

# Biomolekulare Modellierung: Ziele, Probleme, Perspektiven

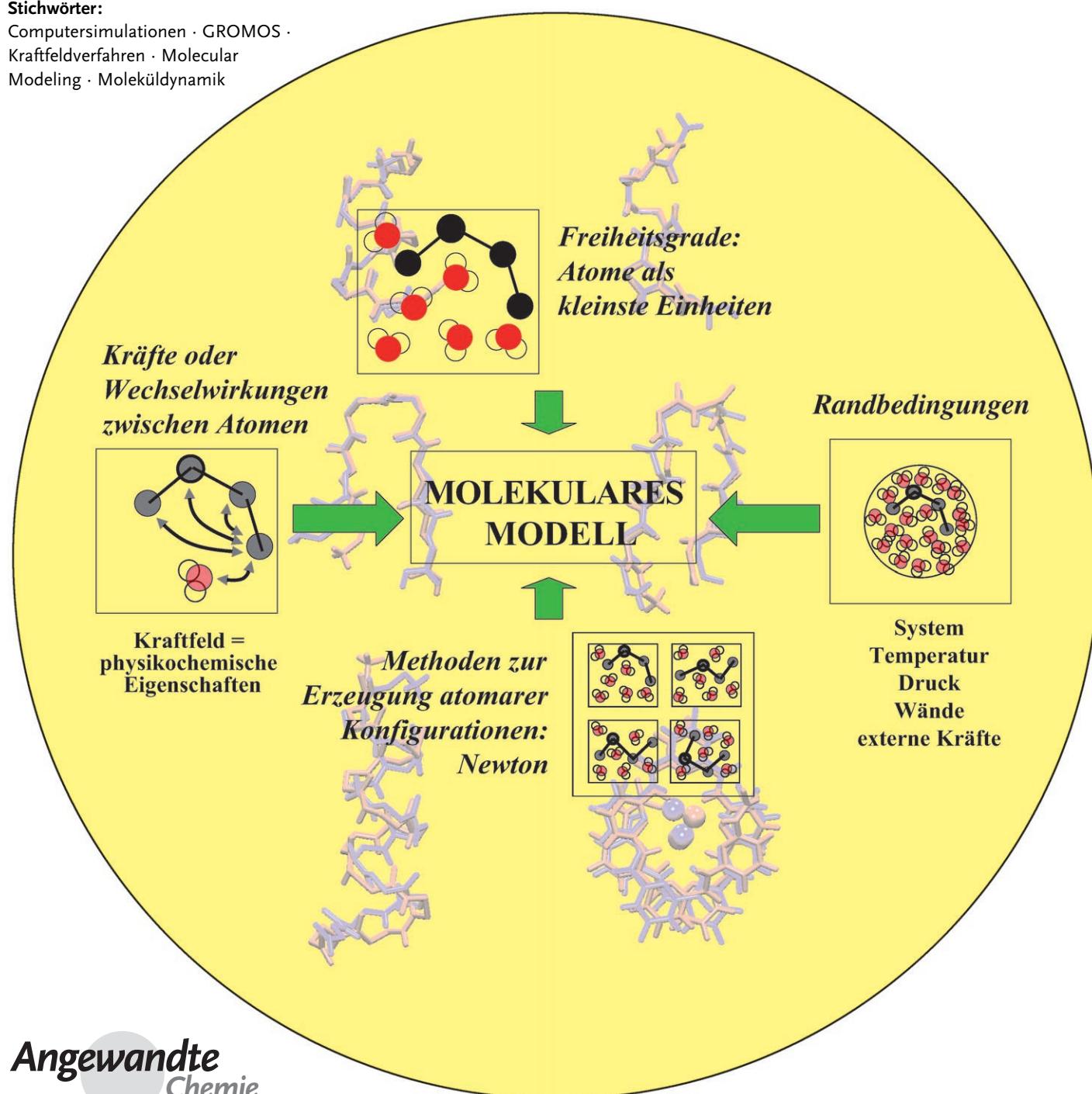
Wilfred F. van Gunsteren,\* Dirk Bakowies, Riccardo Baron, Indira Chandrasekhar, Markus Christen, Xavier Daura, Peter Gee, Daan P. Geerke, Alice Glättli, Philippe H. Hünenberger, Mika A. Kastenholz, Chris Oostenbrink, Merijn Schenk, Daniel Trzesniak, Nico F. A. van der Vegt und Haibo B. Yu

**Stichwörter:**

Computersimulationen · GROMOS ·

Kraftfeldverfahren · Molecular

Modeling · Moleküldynamik



**C**omputerverfahren auf der Grundlage von Molekülmodellen gewinnen in Biologie, biologischer Chemie und Biophysik zunehmend an Bedeutung. Da nur wenige Eigenschaften biomolekularer Systeme durch Messungen zugänglich sind, können Computersimulationen experimentelle Daten ergänzen, indem sie nicht nur Durchschnittswerte, sondern auch Verteilungen und Zeitreihen jeder bestimmbaren Größe liefern, z. B. Konformationsverteilungen oder Wechselwirkungen zwischen Teilen eines Systems. Die Anwendung moderner biomolekulärer Modellierungsverfahren wird zurzeit durch vier grundlegende Probleme eingeschränkt: 1) das Kraftfeldproblem, 2) das Suchproblem, 3) das Ensembleproblem und 4) das Experimentalproblem. Diese vier Probleme werden anhand praktischer Beispiele erläutert, außerdem werden Lösungsperspektiven aufgezeigt.

## 1. Einleitung

In den vergangenen zehn Jahren haben Computerverfahren in der Biologie, biologischen Chemie und Biophysik enorm an Bedeutung gewonnen. Durch die rasant verbesserte Leistungsfähigkeit von Computern (und insbesondere von Personalcomputern) ist es möglich geworden, große und komplexe Datensätze aus Experimenten an biomolekularen Systemen zu analysieren, zu vergleichen und zu charakterisieren. Dies führte im Gegenzug zur Entwicklung von Modellen für biomolekulare Vorgänge, die sich durch Computerverfahren simulieren und analysieren lassen. Entscheidend für jede biomolekulare Modellstudie ist das Modellierungsniveau (Tabelle 1): die räumliche Auflösung, die Zeitskala und die relevanten Freiheitsgrade.

Die Wahl des Modellierungsniveaus hängt von der Art des jeweiligen biomolekularen Vorgangs ab. In diesem Aufsatz besprechen wir drei der vier biomolekularen Vorgänge, die in Abbildung 1 dargestellt sind: 1) Faltung von Polypeptiden, 2) Bildung von Komplexen (Protein-Ligand-, DNA-Ligand-, Protein-DNA-Komplexe etc.), 3) Verteilung von Molekülen in unterschiedlichen Milieus wie Lipidmembranen, Wasser, Gemischen (z. B. Wasser/Harnstoff, Elektrolytlösungen) und apolaren Lösungsmitteln und 4) Bildung von Lipidmembranen oder Micellen aus Mischungen ihrer Komponenten. Diese vier Prozesse spielen eine grundlegende Rolle für das Verhalten von biomolekularen Systemen. Ihnen ist gemeinsam, dass sie durch schwache, nichtbindende Wechselwirkungen zwischen Atomen gelenkt werden, die die thermodynamischen Eigenschaften der kondensierten Phase bestimmen, in der die Prozesse ablaufen. Deshalb sollte die Modellierung dieser Vorgänge auf atomarer oder molekularer Ebene (dritte Reihe in Tabelle 1) am ehesten zum Erfolg führen. Da der relevante Temperaturbereich zwischen Raumtemperatur und physiologischen Temperaturen liegt und Energien zwischen 1 und  $10 k_B T$  beteiligt sind (entsprechend einigen dutzend  $\text{kJ mol}^{-1}$ ,  $k_B$  ist die Boltzmann-Konstante), folgen die Prozesse weitgehend den Gesetzen der klassischen statistischen Mechanik. Obwohl für die Wechselwirkung zwischen den Elektronen der Atome und Mole-

## Aus dem Inhalt

<b>1. Einleitung</b>	4169
<b>2. Das Kraftfeldproblem</b>	4171
<b>3. Das Suchproblem</b>	4178
<b>4. Das Ensembleproblem</b>	4185
<b>5. Das Experimentalproblem</b>	4188
<b>6. Perspektiven in der biomolekularen Modellierung</b>	4193

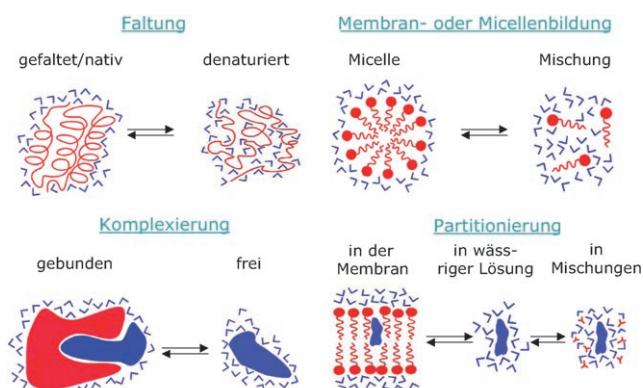
küle und für die Bewegung leichter Teilchen (wie Protonen) quantenmechanische Gesetze gelten, können nichtbindende Wechselwirkungen sehr gut durch eine klassische Potentialenergiefunktion oder ein Kraftfeld als Teil eines klassischen Hamilton-Operators des betreffenden Systems beschrieben werden.<sup>[1]</sup>

Abbildung 2 zeigt die vier Möglichkeiten bei der Auswahl des Modellierungsniveaus. Es wird festgelegt, 1) welche atomaren oder molekularen Freiheitsgrade explizit im Modell berücksichtigt werden, 2) welche Wechselwirkungsfunktion oder welches Kraftfeld verwendet wird, um die Energie des Systems als Funktion der gewählten Freiheitsgrade zu beschreiben, 3) wie der Konfigurationsraum bezüglich der (oft zahlreichen) Freiheitsgrade zu durchmustern ist und 4) wie die räumlichen Systemgrenzen und externen Kräfte modelliert werden. Wir betrachten in erster Linie atomare und molekulare Freiheitsgrade, wobei wir die zugehörigen klas-

[\*] Prof. Dr. W. F. van Gunsteren, Dr. D. Bakowies, R. Baron, Dr. I. Chandrasekhar, M. Christen, Prof. Dr. X. Daura, Dr. P. Gee, D. P. Geerke, Dr. A. Glättli, Dr. P. H. Hünenberger, M. A. Kastenholz, Dr. C. Oostenbrink, Dr. M. Schenk, D. Trzesniak, Dr. N. F. A. van der Vegt, Dr. H. B. Yu  
 Laboratorium für Physikalische Chemie  
 Eidgenössische Technische Hochschule  
 8093 Zürich (Schweiz)  
 Fax: (+41) 44-632-1039  
 E-mail: wfvgn@igc.phys.chem.ethz.ch  
 Prof. Dr. X. Daura  
 ICREA, Institute of Biotechnology and Biomedicine  
 Universitat Autònoma de Barcelona  
 08193 Bellaterra, Barcelona (Spanien)  
 Dr. C. Oostenbrink  
 Pharmaceutical Sciences/Pharmacochemistry  
 Vrije Universiteit  
 De Boelelaan 1083 P262, 1081 HV Amsterdam (Niederlande)  
 Dr. N. F. A. van der Vegt  
 Max-Planck-Institut für Polymerforschung  
 Ackermannweg 10, 55128 Mainz (Deutschland)  
 Dr. H. B. Yu  
 Department of Chemistry  
 University of Wisconsin  
 1101 University Ave, Madison, WI 53706 (USA)

**Tabelle 1:** Beispiele für Modellierungsniveaus in der computergestützten Biochemie und molekularen Biologie.

Methoden	Freiheitsgrade	Eigenschaften, Vorgänge	Zeitskala
Quantendynamik	Atome, Kerne, Elektronen	angeregte Zustände, Relaxation, Reaktionsdynamik	Pikosekunden
Quantenmechanik (Ab-initio-, Dichtefunktional-, semiempirische, Valence-Bond-Methoden)	Atome, Kerne, Elektronen	Grund- und angeregte Zustände, Reaktionsmechanismus	keine Zeitskala
klassische statistische Mechanik (MD, MC, Kraftfelder)	Atome, Lösungsmittel	Ensemble, Durchschnittswerte, Systemeigenschaften, Faltung	Nanosekunden
statistische Methoden (Datenbankanalyse)	Gruppen von Atomen, Aminosäurereste, Basen	Strukturhomologie und -ähnlichkeit	keine Zeitskala
Kontinuumsmethoden (hydrodynamische und elektrostatische)	elektrisches Kontinuum, Fließkontinuum etc.	rheologische Eigenschaften	supramolekular
kinetische Gleichungen	Besetzungs niveaus	Besetzungs dynamik, Signaltransduktion	makroskopisch



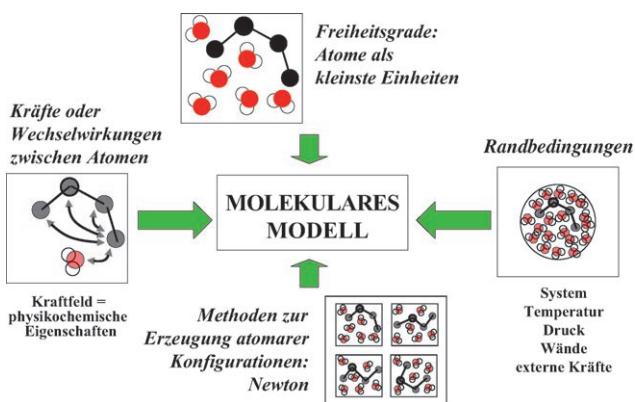
**Abbildung 1.** Vier biomolekulare Prozesse, die von thermodynamischen Gleichgewichten abhängen.



Wilfred F. van Gunsteren wurde 1947 in Wassenaar (Niederlande) geboren. 1968 erwarb er an der Freien Universität Amsterdam einen B.Sc. in Physik, 1974 einen „Meester“ in Jura und 1976 einen Ph.D. in Kernphysik. Nach Postdoc-Aufenthalten an der Universität Groningen und der Harvard University war er von 1980–1987 Senior Lecturer und bis August 1990 Professor für physikalische Chemie an der Universität Groningen. 1990 nahm er einen Ruf auf einen Lehrstuhl für Computerchemie an der ETH Zürich an. Er erhielt eine Goldmedaille für Forschung der Königlichen Chemischen Gesellschaft der Niederlande. Sein Hauptforschungsinteresse gilt den physikalischen Grundlagen der Struktur und Funktion von Biomolekülen.

sischen Kraftfelder und die klassische Newtonsche Dynamik zur Durchmusterung des Konfigurationsraums heranziehen. So können Systeme mit bis zu  $10^5$  oder  $10^6$  Atomen oder Teilchen behandelt werden, was noch immer sehr klein gegenüber der Avogadroschen Zahl ist, die als Größenordnung für makroskopische Systeme dienen kann. Bei sehr kleinen Systemen hat die Modellierung der Systemgrenze oder der Oberfläche einen großen Einfluss auf die berechneten Eigenschaften. Oberflächeneffekte lassen sich mithilfe periodischer Randbedingungen minimieren; in diesem Fall ist der Kasten, der das Molekülsystem enthält, von unendlich vielen identischen Kopien umgeben (Abbildung 3). Auf diese Weise werden zwar Oberflächeneffekte vermieden, dafür treten aber Periodizitätsartefakte auf.<sup>[2–5]</sup>

Moderne biomolekulare Modellierungsverfahren sind durch die in Tabelle 2 angeführten Probleme eingeschränkt:



**Abbildung 2.** Vier grundlegende Auswahlmöglichkeiten zur Festlegung des in der Molekülsimulation verwendeten Modells.

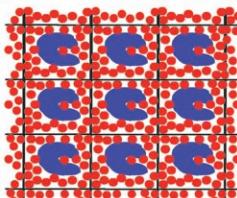
### Vakuum

- Oberflächeneffekte (Oberflächenspannung)
- keine dielektrische Abschirmung

### Tröpfchen

- weiterhin Oberflächeneffekte (an der Wasser-Vakuum-Grenzfläche)
- nur teilweise dielektrische Abschirmung
- Verdampfung des Lösungsmittels

### periodisch: Vervielfältigung des Systems mit identischen Abbildern



- Vorteil: keine Oberflächeneffekte

### Nachteil:

- künstliche Periodizität
- hohe effektive Konzentration

**Abbildung 3.** Drei Arten von räumlichen Randbedingungen, die bei molekularen Simulationen verwendet werden, und ihre Charakteristika.

**Tabelle 2:** Vier grundlegende Probleme in der biomolekularen Modellierung.

<b>1. Kraftfeldproblem</b>	A) sehr kleine (freie) Energiedifferenzen, viele Wechselwirkungen B) entropische Einflüsse C) unterschiedliche Atom- und Molekültypen
<b>2. Suchproblem</b>	A) Konvergenz B) vereinfachende Faktoren C) erschwerende Faktoren
<b>3. Ensembleproblem</b>	A) Entropie B) Mittelungen C) nichtlineare Mittelungen
<b>4. Experimentalproblem</b>	A) Mittelungen B) ungenügende Zahl von Daten C) ungenügende Genauigkeit der Daten

1) das Kraftfeldproblem, 2) das Suchproblem, 3) das Ensembleproblem und 4) das Experimentalproblem. Diese vier Probleme sind das zentrale Thema dieses Aufsatzes und werden in den Abschnitten 2–5 anhand von Arbeiten unserer Gruppe diskutiert. Wir möchten betonen, dass es nicht Ziel dieses Aufsatzes ist, die Beiträge der vielen auf diesem Gebiet forschenden Arbeitsgruppen zusammenzufassen.

Der entscheidende Grund, weshalb Computersimulationen trotz der oben genannten Einschränkungen zur Untersuchung von biomolekularen Systemen eingesetzt werden, ist in Tabelle 3 an vierter Stelle genannt: Nur wenige Eigen-

putersimulationen von unschätzbarem Wert für die Interpretation experimenteller Daten. Sie lassen sich darüber hinaus zur Vorhersage von Eigenschaften in einer Umgebung verwenden, die nur schwer oder mit großem Kostenaufwand realisierbar ist. In den folgenden vier Abschnitten illustrieren wir Einsatzmöglichkeiten, Leistungsvermögen und Grenzen biomolekularer Modellierungsverfahren in Verbindung mit experimentellen Ansätzen bezüglich der in Abbildung 1 gezeigten Prozesse.

## 2. Das Kraftfeldproblem

Ein biomolekulares Kraftfeld besteht im Allgemeinen aus Potentialenergietermen, die zum einen kovalente Wechselwirkungen zwischen Atomen beschreiben (z.B. Bindungsdehnungen, Änderungen von Bindungs- und Diederwinkeln), und zum anderen nichtkovalente Wechselwirkungen zwischen Atomen in unterschiedlichen Molekülen und zwischen Atomen in einem Molekül, die durch mehr als zwei oder drei kovalente Bindungen voneinander getrennt sind.<sup>[6,7]</sup>

Da uns thermodynamische Gleichgewichte und Prozesse interessieren, die von nichtbindenden Wechselwirkungen bestimmt werden (Abbildung 1), befassen wie uns in erster Linie mit der Formulierung und Parametrisierung dieser Potentialenergieterme. Bei der Entwicklung von Kraftfeldern sind vor allem drei Punkte zu beachten (Tabelle 2, Punkt 1, A–C):

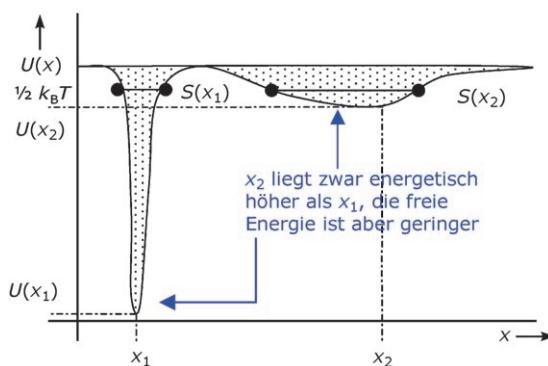
Ein erstes Problem besteht darin, dass den in Abbildung 1 dargestellten Vorgängen (freie) Energiedifferenzen zwischen 1 und  $10 k_B T$  zugrundeliegen (entsprechend einigen Dutzend  $\text{kJ mol}^{-1}$ ). Diese verhältnismäßig kleinen Energien resultieren aus einer Summierung über sehr viele ( $10^6$ – $10^8$ ) Atompaares: Ein System aus  $N=1000$  Atomen enthält etwa  $\frac{1}{2}N(N-1)=500\,000$  Atompaares, die zu den nichtbindenden Wechselwirkungen beitragen. Um die Gesamtenergie der nichtbindenden Wechselwirkungen mit der erforderlichen Genauigkeit zu erhalten, müssen die Einzelerme – die Energien der Atompaares – noch genauer bekannt sein. Das Problem wird noch gravierender, wenn man versucht, ein Kraftfeld hoher Genauigkeit für größere Systeme, d.h. größere Teilchenzahlen  $N$ , abzuleiten.

Das zweite Problem betrifft die Behandlung entropischer Effekte. Da uns biomolekulare Systeme am absoluten Nullpunkt ( $-273.15^\circ\text{C}$ ) nicht interessieren, müssen wir den entropischen Beitrag  $S$  zur freien Energie  $F=U-TS$  eines betreffenden Systems berücksichtigen. Es ist bekannt, dass die Entropie eine entscheidende Rolle bei allen in Abbildung 1 dargestellten Prozessen spielt. Änderungen der freien Energie, die biomolekulare Prozesse antreiben, können von Änderungen der inneren Energie ( $U$ ) oder der Entropie ( $S$ ) herrühren, die entweder zusammenwirken oder sich teilweise kompensieren können, je nach der relativen Stärke der nichtbindenden Wechselwirkungen zwischen den Komponenten (Atomen, Molekülen) des Systems.<sup>[8,9]</sup> Abbildung 4 veranschaulicht das Phänomen der Energie-Entropie-Kompensation: Für zwei Konformationen  $x_1$  und  $x_2$  eines Moleküls kann die Beziehung  $U(x_1) \ll U(x_2)$  gelten und dennoch  $F(x_1) > F(x_2)$  resultieren, wenn bei einer bestimmten Tempe-

**Tabelle 3:** Vier Gründe für den Einsatz von Computersimulationen in der Forschung.

Simulationen können ein Experiment ersetzen oder ergänzen:	
1. Experiment ist unmöglich	Kollision von Sternen oder Galaxien Wettervorhersage
2. Experiment ist gefährlich	Flugsimulation Explosionssimulation
3. Experiment ist teuer	Simulation von Hochdruck Windkanalsimulation
4. Experiment ist blind	viele Eigenschaften können auf sehr kurzen Zeit- und sehr kleinen Raumskalen nicht beobachtet werden.

schaften eines biomolekularen Systems können experimentell erfasst werden, wohingegen eine Computersimulation nicht nur Durchschnittswerte, sondern auch Verteilungen und Zeitreihen jeder definierbaren Größe liefert. Computersimulationen ergänzen daher ein Experiment, indem sie detaillierte Konformations- und andere Verteilungen liefern, die die experimentell ermittelbaren räumlichen und zeitlichen Durchschnittswerte festlegen. In dieser Funktion sind Com-



**Abbildung 4.** Energie-Entropie-Kompensation bei endlichen Temperaturen.

ratur  $S(x_1) \ll S(x_2)$  gilt. Die Entropie ist ein Maß für die Größe des Konformationsraumes ( $x$ ), der für das molekulare System bei der Temperatur  $T$  zugänglich ist.

Abbildung 4 macht außerdem deutlich, dass die Suche nach dem globalen Energieminimum eines biomolekularen Systems bedeutungslos wird, wenn die Entropie des Systems signifikant zur freien Energie beiträgt. Beispielsweise erhält man für flüssiges Wasser bei Raumtemperatur und Normaldruck die Werte  $F = -24 \text{ kJ mol}^{-1}$ ,  $U = -41 \text{ kJ mol}^{-1}$  und  $TS = -17 \text{ kJ mol}^{-1}$ . Deswegen können die Eigenschaften von flüssigem Wasser in der kondensierten Phase nur durch eine Konformationsverteilung beschrieben werden, die sich wiederum durch eine Computersimulation erzeugen lässt. Ähnliche Überlegungen gelten für biomolekulare Systeme: Eine energieminierte Struktur eines Proteins entspricht einer möglichen Konformation bei 0 K und gibt keinerlei Auskunft über die Konformationsverteilung des Proteins bei physiologischen Temperaturen. Dieser Sachverhalt hat Konsequenzen für die Kraftfeldentwicklung: Wenn ein Kraftfeld bei Temperaturen über 0 K in einer Computersimulation verwendet wird, dann sollten bei der Herleitung oder Kalibrierung seiner Parameter aus Konsistenzgründen entropische Effekte berücksichtigt werden. In anderen Worten muss die Kalibrierung von Kraftfeldparametern auch Computersimulationen zur Erzeugung von Konfigurationsensembles einschließen; dadurch wird die Entwicklung aufwändiger, als wenn nur einzelne energieminierte Konformationen oder gemessene Durchschnittsstrukturen verwendet würden.

Ein drittes Problem bei der Entwicklung biomolekularer Kraftfelder ist die enorme Vielfalt von chemischen Verbindungen, für die es adäquate Kraftfeldparameter abzuleiten gilt. Das Problem kann zumindest teilweise entschärft werden, wenn die Kraftfeldparameter zu einem gewissen Grad zwischen Atomen oder Atomgruppen in unterschiedlichen Molekülen übertragbar sind. Dies gelingt im Allgemeinen recht gut, wenn man die Kraftfeldterme auf einer physikalischen statt einer rein statistischen Basis entwickelt, einfach hält und lokal einschränkt. Darüber hinaus macht eine möglichst einfache Berechnung die biomolekulare Simulation effizienter, was die Durchmusterung des Konformationsraumes erleichtert.

## 2.1. Funktionale Form der Kraftfeldterme

Die meisten biomolekularen Kraftfelder setzen sich aus Termen mit relativ einfacher funktionaler Form zusammen.<sup>[6]</sup> Ein Beispiel hierfür sind die Terme des GROMOS-Kraftfelds [Gl. (1)–(8)].<sup>[7,10]</sup>

$$V^{\text{Bindung}}(\mathbf{r}; K_b, b_0) = \sum_{n=1}^{N_b} 1/4 K_{b_n} [b_n^2 - b_{0_n}^2]^2 \quad (1)$$

$$V^{\text{Winkel}}(\mathbf{r}; K_\theta, \theta_0) = \sum_{n=1}^{N_\theta} 1/2 K_{\theta_n} [\cos(\theta_n) - \cos(\theta_{0_n})]^2 \quad (2)$$

$$V^{\text{har}}(\mathbf{r}; K_\xi, \xi_0) = \sum_{n=1}^{N_\xi} 1/2 K_{\xi_n} [\xi_n - \xi_{0_n}]^2 \quad (3)$$

$$V^{\text{trig}}(\mathbf{r}; K_\varphi, \delta, m) = \sum_{n=1}^{N_\varphi} K_{\varphi_n} [1 + \cos(\delta_n) \cos(m_n \varphi_n)] \quad (4)$$

$$V^{\text{LJ}}(\mathbf{r}; C_{12}, C_6) = \sum_{\text{Paare } i,j} \left[ \frac{C_{12}(i,j)}{r_{ij}^{12}} - \frac{C_6(i,j)}{r_{ij}^6} \right] \quad (5)$$

$$V^{\text{C}}(\mathbf{r}; q) = \sum_{\text{Paare } i,j} \frac{q_i q_j}{4\pi\epsilon_0\epsilon_1} \frac{1}{r_{ij}} \quad (6)$$

$$V^{\text{RF}}(\mathbf{r}; q) = \sum_{\text{Paare } i,j} \frac{q_i q_j}{4\pi\epsilon_0\epsilon_1} \frac{(-\frac{1}{2} C_{rf} r_{ij}^2)}{R_{rf}^3} \quad (7)$$

$$V^{\text{RF}_c}(\mathbf{r}; q) = \sum_{\text{Paare } i,j} \frac{q_i q_j}{4\pi\epsilon_0\epsilon_1} \frac{(\frac{1}{2} C_{rf} - 1)}{R_{rf}} \quad (8)$$

Die ersten vier Terme beschreiben die vier schon erwähnten Arten von kovalenten (bindenden) Wechselwirkungen, die letzten vier spezifizieren die nichtbindenden Wechselwirkungen: die Van-der-Waals-Wechselwirkung in Form eines Lennard-Jones-Terms, die elektrostatische Coulomb-Wechselwirkung zwischen Atomladungen (Partialladungen)  $q_i$  sowie die entfernungsabhängigen und entfernungsunabhängigen (konstanten) Wechselwirkungen aufgrund des dipolaren Reaktionsfeldes (RF), das durch die Ladungsverteilung innerhalb der Schnittkugel über das kontinuierliche dielektrische Medium außerhalb der Kugel induziert wird. Dieses Kraftfeld erfasst viele unterschiedliche Molekültypen (unter anderem Polypeptide, Polysaccharide, Nucleinsäuren, Lipide) und enthält deshalb einen umfangreichen Parametersatz:<sup>[7]</sup> 52 Typen von Bindungen [Gl. (1)], 54 Typen von Bindungswinkeln [Gl. (2)], 3 Typen von uneigentlichen (harmonischen) Diederwinkeln [Gl. (3)], 41 Typen von eigentlichen torsionalen (trigonometrischen) Diederwinkeln [Gl. (4)], Van-der-Waals-Parameter von 53 Atomtypen [Gl. (5)] und viele verschiedene Sätze von Atomladungen typischer polarer oder geladener Atomgruppen der oben genannten Moleküle [Gl. (6)–(8)].<sup>[7,10]</sup>

Die funktionale Form wird so gewählt, dass sie leicht zu berechnen ist. Die Terme der nichtbindenden Wechselwirkungen sind ausschließlich Paarterme, und die komplexeren Dreie- und Vierkörperterme der kovalenten Wechselwirkungen [Gl. (3) und (4)] treten sehr viel seltener auf als die nichtbindenden Paarterme. Das Lösungsmittel wird aus-

schließlich durch nichtbindende Terme beschrieben, wobei die intramolekularen Freiheitsgrade der Lösungsmittelmoleküle eingefroren werden. Der Hauptrechenaufwand liegt somit in der Berechnung der nichtbindenden Wechselwirkungen.

## 2.2. Kalibrierung von Kraftfeldparametern

Nachdem die funktionale Form der Wechselwirkungsterme spezifiziert worden ist, verbleibt die anspruchsvolle Aufgabe, konsistente Werte für Hunderte von Kraftfeldparametern zu ermitteln. Diese Aufgabe umfasst die Auswahl des Datentyps, des Systems, der thermodynamischen Phase und der Eigenschaften, die als Kalibrierungssatz für spezifische Kraftfeldparameter dienen. Tabelle 4 zeigt die Auswahl für das GROMOS-Kraftfeld. Da biomolekulare Systeme im Allgemeinen in der kondensierten Phase vorliegen, stützt man sich möglichst auf (experimentelle wie theoretische) Daten für diese Phase. Um Parameter zwischen Atomgruppen in unterschiedlichen Molekülen möglichst gut übertragen zu können, werden außerdem nur Daten für kleine Moleküle verwendet. Bei größeren Molekülen wie Proteinen (z.B. aus der Proteindatenbank) kann die spezielle Umgebung des gefalteten Moleküls die Eigenschaften von Atomgruppen beeinflussen. Davon abgesehen wurden die Strukturen in der Proteindatenbank bei sehr unterschiedlichen thermodynamischen Bedingungen bestimmt (pH-Wert, Ionenstärke etc.). Zuletzt ist noch zu bedenken, dass bestimmte Eigenschaften mit manchen Kraftfeldparametern sehr eng, mit anderen aber nur diffus zusammenhängen. Man kann daher den Aufwand für die Kalibrierung verringern, indem man spezifische Teilsätze von Parametern separat bezüglich eines begrenzten Satzes von Eigenschaften optimiert.

Im Folgenden erläutern wir die Strategie bei der Entwicklung des GROMOS-Kraftfeldes (Tabelle 4).<sup>[7]</sup> Die geometrischen Parameter für die Terme der kovalenten Wechselwirkung wurden aus Kristallstrukturen kleiner Moleküle erhalten; die entsprechenden Schwingungskraftkonstanten stammten aus IR-spektroskopischen Daten kleiner Moleküle in der Gasphase, und die in den Termen für nichtbindende Wechselwirkungen enthaltenen Parameter  $C_{12}$ ,  $C_6$  und  $q$  wurden durch Kalibrierung von Moleküldynamik(MD)-Simulationen mit experimentellen Verdampfungswärmen, Dichten reiner Flüssigkeiten und freie Solvatationsenergien kleiner Moleküle in polaren und apolaren Lösungsmitteln erhalten. Dielektrische Permittivitäten und Diffusionseigenschaften von Flüssigkeiten wurden bei diesen Parametrisierungen als Sekundärdaten genutzt. Elektronendichten aus quantenmechanischen Rechnungen wurden nur verwendet, um einen ersten Schätzwert für atomare Partialladungen zu erhalten, da diese aufgrund von Polarisierungseffekten stark von der Umgebung (Gasphase oder kondensierte Phase) abhängen können. Torsionswinkelparameter wurden durch Angleichen von Torsionsenergieprofilen an quantenmechanische Daten abgeleitet, wobei der Parametersatz für nichtbindende Wechselwirkungen unverändert blieb.

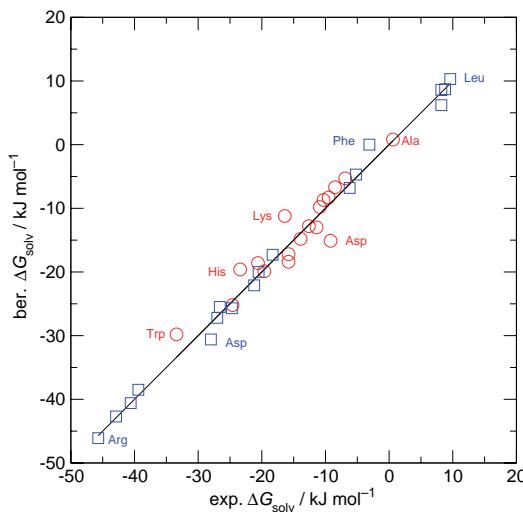
Ein biomolekulares Kraftfeld definiert im Prinzip jeweils einen konsistenten Satz von Parametern sowohl für die gelösten Moleküle (Proteine, Lipide, Saccharide, Nucleotide) als auch für die Lösungsmittelmoleküle (Wasser, Alkohole, Dimethylsulfoxid, Chloroform etc.). Die Verwendung eines Teilsatzes von Parametern anderer Kraftfelder oder Modelle könnte zu Abweichungen führen.

Um Peptid- oder Proteinfaltungen simulieren zu können (Abbildung 1), sollten die aus der Simulation erhaltenen Differenzen der relativen freien Solvatationsenergien der 20 Aminosäurereste, wie sie in einem polaren (Wasser) oder in

**Tabelle 4:** Auswahl der Kalibrierungssätze von Daten, Systemen, Eigenschaften und thermodynamischer Phase zur Herleitung der biomolekularen Parameter des GROMOS-Kraftfeldes.<sup>[7]</sup>

Datentyp	Systemtyp	Phase	Art der Eigenschaften	Kraftfeldparameter
Strukturdaten (exp.)	kleine Moleküle	kristalline Festphase	Molekülstruktur: Bindungslängen, -winkel	$b_0$ , $\theta_0$ , $\xi_0$
spektroskopische Daten (exp.)	kleine Moleküle	Gasphase	Molekülschwingungen: Kraftkonstanten	$K_b$ , $K_\theta$ , $K_\xi$
thermodynamische Daten (exp.)	kleine Moleküle, Mischungen, Lösungen	kondensierte Phase	Verdampfungswärme, Dichte, Verteilungskoeffizienten, freie Solvatationsenthalpie	van der Waals: $C_{12}(i,j)$ , $C_6(i,j)$ , $q_i$ (endgültig)
Dielektrizitätsdaten (exp.)	kleine Moleküle	kondensierte Phase	dielektrische Permittivität, Relaxation	Ladungen $q_i$
Transporteigenschaften (exp.)	kleine Moleküle	kondensierte Phase	Diffusions- und Viskositätskoeffizienten	$C_{12}(i,j)$ , $C_6(i,j)$ , $q_i$
Elektronendichte (theor.)	kleine Moleküle	Gasphase	quantenchemisch berechnete Atomladungen	Ladungen $q_i$ (provisorisch)
Energieprofil (theor.)	kleine Moleküle	Gasphase	quantenchemisch berechnete Torsionswinkel und Rotationsprofile	$K_\phi$ , $\delta$ , $m$

einem unpolaren Lösungsmittel (Cyclohexan) auftreten, gut mit den entsprechenden experimentellen Daten übereinstimmen, da diese Differenzen die wichtigste Triebkraft der Faltung sind. Mit dem GROMOS-53A6-Kraftfeld<sup>[7]</sup> berechnete freie Solvatationsenthalpien sind in Abbildung 5 dargestellt. Bei 18 Aminosäureseitenketten (Gly und Pro fehlen) beträgt die mittlere absolute Abweichung vom Experiment 1.0 kJ mol<sup>-1</sup> in Wasser und 2.0 kJ mol<sup>-1</sup> in Cyclohexan. Da beide Werte kleiner sind als  $k_B T$ , eignet sich das GROMOS-53A6-Kraftfeld zur Untersuchung von Proteinfaltungen.



**Abbildung 5.** Vergleich von berechneten (MD-Simulationen mit dem GROMOS-53A6-Kraftfeld) und experimentellen freien Solvatationsenthalpien von 18 Aminosäureanaloga (Gly und Pro fehlen) in Cyclohexan (Kreise) und in Wasser (Quadrate).<sup>[7]</sup>

### 2.3. Langreichweite Kräfte

Elektrostatische Wechselwirkungen spielen eine wichtige Rolle in biomolekularen Systemen. Im Unterschied zu kovalenten und Van-der-Waals-Wechselwirkungen werden die elektrostatischen Wechselwirkungen zwischen Molekülen oder Molekülteilen mit wachsendem Abstand  $r$  nur langsam schwächer:

1. Die Wechselwirkungsenergie zwischen zwei geladenen Molekülen ist proportional zu  $r^{-1}$ , während die entsprechende Kraft, die (negative) Ableitung der Energie nach dem Abstand, proportional zu  $r^{-2}$  ist.
2. Die Wechselwirkungsenergie zwischen einem neutralen Molekül mit einem dipolaren oder geladenen Molekül ist proportional zu  $r^{-2}$ , die entsprechende Kraft proportional zu  $r^{-3}$ .
3. Die Wechselwirkungsenergie zwischen zwei neutralen Molekülen mit Dipolmoment ist proportional zu  $r^{-3}$ , während die entsprechende Kraft proportional zu  $r^{-4}$  ist.

Setzt man diese Multipolentwicklung für Quadrupolmomente, Octupolmomente etc. fort, dann zeigt sich, dass sogar die Wechselwirkung zwischen zwei dipolfreien Neutralmolekülen mit Quadrupolmomenten eine längere Reichweite hat (proportional zu  $r^{-5}$ ) als die Van-der-Waals-Dispersions-

wechselwirkung, die proportional zu  $r^{-6}$  ist. Wenn wir die elektrostatische Energie einer einzelnen Ladung, eines einzelnen Dipols oder eines einzelnen Quadrupols mit allen umgebenden Ladungen, Dipolen und Quadrupolen betrachten, dann müssen wir die elektrostatische Wechselwirkung,  $V^{el}(r) 4\pi r^2$ , von  $r$  bis unendlich integrieren. Dabei ist  $4\pi r^2 dr$  das Volumen der Kugelschale zwischen  $r$  und  $r+dr$  um die zentrale Ladung, den zentralen Dipol oder den zentralen Quadrupol [Gl. (9)].

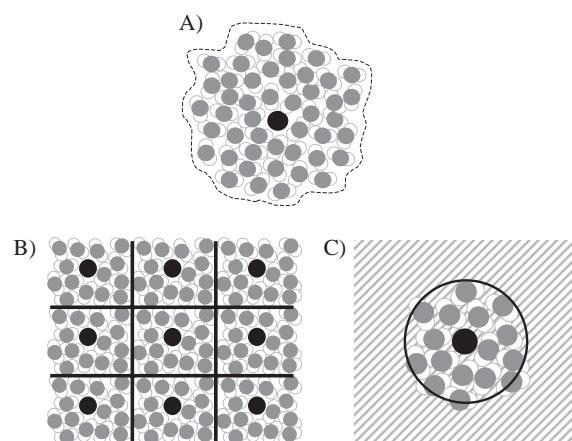
$$\int_0^\infty V^{el}(r) 4\pi r^2 dr \quad (9)$$

Allerdings konvergiert das Integral (9) nur unter der Bedingung in Gleichung (10):

$$V^{el}(r) \sim r^{-n}, n > 3 \quad (10)$$

Dies bedeutet insbesondere, dass die elektrostatische Gesamtenergie ionischer Systeme von den räumlichen Randbedingungen abhängt, die den Bereich des Integrals (9) bei konkreten Rechnungen einschränken. Da die Paarenergien und -kräfte mit dem interatomaren Abstand  $r$  langsam abnehmen, hängen die Ergebnisse von Simulationen davon ab, wie langreichweitige Wechselwirkungen in den Kraft- und Energierechnungen behandelt werden.

Zur Berechnung langreichweiter (elektrostatischer) Wechselwirkungen in biomolekularen Systemen kommen vor allem zwei Techniken zum Einsatz (Abbildung 6). Bei den Gittersummenmethoden wird das System in einen Kasten bestimmter Form gebracht (kubisch, rechteckig, triklin, abgestumpft-oktaedrisch) und von einer unendlichen Zahl identischer Kopien umgeben. Auf diese Weise verlagert man das Problem der Systemgrenzen ins Unendliche, beseitigt es jedoch leider nicht. Hinzu kommt, dass dem System eine künstliche Periodizität aufgezwungen wird. Zu den Gittersummenmethoden gehören die Ewald-Summierung,<sup>[12]</sup> die



**Abbildung 6.** Zwei Methoden zur Berechnung von langreichweiten elektrostatischen Energien und Kräften eines molekularen Systems. A) Reales System mit expliziter Behandlung des Lösungsmittels; B) periodische Anordnung des Systems, wie bei den Ewald-, P<sup>3</sup>M- und PME-Methoden; C) Näherung eines dielektrischen Kontinuums jenseits eines Cutoff-Abstandes.

$P^3M$ -Methode (Particle-Particle/Particle-Mesh; Teilchen-Teilchen/Teilchen-Sieb-Methode)<sup>[13]</sup> und die PME-Methode (Particle-Mesh-Ewald; Teilchen-Sieb-Ewald).<sup>[14]</sup> Bei einem anderen Ansatz wird das Medium jenseits eines Cutoff-Abstandes  $R_{rf}$  von einem spezifischen Atom oder Molekül durch ein dielektrisches Kontinuum einheitlicher Permittivität  $\epsilon_{rf}$  und Ionenstärke  $I_{rf}$  angenähert.<sup>[15,16]</sup> Ein solches dielektrisches Kontinuum erzeugt ein Reaktionsfeld, das der Ladungsverteilung innerhalb der Cutoff-Kugel mit Radius  $R_{rf}$  entspricht und für jedes Atom oder Molekül leicht berechnet werden kann [siehe Gl. (7) und (8)]; die Konstante  $C_{rf}$  ist abhängig von  $R_{rf}$ ,  $\epsilon_{rf}$  und  $I_{rf}$ .<sup>[10,16]</sup>

Beide Methoden sind Näherungen, wenn auch unterschiedlicher Art. Die Reaktionsfeldmethode ist eine Molekularfeldnäherung für die reale Ladungsverteilung jenseits eines Abstandes  $R_{rf}$ . Sie beschreibt das dielektrische Kontinuum als kugelsymmetrisch und führt keine künstliche Periodizität ein. Gittersummenmethoden kommen ohne Mittelwertbildung aus, behandeln aber Wechselwirkungen jenseits des Kastens als periodisch. Beide Näherungen sind in Simulationen verschiedener Systeme getestet worden.<sup>[3–5,17–19]</sup>

Tabelle 5 veranschaulicht, wie unterschiedliche Näherungsmethoden zur Beschreibung der langreichweitigen

elektrostatischen Wechselwirkungen die freie Hydratationsenergie von ionischen Verbindungen in Wasser beeinflussen.<sup>[20]</sup> Die Ergebnisse belegen, dass die Methode, mit der langreichweite Kräfte und ihre Parameter (z.B.  $R_{rf}$ ) gehandhabt werden, bei der Parametrisierung eines Kraftfeldes von größter Bedeutung ist. Beispielsweise hat man die nichtbindenden Parameter des OPLS-Kraftfeldes<sup>[28,29]</sup> gewöhnlich aus Rechnungen mit Cutoff-Radien  $R_c = 0.95\text{--}1.5\text{ nm}$  erhalten und die langreichweitigen Van-der-Waals-Wechselwirkungen durch Korrekturformeln berücksichtigt (siehe z.B. Lit. [30]). Das GROMOS-Kraftfeld wurde mit  $R_c = 1.4\text{ nm}$  und einer Reaktionsfeldkraft kalibriert.

#### 2.4. Erprobung biomolekularer Kraftfelder

Ein biomolekulares Kraftfeld wird zunächst durch Kalibrierung seiner Parameter so eingestellt, dass es die Eigenschaften kleiner Moleküle reproduziert. Die nächste Aufgabe besteht in der Erprobung des Kraftfeldes an größeren Biomolekülen in kondensierter Phase. Zu den Tests sollten Proteine, Saccharide oder Nucleotide in wässriger Lösung herangezogen werden, und simulierte Eigenschaften sollten mit

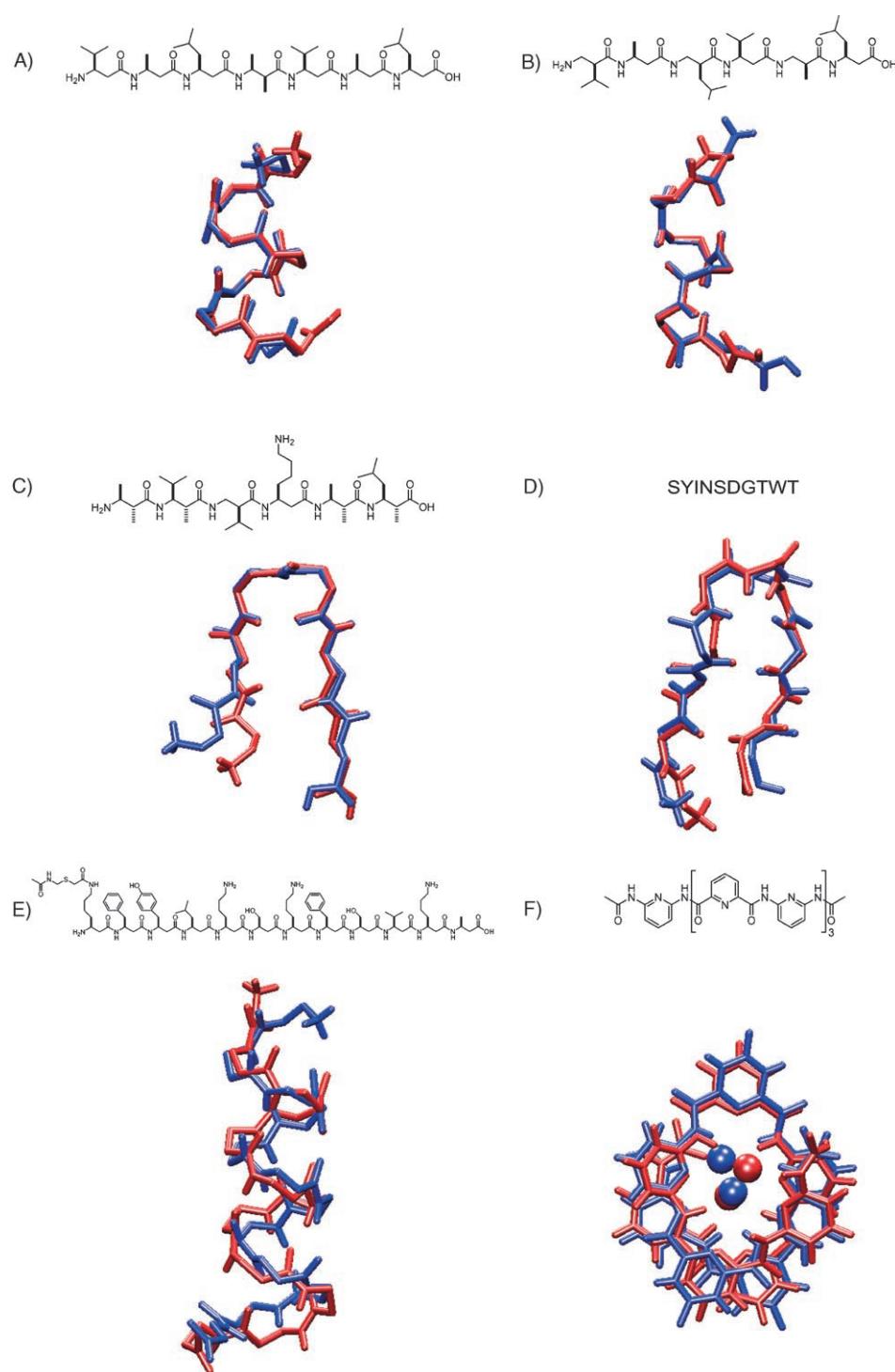
**Tabelle 5:** Berechnung von methodenunabhängigen freien Hydratationsenergien aus Moleküldynamiksimulationen mit explizit berücksichtigtem Lösungsmittel.<sup>[a]</sup>

Methode	$N_w$	$R_{rf}$	$\Delta F_{sim}$	$\Delta F_{ncb}$	$\Delta F_{per}$	$\Delta F_{sum}$	$\Delta F_{srf}$	$\Delta F_{hyd}^0$
$P^3M$	4	—	−45.8	0.0	−293.2	−61.9	−3.54	−390.82
$P^3M$	8	—	−118.7	0.0	−259.7	−69.8	−1.97	−436.62
$P^3M$	16	—	−164.8	0.0	−221.5	−74.5	−1.04	−448.15
$P^3M$	32	—	−209.9	0.0	−183.9	−77.1	−0.54	−457.67
$P^3M$	64	—	−249.0	0.0	−150.0	−78.4	−0.27	−464.02
$P^3M$	128	—	−279.8	0.0	−121.1	−79.1	−0.14	−466.45
$P^3M$	256	—	−303.8	0.0	−97.1	−79.4	−0.07	−466.71
$P^3M$	512	—	−324.8	0.0	−77.6	−79.6	−0.03	−468.35
$P^3M$	1024	—	−340.5	0.0	−61.8	−79.7	−0.02	−468.35
$P^3M$	2048	—	−353.5	0.0	−49.2	−79.7	−0.01	−468.73
RF	512	0.8	−275.9	−128.1	−1.47	−76.8	0.0	−468.55
RF	512	1.0	−298.6	−102.7	−3.76	−77.4	0.0	−468.77
RF	512	1.2	−311.0	−85.7	−6.76	−77.6	0.0	−467.49
RF	1024	0.8	−278.4	−128.1	−0.38	−76.8	0.0	−469.97
RF	1024	1.0	−300.8	−102.7	−1.26	−77.4	0.0	−468.49
RF	1024	1.2	−315.6	−85.7	−2.72	−77.6	0.0	−468.03
RF	2048	0.8	−277.9	−128.1	−0.07	−76.8	0.0	−469.20
RF	2048	1.0	−301.7	−102.7	−0.32	−77.4	0.0	−468.48
RF	2048	1.2	−318.4	−85.7	−0.87	−77.6	0.0	−468.98

[a] Freie Standardhydratationsenergie  $\Delta F_{hyd}$  des Natriumions, berechnet für unterschiedliche Systemgrößen (Zahl der Wassermoleküle  $N_w$ ) mit der  $P^3M$ -Methode<sup>[13,21,22]</sup> oder der Reaktionsfeldmethode (mit unterschiedlichen Cutoff-Radien  $R_{rf}$ )<sup>[15,16]</sup> zur Behandlung elektrostatischer Wechselwirkungen. Verwendet wurde das SPC-Wassermodell<sup>[23]</sup> in Kombination mit den Lennard-Jones-Parametern für Ion-Wasser-Wechselwirkungen nach Straatsma und Berendsen.<sup>[24]</sup> Die Simulationen wurden bei einem konstanten Volumen in periodischen kubischen Kästen der Kantenlänge  $L = [(N_w + 1)\rho^{-1}]^{1/3}$  mit  $\rho = 33.427\text{ nm}^{-3}$  ausgeführt. Bei der  $P^3M$ -Methode wurde eine „Spherical-Hat-Charge-Shaping“-Funktion der Breite 0.4 nm (oder 0.26 nm für  $N_w \leq 32$ ) verwendet,<sup>[25]</sup> zusammen mit einer Zuordnungsfunktion dritter Ordnung, einem Finite-Difference-Operator zweiter Ordnung, drei Alias-Vektoren zur Berechnung der optimalen Einflussfunktion und einem Gitterabstand von 0.05 nm (oder 0.0166 nm für  $N_w \leq 32$ ).<sup>[21]</sup> Bei der RF-Methode wurden der Radius des Reaktionsfeldes auf  $R_{rf}$  und die Lösungsmittelpermittivität auf 66.6 gesetzt. Der Cutoff wurde bezüglich des Sauerstoffatoms als Molekülzentrum vorgenommen. Um zu den intrinsischen Standardwerten  $\Delta F_{hyd}^0$  zu gelangen, werden die Rohdaten der freien Energie  $\Delta F_{sim}$  (berechnet aus Simulationen mit dem von Hummer vorgeschlagenen Schema,<sup>[26]</sup> das auf drei ionischen Ladungszuständen 0, 0.5 und 1 e beruht) bezüglich mehrerer Einflüsse korrigiert (mit einer Lösungsmittelpermittivität von 66.6 und einem ungefähren Ionenradius von 0.2 nm). Die Korrekturterme berücksichtigen die Nicht-Coulomb-Wechselwirkungen ( $\Delta F_{ncb}$ ), die künstliche Periodizität ( $\Delta F_{per}$ ), die Korrektur eines unzulässigen Summierungsschemas für das elektrostatische Potential ( $\Delta F_{sum}$ ; Konvertierung P-Summierungs- in M-Summierungskonvention<sup>[27]</sup>), den Einfluss des Grenzflächenpotentials an der ionischen Oberfläche auf das durchschnittliche Potential im Kasten ( $\Delta F_{srf}$ ), die für die Hohlräumbildung aufzuwendende Arbeit ( $\Delta F_{cav}$ ; 5.67  $\text{kJ mol}^{-1}$ ) und die Kompressionsarbeit  $\Delta F_{cmp}$  (7.95  $\text{kJ mol}^{-1}$ , entsprechend der Korrektur für den Standardzustand eines Gases bei einem Referenzdruck von 1 bar).

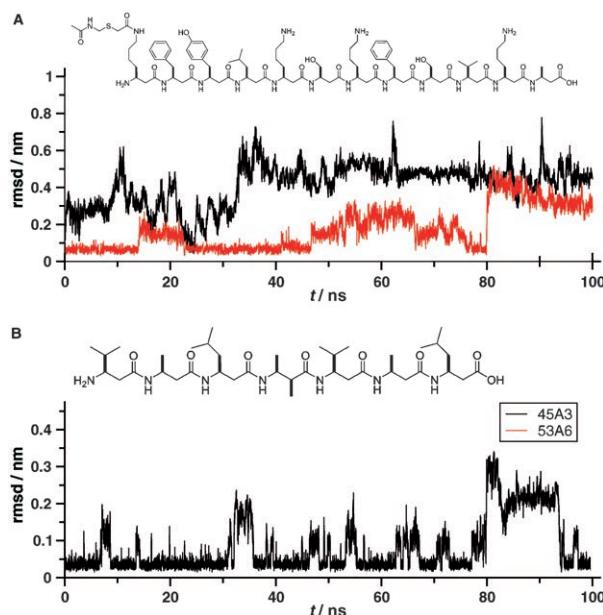
experimentellen Werten messbarer Eigenschaften verglichen werden. Man kann an dieser Stelle auch versuchen, simulierte Konformationsverteilungen in Kristallen mit Durchschnittswerten aus Röntgenbeugungsperimenten zu vergleichen. Weiterhin können aus Simulationen erhaltene NOE-Intensitäten (Kern-Overhauser-Effekt),  $^3J$ -Kopplungskonstanten und chemische Verschiebungen von gelösten Biomolekülen mit entsprechenden Durchschnittswerten aus NMR-Experimenten verglichen werden. Allerdings ist die rechnerische Reproduktion einer experimentell charakterisierten gefalteten Struktur weder eine notwendige<sup>[31]</sup> noch eine hinreichende Bedingung für die Richtigkeit eines Kraftfeldes. Das Kraftfeld sollte in der Lage sein, die Konformationsverteilung des gelösten Biomoleküls als Funktion der thermodynamischen Bedingungen zu reproduzieren; beispielsweise sollte es die korrekte Schmelztemperatur einer bestimmten Faltung vorhersagen können.

Eine besonders schwierige Aufgabe besteht darin, mithilfe eines Kraftfeldes die Faltung eines Polypeptids in Lösung als Funktion der Aminosäurezusammensetzung und der Art des Lösungsmittels vorauszusagen. Die Abbildungen 7 und 8 veranschaulichen, dass das GROMOS-Kraftfeld dieser Aufgabe gewachsen ist. Bei Verwendung der GROMOS-Kraftfeldparametersätze 43A1 und 45A3 wurden unterschiedliche Arten links- oder rechtsgängiger Helices sowie  $\beta$ -Kehren als vorherrschende Konformationen in MD-Simulationen von  $\beta$ - und  $\alpha$ -Peptiden in Methanol oder Wasser identifiziert (Abbildung 7). Die Ergebnisse stimmen mit Befunden auf der Grundlage von NOE-Daten überein.<sup>[32–37]</sup> Da allerdings nachgewiesen wurde, dass diese Parametersätze die Größe der freien Hydratationsenthalpien von Aminosäureanaloga unterschätzen,<sup>[7,38,39]</sup> wurde eine Neuparametrisierung vorgenommen, die dem 53A6-Satz



**Abbildung 7.** Durch MD-Simulationen ermittelte Faltungen von Polypeptiden und Peptoiden in unterschiedlichen Lösungsmitteln. Den gefalteten Strukturen (rot), die mithilfe der verfügbaren experimentellen NMR- oder Röntgendifferenzen modelliert wurden, sind Strukturen (blau) überlagert, die den Konformationen entsprechen, die in MD-Simulationen des Faltungs-Entfaltungs-Gleichgewichts am häufigsten vorgefunden wurden.<sup>[32–37]</sup> Lösungsmittel: A–C, E) Methanol, D) Wasser, F) Wasser oder Chloroform. Verwendete GROMOS-Versionen: A–D) 43A1, E) 53A6, F) 45A3.

ergab.<sup>[7]</sup> Am Beispiel der Faltung eines  $\beta$ -Dodecapeptids mit polaren Seitenketten verdeutlicht Abbildung 8 die Verbesserung, die erzielt wird, wenn freie Solvatisationsenergien polarer Verbindungen in den Kalibrierungssatz aufgenommen



**Abbildung 8.** Mittlere quadratische Abweichungen (rmsd) der Positionen der Rückgratatome zweier  $\beta$ -Peptide in Methanol, bezogen auf die helicalen Modellstrukturen aus NMR-Daten. A) Bei einem Peptid, das polare Seitenketten enthält, wird die experimentelle Faltung ausschließlich durch den neuenen Kraftfeldparametersatz 53A6 reproduziert.<sup>[36]</sup> B) Ein Peptid ohne polare Seitengruppen lässt sich die experimentelle Struktur sowohl durch das alte (45A3) als auch durch neue Kraftfeld (53A6) gleichermaßen reproduzieren (nur Daten für 45A3 gezeigt).

werden. Anders als der 45A3-Satz konnte der 53A6-Satz die experimentell beobachtete Helix reproduzieren. Eigenschaften, die weniger stark von kleinen Energiedifferenzen abhängen, wurden hingegen durch beide Parametersätze gut reproduziert.<sup>[36]</sup>

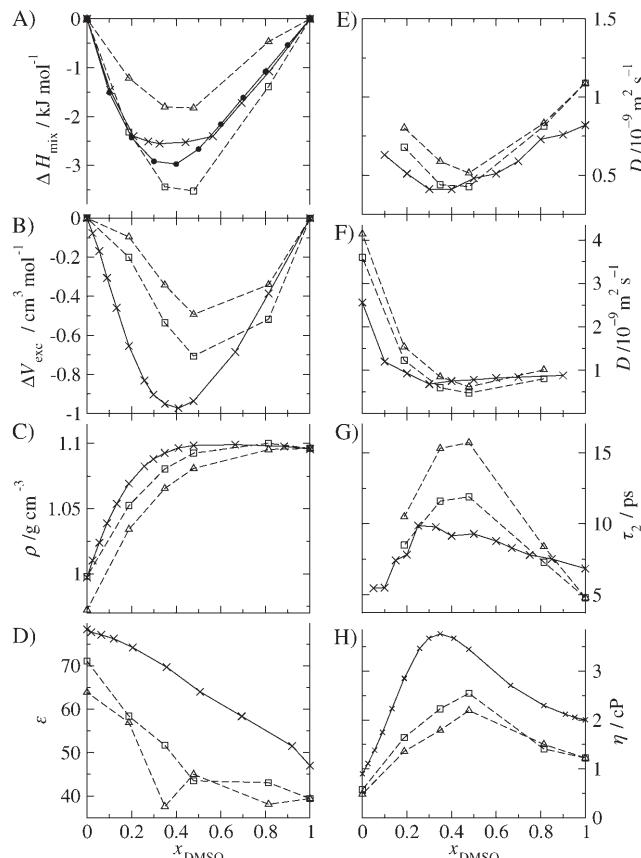
Ein Kraftfeld kann somit nur anhand von Molekül- oder Systemeigenschaften validiert werden, die auf die Simulationsparameter und -bedingungen hinreichend empfindlich reagieren.

## 2.5. Perspektiven bei der Entwicklung von Kraftfeldern

Gegenwärtige biomolekulare Kraftfelder sind durchaus noch verbessungsfähig. Zunächst wäre es wünschenswert, wenn die Van-der-Waals-Parameter und Partialladungsverteilungen geladener Gruppen von freien Solvationsenergien abgeleitet würden, wie dies bei apolaren und polaren neutralen Gruppen bereits geschehen ist.<sup>[7]</sup> Beispiele für solche Gruppen sind die Seitenketten von Arginin-, Lysin-, Asparagin- und Glutaminresten oder Phosphatgruppen in DNA, RNA und Lipiden. In Anbetracht der großen freien Solvationsenergien einzelner Ionen (mehrere hundert  $k_B T$ ) und der Tatsache, dass es technisch schwierig ist, solche Daten experimentell<sup>[40]</sup> oder aus Simulationen<sup>[27]</sup> zu erhalten, dürfte diese Aufgabe allerdings nicht einfach zu lösen sein.

Als zweites sollten die Eigenschaften von Lösungsmittelgemischen, die in Proteindenaturierungsexperimenten verwendet werden, als Funktion ihrer Zusammensetzung

bestimmt werden. Besonders thermodynamische Eigenschaften, wie die Mischungsenthalpie und die Dichte von Mischungen, wären für die Berechnung von freien Solvationsenergien und die Simulation der Faltung gelöster Biomoleküle wichtig.<sup>[8,9,41–43]</sup> Die Eigenschaften eines Lösungsmittelgemisches müssen nicht unbedingt linear von seiner Zusammensetzung abhängen, wie in Abbildung 9 am Beispiel von Wasser-Dimethylsulfoxid(DMSO)-Mischungen gezeigt ist. Lediglich die dielektrische Permittivität  $\epsilon$  zeigt einen linearen Verlauf, alle anderen Eigenschaften verhalten sich auf unterschiedliche Weisen nichtlinear.<sup>[42]</sup>



**Abbildung 9.** Eigenschaften von Wasser/DMSO-Mischungen bei 298 K und 1 atm als Funktion des Stoffmengenanteils an DMSO ( $x_{\text{DMSO}}$ ) aus MD-Simulationen.<sup>[42]</sup>  $\Delta H_{\text{mix}}$ : Mischungsenthalpie,  $\Delta V_{\text{ex}}$ : Exzessvolumen,  $\rho$ : Dichte,  $\epsilon$ : relative dielektrische Permittivität,  $D$ : Diffusionskoeffizient (Diagramm E: DMSO, Diagramm F: Wasser),  $\tau_2$ : Rotationskorrelationszeit (DMSO),  $\eta$ : Scherviskosität. Berechnete Werte:  $\triangle$  SPC-Wassermodell,  $\square$  SPC/L-Wassermodell; experimentelle Werte:  $\times$  und  $\bullet$ .

Derzeit eingesetzte biomolekulare Kraftfelder beschreiben die Elektronenpolarisation der Moleküle nur indirekt, was bei Systemen unter veränderlichen dielektrischen Bedingungen zu Ungenauigkeiten führt.<sup>[44–48]</sup> Dieses Problem ließe sich ausschalten, wenn man die Polarisierbarkeit explizit behandeln würde;<sup>[49,50]</sup> dafür müsste das Kraftfeld allerdings mehr oder weniger vollständig neu parametrisiert und an realistischen Systemen getestet werden. Die ersten Kraftfelder, die die Polarisierung explizit berücksichtigen, wurden

kürzlich vorgestellt,<sup>[51–53]</sup> ihr Leistungsvermögen zur Berechnung freier Solvationsenergien und zur Reproduzierbarkeit von Faltungsgleichgewichten muss allerdings noch überprüft werden.

Um große biomolekulare Systeme und langsame Prozesse wie die Bildung von Membranen oder Micellen effizient zu simulieren, wäre es nützlich, „grobkörnige“ (coarse-grained) Molekülmodelle zu formulieren, bei denen eine Gruppe kovalent gebundener Atome als einzelnes Teilchen behandelt wird.<sup>[54–59]</sup> Modelle dieser Art können um Größenordnungen schneller simuliert werden als atomistische Modelle, allerdings auf Kosten der Detailschärfe. Sie wurden erfolgreich in Simulationen von Membran- und Micellbildungen angewendet.<sup>[60]</sup> Um einschätzen zu können, wie sich die Näherungen und Vereinfachungen auswirken, ist ein Vergleich der mit grobkörnigen und atomistischen Modellen erhaltenen Eigenschaften erforderlich.

### 3. Das Suchproblem

Ein biomolekulares System ist im Allgemeinen durch eine sehr große Zahl von Freiheitsgraden charakterisiert ( $10^4$ – $10^6$  oder mehr). Die Bewegungen entlang dieser Freiheitsgrade sind unterschiedlichster Art. Sie können harmonisch oder anharmonisch, chaotisch oder diffusiv sein. Darüber hinaus kommt es zu einer zeitlichen und räumlichen Korrelation vom Femtosekundenbereich und einigen dutzend Nanometern bis hin zum Millisekundenbereich und mehreren Mikrometern. Die Energiehyperfläche eines solchen Systems, die durch die Potentialenergiefunktion definiert ist [z. B. Gl. (1)–(8)], ist daher und Höhe und räumlicher Ausdehnung auf. Dies macht die Suche nach dem globalen Energieminimum einer solchen mehrdimensionalen Funktion – genauer gesagt: die Suche nach den Bereichen auf der Hyperfläche, die am stärksten zur freien Energie des Systems beitragen – zu einer schwierigen, wenn nicht gar unlösbar Aufgabe.

Wie bereits erwähnt wurde, kann der Zustand eines biomolekularen Systems nicht durch eine einzige, dem globalen Minimum entsprechende Konfiguration oder Struktur beschrieben werden. Vielmehr ist mithilfe der statistischen Mechanik ein Ensemble von Konfigurationen zu ermitteln, in dem die Gewichtung einer Konfiguration  $x$  durch den Boltzmann-Faktor gegeben ist [Gl. (11);  $k_B$  ist die Boltzmann-Konstante und  $T$  die absolute Temperatur]:

$$P(x) \sim \exp[-V(x)/k_B T] \quad (11)$$

Die exponentielle Gewichtung in Gleichung (11) bewirkt, dass Bereiche hoher Energie auf der Hyperfläche keine Konfigurationen beisteuern, die für den Zustand des Systems relevant wären, es sei denn, diese kommen in sehr großer Zahl vor (Entropie). Die Gleichgewichtseigenschaften des Systems werden von denjenigen Bereichen des Konfigurationsraumes bestimmt, in denen  $V(x)$  niedrig ist. Somit besteht eine vorrangige Aufgabe bei der biomolekularen Modellierung darin, Techniken zu entwickeln, mit denen die ausgedehnten biomolekularen Energiehyperflächen nach Bereichen niedriger Energie abgesucht werden können.

Im Folgenden stellen wir die wichtigsten Methoden vor, die gegenwärtig zum Durchsuchen des Konfigurationsraumes verwendet werden.<sup>[61–63]</sup> Es sind weiterhin Suchtechniken im Einsatz, die nicht nur die Molekülkoordinaten  $x$  als Variablen verwenden, sondern auch ihre Boltzmann-Wahrscheinlichkeiten  $P(x)$ . Eine Diskussion dieser als Wahrscheinlichkeitsuche bezeichneten Techniken findet sich in Lit. [61, 64, 65].

#### 3.1. Methoden zum Durchsuchen (Sampling) des Konfigurationsraumes

Alle verfügbaren Suchmethoden haben ihre Stärken und Schwächen. Die Eignung einer Methode für eine bestimmte Aufgabe hängt zum einen von der Form der Potentialfunktion  $V(x)$  ab und zum anderen von Zahl und Art der Freiheitsgrade im System. Man unterscheidet zwei Klassen von Suchmethoden: systematische und heuristische.

Systematische Suchmethoden durchsuchen den gesamten Konfigurationsraum des biomolekularen Systems oder zumindest einen großen Teil davon. Bestimmte Teilaräume können von der Suche ausgeschlossen werden, ohne die Qualität der Lösung zu beeinträchtigen, wenn rigorose Argumente dafür sprechen, dass diese Teilaräume die gewünschte Lösung nicht enthalten.<sup>[66]</sup> Solche Argumente beruhen meist auf physikalisch oder chemisch fundierten Vorgaben für die Struktur des Konfigurationsraumes oder die Energiehyperfläche. Systematische Suchtechniken sind nur bei kleinen Molekülen mit wenigen Freiheitsgraden anwendbar,<sup>[67–70]</sup> da der Rechenaufwand exponentiell mit der Zahl der Freiheitsgrade ansteigt.

Heuristische Suchmethoden suchen nur einen kleinen Teil des Konfigurationsraumes ab und erzeugen einen im Sinne der Boltzmann-Gewichtung repräsentativen Satz von Systemkonfigurationen. Diese Methoden können in drei Kategorien unterteilt werden (siehe auch Tabelle 7):

1. Methoden, bei denen eine Serie von Systemkonfigurationen erzeugt wird, die voneinander unabhängig sind. Ein Beispiel ist die Distanzgeometriemethode mit metrischer Matrix,<sup>[72, 73]</sup> die eine nichtkorrelierte Serie von Zufallskonfigurationen erzeugt, sofern das Suchproblem auf der Grundlage von Abständen ausgedrückt werden kann.
2. Methoden, die die Konfiguration des vollständigen Moleküls oder Systems schrittweise aus Konfigurationen von Molekül- oder Systemfragmenten aufbauen. Beispiele sind die Aufbauprozedur nach Scheraga,<sup>[74, 75]</sup> kombinatorische Aufbaumethoden, die sich dynamische Programmertechniken zu Nutze machen,<sup>[76]</sup> und Monte-Carlo(MC)-Kettenwachstumsmethoden<sup>[77, 78]</sup> wie die „Configurational-Bias-Monte-Carlo“(CBMC)-Technik.<sup>[79]</sup>
3. Energieminimierungs- (EM), Metropolis-Monte-Carlo(MC)-, Moleküldynamik- und stochastisch-dynamische (SD) Methoden,<sup>[80]</sup> die ausgehend von einer bekannten Konfiguration eine neue Konfiguration des vollständigen Systems erzeugen. Diese Methoden können je nach Schrittrichtung und -größe weiter unterteilt werden (siehe Tabelle 6). Energieminimierungen können lediglich auf Energiewerten und Zufallsschritten beruhen (Simplex-Methoden), auf Energien und Energiergradienten (Me-

**Tabelle 6:** Heuristische Methoden zur Durchsuchung des Konfigurationsraumes nach Konfigurationen  $x$  mit niedriger Energie  $V(x)$ .<sup>[a]</sup>

Wechsel basiert auf:	EM	MC	Methode	SD	PEACS
Energie	ja	ja	nein	nein	ja
Energiegradient	ja	nein	ja	ja	ja
2. Ableitung der Energie	ja	nein	nein	nein	nein
Gedächtniseffekt	nein	nein	ja	ja	ja
zufälliger Wechsel	ja	ja	nein	ja	nein

[a] EM: Energieminimierung, MC: Monte-Carlo, MD: Moleküldynamik, SD: stochastische Dynamik, PEACS: Konformationssuche bei getemperter Potentialenergie.<sup>[71]</sup>

thoden des steilsten Abstiegs und konjugierter Gradienten), oder auch Ableitungen zweiter Ordnung der Energie einbeziehen (Hesse-Matrix-Methoden). Bei MC-Methoden ist die Schrittrichtung zufällig, und die Schrittgröße wird durch das Boltzmannsche Akzeptanzkriterium begrenzt. Dieses besagt, dass der Schritt im Konfigurationsraum akzeptiert wird, wenn die Energie des Systems abnimmt ( $\Delta V < 0$ ). Für  $\Delta V > 0$  hingegen wird der Schritt mit der Wahrscheinlichkeit  $\exp(-\Delta V/k_B T)$  akzeptiert. In MD-Simulationen legen zwei Faktoren den Schritt fest: die Kraft (der negative Wert des lokalen Gradienten  $\partial V / \partial x$ ) und die Trägheit der Freiheitsgrade, die als „Kurzzeitgedächtnis“ für den bisher zurückgelegten Weg dient. In SD-Simulationen wird die Kraft durch eine Zufallskomponente ergänzt, deren Größe von der Temperatur des Systems, den Atommassen und den atomaren Reibungskoeffizienten abhängt. In Algorithmen zum Verfolgen von Potentialenergie-Isolinien (potential-energy contour tracing; PECT)<sup>[81]</sup> und zur Konformationssuche bei getemperter Potentialenergie (potential-energy annealing conformational search; PEACS)<sup>[71]</sup> werden die Energiewerte konstant gehalten (PECT) bzw. getempert (PEACS), um Sattelpunkte zu lokalisieren und diese zu überqueren. Viele Suchprozeduren zur schrittweisen Untersuchung des Konfigurationsraumes beruhen auf Kombinationen der fünf genannten Grundprinzipien (Energie, Gradienten, Hesse-Matrix, Gedächtnis, Zufall).<sup>[61]</sup>

Durch die allgemeine Beschaffenheit der Energiehyperflächen  $V(x)$  sind der Wirksamkeit von Suchmethoden natürliche Grenzen gesetzt, denn wegen der Vielzahl hochenergetischer Barrieren zwischen lokalen Minima ist der Konvergenzradius der schrittweisen Methoden (2., 3.) im Allgemeinen sehr klein. Aus diesem Grund hat man verschiedene Techniken entwickelt, um die Leistungsfähigkeit von Suchmethoden zu steigern. In Tabelle 7 sind drei allgemeine Verfahren mit verbessertem Suchverhalten zusammengefasst.<sup>[61]</sup>

### 3.1.1. Deformation oder Glättung der Potentialfläche, um Energiebarrieren zu verringern

a) Im Allgemeinen verkürzt das Glätten der Potentialenergiefunktion  $V(x)$  die Suche nach den Energieminima. Diese Technik ist in verschiedenen Fällen angewendet worden, z.B. bei Strukturaufklärungen auf der Grundlage

**Tabelle 7:** Techniken zur Steigerung der Suchleistung von Simulationsmethoden.<sup>[a]</sup>

#### 1. Deformation oder Glättung der Potentialfläche

- a) Weglassen von hochauflösten Strukturfaktoren bei der Strukturverfeinerung mit Röntgenbeugungsdaten
- b) schrittweises Einführen von Abstandsfixierungen bei Strukturverfeinerungen auf der Basis von NOE-Daten<sup>[82]</sup>
- c) Ersetzen der „harten“ Atomrümpe bei nichtbindenden Wechselwirkungen durch „weiche“ Atomrümpe<sup>[83]</sup>
- d) Glätten der Hyperfläche durch Skalieren mit einer Diffusionsgleichung<sup>[83,84]</sup>
- e) Vermeiden der wiederholten Durchmusterung einer Energiemulde durch lokale Energieanhebung oder Konformationsflutung<sup>[85,86]</sup>
- f) Dämpfung von geometrischen Randbedingungen, die aus experimentellen Daten abgeleitet wurden (NMR, Röntgen), durch zeitliche Mittelung<sup>[87,88]</sup>
- g) Umgehen von Energiebarrieren durch Erhöhen der Dimensionalität des Kartesischen Raumes (4D-MD)<sup>[89]</sup>
- h) Einfrieren von hochfrequenten Freiheitsgraden durch Zwangsbedingungen<sup>[90]</sup>
- i) Erzeugung grobkörniger Modelle durch Verringerung der Zahl der Wechselwirkungszentren<sup>[54–59]</sup>

#### 2. Skalierung der Systemparameter

- a) Tempern<sup>[91]</sup>
- b) Massenskalierung<sup>[92]</sup>
- c) Molekularfeldnäherungen<sup>[93]</sup>

#### 3. Systemvervielfachung

- a) genetische Algorithmen<sup>[94]</sup>
- b) Replika-Austausch und multikanonische Algorithmen<sup>[62]</sup>
- c) kooperative Suche: SWARM<sup>[95]</sup>

[a] Details finden sich in Lit. [60–62] und den darin zitierten Literaturstellen.

von Röntgenbeugungs- oder NMR-spektroskopischen Daten, zur Konformationssuche und zur Vorhersage von Proteinstrukturen. Bei der Methode 1a in Tabelle 7 wird die Elektronendichte eines Biomolekülkristalls geglättet, indem man beim Zurückrechnen aus den Fourier-Transformierten hochauflösten Beugungsintensitäten auslässt. Dieses Glätten erhöht den Konvergenzradius der Strukturverfeinerung.

- b) Beim Aufbau einer Proteinstruktur aus NMR-spektroskopisch ermittelten Atomabständen kann die Konvergenz des Suchprozesses verbessert werden, indem man in der Potentialenergiefunktion nach und nach Abstände zwischen Atomen festlegt, die entlang der Polypeptidkette immer weiter voneinander entfernt sind (Methode der veränderlichen Zielfunktion; „variable-target-function method“).<sup>[82]</sup>
- c) Viele Barrieren auf der Energiehyperfläche eines molekularen Systems werden durch „harte“ Atomrümpe verursacht, d.h. durch die starke Abstoßung zwischen überlappenden Atomen. Diese Barrieren lassen sich beseitigen, indem man die kurzreichweitige Abstoßung zwischen Atomen „aufweicht“.<sup>[96–99]</sup> Durch Atome mit weichem Rumpf erhält man glattere Energiehyperflächen und eine stark verbesserte Suchleistung.
- d) Bei Verformungsmethoden auf der Grundlage der Diffusionsgleichung<sup>[83,84]</sup> wird die Verformung der Energiehyperfläche während einer Simulation als proportional

- zur lokalen Krümmung (d.h. der zweiten Ableitung) der Hyperfläche berechnet. So werden bevorzugt die steilsten Maxima und Minima der Hyperfläche geglättet, und die Suche verläuft sehr effizient.
- e) Eine weitere Möglichkeit zur Erhöhung der Suchleistung besteht darin, die während der Suche gesammelte Information zur Energiehyperfläche in die Potentialenergiefunktion einfließen zu lassen. Sobald ein lokales Energiedreieck gefunden ist, wird es durch eine lokale Verformung der Potentialenergiefunktion von der Energiehyperfläche entfernt. Auf diesem Konzept beruhen Verfahren wie die Deflationsmethode,<sup>[100]</sup> die Methode des lokalen Anstiegs<sup>[85]</sup> („local elevation“; auch als Metadynamik bezeichnet)<sup>[101]</sup> und die Methode der konformativen Flutung.<sup>[86]</sup>
  - f) Bei einer anderen Art der Suche wird eine Gedächtnisfunktion durch einen Potentialenergieterm eingeführt, der anstelle des Istwertes einen gleitenden Mittelwert einer Koordinate über die bis dahin durchlaufene Trajektorie oder das bis dahin erzeugte Ensemble verwendet.<sup>[102]</sup> Die Anwendung dieser zeit- oder ensembleabhängigen Randbedingungen bei der Bestimmung von Proteinstrukturen aus NMR-spektroskopischen oder Röntgenstrukturdaten verbessert die Suchleistung deutlich.<sup>[87,88]</sup>
  - g) Barrieren auf der Energiehyperfläche können umgangen werden, indem man die Dimensionalität des Konfigurationsraumes über die drei Kartesischen Koordinaten hinaus erhöht. Diese Technik lokalisiert eine Konformation niedriger Energie in einem vieldimensionalen Kartesischen Raum und projiziert sie nach und nach in den dreidimensionalen Raum, wobei Energie und Konformation möglichst wenig gestört werden.<sup>[103]</sup> Einige Varianten des ursprünglichen Verfahrens sind vorgestellt worden.<sup>[104–107]</sup> Dynamische Suchmethoden können ebenfalls mit einer Erhöhung der Dimensionalität kombiniert werden. Durch Anwendung der Moleküldynamikmethode im vierdimensionalen Kartesischen Raum lassen sich Energiebarrieren im dreidimensionalen Raum umgehen<sup>[89]</sup> und Änderungen der freien Energie berechnen.<sup>[108]</sup>
  - h) Eine Standardtechnik zum Glätten der Energiehyperfläche besteht darin, in einem System die Freiheitsgrade mit den höchsten Frequenzen durch Zwangsbedingungen einzufrieren.<sup>[90,109–113]</sup> Zwangsbedingungen für Bindungslängen, die in biomolekularen Simulationen standardmäßig verwendet werden, ermöglichen viermal längere Zeitschritte.<sup>[114]</sup> Hochfrequente Bewegungen lassen sich auch mithilfe von so genannten weichen Zwangsbedingungen ausschalten;<sup>[115]</sup> die festgelegten Bindungslängen ändern sich dabei adiabatisch mit den einwirkenden Kräften.

### 3.1.2. Skalierung der Systemparameter

- a) Die Technik des simulierten Temperns<sup>[91]</sup> („simulated annealing“) beruht auf einer Simulation oder Suche bei hoher Temperatur  $T$  und dem anschließenden Abkühlen des Systems. Durch Erhöhen der Temperatur werden Energiebarrieren leichter überwunden, sodass ein größere-

rer Teil des Konfigurationsraumes zugänglich ist. Simulierte Temperiern wird häufig in Kombination mit MC-, MD- und SD-Simulationen angewendet. (Ein Beispiel für das Temperiern der Potentialenergie findet sich in Lit. [71].)

- b) Die Skalierung von Atommassen verbessert die Suchleistung. In der klassischen Verteilungsfunktion ohne Zwangsbedingungen kann die Integration über die Atomimpulse analytisch und getrennt von der Integration über die Koordinaten vorgenommen werden. Damit tauchen die Atommassen nicht im Konfigurationsintegral auf, und die Gleichgewichtseigenschaften des Systems sind folglich unabhängig von den Atommassen. Dies kann auf verschiedene Weise zur Verbesserung der Suchleistung genutzt werden. Indem die Masse eines bestimmten Molekülteils erhöht wird, steigt man dessen relative Trägheit. Dadurch werden Energiebarrieren leichter überwunden,<sup>[92]</sup> und die Zeitschritte können größer gewählt werden.
- c) Eine verbesserte Suchleistung lässt sich auch mithilfe einer Molekularfeldnäherung erreichen, wenn man das biomolekulare System in zwei Teile A und B teilt, die sich jeweils im gemittelten Feld des anderen Teils bewegen. Die Anfangskonfiguration des Systems besteht aus  $N_A$  identischen Abbildern von Teil A und  $N_B$  identischen Abbildern von Teil B. Die Positionen entsprechender Atome in den identischen Abbildern können als identisch gewählt werden. Die Kräfte, die die Atome in allen Abbildern von Teil B auf die Atome in jedem Abbild von Teil A ausüben, werden mit einem Faktor  $N_B^{-1}$  skaliert. Damit erhält man die mittlere Kraft, mit der Teil B auf die einzelnen Atome von Teil A einwirkt. Umgekehrt werden die Kräfte, die die Atome in allen Abbildern von Teil A auf jedes Abbild von Teil B ausüben, mit einem Faktor  $N_A^{-1}$  skaliert. Die Kräfte zwischen unterschiedlichen Abbildern von Teil A sind gleich null, ebenso wie die Kräfte zwischen unterschiedlichen Abbildern von Teil B. Die MD-Simulation beruht auf der simultanen Integration der Newtonschen Bewegungsgleichung,  $\dot{\mathbf{r}} = \mathbf{a}$ , für alle Abbilder der Teile A und B. Auf diese Weise erhält man  $N_A$  einzelne Trajektorien des Teils A im gemittelten Feld des Teils B und umgekehrt. Dabei geht die korrekte Beschreibung der Dynamik verloren, da das dritte Newtonsche Gesetz,  $\mathbf{f}_{AB} = -\mathbf{f}_{BA}$ , verletzt wird. Diese Technik führt nur dann zu einer effizienteren Simulation, wenn das System in Teile sehr unterschiedlicher Größe unterteilt wird (z.B.  $A \ll B$ ) und der größere Teil durch eine einzige Kopie abgebildet wird, d.h.  $N_B = 1$ . Verbesserte Suchprozeduren auf der Grundlage einer Molekularfeldnäherung wurden in verschiedenen Varianten vorgestellt.<sup>[93,116–127]</sup> Angewendet wurden solche Methoden auf die Diffusion von CO-Molekülen im Feld eines Proteinkomplexes,<sup>[93,118]</sup> auf das Konformationsgleichgewicht von Proteineinheitenketten,<sup>[117]</sup> bei der Bestimmung der Konformation von Proteinschleifen,<sup>[123]</sup> zur Suche nach Energiedreiecken von Polypeptidkonformationen<sup>[126,127]</sup> und Nucleinsäuresegmenten<sup>[128]</sup> sowie zur Suche nach Bindungsstellen in Enzymen.<sup>[129–131]</sup>

### 3.1.3. Simulationen mit vielfachen Abbildern (mit Bezug zueinander)

Der Molekularfeldansatz verwendet vielfache Abbilder eines Teils des Systems. Ausgehend von diesem Konzept wurden weitere Methoden zur Verbesserung der Suchleistung entwickelt (Tabelle 7):

- In einem genetischen Algorithmus<sup>[132]</sup> wird eine Gruppe von Abbildern des biomolekularen Systems mit unterschiedlichen Konfigurationen betrachtet. Mutation und Kombination von (Teilen von) Konfigurationen nach vorgegebenen Regeln erzeugen neue und löschen existierende Konfigurationen.
- Beim „Replika-Austausch-Algorithmus“ werden mehrere Abbilder des Systems jeweils bei einer bestimmten Temperatur durch MC, MD oder SD simuliert. Von Zeit zu Zeit werden Abbilder, die aus Simulationen bei ähnlicher Temperatur stammen, unter Berücksichtigung einer Austauschwahrscheinlichkeit auf der Grundlage des Boltzmann-Faktors [Gl. (11)] ersetzt. Auf diese Weise erhält man (bei vollständiger Durchsuchung des Konfigurationsraumes) für jede Temperatur Ensembles mit Boltzmann-Verteilung (kanonische Ensembles).<sup>[133]</sup> Eine Verallgemeinerung dieser Prozedur sind „multikanonische Algorithmen“,<sup>[62]</sup> die zur Simulation von Proteinen im Vakuum eingesetzt wurden.<sup>[133]</sup> Die Einbeziehung von Freiheitsgraden des Lösungsmittels kann die Effizienz des Algorithmus beeinträchtigen.<sup>[134]</sup> Beim Austausch der Abbilder geht dynamische Information verloren, und bei kurzen Simulationszeiten kann der Entropiegehalt hin zu den oberen und unteren Enden des betrachteten Temperaturbereichs verzerrt werden.
- Die SWARM-Moleküldynamikmethode<sup>[95]</sup> beruht auf dem Konzept, eine Anzahl (einen „Schwarm“) von Abbildern, die jeweils durch eine eigene Trajektorie beschrieben werden, in einem kooperativen System aus vielen Abbildern zusammenzufassen, welche den Konfigurationsraum durchsucht. Um ein solches kooperatives System zu konstruieren, wird jedes Abbild zusätzlich zu den physikalischen Kräften des Potentialenergielandes  $V(x)$  künstlichen Kräften unterworfen, die die Trajektorie jedes Abbildes hin zu einem Durchschnittswert der Trajektorien des Schwarms lenken. Dieses Verfahren ist analog zum scheinbar intelligenten Verhalten eines Insektschwarmes, dessen Individuen nicht die Fähigkeit zum Vorausdenken haben. SWARM-Trajektorien werden durch lokale Minima nur schwach angezogen und neigen stattdessen dazu, einem Gesamtenergiegradienten hin zum globalen Energieminimum zu folgen.

Unser Überblick beschreibt die Methoden zur Durchsuchung des Konfigurationsraumes nur in ihren Grundzügen. Ausführlichere Übersichten finden sich in Lit. [61–63]. Da biomolekulare Konfigurationsräume zu groß für systematische Suchmethoden sind, wird man in der biomolekularen Modellierung im Allgemeinen auf heuristische Suchmethoden zurückgreifen. Dieser Überblick (Tabelle 6 und 7) über die Methoden und Tricks, die angewendet und kombiniert werden können, um eine leistungsfähige Suchmethode zu

konstruieren, soll als praktische Hilfe bei der Auswahl des besten Modellierungsverfahrens für eine spezielle Aufgabe oder die interessierende Energiehypersfläche dienen.

### 3.2. Konvergenz der simulierten Eigenschaften

Die Zeitskalen, auf denen sich die unterschiedlichen Eigenschaften von biomolekularen Systemen dynamisch entwickeln, reichen vom Femtosekundenbereich bis über den Sekundenbereich hinaus. Durch die begrenzte Rechenleistung können gegenwärtige MD-Simulationen biomolekulärer Systeme je nach Systemgröße einige Nanosekunden bis zu einige hundert Nanosekunden erfassen. Dies wirft die Frage auf, ob solche Zeiträume ausreichen, um verlässliche Trajektorienmittel für die unterschiedlichen Molekül- oder Systemeigenschaften zu gewinnen. Trajektorienmittel werden gewöhnlich nur dann repräsentativ sein, wenn der Äquilibrierungszeitraum einer Simulation,  $\tau_{\text{Äquil}}$ , länger ist als die Relaxationszeit  $\tau_{\text{Relax}}(Q)$  einer Eigenschaft  $Q$  [Gl. (12)] und wenn die Simulationszeit,  $\tau_{\text{Sim}}$ , viel länger ist als  $\tau_{\text{Relax}}(Q)$  [Gl. (13)].

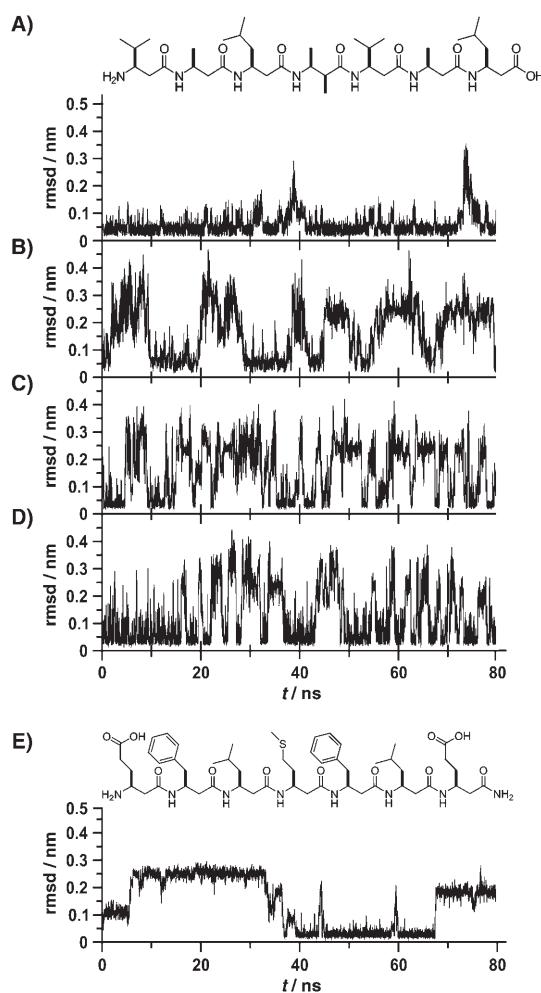
$$\tau_{\text{Äquil}} > \tau_{\text{Relax}}(Q) \quad (12)$$

$$\tau_{\text{Sim}} \gg \tau_{\text{Relax}}(Q) \quad (13)$$

Wenn die Bedingungen (12) und (13) nicht erfüllt sind, wird das Trajektorienmittel  $\langle Q(t) \rangle_t$  der Eigenschaft  $Q$  einen zeitlichen Drift zeigen oder fehlerbehaftet sein.<sup>[135–137]</sup>

Die Zeitskala, auf der sich eine bestimmte physikalische Größe eines biomolekularen Systems ändert und relaxiert, hängt ab von 1) der Art des molekularen Systems, 2) dem thermodynamischen Zustand und 3) der speziellen Größe oder Eigenschaft. Dieser Sachverhalt ist in Abbildung 10 am Beispiel der Dynamik von  $\beta$ -Heptapeptiden in Methanolösung illustriert. Bei 298 K ist das vorherrschende Konformer die  $3_{14}-L$ -Helix, die sich innerhalb von 100 ns nur wenige Male faltet und entfaltet (Diagramm A). Bei 340 K werden sehr viel mehr Faltungs- und Entfaltungseignisse beobachtet (Diagramm B). Durch Verringern der Lösungsmittelviskosität auf 1/3 (Diagramm C) oder 1/10 (Diagramm D) des ursprünglichen Wertes wird der Faltungs- und Entfaltungsprozess deutlich beschleunigt, sodass die Statistik des Faltungs-Entfaltungs-Gleichgewichts besser konvergiert. Das Diagramm E verdeutlicht, dass längere und geladene Seitenketten im Polypeptid den Faltungs- und Entfaltungsprozess verlangsamen und die „Abtastgeschwindigkeit“ entsprechend verringern (vgl. Diagramm B).

Nicht nur unterschiedliche Moleküle und thermodynamische Zustände können sich in unterschiedlichen Relaxationszeiten äußern, sondern auch unterschiedliche Eigenschaften. Die Potentialenergie des gelösten Stoffs und das Quadrat seines Gesamtdipolmomentes (das mit seiner dielektrischen Permittivität in Beziehung steht) relaxieren schneller als etwa die quadratisch gemittelten Fluktuationen der Atompositionen aller Atome.<sup>[140, 141]</sup> Systemeigenschaften wie die freie Faltungsenergie konvergieren sogar noch langsamer. (Ein allgemeines Prinzip besagt, dass Systemeigen-



**Abbildung 10.** Mittlere quadratische Abweichung der Atompositionen im Peptidrückgrat zweier  $\beta$ -Heptapeptide identischer Kettenlänge in Methanol bei 1 atm. A–D) Ein Peptid mit apolaren Seitenketten wurde bei 298 K (A) und bei 340 K (B–D) simuliert.<sup>[32,138]</sup> In den Simulationen (C) und (D) wurde die Viskosität des Lösungsmittels durch Massenskalierung um einen Faktor 3 (C) oder 10 (D) reduziert. Eine Erhöhung der Temperatur oder eine Verringerung der Lösungsmittelviskosität beschleunigt die Faltung und Entfaltung des Peptids. E) Ein Peptid mit einigen polaren Seitenketten wurde bei 340 K in Methanol simuliert. Die polaren Seitenketten verlangsamen die Faltung und Entfaltung.<sup>[139]</sup>

schaften langsamer konvergieren als Moleküleigenschaften.<sup>[1,142]</sup>)

Relaxation und Dynamik der Eigenschaften eines biomolekularen Systems können mithilfe unterschiedlicher Methoden analysiert werden:<sup>[135–137]</sup>

1. Aus Gleichgewichtssimulationen kann man die Zeitreihen einer Eigenschaft  $Q(t)$ , ihres Mittelwertes  $\langle Q(t) \rangle_t$ , oder ihrer Fluktuationen  $\langle (Q(t) - \langle Q(t) \rangle_t)^2 \rangle_t^{1/2}$  ermitteln oder aber die Autokorrelationsfunktion,  $\langle Q(t')Q(t'+t) \rangle_t$ , berechnen. Die Abklingzeit der Autokorrelationsfunktion und die Geschwindigkeit, mit der sich das Trajektorienmittel aufbaut, liefern eine Größenabschätzung für  $\tau_{\text{Relax}}(Q)$ .<sup>[137]</sup>
2. Startet man die Simulation von einem Nichtgleichgewichtszustand aus, dann liefert die Geschwindigkeit, mit der  $Q(t)$  zum Gleichgewicht relaxiert, über mehrere

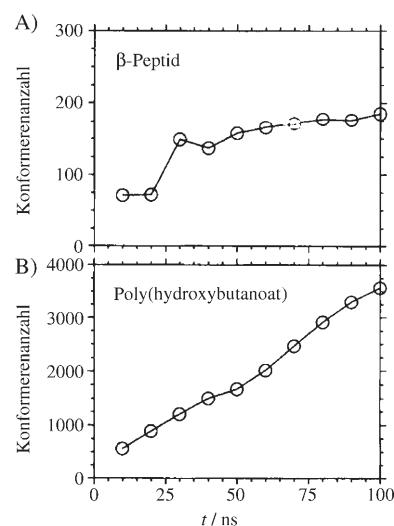
Nichtgleichgewichtstrajektorien gemessen eine Größenangabe für  $\tau_{\text{Relax}}(Q)$ .

3. Konvergieren unterschiedliche MD-Simulationen, die bei unterschiedlichen Anfangszuständen beginnen, nicht auf das gleiche Trajektorienmittel der Eigenschaft  $Q$ , dann ist  $\tau_{\text{Relax}}(Q)$  länger als die Simulationszeit.<sup>[143]</sup>

Bezüglich weiterer Beispiele für Relaxationszeiten von Eigenschaften diverser Systeme verweisen wir auf Lit. [135–137, 140–143].

### 3.3. Vereinfachende Faktoren beim Suchproblem

Das Durchsuchen des Konfigurationsraumes biomolekularer Systeme erscheint auf den ersten Blick wie eine schier unlösbare Aufgabe, doch das Problem kann je nach Beschaffenheit der Energiehyperfläche weniger kompliziert sein. Simulationen der Faltungs-Entfaltungs-Gleichgewichte von Polypeptiden in Lösung mithilfe eines thermodynamisch kalibrierten Kraftfeldes und einer expliziten Darstellung der Lösungsmittelmoleküle haben ergeben, dass im entfalteten oder denaturierten Zustand dieser Polypeptide nur ein geringer Teil der theoretisch möglichen Konformationen im Gleichgewicht vorhanden ist. Im Fall von Peptiden mit 20 Torsionswinkeln im Rückgrat beschränkt die Verwendung von physikalisch realistischen Kraftfeldern, die die (nicht-bindenden) Wechselwirkungen zwischen den Aminosäureresten beschreiben, die Zahl der  $10^9$  theoretisch möglichen Konformere auf  $10^3$  relevante Konformere.<sup>[144–146]</sup> In Abbildung 11 ist die Zahl der Konformationen aufgetragen, die während einer MD-Simulation eines Polypeptids und eines anderen Polymers gleicher Länge generiert werden. In Methanol steigt diese Zahl für das  $\beta$ -Heptapeptid sublinear an und geht in ein Plateau von etwa 200 Konformationen über



**Abbildung 11.** Anzahl generierter Konformationen als Funktion der Laufzeit einer MD-Simulation: A)  $\beta$ -Heptapeptid in Methanol bei 340 K.<sup>[138]</sup> B) (Val-Ala-Leu)<sub>2</sub>-3-Hydroxybutanoat in Chloroform bei 298 K.<sup>[147]</sup> Bezüglich der Definition einer Konformation (Cluster von Strukturen) beziehen wir uns auf Lit. [138].

(Abbildung 11 A). Bei einem Poly(hydroxybutanoat)-Molekül gleicher Länge in Chloroform steigt dagegen die Zahl der generierten (relevanten) Konformationen linear mit der Zeit (Abbildung 11 B), wie man es bei etwa  $10^9$  theoretisch möglichen Konformationen für jedes der beiden Moleküle auch erwarten würde. Der Unterschied ist auf die Gegenwart von Wasserstoffbrückendonoren und -acceptoren im  $\beta$ -Heptapeptid zurückzuführen, die durch Bildung von (energetisch günstigen) Wasserstoffbrücken den zugänglichen Konformationsraum des Moleküls einschränken. Sind wie beim Poly(hydroxybutanoat) keine Wasserstoffbrückendonoren vorhanden, so entfällt diese Beschränkung des Konfigurationsraumes.<sup>[147]</sup>

Bei Verwendung eines realistischen Kraftfeldes wird der größte Teil des Konfigurationsraumes sehr hohen Energien entsprechen, was bedeutet, dass der Teil des Konfigurationsraumes, der zur Vorhersage der stabilsten Faltung eines Polypeptids oder Proteins durchsucht werden muss, nicht exponentiell mit der Systemgröße oder Kettenlänge wächst.<sup>[144]</sup>

### 3.4. Erschwerende Faktoren beim Suchproblem

Die freie Energie  $F$  eines Systems aus  $N$  Teilchen in einem Volumen  $V$  bei der Temperatur  $T$  ist ein  $6N$ -dimensionales Integral über alle Teilchenkoordinaten  $\mathbf{r}$  und Teilchenimpulse  $\mathbf{p}$  des Boltzmann-Faktors mit dem Hamilton-Operator des Systems (kinetische plus Potentialenergie) [Gl. (14);  $h$  ist das Plancksche Wirkungsquantum].

$$F_{NVT} = -k_B T \ln \left( (N! h^{3N})^{-1} \int \int \exp(-H(\mathbf{p}, \mathbf{r})/k_B T) d\mathbf{p} d\mathbf{r} \right) \quad (14)$$

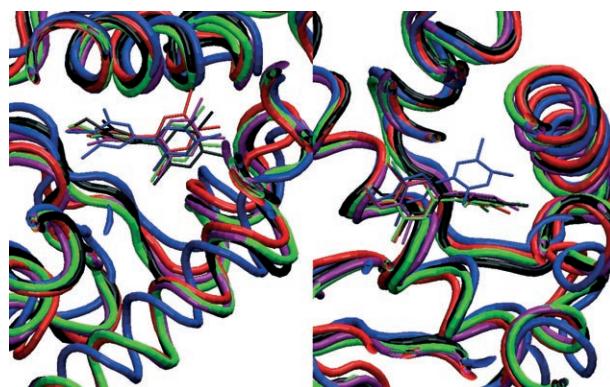
Der Faktor  $N!$  entfällt, wenn unterscheidbare Teilchen betrachtet werden. Da der Integrand – der Boltzmann-Faktor – überall positiv ist, führt das Weglassen von Konformationen im Integral zu systematischen (sich nicht gegenseitig aufhebenden) Fehlern. Die Gleichung (14) verdeutlicht, dass die zum globalen Energieminimum gehörende Struktur weder die freie Energie ganz allein bestimmt noch für das Ensemble von Konformationen repräsentativ ist, sondern dass andere Konformationen höherer Energie und größerer Häufigkeit ebenfalls signifikant beitragen. Mit anderen Worten setzt sich die freie Energie gemäß Gleichung (15) aus der inneren Energie  $U$  und der Entropie  $S$  zusammen.

$$F = U - TS \quad (15)$$

Freiheitsgrade des Lösungsmittels können ebenfalls signifikant zur freien Faltungsenergie beitragen.<sup>[148]</sup> Aus einer 200-ns-MD-Simulation des Faltungs-Entfaltungs-Gleichgewichts eines  $\beta$ -Heptapeptids in Methanol konnten die Enthalpie- ( $\Delta H_{\text{Solut}}$ ) und Entropiedifferenzen ( $\Delta S_{\text{Solut}}$ ) des Peptids zwischen der gefalteten Konformation und den entfalteten Konformationen berechnet werden: Aus  $\Delta H_{\text{Solut}}^{\text{Faltung}} = -64 \text{ kJ mol}^{-1}$  und  $T\Delta S_{\text{Solut}}^{\text{Faltung}} = -157 \text{ kJ mol}^{-1}$  ergibt sich eine Änderung der freien Enthalpie von  $\Delta G_{\text{Solut}}^{\text{Faltung}} = +93 \text{ kJ mol}^{-1}$ . Berechnet man allerdings die freie Faltungsenthalpie des Gesamtsystems, d.h. Peptid plus Lösungsmittel, aus dem

Verhältnis zwischen gefalteten und ungefalteten Konformationen, dann ergibt sich ein Wert von  $\Delta G^{\text{Faltung}} = -8 \text{ kJ mol}^{-1}$ . Änderungen in der freien Enthalpie des gelösten Peptids allein können daher das beobachtete Faltungsverhalten nicht erklären. Dies unterstreicht die wichtige Rolle des Lösungsmittels in der Peptidfaltung und zeigt, dass Entropierechnungen die Freiheitsgrade des Lösungsmittels berücksichtigen müssen. Leider erschwert ein aufwändiges Absuchen der Lösungsmittelfreiheitsgrade das Simulationsproblem.

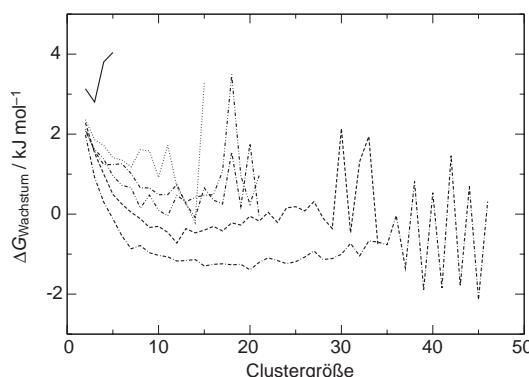
Abbildung 12 illustriert anhand eines Beispiels, dass ein ganzes Ensemble von relevanten Konformationen berück-



**Abbildung 12.** Berechnung der freien Energien für die Bindung von 16 hydroxylierten polychlorierten Biphenylderivaten an die ligandenbindende Domäne des Östrogenrezeptors: Gezeigt sind fünf Protein-Ligand-Strukturen, die am stärksten zur freien Bindungsenergie eines ausgewählten Liganden beitragen. Zur Berechnung wurden MD-Simulationen und die einstufige Störungstechnik (one-step perturbation) verwendet.<sup>[149]</sup>

sichtigt werden muss, um einen guten Schätzwert für die freie Energie zu erhalten. Gezeigt ist eine Überlagerung der Konformationen eines Protein-Ligand-Komplexes, die am stärksten zur freien Bindungsenergie des Biphenylliganden mit dem Östrogenrezeptor beitragen. Die Konformationen weichen beträchtlich voneinander ab, und würde man eine einzelne Konformation als repräsentativ für das gesamte Ensemble annehmen, erhielte man nur einen sehr ungenauen Schätzwert für die freie Bindungsenergie.<sup>[149]</sup> Im Widerspruch zu einer vereinfachenden Annahme, die vielen standardmäßig genutzten Ligand-Protein-Dockingalgorithmen zugrundeliegt, zeigt dieses Beispiel, dass die Freiheitsgrade eines Proteins offenbar im Abtastprozess berücksichtigt werden müssen, um gute Ergebnisse zu erhalten. Leider erschwert auch dies das Such- und Abtastproblem von Dockingalgorithmen.

Ein letztes Beispiel für eine Erschwerung des Suchproblems betrifft die Abhängigkeit des hydrophoben Effekts von der Größe eines hydrophoben Clusters und der Zusammensetzung des Lösungsmittels. Da der hydrophobe Effekt vermutlich entscheidend an der Proteinfaltung beteiligt ist, müssen die Zusammenhänge auf molekularer Ebene aufgeklärt werden. Abbildung 13 zeigt die freie Energie des Clusterwachstums für Cluster aus 2 bis 46 Methanmolekülen in Harnstoff-Wasser-Mischungen bei unterschiedlichen Harn-

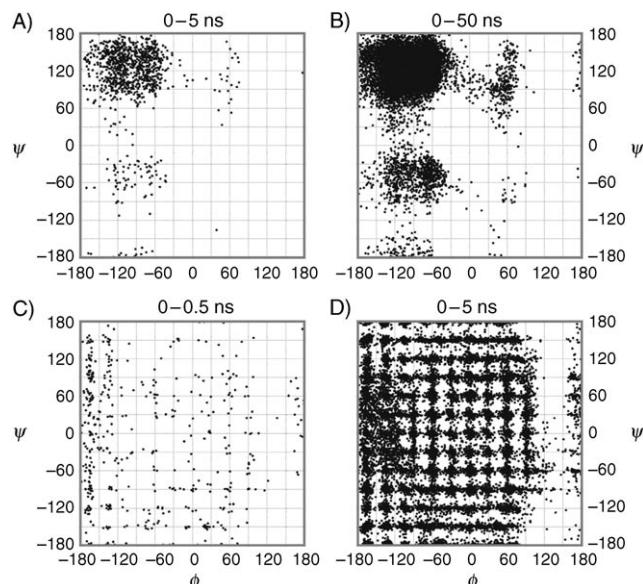


**Abbildung 13.** Freie Enthalpie des Clusterwachstums von Methanclustern als Funktion der Clustergröße gemäß MD-Simulationen für unterschiedliche Zusammensetzungen Methan/Harnstoff/Wasser.<sup>[150]</sup>  $N_m$  ist die Zahl der Methanmoleküle,  $N_u$  die Zahl der Harnstoffmoleküle und  $N_w$  die Zahl der Wassermoleküle. Durchgezogene Linie:  $N_m=10$ ,  $N_u=0$ ,  $N_w=990$ ; gepunktete Linie:  $N_m=40$ ,  $N_u=0$ ,  $N_w=960$ ; Punkt-Strich-Linie:  $N_m=50$ ,  $N_u=150$ ,  $N_w=800$ ; Doppelpunkt-Strich-Linie:  $N_m=50$ ,  $N_u=250$ ,  $N_w=700$ ; gestrichelte Linie:  $N_m=50$ ,  $N_u=0$ ,  $N_w=950$ ; Punkt-Doppelstrich-Linie:  $N_m=64$ ,  $N_u=250$ ,  $N_w=686$ . Die großen Fluktuationen an den Kurvenenden sind auf eine unzureichende statistische Behandlung der großen Cluster zurückzuführen.

stoffkonzentrationen.<sup>[150]</sup> Die Aggregation von Methan ist erst ab Clustergrößen von fünf Methanmolekülen energetisch günstig. Weiterhin zeigt sich, dass hohe Harnstoffkonzentrationen die Bildung von Clustern aus Methanmolekülen erleichtern und die hydrophoben Wechselwirkungen offenbar nicht verringern. Dieses Ergebnis spricht dagegen, dass die Proteindenaturierung durch Harnstoff über einen Mechanismus verläuft, der auf einer Schwächung von hydrophoben Wechselwirkungen beruht.<sup>[151]</sup> Die beobachtete Abhängigkeit der freien Energie von der Clustergröße und der Harnstoffkonzentration erschwert das Suchproblem, weil beide Freiheitsgrade variiert werden müssen, um aussagekräftige Ergebnisse zu erhalten.

### 3.5. Perspektiven zur Vereinfachung des Suchproblems

Es gibt noch reichlich Spielraum für Verbesserungen der Suchleistung biomolekularer Simulationstechniken. Schon die vergangenen zehn Jahre haben ermutigende Fortschritte gebracht, die auch die Erforschung größerer und wichtiger biomolekularer Systeme erleichtern sollten. Insbesondere Techniken zum Glätten der Potentialfläche könnten die Suchleistung verbessern. Hierzu zählen die Methoden der „weichen“ Atomrumpfe und des lokalen Anstiegs (local elevation, LE), die Verformung der Energiehyperfläche durch Anwendung von Diffusionsgleichungen und, auf einem anderen Niveau, die Formulierung von grobkörnigen Modellen. Das Leistungsvermögen von LE-Simulationen<sup>[85]</sup> bei der Durchmusterung des Konformationsraums eines Dipeptids in wässriger Lösung ist in Abbildung 14 illustriert: In einer deutlich kürzeren Simulationszeit erfasst die LE-MD-Methode einen sehr viel größeren Teil der Ramachandran-Karte als Standard-MD-Simulationen.

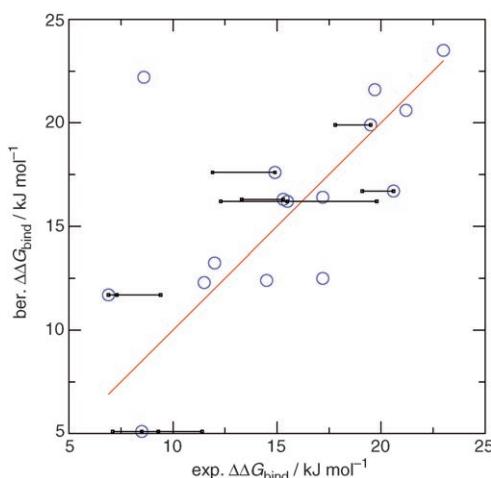


**Abbildung 14.** Ramachandran-Karten aus MD-Simulationen eines Alanindipeptids in Wasser (SPC/E).  $\phi$ - $\psi$ -Verteilungen, die durch Standard-MD über 5 ns (A) und 50 ns (B) oder durch Local-Elevation (LE)-MD über 0.5 ns (C) und 5 ns (D) erhalten wurden,<sup>[85]</sup> belegen die sehr viel bessere Suchleistung des LE-MD-Algorithmus: Wird eine bestimmte Konfiguration ein weiteres Mal generiert, so wird die zugehörige LE-Potentialenergie um  $2.0 \text{ kJ mol}^{-1}$  erhöht.

Die einstufige Störungsmethode ermöglicht die Bestimmung von freien Energien der Ligandenbindung oder freien Solvationsenergien für eine große Zahl von Verbindungen. Die Methode benötigt nur wenige Simulationen, die zur Vergrößerung des Konfigurationsensembles nichtphysikalische Referenzzustände mit „weichen“ Atomrumpfen einzubeziehen. Mit dieser Methode ist die Berechnung von freien Energien um Größenordnungen effizienter als mit Standardtechniken (thermodynamischer Integration oder Störungsmethoden).<sup>[149, 152–156]</sup> Abbildung 15 veranschaulicht dies: Die freien Enthalpien für die Bindung von 16 hydroxylierten polychlorierten Biphenylderivaten an den Östrogenrezeptor wurden mithilfe der einstufigen Störungstechnik aus nur zwei MD-Simulationen berechnet.<sup>[149]</sup> Die mittlere Abweichung der simulierten Werte von den experimentellen Werten ist mit  $2.5 \text{ kJ mol}^{-1}$  kleiner als die Abweichung zwischen experimentellen Werten aus unterschiedlichen Quellen ( $4.2 \text{ kJ mol}^{-1}$ ). Folglich sind das angewendete Kraftfeld und die Suchtechnik auf der Basis „weicher“ Atomrumpfe in der Lage, die experimentellen Werte innerhalb der Messgenauigkeit zu reproduzieren.

Auch eine Verringerung der Lösungsmittelviskosität (siehe Abbildung 10) kann die Suchleistung beträchtlich erhöhen, ohne dass nichtdynamische Gleichgewichtseigenschaften des Systems beeinflusst werden.

Die Beobachtung, dass der ungefaltete Zustand von Polypeptiden sehr viel weniger relevante als theoretisch mögliche Konformationen umfasst, lässt hoffen, dass man schon bald in der Lage sein wird, die reversible Faltung/Entfaltung von kleinen Proteinen zu simulieren (Abbildung 16).



**Abbildung 15.** Vergleich zwischen experimentellen und berechneten relativen freien Enthalpien für die Bindung von 16 hydroxylierten polychlorierten Biphenylderivaten an die Bindungsdomäne des Östrogenrezeptors in Lösung.<sup>[149]</sup> Die horizontalen Linien verbinden unterschiedliche experimentelle Werte für eine Verbindung. Die berechneten Werte wurden aus nur zwei Simulationen und der einstufigen Störungsmethode erhalten. Als Kraftfeld wurde GROMOS43A1 verwendet.

	Zahl der Aminosäuren im Protein	Faltungszeit (exp./sim.) in Sekunden	Anzahl möglicher Strukturen	Anzahl relevanter Strukturen (beobachteter Strukturen)
Peptid	10	$10^{-8}$	$3^{20} \approx 10^9$	$10^3$
Protein	100	$10^{-2}$	$3^{200} \approx 10^{90}$	$10^9$

Unter der Annahme, dass die Zahl der relevanten entfalteten Strukturen proportional zur Faltungszeit ist, müssen nur  $10^9$  anstatt  $10^{90}$  Strukturen simuliert werden.

- ⇒ Der Faltungsmechanismus ist einfacher als allgemein angenommen wird: Durchmusterung von nur  $10^9$  Strukturen
- ⇒ Bis 2010 wird es möglich sein, die Proteinfaltung auf einem Computer zu simulieren

**Abbildung 16.** Ein überraschendes Resultat aus vielen Peptidsimulationen: Die Anzahl entfalteter Konformationen, die in MD-Simulationen von Faltungs-Entfaltungs-Gleichgewichten von Polypeptiden und Peptoiden generiert wird, ist sehr viel kleiner als die Zahl der theoretisch möglichen Konformationen.<sup>[144–146]</sup>

#### 4. Das Ensembleproblem

Biomolekulare Modellierungen werden dadurch erschwert, dass sich biomolekulare Systeme gemäß den Gesetzen der statistischen Mechanik verhalten. Wäre lediglich eine konventionelle Mechanik anzuwenden, könnte man solche Systeme durch Strukturen am globalen Energieminimum charakterisieren. Die statistische Mechanik führt die Entropie eines Systems ein, d.h. die negative Ableitung der freien Energie  $F$  nach der Temperatur [Gl. (16)]:

$$S_{NVT} = -\left(\frac{\partial F}{\partial T}\right)_{NVT} = (U - F)/T \quad (16)$$

Die Entropie legt zusammen mit der inneren Energie des Systems [Gl. (17)], d.h. dem Mittelwert des Hamilton-Ope-

rators des Systems über die Impulse und Koordinaten aller Freiheitsgrade, die freie Energie [Gl. (14)] des Systems fest.

$$U_{NVT} = \left( \frac{\partial (F/T)}{\partial (1/T)} \right)_{NVT} = \langle H(\mathbf{p}, \mathbf{r}) \rangle_{\mathbf{p}, \mathbf{r}} \quad (17)$$

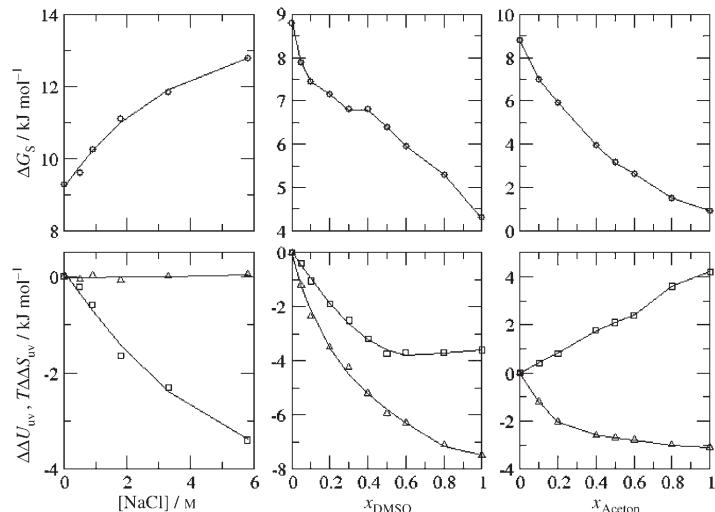
Der Boltzmann-Mittelwert  $\langle \cdots \rangle_{\mathbf{p}, \mathbf{r}}$  einer Größe  $Q(\mathbf{p}, \mathbf{r})$ , die von den Atomkoordinaten und -impulsen abhängt, ist durch Gleichung (18) definiert:

$$\langle Q(\mathbf{p}, \mathbf{r}) \rangle_{\mathbf{p}, \mathbf{r}} = \frac{\int \int Q(\mathbf{p}, \mathbf{r}) \exp[-H(\mathbf{p}, \mathbf{r})/k_B T] d\mathbf{p} d\mathbf{r}}{\int \int \exp[-H(\mathbf{p}, \mathbf{r})/k_B T] d\mathbf{p} d\mathbf{r}} \quad (18)$$

Der Zustand eines Systems wird im Allgemeinen nicht nur durch eine einzelne Konfiguration oder Struktur, sondern durch ein Boltzmann-Ensemble von Konfigurationen oder Strukturen bestimmt. Dies erschwert biomolekulare Modellierungen, da es nicht ausreicht, einzelne Strukturen zu betrachten, sondern ein ganzes Ensemble von Konfigurationen generiert werden muss. Eine Reihe von experimentellen Beobachtungen kann nur verstanden werden, wenn man alternative Strukturen in einem Ensemble sowie entropische Einflüsse in Betracht zieht.

#### 4.1. Freie Energie, innere Energie und Entropie der Solvatation

Weil für die freie Energie die Beziehung  $F = U - TS$  gilt, können unterschiedliche Kombinationen von Werten der inneren Energie ( $U$ ) und der Entropie ( $S$ ) gleiche Werte der freien Energie ergeben. Veranschaulichen lässt sich dies durch eine MD-Simulation der freien Solvationsenergie  $\Delta F_S$  eines Stoffs in einem binären Lösungsmittelgemisch.<sup>[8]</sup> Abbildung 17 zeigt die freie Enthalpie  $\Delta G_S$  (die zu  $\Delta F_S$



**Abbildung 17.** Freie Enthalpie  $\Delta G_S$  für die Solvation von Methan (obere Reihe) sowie Wechselwirkungsenergie  $\Delta \Delta U_{uv}$  (Dreiecke) zwischen Lösungsmittel und gelöstem Stoff und die zugehörige Entropie  $T \Delta \Delta S_{uv}$  (Quadrat) im Verhältnis zu reinem Wasser (untere Diagramme) als Funktion der NaCl-Konzentration (linke Diagramme), des Stoffmengenanteils DMSO (mittlere Diagramme) und des Stoffmengenanteils Aceton (rechte Diagramme).<sup>[8]</sup> Zur Berechnung wurde die „Particle-Insertion“-Technik (Teilcheninsertionstechnik) verwendet.

äquivalente Größe bei konstantem Druck) für die Solvatation eines Methanmoleküls in einem Gemisch aus Wasser und einem Cosolvans als Funktion der Konzentration oder des Stoffmengenanteils des Cosolvans (NaCl, DMSO oder Aceton). Bei Zusatz von NaCl steigt  $\Delta G_S$  mit steigender Konzentration, während  $\Delta G_S$  mit einem steigenden Stoffmengenanteil an DMSO und Aceton sinkt. Die freie Enthalpie  $\Delta G_S$  kann zerlegt werden in einen energetischen Beitrag (die Änderung der Wechselwirkungsenergie bei der Solvatation,  $\Delta U_{uv}$ ) und in einen entsprechenden entropischen Beitrag  $-T\Delta S_{uv}$  [Gl. (19)]:<sup>[8,157–159]</sup>

$$\Delta G_S = \Delta U_{uv} - T\Delta S_{uv} \quad (19)$$

Die untenen Diagramme in Abbildung 17 zeigen diese energetischen und entropischen Beiträge im Verhältnis zu reinem Wasser (angedeutet durch  $\Delta\Delta$ ), ebenfalls als Funktion der Konzentration und des Stoffmengenanteils des Cosolvans. Die Solvatation von Methan in wässriger Lösung wird mit steigender NaCl-Konzentration ungünstiger; Ursache hierfür ist die zunehmend ungünstiger werdende Entropie der Solvatation. Auf der anderen Seite wird die Solvatation von Methan in wässriger Lösung mit steigendem Stoffmengenanteil an DMSO oder Aceton immer günstiger. Die relativen Beiträge von Energie und Entropie sind allerdings deutlich verschieden, obwohl DMSO und Aceton ähnliche Molekülstrukturen haben. Die Solvatation in DMSO-haltigen Lösungen wird mit steigendem DMSO-Gehalt vom günstigen Energieterm bestimmt, während in acetonhaltigen Lösungen der günstige Entropieterm dominierend wird. Die vergleichbar aussehenden Kurven für  $\Delta G_S$  sind also auf unterschiedliche Solvationsmechanismen zurückzuführen. Dieses Beispiel verdeutlicht, dass bei biomolekularen Modellierungen die Entropie sorgfältig berücksichtigt werden sollte.

Die wechselnden Rollen von Energie und Entropie bei der hydrophoben Solvatation sind in Tabelle 8 für einige hydrophobe Moleküle in unterschiedlichen wässrigen Lösungen veranschaulicht. In NaCl-Lösung steigt  $\Delta\Delta G_S$  für größere Molekülen, da  $-T\Delta\Delta S_{uv}$  zunimmt. In einer Harnstoff-Wasser-Mischung kommt es zu einer Energie-Entropie-Kompensation, sodass  $\Delta\Delta G_S$  kaum von der Molekülgröße abhängt. In DMSO wird  $\Delta\Delta G_S$  günstiger, je größer die gelösten Moleküle sind; der Grund hierfür ist ein sehr günstiger energetischer

Beitrag bei einem nur schwach entgegenwirkenden Entropieterm. Beim Übergang von DMSO zu Aceton ändert sich die Situation erneut, da nun der Entropieterm mit dem Solvationsenergieterm zusammenwirkt. Der Effekt scheint sich mit steigendem Acetongehalt zu verstärken. Diese Ergebnisse zeigen, dass ähnliche Änderungen der freien Energie, genauso wie die davon angetriebenen Prozesse, auf sehr unterschiedlichen Mechanismen beruhen können. Um solche Prozesse auf molekularer Ebene erfassen zu können, ist eine geeignete Modellierung der entropischen Einflüsse notwendig.

## 4.2 Temperaturabhängigkeit der Faltungs-Entfaltungs-Gleichgewichte

Das Faltungs-Entfaltungs-Gleichgewicht eines Polypeptids wird durch Änderungen sowohl der Energie als auch der Entropie während des Faltungs- oder Entfaltungsprozesses bestimmt. Daraus folgt, dass das zugehörige Ensemble von Konformationen und die Faltungswäge temperaturabhängig sind. Abbildung 18 zeigt die am häufigsten vorkommenden Konformationen eines  $\beta$ -Heptapeptids in Methanol; die Daten stammen aus MD-Simulationen bei drei unterschiedlichen Temperaturen.<sup>[138]</sup> Die vertikalen Pfeile verweisen auf zusammengehörige Konformationen bei unterschiedlichen Temperaturen; die horizontalen Pfeile markieren die wichtigsten Faltungs- und Entfaltungswege zur bzw. ausgehend von der stabilsten (helicalen) Konformation, C1. Wie erwartet, sind sowohl das Ensemble von Konformationen als auch die Faltungs- und Entfaltungswege temperaturabhängig. Mit MD-Simulationen bietet sich eine Möglichkeit, diese Abhängigkeit zu studieren.

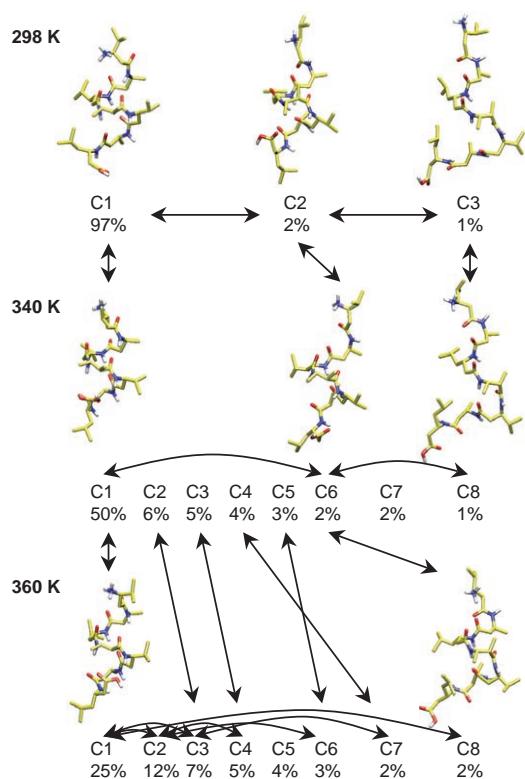
## 4.3 Verschiedene Konformationen können zum Ensemblemittelwert beitragen

Falls in Lösung unterschiedliche Konformationen im Ensemble eines Polypeptids vorliegen, dann kann es durchaus sein, dass der gemessene Mittelwert einer Observablen, z. B. einer NOE-Intensität oder einer  $^3J$ -Kopplungskonstanten, keiner realistischen (d. h. energetisch erreichbaren) Konfor-

**Tabelle 8:** Thermodynamische Daten [ $\text{kJ mol}^{-1}$ ] für den Transfer von gelösten Stoffen aus reinem Wasser in Cosolvans-Wasser-Mischungen bei 298 K und 1 atm (berechnet mit MD-Simulationen).<sup>[9]</sup> Die Prozentangaben bezeichnen Stoffmengenanteile. Für Aceton wurden zwei verschiedene Modelle verwendet (I und II).

gelöster Stoff	NaCl (11%)			Harnstoff (15%)			DMSO (10%) <sup>[a]</sup>			Aceton(I) (10%)			Aceton(I) (50%)			Aceton(II) (10%)		
	$\Delta\Delta G_S$	$\Delta\Delta U_{uv}$	$T\Delta\Delta S_{uv}$	$\Delta\Delta G_S$	$\Delta\Delta U_{uv}$	$T\Delta\Delta S_{uv}$	$\Delta\Delta G_S$	$\Delta\Delta U_{uv}$	$T\Delta\Delta S_{uv}$	$\Delta\Delta G_S$	$\Delta\Delta U_{uv}$	$T\Delta\Delta S_{uv}$	$\Delta\Delta G_S$	$\Delta\Delta U_{uv}$	$T\Delta\Delta S_{uv}$	$\Delta\Delta G_S$	$\Delta\Delta U_{uv}$	$T\Delta\Delta S_{uv}$
Helium	2.0	-0.4	-2.4	1.3	-0.1	-1.4	0.0	-0.2	-0.2	-0.4	-0.2	0.2	-1.7	-0.6	1.1	-0.1	-0.2	-0.1
Neon	2.8	0.2	-2.6	1.1	-0.7	-1.8	-0.3	-0.7	-0.4	-0.7	-0.4	0.3	-2.4	-1.0	1.4	-0.4	-0.6	-0.2
Argon	3.7	-0.1	-3.8	0.8	-2.3	-3.1	-0.9	-1.8	-0.9	-1.3	-1.0	0.3	-4.4	-2.1	2.3	-1.1	-1.7	-0.6
Krypton	4.0	-0.2	-4.2	0.4	-3.2	-3.6	-1.2	-2.3	-1.1	-1.6	-1.2	0.4	-5.2	-2.6	2.6	-1.5	-2.3	-0.8
Xenon	4.4	-1.2	-5.6	-0.5	-4.7	-4.2	-1.5	-2.7	-1.2	-2.2	-1.7	0.5	-6.8	-3.6	3.2	-2.1	-3.4	-1.3
Methan	4.2	-0.2	-4.4	0.5	-3.1	-3.6	-1.1	-2.1	-1.0	-1.7	-1.2	0.5	-5.4	-2.6	2.8	-1.4	-2.2	-0.8
Ethan	5.6	-0.8	-6.4	-0.9	-6.0	-5.1	-2.8	-4.4	-1.6	-2.9	-2.4	0.5	-8.4	-4.6	3.8	-2.6	-4.3	-1.7
Propan	5.4	-0.3	-5.7	-2.2	-7.3	-5.1	-4.8	-5.8	-1.0	-5.8	-3.9	1.9	-13.0	-6.5	6.5	-4.8	-6.2	-1.4
n-Butan	7.4	-1.2	-8.6	-2.7	-10.6	-7.9	-5.8	-7.9	-2.1	-6.4	-5.6	0.8	-15.7	-9.3	6.4	-4.3	-6.0	-1.7
Isobutan	5.7	-1.2	-6.9	-2.3	-10.2	-7.9	-6.5	-6.7	-0.2	-6.7	-4.9	1.8	-15.2	-7.7	7.5	-5.9	-7.7	-1.8
Neopentan	8.2	-0.6	-8.8	-2.2	-10.9	-8.7	-6.0	-8.3	-2.3	-6.7	-4.3	2.4	-15.8	-7.1	8.7	-5.3	-9.7	-4.4

[a] Um den Konfigurationsraum schneller durchsuchen zu können, wurden die MD-Simulationen mit einer Atommasse von 15.9994 amu für die Schwefelatome des DMSO und die Wasserstoffatome des Wassers ausgeführt. Damit war ein MD-Zeitschritt von 4 fs möglich.



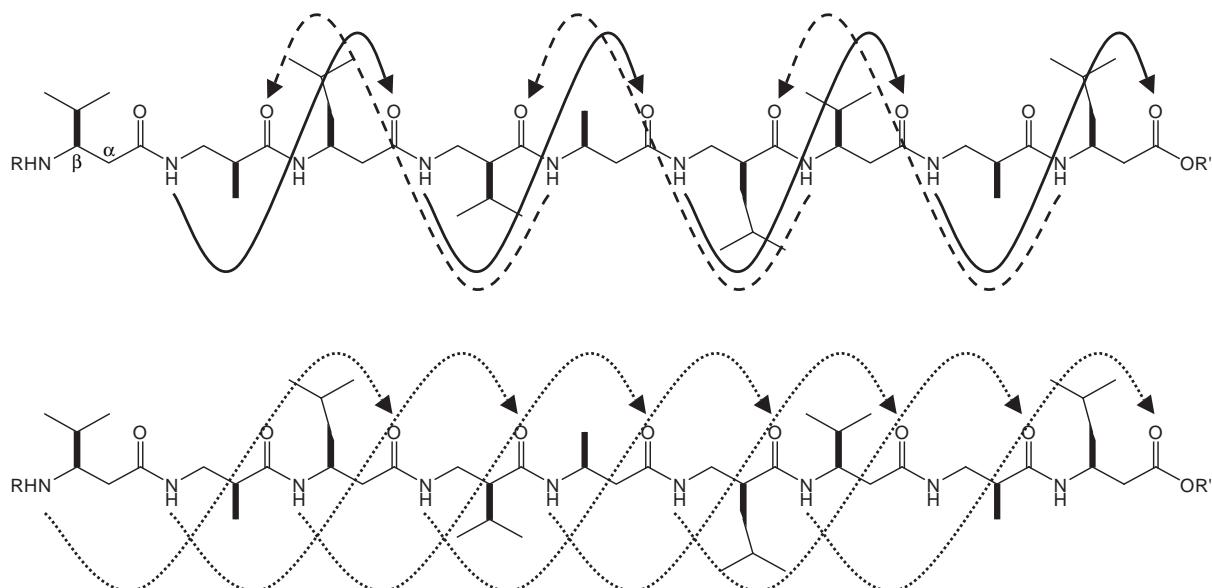
**Abbildung 18.** Häufigste Konformationen eines  $\beta$ -Heptapeptids in Methanol, aus MD-Simulationen bei drei unterschiedlichen Temperaturen.<sup>[138]</sup> Die Pfeile in vertikaler Richtung verbinden zusammengehörige Konformationen bei unterschiedlichen Temperaturen; die horizontalen Pfeile markieren die dominanten Faltungs- und Entfaltungswege zur stabilsten (helicalen) Konformation C1.

mation entspricht. Das Problem ist noch gravierender, wenn die Observable nichtlinear von der Polypeptidkonformation abhängt. In einem solchen Fall kann die Strukturbestimmung aus NMR-Daten nur dann mit den experimentell gemessenen

Werten der Observablen in Einklang gebracht werden, wenn keine Einzelstrukturen, sondern Ensembles von Konformationen betrachtet werden. Ein Beispiel für einen solchen Fall ist das in Abbildung 19 gezeigte  $\beta$ -Nonapeptid. Eine Einzelstrukturverfeinerung auf der Grundlage von 65 NOE-Intensitäten und vier  $^3J$ -Kopplungskonstanten, die für das  $\beta$ -Nonapeptid (ohne Schutzgruppen) in Methanolösung gemessen wurden, ergab eine 12/10-helicale Modellstruktur. Das für diesen Helixtyp charakteristische Wasserstoffbrückenmuster ist in Abbildung 19 dargestellt (volle und gestrichelte Pfeile). Bei der Verfeinerung waren jedoch drei schwache NOE-Intensitäten ( $2\text{H}\text{N}-5\text{H}\beta$ ,  $6\text{H}\text{N}-3\text{H}\beta$ ,  $7\text{H}\text{N}-4\text{H}\beta_{\text{RC}}$ ) weggelassen worden, da sie mit der 12/10-Helix nicht kompatibel waren.<sup>[160]</sup> Eine 100-ns-MD-Simulation des  $\beta$ -Nonapeptids in Methanol (ohne die NOE-Intensitäten und  $^3J$ -Kopplungskonstanten als einschränkende Nebenbedingungen) reproduzierte dagegen alle gemessenen NOE-Intensitäten und  $^3J$ -Werte.<sup>[161]</sup> Eine Analyse der MD-Trajektorie ergab, dass die drei erwähnten NOE-Intensitäten auf einen kleinen Anteil einer alternativen  $3_{14}$ -helicalen Konformation im Ensemble zurückzuführen sind (Abbildung 19, gepunktete Pfeile). Dieses Beispiel belegt, dass MD-Simulationen mit konsistenten, thermodynamisch kalibrierten Kraftfeldern für den gelösten Stoff und das Lösungsmittel bei expliziter Behandlung der Lösungsmittelfreiheitsgrade zu einer korrekten Interpretation von experimentellen Daten auf der Grundlage von Konformationsverteilungen beitragen können, ohne dass die experimentellen Daten von vornherein als Randbedingungen bei der Durchführung der Simulation berücksichtigt werden.

#### 4.4 Perspektiven zur Berechnung von Entropien

Während die Bestimmung von Differenzen der freien Energie mit MD-Simulationen zu einer Standardprozedur



**Abbildung 19.** Strukturformel eines  $\beta$ -Nonapeptids; die Pfeile kennzeichnen das für 12/10- und  $3_{14}$ -Helices charakteristische Wasserstoffbrückenmuster. Die 12/10-Helix (obere Struktur) ist durch 10- (volle Pfeile) und 12-gliedrige (gestrichelte Pfeile) Wasserstoffbrückenringe charakterisiert; die  $3_{14}$ -Helix (untere Struktur) weist 14-gliedrige Wasserstoffbrückenringe auf (gepunktete Pfeile).

gereift ist, für die eine Vielzahl von Techniken entwickelt wurde, sind Entropien und Entropiedifferenzen bisher nur selten berechnet worden. Die Entropie ist allerdings von entscheidender Bedeutung für Phänomene wie hydrophobe Wechselwirkungen, Solvatation, Ligandenbindung etc. Leider ist es keineswegs einfach, absolute Entropien oder Entropiedifferenzen aus MD-Simulationen zu bestimmen, da hierzu im Prinzip der gesamte Phasenraum durchsucht werden muss. Im Allgemeinen kann man zwei Gruppen von Methoden unterscheiden, die aus MD-Simulationen vernünftige Schätzwerke für Entropien ableiten.

Ein Ansatz stützt sich auf Konformationsentropien, die nicht alle, sondern nur die internen (konformativen) nicht-diffusiven Freiheitsgrade einschließen,<sup>[162–165]</sup> z.B. in einem Protein.<sup>[166]</sup>

Der zweite Ansatz bedient sich Techniken, die erfolgreich zur Bestimmung von Differenzen der freien Energie eingesetzt worden sind, um auch Entropiedifferenzen zu berechnen. Differenzen der freien Energie zwischen zwei Zuständen eines Systems oder zwischen zwei Systemen erhält man, ohne die vollständigen Verteilungsfunktionen zu kennen. Vielmehr genügt es, nur die relevanten Teile des Phasen- oder des Konfigurationsraumes gründlich abzusuchen, d.h. diejenigen, in denen sich die beiden Zustände oder Systeme unterscheiden. Die entsprechenden Techniken zur Bestimmung von Entropiedifferenzen sind allerdings mit dem Problem behaftet, dass sie einen genauen Schätzwert für einen Ensemblemittelwert benötigen, der den vollständigen Hamilton-Operator  $H$  des Systems in beiden Zuständen betrifft. Es genügt also nicht, nur den Teil des Hamilton-Operators zu betrachten, in dem sich die beiden Zustände oder Systeme unterscheiden ( $\partial H / \partial \lambda$ ; siehe Abbildung 20).<sup>[142]</sup> Der vollständige Hamilton-Operator ist eine Summe über viele Terme, von denen sich nur einige wenige in den beiden Zuständen des Systems unterscheiden. Daher ist es sehr zeitaufwändig, präzise Ensemblemittelwerte zu erhalten. Am genauesten arbeiten die Methoden 3 und 4 in Abbildung 20. Methode 4 liefert genaue Werte für  $\delta S_{uv}$ , die sämtliche Lösungsmitteltermen einschließen, aber keine Solvationsentropien. Sie wurde verwendet, um die in Tabelle 8 und Abbildung 17 gezeigten Daten zu berechnen.

Der zuerst genannte Ansatz eignet sich nur zur Berechnung von Konformationsentropien des gelösten Moleküls und kann bei diffusiven Systemen wie Lösungsmitteln nicht angewendet werden. Die zur Berechnung der Entropiedifferenzen verwendeten Ansätze (Abbildung 20) wurden kürzlich überprüft und bewertet.<sup>[142]</sup> Trotz der Fortschritte bei der Methodenentwicklung ist die genaue Berechnung von Entropien auch hier schwierig. Keine der untersuchten Techniken scheint für die Berechnung der Entropie des Ligand-Protein-Bindungsvorgangs oder der Entropie der Polypeptidfaltung geeignet zu sein.

## 5. Das Experimentalproblem

Experimentelle Daten spielen eine entscheidende Rolle bei biomolekularen Modellierungen. Zunächst bilden sie die Grundlage für die Entwicklung von klassischen Kraftfeldern

## Berechnung von Energiedifferenzen

mithilfe des Kopplungsparameters  $\lambda$ :

Hamiltonian wird Funktion von  $\lambda$ :

$H_a(\vec{p}, \vec{r}) = H(\vec{p}, \vec{r}, \lambda_a) \rightarrow$  Zustand  $a$

$H_b(\vec{p}, \vec{r}) = H(\vec{p}, \vec{r}, \lambda_b) \rightarrow$  Zustand  $b$

Freie Energie hängt von  $\lambda$  ab:

$$F_{NVT}(\lambda) = -k_b T \ln \left( h^{3N} N! \right)^{-1} \iint \exp(-H(\vec{p}, \vec{r}, \lambda) / k_b T) d\vec{p} d\vec{r}$$

### 1. Entropiedifferenz aus Energiedifferenz und thermodynamischer Integration (TI) der freien Energien

Differenzen von freien Energien und Endzustandsenergien

$$\Delta F_{ba}^T = F(\lambda_b) - F(\lambda_a) = \int_{\lambda_a}^{\lambda_b} \frac{dF}{d\lambda} d\lambda = \int_{\lambda_a}^{\lambda_b} \left\langle \frac{\partial H}{\partial \lambda} \right\rangle_\lambda d\lambda \quad \text{genau}$$

$$\Delta U_{ba} = \langle H \rangle_{\lambda_b} - \langle H \rangle_{\lambda_a} = U(\lambda_b) - U(\lambda_a) = \Delta U_{ba}^{\text{end}} \quad \text{unge nau}$$

$$\Delta S_{ba}^T = \frac{\Delta U_{ba}^{\text{end}} - \Delta F_{ba}^T}{T} \quad \text{weniger exakt wegen Energiegenauigkeit}$$

### 2. Entropiedifferenz direkt aus (TI)

$$\text{mit } S = - \left\langle \frac{\partial F}{\partial T} \right\rangle_{N,V} \Rightarrow \Delta S_{ba}^T = \frac{1}{k_b T^2} \int_{\lambda_a}^{\lambda_b} \left[ \left\langle \frac{\partial H}{\partial \lambda} \right\rangle_\lambda \langle H \rangle_\lambda - \left\langle \frac{\partial H}{\partial \lambda} H \right\rangle_\lambda \right] d\lambda$$

$$\text{und } \left\langle \frac{\partial S}{\partial \lambda} \right\rangle_{N,V,T} \quad \begin{matrix} \text{Korrelation von } \frac{\partial H}{\partial \lambda} \text{ und } H \text{ weniger genau} \\ \uparrow \qquad \uparrow \\ \text{nur } \lambda\text{-abhängige Terme} \qquad \text{alle Terme} \end{matrix}$$

### 3. Entropiedifferenz aus endlichen Temperaturdifferenzen

$$\text{mit } S = - \left\langle \frac{\partial F}{\partial T} \right\rangle_{N,T} \Rightarrow \Delta S_{ba}^{\Delta T} = - \frac{\Delta F_{ba}^T(T + \Delta T) - \Delta F_{ba}^T(T - \Delta T)}{2\Delta T}$$

Unterschied zwischen fast gleichen genauen Werten

### 4. Entropiedifferenz der Solvation aus der Entropiedifferenz von Solvens und gelöstem Stoff (mithilfe von TI) und Solvens-Solvens-Entropiedifferenz im Endzustand

$$\begin{aligned} \Delta S_{ba}^{T,all} &= \frac{1}{k_b T^2} \int_{\lambda_a}^{\lambda_b} \left[ \left\langle \frac{\partial H_{vv}}{\partial \lambda} \right\rangle_\lambda \langle H_{vv} \rangle_\lambda - \left\langle \frac{\partial H_{vv}}{\partial \lambda} H_{vv} \right\rangle_\lambda \right] d\lambda + \frac{1}{T} \left[ \langle H_{vv} \rangle_{\lambda_b} - \langle H_{vv} \rangle_{\lambda_a} \right] \\ &= \Delta S_{ba}^{T,uv} + \frac{\Delta U_{ba}^{\text{end},vv}}{T} \\ &\begin{matrix} \text{genau} & \text{weniger genau} & \text{Solvens: } v \\ \uparrow & \uparrow & \text{gelöster Stoff: } u \\ \text{Terme von Solvens} & \text{reine Solvensterme} & \end{matrix} \end{aligned}$$

**Abbildung 20.** Vier Wege zur Berechnung von Entropiedifferenzen durch Anwendung der Kopplungsparametertechnik im Rahmen von MD-Simulationen.<sup>[142]</sup>

(siehe Tabelle 4). Ohne die dort genannten experimentellen Daten wäre dies praktisch unmöglich, da theoretische, quantenchemische Daten allein nicht ausreichen, um ein Kraftfeld zu konstruieren. In einem weiteren Schritt lassen sich Simulationsmethode und Kraftfeld validieren und testen, indem man simulierte oder berechnete Werte von Molekül- oder Systemeigenschaften mit den experimentellen Daten vergleicht.

Der Einsatz experimenteller Daten in der biomolekularen Modellierung ist jedoch auch mit drei Problemen behaftet:

- Bei fast jedem Experiment wird über die Zeit und den Raum oder über die Moleküle gemittelt, sodass für die Konfigurationen, die sich zu einer Simulationstrajektorie zu-

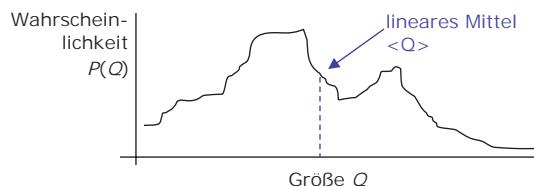
sammensetzen, keine direkten Informationen erhalten werden. 2) Gemessen an der großen Zahl von Freiheitsgraden werden aus Messungen an biomolekularen Systemen nur sehr spärliche experimentelle Daten erhalten. Daher ist das Konformationsensemble eines biomolekularen Systems durch die experimentellen Daten unterbestimmt. Es kann vorkommen, dass viele unterschiedliche Ensembles den gleichen Satz von experimentellen Daten reproduzieren. 3) In einigen Fällen sind experimentelle Daten nicht genau genug, um eine Simulation zu validieren. Diese drei Experimentalprobleme im Zusammenhang mit der biomolekularen Modellierung werden im Folgenden anhand von Beispielen illustriert.

### 5.1. Das Problem der Mittelwertbildung

Die Messung einer Größe  $Q(r)$ , die als Funktion einer Molekül- oder Systemkonfiguration  $r$  ausgedrückt werden kann, liefert keinen Wert  $Q^{\text{obs}} = Q(r)$ , der von einer einzelnen Konfiguration  $r$  abhängt, sondern einen Mittelwert über viele Moleküle im realen (makroskopischen) System und über die Dauer der Messung [Gl. (20); das Symbol  $\langle \dots \rangle$  kennzeichnet eine Mittelwertbildung]:

$$Q^{\text{obs}} = \langle Q(r) \rangle_{\text{Molek\üle, Zeit}} \quad (20)$$

Mit anderen Worten wird nur ein Mittelwert über die Verteilung  $P(Q)$  der Größe  $Q$  gemessen (siehe Abbildung 21), nicht aber die Verteilung selbst. Bei der Mittel-



**Abbildung 21.** Das Mittelwert-Problem: Eine Verteilung von Konformationen, über die ein Mittelwert gebildet wird, kann nicht aus diesem Mittelwert abgeleitet werden.

wertbildung geht die detaillierte Information über die Verteilung verloren, und sehr unterschiedliche Verteilungen können zum gleichen Mittelwert führen. Wie empfindlich der Mittelwert  $\langle Q \rangle$  einer bestimmten Größe  $Q$  von der Form der Verteilung  $P(Q)$  oder der Form der Konformationsverteilung  $P(r)$  abhängt, kann je nach Art der Messgröße  $Q$  variieren.

Als ein Beispiel betrachten wir die Konformationsverteilungen des Octapeptids Aib<sub>6</sub>-Leu-Aib (Aib = Aminoisobuttersäure) in DMSO-Lösung und im Kristall (Abbildung 22). Da dieses Peptid als Hauptkomponente die achiralen helixbildenden Aib-Reste enthält, sollte es in Lösung sowohl eine *R*- als auch eine *L*-helicale Konformation annehmen können, was auch die NOE-Intensitäten und  $^3J$ -Werte aus NMR-Experimenten in Lösung auch bestätigten.<sup>[167]</sup> Dagegen wird im Kristall nur die *R*-helicale Form gefunden.<sup>[168]</sup> Die berechneten  $^3J$ -Werte für die Kristallstruktur sind jedoch mit 4.2 Hz deutlich kleiner als die in Lösung gemessenen Werte von

#### gelöster Stoff:

Polypeptid: (Aib)<sub>6</sub>-Leu-Aib  
 $\uparrow$        $\uparrow$   
 achiral L-Aminosäure

#### Experiment:

**NMR:** NOE-Werte deuten auf eine helicale Struktur in Lösung:  $^3J(\text{H}_N-\text{H}_{C\alpha}) = 6.9 \text{ Hz}$  ← *R*- oder *L*-Helix

**RSA:** *R*-Helix  
 Struktur im Kristall: NOE-Werte bestätigt, aber  $^3J = 4.2 \text{ Hz}$  ?

#### MD-Simulationen:

**DMSO-Lösung:** NOE-Werte bestätigt,  $^3J = 6.8 \text{ Hz}$

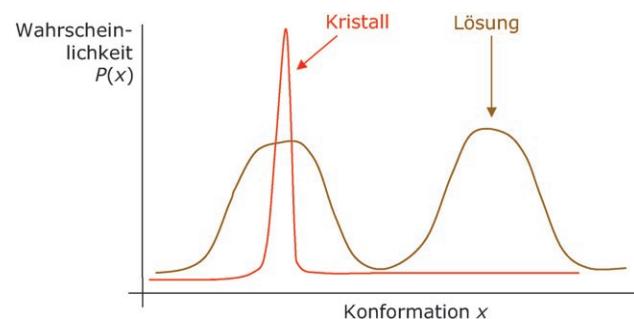
} *L*- und *R*-helicale Fragmente

**Kristall:** *R*-Helix, in Einklang mit RSA  
 NOE-Werte bestätigt  
 $^3J = 4.0 \text{ Hz}$

} nur eine Konformation:  
*R*-Helix

**Abbildung 22.**  $^3J$ -Kopplungskonstanten sind von der Konformationsverteilung eines Octapeptids abhängig, NOE-Intensitäten hingegen nicht.<sup>[169,170]</sup>

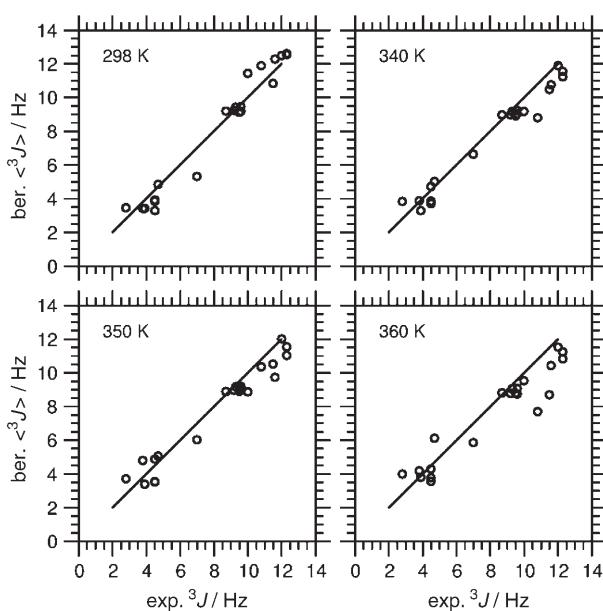
6.9 Hz. Diese Beobachtung weist auf eine unterschiedliche Konformationsverteilung in Lösung und im Kristall hin. MD-Simulationen des Octapeptids in DMSO-Lösung und im Kristall<sup>[169,170]</sup> konnten die gemessenen NOE-Intensitäten,  $^3J$ -Werte und die Röntgenkristallstruktur exakt reproduzieren und zeigten auf, dass in Lösung und im Kristall unterschiedliche Konformationsensembles vorliegen (Abbildung 23). In



**Abbildung 23.** Die Konformationsverteilung in Lösung und im Kristall kann unterschiedlich sein. Im Text wird dies am Beispiel eines Octapeptids erörtert.<sup>[169,170]</sup>

Lösung treten kurzlebige *R*- und *L*-helicale Teilstrukturen auf, die zu einem breiten Ensemble von Konformationen mit dem Wert  $\langle ^3J \rangle = 6.8 \text{ Hz}$  führen, der dem experimentellen Wert gleicht. Im Kristall wird hingegen ein relativ schmales Ensemble von *R*-helicalen Konformationen mit dem Wert  $\langle ^3J \rangle = 4.0 \text{ Hz}$  gefunden, der dem  $^3J$ -Wert der Röntgenkristallstruktur nahekommt. Die Simulationen zeigen, dass die NOE-Intensitäten nur schwach von der Form des Konformationsensembles abhängen, solange helicale Teilstrukturen vorhanden sind. Die NOE-Intensitäten eignen sich daher nicht, um zwischen den Ensembles in Lösung und im Kristall zu unterscheiden. Dagegen spiegeln die  $^3J$ -Kopplungskonstanten in diesem Fall die relativ deutlichen Unterschiede zwischen den beiden Konformationsensembles sehr gut wider.

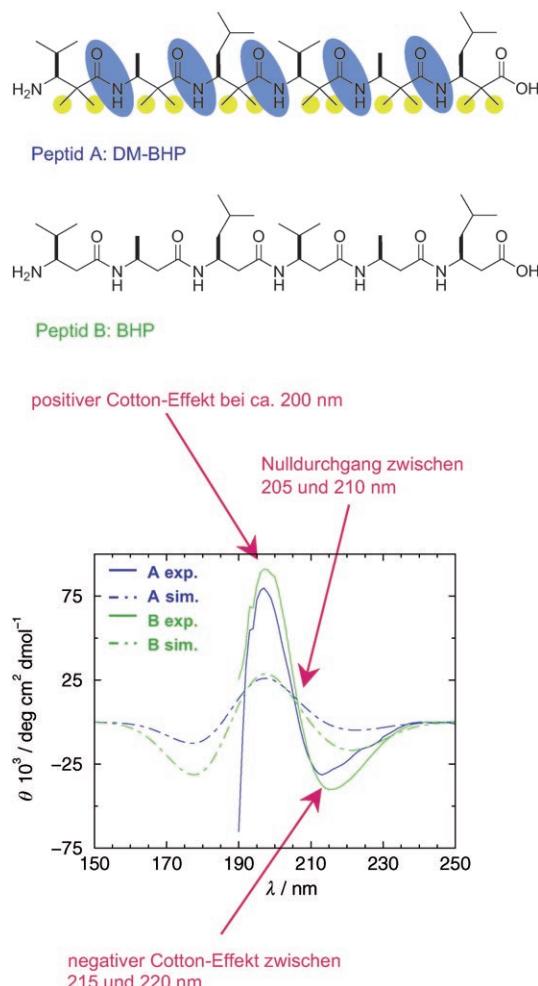
Es kommt auch vor, dass die  $^3J$ -Kopplungskonstanten extrem unempfindlich gegenüber der Konformationsverteilung sind. Dieser Fall ist in Abbildung 24 illustriert, in der simulierte  $^3J$ -Werte für ein  $\beta$ -Heptapeptid in Methanol mit



**Abbildung 24.** Vergleich der 21 gemessenen mittleren  $^3J$ -Kopplungskonstanten bei 298 K mit den gemittelten  $^3J$ -Kopplungskonstanten, die aus den Strukturen eines  $\beta$ -Heptapeptids in Methanol (50-ns-MD-Trajektorien) bei vier Temperaturen berechnet wurden. Die vier Konformationsverteilungen unterscheiden sich relativ deutlich voneinander und enthalten 97, 50, 39 bzw. 25 %  $3_{14}$ -helicale Strukturen.<sup>[171]</sup>

experimentellen Werten verglichen werden.<sup>[171]</sup> Es wurden vier Konformationsverteilungen verwendet (erzeugt für 298, 340, 350 und 360 K), die sich relativ deutlich unterscheiden und 97, 50, 39 bzw. 25 %  $3_{14}$ -helicale Strukturen enthalten.<sup>[171]</sup> Dennoch stimmen die vier Verteilungen gleich gut mit dem Experiment überein.

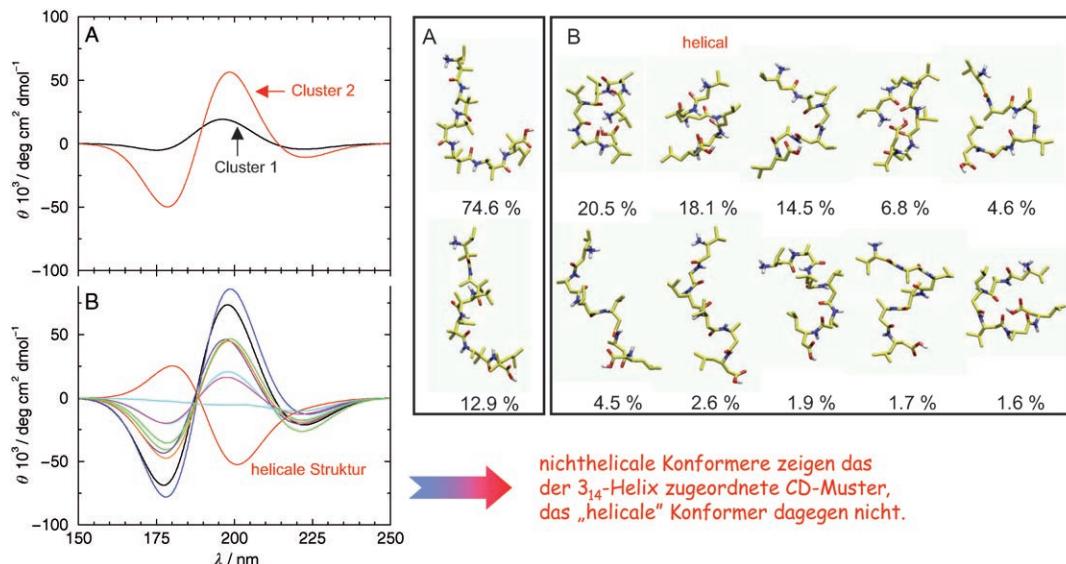
Ein Informationsverlust durch Mittelwertbildung tritt nicht nur bei NMR-Experimenten auf. Ähnliche Beobachtungen wurden auch bei Circulardichroismus(CD)-Spektren gemacht.<sup>[172]</sup> Abbildung 25 zeigt die gemessenen CD-Spektren von zwei  $\beta$ -Hexapeptiden in Methanol. Die Spektren sind sehr ähnlich, obwohl man erwarten sollte, dass sich die Konformationsensembles der beiden Peptide deutlich unterscheiden, verhindert doch die doppelte Methylierung am  $\alpha$ -Kohlenstoffatom des Peptids A die Bildung einer  $3_{14}$ -Helix, wohingegen Peptid B, das sich vom Peptid A nur an dieser Stelle unterscheidet, eine solche Helixform annimmt. Das wurde durch Messungen der NOE-Intensitäten in NMR-Experimenten bestätigt. Um näheren Einblick zu gewinnen, wurden für beide Moleküle 100-ns-Simulationen durchgeführt und mittlere CD-Spektren aus den MD-Trajektorien abgeleitet.<sup>[172]</sup> Die berechneten Spektren ähneln den experimentellen, lediglich die Amplituden waren kleiner. Die wichtigsten Konformationen sind in Abbildung 26 zusammen mit den zugehörigen CD-Spektren gezeigt. Das CD-Spek-



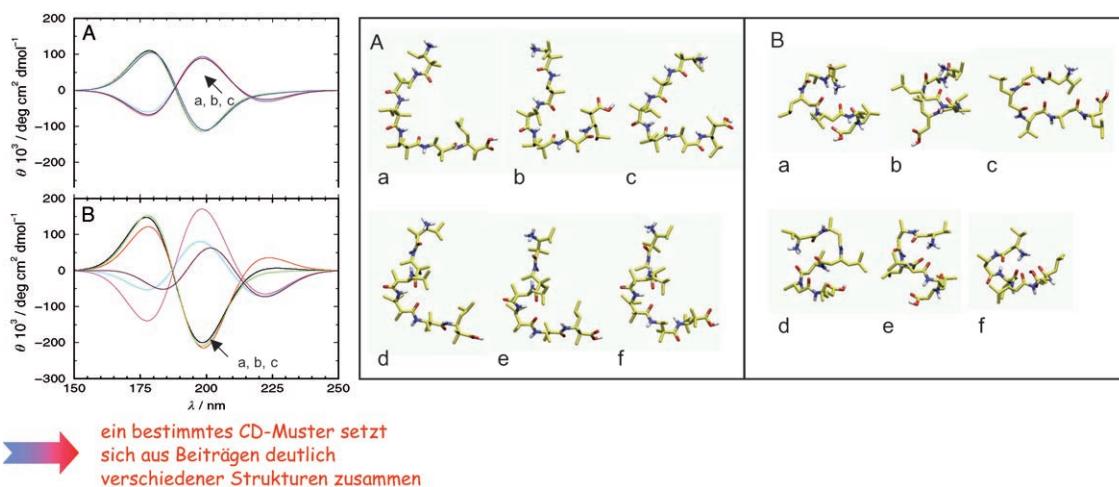
**Abbildung 25.** Experimentelle und durch 100-ns-MD-Simulationen berechnete CD-Spektren von zwei  $\beta$ -Hexapeptiden in Methanol bei 298 K. Peptid B bildet eine  $3_{14}$ -Helix, was durch NMR-Experimente bestätigt wurde. Peptid A ist in  $\alpha$ -Position doppelt methyliert und bildet daher keine  $3_{14}$ -Helix, zeigt aber ein für die  $3_{14}$ -Helix „typisches“ CD-Spektrum.<sup>[172]</sup>

trum von Peptid A wird von zweithäufigsten Konformation bestimmt (13 % des Ensembles), nicht etwa von der häufigsten. Die einzige helicale Konformation des Peptids B (18 %) erzeugt ein Spektrum, das sich von den Spektren anderer Konformationen und vom experimentellen Spektrum deutlich unterscheidet.

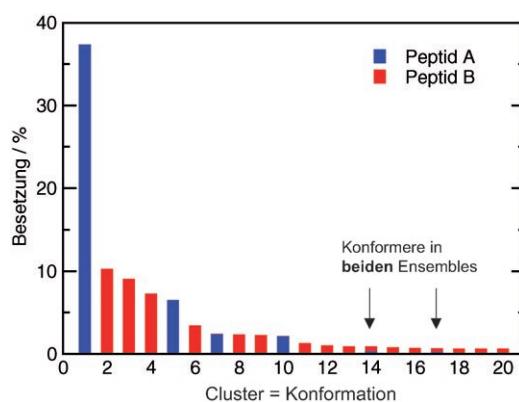
Dies wirft die generelle Frage auf, ob man ein CD-Spektrum überhaupt einer bestimmten Struktur zuordnen kann. Abbildung 27 zeigt für jedes Peptid die sechs CD-Spektren mit den größten Amplituden und die sechs Strukturen auf den MD-Trajektorien, die zu diesen Spektren führen. Es ist klar erkennbar, dass sehr unterschiedliche Strukturen das gleiche Spektrum erzeugen können. Abbildung 28 präsentiert eine Konformations-Clusteranalyse der kombinierten MD-Trajektorien beider Peptide. Zwischen den beiden Konformationsensembles der Peptide gibt es praktisch keine Überlappung, obwohl ihre CD-Spektren sowohl im Experiment als auch in der Simulation sehr ähnlich sind. Diese Beobachtung sollte als Warnung gelten, bei der Interpretation von CD-



**Abbildung 26.** Wichtigste Konformationen aus den MD-Trajektorien der beiden  $\beta$ -Hexapeptide A (methyliert) und B (nicht methyliert; siehe Abbildung 25) sowie die zugehörigen CD-Spektren.<sup>[172]</sup> (Ähnlichkeitskriterium der Clusteranalyse: RMSD der Rückgratatome  $\leq 0.09$  nm; 10000 Strukturen, 10-ps-Intervalle)



**Abbildung 27.** CD-Spektren einzelner Strukturen aus den MD-Trajektorien mit den stärksten CD-Signalen und die zugehörigen Strukturen der beiden  $\beta$ -Hexapeptide A und B.<sup>[172]</sup>

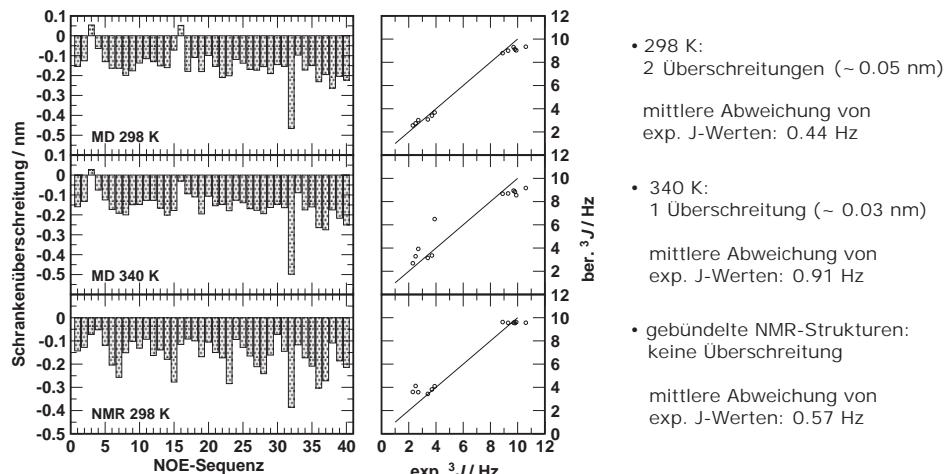


**Abbildung 28.** Die Konformations-Clusteranalyse der kombinierten MD-Trajektorien der  $\beta$ -Hexapeptide A (methyliert, DM-BHP) und B (nicht methyliert, BHP) bestätigt, dass es trotz ähnlicher CD-Spektren praktisch keinerlei Überlappung zwischen den Konformationsensembles der beiden Peptide gibt.<sup>[172]</sup>

Spektren in Hinblick auf Molekülkonformationen vorsichtig zu sein. Wegen des Problems der Mittelwertbildung liefern gemessene CD-Spektren keine verlässlichen Aussagen über bevorzugte Konformationen.

## 5.2. Unzureichende Zahl experimenteller Daten

Hat ein biomolekulares System mehr Freiheitsgrade als experimentelle Daten vorliegen, oder sind die unterschiedlichen experimentellen Daten korreliert, so kann dies dazu führen, dass die vorherrschende Konformation des Biomoleküls im Konformationsensemble nicht eindeutig durch die experimentellen Daten festgelegt ist. Diese Situation illustriert Abbildung 29 anhand eines Beispiels: Für ein  $\beta$ -Hexapeptid in Methanol werden die NOE-Abstände und  $^3J$ -Kopplungskonstanten aus MD-Simulationen<sup>[173]</sup> sowie aus einem Satz von 20 aus NMR-Strukturverfeinerungen abge-



**Abbildung 29.** Überschreitung von NOE-Abstandsschranken sowie  $^3J$ -Kopplungskonstanten bei einem  $\beta$ -Octapeptid in Methanol. Die Parameter wurden aus zwei 100-ns-MD-Trajektorien berechnet (ohne experimentelle Daten)<sup>[173]</sup> sowie aus einem Satz von 20 NMR-Modellstrukturen, die durch Strukturverfeinerung ausgehend von den experimentellen NMR-Daten erhalten wurden.<sup>[174]</sup>

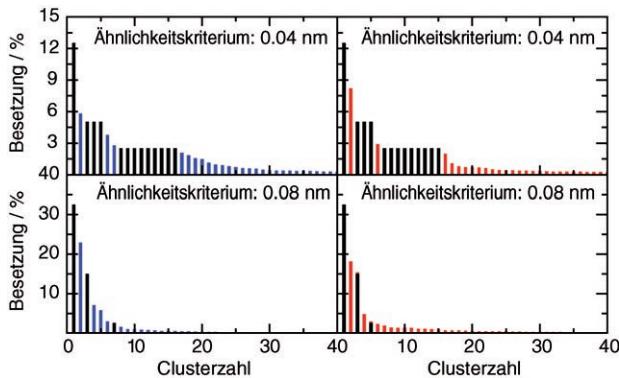
leiteten Modellstrukturen<sup>[174]</sup> mit den Messwerten verglichen. NOE-Abstände und  $^3J$ -Werte aus den 100-ns-MD-Simulationen stimmen mit den experimentellen Werten überein (bei nur zwei geringfügigen Abweichungen). Der Satz von 20 NMR-Modellstrukturen genügt zwangsläufig ebenfalls den experimentellen Daten, aus denen er abgeleitet wurde.<sup>[174]</sup> Zwischen den simulierten Konformationsensembles und dem Satz von NMR-Modellstrukturen gibt es jedoch kaum Überlappungen (Abbildung 30), obwohl beide Sätze die experi-

### 5.3. Genauigkeit experimenteller Daten

Die Genauigkeit von experimentellen Daten ist oft nicht ausreichend, um Simulationsergebnisse zu validieren. Ein Beispiel für diese Situation ist in Abbildung 15 anhand der freien Energien für die Bindung von Liganden an den Östrogenrezeptor dargestellt. Dort sind die Abweichungen zwischen den berechneten und gemessenen freien Bindungsenergien geringer als die Abweichungen zwischen experimentellen Werten eines Liganden.

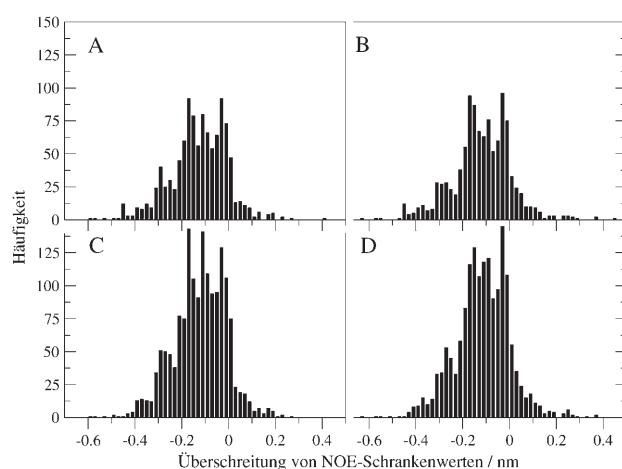
Ein anderer Parameter, dessen Genauigkeit nicht einfach zu bestimmen ist, sind NOE-Abstandsschranken, die aus NMR-Experimenten an Proteinen in wässriger Lösung erhalten werden. Hier taucht die Frage auf, in welchem Ausmaß solche NOE-Abstandsschranken in einer MD-Simulation überschritten werden

müssen, bevor man von einer signifikanten Abweichung zwischen Simulation und Experiment sprechen kann. Abbildung 31 zeigt, wie die Verletzungen von NOE-Schranken für 3.5-ns-MD-Trajektorien des Proteins Hühnereiweiß-Lysozym in wässriger Lösung verteilt sind.<sup>[175]</sup> Zwei MD-Simulationen mit unterschiedlichen Sätzen von Kraftfeldparametern (Diagramme A, C sowie B, D) werden mit zwei Sätzen von experimentell bestimmten NOE-Schranken verglichen (Diagramme A, B sowie C, D). Das Experiment aus dem Jahr 1993 lieferte 1079 Schranken,<sup>[176]</sup> das jüngere Experiment aus dem



**Abbildung 30.** Konformations-Clusteranalysen des Satzes von NMR-Modellstrukturen (schwarz) eines  $\beta$ -Octapeptids und einer MD-Trajektorie bei 298 K (blau) und 340 K (rot) mit zwei unterschiedlichen Ähnlichkeitskriterien für die mittlere quadratische Abweichung zwischen den Rückgratatomen der Peptidstrukturen: Es gibt keine Überlappung zwischen MD-Trajektorie und gebündelten NMR-Strukturen.<sup>[173]</sup>

mentellen Daten reproduzieren. Daher scheint die Zahl der Daten unzureichend zu sein, um das wichtigste Konformer eindeutig festzulegen: Die MD-Trajektorien deuten auf eine  $2.5_{12}$ -P-Helix hin, während der Satz von NMR-Modellstrukturen für eine  $2_8$ -P-Helix spricht.<sup>[173]</sup>



**Abbildung 31.** Verteilung der über  $r^{-3}$  gemittelten  $^1\text{H}$ - $^1\text{H}$ -NOE-Abstände in zwei 3.5-ns-MD-Simulationen des Hühnereiweiß-Lysozyms in wässriger Lösung,<sup>[175]</sup> bezogen auf zwei Sätze von NOE-Abstandsschranken aus NMR-Daten. Positive Werte entsprechen Überschreitungen der oberen NOE-Abstandsschranken. A, B) 1079 NOE-Abstände, berechnet aus den MD-Trajektorien, bezogen auf die in Lit. [176] angegebenen NOE-Schranken. C, D) 1630 NOE-Abstände, berechnet aus denselben MD-Trajektorien, bezogen auf die neueren, in Lit. [177] angegebenen NOE-Schranken. Linke Diagramme: MD-Simulationen mit dem GROMOS-43A1-Kraftfeld. Rechte Diagramme: MD-Simulationen mit dem GROMOS-45A3-Kraftfeld.

Jahr 2001 deutlich mehr (nämlich 1630).<sup>[177]</sup> Ein Vergleich der beiden experimentellen Datensätze mit den MD-Trajektorien belegt, dass die neueren experimentellen Daten etwas besser mit den Simulationen übereinstimmen (geringere mittlere Überschreitung, weniger starke Überschreitungen). Dies zeigt, dass experimentelle Daten sich mit der Zeit theoretischen Daten annähern können. Aus diesem Grund ist Vorsicht geboten, wenn aus beobachteten Abweichungen zwischen berechneten und experimentellen Daten auf die (unzureichende) Qualität von Simulationsergebnissen geschlossen wird.

#### 5.4. Perspektiven für den Vergleich simulierter und gemessener Daten

Simulationsstudien werden normalerweise anhand der gemessenen Eigenschaften des betrachteten Systems validiert.<sup>[178]</sup> Die Ergebnisse eines solchen Vergleichs zwischen Simulation und Experiment können in verschiedene Kategorien fallen.<sup>[137, 179, 180]</sup>

Fall 1: Übereinstimmung zwischen Simulation und Experiment. In diesem Fall kann eine der folgenden Situationen vorliegen:

- Die Simulation gibt das experimentelle System adäquat wieder.
- Die untersuchte Eigenschaft hängt nur wenig von den Details der Simulationstrajektorie ab. Eine Änderung der Simulationsparameter würde die Übereinstimmung nicht ändern.
- Fehlerkompensation: Diese Situation kann leicht eintreten, wenn bei einem System mit vielen Freiheitsgraden nur wenige (globale oder systemische) Eigenschaften berechnet und verglichen werden.

Fall 2: Keine Übereinstimmung zwischen Simulation und Experiment als Folge von einem oder beiden der folgenden Gründe:

- Die Simulation gibt das experimentelle System nicht richtig wieder. Entweder sind Theorie oder Modell falsch, die simulierte Eigenschaft konvergiert nicht, die Software ist fehlerhaft oder sie wird falsch bedient.
- Die experimentellen Daten sind falsch und/oder werden falsch interpretiert.

Bei einem Vergleich von simulierten mit experimentellen Ergebnissen sollten stets dieselben Eigenschaften betrachtet werden. Dies ist nicht immer möglich, und manchmal werden nur verwandte Eigenschaften verglichen. Zum Beispiel sind sowohl die mittleren quadratischen Fluktuationen der Atompositionen als auch die kristallographischen *B*-Faktoren ein Maß für die Atombeweglichkeit und die Atomfehlordnung, sie sind aber unterschiedlich definiert: Die erste Größe misst die räumliche Verteilung eines bestimmten Atoms, die zweite misst, wie häufig eine bestimmte Position im Raum von einem (gleich welchem) Atom eingenommen wird.<sup>[181]</sup> Als weiteres Beispiel führen wir die Geschwindigkeiten von Proteinfaltungen an, die durch Renaturierung bei simulierten Temperaturen ermittelt werden und von den gemessenen

Geschwindigkeiten der Cosolvans-induzierten Faltung abweichen.

Mit der Zeit werden Simulationen mit verbesserten Kraftfeldern und längeren Äquilibrierungs- und Simulationszeiten immer genauere Werte für die verschiedenen molekularen und systemischen Eigenschaften liefern. Es ist zu hoffen, dass die Verbesserungen der experimentellen Genauigkeit durch ausfeilte Messtechniken mit den Entwicklungen in der Modellierung Schritt halten können.

Bezüglich der experimentellen Daten, die in der Kraftfeldentwicklung eingesetzt werden, wäre es für die Verfeinerung von biomolekularen Modellierungsmethoden von vorrangiger Bedeutung, präzise thermodynamische Daten wie Verdampfungswärmen, Mischungswärmen etc. für vielfältige Verbindungen unter physiologischen Bedingungen zu gewinnen.

#### 6. Perspektiven in der biomolekularen Modellierung

Der entscheidende Parameter bei der Entwicklung der biomolekularen Modellierung war, ist und wird auch künftig der rasante Anstieg der verfügbaren Rechenleistung sein. Abbildung 32 verdeutlicht, dass die Computerleistung bisher

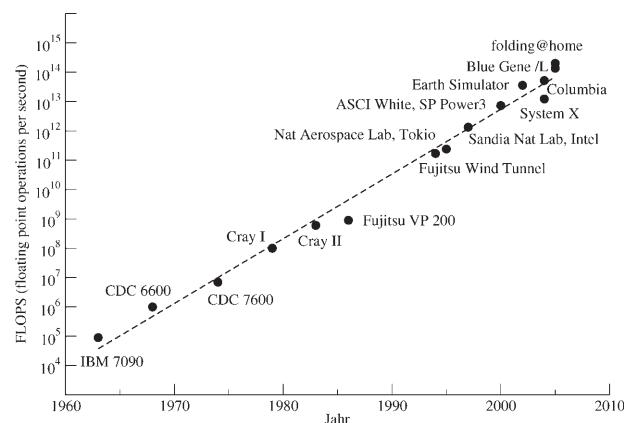


Abbildung 32. Entwicklung der Rechenleistung der weltweit leistungsfähigsten Computer.

alle fünf Jahre um den Faktor 10 gestiegen ist. Nimmt man die gegenwärtigen Parallelrechnern zur Grundlage, dann wird sich dieser Trend in nächster Zukunft wahrscheinlich fortsetzen. Parallelrechentechniken kommen bei biomolekularen Simulationen hervorragend zur Geltung, da der zeitraubendste Teil einer Simulation die Berechnung von Kräften oder Wechselwirkungen ist, die für alle Atome im System parallel ausgeführt werden kann. Insbesondere die Einführung neuer Hardware, die speziell zur Lösung des Proteinfaltungsproblems durch klassische dynamische Simulationen entwickelt wurde, eröffnet Perspektiven für präzisere Simulationen und neue Anwendungen.<sup>[182]</sup>

Eine zweite Triebkraft der biomolekularen Modellierung sind die Fortschritte bei den Modellierungstechniken. Beispielsweise stehen heute zur Berechnung von langreichweiten elektrostatischen Kräften effiziente Algorithmen zur

Verfügung.<sup>[13,21,27]</sup> Zahlreiche Methoden haben die Effizienz bei der Durchmusterung des Konfigurationsraumes verbessert,<sup>[61,63,85,155,183]</sup> und die biomolekularen Kraftfelder sind verfeinert worden.<sup>[7,184]</sup>

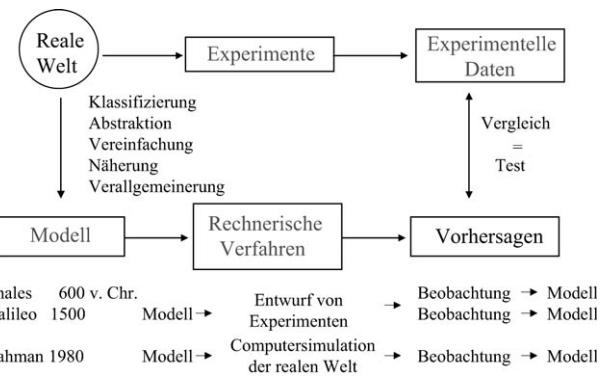
Dank dieser Entwicklungen gelingen Simulationen für immer größere Systeme und über immer längere Zeiträume (siehe Tabelle 9). Praktische Anwendungen von Simulationen

**Tabelle 9:** Geschichte und extrapolierte Zukunft der Moleküldynamik-Simulation. Zukünftige Entwicklungen sind ausgehend von der bisherigen Erfahrung abgeleitet, dass die Computerleistung alle 5 Jahre um den Faktor 10 steigt (siehe Abbildung 32).

Jahr	molekulares System (Art, Größe)	Simulationsdauer [s]
1957	erste Moleküldynamik-Simulation (harte Scheiben)	
1964	atomare Flüssigkeit (Argon)	$10^{-11}$
1971	molekulare Flüssigkeit (Wasser)	$5 \times 10^{-12}$
1977	Protein im Vakuum	$2 \times 10^{-11}$
1983	Protein in Wasser	$2 \times 10^{-11}$
1989	Protein-DNA-Komplex in Wasser	$10^{-10}$
1997	Polypeptidfaltung in Lösung	$10^{-7}$
2001	Micellbildung	$10^{-7}$
200x	Faltung eines kleinen Proteins	$10^{-3}$
Und die Zukunft ...		
2001	Biomoleküle in Wasser (ca. $10^4$ Atome)	$10^{-8}$
2029	Biomoleküle in Wasser (Faltung eher?)	$10^{-3}$
2034	E.-coli-Bakterien (ca. $10^{11}$ Atome)	$10^{-9}$
2056	Säugerzelle (ca. $10^{15}$ Atome)	$10^{-9}$
2080	Biomoleküle in Wasser (so schnell wie in der Natur)	$10^6$
2172	menschlicher Körper (ca. $10^{27}$ Atome)	1

betreffen verschiedenste Systeme und Prozesse: Komplexbildung, Ligandenkoordination, Polypeptidfaltung, Membrantransport, Membranbildung, Kristallisation. Ginge man weiterhin davon aus, dass sich die Effizienz alle 5 Jahre um den Faktor 10 verbessert, so ergeben sich die in Tabelle 9 genannten Prognosen, die aber im Grunde müßig sind. Erstens ist es unwahrscheinlich, dass die Rechenleistung weiterhin mit dem bisherigen Tempo steigt. Zweitens ist zu bedenken, dass zur Simulation immer größerer Systeme auf atomarer Ebene immer mehr Paarwechselwirkungen herangezogen werden müssen. Um ein großes System mit der gleichen Genauigkeit zu berechnen wie ein kleines, müssen die Paarwechselwirkungen des größeren Systems genauer bestimmt sein. Die Genauigkeit wird jedoch durch die Näherungen begrenzt, auf denen eine Kraftfeldbeschreibung des Systems beruht. Drittens kann man den Wert einer detaillierten atomaren Beschreibung makroskopischer Systeme grundsätzlich anzweifeln. Vorrangigste Aufgabe bleibt es daher, einfache und genäherte Modelle zu formulieren (Abbildung 33), die eine interessierende Eigenschaft mithilfe möglichst weniger Freiheitsgrade adäquat wiedergeben.

Es stellt sich die Frage, nach welchen Kriterien biomolekulare Modelle erweitert, verbessert oder vereinfacht werden sollten. Erstens erfordert eine angemessene Beschreibung



**Abbildung 33:** Grundlage der Computerphysik, -chemie und -biologie ist die Formulierung und Überprüfung von (mathematischen) Modellen der realen Welt.

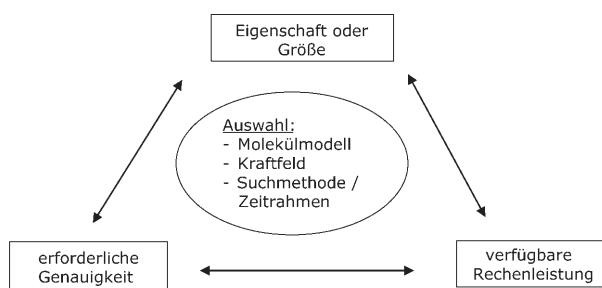
von Enzymreaktionen, dass elektronische Freiheitsgrade berücksichtigt werden, die gemäß Tabelle 1 eine Stufe über dem Theorieniveau klassischer MD-Methoden liegen. Quantenklassische Hybridverfahren (QM/MM) werden in dieser Hinsicht die weitere Entwicklung bestimmen.<sup>[49,185–187]</sup> Zur Simulation von Protonentransferreaktionen könnten quadtodynamische Methoden nötig sein,<sup>[188–191]</sup> die noch höhere Rechenleistungen als QM/MM-Rechnungen erfordern.<sup>[192]</sup> Zweitens werden Verbesserungen im Bereich der klassischen Modellierungsmethoden aus der Berücksichtigung der Polarisierbarkeit in biomolekularen Kraftfeldern folgen<sup>[44–48]</sup> sowie aus der expliziten Simulation von Cosolvenseffekten<sup>[193]</sup> und aus Techniken mit verbesserter Durchmusterung des Konfigurationsraumes.<sup>[61–63]</sup> Drittens wird die Vereinfachung von Modellen durch Mittelwertbildung über atomare Freiheitsgrade (grobkörnige Molekülmodelle)<sup>[54–59]</sup> die Simulation langsamer Prozesse wie der Membranbildung ermöglichen.<sup>[60,194]</sup>

Die Motivation zur Verwendung von Simulationen und Modellstudien wurde in Tabelle 3 dargelegt: Sie sollen ein mikroskopisch genaues Bild mit sonst unerreichbarer Auflösung in Zeit, Raum und Energie erzeugen, das den begrenzten Satz von experimentell zugänglichen Eigenschaften ergänzt. Darüber hinaus können bei einer Modellstudie die Systemparameter nach Belieben verändert werden, um einzelne Ursache-Wirkungs-Beziehungen zu untersuchen und dadurch biomolekulare Systeme besser zu verstehen.

Zur Modellierung eines biomolekularen Systems ist es entscheidend, den richtigen Mittelweg bei der Auswahl der Freiheitsgrade, des Kraftfeldes, der Simulationsmethode und der Randbedingungen zu finden (siehe Abbildung 2). Drei Faktoren sind hier maßgeblich (Abbildung 34):

1. Die interessierenden Eigenschaften des betreffenden biomolekularen Systems sollten benannt werden, und die Größe des zu durchsuchenden Konfigurationsraumes (oder die Zeitskala) sollte abgeschätzt werden.
2. Die erforderliche Genauigkeit sollte festgelegt werden.
3. Die verfügbare Rechenleistung sollte abgeschätzt werden.

Falls das Modell zu einfach gewählt wird, kann es sein, dass das interessierende Phänomen verborgen bleibt oder die Genauigkeit unzureichend ist. Bei zu aufwändigen Modellen

**Ziele:**

längere Zeitskalen      } Nutze alle verfügbaren Techniken zur Zeitersparnis, höhere Suchleistung      } die nicht die erforderliche Genauigkeit beeinträchtigen!

**Abbildung 34.** Die Auswahl des Molekülmodells (Freiheitsgrade), des Kraftfeldes und des Simulationsverfahrens hängen von der betreffenden Eigenschaft ab, die berechnet werden soll, der benötigten Genauigkeit des Ergebnisses und der verfügbaren Rechenleistung zur Erzeugung des Boltzmann-Ensembles.

kann es andererseits unmöglich werden, den Konfigurationsraum im geforderten Umfang zu durchsuchen. Hier den richtigen Mittelweg zu finden – darin besteht die Kunst bei der biomolekularen Modellierung.

*Wir danken dem Nationalen Forschungsschwerpunkt (NFS) „Strukturbioologie“ der Schweizerischen Forschungsgesellschaft für finanzielle Unterstützung. Daniela Kalbermatter danken wir für das Editieren des Bildmaterials und des Manuskripts.*

Ein eingegangen am 28. Juli 2005  
Übersetzt von Dr. Frank Maaß, Weinheim

- [1] M. P. Allen, D. J. Tildesley, *Computer Simulation of Liquids*, Oxford University Press, New York, **1987**.
- [2] B. A. Luty, W. F. van Gunsteren, *J. Phys. Chem.* **1996**, *100*, 2581–2587.
- [3] P. H. Hünenberger, J. A. McCammon, *J. Chem. Phys.* **1999**, *110*, 1856–1872.
- [4] P. H. Hünenberger, J. A. McCammon, *Biophys. Chem.* **1999**, *78*, 69–88.
- [5] W. Weber, P. H. Hünenberger, J. A. McCammon, *J. Phys. Chem. B* **2000**, *104*, 3668–3675.
- [6] P. H. Hünenberger, W. F. van Gunsteren in *Computer Simulation of Biomolecular Systems. Theoretical and Experimental Applications*, Vol. 3 (Hrsg.: W. F. van Gunsteren, P. K. Weiner, A. J. Wilkinson), Kluwer, Dordrecht, **1997**, S. 3–82.
- [7] C. Oostenbrink, A. Villa, A. E. Mark, W. F. van Gunsteren, *J. Comput. Chem.* **2004**, *25*, 1656–1676.
- [8] N. F. A. van der Vegt, W. F. van Gunsteren, *J. Phys. Chem. B* **2004**, *108*, 1056–1064.
- [9] N. F. A. van der Vegt, D. Trzesniak, B. Kasumaj, W. F. van Gunsteren, *ChemPhysChem* **2004**, *5*, 144–147.
- [10] W. F. van Gunsteren, S. R. Billeter, A. A. Eising, P. H. Hünenberger, P. Krüger, A. E. Mark, W. R. P. Scott, I. G. Tironi, *Biomolecular Simulation: The GROMOS96 Manual and User Guide*, Vdf Hochschulverlag, Zürich, **1996**.
- [11] J. D. Jackson, *Classical Electrodynamics*, Wiley, New York, **1962**.
- [12] P. Ewald, *Ann. Phys.* **1921**, *64*, 253–287.
- [13] R. W. Hockney, J. W. Eastwood, *Computer Simulation Using Particles*, 2. Aufl., Institute of Physics Publishing, Bristol, **1988**.
- [14] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, L. G. Pedersen, *J. Chem. Phys.* **1995**, *103*, 8577–8593.
- [15] J. A. Barker, R. O. Watts, *Mol. Phys.* **1973**, *26*, 789–792.
- [16] I. G. Tironi, R. Sperb, P. E. Smith, W. F. van Gunsteren, *J. Chem. Phys.* **1995**, *102*, 5451–5459.
- [17] S. Boresch, O. Steinhauser, *J. Chem. Phys.* **2001**, *115*, 10793–10807.
- [18] M. Bergdorf, C. Peter, P. H. Hünenberger, *J. Chem. Phys.* **2003**, *119*, 9129–9144.
- [19] C. Peter, W. F. van Gunsteren, P. H. Hünenberger, *J. Chem. Phys.* **2003**, *119*, 12205–12223.
- [20] M. A. Kastenholz, P. H. Hünenberger, *J. Phys. Chem. B* **2004**, *108*, 774–778.
- [21] P. H. Hünenberger in *Simulation and Theory of Electrostatic Interactions in Solution: Computational Chemistry, Biophysics and Aqueous Solution* (Hrsg.: L. R. Pratt, G. Hummer), AIP, New York, **1999**, S. 17–83.
- [22] B. A. Luty, I. G. Tironi, W. F. van Gunsteren, *J. Chem. Phys.* **1995**, *103*, 3014–3021.
- [23] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, J. Hermans in *Intermolecular Forces* (Hrsg.: B. Pullman), Reidel, Dordrecht, **1981**, S. 331–342.
- [24] T. P. Straatsma, H. J. C. Berendsen, *J. Chem. Phys.* **1988**, *89*, 5876–5886.
- [25] P. H. Hünenberger, *J. Chem. Phys.* **2000**, *113*, 10464–10476.
- [26] G. Hummer, L. R. Pratt, A. E. Garcia, *J. Phys. Chem.* **1996**, *100*, 1206–1215.
- [27] M. A. Kastenholz, P. H. Hünenberger, *J. Chem. Phys.*, im Druck.
- [28] W. L. Jorgensen, J. D. Madura, C. J. Swenson, *J. Am. Chem. Soc.* **1984**, *106*, 6638–6646.
- [29] W. L. Jorgensen, D. S. Maxwell, J. Tirado-Rives, *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236.
- [30] L. D. Schuler, W. F. van Gunsteren, *Mol. Simul.* **2000**, *25*, 301–319.
- [31] A. Glättli, W. F. van Gunsteren, *Angew. Chem.* **2004**, *116*, 6472–6476; *Angew. Chem. Int. Ed.* **2004**, *43*, 6312–6316.
- [32] X. Daura, B. Jaun, D. Seebach, W. F. van Gunsteren, A. E. Mark, *J. Mol. Biol.* **1998**, *280*, 925–932.
- [33] X. Daura, K. Gademann, B. Jaun, D. Seebach, W. F. van Gunsteren, A. E. Mark, *Angew. Chem.* **1999**, *111*, 249–253; *Angew. Chem. Int. Ed.* **1999**, *38*, 236–240.
- [34] X. Daura, K. Gademann, H. Schäfer, B. Jaun, D. Seebach, W. F. van Gunsteren, *J. Am. Chem. Soc.* **2001**, *123*, 2393–2404.
- [35] C. M. Santiveri, M. A. Jiménez, M. Rico, W. F. van Gunsteren, X. Daura, *J. Pept. Sci.* **2004**, *10*, 546–565.
- [36] C. Oostenbrink, T. A. Soares, N. F. A. van der Vegt, W. F. van Gunsteren, *Eur. Biophys. J.* **2005**, *34*, 273–284.
- [37] I. Huc, V. Maurizot, H. Gornitzka, J.-M. Leger, *Chem. Commun.* **2002**, 578–579.
- [38] A. Villa, A. E. Mark, *J. Comput. Chem.* **2002**, *23*, 548–553.
- [39] M. R. Shirts, J. W. Pitera, W. C. Swope, V. S. Pande, *J. Chem. Phys.* **2003**, *119*, 5740–5761.
- [40] Y. Marcus, *Ionic Solvation*, Wiley, Chichester, **1985**.
- [41] L. J. Smith, H. J. C. Berendsen, W. F. van Gunsteren, *J. Phys. Chem. A* **2004**, *108*, 1065–1071.
- [42] D. P. Geerke, C. Oostenbrink, N. F. A. van der Vegt, W. F. van Gunsteren, *J. Phys. Chem. B* **2004**, *108*, 1436–1445.
- [43] D. Trzesniak, N. F. A. van der Vegt, W. F. van Gunsteren, *Phys. Chem. Chem. Phys.* **2004**, *6*, 697–702; Erratum: D. Trzesniak, N. F. A. van der Vegt, W. F. van Gunsteren, *Phys. Chem. Chem. Phys.* **2004**, *6* (3. August 2004).
- [44] T. A. Halgren, W. Damm, *Curr. Opin. Struct. Biol.* **2001**, *11*, 236–242.
- [45] S. W. Rick, S. J. Stuart, *Rev. Comput. Chem.* **2002**, 89–146.

- [46] B. Guillot, *J. Mol. Liq.* **2002**, *101*, 219–260.
- [47] J. W. Ponder, D. A. Case, *Adv. Protein Chem.* **2003**, *56*, 27–85.
- [48] H. B. Yu, W. F. van Gunsteren, *Comput. Phys. Commun.* **2005**, *172*, 69–85.
- [49] A. Warshel, M. Levitt, *J. Mol. Biol.* **1976**, *103*, 227–249.
- [50] D. Van Belle, I. Couplet, M. Prevost, S. J. Wodak, *J. Mol. Biol.* **1987**, *198*, 721–735.
- [51] J. L. Banks, G. A. Kaminski, R. H. Zhou, D. T. Mainz, B. J. Berne, R. A. Friesner, *J. Chem. Phys.* **1999**, *110*, 741–754.
- [52] S. Patel, C. L. Brooks, *J. Comput. Chem.* **2004**, *25*, 1–15.
- [53] S. Patel, A. D. MacKerell, C. L. Brooks, *J. Comput. Chem.* **2004**, *25*, 1504–1514.
- [54] A. Liwo, M. R. Pincus, R. J. Wawak, S. Rackofsky, H. A. Scheraga, *Protein Sci.* **1993**, *2*, 1715–1731.
- [55] J. C. Shelley, M. Y. Shelley, *Curr. Opin. Colloid Interface Sci.* **2000**, *5*, 101–110.
- [56] J. Barschnagel, K. Binder, P. Doruker, A. A. Gusev, O. Hahn, K. Kremer, W. L. Mattice, F. Müller-Plathe, M. Murat, W. Paul, S. Santos, U. W. Suter, V. Tries, *Adv. Polym. Sci.* **2000**, *152*, 41–156.
- [57] M. Müller, K. Katsov, M. Schick, *J. Polym. Sci. Part B* **2003**, *41*, 1441–1450.
- [58] A. Liwo, M. Khalili, H. A. Scheraga, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 2362–2367.
- [59] S. J. Marrink, A. H. de Vries, A. E. Mark, *J. Phys. Chem. B* **2004**, *108*, 750–760.
- [60] S. J. Marrink, H. J. Risselada, A. E. Mark, *Chem. Phys. Lip.* **2005**, *135*, 223–244.
- [61] W. F. van Gunsteren, T. Huber, A. E. Torda in *Proc. Eur. Conf. Comput. Chem. (ECCC 1)*, American Institute of Physics, Conference Proceedings **1995**, *330*, S. 253–268.
- [62] Y. Okamoto, *J. Mol. Graphics Modell.* **2004**, *22*, 425–439.
- [63] K. Tai, *Biophys. Chem.* **2004**, *107*, 213–220.
- [64] P. Koehl, M. Delaure, *J. Mol. Biol.* **1994**, *239*, 249–275.
- [65] T. Huber, A. E. Torda, W. F. van Gunsteren, *Biopolymers* **1996**, *39*, 103–114.
- [66] J. Desmet, M. DeMaeyer, B. Hazes, I. Lasters, *Nature* **1992**, *356*, 539–542.
- [67] M. Lipton, W. C. Still, *J. Comput. Chem.* **1988**, *9*, 343–355.
- [68] „Protein Structure and Engineering“: D. D. Bensen, G. R. Marshall, *NATO ASI Ser. Ser. A* **1989**, *183*, 97–109.
- [69] M. Saunders, K. N. Houk, Y.-D. Wu, W. C. Still, M. Lipton, G. Chang, W. C. Guida, *J. Am. Chem. Soc.* **1990**, *112*, 1419–1427.
- [70] D. G. Covell, R. L. Jernigan, *Biochemistry* **1990**, *29*, 3287–3294.
- [71] R. C. van Schaik, W. F. van Gunsteren, H. J. C. Berendsen, *J. Comput.-Aided Mol. Des.* **1992**, *6*, 97–112.
- [72] G. M. Crippen, T. F. Havel, *Distance Geometry and Molecular Conformation*, Wiley, New York, **1988**.
- [73] T. F. Havel, *Biopolymers* **1990**, *29*, 1565–1585.
- [74] K. D. Gibson, H. A. Scheraga, *J. Comput. Chem.* **1987**, *8*, 826–834.
- [75] H. A. Scheraga in *Computer Simulation of Biomolecular Systems, Theoretical and Experimental Applications*, Vol. 2 (Hrsg.: W. F. van Gunsteren, P. K. Weiner, A. J. Wilkinson), Escom Science, Leiden, **1993**, S. 231–248.
- [76] S. Vajda, C. DeLisi, *Biopolymers* **1990**, *29*, 1755–1772.
- [77] J. Harris, S. A. Rice, *J. Chem. Phys.* **1988**, *88*, 1298–1306.
- [78] B. Velikson, T. Garel, J.-C. Niel, H. Orland, J. C. Smith, *J. Comput. Chem.* **1992**, *13*, 1216–1233.
- [79] D. Frenkel, G. C. A. M. Mooij, B. Smit, *J. Phys. Condens. Matter* **1992**, *4*, 3053–3076.
- [80] W. F. van Gunsteren in *Computer Simulation of Biomolecular Systems, Theoretical and Experimental Applications*, Vol. 2 (Hrsg.: W. F. van Gunsteren, P. K. Weiner, A. J. Wilkinson), Escom Science, Leiden, **1993**, S. 3–36.
- [81] R. M. J. Cotterill, J. K. Madsen in *Characterising Complex Systems* (Hrsg.: H. Bohr), World Scientific, Singapur, **1990**, S. 177–191.
- [82] W. Braun, N. Go, *J. Mol. Biol.* **1985**, *186*, 611–626.
- [83] T. Huber, A. E. Torda, W. F. van Gunsteren, *J. Phys. Chem. A* **1997**, *101*, 5926–5930.
- [84] A. Piela, J. Kostrowicki, H. A. Scheraga, *J. Phys. Chem.* **1989**, *93*, 3339–3346.
- [85] T. Huber, A. E. Torda, W. F. van Gunsteren, *J. Comput.-Aided Mol. Des.* **1994**, *8*, 695–708.
- [86] H. Grubmüller, *Phys. Rev. E* **1995**, *52*, 2893–2906.
- [87] A. E. Torda, R. M. Scheek, W. F. van Gunsteren, *J. Mol. Biol.* **1990**, *214*, 223–235.
- [88] W. F. van Gunsteren, R. M. Brunne, P. Gros, R. C. van Schaik, C. A. Schiffer, A. E. Torda in *Methods in Enzymology: Nuclear Magnetic Resonance*, Vol. 239 (Hrsg.: T. L. James, N. J. Oppenheimer), Academic Press, New York, **1994**, S. 619–654.
- [89] R. C. van Schaik, H. J. C. Berendsen, A. E. Torda, W. F. van Gunsteren, *J. Mol. Biol.* **1993**, *234*, 751–762.
- [90] W. F. van Gunsteren, H. J. C. Berendsen, *Mol. Phys.* **1977**, *34*, 1311–1327.
- [91] S. Kirkpatrick, C. D. Gelatt, M. P. Vecchi, *Science* **1983**, *220*, 671–680.
- [92] B. Mao, A. R. Friedmann, *Biophys. J.* **1990**, *58*, 803–805.
- [93] R. Elber, M. Karplus, *J. Am. Chem. Soc.* **1990**, *112*, 9161–9175.
- [94] R. Unger, J. Moult, *J. Mol. Biol.* **1993**, *231*, 75–81.
- [95] T. Huber, W. F. van Gunsteren, *J. Phys. Chem. A* **1998**, *102*, 5937–5943.
- [96] M. Levitt, *J. Mol. Biol.* **1983**, *170*, 723–764.
- [97] T. C. Beutler, A. E. Mark, R. C. van Schaik, P. R. Gerber, W. F. van Gunsteren, *Chem. Phys. Lett.* **1994**, *222*, 529–539.
- [98] M. Zacharias, T. P. Straatsma, J. A. McCammon, *J. Chem. Phys.* **1994**, *100*, 9025–9031.
- [99] V. Hornak, C. Simmerling, *J. Mol. Graphics Modell.* **2004**, *22*, 405–413.
- [100] G. M. Crippen, H. A. Scheraga, *Proc. Natl. Acad. Sci. USA* **1969**, *64*, 42–49.
- [101] A. Laio, M. Parrinello, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 12562–12566.
- [102] A. E. Torda, R. M. Scheek, W. F. van Gunsteren, *Chem. Phys. Lett.* **1989**, *157*, 289–294.
- [103] G. M. Crippen, *J. Comput. Chem.* **1982**, *3*, 471–476.
- [104] E. O. Purisima, H. A. Scheraga, *Proc. Natl. Acad. Sci. USA* **1986**, *83*, 2782–2786.
- [105] G. M. Crippen, *J. Phys. Chem.* **1987**, *91*, 6341–6343.
- [106] G. M. Crippen, T. F. Havel, *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 222–227.
- [107] P. L. Weber, R. Morrison, D. L. Hare, *J. Mol. Biol.* **1988**, *204*, 483–487.
- [108] T. C. Beutler, W. F. van Gunsteren, *J. Chem. Phys.* **1994**, *101*, 1417–1422.
- [109] J.-P. Ryckaert, G. Ciccotti, H. J. C. Berendsen, *J. Comput. Phys.* **1977**, *23*, 327–341.
- [110] H. C. Andersen, *J. Comput. Phys.* **1983**, *52*, 24–34.
- [111] S. Miyamoto, P. A. Kollman, *J. Comput. Chem.* **1992**, *13*, 952–962.
- [112] B. Hess, H. Bekker, H. J. C. Berendsen, J. G. E. M. Fraaije, *J. Comput. Chem.* **1997**, *18*, 1463–1472.
- [113] V. Kräutler, W. F. van Gunsteren, P. H. Hünenberger, *J. Comput. Chem.* **2001**, *22*, 501–508.
- [114] W. F. van Gunsteren, M. Karplus, *Macromolecules* **1982**, *15*, 1528–1544.
- [115] M. Christen, W. F. van Gunsteren, *J. Chem. Phys.* **2005**, *122*, 144106.
- [116] J. E. Straub, M. Karplus, *J. Chem. Phys.* **1991**, *94*, 6737–6739.
- [117] A. Roitberg, R. Elber, *J. Chem. Phys.* **1991**, *95*, 9277–9287.

- [118] G. Verkhivker, R. Elber, Q. H. Gibson, *J. Am. Chem. Soc.* **1992**, *114*, 7866–7878.
- [119] G. Verkhivker, R. Elber, W. Nowak, *J. Chem. Phys.* **1992**, *97*, 7838–7841.
- [120] A. Ulitsky, R. Elber, *J. Chem. Phys.* **1993**, *98*, 3380–3388.
- [121] P. Amara, D. Hsu, J. E. Straub, *J. Phys. Chem.* **1993**, *97*, 6715–6721.
- [122] J. P. Ma, D. Hsu, J. E. Straub, *J. Chem. Phys.* **1993**, *99*, 4024–4035.
- [123] Q. Zheng, R. Rosenfeld, S. Vajda, C. DeLisi, *Protein Sci.* **1993**, *2*, 1242–1248.
- [124] Q. Zheng, R. Rosenfeld, D. J. Kyle, *J. Chem. Phys.* **1993**, *99*, 8892–8896.
- [125] K. A. Olszewski, L. Piela, H. A. Scheraga, *J. Phys. Chem.* **1992**, *96*, 4672–4676.
- [126] K. A. Olszewski, L. Piela, H. A. Scheraga, *J. Phys. Chem.* **1993**, *97*, 260–266.
- [127] K. A. Olszewski, L. Piela, H. A. Scheraga, *J. Phys. Chem.* **1993**, *97*, 267–270.
- [128] C. Simmerling, J. L. Miller, P. A. Kollman, *J. Am. Chem. Soc.* **1998**, *120*, 7149–7155.
- [129] H. Y. Liu, Z. H. Duan, Q. M. Luo, Y. Y. Shi, *Proteins Struct. Funct. Genet.* **1999**, *36*, 462–470.
- [130] J. Zhu, H. B. Yu, H. Fan, H. Y. Liu, Y. Y. Shi, *J. Comput.-Aided Mol. Des.* **2001**, *15*, 447–463.
- [131] J. Zhu, H. Fan, H. Y. Liu, Y. Y. Shi, *J. Comput.-Aided Mol. Des.* **2001**, *15*, 979–996.
- [132] D. E. Goldberg, *Genetic Algorithms in Search, Optimisation and Machine Learning*, Addison-Wesley, Reading, **1989**.
- [133] Y. Sugita, Y. Okamoto, *Chem. Phys. Lett.* **1999**, *314*, 141–151.
- [134] R. Zhou, B. J. Berne, R. Germain, *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 14931–14936.
- [135] W. F. van Gunsteren, P. H. Hünenberger, A. E. Mark, P. E. Smith, I. G. Tironi, *Comput. Phys. Commun.* **1995**, *91*, 305–319.
- [136] W. F. van Gunsteren, A. E. Mark, *J. Chem. Phys.* **1998**, *108*, 6109–6116.
- [137] „Dynamics, Structure and Function of Biological Macromolecules“: W. F. van Gunsteren, D. Bakowies, W. Damm, T. Hansson, U. Stocker, X. Daura, *NATO ASI Ser. Ser. A* **2001**, *315*, 1–26.
- [138] X. Daura, W. F. van Gunsteren, A. E. Mark, *Proteins Struct. Funct. Genet.* **1999**, *34*, 269–280.
- [139] A. Glärtli, X. Daura, P. Bindschädler, B. Jaun, Y. R. Mahajan, R. I. Mathad, M. Rueping, D. Seebach, W. F. van Gunsteren, *Chem. Eur. J.* **2005**, *11*, 7276–7293.
- [140] J. W. Pitera, M. Falta, W. F. van Gunsteren, *Biophys. J.* **2001**, *80*, 2546–2555.
- [141] T. N. Heinz, W. F. van Gunsteren, P. H. Hünenberger, *J. Chem. Phys.* **2001**, *115*, 1125–1136.
- [142] C. Peter, C. Oostenbrink, A. van Dorp, W. F. van Gunsteren, *J. Chem. Phys.* **2004**, *120*, 2652–2661.
- [143] L. J. Smith, X. Daura, W. F. van Gunsteren, *Proteins Struct. Funct. Genet.* **2002**, *48*, 487–496.
- [144] W. F. van Gunsteren, R. Bürgi, C. Peter, X. Daura, *Angew. Chem.* **2001**, *113*, 363–367; *Angew. Chem. Int. Ed.* **2001**, *40*, 351–355.
- [145] W. F. van Gunsteren, D. Bakowies, R. Bürgi, I. Chandrasekhar, M. Christen, X. Daura, P. Gee, A. Glärtli, T. Hansson, C. Oostenbrink, C. Peter, J. Pitera, L. Schuler, T. Soares, H. B. Yu, *Chimia* **2001**, *55*, 856–860.
- [146] X. Daura, A. Glärtli, P. Gee, C. Peter, W. F. van Gunsteren, *Adv. Protein Chem.* **2002**, *62*, 341–360.
- [147] P. J. Gee, F. A. Hamprecht, L. D. Schuler, W. F. van Gunsteren, E. Duchardt, H. Schwalbe, M. Albert, D. Seebach, *Helv. Chim. Acta* **2002**, *85*, 618–632.
- [148] H. Schäfer, X. Daura, A. E. Mark, W. F. van Gunsteren, *Proteins Struct. Funct. Genet.* **2001**, *43*, 45–56.
- [149] C. Oostenbrink, W. F. van Gunsteren, *Proteins Struct. Funct. Genet.* **2004**, *54*, 237–246.
- [150] C. Oostenbrink, W. F. van Gunsteren, *Phys. Chem. Chem. Phys.* **2005**, *7*, 53–58.
- [151] L. J. Smith, R. M. Jones, W. F. van Gunsteren, *Proteins Struct. Funct. Genet.* **2005**, *58*, 439–449.
- [152] H. Liu, A. E. Mark, W. F. van Gunsteren, *J. Phys. Chem.* **1996**, *100*, 9485–9494.
- [153] J. W. Pitera, W. F. van Gunsteren, *J. Phys. Chem. B* **2001**, *105*, 11264–11274.
- [154] C. Oostenbrink, W. F. van Gunsteren, *J. Comput. Chem.* **2003**, *24*, 1730–1739.
- [155] C. Oostenbrink, W. F. van Gunsteren, *Chem. Eur. J.* **2005**, *11*, 4340–4348.
- [156] C. Oostenbrink, W. F. van Gunsteren, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 6750–6754.
- [157] H.-A. Yu, M. Karplus, *J. Chem. Phys.* **1988**, *89*, 2366–2379.
- [158] B. Guillot, Y. Guissani, *J. Chem. Phys.* **1993**, *99*, 8075–8094.
- [159] E. Gallicchio, M. M. Kubo, R. M. Levy, *J. Phys. Chem. B* **2000**, *104*, 6271–6285.
- [160] M. Rueping, J. V. Schreiber, G. Lelais, B. Jaun, D. Seebach, *Helv. Chim. Acta* **2002**, *85*, 2577–2593.
- [161] D. Trzesniak, A. Glärtli, B. Jaun, W. F. van Gunsteren, *J. Am. Chem. Soc.* **2005**, *127*, 14320–14329.
- [162] M. Karplus, J. N. Kushick, *Macromolecules* **1981**, *14*, 325–332.
- [163] J. Schlitter, *Chem. Phys. Lett.* **1993**, *215*, 617–621.
- [164] H. Schäfer, A. E. Mark, W. F. van Gunsteren, *J. Chem. Phys.* **2000**, *113*, 7809–7817.
- [165] J. Carlsson, J. Aqvist, *J. Phys. Chem. B* **2005**, *109*, 6448–6456.
- [166] H. Schäfer, L. J. Smith, A. E. Mark, W. F. van Gunsteren, *Proteins Struct. Funct. Genet.* **2002**, *46*, 215–224.
- [167] M. Bellanda, E. Peggion, R. Bürgi, W. F. van Gunsteren, S. Mammi, *J. Pept. Res.* **2001**, *57*, 97–106.
- [168] A. Bavoso, E. Benedetti, B. DiBlasio, V. Pavone, C. Pedone, C. Toniolo, G. M. Bonora, F. Formaggio, M. Crisma, *J. Biomol. Struct. Dyn.* **1988**, *6*, 803–817.
- [169] R. Bürgi, X. Daura, A. Mark, M. Bellanda, S. Mammi, E. Peggion, W. F. van Gunsteren, *J. Pept. Res.* **2001**, *57*, 107–118.
- [170] H. B. Yu, M. Ramseier, R. Bürgi, W. F. van Gunsteren, *Chem. Phys. Chem.* **2004**, *5*, 633–641.
- [171] X. Daura, I. Antes, W. F. van Gunsteren, W. Thiel, A. E. Mark, *Proteins Struct. Funct. Genet.* **1999**, *36*, 542–555.
- [172] A. Glärtli, X. Daura, D. Seebach, W. F. van Gunsteren, *J. Am. Chem. Soc.* **2002**, *124*, 12972–12978.
- [173] A. Glärtli, W. F. van Gunsteren, *Angew. Chem.* **2004**, *116*, 6472–6476; *Angew. Chem. Int. Ed.* **2004**, *43*, 6312–6316.
- [174] K. Gademann, A. Häne, M. Rueping, B. Jaun, D. Seebach, *Angew. Chem.* **2003**, *115*, 1573–1575; *Angew. Chem. Int. Ed.* **2003**, *42*, 1534–1537.
- [175] T. A. Soares, X. Daura, C. Oostenbrink, L. J. Smith, W. F. van Gunsteren, *J. Biomol. NMR* **2004**, *30*, 407–422.
- [176] L. J. Smith, M. J. Sutcliffe, C. Redfield, C. M. Dobson, *J. Mol. Biol.* **1993**, *229*, 930–944.
- [177] H. Schwalbe, S. B. Grimshaw, A. Spencer, M. Buck, J. Boyd, C. M. Dobson, C. Redfield, L. J. Smith, *Protein Sci.* **2001**, *10*, 677–688.
- [178] W. F. van Gunsteren, A. E. Mark, *J. Chem. Phys.* **1998**, *108*, 6109–6116.
- [179] „Modelling of Molecular Structures, Properties“: W. F. van Gunsteren in *Studies in Physical Theoretical Chemistry*, Vol. 71 (Hrsg.: J.-L. Rivail), Elsevier, Amsterdam, **1990**, S. 463–478.
- [180] W. F. van Gunsteren, A. E. Mark, *Eur. J. Biochem.* **1992**, *204*, 947–961.
- [181] P. H. Hünenberger, A. E. Mark, W. F. van Gunsteren, *J. Mol. Biol.* **1995**, *252*, 492–503.

- [182] B. G. Fitch, R. S. Germain, M. Mendell, J. Pitera, M. Pitman, A. Rayshubskiy, Y. Sham, F. Suits, W. Swope, T. J. C. Ward, Y. Zhestkov, R. Zhou, *J. Parall. Distr. Comp.* **2003**, *63*, 759–773.
- [183] J. Norberg, L. Nilsson, *Q. Rev. Biophys.* **2003**, *36*, 257–306.
- [184] A. D. MacKerell, *J. Comput. Chem.* **2004**, *25*, 1584–1604.
- [185] M. J. Field, P. A. Bash, M. Karplus, *J. Comput. Chem.* **1990**, *11*, 700–733.
- [186] D. Bakowies, W. Thiel, *J. Phys. Chem.* **1996**, *100*, 10580–10594.
- [187] A. Warshel, *Computer Modelling of Chemical Reactions in Enzymes and Solution*, Wiley, New York, **1991**.
- [188] H. J. C. Berendsen, J. Mavri, *J. Phys. Chem.* **1993**, *97*, 13464–13468.
- [189] S. R. Billeter, W. F. van Gunsteren, *Comput. Phys. Commun.* **1997**, *107*, 61–91.
- [190] H. J. C. Berendsen, J. Mavri in *Theoretical Treatments of Hydrogen Bonding* (Hrsg.: D. Hadzi), Wiley, New York, **1997**, S. 119–141.
- [191] S. R. Billeter, S. P. Webb, T. Iordanov, P. K. Agarwal, S. Hammes-Schiffer, *J. Chem. Phys.* **2001**, *114*, 6925–6936.
- [192] S. J. Benkovic, S. Hammes-Schiffer, *Science* **2003**, *301*, 1196–1202.
- [193] T. Hansson, C. Oostenbrink, W. F. van Gunsteren, *Curr. Opin. Struct. Biol.* **2002**, *12*, 190–196.
- [194] J. C. Shillcock, R. Lipowsky, *J. Chem. Phys.* **2002**, *117*, 5048–5061.

The screenshot shows the Wiley InterScience homepage. At the top left is the Wiley logo and the text "WILEY InterScience® DISCOVER SOMETHING GREAT". Below this is a section titled "Discover some of the finest full text journals, reference works, books, and databases from around the globe. It's just what you need to make some important discoveries of your own." To the right are links for "ABOUT US", "VIEW DEMO", "CONTACT US", and "HELP". On the far right is a login form with fields for "USER NAME" and "PASSWORD", a "Remember Me" checkbox, and a "Go" button. Below the login form are links for "Register Now", "Athens Login", and "Forgot My Password".

## Manage your access easily with “MY PROFILE”

Simply register. Registration is fast and free to all internet users.

### Easy Access

- Save Titles, Articles & Queries for quick access
- Set up roaming access to access content outside of your institutions network
- Get free online sample copies
- Get free online trial subscriptions
- View a complete list of your subscriptions and accessible products

### Enhanced Tools

- Receive E-Mail Alerts when new content is available
- Purchase Article Select Tokens online
- Purchase individual articles online with Pay-Per-View

A grid of journal covers including *Angewandte Chemie*, *Chemistry - A European Journal*, *EurJIC*, *EurJOC*, *ZAAC*, *CHEMPHYS CHEM*, and *Chinese Journal of Chemistry*.

[www.interscience.wiley.com](http://www.interscience.wiley.com)

1734502\_v0