Molekulardynamik von Proteinen

auf langen Zeitskalen

Physik-Department
Technische Universität München
Theoretische Physik

Molekulardynamik von Proteinen auf langen Zeitskalen

Helmut Grubmüller

Vollständiger Abdruck der von der Fakultät für Physik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften

genehmigten Dissertation

Vorsitzender: ....
1. Prüfer: ....
2. Prüfer: ....
3. Prüfer: ....

Die Dissertation wurde am .... bei der Technischen Universität München eingereicht und durch die Fakultät für Physik am .... angenommen

Inhaltsverzeichnis

'The problem faced by a beginner today is enormous.'

C. P. Slichter, 1961

Kapitel 1

Einleitung

Proteine sind die wesentlichen biochemischen Funktionsträger der Zelle. Sie erfüllen dort eine Vielzahl von Aufgaben wie die enzymatische Katalyse biochemischer Reaktionen, Erkennung und Signalübertragung, etwa durch Immunglobuline oder Hormone, Transportfunktionen, welche häufig von Membranproteinen wahrgenommen werden, Effektorfunktionen oder den Aufbau von Strukturen [1]. Der komplexe räumliche Aufbau von Proteinen soll hier kurz skizziert werden (vgl. Abbildung 1.1). Eine genauere Darstellung findet sich z.B. in [2].

1.1 Proteinstruktur

Proteine sind Polypeptidketten, welche aus einer Abfolge von etwa 20 verschiedenen Aminosäuren bestehen und deren Länge zwischen einigen zehn und mehreren hundert Monomeren beträgt. Die proteinspezifische Sequenz dieser Bausteine entlang der Kette ist genetisch festgelegt und wird als *Primärstruktur* bezeichnet. Abbildung 1.1 (oben links) zeigt einen kurzen Ausschnitt einer derartigen Polypeptidkette[1]; die grau hervorgehobenen Atome bilden das 'Proteinrückgrat'.

Teile des Polypeptidrückgrats sind lokal geordnet. Solche 'Sekundärstukturelemente' werden wesentlich durch Wasserstoffbrückenbindungen stabilisiert [3]. Beispiele solcher Strukturelemente sind die rechts oben in Abbildung 1.1 dargestellte α-Helix, welche ihren Namen aus der helikalen Struktur des Proteinrückgrats (hier schwarz dargestellt) ableitet, oder sogenannte β-Faltblätter. Wie in der Abbildung zu erkennen ist, sind die weiß gezeichneten Seitengruppen der verschiedenen Aminosäuren im

[1] unpolare Wasserstoffatome sind nicht gezeigt

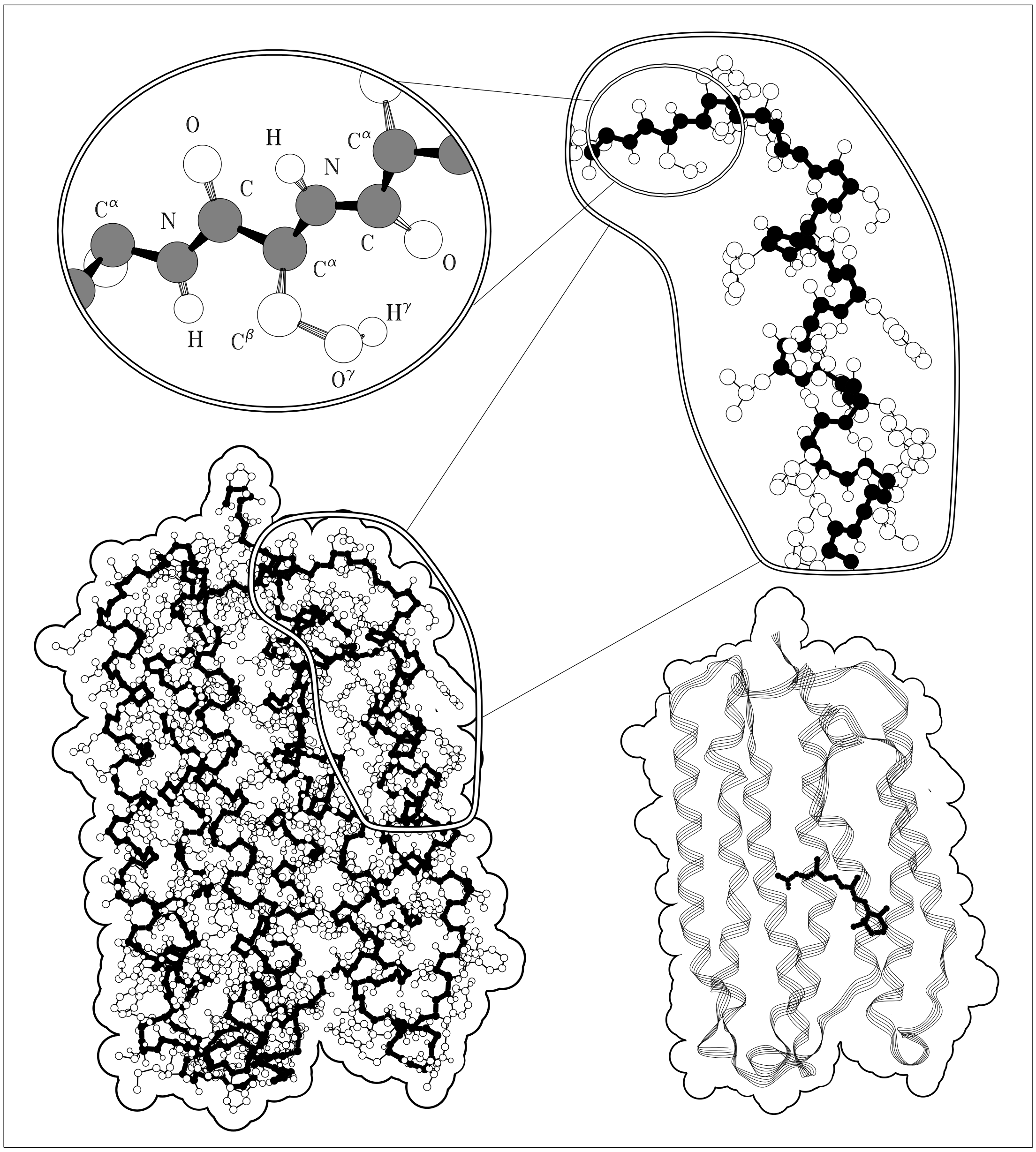

Abbildung 1.1: Proteine sind hierarchisch aufgebaut: Sie bestehen aus einer linearen Kette kovalent verbundener Aminosäuren (links oben). Diese Ketten bilden sekundäre Strukturelemente, wie etwa die rechts oben dargestellte α-Helix. Die reguläre Struktur der α-Helix wird durch Hervorhebung des 'Proteinrückgrats' (schwarz) erkennbar. In nativer Umgebung bestimmt die Abfolge der Aminosäuren entlang der Kette die Tertiärstruktur eines Proteins; als Beispiel ist links unten ein Modell der atomaren Struktur von Bakteriorhodopsin gezeigt. Ein 'Bandmodell' (rechts unten im Bild) verdeutlicht dessen Aufbau aus sieben annähernd parallelen α-Helizes.

Gegensatz zu den Atomen des Proteinrückgrats *unregelmäßig*, aber, wie Strukturuntersuchungen zeigen, in weitgehend festgelegter Weise angeordnet. Diese strukturelle Heterogenität ist ein wesentlicher Aspekt der Komplexität von Proteinen.

Unter nativen Bedingungen nimmt ein Protein eine eindeutig bestimmte globale räumliche Struktur ein, welche als 'Tertiärstruktur' bezeichnet wird. Abbildung 1.1 (links unten) zeigt als Beispiel ein Modell der atomaren Struktur von Bakteriorhodopsin [4], einer lichtgetriebenen Protonenpumpe [5]. Rechts daneben ist dieselbe Struktur als 'Bandmodell' gezeichnet, welches den helikalen Verlauf des Proteinrückgrats hervorhebt. Wie zu erkennen ist, ergibt sich die Tertiärstruktur des Proteins als kompakte Anordnung abwechselnd geordneter und ungeordneter Sekundärstrukturelemente. Im Zentrum des Proteins ist ein Farbstoffmolekül (Retinal) eingelagert. Derartige 'prosthetische Gruppen' sind in zahlreichen Proteinen vorhanden und tragen oft zu ihrer biochemischen Funktion bei.

1.2 Proteinfunktion

Die Proteinfunktion wird durch die physikalisch-chemischen Eigenschaften der Aminosäuren sowie deren räumliche Anordnung und Bewegung im Protein bestimmt. Wesentliche Anstrengungen der molekularen Biophysik von Proteinen zielen auf die Beantwortung (mindestens) einer der beiden Fragen ab:

(1) Wie funktionieren Proteine?

(2) Wie läßt sich aus der Kenntnis der Primärstruktur eines Proteins seine Tertiärstruktur vorhersagen?

Wie deutlich werden wird, muß die Beantwortung der zweiten Frage als Fernziel angesehen werden. Demgegenüber wurden hinsichtlich des Verständnisses von Proteinfunktion auf atomarer Ebene in den letzten Jahren wesentliche Fortschritte erzielt, nicht zuletzt dank der rasch zunehmenden Zahl von Tertiärstrukturen, welche mit atomarer Auflösung bestimmt werden konnten [4, 6–9].

Dabei zeigte sich in vielen Fällen, daß ein physikalisch-chemisches Verständnis der Proteinfunktion im allgemeinen eine Beschreibung der *Proteindynamik* auf atomarer Ebene erfordert [10–15]. Entsprechend sollten adäquaten Beschreibungen dieser Molekulardynamik (MD) wesentlich zur Aufklärung von Struktur-Funktionsbeziehungen in Proteinen beitragen. Die hohe Komplexität der heterogen geordneten Proteinstruktur erschwert jedoch die Entwicklung geeigneter Beschreibungen außerordentlich. Insbesondere verhindert das vollständige Fehlen von Symmetrien die Anwendung herkömmlicher Methoden der Vielteilchenphysik. Neue Konzepte werden zur Behandlung der Proteindynamik benötigt, entsprechend steht deren Entwicklung im Zentrum der Aufmerksamkeit.

Eine wesentliche Konsequenz der strukturellen Komplexität ist die Vielfalt funktioneller Vorgänge unterschiedlichster Dauer, die an der rasch wachsenden Zahl bekannter Proteine beobachtet werden und welche nur unter Einbeziehung der Proteindynamik verstanden werden können. Das breite Spektrum bislang untersuchter Funktionsprozesse soll anhand einiger Beispiele illustriert werden:

Ein sehr schneller biochemischer Prozeß ist der Primärprozeß des lichtinduzierten aktiven Protonentransports im bereits angesprochenen Bakteriorhodopsin: Innerhalb einiger hundert Femtosekunden führt dort die Absorption eines Photons zu einer Isomerisierungsreaktion des im Protein eingelagerten Farbstoffs [16]. Die Dynamik dieses schnellen Prozesses wird wesentlich durch die thermische Bewegung der beteiligten Atome um ihre Ruhelage, durch die Form der Isomerisierungsbarrieren im isolierten Farbstoff sowie durch deren Modifikation aufgrund sterischer und elektrostatischer Wechselwirkungen des Farbstoffs mit der Proteinmatrix bestimmt [17]; bei diesem schnellen Vorgang verhält sich das den Farbstoff umgebende Protein als passives und, von thermischen Atombewegungen abgesehen, starres Gerüst.

In der Regel sind Proteine jedoch *aktiv* an dynamischen Prozessen beteiligt: Sie besitzen insbesondere die Fähigkeit, Entropie für einen lokalen Prozeß bereitzustellen und so Reaktionen mit hoher Aktivierungsenthalpie zu katalysieren [18]. Abbildung 1.2 a stellt die Thermodynamik einer solchen *entropischen Katalyse* schematisch dar. Die relativ zur thermischen Energie hohe Aktivierungsenthalpie (obere Kurve, ΔH) einer biochemischen Reaktion wird hier durch Ankopplung des aktivierten Komplexes an viele Freiheitsgrade des Proteins und durch den damit verbundenen Entropiegewinn (mittlere Kurve, $T\Delta S$) hinreichend weit abgesenkt, so daß die resultierende Aktivierungsbarriere der freien Energie (untere Kurve, ΔG) thermisch überwunden werden kann. Beispielsweise ermöglicht eine derartige entropische Katalyse des Methämoglobins die intramolekulare Dissoziation einer Histidin-Seitengruppe vom Eisenatom der Hämgruppe: Der Entropiebeitrag der Proteinmatrix von 112 J/(mol K) bei Zimmertemperatur reduziert dort die Aktivierungs-Enthalpie von 87 kJ/mol auf 54 kJ/mol und beschleunigt so die Reaktion um den Faktor 10^6. Gleichzeitig wird dadurch der dissoziierte Zustand stabilisiert [18].

Abbildung 1.2 b zeigt den dazu inversen Vorgang: Das Protein verringert kurzzeitig seinen Entropiegehalt und ist somit in der Lage, etwa die im oben beschriebenen Primärprozeß des Bakteriorhodopsins gewonnene Lichtenergie transient zu speichern [19–21]. In einer Folgereaktion erzwingt dieser vom Protein zwischengespeicherte entropische Beitrag zur freien Energie eine endotherme Ladungstrennung und treibt so den Protonentransport an.

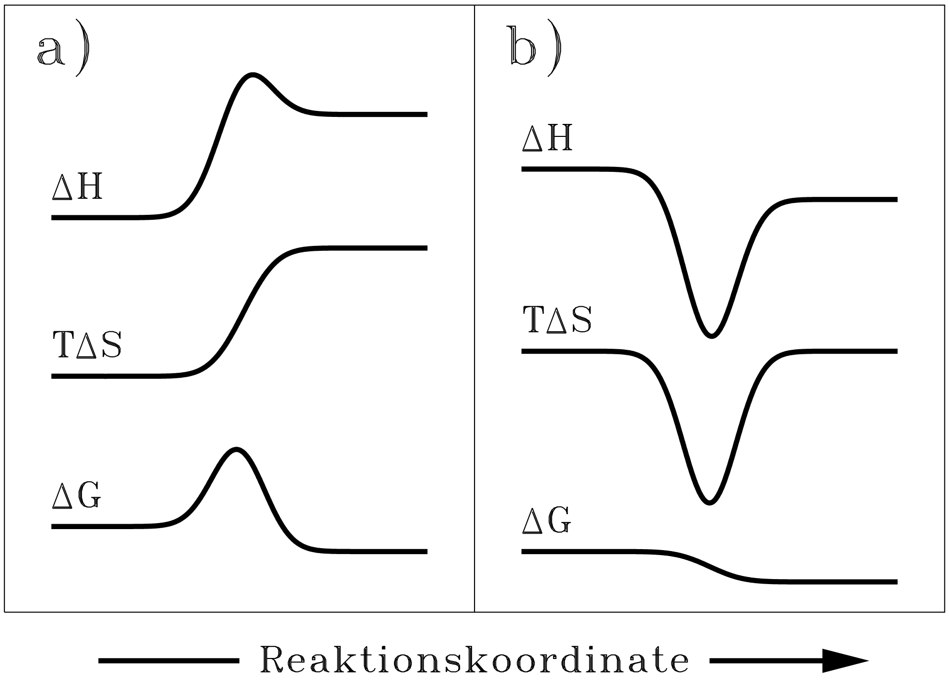

Abbildung 1.2: Proteine können als 'Entropie-Puffer' dienen: Bild a) stellt die Enthalpie-Barriere (ΔH) einer typischen biochemischen Reaktion schematisch dar. Sie ist wesentlich größer als die verfügbare thermische Energie und kann deshalb zunächst nicht überwunden werden. Die Ankopplung an die Freiheitsgrade der Proteinmatrix als 'Entropiequelle' ($T\Delta S$) senkt die freie Energie ΔG jedoch hinreichend weit ab, um die Barriere passierbar zu machen und den Endzustand zu stabilisieren. In Bild b) dient die Proteinmatrix analog als 'Entropiesenke' und ist so in der Lage, Energie einer exothermen Reaktion zwischenzuspeichern. Das Protein setzt diesen entropischen Energiebeitrag anschließend gezielt in einer endothermen Reaktion ein.

Entropieaktivierte Prozesse laufen im allgemeinen um mehrere Größenordnungen langsamer ab als die oben erwähnte enthalpiedominierte Primärreaktion: Während letztere auf einer Zeitskala von Pikosekunden stattfindet, liegen die Reaktionszeiten entropisch katalysierter Vorgänge im Mikro- bis Millisekundenbereich.

Größere strukturelle Änderungen von Proteinen, sogenannte 'Konformationsübergänge', bilden häufig die Elementarprozesse der Proteinfunktion [22] und finden ebenfalls in diesem Zeitbereich statt. Beispielsweise induziert die Absorption eines Photons durch einen eingelagerten Farbstoff eine Änderung der Tertiärstuktur von Rhodopsin, dem Chromoprotein der Sehzellen von Wirbeltieren. Das G-Protein Transduzin erkennt diese Konformationsänderung und löst eine Signalkette aus, welche in einem Nervenimpuls endet [23].

Ein weiteres, prominentes Beispiel für funktionelle Konformationsübergänge ist das Öffnen und Schließen von Ionenkanalproteinen, das auf der Zeitskala von Millisekunden stattfindet. Wie Abbildung 1.3 zeigt, lassen sich mit Hilfe der 'Patch-Clamp'-Technik (vgl. die mit dem Nobelpreis ausgezeichnete Arbeit von Neher und Sakmann [24]) solche Konformationsübergänge einzelner Proteine direkt sichtbar machen: Dargestellt ist der zeitliche Verlauf des Stroms durch einen einzelnen, in der Mitochondrienmembran eingelagerten Ionenkanal; die erkennbaren abrupten Änderungen ('o': offen, 'c': geschlossen) werden hier durch Konformationsübergänge des Kanalproteins hervorgerufen.

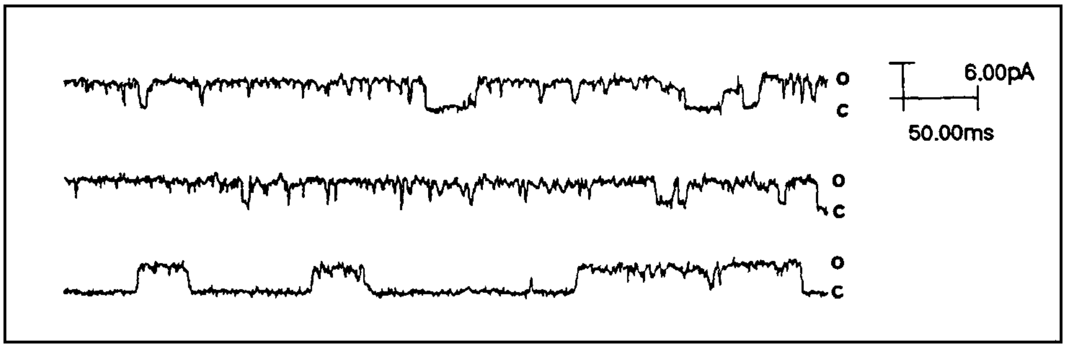

Abbildung 1.3: Der zeitliche Verlauf des Stroms durch ein einzelnes, in die Mitochondrienmembran von Leberzellen eingelagertes Ionenkanalprotein macht Konformationsübergänge des Proteins, die auf der Zeitskala von Millisekunden stattfinden, direkt sichtbar (die drei gezeigten Kurven sind als Sequenz zu verstehen); die Abbildung wurde [25] entnommen.

Eine Reihe von Proteinen, etwa BPTI (Bovine Pancreatic Trypsin Inhibitor), bildet unter physiologischen Bedingungen selbständig seine native Tertiärstruktur aus [26]. Dies geschieht auch dann, wenn diese zuvor gestört (denaturiert) wurde, etwa durch Änderung des pH-Werts oder durch Erwärmung des Lösungsmittels. Dieser Prozeß der 'Proteinfaltung' benötigt zwischen 10^{-1} und 10^{3} Sekunden und zählt damit zu den langsamsten Bewegungen einzelner Proteine [27]. Im Verlauf des Faltungsvorgangs durchläuft ein Protein eine Fülle struktureller Änderungen; einige davon vielfach [26].

1.3 Proteindynamik

Die große Spanne von Zeitskalen funktioneller Prozesse in Proteinen findet ihre Entsprechung in einer Hierarchie *dynamischer* Prozesse, welche ebenfalls einen großen Zeitbereich überdeckt: Hochfrequente, thermische Schwingungen einzelner Atome um ihre Ruhelage sind durch Perioden zwischen einigen zehn und einigen hundert Femtosekunden gekennzeichnet; kollektive Bewegungen kleiner Atomgruppen benötigen einige zehn Pikosekunden. Größere strukturelle Umlagerungen durch eine Hierarchie von Raten oberhalb des Nanosekundenbereichs gekennzeichnet. So weisen experimentelle Befunde [28–33] mit zunehmender Sicherheit die Existenz 'hierarchischer Subzustände' [15, 22] der Proteinstruktur nach. Zwischen diesen sogenannten 'Konformationssubzuständen' [34] finden auf unterschiedlichsten Zeitskalen Konformationsübergänge statt [29, 30, 32, 35], welche anscheinend ein spezifisches Merkmal der Dynamik von Proteinen und ähnlich komplexer Systeme darstellen [36, 37]. Reaktionsketten großer struktureller Umlagerungen sind durch Zeitkonstanten von bis zu 10^3 Sekunden charakterisiert.

Um einem Verständnis der Proteindynamik zu dienen, sollten Beschreibungen diesen Zeitbereich von etwa 17 Größenordnungen überdecken. Aufgrund der Komplexität von Proteinstruktur und -dynamik ist die Entwicklung geeigneter effektiver Beschreibungen eine schwierige und, zum gegenwärtigen Zeitpunkt, noch weitgehend ungelöste Aufgabe.

Die hierarchische Verteilung von Konformationssubzuständen ist Gegenstand theoretischer Untersuchungen [38, 39]. In diesen Arbeiten wird die Verteilung thermisch zugänglicher lokaler Minima der *potentiellen Energie* im Konfigurationsraum betrachtet, wobei angenommen wird, daß diese Minima Information über die experimentell beobachteten Konformationssubzustände enthalten.

Diese Annahme ist sicherlich für niedrige Temperaturen gerechtfertigt, da in diesem Fall der entropische Beitrag zur freien Energie klein ist und vernachlässigt werden kann, so daß die Fläche *potentieller Energie* tatsächlich zur Analyse von Konformationssubzuständen herangezogen werden kann. Bei physiologischen Temperaturen ist die vorgeschlagene Beziehung zwischen zugänglichen Minima der potentiellen Energie und der Lage der experimentell charakterisierten Konformationssubzustände jedoch nicht gesichert. Darüberhinaus kann aus der statischen Struktur des Konformationsraums nicht ohne weiteres auf die Dynamik von Konformationsübergängen geschlossen werden. Sollen entropische Beiträge berücksichtigt werden, so können aufgrund der Größe des abzusuchenden Konfigurationsraums zum gegenwärtigen Zeitpunkt lediglich vergleichsweise einfache Moleküle, wie etwa ein Dipeptid [40], untersucht werden.

Die explizite Methode der Molekulardynamiksimulation [41–43] berücksichtigt entropische Beiträge zur freien Energie und ist heute das einzige Verfahren zur Beschreibung der Proteindynamik, dem eine gewisse Verläßlichkeit zugesprochen werden kann. Dem Anwendungsbereich dieser Methode sind jedoch ebenfalls Grenzen gesetzt [44]: Diese sind sowohl physikalischen Ursprungs — so ist eine Reihe molekularer Eigenschaften durch MD-Simulationen nicht beschreibbar —, als auch rein technischer Natur, so können nur kleine Systeme auf kurzen Zeitskalen behandelt werden.

Die physikalischen Beschränkungen folgen aus der *klassischen* Behandlung der Atombewegungen. Entsprechend werden die Atome eines Makromoleküls als System von Massenpunkten beschrieben, das den klassischen Bewegungsgesetzen gehorcht. Dies verhindert die explizite Beschreibung von Kräften intrinsisch quantenmechanischen Charakters, wie etwa von chemischen Bindungskräften. Stattdessen werden die elektronischen Freiheitsgrade im Rahmen der bekannten Born-Oppenheimer-Näherung in Form eines effektiven Wechselwirkungspotentials für die Kernbewegung behandelt. Dieses 'Kraftfeld' wird semi-empirisch bestimmt, und die Atombewegungen werden näherungsweise durch Lösen der Newtonschen Gleichungen beschrieben. Anwendbarkeit und Qualität all dieser Näherungen werden kontrovers diskutiert [10, 45–51], sind jedoch nicht Gegenstand der vorliegenden Arbeit.

Vielmehr sollen Strategien zur Überwindung der Begrenzungen *technischer* Natur behandelt werden. MD-Simulationen biologischer Makromoleküle, insbesondere von Proteinen, erfordern eine extrem große Anzahl von Rechenoperationen: Die Simulation der Dynamik eines mittelgroßen Systems, bestehend aus 30 000 Atomen, über den Zeitraum einer Nanosekunde benötigt etwa $2 \cdot 10^{16}$ Fließkommaoperationen, wenn auch langreichweitige intramolekularen Wechselwirkungen berücksichtigt werden sollen. Ein Supercomputer mit einer Rechenleistung von 1 GFLOPS[2] benötigt etwa 200 Tage für diese Aufgabe. Aus diesem Grund begrenzt die heute verfügbare Rechenleistung die Anwendbarkeit von MD-Simulationen der Proteindynamik auf den Nanosekundenbereich, so daß nur ein kleiner Teil der oben angesprochenen Funktionenvielfalt beschreibbar ist.

Die beiden geschilderten Beschränkungen betreffen sowohl den MD-Simulationen zugänglichen Bereich physikalischer Phänomene, als auch die Qualität der Beschreibung. Hinsichtlich des approximativen Charakters des verwendeten physikalischen Modells ist dies unmittelbar einsichtig; der Einfluß der technischen Beschränkungen bedarf jedoch der Erörterung.

Die computertechnische Beschränkung der Methode der MD-Simulation bewirkt, daß Observablen, die mit experimentellen Beobachtungen verglichen werden können, aus Kurzzeit-Simulationen eines einzelnen Proteins berechnet werden müssen.

[2] 10^9 **FL**oating point **O**perations **P**er **S**econd (Fließkommaoperationen pro Sekunde)

Wie oben geschildert, finden wichtige biochemische Prozesse jedoch weit oberhalb der durch die heute verfügbare Computertechnologie gesetzten Grenze von einigen Nanosekunden statt. Darüberhinaus sind die meisten experimentellen Beobachtungsgrößen, wie mittlere Atompositionen oder Schwingungsspektren, Mittelwerte über große Ensembles und lange Zeitspannen. Schließlich erfordert die zuverlässige Bestimmung statistischer Größen, wie von Ratenkonstanten oder Differenzen der freien Energie, typischerweise sehr lange Simulationen, da entsprechende Rechnungen aufgrund der Größe des zu untersuchenden Phasenraums nur langsam konvergieren [48].

Letzterem Aspekt kommt aufgrund der angesprochenen Konformationssubzustände besondere Bedeutung zu, da typische Übergangszeiten zwischen verschiedenen Konformationen bei Zimmertemperatur im Bereich oberhalb von Nanosekunden liegen [52, 37] und deshalb davon ausgegangen werden muß, daß sich Mittelwerte, welche aus MD-Simulationen von lediglich einigen 100 Pikosekunden Dauer abgeleitet werden, grundsätzlich von den experimentell beobachteten Ensemble- oder Langzeit-Mittelwerten unterscheiden.

Es sei betont, daß die Verfeinerung molekularer Modelle auf den Vergleich von Simulationsergebnissen mit experimentellen Daten angewiesen ist und daher die skizzierte Inkompatibilität zwischen Theorie und Experiment die Weiterentwicklung molekularer Modelle behindern oder sogar in die Irre führen kann. Deshalb wird jede Steigerung der Simulationsleistung nicht nur dazu dienen, die der Beschreibung zugängliche Menge biochemischer Prozesse zu erweitern, sondern gleichzeitig dazu beitragen, die *Qualität* der Beschreibungen zu verbessern.

1.4 Effiziente Simulationsmethoden

Daher wurden in den letzten Jahren umfangreiche Anstrengungen unternommen, höchstmögliche Rechenleistung für MD-Simulationen verfügbar zu machen, etwa durch effiziente Implementierung von Molekulardynamikprogrammen auf Vektorrechnern [53, 54] oder auf Parallelrechnern [55–60] (vgl. hierzu auch die Diplomarbeit des Autors [61]). Unter der sicher optimistischen Annahme, daß solche Anstrengungen auch weiterhin die verfügbarer Computerleistung um den Faktor 1000 pro Dekade steigern werden, muß jedoch davon ausgegangen werden, daß mit herkömmlichen MD-Methoden eine Vielzahl biochemischer Prozesse nicht vor dem Jahr 2030 beschreibbar sein wird (vgl. auch [44]).

Um den zugänglichen Zeitbereich bereits heute ausdehnen zu können, muß deshalb zusätzlich der zur Beschreibung von Proteindynamik erforderliche Rechenaufwand drastisch verringert werden. Auch in diese Richtung wurden Anstrengungen unternommen, welche sich methodisch in drei Gruppen gliedern:

1) *Mathematische und numerische Methoden* streben eine Einsparung benötigter Rechenoperationen an, ohne die physikalische Beschreibung wesentlich zu verändern. Beispiele hierfür sind Integrationsalgorithmen höherer Ordnung [62, 63], die Verwendung verallgemeinerter interner Koordinaten [64, 65], symplektische Integrationsalgorithmen [63, 66], schnelle Multipolmethoden (fast multipole methods, FMM) [67–70], Methoden mit veränderlicher Integrationsschrittweite [71–74] sowie eine Reihe von Mehrschrittverfahren [71, 73, 75–78].

2) Geeignete *Näherungen* verändern das den Simulationen zugrundeliegende Molekülmodell und verringern so den Rechenaufwand. Dabei muß sichergestellt werden, daß die verwendeten Näherungen die Physik der Proteindynamik nicht allzusehr verfälschen [79]. Beispiele solcher Näherungen sind die Vernachlässigung langreichweitiger Wechselwirkungen mittels einer ‘Abschneidefunktion’ [76], die Unterdrückung schneller Freiheitsgrade wie der Bindungsstreckschwingungen [42] und die sogenannte ‘Massentensor’-Molekulardynamik [80].

3) *Effektive Modelle*, welche das ursprüngliche Molekülmodell ersetzen sollen, basieren auf einer Unterscheidung zwischen ‘relevanten’ und ‘irrelevanten’ Systemeigenschaften. Effektive Modelle beschränken die explizite Beschreibung auf die relevanten Eigenschaften, während irrelevante Eigenschaften implizit durch renormierte Wechselwirkungen oder andere statistische Größen behandelt werden. Insbesondere wird im Rahmen stochastischer Modelle angenommen, daß Details der Dynamik schneller Freiheitsgrade wie Bindungsstreck- oder Bindungswinkelschwingungen nicht funktionsrelevant sind. Gestützt wird diese Annahme durch die erfolgreiche Anwendung stochastischer Beschreibungen, etwa von Monte-Carlo-Simulationen [81, 82], Langevin-Dynamiken [83], verallgemeinerter Langevin-Dynamiken mit Gedächtniskernen [84,85] oder der Verwendung statistischer Potentiale [86, 87].

Eine solche Einteilung ist sicher nicht eindeutig, wie am Beispiel der Vernachlässigung langreichweitiger Wechselwirkungen erkennbar wird: Diese Methode, oben als Näherung klassifiziert, könnte auch als ein effektives Modell angesehen werden, welches die durch atomare Polarisierbarkeiten verursachten Abschirmeffekte implizit berücksichtigt (vgl. auch [88]).

Die oben genannten Verfahren wurden überwiegend für eine große Klasse verschiedenster Vielteilchensysteme entwickelt; sie sind in diesem Sinne allgemeiner Natur. Ihre Anwendung in der Proteindynamik beschleunigt Simulationsrechnungen typischerweise um eine Größenordnung. Im Hinblick auf den Wunsch, erreichbare Zeitskalen um sechs bis zwölf Größenordnungen auszudehnen, repräsentieren diese allgemeinen Methoden jedoch nur einen begrenzten Fortschritt.

1.5 Proteinspezifische Methoden

Weitaus größere Effizienzsteigerungen können von Methoden und effektiven Modellen erwartet werden, die an *proteinspezifische* strukturelle und dynamische Eigenschaften angepaßt sind und diese ausnützen. In der Tat zeichnet sich ab, daß Proteine charakteristische Eigenschaften besitzen [37], welche sie qualitativ von anderen Vielteilchensystemen unterscheiden [27].

Der Leitgedanke der vorliegenden Arbeit ist deshalb die Erkenntnis, daß die physikalischen und numerischen Aspekte der Beschreibung von Proteindynamik aufs engste verknüpft sind: Verbesserte Simulationsmethoden haben bereits wertvolle Beiträge zum Verständnis der Proteindynamik erbracht; eine rückwirkende Anwendung dieser Erkenntnisse zur Verfeinerung der Simulationsmethoden hat jedoch kaum stattgefunden. Diese Arbeit ist in der Überzeugung entstanden, daß eine genaue Charakterisierung proteinspezifischer dynamischer Eigenschaften zur Entwicklung effizienter Methoden der Beschreibung von Proteindynamik insbesondere auf langen Zeitskalen verwendet werden kann und muß.

Sicherlich wird ein derartiges Programm den Rahmen einer einzelnen Dissertation sprengen. Mein Beitrag soll daher darin bestehen, den vorgezeichneten Weg zu erkunden. Wir werden dabei in drei Schritten vorgehen.

a) Effiziente MD-Verfahren

Zunächst sollen in Kapitel 2, im Anschluß an eine Skizze der Methode der Moleküldynamiksimulation, übliche Näherungsverfahren beschrieben werden, die eine effiziente Bestimmung langreichweitiger intramolekularer Wechselwirkungen erlauben. Insbesondere wird auf die weitverbreitete Methode der Abschneidefunktionen näher eingegangen. Anschließend soll die Methode der 'Entfernungsklassen' [89] dargestellt und erweitert werden. Diese Methode wurde 1978 erstmals von W.B. Streett et al. für MD-Simulationen eingesetzt [90], von A. Windemuth zur Simulation eines Membranproteins verwendet [91], im Rahmen der Diplomarbeit des Autors verallgemeinert [61, 77] und schließlich, zu Beginn dieser Dissertation, verfeinert und den Erfordernissen der Proteindynamiksimulation angepaßt. Die dabei entwickelten Algorithmen stützen sich auf spezielle Eigenschaften der Proteindynamik auf kurzen Zeitskalen und sind somit ein erster Schritt auf dem Weg der Entwicklung proteinspezifischer Methoden.

b) Beurteilung der MD-Verfahren

In einem zweiten Schritt (Kapitel 4) soll geprüft werden, wie gut sich die oben erwähnten effizienten MD-Methoden zur Beschreibung von Proteindynamik und -funktion eignen. Dazu wird zunächst die Frage untersucht, welchen Kriterien hierzu verwendet werden können und sollen. Es könnte scheinen, als ob MD-Methoden die Bewegungsgleichungen möglichst effizient *und möglichst genau* lösen sollten. In der Tat wird die Qualität numerischer Methoden zur Lösung von Vielkörperproblemen häufig nach diesem Kriterium beurteilt. Für die Aufklärung von Struktur-Funktionszusammenhängen in Proteinen ist es jedoch üblicherweise nicht nötig, alle Details der Atombewegungen genau zu bestimmen. MD-Simulationen dienen vielmehr dem Ziel, *funktionsrelevante* Größen zu berechnen, d.h. Größen, die geeignet sind, zum Verständnis biochemischer Prozesse beizutragen. Wir werden einen anwendungsorientierten Standpunkt einnehmen und fordern, daß MD-Methoden solche *relevanten* Observablen zuverlässig beschreiben sollten. Da bei MD-Algorithmen eine Verbesserung der Genauigkeit typischerweise einen erhöhten Rechenaufwand erfordert, sollten mithin sogar Methoden verwendet werden, welche irrelevante Freiheitsgrade *möglichst ungenau* beschreiben.

Entsprechend dieser Forderung soll eine Reihe funktionsrelevanter Observablen vorgeschlagen, *problemangepaßte* Qualitätskriterien entwickelt und mit herkömmlichen Verfahren der Qualitätsbestimmung verglichen werden. Anhand von Testsimulationen wird anschließend untersucht, wie genau die effizienten Abschneide- und Entfernungsklassenmethoden diese Observablen der Proteindynamik beschreiben. Der erhaltene 'Steckbrief' algorithmischer Artefakte soll die Frage beantworten helfen, welcher der untersuchten Algorithmen zur MD-Simulation von Proteinen vorzuziehen ist.

c) Konformationsdynamik

Mit Blick auf die Entwicklung effektiver Modelle sollen in einem dritten Schritt Eigenschaften der Proteindynamik auf langen Zeitskalen untersucht werden (Kapitel 5). Während die Atombewegung in Proteinen auf kurzen Zeitskalen durch MD-Simulationen sowie durch Normalmodenanalysen gut charakterisiert werden kann [92–95], ist das theoretische Verständnis der Dynamik auf längeren Zeitskalen noch wenig entwickelt. Wir möchten deshalb durch Untersuchung *langsamer* struktureller Änderungen, für die in Abschnitt 1.2 Beispiele angeführt wurden und die wir als ‘Konformationsdynamik’ bezeichnen wollen, zur Konstruktion effektiver Beschreibungen beitragen.

Auf *kurzen* Zeitskalen unterhalb des Pikosekundenbereichs treten in der Proteindynamik ‘Gedächtniseffekte’ in Erscheinung [84], welche zu nichtverschwindenden Autokorrelationsfunktionen atomarer Bewegungen führen [95]. Derartige Kurzzeitkorrelationen müssen bei stochastischen Beschreibungen schneller Freiheitsgrade in geeigneter Weise berücksichtigt werden [96, 97]. Ganz analog erfordert die von uns angestrebte Entwicklung effektiver Modelle zur Beschreibung *langsamer* Freiheitsgrade, welche die Konformationsdynamik definieren, Kenntnis über Existenz und Eigenschaften von Langzeit-Korrelationen. Wir werden die Konformationsdynamik daher unter diesem Aspekt untersuchen und fragen, inwieweit Gedächtniseffekte in der niederfrequenten Konformationsdynamik von Proteinen auftreten.

d) Langzeitsimulationen, aber wie ?

Zur Charakterisierung dieser Konformationsdynamik bei Zimmertemperatur muß, im Gegensatz zu den üblichen Untersuchungen der potentiellen Energie, die Fläche *freier Energie* mit Hilfe ausgedehnter MD-Simulationen direkt studiert und damit entropische Beiträge berücksichtigt werden. Mit diesem Verfahren ist ein wesentlich höherer Rechenaufwand verbunden: Außerordentlich lange Simulationen, welche bei Verwendung üblicher, komplexer Proteinmodelle die zum gegenwärtigen Zeitpunkt zugänglichen Zeitskalen im Bereich einer Nanosekunde um wenigstens zwei Größenordnungen übersteigen, werden benötigt, um einen hinreichend großen Teil des Phasenraums abzudecken.

Die Entwicklung effektiver Beschreibungen der Proteindynamik auf langen Zeitskalen scheint also in einem Teufelskreis gefangen zu sein: Einerseits erfordert die

Entwicklung geeigneter Beschreibungen der Proteindynamik aufgrund der hohen Komplexität der zu untersuchenden Systeme ausgedehnte Langzeitsimulationen; andererseits verhindert gerade das Fehlen hinreichend effizienter Methoden deren Durchführung.

In ähnlicher Weise erschwert die Begrenzung erreichbarer Simulationszeiten auch die angesprochene *problemangepaßte* Beurteilung von MD-Algorithmen, die ebenfalls als Voraussetzung zur Optimierung effektiver Beschreibungen der Dynamik angesehen werden muß: Auch hier werden ausgedehnte Simulationen benötigt, da funktionsrelevante Observablen in der Regel aus *statistischen* Eigenschaften wie mittleren Atomlagen und -bewegungen, Spektren oder korrelierten Bewegungen von Atomgruppen abgeleitet und daher mit Fehlern behaftet sind. Die Berechnung anderer relevanter Größen, wie etwa der freien Energie, konvergiert oft nur langsam und wird ebenfalls durch statistische Fehler erschwert. Um die statistischen Fluktuationen klein zu halten und so algorithmische Ungenauigkeiten bei der Berechnung relevanter Observabler identifizieren zu können, sind auch hier umfangreiche Testsimulationen erforderlich, welche die zum gegenwärtigen Zeitpunkt zugänglichen Zeitskalen deutlich übersteigen.

e) Proteinmodell vereinfachen!

Um trotz dieser Schwierigkeit die beabsichtigten Untersuchungen der Proteindynamik auf langen Zeitskalen zu ermöglichen, wurde ein vereinfachtes, aber dennoch in wesentlichen strukturellen und dynamischen Aspekten realistisches Proteinmodell entwickelt. Dieses 'realistische Minimalmodell' wird in Kapitel 3 vorgestellt. Ähnlich vereinfachte Modelle, wie z.B. Gittermodelle [27, 98] oder 'Perlenkettenmodelle' ('bead models', [99]), leisteten bereits wertvolle Dienste bei der theoretischen Untersuchung anderer langsamer Vorgänge der Proteindynamik, insbesondere von Faltungsprozessen [100, 101].

Selbstverständlich erfordern die solchen Modellsystemen innewohnenden einschneidenden Vereinfachungen größte Vorsicht bei der Interpretation der Ergebnisse, falls Aussagen über dynamische Eigenschaften von Proteinen getroffen werden sollen. Deshalb wird in Kapitel 3 zunächst die Dynamik unseres Proteinmodells untersucht und mit dem Verhalten detaillierter Proteinmodelle verglichen um sicherzustellen, daß das Modell trotz seiner vergleichsweise einfachen Struktur echten Proteinen hinreichend ähnelt.

Wie sich zeigen wird, lassen sich aus den am vereinfachten Modell durchgeführten

Langzeitsimulationen tatsächlich neue Aussagen über die Konformationsdynamik von Proteinen ableiten. Ebenso sollten anhand des Modells abgeleitete Aussagen über die Qualität der untersuchten MD-Algorithmen allgemeinen Charakter besitzen: Die sich aus den Testsimulationen am vereinfachten Proteinmodell ergebende problemangepaßte Bewertung sollte auch für MD-Simulationen realistischerer Proteinmodelle ihre Gültigkeit behalten.

Die Untersuchung geeignet vereinfachter Minimalmodelle der Proteindynamik birgt damit den Schlüssel zur Überwindung des benannten Teufelskreises, da nur auf diese Weise die benötigten ausgedehnten Simulationsrechnungen ermöglicht werden. So erlaubte das in der vorliegenden Arbeit vorgeschlagene Proteinmodell erstmals die Durchführung außerordentlich langer Molekulardynamiksimulationen über eine Zeitspanne von insgesamt $0,232\,\mu$s. Diese ermöglichten Untersuchungen der Konformationsdynamik, die zur Formulierung effizienter Beschreibungen der Proteindynamik beitragen dürften.

Teile der in Kap. 4 dargestellten Ergebnisse wurden in [102] vorgestellt und zur Publikation eingereicht [103]; das vereinfachte Proteinmodell und die in Kap. 5 dargestellten Ergebnisse werden in [104] veröffentlicht.

'The gravitational n-body problem poses a formidable challenge to the numerical analyst.'

Sverre J. Aarseth in **Multiple Time Scales**, 1985

Kapitel 2

Molekulardynamik (MD)

Die Methode der Molekulardynamiksimulation zur Beschreibung atomarer Bewegungen in molekularen Systemen findet zunehmende Verbreitung in einer Vielzahl verschiedener Anwendungen. Seit der ersten Dynamiksimulation eines Ensembles zweidimensionaler harter Scheiben im Jahr 1957 [105] und den anschließenden MD-Rechnungen an flüssigem Argon über Zeitspannen von einigen zehn Pikosekunden [106, 107] wurden beachtliche methodische und rechentechnische Fortschritte erzielt. Die erste Dynamiksimulation eines Proteins (BPTI), noch ohne Einbeziehung seiner atomaren Umgebung, wurde 1977 durchgeführt [43]. Seitdem konnten dank der schnell wachsenden Zahl bekannter Proteinstrukturen, verfeinerter Methoden und dank der zunehmenden Leistungsfähigkeit verfügbarer Computerhard- und Software eine Vielzahl komplexer Systeme wie photosynthetische Proteine [17,57,61, 93,108,109] oder Protein-DNS-Komplexe [110] untersucht werden. Jüngste Beispiele außerordentlich umfangreicher Rechnungen sind die Simulation einer Lipidmembran in wäßriger Lösung [111] sowie des bereits erwähnten Membranproteins Bakteriorhodopsin in nativer Umgebung [112]. Hier erstreckten sich die Simulationen über einige hundert Pikosekunden.

All diesen Untersuchungen liegt eine Reihe vereinfachender Annahmen zugrunde, welche einige physikalisch/chemische Phänomene von der Behandlung ausschließen. Die Zuverlässigkeit, mit der MD-Simulationen zugängliche Phänomene beschrieben werden können, hängt entscheidend von Details des verwendeten molekularen Modells ab, etwa von der Behandlung der langreichweitigen elektrostatischen Wechselwirkungen[1] oder der molekularen Umgebung. Nach einer kurzen Einführung

[1] Eine Übersicht gängiger Methoden ist in [113] gegeben.

der Grundlagen der MD-Methode[2] wollen wir uns deshalb Verfahren zur Berechnung elektrostatischer Kräfte zuwenden.

2.1 Die Methode der MD-Simulation

Die exakte Beschreibung der atomaren Dynamik eines Moleküls erfordert die Lösung der zeitabhängingen Schrödingergleichung für das aus Atomkernen und Elektronen bestehende Vielteilchensystem. Dessen Hamiltonoperator setzt sich aus der kinetischen Energie der Kerne und der Elektronen sowie aus einem elektrostatischen Beitrag zusammen, der die Coulomb-Wechselwirkung der Teilchen beschreibt. Schon für kleine Moleküle mit wenigen Atomen ist der Rechenaufwand zur Lösung dieser Gleichung jedoch immens, so daß die Behandlung größerer Systeme, wie sie Proteine darstellen, eine Reihe von Näherungen erfordert.

Zunächst wird eine Beschreibung der schnellen elektronischen Freiheitsgrade angestrebt. Die im Vergleich zur Dynamik der leichten Elektronen langsame Bewegung der schweren Kerne erlaubt im Rahmen der bekannten Born-Oppenheimer-Näherung die Abspaltung einer zeitunabhängigen Schrödingergleichung für die elektronische Dynamik. Diese definiert die Eigenzustände des Elektronensystems bei adiabatisch veränderlich angenommener Kernkonfiguration; die zugeordneten Energieniveaus beschreiben den mittleren Einfluß der elektronischen Freiheitsgrade auf die Kernbewegung in Form eines *effektiven* Potentials. Ist dieses bekannt, so reduziert sich die Beschreibung der molekularen Dynamik auf die Bestimmung der Kernbewegung auf der effektiven Potentialfläche.

Für Systeme mit einigen zehn Atomen können die benötigten quasistationären elektronischen Zustände bei gegebener Kernkonfiguration mit Hilfe sogenannter 'ab initio'-Methoden in guter Näherung berechnet werden. Zur Behandlung von Systemen mit mehreren hundert Atomen haben sich halbempirische Verfahren wie die MNDO-Methode [114] bewährt. Es ist jedoch zum gegenwärtigen Zeitpunkt nicht möglich, auf diese Weise die Elektronensysteme ganzer Proteine zu beschreiben und daraus die Potentialfläche für die Kernbewegung direkt abzuleiten.

Man ersetzt das effektive Potential deshalb durch ein 'molekulares Kraftfeld', dessen funktionale Abhängigkeit von der Kernkonfiguration *empirisch* begründet ist. Darin wird der elektronische Wechselwirkungsbeitrag als Summe intramolekularer Wechselwirkungen modelliert. Diese umfassen üblicherweise sowohl chemisch motivierte

[2]Eine ausführlichere Darstellung der Methode und ihrer Anwendung findet sich z.B. in [44].

lokale Terme wie Bindungsstreckpotentiale, Bindungswinkelpotentiale, Torsionspotentiale und Kräfte zur Beschreibung von Wasserstoffbrückenbindungen als auch nichtlokale Wechselwirkungen wie die langreichweitigen Coulomb-Kräfte zwischen geladenen Atomen und die Van der Waals-Anziehung. Jeder dieser Terme ist von einfacher Form; beispielsweise kann der Energiebeitrag der Längenänderung einer chemischen Bindung als harmonisch angenommen werden. Eine Reihe verschiedener Kraftfelder wurde entwickelt [76, 110, 115–118] und mit Hilfe quantenmechanischer Rechnungen an kleinen Molekülen sowie durch Vergleich von MD-Simulationen mit experimentellen Daten schrittweise verfeinert. Diese Vorgehensweise ist z.B. in [119] beschrieben.

Zur weiteren Vereinfachung wird im Rahmen der MD-Methode die Kernbewegung *klassisch* behandelt. Entsprechend werden die N Atome eines Moleküls als Massenpunkte der Masse m_α aufgefaßt[3] und die interatomaren Kräfte $\mathbf{F}_\alpha$ aus der empirischen Energiefunktion $E(\mathbf{r}_1, \ldots, \mathbf{r}_N)$ durch Gradientenbildung gewonnen. Die molekulare Dynamik, d.h. die Zeitentwicklung der Atompositionen $\mathbf{r}_\alpha$, ergibt sich aus der numerischen Lösung der Bewegungsgleichungen

$$m_\alpha \frac{d^2}{dt^2}\mathbf{r}_\alpha = \mathbf{F}_\alpha(\mathbf{r}_1, \ldots, \mathbf{r}_N) \quad \text{mit} \quad \alpha = 1, \ldots, N\,. \tag{2.1}$$

Die erwähnten Näherungen, insbesondere die klassische Behandlung der Kernbewegungen in einem effektiven elektronischen Potential, schließen eine Reihe physikalischer Phänomene von der Beschreibung aus: So können chemische Reaktionen sowie die detaillierte Dynamik der leichten Wasserstoffkerne mit der Methode der MD-Simulation nicht behandelt werden. Auch deren Anwendung auf Vorgänge bei tiefen Temperaturen ist wegen des quantenmechanischen Charakters niederenergetischer Kernbewegungen problematisch. Bei Berücksichtigung dieser Effekte stellen MD-Simulationen jedoch eine für viele Aspekte gültige Beschreibung der Proteindynamik dar [44].

Zur numerische Integration der Bewegungsgleichungen (2.1) hat sich die Nyström-Methode [120] zweiter Ordnung, $\mathcal{N}^{II}$, als geeignet erwiesen. Dieses Verfahren wurde erstmals 1967 von L. Verlet für MD-Simulationen eingesetzt [107] und ist deshalb in diesem Gebiet als 'Verlet-Algorithmus' bekannt. Der Algorithmus berechnet die durch (2.1) definierte Bewegung $\mathbf{r}_i := \mathbf{r}(t_i)$ jedes Atoms[4] zu diskreten Zeitpunkten

[3] Wie in der Molekülphysik üblich, bezeichnen griechische Indizes Atomkerne.

[4] Der Übersichtlichkeit halber wird generell bei der Behandlung numerischer Integrationsverfahren auf den Atomindex α verzichtet.

$t_i := i\Delta_t$ mit Hilfe der Rekursionsformel

$$\mathbf{r}_{i+1} := \mathcal{N}^{II}(\mathbf{r}_{i-1}, \mathbf{r}_i) \equiv 2\mathbf{r}_i - \mathbf{r}_{i-1} + \frac{(\Delta_t)^2}{m}\mathbf{F}_i\,, \tag{2.2}$$

wobei Δ_t die Integrationsschrittweite bedeutet. Die Bezeichnung des Integrationsschritts mit *lateinischen* Indizes soll hier wie in allen folgenden Gleichungen eine Verwechslung mit der griechisch indizierten Atomnummer vermeiden. Der Verlet-Algorithmus ist robust genug, um relativ große Integrationsschrittweiten im Bereich einer Femtosekunde zu erlauben, darüberhinaus ist im Gegensatz zu den bekannten Runge-Kutta-Verfahren pro Integrationsschritt nur *eine* rechenzeitbestimmende Kraftberechnung erforderlich. Ferner ist (2.2) invariant gegenüber einer Zeitumkehr, da ein Integrationsschritt 'vorwärts', gefolgt von einem Schritt 'rückwärts' auf den Ausgangszustand zurückführt:

$$\mathcal{N}^{II}(\mathbf{r}_{i+1}, \mathbf{r}_i) = \mathcal{N}^{II}(\mathcal{N}^{II}(\mathbf{r}_{i-1}, \mathbf{r}_i), \mathbf{r}_i) = \mathbf{r}_{i-1}\,. \tag{2.3}$$

Als wesentliche Konsequenz dieser Eigenschaft hält diese Methode die Gesamtenergie konservativer Vielteilchensysteme in sehr guter Näherung konstant (vgl. hierzu auch [121] sowie Arbeiten zu 'symplektischen' Algorithmen [66, 122]). Alle in Abschnitt 2.3 beschriebenen effizienten Algorithmen bauen deshalb auf dem Verlet-Algorithmus auf.

2.2 Langreichweitige Kräfte

Viele Proteine denaturieren bei Änderung des Säuregrads des sie umgebenden Lösungsmittels. Da der pH-Wert den Protonierungszustand saurer oder basischer Aminosäuren und damit deren Ladung bestimmt, muß man schließen, daß intramolekulare Coulomb-Kräfte wesentlich zur Stabilisierung der Proteinstruktur beitragen. In der Tat zeigen MD-Simulationen, daß etwa die Vernachlässigung der langreichweitigen Anteile dieser Kräfte die Proteinstruktur signifikant verändern kann [79, 123].

Auch vom numerischen Standpunkt aus betrachtet muß langreichweitigen Kräften besondere Aufmerksamkeit geschenkt werden. Während bei der numerischen Integration der Bewegungsgleichungen der Aufwand zur Berechnung der *kurzreichweitigen* chemischen Bindungskräfte annähernd proportional zur Anzahl der Atome im System ist, müssen zur Berechnung der *langreichweitigen* Coulomb-Energie

$$E_{\text{Coulomb}} = \sum_{\langle\alpha,\beta\rangle} E^{\alpha\beta}_{\text{Coulomb}} \sim \sum_{\langle\alpha,\beta\rangle} \frac{q_\alpha q_\beta}{r_{\alpha\beta}} \tag{2.4}$$

alle interatomaren *Paar*wechselwirkungen berücksichtigt werden, und der Rechenaufwand skaliert *quadratisch* mit der Anzahl der Atome. Entsprechend erstreckt sich die Summe in (2.4) über alle Atompaare $\langle\alpha,\beta\rangle$, wobei $r_{\alpha\beta}$ deren Abstand und q_α die Partialladung von Atom α ist. Aufgrund des beschriebenen Skalierungsverhaltens wird bei zunehmender Systemgröße die Berechnung der Coulombkräfte rasch zum rechenzeitbestimmenden Schritt. Aus diesem Grund zielt die Mehrzahl der Methoden zur Beschleunigung von MD-Simulationen auf eine möglichst effiziente Approximation der Coulombkräfte (2.4) ab.

2.3 Effiziente MD-Methoden

Eine weitverbreitete Näherung ist die Methode der 'Abschneidefunktionen' ('cut-off'), die im weiteren mit 'CUT' bezeichnet werden soll. Diese vernachlässigt Paarwechselwirkungen zwischen Atomen, deren Abstand einen bestimmen festzulegenden Wert, den 'Abschneideradius' r_{cut}, übersteigt [76]. Dazu wird das Coulomb-Potential (2.4) mit einer abstandsabhängigen Abschneidefunktionen $f(r)$ gewichtet. Abbildung 2.1 (oben) zeigt als Beispiel den Verlauf der Abschneidefunktion $f(r) := (1 - r^2/r_{cut}^2)^2$ in Einheiten von r_{cut}: Für $r = 0$ beträgt ihr Wert Eins, und für wachsende Abstände reduziert sie die Kräfte stetig differenzierbar auf Null. Abbildung 2.1 (unten) vergleicht den Verlauf des entsprechend genäherten Coulomb-Potentials $f(r) \cdot E^{\alpha\beta}_{\text{Coulomb}}$ mit dessen exakten Verlauf $E^{\alpha\beta}_{\text{Coulomb}}$ (2.4). Wie zu erkennen ist, verschwinden in dieser Näherung alle Kräfte zwischen Atompaaren, deren Abstand r_{cut} übersteigt, wodurch sich die Summe (2.4) auf die verbleibenden Paarwechselwirkungen innerhalb r_{cut} reduziert. Für MD-Simulationen von Proteinen werden üblicherweise Abschneideradien im Bereich 8 Å – 15 Å verwendet.

Wie in Kapitel 4 gezeigt wird, kann die Verwendung einer Abschneidefunktion in MD-Simulationen jedoch nicht nur die räumliche Struktur des untersuchten Systems verändern, sondern auch dessen dynamische Eigenschaften. Es soll betont werden, daß diese Aussage *nicht* impliziert, daß die 'exakte' Berechnung der Coulomb-Kräfte gemäß (2.4) in jedem Fall zu einer gegenüber der Abschneide-Methode verbesserten Beschreibung der Proteindynamik führt. Tatsächlich läßt sich das Abschneiden der langreichweitigen Coulomb-Anteile, entsprechend einer entfernungsabhängigen Dielektrizitätskonstanten, als effektive Beschreibung von Abschirmeffekten, die durch atomare Polarisierbarkeiten hervorgerufen werden, auffassen. Dies darf jedoch nur als grobe Näherung gelten: Zum einen können Polarisierungseffekte in heterogen polarisierbaren Systemen, wie es Proteine darstellen, vermittels ei-

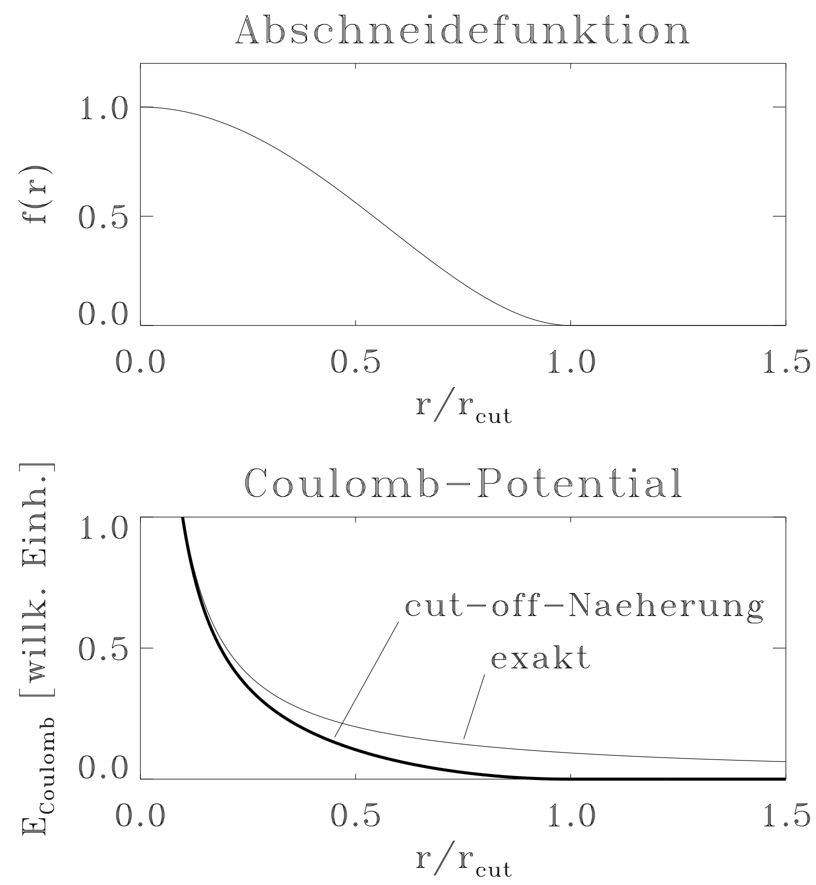

Abbildung 2.1: Beispiel einer Abschneidefunktion, wie sie etwa im Progamm XPLOR [124] verwendet wird; oben: $f(r)$ hat den Wert Eins für $r = 0$ und fällt im Bereich $0 < r < r_{cut}$ stetig differenzierbar auf Null ab; unten: Die Abschneidefunktion verändert den Verlauf des Coulomb-Potentials; für Atomabstände $r > r_{cut}$ verschwindet das Potential, und es wirkt keine Kraft.

ner entfernungsabhängigen Dielektrizitätskonstanten nur ungenügend beschrieben werden [88], vielmehr erfordert deren exakte Behandlung die Lösung der Poisson-Boltzmann-Gleichung [125,126]. Zum anderen sprechen keine physikalischen Gründe dafür, die elektrostatische Wechselwirkung bereits innerhalb der vergleichsweise kurzen Distanz r_{cut} völlig verschwinden zu lassen; ein exponentiell abnehmender Verlauf scheint hier realistischer. Eine verbesserte Einbeziehung atomarer Polarisierungen ist also in jedem Fall anzustreben (vgl. die diesbezüglichen Arbeiten [127–129]). Effiziente Methoden zur Behandlung langreichweitiger Wechselwirkungen sind hierfür Voraussetzung.

Im folgenden sollen zwei Methoden zur Reduzierung des mit der Bestimmung der Coulomb-Kräfte verbundenen Rechenaufwands geschildert werden, welche langreichweitige Anteile näherungsweise berücksichtigen und so Artefakte vermeiden sollen.

Dies sind eine Gruppe von Entfernungsklassenalgorithmen ('distance class methods', DC) sowie schnelle Multipolverfahren ('fast multipole methods', FMM). Ein Großteil der Aufmerksamkeit soll dabei den Entfernungsklassenmethoden gelten, welche in der Diplomarbeit des Autors verallgemeinert [61, 77] und im Rahmen der vorliegenden Dissertation weiter verfeinert und den Eigenschaften der Proteindynamik angepaßt wurden. Die Qualität dieser Verfahren wird in Kapitel 4 anhand der ebenfalls in dieser Arbeit entwickelten problemangepaßten Bewertungskriterien mit der Qualität der Abschneide-Methode verglichen.

2.3.1 Entfernungsklassenmethoden

Entfernungsklassenmethoden sind Mehrschrittverfahren [73, 89, 90, 130, 131], welche auf der Beobachtung beruhen, daß in Proteinen die Zeitableitung interatomarer Paarwechselwirkungen mit wachsendem Atomabstand im allgemeinen abnimmt. Abbildung 2.2 stellt diese Situation ist schematisch in dar: Die mit $\mathbf{F}^{(0)}$ bezeichnete Kurve illustriert den zeitlichen Verlauf der Kraft, die auf ein bestimmtes Atom von allen unmittelbar benachbarten Atomen im Protein ausgeübt wird. Darunter, mit $\mathbf{F}^{(1)}$ bezeichnet, ist die von weiter entfernt liegenden Atomen ausgehende Kraft dargestellt und im unteren Bild ($\mathbf{F}^{(2)}$) der Einfluß der entferntesten Atome im System. Man erkennt, daß in dieser Reihe entfernungsabhängiger Kräfte sowohl die Amplitude als auch die Frequenz der Kraftfluktuationen abnimmt.

Entsprechend werden im Rahmen eines Entfernungsklassenschemas (vgl. [71]) die langsam veränderlichen langreichweitigen Kraftbeiträge im Verlauf der numerischen Integration der Bewegungsgleichungen weniger häufig exakt berechnet als die schnell und stark fluktuierenden kurzreichweitigen Anteile (in Abb. 2.2 sind die expliziten Berechnungen durch schwarze Kreise gekennzeichnet). Stattdessen wird die überwiegende Mehrzahl der benötigten Kräfte durch Extrapolationen abgeschätzt (weiße Kreise) und mit den so eingesparten expliziten Kraftberechnungen der Rechenaufwand stark reduziert.

Die Frequenz expliziter Kraftberechnungen wird für jedes Atompaar durch Zuordnung zu einer 'Entfernungsklasse' festgelegt, die gemäß seines Abstands $r_{\alpha\beta}$ wie folgt gewählt wird: Sei $\{R_0, \ldots, R_J\}$ eine ansteigende Folge von $J+1$ Radien mit $R_0 = 0$. Dann wird die Menge der Atompaare $\langle\alpha, \beta\rangle$, für die $R_j \leq r_{\alpha\beta} < R_{j+1}$ erfüllt ist, Entfernungsklasse j genannt. Abbildung 2.2 (links) zeigt dieses Entfernungsklassenschema: Bezüglich des schwarzen Atoms in der Mitte befinden sich die dunkelgrauen Atome in Entfernungsklasse 0, die hellgrauen in Klasse 1, und die weißen in Klasse 2. Die konzentrischen Kreise markieren dabei die Abstände R_1, R_2 und R_3. Die

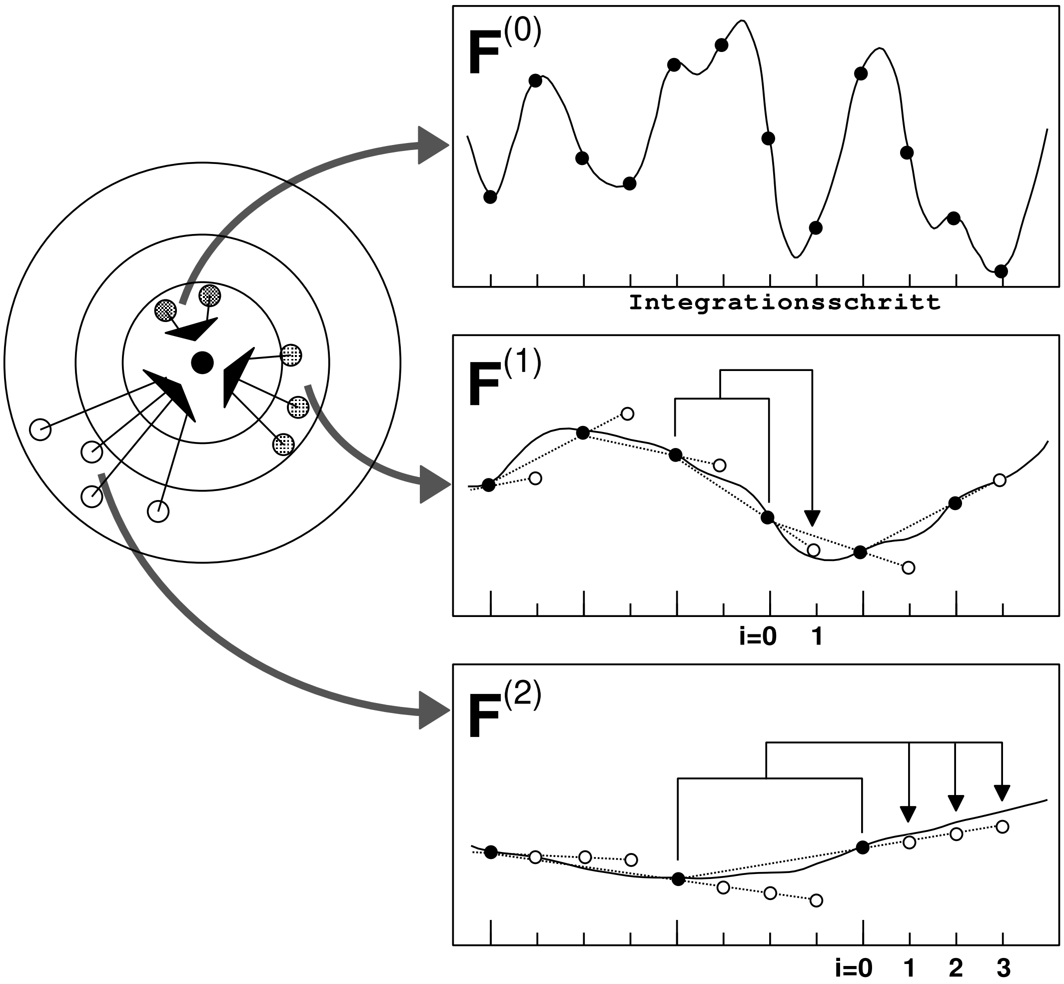

Abbildung 2.2: Entfernungsklassenmethoden nützen Eigenschaften interatomarer Paarwechselwirkungen. In Proteinen nimmt deren Zeitableitung mit wachsendem Abstand ab. Die mit $\mathbf{F}^{(0)}$, $\mathbf{F}^{(1)}$ und $\mathbf{F}^{(2)}$ bezeichneten Bilder stellen entsprechende Kraftfluktuationen schematisch dar; oben: zeitlicher Verlauf der auf ein bestimmtes Atom (links in der Abbildung schwarz dargestellt) von den nächstgelegenen Atomen (dunkelgrau) ausgeübten Kraft; Mitte: von weiter entfernten Atomen (hellgrau) ausgehende Kraft; unten: Kraft der entferntesten Atome im System (weiß). Bei der numerischen Integration der Bewegungsgleichungen kann die exakte Berechnung der Kräfte $\mathbf{F}^{(0)}$, $\mathbf{F}^{(1)}$ und $\mathbf{F}^{(2)}$ mit abnehmender Häufigkeit erfolgen: Die schnell fluktuierende Kraft $\mathbf{F}^{(0)}$ wird in jedem Integrationsschritt exakt ermittelt (schwarze Punkte auf der Kurve), die langsamer veränderliche Kraft $\mathbf{F}^{(1)}$ wird lediglich jeden zweiten Schritt berechnet, $\mathbf{F}^{(2)}$ nur jeden vierten usw. Die zur Integration benötigten übrigen Kräfte (weiße Kreise auf den Kurven) werden unter Verwendung zurückliegender exakter Kräfte abgeschätzt (Pfeile), etwa, wie in der Abbildung dargestellt, durch lineare Extrapolation (vgl. Methode DC-1b).

Summen (2.4) und die Gradienten zur Ermittlung der auf jedes einzelne Atom von allen Atomen aus Entfernungsklasse j ausgeübten Kraft $\mathbf{F}^{(j)}$ werden lediglich jeden 2^j-ten Integrationsschritt explizit berechnet. Alle anderen in (2.2) zur Ausführung der $2^j - 1$ dazwischenliegenden Integrationsschritte benötigten Kraftbeiträge werden stattdessen unter Verwendung älterer Kräfte extrapoliert. Die Zahl möglicher Extrapolationsmethoden ist unbeschränkt; die Wahl einer bestimmten Extrapolationsmethode definiert den zugehörigen Entfernungsklassenalgorithmus. Sechs verschiedene Extrapolationsverfahren sollen im folgenden diskutiert werden. Die resultierenden Algorithmen werden in Kapitel 4 ebenfalls beurteilt und vergleichen.

In [91] wurde vorgeschlagen, die Abweichung der intermediären Kräfte $\mathbf{F}^{(j)}$ von der zuletzt exakt berechneten Kraft als gering anzunehmen und diese Kräfte deshalb im Sinne einer Extrapolation nullter Ordnung durch die zuletzt exakt berechnete Kraft zu ersetzen. Dieses Verfahren soll entsprechend als DC-0 (distance class, 0^{th} order extrapolation) bezeichnet werden. Abbildung 2.3 (DC-0) stellt den sich ergebenden zeitlichen Verlauf der Kraft $\mathbf{F}^{(j)}$ für den Fall $j = 3$ schematisch dar: Der exakte Verlauf ist fett gezeichnet, die dünnen, senkrechten Linien repräsentieren die im Rahmen der Entfernungsklassennäherung abgeschätzten und in (2.2) eingehenden Kräfte. Alle $2^j = 8$ Integrationsschritte stimmen die verwendeten Kräfte mit den exakten überein; die Extrapolation nullter Ordnung führt zu dem in der Abbildung erkennbaren treppenstufenartigen Verlauf der genäherten Kräfte.

Bei Anwendung von Methode DC-0 auf konservative Systeme, etwa für MD-Simulationen *in vacuo*, tritt eine unphysikalische Drift der Gesamtenergie [61] auf (siehe auch Kapitel 4), die in den durch die Kraftextrapolation verursachten Abweichungen von den exakten Kräften begründet liegt. Diese in Abbildung 2.3 (DC-0) erkennbaren *algorithmischen Rauschkräfte* erhöhen im allgemeinen die kinetische Energie der Atome, ohne deren mittlere potentielle Energie entsprechend zu vermindern, und führen so zu der beobachteten Energiedrift.

Entgegen dem ersten Anschein führen derartige Rauschkräfte jedoch nicht *notwendig* zu einer Verletzung der Energieerhaltung: Wie in der Diplomarbeit des Autors [61] gezeigt wurde, kann in Spezialfällen durch geeignete Wahl der Kraftextrapolation die Energiedrift *trotz des Auftretens algorithmischer Rauschkräfte* zum Verschwinden gebracht werden. Diese Beobachtung führte dort zur Entwicklung zweier Entfernungsklassenalgorithmen, die hier mit DC-i bzw. DC-1a bezeichnet werden. Die dieser Entwicklung zugrundeliegende Argumentation ist zum Verständnis der im Rahmen der vorliegenden Arbeit entwickelten Verfeinerung notwendig und soll deshalb kurz zusammengefaßt werden.

Zunächst bemerken wir, daß DC-0 — im Gegensatz zum Verlet-Algorithmus — *nicht* zeitumkehrinvariant ist. Tatsächlich ist es prinzipiell unmöglich, eine Extrapolation anzugeben, die zu einem zeitumkehrinvarianten Entfernungsklassenalgorithmus

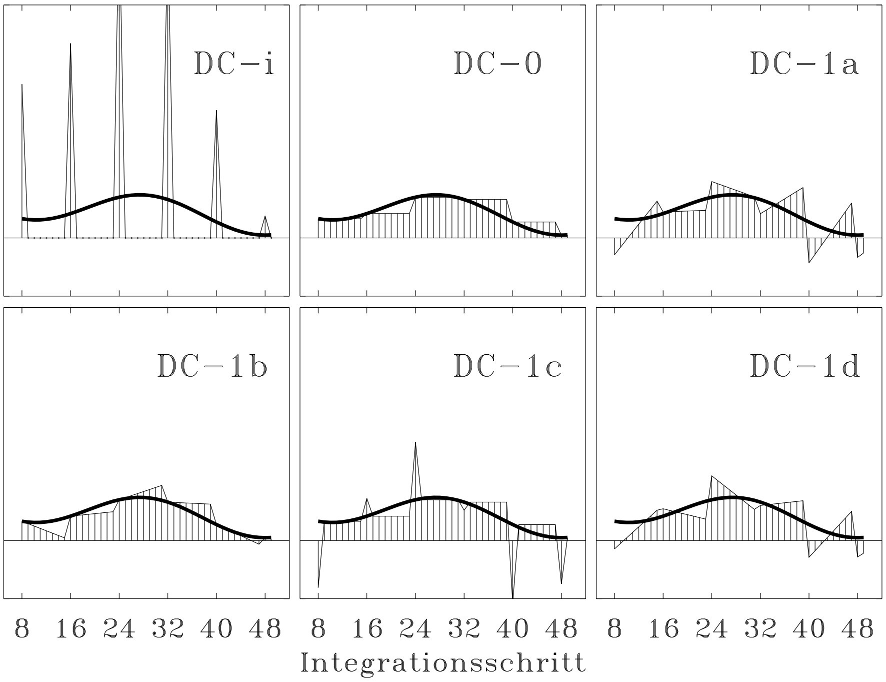

Abbildung 2.3: Vergleich eines exakten Kraftverlaufs (fett gezeichnet) mit im Rahmen der sechs im Text beschriebenen Entfernungsklassenalgorithmen extrapolierten Kraftbeiträgen (dünne, senkrechte Linien); diese Näherungen werden bei der Integration der Newton'schen Bewegungsgleichungen verwendet.

führt. Dies liegt im intrinsisch zeitgerichteten Charakter einer approximativen Vorhersage begründet: Der Wert einer auf der Grundlage *vergangener* Größen vorhergesagten, aber durch diese nicht eindeutig bestimmten Observablen wird sich im allgemeinen von dem entsprechenden Wert unterscheiden, der unter Verwendung *zukünftiger* Größen 'nachhergesagt'[5] wurde. Eine (2.3) entsprechende Gleichung, welche die Voraussetzung für eine Zeitspiegelungssymmetrie bildet, ist damit ausgeschlossen.

Die vorgeschlagene Wahl der Kraftextrapolation gründet sich auf die Erwartung, daß durch Vermeiden einer Energiedrift in den angesprochenen Spezialfällen die Drift im Allgemeinfall zumindest reduziert werden kann. Ferner wird angenommen, daß dies tatsächlich eine Verbesserung des Algorithmus bewirkt (in Kapitel 4 soll letztere Erwartung kritisch hinterfragt werden). Der angesprochene Spezialfall wird durch die Besetzung lediglich einer einzigen Entfernungsklasse j definiert, d.h., für alle Atompaare α, β gilt $R_j \leq r_{\alpha\beta} < R_{j+1}$. Im Trivialfall $j = 0$ geht der Entfernungs-

[5]Dieser Begriff wurde [132] entnommen; dort wird er in ähnlichem Kontext zur Diskussion des zweiten Hauptsatzes der Thermodynamik und dessen Zeitumkehrinvarianz herangezogen.

klassenalgorithmus in den konventionellen Verlet-Algorithmus über, da dann *alle* Kräfte (jede 2^0-te) exakt berechnet werden. Die Konstanz der Gesamtenergie für den Fall $j \neq 0$ wird nun durch die Forderung erzwungen, daß in dieser Situation der Entfernungsklassenalgorithmus in einen Verlet-Algorithmus der 2^j-fachen Integrationsschrittweite übergeht. Diese Äquivalenz wird als 'Verlet-Kriterium' bezeichnet.

Zur Ableitung dieses Kriteriums wurden in der Diplomarbeit des Autors [61] Extrapolationen betrachtet, die eine *lineare* Funktion der $E+1$ zurückliegenden exakt berechneten Kräften sind, d.h.,

$$\mathbf{F}^{(j)}_{kn_j+i} = \sum_{r=0}^{E} a^{(j)}_{i,r} \mathbf{F}^{(j)}_{(k-r)n_j} \,, \tag{2.5}$$

wobei die $a^{(j)}_{i,r}$ festzulegende Extrapolationskoeffizienten sind und $n_j := 2^j$ gesetzt wurde. Für den angegebenen Spezialfall läßt sich die Rekursion (2.2) umschreiben in

$$\left.\begin{aligned} \mathbf{r}_{(k+2)n_j} = 2\mathbf{r}_{(k+1)n_j} - \mathbf{r}_{kn_j} + \frac{(\Delta_t)^2}{m} \Bigg\{ & \mathbf{F}^{(j)}_{(k+1)n_j} \sum_{i=0}^{n_j-1} (n_j - i) a^{(j)}_{i,0} \\ + & \sum_{r=0}^{E-1} \mathbf{F}^{(j)}_{(k-r)n_j} \sum_{i=0}^{n_j-1} \left[i a^{(j)}_{i,r} + (n_j - i) a^{(j)}_{i,r+1} \right] \\ + & \mathbf{F}^{(j)}_{(k-E)n_j} \sum_{i=0}^{n_j-1} i a^{(j)}_{i,E} \Bigg\} \quad , \quad k \in \mathcal{Z} \,. \end{aligned}\right\} \tag{2.6}$$

Wie in [61] gezeigt wurde, ist diese Rekursion genau dann äquivalent zu einem Verlet-Algorithmus, wenn der Kraftterm in (2.6) gleich $(n_j \Delta_t)^2 \mathbf{F}^{(j)}_{(k+1)n_j}/m$ ist und (2.6) damit in (2.2) mit Integrationsschrittweite $n_j \Delta_t$ übergeht. Somit nimmt das Verlet-Kriterium die Form eines linearen Systems mit $E+2$ Bestimmungsgleichungen für die $(E+1)n_j$ Extrapolationskoeffizienten $a^{(j)}_{i,r}$ an:

$$\left.\begin{aligned} \sum_{i=0}^{n_j-1} (n_j - i) a^{(j)}_{i,0} &= n_j^2 \\ \sum_{i=0}^{n_j-1} \left(i a^{(j)}_{i,r} + (n_j - i) a^{(j)}_{i,r+1} \right) &= 0 \quad , \quad r = 1, \ldots, E-1 \\ \sum_{i=0}^{n_j-1} i a^{(j)}_{i,E} &= 0 \quad . \end{aligned}\right\} \tag{2.7}$$

In [61] wird zunächst der Fall $E=0$ betrachtet, d.h., nur die jeweils zuletzt exakt berechnete Kraft wird zur Abschätzung verwendet. Werden nur positive Extrapolationskoeffizienten $a_{i,0}^{(j)}$ zugelassen, so legt das Verlet-Kriterium die Lösung

$$a_{i,0}^{(j)} = \begin{cases} 2^j & \text{für} \quad i = 0 \\ 0 & \text{für} \quad i = 1, 2, \ldots, 2^j - 1 \end{cases} \tag{2.8}$$

und damit den Entfernungsklassenalgorithmus DC-i ('impulse') eindeutig fest. Dieser Algorithmus besitzt große Ähnlichkeit zu dem in [133] auf anderem Weg abgeleiteten RESPA-Algorithmus. Tatsächlich findet in DC-i keine Extrapolation im eigentlichen Sinne statt. Vielmehr werden die Kräfte aller n_j aufeinanderfolgenden Integrationsschritte kombiniert und während eines einzigen Integrationsschrittes gleichzeitig ausgeübt. Abbildung 2.3 (DC-i) zeigt den sich ergebenden Verlauf, der hohe Kraftspitzen aufweist. Da die Stärke dieser Kraftspitzen exponentiell mit der Anzahl verwendeter Entfernungsklassen zunimmt, besteht hier — im Gegensatz zu DC-0 — die Gefahr von Resonanzinstabilitäten, falls mehr als drei Entfernungsklassen verwendet werden. In [77] wird dies anhand eines einfachen Beispiels näher erläutert. Ganz allgemein kann angenommen werden, daß eine möglichst gleichmäßige Verteilung der Kräfte von n_j aufeinanderfolgenden Integrationsschritten die Stabilität des Entfernungsklassenalgorithmus erhöht, während Diskontinuitäten im Kraftverlauf die Gefahr von Instabilitäten in sich bergen. Ausgehend von dieser Annahme ist es sinnvoll, ein *Glattheitskriterium* für Extrapolationen zu fordern, das Kraftverläufe mit geringen Unstetigkeiten favorisiert.

Die Extrapolation (2.5) kann durch Miteinbeziehung weiterer, zu früheren Zeitpunkten exakt berechneter Kräfte verbessert werden. Vier Methoden erster Ordnung ($E = 1$), DC-1a, DC-1b, DC-1c und DC-1d, sollen hier betrachtet werden. DC-1a wurde in der Diplomarbeit des Autors [61] durch Variation der Koeffizienten abgeleitet; darauf aufbauend hat R. Skeel die Methode DC-1c vorgeschlagen [134]. Die vorliegende Arbeit kombiniert die gewonnenen Erkenntnisse und leitet den optimierten Entfernungsklassenalgorithmus DC-1d ab. Tabelle 2.1 führt die Extrapolationskoeffizienten $a_{i,r}^{(j)}$ ($r \in \{0, 1\}$), welche die vier Methoden nach (2.5) definieren, zusammen mit den Koeffizienten für DC-i und DC-0 auf.

Eine entsprechend dem oben vorgeschlagenen Glattheitskriterium optimale Wahl der Extrapolationskoeffizienten besteht in der Extrapolation

$$\left.\begin{aligned} a_{i,0}^{(j)} &= 1 + i/n_j \\ a_{i,1}^{(j)} &= -i/n_j \,, \end{aligned}\right\} \tag{2.9}$$

welche die durch zwei aufeinanderfolgende exakte Kräfte gegebene Zeitentwicklung *linear* fortsetzt. Der entsprechende Entfernungsklassenalgorithmus, der mit DC-1b bezeichnet werden soll, wurde, soweit bekannt, noch nicht in MD-Simulationen eingesetzt. In Abbildung 2.3 (DC-1b) ist der glatte Verlauf der extrapolierten Kräfte zu erkennen.

Algorithmus	$a_{i,0}^{(j)}$	$a_{i,1}^{(j)}$
DC-i	n_j für $i = 0$; 0 sonst	–
DC-0	1	–
DC-1a	$6n_j \dfrac{(n_j^2 - n_j + 1) - i(n_j - 2)}{(n_j + 1)(2n_j^2 + 1)}$	$\dfrac{2(-2n_j^2 + 3n_j - 1) + 6i(n_j - 1)}{2n_j^2 + 1}$
DC-1b	$1 + i/n_j$	$-i/n_j$
DC-1c	$(n_j + 1)/2$ für $i = 0$; 1 sonst	$(1 - n_j)/2$ für $i = 0$; 0 sonst
DC-1d	$\dfrac{3n_j^2 - 2n_j + 1}{n_j(n_j + 1)} - 3i\dfrac{n_j - 1}{n_j(n_j + 1)}$	$1 - \dfrac{3n_j^2 - 2n_j + 1}{n_j(n_j + 1)} + 3i\dfrac{n_j - 1}{n_j(n_j + 1)}$

Tabelle 2.1: Die im Text beschriebenen Entfernungsklassenalgorithmen unterscheiden sich in ihren Extrapolationskoeffizienten $a_{i,r}^{(j)}$. Die mit DC-i und DC-0 bezeichneten Algorithmen verwenden lediglich die zuletzt exakt berechneten Kräfte; entsprechend ist hier nur $a_{i,0}^{(j)}$ aufgeführt. Die übrigen Algorithmen, DC-1a, DC-1b, DC-1c und DC-1d verwenden Extrapolationen erster Ordnung ($E = 1$); entsprechend sind für diese Algorithmen die Koeffizienten $a_{i,0}^{(j)}$ und $a_{i,1}^{(j)}$ angegeben. Zur Abkürzung wurde $n_j := 2^j$ gesetzt; i ist die Anzahl der seit der letzten exakten Kraftberechnung für Entfernungsklasse j ausgeführten Integrationsschritte.

Diese einfache Extrapolation erfüllt jedoch nicht das Verlet-Kriterium, wovon man sich durch Einsetzten der Koeffizienten in (2.7) überzeugt. Diese Tatsache gab Anlaß zur Entwicklung von Algorithmus DC-1a, der einen Kompromiß zwischen den beiden offenbar widersprüchlichen Forderungen, Glattheitskriterium und Verlet-Kriterium, darstellt. Entsprechend wurde unter der Vielzahl möglicher Sätze von Extrapolationskoeffizienten erster Ordnung derjenige ausgewählt, welcher (1) dem Verlet-Kriterium genügt und (2) von den in DC-1b verwendeten Koeffizienten möglichst wenig abweicht. Die Lösung dieses Optimierungsproblems ist in Tabelle 2.1, DC-1a, angegeben. Wie man im entsprechenden Bild der Abb. 2.3 erkennt, weist diese Extrapolation erheblich größere Kraftsprünge auf als DC-1b. Insbesondere ist der Kraftverlauf dem exakten Zeitverlauf (fette Kurve) wie auch dem durch DC-1b gegebenen, genäherten Verlauf entgegengerichtet: Eine in der Zeit zunehmende exakte

Kraft ruft stückweise abnehmende extrapolierte Kräfte hervor und umgekehrt. Der so entstehende sägezahnartige Kraftverlauf trägt wesentlich zur Erhöhung des oben beschriebenen ‘algorithmischen Rauschens’ bei. Es wird deshalb von Interesse sein, ob dieser Nachteil erhöhter Rauschkräfte durch das Verlet-Kriterium kompensiert werden kann.

Die von Skeel angegebene dritte Klasse von Extrapolationskoeffizienten, DC-1c, (vgl. Tab. 2.1) erfüllt ebenfalls das Verlet-Kriterium, besitzt aber im Gegensatz zu DC-1a eine zusätzliche Eigenschaft, die als *Kontinuitätskriterium* bezeichnet werden soll. Diesem im Rahmen der vorliegenden Arbeit formulierten Kriterium liegt die Beobachtung zugrunde, daß die Zuordnung eines gegebenen Atompaars zu einer Entfernungsklasse aufgrund der im Prinzip frei wählbaren Entfernungsklassenradien R_j einer gewissen Willkür unterliegt. Darüberhinaus muß diese Zuordnung in der Regel während einer MD-Simulation von Zeit zu Zeit geändert werden, etwa dann, wenn größere intramolekulare Umlagerungen eine starke Verschiebung interatomarer Abstände nach sich ziehen. Es ist deshalb vernünftig zu verlangen, daß extrapolierte Kräfte, wenigstens annähernd, unabhängig von der speziellen Entfernungsklasse des betrachteten Atompaars sein sollten.

Für zwei benachbarte Entfernungsklassen j und $j+1$ nimmt diese Kontinuitätsbedingung die Form

$$\sum_{r=0}^{E} a_{i,r}^{(j)} \mathbf{F}_{(k-r)n_j}^{(j)} = \sum_{r=0}^{E} a_{i,r}^{(j+1)} \mathbf{F}_{(k-r)n_{j+1}}^{(j+1)} \tag{2.10}$$

an, wobei (2.5) zum Zeitpunkt t_{kn_j+i} verwendet wurde. Aufgrund des approximativen Charakters der Kraftextrapolation ist (2.10) nicht exakt erfüllbar. In einer nullten Näherung sollen die in (2.10) eingehenden Kräfte über den betrachteten Zeitraum als unverändert angenommen werden und man erhält:

$$\sum_{r=0}^{E} a_{i,r}^{(j)} = \sum_{r=0}^{E} a_{i,r}^{(j+1)} \ . \tag{2.11}$$

Da die linke Summe in (2.11) im Trivialfall $j = 0$ gleich Eins ist, ergibt die rekursive Anwendung von (2.11) das Kontinuitätskriterium:

$$\sum_{r=0}^{E} a_{i,r}^{(j)} = 1 \quad \text{für alle } i, j \ . \tag{2.12}$$

Die in Tab. 2.1 angegebenen Koeffizienten von DC-1c erfüllen diese Bedingung. Der in Abb. 2.3 (DC-1c) gezeigte resultierende Kraftverlauf weist jedoch zu DC-i ähnliche, wenn auch niedrigere, Kraftspitzen auf. Es kann deshalb im allgemeinen nicht ausgeschlossen werden, daß diese Verletzung des Glattheitskriteriums auch bei DC-1c zu Resonanzinstabilitäten führt.

Aufbauend auf diesen Beobachtungen wurde mit der Entwicklung von DC-1d das Ziel verfolgt, alle drei erwähnten Kriterien zu erfüllen. Wie sich herausstellt, können tatsächlich sowohl das Verlet-Kriterium als auch das Kontinuitätskriterium exakt erfüllt und gleichzeitig ein relativ glatter Kraftverlauf erreicht werden. Abbildung 2.3 (DC-1d) zeigt diesen Verlauf. Er ist, wie man erkennt, ähnlich dem sägezahnartigen Verlauf von DC-1a.

Zur Ableitung des Algorithmus bemerken wir, daß alle Extrapolationsverfahren, deren Koeffizienten *linear* innerhalb der n_j aufeinanderfolgenden Integrationsschritte variieren,

$$\left.\begin{array}{rcl} a_{i,0}^{(j)} & = & c_1 + ic_2 \\ a_{i,1}^{(j)} & = & 1 - (c_1 + ic_2)\,, \end{array}\right\} \qquad (2.13)$$

(a) dem Kontinuitätskriterium genügen und (b) Kraftspitzen wie in DC-i oder DC-1c vermeiden, da die Beträge der Koeffizienten nach oben beschränkt sind. Tatsächlich können die zunächst unbestimmten Koeffizienten c_1 und c_2 so gewählt werden, daß auch das Verlet-Kriterium exakt erfüllt ist. Die resultierenden Extrapolationskoeffizienten sind in Tabelle 2.1 (DC-1d) aufgeführt.

Die mit den bisher beschriebenen MD-Methoden erzielbaren Effizienzsteigerungen gegenüber der expliziten Berechnung langreichweitiger Kräfte lassen sich durch Betrachtung der pro Integrationsschritt im Mittel zu berechnenden Paarwechselwirkungen abschätzen. Für ein System mit N Atomen ist die Zahl N_{explizit} der pro Integrationsschritt explizit nach (2.4) zu berechnenden Paarwechselwirkungen gleich $N(N-1)/2$. Bei Verwendung einer Abschneidefunktion reduziert sich diese Zahl zunächst auf die Anzahl N_{cut} derjenigen Atompaare, deren Abstand kleiner als der gewählte Abschneideradius r_{cut} ist. Die Bestimmung der Menge dieser Paare erfordert dabei einen nicht vernachlässigbaren zusätzlichen Rechenaufwand. Diese Menge wird in der Regel nicht während jedes Integrationsschritts neu bestimmt, da die damit verbundenen $N(N-1)/2$ Abstandsberechnungen die Effizienz stark einschränken würden. Vielmehr werden die Indizes dieser Paare in einer 'Paarliste' gespeichert, welche in regelmäßigen Intervallen neu erstellt wird, um größeren Umlagerungen im Protein Rechnung zu tragen. Der Einfachheit halber soll angenommen werden, daß die Erstellung dieser Liste, die $N(N-1)/2$ Abstandsberechnungen erfordert, mit einem der expliziten Kraftberechnung vergleichbaren Rechenaufwand verbunden ist. Wird die Liste alle N_{update} Schritte neu bestimmt, so ist die pro Integrationsschritt im Mittel zu berücksichtigende Anzahl von Paaren gleich $N_{\mathrm{CUT}} = N_{cut} + N(N-1)/(2N_{update})$.

Im Falle der Entfernungsklassenalgorithmen ist die Anzahl der pro Integrationsschritt explizit berücksichtigten Atompaare nicht konstant (vgl. (2.5)), da Paarwechselwirkungen der Entfernungsklasse j lediglich jeden 2^j-ten Integrationsschritt exakt berechnet werden. Entsprechend läßt sich die mittlere Zahl zu berechnender Paarkräfte mit $N_{\mathrm{DC}} = \sum_{j=0}^{J-1} N_j/2^j$ angeben, wobei N_j die Anzahl der Atompaare in Entfernungsklasse j und J die Anzahl verwendeter Entfernungsklassen ist. Jeden 2^{J-1}-ten Integrationsschritt werden alle Paarkräfte exakt berechnet; zu diesen Zeitpunkten können also ohne zusätzlichen Rechenaufwand neue Paarlisten erstellt werden (Strategien zur effizienten Bestimmung von Paarlisten sind z.B. in [61, 135] beschrieben).

Proteinsystem	Atomzahl N	$N_{\mathrm{exakt}}/N_{\mathrm{CUT}}$	$N_{\mathrm{exakt}}/N_{\mathrm{DC(a)}}$	$N_{\mathrm{exakt}}/N_{\mathrm{DC(b)}}$
Crambin	327	2,9	3,5	4,6
BPTI	568	3,7	4,1	5,6
BR	2365	11,1	7,7	12,4
RC	12637	41,0	19,2	37,0
BR+Umgebung	23208	52,2	24,1	47,6

Tabelle 2.2: Reduktion zu berechnender Paarwechselwirkungen bei Einsatz effizienter MD-Algorithmen; gezeigt ist das Verhältnis der Zahl Paarwechselwirkungen, welche pro Integrationsschritt bei exakter Kraftberechnung berechnet werden müssen, zur entsprechenden Zahl bei Einsatz von CUT bzw. DC; im Fall von DC sind Werte für zwei verschiedene Entfernungsklasseneinteilungen angegeben (siehe Text).

Tabelle 2.2 zeigt die für eine Reihe von Proteinsystemen verschiedener Größe nach dem beschriebenen Verfahren bestimmten Leistungssteigerungen, $N_{\mathrm{explizit}}/N_{\mathrm{CUT}}$ und $N_{\mathrm{explizit}}/N_{\mathrm{DC}}$. Dabei legten wir für N_{CUT} einen Abschneideradius $r_{cut} = 10\,\text{Å}$ zugrunde; N_{DC} wurde für zwei verschiedene Entfernungsklasseneinteilungen mit je acht Entfernungsklassen berechnet, die durch (a) $\{R_i\} = \{0; 5; 8; 14; 23; 35; 50; 68\}$ bzw. (b) $\{R_i\} = \{0; 4; 7; 11; 16; 22; 29; 37\}$ bestimmt sind (Angaben in Å). Einteilung (a) wurde in [77, 111] verwendet. Einteilung (b) führt zu einer weiteren Effizienzsteigerung gegenüber (a), gleichzeitig treten aber größere Rauschkräfte auf. Die in Kapitel 4 getesteten Entfernungsklassenalgorithmen benützen Einteilung (b). Die Wahl von acht Entfernungsklassen erlaubt alle $2^{8-1} = 128$ Integrationsschritte eine Neueinteilung der benötigten Listen. Um den Vergleich mit CUT zu ermöglichen, wurde in Tabelle 2.2 entsprechend $N_{update} = 128$ angenommen. Die zur Bestimmung der Zahlen N_{cut} bzw. N_j verwendeten atomaren Strukturen der Proteine Crambin, Bovine Pancreatic Trypsin Inhibitor (BPTI) und des photosynthetischen Reaktionszentrums von *Rhodopseudomonas viridis* (RC) wurden der *Brookhaven Protein*

Data Bank entnommen; im Falle von Bakteriorhodopsin (BR) wurde das von R. Henderson angegebene Strukturmodell verwendet [4]; das atomare Modell für ein BR-Membran-Wasser-System (BR+Umgebung) wurde von H. Heller zur Verfügung gestellt.

Wie man erkennt, erreichen Entfernungsklassenalgorithmen bei Anwendung auf große Systeme etwa die Effizienz von Abschneide-Methoden; bei kleinen Proteinen sind sie sogar überlegen. Tatsächlich wird für beide Algorithmen im allgemeinen die real erzielbare Leistungssteigerung bei MD-Simulationen geringer ausfallen, da bei der angegebenen Abschätzung (1) der mit der Auswahl der zu berechnenden Paarkräfte verbundene Verwaltungsaufwand vernachlässigt wurde, da (2) der Aufwand zur Berechnung aller kurzreichweitigen Wechselwirkungen nicht berücksichtigt wurde und da (3) neben der eigentlichen Kraftberechnung zusätzlich die numerische Integration in die benötigte Rechenzeit eingeht. So konnte etwa durch den Einsatz von Entfernungsklassenalgorithmen auf einem Parallelrechner die zur MD-Simulation von RC benötigte Rechenzeit gegenüber expliziter Kraftberechnung um den Faktor acht reduziert werden [77].

2.3.2 Schnelle Multipolverfahren

Wie beschrieben wurde, nützen Entfernungsklassenalgorithmen Zeitkorrelationen interatomarer Kräfte und reduzieren so die *Anzahl* benötigter exakter Kraftberechnungen. Im Gegensatz dazu reduzieren Multipolmethoden den Aufwand zur Berechnung der Summe (2.4) mit Hilfe einer geeigneten Multipolentwicklung der langreichweitigen Kraftanteile. Hierzu werden benachbarte Ladungen zu Gruppen zusammengefaßt; die Wechselwirkung zwischen einer Ladung und einer entfernten Ladungsgruppe kann dann in eine Multipolreihe entwickelt werden, die bei hinreichender Genauigkeit abgebrochen wird. Die rekursive Anwendung dieses Verfahrens erfordert eine Hierarchie von Gruppen und verringert den Rechenaufwand von $O(N^2)$ auf $O(N \log N)$. Im einfachsten Fall wird hierzu ein würfelförmiges Simulationsvolumen in acht Würfel halber Kantenlänge unterteilt, ebenso jeder dieser acht Würfel und so fort [68]. Die Unterteilung wird abgebrochen, wenn auf jeden Würfel der untersten Hierarchiestufe nur noch wenige Atome entfallen.

Die ersten derartigen Algorithmen wurden zur Berechnung der Gravitationswechselwirkung in Sternhaufen oder Galaxienclustern eingesetzt, wobei die Multipolentwicklung beim Monopolterm [136] oder nach einer festzulegenden Anzahl von Termen abgebrochen wurde [67]. Einen wesentlichen Fortschritt stellt die sogenannte ‘schnelle Multipolmethode’ (FMM, <u>F</u>ast <u>M</u>ultipole <u>M</u>ethod) dar [68, 137]: Eine geschickte rekursive Berechnung der Multipolmomente drückt hier den Rechenaufwand auf $O(N)$ herab, wodurch die Berechnung der Coulomb-Kräfte für 10 000 Teilchen um etwa den Faktor zehn gegenüber der exakten Rechnung beschleunigt wird [69, 137].

Eine weitere Leistungssteigerung für MD-Simulation von Proteinen läßt sich erzielen, wenn statt einer vorgegebenen, statischen Gruppenhierarchie eine Unterteilung vorgenommen wird, die sich der spezifischen Struktur des zu simulierenden Systems dynamisch anpaßt. Mit diesem Ziel wird in [138] ein neuronaler Algorithmus eingesetzt, der während einer Dynamiksimulation adaptiv eine bezüglich der jeweiligen Proteinstruktur optimierte Gruppeneinteilung ermittelt, die es gestattet, die Multipolreihe bereits nach dem Dipolterm abzubrechen (siehe auch [139]).

Die Komplementarität der beiden Methoden — Entfernungsklassen und Multipolentwicklung — ermöglicht die Kombination beider Verfahren. Für Systeme mit 10 000 Atomen ist dann eine Reduzierung der Rechenzeit gegenüber der exakten Kraftberechnung um etwa zwei Größenordnungen zu erwarten.

„Das beste Modell für eine Katze ist eine Katze. Möglichst dieselbe Katze.“

Norbert Wiener

Kapitel 3

Ein vereinfachtes Proteinmodell

In der Einleitung wurde betont, daß die Charakterisierung der Proteindynamik auf langen Zeitskalen Voraussetzung zur Entwicklung effektiver Modelle zu deren Beschreibung ist. Entsprechende Untersuchungen im Bereich niedriger Frequenzen erfordern MD-Simulationen von einigen hundert Nanosekunden. Da die derzeit verfügbare Rechenleistung jedoch lediglich die Simulation realistischer Proteinmodelle über wenige Nanosekunden erlaubt, soll hier ein *vereinfachtes* Proteinmodell entwickelt und untersucht werden. Dieses sollte einerseits Proteinen im Hinblick auf seine strukturellen und dynamischen Eigenschaften so weit wie möglich gleichen; andererseits sollte das Modell eine möglichst geringe Anzahl von Freiheitsgraden aufweisen, also einfach strukturiert sein, um ausgedehnte Simulationen zu ermöglichen. In diesem Sinne sollte das Modell also ein 'Minimal-Modell' darstellen. Ein solch vereinfachtes Modell kann natürlich nicht *alle* Eigenschaften von Proteinen beinhalten; ebensowenig wird es geeignet sein, die Dynamik eines *bestimmten* Proteins zu beschreiben. Vielmehr sollten lediglich solche Strukturelemente in das Modell übernommen werden, welche zur Beschreibung der Niederfrequenzdynamik von Proteinen unabdingbar erscheinen. Von diesen Vorgaben ausgehend haben wir ein vereinfachtes Proteinmodell entwickelt, dessen Aufbau und Eigenschaften nachfolgend näher beschrieben werden sollen.

3.1 Aufbau des Modells

Der Entwurf des Proteinmodells gliederte sich in zwei Schritte: Zuerst legten wir dessen 'Primärstruktur', bestehend aus 100 Bausteinen, fest und definierten die

Wechselwirkung der Bausteine in Form eines Kraftfelds. Anschließend wurde ein ‘Faltungsvorgang’ des Modells simuliert und so eine auf der Nanosekundenzeitskala stabile Tertiärstruktur erzeugt.

Um die oben beschriebenen widersprüchlichen Zielvorgaben möglichst gut erfüllen zu können, wurde von einer detaillierten internen Struktur der Bausteine abgesehen. Stattdessen modellierten wir, wie Abbildung 3.1 zeigt, die Polypeptidkette durch 100 in einer Reihe angeordnete Van der Waals-Kugeln. Jeweils benachbarte Bausteine sind durch Wechselwirkungen verbunden, welche chemische Einfachbindungen repräsentieren. Entsprechend enthält das verwendete Kraftfeld Bindungsstreck-, Bindungswinkel-, Van der Waals- und Coulomb-Wechselwirkungen,

$$E = E_{\text{bond}} + E_{\text{angle}} + E_{\text{vdW}} + E_{\text{Coulomb}} \,. \tag{3.1}$$

Die Form der einzelnen Energiebeiträge sowie die benötigten Kraftparameter wurden dem CHARMm-Kraftfeld [76] entnommen:

$$\begin{aligned}
E_{\text{bond}} &= \sum_{\text{Bindungen } i} k_b (r_i - r_0)^2 \,, && (3.2)\\
E_{\text{angle}} &= \sum_{\substack{\text{Bindungs-}\\ \text{winkel } j}} k_\Theta (\Theta_j - \Theta_0)^2 \,, && (3.3)\\
E_{\text{vdW}} &= \sum_{\langle \alpha,\beta \rangle} \left(\frac{A}{r_{\alpha\beta}^{12}} - \frac{B}{r_{\alpha\beta}^{6}} \right) . && (3.4)
\end{aligned}$$

Die Coulomb-Wechselwirkung ist in (2.4) definiert; auf Dihedralkräfte wurde verzichtet, so daß die Einfachbindungen der Kette frei drehbar sind. Die Teilchenmassen und Kraftparameter entsprechen denen von CH_2-Gruppen ($m_\alpha = 14\,\text{u}$) und zugehörigen C—C-Einfachbindungen. Als Kraftparameter wurden verwendet: $k_b = 225\,\text{kcal}\,\text{mol}^{-1}\,\text{Å}^{-2}$, $r_0 = 1,52\,\text{Å}$, $k_\Theta = 45\,\text{kcal}\,\text{mol}^{-1}\,\text{rad}^{-2}$, $\Theta_0 = 110^{\text{o}}$, $A = 6,8 \cdot 10^6\,\text{kJ}\,\text{mol}^{-1}\,\text{Å}^{12}$ und $B = 1705\,\text{kJ}\,\text{mol}^{-1}\,\text{Å}^{6}$.

Das so definierte Modell beschreibt freilich noch kein Protein, sondern vielmehr ein 100-Alkan (‘Hektan’). Aus diesem Grund müssen zusätzliche Eigenschaften, die geeignet sind, ein der Dynamik von Proteinen ähnliches Verhalten zu induzieren, hinzugefügt werden. Zur Identifikation derartiger Eigenschaften bemerken wir, daß ein charakteristisches Kennzeichen von Proteinen ihre Fähigkeit ist, in nativer Umgebung eine eindeutige Tertiärstruktur einzunehmen[1] und zu erhalten [140]. Diese Eigenschaft bildet die Voraussetzung für ihre spezifische biochemische Funktion. Sowohl spezifische Wechselwirkungen von Aminosäureseitengruppen, etwa Wasserstoffbrücken oder Disulfidbindungen, wie auch weniger spezifische langreichweitige

[1] genauer, eine Menge wechselseitig ähnlicher Konformationssubzustände, vgl. Kap. 5.

Wechselwirkungen wie elektrostatische, hydrophile und hydrophobe Wechselwirkungen, bestimmen in ihrer Kombination diese Tertiärstruktur. Letztere Wechselwirkungen, insbesondere hydrophobe Kräfte, tragen wesentlich zur Stabilität gefalteter Proteine bei [141, 142]. Aus diesem Grund wurden derartige langreichweitige Wechselwirkungen in unser Modell einbezogen.

Primärstruktur:

Zur Definition einer heterogenen 'Primärstruktur' versahen wir die 100 Van der Waals-Bausteinen des Modells mit unterschiedlichen Partialladungen. Abbildung 3.1 zeigt die gewählte Ladungsverteilung entlang der gestreckten Kette. Wie man erkennt, unterteilt diese Ladungsverteilung das Proteinmodell in fünf Abschnitte, wovon drei positiv und die verbleibenden zwei negativ geladen sind. Der Ausschnitt in Abb. 3.1 zeigt Details der Modellstruktur.

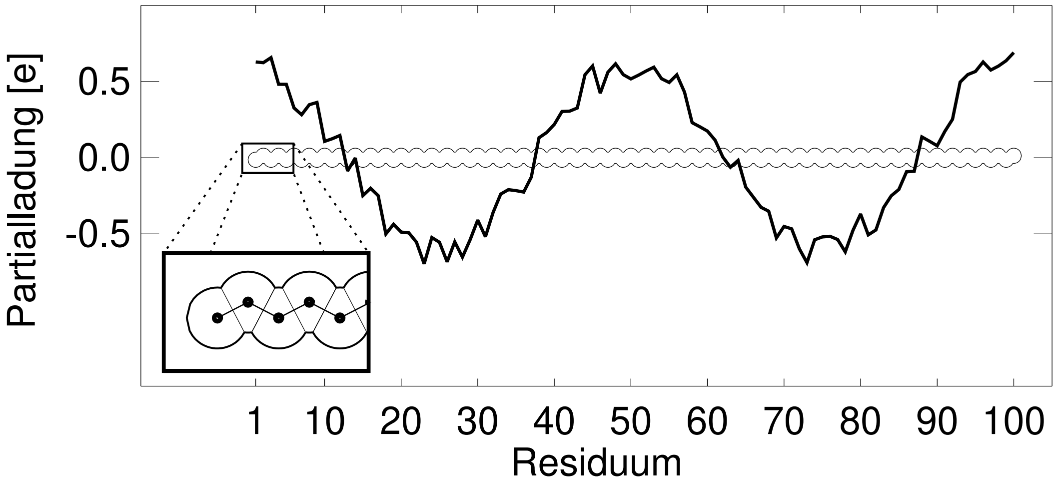

Abbildung 3.1: Das im Text vorgestellte Proteinmodell, bestehend aus 100 CH_2-ähnlichen 'Bausteinen', in gestreckter, entfalteter Konfiguration; die Werte der Partialladungen der einzelnen Bausteine entlang der Kette sind durch die dicke, wellige Kurve dargestellt; der Ausschnitt links unten zeigt Details der Struktur.

Der mittlere Ladungsbetrag wurde mit $0,25\,\mathrm{e}$ so gewählt, daß die resultierende Coulomb-Energie den entropischen Beitrag zur freien Energie des Systems bei Zimmertemperatur geringfügig übertrifft und auf diese Weise gefaltete Zustände der Struktur stabilisiert. Variation der Durchschnittsladung ergab, daß erst Werte im Bereich oberhalb von etwa $0,25\,\mathrm{e}$ stabile Strukturen erzwangen, ein Wert, der — zunächst unerwartet — dem realer Proteine entspricht.

Bezüglich der so erzielten stabilisierenden Funktion imitiert die gewählte Coulomb-

Wechselwirkung hydrophobe Wechselwirkungen zwischen Protein und Lösungsmittel, die denen des in [99] entwickelten 'bead'-Modells ähneln: Die Anziehung entgegengesetzt geladener 'Bausteine' entspricht hydrophoben Kräften, während die Abstoßung von 'Bausteinen' gleicher Ladung die Neigung zur Solvatisierung hydrophiler Gruppen modelliert. Nach dieser Interpretation beinhaltet das Proteinmodell eine effektive Beschreibung der elektrostatischen Eigenschaften des umgebenden Lösungsmittels.

Faltungsprozeß:

Mit Hilfe einer MD-Simulation 'falteten' wir in einem zweiten Schritt das Proteinmodell in eine stabile Tertiärstruktur. Zur Integration der Bewegungsgleichungen wurde der in Kapitel 2 skizzierte Verlet-Algorithmus mit einer Integrationsschrittweite von einer Femtosekunde verwendet. Entfernungsklassenalgorithmen zur effizienten Berechnung der Coulomb-Wechselwirkung wurden aufgrund der geringen Zahl von 100 Teilchen zunächst nicht benötigt.

Ausgangspunkt des simulierten Faltungsprozesses war die in Abb. 3.1 gezeigte gestreckte Konfiguration des Modells, welche auch in Abbildung 3.2 (a) als Bandmodell dargestellt ist. Simuliert wurde dessen freie Bewegung unter dem Einfluß der Bindungs-, Van der Waals- und Coulomb-Kräfte. Die Bilderfolge (b-f) in Abb. 3.2 zeigt Momentaufnahmen der Modellstruktur im Verlauf der ersten Phase der Faltung. Wie zu erkennen ist, nimmt das Modell schnell eine kompakte Struktur an, deren Packungsdichte bereits nach etwa 10 Pikosekunden derjenigen der Endkonfiguration (Abbildung 3.2 (g)) gleicht. Zu diesem Zeitpunkt ist die Tertiärstruktur des Modells jedoch noch nicht stabil und unterliegt häufigen Konformationsänderungen.

Innerhalb der ersten Femtosekunden des Faltungsprozesses schnellte die Temperatur des Modells von 300 K auf über 10 000 K. Dieser große Temperatursprung lag in einer starken Verspannung und der damit verbundenen hohen Konformationsenergie der Ausgangsstruktur begründet; entsprechend wurde ein Teil dieser Energie rasch in kinetische Energie umgewandelt. Die Hochtemperaturdynamik wurde zwei Nanosekunden weitergeführt, um einen möglichst großen Teil des Konfigurationsraums abzusuchen und so das Auffinden einer stabilen Tertiärstruktur zu erleichtern. Schließlich kühlten wir das Modell durch Einführung einer geringen, geschwindigkeitsunabhängigen Reibung für die Atombewegung langsam auf 300 K ab. Dabei nahm es eine wohldefinierte Tertiärstruktur an, die sich auch während der darauffolgenden Equilibrierungsphase von einer Nanosekunde Dauer nicht veränderte. Um das System in die Nähe eines thermischen Gleichgewichts zu bringen, wurde

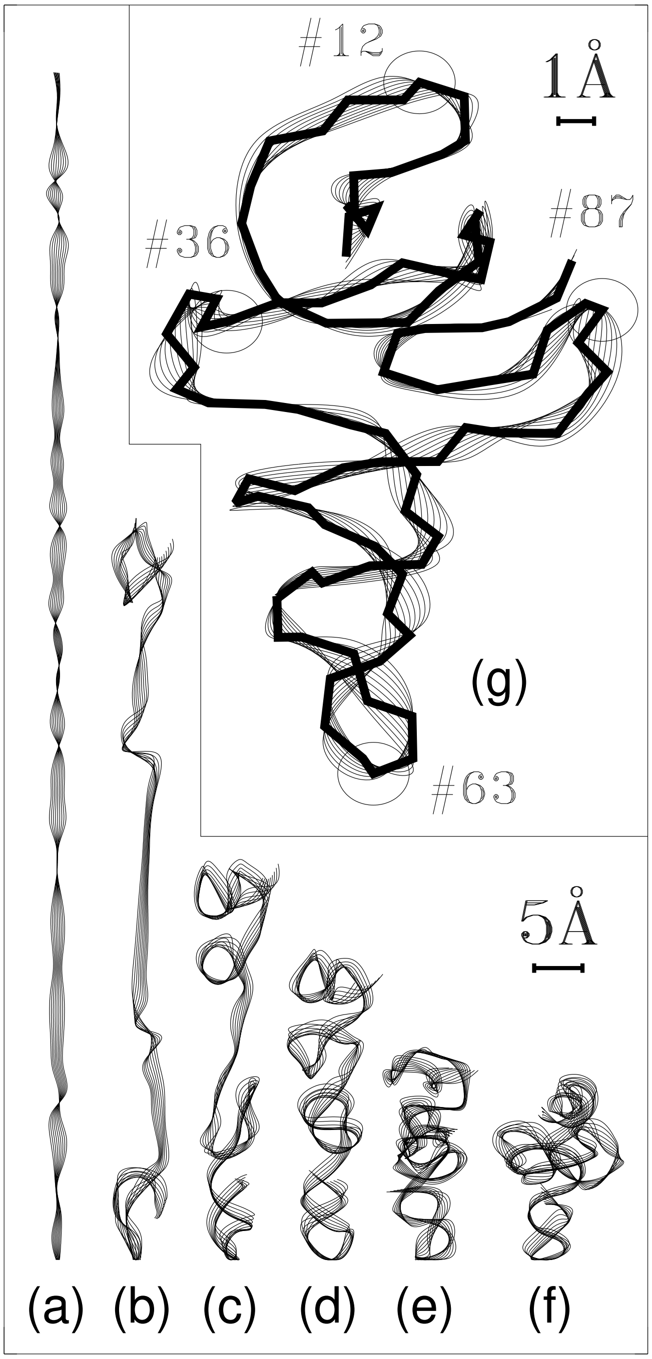

Abbildung 3.2: Momentaufnahmen im Verlauf der im Text beschriebenen Faltungssimulation des Proteinmodells; dessen Struktur ist in Form von Banddarstellungen gezeigt (a–f): vollständig entfaltete Ausgangsstruktur (a), aufeinanderfolgende Konfigurationen nach 3 ps (b), 6 ps (c), 7 ps (d), 8 ps (e), und 100 ps (f); (g): gefaltete Tertiärstruktur des Modells nach 5 ns, welche die Grundlage für die in den beiden folgenden Kapiteln beschriebenen Untersuchungen bildete; die fett eingezeichneten Linien stellen chemische Bindungen dar; vier mit Nummern versehene Kreise bezeichnen Bausteine, auf die im Text Bezug genommen wird.

die ungehinderte Bewegung des Systems für weitere zwei Nanosekunden simuliert. Dabei trat keine Temperaturänderung auf, was darauf hindeutet, daß dieses Ziel auch tatsächlich erreicht wurde. Die in Abb. 3.2 (g) gezeigte Struktur ist das Ergebnis der simulierten Faltung und wird für die in den beiden folgenden Kapiteln geschilderten Untersuchungen verwendet. Zuvor jedoch muß sichergestellt werden, daß das auf diese Weise konstruierte Proteinmodell auch wirklich zur Untersuchung dynamischer Eigenschaften von Proteinen geeignet ist.

3.2 Relevanz des Proteinmodells

Die einschneidenden Vereinfachungen, die im Rahmen der im vorigen Abschnitt beschriebenen Modellbildung vorgenommen wurden, erzwingen eine genaue Analyse im Hinblick auf den Realitätsbezug des Proteinmodells. Insbesondere soll dargelegt werden, worauf bei der Interpretation der Modellstudie geachtet werden muß.

Die Vernachlässigung jeglicher interner Struktur der 100 Bausteine, aus denen das Proteinmodell besteht, stellt wohl die einschneidenste Vereinfachung dar. Da die einzelnen Bausteine durch Van der Waals-Kugeln repräsentiert werden, ist die überwiegende Anzahl kurzreichweitiger Wechselwirkungen, die in Proteinen zwischen den Peptidgruppen des Rückgrats und zwischen den verschiedenen Seitengruppen auftreten, im Modell nicht vorhanden. Dies betrifft insbesondere Wasserstoffbrückenbindungen, welche wesentlich zur Stabilität von Sekundärstrukturelementen beitragen. Da derartige Strukturelemente relativ starr sind, ist damit zu rechnen, daß das Modell eine im Vergleich zu echten Proteinen erhöhte Flexibilität aufweist. Dieser Effekt wird durch das Fehlen strukturierter Seitengruppen und der entsprechenden sterischen und elektrostatischen Wechselwirkungen noch verstärkt. Die Vernachlässigung der Torsionsbarrieren der chemischen Bindungen zwischen den als Van der Waals-Kugeln modellierten Peptidgruppen sollte ebenfalls zur erhöhten Flexibilität des Modells beitragen.

Des weiteren wurde die Masse der Bausteine um etwa eine Größenordnung kleiner als die von Aminosäuren gewählt. Dieser Unterschied hat freilich keinen Einfluß auf die *Qualität* der Dynamik: Vielmehr bewirkt, entsprechend den Newtonschen Gesetzen, die Skalierung der Massen lediglich eine Verschiebung von Zeitskalen um einen Faktor von etwa $2,8$. Insbesondere findet damit die Konformationsdynamik des Modells auf einer im Vergleich zu echten Proteinen entsprechend schnelleren Zeitskala statt.

Schließlich beinhaltet das im vorigen Abschnitt angesprochene effektive Lösungsmittelmodell weder Rausch- noch Reibungskräfte, welche von Lösungsmittelmolekülen ausgehen und an Proteinoberflächen angreifen. Vergleiche von MD-Simulationen realistischer Proteinmodelle *in vacuo* mit Simulationen im Lösungsmittel haben gezeigt, daß die vom Lösungsmittel ausgehenden Kräfte dazu beitragen, trägheitsdominierte Moden der Dynamik zu dämpfen und zu entkoppeln [95]. Demzufolge sollten unsere Simulationen *in vacuo* einen wesentlich langsameren Abfall der Kurzzeit-Autokorrelationsfunktion von Atompositionen der Proteinoberfläche aufweisen als im Fall einer realistischeren Modellierung von Lösungsmitteleffekten zu erwarten wäre.

Zusammengefaßt sollte der kombinierte Effekt aller vorgenommenen Vereinfachungen darin bestehen, Barrieren der freien Energie für Konformationsübergänge zu verkleinern und die Konformationsdynamik zu beschleunigen, d.h., die Raten von Konformationsübergängen zu erhöhen. Zusätzlich ist ein verstärktes Auftreten von Gedächtniseffekten auf kurzen Zeitskalen zu erwarten.

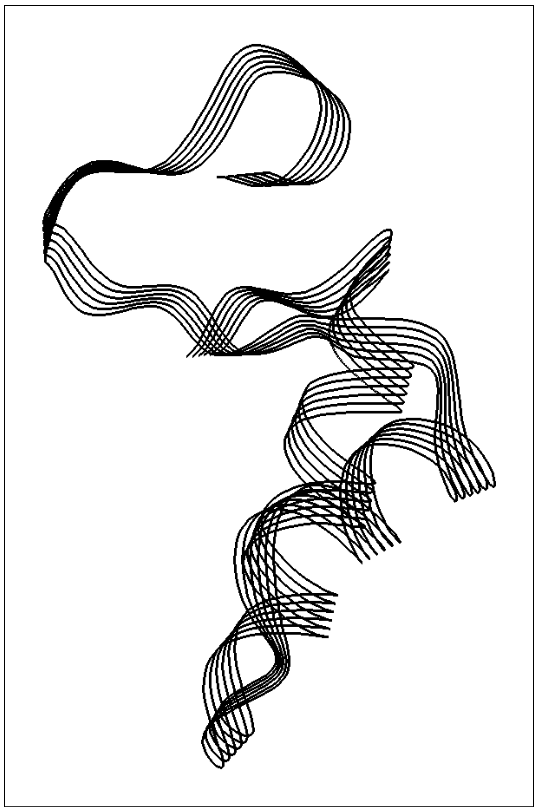

Abbildung 3.3: Die Banddarstellung der Tertiärstruktur von Crambin, eines globulären Proteins, weist Ähnlichkeiten mit dem im Text diskutierten Proteinmodell auf (vgl. Abb. 3.2). In beiden Strukturen sind Sekundärstrukturelemente wie etwa α-Helizes (unten im Bild) und Schleifen ohne reguläre Struktur, sogenannte 'loops' (oben) erkennbar.

Die in Abbildung 3.2 (g) dargestellte gefaltete Tertiärstruktur des Proteinmodells scheint trotz des Fehlens seitengruppenspezifischer Wechselwirkungen Strukturelemente aufzuweisen, die mit Sekundärstrukturelementen von Proteinen vergleichbar sind. Beispiele hierfür sind die 'Helix' an der Unterseite des Modells und die links, rechts und oben im Bild erkennbaren Schleifen ('loops'). Die räumliche Anordnung dieser Elemente gleicht der Tertiärstruktur kleiner globulärer Proteine, wie etwa der in Abbildung 3.3 ebenfalls in Form eines Bandmodells dargestellten Struktur von Crambin, die der *Brookhaven Protein Data Bank* entnommen wurde. Es sei betont, daß diese überraschende strukturelle Ähnlichkeit nicht *explizit* konstruiert wurde, sondern vielmehr ein Ergebnis der speziellen Wahl intramolekularer Wechselwirkungen ist. Entsprechend kann erwartet werden, daß das Modell auch realistische *dynamische* Eigenschaften aufweist.

Um diese Erwartung zu überprüfen, soll eine Reihe verschiedener Eigenschaften, deren Kombination für Proteindynamik charakteristisch ist, berechnet und mit MD-Simulationen realistischer Proteinmodelle oder experimentellen Befunden verglichen werden. Wir werden dabei bei kurzen Zeitskalen (Femtosekunden) beginnen und sukzessive bis in den Nanosekundenbereich fortschreiten.

Wie im vorangegangenen Abschnitt beschrieben, definierten wir die kurzreichweitigen Wechselwirkungen des Proteinmodells, welche wesentlich dessen dynamisches Verhalten bei hohen Frequenzen bestimmen, in enger Anlehnung an das CHARMm-Kraftfeld für Kohlenwasserstoffe. Es ist demnach zu erwarten, daß das Modell auf der Zeitskala unterhalb einiger hundert Femtosekunden vergleichsweise realistische Eigenschaften aufweisen sollte.

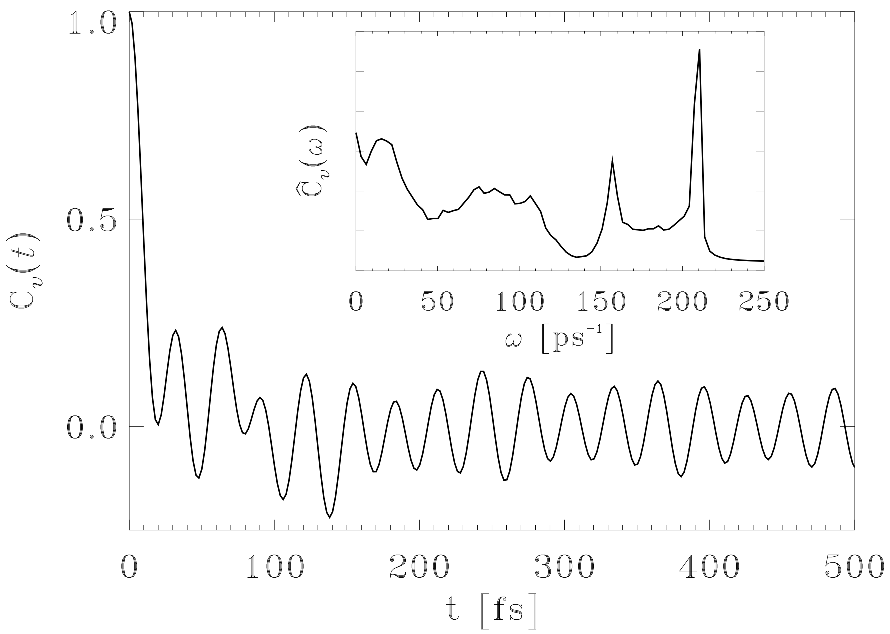

Abbildung 3.4: Normierte Autokorrelationsfunktion $C_v(t) = \langle \mathbf{v}(0) \cdot \mathbf{v}(t) \rangle / \langle \mathbf{v}(0)^2 \rangle$ der Atombewegung, berechnet durch Mittelwertbildung aus einer 1 ns-Trajektorie, wobei der Geschwindigkeitsvektors $\mathbf{v}$ eines willkürlich ausgewählten Teilchens des Proteinmodells verwendet wurde. Rechts oben ist das entsprechende Spektrum der Autokorrelationsfunktion eingeblendet. Dieses Spektrum ähnelt denen realistischer Proteinmodelle: Die zwei im Bild erkennbaren scharfen Spitzen rühren von schnellen Bindungslängenschwingungen her, während die breiten Banden im niedrigeren Frequenzbereich überwiegend vom stochastischen Charakter interatomarer Van der Waals-Kontakte verursacht werden.

Als Beispiel zeigt Abbildung 3.4 eine Geschwindigkeits-Autokorrelationsfunktion $C_v(t) = \langle \mathbf{v}(0) \cdot \mathbf{v}(t) \rangle / \langle \mathbf{v}(0)^2 \rangle$, welche aus einer 1 ns-Trajektorie berechnet wurde, wobei $\mathbf{v}$ der Geschwindigkeitsvektors eines willkürlich ausgewählten Teilchens des Proteinmodells ist. Rechts oben im Bild ist das zugehörige Spektrum eingeblendet. Um einen Vergleich mit den Ergebnissen von MD-Simulationen eines detaillierten

Modells von RNAse-A in [95] zu ermöglichen, ist das Spektrum in gleicher Weise wie dort berechnet.

Beide Spektren weisen scharfe Spitzen im Frequenzbereich zwischen 150 und 250 ps^{-1} auf, welche von schnellen Bindungsstreckschwingungen herrühren. Unterhalb von 150 ps^{-1} sind in beiden Fällen breitere Bänder zu erkennen, die durch Schwingungen der Bindungswinkel und insbesondere durch unregelmäßige interatomare Van der Waals-Kontakte hervorgerufen werden. Die Hochfrequenzdynamik realistischer Proteinmodelle unterscheidet sich von der Dynamik des vereinfachten Modells dadurch, daß erstere aufgrund der Heterogenität der mit den Bindungsstreckschwingungen verbundenen Frequenzen eine größere Zahl derartiger Resonanzspitzen zeigt. Im Gegensatz dazu ist in unserem Modell lediglich ein einziger Bindungstyp vorhanden.

Die in Abb. 3.4 erkennbaren Hochfrequenzmoden stellen die unterste Ebene einer ganzen Reihe dynamischer Vorgänge in Proteinen dar, die auf einer Hierarchie von Zeitskalen ablaufen. Auf jeder dieser Zeitskalen sind spezifische Vorgänge beobachtbar [22,143–145]. Diese beinhalten die erwähnten Bindungsstreckschwingungen ($\approx$ 30 fs), Schwingungen der Winkel, die von jeweils zwei chemischen Bindungen eingeschlossen werden, sowie Torsionsschwingungen der Bindungen (einige 100 fs), kollektive Bewegungen, welche ganze Atomgruppen miteinbeziehen (einige 10 ps), sowie Konformationsübergänge, die oberhalb einiger 100 ps einsetzten und in einem weiten Zeitbereich auftreten.

Eine Analyse dynamischer Details zeigt, daß sich eine derartige Hierarchie tatsächlich in der Dynamik unseres Proteinmodells über mehr als sechs Größenordnungen beobachten läßt. In Abbildung 3.5 soll dies anhand eines Beispiels illustriert werden. Dargestellt ist das zeitliche Verhalten des Abstands zweier willkürlich gewählter Bausteine des Modells bei abnehmender Zeitauflösung. Die Bildsequenz zeigt auf jeweils um einen Faktor zehn zunehmenden Zeitskalen charakteristische Abstandsfluktuationen, die ihren Ursprung in den vier oben genannten dynamischen Prozessen haben.

Die Existenz einer Reihe hierarchisch angeordneter Konformationssubzustände, die durch Konformationsübergänge verbunden sind, ist ein charakteristisches Kennzeichen der Proteindynamik. Derartige Konformationsübergänge erscheinen auch in der Dynamik unseres vereinfachten Proteinmodells. Sie treten durch plötzliche Änderungen der Proteinstruktur in Erscheinung. So werden etwa die abrupten Abstandsänderungen, die in den unteren drei Bildern der Abb. 3.5 auf der Zeitskala einiger hundert Pikosekunden erkennbar sind, von solchen Konformationsübergängen

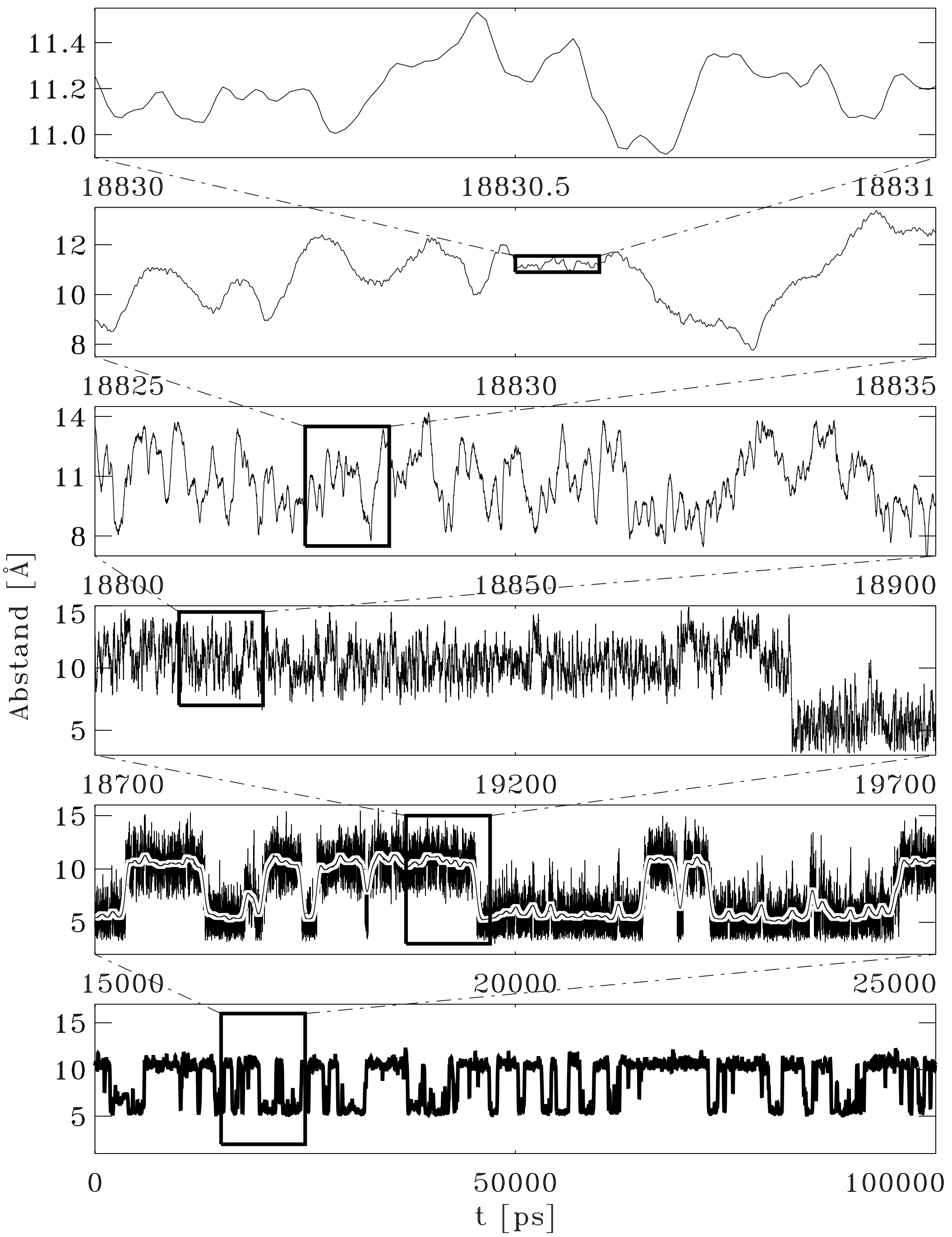

Abbildung 3.5: Fluktuation des Abstands (in Å) zwischen den Bausteinen # 12 und # 36 (vgl. Abb. 3.2 (g)) im Verlauf einer MD-Simulation der Länge 100 Nanosekunden. Die Bildsequenz zeigt die Abstandsfluktuationen auf Zeitskalen, die von Bild zu Bild um den Faktor zehn zunehmen; die fett eingezeichneten Kurven der unteren beiden Bilder repräsentieren geglättete Daten.

hervorgerufen. Tatsächlich gleicht diese Konformationsdynamik der von realen Proteinen, wie am Beispiel des in der Einleitung beschriebenen Ionenkanalproteins deutlich wird: Wie ein Vergleich der Abb. 3.5 mit Abb. 1.3 zeigt, treten auch dort ähnlich abrupte Konformationsänderungen auf, welche die dargestellten Sprünge der elektrischen Leitfähigkeit hervorrufen, in diesem Protein allerdings auf der Zeitskala von Millisekunden.

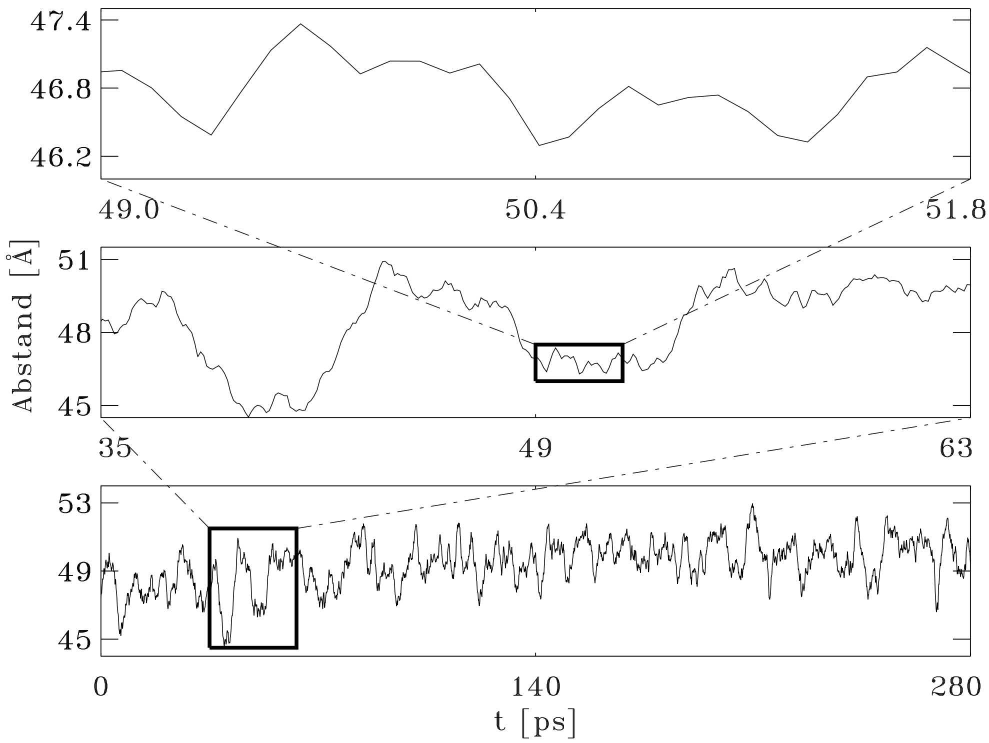

Abbildung 3.6: Fluktuation des Abstands (in Å) zwischen zwei ausgewählten Atomen an entgegengesetzten Seiten eines atomaren Modells von Bakteriorhodopsin (vgl. Abb. 1.1) im Verlauf einer 280 Pikosekunden langen MD-Simulation; die Bildsequenz zeigt die Abstandsfluktuationen auf Zeitskalen, die von Bild zu Bild um den Faktor zehn zunehmen; berücksichtigt man die im Text diskutierte Zeitskalendehnung um den Faktor $2,8$, so entsprechen die gezeigten Ausschnitte ($2,8$ ps, 28 ps und 280 ps) den oberen drei Ausschnitten in Abbildung 3.5 ($1,0$ ps, 10 ps und 100 ps).

Um einen Vergleich mit einem realistischen Proteinmodell zu ermöglichen, zeigt Abbildung 3.6 das Resultat einer MD-Simulation von Bakteriorhodopsin *in vacuo* [112]. Auch hier wurde der Abstand zweier Atome im Protein über der Zeit aufgetragen. Um den Vergleich mit der Simulation am vereinfachten Proteinmodell zu ermöglichen, wurde eine der Abb. 3.5 entsprechende Bildsequenz gewählt, wobei die angesprochene Zeitskalierung um den Faktor $2,8$ zu beachten ist. Die Sequenz bricht allerdings aufgrund der hohen Komplexität des Bakteriorhodopsin-Modells und des damit verbundenen Rechenaufwands bereits im Bereich einiger hundert Pikosekunden ab. Die in der Abbildung erkennbaren Fluktuationen sind durch kollektive Bewegungen der Aminosäuren bestimmt. Konformationsübergänge wurden hier aufgrund des beschränkten Simulationszeitraums nicht beobachtet.

All die geschilderten Beobachtungen untermauern die Annahme, daß das in diesem Kapitel vorgestellte Proteinmodell *trotz der vorgenommenen Vereinfachungen* eine Reihe struktureller und dynamischer Eigenschaften aufweist, die für Proteine charakteristisch sind. Insbesondere sollte sich das Proteinmodell sowohl für die im Kapitel 5 geschilderte Analyse der Konformationsdynamik auf langen Zeitskalen, als auch für die im folgenden vorgestellte problemangepaßte Bewertung effizienter MD-Methoden eignen.

'This came directly from a computer and is not to be doubted or disbelieved.'

Lisa Balbes, 1993

Kapitel 4

Beurteilung effizienter MD-Methoden

Im Abschnitt 2.3 wurde eine Reihe von Methoden eingeführt, welche die Coulomb-Wechselwirkung in Proteinen approximieren und so die Effizienz von MD-Simulationen um etwa eine Größenordnung steigern. Alle diese Methoden zielen darauf ab, den mit der Approximation verbundenen Fehler nicht zu groß werden zu lassen. Die Abschneide-Methode ('CUT') gewichtet des Coulomb-Potential deshalb mit einer entfernungsabhängigen, stetig differenzierbare Abschneidefunktion. Der so erreichte kontinuierliche Potentialverlauf unterdrückt Rauschkräfte, die etwa entstehen würden, falls die Coulomb-Kräfte am Abschneideradius *abrupt* abgeschnitten würden. Der Abschneideradius selbst wird so gewählt, daß die auf jedes Atom wirkenden Coulomb-Kräfte im Mittel um nicht mehr als einen festgelegten Prozentsatz von den exakten Coulomb-Kräften abweichen.

In gleicher Weise stützte sich die geschilderte Entwicklung von Entfernungsklassenalgorithmen ('DC') auf eine Kombination verschiedener Kriterien wie des Verlet-Kriteriums, des Glattheitskriteriums oder des Kontinuitäts-Kriteriums. Auch diese Kriterien dienen dem Zweck, die Abweichung der genäherten Kräfte von den exakten Kräften möglichst gering zu halten und insbesondere die genäherten Trajektorien der Atome den mit exakten Coulomb-Kräften bestimmten Referenztrajektorien anzugleichen.

Sollen mit Hilfe derartige Approximationen dynamische Vorgänge in Proteinen studiert und deren Funktion aufgeklärt werden, stellt sich jedoch die Frage, inwiefern die eingesetzten Kriterien in der Lage sind, Eignung und Qualität von MD-

Algorithmen *für diesen hier interessierenden Anwendungsbereich* zu beurteilen. In der Einleitung wurde bereits festgestellt, daß nicht alle Details atomarer Bewegungen für die Proteinfunktion relevant sind. Leider ist *a priori* nicht entscheidbar, mit welcher Genauigkeit interatomare Kräfte berechnet werden müssen, um die zuverlässige Bestimmung *funktionsrelevanter* Observabler zu gewährleisten. Dies liegt zum einen in der Schwierigkeit begründet, funktionsrelevante Größen zu identifizieren, zum anderen ist die Wirkung algorithmisch bedingter, häufig 'mikroskopischer' Artefakte in der atomaren Bewegung auf relevante Observable, welche häufig mesoskopischen oder makroskopischen Charakter aufweisen, nicht bekannt. Es ist somit nicht möglich, allein aufgrund der oben angesprochenen oder ähnlicher Kriterien die vorgestellten MD-Methoden zu vergleichen. Um diese Lücke zu füllen, wollen wir in diesem Kapitel *problemangepaßte Beurteilungsverfahren* vorschlagen und einsetzen um die Qualität der beschriebenen MD-Algorithmen zu ermitteln und insbesondere die neuen Entfernungsklassenalgorithmen auf eine solide Grundlage zu stellen. Dazu werden wir einige der Prinzipien diskutieren, welche der Methode der MD-Simulation zugrunde liegen. Eine daraus abgeleitete Klassifikation möglicher Fehlerquellen, die bei approximativen MD-Simulationen auftreten können, wird es dann erlauben, den Begriff der 'problemangepaßten Genauigkeit' näher zu präzisieren.

4.1 Fehlerabschätzung bei MD-Simulationen

Abbildung 4.1 zeigt schematisch die grundlegenden Konzepte der Methode der MD-Simulation und deren wechselseitige Beziehung. Wie die beiden grauen Pfeile im Bild verdeutlichen, beinhaltet die Methode der MD-Simulation *zwei* aufeinanderfolgende Schritte, einen physikalischen und einen numerischen. Beiden Schritten liegen bestimme Annahmen und Näherungen zugrunde.

Der erste Schritt besteht in der Beschreibung der realen Proteindynamik, die durch den obersten Kasten dargestellt ist, durch ein *physikalisches Modell.* Diese Modellbildung wurde in Abschnitt 2.1 skizziert und soll hier nicht näher erläutert werden. Jedes derartige Modell besteht aus einem großen System gekoppelter Differentialgleichungen zweiter Ordnung, den Bewegungsgleichungen, sowie vorgegebenen Anfangsbedingungen, etwa Atompositionen und -geschwindigkeiten. Die Spezifikation des Modells definiert als eindeutige Lösung dieses Gleichungssystems detaillierte Atombewegungen, dargestellt durch den Kasten in der Mitte von Abb. 4.1. Diese Lösung ist durch die 'exakte' Trajektorie $\mathbf{r}_1(t), \ldots, \mathbf{r}_N(t)$ gegeben, die alle N Atompositionen $\mathbf{r}_\alpha$ zu jedem Zeitpunkt t festlegt.

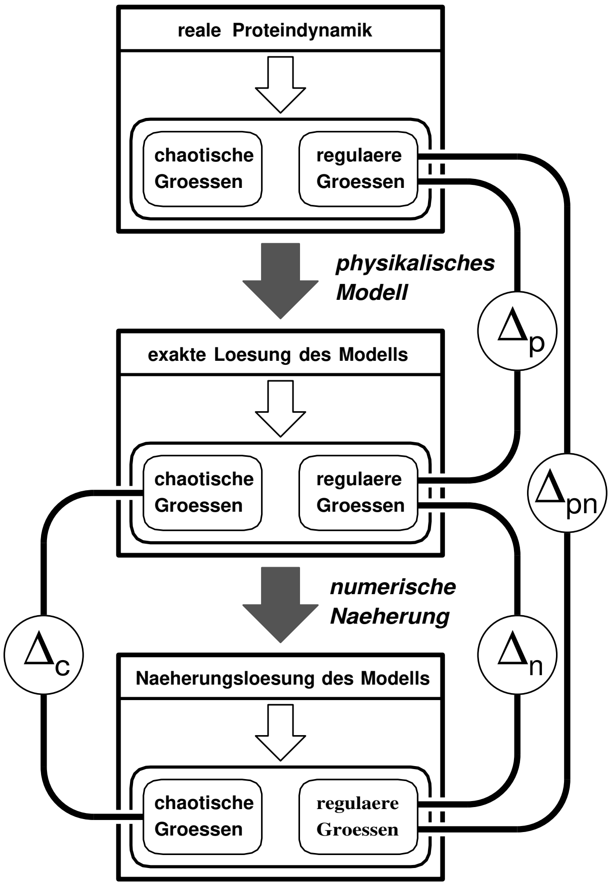

Abbildung 4.1: Grundlegende Konzepte der MD-Simulation; zur Erläuterung siehe Text.

Da die Bewegungsgleichungen nicht analytisch lösbar sind, muß die Trajektorie in einem zweiten Schritt *numerisch genähert*[1] werden; der untere Kasten im Bild repräsentiert diese Näherungslösung. In Abschnitt 2.1 wurde am Beispiel des Verlet-Algorithmus erläutert, daß dies dies die numerische Integration der Newtonschen Gleichungen (2.1) sowie die wiederholte Bestimmung der auf alle Atome im Protein wirkenden Kräfte erfordert. Ein Verfahren, das beides bewerkstelligt — numerische Integration und Kraftberechnung — wird 'MD-Algorithmus' genannt.

Aufgrund der vorgenommenen Näherungen ist jeder der beiden aufgeführten Schritte mit Fehlern verbunden. Die der physikalischen Modellbildung inhärenten Näherungen, etwa die klassische Behandlung der atomaren Bewegung, ein ungenaues Kraftfeld oder das Vernachlässigen von Umgebungseffekten, verursachen Abweichungen der exakten Lösung des Modells von der real stattfindenden Proteindynamik. *Zusätzliche* Abweichungen numerischen Ursprungs werden im zweiten Schritt eingeführt, wobei drei Arten numerischer Fehler unterschieden werden können: Zwei

[1] Die Bezeichnung 'numerische Näherung' wurde gewählt, um diesen Schritt von der *physikalischen* Modellbildung zu unterscheiden. Näherungen, die auf speziellen physikalischen Eigenschaften von Proteinen beruhen, sind dabei selbstverständlich eingeschlossen.

davon, Rundungsfehler, bedingt durch die beschränkte Genauigkeit arithmetischer Operationen, sowie Diskretisierungsfehler, die bei der numerischen Integration entstehen, sind in der Regel klein und werden hier nicht betrachtet. Die weitaus größten numerischen Fehler werden typischerweise durch Approximationen hervorgerufen, die, wie in Abschnitt 2.3 beschrieben, der effizienten Berechnung langreichweitiger Kräfte dienen. Dies betrifft insbesondere die Coulomb-Kräfte, deren Ermittlung wie in Abschnitt 2.2 gezeigt wurde, bei MD-Simulationen rechenzeitbestimmend ist. Wir wollen deshalb die Qualität von MD-Algorithmen beurteilen, welche eine effiziente Berechnung langreichweitiger Kräfte anstreben. Die in Abschnitt 2.3 beschriebenen Verfahren (Abschneide-Methoden, Entfernungsklassenalgorithmen und Multipolmethoden) sind von dieser Art.

Es soll hier nochmals hervorgehoben werden, daß eine Unterscheidung zwischen physikalischer Modellbildung und numerischer Behandlung tatsächlich notwendig ist; nur so können die zugehörigen Fehlertypen getrennt behandelt werden. Würde diese Trennung nicht beachtet, könnten Artefakte, welche durch *numerische* Approximationen hervorgerufen sind, fälschlicherweise einer ungenauen physikalischen Modellbildung zugeschrieben werden, wodurch die Verbesserung molekularer Modelle wie auch die Weiterentwicklung effizienter MD-Methoden behindert oder gar irregeleitet würden. Um beide zur Durchführung von MD-Simulationen erforderlichen Schritte unabhängig voneinander optimieren zu können, müssen daher mögliche Fehlerquellen getrennt werden.

Für die weitere Analyse unterteilen wir alle Größen, welche aus MD-Simulationen abgeleitet werden können, in zwei Gruppen, *chaotische* Größen, in Abb. 4.1 links angeordnet, und *reguläre* Größen (rechte Hälfte). Nahegelegt wird eine derartige Unterscheidung durch die Beobachtung, daß die molekulare Dynamik eines Proteins bei Zimmertemperatur chaotisches Verhalten zeigt: Identische Systeme, die sich zu einem Zeitpunkt an nahezu gleicher Position im Phasenraum befinden, entfernen sich rasch voneinander und sind bereits nach wenigen Pikosekunden weitgehend dekorreliert. Abbildung 4.2 demonstriert dieses Verhalten am Beispiel des im vorigen Kapitel vorgestellten Proteinmodells: Hier wurde die in Abbildung 3.5 gezeigte Zeitentwicklung des Abstands zweier Atome für zwei nahezu identische Systeme verfolgt, die sich lediglich in der Position eines einzigen Atoms unterschieden; dieses Atom wurde im zweiten System um die winzige Strecke von 10^{-6} Å verschoben. Abbildung 4.2 zeigt die schnell zunehmende Abweichung der gewählten Observable in den beiden Systemen. Bereits nach zwei Pikosekunden hat die Abweichung ihren Maximalwert erreicht; die beiden Abstandsfluktuationen sind hier vollständig dekorreliert.

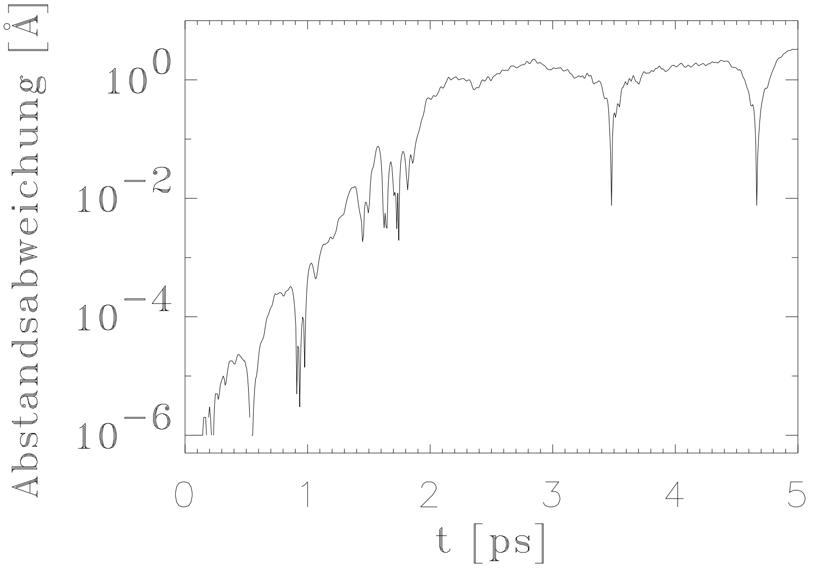

Abbildung 4.2: Die molekulare Dynamik von Proteinen ist chaotisch. Zwei identische Proteinmodelle, die an nahezu identischen Positionen im Phasenraum starten, entfernen sich schnell voneinander. Gezeigt ist die Zeitentwicklung der Abweichung des Abstands zweier ausgewählter Atome des einen Systems vom entsprechenden Atomabstand des zweiten Systems. Nach etwa zwei Pikosekunden molekularer Dynamik sind die beobachteten Abstandsfluktuationen vollständig dekorreliert; das System hat seinen Startpunkt im Phasenraum 'vergessen'.

Alle Größen, die ähnlich empfindlich auf geringste Störungen reagieren, nennen wir im folgenden 'chaotische Größen'. Die wenigen übrigen Größen, die kein solches chaotisches Verhalten zeigen, bezeichnen wir als 'reguläre Größen'. Letztere sind oftmals Mittelwerte, etwa mittlere Atompositionen, Gyrationsradien, Durchschnittstemperatur, Schwingungsspektren oder Differenzen der freien Energie.

Offensichtlich ist dies keine scharfe Unterteilung; die Zuordnung zu einer der beiden Gruppen wird in der Regel davon abhängen, welches molekulare Modell gewählt und über welche Zeitspanne simuliert wird. Dennoch gibt es auf einer gegebenen Zeitskala erfahrungsgemäß immer Größen, die sicher chaotisch oder regulär sind; nur solche sollen hier behandelt werden.

Algorithmische Genauigkeit wird durch den Vergleich ausgewählter 'Testgrößen' ermittelt. Kleine Abweichungen der Testgrößen von den Referenzwerten lassen entsprechend auf hohe Genauigkeit des untersuchten Algorithmus bezüglich der betrachteten Größe schließen. Da eine analytische Abschätzung nur selten möglich und sinnvoll ist, wird man sich in der Regel auf Referenzgrößen stützen, die zwar nicht exakt sind, deren Genauigkeit aber als hoch im Vergleich zu den Testgrößen angenommen werden kann.

In Abb. 4.1 wurde die oben angesprochene Unterteilung in chaotische und reguläre Größen auf allen drei Ebenen — reale Proteindynamik, exakte Lösung und Näherungslösung — vorgenommen. Daraus ergeben sich insgesamt sechs Klassen von Größen, die zur Ermittlung algorithmischer Genauigkeit verglichen werden können. Kriterien zur Beurteilung algorithmischer Genauigkeit können nach diesem Schema klassifiziert werden.

Von den angesprochenen sechs prinzipiell möglichen Vergleichstypen müssen wir diejenigen beiden als nicht realisierbar ausschließen, die chaotische Größen der realen Proteindynamik einbeziehen, da diese Größen grundsätzlich nicht meßbar sind. Die verbleibenden vier Kombinationen, die in Abb. 4.1 durch die fett gezeichneten Verbindungen symbolisiert und mit Δ_c (chaotisch), Δ_p (physikalisch), Δ_n (numerisch), und Δ_pn bezeichnet sind, definieren alle Vergleichstypen, mit denen algorithmische Genauigkeit untersucht werden kann.

Alle vier Beurteilungsverfahren werden häufig eingesetzt. Es sei bereits an dieser Stelle vorweggenommen, daß die Genauigkeit von Algorithmen, welche zur Proteindynamiksimulation eingesetzt werden sollen, unserer Auffassung nach durch Vergleiche des Typs Δ_n beurteilt werden sollte. Bevor wir diese Behauptung begründen, sollen die vier Vergleichstypen anhand einiger Beispiele illustriert werden:

1. *Vergleich chaotischer Größen der Näherungslösung mit chaotischen Größen der exakten Lösung* (Δ_c):

   Typische Vertreter dieses Beurteilungsverfahrens vergleichen Trajektorien oder die Zeitentwicklung ausgewählter Atompositionen, Bindungs- oder Dihedralwinkel. In diesen Fällen wird die Abweichung der Näherungslösung von der exakten — oder als genauer angenommenen — Lösung als Genauigkeitskriterium herangezogen.

   Diese Methode wurde auf kleine Systeme mit wenigen Teilchen, wie etwa das Keplersche Zweikörperproblem, angewandt, für welche die exakte Lösung bekannt ist. Um explizite kanonische Integrationsverfahren zu testen, haben D. Okunbor und R. D. Skeel [122] die Zeitentwicklung des globalen Fehlers $\Delta(t) = |\mathbf{r}_\mathrm{approx}(t) - \mathbf{r}_\mathrm{exakt}(t)|/|\mathbf{r}_\mathrm{exakt}(t)|$ untersucht, wobei $\mathbf{r}_\mathrm{approx}(t)$ und $\mathbf{r}_\mathrm{exakt}(t)$ die zusammengesetzten genäherten bzw. exakten Ortsvektoren der beiden Körper sind. In ähnlicher Weise verwendeten G. D. Quinlan und S. Tremaine Abweichungen der Positionen, um zu untersuchen, wie gut sich verschiedene Integrationsverfahren höherer Ordnung zur Lösung der Bewegungsgleichungen der zehn massereichsten Körper des Sonnensystems eignen [131].

K. D. Gibson und H. A. Scheraga [146] verwendeten Testsimulationen an einem kleinen Peptid (*N*-acetyl-Alanin), um die Abweichung bestimmter Dihedralwinkel, welche die Konformation des Moleküls charakterisieren, bestimmen zu können. In Übereinstimmung mit unseren MD-Simulationen am Proteinmodell stellten sie fest, daß Trajektorien, die länger als einige Pikosekunden sind, aufgrund der Chaotizität solcher Systeme (vgl. Abb. 4.2) nicht sinnvoll verglichen werden können. Abweichungen der detaillierten Atombewegung wurden auch an größeren Systemen untersucht, wie etwa an Lennard-Jones-Flüssigkeiten [107, 147].

In [136] wird der durch eine hierarchische Monopol-Approximation verursachte Fehler diskutiert, wobei ein großer Galaxienhaufen mit Gravitationswechselwirkung als Testsystem verwendet wurde. Für den schnellen Multipolalgorithmus haben Greengard et al. auf analytischem Weg eine obere Schranke für den Fehler in den Coulomb-Kräften abgeleitet [68].

2. *Vergleich regulärer Größen der Näherungslösung mit regulären, experimentell gewonnenen Größen* (Δ_{pn}):

   Dieses Verfahren wurde verwendet, um den Einfluß des Abschneidens langreichweitiger Kräfte (vgl. Abschnitt 2.3) auf die molekulare Dynamik von Flüssigkeiten [148] und auf die Struktur von Proteinen [79] zu ermitteln. In diesen Fällen waren die betrachteten Referenzwerte experimentell bestimmte Größen, wie etwa mittlere Atompositionen und -beweglichkeiten sowie thermodynamische Eigenschaften wie Kompressibilität, spezifische Wärme oder Diffusionskoeffizienten. Wir ordnen derartige Vergleiche der Gruppe Δ_{pn} zu, da wir die Vernachlässigung langreichweitiger Wechselwirkungen wie der Coulomb-Kräfte nicht als Bestandteil des molekularen Modells betrachten (vgl. hierzu auch Abschnitt 1.4), sondern vielmehr als numerische Näherung, die auf eine Effizienzsteigerung abzielt; physikalische Gründe für eine *vollständige* Vernachlässigung langreichweitiger Anteile der Coulomb-Wechselwirkung in Proteinen lassen sich nicht angeben.

3. *Vergleich regulärer Größen der Näherungslösung mit regulären Größen der exakten Lösung* (Δ_{n}):

   Der am häufigsten verwendete Test dieses Typs betrachtet die Gesamtenergie des Testsystems, welche für konservative Vielteilchensysteme konstant ist. Dabei werden hauptsächlich zwei Meßgrößen untersucht [42, 76, 146]: (1) das

Verhältnis der mittleren quadratischen Abweichung der in der numerischen Näherung fluktuierenden Gesamtenergie zur Fluktuation der kinetischen Energie des Systems und (2) der Betrag der Energiedrift, die im Verlauf einer Simulation auftritt und den Energieerhaltungssatz verletzt.

Andere Testmethoden des Typs Δ_{n} vergleichen die Temperatur oder den Druck einer Van der Waals-Flüssigkeit [107] sowie Geschwindigkeits-Autokorrelationsfunktionen atomarer Bewegungen [133]. Um verschiedene molekulare Modelle für Wasser anhand von Testsimulationen zu beurteilen, berechneten M. Prevost et al. atomare Abstandsverteilungen, Kompressibilitäten, spezifische Wärmen und Selbstdiffusionskoeffizienten und verglichen die erhaltenen Werte mit experimentellen Größen wie auch mit Referenzsimulationen [148].

Der Einfluß der Unterdrückung schneller Freiheitsgrade mit Hilfe verallgemeinerter Langevin-Gleichungen wurde von P. Romiszowski und R. Yaris mit Hilfe von Testsimulationen an einer zweidimensionalen regelmäßigen Anordnung von Ellipsoiden studiert, welche durch 'nächste-Nachbar'-Wechselwirkungen verbunden waren [149]. Hier wurde der zeitliche Verlauf der Orientierung eines ausgewählten Ellipsoids mit Hilfe von Autokorrelationsfunktionen untersucht. Zusätzlich wurde die Rate gelegentlich auftretender abrupter Orientierungsänderungen des Ellipsoids bestimmt und mit einer expliziten Referenzsimulation verglichen, welche die schnellen Freiheitsgrade berücksichtigte.

4. *Vergleich regulärer Größen der exakten Lösung mit regulären, experimentell gewonnenen Größen* (Δ_{p}):

   Dieser Vergleichstyp ist hier von untergeordneter Bedeutung, da er keine Aussagen über algorithmische Genauigkeit erlaubt; vielmehr wird damit die Qualität molekularer Modelle beurteilt. Es sei angemerkt, daß bei Proteinen Δ_{p}-Vergleiche mangels exakter Größen nicht unmittelbar durchführbar sind; vielmehr muß auch hier auf *Simulations*ergebnisse zurückgegriffen werden, d.h., es wird in jedem Fall ein Δ_{pn}-Vergleich durchgeführt. Daraus abgeleitete Aussagen über die Qualität von Modellen sind jedoch nur dann gültig, wenn sichergestellt ist, daß Δ_{n} deutlich kleiner als Δ_{p} ist, wenn also ein zuverlässiger MD-Algorithmus verwendet wird. An dieser Stelle tritt die eingangs angesprochene Wechselbeziehung zwischen algorithmischer Genauigkeit und der Qualität physikalischer Modellbildung klar hervor.

4.2 Problemangepaßte Beurteilungskriterien

Die angeführten Beispiele machen deutlich, daß die Qualität von MD-Algorithmen häufig anhand von Größen beurteilt wird, die zur Erklärung von Proteinfunktion nicht geeignet sind, wie etwa Details atomarer Bewegungen, der detaillierte Verlauf der Kräfte auf einzelne Atome, oder Drift und Fluktuation der Gesamtenergie des isolierten Systems. Größen, die sich zur Beschreibung von Proteinfunktion eignen, werden in der Regel nicht zur Beurteilung algorithmischer Genauigkeit herangezogen. Wir wollen die vier Vergleichsklassen Δ_c, Δ_pn, Δ_n und Δ_p genauer betrachten.

Beurteilungskriterien des Typs Δ_c können für die spezielle Anwendung der Proteindynamiksimulation sogleich als ungeeignet verworfen werden, da chaotische Größen *per definitionem* irrelevant sind: Geringfügige Störungen des Systems finden fortwährend statt (etwa aufgrund von Umgebungseinflüssen) und dürfen nicht über Funktion oder Nichtfunktion eines Proteins entscheiden. Nur unnötig genaue und entsprechend ineffiziente MD-Algorithmen würden eine harte Δ_c-Prüfung bestehen; die meisten effizienten und zur Beschreibung von Proteindynamik besonders geeigneten Methoden dagegen würden zurückgewiesen. In einigen Fällen könnte sich ein Vergleich chaotischer Größen sogar als völlig ungeeignet erweisen, da bereits kleinste Abweichungen der Dynamik, etwa bedingt durch Rundungsfehler, rasch und unkontrolliert verstärkt und dadurch die interessierenden Abweichungen überdeckt werden könnten. Wir folgern, daß ausschließlich *reguläre* Größen die Beurteilung von MD-Algorithmen erlauben. Es sei betont, daß die hier vorgebrachte Argumentation nur für die spezielle Anwendung der MD-Simulation von Proteinen oder vergleichbaren Makromolekülen gültig ist; in anderen Fällen sind Δ_c-Beurteilungen erforderlich und nützlich, wie etwa für die oben angeführte Simulation des Sonnensystems, bei der kleinste Details der Trajektorie von entscheidender Bedeutung sind.

Es verbleiben die Vergleiche des Typs Δ_n und Δ_pn. Wie bereits erläutert, beinhaltet das zweite Verfahren, Δ_pn, den Vergleich von Größen, die durch *zwei* Näherungsschritte getrennt sind — den physikalischen und den numerischen —, so daß dieses Verfahren keine Möglichkeit bietet, die mit beiden Schritten verbundenen Fehler zu trennen und so die numerischen Näherungen unabhängig von den physikalischen Näherungen und Annahmen zu bewerten.

Der Beurteilungstyp Δ_n vermeidet diesen Nachteil: Allein der Vergleich regulärer Testgrößen, die mit dem zu beurteilenden MD-Algorithmus berechnet werden, mit exakten Referenzgrößen erlaubt es, (i) die durch die verwendete numerische Näherung verursachten Artefakte von Inkompatibilitäten zwischen Simulationsergebnissen und experimentellen Daten zu trennen und (ii) effiziente MD-Algorithmen unabhängig von der Qualität des verwendeten physikalischen Modells zu bewerten.

An dieser Stelle ist eine Anmerkung zum bereits mehrmals verwendeten Begriff der 'exakten Referenzgrößen' nötig: Die angesprochene prinzipielle Unmöglichkeit, Δ_p direkt zu bestimmen, könnte dazu verleiten zu glauben, daß ganz analog ein Δ_n-Vergleich nicht durchführbar ist, da auch dieser im allgemeinen nicht verfügbare *exakte* Referenzgrößen erfordert[2]. In diesem Fall genügt es jedoch, Referenzgrößen zu verwenden, deren Abweichung vom exakten Wert als klein gegen den zu bestimmenden Fehler der Testgrößen angenommen werden kann. Insbesondere kann ein numerisches Näherungsverfahren durch Vergleich mit Simulationen beurteilt werden, die ohne diese Näherung durchgeführt wurden. Da wir verschiedene Näherungsverfahren zur effizienten Berechnung der Coulomb-Summe (2.4) beurteilen wollen, sollen die Testgrößen entsprechend mit Referenzsimulationen verglichen werden, in denen die Coulomb-Kräfte explizit nach (2.4) berechnet wurden.

Es sei weiterhin angemerkt, daß die hier vorgestellte Vorgehensweise nicht geeignet ist, *den* besten MD-Algorithmus zu bestimmen. Vielmehr ist jede derartige Beurteilung nur *in Bezug auf die verwendeten Größen* gültig, und sie wird im allgemeinen auch vom speziell für die zugrundeliegenden Testsimulationen verwendeten Proteinmodell abhängen. Wir gehen jedoch davon aus, daß Testsimulationen an einander ähnlichen Proteinmodellen vergleichbare Beurteilungsresultate ergeben. Entsprechend sollte das für die im folgenden geschilderten Testsimulationen herangezogene, in Kapitel 3 vorgestellte Proteinmodell aufgrund seiner realistischen dynamischen Eigenschaften geeignet sein, allgemeine Aussagen über die Qualität von MD-Algorithmen zur Untersuchung von Proteindynamik zu gewinnen.

4.3 Durchgeführte Testsimulationen

An dem in Kapitel 3 vorgestellten vereinfachten Proteinmodell wurden Testsimulationen von je einer Nanosekunde Länge mit Hilfe der Abschneide-Methode 'CUT' sowie mit den sechs in Abschnitt 2.3.1 vorgestellten Entfernungsklassenalgorithmen DC-i, DC-0, DC-1a, DC-1b, DC-1c und DC-1d (vgl. Tabelle 2.1 und Abbildung 2.3) durchgeführt. Für die Abschneide-Methode wurde ein üblicher und, angesichts der geringen Ausdehnung des Testsystems großer Abschneideradius von

[2]Eine Ausnahme hiervon bildet die Drift der Gesamtenergie im Falle eines abgeschlossenen konservativen Systems, die exakt verschwindet.

$r_{cut} = 10$ Å gewählt; die Entfernungsklassenradien wurden entsprechend der am Ende von Abschnitt 2.3.1 angeführten Einteilung (b), $\{R_i\} = \{0; 4; 7; 11; 16; 22; 29; 37\}$ definiert. Zur Untersuchung des Einflusses der Integrationsschrittweite wurden drei verschiedene Schrittweiten, 0,5 fs, 1,0 fs und 2,0 fs, verwendet. Die für die Δ_n-Vergleiche benötigten Referenzgrößen wurden mit Hilfe 'exakter' Referenzsimulationen bestimmt, bei denen alle langreichweitigen Kräfte explizit gemäß (2.4) berechnet wurden.

Alle insgesamt 232 durchgeführten Testsimulationen wurden mit nahezu identischen Anfangsbedingungen gestartet, gegeben durch die atomaren Orte und Geschwindigkeiten der in Abbildung 3.2 (g) gezeigten Struktur des Modellsystems. Zu Beginn einer jeden Simulation wurden einige Atompositionen um einen kleinen, zufälligen Betrag der Größenordnung 10^{-6} Å verschoben und so aufgrund der in Abschnitt 4.1 erläuterten Chaotizität der Dynamik die weitgehende statistische Unabhängigkeit der 232 berechneten Trajektorien erreicht. Aus den erhaltenen 1ns-Trajektorien wurden ausgewählte Observable, die im folgenden Abschnitt beschrieben werden, berechnet und verglichen.

Trotz der beträchtlichen Länge jeder einzelnen Testsimulation waren einige der betrachteten Observablen mit statistischen Fehlern behaftet. In diesen Fällen trugen zu den beobachteten Abweichungen der Testgrößen von den zugehörigen Referenzwerten nicht nur die interessierenden algorithmischen Artefakte bei, sondern zusätzlich statistische Fehler. Um die algorithmischen Artefakte identifizieren zu können, wurde deshalb der statistische Beitrag durch den Vergleich voneinander unabhängiger Referenzsimulationen, in denen keine Artefakte auftreten, abgeschätzt.

Wie in Abschnitt 3.2 beschrieben wurde und in Abbildung 3.5 erkennbar ist, zeigt die Dynamik des Proteinmodells seltene Konformationsübergänge zwischen einigen wenigen Konformationssubzuständen. Zahlreiche Experimente zeigen (siehe z.B. [30,32,33]), daß die durch eine solche Konformationsdynamik bedingte strukturelle Heterogenität von Proteinen auch makroskopische Observable beeinflußt. Auch an dem hier verwendeten vereinfachten Proteinmodell ist ein solcher Effekt bei eine Reihe verschiedener Beobachtungsgrößen nachweisbar, soll aber hier nicht näher beschrieben werden. Ebenso ist bekannt, daß Näherungen bei der Beschreibung langreichweitiger Kräfte die Proteinstruktur artifiziell verändern können [79]; auch dies soll hier nicht untersucht werden. Wir wollen vielmehr den Einfluß der beschriebenen unterschiedlichen Näherungsverfahren auf die *Dynamik* von Proteinen aufdecken, sodaß lediglich diejenigen Fälle betrachtet werden dürfen, in denen die Molekülstruktur ungestört und deren Einfluß auf die beobachteten Testgrößen somit ausgeschlossen ist.

Aus diesem Grund wählten wir aus den 232 Testsimulationen diejenigen aus, in denen *keine* Konformationsübergänge auftraten. Dazu wurde sichergestellt, daß (a) die gemittelte Proteinstruktur keiner der verwendeten Testsimulationen von der Referenzstruktur allzusehr abweicht und (b) keine abrupten Änderungen ausgewählter Atomabstände im Verlauf der Simulation auftraten, die auf einen Konformationsübergang hindeuten würden.

Entsprechend dem ersten Kriterium bestimmten wir für jede Testsimulation die *maximale* Abweichung mittlerer Atompositionen von den Referenzwerten. Simulationen, die eine Abweichung von mehr als $2,5\,\text{Å}$ zeigten, wurden verworfen. Für das zweite Kriterium wurden vier Bausteine des Proteinmodells ausgewählt. Die Lage dieser vier Bausteine, #12, #36, #63 und #87, ist in Abbildung 3.2 (g) durch Kreise angedeutet. Da, wie Abbildung 3.5 zeigt, Konformationsübergänge charakteristische, abrupte Änderungen des gegenseitigen Abstands dieser Atome bewirken, wurden nur solche 1ns-Trajektorien zugelassen, die keine derartigen Sprünge aufweisen.

Die getroffene Auswahl verhindert, daß beobachtete Abweichungen von der Referenzsimulation lediglich die Folge zufällig auftretender Konformationsübergänge oder artifizieller struktureller Änderungen sind. Jede beobachtete, statistisch signifikante Abweichung ist somit ein zuverlässiger Indikator für eine algorithmisch bedingte Verfälschung der *Dynamik* unseres Modells.

4.4 Relevante Observable

Zur Beurteilung der angesprochenen sieben MD-Algorithmen wurde eine Reihe regulärer Größen ausgewählt. Auswahlkriterium war die *Relevanz* der verwendeten Größen, d.h., sie sollten geeignet sein, Proteinfunktion zu beschreiben und zu erklären. Die ausgewählten Observablen werden im folgenden definiert und erläutert. Sie beinhalten atomare Beweglichkeiten sowie Spektren, Autokorrelationsfunktionen und Kovarianzen atomarer Bewegungen. Neben diesen häufig zur Charakterisierung atomarer Dynamik verwendeten Observablen wollen wir auch Konfigurationsraumdichten, die von der Proteindynamik erzeugt werden, untersuchen. Wie näher ausgeführt wird, stehen derartige Dichten in enger Beziehung zur freien Energie sowie zu Reaktionsraten und sind deshalb funktionsrelevant.

Der Grad der in Abschnitt 4.1 angesprochenen Verletzung der Energieerhaltung ist ein weitverbreitetes Maß algorithmischer Genauigkeit. Um den Vergleich mit den im Rahmen der vorliegenden Arbeit vorgeschlagenen und eingesetzten problemangepaßten Beurteilungskriterien zu ermöglichen, wurde die in den Testsimulationen auftretende Drift der Gesamtenergie untersucht.

Zur Definition der verwendeten relevanten Größen werden folgende Abkürzungen benützt: $\mathbf{r}_\alpha(t_i)$ ist der Ort von Atom α zur Zeit t_i, dem i-ten Integrationsschritt einer Simulation[3]; $\langle \mathbf{r}_\alpha \rangle := \sum_{t_i} \mathbf{r}_\alpha(t_i)/N_t$ ist die mittlere Position von Atom α, berechnet aus einer durch N_t Koordinatensätze repräsentierten Trajektorie der Länge $T_t = 1,048,576\,\text{fs}$, wobei $N_t := T_t/\Delta_t$ die Anzahl der ausgeführten Integrationsschritte und Δ_t die verwendete Integrationsschrittweite ist. Mit *'ref'* indizierte Observable sind aus einer Referenzsimulation abgeleitet.

a) Drift der Gesamtenergie

Die durch Entfernungsklassenalgorithmen verursachten Rauschkräfte ließen eine teilweise beträchtliche resultierende Drift der Gesamtenergie erwarten. Aufgrund der langen Simulationszeit von einer Nanosekunde mußte mit Werten von einigen hundert kcal/mol gerechnet werden. Eine derartige artifizielle Verschiebung der Gesamtenergie des Modellsystems würde dessen Temperatur nachhaltig ändern und dadurch weitere Vergleiche relevanter Observabler verhindern. Aus diesem Grund wurde bei allen Simulationen die Konstanz der Gesamtenergie, und damit der Temperatur, durch schwache Kopplung des Systems an ein Wärmebad erzwungen [150–152]. Der Energiefluß vom oder zum Wärmebad wurde als Maß für die algorithmisch bedingte Energiedrift verwendet.

b) Gemitteltes Spektrum atomarer Bewegungen

Spektren atomarer Bewegungen $\hat{\mathbf{r}}_\alpha(\omega)$ wurden durch Fourier-Transformation der Trajektorien einzelner Atome α gewonnen. Derartige Spektren bestimmen den Beitrag der Bewegung des betrachteten Atoms zum Schwingungsspektrum eines Proteins. Als Maßstab algorithmischer Genauigkeit verwendeten wir das über alle Atome gemittelte Leistungsspektrum $S(\omega)$:

$$S(\omega) := \sum_\alpha |\hat{\mathbf{r}}_\alpha(\omega)|^2 \,. \tag{4.1}$$

Aus den Testsimulationen wurden Spektren in zwei verschiedenen Frequenzbereichen berechnet: Hochfrequenzspektren im Bereich $0,5 \ldots 250\,\text{ps}^{-1}$ und Niederfrequenzspektren im Bereich $0,001 \ldots 4\,\text{ps}^{-1}$.

[3] Aus technischen Gründen wurde in allen Summen über α lediglich jedes fünfte der hundert Atome berücksichtigt; d.h. $\alpha \in \{1, 6, 11, \ldots, 91, 96\}$.

b1) Hochfrequenzspektren atomarer Bewegungen

Zur Berechnung der Hochfrequenzspektren wurde jede 1ns-Trajektorie in $M = 511$ aufeinanderfolgende Abschnitte $\{\mathbf{r}(m\Delta_M), \ldots, \mathbf{r}((m+2)\Delta_M);\ m = 0 \ldots M-1\}$ der Länge $2\Delta_M = 4096\,\mathrm{fs}$ zerlegt. Entsprechend dem in [153] angegebenen Verfahren zur Minimierung statistischer Fehler wurden alle Abschnitte mit einer 'Hanning-Funktion' gewichtet und anschließend einer Fourier-Transformation unterworfen. Die erhaltenen M Hochfrequenzspektren wurden gemittelt.

b2) Niederfrequenzspektren atomarer Bewegungen

Die Niederfrequenzspektren wurden mit Hilfe geglätteter Trajektorien $\bar{\mathbf{r}}_\alpha(t)$,

$$\bar{\mathbf{r}}_\alpha(t) = \frac{1}{Z} \sum_{k=-10\sigma/\Delta_t}^{10\sigma/\Delta_t} \mathbf{r}_\alpha(t)\ \exp\left(-\frac{(t-k\Delta_t)^2}{2\sigma^2}\right)$$

mit

$$Z := \sum_{k=-10\sigma/\Delta_t}^{10\sigma/\Delta_t} \exp\left(-\frac{(t-k\Delta_t)^2}{2\sigma^2}\right)$$

gemäß

$$\hat{\mathbf{r}}_\alpha(\omega) = \sum_{k=0}^{N_{\mathrm{sm}}-1} \bar{\mathbf{r}}_\alpha(t_k)\exp(i\omega t_k)\,, \quad t_k = k\sigma\,,$$

berechnet. Die Abtastfrequenz der geglätteten Trajektorien wurde gleich der Breite $\sigma = 128\,\mathrm{fs}$ der Glättungsfunktion gewählt; $N_{\mathrm{sm}} := T_t/N_t$ bedeutet hier die Anzahl geglätteter Koordinatensätze.

c) Atomare Beweglichkeiten

Der mittlere Abstand σ_α des Atoms α von seinem mittleren Ort $\langle\mathbf{r}_\alpha\rangle$ ist durch

$$\sigma_\alpha = \sqrt{\sum_{t_i}((\mathbf{r}_\alpha(t_i) - \langle\mathbf{r}_\alpha\rangle)^2)/N_t} \tag{4.2}$$

gegeben. Diese 'rms-Abweichung' σ_α ist ein Maß atomarer Beweglichkeit. Im allgemeinen hängt diese Beweglichkeit von der molekularen Umgebung des betrachteten Atoms ab: Beispielsweise zeigen die Atome von Proteinseitengruppen, die an der Proteinoberfläche dem Lösungsmittel ausgesetzt sind, eine größere Beweglichkeit als Atome des Proteinrückgrats [95]. Da atomare Beweglichkeiten etwa mit Hilfe von Röntgen- oder Neutronenstreuversuchen gemessen werden können, werden sie häufig zur Überprüfung der Qualität molekularer Modelle verwendet.

d) Korrelationen atomarer Bewegungen

Korrelationen atomarer Bewegungen ermöglichen es, Proteindynamik genauer zu charakterisieren. Autokorrelationsfunktionen atomarer Orte oder Geschwindigkeiten beschreiben den diffusiven Charakter der Bewegungen und erlauben die Bestimmung von Reibungskoeffizienten [95]. Kovarianzen der Bewegung von Atompaaren beschreiben kollektive Bewegungen und ermöglichen die Untersuchung dynamisch-funktionaler, möglicherweise kausaler Beziehungen zwischen entfernten Regionen im Protein (vgl. z.B. [154]). Aufgrund der weiten Verbreitung dieser Größen untersuchten wir Orts-Autokorrelationsfunktionen $C_\alpha^{\mathrm{dis}}(\tau)$, Geschwindigkeits-Autokorrelationsfunktionen $C_\alpha^{\mathrm{vel}}(\tau)$ bezüglich Atom α sowie Kovarianzen $K_{\alpha\beta}$ der Bewegung der Atome α und β als relevante Größen. Diese wurden gemäß

$$C_\alpha^{\mathrm{dis}}(\tau_j) = \frac{\sum_{i=0}^{N_j-1}(\mathbf{r}_\alpha(t_i) - \langle\mathbf{r}_\alpha\rangle)(\mathbf{r}_\alpha(t_i+\tau_j) - \langle\mathbf{r}_\alpha\rangle)}{\sum_{i=0}^{N_j-1}(\mathbf{r}_\alpha(t_i) - \langle\mathbf{r}_\alpha\rangle)^2} \tag{4.3}$$

und

$$C_\alpha^{\mathrm{vel}}(\tau_j) = \frac{\sum_{i=0}^{N_j-1}\mathbf{v}_\alpha(t_i)\,\mathbf{v}_\alpha(t_i+\tau_j)}{\sum_{i=0}^{N_j-1}\mathbf{v}_\alpha(t_i)^2} \tag{4.4}$$

sowie

$$K_{\alpha\beta} = \frac{k_{\alpha\beta}}{\sqrt{k_{\alpha\alpha}\,k_{\beta\beta}}} \tag{4.5}$$

berechnet, wobei in (4.5) die Kovarianzmatrix

$$k_{\alpha\beta} = \sum_{t_i}(\mathbf{r}_\alpha(t_i) - \langle\mathbf{r}_\alpha\rangle)(\mathbf{r}_\beta(t_i) - \langle\mathbf{r}_\beta\rangle)$$

verwendet wurde. In (4.3–4.5) ist $N_j = (T_t - \tau_j)/\Delta_t$ und $\tau_j = j\,\Delta_t$ gesetzt.

e) Projizierte Konfigurationsraumdichte

Eine grundlegende thermodynamische Größe ist die Phasenraumdichte $\rho(\mathbf{r}^N, \mathbf{p}^N)$, welche durch die Dynamik eines Vielteilchensystems erzeugt wird. Im Fall der Proteindynamik bezeichnen $\mathbf{r}^N$ und $\mathbf{p}^N$ $3N$-dimensionale Vektoren, die aus den N Atompositionen bzw. -impulsen zusammengesetzt sind. Die Dichte ρ wird selten direkt bestimmt und verwendet, sie ist vielmehr wegen ihrer engen Beziehung zu statistischen Observablen wie Entropie, freier Energie oder Reaktionsraten von Bedeutung. Sie kann mit Hilfe eines Boltzmann-Ensembles molekularer Konfigurationen, das mit Monte-Carlo-Verfahren oder MD-Simulationen gewonnen werden kann, näherungsweise bestimmt werden, indem der Phasenraum in Volumenelemente V^k unterteilt wird: Der Anteil N^k aller in V^k liegenden Trajektorienpunkte konvergiert für lange Simulationszeiten gegen die Phasenraumdichte ρ am Ort von V^k.

Solche Näherungen der Phasenraumdichte sind jedoch aufgrund der begrenzten Länge von MD-Simulationen sowie aufgrund der hohen Dimensionalität des Phasenraums in der Regel keine regulären Größen. Schon für unser kleines Proteinmodell ist der zugehörige Phasenraum 600-dimensional. Die Zahl K der zur Bestimmung der Phasenraumdichte benötigten Volumenelemente V^k ist deshalb selbst bei grober Auflösung außerordentlich groß und entsprechend entfallen nur wenige Trajektorienpunkte auf die einzelnen Volumenelemente. Der resultierende statistische Fehler ist beträchtlich und verhindert die direkte Verwendung von Phasenraumdichten als Maß algorithmischer Genauigkeit.

Wir umgehen dieses Problem, indem wir statt *aller* $6N$ Freiheitsgrade des Systems lediglich *einige wenige* berücksichtigen. Dazu betrachte man die *Projektion* $\rho_c(\mathbf{c})$ der Phasenraumdichte ρ auf einen niedrigdimensionalen Unterraum, der durch die M Freiheitsgrade $\mathbf{c} := \{c_i(\mathbf{r}^N, \mathbf{p}^N)\}$, $i = 1, \ldots, M$ aufgespannt wird,

$$\rho_c(\mathbf{c}) = \int d\mathbf{r}'^N d\mathbf{p}'^N \rho(\mathbf{r}'^N, \mathbf{p}'^N) \delta(\mathbf{c} - \mathbf{c}') \,.$$

Die so definierte 'projizierte Phasenraumdichte' ρ_c ist genau dann eine reguläre Größe, wenn die Anzahl von Trajektorienpunkten pro Volumenelement *des Unterraums* groß genug ist, um statistisch signifikante Aussagen zu erlauben. Wir beschränken uns hier auf Projektionen auf *Konformations*freiheitsgrade $\mathbf{c} = \{c_i(\mathbf{r}^N)\}$, welche auch als *Konformationskoordinaten* bezeichnet werden [34, 22].

Zur Auswertung der durchgeführten Testsimulationen wählten wir eine Reihe verschiedener Projektionen auf zwei-, drei- und vier-dimensionale Konformationsunterräume. Insgesamt wurden aus jeder Simulation $P = 50$ projizierte Konfigurationsraumdichten bestimmt, die durch Paare, Tripel bzw. Quadrupel interatomarer Abstände definiert sind. Die zwei-dimensionalen Unterräume wurden durch Abstandspaare $\mathbf{c} = (r_{\alpha\beta}, r_{\gamma\delta})$ zwischen vier ausgewählten Atomen $\alpha = 12$, $\beta = 36$, $\gamma = 63$ und $\delta = 87$ definiert; die Lage der vier gewählten Atome ist in der in Abbildung 3.2 (g) gezeigten Banddarstellung des Proteinmodells durch vier Kreise markiert. Alle 15 möglichen Projektionen, die mit Hilfe der sechs Abstände $r_1 = r(\#12; \#36)$, $r_2 = r(\#12; \#63)$, $r_3 = r(\#12; \#87)$, $r_4 = r(\#36; \#63)$, $r_5 = r(\#36; \#87)$ und $r_6 = r(\#63; \#87)$ gebildet werden können, wurden berechnet und daraus 15 verschiedene zweidimensionale Dichteverteilungen bestimmt. In gleicher Weise wurden 20 unterschiedliche Projektionen auf drei-dimensionale Unterräume sowie 15 vier-dimensionale Projektionen definiert und für jede Testsimulation bestimmt.

Die Entscheidung, algorithmische Genauigkeit anhand der beschriebenen Projektionen zu beurteilen, die durch interatomare Abstände definiert sind, erscheint zunächst willkürlich und soll daher näher erläutert werden. Wie bereits erwähnt, sind Differenzen der Entropie oder der freien Energie Funktionen der Phasenraumdichte ρ. Insbesondere sind konformationsbestimmte Reaktionsraten Funktionen *projizierter Phasenraumdichten* ρ_c. Erst die zuverlässige Berechnung niedrigdimensionaler Projektionen der betrachteten Art ermöglicht es, solche Raten zuverlässig zu bestimmen und zur Erklärung der Kinetik von Proteinfunktion mittels geeigneter MD-Simulationen zu verwenden.

Abbildung 4.3 zeigt ein vereinfachte Beispiel, das diesen Sachverhalt verdeutlichen soll. Hierzu wird eine hypothetische Bindungsreaktion zweier Proteine **A** und **B** betrachtet; ein solcher Prozeß ist etwa die in der Einleitung angesprochene Bindung von G-Proteinen an Photorezeptoren. Wir nehmen an, daß die beiden Proteine jeweils drei spezifische Bindungsstellen besitzen (in der Abbildung durch die Dreiecks-, Halbkreis- und Rechteckssymbole dargestellt), deren gegenseitige Abstände, wie abgebildet, mit r_{12} und r_{23} bzw. R_{12} und R_{23} bezeichnet sind. Der Einfachheit halber soll Protein **A** als flexibel angenommen werden, so daß die Abstände r_{12} und r_{23} beträchtliche Fluktuationen aufweisen, während Protein **B** vergleichsweise starr sein soll, so daß R_{12} bzw. R_{23} als zeitlich konstant angenommen werden können. In Abb. 4.3 sind diese Eigenschaften durch die Form der beiden Proteine angedeutet. Während sich das System bewegt und die betrachteten Abstände fluktuieren, wird die Wahrscheinlichkeit (pro Zeiteinheit) einer Bindungsreaktion der beiden Proteine variieren: Passen die drei Bindungsstellen gut aufeinander, d.h., $r_{12} \approx R_{12}$ und $r_{23} \approx R_{23}$, so ist die Bindungswahrscheinlichkeit hoch, andernfalls gering. Die resultierende Reaktionsrate wird demnach davon abhängen, wie oft die drei Bindungsstellen geeignet angeordnet sind; sie ist also in dem betrachteten fiktiven Beispiel durch den Wert der projizierten Konfigurationsraumdichte $\rho_c(r_{12},r_{23})$ in der Umgebung von (R_{12}, R_{23}) bestimmt.

In vergleichbarer Weise wurde in [155] die Rate eines Elektronentransfers, der in einem photosynthetischen Protein auftritt, mit Hilfe von MD-Rechnungen abgeschätzt. Die Autoren untersuchten die Wahrscheinlichkeit des betrachteten Elektronentransfers im Verlauf der molekularen Dynamik; diese Wahrscheinlichkeit kann für jeden Zeitpunkt im Rahmen der Marcus-Theorie [156] aus der augenblicklichen molekularen Konfiguration bestimmt werden: Sie ist genau dann groß, wenn die mit dem als instantan angenommenen Elektronentransfer verbundene Änderung der elektrostatischen Energie des Proteins verschwindet. Da diese Energiedifferenz eine Funktion aller Atompositionen des Proteins ist, ist auch in diesem Fall die betrachtete Transferrate durch eine projizierte Konfigurationsraumdichte bestimmt, wobei ρ hier auf einen eindimensionalen Unterraum zu projizieren ist.

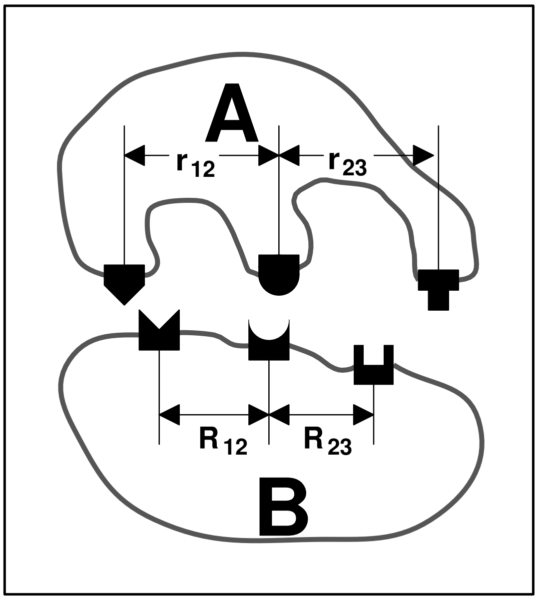

Abbildung 4.3: Eine schematisch dargestellte Bindungsreaktion zweier Proteine **A** und **B** verdeutlicht die Relevanz projizierter Konfigurationsraumdichten zur Beschreibung von Proteinfunktion; zur näheren Erläuterung siehe Text.

Die beiden angeführten Beispiele machen deutlich, daß die zur Beurteilung algorithmischer Genauigkeit herangezogenen projizierten Konfigurationsraumdichten in der Tat für eine große Klasse verschiedener Anwendungen von MD-Simulationen als relevant bezeichnet werden können.

Um die benötigten $P = 50$ projizierten Konfigurationsraumdichten zu bestimmen, wurde aus jeder Testsimulation, durchgeführt mit Algorithmus a, ein Ensemble S_a, bestehend aus $N_c = 131\,072$ Konfigurationen, gewonnen. Jeder der P Unterräume wurde in K rechteckige Volumenelemente $V^{k,p}$, $k = 1 \ldots K$, $p = 1 \ldots P$, unterteilt. Die Volumenelemente wurden so gewählt, daß auf alle $V^{k,p}$ der gleiche Anteil $N^{ref} := N_c/K$ von Konfigurationen Referenzensembles R_1 entfällt. Zur Veranschaulichung zeigt Abbildung 4.4 eine derartige Unterteilung eines zwei-dimensionalen Unterraums in $K = 16$ rechteckige Bereiche $V^{16,p}$, die durch Linien getrennt sind. Das Ensemble R_1 ist als Punktwolke eingezeichnet, wobei jeder Punkt eine Konfiguration des Ensembles repräsentiert[4].

Unter Verwendung der beschriebenen Unterteilungen ermittelten wir die Abweichungen $\Delta^p_{S_a - R_1}$ der mit dem Algorithmus a berechneten projizierten Konfigurationsraumdichte von der Referenzdichte gemäß

$$\Delta^p_{S_a - R_1} = \sqrt{\frac{1}{K-1} \sum_{k=1}^{K} \left(\frac{N^{k,p}_{S_a} - N^{ref}}{N^{ref}} \right)^2}\,, \tag{4.6}$$

wobei $N^{k,p}_{S_a}$ die Anzahl derjenigen Konfigurationen des Ensembles S_a bezeichnet, die

[4]Der Übersichtlichkeit halber zeigt die Abbildung nur einen Bruchteil aller N_c verwendeten Konfigurationen.

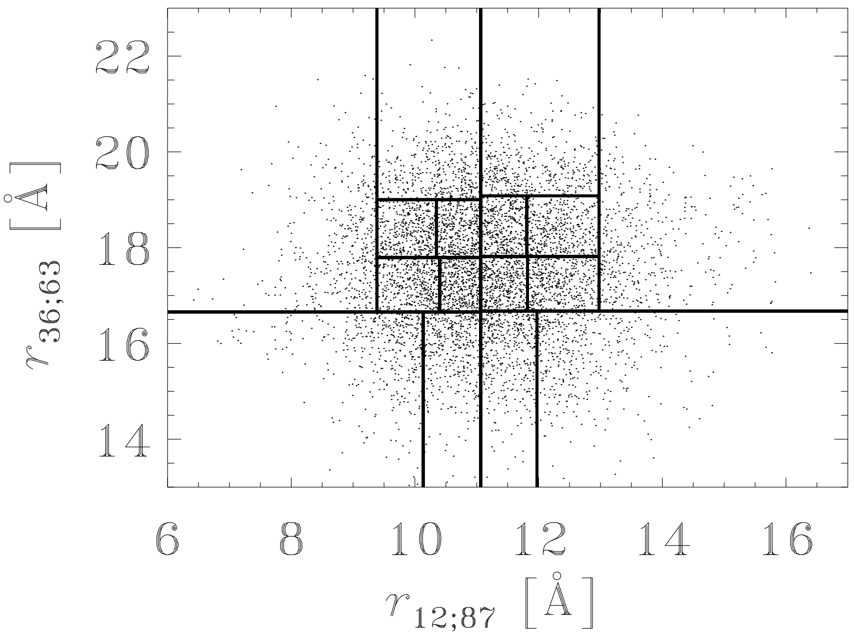

Abbildung 4.4: Ensemble R_1 von Konfigurationen, erzeugt aus einer Referenzsimulation des vereinfachten Proteinmodells; jeder der Phasenraumpunkte wurde auf den durch $\mathbf{c} = (r_{12;87}, r_{36;63})$ definierten zwei-dimensionalen Konformationsunterraum projiziert und ist durch einen Punkt dargestellt; weiterhin ist die im Text diskutierten Unterteilung $V^{16,p}$ des Unterraums (Linien) eingezeichnet.

nach Projektion p in den Bereich $V^{k,p}$ fallen.

Die Abweichungen $\Delta^p_{S_a-R_1}$ setzen sich im allgemeinen aus zwei Anteilen zusammen: Ein Anteil resultiert aus statistischen Fluktuationen der Häufigkeiten $N^{k,p}_{S_a}$ und tritt immer auf. Darüberhinaus können beobachtete Abweichungen aber auch aus einer Verfälschung der Dynamik durch den untersuchten MD-Algorithmus entstehen. Um solche Artefakte signifikant nachweisen zu können, muß der statistische Beitrag abgeschätzt werden. Wir weisen darauf hin, daß die mittlere statistische Schwankung der $N^{k,p}_{S_a}$ sicherlich deutlich größer als $(N^{k,p}_{S_a})^{1/2}$ sein wird, da aufeinanderfolgende Konfigurationen, deren zeitlicher Abstand lediglich $T_t/N_c = 8\,\mathrm{fs}$ beträgt, stark korreliert sind. Aus diesem Grund ist eine einfache Abschätzung unter der Annahme statistisch unabhängiger Ereignisse sicher nicht möglich.

Deshalb ermittelten wir den statistischen Beitrag zur Abweichung $\Delta^p_{S_a-R_1}$ empirisch, indem wir die vom Ensemble R_1 erzeugten projizierten Konfigurationsraumdichten mit einem zweiten, unabhängigen Referenzensemble R_2 verglichen. Beide Referenzensembles wurden mit dem selben Verlet-Algorithmus erzeugt, wobei, wie bereits beschrieben, die Anfangsbedingungen der zweiten Referenzsimulation geringfügig gegenüber der ersten Simulation verändert wurden, um so eine weitgehend unabhängige Trajektorie zu gewinnen. Entsprechend sollte der Mittelwert der bei den 50 verschiedenen Projektionen beobachteten Abweichung $\Delta^p_{R_2-R_1}$ der beiden Referenzsimulationen,

$$\langle \Delta_{R_2-R_1} \rangle := (\sum_{p=1}^{P} \Delta^p_{R_2-R_1})/P\,, \tag{4.7}$$

eine Näherung für die zu erwartenden statistischen Fluktuationen ergeben. Als Maß des mit dieser Abschätzung verbundenen Fehlers wurde die Varianz $\sigma_{R_2-R_1}$ der Abweichungen $\Delta^p_{R_2-R_1}$,

$$\sigma_{R_2-R_1} := \sqrt{\frac{1}{P-1}\sum_{p=1}^{P}(\Delta^p_{R_2-R_1} - \langle \Delta_{R_2-R_1} \rangle)^2}\,, \tag{4.8}$$

bestimmt. In gleicher Weise wurde für jeden getesteten Algorithmus die Varianz $\sigma_{S_a-R_1}$ der Abweichungen $\Delta^p_{S_a-R_1}$ ermittelt. Der statistisch signifikante Anteil zur mittleren Abweichung $\langle \Delta^p_{S_a-R_1} \rangle$, die interessierende algorithmisch bedingte Abweichung Δ_a, kann somit durch

$$\Delta_a = \langle \Delta_{S_a-R_1} \rangle - \langle \Delta_{R_2-R_1} \rangle \pm \sigma_a \tag{4.9}$$

und ein Fehlerintervall σ_a angegeben werden.

Wären die $P = 50$ Konfigurationsraumprojektionen statistisch unabhängig voneinander, dann wäre

$$\sigma_a \approx \sigma_a^{\min} := \sqrt{(\sigma^2_{R_2-R_1} + \sigma^2_{S_a-R_1})/P}\,. \tag{4.10}$$

Da jedoch die P Projektionen nicht als *vollständig* statistisch unabhängig angenommen werden dürfen, muß als obere Schranke der Fall vollständiger Korrelation der Projektionen,

$$\sigma_a^{\max} := \sqrt{(\sigma^2_{R_2-R_1} + \sigma^2_{S_a-R_1})}\,, \tag{4.11}$$

berücksichtigt werden. Der tatsächliche Fehler σ_a liegt dann sicher im Bereich zwischen $\sigma_a^{\min}$ und $\sigma_a^{\max}$.

Es wurde noch nicht angesprochen, mit welcher Auflösung K (vgl. (4.6)) die von den verschiedenen Testsimulationen abgeleiteten projizierten Konfigurationsraumdichten ρ_c mit den jeweiligen Referenzdichten verglichen wurden. Tatsächlich ist es nicht leicht, *von vorneherein* zu entscheiden, in wieviele Volumenelemente $V^{k,p}$ der betrachtete Konformationsunterraum zu zerlegen ist: Eine grobe Unterteilung in einige wenige Bereiche führt zwar einerseits zu kleinen statistischen Fehlern $\langle \Delta_{R_2-R_1} \rangle$, da in diesem Fall auf jedes Volumenelement viele Konfigurationen entfallen; Details der Dynamik können so aber nicht verglichen werden. Im Gegensatz dazu erlaubt eine große Anzahl kleiner Bereiche einen Vergleich hoher Auflösung, der aber wegen der auftretenden großen statistischen Fehler nicht signifikant ist. Um eine vorurteilsfreie Analyse zu gewährleisten, wurde die beschriebene Vergleichsprozedur auf einer Reihe unterschiedlicher Auflösungsstufen durchgeführt.

4.5 Ergebnisse

Die 232 Testsimulationen von je einer Nanosekunde Dauer wurden unter Verwendung der oben definierten und beschriebenen relevanten Observablen ausgewertet. Entsprechend den in Abschnitt 4.3 beschriebenen Kriterien zur Vermeidung struktureller Einflüsse wurden für jeden der acht untersuchten MD-Algorithmen[5] und für jede der verwendeten Integrationsschrittweiten, 0,5 fs, 1,0 fs, und 2,0 fs, eine geeignete 1ns-Trajektorie zur weiteren Analyse ausgewählt. In allen Fällen waren die bei verschiedenen Integrationsschrittweiten bestimmten algorithmischen Genauigkeiten konsistent und spiegelten die bekannte Tatsache wider, daß mit der numerischen Integration verbundene Diskretisierungsfehler mit wachsender Integrationsschrittweite zunehmen. Aus diesem Grund werden wir häufig auf die Darstellung der Ergebnisse aller drei Integrationsschrittweiten verzichten; stellvertretend sind dann lediglich die Resultate der 1,0 fs-Simulationen gezeigt. In einigen Fällen, wie etwa bei der Analyse der Niederfrequenzspektren, erwies es sich als vorteilhaft, den statistischen Fehler der betrachteten Observablen durch Mittelung über alle Trajektorien verschiedener Integrationsschrittweiten zu reduzieren.

a) Drift der Gesamtenergie

Abbildung 4.5 zeigt die Drift der Gesamtenergie, die bei den 24 ausgewählten 1ns-Simulationen auftrat und kompensiert werden mußte. Bei allen betrachteten Algorithmen läßt sich eine mit der Integrationsschrittweite zunehmende Energiedrift feststellen, die durch die entsprechende Zunahme des Diskretisierungsfehlers hervorgerufen wird. Die Drift der Gesamtenergie des Systems ist im Falle der Referenzsimulation REF, also bei Verwendung des herkömmlichen Verlet-Algorithmus, äußerst gering; auch CUT verletzt die Energieerhaltung nur wenig. Im Gegensatz dazu erzeugen Entfernungsklassenalgorithmen relativ große Wärmebeträge. Eine derartige artifizielle Energiezufuhr würde das System ohne eine entsprechende Kompensation

[5] REF, CUT, DC-i, DC-0, DC-1a, DC-1b, DC-1c und DC-1d

durch schwache Kopplung an ein Wärmebad innerhalb der Simulationszeit von einer Nanosekunde denaturieren.

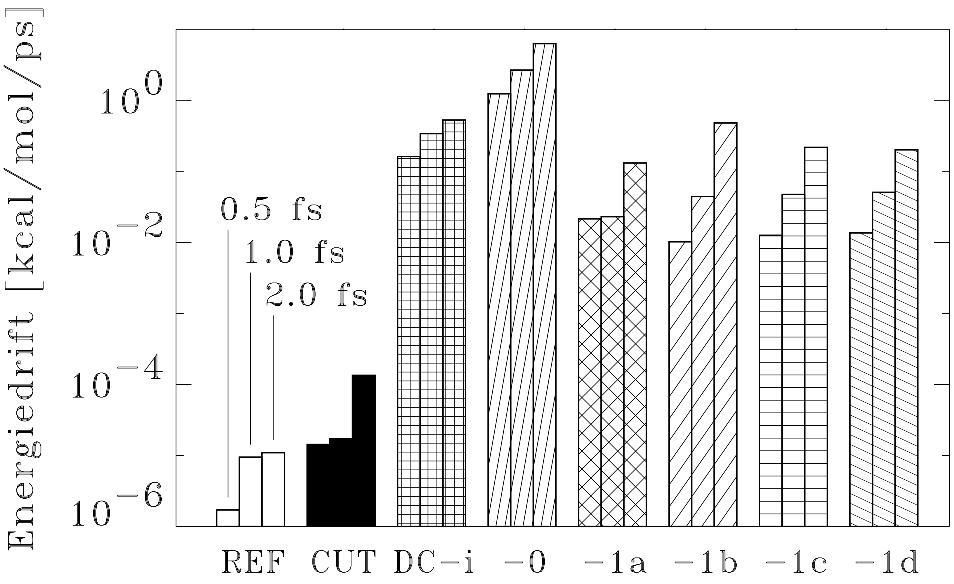

Abbildung 4.5: Drift der Gesamtenergie pro Pikosekunde für die Referenzsimulation REF sowie für die sieben untersuchten MD-Algorithmen. Alle acht Algorithmen wurden mit den Integrationsschrittweiten 0, 5 fs, 1, 0 fs und 2, 0 fs eingesetzt.

Diese große Drift der Gesamtenergie wird von Fehlern in der Berechnung der langreichweitigen Kräfte hervorgerufen, die mit der in Abbildung 2.3 veranschaulichten Kraftextrapolation verbunden sind. Wie in Abschnitt 2.3.1 diskutiert wurde, stellen diese Abweichungen von den exakten Coulombkräften Rauschkräfte dar, welche kinetische Energie auf das System übertragen. Vergleicht man die in Abb. 4.5 gezeigten Beträge der übertragenen Energie mit der Größe der in Abb. 2.3 dargestellten Rauschkräfte, so fällt jedoch auf, daß große Rauschkräfte nicht notwendig zu einer starken Verletzung der Energieerhaltung führen: So sind etwa mit den Extrapolationen DC-i, DC-1c, DC-1a und DC-1d größere Rauschkräfte verbunden als mit den übrigen beiden Algorithmen; die letzten drei der genannten Extrapolationen, welche alle dem Verlet-Kriterium genügen, zeigen jedoch eine vergleichsweise geringe Energiedrift. Umgekehrt wird die größte beobachtete Drift durch die relativ kleinen Rauschkräfte des Algorithmus DC-0 hervorgerufen; dieser Algorithmus erfüllt das Verlet-Kriterium nicht. Dieses Ergebnis zeigt, daß das mit der Formulierung des Verlet-Kriteriums verfolgte Ziel, einen dem Verlet-Algorithmus möglichst ähnlichen Entfernungsklassenalgorithmus zu formulieren, zumindest im Hinblick auf das Verhalten der Gesamtenergie erreicht wurde.

Es sei erneut darauf hingewiesen, daß dieses Ergebnis für sich genommen *keine* problemangepaßte Beurteilung der Algorithmen impliziert. Es ist vielmehr zu fragen, ob die durch das Verlet-Kriterium erreichte Verringerung der Energiedrift zu einer verbesserten Beschreibung *relevanter* Observabler führt. Die im folgenden dargestellten Ergebnisse sollen diese Frage beantworten.

b1) Hochfrequenzspektren atomarer Bewegungen

Wie in der Abbildung 2.3 erkennbar ist, weist der zeitliche Verlauf der durch Entfernungsklassenalgorithmen extrapolierten Kräfte Unstetigkeiten auf. Da vier Entfernungsklassen besetzt sind, treten diese Kraftsprünge alle $2^{4-1} = 8$ Integrationsschritte auf. Entsprechend sollte sich deren Einfluß auf die Atombewegung in den gemittelten Fourierspektren der atomaren Bewegung bei *hohen* Frequenzen widerspiegeln. Abbildung 4.6 zeigt die aus den Testsimulationen bestimmten Hochfrequenzspektren, die gemäß (4.1) für die Integrationsschrittweiten $0,5\,\mathrm{fs}$ und $1,0\,\mathrm{fs}$ berechnet wurden. Links oben in der Abbildung sind die beiden Spektren der Referenzsimulation (REF) gestrichelt eingezeichnet. Da hier keine algorithmischen Rauschkräfte vorhanden sind, verschwinden die Spektren oberhalb $100\ \mathrm{ps}^{-1}$. Im Gegensatz dazu zeigen die entsprechenden beiden Spektren der Simulationen DC-i (durchgezogene Linien) scharfe Resonanzspitzen bei $250\ \mathrm{ps}^{-1}$ und, für die Schrittweite von $1,0\,\mathrm{fs}$, zusätzlich bei $125\ \mathrm{ps}^{-1}$. Dies sind tatsächlich die erwarteten Frequenzen, die sich aus der Dauer von acht Integrationsschritten ergeben.

Rechts oben in Abb. 4.6 sind dieselben Resultate aufgetragen, diesmal als relative Abweichung des Spektrums DC-i gegenüber REF. Auch hier, wie in allen übrigen Bildern der Abbildung, sind Resultate für die Integrationsschrittweiten von $0,5\,\mathrm{fs}$ bzw. $1,0\,\mathrm{fs}$ gezeigt. Die vier Abbildungen der unteren Bildhälfte zeigen in gleicher Weise die mit den Algorithmen DC-0, DC-1a, DC-1b und DC-1c verbundenen relativen Fehler. Man beachte hier jedoch den vergrößerten Maßstab, der zur Darstellung der relativ zu DC-i kleineren Fehler erforderlich war. Nicht gezeigt sind die Hochfrequenzspektren der Simulationen CUT sowie DC-1d: Erstere gleichen zu den entsprechenden Referenzspektren, letztere den Spektren von DC-1c. In allen Fällen verhalten sich die Höhen der resonanten Rauschspitzen wie die Größen der in Abb. 2.3 dargestellten Kraftsprünge. Aufgrund der Länge der Simulationen zeigten die betrachteten Hochfrequenzspektren keine statistischen Fluktuationen; sie sind also reguläre Größen.

Sind diese teilweise deutlich hervortretenden Artefakte relevant? Wir weisen zunächst auf die Bande des im linken oberen Teil der Abb. 4.6 dargestellten Spektrums bei etwa $30\,\mathrm{ps}^{-1}$ hin, welche von Bindungsstreckschwingungen herrührt. Die-

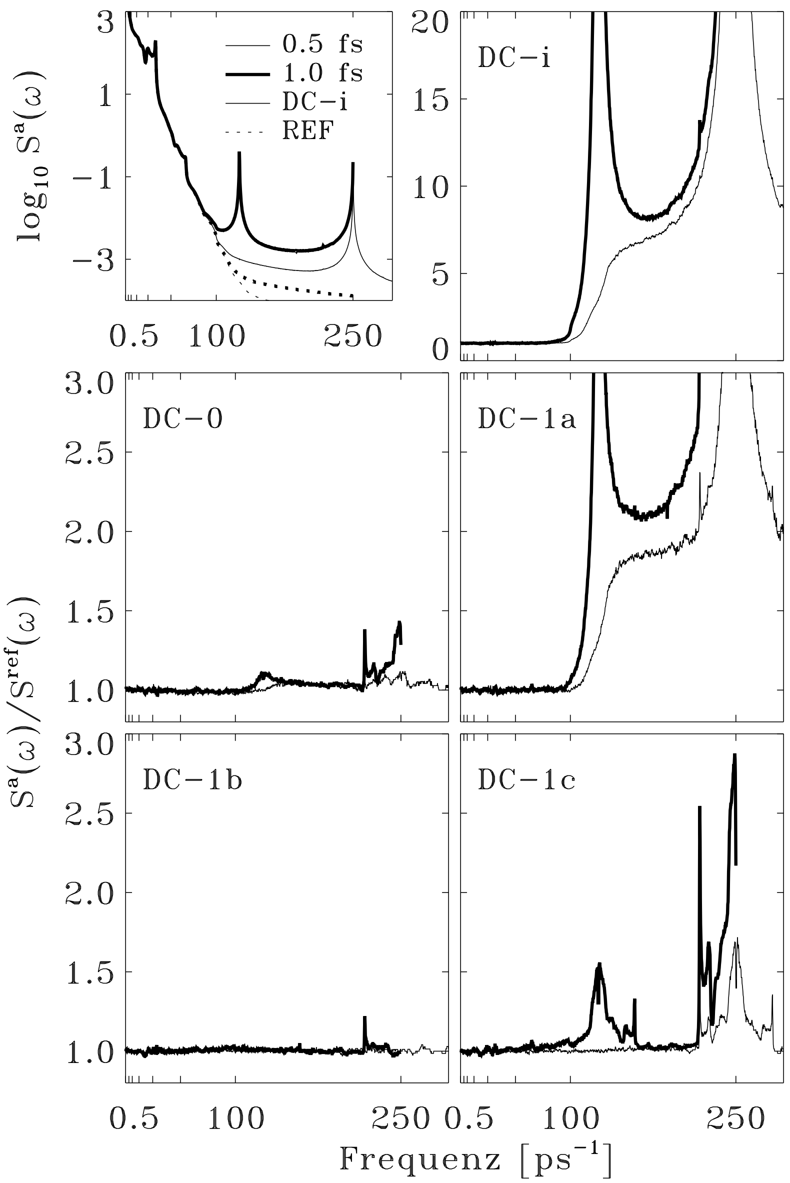

Abbildung 4.6: Hochfrequenzspektren atomarer Bewegungen; *links oben:* Referenzspektrum (gestrichelt) und DC-i-Spektrum (durchgezogene Kurven), in beiden Fällen dargestellt für Integrationsschrittweiten 0,5 fs (dünn gezeichnet) und 1,0 fs (fett gezeichnet); *übrige Bilder:* relative Fehler S^a/S^{ref} der Spektren S^a, die aus den Simulationen $a \in \{\text{DC-i}, \text{DC-0}, \text{DC-1a}, \text{DC-1b}, \text{DC-1c}\}$ mit der Integrationsschrittweiten 0,5 fs (dünn gezeichnet) und 1,0 fs (fett gezeichnet) ermittelt wurden; das Spektrum DC-i (rechts oben) ist mit gestauchter Ordinate wiedergegeben; der Algorithmus CUT zeigte keine signifikante Abweichung von REF (Ergebnisse nicht gezeigt); die relative Abweichung für DC-1d (ebenfalls nicht gezeigt) gleicht der von DC-1c.

se Schwingungen entsprechen den schnellsten *physikalischen* Freiheitsgraden im System. Somit stehen die gezeigten Spektren im Bereich oberhalb dieser Frequenz in keiner direkten Beziehung zu dessen physikalischen Eigenschaften. Da im physikalisch relevanten Bereich unterhalb 30 ps^{-1} keine Abweichungen von den Referenzspektren beobachtet wurden, sind die gezeigten Artefakte, ähnlich der oben diskutierten Drift der Gesamtenergie, nicht von *unmittelbarem* Belang. Noch immer ist unklar, ob das algorithmische Hochfrequenzrauschen Fehler in der Berechnung *relevanter* Observabler verursacht. Die Abschneide-Methode ‘CUT’ zeigte bisher keine Artefakte.

b2) Niederfrequenzspektren atomarer Bewegungen

Die ausgedehnten Testsimulationen gestatteten die Berechnung von Spektren atomarer Bewegungen in einem MD-Simulationen üblicherweise nicht zugänglichen Bereich kleiner Frequenzen bis zu $0,05\,\mathrm{ps}^{-1}$. Das oberen Bild der Abbildung 4.7 zeigt drei derartige Niederfrequenzspektren, wobei jeweils zur Reduktion statistischer Fehler drei Simulationen mit den Integrationsschrittweiten $0,5\,\mathrm{fs}$ $1,0\,\mathrm{fs}$ und $2,0\,\mathrm{fs}$ kombiniert wurden. Das Spektrum der Referenzsimulation ist gestrichelt gezeichnet.

Keiner der sechs untersuchten Entfernungsklassenalgorithmen verursachte hier signifikante Abweichungen vom Referenzspektrum; stellvertretend zeigt die Abbildung das Spektrum DC-1d (durchgezogen, dünn). Im Gegensatz dazu unterdrückt das Vernachlässigen langreichweitiger Kräfte durch CUT atomare Bewegungen im Frequenzbereich unterhalb $0,2\,\mathrm{ps}^{-1}$: Der in Abb. 4.7 unten dargestellte relative Fehler S^a/S^{ref} zeigt eine beträchtliche Absenkung des Spektrums gegenüber der Referenzsimulation (fett gezeichnete Kurve) um bis zu 40%. Eine bemerkenswerte Konsequenz dieser Beobachtung ist, daß offensichtlich die langreichweitigen Kraftbeiträge oberhalb 10 Å, die in CUT vernachlässigt und in REF berücksichtigt werden, *bei gleicher mittlerer Proteinstruktur* deutlichen Einfluß auf die Niederfrequenzdynamik unseres Proteinmodells nehmen.

c) Atomare Beweglichkeiten

Zur Reduktion statistischer Fehler in den betrachteten Niederfrequenzspektren wurde über die Spektren der Bewegung *aller* Atome gemittelt (vgl. 4.1), wodurch zwar eine hohe spektrale Auflösung erreicht, aber gleichzeitig atomspezifische algorithmische Artefakte möglicherweise verdeckt wurden. Diesbezüglich detailliertere Beobachtungsgrößen sind die durch (4.1) definierten atomaren Beweglichkeiten σ_α. Die Zuverlässigkeit der Berechnung der einzelnen Beweglichkeiten wurde anhand der mittleren quadratischen Abweichung Δ_{rms} der σ_α von den entsprechenden Referenzwerten σ_α^{ref}, $\Delta_{\mathrm{rms}} := (\langle(\sigma_\alpha - \sigma_\alpha^{ref})^2\rangle_\alpha)^{1/2}$, beurteilt.

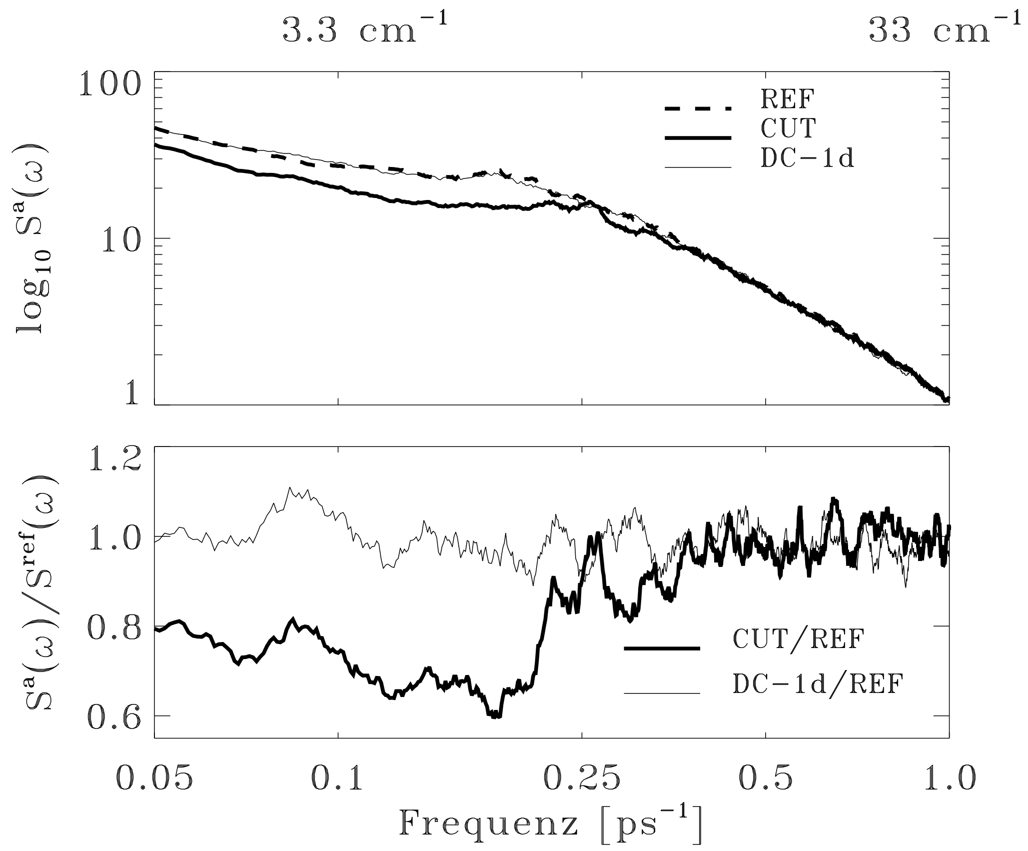

Abbildung 4.7: Niederfrequenzspektren atomarer Bewegungen; *oben:* Spektren der Simulationen REF (gestrichelt), CUT (durchgezogen, fett) und DC-1d (durchgezogen, dünn); *unten:* relative Fehler S^a/S^{ref} der Spektren S^a, für die Algorithmen $a \in \{\mathrm{CUT}, \mathrm{DC\text{-}1d}\}$.

Abbildung 4.8 zeigt die für die sieben MD-Algorithmen gewonnenen Ergebnisse, wobei eine Integrationsschrittweite von 1 fs verwendet wurde. Bei der Interpretation ist zu beachten, daß die gezeigten Abweichungen nicht ausschließlich durch die jeweiligen algorithmischen Näherungen hervorgerufen werden; auch statistische Fluktuationen tragen aufgrund der begrenzten Länge der Testsimulationen zu den beobachteten Δ_{rms} bei. Dieser statistische Beitrag wurde durch den Vergleich zweier unabhängiger Referenzsimulationen abgeschätzt. Der erhaltene Schätzwert ist in der Abbildung links (mit REF bezeichnet) eingetragen und durch die gestrichelte Linie markiert. Lediglich Abweichungen, die deutlich über diesen statistischen Beitrag hinausgehen, lassen auf algorithmisch bedingte Artefakte schließen. Dies ist für die Algorithmen DC-i und DC-0 der Fall; die Abweichungen der anderen Testsimulationen sind nicht signifikant.

Man könnte vermuten, daß die bei DC-i und DC-0 auftretenden großen Rauschkräfte (vgl. Abbildung 2.3) atomare Beweglichkeiten vergrößern sollten. Eine genauere

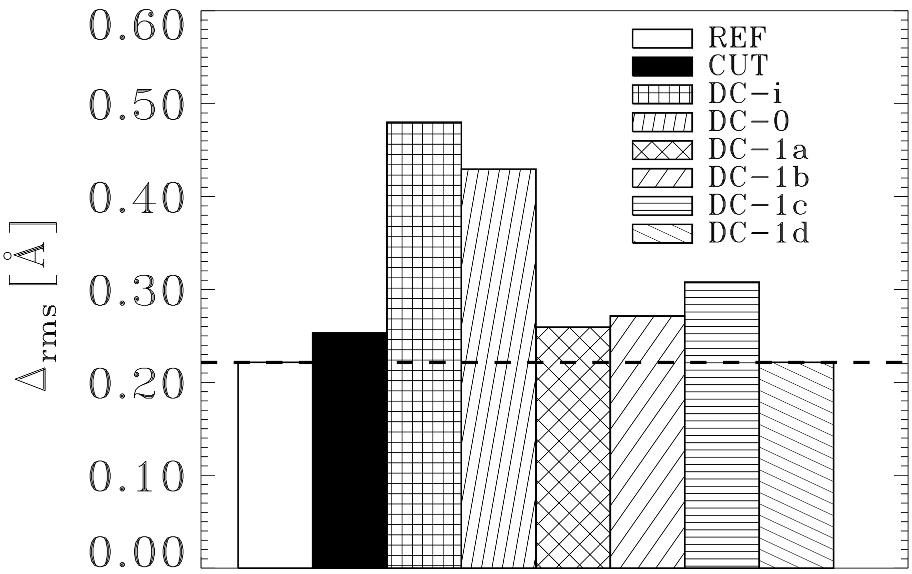

Abbildung 4.8: Mittlere quadratische Abweichung atomarer Beweglichkeiten von den aus der Referenzsimulation bestimmten Beweglichkeiten; dargestellt sind die für die sieben betrachteten Algorithmen beobachteten Abweichungen; der mit REF bezeichnete Wert, der auch durch die gestrichelte Linie markiert ist, wurde aus dem Vergleich zweier Referenzsimulationen gewonnen und dient zur Abschätzung des statistischen Fehlers.

Analyse der Daten (nicht dargestellt) zeigt jedoch, daß diese beiden Methoden im Gegenteil die Beweglichkeiten σ_α *verringern.* Wir nehmen deshalb an, daß diese Algorithmen atomare Bewegungen dekorrelieren, den Trägheitscharakter der Dynamik verringern und die zugehörigen Amplituden entsprechend verkleinern.

d) Korrelationen atomarer Bewegungen

Diese Hypothese läßt sich überprüfen, indem Korrelationen atomarer Bewegungen direkt untersucht werden. Zu diesem Zweck wurden aus den Testsimulationen Orts-Autokorrelationsfunktionen C_α^{dis} gemäß (4.3) abgeleitet. Abbildung 4.9 zeigt Korrelationsfunktionen der sieben betrachteten Algorithmen, wobei über zwanzig Atome des Testsystems gemittelt wurde. Auch diese Observablen unterliegen, gleich den Δ_{rms}, statistischen Schwankungen. Ein Vergleich zweier Referenzsimulationen, beide in der Abbildung als dünne, durchgezogene Linien dargestellt und mit REF bezeichnet, belegt, daß hier nur geringe statistischen Fehler auftreten.

Entsprechend dürfen die durch sechs der sieben untersuchten Algorithmen verursachten Abweichungen als signifikant betrachtet werden. Danach gliedern sich die

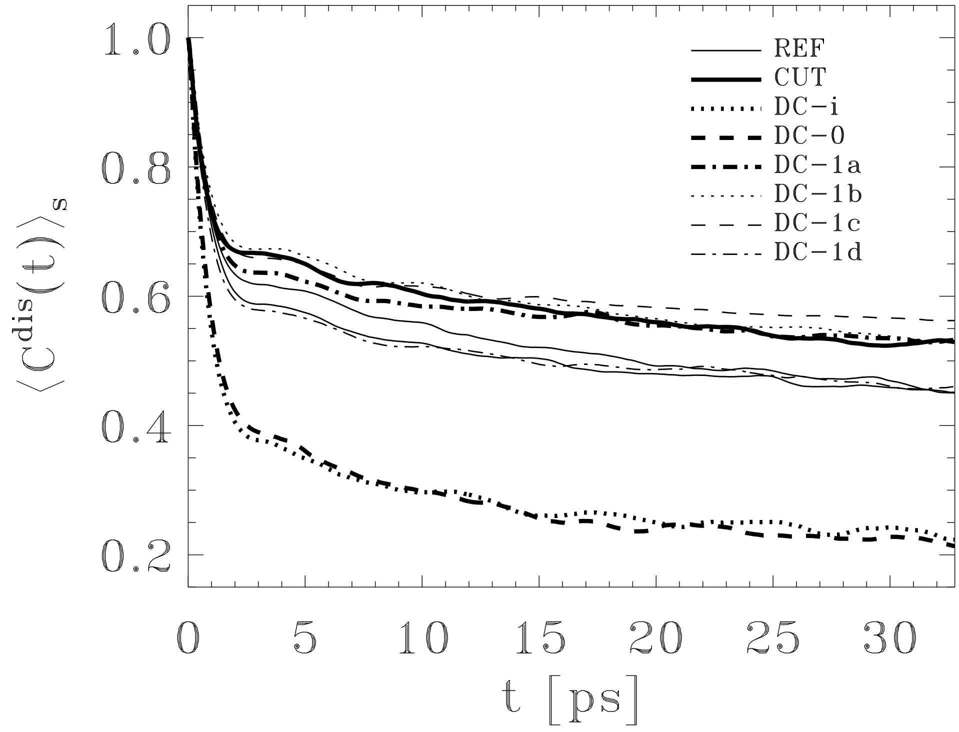

Abbildung 4.9: Gemittelte Autokorrelationsfunktionen $\langle C^{\mathrm{dis}}(t)\rangle_\alpha$ atomarer Positionen; die aus zwei Referenzsimulationen bestimmten Korrelationsfunktionen (dünne, durchgezogene Kurven) dienen der Abschätzung des statistischen Fehlers.

beobachteten und in Abb. 4.9 erkennbaren Abweichungen in drei Gruppen: In Übereinstimmung mit der oben angeführten Hypothese ist zu erkennen, daß die Entfernungsklassenalgorithmen DC-i und DC-0 in der Tat Korrelationen unterdrücken. Dagegen bewirken die Algorithmen DC-1a, DC-1b, DC-1c und CUT offensichtlich einer leichte Anhebung dieser Korrelationen. Die Methode DC-1d zeigt keine signifikante Abweichung.

Geschwindigkeits-Autokorrelationsfunktionen C_α^{vel} wurden entsprechend (4.4) aus den Testsimulationen abgeleitet. Keiner der untersuchten Algorithmen zeigte dabei signifikante Abweichungen von der Referenzsimulation. Die Erklärung für dieses Ergebnis besteht in der engen Beziehung von Autokorrelationsfunktionen zu Spektren atomarer Bewegungen, welche mit Hilfe einer Fourier-Transformation aus C_α^{vel} gewonnen werden können: Die betrachteten Autokorrelationsfunktionen sind im untersuchten Zeitbereich wesentlich durch den physikalisch relevanten Teil des unter *b1)* diskutierten Hochfrequenzspektrums bestimmt, der ebenfalls keine Artefakte aufwies.

Wir halten kurz inne und bemerken, daß unsere Methode DC-1d bei allen bisher aufgeführten Tests keine Artefakte zeigte.

Bisher wurden lediglich die dynamischen Eigenschaften einzelner Atome untersucht, jedoch keine Beziehungen dieser Bewegungen untereinander. Um zu untersuchen, ob *korrelierte* Bewegungen durch die betrachteten Algorithmen korrekt beschrieben werden, bestimmten wir die Kovarianzen $K_{\alpha\beta}$ der Bewegung von 190 Atompaaren (α, β) gemäß (4.5). Diese Kovarianzen sind in Abbildung 4.10 gezeigt: In den acht Bildern repräsentiert jede Raute eines der 190 Atompaare. Die vertikale Position der Rauten mißt die Kovarianz $K_{\alpha\beta}$ des Atompaars, die horizontale Position gibt den mittleren Abstand $r_{\alpha\beta}$ der beiden Atome α und β an. Zusätzlich ist in jedem der acht Bilder als Maß für die Abweichung der 190 Kovarianzwerte von den entsprechenden Referenzwerten der mittlere quadratische Fehler ('rms') als Zahlenwert angegeben. Auch hier wurde der mit den eingetragenen rms-Werten verbundene statistische Fehler durch Vergleich zweier Referenzsimulationen abgeschätzt; er ist im Bild oben links (mit 'REF' bezeichnet) zusammen mit den Kovarianzen einer der beiden Referenzsimulationen eingetragen.

Aus Abb. 4.10 ist ersichtlich, daß in dem verwendeten Proteinmodell die Bewegung nahe beieinander liegender Atome korreliert ist (positive Kovarianzen), während entferntere Atome zu anti-korrelierter Bewegung neigen (negative Kovarianzen). Während dieses Ergebnis qualitativ von allen betrachteten Algorithmen richtig wiedergegeben wird, sind deutliche quantitative Unterschiede erkennbar: Insbesondere für große Atomabstände neigen die Algorithmen DC-i und DC-0, welche beide starke Rauschkräfte aufweisen, dazu, (Anti-)Korrelationen zu unterdrücken. Die unten in den Bildern angegebenen quadratischen Abweichungen von der Referenzsimulation machen darüberhinaus deutlich, daß neben diesen beiden Algorithmen noch zwei weitere, CUT und DC-1c, von den Referenzwerten signifikant abweichende Kovarianzen verursachen. Die Algorithmen DC-1a, DC-1b und DC-1c zeigen dagegen keine über die statistischen Fehler hinausgehenden Abweichungen. Die mit den zuletzt aufgeführten drei Entfernungsklassenalgorithmen berechneten Trajektorien können durch Analyse der Kovarianzen allein also nicht von der Referenzsimulation unterschieden werden. Es wird sich zeigen, daß die in Abschnitt 4.4 e eingeführten projizierten Konfigurationsraumdichten ρ_c ein empfindlicheres Beurteilungskriterium darstellen.

e) Projizierte Konfigurationsraumdichte

Die spezielle Wahl der Konformationskoordinaten $\mathbf{c}$, welche die untersuchten projizierten Konfigurationsraumdichten ρ_c festlegt, zielte darauf ab, die Niederfrequenz-

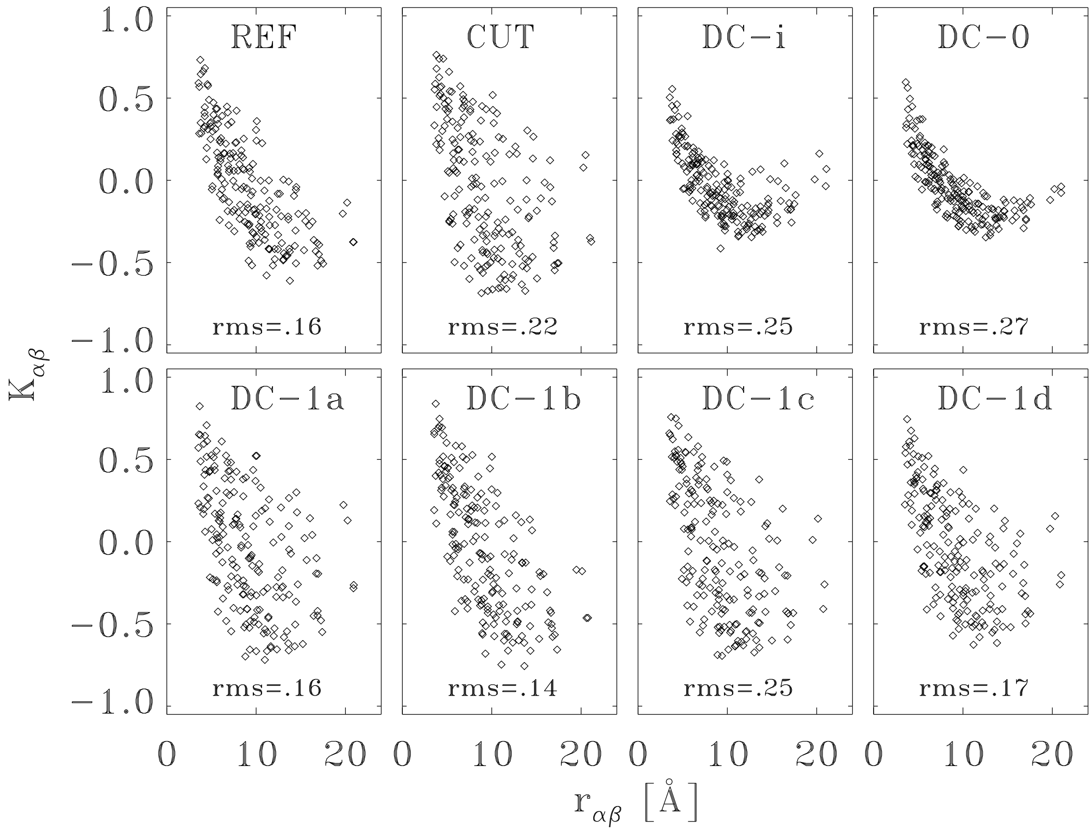

Abbildung 4.10: Kovarianzen atomarer Bewegungen; in den acht Bildern repräsentiert jede der 190 Rauten ein Atompaar $\langle\alpha\beta\rangle$; die vertikale Position der Rauten gibt die aus den Testsimulationen bestimmte Kovarianz $K_{\alpha\beta}$ der Bewegung des jeweiligen Atompaars an, die horizontale Position dessen mittleren Abstand $r_{\alpha\beta}$. In jedem der acht Bilder ist als Maß für die Abweichung der 190 Kovarianzwerte von den entsprechenden Referenzwerten der mittlere quadratische Fehler ('rms') als Zahlenwert angegeben; die statistische Abweichung der Kovarianzwerte zweier Referenzsimulationen voneinander ist im Bild oben links (mit 'REF' bezeichnet) zusammen mit den Kovarianzen einer der beiden Referenzsimulationen eingetragen.

dynamik unseres Proteinmodells zu überprüfen. Entsprechend wurden zur Definition der Konformationskoordinaten solche Atome ausgewählt, die in der Tertiärstruktur der Proteinmodells weit voneinander entfernt sind (vgl. Abb. 3.2). Die zeitliche Entwicklung dieser Konformationskoordinaten sollte deshalb wesentlich durch niederfrequente, kollektive Bewegungen des 'Proteinrückgrats' bestimmt sein Wie bereits die Analyse der Niederfrequenzspektren atomarer Bewegungen in Abschnitt 4.4 b2 nahelegte, werden gerade diese langsamen Freiheitsgrade in besonders hohem Maße von langreichweitigen Coulomb-Anteilen beeinflußt und sollten deshalb besonders geeignet sein, algorithmische Artefakte aufzudecken.

Für alle sieben zu beurteilenden Algorithmen sowie für zwei Referenzsimulationen wurden je $P = 50$ Dichteverteilungen ρ_c bestimmt, wobei nur die mit der Integrationsschrittweite von 1 fs berechneten Trajektorien berücksichtigt wurden. Die Dichteverteilungen wurden mit Hilfe von neun verschiedenen Unterteilungen $\{V^k\}$ der 50 Konfigurationsunterräume verglichen, ausgehend von grober Auflösung ($K = 2$) bis hin zu feiner Auflösung ($K = 512$).

Wie in Abschnitt 4.4 e beschrieben, diente ein Vergleich zweier Referenzsimulationen entsprechend (4.6) und (4.7) der Bestimmung statistischer Fehler. Das obere linke Bild in Abbildung 4.11 zeigt die beobachtete Abweichung $\langle\Delta_{R_2-R_1}\rangle$ (durchgezogene Kurve); der Auflösungsgrad K ist dabei an der Abszisse angegeben. Die mit wachsender Auflösung zunehmende Abweichung $\langle\Delta_{R_2-R_1}\rangle$ spiegelt die mit der Zahl K verwendeter Volumenelemente $\{V^k\}$ zunehmenden statistischen Fluktuationen in der Anzahl der auf jedes Element entfallenden Phasenraumpunkte wider. Die Standardabweichung $\sigma_{R_2-R_1}$ dieser Fehlerabschätzung wurde nach (4.8) bestimmt und ist im Bild als Fehlerintervall um $\langle\Delta_{R_2-R_1}\rangle$ gestrichelt eingezeichnet. Das schmale Fehlerintervall zeigt, daß im vorliegenden Fall, in dem zugleich ein vereinfachtes Proteinmodell und außergewöhnlich lange Testsimulationen eingesetzt wurden, die betrachteten Konfigurationsraumdichten in der Tat reguläre Größen darstellen.

Die übrigen sieben Bilder der Abb. 4.11 zeigen die gemittelten Abweichungen Δ_a der für die sieben Algorithmen a bestimmten projizierten Konfigurationsraumdichten von der Referenzdichte REF. Gemäß (4.9) wurde hier der statistische Anteil $\langle\Delta_{R_2-R_1}\rangle$ von den beobachteten Abweichungen $\langle\Delta_{S_a-R_1}\rangle$ abgezogen. Zwei Fehlerintervalle sind eingezeichnet, die durch (4.10) gegebene untere Grenze (gestrichelt) sowie die durch (4.11) gegebene obere Grenze (Strich-Punkt).

In allen sieben Fällen ist zu erkennen, daß die mit mittlerer Auflösung bestimmten Abweichungen die höchste Signifikanz aufweisen. Dort scheinen die verwendeten Volumenelemente die Konfigurationsraumdichten ρ_c hinreichend detailliert zu beschreiben, um algorithmische Artefakte aufzudecken. Demgegenüber verhindern sowohl die grobe Abbildung von ρ_c bei kleinen K-Werten, wie auch die mit großen K-Werten verbundenen erhöhten statistischen Fehler die Unterscheidung der verschiedenen Algorithmen (vgl. dazu auch die Bemerkung am Ende von Abschnitt 4.4).

Abweichungen Δ_a hoher Signifikanz wurden bei den Algorithmen CUT und DC-1b festgestellt. Die maximalen Abweichungen für DC-i, DC-0, DC-1a und DC-1c sind demgegenüber vier- bis fünfmal kleiner und entsprechen in etwa dem maximal zu erwartenden statistischen Fehler $\sigma_a^{\max}$, also dem siebenfachen Erwartungswert der unteren Fehlerschranke $\sigma_a^{\min}$. Zur Beurteilung dieses Ergebnisses sei angemerkt, daß das durch $\sigma_a^{\max}$ gegebene Fehlerintervall unter der Annahme des ungünstigsten

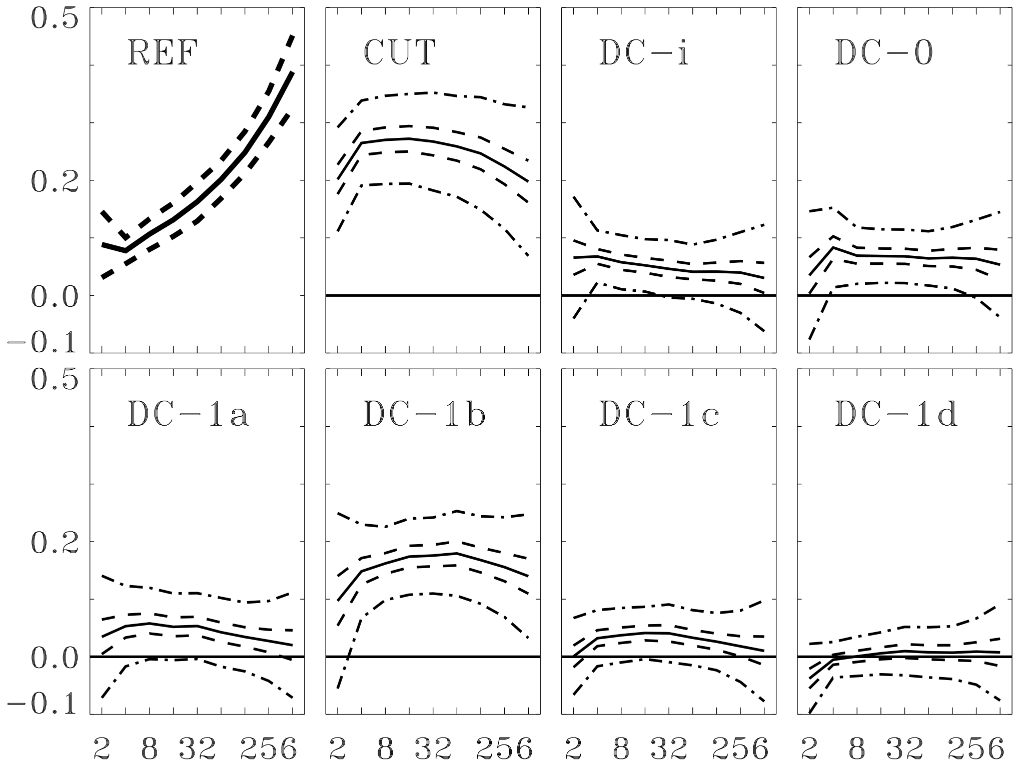

Abbildung 4.11: Algorithmisch bedingte Anteile der Abweichung projizierter Konfigurationsraumdichten von den Referenzwerten und deren statistische Signifikanz; das mit 'REF' bezeichnete Bild oben links zeigt eine Abschätzung des mit den Vergleichen verbundenen statistischen Fehlers; an der Abszisse ist der Auflösungsgrad K, mit dem die dargestellten Abweichungen bestimmt wurde, aufgetragen; zur näheren Erläuterung siehe Text.

Falls *vollständiger* Korrelation aller P Dichteverteilungen abgeleitet wurde. Es darf also davon ausgegangen werden, daß das tatsächlich vorliegende Fehlerintervall σ_a deutlich kleiner als $\sigma_a^{\max}$ ist, und somit dürfen auch die für die Algorithmen DC-i, DC-0, DC-1a und DC-1c beobachteten ρ_c-Abweichungen als statistisch signifikant eingestuft werden. So kann unter der sicher vernünftigen Annahme $\sigma_a < \sigma_a^{\max}/2$ die Nullhypothese *'Es treten keine algorithmisch bedingten Artefakte auf'* mit einem Signifikanzniveau unter 5% verworfen werden.

Wir sind am Ende unserer ausführlichen Beurteilung angelangt und stellen fest, daß der in dieser Arbeit vorgeschlagene Entfernungsklassenalgorithmus DC-1d als einzige effiziente Methode keine Artefakte der Proteindynamik zeigte.

4.6 Diskussion

Tatsächlich beschreibt der DC-1d die Dynamik des Proteinmodells auf allen untersuchten Zeitskalen so genau, daß bei keiner der betrachteten relevanten Observablen signifikante Abweichungen von der 'exakten' Referenzsimulation nachweisbar waren. Demgegenüber traten bei allen anderen untersuchten MD-Methoden Abweichungen in der einen oder anderen Observablen auf. Um deren Diskussion übersichtlicher zu gestalten, faßt Tabelle 4.1 die im vorigen Abschnitt gezeigten Ergebnisse qualitativ zusammen. Die Tabelle zeigt für jeden Algorithmus und für alle untersuchten Observablen den Grad der beobachteten Abweichung von den entsprechenden Referenzwerten. Dabei wird unterschieden zwischen 'keine signifikante Abweichung' (–), 'es trat eine Abweichung auf' (•) und 'starke Abweichung beobachtet' (••). Die durch die verschiedenen Algorithmen verursachten und in der Zusammenfassung aller Observabler erkennbaren 'Artefaktmuster' legen eine Klassifizierung der MD-Methoden in vier Gruppen nahe, die in Tab. 4.1 durch senkrechte Linien angedeutet ist und im folgenden diskutiert werden soll.

Observable	CUT (Gruppe 1)	DC-i (Gruppe 2)	DC-0	DC-1a (Gruppe 3)	DC-1b	DC-1c	DC-1d (Gruppe 4)
ΔE	–	••	••	•	•	•	•
S^{HF}	–	••	•	••	•	••	••
S^{LF}	••	–	–	–	–	–	–
σ_α	–	•	•	–	–	–	–
C_α^{dis}	•	••	••	•	•	•	–
C_α^{vel}	–	–	–	–	–	–	–
$K_{\alpha\beta}$	•	•	•	–	–	•	–
ρ_c	••	•	•	•	••	•	–

Tabelle 4.1: Die Zusammenfassung der bei den untersuchten MD-Methoden beobachteten Artefakte in den Observablen ΔE (Energiedrift), S^{HF} (Hochfrequenzspektren), S^{LF} (Niederfrequenzspektren), σ_α (atomare Beweglichkeiten), C_α^{dis} (Orts-Autokorrelationen), C_α^{vel} (Geschwindigkeits-Autokorrelationen), $K_{\alpha\beta}$ (korrelierte Bewegungen) sowie ρ_c (projizierte Konfigurationsraumdichten) erlaubt eine Beurteilung dieser Algorithmen, die dem Problem der Proteindynamiksimulation angepaßt ist. Die Symbole '–', '•' und '••' bedeuten 'keine signifikante Abweichung', 'es trat eine Abweichung auf' bzw. 'starke Abweichung'. Faßt man alle Observablen zusammen, sind charakteristische 'Artefaktmuster' erkennbar, welche eine Klassifizierung der sieben Algorithmen in vier Gruppen nahelegen (senkrechte Striche).

Die Abschneide-Methode CUT (Gruppe 1) verletzt die Erhaltung der Gesamtenergie nur geringfügig; ebenso wird die molekulare Hochfrequenzdynamik, die durch kurzreichweitige Wechselwirkungen bestimmt ist, richtig wiedergegeben. Überraschend große Fehler treten dagegen in der Beschreibung langsamer Prozesse auf. Hier ist insbesondere die Absenkung des Bewegungsspektrums im Zeitbereich oberhalb einiger Pikosekunden sowie starke Abweichungen der projizierten Konfigurationsraumdichten hervorzuheben. Neben dem bekannten Einfluß auf die Proteinstruktur ist die beobachtete Unterdrückung kollektiver Bewegungen ein bislang unbekanntes Artefakt der Vernachlässigung langreichweitiger Anteile von Coulomb-Kräften. Offensichtlich werden gerade die häufig an die Proteinfunktion gekoppelten langsamen, kollektiven Atombewegungen entscheidend durch langreichweitige Wechselwirkungen bestimmt. Dieses Resultat unterstreicht die besondere Wichtigkeit genauer Beschreibungen der Elektrostatik in Proteinen.

Keiner der untersuchten Entfernungsklassenalgorithmen zeigte eine vergleichbare artifizielle Unterdrückung niederfrequenter Atombewegungen, was darauf hinweist, daß diese Algorithmen die langreichweitigen Coulomb-Kräfte tatsächlich hinreichend genau berücksichtigen, um derartig gravierende Artefakte zu vermeiden. Stattdessen müssen jedoch algorithmisch bedingte Rauschkräfte in Kauf genommen werden, die generell zu einer Verletzung der Energieerhaltung führen. Die damit verbundene Temperaturdrift muß durch Ankopplung des Systems an ein Wärmebad kompensiert werden.

Die durch die Algorithmen DC-i und DC-0 (Gruppe 2) verursachten algorithmischen Rauschkräfte, welche in den Hochfrequenzspektren atomarer Bewegungen zutage treten, verfälschen fast alle untersuchten relevanten Observablen. Inbesondere dekorrelieren die Rauschkräfte atomare Bewegungen. Die von der Dynamik erzeugte Konfigurationsraumdichte wurde dadurch signifikant verändert. Die beiden Algorithmen zeigen eine höhere Energiedrift als alle anderen Algorithmen. DC-i, der die weitaus größten Rauschkräfte aufweist (vgl. Abb. 2.3), verursacht jedoch eine deutlich *kleinere* Energiedrift als DC-0. Da DC-i im Gegensatz zu DC-0 dem Verlet-Kriterium genügt, ist dies als Indiz für die Wirksamkeit dieses Kriteriums anzusehen.

Abzielend auf eine verbesserte Kraftextrapolation werden bei den Entfernungsklassenalgorithmen erster Ordnung (Gruppen 3 und 4) weitere, zeitlich zurückliegende Kräfte zur Abschätzung der nicht exakt berechneten Kräfte verwendet. Erwartungsgemäß wird deshalb die Beschreibung der Proteindynamik generell verbessert.

Der detaillierte Vergleich der betrachteten Observablen erlaubt eine differenzierende

Bewertung: Zunächst fällt auf, daß auch die Algorithmen erster Ordnung erhebliche Rauschkräfte aufweisen. Diese verletzen, im Gegensatz zur Gruppe 2, die Energieerhaltung nur mäßig. Im Falle von DC-1b ist zu vermuten, daß dies durch den relativ glatten Kraftverlauf dieses Algorithmus (vgl. Abb. 2.3) erreicht wurde. Die übrigen drei Methoden der Gruppen 3 und 4 zeigen keinen vergleichbar glatten Kraftverlauf, erfüllen aber stattdessen das Verlet-Kriterium. Die Tatsache, daß auch hier eine nur mäßige Energiedrift auftritt, demonstriert ebenfalls die Wirksamkeit des Verlet-Kriteriums.

Bei der Beurteilung von CUT fiel auf, daß eine genaue Beschreibung irrelevanter Größen nicht notwendig die genaue Beschreibung relevanter Observabler impliziert. Auch in Gruppe 3 ist hierfür ein Beispiel zu finden: Zwar wird der irrelevante Anteil des Hochfrequenzspektrums atomarer Bewegungen von DC-1b recht genau wiedergegeben, dennoch zeigt diese Methode große Abweichungen in der (relevanten) Konfigurationsraumdichte. Die beiden anderen Algorithmen dieser Gruppe verhalten sich diesbezüglich entgegengesetzt: Trotz artifizieller Resonanzspitzen im irrelevanten Teil des Hochfrequenzspektrums wird die Konfigurationsraumdichte wesentlich genauer beschrieben. Auch dieser Befund unterstreicht die Bedeutung des Verlet-Kriteriums.

Die Entwicklung von Methode DC-1d (vgl. Abschnitt 2.3.1) wurde von der Erwartung geleitet, daß die Kombination verschiedener algorithmischer Kriterien eine Verbesserung der Entfernungsklassenmethode ermöglichen sollte. Entsprechend erfüllt DC-1d das Verlet-Kriterium, das Kontinuitätskriterium und das Glattheitskriterium. Wie Tab. 4.1 belegt, hat sich die angesprochene Erwartung in vollem Umfang erfüllt: Tatsächlich zeigt DC-1d als einziger Algorithmus in keiner der betrachteten relevanten Observablen signifikante Artefakte.

Wir bemerken, daß für diesen Fall unsere problemangepaßte Beurteilung nicht nur die hohe Qualität von DC-1d nachgewiesen hat. Sie zeigte gleichzeitig, daß trotz der notwendige Kopplung des Systems an ein Wärmebad eine korrekte Beschreibung molekularer Dynamik möglich war.

'Little did they know how much they didn't know.'

Max Delbrück in **Mind From Matter?**, 1986

Kapitel 5

Konformationsdynamik von Proteinen

Die beschriebenen Testsimulationen an dem in Kapitel 3 vorgestellten Proteinmodell repräsentieren eine Trajektorie von insgesamt 232 Nanosekunden Länge[1]. Wie man bereits in Abbildung 3.5 erkennt, tritt im Verlauf einer derartig langen MD-Simulation an dem in Kapitel 3 vorgeschlagenen vereinfachten Proteinmodell eine Vielzahl von Konformationsübergängen zwischen verschiedenen Konformationssubzuständen auf, die der experimentell beobachteten Konformationsdynamik von Proteinen ähneln. Diese besondere Eigenschaft des Proteinmodells bietet die Chance, dessen Konformationsdynamik anhand der *expliziten* MD-Simulation zu studieren.

Es wurde bereits in der Einleitung angesprochen, daß die atomare Bewegung in Proteinen auf kurzen Zeitskalen Gedächtniseffekte aufweist, die bei effektiven Beschreibungen berücksichtigt werden müssen. Sollen etwa die schnellen Freiheitsgrade eines Lösungsmittels stochastisch behandelt werden, so müssen deren Korrelationen in den verwendeten Rausch- und Reibungskräften enthalten sein. Entsprechend erfordert die Entwicklung effektiver Modelle für die langsamere Konformationsdynamik von Proteinen Kenntnisse über Existenz und Charakter von Gedächtniseffekten auf längeren Zeitskalen. Um die Dynamik des Proteinmodells unter diesem Gesichtspunkt zu untersuchen, sollen im folgenden zwei einfache effektive Modelle dieser Dynamik angegeben und mit der expliziten Langzeitsimulation verglichen

[1] Präzise gesagt handelt es sich, wie im vorigen Kapitel beschrieben, um 232 Trajektorien von je einer Nanosekunde Länge. Wie näher begründet wurde, können diese 232 Trajektorien jedoch aufgrund des chaotischen Charakters der Dynamik als weitgehend statistisch unabhängig angenommen werden, so daß tatsächlich von einer Simulation der Länge 232 ns gesprochen werden darf.

werden. Aufgrund ihrer Konstruktion als rein stochastische Modelle, die in den Abschnitten 5.2 und 5.4 genauer dargestellt wird, sollten beide Modelle *keine* Gedächtniseffekte enthalten, so daß der Vergleich mit der expliziten Simulation Aufschluß über die Existenz von Gedächtniseffekten ergeben kann.

Zur Definition der beiden angesprochenen effektiven Modelle ist es jedoch zunächst notwendig, den anhand der in Abb. 3.5 erkennbaren abrupten strukturellen Änderungen qualitativ eingeführten Begriff des ‘Konformationssubzustands’ genauer einzugrenzen.

5.1 Theorie der Konformationssubzustände

Es soll ein klassisches Vielteilchensystem im mikrokanonischen Ensemble betrachtet werden, das durch die Hamiltonfunktion $\mathcal{H}(\mathbf{q}^N, \mathbf{p}^N)$ definiert ist, wobei $\mathbf{q}^N$ die $3N$ kartesischen Koordinaten der N Teilchen und $\mathbf{p}^N$ deren Impulse bezeichnen. Dieses System kann effektiv beschrieben werden, indem lediglich einige wenige Freiheitsgrade $c_i(\mathbf{q}^N, \mathbf{p}^N)$, die bereits in Abschnitt 4.4 e als Konformationskoordinaten eingeführt wurden, explizit behandelt werden. Eine solche dimensionsreduzierte Beschreibung erfordert die *implizite* Berücksichtigung der übrigen Freiheitsgrade in Form eines Wärmebads. Das so erhaltene effektive Modell ist dann Element eines *kanonischen Ensembles* [157]. Wir betonen, daß in dieser allgemeinen Formulierung das Wärmebad als Menge von *Freiheitsgraden* definiert ist und nicht, wie häufig implizit angenommen wird, als Menge von Teilchen.

Die Fläche freier Energie dieses Untersystems ist ein effektives Potential (‘potential of mean force’ [158]) $W(c_i)$, welches, in Verbindung mit dem Wärmebad, die Dynamik der Konformationskoordinaten $\mathbf{c}$ bestimmt. $W(\mathbf{c})$ kann mit Hilfe der *kanonischen* (projizierten) Phasenraumdichte $\rho_c(\mathbf{c})$ (vgl. Abschnitt 4.4 e),

$$\rho_c(\mathbf{c}) = \int d\mathbf{q}'^N d\mathbf{p}'^N \rho(\mathbf{q}'^N, \mathbf{p}'^N)\delta(\mathbf{c} - \mathbf{c}')\,, \tag{5.1}$$

aus der vom *Gesamtsystem* erzeugten Phasenraumdichte $\rho(\mathbf{q}^N, \mathbf{p}^N)$ abgeleitet werden:

$$W(\mathbf{c}) = -k_b T \ln \rho_c(\mathbf{c})\,. \tag{5.2}$$

Für unser Proteinmodell läßt sich eine Fläche freier Energie, welche die Konformationsdynamik der Proteinmodells bestimmt, durch Wahl geeigneter Konformationskoordinaten veranschaulichen. Abbildung 5.1 zeigt in Höhenliniendarstellung eine solche, aus der kanonischen Phasenraumdichte $\rho_c(\mathbf{c}) = \rho_c(r_{12,36}, r_{12,87})$ nach (5.2) abgeleitete Energielandschaft, wobei hier die beiden Abstände $r_{\alpha\beta}$ zwischen den Bausteinen #12 und #36 bzw. #12 und #87, die bereits in Abschnitt 4.4 e verwendet wurden, als Konformationskoordinaten dienen. Die auf den durch die beiden Konformationskoordinaten $c_1 := r_{12,36}$ und $c_2 := r_{12,87}$ aufgespannten Unterraum projizierte Konfigurationsraumdichte $\rho_c(c_1, c_2)$ wurde unter Verwendung der gesamten verfügbaren 232-ns-Trajektorie bestimmt, indem im Verlauf der Simulation die beiden Abstände alle 8 fs aufgezeichnet wurden. Nachdem der zweidimensionale Unterraum in Quadrate der Länge $0,5$ Å eingeteilt wurde, ergab die Abzählung der auf die einzelnen Quadrate entfallenden Phasenraumpunkte eine Näherung für $\rho_c(\mathbf{c})$. Auch hier stellt die Wahl der Größe der verwendeten Volumenelemente, wie bereits in Abschnitt 4.4 e diskutiert, einen Kompromiß zwischen hoher Auflösung und niedrigen statistischen Schwankungen dar, wobei aufgrund des hier verwendeten wesentlich größeren Ensembles eine deutlich höhere Auflösung erzielt werden konnte. Die (logarithmische) Höhenliniendarstellung in Abb. 5.1 gibt das so gewonnene zwei-dimensionale Histogramm geglättet wieder. Der obere Teil der Abbildung zeigt einen Schnitt durch diese Fläche freier Energie $W(c_1, c_2)$, welcher deren Struktur verdeutlicht. Dazu wurde der Verlauf von W entlang einer willkürlich gewählten Reaktionskoordinate E-B-D-C-B-A (fett gezeichnet), in Einheiten von k_bT (bei $T = 300K$) aufgetragen. Die Kurve wurde mit Hilfe einer kubischen Spline-Funktion geglättet.

Die in Abb. 5.1 dargestellte effektive Energiefläche weist drei Minima auf, die mit 'B', 'C' und 'D' bezeichnet wurden. Die zugehörigen Bereiche niedriger freier Energie im Konformationsraum sind erkennbar durch Energiebarrieren voneinander getrennt. Jeder dieser Bereiche definiert, wie von Frauenfelder vorgeschlagen wurde [34], einen *Konformationssubzustand*. Der in dieser Betrachtungsweise stochastische Charakter der zahlreichen nicht gezeigten Freiheitsgrade des Wärmebads führt zu einer diffusiven Bewegung des Systems in der dargestellten Energielandschaft, entsprechend der Brown'schen Bewegung eines Teilchens in einem Potential. Wegen der vom Wärmebad auf das 'Teilchen' ausgeübten Rauschkräfte überquert das System von Zeit zu Zeit eine der angesprochenen Energiebarrieren. Dies führt zu einer abrupten Änderung der Proteinstruktur, die als Konformationsübergang beobachtet wird. An einer Reihe verschiedener Proteine wurde eine ganze Hierarchie von Konformationssubzuständen beobachtet, welche durch Energiebarrieren unterschiedlichster Höhe getrennt sind, so daß die zugehörigen Konformationsübergänge durch ein breites Spektrum temperaturabhängiger Raten gekennzeichnet sind [33, 52, 159].

Es sei an dieser Stelle erneut (vgl. Abschnitt 1.3) darauf hingewiesen, daß sich die gegebene Definition für Konformationssubzustände von früher gewählten theoreti-

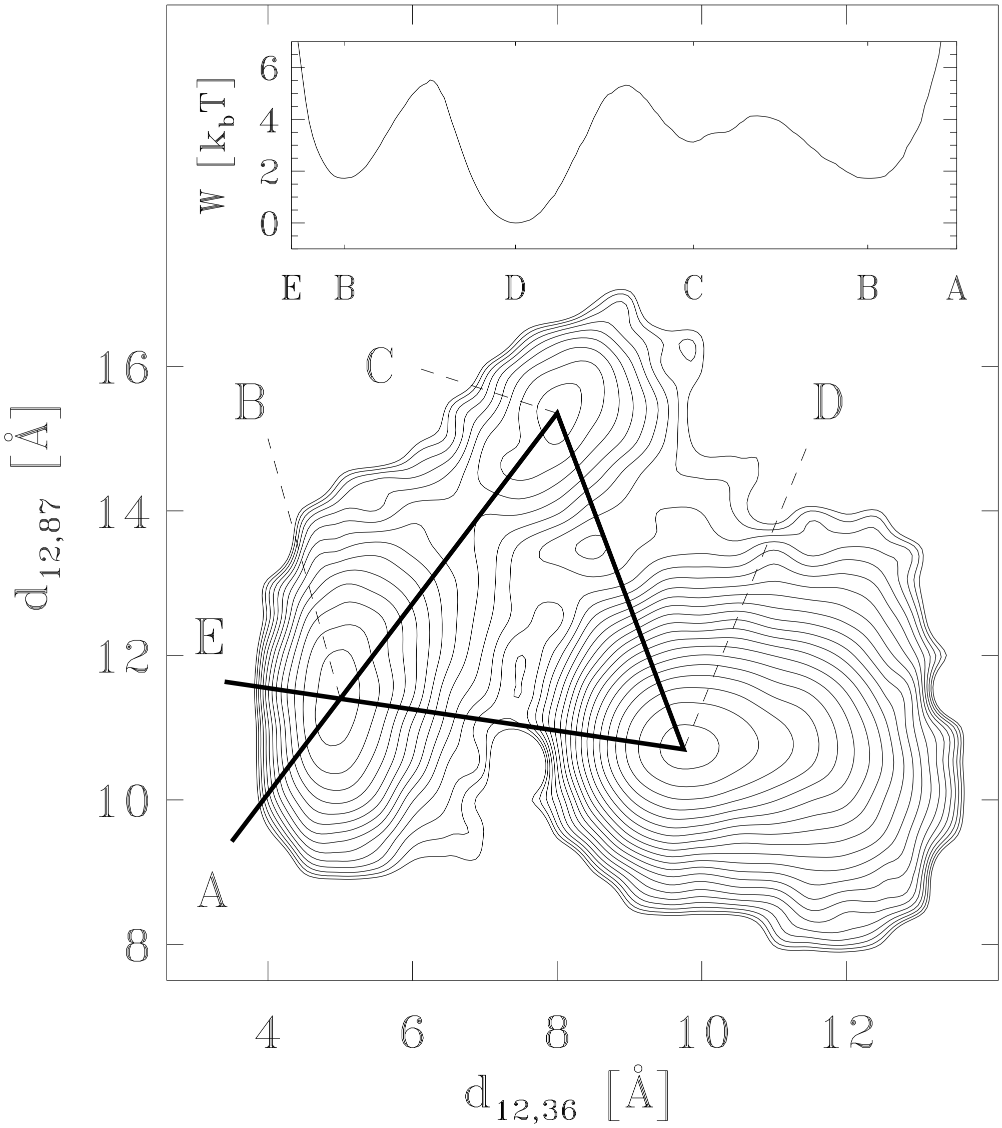

Abbildung 5.1: Höhenliniendarstellung der aus der kanonischen Phasenraumdichte $\rho_c(c_1, c_2)$ abgeleiteten Landschaft freier Energie; als Konformationskoordinaten wurden die beiden Abstände zwischen den Bausteinen #12 und #36 bzw. #12 und #87 (vgl. Abbildung 3.2) gewählt. Die Form dieser effektiven Energiefläche wird durch den im oberen Teil der Abbildung gezeigten Schnitt illustriert; dort ist deren Verlauf entlang einer hypothetischen Reaktionskoordinate E-B-D-C-B-A (in der Höhenliniendarstellung fett eingezeichnet) in Einheiten von k_bT (bei $T = 300K$) aufgetragen.

schen Ansätzen [38, 39] unterscheidet: Da dort die Verteilung thermisch zugänglicher lokaler Minima der *potentiellen Energie* des Systems im Konfigurationsraum untersucht wurde, beruhen derartige Untersuchungen auf der Annahme, daß diese Minima, bzw. Haufen derartiger Minima, Information über die Verteilung der experimentell beobachteten *Bereiche* niedriger *freier* Energie enthalten. Es wurde bereits in der Einleitung bemerkt, daß diese Annahme im allgemeinen lediglich bei tiefen Temperaturen gültig ist, da dabei entropische Beiträge zur freien Energie vernachlässigt werden. Bei physiologischen Temperaturen ist die angenommene Beziehung zwischen zugänglichen Minima der potentiellen Energie und der Lage von Konformationssubzuständen jedoch im allgemeinen nicht gesichert.

Da wir vermittels *expliziter* MD-Simulationen die Fläche *freier* Energie untersuchen und somit entropische Beiträge berücksichtigen, erlaubt unser Ansatz die Analyse der Konformationsdynamik auch bei Zimmertemperatur, ohne auf unsichere Zusatzannahmen angewiesen zu sein. Theorie und Experiment können hier im Prinzip verglichen werden, was andernfalls zumindest problematisch erscheint. Da die Bestimmung von ρ_c außerordentlich lange Simulationen erfordert, können mit unserer Methode allerdings zum gegenwärtigen Zeitpunkt lediglich *vereinfachte* Proteinmodelle untersucht werden.

Wie Abbildung 3.5 ausweist, sind die Konformationssubzustände unseres Proteinmodells auf der Zeitskala von einigen zehn Pikosekunden stabil, Konformationsübergänge treten erst auf einer Zeitskala von einigen hundert Pikosekunden auf. Die beiden angesprochenen effektiven Modelle sollen der Untersuchung von Gedächtniseffekten auf diesen beiden Zeitskalen dienen. Entsprechend soll sich das erste der beiden effektiven Modelle auf die Proteindynamik *innerhalb* eines Konformationssubzustands beziehen und aus ihr abgeleitet werden. Wir wählen ein ein-dimensionales Langevin-Modell, das diese Dynamik als Brown'sche Bewegung eines Teilchens in einem effektiven Potential auffaßt. Wir werden dieses effektive Potential, analog zum Vorgehen bei der Berechnung der in Abb. 5.1 dargestellten Energiefläche $W(c_1, c_2)$, aus unserer expliziten Simulation bestimmen. Für das zweite Modell, das die Konformationsdynamik der Übergänge zwischen den verschiedenen Konformationssubzuständen stochastisch beschreiben soll und auf einer entsprechend längeren Zeitskala angesiedelt ist, wählen wir ein Markov-Modell, in dem die Proteindynamik innerhalb der Konformationssubzustände vernachlässigt wird.

5.2 Ein einfaches Langevin-Modell der Proteindynamik

Wir wollen nun zunächst versuchen, die Dynamik des Proteinmodells in der Umgebung der Energieminimums 'D' (vgl. Abb. 5.1) als Brown'sche Bewegung eines Teilchens entlang einer einzigen Konformationskoordinate zu beschreiben. Als Konformationskoordinate c wählen wir den Abstand $r_{12,36}$, welcher es erlaubt, die Konformationssubzustände 'B' und 'D' zu unterscheiden (vgl. Abb. 5.1). Die Brown'sche

Dynamik in der Umgebung von 'D' bestimmt die Häufigkeit von Konformationsübergängen nach 'B', also die mittlere Verweildauer im Zustand 'D'. Durch Vergleich der zugehörigen Rate die sich aus dem abzuleitenden Langevin-Modell ergibt, mit der expliziten Simulation soll anschließend die Gültigkeit des Langevin-Modells überprüft werden.

Im skizzierten Szenario werden also alle übrigen Freiheitsgrade des Proteinmodells in Form eines Wärmebads berücksichtigt, und unser effektives Modell repräsentiert die diffusive Bewegung eines Teilchens in dem durch (5.2) gegebenen effektiven Potential $W(c)$. Die Bewegung wird durch die Langevin-Gleichung

$$m\ddot{c} = -m\gamma\dot{c} - \nabla W(c) + \xi(t) \tag{5.3}$$

beschrieben, wobei γ einen Reibungskoeffizienten und m eine effektive Masse bezeichnet [160]. Sowohl γ als auch die Rauschkraft $\xi(t)$ modellieren den Einfluß des Wärmebads, also aller implizit behandelten Freiheitsgrade des Systems. Wir nehmen nun an, daß $\xi(t)$ durch Gaußsches (weißes) Rauschen beschrieben werden kann und wollen untersuchen, ob diese Annahme, die das Fehlen von Gedächtniseffekten beinhaltet, gerechtfertigt ist.

Abbildung 5.2 (links) zeigt das effektive Potential $W(c) = W(r_{12,36})$ (dünne Kurve), das aus der 232ns-Trajektorie gemäß (5.2) und (5.1) bestimmt wurde. Dabei wurden lediglich diejenigen Konformationen berücksichtigt, die den Zuständen 'D' oder 'B' zuzuordnen sind; der gering besetzte Zustand 'C' wurde außer Acht gelassen.

Wie zu erwarten, ähnelt der Energieverlauf dem in der linken Hälfte der Abb. 5.1 gezeigten Schnitt; im Gegensatz zur dortigen Energiefläche wurde hier jedoch auf einen *ein*dimensionalen Unterraum c projiziert.

Um die Gültigkeit des Modells zu überprüfen, greifen wir den speziellen Konformationsübergang vom tiefsten Minimum (in Abb. 5.1 mit 'D' bezeichnet) nach links in den Konformationssubzustand 'B' heraus und bestimmen dessen Rate. In Abb. 5.2 (links) entspricht dieser Übergang einem Verlassen des tiefen Potentialminimums über die angrenzende Barriere. Zur Berechnung der Rate muß das Langevin-Modell durch Angabe der effektiven Masse m und des Reibungskoeffizienten γ, die beide als zunächst unbekannte Parameter in Gleichung (5.3) eingehen, spezifiziert werden. Die Amplitude der Rauschkraft $\xi(t)$ ergibt sich dann aus dem Dissipations-Fluktuations-Theorem.

Hierzu bemerken wir, daß der Verlauf des in Abb. 5.2 (links) erkennbare Potentialminimums sowie der Energiebarriere annähernd harmonisch ist. Die gestrichelt eingezeichneten harmonischen Näherungen machen deutlich, daß die Energiefläche

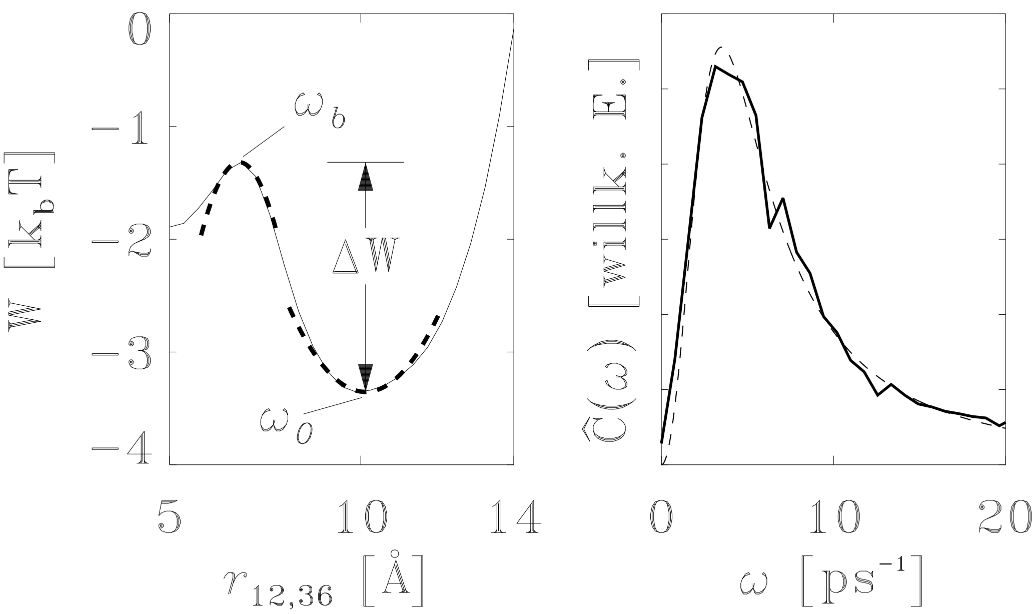

Abbildung 5.2: Linke Seite: effektives Potential $W(c)$ für die Konformationskoordinate c, aufgetragen in Einheiten k_bT (durchgezogene Kurve); die im Text beschriebenen harmonischen Näherungen sind gestrichelt eingezeichnet; rechte Seite: niederfrequenter Teil des Spektrums $\hat{C}(\omega)$ der Geschwindigkeits-Autokorrelationsfunktion $C(t) = \langle \dot{c}(0) \cdot \dot{c}(t) \rangle$, abgeleitet aus einer Trajektorie von einer Nanosekunde Länge (durchgezogene Kurve); aus dem Modell eines harmonischen Oszillators wurde ein entsprechenden Spektrum abgeleitet und an das beobachtete Spektrum angepaßt (gestrichelte Kurve).

durch diese Näherungen tatsächlich sehr gut beschrieben werden kann. Zur Bestimmung von γ verwenden wir einen 1 ns langen Ausschnitt der MD-Trajektorie, in dem das System den Zustand 'D' nicht verläßt, also die in Abb. 5.2 (links) erkennbare Barriere nicht überquert. In diesem Fall 'sieht' das System lediglich den annähernd harmonischen Bereich in der Umgebung des Potentialminimums, und wir können deshalb $W(c) \approx \frac{1}{2} m\omega_0^2 (c - c_0)^2$ annehmen. In dieser harmonischen Näherung ist die Brown'sche Bewegung analytisch beschreibbar, und γ kann durch Vergleich mit der Simulationsrechnung ermittelt werden.

Hierzu verwenden wir die Geschwindigkeits-Autokorrelationsfunktion $C(t) = \langle \dot{c}(0) \cdot \dot{c}(t) \rangle$, die aus dem ausgewählten Ausschnitt der Trajektorie berechnet wurde. Die Geschwindigkeits-Autokorrelationsfunktion wurde einer Orts-Autokorrelationsfunktion vorgezogen, (a) da sie schneller gegen Null strebt, (b) da dieser Abfall empfindlich vom Reibungskoeffizienten abhängt und (c) da sie relativ unempfindlich gegenüber Details des effektiven Potentials ist [95]. Abb. 5.2 (rechts) zeigt den niederfrequenten Teil des Spektrums dieser Autokorrelationsfunktion (durchgezogene Kurve).

Für einen schwach bis mäßig gedämpften ($\omega_0 > \gamma/2$) diffusiven harmonischen Oszillator ist die spektrale Dichte $\hat{C}(\omega)$ der Geschwindigkeits-Autokorrelationsfunktion durch

$$\hat{C}(\omega) \propto \frac{\gamma\omega^2}{(\omega_0^2 - \omega^2)^2 + \gamma^2\omega^2} \tag{5.4}$$

gegeben [160]. Abbildung 5.2 (rechts) zeigt (gestrichelt) die Anpassung dieses Ausdrucks (5.4) an das aus der MD-Simulation bestimmte Spektrum. Die ausgezeichnete Übereinstimmung mit dem beobachteten Spektrum (durchgezogene Kurve) belegt erneut die Anwendbarkeit der harmonischen Näherung. Tatsächlich scheint das Langevin-Modell die Dynamik das Proteinmodells richtig wiederzugeben. Darüberhinaus zeigen die aus der Anpassung erhaltenen Werte $\gamma = 6,0\,\mathrm{ps}^{-1}$ und $\omega_0 = 3,5\,\mathrm{ps}^{-1}$, daß die Annahme mäßiger Reibung ($\omega_0 \gtrsim \gamma/2$) gerechtfertigt war. Im Modell des harmonischen Oszillators entspricht die erhaltene Frequenz ω_0 einer effektiven Masse von $m = 7,4$ atomaren Einheiten.

Der Kramers'schen Ratentheorie [161] folgend, läßt sich für das betrachtete Langevin-Modell eine obere Schranke $k_{D\to B}$ für die Übergangsrate von Zustand 'D' in den Zustand 'B' angeben. Die Voraussetzung hierfür, eine gegenüber ω_0 kleine Übergangsrate, ist wegen der gegenüber k_bT hohen harmonischen Barriere gegeben. Wird die Barriere durch $W(c) \approx -\frac{1}{2}m\omega_b^2(c - c_b)^2$ harmonisch genähert (gestrichelte Kurve in Abb. 5.2 (links), dann liegt für den Übergangszustand wegen $\gamma \approx \omega_b$ der Fall mittelstarker Reibung vor, und $k_{D\to B}$ ist durch den bekannten Ausdruck [161]

$$k_{D\to B} = \frac{\sqrt{\gamma^2/4 + \omega_b^2} - \gamma/2}{\omega_b} \, \frac{\omega_0}{2\pi} \, \exp(-\Delta W/k_bT) \tag{5.5}$$

gegeben, wobei ΔW die Höhe der Barriere ist (vgl. Abb. 5.2 (links)). Aus $\omega_b = 6,5\,\mathrm{ps}^{-1}$ folgt die Rate $k_{D\to B} = 46,5\,\mathrm{ns}^{-1}$. Diese Vorhersage überschätzt jedoch die aus den MD-Simulationen ermittelte Rate bei weitem: Diese ist mit $1,9\,\mathrm{ns}^{-1}$ um annähernd den Faktor 25 kleiner!

5.3 Diskussion

Offensichtlich ist das Langevin-Modell nicht in der Lage, die aus den expliziten MD-Simulationen abgeleiteten Übergangsraten richtig wiederzugeben. Dieses Versagen muß der einzigen nicht begründeten Annahme, die bei der Definition des Modells verwendet wurde, zugeschrieben werden: Es wurde angenommen, daß der Einfluß der vielen nicht explizit beschriebenen Freiheitsgrade des Systems auf die Dynamik der

Konformationskoordinate c durch *gedächtnisfreies*, Gaußsches Rauschen beschreibbar ist. Wir folgern, daß dies nicht zulässig ist, daß also Gedächtniseffekte, bedingt durch vernachlässigte Korrelationen vieler Freiheitsgrade des Systems, die schnelle, im Pikosekundenbereich stattfindende Dynamik innerhalb von Konformationssubzuständen nachhaltig beeinflussen. Diese Gedächtniseffekte verlangsamen die Rate des betrachteten Konformationsübergangs um den beobachteten Faktor.

Die nähere Untersuchung der 232 ns-Trajektorie zeigte, daß die in Abb. 5.2 (links) dargestellte Barriere in der Tat häufig überquert wird, in völliger Übereinstimmung mit dem durch (5.5) vorhergesagten Ergebnis. Der überwiegenden Zahl dieser Ereignisse folgte jedoch *kein* Übergang in den anderen Konformationszustand, d.h., das System 'erinnerte' sich an seine Herkunft und kehrte dorthin zurück. Es sei betont, daß dieses 'Gedächtnis' *nicht* auf den Trägheitscharakter der durch (5.3) definierten Brown'schen Bewegung zurückzuführen ist, denn dieser wird in (5.5) ja vollständig berücksichtigt.

Es ist zu erwarten, daß derartige Übergangsraten mit Hilfe der Methode des 'reaktiven Flusses' [75] genauer abgeschätzt werden können (eine Übersicht findet sich in [162]). Dazu müßte jedoch zunächst ein nahe am Übergangszustand gelegenes Ensemble erzeugt werden. Ob das hierzu erforderlich Absuchen des Phasenraums einen geringeren Aufwand als die hier durchgeführten ausgedehnten MD-Simulationen darstellt, ist an dieser Stelle nicht entscheidbar.

5.4 Ein einfaches Markov-Modell der Konformationsdynamik

Zeigen sich Gedächtniseffekte auch auf längeren Zeitskalen, etwa im Bereich einiger hundert Pikosekunden? In diesem Zeitbereich ist die Dynamik des Systems wesentlich durch Übergänge zwischen den drei in Abbildung 5.1 erkennbaren Konformationssubzuständen bestimmt. Deshalb soll nun, im Gegensatz zum im vorigen Abschnitt untersuchten Langevin-Modell die innerhalb der Konformationssubzustände stattfindende Proteindynamik *nicht* explizit beschrieben und stattdessen lediglich die (diskrete) Konformationsdynamik zwischen den drei beobachteten Zuständen untersucht werden. Um Gedächtniseffekte dieser Konformationsdynamik aufzuspüren, soll überprüft werden, ob die beobachtete statistische Verteilung von Verweildauern in den drei Konformationssubzuständen durch eine gedächtnisfreie stochastische Dynamik beschreibbar ist. Dazu soll die aus der 232 ns-Trajektorie abgeleitete Konformationsdynamik mit dem in Abbildung 5.3 dargestellten kontinuierlichen

Markov-Prozeß verglichen werden. Dieser Markov-Prozeß besitzt drei Zustände, entsprechend den in Abb. 5.1 erkennbaren Konformationssubzuständen mit 'B', 'C' und 'D' bezeichnet, zwischen denen Übergänge (Pfeile) stattfinden können.

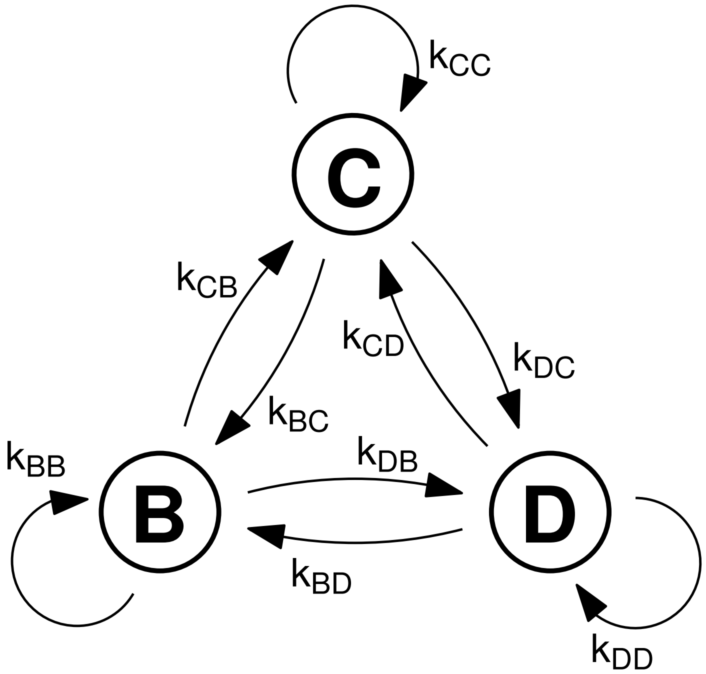

Abbildung 5.3: Das zur Beschreibung der Konformationsdynamik des Proteinmodells verwendete Markov-Modell besitzt drei Zustände $S \in \{B, C, D\}$. Diese Zustände entsprechen den in der Abbildung 5.1 erkennbaren Konformationssubzuständen und sind durch Konformationsübergänge (Pfeile) verbunden. Die neun zugehörigen Übergangsraten k_{ij} definieren nach (5.6) die gedächtnisfreie, stochastische Dynamik dieses einfachen Modells.

Im diesem Szenario gehorchen die Wahrscheinlichkeiten $P_i(t)$, das Proteinmodell zur Zeit t im Zustand $S_t = i$ vorzufinden, der Mastergleichung

$$\frac{d}{dt}P_i(t) = \sum_{j=B,C,D} k_{ij}P_j(t) - \sum_{j=B,C,D} k_{ji}P_i(t) \quad , \quad i = B, C, D\,. \tag{5.6}$$

Die zeitunabhängigen Übergangsraten k_{ij} $(i, j \in \{B, C, D\})$ sind dabei durch die bedingten Wahrscheinlichkeiten gegeben, das System zu einem beliebigen Zeitpunkt $t = 0$ im Zustand i zu finden, falls es zuvor, zur Zeit $t = -\Delta$, im Zustand j war:

$$k_{ij} = \lim_{\Delta \to 0} \frac{1}{\Delta} P(S_0 = i | S_{-\Delta} = j) \quad , \quad i \neq j\,, \quad k_{ii} = 1 - \sum_{j \neq i} k_{ji}\,.$$

Um entscheiden zu können, ob die beobachtete Konformationsdynamik Markov-Charakter besitzt, muß überprüft werden (vgl. [158]), ob die aus der expliziten MD-Simulation abgeleiteten Übergangswahrscheinlichkeiten von der Vorgeschichte des Systems abhängen, d.h., ob die bedingten Wahrscheinlichkeiten

$$P_{ij}(\Delta_1) \;:=\; P(S_0 = i | S_{-\Delta_1} = j)$$

$$\begin{aligned}
P_{ij_1j_2}(\Delta_1,\Delta_2) &:= P(S_0\!=\!i|S_{-\Delta_1}\!=\!j_1 \wedge S_{-\Delta_1-\Delta_2}\!=\!j_2)\\
P_{ij_1j_2j_3}(\Delta_1,\Delta_2,\Delta_3) &:= P(S_0\!=\!i|S_{-\Delta_1}\!=\!j_1 \wedge S_{-\Delta_1-\Delta_2}\!=\!j_2 \wedge S_{-\Delta_1-\Delta_2-\Delta_3}\!=\!j_3)\\
&\vdots \qquad\qquad \vdots
\end{aligned} \tag{5.7}$$

unabhängig vom gewählten zeitlichen Abstand Δ_k, $k = 0, 1, 2, \ldots$, sind. Freilich können aufgrund der begrenzten Menge verfügbarer Daten (insgesamt fanden im Verlauf der 232 ns-Simulationen 238 Konformationsübergänge statt) nicht alle Bedingungen in (5.7) überprüft werden. Stattdessen sollen lediglich die bedingten Wahrscheinlichkeiten erster und zweiter Ordnung, $P_{ij}(\Delta_1)$ bzw. $P_{ij_1j_2}(\Delta_1, \Delta_2)$, miteinander verglichen werden. Weiterhin sollen nur Wahrscheinlichkeiten $P_{ij_1j_2}(\Delta) := P_{ij_1j_2}(\Delta, \Delta)$, ermittelt aus zeitlich äquidistant getrennten Zuständen des Systems, betrachtet werden. Indem wir diesen geringfügig milderen Test anwenden, nehmen wir an, das keine 'teilweise Amnesie' auftritt; die Analyse ist also nicht gültig für Systeme, die Gedächtniseffekte auf *langen* Zeitskalen zeigen, jedoch gleichzeitig *kein* Kurzzeitgedächtnis besitzen. Unsere Analyse wird jedoch gestützt durch die Ergebnisse des vorangegangenen Abschnitts, in dem gezeigt wurde, daß unser System tatsächlich ein 'Kurzzeitgedächtnis' aufweist.

Die benötigten bedingten Wahrscheinlichkeiten erster und zweiter Ordnung, $P_{ij}(\Delta)$ und $P_{ij_1j_2}(\Delta)$, wurden aus der Zahl der in der 232-ns-Trajektorie auftretenden Konformationsübergänge gemäß

$$\begin{aligned}
P_{ij}(\Delta) &\approx \frac{N_{ij}(\Delta)}{N_j(\Delta)}\\
P_{ij_1j_2}(\Delta) &\approx \frac{N_{ij_1j_2}(\Delta)}{N_{j_1j_2}(\Delta)}
\end{aligned}$$

abgeschätzt. Die Anzahlen N_j, N_{ij} und $N_{ij_1j_2}$ sind dabei in Analogie zu den Wahrscheinlichkeiten in (5.7) definiert, wobei $i, j, j_1, j_2 \in \{B, C, D\}$ gilt. Entsprechend bezeichnet $N_j(\Delta)$ die Zahl der Zeitpunkte während der Simulation, zu denen sich das System im Zustand j befand; $N_{ij}(\Delta)$ bezeichnet die Zahl der Zeitpunkte, zu denen es sich im Zustand i befand, *falls* es eine gewisse Zeitspanne Δ zuvor in Zustand j war; und schließlich zählt $N_{ij_1j_2}$, wie oft das System in Zustand i war, nachdem es zuvor in Zustand j_1 war *und*, wiederum davor, in Zustand j_2:

$$\begin{aligned}
N_j &= \sum_k \delta_{j,S(k\Delta')}\,,\\
N_{ij} &= \sum_k \delta_{i,S(k\Delta'+\Delta)}\delta_{j,S(k\Delta')}\,,\\
N_{ij_1j_2} &= \sum_k \delta_{i,S(k\Delta'+2\Delta)}\delta_{j_1,S(k\Delta'+\Delta)}\delta_{j_2,S(k\Delta')}\,,
\end{aligned}$$

wobei das Kronecker-Symbol $\delta_{i,S(t)}$ Eins ist, falls das Proteinmodell zur Zeit t im Zustand i ist, andernfalls Null. Die Zählung wurde mit der Abtastrate $\Delta' = 0,5\,\mathrm{ps}$ durchgeführt.

Falls sich entsprechende bedingte Wahrscheinlichkeiten erster und zweiter Ordnung gleichen, hätte die beobachtete Konformationsdynamik tatsächlich Markov-Charakter, d.h., es wäre kein 'Gedächtnis' beobachtbar. Um die Hypothese gedächtnisfreier Dynamik zu testen, wurde die zu erwartende mittlere statistische Abweichung der Wahrscheinlichkeitsschätzungen $P_{ij_1j_2}$ von den nach (5.7) zugehörigen Werten P_{ij} bestimmt. Die entsprechende Standardabweichung $\sigma_{ij_1j_2}$ ergibt sich aus den beobachteten Übergangshäufigkeiten gemäß

$$\sigma_{ij_1j_2}(\Delta) \approx \sqrt{P_{ij_1j_2}(\Delta)(1 - P_{ij_1j_2}(\Delta))/N_{j_1j2}(\Delta)}\,, \tag{5.8}$$

wobei die auftretenden Binomialverteilungen durch Gaußfunktionen genähert wurden [163].

Abbildung 5.4 zeigt eine Auswahl von sechs typischen Vertretern der insgesamt 27 untersuchten Verteilungsfunktionen $P_{ij_1j_2}(\Delta)$; $i, j_1, j_2 \in \{B, C, D\}$. Dargestellt sind die geschätzten Wahrscheinlichkeiten zweiter Ordnung P_{DDD}, P_{DDB}, P_{DDC}, P_{BCD}, P_{BDD} und P_{CCD} (fette, durchgezogene Kurven) welche für Zeitintervalle im Bereich $\Delta = 50 \ldots 500\,\mathrm{ps}$ bestimmt wurden. Diese sind zu vergleichen mit den fett gestrichelt eingezeichneten zugehörigen Wahrscheinlichkeiten erster Ordnung, $P_{ij}(\Delta)$. Die nach (5.8) bestimmte statistische Abweichung $\sigma_{ij_1j_2}(\Delta)$ ist als 1σ-Intervall (dünne, durchgezogene Kurven) und als 2σ-Intervall (dünn, gestrichelt) eingezeichnet.

5.5 Diskussion

Als wesentliches Ergebnis ist in den sechs Bildern der Abb. 5.4 zu erkennen, daß die Abweichungen der bedingten Wahrscheinlichkeiten zweiter Ordnung von denen erster Ordnung in keinem Fall das 2σ-Intervall verlassen. Dies gilt auch für die übrigen 21, nicht in der Abbildung gezeigten Wahrscheinlichkeitsverteilungen. Die beobachteten Konformationsübergänge des Proteinmodells sind also auf Signifikanzniveaus unter 5% mit der Hypothese verträglich, daß die Konformationsdynamik tatsächlich Markov-Charakter besitzt.

Eine genauere Betrachtung zeigt, daß sich die drei in der oberen Hälfte der Abbildung gezeigten Wahrscheinlichkeiten zweiter Ordnung, P_{DDD}, P_{DDB} und P_{DDC}, welche der gemeinsamen bedingten Wahrscheinlichkeit erster Ordnung P_{DD} zuzuordnen sind, in ihren jeweiligen Fehlerintervallen beträchtlich unterscheiden. Dies liegt

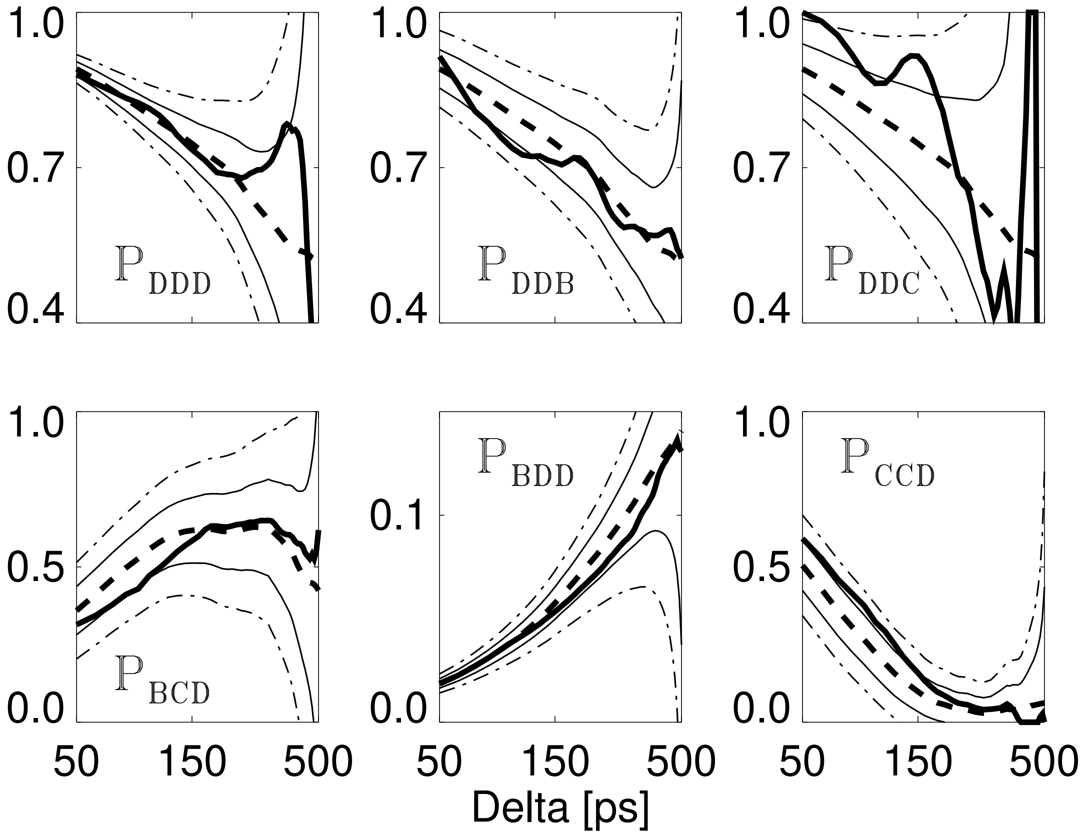

Abbildung 5.4: Sechs ausgewählte bedingte Wahrscheinlichkeiten zweiter Ordnung $P_{ij_1j_2}$ (fette, durchgezogene Kurven) werden, wie im Text genauer ausgeführt, mit den zugeordneten bedingten Wahrscheinlichkeiten erster Ordnung P_{ij} (fett, gestrichelt) verglichen. Die aufgrund des begrenzten Datenmaterials zu erwartende statistische Abweichung ist als 1σ-Intervall (dünne, durchgezogene Kurven) wie auch als 2σ-Intervall (dünn, gestrichelt) eingezeichnet.

in den unterschiedlichen Aufenthaltswahrscheinlichkeiten des Modells in den drei Zuständen begründet: Während sich das Modell relativ selten im Konformationssubzustand ‘C’ befindet (vgl. hierzu auch Abbildung 5.1) und die daraus resultierenden großen statistischen Fehler bei der Bestimmung von P_{DDC} oder P_{BCD} nur eine geringe Signifikanz zulassen, ermöglichen die schmalen Fehlerschranken im Falle von P_{DDD} und P_{BDD}, die von der großen Zahl beobachteter Konformationsübergänge der entsprechenden, dicht bevölkerten Zustände herrühren, einen empfindlichen Test zur Aufdeckung von Gedächtniseffekten. Die größte Abweichung einer bedingten Wahrscheinlichkeit zweiter Ordnung von der zugehörigen Wahrscheinlichkeit erster Ordnung ($\approx 2\sigma$) wurde im Fall von P_{DDC} auf Zeitskalen unterhalb 60 ps beobachtet. Man könnte vermuten, daß diese spezielle Abweichung tatsächlich einen Gedächtniseffekt anzeigt, der durch einen möglichen Trägheitscharakter der schnellen Zerfallsdynamik des Konformationszustands C nach D begründbar wäre.

Sieht man von diesem Einzelfall ab, zeigt die durchgeführte Analyse der Verteilung von Verweildauern in den drei Konformationssubzuständen des Proteinmodells, daß die Konformationsdynamik dieses vereinfachten Modells in der Tat sehr gut durch ein stochastisches Markov-Modell beschreibbar ist. Dies legt den Schluß nahe, daß auf der entsprechenden Zeitskala von einigen hundert Pikosekunden keine Gedächtniseffekte mehr auftreten.

„Before I came here I was confused about this subject.
Having listened to your lecture, I am still confused,
but on a higher level.“

Enrico Fermi, 1938

Kapitel 6

Zusammenfassung, Diskussion und Ausblick

Die in der vorliegenden Arbeit durchgeführten Untersuchungen der Proteindynamik auf langen Zeitskalen dienten dem Ziel, zur Entwicklung und Verfeinerung von Beschreibungen dieser komplexen Dynamik beizutragen. Ein geeignet vereinfachtes, aber gleichzeitig hinreichend realistisches Proteinmodell ermöglichte dabei die Charakterisierung dynamischer Prozesse auf Zeitskalen, welche die mit Hilfe üblicher, detaillierter Proteinmodelle beschreibbaren Zeitspannen um mehr als zwei Größenordnungen übertreffen. Besonderes Augenmerk galt dabei der zum gegenwärtigen Zeitpunkt theoretisch kaum untersuchten Konformationsdynamik, die oberhalb des Nanosekundenbereichs stattfindet und häufig direkt mit der Proteinfunktion verbunden ist, wie etwa bei den in der Einleitung angesprochenen G-Proteinen oder Ionenkanälen.

Ein erster Beitrag bestand in der Weiterentwicklung und Verfeinerung von Entfernungsklassenalgorithmen, die zur Familie der Vielschrittverfahren zu rechnen sind und effiziente Molekulardynamiksimulationen erlauben. Diese Algorithmen nützen proteinspezifische Eigenschaften interatomarer Kraftfluktuationen aus und reduzieren so den mit Simulationen der molekularen Proteindynamik verbundenen Rechenaufwand. In der Diplomarbeit des Autors wurden bereits verschiedene Entfernungsklassenmethoden entwickelt; daran anknüpfend wurde in der vorliegenden Arbeit eine Reihe von Forderungen, welche auf eine Minimierung der durch die verwendeten Näherungen langreichweitiger Kräfte hervorgerufenen Artefakte abzielten, kombiniert und zur Konstruktion der Entfernungsklassenmethode DC-1d verwendet. Diese Methode ist im parallelen Molekulardynamikprogramm EGO [57] implementiert und wurde bereits erfolgreich zur Untersuchung makromolekularer Systeme eingesetzt [111].

Um die Qualität der vorgestellten Verfahren zur Simulation der Proteindynamik zu ermitteln und so die durchgeführten algorithmischen Entwicklungen auf eine solide Grundlage zu stellen, wurden in einem zweiten Beitrag *problemangepaßte* Kriterien algorithmischer Genauigkeit erarbeitet und eingesetzt. Es wurde festgestellt, daß übliche Beurteilungskriterien algorithmischer Genauigkeit für das Verständnis von Proteinfunktion irrelevante Größen betrachten und daher in der Regel wenig darüber aussagen, inwiefern sich eine bestimmte MD-Methode zur Beschreibung der Proteindynamik eignet. Nach genauerer Betrachtung der MD-Simulationen zugrundeliegenden Prinzipien stellten wir fest, daß die benötigten Eignungskriterien vielmehr auf funktionsrelevante und gleichzeitig reguläre, also nicht-chaotischen Größen achten sollten. MD-Methoden und physikalische Modelle müssen dabei *getrennt* betrachtet werden; entsprechend ist der Vergleich von Simulationsergebnissen mit experimentellen Beobachtungsgrößen *nicht* zur Beurteilung einer MD-Methode geeignet. Stattdessen sollten Größen, die mit Hilfe des zu beurteilenden Verfahrens bestimmt wurden, mit hinreichend genauen *Simulationsergebnissen* als Referenzgrößen verglichen werden. Das vorgeschlagene problemangepaßte Verfahren wurde zur Beurteilung der beschriebenen Entfernungsklassenalgorithmen eingesetzt. Zusätzlich wurde die konventionelle Methode des Abschneidens langreichweitiger Kraftbeiträge untersucht.

Reguläre Größen bestehen im allgemeinen aus Ensemble- oder Zeit-Mittelwerten und konvergieren deshalb langsam. Um diese Größen hinreichend genau bestimmen zu können, waren ausgedehnte Testsimulationen nötig, und es mußte ein vereinfachtes Proteinmodell als Testsystem verwendet werden. Als *relevante* Größen wählten wir eine Gruppe von Observablen, die sich unserer Meinung nach zur Erklärung von Proteinfunktion eignen. Diese umfassen atomare Beweglichkeiten, Schwingungsspektren, verschiedene Korrelationsfunktionen der Bewegung sowie Projektionen der durch die molekulare Dynamik erzeugten Phasenraumdichte. Als Beispiel eines häufig eingesetzten Kriteriums algorithmischer Genauigkeit betrachteten wir darüberhinaus die Erhaltung der Gesamtenergie, die jedoch unserer Meinung nach eine für die Proteindynamik irrelevante Größe ist.

Die durchgeführten umfangreichen Vergleichssimulationen demonstrierten eine deutliche Überlegenheit unserer Entfernungsklassenalgorithmen gegenüber der Methode der Abschneidefunktionen. Insbesondere konnte gezeigt werden, daß die Vernachlässigung langreichweitiger Anteile der Coulomb-Wechselwirkung niederfrequente Atombewegungen im Bereich von $0,1\,\mathrm{ps}^{-1}$ um bis zu 40% unterdrückt — ein bislang unbekanntes Artefakt. Keine der untersuchten Entfernungsklassenmethoden verfälschte die Proteindynamik in vergleichbarer Weise.

Entfernungsklassenmethoden approximieren langreichweitige Kraftbeiträge und verursachen daher ein 'algorithmisches Rauschen', das die Energieerhaltung verletzt. Im Gegensatz zur aufwendigen exakten Kraftberechnung tritt eine Energiedrift auf, die durch schwache Kopplung des Systems an ein Wärmebad kompensiert werden muß. Die durchgeführten Testsimulationen zeigten, daß die in der vorliegenden Arbeit entwickelte Methode DC-1d von allen untersuchten Algorithmen als einzige keine relevanten Artefakte verursacht: Weder die algorithmischen Rauschkräfte noch die Kopplung an ein Wärmebad führte zu einer Verfälschung der Proteindynamik. Da Entfernungsklassenmethoden im allgemeinen ähnlich effizient wie Abschneide-Methoden sind, sollte die Methode DC-1d zur Proteindynamiksimulation in jedem Fall vorgezogen werden.

Projizierte Konfigurationsraumdichten erwiesen sich als besonders empfindlicher Indikator algorithmischer Genauigkeit. Aufgrund ihres direkten Bezugs zu thermodynamischen Observablen wie Entropie, freie Energie oder Reaktionsraten halten wir derartige Dichten im Hinblick auf ein Verständnis von Proteinfunktion für besonders relevant; MD-Methoden sollten deshalb solche Dichten korrekt beschreiben. Unsere Methode DC-1d zeigte als einzige bezüglich der untersuchten Konfigurationsraumdichten keine Abweichungen von der 'exakten' Simulation.

Unsere Untersuchungen bestätigten den eingangs geäußerte Verdacht, übliche Kriterien algorithmischer Genauigkeit seien im allgemeinen nicht in der Lage, über algorithmische Eignung für Proteindynamiksimulationen zu befinden. Kriterien wie die Erhaltung der Gesamtenergie, welche auf anderen Gebieten wertvolle Dienste leisten, sollten deshalb hier nicht herangezogen werden.

Multipolalgorithmen zur effizienten Berechnung der Coulomb-Wechselwirkung wurden nicht untersucht. Wir beabsichtigen jedoch, das entworfene Beurteilungsverfahren auch auf diese Algorithmenklasse anzuwenden, insbesondere auf eine kürzlich vorgeschlagene strukturadaptive Multipolmethode, die, ähnlich den Entfernungsklassenalgorithmen, proteinspezifische Merkmale ausnützt. Im Hinblick auf die Weiterentwicklung effizienter MD-Algorithmen sei angemerkt, daß die erreichte Effizienz der beiden Verfahren, Entfernungsklassenalgorithmen und strukturadaptive Multipolmethoden, auf voneinander unabhängigen Proteineigenschaften beruht, so daß deren Kombination eine weitere Effizienzsteigerung um etwa eine Größenordnung erwarten läßt.

Um hinreichend ausgedehnte Testsimulationen durchführen zu können, wurde ein vereinfachtes Proteinmodell entworfen. Dieses 'Minimalmodell' zeigte eine überraschend komplexe Dynamik auf einer Hierarchie von Zeitskalen, die über einen Bereich von annähernd fünf Größenordnungen, von zehn Femtosekunden bis zu Bruchteilen einer Mikrosekunde, untersucht werden konnte. Eine bemerkenswerte Eigenschaft des Proteinmodells ist das Auftreten gelegentlicher Konformationsübergänge zwischen beobachtbaren Konformationssubzuständen. Diese langsame Konformationsdynamik ist ein noch wenig theoretisch untersuchter Bereich der Proteindynamik. Der diesbezüglich erstaunliche Realitätsbezug des verwendeten Modells sowie die Länge der verfügbaren Dynamiksimulation von 232 Nanosekunden Länge erlaubten erstmals deren direkte Untersuchung. Unser Interesse galt dabei Gedächtniseffekten der Proteindynamik, welche auf kurzen Zeitskalen bekannt sind, deren Existenz auf den langen Zeitskalen der Konformationsdynamik jedoch ungeklärt ist. Um das Proteinmodell unter diesem Aspekt zu untersuchen, wurden für dessen Dynamik zwei effektive Beschreibungen angegeben. In den beiden effektiven Modellen wurden Gedächtniseffekte vernachlässigt, so daß ein Vergleich mit der expliziten Molekulardynamiksimulation Aussagen über das Auftreten von Gedächtniseffekten auf unterschiedlichen Zeitskalen der Proteindynamik gestattete.

Die Konstruktion dieser Modelle erforderte eine genaue Definition des Konzepts der Konformationssubzustände, die aus der freien Energie eines durch geeignete Phasenraumprojektion gewonnenen dimensionsreduzierten Systems abgeleitet wurde. Dieses Konzept unterscheidet sich von üblicherweise zum theoretischen Studium der Struktur hierarchischer Konformationszustände herangezogenen Verfahren; es erlaubt insbesondere, die Konformationsdynamik von Proteinen auch bei endlichen Temperaturen zu untersuchen, ohne auf Zusatzannahmen angewiesen zu sein.

Das erste effektive Modell, ein gedächtnisfreies Langevin-Modell, zielte auf die Beschreibung der Proteindynamik im Pikosekundenbereich ab. Es wurde gezeigt, daß dieses einfache Modell die in der Molekulardynamiksimulation beobachteten Konformationsübergangsraten *nicht* korrekt vorhersagen kann. Dieses Ergebnis bestätigte die für komplexe Proteinmodelle bereits bekannte Tatsache, daß auf dieser Zeitskala Gedächtniseffekte, die aus Korrelationen vieler Freiheitsgrade entstehen, die Proteindynamik wesentlich beeinflussen und deshalb nicht vernachlässigt werden dürfen. Auch in dieser Hinsicht verhält sich also unser vereinfachtes Proteinmodell realistisch.

Das untersuchte Langevin-Modell ist der Prototyp einer Klasse stochastischer Modelle, welche irrelevante Freiheitsgrade *implizit* berücksichtigen und das System auf diese Weise effektiv beschreiben. Vielversprechende Vertreter dieser Klasse sind verallgemeinerte Langevin-Modelle mit Gedächtniskernen, welche einen Teil der in der Analyse des einfachen Langevin-Modells aufgezeigten Gedächtniseffekte berücksichtigen. Aufgrund seiner komplexen und realistischen dynamischen Eigenschaften sollte sich das vorgestellte Proteinmodell als Modellsystem zur Untersuchung und Weiterentwicklung derartiger effektiver Beschreibungen der Proteindynamik auf langen Zeitskalen sehr gut eignen. Diesbezügliche Arbeiten, die zunächst auf eine genaue Charakterisierung der Konformationsdynamik des Proteinmodells abzielen, wurden bereits begonnen [164].

Die zweite Beschreibung, ein Markov-Modell, diente der direkten Untersuchung der Konformationsdynamik unseres Proteinmodells, welche auf der wesentlich längeren Zeitskala von einigen hundert Pikosekunden stattfindet. Die gefundene Verteilung von Verweildauern in den verschiedenen Konformationssubzuständen zeigte, daß die Konformationsdynamik des verwendeten Proteinmodells *keine* Gedächtniseffekte aufweist.

Beim Versuch, die Ergebnisse dieser Modellstudie auf das dynamische Verhalten echter Proteine zu übertragen, ist Vorsicht geboten. Wie im Zusammenhang mit der Konstruktion des vereinfachten Modells erläutert wurde, dürfte die Polypeptidkette von Proteinen starrer als unser Modell sein. Entsprechend sollte die durch kollektive Bewegungen der Polypeptidkette bestimmte Konformationsdynamik realer Proteine im allgemeinen auf längeren Zeitskalen stattfinden. Aus diesem Grund wäre es sicher verfehlt zu schließen, daß die Dynamik von Proteinen tatsächlich keine Gedächtniseffekte auf der für die Konformationsdynamik unseres Modells charakteristischen Zeitskala von einigen hundert Pikosekunden aufweist. Die Resultate lassen jedoch vermuten, daß Gedächtniseffekte für die Konformationsdynamik von Proteinen eine untergeordnete Rolle spielen. Darüberhinaus darf die Zeitskala einiger hundert Pikosekunden aufgrund der erhöhten Flexibilität des vereinfachten Proteinmodells als *untere Schranke* für das Verschwinden von Gedächtniseffekten angesehen werden.

Die aufgeführten Resultate lassen *qualitative* Veränderungen der Proteindynamik erwarten, falls der MD-Simulationen an realistischen, komplexen Proteinmodellen zugängliche Zeitbereich verlassen wird. Während auf kurzen Zeitskalen etwa Gedächtniseffekte die Proteindynamik wesentlich beeinflussen und deshalb in entsprechenden stochastischen Beschreibungen nicht vernachlässigt werden dürfen, scheinen derartige Korrelationen im allgemeinen auf längeren Zeitskalen an Bedeutung zu verlieren.

Entsprechend dem Leitgedanken dieser Arbeit, wonach das enge Zusammenspiel von physikalischer Modellbildung und der Entwicklung effizienter Simulationsmethoden Voraussetzung für ein weitergehendes theoretisches Verständnis von Proteinfunktion ist, geben die Resultate unserer Modellstudien Hinweise zur zukünftigen Entwicklung geeigneter effektiver Beschreibungen der Proteindynamik auf langen Zeitskalen. Dem hier eingeführten Begriff der 'Konformationsdynamik' fällt dabei eine zentrale Rolle zu: Offensichtlich repräsentiert die Einteilung des Konfigurationsraums eines Proteins in Konformationssubzustände eine 'Grobkörnung' des Phasenraums, ähnlich der zur molekularen Beschreibung von Flüssigkeiten erfolgreich herangezogenen Unterteilung [158]. Das mutmaßliche Fehlen von Gedächtniseffekten in der zugeordneten Konformationsdynamik scheint deren Beschreibung — zumindest prinzipiell — zu vereinfachen.

Deren praktische Realisierung wird jedoch durch die im allgemeinen komplexe Struktur der Konformationsraumdichte sowie durch die unter Umständen unüberschaubare Zahl zu berücksichtigender Konformationssubzustände außerordentlich erschwert. Es wird daher nötig sein, solche Zustände hierarchisch zu klassifizieren und Konformationskoordinaten zu definieren, die den einzelnen Konformationssubzuständen angepaßt sind. Nach ersten Untersuchungen der Konformationsraumdichte unseres Proteinmodells [164] halten wir hierfür neuronale Algorithmen für besonders geeignet (vgl. hierzu [165] sowie die dort zitierten Arbeiten). Aufgrund seiner relativ einfachen Struktur kann das vorgeschlagene Proteinmodell den weiteren methodischen Entwicklungen als Testsystem dienen. Komplexere Proteinmodelle müssen freilich nach Möglichkeit einbezogen werden.

Als nützliches Hilfsmittel zur Analyse komplexer biochemischer Vorgänge in Proteinen hat sich im Verlauf dieser Arbeit die bewegte Computergraphik ('Computeranimation') erwiesen. So wurde in Zusammenarbeit mit Marco Nonella ein atomares Funktionsmodell des eingangs erwähnten und in Abb. 1.1 dargestellten Membranproteins Bakteriorhodopsin, abgeleitet aus MD-Simulationen, in einem 13-minütigen photorealistischen Lehrfilm veranschaulicht [166, 167]. Zu unserer Überraschung zeigte die Visualisierung funktionsrelevante Details der Dynamik, welche ansonsten, bei ausschließlicher Verwendung konventioneller Analysemethoden, unentdeckt geblieben wären.

Warum sollte man sich die Mühe machen, den angedeuteten langen, teilweise nur schemenhaft erkennbaren Weg der effektiven, hierarchischen Beschreibung der Proteindynamik zu beschreiten? Warum nicht einfach abwarten, bis Computer mit hinreichend hoher Rechenleistung die explizite Simulation enzymatischer Reaktionen oder sogar von Proteinfaltungsprozessen erlauben[1]?

Wahrscheinlich würden 'Sekunden-Simulationen' der Proteindynamik eine Vorhersage von Proteinstruktur und -funktion tatsächlich ermöglichen — vorausgesetzt, ein hinreichend genaues molekulares Modell wäre verfügbar. Wir glauben jedoch, die Analyse von solchen Mammut-Simulationen allein würde wenig zum Verständnis der *Physik* der Proteindynamik beitragen. Neben dem grundlegenden Wunsch nach physikalischem Verstehen ist ein solches Verständnis insbesondere erforderlich, falls etwa Proteine mit gewünschter Funktion 'maßgeschneidert' werden sollen ('Protein-Design'). Wir sind überzeugt, daß effektive Modelle, welche auf eine explizite Beschreibung möglichst vieler irrelevanter Details der atomaren Bewegung verzichten, die Komplexität der Proteindynamik hinreichend reduzieren können, um grundlegende Prinzipien der Proteinfunktion erkennbar zu machen.

[1] Wie bereits erwähnt, muß freilich eine Wartezeit von einigen Jahrzehnten in Kauf genommen werden.

„Symbole sind Mystifikationen."

Novalis, Ende 18. Jh.

Anhang A

Verwendete Formelzeichen

In der folgenden Tabelle sind die wesentlichen in dieser Arbeit verwendeten Formelzeichen aufgeführt:

t	Zeit
ν	Frequenz
ω	Kreisfrequenz
T	Temperatur
k_b	Boltzmannkonstante
N	Anzahl der Atome eines Proteins
$\alpha, \beta, \gamma, \delta$	griechische Indizes bezeichnen Atome
$\mathbf{r}_\alpha$	Ortsvektor des Atoms α
$\mathbf{v}_\alpha$	Geschwindigkeitsvektor des Atoms α
$\mathbf{p}_\alpha$	Impuls des Atoms α
m_α	Masse des Atoms α
q_α	elektrische Ladung des Atoms α
$r_{\alpha\beta}$	Abstand der Atome α und β
$\mathbf{F}_\alpha$	auf das Atom α ausgeübte Kraft
$\mathbf{r}^N$	$3N$-dimensionaler Vektor atomarer Orte
$\mathbf{v}^N$	$3N$-dimensionaler Vektor atomarer Geschwindigkeiten

Tabelle A.1: Verwendete Formelzeichen

Δ_t	Integrationsschrittweite
$t_i := i\Delta_t$	i-ter Integrationsschritt
$\mathbf{r}_i$	Ortsvektor eines Atoms zur Zeit t_i
$\mathbf{F}_i$	auf ein Atom zur Zeit t_i ausgeübte Kraft
E_{Coulomb}	elektrostatische Energie
E_{bond}	Bindungsstreckenergie
E_{angle}	Bindungswinkelenergie
E_{vdW}	Van der Waals-Energie
r_{cut}	Abschneideradius
$f(r)$	Abschneidefunktion
R_j	Entfernungsklassenradien
$\mathbf{F}_i^{(j)}$	auf ein Atom zur Zeit t_i von Atomen der Entfernungsklasse j ausgeübte Kraft
$a_{i,r}^{(j)}$	Extrapolationskoeffizienten von Entfernungsklassenalgorithmen
$E+1$	Anzahl der zur Extrapolation verwendeten zurückliegenden Kräfte
$C_v(t)$	Geschwindigkeits-Autokorrelationsfunktion
N_t	Anzahl durchgeführter Integrationsschritte
N_c	Anzahl verwendeter Konfigurationen
T_t	zeitliche Länge einer Trajektorie
$S(\omega)$	(Leistungs-)Spektrum
σ_α	Beweglichkeit von Atom α
$C_\alpha^{\text{dis}}(\tau)$	Orts-Autokorrelationsfunktion des Atoms α
$C_\alpha^{\text{vel}}(\tau)$	Geschwindigkeits-Autokorrelationsfunktion des Atoms α
$K_{\alpha\beta}$	Kovarianz der Bewegung des Atompaars α, β

Tabelle A.2: Verwendete Formelzeichen (Fortsetzung)

ρ	Phasenraumdichte
ρ_c	projizierte Phasenraumdichte
V^k	(diskretes) Volumenelement eines Konformations(unter-)raums
K	Anzahl der Volumenelemente V^k
P	Anzahl verwendeter Phasenraumprojektionen
$V^{k,p}$	Volumenelement k des durch Projektion p definierten Unterraums
M	Dimension eines Konformationsraums
$\mathbf{c} := \{c_i\,;\, i = 1 \ldots M\}$	Konformationskoordinaten
$\mathcal{H}(\mathbf{q}^N, \mathbf{p}^N)$	Hamiltonfunktion
$W(\mathbf{c})$	effektives Potential
γ	Reibungskoeffizient
ξ	Rauschkraft
k_{BC}	Übergangsrate von Zustand B nach C
$P(t)$	Wahrscheinlichkeit zur Zeit t
$P(E_i\|E_j) = P_{ij}$	bedingte Wahrscheinlichkeit
B, C, D	Konformationssubzustände
S_a	Testsimulation mit MD-Algorithmus a
R_1, R_2	Referenzsimulationen
rms	mittleres Fehlerquadrat (root mean square)

Tabelle A.3: Verwendete Formelzeichen (Fortsetzung)

„Gedenke der Quelle,
wenn du trinkst!“

Chinesisches Sprichwort

Anhang B

Danksagung

Viele haben zum Gelingen dieser Arbeit beigetragen. Ihnen allen möchte ich herzlich danken.

In den vergangenen drei Jahren meiner Promotion stand mir Paul Tavan, unter dessen Leitung diese Arbeit entstand, als Betreuer stets hilfreich und mit großem Engagement zur Seite. Sowohl während der Aufbauphase der Arbeitsgruppe, mit ungewisser Zukunft, als auch in letzter Zeit, in der die Anzahl der Mitarbeiter beachtlich angestiegen ist, fand ich in ihm, wann immer es nötig war, einen konstruktiven Kritiker, der mir half, den richtigen Weg zu finden. Für die hervorragende Zusammenarbeit, für zahlreiche interessante Diskussionen, die häufig ein neues Licht auf die Dinge warfen, für das, was ich bei ihm lernte, sowie für die Schaffung eines ausgezeichneten materiellen Umfelds ein herzliches Dankeschön.

Alle Mitglieder der Arbeitsgruppe ‘Theoretische Biophysik’, neuerdings auch bekannt als ‘Paulaner’, haben dazu beigetragen, ‘daß der Laden läuft’. Für die gute Arbeitsatmosphäre sowie hilfreiche Diskussionen und Beiträge zur Dynamik von Proteinen, Computern, Programmen, Neuronen und Gruppen bedanke ich mich bei den ‘Molis’ Christoph Niedermeier, Freddy Kraxenberger („Ned schwäch’ln!“), Klaus Döhring, Helmut Heller, Nico Ehrenhofer, Werner Bauer und Christian Storch sowie bei den ‘Neuros’ Hans Kühnel, Niels Kunstmann, Dominik Dersch, Claus Hillermeier, Rolf Huber, Ute Adolphs und Bernhard Rabus. Besonders die Zusammenarbeit mit den ‘Altgedienten’ Hans Kühnel und Christoph Niedermeier war von gegenseitiger Unterstützung und freundschaftlichem Umgang geprägt. Die in Abbildung 1.1 zusammengestellten Proteinstrukturen wurden mit Hilfe des von Christoph Niedermeier erstellten Programms `molplot` gezeichnet.

Daß der Kontakt mit Helmut Heller während seines Aufenthalts bei Klaus Schulten nie abgerissen ist, hat mich gefreut. Für die trotz der 'heißen Phase' seiner und meiner Promotion gute Zusammenarbeit und sein engagierteres Bemühen um ein geeignetes Parallelrechnersystem für Molekulardynamiksimulationen (T9000 in virtuellem Gehäuse?) möchte ich ihm herzlich danken.

Für die gute Zusammenarbeit bei der Strukturaufklärung von Bakteriorhodopsin sei Hermann Eckbringhoff gedankt, der mit einer Unmenge von Dichroismusmessungen an einer Unmenge von BR-Kristallen einen wesentlichen Beitrag leistete. Gespräche mit Dieter Oesterhelt waren stets von großem Nutzen und sind mir in angenehmer Erinnerung.

Die Zusammenarbeit mit Marco Nonella bei der Erstellung der Computeranimation 'BR AT WORK' hat mir großen Spaß gemacht. Zu diesem Projekt haben auch Paul Tavan, Dieter Oesterhelt und Klaus Döhring beigetragen. Bei Craig Colb möchte ich mich bedanken für die Überlassung seines 'Ray-Tracing'-Programms und für die schnelle Hilfe, als mir ein Fehler im Programm zu schaffen machte. An dieser Stelle seien auch alle Mitarbeiter des Leibniz-Rechenzentrums, welche die benötigte umfangreiche Rechenzeit (über 2500 h) auf dem HP-Cluster verfügbar machten und dabei halfen, die berechneten Bilder aufs Videoband zu bringen, dankend erwähnt: Herr Sarreiter, Frau Brunner, Herr Weidner und Herr Kaiser.

Der Besuch von Konferenzen und Seminaren ist erforderlich, um eigene Arbeiten und Ideen in internationalem Kontext einzuordnen, diese aus unterschiedlichen Blickwinkeln betrachten zu können und 'über den Tellerrand hinauszuschauen'. Für hilfreiche Diskussionen bedanke ich mich bei Hans Bartunik, Jacques Breton, Charles Brooks, G. Büldt, Andrew McCammon, Enrico Clementi, David Epstein, Sighart Fischer, Hans Frauenfelder, Peter Fromherz, Cornelius Frömmel, Hermann Gaub, Wilfried van Gunsteren, Norbert Hampp, Bernhard Hassenstein, Douglas Heggie, M. Hibert, Tom Kirkwood, Jutta Köhler, Michał Kurzyński, Shneior Lifson, Steven Lustig, Martin Karplus, Peter Kollman, Fritz Parak, Frank Plab, Rudolf Riegler, H. Rüterjans, Chris Sander, Jörg Schlitter, Friedrich Siebert, Robert Skeel, Jeremy Smith, Jörg Tittor, J. van der Waals und H. P. Weber. Für Einladungen zu Vorträgen danke ich Klaus Schulten, Dieter Oesterhelt sowie Manfred Eigen und Ruthild Winkler-Oswatitsch. Die Teilnahme an zahlreichen Seminaren und Konferenzen war nur durch die finanzielle Unterstützung des Sonderforschungsbereichs 143 der Deutschen Forschungsgemeinschaft möglich, in dessen Rahmen diese Arbeit entstand. Dies gab mir die Gelegenheit, meine Arbeiten einer breiteren wissenschaftlichen Öffentlichkeit vorzustellen. Als das Geld nicht reichte, erhielt ich aus den Reisemitteln von Wolfgang Zinth und von Dieter Oesterhelt die notwendigen

Zuschüsse.

Dem Höchstleistungsrechenzentrum der Kernforschungsanlage Jülich sowie dem Rechenzentrum des Max-Planck-Instituts für Biochemie (Martinsried), vertreten durch Herrn Steigemann, sei für die zur Verfügung gestellte Cray-Rechenzeit gedankt.

Ganz besonderer Dank gebührt meiner Ehefrau Brigitte. Ihre Liebe, ihr unermüdlicher Einsatz im Haushalt und bei der Betreuung unserer beiden Töchter sowie ihre Geduld mit einem 'Physik-Vernarrten' ist zu bewundern und gab mir die Möglichkeit, die Ruhe und die Kraft, mich dieser Promotion zu widmen. Auch für das Korrekturlesen einiger Kapitel dieser Arbeit bedanke ich mich.

Wenn es 'eng' wurde, fanden wir beide stets umfangreiche Unterstützung bei unseren Eltern und Schwiegereltern, dafür ebenfalls mein herzlicher Dank.

Literaturverzeichnis

[1] L. Stryer. *Biochemistry.* W. H. Freeman and Company, San Francisco, 1988.

[2] C. R. Cantor and P. R. Schimmel. *Biophysical Chemistry.* W. H. Freeman and Company, New York, 1980.

[3] L. Pauling, R. B. Corey, and H. R. Branson. The structure of proteins: Two hydrogen-bonded helical configurations of polypeptide chain. *Proc. Natl. Acad. Sci. USA*, 37:205, 1951.

[4] R. Henderson, J. M. Baldwin, T. A. Ceska, F. Zemlin, E. Beckmann, and K. H. Downing. Model for the structure of bacteriorhodopsin based on high-resolution electron cryo-microscopy. *J. Molec. Biol.*, 213:899–929, 1990.

[5] Dieter Oesterhelt and Walther Stoeckenius. Functions of a new photoreceptor membrane. *Proc. Natl. Acad. Sci. USA*, 70(10):2853–2857, 1973.

[6] R. Kaptein, R. Boelens, R. M. Scheek, and W. F. van Gunsteren. Protein structure from NMR. *Biochemistry*, 27(15):5389–5395, Jul. 1988.

[7] Axel T. Brünger, John Kuriyan, and Martin Karplus. Crystallographic R-factor refinement by molecular dynamics. *Science*, 235:458–460, Jan. 1987.

[8] Axel T. Brünger, Robert L. Campbell, G. Marius Clore, Angela M. Gronenborn, Martin Karplus, Gregory A. Petsko, and Martha M. Teeter. Solution of a protein crystal structure with a model obtained from NMR interproton distance restraints. *Science*, 235:1049–1053, Feb. 1987.

[9] J. Deisenhofer and H. Michel. The crystal structure of the photosynthetic reaction center from Rps. viridis. In J. Breton and A. Vermeglio, editors, *The Photosynthetic Bacterial Reaction Center: Structure and Dynamics*, Seiten 1–3, London, 1988. Plenum Press.

[10] M. Karplus and A. McCammon. Dynamics of proteins: Elements and function. *Ann. Rev. Biochem.*, 53:263–300, 1983.

[11] J. A. McCammon. Protein dynamics. *Rep. Prog. Phys.*, 47:1–46, 1984.

[12] M. Karplus and J. A. McCammon. The dynamics of proteins. *Scient. Am.*, 4:30, 1986.

[13] M. Levitt. Molecular dynamics of native proteins. II. Analysis and nature of motion. *J. Molec. Biol.*, 168:621, 1983.

[14] J. A. McCammon and M. Karplus. Dynamics of activated processes in globular proteins. *Proc. Natl. Acad. Sci. USA*, 76:3585, 1987.

[15] H. Frauenfelder. Protein states and motion. *Biophys. J.*, 47:35, 1985.

[16] M. C. Nuss, W. Zinth, W. Kaiser, E. Kölling, and D. Oesterhelt. Femtosecond spectroscopy of the first events of the photochemical cycle in bacteriorhodopsin. *Chem. Phys. Lett.*, 117(1):1–7, 1985.

[17] Feng Zhou, Andreas Windemuth, and Klaus Schulten. Molecular-dynamis study of the proton pump cycle of bacteriorhodopsin. *Biochemistry*, 32(9):2291–2306, 1993.

[18] Heinz-Jürgen Steinhoff, Jürgen Schlitter, Albrecht Redhardt, and Dirk Husmeier. Structural fluctuations and conformational entropy in proteins: entropy balance in an intramolecular reaction in methemoglobin. *Biochim. Biophys. Acta*, 1121:189–198, 1992.

[19] Klaus Schulten, Zan Schulten, and Paul Tavan. An isomerization model for the pump cycle of bacteriorhodopsin. In L. Bolis, E. J. M. Helmreich, and H. Passow, editors, *Information and Energy Transduction in Biological Membranes*, Seiten 113–131. Allan R. Liss, Inc., 1984.

[20] György Váró and Janos K. Lanyi. Thermodynamics and energy coupling in the bacteriorhodopsin photocycle. *Biochemistry*, 30:5016–5022, 1991.

[21] György Váró and Janos K. Lanyi. Distorsions in the photocycle of bacteriorhodopsin at moderate dehydration. *Biophys. J.*, 59:313–322, Feb. 1991.

[22] A. Ansari, J. Berendzen, S. F. Browne, H. Frauenfelder, I. E. T. Iben, T. B. Sauke, E. Shyamsunder, and R. D. Young. Protein states and protein quakes. *Proc. Natl. Acad. Sci. USA*, 82:5000, 1985.

[23] K. P. Hofmann. Rhodopsin/G-protein interaction. In B. Dickey and L. Birnbaumer, editors, *GTPases in Biology; Handbook of Experimental Pharmacology*, Seiten 955–978, Berlin, Heidelberg, New York. Springer-Verlag. Im Druck.

[24] Erwin Neher and Bert Sakmann. Single-channel events recorded from membrane of denervated frog muscle fibres. *Nature*, 260:799–802, 1976.

[25] German Costa, Kathleen W. Kinnally, and Joyce J. Diwan. Patch clamp analysis of a partially purified ion channel from rat liver mitochondria. *Biochem. Biophys. Research Comm.*, 175:305–310, 1991.

[26] Thomas E. Creighton. Protein folding. *Biochem. J.*, 270:1–16, 1990.

[27] Hue Sun Chan and Ken A. Dill. The protein folding problem. *Physics Today*, 95:24–32, 1993.

[28] R. H. Austin, K. W. Beeson, L. Eisenstein, H. Frauenfelder, and I.C. Gunsalus. Dynamics of ligand binding to myoglobin. *Biochemistry*, 14(24):5355–5373, 1975.

[29] Anjum Ansari, Joel Berendzen, David Braunstein, Benjamin R. Cowen, Hans Frauenfelder, Mi Kyung Hong, Icko E. T. Iben, J. Bruce Johnson, Pál Ormos, Todd B. Sauke, Reinhard Scholl, Alfons Schulte, Peter J. Steinbach, Joseph Vittitow, and Robert D. Young. Rebinding and relaxation in the myoglobin pocket. *Biophys. Chem.*, 26:337–355, 1987.

[30] Hans Frauenfelder et al. Proteins and pressure. *J. Phys. Chem.*, 94:1024–1037, 1990.

[31] G. Wagner and N. R. Nirmala. Studies of protein dynamics by heteronuclear NMR: Individual ^{13}C relaxation times and evidence for multiple conformations in the reactive site of BPTI. *Chem. Scr.*, 29A:27–30, 1989.

[32] G. Ulrich Nienhaus, Judith R. Mourant, and Hans Frauenfelder. Spectroscopic evidence for conformational relaxation in myoglobin. *Proc. Natl. Acad. Sci. USA*, 89:2902–2906, 1992.

[33] J. Zollfrank, J. Friedrich, and F. Parak. Spectral hole burning study of protoporphyrin IX substituted myoglobin. *Biophys. J.*, 61:716–724, 1992.

[34] Hans Frauenfelder, Gregory A. Petsko, and Demetrius Tsernoglou. Temperature-dependent X-ray diffraction as a probe of protein structural dynamics. *Nature*, 280:558–563, August 1979.

[35] Mi Kyung Hong, David Braunstein, Benjamin R. Cowen, Hans Frauenfelder, Icko E. T. Iben, Judith R. Mourat, Pál Ormos, Reinhard Scholl, Alfons Schulte, Peter J. Steinbach, Ai-Hua Xie, and Robert D. Young. Conformational substates and motions in myoglobin. *Biophys. J.*, 58:429–436, 1990.

[36] Hans Frauenfelder, Stephen G. Sligar, and Peter G. Wolynes. The energy landscape and motions of proteins. *Science*, 254:1598–1603, 1991.

[37] Hans Frauenfelder, Kelvin Chu, and Robert Philipp. Physics from proteins. In L. Peliti, editor, *Biologically Inspired Physics*. Plenum Press, New York, 1991.

[38] R. Elber and M. Karplus. Multiple conformational states of proteins: A molecular dynamics analysis of myoglobin. *Science*, 235:318–321, 1987.

[39] Nobuhiro Go and Tosiyuki Noguti. Structural basis of hierarchical multiple substates of a protein. *Chem. Scr.*, 29A:151–164, 1989.

[40] Ron Elber. Calculation of the potential of mean force using molecular dynamics with linear constraints: An application to a conformational transition in a solvated dipeptide. *J. Chem. Phys.*, 93(6):4312–4321, 1990.

[41] M. Levitt and Shneior Lifson. Refinement of protein conformation using a macromolecular energy minimization procedure. *J. Molec. Biol.*, 46:269–279, 1969.

[42] W. F. van Gunsteren and H. J. C. Berendsen. Algorithms for macromolecular dynamics and constraint dynamics. *Molec. Phys.*, 34(5):1311–1327, 1977.

[43] J. A. McCammon, B. R. Gelin, and M. Karplus. Dynamics of folded proteins. *Nature*, 267:585–590, 1977.

[44] Wilfried F. van Gunsteren and Herman J. C. Berendsen. Computer simulation of molecular dynamics: Methodology, applications, and perspectives in chemistry. *Angew. Chem. Int. Ed. Engl.*, 29:992–1023, 1990.

[45] J. A. McCammon and S. C. Harvey. *Dynamics of Proteins and Nucleic Acids.* Cambridge University Press, Cambridge, 1987.

[46] G. Lipari, A. Szabo, and R. M. Levy. Protein dynamics and NMR relaxation: Comparison of simulations with experiment. *Nature*, 300:197, 1982.

[47] Jeremy Smith and Martin Karplus. Empirical force field study of geometries and conformational transitions of some organic molecules. *J. Am. Chem. Soc.*, 114(3):801–812, 1992.

[48] Michael J. Mitchell and J. Andrew McCammon. Free energy difference calculations by thermodynamic integration: Difficulties in obtaining a precise value. *J. Comp. Chem.*, 12(2):271–275, 1991.

[49] A. Jean-Charles, A. Nicholls, K. Sharp, B. Honig, A. Tempezyk, T. F. Hendrickson, and W. C. Still. Electrostatic contributions to solvation energies: Comparison of free energy perturbation and continuum calculations. *J. Am. Chem. Soc.*, 113:1454–1455, 1991.

[50] J. R. Hart and A. K. Rappé. Van der Waals functional forms for molecular simulations. *J. Chem. Phys.*, 97(2):1109–1115, July 1992.

[51] J. Pleiss and F. Jähnig. Conformational transition of an α-helix studied by molecular dynamics. *Eur. Biophys. J.*, 21:63–70, 1992.

[52] Robert D. Young et al. Time- and temperature dependence of large-scale conformational transitions in myoglobin. *Chem. Phys.*, 158:315–327, 1991.

[53] Olle Teleman and Bo Jönsson. Vectorizing a general purpose molecular dynamics simulation program. *J. Comp. Chem.*, 7(1):58–66, 1986.

[54] John E. Mertz, Douglas J. Tobias, Charles L. Brooks, and U. C. Singh. Vector and parallel algorithms for the molecular dynamics simulation of macromolecules on shared-memory computers. *J. Comp. Chem.*, 12(10):1270–1277, 1991.

[55] N. S. Ostlund and R. A. Whiteside. A machine architecture for molecular dynamics: the systolic loop. In B. Venkataraghavan and R. J. Feldman, editors, *Macromolecular Structure and Specificity: Computer-Assisted Modeling and Applications*, Seiten 195–208. Annals of the N. Y. Acad. of Sciences 439, New York, 1985.

[56] D. C. Rapaport. Large-scale molecular dynamics simulation using vector and parallel computers. *Computer Physics Reports*, 9:1–53, 1988.

[57] Helmut Heller, Helmut Grubmüller, and Klaus Schulten. Molecular dynamics simulation on a parallel computer. *Molec. Sim.*, 5:133–165, 1990.

[58] Eugene D. Brooks III. Massive parallelism overcomes shared-memory limitations. *Computers in Physics*, 6(2):139–145, March/April 1992.

[59] Bernhard R. Brooks and Milan Hodošček. Parallelization of CHARMM for MIMD machines. *Chem. Des. Autom. News*, 7(12):16–22, December 1992.

[60] W. Scott, A. Gunzinger, B. Bäumle, P. Kohler, U. A. Müller, H-R. Vonder Mühll, A. Eichenberger, W. Guggenbühl, N. Ironmonger, F. Müller-Plathe, and W. F. van Gunsteren. Parallel molecular dynamics on a multi signalprocessor system. *Computer Phys. Comm.*, 75(1-2):65–86, 1993.

[61] H. Grubmüller. Dynamiksimulation sehr großer Makromoleküle auf einem Parallelrechner. Diplomarbeit, Technische Universität München, 1989.

[62] E. Hairer, S. P. Nørsett, and G. Wanner. Solving Ordinary Differential Equations I. Springer Series in Computational Mathematics. Springer-Verlag, Berlin, 1987.

[63] Daniel Irowa Okunbor. *Canonical Integration Methods for Hamiltonian Dynamical Systems.* Doktorarbeit, University of Illinois at Urbana-Champaign, Department of Computer Science, 1304 W. Springfield Avenue, Urbana, Il 61801-2987, 1992.

[64] Michael Buchner and Branka M. Ladanyi. Molecular dynamics algorithm for flexible molecules using normal coordinates. *Molec. Phys.*, 73(5):1127–1143, 1991.

[65] Jim Turner, Hon Chun, Victor Lupi, Paul Weiner, Steve Gallion, and Chandra Singh. Researchers apply variable reduction technices to molecular dynamics simulations. *Chem. Des. Autom. News*, 7(12):33–39, 1992.

[66] R. D. Ruth. A canonical integration technique. *IEEE Trans. Nuc. Sci.*, 30(4):2669–2671, 1983.

[67] Josh Barnes and Piet Hut. A hierarchical $O(N \log N)$ force-calculation algorithm. *Nature*, 324:446–449, December 1986.

[68] L. Greengard and V. Rokhlin. A fast algorithm for particle simulations. *J. Comp. Phys.*, 73:325–348, 1987.

[69] L. Greengard and V. Rokhlin. On the evaluation of electrostatic interactions in molecular modeling. *Chem. Scr.*, 29A:139–144, 1989.

[70] James F. Leathrum and John A. Board. The parallel fast multipole algorithm in three dimensions. Technical report, Dept. of Electrical Engineering, Duke Univerity, Durham, 1992.

[71] S. J. Aarseth. *Direct Methods for N-Body Simulations*, Seiten 378–418. Multiple Time Scales. Academic Press, 1st edition, 1985.

[72] Zahari Zlatev. Advances in the theory of variable stepsize variable formula methods for ordinary differential-equations. *Applied Mathematics and Computation*, 31:209–250, May 1989.

[73] R. C. Y. Chin, G. W. Hedstrom, and F. A. Howes. *Considerations on Solving Problems with Multiple Scales*, Seiten 1–27. Multiple Time Scales. Academic Press, 1st edition, 1985.

[74] R. Chicon and E. Martin. Variable step methods for the integration of Newton's equation. *American Journal of Physics*, 59(8):759–760, Aug. 1991.

[75] B. J. Berne. *Molecular Dynamics and Monte Carlo Simulations of Rare Events*, Seiten 419–436. Multiple Time Scales. Academic Press, 1st edition, 1985.

[76] Bernhard R. Brooks, Robert E. Bruccoleri, Barry D. Olafson, David J. States, S. Swaminathan, and Martin Karplus. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comp. Chem.*, 4(2):187–217, 1983.

[77] Helmut Grubmüller, Helmut Heller, Andreas Windemuth, and Klaus Schulten. Generalized Verlet algorithm for efficient molecular dynamics simulations with long-range interactions. *Molec. Sim.*, 6:121–142, 1991.

[78] Mark E. Tuckerman and Bruce J. Berne. Molecular dynamics in systems with multiple time scales: Systems with stiff and soft degrees of freedom and with short and long range forces. *J. Chem. Phys.*, 95(11):8362–8364, December 1991.

[79] Richard J. Loncharich and Bernard R. Brooks. The effects of truncating long-range forces on protein dynamics. *Proteins*, 6:32–45, 1989.

[80] Charles H. Bennet. Mass tensor molecular dynamics. *J. Comp. Phys.*, 19:297–279, 1975.

[81] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, and E. Teller. Equation of state calculations for fast computing machines. *J. Chem. Phys.*, 21:1087–1092, 1953.

[82] C. Pangali, M. Rao, and B. J. Berne. On a novel Monte Carlo scheme for simulating water and aqueous solutions. *Chem. Phys. Lett.*, 55(3):413–417, 1978.

[83] R. M. Levy, M. Karplus, and J. A. McCammon. Diffusive langevin dynamics of model alkanes. *Chem. Phys. Lett.*, 65(1):4–11, 1979.

[84] Mark E. Tuckerman and Bruce J. Berne. Stochastic molecular dynamics in systems with multiple time scales and memory friction. *J. Chem. Phys.*, 95(6):4389–4396, September 1991.

[85] Ram Brustein, Shlomo Marianer, and Moshe Schwartz. Langevin memory kernel and noise from Lagrangian dynamics. *Physica A*, 175:47–58, 1991.

[86] J.U. Bowie, R. Luthy, and D. Eisenberg. A method to identify protein sequences that fold into a known 3-dimensional structure. *Science*, 253:164–170, 1991.

[87] Sanzo Miyazawaa and Robert L. Jernigan. Estimation of effective interresidue contact energies from protein crystal structures: Quasi-chemical approximation. *Macromolecules*, 18:534–552, 1985.

[88] S. C. Harvey. Treatment of electrostatic effects in macromolecular modeling. *Proteins*, 5:78–92, 1989.

[89] A. Ahmad and L. Cohen. A numerical integration scheme for the N-body gravitational problem. *J. Comp. Phys.*, 12:389–402, 1973.

[90] W. B. Streett, D. J. Tildesley, and G. Saville. Multiple time step methods in molecular dynamics. *Molec. Phys.*, 35:639–648, 1978.

[91] Andreas Windemuth. Dynamiksimulation von Makromolekülen. Diplomarbeit, Technische Universität München, August 1988.

[92] C. L. Brooks, A. Brünger, and M. Karplus. Active site dynamics in protein molecules: A stochastic boundary molecular-dynamics approach. *Biopolymers*, 24:843, 1985.

[93] Herbert Treutlein, Klaus Schulten, J. Deisenhofer, H. Michel, Axel Brünger, and Martin Karplus. Chromophore-protein interactions and the function of the photosynthetic reaction center: A molecular dynamics study. *Proc. Natl. Acad. Sci. USA*, 89:75–79, 1991.

[94] S. Swaminathan, T. Ichiye, W. van Gunsteren, and M. Karplus. Time dependence of atomic fluctuations in proteins: Analysis of local and collective motions in bovine pancreatic trypsin inhibitor. *Biochemistry*, 21(21):5230–5241, 1982.

[95] Walter Nadler, Axel T. Brünger, Klaus Schulten, and Martin Karplus. Molecular and stochastic dynamics of proteins. *Proc. Natl. Acad. Sci. USA*, 84:7933–7937, Nov. 1987.

[96] D. E. Smith and C. B. Harris. Generalized Brownian dynamics. I. Numerical integration of the generalized Langevin equation through autoregressive modeling of the memory function. *J. Chem. Phys.*, 92(2):1304–1311, Jan. 1990.

[97] T. Xiang, Fang Liu, and David M. Grant. Generalized Langevin equations for molecular dynamics in solution. *J. Chem. Phys.*, 94(6):4463–4471, Mar. 1991.

[98] Robert Miller, Craig A. Danko, Michael J. Fasolka, and Anna C. Balazs. Folding kinetics of proteins and copolymers. *J. Chem. Phys.*, 96(1):768–780, Jan. 1992.

[99] J. D. Honeycutt and D. Thirumalai. The nature of folded states of globular proteins. *Biopolymers*, 32:695–709, 1992.

[100] Zhuyan Guo, D. Thirumalai, and J. D. Honeycutt. Folding kinetics of proteins: A model study. *J. Chem. Phys.*, 97(1):525–535, July 1992.

[101] J. D. Honeycutt and D. Thirumalai. Metastability of the folded states of globular proteins. *Proc. Natl. Acad. Sci. USA*, 87:3526–3529, May 1990.

[102] Helmut Grubmüller. On the suitability of efficient many-body algorithms for molecular dynamics simulations of biological macromolecules. In K. I. M. McKinnon and F. Plab, editors, *Proceedings of the Second Parallel Numerical Analysis Workshop, June 25–26, 1992, Edinburgh, UK*, Seiten 226–238. Edinburgh Parallel Computing Centre, University of Edinburgh, 1992.

[103] Helmut Grubmüller and Paul Tavan. Efficient algorithms for molecular dynamics simulations of proteins: How good are they? *J. Chem. Phys.*, 1993. Eingesandt.

[104] Helmut Grubmüller and Paul Tavan. Molecular dynamics of conformational substates for a simplified protein model. *J. Chem. Phys.*, 1993. Im Druck.

[105] B. J. Alder and T. E. Wainwright. Phase transition for a hard sphere system. *J. Chem. Phys.*, 27:1208–1209, 1957.

[106] A. Rahman. Correlations in the motion of atoms in liquid argon. *Phys. Rev.*, 136(2A):A405–A411, 1964.

[107] Loup Verlet. Computer "experiments" on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules. *Phys. Rev.*, 159(1):98–103, Jul. 1967.

[108] Herbert Treutlein. *Molekulardynamiksimulation des photosynthetischen Reaktionszentrums von Rhodopseudomonas virids.* Doktorarbeit, Technische Universität München, Dez. 1988.

[109] Marco Nonella and Dieter Oesterhelt. Structure of bacteriorhodopsin and the effect of mutation of ASP-85 or ASP-212 on structure, spectrum, and electrostatics: A computational study. *J. Molec. Biol.*, 1994. Eingesandt.

[110] W. F. van Gunsteren and H. J. C. Berendsen. *GROMOS Manual.* BIOMOS b. v., Biomolecular Software, Groningen, The Netherlands.

[111] Helmut Heller, Michael Schaefer, and Klaus Schulten. Molecular dynamics simulation of a bilayer of 200 lipids in the gel and in the liquid crystal phase. *J. Phys. Chem.*, 97(31):8343–8360, 1993.

[112] H. Heller. Private Mitteilung.

[113] Malcolm E. Davis and J. Andrew McCammon. Electrostatics in biomolecular structure and dynamics. *Chem. Rev.*, 90(3):509–521, 1990.

[114] G. Klopman and R. C. Evans. The neglect-of-differential-overlap methods of molecular orbital theory. In G. A. Segal, editor, *Semiempirical methods of electronic structure calculation, Part A: Techniques.* Plenum Press, New York, 1977.

[115] Scott J. Weiner, Peter A. Kollman, David A. Case, U. Chandra Singh, Caterina Ghio, Giuliano Alagona, Jr. Salvatore Profeta, and Paul Weiner. A new force field for molecular mechanical simulation of nucleic acids and proteins. *J. Am. Chem. Soc.*, 106:765–784, 1984.

[116] Scott J. Weiner, Peter A. Kollman, Dzung T. Nguyen, and David A. Case. An all atom force field for simulations of proteins and nucleic acids. *J. Comp. Chem.*, 7(2):230–252, 1986.

[117] A. Wallquist and G. Karlström. A new non-empirical force field for computer simulations. *Chem. Scr.*, 29A, 1989.

[118] A. Mackerell, D. Bashford, M. Bellott, R. Dunbrack, M. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph, L. Kuchnir, K. Kuczera, F. Lau, C. Mattos, S. Michnick, D. Nguyen, T. Ngo, B. Prodhom, B. Roux, M. Schlenkrich, J. Smith, R. Stote, J. Straub, J. Wiorkiewiczkuczera, and M. Karplus. Self-consistent parametrization of biomolecules for molecular modeling and condensed phase simulations. *The FASEB Journal*, 6(1):143, 1992.

[119] A. Warshel. The consistent force field and its quantum mechanical extensions. In G. A. Segal, editor, *Semiempirical Methods of Electronic Structure Calculation, Part A: Techniques*, New York, 1977. Plenum Press.

[120] E. J. Nyström. Über die numerische Integration von Differentialgleichungen. *Acta Soc. Sci. Fenn.*, 50(13):1–54, 1925.

[121] R. H. Miller. A horror story about integration methods. *J. Chem. Phys.*, 93:469–476, 1991.

[122] D. Okunbor and R. D. Skeel. Explicit canonical methods for Hamiltonian systems. *Working Document, Numerical Computing Group, University of Illinois at Urbana-Champaign*, (1), 1991.

[123] Charles L. Brooks III, B. Montgomery Pettitt, and Martin Karplus. Structural and energetic effects of truncating long ranged interactions in ionic and polar fluids. *J. Chem. Phys.*, 83(11):5897–5908, December 1985.

[124] A. T. Brünger. *X-PLOR*. The Howard Hughes Medical Institute and Department of Molecular Biophysics and Biochemistry, Yale University, 260 Whitney Avenue, P.O. Box 6666, New Haven, CT 06511, May 1988.

[125] J. Warwicker and H. C. Watson. Calculation of the electric potential in the active site cleft due to α-helix dipoles. *J. Molec. Biol.*, 157:671–679, 1982.

[126] I. Klapper, R. Hagstrom, R. Fine, K. Sharp, and B. Honig. Focusing of electric fields in the active site of Cu-Zn superoxide dismutase: Effects of ionic strength and amino-acid modification. *Proteins*, 1:47–59, 1986.

[127] A. Warshel and S. T. Russel. Calculations of electrostatic interactions in biological systems and in solution. *Q. Rev. Biophys.*, 17:283–422, 1984.

[128] Anthony Nicholls and Barry Honig. A rapid finite difference algorithm, utilizing successive over-relaxation to solve the Poisson-Boltzmann equation. *J. Comp. Chem.*, 12(4):435–445, 1991.

[129] Thomas Simonson, David Perahia, and Axel T. Brünger. Microscopic theory of the dielectric properties of proteins. *Biophys. J.*, 59:670–690, Mar. 1991.

[130] Zdzislav Jackiewicz. Quasilinear multistep methods and variable step predictor-corrector methods for neutral functional differential equations. *SIAM J. Numer. Anal.*, 23(2):423–452, Apr. 1986.

[131] Gerald D. Quinlan and Scott Tremaine. Symmetric multistep methods for the numerical integration of planetary orbits. *The Astronomical Journal*, 100(5):1694–1700, Nov. 1990.

[132] Roger Penrose. *The Emperor's New Mind.* Oxford University Press, New York, 1989.

[133] Mark E. Tuckerman and Bruce J. Berne. Molecular dynamics algorithm for multiple time scales: Systems with disparate masses. *J. Chem. Phys.*, 94(2):1465–1469, January 1991.

[134] R. D. Skeel. Private Mitteilung.

[135] W. F. van Gunsteren, H. J. C. Berendsen, F. Colonna, D. Perahia, J. P. Hollenberg, and D. Lellouch. On searching neighbors in computer simulations of macromolecular systems. *J. Comp. Chem.*, 5(3):272–279, 1984.

[136] Andrew W. Appel. An efficient program for many-body simulation. *Siam J. Sci. Stat. Comput.*, 6(1):85–103, January 1985.

[137] L. Greengard and V. Rokhlin. On the efficient implementation of the fast multipole algorithm. *Research Report of the Yale University, Department of Computer Science*, RR-602, Feb. 1988.

[138] C. Niedermeier and P. Tavan. A structure adapted multipole method for electrostatic interactions in protein dynamics. *J. Chem. Phys.*, 1994. Eingesandt.

[139] Christoph Niedermeier. *Modellierung elektrostatischer Wechselwirkungen in Proteinen: Eine strukturadaptierte Multipolmethode.* Doktorarbeit, Ludwig-Maximilians-Universität, München, 1994. In Vorbereitung.

[140] T. E. Creighton. *Proteins.* W. H. Freeman and Company, San Francisco, 1984.

[141] Kan A. Dill. Dominant forces in protein folding. *Biochemistry*, 29(31):7133–7155, Aug. 1990.

[142] M. Prévost, S. J. Wodak, B. Tidor, and M. Karplus. Contribution of the hydrophobic effect to protein stability: Analysis based on simulations of the Ile-96 $\rightarrow$ Ala mutation in barnase. *Proc. Natl. Acad. Sci. USA*, 88:10880–10884, 1991.

[143] J. C. Smith. Protein dynamics: comparison of simulations with inelastic neutron scattering experiments. *Q. Rev. Biophys.*, 24(3):227–291, 1991.

[144] Michał Kurzyński. Enzymatic catalysis as a process controlled by protein conformational relaxation. *Accounts of Chemical Research*. Eingesandt.

[145] Hans Frauenfelder. New looks at protein motions. *Nature*, 338:623–624, April 1989.

[146] Kenneth D. Gibson and Harold A. Scheraga. Variable step molecular dynamics: An exploratory technique for peptides with fixed geometry. *J. Comp. Chem.*, 11(4):468–486, 1990.

[147] Mark E. Tuckerman, Glenn J. Martyna, and Bruce J. Berne. Molecular dynamics algorithm for condensed systems with multiple time scales. *J. Chem. Phys.*, 93(2):1287–1291, Jul. 1990.

[148] Martine Prévost, Daniel van Belle, Guy Lippens, and Shoshana Wodak. Computer simulations of liquid water: Treatment of long-range interactions. *Molec. Phys.*, 71(3):587–603, 1990.

[149] Piotr Romiszowski and Robert Yaris. A dynamic simulation method suppressing uninteresting degrees of freedom. *J. Chem. Phys.*, 94(10):6751–6761, May 1991.

[150] Hans C. Anderson. Molecular dynamics simulations at constant pressure and/or temperature. *J. Chem. Phys.*, 72(4):2384–2393, Feb. 1980.

[151] J. M. Haile and S. Gupta. Extensions of the molecular dynamics simulation method. II. Isothermal systems. *J. Chem. Phys.*, 79(6):3067–3076, Sep. 1983.

[152] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.*, 81(8):3684–3690, Oct. 1984.

[153] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling. *Numerical Recipes in C*. Cambridge University Press, Cambridge, 1988.

[154] H. Treutlein, K. Schulten, J. Deisenhofer, H. Michel, A. Brünger, and M. Karplus. Molecular dynamics simulation of the primary processes in the photosynthetic reaction center of rhodpseudomonas viridis. In J. Breton and A. Vermeglio, editors, *The Photosynthetic Bacterial Reaction Center*, volume 149 of *Nato ASI Series: Life Sciences*, Seiten 139–150, New York, 1988. Plenum Publishing Corporation.

[155] Klaus Schulten and Markus Tesch. Coupling of protein motion to electron transfer: Molecular dynamics and stochastic quantum mechanics study of photosynthetic reaction centers. *Chem. Phys.*, 158:421–446, 1991. [Beckman Institute Technical Report TB-91-15].

[156] R. A. Marcus and N. Sutin. Electron transfers in chemistry and biology. *Biochim. Biophys. Acta*, 811:265–322, 1985.

[157] Frederick Reif. *Statistische Physik und Theorie der Wärme.* Walter de Gruyter, Berlin, 2nd edition, 1985.

[158] Stuart A. Rice and Peter Gray. *The Statistical Mechanics of Simple Liquids.* Interscience Publishers, John Wiley & Sons, Inc., New York, 1965.

[159] H. Frauenfelder, P. J. Steinbach, and R. D. Young. Conformational relaxation in proteins. *Chem. Scr.*, 29A:145–150, 1989.

[160] S. Chandrasekhar. Stochastic problems in physics and astronomy. *Rev. Mod. Phys.*, 15(1):1–89, 1943.

[161] H. A. Kramers. Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica*, VII(4):284–304, 1940.

[162] Peter Hänggi, Peter Talkner, and Michal Borkovec. Reaction-rate theory: fifty years after Kramers. *Rev. Mod. Phys.*, 62(2):251–341, Apr. 1990.

[163] Bronstein and Semendjajew. *Taschenbuch der Mathematik.* Harry Deutsch, Thun, Frankfurt (Main), 21th edition, 1981.

[164] N. Ehrenhofer. Untersuchung der Konformationsdynamik eines vereinfachten Proteinmodells. Diplomarbeit, Ludwig-Maximilians-Universität München, 1994. In Vorbereitung.

[165] Paul Tavan, Helmut Grubmüller, and Hans Kühnel. Self-organization of associative memory and pattern classification: Recurrent signal processing on topological feature maps. *Biological Cybernetics*, 64(2):95–105, 1990.

[166] Helmut Grubmüller, Klaus Döhring, Paul Tavan, Marco Nonella, and Dieter Oesterhelt. BR AT WORK: A computeranimation for the 13-14-cis-model of the photochemical cycle of bacteriorhodopsin. *Journal of Molecular Graphics*, 1994. Im Druck.

[167] Helmut Grubmüller, Klaus Döhring, Paul Tavan, Marco Nonella, and Dieter Oesterhelt. Ein Protein bei der Arbeit: Computervisualisierung eines atomaren Funktionsmodells von Bakteriorhodopsin. *Spektrum der Wissenschaft*, 1993. Eingesandt.