5	1.01	0.8068
TYR 56	131.7	10.9	0.54	
TYR 56	115.7	15.2	0.46	0.4113
TYR 57	131.2	13.7	1.01	0.4606
GLU 58	130.8	11.5	1.00	0.3563
TRP 59	131.7	9.8	0.99	0.4652
PRO 60	135.4	13.7	1.00	0.3829
ILE 61	98.7	16.5	0.97	0.7483
LEU 62	-37.1	12.2	0.11	
LEU 62	82.5	21.8	0.64	
LEU 62	129.9	20.7	0.24	0.5537
SER 63	-51.3	17.7	0.71	
SER 63	116.1	44.5	0.28	0.9091
SER 64	-40.6	16.4	0.19	
SER 64	106.2	30.5	0.80	0.5997
GLY 65	-178.4	26.6	0.39	
GLY 65	-89.7	31.5	0.61	0.5299
ASP 66	-68.4	15.3	0.18	
ASP 66	151.3	18.0	0.78	0.6568
VAL 67	-48.5	15.6	0.99	0.3565
TYR 68	113.5	19.3	0.25	
TYR 68	82.2	34.0	0.48	

Aminosäurerest	$\langle\psi_i\rangle$	σ_i	A_k	$T_{Gau}/10^4$
TYR 68	-44.6	13.5	0.27	0.4902
SER 69	104.0	25.4	0.92	
SER 69	-53.2	20.1	0.05	0.8844
GLY 70	-101.8	10.7	0.11	
GLY 70	-81.7	33.1	0.29	
GLY 70	51.9	16.5	0.22	
GLY 70	104.6	15.1	0.19	
GLY 70	132.5	65.9	0.19	
GLY 70	132.5	65.9	0.19	0.3433
GLY 71	-75.9	33.8	0.70	
GLY 71	53.3	37.5	0.30	0.4749
SER 72	119.8	14.2	0.99	0.6742
PRO 73	-42.8	14.8	0.14	
PRO 73	106.7	20.1	0.27	
PRO 73	144.0	16.0	0.59	0.3345
GLY 74	133.5	23.6	0.95	0.9775
ALA 75	-51.8	14.9	0.98	0.6262
ASP 76	120.8	13.0	1.00	0.3822
ARG 77	141.3	12.7	1.00	0.4289
VAL 78	120.3	13.2	1.00	0.2716
VAL 79	119.0	10.9	1.00	0.3699
PHE 80	142.5	10.9	0.55	
PHE 80	156.0	14.4	0.45	0.5540
ASN 81	121.7	17.9	0.92	
ASN 81	-161.9	17.8	0.08	0.6678
GLU 82	-66.3	17.6	0.35	

Aminosäurerest	$\langle\psi_i\rangle$	σ_i	A_k	$T_{Gau}/10^4$
GLU 82	-80.6	11.0	0.65	0.4330
ASN 83	-57.5	12.9	0.79	
ASN 83	79.9	15.1	0.11	
ASN 83	156.0	12.9	0.09	0.4821
ASN 84	-45.0	14.2	0.70	
ASN 84	68.1	24.0	0.12	
ASN 84	175.4	23.5	0.15	0.8903
GLN 85	121.2	14.6	0.99	0.4538
LEU 86	120.8	13.3	1.01	0.4415
ALA 87	-47.6	11.0	0.99	0.4118
GLY 88	167.8	16.3	1.00	0.3824
VAL 89	113.2	14.4	1.00	0.5867
ILE 90	122.8	15.6	1.00	0.3339
THR 91	140.1	10.2	0.99	0.4288
HISH 92	87.7	25.1	0.81	
HISH 92	81.0	9.9	0.19	0.4563
THR 93	-54.6	20.4	0.69	
THR 93	123.5	44.4	0.32	0.9421
GLY 94	-166.9	21.9	0.21	
GLY 94	-48.8	56.9	0.24	
GLY 94	84.7	25.9	0.55	0.5833
ALA 95	133.7	25.4	1.00	0.8438
SER 96	-62.0	25.5	0.12	
SER 96	74.8	16.4	0.29	
SER 96	125.5	21.8	0.59	0.4770
GLY 97	-71.6	21.5	0.66	

Aminosäurerest	$\langle\psi_i\rangle$	σ_i	A_k	$T_{Gau}/10^4$
GLY 97	-83.4	69.2	0.35	0.4870
ASN 98	74.7	17.9	0.55	
ASN 98	128.7	18.9	0.42	0.6333
ASN 99	-53.7	11.9	0.84	
ASN 99	94.8	19.8	0.14	0.5293
PHE 100	93.4	12.9	0.79	
PHE 100	128.6	18.4	0.22	0.4247
VAL 101	139.9	11.9	0.99	0.3450
GLU 102	180.0	16.5	0.14	
GLU 102	135.6	15.3	0.85	0.3827
CYS 103	131.3	16.7	0.45	
CYS 103	-33.1	17.9	0.22	
CYS 103	71.9	29.3	0.32	0.3852

Aminosäurerest	$\langle\chi_{1,i}\rangle$	σ_i	A_k	$T_{Gau}/10^4$
CYS 2	65.3	10.4	0.09	
CYS 2	178.7	9.4	0.69	
CYS 2	176.8	13.5	0.22	0.3065
ASP 3	-67.9	9.9	0.96	0.5580
TYR 4	-67.9	9.5	0.97	0.6863
THR 5	39.0	12.1	0.22	
THR 5	49.9	8.3	0.73	0.8555
CYS 6	-176.5	14.3	0.14	
CYS 6	-61.3	11.3	0.66	
CYS 6	-67.5	18.5	0.14	0.9941
SER 8	180.0	12.4	0.06	
SER 8	54.1	8.5	0.79	
SER 8	55.5	15.2	0.09	0.8511
ASN 9	-178.5	11.6	0.71	
ASN 9	41.2	4.2	0.02	
ASN 9	31.6	5.6	0.01	
ASN 9	-61.9	12.5	0.16	
ASN 9	52.2	10.7	0.08	0.9644
CYS 10	-173.5	8.6	1.00	0.4902
TYR 11	-67.3	9.4	1.00	0.6163
SER 12	-169.9	10.1	0.16	
SER 12	70.5	7.1	0.55	
SER 12	73.5	10.9	0.26	0.5092
SER 13	-176.0	11.2	0.24	
SER 13	-67.1	11.3	0.33	
SER 13	54.1	12.2	0.42	0.3777

Aminosäurerest	$\langle\chi_{1,i}\rangle$	σ_i	A_k	$T_{Gau}/10^4$
SER 14	-60.0	12.9	0.40	
SER 14	54.5	11.5	0.33	
SER 14	-176.2	12.3	0.24	0.4498
ASP 15	-62.8	9.5	0.99	0.5586
VAL 16	174.0	9.7	0.98	0.4627
SER 17	-66.5	11.8	0.42	
SER 17	61.2	11.8	0.14	
SER 17	-176.2	11.8	0.42	0.5368
THR 18	55.1	9.5	0.56	
THR 18	-51.6	10.5	0.42	0.5966
GLN 20	-80.3	14.1	0.16	
GLN 20	-71.3	9.3	0.48	
GLN 20	-164.3	10.6	0.35	0.4963
TYR 24	159.6	8.6	0.35	
TYR 24	180.0	14.0	0.65	0.6100
LYSH 25	-74.9	10.7	0.60	
LYSH 25	-170.3	13.2	0.37	0.6982
LEU 26	-110.4	32.7	0.11	
LEU 26	-159.6	12.4	0.14	
LEU 26	-84.8	9.6	0.38	
LEU 26	-64.2	10.2	0.36	0.4245
HISH 27	-175.5	11.0	0.98	0.6218
GLU 28	-73.4	11.0	0.40	
GLU 28	-168.2	11.5	0.57	0.7947
ASP 29	-166.6	13.5	0.49	
ASP 29	-75.1	12.9	0.48	0.9056

Aminosäurerest	$\langle\chi_{1,i}\rangle$	σ_i	A_k	$T_{Gau}/10^4$
GLU 31	-171.0	10.8	0.49	
GLU 31	-129.6	12.3	0.44	
GLU 31	-74.7	16.0	0.08	0.4777
THR 32	-49.7	11.2	0.15	
THR 32	56.6	7.8	0.55	
THR 32	46.8	10.4	0.28	0.6511
VAL 33	-176.6	11.0	0.27	
VAL 33	-70.4	8.9	0.29	
VAL 33	68.8	8.8	0.43	0.4606
SER 35	-175.2	12.3	0.29	
SER 35	-62.9	13.0	0.19	
SER 35	56.0	10.8	0.50	0.4303
ASN 36	-164.1	12.6	0.25	
ASN 36	-68.6	12.1	0.53	
ASN 36	49.8	10.8	0.20	0.4558
SER 37	-170.3	13.0	0.12	
SER 37	-62.2	11.8	0.64	
SER 37	55.8	11.1	0.20	0.6470
TYR 38	-85.9	7.0	0.63	
TYR 38	-74.6	11.1	0.37	0.3258
PRO 39	30.4	6.1	0.77	
PRO 39	-3.0	18.1	0.26	1.6777
LYSH 41	-167.7	9.7	0.56	
LYSH 41	-156.0	18.5	0.17	
LYSH 41	60.2	9.6	0.04	
LYSH 41	-74.0	12.3	0.24	0.5409

Aminosäurerest	$\langle\chi_{1,i}\rangle$	σ_i	A_k	$T_{Gau}/10^4$
TYR 42	-174.2	9.5	0.60	
TYR 42	-180.0	6.4	0.40	0.3000
ASN 43	-70.8	12.2	0.94	
ASN 43	-172.0	11.5	0.05	0.3015
ASN 44	-170.1	11.7	0.63	
ASN 44	-59.6	10.0	0.29	
ASN 44	56.5	9.7	0.07	0.4586
TYR 45	-65.1	12.0	0.84	
TYR 45	-170.0	15.2	0.14	0.9565
GLU 46	-146.5	17.4	0.12	
GLU 46	-165.5	11.0	0.47	
GLU 46	-82.2	16.4	0.13	
GLU 46	-70.1	10.3	0.19	
GLU 46	55.1	10.2	0.08	0.3590
PHE 48	-164.8	12.0	0.56	
PHE 48	-85.5	11.3	0.24	
PHE 48	-136.8	24.0	0.17	0.3811
ASP 49	58.9	10.0	0.30	
ASP 49	-162.0	12.9	0.65	0.6018
PHE 50	-69.3	10.4	0.96	0.7266
SER 51	-172.1	12.1	0.34	
SER 51	-66.9	12.6	0.43	
SER 51	55.7	10.9	0.20	0.3920
VAL 52	-169.4	10.5	0.63	
VAL 52	-72.3	10.0	0.27	
VAL 52	73.1	9.2	0.06	0.7883

Aminosäurerest	$\langle\chi_{1,i}\rangle$	σ_i	A_k	$T_{Gau}/10^4$
SER 53	-170.8	11.9	0.55	
SER 53	-65.6	13.5	0.38	
SER 53	55.0	10.9	0.04	0.7574
SER 54	-175.1	12.5	0.70	
SER 54	-68.3	14.6	0.13	
SER 54	57.6	11.9	0.14	0.5475
PRO 55	-15.0	9.7	0.40	
PRO 55	13.0	12.2	0.16	
PRO 55	31.0	6.3	0.45	0.5631
TYR 56	-56.3	7.9	0.58	
TYR 56	42.2	9.0	0.41	0.4939
TYR 57	-61.5	6.7	0.24	
TYR 57	-72.4	12.8	0.75	0.3504
GLU 58	180.0	13.9	0.06	
GLU 58	55.6	8.3	0.65	
GLU 58	54.4	12.6	0.27	0.5854
TRP 59	-178.6	8.1	1.00	0.5219
PRO 60	23.6	8.3	0.24	
PRO 60	-17.6	10.2	0.18	
PRO 60	-28.0	7.1	0.58	0.4240
ILE 61	-174.4	8.4	0.98	0.7492
LEU 62	-159.0	20.8	0.17	
LEU 62	-69.5	12.1	0.45	
LEU 62	-82.2	15.1	0.34	
LEU 62	54.6	9.8	0.06	0.4136
SER 63	-172.5	11.9	0.52	

Aminosäurerest	$\langle\chi_{1,i}\rangle$	σ_i	A_k	$T_{Gau}/10^4$
SER 63	-63.5	13.7	0.24	
SER 63	57.5	11.1	0.20	0.5558
SER 64	-173.2	12.1	0.35	
SER 64	55.9	11.6	0.31	
SER 64	-63.1	13.2	0.31	0.4477
ASP 66	-173.2	9.5	0.45	
ASP 66	65.9	9.0	0.07	
ASP 66	-71.6	14.2	0.18	
ASP 66	-165.9	15.8	0.27	0.3678
VAL 67	-71.4	9.0	0.51	
VAL 67	-177.7	12.1	0.44	0.7514
TYR 68	-178.0	12.6	0.65	
TYR 68	-73.5	11.4	0.32	0.9817
SER 69	-173.4	12.0	0.43	
SER 69	-63.5	13.6	0.33	
SER 69	56.3	11.7	0.22	0.4910
SER 72	-175.4	11.5	0.52	
SER 72	54.9	11.3	0.19	
SER 72	-64.6	13.0	0.27	0.4549
PRO 73	8.6	13.3	0.12	
PRO 73	-25.2	8.7	0.39	
PRO 73	26.3	6.9	0.49	0.5240
ASP 76	-55.7	11.8	0.99	0.6065
ARG 77	-69.8	10.8	0.97	0.5804
VAL 78	174.1	11.9	0.07	
VAL 78	-70.5	9.3	0.80	

Aminosäurerest	$\langle\chi_{1,i}\rangle$	σ_i	A_k	$T_{Gau}/10^4$
VAL 78	65.7	9.4	0.11	0.5403
VAL 79	-179.6	9.3	0.88	
VAL 79	66.8	9.1	0.09	0.4890
PHE 80	60.5	10.6	0.99	0.8942
ASN 81	-180.0	20.2	0.08	
ASN 81	-60.6	9.7	0.70	
ASN 81	57.8	9.8	0.22	0.3267
GLU 82	-66.6	11.6	0.44	
GLU 82	-176.9	7.4	0.18	
GLU 82	-152.8	24.9	0.37	0.9012
ASN 83	-69.5	10.3	0.33	
ASN 83	-169.1	13.2	0.40	
ASN 83	-89.9	23.7	0.28	0.4173
ASN 84	-144.4	24.6	0.12	
ASN 84	-71.8	11.2	0.38	
ASN 84	-169.0	11.8	0.46	
ASN 84	56.7	8.8	0.04	0.4806
GLN 85	-164.2	13.2	0.63	
GLN 85	-73.9	13.2	0.35	0.4696
LEU 86	-169.6	8.9	0.83	
LEU 86	-159.3	17.9	0.13	0.5274
VAL 89	-174.1	8.2	0.99	0.2797
ILE 90	44.2	9.3	0.99	0.4405
THR 91	-175.5	7.4	1.00	0.5901
HISH 92	-171.5	9.9	0.98	0.9671
THR 93	-52.1	11.8	0.10	

Aminosäurerest	$\langle\chi_{1,i}\rangle$	σ_i	A_k	$T_{Gau}/10^4$
THR 93	52.0	9.6	0.80	
THR 93	46.7	15.8	0.10	0.3417
SER 96	-173.0	11.7	0.37	
SER 96	-65.3	13.0	0.37	
SER 96	57.6	12.4	0.22	0.5659
ASN 98	-167.5	12.9	0.23	
ASN 98	-67.7	11.5	0.73	0.8449
ASN 99	-155.1	14.8	0.59	
ASN 99	-81.2	29.5	0.09	
ASN 99	52.8	10.0	0.28	0.6648
PHE 100	-73.0	10.9	0.99	0.4958
VAL 101	-70.2	9.3	0.36	
VAL 101	-178.0	10.6	0.40	
VAL 101	69.1	9.8	0.22	0.4962
GLU 102	-83.0	14.9	0.16	
GLU 102	-142.8	7.1	0.10	
GLU 102	-152.9	18.1	0.75	0.4933
CYS 103	-170.1	13.0	0.14	
CYS 103	-64.4	11.0	0.62	
CYS 103	59.6	10.2	0.23	0.6869
THR 104	54.6	9.3	0.46	
THR 104	-44.3	13.2	0.21	
THR 104	39.5	15.1	0.32	0.4584

Lebenslauf

Name	Schieborr
Vorname	Ulrich
Geburtsdatum	05.07.1968
Geburtsort	Essen
Staatsangehörigkeit	deutsch
Familienstand	verheiratet, 2 Kinder

Schulbildung

1975-1979	Grundschule
1979-1988	Gymnasium
07/1988	Abitur

Wehrdienst

10/1988-12/1989	in Düsseldorf und Essen
-----------------	-------------------------

Studium

01/1990	Beginn des Chemiestudiums in Dortmund
03/1996	Abschluß der Hauptprüfung zum Diplom-Chemiker
04/1996-10/1996	Diplomarbeit am Institut für Physikalische Chemie der Universität Dortmund (Prof. Dr. Roland Winter und Prof. Dr. Geiger), Durchführung der Diplomarbeit am Institut für Biophysikalische Chemie der Johann Wolfgang Goethe-Universität Frankfurt am Main
11/1996	Beginn der Doktorarbeit am Institut für Biophysikalische Chemie der Johann Wolfgang Goethe-Universität Frankfurt am Main unter der Anleitung von Herrn Prof. Dr. Heinz Rüterjans

Veröffentlichungen

Proteins 45 (3), 207-218 (2001)	Ulrich Schieborr and Heinz Rüterjans „Bias-free separation of internal and overall motion of biomolecules”
------------------------------------	---

in Präparation	Ulrich Schieborr and Heinz Rüterjans „A multi harmonic approach to describe the potential hyper surface of proteins in dihedral angle space.”
----------------	--