Rechnergestützte Beschreibung der Struktur und Dynamik von Peptiden und ihren Bausteinen

Rudolf Reichold

SIGILL. UNIVERSITAT. LUDOVIC. MAXIM.

München 2009

Rechnergestützte Beschreibung der Struktur und Dynamik von Peptiden und ihren Bausteinen

Rudolf Reichold

Dissertation
an der Fakultät für Physik
der Ludwig-Maximilians-Universität
München

vorgelegt von
Rudolf Reichold
aus München

München, im Januar 2009

Erstgutachter: Prof. Paul Tavan
Zweitgutachter: Prof. Ulrich Gerland
Tag der mündlichen Prüfung: 24. April 2009

Zusammenfassung

Bedingt durch Fortschritte der Computertechnik und bei der Entwicklung schnellerer Algorithmen sowie verbesserter physikalischer Näherungen gewinnen numerische Simulationen für die Untersuchung biomolekularer Systeme zunehmend an Bedeutung. Ausschlag gebend für den Erfolg derartiger Untersuchungen ist es, die theoretische Methode auf das behandelte System und die konkrete Fragestellung abzustimmen.

Im ersten Teil dieser Arbeit werden die Konformationseigenschaften des nichtnatürlichen, synthetischen 4-Mercaptoprolins, einem Chalkogen-Analogon des in der Natur oft anzutreffenden Hydroxyprolins, untersucht. Durch NMR-Messungen wurde festgestellt, dass die 4*R*- und 4*S*-Epimere des Mercaptoprolins deutlich andere Konformationspräferenzen zeigen als die 4*R*- und 4*S*-Epimere des Hydroxyprolins. Mit Hilfe der Dichtefunktionaltheorie wird aufgeklärt, wodurch die Konformationseigenschaften des 4-Mercaptoprolins bestimmt werden.

Im Mittelpunkt des zweiten Teils steht der Faltungsprozess α-helikaler Peptide in Lösung. Wegen der Größe der dabei betrachteten Systeme ist zu deren Beschreibung die Dichtefunktionaltheorie nicht mehr geeignet. Es muss auf molekularmechanische Kraftfelder zurückgegriffen werden. In dieser Arbeit wird nun der Frage nachgegangen, inwieweit CHARMM22 (C22), ein weit verbreitetes molekularmechanisches Kraftfeld, in Verbindung mit der Rechenleistung heute verfügbarer Computer geeignet ist, die Faltung α-helikaler Peptide in Molekulardynamik-Simulationen zu untersuchen. Dabei werden mit dem Sampling-Problem und dem Kraftfeld-Problem die beiden wesentlichen Hindernisse auf dem Weg zu realistischen Simulationen dieses Prozesses identifiziert. Im Anschluss wird gezeigt, dass die von Liu et al. im Jahr 2005 vorgeschlagene Methode REST [P. Liu et al., *Proc. Natl. Acad. Sci. U.S.A.* **102**, 13749-13754 (2005)] ein effizientes Abtasten des Konfigurationsraums helikaler Peptide in Lösung ermöglicht. Zumindest für die Untersuchung der Gleichgewichtseigenschaften der von uns betrachteten Peptide löst REST also das Sampling-Problem. Durch den Einsatz dieser Technik konnten wir dann bestätigen, dass das um den Energiebeitrag CMAP erweiterte Kraftfeld C22/CMAP im Gegensatz zu C22 eine qualitativ korrekte Beschreibung der Struktureigenschaften α-helikaler Peptide erlaubt. Andererseits wird durch Vergleiche von Simulationsergebnissen mit den Ergebnissen aus spektroskopischen Messungen zum ersten Mal klar gezeigt, dass das Kraftfeld C22/CMAP α-helikale Peptidstrukturen deutlich überstabilisiert.

Prozesse wie die Faltung einer Helix sind stark temperaturabhängig. Deshalb spielt die Temperaturkontrolle in MD-Simulationen eine große Rolle. Im dritten Teil dieser Arbeit werden daher einige gängige Strategien zur Temperaturkontrolle detailliert analysiert. Aus der Diskussion ihrer Mängel wird eine neue Strategie entwickelt.

Inhaltsverzeichnis

1 Einführung **1**
1.1 Proteine — Aufbau und Struktur 1
1.2 Experimentelle Methoden zur Strukturaufklärung 6
1.2.1 Circulardichroismus-Spektroskopie 7
1.2.2 Kernspinresonanz-Spektroskopie 8
1.3 Theoretische Methoden 9
1.3.1 Dichtefunktionaltheorie 10
1.3.2 Molekularmechanik 11
1.3.3 Molekulardynamik 14
1.4 Ziele und Gliederung dieser Arbeit 15

2 Konformationseigenschaften des 4-Mercaptoprolins und verwandter Derivate **19**

3 Molekulardynamik-Simulationen der Faltungs/Entfaltungs-Gleichgewichte α-helikaler Peptide **31**

4 Temperaturkontrolle in Molekulardynamik-Simulationen heterogener Systeme **61**

5 Zusammenfassung und Ausblick **77**

A CMAP — der neue Energieterm in CHARMM **85**
A.1 Das Prinzip 86
A.2 Spline-Interpolation 87
A.3 Berechnung der CMAP-Kräfte 89
A.4 Die Energiefläche 91

Literaturverzeichnis **95**

Danksagung **105**

Lebenslauf **107**

1 Einführung

Leben wie wir es kennen ist ohne Proteine nicht denkbar. Als hochspezialisierte und erstaunlich effiziente molekulare Maschinen nehmen Proteine in allen lebenden Organismen vielfältige Aufgaben wahr. Strukturproteine sind die Bausteine, aus denen das Grundgerüst der Zellen und des Gewebes der Lebewesen gebildet wird. So stellt das Faserprotein Kollagen, das bei Säugetieren das häufigste Protein ist und dort etwa ein Viertel des Gesamtproteingewichts ausmacht [1], den wesentlichen Bestandteil des Bindegewebes dar und ist damit z.B. der wichtigste Bestandteil von Haut und Knochen. Aus Aktin, einem weiteren Strukturprotein, entstehen Filamente, die zum einen Zellen stabilisieren, zugleich aber auch als intrazelluläre Transportwege genutzt werden. Auf diesen Fasern können bestimmte Mitglieder der Motorproteinfamilie des Myosins „entlangwandern“ und dabei Lasten transportieren [1, 2]. Ähnliche Transportfunktionen übernehmen die Motorproteingruppen Kinesin und Dynein im Zusammenwirken mit einer anderen, als Mikrotubuli bezeichneten, Klasse von Strukturproteinen [1, 3]. Ein weiteres Beispiel für den molekularen Maschinenpark der Proteine ist das Membranprotein Bacteriorhodopsin, das im Halobacterium Salinarum als lichtgetriebene Protonenpumpe wirkt [4]. Auch als Katalysatoren für biochemische Reaktionen spielen Proteine – man bezeichnet sie dann als Enzyme – im Stoffwechsel aller lebenden Organismen eine tragende Rolle [5].

Für die Wahrnehmung der spezifischen Funktion eines jeden Proteins sind dessen Struktur und Beweglichkeit von entscheidender Bedeutung [6–8]. Dieser Zusammenhang zwischen Struktur und Funktion geht sogar so weit, dass bestimmte Proteine, wenn sie in einer anderen als ihrer nativen Struktur vorliegen, nicht nur ihre eigentliche Funktion verlieren, sondern sogar schwere Krankheiten auslösen. Beispiele hierfür sind verschiedene spongiforme Enzephalopathien wie CJK und Kuru beim Menschen oder Scrapie und BSE bei Tieren [9].

Dieser kurze Überblick über die Bedeutung der Proteine für unser und das uns umgebende Leben macht deutlich, dass eine Beschäftigung mit diesen Makromolekülen eine reizvolle und lohnende Aufgabe sein kann. Im nächsten Abschnitt sollen diese Systeme deshalb näher beschrieben werden.

1.1 Proteine — Aufbau und Struktur

Die Grundbausteine der Proteine sind die α-Aminosäuren. Wie in Abbildung 1.1 dargestellt, besteht jede dieser Aminosäuren aus einem zentralen Kohlenstoffatom

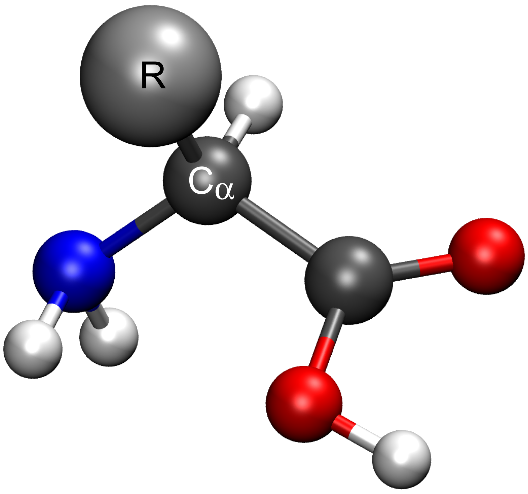


Abbildung 1.1: Aminosäuren sind die Grundbausteine der Proteine. Jede Aminosäure besteht aus einem zentralen Kohlenstoffatom C_α, an das neben einem Wasserstoffatom (H) eine Aminogruppe (NH_2), eine Carboxylgruppe (COOH) und eine Seitengruppe (R) gebunden sind. Die Aminosäuren unterscheiden sich untereinander nur durch ihre Seitengruppe. Bis auf die einfachste Aminosäure (Glycin, hier gilt R=H) besitzen alle Aminosäuren mit ihrem C_α-Atom ein chirales Zentrum, weshalb je zwei Enantiomere existieren. Für die Proteinsynthese werden ausschließlich die hier abgebildeten L-Aminosäuren verwendet.

(C_α), an das neben einem Wasserstoffatom (H) eine Aminogruppe (NH_2), eine Carboxylgruppe (COOH) sowie eine, die Aminosäure spezifizierende, Seitengruppe (R) gebunden sind. Die verschiedenen Aminosäuren unterscheiden sich also nur durch ihre Seitengruppe und können nach deren chemischen Eigenschaften (z.B. polar, apolar, sauer, basisch) in verschiedene Klassen eingeteilt werden [1]. Im genetischen Code der DNS sind 20, die sogenannten *kanonischen*, Aminosäuren kodiert [1].

Das zentrale Kohlenstoffatom C_α stellt bei allen Aminosäuren außer Glycin (hier besteht die Seitengruppe aus einem einfachen Wasserstoffatom) ein chirales Zentrum dar, weshalb je zwei Enantiomere existieren, die L- und die D-Aminosäuren. Für die Biosynthese der Proteine werden nur die L-Aminosäuren verwendet. Proteine sind lineare Ketten aus einigen 100 bis einigen 1000 Aminosäureresten, die auch Residuen genannt werden. Diese sind über die sogenannte Peptidbindung miteinander verbunden, bei der je zwei Aminosäuren unter Abspaltung eines Wassermoleküls eine kovalente Bindung ausbilden (vgl. Abbildung 1.2). Eine kürzere Aminosäurekette nennt man Peptid. Durch die Peptidbindung zweier Aminosäuren ensteht als Bindeglied der beiden C_α-Atome ein sogenanntes Peptidplättchen. Aufgrund mesomerer Effekte ist die zentrale Bindung dieses Komplexes sehr torsionsstabil, weshalb das Peptidplättchen eine starre Ebene bildet. Durch die hohen Elekronegativitäten seines Sauerstoff- und Stickstoffatoms (O und N) besitzt das Peptidplättchen zudem ein

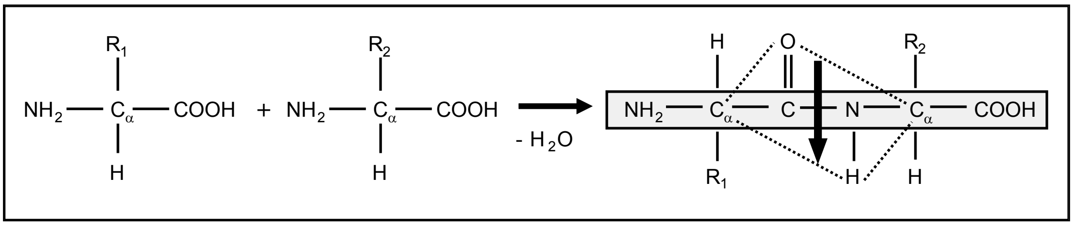


Abbildung 1.2: Die Synthese einer Polypeptidkette aus Aminosäuren. Bei der Peptidsynthese bilden je zwei Aminosäuren (links) unter Wasserabspaltung eine kovalente Bindung, die sogenannte Peptidbindung, aus. Das dabei entstehende torsionsstabile ebene Peptidplättchen trägt ein ausgeprägtes Dipolmoment (senkrechter Pfeil) und ist in der Abbildung durch einen gestrichelten Rahmen hervorgehoben. Die grau unterlegte Atomkette wird als Protein*rückgrat* (engl. *backbone*) bezeichnet. Die Abbildung basiert auf einer Abbildung in [10].

großes Dipolmoment, da diese Atome Elektronen von dem benachbarten Kohlenstoff- und Wasserstoffatom (C und H) abziehen. Wegen der Planarität des Peptidplättchens weist das Proteinrückgrat pro Aminosäurerest nur zwei Torsionsfreiheitsgrade auf: eine Rotation um die NC_α-Bindung und eine Rotation um die C_αC-Bindung. Die entsprechenden Torsions- bzw. Diederwinkel werden üblicherweise ϕ und ψ genannt.

In den Zellen findet die Synthese von Proteinen an großen DNS/Protein-Komplexen [1], den sogenannten Ribosomen, statt. Dieser Vorgang wird *Translation* genannt. In posttranslationalen Prozessen werden die Aminosäurereste zum Teil noch z.B. durch Hydroxylierung, Carboxylierung oder Phosphorylierung modifiziert [1]. Ein Beispiel hierfür ist das bereits erwähnte Faserprotein Kollagen. Dieses Protein besteht aus drei zu einer Dreifachhelix gewundenen Polypeptidsträngen, die jeweils bis zu 300 Wiederholungen der Sequenz Gly-X-Y enthalten [11, 12]. Hier ist X meist ein Prolinrest, wohingegen an der Y-Position oft das Prolinderivat $(2S, 4R)$-Hydroxyprolin sitzt [13], das durch posttranslationale Hydroxylierung aus einem Prolinrest entsteht [1]. Auf das Prolin und verschiedene von ihm abgeleitete Derivate werden wir in Kapitel 2, das eine in der Zeitschrift *Angewandte Chemie* publizierte Arbeit [14] vorstellt, noch einmal zu sprechen kommen.

Die Abfolge der Aminosäuren in einem Protein wird als dessen *Primärstruktur* bezeichnet. Nach der Synthese nimmt jedes Protein in seiner nativen Umgebung eine ganz spezifische, durch seine Primärstruktur bestimmte [15], dreidimensionale Struktur, seine *Tertiärstruktur*, an. Im Sinne der Thermodynamik minimiert diese Struktur die freie Energie des Proteins und seiner Umgebung [16, 17]. Die Details dieses als *Proteinfaltung* bezeichneten Prozesses werden seit vielen Jahren intensiv untersucht und diskutiert [18–23]. Eine wichtige Rolle spielt dabei sicherlich die Ausbildung von *Sekundärstruktur*elementen, von lokalen Strukturmotiven also, die haupsächlich durch Wechselwirkungen der Dipolmomente der Peptidplättchen sta-

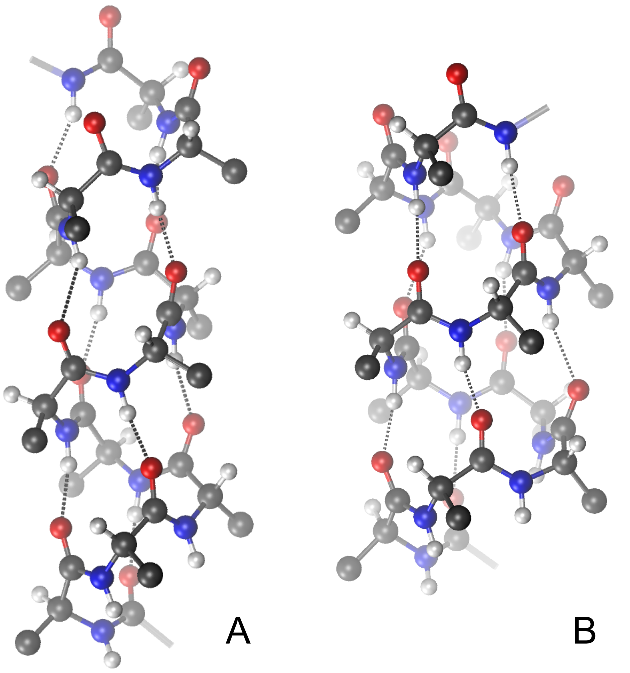


Abbildung 1.3: Die beiden Sekundärstrukturelemente α-Helix (A) und π-Helix (B). In beiden Fällen bildet das Proteinrückgrat eine rechtsgängige Spirale, die durch Wasserstoffbrücken (gestrichelte Linien) stabilisiert wird. In der α-Helix treten diese zwischen der CO-Gruppe eines Residuums i und der NH-Gruppe des viertnächsten Residuums $i+4$ auf. Die π-Helix dagegen ist weniger eng gewunden und ist durch Wasserstoffbrücken vom Typ $(i, i+5)$ gekennzeichnet. Die hier vereinfacht als je ein Atom dargestellten Seitengruppen zeigen in beiden Fällen nach außen.

bilisiert werden. In einer äquivalenten Sichtweise werden Sekundärstrukturelemente durch Wasserstoffbrücken zwischen der CO-Gruppe eines und der NH-Gruppe eines anderen Peptidplättchens stabilisiert.

Linus Pauling, Robert Corey und Herman Branson schlugen 1951 für den Aufbau von Proteinen zwei helikale Sekundärstrukturelemente vor. In diesen bildet das Proteinrückgrat eine Spirale, die durch Wasserstoffbrücken zwischen der CO-Gruppe eines Residuums i und der NH-Gruppe des Residuums $i+4$ bzw. $i+5$ stabilisiert wird [24] (vgl. Abbildung 1.3). Inzwischen haben sich für diese Strukturmotive die Bezeichnungen α-Helix bzw. π-Helix durchgesetzt. Wie sich später herausstellte, ist die α-Helix das in Proteinen am häufigsten auftauchende Sekundärstrukturelement [25], wohingegen die π-Helix in Proteinstrukturen kaum anzutreffen ist [26, 27]. Für die vorliegende Arbeit spielen jedoch beide Strukturmotive eine große Rolle (vgl. Kapitel 3).

Neben den eben eingeführten gibt es eine ganze Reihe weiterer Sekundärstrukturelemente, z.B. den β-Strang, die PolyprolinII (PPII)-Helix, oder die linksgängi-

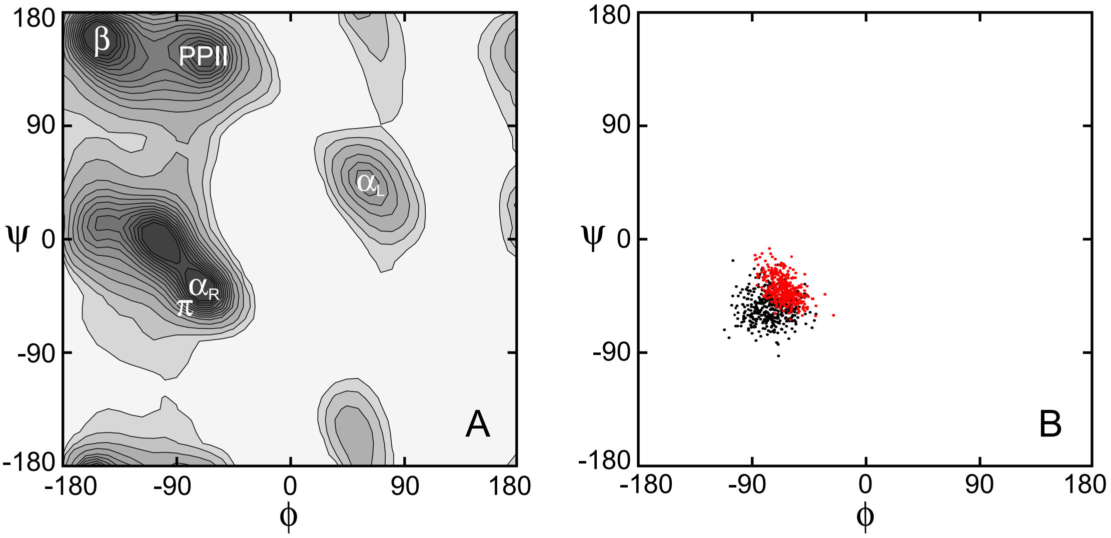


Abbildung 1.4: Konturdarstellung einer Histogrammschätzung der Punktdichte der Verteilung der ϕ/ψ-Winkel aus einer MD-Simulation des Alanin-Dipeptids (A). Je dunkler ein Bereich dargestellt ist, desto häufiger wurde er während der Simulation besucht. Einige Sekundärstrukturelemente (α_R, π, β, PPII, α_L), die durch entsprechende Werte der ϕ/ψ-Winkel gekennzeichnet sind, sind eingezeichnet. In (B) zum Vergleich die Punktdichte der Winkelverteilungen aus einer MD-Simulation eines kurzen Peptids (Simulation $\mathrm{P_E}$.1 aus Kapitel 3), in der sowohl α-helikale (rot) als auch π-helikale (schwarz) Strukturen beobachtet wurden. Hier ist deutlich zu erkennen, dass die Zentren der α- und π-helikalen Bereiche zwar, wie in (A) angedeutet, an unterschiedlichen Punkten der ϕ/ψ-Ebene liegen, die beiden Bereiche sich aber deutlich überlappen. Dies bedeutet unter anderem, dass die ϕ/ψ-Winkel keine geeignete Observable zur Unterscheidung α- und π-helikaler Strukturen darstellen. Für eine solche Unterscheidung muss die Struktur der Wasserstoffbrücken analysiert werden.

ge α-Helix. Diese unterscheiden sich voneinander nicht nur durch ihre Wasserstoffbrückenbindungsmuster, sondern, damit verbunden, auch durch die Geometrie des Proteinrückgrats. Diese kann, wie bereits erwähnt, durch die Koordinaten ϕ und ψ beschrieben werden. Die Darstellung der Verteilung der ϕ/ψ-Winkel in einem Protein oder Peptid in der von diesen Winkeln aufgespannten Ebene bezeichnet man als Ramachandran-Plot [28]. Abbildung 1.4 zeigt Beispiele solcher Darstellungen.

Im gefalteten Protein stellen die Sekundärstrukturelemente relativ starre Einheiten dar, die dem Protein seine Struktur verleihen. Oft sind sie über flexible Schleifen (engl. *loops*) miteinander verbunden, die für die Beweglichkeit des Proteins sorgen. Sowohl die Struktur als auch die Beweglichkeit eines Proteins in seiner nativen Umgebung, die, wie weiter oben bereits erwähnt wurde, letztlich entscheidend für seine

Funktion sind, sind also in der Primärstruktur des Proteins kodiert. Diese Feststellung bietet sich als Ausgangspunkt für vielfältige Untersuchungen an. Zwei Aspekte sollen in der vorliegenden Arbeit behandelt werden.

Wie oben dargelegt wurde, sollte sich prinzipiell allein aus der Kenntnis der Primärstruktur eines Proteins dessen Tertiärstruktur vorhersagen lassen. Obwohl es vielversprechende Ansätze gibt, ist man von der endgültigen Lösung dieser Aufgabe noch weit entfernt. Eines aber ist klar: Um vorhersagen zu können, welche Struktur eine gegebene Aminosäuresequenz einnimmt, ist eine genaue Kenntnis der von den beteiligten Aminosäuren bervorzugten Konformationen unabdingbar. In Kapitel 2 soll deshalb eine Arbeit vorgestellt werden, in der die Konformationseigenschaften bestimmter Aminosäuren untersucht werden.

Der zweite Aspekt ist der Prozess der Faltung. Es ist schließlich alles andere als offensichtlich, wie ein Protein aus einer ungefalteten in seine native Struktur findet. Man hat schon früh erkannt, dass ein zufälliges Ausprobieren aller möglichen Konformationen viel zu lange dauern würde [29]. Immerhin läuft der Faltungsprozess kleinerer Proteine auf Zeitskalen von Mikrosekunden bis Sekunden ab [30]. Ein wichtiger Schritt auf dem Weg vom ungefalteten zum gefalteten Protein ist aber zweifellos die Ausbildung lokaler Sekundärstrukturelemente. Für einige Proteine konnte sogar gezeigt werden, dass z.B. die Faltung von α-Helizes unabhängig von anderen Faltungsprozessen und vor der Ausbildung der Tertiärstruktur erfolgt [20, 31]. Die Bildung von α-Helizes kann also als elementarer Schritt der Proteinfaltung betrachtet und an Modellpeptiden untersucht werden [32]. Dies ist das Ziel der in Kapitel 3 vorgestellten Arbeit [33].

Sowohl in der in Kapitel 2 als auch in der in Kapitel 3 vorgestellten Arbeit kamen experimentelle sowie theoretische, rechnergestützte Methoden zum Einsatz. Diese Methoden sollen im Folgenden vorgestellt werden. Da ich an der Anwendung der experimentellen Methoden nicht beteiligt war, beschränkt sich die Darstellung dieser Methoden auf deren wesentliche Grundzüge. Die theoretischen Methoden, die im Zentrum dieser Arbeit standen, werden dagegen deutlich detaillierter eingeführt.

1.2 Experimentelle Methoden zur Strukturaufklärung

In den in den Kapiteln 2 und 3 vorgestellten Arbeiten kamen zwei experimentelle Methoden zur Strukturaufklärung von Proteinen oder Peptiden zum Einsatz. Dabei handelt es sich um die Circulardichroismus- und die Kernspinresonanz-Spektroskopie. Die beiden Methoden basieren auf völlig unterschiedlichen Prinzipien. In den folgenden zwei Abschnitten werden beide Methoden kurz vorgestellt und ihre Vor- und Nachteile diskutiert.

1.2.1 Circulardichroismus-Spektroskopie

Die Circulardichroismus (CD)-Spektroskopie ist eine absorptionsspektroskopische Methode [34]. Sie nutzt aus, dass die L-Aminosäuren, aus denen Proteine und Peptide aufgebaut sind (vgl. Kapitel 1), mit zirkular polarisiertem Licht je nach dessen Drehsinn unterschiedlich wechselwirken [35]. Bei der CD-Spektroskopie werden in einem bestimmten Wellenlängenbereich (meist im UV-Bereich) die Absorptionskoeffizienten einer Probe für links- und rechts-zirkular polarisiertes Licht gemessen. Daraus wird die Differenz von deren Extinktionskoeffizienten ϵ für links- bzw. rechts-zirkular polarisiertes Licht in Abhängigkeit der Wellenlänge λ bestimmt:

$$\Delta\epsilon(\lambda) = \epsilon_L(\lambda) - \epsilon_R(\lambda). \tag{1.1}$$

Meistens wird aus historischen Gründen in einem CD-Spektrum nicht direkt $\Delta\epsilon(\lambda)$ gegen λ aufgetragen, sondern stattdessen die molare Elliptizität $[\theta]$, die in $\mathrm{deg} \cdot \mathrm{dmol}^{-1} \cdot \mathrm{cm}^2$ angegeben wird (vgl. Kapitel 3). Für den Zusammenhang zwischen diesen beiden Größen gilt [34]

$$\Delta\epsilon = \frac{[\theta]}{3298}. \tag{1.2}$$

Wie sich herausgestellt hat, liefern die verschiedenen Sekundärstrukturelemente ganz unterschiedliche CD-Spektren. Abbildung 1.5 zeigt die typischen CD-Spektren der Sekundärstrukturelemente α-Helix und β-Faltblatt sowie ein Spektrum, das typisch für ungefaltete Strukturen ist.

Durch den Vergleich des CD-Spektrums eines Peptids oder Proteins, dessen Struktur unbekannt ist, mit derartigen Referenzspektren kann der Sekundärstrukturgehalt des Moleküls bestimmt werden [37–40]. Merkmale des α-helikalen Spektrums sind ein doppeltes Minimum bei Wellenlängen von 210 nm und 222 nm sowie ein ausgeprägtes Maximum bei 192 nm. Für α-helikale Peptide besteht ein linearer Zusammenhang zwischen der Elliptizität $[\theta]_{222}$ bei 222 nm und dem Helixgehalt des Peptids [32, 41], weshalb sich dieser durch eine einfache Messung von $[\theta]_{222}$ bestimmen lässt.

Schon lange gibt es Bestrebungen, die CD-Spektren von Peptiden zu berechnen, um so ein besseres Verständnis für den Zusammenhang zwischen Struktur und Spektrum zu erreichen [42–46]. Die entsprechenden Methoden sind allerdings noch nicht ganz ausgereift und konnten sich bisher nicht allgemein durchsetzen.

Ein großer Vorteil der CD-Spektroskopie ist, dass sie auf Moleküle in Lösung angewendet werden kann, so dass Proteine und Peptide in nativer Umgebung untersucht werden können. Außerdem lassen sich CD-Messungen schnell und einfach durchführen und benötigen nur geringe Probenmengen [47]. Allerdings können mit dieser Methode nur die globalen strukturellen Eigenschaften eines Moleküls bestimmt werden. Deutlich detailliertere Einblicke in die Struktur eines Peptids oder Proteins gestattet hingegen die Kernspinresonanz-Spektroskopie, die allerdings auch mit einem deutlich erhöhten technischen Aufwand verbunden ist.

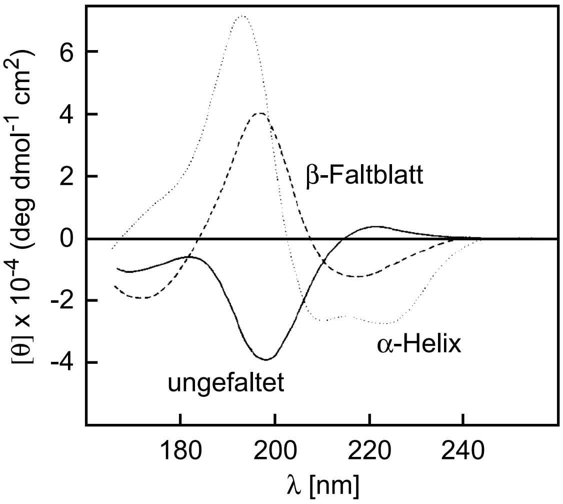


Abbildung 1.5: Charakteristische CD-Spektren für die Sekundärstrukturelemente α-Helix und β-Faltblatt sowie für ungefaltete Strukturen. Die Aufnahme eines CD-Spektrums für ein Peptid, dessen Struktur unbekannt ist, erlaubt Rückschlüsse auf dessen Sekundärstrukturgehalt. Die Abbildung basiert auf einer Vorlage aus [36].

1.2.2 Kernspinresonanz-Spektroskopie

Atomkerne, die einen Spin tragen, wechselwirken mit magnetischen Feldern. Diese Wechselwirkung ist die Grundlage der Kernspinresonanz (NMR, engl. *nuclear magnetic resonance*)-Spektroskopie [48]. Bei dieser Technik befindet sich die Probe in einem externen Magnetfeld. Die Energie eines Atomkerns in diesem Feld hängt nun von der Orientierung seines Spins relativ zu dem Feld ab, so dass je nach Spin eine unterschiedliche Anzahl von Energienieveaus entsteht. Durch das Einstrahlen elektromagnetischer Pulse können Übergänge zwischen diesen Energieniveaus angeregt werden und anschließend kann die Relaxation der Besetzung der unterschiedlichen Niveaus ins Gleichgewicht beobachtet werden [35]. Die Resonanzfrequenz, bei der ein bestimmter Atomkern Energie absorbiert, hängt von dem Magnetfeld am Ort des Kerns ab. Dieses setzt sich aus dem externen Feld und einem lokalen, von den Elektronen in der Umgebung des Kerns erzeugten, Feld zusammen. Dies hat zur Folge, dass sich z.B. die Absorptionsfrequenzen der Wasserstoffkerne einer Probe je nach deren jeweiliger Umgebung unterscheiden. Dieses Phänomen wird *chemische Verschiebung* (engl. *chemical shift*) genannt [49] und erlaubt es, aus den Absorptionsfrequenzen Rückschlüsse auf die chemische Umgebung der Atomkerne zu ziehen [35].

Um jedoch Informationen über die räumliche Anordnung der Atomkerne zueinander zu bekommen, kann die Wechselwirkung verschiedener Kernspins ausgenutzt werden. Hierbei werden zwei Prozesse unterschieden [50]. Der eine betrifft Kerne von

Atomen, die kovalent miteinander verbunden sind. Hier wird eine Spin-Spin Wechselwirkung durch die Bindungselektronen vermittelt (*through bond*). Diese skalare Wechselwirkung wird auch J-Kopplung (*J coupling*) genannt [50]. Für den Fall von zwei Atomen, die über drei kovalente Bindungen miteinander verbunden sind, kann aus der entsprechenden Kopplungskonstante ^{3}J der Diederwinkel dieses Bindungskomplexes bestimmt werden [51–53]. Zusätzlich sind die Kernspins durch die magnetische Dipol-Dipol Wechselwirkung miteinander gekoppelt (*through space*) [50]. Über diese auch *Nuclear Overhauser Effect* (NOE) [54] genannte Wechselwirkung lassen sich Obergrenzen für den Abstand zweier Atomkerne bestimmen, falls dieser kleiner als 5 Å ist [55].

Da die NMR-Spektroskopie auf der Wechselwirkung von Kernspins untereinander und mit magnetischen Feldern beruht, sieht sie nur Kerne, die einen Spin tragen. In Biomolekülen sind dies vor allem die Kerne der Wasserstoffatome. Die Kenntnis der per NMR bestimmten Atomabstände und Diederwinkel reicht nicht aus, um die dreidimensionale Struktur eines Moleküls eindeutig festzulegen. Deshalb schließt sich an die NMR-Messung oft ein Strukturverfeinerungsprozess (engl. *structure refinement*) an, in dem zusätzliche Informationen (z.B. die Topologie der Aminosäuren und typische Werte für Bindungslängen und -winkel) dazu verwendet werden, ein Ensemble aus Strukturen zu generieren, die jeweils sowohl mit diesen Informationen als auch mit den experimentell bestimmten Werten für Abstände und Diederwinkel vereinbar sind [56, 57]. Diese zusätzlichen Informationen liegen meist in Form eines molekularmechanischen Kraftfelds vor. Auf derartige Kraftfelder wird in Abschnitt 1.3.2 ausführlich eingegangen werden.

Abgesehen davon, dass die NMR-Spektroskopie sehr detaillierte Einblicke in die Struktur von Molekülen ermöglicht, ist ein großer Vorteil dieser Technik, dass auch hier, ähnlich wie bei der CD-Spektroskopie, Moleküle in Lösung untersucht werden können. Proteine und Peptide können also in ihrer nativen Umgebung studiert werden, so dass sich ihre native, funktionale Struktur bestimmen lässt. Allerdings erfordert die NMR- im Vergleich zur CD-Spektroskopie einen viel höheren technischen Aufwand.

1.3 Theoretische Methoden

Wie bereits in Abschnitt 1.1 erwähnt wurde, stand die Anwendung theoretischer Methoden zur Untersuchung von Struktur und Dynamik von Biomolekülen im Zentrum dieser Arbeit. Dabei wurden die Ergebnisse dieser theoretischen Untersuchungen jeweils mit den Ergebnissen experimenteller Untersuchungen an den gleichen Systemen verglichen. Nachdem in den Abschnitten 1.2.1 und 1.2.2 die dabei verwendeten experimentellen Methoden vorgestellt wurden, sollen in den nächsten Abschnitten mit der Dichtefunktionaltheorie, dem molekularmechanischen Kraftfeld und der Molekulardynamik die theoretischen Methoden eingeführt werden, die in dieser Arbeit zur Anwendung kamen.

1.3.1 Dichtefunktionaltheorie

Die Dichtefunktionaltheorie (DFT) ist eine quantenmechanische Methode, die es ermöglicht, mit moderatem Rechenaufwand die Grundzustandsenergie von kleineren bis mittelgroßen Molekülen (mit bis zu 100 Atomen) zu bestimmen [58]. Die DFT kann deshalb zur Geometrieoptimierung derartiger Moleküle verwendet werden. Dabei wird, ausgehend von einer Startstruktur, die Grundzustandsenergie des Moleküls durch gezieltes Verändern der Geometrie des Moleküls minimiert.

Die Ursprünge der DFT reichen in die 1920er Jahre zurück. Damals wurde erkannt, dass die Verteilung der Elektronen in einem Atom durch statistische Betrachtungen näherungsweise bestimmt werden kann [58–61]. Die eigentliche Basis der DFT sind aber die Hohenberg-Kohn-Theoreme [62]. Diese besagen, dass zur Beschreibung des Grundzustands eines Moleküls mit N Elektronen nicht auf die Wellenfunktion $\psi(\mathbf{r}_1, s_1, \mathbf{r}_2, s_2, \ldots, \mathbf{r}_N, s_N)$ zurückgegriffen werden muss, die eine Funktion der Elektronenorte $\mathbf{r}_i$ und Elektronenspins s_i ist, sondern dass dafür die Betrachtung der Elektronendichte $n(\mathbf{r})$ genügt. Die Elektronendichte $n(\mathbf{r})$ lässt sich durch Anwendung der Kohn-Sham-Gleichungen [63] über ein selbstkonsistentes Verfahren berechnen. Neben der Betrachtung der Elektronendichte ist ein wichtiger Aspekt der DFT die Verwendung empirisch parametrisierter Austausch- und Korrelationsfunktionale, die für eine im Vergleich zum Hartree-Fock-Verfahren [64] verbesserte Berücksichtigung der Elektronenkorrelation sorgen. In der vorliegenden Arbeit kamen das Funktional BP86 [65, 66] von Becke und Perdew sowie das Hybridfunktional B3LYP [67, 68] von Becke, Lee, Yang und Parr zum Einsatz.

Es gibt eine Vielzahl von Programmen, in denen die DFT-Methode implementiert ist (z.B. Gaussian [69], CPMD [70], TURBOMOLE [71, 72], ORCA [73], Gamess [74, 75] und Quickstep [76]). Diese Programme unterscheiden sich unter anderem durch die Art der Basisfunktionen, in denen die Molekülorbitale entwickelt werden. Neben einer Entwicklung in ebenen Wellen [70] ist vor allem die Verwendung von Gaußfunktionen üblich [69, 71, 72]. Auch die Kombination beider Ansätze ist möglich [77, 78]. Das in dieser Arbeit eingesetzte Programmpaket TURBOMOLE 5.6 verwendet gaußförmige Basissätze wie z.B. den Basissatz TZVP [79], der in den in den Kapiteln 2 und 3 vorgestellten Rechnungen eingesetzt wurde.

Im Zusammenhang mit der Entwicklung eines polarisierbaren Kraftfelds zur Berechnung von Infrarotspektren des Proteinrückgrats wurde die DFT von mir dazu verwendet, Schwingungsspektren und Potentialkurven eines kleinen Modellmoleküls zu berechnen. Die Ergebnisse dieser Rechnungen sind Teil einer Veröffentlichung [80]. In der vorliegenden Arbeit wurde die DFT sowohl dazu verwendet, die Konformationseigenschaften von einigen kleineren Molekülen zu untersuchen (Kapitel 2), als auch, um Parameter eines molekularmechanischen Kraftfelds zu bestimmen (Kapitel 3). Das Konzept des molekularmechanischen Kraftfelds wird im folgenden Abschnitt vorgestellt.

1.3.2 Molekularmechanik

Wie in Abschnitt 1.3.1 erwähnt wurde, ist die Anwendung der DFT auf Systeme beschränkt, die bis zu etwa 100 Atome enthalten. Größere Systeme können wegen des schnell mit der Systemgröße wachsenden Rechenaufwands der DFT nicht mehr mit dieser Methode behandelt werden. Für die Untersuchung größerer Systeme muss daher auf sogenannte *molekularmechanische* (MM) Kraftfelder zurückgegriffen werden.

In dem MM-Ansatz werden Atome durch Massenpunkte repräsentiert und Wechselwirkungen zwischen den Atomen durch einfache analytische Funktionen beschrieben. Die Gesamtheit dieser Funktionen zusammen mit den bei ihrer Berechnung verwendeten Parametern nennt man MM-Kraftfeld. Für jede Konfiguration $\mathbf{R} = \{\mathbf{r}_1, \ldots, \mathbf{r}_N\}$ eines Systems aus N Atomen liefert ein MM-Kraftfeld also einen Wert $E_{\mathrm{MM}}(\mathbf{R})$ für dessen potentielle Energie. Es gibt eine ganze Reihe weit verbreiteter MM-Kraftfelder [81, 82], die zum Teil in diversen Varianten vorliegen. Beispiele für oft eingesetzte Kraftfelder sind AMBER [83–93], GROMOS [94, 95], OPLS [96–98] und CHARMM [99–101]. In der vorliegenden Arbeit wurde das Kraftfeld CHARMM22 (C22) [100] eingesetzt, zum Teil mit, zum Teil ohne CMAP-Erweiterung [101–104] (vgl. Kapitel 3 und Anhang A). Deshalb soll dieses Kraftfeld nun vorgestellt werden [57, 100, 105].

Im C22-Kraftfeld wird die potentielle Energie $E_{\mathrm{C22}}(\mathbf{R})$ in zwei Beiträge

$$E_{\mathrm{C22}}(\mathbf{R}) = E_{\mathrm{bonded}}(\mathbf{R}) + E_{\mathrm{nonbonded}}(\mathbf{R}) \tag{1.3}$$

aufgeteilt. Dabei fasst E_{bonded} die Wechselwirkungen zusammen, die im Zusammenhang mit kovalenten Bindungen stehen. Man unterscheidet hier drei Beiträge

$$E_{\mathrm{bonded}}(\boldsymbol{b}, \boldsymbol{\theta}, \boldsymbol{\phi}) = E_{\mathrm{bond}}(\boldsymbol{b}) + E_{\mathrm{angle}}(\boldsymbol{\theta}) + E_{\mathrm{dihedral}}(\boldsymbol{\phi}), \tag{1.4}$$

die als Funktionen der sogenannten *internen Koordinaten* $\boldsymbol{b}$, $\boldsymbol{\theta}$ und $\boldsymbol{\phi}$ definiert sind. Ein vierter Beitrag E_{improper}, der eng mit E_{dihedral} verwandt ist, soll hier der Übersichtlichkeit halber vernachlässigt werden. Bei den internen Koordinaten bezeichnet $\boldsymbol{b} = (b_1, \ldots, b_i)$ die Abstände je zweier kovalent gebundener Atome, $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_j)$ die Winkel, die von je zwei solchen Bindungen gebildet werden, und $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_k)$ die sogenannten Dieder-Winkel, die Winkel zwischen zwei von je drei Atomen aufgespannten Flächen beschreiben.

Zur Charakterisierung der in einem Simulationssystem verwendeten Atome setzt C22 das Konzept der Atomtypen ein. Jedem Atom wird je nach seinem chemischen

Charakter ein Atomtyp zugeordnet. So wird z.B. bei den Wasserstoffatomen unter anderem zwischen polaren, apolaren und aromatischen Atomen unterschieden. Die Wahl der Parameter für die Wechselwirkungspotentiale (Gl. 1.5–1.7, 1.10) wird dann je nach den Atomtypen der beteiligten Atome getroffen.

Unter der Annahme kleiner Auslenkungen wird die Bindungsenergie zweier kovalent gebundener Atome in C22 durch harmonische Potentiale

$$E_{\text{bond}}(\boldsymbol{b}) = \sum_{\text{bonds}} k_b (b - b_0)^2 \tag{1.5}$$

beschrieben, wobei b der aktuelle Abstand zweier Atome ist, b_0 deren Gleichgewichtsabstand und k_b die mit den entsprechenden Atomtypen assoziierte Kraftkonstante. Die Summe läuft über alle in dem betrachteten System definierten Bindungen. Um die Übersichtlichkeit der Darstellung zu erhöhen, wurde hier und in den folgenden Gleichungen auf eine entsprechende Indizierung der Parameter verzichtet.

Ähnlich wie die Bindungspotentiale werden auch die Winkelpotentiale gemäß

$$E_{\text{angle}}(\boldsymbol{\theta}) = \sum_{\text{angles}} k_\theta (\theta - \theta_0)^2 \tag{1.6}$$

harmonisch genähert. Diese Näherung kann für die Torsionspotentiale E_{dihedral} nicht mehr angewendet werden, da hier nicht von kleinen Auslenkungen aus der Gleichgewichtslage ausgegangen werden kann. Wegen der Periodizität dieser Potentiale ist für ihre Parametrisierung eine Kosinus-Entwicklung gut geeignet:

$$E_{\text{dihedral}}(\boldsymbol{\phi}) = \sum_{\text{dihedrals}} \sum_{n} k_{\phi,n} \left[1 + \cos(n\,\phi + \delta_n)\right]. \tag{1.7}$$

Während E_{bonded} die Wechselwirkungen berücksichtigt, die im Zusammenhang mit kovalenten Bindungen stehen (Gl. 1.5–1.7), werden die restlichen Wechselwirkungen in $E_{\text{nonbonded}}$ zusammengefasst:

$$E_{\text{nonbonded}}(\mathbf{R}) = E_{\text{VDW}}(\mathbf{R}) + E_{\text{elec}}(\mathbf{R}). \tag{1.8}$$

Beide Beiträge zu $E_{\text{nonbonded}}$ sind Paarwechselwirkungen, die von den paarweisen Abständen der Atome abhängen. Bezeichnet

$$r_{ij} = |\mathbf{r}_i - \mathbf{r}_j| \tag{1.9}$$

den Abstand der Atome i und j, so gilt für die Van der Waals (VDW)-Energie

$$E_{\text{VDW}}(\mathbf{R}) = \sum_{i<j} 4\,\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6} \right]. \tag{1.10}$$

Der abstoßende Anteil ($\sim r_{ij}^{-12}$) dieses sogenannten Lennard-Jones Potentials modelliert die Pauli-Abstoßung zwischen Atomen, der anziehende Anteil ($\sim r_{ij}^{-6}$) die auf Induktionseffekten beruhende attraktive Dispersionswechselwirkung.

Die elektrostatische Wechselwirkung der in der Simulation betrachteten Atome wird über ihre Coulomb-Wechselwirkung berücksichtigt. Wie in den anderen weiter oben erwähnten Kraftfeldern auch, tragen die Atome in C22 statische Partialladungen. Diese wurden von den Kraftfeldentwicklern so gewählt, dass die Effekte der elektronischen Polarisierbarkeit in einer *mean field*-Näherung berücksichtigt werden. Für Proteine und Peptide in Lösung ist diese Näherung jedoch ungeeignet [81, 106, 107]. Daher wird seit längerem an der Entwicklung polarisierbarer Kraftfelder gearbeitet [98, 108–114], die sich jedoch, hauptsächlich wegen ihres hohen Rechenaufwands, bisher nicht auf breiter Front durchsetzen konnten. Sind q_i und q_j also die Partialladungen der Atome i und j, gilt für den elektrostatischen Beitrag zu $E_{\text{nonbonded}}$

$$E_{\text{elec}}(\mathbf{R}) = \sum_{i<j} \frac{q_i\, q_j}{r_{ij}}. \tag{1.11}$$

Die Summen in der Berechnung von E_{VDW} (Gl. 1.10) und E_{elec} (Gl. 1.11) laufen jeweils über alle Atompaare im Simulationssystem mit Ausnahme derjenigen Paare, die kovalent oder über einen Bindungswinkel miteinander verbunden sind. Die Wechselwirkung dieser Atome wird ja bereits durch E_{bonded} erfasst.

Vor einigen Jahren wurde das CHARMM-Kraftfeld um einen zusätzlichen Energieterm E_{CMAP} erweitert [101–104]. Einer der Anlässe für diese Erweiterung war die häufige Beobachtung unrealistischer π-helikaler Strukturen in Simulationen α-helikaler Peptide [101]. Abbildung 1.3 zeigt diese beiden Sekundärstrukturelemente. Während die α-Helix durch Wasserstoffbrücken vom Typ $(i, i+4)$ stabilisiert wird, liegen in der π-Helix Wasserstoffbrücken vom Typ $(i, i+5)$ vor. Dieser Unterschied führt dazu, dass das Peptidrückgrat einer π-Helix weniger eng gewunden ist als das einer α-Helix. Dies äußert sich schließlich in leicht unterschiedlichen Positionen dieser beiden Sekundärstrukturelemente in der Ramachandran-Ebene (vgl. Abbildung 1.4).

Durch eine verbesserte Modellierung der Energetik des Peptid- bzw. Proteinrückgrats sollte sich die unerwünschte Vorliebe von C22 für π-Helizes beseitigen lassen. Genau dies ist der Ansatzpunkt des Energieterms E_{CMAP}. Dieser Energieterm, der einen weiteren Beitrag zu E_{bonded} (Gl. 1.4) liefert, beruht auf einem Grid-basierten Interpolationsverfahren, das eine sehr flexible Anpassung der MM-Potentialfläche in der Ramachandran-Ebene an quantenmechanisch gewonnene Resultate ermöglicht:

$$E_{\text{CMAP}}(\phi, \psi) = \sum_{i,j=0}^{3} c_{i,j} \left(\frac{\phi - \phi_L}{h}\right)^i \left(\frac{\psi - \psi_L}{h}\right)^j. \tag{1.12}$$

Die technischen Details dieses Energieterms sowie Angaben zu seiner Parametrisierung werden in Anhang A diskutiert. Im Verlauf der hier vorgestellten Arbeit zeigte sich, dass diese Kraftfelderweiterung für uns von großer Bedeutung ist. Deshalb wurde unser MD Programmpaket EGO-MMVI im Rahmen dieser Arbeit um die Möglichkeit erweitert, das neue Kraftfeld C22/CMAP einzusetzen. Für die Details dieser Implementierung sei noch einmal auf Anhang A verwiesen.

Die anderen weiter oben erwähnten MM-Kraftfelder verwenden ganz ähnliche Funktionen zur Modellierung der atomaren Wechselwirkungen wie das soeben vorgestellte C22. Nur der Energieterm E_{CMAP} ist bislang eine Spezialität des CHARMM-Kraftfelds. Von diesem Term abgesehen unterscheiden sich die gängigen MM-Kraftfelder also hauptsächlich durch die Werte der Parameter, die bei der Auswertung dieser Funktionen verwendet werden. Die Kraftfelder stellen Parameter zur Simulation üblicher Biomoleküle zur Verfügung. So umfassen sie z.B. die kanonischen Aminosäuren (zum Teil in verschiedenen Protonierungszuständen) und diverse Gruppen zur Terminierung von Polypeptidketten. Möchte man jedoch mit einem MM-Kraftfeld ein Molekül untersuchen, das nicht ausschließlich aus Standardbausteinen aufgebaut ist, müssen die für die Behandlung der exotischen Teile benötigten Parameter (Gl. 1.5–1.7, 1.10, 1.11) erst bestimmt werden. Wie bereits erwähnt wurde, können dafür z.B. DFT-Rechnungen eingesetzt werden (vgl. Kapitel 3).

1.3.3 Molekulardynamik

Im vorigen Abschnitt wurde das Konzept des molekularmechanischen Kraftfelds eingeführt. Ein solches Kraftfeld liefert auf effiziente Weise einen Näherungswert für die potentielle Energie $E_{\mathrm{MM}}(\mathbf{R})$ eines molekularen Systems als Funktion der Ortskoordinaten der betrachteten Atome. Durch Gradientenbildung bezüglich der Atomorte lassen sich aus dieser Energie Kräfte berechnen, die auf die Atome wirken und diese beschleunigen. Diese Beschleunigung führt zu einer Positionsänderung der Atome, wodurch sich wiederum $E_{\mathrm{MM}}(\mathbf{R})$ ändert und damit auch die auf die Atome wirkenden Kräfte. Auf diesem Prinzip beruht die Methode der Molekulardynamik (MD)-Simulation [107, 115–119].

In einer MD-Simulation werden also die Newtonschen Bewegungsgleichungen numerisch integriert. Dabei kommt oft, so z.B. auch in EGO-MMVI, der Verlet-Algorithmus [120]

$$\mathbf{r}_i(t+\Delta t) = 2\mathbf{r}_i(t) - \mathbf{r}_i(t-\Delta t) - \frac{\Delta t^2}{m_i}\boldsymbol{\nabla}_i E_{\mathrm{MM}}(\mathbf{R}(t)) \tag{1.13}$$

zum Einsatz. Hier bezeichnet $\mathbf{r}_i(t)$ den Ort des Atoms i mit der Masse m_i zum Zeitpunkt t. Die Zeit t ist diskretisiert, so dass sich zwei aufeinander folgende Zeitpunkte um den Zeitschritt Δt unterscheiden. Der Verlet-Algorithmus bestimmt also aus dem aktuellen und dem letzten Ort eines Atoms [$\mathbf{r}_i(t)$ und $\mathbf{r}_i(t-\Delta t)$] sowie aus dem Gradienten der Energie E_{MM} den Ort $\mathbf{r}_i(t+\Delta t)$ des Atoms im nächsten Zeitschritt. Aus diesem Grund ist es wichtig, dass E_{MM} eine bezüglich der Atomkoordinaten stetig differenzierbare Funktion ist. Für die Wechselwirkungen des C22-Kraftfelds (Gl. 1.5–1.7, 1.10, 1.11) ist dies der Fall. Auch die CMAP-Energie (Gl. 1.12) ist stetig differenzierbar, falls die Koeffizienten $c_{i,j}$ geeignet gewählt werden. Dieser Punkt wird in Anhang A noch einmal aufgegriffen. Um sicherzustellen, dass in einer MD-Simulation auch die schnellsten atomaren Bewegungen noch hinreichend genau abgetastet werden, darf der Zeitschritt Δt nicht zu groß gewählt werden. Dies gilt insbesondere beim Einsatz eines sogenannten Mehrschrittverfahrens [121–123] zur effizienten Berechnung der langreichweitigen Wechselwirkungen. Üblicherweise werden Zeitschritte im Bereich von 1 fs verwendet [124].

Als erste MD-Simulation eines Proteins gilt die 1977 veröffentlichte Untersuchung des kleinen Proteins BPTI [125]. Obwohl die Dynamik des Moleküls nur über einen kurzen Zeitraum (wenige ps) verfolgt werden konnte, das Protein im Vakuum simuliert wurde und die intramolekularen Wechselwirkungen nur grob genähert wurden, lieferte diese Simulation doch neue Einsichten in die Dynamik von Proteinen [126]. 1984 war es dann dank der Entwicklung leistungsfähigerer Computer und Algorithmen möglich, dieses Protein in wässriger Lösung zu simulieren [127]. Allerdings war auch diese Simulation auf wenige ps beschränkt. Seit diesen Pionierarbeiten ist viel Zeit vergangen, während der die Rechenkraft der Computer enorm gewachsen ist und die Methodik der MM-MD-Simulation stark verbessert wurde. Dies schlägt sich in der ersten $1\,\mu$s-Simulation eines Proteins in Wasser im Jahr 1998 [128], gefolgt von der ersten $10\,\mu$s-Simulation im Jahr 2008 [129], nieder. Die Simulation von kleinen Proteinen oder Peptiden unter nativen oder experimentellen Bedingungen über Zeiträume von vielen Nanosekunden ist heute ein Standardwerkzeug zur Untersuchung von deren Struktur und Dynamik [130–137]. Die Fortschritte in der Computertechnik und bei der Entwicklung effizienter Simulationsmethoden bringen jedoch auch Unzulänglichkeiten der etablierten MM-MD-Simulationstechniken ans Tageslicht, die erst durch umfassende Statistik aufgedeckt werden können. Die in den letzten Jahren entwickelten sogenannten *enhanced sampling*-Methoden (z.B. die *replica exchange*-Methode [138–140]) erfreuen sich zwar großer Beliebtheit, halten jedoch nicht immer das, was sie versprechen [141], und müssen deshalb mit Bedacht verwendet werden [142, 143]. Eine dieser Methoden [144] wird in Kapitel 3 eingesetzt und diskutiert.

Nachdem nun in Abschnitt 1.1 die Systeme eingeführt wurden, die in dieser Arbeit untersucht wurden, und die dabei eingesetzten experimentellen und theoretischen Methoden in den Abschnitten 1.2 und 1.3 vorgestellt wurden, soll im folgenden Abschnitt dargestellt werden, welche Fragen in dieser Arbeit nun konkret untersucht wurden.

1.4 Ziele und Gliederung dieser Arbeit

In Kapitel 2 wird eine Arbeit [14] vorgestellt, in der die Konformationseigenschaften des 4-Mercaptoprolins und verwandter Derivate untersucht werden. Wie bereits in

Abschnitt 1.1 erwähnt wurde, spielen Prolinreste eine wichtige Rolle für die Struktureigenschaften von Peptiden und Proteinen. Oft taucht das Prolin in modifizierter Form auf. Die in der Natur häufigste Modifikation ist das $(2S, 4R)$-Hydroxyprolin (Hyp), das einen wesentlichen Bestandteil der Kollagendreifachhelix darstellt. Die stereoelektronischen Einflüsse der elektronegativen Substituierung am C4 des Pyrrolidinrings wurden in der Vergangenheit in synthetischen Modellverbindungen vergleichend analysiert, insbesondere für Fluor-Substituenten (Flp).

In der hier vorgestellten Arbeit, die gemeinsam mit experimentell arbeitenden Kollegen aus der Arbeitsgruppe Moroder vom MPI für Biochemie in Martinsried publiziert wurde, wird nun über einen Strukturvergleich der $(2S, 4R)$- und $(2S, 4S)$-Epimere der 4-Mercaptopyrrolidin-2-carbonsäuren (Mpc) berichtet. Mpc ist ein nichtnatürliches, synthetisches Chalkogen-Analogon des Hydroxyprolins, das genutzt werden kann, um durch Seitenkettenverknüpfung von Peptiden durch Tioether- oder Disulfidbrücken den Konformationsraum von peptidischen Makrozyklen einzuschränken. Um es gezielt zum Design von Peptiden und Proteinen einsetzen zu können, ist es wichtig, seine Konformationseigenschaften genau zu kennen. Es wird sich zeigen, dass sich die Konformationspräferenzen des Mpc drastisch von denen des Hyp oder Flp unterscheiden. Um die Ursachen dieser experimentell durch NMR-Spektroskopie bestimmten Konformationspräferenzen des Mercaptoprolins aufzuklären, wurden von mir für einige der experimentell untersuchten Modellmoleküle DFT-Rechnungen durchgeführt.

Während in Kapitel 2 also die Konformationseigenschaften einer einzelnen Aminosäure analysiert wurden, wendet sich Kapitel 3, das den Hauptbestandteil meiner Arbeit darstellt, größeren Systemen zu. Wie in Abschnitt 1.3.2 erläutert wurde, muss zu deren Beschreibung auf MM-Kraftfelder zurückgegriffen werden. Kapitel 3 besteht aus einem Manuskript, das meinen Beitrag zu einem von der Volkswagenstiftung geförderten Verbundprojekt zusammenfasst. Dieses Verbundprojekt war von Thomas Kiefhaber und Paul Tavan initiiert worden. Ziel des Projekts war die Untersuchung des lichtinduzierten Faltungsprozesses eines α-helikalen Peptids durch eine Kombination experimenteller und theoretischer Methoden. Dazu wurde ein Molekül entwickelt, das eine intramolekulare Verbrückung enthält, die durch einen kurzen Laserblitz geöffnet werden kann. Das Peptid kann also in zwei Zuständen vorliegen: offen oder geschlossen. Um einen Faltungsprozess beobachten zu können, muss sichergestellt sein, dass das geschlossene Peptid ungefaltet ist, wohingegen das offene Peptid einen hohen Helixgehalt aufweisen sollte. Um den Rechenaufwand der MD-Simulationen begrenzt zu halten, darf das Peptid außerdem nicht zu groß sein.

Da das Peptid mit der lichtspaltbaren Verbrückung eine chemische Gruppe enthält, für deren Beschreibung in den gängigen MM-Kraftfeldern keine Parameter enthalten sind, war der erste Schritt zur Realisierung dieses Projekts die zeitaufwändige Entwicklung eines MM-Kraftfelds für diese Gruppe. Um währenddessen schon mit der Auswahl geeigneter MD-Simulationsmethoden beginnen zu können, wurden einige Modellpeptide untersucht, in denen auf den Einbau der lichtspaltbaren Verbrückung verzichtet wurde. In Kapitel 3 wird nun zum einen die Entwicklung des MM-Kraftfelds für die Verbrückung beschrieben und zum anderen eine Reihe von MD-Simulationen diverser helikaler Peptide präsentiert. Mit der Untersuchung dieser Peptide soll der Frage nachgegangen werden, inwieweit die Qualität eines heute gängigen MM-Kraftfelds und die erzeugbare Menge an statistischen Daten ausreichen, um experimentelle Untersuchungen schnell faltender α-helikaler Peptide durch MD-Simulationen zu ergänzen.

Bei Simulationen von Peptiden in Lösung, wie z.B. denen in Kapitel 3, ist es wichtig, die Temperatur des Simulationssystems auf geeignete Weise zu kontrollieren. Schließlich hat die Temperatur einen großen Einfluss auf die Konformationsdynamik des Peptids. Soll in der Simulation eine bestimmte experimentelle Situation wiedergegeben werden, ist also darauf zu achten, dass in der Simulation das gleiche statistische Ensemble betrachtet wird wie im Experiment. Es existieren etliche Strategien zur Temperaturkontrolle in MD-Simulationen, die immer in die Dynamik des Systems eingreifen.

In der in Kapitel 4 abgedruckten Publikation [145], die aus einer Zusammenarbeit mit meinen Doktorandenkollegen Martin Lingenheil und Robert Denschlag entstanden ist, wird nun eine dieser Strategien aufgegriffen und genau analysiert. Ausgehend von der Diskussion ihrer Mängel wird eine neue Strategie zur Temperaturkontrolle inhomogener Simulationssysteme vorgeschlagen, die einerseits für eine homogene Temperaturverteilung im gesamten System sorgt und dabei andererseits die Eingriffe des Thermostaten in die Dynamik des Systems auf ein Minimum begrenzt.

2 Konformationseigenschaften des 4-Mercaptoprolins und verwandter Derivate

Wir wenden uns zunächst den Konformationseigenschaften des 4-Mercaptoprolins und diverser verwandter Derivate zu. Diese Eigenschaften wurden mit Hilfe sowohl experimenteller als auch theoretischer Methoden untersucht. Das Kapitel enthält einen Abdruck[1] des Artikels

> Sergio A. Cadamuro, Rudolf Reichold, Ulrike Kusebauch, Hans-Jürgen Musiol,
> Christian Renner, Paul Tavan und Luis Moroder:
> „Conformational Properties of 4-Mercaptoproline and Related Derivatives“
> *Angewandte Chemie International Edition* **47**, 2143-2146 (2008),

den ich gemeinsam mit Paul Tavan aus unserer Arbeitsgruppe, sowie mit Sergio A. Cadamuro, Ulrike Kusebauch, Hans-Jürgen Musiol, Christian Renner und Luis Moroder aus der Forschungsgruppe Bioorganische Chemie des Max-Planck-Instituts für Biochemie verfasst habe. Die auf der Internetseite der Angewandten Chemie[2] verfügbaren Hintergrundinformationen zu dieser Veröffentlichung enthalten die technischen Details der durchgeführten Experimente und Simulationen sowie zusätzliche Ergebnisse und sind in dieser Arbeit im Anschluss an die Originalveröffentlichung abgedruckt.

Diese Arbeit habe ich als Erstautor des „Theorie-Teils“ wesentlich mitgestaltet.

Es sei angemerkt, dass der nachfolgend abgedruckte Artikel zusätzlich auch noch in deutscher Übersetzung erschienen ist:

> Sergio A. Cadamuro, Rudolf Reichold, Ulrike Kusebauch, Hans-Jürgen Musiol,
> Christian Renner, Paul Tavan und Luis Moroder:
> „Konformationseigenschaften des 4-Mercaptoprolins und verwandter Derivate“
> *Angewandte Chemie* **120**, 2174-2177 (2008).

---

[1] Mit freundlicher Genehmigung des Verlags Wiley-VCH

[2] http://www.angewandte.de

Angewandte
Chemie

DOI: 10.1002/anie.200704310

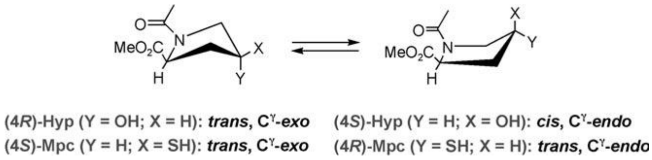
Proline Conformations

Conformational Properties of 4-Mercaptoproline and Related Derivatives[**]

Sergio A. Cadamuro, Rudolf Reichold, Ulrike Kusebauch, Hans-Jürgen Musiol, Christian Renner, Paul Tavan, and Luis Moroder[*]

Proline residues have long been recognized to play a unique and important role in the structural properties of peptides and proteins. The *cis*/*trans* isomerization of the aminoacyl–proline bonds is critically involved in folding and stabilizing protein structures.[1] This general notion has inspired an intensive search for proline analogues that influence the equilibrium conformational populations of the *cis*/*trans* prolyl bonds and of the pyrrolidine ring-pucker isomers (*endo*/*exo* of C4) in order to possibly restrict at will the degree of conformational freedom of polypeptide chains and thus to modulate the thermodynamic stability of peptide and protein structures.[2] Of the various modified prolines found in nature, the most common are (2*S*,4*R*)-hydroxyproline (Hyp) and (2*S*,3*S*)-Hyp, which are generated in post-translational processes exclusively in Y and X positions of the collagen (Xaa-Yaa-Gly) repeats, respectively, with the enzymatic 4*R* hydroxylation being by far the dominant modification.[3] The stereoelectronic effects of this electronegative substituent at C4 or C3 of the pyrrolidine ring have been the subject of comparative analysis, particularly in synthetic model compounds with the fluorine substituent, in terms of (de)stabilization of the collagen triple helix.[2a,h,i,k,4] These studies on collagen model peptides have been extended to other proteins by exploiting the strong effects of 4-fluoroprolines (Flp).[2g,5]

Rather surprisingly, the non-natural synthetic 3- and 4-mercaptopyrrolidine-2-carboxylic acids (Mpc),[6] chalcogen analogues of hydroxyprolines, have been used only sporadically for side chain/side chain cyclization of peptides through thioether or disulfide bridges in attempts to restrict the conformational space of peptidic macrocycles.[7] Herein we report a structural comparison of the (2*S*,4*R*)- and (2*S*,4*S*)-Mpc epimers, in which replacement of the hydroxy group with the less electronegative thiol group results in altered conformational preferences: The 4*R* epimer of Mpc induces a C^{γ}-*endo* pucker while (4*R*)-Hyp and similar 4*R* substitutions generate the *exo* pucker (Scheme 1). This may have interesting structural implications for the design of peptides and proteins, particularly when the thiol group is exploited for thioether or disulfide intra- and interchain cross-bridging of polypeptide chains.


O
MeO2C
N
H
X
Y
(4R)-Hyp (Y = OH; X = H): trans, Cγ-exo
(4S)-Mpc (Y = H; X = SH): trans, Cγ-exo
(4S)-Hyp (Y = H; X = OH): cis, Cγ-endo
(4R)-Mpc (Y = SH; X = H): trans, Cγ-endo

Scheme 1. Conformational equilibria of Ac-(4*R*/*S*)-Mpc-OMe (**1**) and Ac-(4*R*/*S*)-Hyp-OMe (**5**).

To evaluate the effect of 4-mercapto substitution on the pyrrolidine ring conformation the epimeric Ac-(2*S*,4*R*/*S*)-Mpc-OMe derivatives **1** were synthesized (Figure 1). In addition, to mimic the effect of side-chain bridging of such Mpc residues by thioethers or disulfides in peptides, the related alkyl and alkylthio epimer pairs Ac-(2*S*,4*R*/*S*)-Mpc(Me)-OMe (**2**) and Ac-(2*S*,4*R*/*S*)-Mpc(SMe)-OMe (**3**) were prepared, and their conformational preferences were compared with those of the known Ac-Pro-OMe (**4**), Ac-(2*S*,4*R*/*S*)-Hyp-OMe (**5**), and Ac-(2*S*,4*R*/*S*)-Flp-OMe (**6**) by NMR structural analysis in aqueous solution. In the model compounds the known effect of pH on the isomerization of Xaa–Pro bonds was prevented by the N and C derivatization.[8] Similarly, complications arising from hydrogen bonding in amide derivatives,[9] although weak in aqueous environments, are suppressed with C-terminal esters.

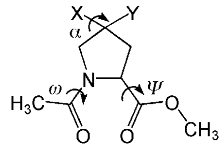


Figure 1. Chemical structure of *N*-acetylproline methyl ester (**4**, X = Y = H) and various derivatives, in which either X or Y is replaced by SH (**1**), SMe (**2**), SSMe (**3**), OH (**5**), or F (**6**).

trans/*cis* Equilibrium constants ($K_{t,c}$) and predominant ring puckerings were extracted from NMR spectral data to estimate the stereoelectronic effect of the thiol group on the conformational preference of the prolyl bond and on the ring puckering. These values are reported in Table 1 and compared to those known for the 4-hydroxy- and 4-fluoroproline epimer pairs. The equilibrium constants show that the 4*R*/*S* thiol group has a much weaker effect on the *trans*/*cis*

---

[*] Dr. S. A. Cadamuro, Dr. U. Kusebauch, H.-J. Musiol, Priv.-Doz. Dr. C. Renner,[+] Prof. Dr. L. Moroder
Max-Planck-Institut für Biochemie
Am Klopferspitz 18, 82152 Martinsried (Germany)
Fax: (+49) 89-8578-2847
E-mail: moroder@biochem.mpg.de
Homepage: http://www.biochem.mpg.de/en/rg/moroder/
R. Reichold, Prof. Dr. P. Tavan
Lehrstuhl für Biomolekulare Optik
Ludwig-Maximilians-Universität
Oettingenstrasse 67, 80538 München (Germany)
[+] Current address: Deutsche Forschungsgemeinschaft
Kennedyallee 40, 53170 Bonn/Bad-Godesberg (Germany)

[**] This work was supported by the Deutsche Forschungsgemeinschaft (SFB533, A8, and C3).

Supporting information for this article is available on the WWW under http://www.angewandte.org or from the author.

© 2008 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

Table 1: Thermodynamic parameters and conformational preferences of compounds **1**–**6**.

Cmpd	$K_{t/c}$[a]	Dominant conformation	$-\Delta H^{0}$[b] [kJ mol^{-1}]	ΔS^{0}[b] [J mol^{-1} K^{-1}]
1(4*R*)	5.4	*trans*, C$^\gamma$-*endo*	4.84 ± 0.8 %	−2.19 ± 5.9 %
1(4*S*)	4.7	*trans*, C$^\gamma$-*exo*	3.30 ± 1.0 %	1.89 ± 5.4 %
2(4*R*)	4.1	*trans*, C$^\gamma$-*endo*	1.93 ± 4.8 %	2.86 ± 2.8 %
2(4*S*)	3.1	*trans*, C$^\gamma$-*exo*	4.99 ± 2.5 %	−5.08 ± 5.0 %
3(4*R*)	4.3	*trans*, C$^\gamma$-*endo*	4.88 ± 2.6 %	−4.32 ± 4.3 %
3(4*S*)	3.6	*trans*, C$^\gamma$-*exo*	1.62 ± 4.6 %	5.25 ± 5.2 %
4[c]	4.8	*trans*, C$^\gamma$-*endo*	5.04 ± 1.0 %	−3.82 ± 4.3 %
5(4*R*)[d]	6.1	*trans*, C$^\gamma$-*exo*	7.84 ± 1.0 %	−10.7 ± 1.0 %
5(4*S*)[d]	2.4	*trans*, C$^\gamma$-*endo*	n.a.[e]	n.a.[e]
6(4*R*)[c]	7.3	*trans*, C$^\gamma$-*exo*	7.73 ± 3.3 %	−9.81 ± 8.3 %
6(4*S*)[c]	2.6	*trans*, C$^\gamma$-*endo*	3.04 ± 1.1 %	−2.47 ± 4.3 %

[a] Determined by integration of well-resolved signals in the ^{1}H NMR spectra in D_2O at 298 K. [b] The enthalpy (ΔH^{0}) and entropy (ΔS^{0}) contributions to the free energy difference between the *trans* and *cis* conformers were derived from van't Hoff plots (see Figure 1 in the Supporting Information); error limits were obtained from the residuals of the linear least-squares fitting. [c] Values from Ref. [2g]. [d] Values from Ref. [2k]. [e] Not available.

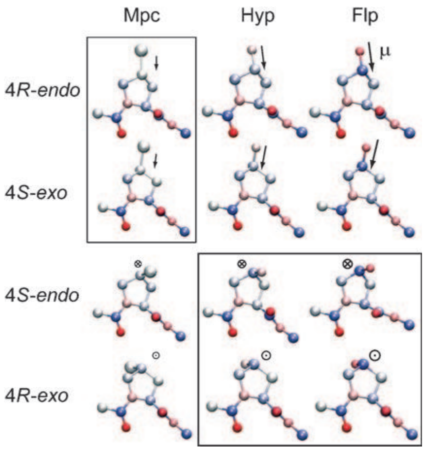


Figure 2. Optimized geometries of the L-proline derivatives as *trans* amide conformers: The strength and direction of the X–C$^\gamma$ bond dipoles μ are indicated by arrows. The atomic partial charges are indicated by a color code ranging from red (negative) through white (neutral) to blue (positive); adjacent red and blue atoms represent dipoles. The spheres represent "compound atoms"; the charges of hydrogen atoms are added to those of the neighboring heavy atoms. The experimentally observed conformers are marked by boxes.

conformational preference than the significantly more electronegative hydroxy and fluorine substituents. Indeed the $K_{t,c}$ values of **1**(4*R*) and **1**(4*S*) differ only slightly from that of unsubstituted proline. However, the ring-pucker preferences were reversed between the ring-substituted Ac-Pro-OMe species with similar *anti* (**1**(4*R*), **5**(4*R*), and **6**(4*R*)) and *syn* (**1**(4*S*), **5**(4*S*), and **6**(4*S*)) orientation of the electronegative substituent relative to the fixed L configuration of the C$^\alpha$ atom. Thus, an *anti* orientation of the 4-substituent resulted in a predominant C$^\gamma$-*endo* pucker for Mpc and in the known predominant C$^\gamma$-*exo* pucker for Hyp and Flp, whereas the opposite was true for the respective *syn*-oriented species.

The transformation of the thiol group into a methylthioether or methyldisulfide shifted the *trans*/*cis* equilibrium of the prolyl bond toward the *cis* conformation with the effect of the thioether being milder than that of the methyl disulfide. The ring-pucker preference was not affected by this type of derivatizations of the thiol group.

Quantum chemical calculations using density functional theory (DFT) were used to explain the experimental data obtained for compound **1** and **4**–**6**. In the case of Mpc (**1**) and Hyp (**5**) derivatives, potential curves were calculated for the dihedral angles ω, ψ, and α (Figure 1). The gas-phase energies and molecular geometries obtained by our DFT descriptions of **4**, **6**(4*R*), and **6**(4*S*) (see Tables II and III in the Supporting Information) agree very well with the results of previous DFT calculations,[2h] confirming our theoretical approach. The populations of the different conformers derived from NMR experimental data and those computed from DFT results are congruent with few exceptions (see Table IV in the Supporting Information). Most importantly, the calculations predict the initially unexpected but experimentally observed preference of **1**(4*R*) for the C$^\gamma$-*endo* pucker and of **1**(4*S*) for the C$^\gamma$-*exo* pucker.

The DFT calculations reveal that the pucker preference of a given substitution at the 4-position is mainly determined by the interaction between the bond dipole at the substitution site and the dipole of the preceding amide bond. Figure 2 shows how the combination of (4*R*)-Hyp or (4*R*)-Flp with a C$^\gamma$-*endo* pucker leads to an unfavorable antiparallel orientation of these dipoles, while the C$^\gamma$-*exo* pucker results in close to perpendicular dipole moments with an almost neutral contribution to the total energy. This results in the C$^\gamma$-*exo* pucker being more favorable for the (4*R*)-Hyp and the (4*R*)-Flp derivatives. In contrast, the dipole moment at the substitution site of Mpc is very weak and causes almost no energetic penalty for the 4*R*-*endo* combination. The experimentally observed preference of Mpc for the 4*R*-*endo* geometric variant implies that the net contributions of all interactions (electrostatic, van der Waals, bond geometries, etc.) except the decisive dipole–dipole interaction discussed before are slightly in favor of the C$^\gamma$-*endo* pucker for the (4*R*)-Mpc derivative **1**. Therefore, the preference of the (4*R*)-Hyp (**5**) and (4*R*)-Flp (**6**) derivatives for the C$^\gamma$-*exo* pucker is explained by a reduced unfavorable dipole–dipole orientation rather than a specific favorable interaction. One would expect that other 4*R* or 4*S* substitutions that introduce small dipole moments will also prefer the 4*R*-*endo* and 4*S*-*exo* combinations that had been hitherto considered unfavorable and unusual, unless steric effects prevail over electrostatic interaction as it is the case in the pair of 4-methylproline epimers.[2n,9]

In our preceding synthetic efforts to control the folding/unfolding of a collagen triple helix by applying light we introduced two (4*S*)-Mpc residues, as the synthetically more readily accessible epimer, into the Ac-(Gly-Pro-Hyp)$_7$-Gly-Gly-NH$_2$ model collagen peptide for side chain/side chain cross-bridging of the two thiol groups with a suitable thiol-reactive azobenzene derivative (Figure 3).[11] Optimal locations for the Mpc residues according to modeling experiments are the i and $i+7$ positions corresponding to Xaa and Yaa residues of the classical (Xaa-Yaa-Gly) collagen triplets. High-resolution X-ray analysis of collagen model peptides[12]

© 2008 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

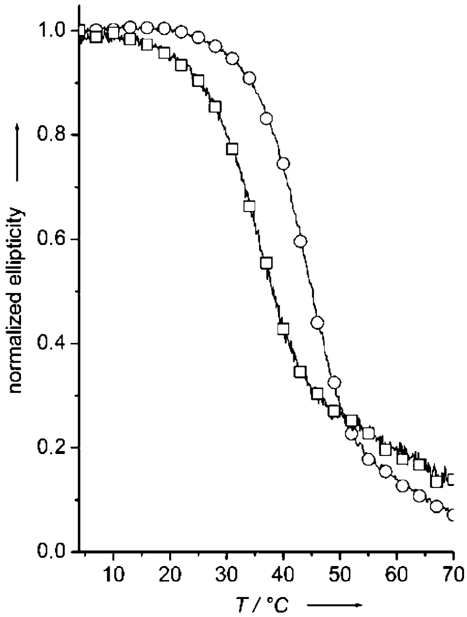


Figure 3. Thermal unfolding of the triple-helical Ac-(Gly-Pro-Hyp)$_7$-Gly-Gly-NH$_2$ (○) ($T_m = 43\,°C$) and of its analogue Ac-(GPO)$_2$-(G-(4*S*)-Mpc-O)-(GPO)-(GP-(4*S*)-MPC)-(GPO)$_2$-GG-NH$_2$ with (4*S*)-Mpc residues located in the X and Y positions (□) ($T_m = 34.5\,°C$). The unfolding was monitored by CD at 225 nm in aqueous solution at a peptide concentration of 1 mm; O = (4*R*)-hydroxyproline.

and computational analysis[13] revealed alternate C$^\gamma$-*endo* and C$^\gamma$-*exo* puckers of the Pro and Hyp residues in the X and Y positions of the triplets as a repetitive motif possibly involved in stabilizing the triple helix. Extensive comparative studies of synthetic collagen peptides containing (4*R*)-Hyp or (4*S*)-Hyp, and (4*R*)- or (4*S*)-Flp have shown that occupancy of the X and Y positions with proline analogues characterized by preferences for C$^\gamma$-*endo* and C$^\gamma$-*exo* puckers, respectively, leads to markedly increased thermal stabilities of the triple helix, while the opposite effect is induced with reversed ring puckerings.[2h,i,k,4a,b,12] However, more recently, in X-ray structures of [Gly-(4*R*)-Hyp-(4*R*)-Hyp]$_n$ peptides both Hyp residues assume *exo* conformations.[4g,j] In view of the still-evolving understanding of the mechanism of triple-helix stabilization and the weak electron-withdrawing property of the thiol group we assumed that replacement of a Pro and Hyp residue in Ac-(Gly-Pro-Hyp)$_7$-Gly-Gly-NH$_2$ with (4*S*)-Mpc should affect only marginally the triple-helix stability. In contrast, a rather substantial decrease in the thermal stability was observed as shown in Figure 3.

Taking into account that Pro and/or Hyp replacements in single triplets of (Gly-Pro-Hyp)$_n$ peptides, that is, in host–guest peptides, lead to results that are significantly different than those from repetitive replacements,[4k] a rational interpretation of the drop of the T_m value by about 8 °C can be put forward. The stereoelectronic effects of substituted prolines in the (Gly-Pro-Hyp) collagen repeats are not additive in terms of triple-helix (de)stabilization,[4i] but the steric effects seem to be so.[2n] By the single substitution of a Hyp residue in Y with (4*S*)-Mpc the *trans* peptide bond is less favored and similar to that of a Pro residue (Table 1). This negative effect, which leads to a 2 °C lower T_m in a (Gly-Pro-Hyp)$_8$ host peptide containing one Gly-Pro-Pro repeat,[4k] should be fully compensated by the favored C$^\gamma$-*exo* pucker of the (4*S*)-Mpc. Similarly, like a Pro residue in X position combined with Hyp in Y, (4*S*)-Mpc should marginally affect the triple-helical fold despite its favored C$^\gamma$-*exo* pucker. The rather strong experimentally observed destabilization must therefore be assigned mainly to steric effects, fully supporting a strong interplay between stereoelectronic and steric effects in the assembly of collagen triple helices.

Despite the limitations of the predictive power of the simple Ac-Pro-OMe system for values in protein environments, the results of this study offer a more general view of the relation between 4-substitutions of proline and resulting conformational properties of the amide bond and, especially, of the proline ring pucker. The increased understanding of the determinants of proline geometry together with the decisive role of proline residues and related analogues in peptide and protein structures can provide a powerful tool in the design and folding studies of polypeptides. In contrast, their application in proteins must await improved methodologies for an efficient incorporation of such non-natural amino acids into expressed proteins unless synthetic and semisynthetic ligation strategies suffice for the purpose.

Experimental Section

Details of the synthesis of Ac-(4*R*/*S*)-Mpc-OMe (**1**), Ac-(4*R*/*S*)-Mpc(Me)-OMe (**2**), and Ac-(4*R*/*S*)-Mpc(SMe)-OMe (**3**) are reported in Supporting Information. Solutions in D$_2$O were used for NMR measurements, and in the case of Ac-Mpc-OMe tris-(2-carboxyethyl)-phosphine hydrochloride (1 equiv) was added to prevent oxidation. NMR experiments were performed with a Bruker DRX500 spectrometer using a triple-resonance (^{15}N/^{13}C/^{1}H) inverse probe. Assignment of ^{1}H and ^{13}C NMR resonances was based on homonuclear 2D ^{1}H-^{1}H NOESY and TOCSY experiments and heteronuclear 2D ^{13}C-^{1}H COSY experiments.[15]

Thermodynamics and kinetics of amide-bond isomerization: Equilibrium constants ($K_{t/c}$) for the *trans*/*cis* conformer ratios at various temperatures were determined by integration of the signals of the α and γ protons in 1D ^{1}H NMR spectra. The enthalpic and entropic contributions to the free energy difference between the *cis* and *trans* conformers were obtained from van't Hoff plots according to $\ln(K_{ZE}) = (-\Delta H^0/R)(1/T) + \Delta S^0/R$.

NMR conformation analysis: The pucker of the proline ring was identified by the method of Gerig and McLeod[16] by means of the distinct pattern of the ^{1}H–^{1}H coupling constants observed in 1D ^{1}H NMR spectra; for example, the C$^\gamma$-*exo* pucker results in large and similar coupling constants $J(H_\alpha,H_{\beta1})$ and $J(H_\alpha,H_{\beta2})$, whereas for the C$^\gamma$-*endo* pucker one of the coupling constants is large and the other small. The *cis*-to-*trans* isomerization of the amide bond is a slow process on the NMR time scale, and therefore two distinct signals are observed for the two conformers. On the other hand, the pucker inversion is a fast process, and therefore the signal obtained for the *trans* conformer is an averaging of the *trans*, C$^\gamma$-*endo* and *trans*, C$^\gamma$-*exo* pucker; the same is true for the *cis* conformer. An estimation of the ratio of the two puckerings for the *trans* and *cis* conformers was obtained by the equation $\Delta J_{exp} = x\Delta J_{endo} + y\Delta J_{exo}$ where ΔJ refers to the difference between $J(H_\alpha,H_{\beta1})$ and $J(H_\alpha,H_{\beta2})$, while ΔJ_{endo} and ΔJ_{exo} were calculated with the program MestRe-J[17] using the dihedral angles obtained from the lowest energy conformers produced by quantum chemical calculations (see Table I in the Supporting Information).

Computational methods: Density functional theory (DFT) calculations were carried out with the program TURBOMOLE 5.6.[18] Here, the B3LYP functional[19] and the TZVP basis set[20] (a Gaussian basis set of triple-zeta valence quality augmented by polarization functions) were employed. Various isomeric structures of compound **4** as well as of its mercapto- (**1**), hydroxy- (**5**), and fluoro dervatives (**6**) were calculated both for the gas phase and with the continuum solvent model COSMO[21] (dielectric constant $\varepsilon = 80$) to account for the dielectric shielding by the D_2O solvent used in the NMR experiments. Mulliken population analyses[22] were performed to determine partial atomic charges. For further details see the Supporting Information.

Received: September 18, 2007
Published online: January 28, 2008

Keywords: conformation analysis · peptides · proline · pyrrolidine · stereoelectronic effects

---

[1] a) J. F. Brandts, H. R. Halvorson, M. Brennan, *Biochemistry* **1975**, *14*, 4953–4963; b) W. J. Wedemeyer, E. Welker, H. A. Scheraga, *Biochemistry* **2002**, *41*, 14637–14644; c) G. Fischer, *Chem. Soc. Rev.* **2000**, *29*, 119–127; d) F. X. Schmid in *Protein Folding Handbook*, *Vol. 2* (Eds: J. Buchner, T. Kiefhaber), Wiley-VCH, Weinheim, **2005**, pp. 916–945.

[2] a) E. S. Eberhardt, N. Panasik, R. T. Raines, *J. Am. Chem. Soc.* **1996**, *118*, 12261–12266; b) N. Panasik, E. S. Eberhardt, A. S. Edison, D. R. Powell, R. T. Raines, *Int. J. Pept. Protein Res.* **1994**, *44*, 262–269; c) D. Kern, M. Schutkowski, T. Drakenberg, *J. Am. Chem. Soc.* **1997**, *119*, 8403–8408; d) M. Keller, C. Sager, P. Dumy, M. Schutkowski, G. S. Fischer, M. Mutter, *J. Am. Chem. Soc.* **1998**, *120*, 2714–2720; e) E. Beausoleil, R. Sharma, S. W. Michnick, W. D. Lubell, *J. Org. Chem.* **1998**, *63*, 6572–6578; f) S. S. A. An, C. C. Lester, J. L. Peng, Y. J. Li, D. M. Rothwarf, E. Welker, T. W. Thannhauser, L. S. Zhang, J. P. Tam, H. A. Scheraga, *J. Am. Chem. Soc.* **1999**, *121*, 11558–11566; g) C. Renner, S. Alefelder, J. H. Bae, N. Budisa, R. Huber, L. Moroder, *Angew. Chem.* **2001**, *113*, 949–951; *Angew. Chem. Int. Ed.* **2001**, *40*, 923–925; h) M. L. DeRider, S. J. Wilkens, M. J. Waddell, L. E. Bretscher, F. Weinhold, R. T. Raines, J. L. Markley, *J. Am. Chem. Soc.* **2002**, *124*, 2497–2505; i) J. A. Hodges, R. T. Raines, *J. Am. Chem. Soc.* **2003**, *125*, 9262–9263; j) Y. Che, G. R. Marshall, *J. Org. Chem.* **2004**, *69*, 9030–9042; k) L. E. Bretscher, C. L. Jenkins, K. M. Taylor, M. L. DeRider, R. T. Raines, *J. Am. Chem. Soc.* **2001**, *123*, 777–778; l) C. M. Taylor, R. Hardre, P. J. B. Edwards, *J. Org. Chem.* **2005**, *70*, 1306–1315; m) W. Kim, K. I. Hardcastle, V. P. Conticello, *Angew. Chem.* **2006**, *118*, 8321–8325; *Angew. Chem. Int. Ed.* **2006**, *45*, 8141–8145; n) M. D. Shoulders, J. A. Hodges, R. T. Raines, *J. Am. Chem. Soc.* **2006**, *128*, 8112–8113; o) S. K. M. Umashankara, I. R. Babu, K. N. Ganesh, *Chem. Commun.* **2003**, 2606–2607; p) I. R. Babu, K. N. Ganesh, *J. Am. Chem. Soc.* **2001**, *123*, 2079–2080.

[3] a) I. Wagner, H. Musso, *Angew. Chem.* **1983**, *95*, 827–839; *Angew. Chem. Int. Ed. Engl.* **1983**, *22*, 816–828; b) A. B. Mauger, *J. Nat. Prod.* **1996**, *59*, 1205–1211; c) L. Moroder, C. Renner, J. J. Lopez, M. T. Mutter, G. Tuchscherer in *Cis-trans Isomerization in Biochemistry* (Ed.: C. Dugave), Wiley-VCH, Weinheim, **2006**, pp. 225–259.

[4] a) S. K. Holmgren, K. M. Taylor, L. E. Bretscher, R. T. Raines, *Nature* **1998**, *392*, 666–667; b) S. K. Holmgren, L. E. Bretscher, K. M. Taylor, R. T. Raines, *Chem. Biol.* **1999**, *6*, 63–70; c) C. L. Jenkins, L. E. Bretscher, I. A. Guzei, R. T. Raines, *J. Am. Chem. Soc.* **2003**, *125*, 6422–6427; d) J. A. Hodges, R. T. Raines, *J. Am. Chem. Soc.* **2005**, *127*, 15923–15932; e) M. Doi, Y. Nishi, S. Uchiyama, Y. Nishiuchi, H. Nishio, T. Nakazawa, T. Ohkubo, Y. Kobayashi, *J. Pept. Sci.* **2005**, *11*, 609–616; f) M. Doi, Y. Nishi, S. Uchiyama, Y. Nishiuchi, T. Nakazawa, T. Ohkubo, Y. Kobayashi, *J. Am. Chem. Soc.* **2003**, *125*, 9922–9923; g) K. Kawahara, Y. Nishi, S. Nakamura, S. Uchiyama, Y. Nishiuchi, T. Nakazawa, T. Ohkubo, Y. Kobayashi, *Biochemistry* **2005**, *44*, 15812–15822; h) Y. Nishi, S. Uchiyama, M. Doi, Y. Nishiuchi, T. Nakazawa, Z. Ohkubo, Y. Kobayashi, *Biochemistry* **2005**, *44*, 6034–6042; i) D. Barth, A. G. Milbradt, C. Renner, L. Moroder, *ChemBioChem* **2004**, *5*, 79–86; j) M. Schumacher, K. Mizuno, H. P. Bächinger, *J. Biol. Chem.* **2005**, *280*, 20397–20403; k) A. V. Persikov, J. A. M. Ramshaw, A. Kirkpatrick, B. Brodsky, *J. Am. Chem. Soc.* **2003**, *125*, 11500–11501.

[5] a) W. Kim, R. A. McMillan, J. P. Snyder, V. P. Conticello, *J. Am. Chem. Soc.* **2005**, *127*, 18121–18132; b) R. Golbik, C. Yu, E. Weyher-Stingl, R. Huber, L. Moroder, N. Budisa, C. Schiene-Fischer, *Biochemistry* **2005**, *44*, 16026–16034; c) S. C. R. Lummis, D. L. Beene, L. W. Lee, H. A. Lester, R. W. Broadhurst, D. A. Dougherty, *Nature* **2005**, *438*, 248–252.

[6] a) V. Eswarakrishnan, L. Field, *J. Org. Chem.* **1981**, *46*, 4182–4187; b) A. J. Verbiscar, B. Witkop, *J. Org. Chem.* **1970**, *35*, 1924–1927.

[7] a) D. S. Kemp, J. H. Rothman, T. C. Curran, D. E. Blanchard, *Tetrahedron Lett.* **1995**, *36*, 3809–3812; b) G. V. Nikiforovich, J. L. F. Kao, K. Plucinska, W. J. Zhang, G. R. Marshall, *Biochemistry* **1994**, *33*, 3591–3598; c) K. Plucinska, T. Kataoka, M. Yodo, W. L. Cody, J. X. He, C. Humblet, G. H. Lu, E. Lunney, T. C. Major, R. L. Panek, P. Schelkun, R. Skeean, G. R. Marshall, *J. Med. Chem.* **1993**, *36*, 1902–1913; d) T. Kataoka, D. D. Beusen, J. D. Clark, M. Yodo, G. R. Marshall, *Biopolymers* **1992**, *32*, 1519–1533.

[8] C. Grathwohl, K. Wüthrich, *Biopolymers* **1976**, *15*, 2025–2041.

[9] G. B. Liang, C. J. Rito, S. H. Gellman, *J. Am. Chem. Soc.* **1992**, *114*, 4440–4442.

[10] J. L. Flippen-Anderson, R. Gilardi, I. L. Karle, M. H. Frey, S. J. Opella, L. M. Gierasch, M. Goodman, V. Madison, N. G. Delaney, *J. Am. Chem. Soc.* **1983**, *105*, 6609–6614.

[11] U. Kusebauch, S. A. Cadamuro, H. J. Musiol, M. O. Lenz, J. Wachtveitl, L. Moroder, C. Renner, *Angew. Chem.* **2006**, *118*, 7170–7173; *Angew. Chem. Int. Ed.* **2006**, *45*, 7015–7018.

[12] a) V. Nagarajan, S. Kamitori, K. Okuyama, *J. Biochem.* **1999**, *125*, 310–318; b) R. Z. Kramer, L. Vitagliano, J. Bella, R. Berisio, L. Mazzarella, B. Brodsky, A. Zagari, H. M. Berman, *J. Mol. Biol.* **1998**, *280*, 623–638; c) R. Berisio, L. Vitagliano, L. Mazzarella, A. Zagari, *Protein Sci.* **2002**, *11*, 262–270; d) L. Vitagliano, R. Berisio, A. Mastrangelo, L. Mazzarella, A. Zagari, *Protein Sci.* **2001**, *10*, 2627–2632; e) L. Vitagliano, R. Berisio, L. Mazzarella, A. Zagari, *Biopolymers* **2001**, *58*, 459–464.

[13] R. Improta, C. Benzi, V. Barone, *J. Am. Chem. Soc.* **2001**, *123*, 12568–12577.

[14] R. T. Raines, *Protein Sci.* **2006**, *15*, 1219–1225.

[15] K. Wütrich, *NMR of Proteins and Nucleic Acids*, Wiley, New York, **1986**.

[16] J. T. Gerig, R. S. McLeod, *J. Am. Chem. Soc.* **1973**, *95*, 5725–5729.

[17] A. Navarro-Vazquez, J. C. Cobas, F. J. Sardina, J. Casanueva, E. Diez, *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1680–1685.

[18] a) R. Ahlrichs, M. Bär, M. Häser, H. Horn, C. Kölmel, *Chem. Phys. Lett.* **1989**, *162*, 165–169; b) O. Treutler, R. Ahlrichs, *J. Chem. Phys.* **1995**, *102*, 346–354.

[19] a) A. D. Becke, *J. Chem. Phys.* **1993**, *98*, 5648–5652; b) C. T. Lee, W. T. Yang, R. G. Parr, *Phys. Rev. B* **1988**, *37*, 785–789.

[20] A. Schäfer, C. Huber, R. Ahlrichs, *J. Chem. Phys.* **1994**, *100*, 5829–5835.

[21] A. Klamt, G. Schüürmann, *J. Chem. Soc. Perkin Trans. 2* **1993**, 799–805.

[22] R. S. Mulliken, *J. Chem. Phys.* **1955**, *23*, 1833–1840.

© 2008 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

Conformational Preferences of 4-Mercaptoproline and Related Derivatives

Sergio A. Cadamuro,[a] Rudolf Reichold,[b] Ulrike Kusebauch,[a] Hans-Jürgen Musiol,[a] Christian Renner,[a] Paul Tavan,[b] Luis Moroder[a]*

[a]Max-Planck-Institut für Biochemie, Am Klopferspitz 18, D-82152 Martinsried; [b]Lehrstuhl für Biomolekulare Optik, Ludwig-Maximilians-Universität, Oettingenstr. 67, D-80538 München; Germany

Materials and Methods

All solvents and reagents used were of highest quality commercially available. Fmoc-(2*S*,4*R*)-Mpc(Trt)-OH and Fmoc-(2*S*,4*S*)-Mpc(Trt)-OH were purchased from NeoMPS (Strassbourg, France) and the chemicals from Roth (Karlsruhe, Germany), Fluka (Buchs, CH), Sigma-Aldrich (Taufkirchen, Germany), Bachem (Bubendorf, CH) and Merck (Darmstadt, Germany). High resolution mass spectra were performed on a micro-TOF-LC-MS from Bruker Daltonics (Bremen, Germany) in combination with a series 1100 HPLC system from Agilent (Palo Alto, USA) including the Compass 1.1 software (HR-LC-MS). To analyze the samples a C18 Hypersil Gold column 100×2.1 mm, 5 µm from Thermo Electron Corporation (Waltham, USA) was used with a flow rate of 250 µl min^{-1} and a gradient from 10 % eluent B (0.05 % TFA in water) to 90 % eluent B (0.05 % TFA in MeCN) in 17 min. NMR spectra were recorded on a DRX 500 NMR spectrometer from Bruker Biospin (Rheinstetten, Germany); D_2O was purchased from Eurisotop (Saint-Aubin, France).

Synthesis

Ac-(2*S*,4*S*)-Mpc-OMe, 1(*S*): Fmoc-(2*S*,4*S*)-Mpc(Trt)-OH (1 eq) was esterified with trimethylsilyl diazomethane (2 M in hexane, 10 eq) in cooled benzene/methanol (5/1). After 90 min at rt, AcOH was added until the reaction changed to a colorless solution. The product was precipitated with MTB/hexane. The Fmoc-group was cleaved with 20 % diethylamine in DMF for 1h, the solvent was evaporated and the resulting H-(2*S*,4*S*)-Mpc(Trt)-OMe was purified by silica gel chromatography ($CHCl_3$/2-propanol/AcOH, 85/10/5). Acetylation was performed with Ac_2O (4 eq) and *N*-ethyldiisopropylamine (5 eq) in EtOAc for 3 h. The solution was concentrated to small volume and diluted with MTB, washed with water, dried over Na_2SO_4 and taken to dryness. The S-trityl group was removed TFA/H_2O/triisopropylsilane (96/2/2) for 1 h and the product was purified by silica gel chromatography ($CHCl_3$/cyclohexane/AcOH, 80/15/5); yield: 40 mg (47 %); colorless oil; TLC: R_f 0.2 (cyclohexane/$CHCl_3$/AcOH, 15/80/5); HR-LC-MS: t_R = 5.8 min; *m/z* = 204.06485 $[M+H]^+$; calc. for $C_8H_{13}NO_3S + H^+$: 204.0694; ^{1}H-NMR (D_2O, 300 K, 500 MHz): δ = 1.98 (s, 3H, $COCH_3$); 2.09 - 2.19 (m, 1H, $NCHCH_2$); 2.27 - 2.34 (m, 1H, $NCHCH_2$); 3.36 - 3.43 (m, 1H, NCH_2); 3.47 - 3.53 (m, 1H, CHSH); 3.62 (s, 3H, OCH_3); 3.93 (dd, 1H, *J* = 6.42 Hz, 4.40 Hz, NCH_2); 4.47 (dd, 1H, *J* = 5.14 Hz, 3.67 Hz, NCH), ^{13}C-NMR (D_2O, 300 K, 125 MHz): δ = 21.20 (CH_3CO); 34.61 (CHSH); 38.92 ($NCHCH_2$); 52.99 (CH_3O); 56.60 (NCH_2); 58.73 (NCH); 172.72 ($COCH_3$); 174.19 ($COOCH_3$).

Ac-(2*S*,4*R*)-Mpc-OMe, 1(*R*): the compound was synthesized as described for Ac-(2*S*,4*S*)-Mpc-OMe; yield: 70 mg (83 %); colorless oil; TLC: R_f = 0.3 (cyclohexane/$CHCl_3$/AcOH, 15/80/5); HR-LC-MS: t_R = 5.7 min; m/z = 204.06775 $[M+H]^+$; calc. for $C_8H_{13}NO_3S + H^+$:204.0694; ^{1}H-NMR (D_2O, 300 K, 500 MHz) δ = 2.07 (s, 3H, $COCH_3$); 2.20 - 2.28 (m, 1H, $NCHCH_2$); 2.37 - 2.44 (m, 1H, $NCHCH_2$); 3.49 (dd, 1H, *J* = 6.41 Hz, 4.27 Hz, NCH_2); 3.56 - 3.63 (m, 1H, CHSH); 3.71 (s, 3H, OCH_3); 4.02 (dd, 1H, *J* = 6.71 Hz, 4.27 Hz, NCH_2); 4.56 (dd, 1H, *J* = 4.88 Hz, 3.97 Hz, NCH), ^{13}C-NMR (D_2O, 300 K, 125 MHz): δ = 21.15 (CH_3CO); 34.84 (CHSH); 38.55 ($NCHCH_2$); 53.00 (CH_3O); 56.57 (NCH_2); 58.32 (NCH); 172.90 ($COCH_3$); 174.13 ($COOCH_3$).

Ac-(2*S*,4*S*)-Mpc(Me)-OMe, 2(*S*): methylation of the thiol group of Ac-(2*S*,4*S*)-Mpc-OMe was performed with CH_3I (2.4 eq) and a tip of a spatula of K_2CO_3 in dry acetone over night under argon atmosphere. The mixture was filtered off to remove residual K_2CO_3 and evaporated. The product was purified by silica gel chromatography ($CHCl_3$/2-propanol/AcOH, 85/10/5); yield: 20 mg (56 %); brown oil; TLC: R_f 0.8 ($CHCl_3$/2-propanol/AcOH, 85/10/15); HR-LC-MS: t_R = 7.8 min; *m/z*: = 218.08539 $[M+H]^+$; calc. for $C_9H_{15}NO_3S + H^+$: 218.0851; ^{1}H-NMR (D_2O, 300 K, 500 MHz): δ = 2.07-2.13 (m, 4H, Hβa and $COCH_3$); 2.15 (s, 3H, SCH_3); 2.65-2.73 (m, 1H, Hβb); 3.47-3.58 (m, 2H, Hγ and Hδa); 3.76 (s, 3H, OCH_3); 4.02-4.11 (m, 1H, Hδb); 4.54 (dd, J = 8.24 and 6.41, Hα), 13C-NMR δ = 13.288 (SCH_3); 21.225 ($COCH_3$); 34.727 (Cβ); 42.263 (Cγ); 53.004 (OCH_3); 53.618 (Cγ); 58.477 (Cα); 173.000 (COO); 174.217 (CON).

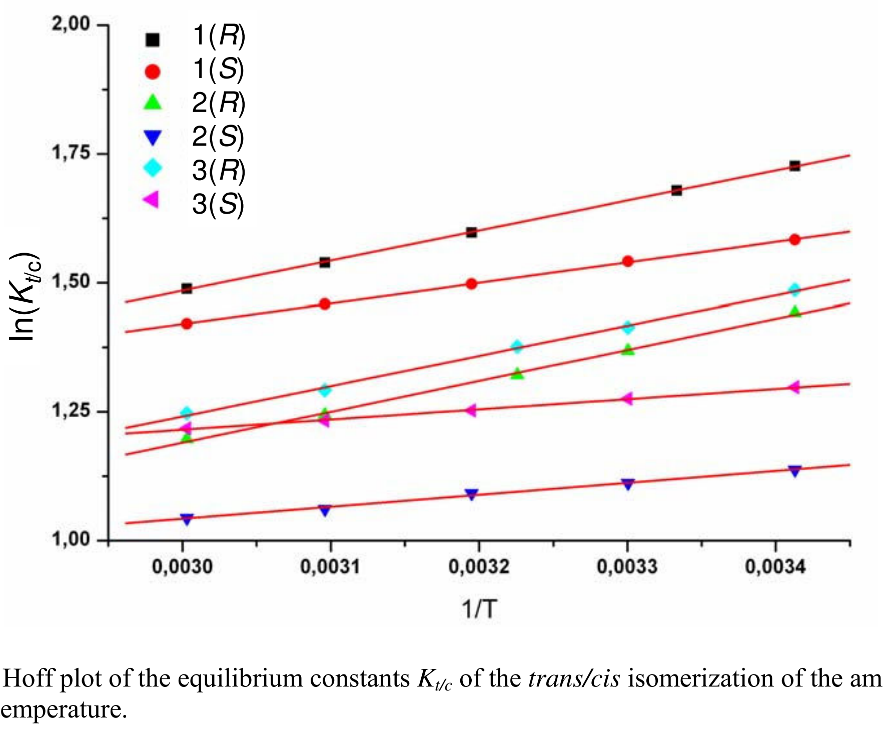
Ac-(2*S*,4*R*)-Mpc(Me)-OMe, 2(*R*): the compound was synthesized as described for Ac-(2*S*,4*R*)-Mpc(Me)-OMe and purified by silica gel chromatography; yield: 10 mg (38 %); colorless oil; TLC: R_f 0.7 (cyclohexane/$CHCl_3$/2-propanol/AcOH, 10/75/13/2); HR-LC-MS: t_R = 8.1 min; *m/z* = 218.08482 $[M+H]^+$; calc. for $C_9H_{15}NO_3S + H^+$: 218.0851; ^{1}H-NMR (D_2O, 300 K, 500 MHz): δ = 2.07-2.13 (m, 4H, Hβa and $COCH_3$); 2.15 (s, 3H, SCH_3); 2.65-2.73 (m, 1H, Hβb); 3.47-3.58 (m, 2H, Hγ and Hδa); 3.76 (s, 3H, OCH_3); 4.02-4.11 (m, 1H, Hδb); 4.54 (dd, J = 8.24 and 6.41, Hα). 13C-NMR δ = 13.288 (SCH_3); 21.225 ($COCH_3$); 34.727 (Cβ); 42.263 (Cγ); 53.004 (OCH_3); 53.618 (Cδ); 58.477 (Cα); 173.000 (COO); 174.217 (CON).

Ac-(2*S*,4*S*)-Mpc(SMe)-OMe, 3(*S*): Ac-(2S,4S)-Mpc-OMe and methanethiosulfonic acid-S-methyl ester (1.7 eq) were stirred in absolute EtOH (1 ml) over night. After evaporation of the solvent, the product was dissolved in EtOAc, washed with $NaHCO_3$, with water and brine. The solution was taken to dryness and the product was purified by silica gel chromatography ($CHCl_3$/cyclohexane/AcOH, 78/20/2); yield: 10 mg (33 %); colorless oil; TLC: R_f 0.2 ($CHCl_3$/cyclohexane/AcOH, 78/20/2); HR-LC-MS: t_R = 9.8 min; *m/z* = 250.05667 $[M+H]^+$; calc. for $C_9H_{15}NO_3S_2 + H^+$: 250.0571; ^{1}H-NMR (D_2O, 300 K, 500 MHz): δ = 2.15 (s, 3H, $COCH_3$); 2.23-2.33 (m, 1H, Hαa) 2.47 (s, 3H, SCH_3); 2.67-2.78 (m, 1H, Hβb); 3.65-3.83 (m, 5H, Hγ, Hδa, OCH_3); 4.05-4.13 (m, 1H, Hδb); 4.57 (dd, J = 6.10 and 8.54, 1H, Hα). 13C-NMR δ = 21.323 ($COCH_3$); 23.269 (SCH_3); 34.283 (Cβ); 45.711 (Cγ); 52.979 (OCH_3); 53.034 (Cδ); 58.362 (Cα); 172.873 (COO); 173.888 (CON).

Ac-(2*S*,4*R*)-Mpc(SMe)-OMe, 3(*R*): the compound was prepared as described for Ac-(2S,4S)-Mpc(SMe)-OMe. Since the product was obtained in pure quality, purification by silica gel chromatography was not required; yield: 35 mg (83 %); colorless oil; TLC: R_f 0.2 ($CHCl_3$/cyclohexane/AcOH, 78/20/2); HR-LC-MS: t_R = 10.1 min; *m/z* = 250.05727 $[M+H]^+$; calc. for $C_9H_{15}NO_3S_2 + H^+$: 250.0571; ^{1}H-NMR (D_2O, 300 K, 500 MHz): δ = 2.14 (s, 3H, $COCH_3$); 2.35-2.43 (m, 1H, Hβa); 2.46 (s, 3H, SCH_3); 2.48-2.545 (m, 1H, Hβb); 3.73-3.85 (m, 5H, Hγ, Hδa and OCH_3); 4.01-4.10 (m, 1H, Hδb); 4.59 (dd, J = 5.80 and 8.54, 1H, Hα). ^{13}C-NMR δ = 21.219 ($COCH_3$); 23.318 (SCH_3); 34.131 (Cβ); 45.498 (Cγ); 52.833 (OCH_3); 53.064 (Cδ); 58.264 (Cα); 173.049 (COO); 174.150 (CON).

NMR Measurements

All NMR experiments were performed on a Bruker DRX500 spectrometer using a triple resonance $^{15}N/^{13}C/^1H$ inverse probe. ^{1}H and ^{13}C resonances were assigned by homonuclear 2D ^1H-^1H NOESY and TOCSY experiments, and 2D ^{13}C-^1H COSY experiments.[1] For assignment of the *trans* and *cis* prolyl bonds of compounds 1-3 at various temperatures (Suppl. Figure 1) the α and γ protons signals in 1D-^{1}H NMR spectra were integrated. The pucker of the proline ring was identified by the distinct pattern of the ^1H-^1H coupling constants observed in ^{1}H 1D NMR spectra according to Gerig and McLeod.[2]



Suppl. Figure 1. Van't Hoff plot of the equilibrium constants $K_{t/c}$ of the *trans/cis* isomerization of the amide bond of compounds 5-7 versus temperature.

Quantum Chemical Calculations

For the computations initial structures of 1(*R*) and 1(*S*) were modeled with the program MOLDEN[3] and energetically optimized by DFT leading to *trans* conformations with an *endo* ring pucker in the case of 1(*R*) and an *exo* ring pucker in the case of 1(*S*). Starting structures for hydroxyproline, fluoroproline and proline were generated by replacing the SH group of the mercaptoprolines by an OH group, a fluorine or a hydrogen atom. Changing the positions of the substituents (*4R* → *4S*, *4S* → *4R*) in the optimized structures or replacing them by a hydrogen atom resulted in structures with both ring puckers for each of the seven molecules that were used as starting points for further energy minimizations. To create the corresponding *cis* prolyl bonds the N-terminal COMe groups were rotated by 180° around the N–C bond.

For comparison with the NMR data estimates on the room temperature conformational ensembles had to be determined. For this purpose the orientation of the C-terminal CO_2Me with respect to the remainder of the molecules, i.e. the dihedral angle ψ (Figure 1), was varied by 360° in steps of 10° and corresponding potential curves were calculated by DFT for various structures of proline and of its derivatives. Because these calculations generally revealed that an extended (PPII) structure with $\psi \approx 150°$ is energetically preferred over a helical configuration ($\psi \approx -30°$), the configurational ensemble of the molecules was restricted to this extended configuration. Similar potential curves were also recorded for the dihedral angle α at fixed orientations of the SH and OH substituents. These curves indicated three angular ranges within which local minima are found. Dihedral angles taken from these ranges were subsequently used as starting values for energy minimizations of the respective conformers. The resulting optimal angles α_1, α_2, and α_3 distinguish three energetically close conformers of the SH and OH derivatives, which we denote by the symbols g^+, a, and g^- (for gauche$^+$, anti, and gauche$^-$).[4, 5] The barriers separating these conformers were found to measure only a few kcal/mol, such that they can be rapidly crossed at room temperature upon thermal excitation and cannot be distinguished by NMR. Further conformers of the proline derivatives are obtained by changing the ring pucker from *exo* to *endo*. For this change no potential curves were calculated. The variation of the angle ω (Figure 1) from its *trans* to its *cis* value revealed large barriers in the range of 20 kcal/mol, which cannot be thermally crossed on the NMR time scale. As a result we obtained for all proline derivatives four conformations possibly distinguishable by NMR. Denoting the energies of the various configurations by $E_{i,j}$ in the case of Ac-Pro-OMe (4) or Ac-(4*R/S*)-Flp-OMe (6) and by $E^k_{i,j}$ in the case of Ac-(4*R/S*)-Mpc-OMe (1) or Ac-(4*R/S*)-Hyp-OMe (5) ($i \in \{trans, cis\}, j \in \{endo, exo\}, k \in \{g^+, a, g^-\}$) and assuming a Boltzmann ensemble, the populations $p_{i,j}$ of the four putative conformeric states of Ac-Pro-OMe (4) or Ac-(4*R/S*)-Flp-OMe (6) are given by

$$p_{i,j} = \frac{\exp(-E_{i,j}/k_B T)}{\sum_{i \in \{trans,cis\}} \sum_{j \in \{endo,exo\}} \exp(-E_{i,j}/k_B T)} \qquad (1a)$$

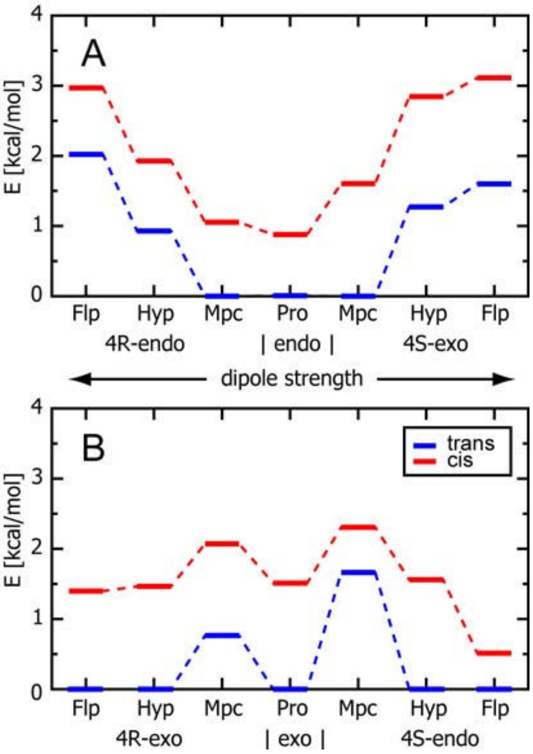
and by

$$p_{i,j} = \frac{\sum_{k \in \{g^+,a,g^-\}} \exp(-E^k_{i,j}/k_B T)}{\sum_{i \in \{trans,cis\}} \sum_{j \in \{endo,exo\}} \sum_{k \in \{g^+,a,g^-\}} \exp(-E^k_{i,j}/k_B T)} \qquad (1b)$$

for Ac-(4*R/S*)-Mpc-OMe (1) or Ac-(4*R/S*)-Hyp-OMe (5). Here k_B is the Boltzmann constant and T is the temperature. Corresponding to the experimental setup the temperature $T = 293$ K was chosen.

Suppl. Table I. Comparison between theoretical ΔJ and experimental values to quantify the relative abundance of *tran-exo, trans-endo, cis-exo* and *cis-endo* conformations present at equilibrium in D_2O at 25 °C. The dihedral angles were derived from the structures at minimum energy level obtained by q.c.c. The theoretical J_3 were derived from MestRe-J.[6]

	dihedral Hα-Hβ1	D dihedral Hα-Hβ2	J_3 calcd Hα-Hβ1	J_3 calcd Hα-Hβ2	ΔJcalcd	J_3 meas Hα-Hβ1	J_3 meas Hα-Hβ2	ΔJmeas	% exp
1(*R*)									
trans, exo	329.31	208.81	7.19	9.94	0.51	8.85	4.88	3.97	37
trans, endo	29.88	269.63	7.84	1.24	6.6				48
cis, exo	332.12	212.02	8.01	9.93	1.92	8.85	3.36	5.49	4
cis, endo	34.51	274.08	7.49	0.78	6.71				11
									100
1(*S*)									
trans, exo	332.07	211.12	8.01	10.07	2.06	6.53	8.73	2.2	82
trans, endo	22.31	262.83	9.16	1.46	7.7				2
cis, exo	335.51	214.97	7.82	8.59	0.77	8.85	3.97	4.88	3
cis, endo	35.28	275.35	7.35	1.35	6				13
									100



Suppl. Figure 2. Comparison of the relative energies of the *cis/trans/endo/exo* conformers for the various proline derivatives as calculated by DFT with the COSMO solvation model. For each derivative the energetically lowest conformer defines the point of energy zero. The energy levels of the *cis* isomers are drawn in red, those of the *trans* isomers in blue. The upper graph shows results for the compounds depicted in the first two lines of Fig. 2 exhibiting a X-C$^\gamma$ bond dipole μ within the molecular plane, , the lower graph refers to the compounds in the following two lines featuring a perpendicular orientation of μ.

Suppl. Table II. Gas phase energy differences determined by DFT calculations with B3LYP/TZVP. The numbers for Pro (4) and Flp (6) are similar to corresponding numbers published in Table 3 of Ref. [7], which however include a zero point energy correction. For the Mpc (1) and Hyp (5) solely data for the lowest-energy S-H and O-H rotamers are given.

	Δ(E_{endo}-E_{exo}) [kcal/mol]		Δ(E_{trans}-E_{cis}) [kcal/mol]	
	trans	*cis*	*endo*	*exo*
1(*R*)	-1.04	-1.01	-1.64	-1.61
1(*S*)	0.41	0.29	-1.64	-1.76
4	-0.42	-0.58	-1.31	-1.47
5(*R*)	-0.02	0.78	-1.58	-0.78
5(*S*)	-3.85	-3.05	-2.45	-1.65
6(*R*)	0.76	1.13	-1.70	-1.33
6(*S*)	-0.24	-1.79	-0.11	-1.66

Suppl. Table III. Geometric parameters of the lowest energy *trans* conformer of the Mpc (1), Pro (4), Hyp (5) and Flp (6) derivatives.

	conformer	ϕ [deg]	ψ [deg]	τ_{BD} [deg]	δ_{BD} [Å]
1(*S*)	*exo*	-60.67	145.20	98.96	2.89
4	*exo*	-60.33	146.16	98.08	2.89
5(*R*)	*exo*	-61.16	144.13	99.39	2.89
6(*R*)	*exo*	-61.30	143.93	99.64	2.89
1(*R*)	*endo*	-69.69	150.88	100.16	3.05
4	*endo*	-71.20	153.80	98.76	3.08
5(*S*)	*endo*	-63.58	139.01	104.73	2.92
6(*S*)	*endo*	-77.52	172.49	88.84	3.23

Suppl. Table IV. Populations p of the four conformeric states of proline analogues calculated according to Eqs. (1) compared with the relative abundance estimated from the experimental data.

		continuous solvent model (COSMO)		gas phase calculation		Experimental	
		trans	*cis*	*trans*	*cis*	*trans*	*cis*
1(*R*)	*endo*	63 %	11 %	82 %	5 %	48%	11%
	exo	24 %	2 %	12 %	1 %	37%	4%
1(*S*)	*endo*	4 %	2 %	22 %	2 %	2%	13%
	exo	88 %	6 %	71 %	4 %	82%	3%
4	*endo*	43 %	10 %	61 %	6 %	83% (>*endo*)	17% (>*endo*)
	exo	44 %	3 %	30 %	2 %		
5(*R*)	*endo*	21 %	3 %	44 %	3 %	87% (>*exo*)	13%
	exo	70 %	6 %	46 %	7 %		
5(*S*)	*endo*	70 %	10 %	98 %	2 %	71% (>*endo*)	29%
	exo	19 %	1 %	0 %	0 %		
6(*R*)	*endo*	3 %	1 %	20 %	1 %	88% (>*exo*)	12%
	exo	89 %	8 %	72 %	7 %		
6(*S*)	*endo*	67 %	28 %	40 %	33 %	72% (>*endo*)	28%
	exo	4 %	0 %	26 %	2 %		

References

[1] K. Wüthrich; *NMR of Proteins and Nucleic Acids*, Wiley, New York, 1986.

[2] J. T. Gerig, R. S. McLeod, *J. Am. Chem. Soc.* 1973, *95*, 5725-5729.

[3] G. Schaftenaara, J. H. Noordik, *J. Comput.-Aided Mol. Des.* 2000, *14*, 123-134.

[4] J. S. W. Lam, J. C. P. Koo, I. Hud ky, A. Varro, J. G. Papp, B. Penke, I. G. Csizmadia, *Theochem-J. Mol. Struct.* 2003, *666*, 285-289.

[5] I. K. Song, Y. K. Kang, *J. Phys. Chem. B* 2006, *110*, 1915-1927.

[6] A. Navarro-Vázquez, J. C. Cobas, F. J. Sardina, *J. Chem. Inf. Comput. Sci.*, 2004, *44*, 1680-1685.

[7] M. L. DeRider, S. J. Wilkens, M. J. Waddell, L. E. Bretscher, F. Weinhold, R. T. Raines, J. L. Markle; *J. Am. Chem. Soc.* 2002, *124*, 2497-2505.

3 Molekulardynamik-Simulationen der Faltungs/Entfaltungs-Gleichgewichte α-helikaler Peptide

Im Zentrum der nun folgenden Ausführungen stehen die Struktureigenschaften α-helikaler Peptide. Es wird analysiert, inwieweit das molekularmechanische Kraftfeld CHARMM22 (C22) bzw. dessen um den Energiebeitrag CMAP erweiterte Variante C22/CMAP geeignet ist, die Faltung α-helikaler Peptide in Molekulardynamik-Simulationen zu untersuchen. Das Kapitel enthält einen Abdruck des Manuskripts

> Rudolf Reichold, Beat Fierz, Thomas Kiefhaber und Paul Tavan:
> „Toward MD Descriptions of α-Helix Folding“,

das ich gemeinsam mit Paul Tavan aus unserer Arbeitsgruppe sowie mit Beat Fierz und Thomas Kiefhaber vom Lehrstuhl für Biophysikalische Chemie an der TU München verfasst habe. Im Anschluss an das Manuskript ist Zusatzmaterial zu diesem Artikel abgedruckt. Das Manuskript und das Zusatzmaterial stellen die Hauptergebnisse meiner Arbeit dar.

Toward MD Descriptions of α-Helix Folding

Rudolf Reichold[1], **Beat Fierz**[2†], **Thomas Kiefhaber**[2], **and Paul Tavan**[1*]

[1] *Lehrstuhl für Biomolekulare Optik, LMU München, Oettingenstr. 67, 80538 München, Germany*
[2] *Lehrstuhl für Biophysikalische Chemie & Munich Center for Integrated Protein Science,*
TU München, Lichtenbergstr. 4, 85748 Garching, Germany
[†] *present address: Laboratory of Synthetic Protein Chemistry, Rockefeller University, New York, New York 10065, USA*
[*] *corresponding author, email: tavan@physik.uni-muenchen.de, phone: +49-89-2180-9220, fax: +49-89-2180-9202*

Light-switchable peptides are ideally suited as minimal models, if one wants to study processes of peptide folding by combined experimental and theoretical methods. Here we address the question to what extent the accuracy of a popular molecular mechanics (MM) force field and the computationally accessible amount of statistical sampling suffice to complement experimental studies of rapidly folding α-helical peptides by molecular dynamics simulations. We show that the CHARMM22 force field, when extended by the CMAP correction (MacKerell et al., 2004, J. Comput. Chem. 25, 1400) actually produces α-helical models for α-helical peptides. Applying safely converging replica exchange simulations to carefully chosen peptide sequences we derive α-helix melting curves. According to these results the CHARMM22/CMAP force field sizeably overestimates the stability of α-helical structures. We explain why the CHARMM22/CMAP force field is nevertheless suitable for simulations of α-helix folding. Furthermore we demonstrate that the computer technology currently has reached a state at which light-triggered processes of α-helix folding have become accessible to MD simulation.

1. Introduction

In principle, molecular dynamics (MD) simulations can offer lively insights into the processes of protein and peptide folding that are inaccessible to other methods.[1] In fact, many MD studies (e.g. refs 1–4) succeeded to fold a disordered polypeptide into a near-native state and, thus, enabled a detailed monitoring of the corresponding folding process. However, the native conformation, i.e., a conformation matching a crystal structure with a C_α root mean square deviation smaller than 2 Å, was hardly ever found.[1,5]

The main obstacles toward a better preformance of folding simulations are the problems posed by (i) statistical sampling and (ii) force field accuracy. These problems are intimately connected because they represent the two sides of the well-known contradiction between accuracy and computational efficiency.

If one sacrifices accuracy in favor of enhanced computational speed, as has been done, e.g., by modeling the aqueous environment with so-called implicit solvent models of the "generalized Born" type,[6] then even for simple peptides such as α-helices[7] and β-hairpins[8,9] the resulting free energy landscapes were found to differ from results obtained with the much more costly explicit solvent models. Furthermore, the native state of many proteins did not appear to be the state of lowest free energy,[10] when such a generalized Born modeling was used for the solvent in combination with an established molecular mechanics (MM) force field for the peptide (AMBER[11]).

But even if one sacrifices sampling efficiency by resorting to a supposedly more accurate all-atom model for the solvent, the success of a folding simulation is not necessarily guaranteed. In a heroic attempt to simulate the folding of a small β-sheet protein in explicit solvent Freddolino et al.[5] most recently computed a huge MM-MD trajectory covering a duration of 10 μs (the required computational effort may currently represent a world-record). Experimentally the simulated protein finds its native state within about 10 μs. However, within the single 10 μs trajectory the polypeptide chain did not even remotely approach the native state and sampled several predominantly α-helical states instead. This apparent failure may simply represent a case of bad luck. It may well be that an extension of the sampling, e.g., to 10.1 μs might have shown the desired folding event. But it may also be that a (currently unfeasible) extension to the millisecond time scale would have continued to offer α-helical states due to inaccuracies of the employed force field. In the given case the protein was described by the popular all-atom force field CHARMM22 (C22)[12] in combination with the recent "correction map (CMAP)" extension.[13,14]

The quoted example highlights yet another connection between the sampling and force field problems. Because every MD simulation is a random experiment, extended statistics is required if one wants to judge whether a given force field correctly describes the free energy landscape of a given polypeptide sequence under physiological conditions. Correspondingly, any attempt to improve the quality of a given force field has to be based on extended statistics for many relevant examples. Therefore, in the context of force field development the sampling problem dictates that one cannot stick to large proteins in explicit solvent. Instead, to reduce the size and the complexity of the sample systems one is forced to consider the folding dynamics of small peptides in explicit solvent (see e.g. refs 15–17). If one should succeed to improve

MM force fields to a point at which these folding processes are adequately and reliably covered by all-atom simulations, one will have a data base allowing to judge and tune the more efficient implicit solvent models.[18–21]

Having just spelled the general purpose to which also our current work is devoted, a follow-up problem appears that may be summarized by the question as to how one can gain reliable and sufficiently detailed experimental data on the folding dynamics of peptides in solution that can provide references for computational descriptions. As boundary conditions one has to require that (i) the respective peptides should be as small as possible, that (ii) the folding dynamics can be elicited at a well-defined point in time and conformational ensemble, that (iii) it can be monitored in a time-resolved fashion, that (iv) it proceeds as rapidly as possible to an equally well-structured equilibrium state, and that (v) it nevertheless can serve as a model for physiological processes of peptide folding.

The construction of a model system fulfilling most (if not all) of the above conditions is by no means simple and several of the emerging problems will be discussed in this paper. Quite clearly, the condition (ii) of a most well-defined starting time for the folding process suggests the use of a light-trigger. Photocleavage of a disulphide bridge is an early example of such a technique.[22] Although this approach is hampered by the high recombination rate of the thiyl radicals,[22,23] it has nevertheless been successfully applied to study the hydrogen-bond dynamics in a β-turn.[24] In a similar fashion, azobenzene dyes either covalently integrated into the backbone of cyclic[15,16,25] and β-hairpin[17] peptides or connecting side-chains of an α-helical peptide[23,26] have been employed as ultrafast (sub-picosecond) triggers of peptide refolding. However, in the case of the small cyclic peptides[15,16,25] the conformational space was quite constrained thus violating condition (v). In the α-helical construct[23,26] the light-induced change of the azobenzene geometry was not large enough to induce a complete folding-unfolding transition thus also contradicting condition (v). In the case of the light-triggered β-hairpin[17] the folding time was quite extended (μs) and, therefore, at variance with condition (iv). Here, the sizeable folding time restricted the simulation descriptions to the much faster process of enforced unfolding. Note in this connection that for β-hairpin peptides the simulations of the folding-unfolding equilibria are generally hampered by the fact that supposedly more efficient sampling techniques like replica exchange[27–29] may fail and even lead to efficiency reduction.[30]

With the aim of constructing a light-triggered system that exhibits a sufficiently rapid folding [condition (iv)] and remains unconstrained during this process [condition (v)] we decided to construct a peptide that starts at an artificially cyclic and, thereby, covalently restrained conformation. Upon light excitation the constraint should be rapidly removed thus allowing the peptide to quickly fold into an α-helix. We wanted to study this process in parallel by spectroscopic and MD simulation techniques. This was the plan.

In this paper we report how far we meanwhile proceeded on this way mainly concentrating on the simulation aspects. It is organized as follows. First we outline the design- and selection principles of the model peptides thus introducing the systems studied by MD simulation. Subsequently we introduce the computational methods covering issues of force fields, sampling, and data analysis. The presentation and discussion of the results will be concluded by a short summary and outlook.

2. Peptide Design

When converting the above design principles, which aim at a peptide suited for studies of α-helix folding (i.e. at a peptide that contains a photo-cleavable crosslink and lacks helical contacts upon cyclic closure but is highly helical in its open form), into a specific molecular construct one is confronted with two main challenges. One has to choose firstly an appropriate photo-cleavable intrachain crosslink and secondly a suitable peptide sequence.

As our cross-link we chose the p-hydroxyphenacyl (pHP) derivative[31] shown in Figure 1. It consists of a carboxylated pHP molecule N-terminally attached to a peptide X_1-X_2-X_3-X_4-E-R composed of four residues X_i, $i = 1, \ldots, 4$ preceding a glutamate E and an additional chain R of C-terminal amino acids. The carboxylated pHP can form a photocleavable crosslink with the carboxyl group of the glutamate in the i,$i + 5$-position which should strongly destabilize a helix. As shown for various other derivatives of pHP, irradiation at 300-350 nm should release the depicted product in an irreversible rearrangement reaction at a high yield ($\Phi \sim 0.5$) and with a fast kinetics on the nanosecond time scale.[32,33]

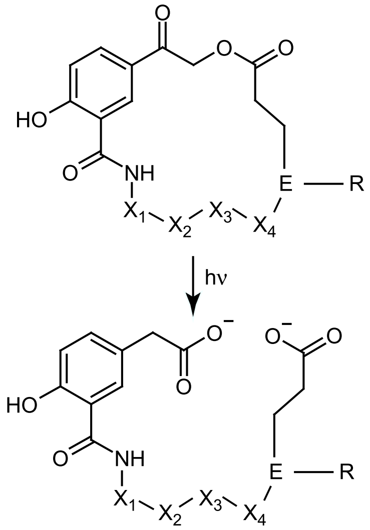


Figure 1. Chemical structures of the closed and open forms of our pHP peptides (X_i, $i = 1, \ldots, 4$, N-terminal residues, R = different C-terminal peptide sequences, see the text for details).

Because of the high helical propensity of alanine residues (see e.g. refs 34 and 35) choices of alanine rich sequences seemed to be reasonable. Furthermore, to enhance the water solubility, the addition of two or more lysine residues appeared to be wise. By placing one of them one helix turn distant from the glutamate, which is set free by photolysis, one could expect to stabilize the helix by the formation of a favorable intramolecular salt-bridge.[36,37] Finally one could

TABLE 1: The Peptides Considered in this Study[a]

$\tilde{\mathrm{P}}_{\mathrm{pHP}}$:	pHP-AAAKEAAAKK-NH_2
$\tilde{\mathrm{P}}$:	Ace-AAAKEAAAKK-NH_2
P_{E}:	Ace-EAAAAEAAAKKGY-NH_2
$\mathrm{P}_{\mathrm{pHP}}$:	pHP-AAAAEAAAKK-NH_2
$\mathrm{P}^{\mathrm{XL}}_{\mathrm{pHP}}$:	pHP-AAAAEAAAKAAAAK-NH_2
$\mathrm{P}_{\mathrm{pHP}*}$:	IPA-AAAAEAAAKK-NH_2
$\mathrm{P}^{\mathrm{XL}}_{\mathrm{pHP}*}$:	IPA-AAAAEAAAKAAAAKGY-NH_2

[a] Ace designates an acetylated N-terminus. At a pH of 7 the lysines are protonated and the glutamic acids as well as the open pHP phototrigger are deprotonated. The peptides $\mathrm{P}_{\mathrm{pHP}*}$ and $\mathrm{P}^{\mathrm{XL}}_{\mathrm{pHP}*}$ are cost effective models for the expensive peptides $\mathrm{P}_{\mathrm{pHP}}$ and $\mathrm{P}^{\mathrm{XL}}_{\mathrm{pHP}}$ with the open phototrigger. In these models the pHP group is replaced by an isophtalic acid (IPA). Figure 18 in the SI compares the structures of pHP and IPA.

expect that C-terminally placed lysines lead to an additional stabilization through charge/helix-dipole interactions.[38] With these design principles in mind we came up with the series of peptides presented in Table 1.

The peptide $\tilde{\mathrm{P}}_{\mathrm{pHP}}$ was the original proposal. In addition to the salt-bridge between the residues E_5 and K_9 this peptide can form a second helix-favoring salt-bridge between the cation K_4 and the phototrigger pHP, which carries a negative charge in its open form (cf. Figure 1). Furthermore, a negatively charged residue like pHP near the N-terminus is known to favor helix formation through charge/helix-dipole interactions.[38] Being composed of only 10 residues, the molecule $\tilde{\mathrm{P}}_{\mathrm{pHP}}$ clearly meets requirement (i) stated in the Introduction, i.e. it is very small thus reducing the computational effort in simulations. Finally, with a closed cross-link the peptide $\tilde{\mathrm{P}}_{\mathrm{pHP}}$ is expected to be completely unfolded, because the C-terminal part R = AAAKK-NH_2 contains only five residues and, therefore, should be too short for stably folding into a helix.

After this design was drawn on a paper the first task for the theoreticians in the joint project was defined: Because the pHP cross-link is an unusual chemical motif, which is neither contained in the force field C22 employed by us nor in any other standard peptide force field such as AMBER[11] or GROMOS,[39] suitable force fields for open and cyclically closed pHP had to be derived using quantum chemical methods. However, because this task is time-consuming, we thought that it might be a good idea to immediately start extended folding simulations for a model of the peptide $\tilde{\mathrm{P}}_{\mathrm{pHP}}$, which can be constructed from the chemical building blocks covered by C22. The peptide $\tilde{\mathrm{P}}$ characterized in Table 1 is the result of this initial model design.

When carefully inspecting the sequence of $\tilde{\mathrm{P}}$ one recognizes that it differs from $\tilde{\mathrm{P}}_{\mathrm{pHP}}$ in a physically important detail: The peptide $\tilde{\mathrm{P}}$ exhibits an acetylated, i.e. electrostatically neutral, N-terminus and therefore lacks the negative charge introduced by the open pHP terminus into the parent peptide $\tilde{\mathrm{P}}_{\mathrm{pHP}}$. As explained above, this negative charge was supposed to additionally stabilize the expected $\tilde{\mathrm{P}}_{\mathrm{pHP}}$ helix through electrostatic interactions. Due to the lacking N-terminal anion the helix-propensity of $\tilde{\mathrm{P}}$ should be smaller than that of the open $\tilde{\mathrm{P}}_{\mathrm{pHP}}$.

Unfortunately, $\tilde{\mathrm{P}}_{\mathrm{pHP}}$ could only be synthesized with low yields. Moreover, aqueous solutions of $\tilde{\mathrm{P}}_{\mathrm{pHP}}$ with an open pHP trigger exhibited an unexpectedly low helix content (data not shown). These findings were the reason for a revision of the original design and led to the peptide family P characterized in the lower part of Table 1. A sequence comparison of the families P and $\tilde{\mathrm{P}}$ shows the key difference: In family P the residue X_4 is an alanine replacing the lysine present in family $\tilde{\mathrm{P}}$. This variation was chosen to enable the cyclization during synthesis of the closed pHP phototrigger.

The new peptide family P contains besides the short peptide $\mathrm{P}_{\mathrm{pHP}}$ two further members: A modified and slightly elongated model P_{E} and a C-terminally more strongly elongated photocleavable peptide $\mathrm{P}^{\mathrm{XL}}_{\mathrm{pHP}}$. In the P_{E} model for $\mathrm{P}_{\mathrm{pHP}}$ the additional glutamate at the N-terminus introduces a negative charge mimicking the charge of the open pHP phototrigger. Furthermore, the C-terminus has been elongated by the two residues GY because optical concentration measurements require at least one aromatic residue in the sequence. The C-terminal GY addition is known to leave the α-helical content of the remaining peptide invariant.[34] In the photocleavable compounds the required aromatic residue is automatically present through pHP. Thus, one sees that the initially studied small model peptide $\tilde{\mathrm{P}}$, which lacks an aromatic residue, could solely be used for purposes of simulation while escaping a simple experimental characterization. Because P_{E} and $\mathrm{P}^{\mathrm{XL}}_{\mathrm{pHP}}$ are much larger than the initial model $\tilde{\mathrm{P}}$, the corresponding simulations require a substantially enhanced computational effort thus conflicting with the design requirement (i) stated in the Introduction. In contrast, the computational effort required for the short peptide $\mathrm{P}_{\mathrm{pHP}}$ seems to be better manageable.

Before one can actually study the light-triggered folding dynamics using the peptide constructs introduced above in parallel by spectroscopic and simulation techniques, one must make sure, of course, that aqueous solutions of these peptides with open phototriggers actually contain stably folded α-helices at temperatures compatible with the liquid state of water. The α-helix content should be high not only in the experimental situation but also in the simulation model. Furthermore, the experimentally accessible thermal unfolding behavior should be reproducible on the computer, because a corresponding match represents a sensitive check on the quality of the employed force field. With such a check at hand one will have gained a solid basis to study also the nonequilibrium folding dynamics. Thus, on the simulation side, the first task was to characterize the conformational equilibria of the various peptides in aqueous solution as functions of the temperature.

For the actual spectroscopic measurements, model peptides corresponding to the simulated sequences were synthesized (cf. Table 1). As a more cost effective way to gauge the effect of the open phototrigger on helix propensity, isophtalic acid (IPA) was coupled to the N-termini of the synthesized peptides, which mimics the open pHP structure (cf. Figure 18, SI).

3. Methods

Here we describe the procedures and, particularly, the quantum chemical methods employed to derive a suitable force field for the pHP light trigger. Subsequently, we sketch the setup of the MD simulation systems used to characterize the conformational ensembles of the peptides $\tilde{\mathrm{P}}$, $\mathrm{P_E}$, $\mathrm{P_{pHP}}$, and $\mathrm{P_{pHP}^{XL}}$ in aqueous solution at various temperatures. After a specification of the MD methods we provide the equilibration procedures preceding the simulations serving for data acquisition. Next, the various types and details of these data production runs are explained and listed. Finally we outline the observables used in the statistical data analysis for the identification of secondary structures.

3.1. A Force Field for pHP. As indicated above, we intended to use the C22 force field[12] for the light-switchable peptides. C22 assigns names and types to the atoms occurring in a chemical compound. The atom type specifies the bonding motif surrounding a given atom. C22 describes the van der Waals interactions of distant atoms by Lennard-Jones potentials and their electrostatic interactions by Coulomb potentials of atomic partial charges. Thus, to define force fields for the open (pHPo) and closed (pHPc) trigger molecules, one has to choose suitable parameters specifying these so-called nonbonded potentials.

Figures 15 and 16 in the Supporting Information (SI) show the atom names introduced by us for the compounds pHPo and pHPc. Tables 5 and 6 (SI) assign atom types to these names. Some of these types are C22 standard (characterized by asterisks in the tables), some of them are new. To enable a selection of suitable Lennard-Jones parameters from the C22 data base, Table 7 in the SI assigns the newly introduced atom types to standard C22 types. This assignment has been based on a comparison of the local bonding structures found in pHPo and pHPc with structures covered by C22.

In C22, the parametrization of the nonbonded interactions is complete as soon as also partial charges are assigned to the various atoms. The charges employed for the pHP molecules are also listed in the SI Tables 5 and 6. In part they have been adopted from equivalent chemical structures covered by C22. Others have been derived from quantum chemical calculations on model compounds using density functional theory (DFT). These DFT calculations have also been used to derive parameters for the so-called bonded potentials by which C22 generally treats the stereo-chemical bonding forces within molecules. The latter comprise harmonic bond stretching potentials E_b, harmonic bond angle deformation potentials E_a, harmonic improper dihedral potentials E_i, and nonharmonic proper dihedral potentials E_d which are usually modeled by strongly truncated Fourier expansions.[12] The force constants and equilibrium constants entering the harmonic potentials, together with the Fourier expansion coefficients occurring in E_d, form the parameters to be determined by DFT for pHPo and pHPc.

The compounds treated by DFT, which serve to model the various parts of pHPo and pHPc, are depicted in Figure 2. Here, compound A covers the chemical core of pHPo with $\mathrm{Meth_1}$ modeling the C_α atom of residue $\mathrm{X_1}$ in the peptide and $\mathrm{Meth_2}$ replacing the $\mathrm{CH_2{-}COO^-}$ group (cf. Figure 1). The latter group is equivalent to the side chain of a glutamate and, therefore, described by C22. Similarly compounds B and C cover two different parts of pHPc with the various methyl groups terminating the molecules.

The DFT calculations were carried out with the program TURBOMOLE 5.6.[40,41] Here, the BP86[42,43] functional and the TZVP basis set[44] (a Gaussian basis set of triple-zeta valence quality augmented by polarization functions) were used. Values for the parameters of the bonded potentials E_b, E_a, E_i and E_d were derived as follows:

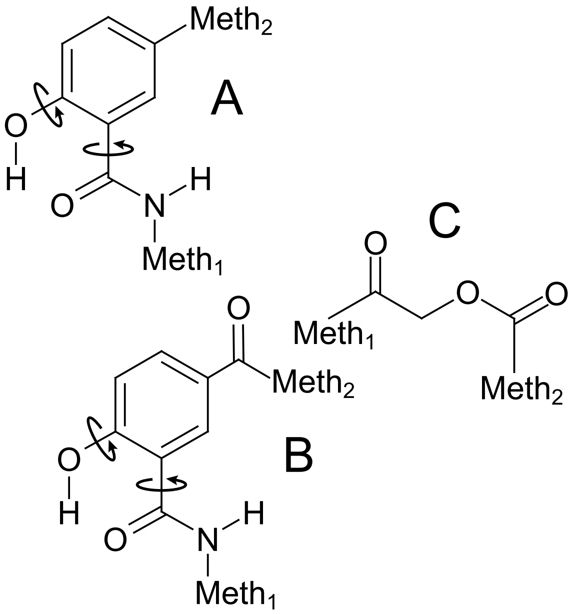


Figure 2. Chemical structures of model compounds for pHPo (A) and pHPc (B,C). The arrows indicate essential torsional degrees of freedom.

For each associated internal coordinate, potential energy curves were calculated by DFT while relaxing all other internal coordinates within the model compounds A, B, and C, respectively. Using trial values for the parameters of the MM potentials, for each internal coordinate also an MM energy curve was calculated. In these calculations the DFT geometries which characterize the various points of the respective DFT curve were adopted. The MM parameters were subsequently adjusted to minimize the differences between the DFT and MM energy curves and this process was iterated over the set of internal coordinates until approximate self-consistency was achieved. The resulting values for all these parameters are listed in the Tables 8 - 11 (SI).

Special care was taken in the parametrization of the two torsional degrees of freedom highlighted by arrows in Figure 2. These degrees of freedom are expected to be important for the conformational dynamics of the pHP peptides, because they are localized at the interface between the phenyl moiety marking the beginning of the pHP trigger and the first amide group in the backbone of the pHP peptide. To estimate suitable torsional potentials we computed the DFT energy surface on the plane spanned by the two torsional angles

TABLE 2: Sizes of Simulation Systems Constructed for Peptides Specified in Table 1[a]

	peptide			
	$\tilde{\mathrm{P}}$	$\mathrm{P_E}$	$\mathrm{P_{pHP}}$	$\mathrm{P_{pHP}^{XL}}$
atoms per peptide	150	181	151	191
TIP3P molecules	1809	5757	1809	5768
r_i [Å]	21.1	30.9	21.2	31.1

[a] r_i is the radius of the largest sphere that fits into the respective orthorhombic, dodecahedral, and periodic water box.

using a suitable discretization and optimized the associated potentials E_d for an optimal match of the DFT and MM energy surfaces. As is common practice in C22, we employed redundant sets of dihedral angles for the two torsions. These sets cover each possible selection of atoms neighboring the central bond. Each of the associated dihedral potentials was modeled by a Fourier expansion truncated after the first term (cf. Table 11, SI).

3.2. Setup of the Simulation Systems. Ideal α-helical models of the four peptides $\tilde{\mathrm{P}}$, $\mathrm{P_E}$, $\mathrm{P_{pHP}}$, and $\mathrm{P_{pHP}^{XL}}$ were built with the program MOLDEN 4.0.[45] These α-helices were defined by choosing the values $(-58°, -47°)$ for the pair of dihedral angles (ϕ, ψ) which characterize the backbone geometry at the C_α atoms. To enable MD simulations of the peptides in explicit solvent, periodic boxes shaped as orthorhombic dodecahedra were filled with water molecules modeled by the C22 variant[12,46] of Jorgensen's[47] three point transferable interaction potential (TIP3P). The periodic water boxes were equilibrated by MD simulations in the NpT ensemble to assume the target pressure $p = 1.0$ atm and the target temperature $T = 300$ K. Now, the peptide models were transferred into the centers of the boxes, and water molecules with distances smaller than 2.0 Å of any water atom from any peptide atom were removed. Subsequently, the systems were equilibrated toward the target pressure and temperature as described further below. Details of the resulting simulation systems are collected in Table 2.

3.3. Simulation Methods. All MD simulations were performed with the program package EGO-MMVI[48] which was modified to allow the use of the CMAP extension[14] to the C22 force field.[12] The CMAP topology information and parameters were taken from the file "toppar_c31b1.tar.gz" found on A. D. MacKerell's website. Apart from the pHP group, the peptides were described by the C22 force field. Some simulations were carried out with (C22/CMAP) and others without (C22) the CMAP extension. Most simulations employed for the water molecules the C22 variant of the TIP3P model introduced above, some simulations used also Berendsen's extended simple point charge model (SPC/E).[49]

As is common in EGO-MMVI, the long-range electrostatic interactions were treated by a particular combination of structure-adapted multipole expansions[50] with a moving boundary reaction-field approach and a multiple time step integrator.[48,51] Here, the default values specified in ref 48 were chosen and the basic time step was set to 1 fs. The van der Waals interactions were calculated explicitly up to 10 Å while at larger distances a mean field approach[52] was applied. Covalent bonds involving hydrogen atoms were kept fixed by the M-SHAKE algorithm.[53]

3.4. Equilibration Procedures. Above we have mentioned equilibrations of the simulation systems. These equilibrations started with short energy minimizations. Subsequently, while keeping the peptide atoms and the box size fixed, in a 10 ps MD simulation the solvent was heated to $T = 300\,\mathrm{K}$ by coupling the water molecules to a Berendsen thermostat[49] with a relaxation time $\tau = 0.1\,\mathrm{ps}$. Next we switched to the NpT ensemble for the following 100 ps of MD simulation. Here all atoms in the system were coupled to the thermostat and to a Berendsen barostat[49] characterized by the target pressure $p = 1.0\,\mathrm{atm}$, by the relaxation time $\tau = 1\,\mathrm{ps}$, and by the isothermal compressibility $\beta = 4.9 \cdot 10^{-5}\,\mathrm{atm}^{-1}$ of water at 278 K.[54] During this equilibration step the peptide atoms were allowed to move more or less freely. Solely the positions of the C_α atoms were bound by harmonic potentials [with force constants of 50 kcal/(mol Å^2)] to the ideal α-helical positions. Next a 100 ps equilibration toward the NVT ensemble was carried out. Here, the barostat was switched off and the system size was chosen as the average over the last 90 ps of the preceding equilibration step. Whenever different, statistically independet, and α-helical starting structures of the peptides were required, another 100 ps simulation was added for each new starting structure. These starting structures were then obtained by a two step procedure. First, a 50 ps MD simulation in the NVT ensemble was carried out during which the harmonic restraints on the C_α atoms were slowly reduced to zero. Next the system was simulated for 50 ps in the NVT ensemble and, here, solely the water molecules remained coupled to the thermostat.

The exclusive coupling of the solvent to the thermostat was maintained during all MD simulations that served for data acquisition. This strategy of temperature control had been motivated by the results of ref 55. As explained in the quoted paper, the thermal contact between the solvent and the peptide can suffice to keep a solute peptide at the desired temperature. For our peptides this expectation turned out to hold very well (i.e. with a root mean square deviation of 1% relative to the target temperature and averaged over all REST simulations listed in Table 3). Furthermore, as also explained in that paper, a coupling of a thermostat to the peptide may perturb the dynamics of this molecule. Thus, to avoid such computational artifacts, the thermostat was coupled only to the water.

3.5. Enhanced Sampling. In many cases, we chose to sample the conformational space of the solvated peptides by standard 300 K MD simulations. However, in some cases this straight forward sampling method turned out to be too slow for achieving sufficient statistics on the respective 300 K equilibrium ensemble. Therefore, we additionally applied the "replica exchange with solute tempering (REST)" technique suggested in ref 56. For systems like ours, which contain only a few peptide degrees of freedom surrounded by large numbers of water molecules, REST can span a given temperature range (e.g. from 300 K to 500 K) with much fewer replicas than the conventional "temperature" replica exchange method (TREM)[27–29] and, nevertheless, can guarantee a high exchange probability.

The central idea of REST is the division of the simulation system into two parts, e.g., the peptide and the solvent. Then the potential energy $E_0(X)$ of the system in a configuration X is composed of three contributions

$$E_0(X) = E_{pp}(X) + E_{ps}(X) + E_{ss}(X) \tag{1}$$

where E_{pp}, E_{ps}, and E_{ss} are the internal energy of the peptide, the interaction energy between the peptide and the solvent, and the energy of the solvent, respectively. In REST, only the replica at the lowest temperature T_0 is described by eq (1) whereas the potential energies of replicas at elevated temperatures T_i are scaled. Choosing a scaling which slightly differs from the one used in ref 56 but already suggested therein, the potential energy of a replica at temperature T_i is given by

$$E_i(X) = E_{pp}(X) + \sqrt{\frac{T_i}{T_0}}E_{ps}(X) + \frac{T_i}{T_0}E_{ss}(X), \tag{2}$$

i.e. E_{pp} remains unscaled whereas the other two terms are differently scaled. In regular time intervals, all pairs of replicas simulated at neighboring temperatures T_i and T_j (with configurations X_i and X_j) exchange temperatures with the probability

$$P_{ij} = \min[1, \exp(\Delta_{ij})]. \tag{3}$$

Here,

$$\begin{aligned}\Delta_{ij} &= (\beta_i - \beta_j)\,[E_{pp}(X_j) - E_{pp}(X_i)] \\ &+ \left(\sqrt{\beta_0\beta_i} - \sqrt{\beta_0\beta_j}\right)[E_{ps}(X_j) - E_{ps}(X_i)]\end{aligned} \tag{4}$$

with $\beta_i = 1/k_B T_i$ and with the Boltzmann constant k_B. The form of the exchange probability defined in eqs (3) and (4) guarantees[56] that the replica at the lowest temperature T_0 samples the canonical ensemble associated with the physical energy function eq (1). The replicas at higher temperatures, in contrast, sample ensembles which are associated with the modified energy functions (2). Therefore, in contrast to the properties of TREM, they cannot accurately describe the physical behavior of the solvated peptide at elevated temperatures. Instead they describe something like a hot peptide (at T_i) in a cool solvent (at T_0).

To avoid the latter drawback of REST and, concurrently, to take advantage of its efficient sampling, we designed a new combination of TREM and REST. In this TREM/REST combination, one part of the replicas, covering the temperature range $[T_l; T_m]$, is simulated following the TREM protocol while another part, covering $[T_m; T_h]$, is treated by REST. Thus, the TREM replica at the highest temperature T_m additionally has the role of the lowest REST replica, i.e. T_m is the target temperature of the REST sampling. TREM/REST allows one to cover the temperature range $[T_l; T_m]$ by replicas with an unscaled energy function. Correspondingly one samples a set of equilibrium ensembles within this range. The REST ensemble simulated at much higher temperatures then can speed up the convergence of sampling in the TREM temperature range.

TABLE 3: Simulations Carried out in this Study

sys^a	ff^b	type^c	duration^d	sc^e
$\tilde{\mathrm{P}}$.1	C/T	300 K	4 x 50 ns	α
$\tilde{\mathrm{P}}$.2	CC/T	300 K	4 x 50 ns	α
$\tilde{\mathrm{P}}$.3	CC/T	300 K	10 x 50 ns	u
$\tilde{\mathrm{P}}$.4	CC/T	REST	25 ns	α
$\tilde{\mathrm{P}}$.5	CC/T	REST	25 ns	u
$\tilde{\mathrm{P}}$.6	CC/T	T/R	25 ns	α/u
$\mathrm{P_E}$.1	C/T	300 K	5 x 10 ns	α
$\mathrm{P_E}$.2	C/S	300 K	5 x 10 ns	α
$\mathrm{P_E}$.3	CC/T	300 K	5 x 10 ns	α
$\mathrm{P_E}$.4	CC/T	REST	25 ns	α
$\mathrm{P_{pHP}}$.1	CC/T	300 K	3 x 50 ns	α
$\mathrm{P_{pHP}}$.2	CC/T	REST	25 ns	α
$\mathrm{P_{pHP}}$.3	CC/T	REST	25 ns	u
$\mathrm{P_{pHP}^{XL}}$.1	CC/T	300 K	3 x 25 ns	α

[a] Name characterizing a simulation of a peptide (cf. Table 2). [b] Force field for peptide/solvent system; C stands for C22, CC for C22/CMAP, T for TIP3P, and S for SPC/E. [c] Sampling method; 300 K denotes conventional MD simulations in the NVT ensemble; T/R denotes TREM/REST. [d] Duration of the simulation(s), e.g., 3 x 25 ns means that 3 trajectories spanning 25 ns each were calculated for different starting configurations. [e] Conformation of starting structures which may be α-helical (α), unfolded (u), or a mixture (α/u).

3.6. Simulations. Table 3 specifies the simulations carried out in this study. The table assigns to each simulation a unique name which, concurrently, is a pointer to one of the simulation systems listed in Table 2. Furthermore it characzterizes the force field employed in the respective simulation for the peptide/water system, the sampling method applied, the durations and multiplicities of the trajectories collected under the given name, and finally the employed starting structures which have been mostly α-helical but sometimes also unfolded or a mixture of the two.

The setup of the REST and TREM/REST simulations listed in Table 3 deserves a more detailed characterization. Each of the listed REST simulations employed 10 replicas distributed among the temperatures 300 K, 321 K, 343 K, 366 K, 390 K, 416 K, 444 K, 474 K, 506 K, and 540 K. This setup resulted in exchange probabilities (cf. eq 3) in the range between 30% and 40%. Exchanges between neighboring replicas were attempted every 10 ps.

For each of the peptides $\tilde{\mathrm{P}}$ and $\mathrm{P_{pHP}}$ we have carried out two REST simulations with different initial conditions, one starting with 10 independent α-helical starting structures (simulations $\tilde{\mathrm{P}}$.4 and $\mathrm{P_{pHP}}$.3) and the other with 10 independent unfolded structures (simulations $\tilde{\mathrm{P}}$.5 and $\mathrm{P}_{\mathrm{pHP}}$.4). By this strategy we wanted to enable an assessment to what extent the REST sampling has reached convergence.[57] The required unfolded starting structures were randomly selected from the 540 K replicas of the simulations $\tilde{\mathrm{P}}$.4 and $\mathrm{P}_{\mathrm{pHP}}$.3, respectively. Because the computational cost for peptide P_{E} is much higher than that for $\tilde{\mathrm{P}}$ or $\mathrm{P}_{\mathrm{pHP}}$, only one REST simulation (P_{E}.4) starting from folded structures was manageable for us in this case.

For peptide $\tilde{\mathrm{P}}$ we additionally performed the TREM/REST simulation $\tilde{\mathrm{P}}$.6. The TREM temperature range [340.0 K; 383.2 K] was covered by 10 replicas and 4 REST replicas were added to reach the topmost temperature of $T_h = 496.0$ K. Replica exchanges were attempted every 5 ps. The exchange probabilities were in the range given above. The mixture of 10 folded, partially folded, and unfolded starting structures used for the TREM initialization was adopted from an intermediate (20 ns) state of the REST simulation $\tilde{\mathrm{P}}$.4, whereas the 4 starting structures required for the REST part were selected as different unfolded snapshots from the 540 K replica of simulation $\tilde{\mathrm{P}}$.4.

3.7. Secondary Structure Analysis. We measure the local secondary structures of the peptides using the dictionary of protein secondary structure (DSSP).[58] Based mainly on the analysis of hydrogen bonding patterns, DSSP classifies each residue of a peptide into one of eight classes, i.e. as locally α- or π-helical, β-strand, etc.

If N_α is the DSSP number of α-helical and N the total number of residues in a given structure then one may define the *helix fraction* by

$$H_\alpha = \frac{N_\alpha}{N-2} \tag{5}$$

where the reduction of N by 2 is due to the fact that DSSP does neither classify the first nor the last residue of a peptide. Correspondingly one can define such fractions for other secondary structure motifs. Evaluating H for every MD time step t gives a helix fraction trajectory $H(t)$ whose visualization monitors the change of α-helical content during a simulation.

3.8. Peptide Synthesis. Peptides were synthesized using standard Fmoc-chemistry on an Applied Biosystems 433A peptide synthesizer. To modify the peptide N-termini with isophtalic acid (IPA), t-butyl protected IPA (3-(tert-butoxycarbonyl)benzoic acid) was coupled to the resin bound peptides using PyBOP (benzotriazol-1-yl-oxytripyrrolidinophosphonium hexafluorophosphate) activation followed by TFA (trifluoroacetic acid) deprotection. All peptides were purified to $> 95\%$ purity by preparative HPLC on a RP-8 column. Purity was checked by analytical HPLC and the mass was determined by MALDI mass spectrometry.

3.9. CD Spectroscopy. Circular dichroism measurements were performed in 10 mM phosphate buffer, pH 7.0 at 5 °C on an Aviv DS62 spectropolarimeter. All solutions were degassed prior to the measurements. Peptide concentrations were between 100 and 200 μM as determined by UV absorption. A Gly-Tyr sequence was added to the C-terminus of peptides containing no aromatic side chains. The addition of Gly-Tyr at the C-terminus allows spectrometric concentration determination ($\epsilon_{275} = 1450\ \mathrm{M}^{-1}\ \mathrm{cm}^{-1}$) and does not significantly affect the far-UV CD spectrum.[34] The helix content of the peptides was calculated from measured CD data at 222 nm using values of $[\Theta]^{\mathrm{H}}_{\mathrm{MRW}} = -44000 \cdot (1 - 3/N)$ deg cm^2 dmol^{-1} (with N the number of residues) for 100% helical state and $[\Theta]^{\mathrm{C}}_{\mathrm{MRW}} = 2220$ deg cm^2 dmol^{-1} for 100% unfolded state.[59] Here, the subscript MRW indicates that the ellipticity was calculated per mean residue weight. Experimental data were analyzed using ProFit software (QuantumSoft, Zürich, Switzerland).

4. Results and Discussion

As motivated in section 2, we started our project with extended simulations on the model $\tilde{\mathrm{P}}$ of our original target peptide $\tilde{\mathrm{P}}_{\mathrm{pHP}}$, because this model, unlike $\tilde{\mathrm{P}}_{\mathrm{pHP}}$, exclusively consists of chemical building blocks covered by C22 and, therefore, does not have a time-consuming force field computation as its prerequisite.

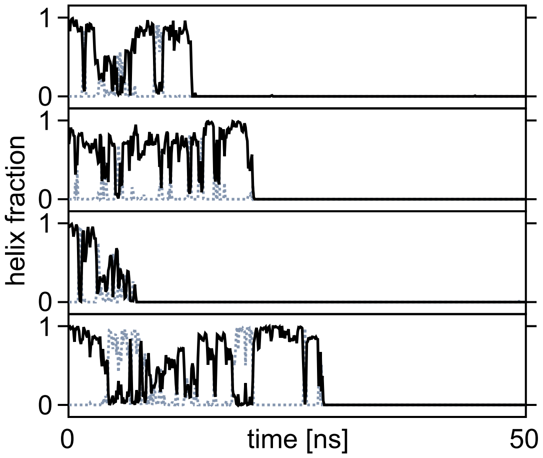


Figure 3. Trajectories of DSSP helix contents obtained by simulation $\tilde{\mathrm{P}}$.1; α-helix fraction $H_\alpha(t)$: black solid; π-helix fraction $H_\pi(t)$: gray dotted.

Figure 3 shows the helix fractions observed in the 4 trajectories (50 ns, 300 K) of simulation $\tilde{\mathrm{P}}$.1. After 8 - 28 ns the initial α-helices are completely unfolded and do not refold within the simulated time spans. Before complete unfolding the peptide is seen to exhibit occasionally π-helical contributions which also disappear upon unfolding.

Simulation $\tilde{\mathrm{P}}$.1 thus suggests that the model $\tilde{\mathrm{P}}$ of our original target peptide $\tilde{\mathrm{P}}_{\mathrm{pHP}}$ is unexpectedly unfolded at 300 K in aqueous solution — at least if one trusts the C22 force field in combination with the water model TIP3P. To what extent this prediction actually holds for the model peptide $\tilde{\mathrm{P}}$ in water at room temperature cannot be easily checked experimentally because $\tilde{\mathrm{P}}$ unfortunately does not contain the aromatic residue required for optical measurements. The lacking helical propensity of $\tilde{\mathrm{P}}$ could, of course, be caused by another shortcoming of its design: As mentioned already in section 2, the model $\tilde{P}$ lacks a negative charge near the N-terminus which, following the helix-dipole model,[38] can contribute to the stabilization of an α-helical conformation.

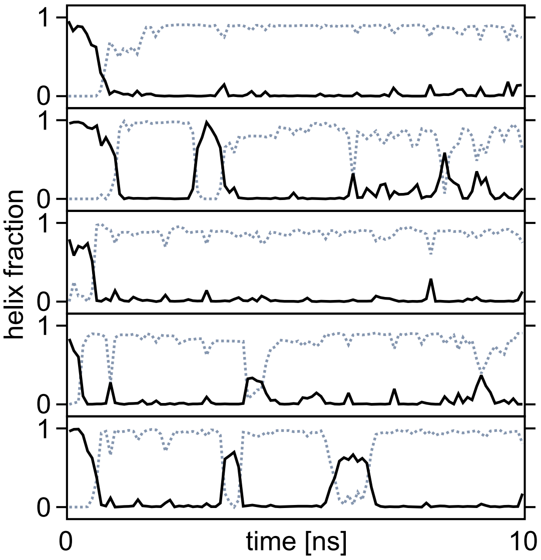


Figure 4. Trajectories of DSSP helix fractions resulting from simulation $P_E.1$; linestyles are explained in the caption to Figure 3.

To check whether adding a negative charge near the N-terminus increases the α-helical propensity we next considered the peptide P_E, which models the revised target peptide P_{pHP} and is accessible to spectroscopic helix content measurements (cf. section 2). Figure 4 shows the 5 trajectories $H_\alpha(t)$ and $H_\pi(t)$ collected in simulation $P_E.1$. In each of these 10 ns trajectories the peptide rapidly refolds from its inital α-helical conformation into a π-helix which is apparently quite stable.

Thus our simulation assigns to the sequence chosen for peptide P_E a π-helical instead of the expected α-helical conformation. However, π-helices are quite unusual secondary structure motifs in proteins and peptides.[60,61] Furthermore, the prediction of π- instead of α-helices is a well-known artifact of the C22 force field.[62] Therefore, it seemed likely that also the π-helical propensity observed for peptide P_E is an artifact of C22. To scrutinize this issue we repeated simulation $P_E.1$ replacing, however, the water model TIP3P by the model SPC/E (simulation $P_E.2$). As demonstrated by Figure 17 in the SI, which presents the results of simulation $P_E.2$, a different water model does not change the π-helical propensity of the peptide P_E. Hence, this π-helical character must be attributed exclusively to the C22 force field describing the peptide.

The observation of the frequent occurrence of π-helices in MD simulations employing C22 and other empirical force fields has prompted the developers of C22 to improve the treatment of the peptide backbone energetics in their force field.[14] Correspondingly, the C22 energy function was complemented by the CMAP extension $E_{CMAP}(\phi, \psi)$. Based on quantum chemical calculations of the ϕ/ψ energy surface of the alanine, glycine, and proline dipeptides, E_{CMAP} includes correlations of the ϕ/ψ dihedral potentials. It is implemented through a grid-based interpolation that reproduces the difference between the C22 and the QM target energy surfaces almost exactly.[14] Aiming at the reproduction of experimental macromolecular target data, this QM derived energy correction was then empirically optimized to take into account condensed phase contributions to the energy surface not included in the QM energy surfaces. For this purpose MD simulations of a number of proteins in their crystal environments were performed. The C22/CMAP force field was finally tested by short (5 ns) MD simulations of three proteins in solution which demonstrated that the crystallographic starting structures were now maintained much better.[14]

As shown by our short summary of the CMAP development and validation,[14] only dipeptides and proteins but no helical peptides were considered here. This choice of test systems is somewhat astonishing, because the observation of π-helices in MD simulations of peptides with C22 was one of the key reasons[14] for the development of CMAP. We are aware of only two studies that compare C22 with C22/CMAP for helical peptides.[62,63] In ref 62, an alanine-rich peptide with 15 residues was studied by replica exchange MD (8 replicas from 250 K to 450 K, 1.4 ns) using a (generalized Born) continuum model for the solvent. In ref 63, conventional MD simulations (up to 5 ns) of two alanine-rich peptides with 15 and 20 residues, respectively, were performed using explicit solvent. Both studies showed that the peptides prefer π-helical conformations when treated with C22 and α-helical conformations when simulated with C22/CMAP. The authors conclude that the C22 force field is significantly improved by CMAP.

4.1. Removal of a C22 Artifact.

All these results convinced us that the π-helical propensity predicted for peptide P_E by Figures 4 and 17 (SI) must be an artifact of the C22 force field. Therefore we implemented the CMAP extension into our parallelized MD program package EGO-MMVI and checked how it modifies the description of peptide P_E. For this purpose we simply repeated simulation $P_E.1$ now using C22/CMAP.

Figure 5 shows the trajectories of helix fractions resulting from the corresponding simulation $P_E.3$. The 5 trajectories exhibit now a rock-solid α-helix over the whole time span of 10 ns moving in the water like a rigid rod. The π-helix fraction is zero everywhere. Thus, the inclusion of the CMAP has completely removed the π-helix artifact of C22.

4.2. Yet Another Artifact.

However, now the important question arises whether the shown rock-solid α-helix is actually a property of the petide P_E in water at 300 K and ambient pressure. This question can be answered experimentally because the α-helix content of P_E in aqueous solution can be characterized by circular dichroism (CD) measurements.

Figure 6 shows that the peptide P_E in water has at 278 K an α-helical content of only about 40% (see Table 4). At these conditions the α-helical content of the experimental model P_{pHP^*} for the open form of the light-switchable peptide P_{pHP} is even lower ($\sim$ 30 %). Thus the original concept of employing this small light-switchable peptide for experimental studies of α-helix folding will not work due to low helix content.

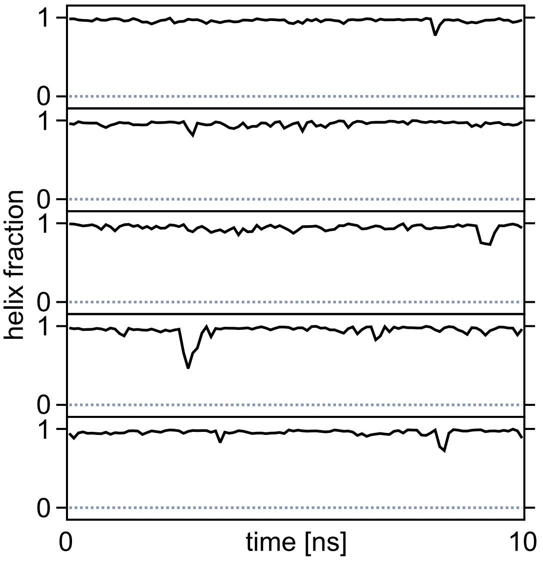


Figure 5. Trajectories of DSSP helix fractions resulting from simulation $\mathrm{P_E}$.3 which differs from $\mathrm{P_E}$.1 (Figure 4) solely by the CMAP extension to the C22 force field; the linestyles are explained in the caption to Figure 3.

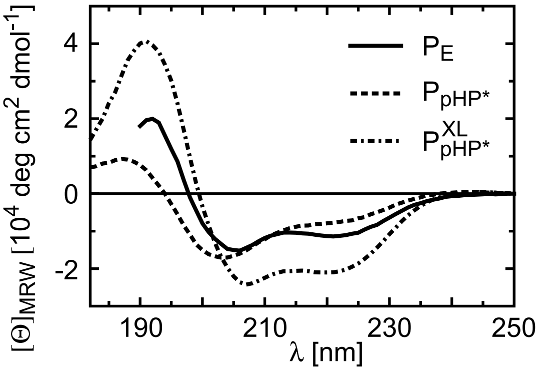


Figure 6. CD spectra of the peptides $\mathrm{P_E}$, $\mathrm{P_{pHP*}}$, and $\mathrm{P^{XL}_{pHP*}}$ at 278 K and at pH 7.

The elongated peptide $\mathrm{P^{XL}_{pHP*}}$ exhibits a somewhat larger α-helical content of about 60 % suggesting that the midpoint of the folding-unfolding transition is above 278 K. Although an α-helical content of 60 % signals some disorder it should suffice for experimental studies of light-triggered folding.

As a result of the rather small (40 %) helix content measured at 278 K for the peptide $\mathrm{P_E}$ in water, the apparent prediction of Figure 5 that this peptide is a rock-solid α-helix in water at 300 K seems to be somewhat off. Now there are two possible sources of this discrepancy between the simulation $\mathrm{P_E}$.3 and the CD data. As always, the discrepancy may either be caused by the applied C22/CMAP force field or by insufficient sampling. In the latter case, the stable α-helical state displayed by Figure 5 would represent a metastable state separated by a high free-energy barrier from the nearly unfolded native state. This issue can be computationally checked. Extended sampling methods of the replica exchange type are ideally suited for an answer.[64,65] Therefore we have performed the REST simulation $\mathrm{P_E}$.4.

TABLE 4: Comparison of Helix Contents from CD Data

peptide	$[\Theta]^{222}_{\mathrm{MRW}}$ [deg cm^2 dmol^{-1}]	f_{H}[a] [%]	Agadir[b] [%]
$\mathrm{P_E}$	-11259[c]	39 ± 8	45 ± 5
$\mathrm{P_{pHP*}}$	-7541	28 ± 8	36 ± 5
$\mathrm{P^{XL}_{pHP*}}$	-20691[c]	61 ± 9	67 ± 5

[a] Calculated according to the prescription given in Methods assuming a 5% error for measured data. [b] Calculated by the Agadir algorithm.[36,37] [c] For calculation of $[\Theta]_{\mathrm{MRW}}$ values the C-terminal GY residues were not included.

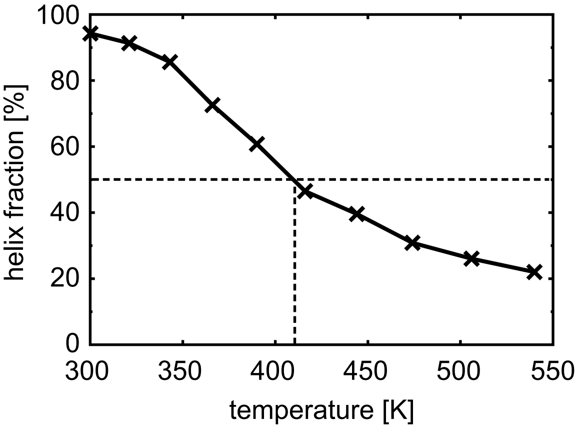


Figure 7. Temperature dependence of average DSSP helix fractions $\langle H_\alpha \rangle$ extracted from the REST simulation $\mathrm{P_E}$.4. The depicted values $\langle H_\alpha \rangle (T)$ are averages over the last 15 ns of the simulation sampling the replica at the temperature T and over the residues in the central part of peptide $\mathrm{P_E}$ ranging from the second alanine to the first lysine in the sequence (cf. Table 1). The dashed lines indicate the midpoint of the sigmoidal curve.

Figure 7 shows the temperature dependence of average DSSP helix fractions $\langle H_\alpha \rangle$ determined from the last 15 ns of the REST simulation $\mathrm{P_E}$.4. At the target temperature (300 K) of the REST sampling the helix content of $\mathrm{P_E}$ is 94 % in complete agreement with simulation $\mathrm{P_E}$.3 which indicated that the C22/CMAP model of $\mathrm{P_E}$ is a rock-solid α-helix in TIP3P water at 300 K. With increasing temperature the replicas are seen to have successively smaller helical contents. This REST result suggests that the rigid α-helical rod described by the conventional 300 K MD simulation $\mathrm{P_E}$.3 is not an artifact of insufficient sampling but is the equilibrium state of our C22/CMAP model for the peptide $\mathrm{P_E}$ at the given conditions.

Thus C22/CMAP, while successfully removing the π-helix artifact, apparently has introduced a new one, i.e. an overstabilization of α-helical structures for the given peptide sequence. If generalized to other sequences this finding may explain the α-helical conformations found in the extended C22/CMAP folding simulation[5] of a β-sheet protein quoted in the Introduction. Note here that recent C22/CMAP simulations of α-helical peptides, in which the generalized Born approximation was applied to model the surrounding dielectric solvent continua,[66] led the authors to the vague suspicion that "helical structures are slightly too favorable compared to experimental data, although further experimental and computational studies are needed to more clearly address this issue". Furthermore also a C22/CMAP MD study of dipeptides in explicit water suggested "a slight bias toward α-helical conformations".[67]

The question then is: How serious is this force field artifact for peptide simulations, i.e. does it mean that C22/CMAP simulations of α-helical peptides cannot give any insights into the processes of peptide folding?

In the following we will argue that the situation is not as bad as one might think. To elaborate these arguments reconsider the dependence of the average helix fraction $\langle H_\alpha \rangle$ on the temperature T of the REST replica shown in Figure 7. This dependence somehow resembles a melting curve of an α-helical peptide. In fact, if we had used a sufficiently extended TREM sampling instead of REST, $\langle H_\alpha \rangle(T)$ would actually be a melting curve. By comparing a computed with an experimental melting curve one could then find out, how the energy function used in the MD simulation must be scaled for an optimal match. For instance, in Figure 7 the transition temperature of helix melting lies at 410 K (cf. the dashed lines) whereas for peptide $\mathrm{P_E}$ in water we would estimate from the experimental data in Table 4 a transition near 270 K indicating that the forces stabilizing the α-helical conformation are overestimated by a factor of 1.5. A corresponding scaling of the peptide force field, particularly of its electrostatics, could then shift the transition temperature of our $\mathrm{P_E}$ model to that of $\mathrm{P_E}$.

However, we did not employ TREM but instead the more efficient REST sampling for the computation of $\langle H_\alpha \rangle(T)$ raising the question to what extent the curve in Figure 7 approximates the melting curve despite the REST energy scaling [cf. eq (2)]. This question has two aspects: (i) how much does the scaled interaction with and within the solvent modify the peptide ensemble at the given REST temperature, and (ii) to what extent are the REST ensembles actually converged. To address these questions, the rather large peptide $\mathrm{P_E}$ would have been a bad choice because of the enhanced computational cost associated with this peptide. Much better suited are here the smaller peptides $\tilde{\mathrm{P}}$ and $\mathrm{P_{pHP}}$.

4.3. Convergence of REST Simulations. Turning first to the issue of convergence we have calculated two REST melting curves $\langle H_\alpha \rangle(T)$ for the peptide $\tilde{\mathrm{P}}$ using different initial conditions. In the REST simulation $\tilde{\mathrm{P}}$.4 the initial structures of $\tilde{\mathrm{P}}$ were α-helical corresponding to $H_\alpha(T) = 1$ for all replica temperatures T whereas the simulation $\tilde{\mathrm{P}}$.5 started at unfolded structures corresponding to $H_\alpha(T) = 0$. With these two sets of initial conditions the approach toward the folding-unfolding equilibrium can be monitored by the mutual approach of the computed curves $\langle H_\alpha \rangle(T)$. If they match, the equilibrium has been reached and has been sufficiently sampled. Otherwise, the remaining differences can indicate the degree of convergence and the quality of the statistics.

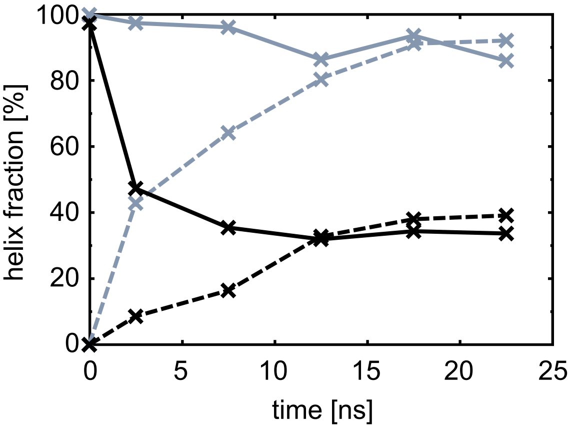


Figure 8. DSSP helix fractions $\langle H_\alpha \rangle$ averaged over 5 ns periods within the two 25 ns REST simulations $\tilde{\mathrm{P}}$.4 (solid curves) and $\tilde{\mathrm{P}}$.5 (dashed curves) and over the residues of $\tilde{\mathrm{P}}$. $\tilde{\mathrm{P}}$.4 started from α-helical, and $\tilde{\mathrm{P}}$.5 from unfolded structures. The gray curves refer to $\langle H_\alpha \rangle$ measured for the replica at the target temperature of 300 K. The black curves are the overall $\langle H_\alpha \rangle$ in the two generalized REST ensembles made up of all replicas T.

Figure 8 demonstrates that the REST simulations $\tilde{\mathrm{P}}$.4 and $\tilde{\mathrm{P}}$.5 are actually well converged both at the target temperature of 300 K (gray curves) and in the generalized ensemble (black curves). The convergence is proven by the fact that after 25 ns the helical content has become independent of the initial conditions in both ensembles. Apparently, the convergence to the respective equilibria has already been reached after about 10 ns. As demonstrated by Figure 19 in the SI, which displays the results of the analogous simulations $\mathrm{P_{pHP}}$.2 and $\mathrm{P_{pHP}}$.3, a similarly rapid convergence is obtained for the open form of the light-switchable peptide $\mathrm{P_{pHP}}$. Therefore, averages over the last 15 ns of our 25 ns REST simulations should represent equilibrium properties. Such averages are, of course, still subject to certain statistical fluctuations which could be diminished only upon a substantial extension of the sampling period starting after the initial 10 ns period of relaxation. Within the generalized REST ensemble an important equilibrium quantity is the temperature dependence $\langle H_\alpha \rangle(T)$ of the average helix fraction.

Figure 9 shows the equilibrium temperature dependences $\langle H_\alpha \rangle(T)$ of the α-helix fractions calculated from the simulations $\tilde{\mathrm{P}}$.4 and $\tilde{\mathrm{P}}$.5 together with the associated average. Interestingly, the dashed curve, which belongs to the simulation $\tilde{\mathrm{P}}$.4 featuring α-helical intial structures, is found nearly everywhere below the dot-dashed curve belonging to the unfolded starting structures. The remaining small deviations thus indicate the considerable quality of the underlying statistics. The average curve $\langle H_\alpha \rangle(T)$ should be associated with an even smaller statistical uncertainty.

4.4. Not all Peptides are Rods with C22/CMAP. According to Figure 9 the simulations $\tilde{\mathrm{P}}$.4/$\tilde{\mathrm{P}}$.5 predict for our peptide model $\tilde{\mathrm{P}}$ at 300 K an α-helix content of 89 %. Thus, our C22/CMAP model of peptide $\tilde{\mathrm{P}}$ seems to be a little less rod-like than the corresponding model of peptide $\mathrm{P_E}$ (94 %) discussed above in connection with Figures 5 and 7. This ex-

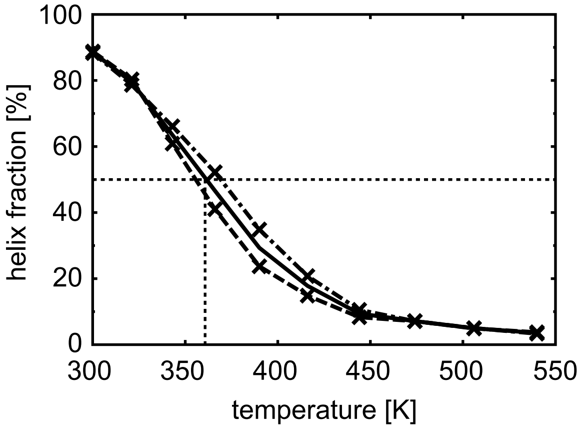


Figure 9. Equilibrium DSSP helix fractions $\langle H_\alpha \rangle(T)$ averaged over the last 15 ns of the two 25 ns REST simulations $\tilde{\mathrm{P}}$.4 (dashed curve) and $\tilde{\mathrm{P}}$.5 (dot-dashed curve) and over the residues of $\tilde{\mathrm{P}}$. Also given is the average of the two results (solid curve).

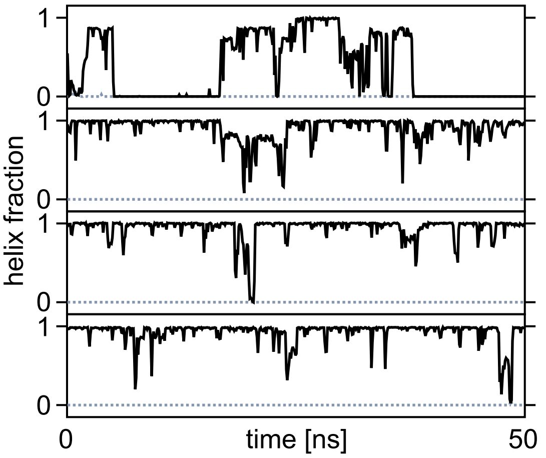


Figure 10. Trajectories $H_{\alpha/\pi}(t)$ of DSSP helix fractions resulting from the 300 K simulation $\tilde{\mathrm{P}}$.2 which differs from simulation $\tilde{\mathrm{P}}$.1 (Figure 3) solely by the CMAP extension to the C22 force field (cf. Figure 3 for the linestyles).

pectation is confirmed by Figure 10 which depicts the trajectories $H_\alpha(t)$ (solid) and $H_\pi(t)$ (dotted) resulting from simulation $\tilde{\mathrm{P}}$.2. This simulation employed the same initial conditions as $\tilde{\mathrm{P}}$.1 (Figure 3) but applied the C22/CMAP instead of the C22 force field to the peptide. The last three trajectories of $\tilde{\mathrm{P}}$.2 are similar: the initial α-helical conformation of the peptide is essentially maintained during the 50 ns covered by these simulations. The frequent fluctuations indicate short and incomplete events of unfolding which are immediately followed by refolding. According to the first trajectory, however, the peptide instantly unfolds at the beginning of the simulation, refolds after 1 ns, unfolds again at 5 ns, and stays unfolded until it refolds again (17 ns). Fluctuations of the α-helical content similar to those in the other three trajectories are then observed for 21 ns. Finally, the peptide unfolds and stays unfolded until the end of the simulation. All in all, the helix content of $\tilde{\mathrm{P}}$ (Figure 10) is seen to fluctuate much more than that of $\mathrm{P_E}$ (Figure 5) confirming the conclusion from Figure 9 that not every peptide is a rod-like α-helix when simulated with C22/CMAP at 300 K.

4.5. Does REST Yield Melting Curves? In the context of the temperature dependent helix content computed by REST for the peptide $\mathrm{P_E}$ (cf. Figure 7) we raised two questions. The question addressing the convergence of our REST simulations has been meanwhile answered by showing that the simulations $\tilde{\mathrm{P}}$.4/$\tilde{\mathrm{P}}$.5 (like the analogous simulations $\mathrm{P_{pHP}}$.2/$\mathrm{P_{pHP}}$.3) are well converged (section 4.3). Therefore, these simulations are ideally suited to address the other question, namely, how the energy scaling underlying REST influences the peptide ensembles at the various temperatures. Because the simulations $\tilde{\mathrm{P}}$.4/$\tilde{\mathrm{P}}$.5 have accurately determined the equilibrium α-helix content within the REST ensemble of $\tilde{\mathrm{P}}$ we simply have to check whether a TREM simulation yields a similar dependence. If this should turn out to be the case, the REST curves $\langle H_\alpha \rangle(T)$ would represent approximate melting curves.

For a corresponding check we performed simulation $\tilde{\mathrm{P}}$.6 with the TREM/REST method introduced in section 3.5. Such simulations efficiently sample a set of ensembles in a given temperature range (here: [340.0 K; 383.2 K]) with TREM while further speeding up the sampling by the inclusion of a few REST replicas simulated within a range of much more elevated temperatures (here: [383.2 K; 496 K]). In $\tilde{\mathrm{P}}$.6 the TREM temperature range is roughly centered around the unfolding temperature of peptide $\tilde{\mathrm{P}}$ identified by REST (360 K, cf. Figure 9). Because peptide $\tilde{\mathrm{P}}$ had shown a rapid convergence toward equilibrium in the REST setting, we did not execute two simulations with opposite initial conditions in the related TREM/REST approach. Instead, as starting structures we used peptide/water configurations drawn from the ensembles generated by the REST simulation $\tilde{\mathrm{P}}$.4 (for details see Section 3.6).

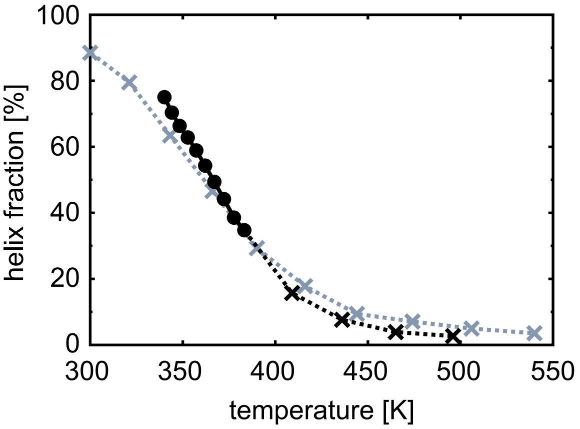


Figure 11. Temperature dependence of the equilibrium DSSP helix fractions $\langle H_\alpha \rangle(T)$ resulting from the TREM/REST simulation $\tilde{\mathrm{P}}$.6 (black) and the REST simulations $\tilde{\mathrm{P}}$.4/$\tilde{\mathrm{P}}$.5 (gray). The values for TREM replicas are indicated by dots, those for the REST replicas by crosses.

Figure 11 compares the TREM/REST melting curve from simulation $\tilde{\mathrm{P}}$.6 with the REST melting curve obtained from

$\tilde{\mathrm{P}}$.4/$\tilde{\mathrm{P}}$.5 (Figure 9). Despite the scaling of solvent-solvent and peptide-solvent interactions applied in REST, the curves almost match. This finding indicates that the REST energy scaling has only a quite small influence on the helical content of $\tilde{\mathrm{P}}$. Although the similarity of the $\langle H_\alpha \rangle(T)$ curves resulting from TREM and REST, respectively, was solely shown for our peptide model $\tilde{\mathrm{P}}$ we expect that it can be generalized to other (helical) peptides. Therefore, if one wants to efficiently calculate an approximate melting curve of an α-helical peptide over an extended temperature range, one can use REST instead of the much more costly TREM sampling.

Note here that simulations $\tilde{\mathrm{P}}$.4/$\tilde{\mathrm{P}}$.5 and $\tilde{\mathrm{P}}$.6 nicely illustrate the efficiency gain of REST vs. TREM. In $\tilde{\mathrm{P}}$.4/$\tilde{\mathrm{P}}$.5 a temperature range of 240 K is covered by 10 replicas whereas in the TREM part of $\tilde{\mathrm{P}}$.6 the same number of replicas covers only a range of 43 K. Because both setups operate with comparable exchange probabilities, the REST sampling is nearly six times more efficient than TREM for the given system.

4.6. Enhanced Sampling is Mandatory. The above results have shown that REST is a powerful sampling technique for α-helical peptides. Besides thoroughly characterizing the conformational ensemble of such a peptide at the target temperature (in our case 300 K), the method additionally provides an approximate thermal melting curve and transition temperature for helix unfolding that can be compared with experimental data. In this way we derived transition temperatures of about 360 K for our model peptide $\tilde{\mathrm{P}}$ and of about 410 K for peptide $\mathrm{P_E}$. Although the latter transition temperature lies well above the experimental value ($<$ 278 K, cf. Figure 6) our C22/CMAP simulations nevertheless describe, in qualitative agreement with the actual peptides, a temperature driven transition from an α-helical to an unfolded state. The older C22 force field, in contrast, would describe (e.g. in the case of $\mathrm{P_E}$) the unfolding of a qualitatively different π-helix. Furthermore, the overestimate of the transition temperatures connected with C22/CMAP should be repairable by globally scaling the peptide force field. Thus, as far as the equilibrium properties of α-helical peptides are concerned, the development of C22/CMAP has brought MD simulations matching the experimentally observed properties within reach. This conclusion holds to the extent that also the sampling problem is satisfactorily solved.

Addressing the sampling problem we will now argue that even quite extended MD simulations using the conventional setup may lead to erroneous conclusions on the equilibrium properties of a helical peptide in water at 300 K, because the associated statistical sampling can show a very slow convergence. Therefore, the application of an enhanced sampling technique such as REST is mandatory for characterizing the equilibrium properties of such a peptide with a reliable statistics.

The above claim is substantiated by comparing the results displayed in Figure 12 with data presented further above. The figure shows several extended (50 ns) trajectories $H_\alpha(t)$ of the helix fraction which were obtained from the simple 300 K MD simulations $\mathrm{P^{XL}_{pHP}}$.1 and $\mathrm{P_{pHP}}$.1.

A visual comparison of Figure 12A with Figure 5, for instance, would suggest that the MM models of the peptides $\mathrm{P^{XL}_{pHP}}$ and $\mathrm{P_E}$ are likewise rock-solid α-helical rods in TIP3P water at 300 K. In contrast, the α-helical contents derived from the 278 K CD data in Figure 6 demonstrate that the experimental model $\mathrm{P^{XL}_{pHP*}}$ of $\mathrm{P^{XL}_{pHP}}$ is somewhat more stable (61 %) than the quite disordered (39 %) peptide $\mathrm{P_E}$. Thus one would expect a similar difference also for the respective models. Unfortunately, we did not have the computational resources to compute safely converged REST melting curves for the large peptides $\mathrm{P^{XL}_{pHP}}$ and $\mathrm{P_E}$. For $\mathrm{P_E}$ we obtained only a single $\langle H_\alpha \rangle(T)$ curve which was based on α-helical initial conditions (Figure 7). Here, a convergence proof would have required an additional very costly REST simulation starting from an unfolded ensemble. For $\mathrm{P^{XL}_{pHP}}$ we could not generate any such data because of computational limitations. However, our resources did suffice to study the predictive power of quite extended conventional 300 K MD simulations for the smaller peptides $\mathrm{P_{pHP}}$ and $\tilde{\mathrm{P}}$.

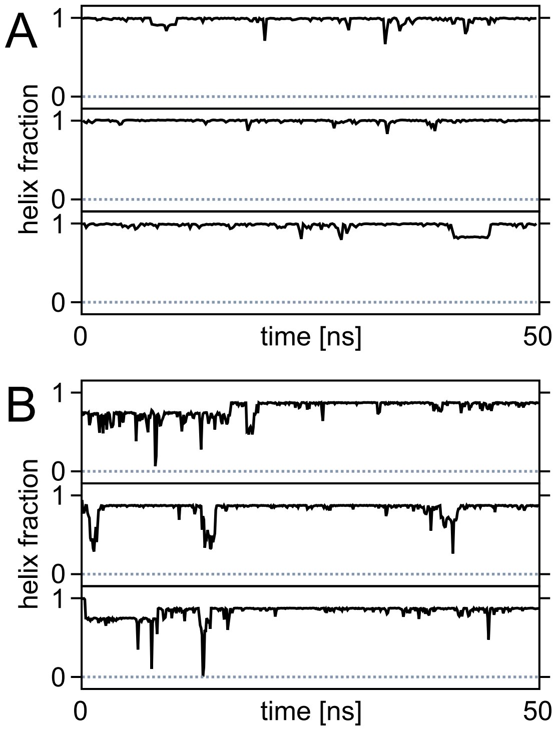


Figure 12. Trajectories $H_{\alpha/\pi}(t)$ of DSSP helix fractions resulting from the 300 K simulations $\mathrm{P^{XL}_{pHP}}$.1 (A) and $\mathrm{P_{pHP}}$.1 (B) (cf. Figure 3 for the linestyles).

Here we compare first Figure 12B with Figure 10. The obvious similarity of the $H_\alpha(t)$ trajectories seems to imply that the model peptides $\mathrm{P_{pHP}}$ and $\tilde{\mathrm{P}}$ are similarly stable in TIP3P water at 300 K. Due to the lack of CD data on $\tilde{\mathrm{P}}$ we have to resort to REST simulations for checking this prediction of straight forward MD.

Figure 13 shows the REST melting curve $\langle H_\alpha \rangle(T)$ obtained for the peptide $\mathrm{P_{pHP}}$ predicting that the α-helical content is 51 % at 300 K. The REST melting curve in Figure 9, however, assigns at 300 K a much larger α-helical content of 89 % to the peptide $\tilde{\mathrm{P}}$ despite the noted similarity of the 300 K helix fraction trajectories. Correspondingly, the helix transition of $\mathrm{P_{pHP}}$ is predicted at a much lower temperature

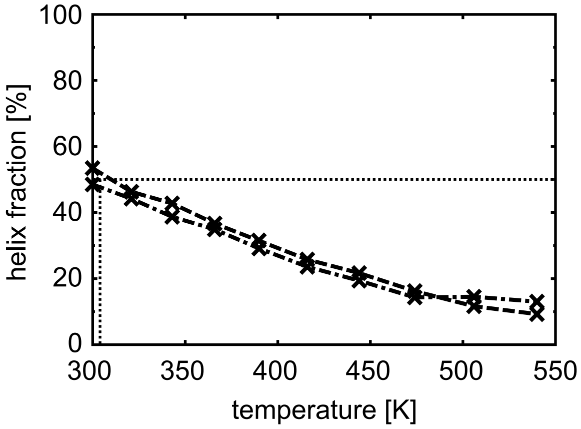


Figure 13. DSSP helix fractions $\langle H_\alpha \rangle (T)$ averaged over the last 15 ns of the two 25 ns REST simulations P_{pHP}.2 (dashed curve) and P_{pHP}.3 (dot-dashed curve) and over the residues of P_{pHP}. P_{pHP}.2 started from α-helical, and P_{pHP}.3 from unfolded structures.

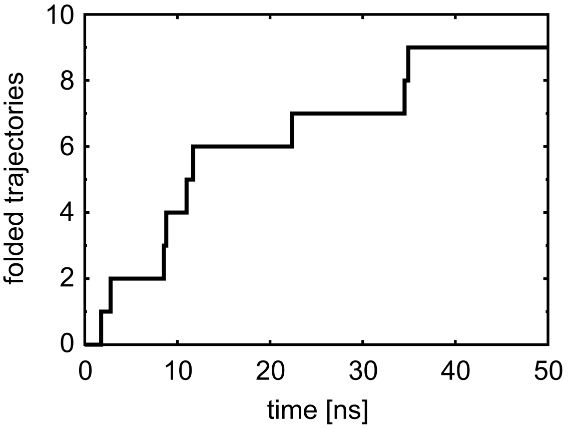


Figure 14. Number of trajectories that show a folding event up to the respective moment of simulation $\tilde{P}$.3.

(300 K) than that of $\tilde{P}$ (360 K). Note here that the 300 K transition temperature calculated for the peptide P_{pHP} represents a second example for the overstabilization of α-helical structures induced by the C22/CMAP force field, because the experimental transition temperature must be much smaller than 278 K. This conclusion derives from the fact that at 278 K the helix content of the peptide $P_{\mathrm{pHP^*}}$ is only about 28 % (cf. Table 4).

The wrong conclusion on a seemingly comparable α-helical stability of the peptides P_{pHP} and $\tilde{P}$, which was derived from several extended (50 ns) MD simulations at 300 K, is most likely caused by a quite small unfolding rate of the peptide P_{pHP} at this temperature. As a result, P_{pHP} happens to largely maintain its metastable initial conformation in all three trajectories displayed by Figure 12B without showing major events of α-helix unfolding. Note that slower folding-unfolding rates of P_{pHP} may also explain the enhanced fluctuations visible in Figure 19 (SI) which, by the way, documents the excellent convergence of the P_{pHP} melting curves in Figure 13.

The stated failure of the extended 300 K simulations in Figure 12B to show the unfolding of the petide P_{pHP} from the initial α-helical state into a partially unfolded ensemble raises the question whether the 50 ns time scale, which is accessible by standard MD simulations, can suffice to simulate the reverse process of α-helix folding.

4.7. On the Timescale of Helix Folding. To scrutinize this issue, we returned to our model peptide $\tilde{P}$ and performed the simulation $\tilde{P}$.3 (cf. Table 3). In this simulation, ten conventional MD trajectories were calculated for the peptide $\tilde{P}$ at 300 K starting from unfolded peptide structures (the same structures that were used as starting structures in simulation $\tilde{P}$.5).

Figure 14 shows in how many of the ten trajectories of simulation $\tilde{P}$.3 at least one folding event was observed up to a given moment. We consider a peptide to be folded if more than half of its residues are classified by DSSP to be α-helical. In the given case of peptide $\tilde{P}$ we thus count a folding event if at least five residues are α-helical. According to Figure 14, in five out of the ten trajectories at least one folding event has occured after 11 ns.

The value of 11 ns characterizes the folding time of our C22/CMAP model for $\tilde{P}$ and falls well within the time spans currently accessible to MD simulations. For a statistically sound characterization of the various folding pathways, however, a few hundred of such simulations will be required starting from initial structures representative for the equilibrium ensemble. Furthermore, the peptide $\tilde{P}$ studied by us may have been a particularly rapidly folding peptide. Therefore, the non-equilibrium processes of α-helical peptide folding, at which our project initially was aiming, pose a huge computational task (currently exceeding our limited resources) although they are meanwhile in principle accessible to MD descriptions.

5. Conclusion

In the Introduction we identified the problems of sufficient statistical sampling and force field accuracy as the two main obstacles toward physically realistic simulations of folding-unfolding equilibria as well as of light-triggered folding processes. Whenever, for reasons of accuracy, an all-atom solvent model is used in the simulations, the sampling issue becomes a challenge. As shown by our results, for the folding-unfolding equilibria of α-helical peptides this challenge can be met if the REST technique is applied. Then this technique yields approximate melting curves and enables clear-cut convergence checks by utilizing opposite, i.e. folded and unfolded, initial conditions.

Once sufficient sampling is guaranteed also the problems of the force field accuracy can be tackled. Using REST we thus were able to show that the CMAP extension to the C22 force field represents a major progress toward an empirical force field which is accurate enough to realistically describe the processes of α-helix folding. After all, using the C22/CMAP force field in REST simulations of three differently stable peptides we obtained melting curves exhibiting the typical shapes known for α-helix folding-unfolding equilibria. In contrast, the predecessor force field assigned π-helical or unfolded structures to these peptides. However, comparing for two of these peptides experimentally measured α-helix contents with corresponding REST results revealed a remaining deficiency of the C22/CMAP force field. Confirming earlier suspicions[66,67] we have now clearly demonstrated that C22/CMAP sizeably overestimates the stability of α-helical structures. As a result, for a particular α-helical peptide covering 13 residues (i.e. peptide P_{E} in Table 1) the midpoint of the thermal unfolding transition in aqueous solution is calculated at the elevated temperature of 410 K whereas experimentally it is found below 278 K.

Initially this project had been motivated by our desire to bridge the gap between experimental and simulation approaches toward the understanding of peptide folding. For this purpose we tried to design minimal molcular models that should enable the light-triggered initialization of α-helix folding processes. Although we were able to show that such a bridging is now within reach we did not quite accomplish our goal for two reasons.

First, the designed molecular models turned out to be largely unfolded as long as they covered only about 11 residues. Larger peptides covering 15 residues started to show some α-helical structure, were however too large for thorough simulation descriptions because of the limited computational resources accessible to us.

Secondly, we had to scrutinize the sampling and force field issues before a simulation of folding processes deemed reasonable. The results then provided a clear insight into the additional steps that have to be taken and the computational effort that has to be spent for physically realistic descriptions. With enhanced computational resources such projects can now be tackled.

Concluding we would like to stress that simulations of light-triggered peptide folding are currently within reach solely if these peptides fold into α-helices. Only then the folding-unfolding equilibria can be efficiently sampled with rapid techniques like REST, whereas for β-hairpins such techniques are likely to fail.[30,65] Furthermore, only for α-helices the non-equilibrium folding processes will be rapid enough, i.e. proceed on sub-microsecond time-scales, for a statistically sufficient sampling of the folding pathways.

Acknowledgment. This work was supported by the VolkswagenStiftung (I/79 884). RR and PT would like to thank Robert Denschlag and Martin Lingenheil for countless fruitful discussions on the intricacies of simulation methods and for technical support in the setup of the REST simulations.

Supporting Information Available: The SI contains the details of the MM force field for the phototrigger pHP in form of figures and tables and presents some additional results.

References

1. Lei, H.; Duan, Y. *J. Mol. Biol.* **2007,** *370,* 196-206.
2. Lei, H.; Duan, Y. *J. Phys. Chem. B* **2007,** *111,* 5458-5463.
3. Duan, Y.; Kollman, P. A. *Science* **1998,** *282,* 740-744.
4. Snow, C. D.; Nguyen, H.; Pande, V. S.; Gruebele, M. *Nature* **2002,** *420,* 102-106.
5. Freddolino, P. L.; Liu, F.; Gruebele, M.; Schulten, K. *Biophys. J.* **2008,** *94,* L75-L77.
6. Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. *J. Am. Chem. Soc.* **1990,** *112,* 6127-6129.
7. Nymeyer, H.; Garcia, A. E. *Proc. Natl. Acad. Sci. U.S.A.* **2003,** *100,* 13934-13939.
8. Zhou, R. H.; Berne, B. J. *Proc. Natl. Acad. Sci. U.S.A.* **2002,** *99,* 12777-12782.
9. Zhou, R. H. *Proteins* **2003,** *53,* 148-161.
10. Wroblewska, L.; Skolnick, J. *J. Comput. Chem.* **2007,** *28,* 2059-2066.
11. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, Jr., K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995,** *117,* 5179-5197.
12. MacKerell, Jr., A. D. *et al. J. Phys. Chem. B* **1998,** *102,* 3586-3616.
13. MacKerell, Jr., A. D.; Feig, M.; Brooks, III, C. L. *J. Am. Chem. Soc.* **2004,** *126,* 698-699.
14. MacKerell, Jr., A. D.; Feig, M.; Brooks, III, C. L. *J. Comput. Chem.* **2004,** *25,* 1400-1415.
15. Spörlein, S.; Carstens, H.; Satzger, H.; Renner, C.; Behrendt, R.; Moroder, L.; Tavan, P.; Zinth, W.; Wachtveitl, J. *Proc. Natl. Acad. Sci. U.S.A.* **2002,** *99,* 7998-8002.
16. Carstens, H.; Renner, C.; Milbradt, A. G.; Moroder, L.; Tavan, P. *Biochemistry* **2005,** *44,* 4829-4840.
17. Schrader, T. E.; Schreier, W. J.; Cordes, T.; Koller, F. O.; Babitzki, G.; Denschlag, R.; Renner, C.; Dong, S.-L.; Löweneck, M.; Moroder, L.; Tavan, P.; Zinth, W. *Proc. Natl. Acad. Sci. U.S.A.* **2007,** *104,* 15729-15734.
18. Egwolf, B.; Tavan, P. *J. Chem. Phys.* **2003,** *118,* 2039-2056.
19. Tavan, P.; Carstens, H.; Mathias, G. Molecular dynamics simulations of proteins and peptides: Problems, achievements, and perspectives. In *Protein Folding Handbook. Part 1.*; Buchner, J.; Kiefhaber, T., Eds.; Wiley-VCH: Weinheim, 2005.
20. Stork, M.; Tavan, P. *J. Chem. Phys.* **2007,** *126,* 165105.
21. Stork, M.; Tavan, P. *J. Chem. Phys.* **2007,** *126,* 165106.
22. Volk, M.; Kholodenko, Y.; Lu, H. S. M.; Gooding, E. A.; DeGrado, W. F.; Hochstrasser, R. M. *J. Phys. Chem. B* **1997,** *101,* 8607-8616.
23. Bredenbeck, J.; Helbing, J.; Kumita, J. R.; Woolley, G. A.; Hamm, P. *Proc. Natl. Acad. Sci. U.S.A.* **2005,** *102,* 2379-2384.
24. Kolano, C.; Helbing, J.; Kozinski, M.; Sander, W.; Hamm, P. *Nature* **2006,** *444,* 469-472.
25. Bredenbeck, J.; Helbing, J.; Sieg, A.; Schrader, T.; Zinth, W.; Renner, C.; Behrendt, R.; Moroder, L.; Wachtveitl, J.; Hamm, P. *Proc. Natl. Acad. Sci. U.S.A.* **2003,** *100,* 6452-6457.
26. Chen, E. F.; Kumita, J. R.; Woolley, G. A.; Kliger, D. S. *J. Am. Chem. Soc.* **2003,** *125,* 12443-12449.
27. Hukushima, K.; Nemoto, K. *J. Phys. Soc. Jpn* **1996,** *65,* 1604-1608.
28. Hansmann, U. H. E. *Chem. Phys. Lett.* **1997,** *281,* 140-150.
29. Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **1999,** *314,* 141-151.
30. Denschlag, R.; Lingenheil, M.; Tavan, P. *Chem. Phys. Lett.* **2008,** *458,* 244-248.
31. Givens, R. S.; Weber, J. F. W.; Conrad, P. G.; Orosz, G.; Donahue, S. L.; Thayer, S. A. *J. Am. Chem. Soc.* **2000,** *122,* 2687-2697.
32. Conrad, P. G. I.; Givens, R. S.; Hellrung, B.; Rajesh, C. S.; Ramseier, M.; Wirz, J. *J. Am. Chem. Soc.* **2000,** *122,* 9346 - 9347.
33. Wirz, J. Unpublished data.
34. Chakrabartty, A.; Kortemme, T.; Baldwin, R. L. *Protein Sci.* **1994,** *3,* 843-852.
35. Rohl, C. A.; Chakrabartty, A.; Baldwin, R. L. *Protein Sci.* **1996,** *5,* 2623-2637.
36. Munoz, V.; Serrano, L. *Nature Struct. Biol.* **1994,** *1,* 399-409.
37. Munoz, V.; Serrano, L. *J. Mol. Biol.* **1995,** *245,* 275-296.
38. Shoemaker, K. R.; Kim, P. S.; York, E. J.; Stewart, J. M.; Baldwin, R. L. *Nature* **1987,** *326,* 563-567.
39. Scott, W. R. P.; Hünenberger, P. H.; Tironi, I. G.; Mark, A. E.; Billeter, S. R.; Fennen, J.; Torda, A. E.; Huber, T.; Krüger, P.; van Gunsteren, W. F. *J. Phys. Chem. A* **1999,** *103,* 3596-3607.
40. Ahlrichs, R.; Bär, M.; Häser, M.; Horn, H.; Kölmel, C. *Chem. Phys. Lett.* **1989,** *162,* 165-169.

41. Treutler, O.; Ahlrichs, R. *J. Chem. Phys.* **1995,** *102,* 346-354.

42. Perdew, J. P. *Phys. Rev. B* **1986,** *33,* 8822-8824.

43. Becke, A. D. *Phys. Rev. A* **1988,** *38,* 3098-3100.

44. Schäfer, A.; Huber, C.; Reinhart, A. *J. Chem. Phys.* **1994,** *100,* 5829-5835.

45. Schaftenaar, G.; Noordik, J. H. *J. Comput. Aid. Mol. Des.* **2000,** *14,* 123-134.

46. Durell, S. R.; Brooks, B. R.; Ben-Naim, A. *J. Phys. Chem.* **1994,** *98,* 2198-2202.

47. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Phys. Chem.* **1983,** *79,* 926-935.

48. Mathias, G.; Egwolf, B.; M., N.; P., T. *J. Chem. Phys.* **2003,** *118,* 10847-10860.

49. Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984,** *81,* 3684-3690.

50. Niedermeier, C.; Tavan, P. *J. Chem. Phys.* **1994,** *101,* 734-748.

51. Mathias, G.; Tavan, P. *J. Chem. Phys.* **2004,** *120,* 4393-4403.

52. Allen, M. P.; Tildesley, D. J. *Computer simulation of liquids;* Clarendon Press: Oxford, 1987.

53. Kräutler, V.; van Gunsteren, W. F.; Hünenberger, P. H. *J. Comput. Chem.* **2001,** *22,* 501-508.

54. Eisenberg, D.; Kauzmann, W. *The structure and properties of water;* Clarendon Press: Oxford, 1969.

55. Lingenheil, M.; Denschlag, R.; Reichold, R.; Tavan, P. *J. Chem. Theory Comput.* **2008,** *4,* 1293-1306.

56. Liu, P.; Kim, B.; Friesner, R. A.; Berne, B. J. *Proc. Natl. Acad. Sci. U.S.A.* **2005,** *102,* 13749-13754.

57. Garcia, A. E.; Sanbonmatsu, K. Y. *Proc. Natl. Acad. Sci. U.S.A.* **2002,** *99,* 2782-2787.

58. Kabsch, W.; Sander, C. *Biopolymers* **1983,** *22,* 2577–2637.

59. Luo, P.; Baldwin, R. L. *Biochemistry* **1997,** *36,* 8413-8421.

60. Weaver, T. M. *Protein Sci.* **2000,** *9,* 201-206.

61. Fodje, M. N.; Al-Karadaghi, S. *Protein Eng.* **2002,** *15,* 353-358.

62. Feig, M.; MacKerell, Jr., A. D.; Brooks, III, C. L. *J. Phys. Chem. B* **2003,** *107,* 2831-2836.

63. Freedberg, D. I.; Venable, R. M.; Rossi, A.; Bull, T. E.; Pastor, R. W. *J. Am. Chem. Soc.* **2004,** *126,* 10478-10484.

64. Periole, X.; Mark, A. E. *J. Chem. Phys.* **2007,** *126,* 014903.

65. Huang, X.; Hagen, M.; Kim, B.; Friesner, R. A.; Zhou, R.; Berne, B. J. *J. Phys. Chem. B* **2007,** *111,* 5405-5410.

66. Tanizaki, S.; Clifford, J.; Connelly, B. D.; Feig, M. *Biophys. J.* **2008,** *94,* 747-759.

67. Feig, M. *J. Chem. Theory Comput.* **2008,** *4,* 1555-1564.

Supporting information to the manuscript

Toward MD Descriptions of α-Helix Folding

R. Reichold[1], B. Fierz[2†], T. Kiefhaber[2], and P. Tavan[1*]

[1]Theoretische Biophysik, Lehrstuhl für Biomolekulare Optik
Ludwig-Maximilians-Universität
Oettingenstr. 67, 80538 München, Germany

[2]Lehrstuhl für Biophysikalische Chemie & Munich Center for Integrated Protein Science
TU München
Lichtenbergstr. 4, 85748 Garching, Germany

[†]present address: Laboratory of Synthetic Protein Chemistry,
The Rockefeller University, New York, New York 10065, USA.

[*]corresponding author, email: tavan@physik.uni-muenchen.de,
phone: +49-89-2180-9220, fax: +49-89-2180-9202

MM Force Field for pHP

Figures 15 and 16 show chemical structures of molecular models for the open and closed phototriggers pHPo and pHPc, respectively. The figures assign CHARMM-type names to the various atoms.

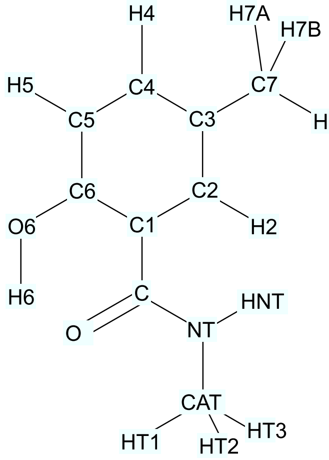

H4
H7A
H7B
H5
C4
C7
C5
C3
H7C
C6
C2
O6
C1
H2
H6
C
HNT
O
NT
CAT
HT3
HT1
HT2

Figure 15: Chemical structure of the molecule that was used to develop force field parameters for the open form of the photo-trigger pHP (pHPo).

In the case of pHPo, the structure drawn in Figure 15 is identical to the model compound A depicted by Figure 2 in the paper which was used for the quantum chemical calculation of a CHARMM-type force field.

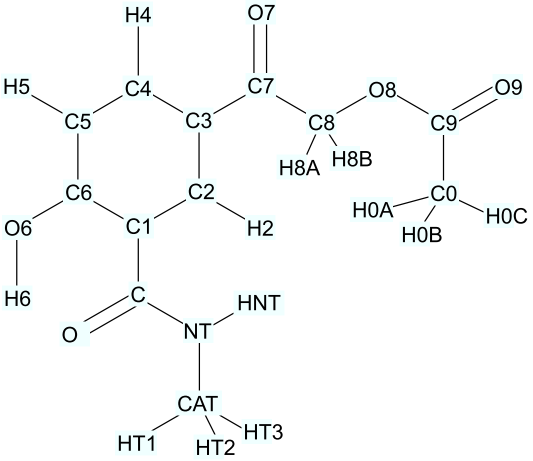


Figure 16: Chemical structure of the closed photo-trigger pHP (pHPc).

In contrast, a force field for pHPc was determined from quantum chemical treatments of the two model compounds B and C shown by Figure 2 in the paper, because these compounds represent the key substructures of the pHPc model drawn in Figure 16.

The following Tables 5 and 6 assign to each of the atom names introduced in Figures 15 and 16 an atom type and a partial charge. Some of the listed charges are results of the quantum chemical calculations described in the paper, others are adopted from C22. The charges adopted from C22 are marked by asterisks in the tables. Similarly, some of the listed atom types were newly introduced for the pHP compounds, whereas others represent types previously contained in C22. The latter types are also marked by asterisks.

Note that the two methyl groups around the carbon atom named CAT carry unsual atom types. Our choice was motivated by the fact that, after integration of the trigger molecules into a peptide, the carbon atom CAT becomes a C_α atom of the backbone which in C22 is of type CT1. Thus also the two atoms CAT are of this type. The types of the attached hydrogen atoms were chosen upon similar considerations.

Name	Type	Charge	Name	Type	Charge
C3	CCL	0.00	C7	CT3*	-0.27*
C2	CBL	-0.11	H7A	HA*	0.09*
H2	HPL	0.11	H7B	HA*	0.09*
C4	CCL	-0.11	H7C	HA*	0.09*
H4	HPL	0.11	C	CL	0.51*
C1	CCL	-0.03	O	OL	-0.51*
C5	CCL	-0.15	NT	NHL	-0.47*
H5	HPL	0.11	HNT	HL	0.31*
C6	CCL	0.08	CAT	CT1*	-0.11*
O6	OHL	-0.38	HT1	HB*	0.09*
H6	HL	0.37	HT2	HB*	0.09*
			HT3	HB*	0.09*

Table 5: Atom names, types, and partial charges used for pHPo.

Name	Type	Charge	Name	Type	Charge	Name	Type	Charge
C3	CCL	0.00	C7	CC1	0.51*	C	CL	0.51*
C2	CBL	-0.11	O7	OC1	-0.51*	O	OL	-0.51*
H2	HPL	0.11	C8	CC2	-0.09	NT	NHL	-0.47*
C4	CCL	-0.11	H8A	HA*	0.09	HNT	HL	0.31*
H4	HPL	0.11	H8B	HA*	0.09	CAT	CT1*	-0.11*
C1	CCL	-0.03	O8	OC2	-0.37	HT1	HB*	0.09*
C5	CCL	-0.15	C9	CC1	0.79	HT2	HB*	0.09*
H5	HPL	0.11	O9	OC1	-0.51	HT3	HB*	0.09*
C6	CCL	0.08	C0	CT2*	-0.27*			
O6	OHL	-0.38	H0A	HA*	0.09*			
H6	HL	0.37	H0B	HA*	0.09*			
			H0C	HA*	0.09*			

Table 6: Atom names, types, and partial charges used for pHPc.

pHP	C22	Description
HPL	HP	aromatic H
HL	H	polar H
CC1, CL	C	polar C
CC2	CT2	aliphatic sp^3 C for CH$_2$
CBL, CCL	CA	aromatic C
NHL	NH1	peptide nitrogen
OC1, OL	O	carbonyl oxygen
OHL	OH1	hydroxyl oxygen
OC2	OS	ester oxygen

Table 7: Atom types newly introduced for pHP (left column) are associated to standard C22 atom types (central column) for transfer of existing Lennard-Jones parameters to the new atom types. The Lennard-Jones parameters for the C22 atom types can be found in the Supporting Information to ref 1.

To select suitable Lennard-Jones potentials for the pHP atoms classified by newly introduced atom types we selected from the C22 data base similar chemical motifs and associated C22 atom types. We then assigned the Lennard-Jones parameters of these C22 atom types to atoms with newly introduced atom types. Table 7 lists these associations.

A1	A2	k_b	r_0	A1	A2	k_b	r_0
OL	CL	480.00	1.250	CCL	CCL	330.00	1.360
CCL	CBL	330.00	1.360	HPL	CCL	370.00	1.099
HPL	CBL	370.00	1.099	OHL	CCL	334.30	1.330
OHL	HL	260.00	1.019	CT3	CCL	230.00	1.500
NHL	CL	350.00	1.355	NHL	HL	440.00	1.023
NHL	CT1	320.00	1.450	CL	CCL	230.00	1.430
CT2	CCL	230.00	1.500	CCL	CC1	230.00	1.520
CC1	OC1	813.00	1.210	CC1	CC2	203.50	1.536
CC2	HA	330.00	1.110	CC2	OC2	255.00	1.410
OC2	CC1	273.70	1.370	CC1	CT2	241.30	1.510
CCL	HA	300.00	1.090				

Table 8: Harmonic bond stretching parameters (force constants k_b [kcal/(molÅ^2)] and equilibrium bondlengths r_0 [Å]) for the new atom types.

For the so-called bonded potentials covering harmonic functions for bond stretches, angle deformations, and improper dihedrals as well as truncated Fourier series for proper torsions we determined part of the required parameters through quantum chemistry. Others, which are marked by asterisks in the following tables 8, 9, 10, and 11, were adopted from analogous binding motifs contained in the C22 data base.

A1	A2	A3	A4	k_ϕ	ϕ_0
NHL	X	X	HL	15.00	0.00
OL	X	X	CL	90.00	0.00
CC1	X	X	OC1	96.00	0.00
CC1	X	X	OC2	96.00	0.00

Table 9: Parameters of harmonic potentials for improper dihedral angles involving atoms with new types (force constants k_ϕ [kcal/(mol rad^2)] and equilibrium angles ϕ_0 [deg]). X indicates a wildcard.

A1	A2	A3	k_θ	θ_0	A1	A2	A3	k_θ	θ_0
CCL	CCL	CCL	*40.00	*120.00	CT2	CCL	CBL	39.30	120.00
CBL	CCL	CCL	*40.00	*120.00	HA	CT2	CCL	*49.30	*107.50
CCL	CBL	CCL	*40.00	*120.00	CC	CT2	CCL	52.00	108.00
HPL	CCL	CCL	35.00	120.00	NHL	CT1	CT3	*70.00	*113.50
HPL	CBL	CCL	35.00	120.00	NHL	CT1	C	*50.00	*107.00
HPL	CCL	CBL	35.00	120.00	HA	CCL	HA	48.00	109.00
CT3	CCL	CCL	39.30	120.00	HA	CCL	CC1	42.00	109.00
CT3	CCL	CBL	39.30	120.00	CCL	CC1	CC2	48.50	114.00
HA	CT3	CCL	33.00	111.00	CCL	CC1	OC1	60.50	123.00
OHL	CCL	CCL	35.20	120.00	OC1	CC1	CC2	60.50	123.00
HL	OHL	CCL	63.60	105.60	CC1	CC2	OC2	68.59	110.00
CL	CCL	CCL	42.00	116.00	CC1	CC2	HA	35.00	109.00
CL	CCL	CBL	42.00	124.00	HA	CC2	HA	*35.50	*109.00
OL	CL	CCL	60.00	121.50	HA	CC2	OC2	35.00	109.00
NHL	CL	CCL	46.00	117.00	CC2	OC2	CC1	42.80	112.00
OL	CL	NHL	80.00	121.50	OC2	CC1	OC1	95.00	119.00
HL	NHL	CL	*34.00	*123.00	OC2	CC1	CT2	53.00	115.00
HL	NHL	CT1	*35.00	*117.00	OC1	CC1	CT2	53.00	126.00
CT1	NHL	CL	48.00	120.00	CC1	CT2	HA	42.00	109.00
NHL	CT1	HB	51.50	109.50	CBL	CCL	CC1	39.30	120.00
HB	CT1	HB	35.50	108.40	CCL	CCL	CC1	39.30	120.00
CT2	CCL	CCL	39.30	120.00	CC1	CT2	CT2	*52.00	*108.00

Table 10: Harmonic angle deformation parameters (force constants k_θ [kcal/(mol rad^2)] and equilibrium angles θ_0 [deg]) for the new atom types. Standard C22 values are marked by asterisks.

A1	A2	A3	A4	k_ϕ	n	δ	A1	A2	A3	A4	k_ϕ	n	δ
CCL	CCL	CCL	CCL	*3.10	*2	*180	HPL	CCL	CCL	CCL	2.20	2	180
CT3	CCL	CCL	CCL	2.00	2	180	CT3	CCL	CBL	CCL	2.00	2	180
CL	CCL	CCL	CCL	2.00	2	180	CL	CCL	CBL	CCL	2.00	2	180
OHL	CCL	CCL	CCL	*3.10	*2	*180	OHL	CCL	CCL	HPL	*4.20	*2	*180
OHL	CCL	CCL	CL	3.10	2	180	HPL	CCL	CCL	HPL	*2.40	*2	*180
HPL	CCL	CCL	CT3	*4.20	*2	*180	CCL	CBL	CCL	CCL	*3.10	*2	*180
CBL	CCL	CCL	CCL	*3.10	*2	*180	CBL	CCL	CCL	HPL	2.20	2	180
CBL	CCL	CCL	OHL	*3.10	*2	*180	HPL	CBL	CCL	CCL	2.20	2	180
HPL	CBL	CCL	CL	4.20	2	180	HPL	CBL	CCL	CT3	*4.20	*2	*180
HPL	CCL	CBL	CCL	2.20	2	180	HPL	CCL	CBL	HPL	*2.40	*2	*180
HL	OHL	CCL	CCL	0.60	2	180	HA	CT3	CCL	CCL	.03	3	0
HA	CT3	CCL	CBL	.03	3	180	OL	CL	CCL	CCL	1.20	2	180
OL	CL	CCL	CBL	1.20	2	180	NHL	CL	CCL	CCL	1.20	2	180
NHL	CL	CCL	CBL	1.20	2	180	HL	NHL	CL	CCL	1.00	2	180
CT1	NHL	CL	CCL	1.50	1	0	CT1	NHL	CL	CCL	0.60	2	180
OL	CL	NHL	CT1	0.60	2	180	OL	CL	NHL	HL	0.60	2	180
HB	CT1	NHL	CL	.12	3	180	HB	CT1	NHL	HL	.12	3	0
HL	NHL	CT1	CT3	*.00	*1	*0	HL	NHL	CT1	C	*.00	*1	*0
NHL	CT1	C	O	*.00	*1	*0	NHL	CT1	C	NH1	*.60	*1	*0
CL	NHL	CT1	CT3	*1.80	*1	*0	CL	NHL	CT1	C	*.20	*1	*180
CT2	CCL	CCL	HPL	*4.20	*2	*180	CT2	CCL	CCL	CCL	2.00	2	180
CT2	CCL	CBL	HPL	*4.20	*2	*180	CT2	CCL	CBL	CCL	2.00	2	180
HA	CT2	CCL	CCL	.03	3	0	HA	CT2	CCL	CBL	.03	3	180
CC	CT2	CCL	CCL	.03	3	0	CC	CT2	CCL	CBL	.03	3	180
HA	CCL	CC1	CC2	0.00	3	180	HA	CCL	CC1	OC1	0.00	3	0
CCL	CC1	CC2	OC2	1.50	1	180	CCL	CC1	CC2	OC2	0.55	2	180
CCL	CC1	CC2	HA	0.01	3	0	OC1	CC1	CC2	OC2	0.55	2	180
OC1	CC1	CC2	HA	0.01	3	180	CC1	CC2	OC2	CC1	1.80	2	0
HA	CC2	OC2	CC1	0.30	3	180	CC2	OC2	CC1	CT2	2.80	1	0
CC2	OC2	CC1	CT2	2.90	2	180	CC2	OC2	CC1	OC1	2.50	2	180
OC2	CC1	CT2	HA	0.01	3	0	OC1	CC1	CT2	HA	0.01	3	180
CCL	CBL	CCL	CC1	2.00	2	180	HPL	CBL	CCL	CC1	4.20	2	180
CCL	CCL	CCL	CC1	2.00	2	180	HPL	CCL	CCL	CC1	4.20	2	180
CBL	CCL	CC1	OC1	0.10	1	0	CBL	CCL	CC1	CC2	2.90	2	180
CCL	CCL	CC1	OC1	2.90	2	180	CCL	CCL	CC1	CC2	0.10	1	0
X	CT2	CC1	X	0.05	6	180							

Table 11: Parameters of dihedral potentials $E(\phi) = k_\phi\,[1 + \cos(n\phi + \delta)]$ for the new atom types (k_ϕ [kcal/mol], δ [deg]). Whenever atom type quadruples are listed twice, the Fourier expansion contains two contributions $E(\phi)$. Asterisks mark standard C22 values.

An Artifact of C22

The strange π-helical propensity of peptide P_E displayed by Figure 4 has led us to the question whether the model employed for the solvent can contribute to this artifact or whether it has to be exclusively attributed to C22 as claimed by refs 2–4. Simulation P_E.2 served us to address this question.

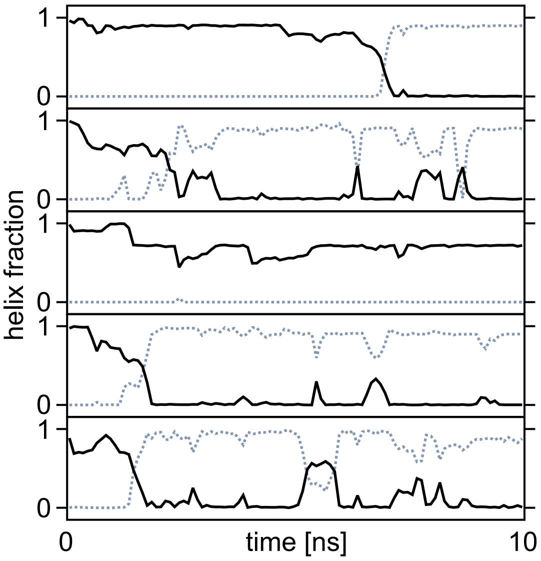


Figure 17: Trajectories of DSSP helix fractions resulting from simulation P_E.2 which differs from simulation P_E.1 (Figure 4) by choosing the SPC/E instead of the TIP3P model for the water molecules; linestyles are explained in the caption to Figure 3.

A visual comparison of Figure 17 with Figure 4 immediately demonstrates that the peptide P_E behaves identically for both solvent models and exhibits the same π-helical propensity.

Cost effective models for the peptides $\mathrm{P}_{\mathrm{pHP}}$ and $\mathrm{P}^{\mathrm{XL}}_{\mathrm{pHP}}$

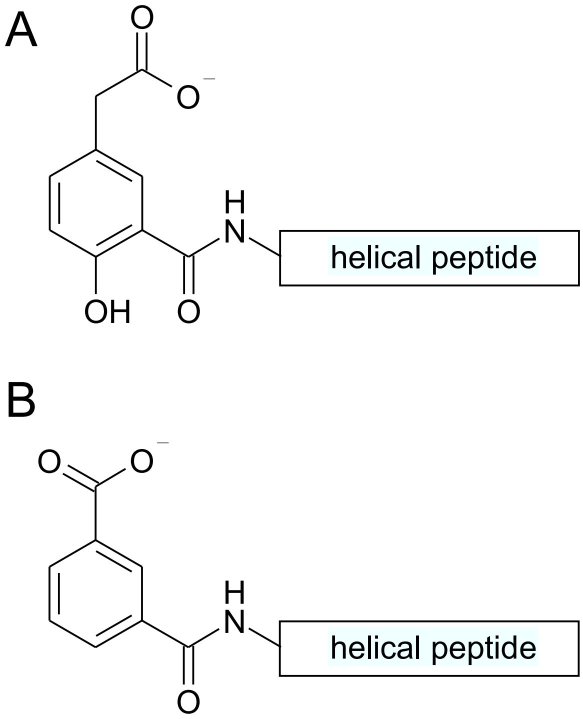


Figure 18: A) Open form of the pHP-phototrigger. B) The IPA-moiety closely mimics the pHP structure and thus IPA-modified peptides serve as a more cost-effective model of pHP-peptides.

For the CD measurements of helix contents the costly pHP peptides were replaced by model compounds featuring instead of the pHP group the closely related isophtalic acid (IPA) group at the N-terminus. Figure 18 demonstrates the close similarity of the pHP and IPA groups in terms of structure and physicochemical properties. Our notation distinguishes the IPA model peptides from the pHP originals by an attached asterisk. Thus $\mathrm{P}_{\mathrm{pHP}^*}$ designates the IPA model of $\mathrm{P}_{\mathrm{pHP}}$, for instance.

Convergence of the Simulations $\mathrm{P}_{\mathrm{pHP}}$.2/$\mathrm{P}_{\mathrm{pHP}}$.3

Quite like Figure 8 in the paper, which addressed the question of convergence for the REST simulations $\tilde{\mathrm{P}}$.4 and $\tilde{\mathrm{P}}$.5 of peptide $\tilde{\mathrm{P}}$, the following Figure 19 allows us to assess the convergence of the REST simulations $\mathrm{P}_{\mathrm{pHP}}$.2 and $\mathrm{P}_{\mathrm{pHP}}$.3 on the open form of the light-switchable peptide $\mathrm{P}_{\mathrm{pHP}}$.

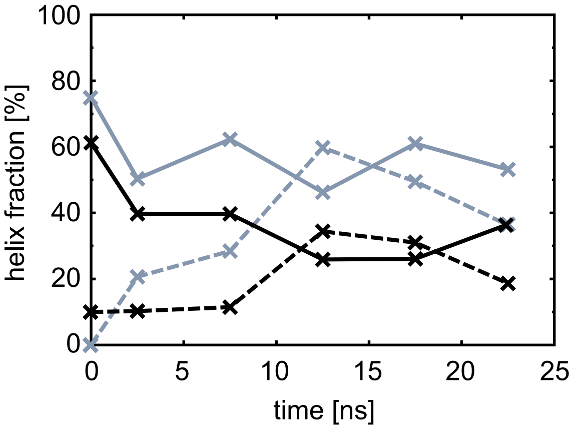


Figure 19: DSSP helix fractions $\langle H_\alpha(t)\rangle_{5\mathrm{ns}}$ averaged over 5 ns periods within the two 25 ns REST simulations $\mathrm{P}_{\mathrm{pHP}}$.2 (solid curves) and $\mathrm{P}_{\mathrm{pHP}}$.3 (dashed curves) and over the residues of $\mathrm{P}_{\mathrm{pHP}}$. The simulation $\mathrm{P}_{\mathrm{pHP}}$.2 started from predominantly α-helical, and $\mathrm{P}_{\mathrm{pHP}}$.3 from predominantly unfolded structures. The gray curves refer to $\langle H_\alpha\rangle$ measured for the replica at the target temperature of 300 K. The black curves are the overall $\langle H_\alpha\rangle$ in the two generalized REST ensembles made up of all replicas T.

According to Figure 19 the initial conditions are apparently forgotten both in the 300 K and in the generalized REST ensembles. As compared to the case of peptide $\tilde{\mathrm{P}}$, the fluctuations of the helix fraction $\langle H_\alpha(t)\rangle_{5\mathrm{ns}}$ observed after the initial 10 ns equilibration phase seem to be somewhat larger for the peptide $\mathrm{P}_{\mathrm{pHP}}$. This observation could indicate that the folding-unfolding processes proceed at a slighly smaller rate for $\mathrm{P}_{\mathrm{pHP}}$ than for $\tilde{\mathrm{P}}$. However, a clarification of this isssue would require a substantially more extended statistics focussing on the speed of folding-unfolding events.

References

1. MacKerell, Jr., A. D. *et al.* *J. Phys. Chem. B* **1998,** *102,* 3586-3616.

2. Feig, M.; MacKerell, Jr., A. D.; Brooks, III, C. L. *J. Phys. Chem. B* **2003,** *107,* 2831-2836.

3. MacKerell, Jr., A. D.; Feig, M.; Brooks, III, C. L. *J. Am. Chem. Soc.* **2004,** *126,* 698-699.

4. MacKerell, Jr., A. D.; Feig, M.; Brooks, III, C. L. *J. Comput. Chem.* **2004,** *25,* 1400-1415.

4 Temperaturkontrolle in Molekulardynamik-Simulationen heterogener Systeme

In diesem Kapitel werden Methoden zur Temperaturkontrolle in MD-Simulationen diskutiert. Dabei wird insbesondere auf die Problematik heterogener Systeme, wie z.B. der in Kapitel 3 betrachteten Peptid/Wasser-Systeme, eingegangen. Aus der Analyse der Mängel gängiger Verfahren zur Temperaturkontrolle in solchen Systemen wird eine neue Strategie entwickelt, die diese Mängel nicht aufweist.

Das Kapitel ist ein Abdruck[1] des Artikels

> Martin Lingenheil, Robert Denschlag, Rudolf Reichold und Paul Tavan:
> „The "Hot-Solvent/Cold-Solute" Problem Revisited“
> *Journal of Chemical Theory and Computation* **4**, 1293-1306 (2008),

den ich gemeinsam mit Martin Lingenheil, Robert Denschlag und Paul Tavan verfasst habe.

Zu diesem Projekt habe ich durch die Planung, Durchführung und Auswertung eines Teils der vorgestellten Simulationen beigetragen.

---

[1] Reproduced with permission from The Journal of Chemical Theory and Computation, 4, 1293-1306, 2008. Copyright 2008 American Chemical Society.

J. Chem. Theory Comput. 2008, 4, 1293–1306

JCTC Journal of Chemical Theory and Computation

The "Hot-Solvent/Cold-Solute" Problem Revisited

M. Lingenheil, R. Denschlag, R. Reichold, and P. Tavan*

Lehrstuhl für Biomolekulare Optik, Ludwig-Maximillians-Universität München,
Oettingestrasse 67, 80538 München, Germany

Received February 1, 2008

Abstract: The temperature steers the equilibrium and nonequilibrium conformational dynamics of macromolecules in solution. Therefore, corresponding molecular dynamics simulations require a strategy for temperature control which should guarantee that the experimental statistical ensemble is also sampled *in silico*. Several algorithms for temperature control have been proposed. All these thermostats interfere with the macromolecule's "natural" dynamics as given by the Newtonian mechanics. Furthermore, using a single thermostat for an inhomogeneous solute−solvent system can lead to stationary temperature gradients. To avoid this "hot solvent/cold solute" problem, two separate thermostats are frequently applied, one to the solute and one to the solvent. However, such a separate temperature control will perturb the dynamics of the macromolecule much more strongly than a global one and, therefore, can introduce large artifacts into its conformational dynamics. Based on the concept that an explicit solvent environment represents an ideal thermostat concerning the magnitude and time correlation of temperature fluctuations of the solute, we propose a temperature control strategy that, on the one hand, provides a homogeneous temperature distribution throughout the system together with the correct statistical ensemble for the solute molecule while, on the other hand, minimally perturbing its dynamics.

1. Introduction

Molecular dynamics (MD) simulations using molecular mechanics (MM) force fields have become a widespread tool to study the equilibrium conformational dynamics of proteins and peptides in solution,[1] including processes of folding and refolding.[2] More recently, also nonequilibrium processes have been simulated in which a protein or peptide is destabilized, for example by applying an external force mimicking the action of an atomic-force microscope,[3–5] by exerting internal mechanical strain,[6,7] by introducing point mutations into the protein sequence,[8,9] or simply by elevating the temperature.[9,10]

The behavior of proteins in solution is steered by the thermodynamic conditions, notably by the temperature. The native state is stable only within a certain temperature range; processes of hot and cold unfolding have been observed.[11] The temperature influences the stability and function of proteins not only directly by changing the relative importance of the entropy but also indirectly via certain temperature dependent solvent properties such as the dielectric constant[12] or the viscosity.[13] Therefore, if one wants to describe experiments on proteins by MD simulations, the temperature must be properly controlled.

Clearly, an adequate method for temperature control is not the only precondition if one aims at quantitative descriptions of experimental data. In this respect, the quality of the employed force field, the sufficiency of statistical sampling achieved by finite simulation times, and other technical issues are also questions of concern.[14] However, the temperature is of key importance because many experimental observables that can be compared with the information obtained from MD simulations sensitively depend on this parameter. Examples are the temperature factors in X-ray crystallography,[15] the proton exchange and spin relaxation rates in nuclear magnetic resonance spectroscopy [see ref 16 and references therein], and the fluorescence depolarization rates[17] as well as the thermodynamical measures of protein stability.[18]

---

* Corresponding author e-mail: tavan@physik.uni-muenchen.de.

10.1021/ct8000365 CCC: $40.75 © 2008 American Chemical Society
Published on Web 07/08/2008

The requirement of proper control does not only apply to the temperature, i.e. the average kinetic energy of the system, but also to other ensemble properties (e.g., energy fluctuations) associated with experimental observables. Thus, in a broader sense, the problem of temperature control in MD simulations is also that of generating the correct statistical ensemble (usually canonical or isothermal−isobaric). The accurate generation of a specific statistical ensemble by means of a MD simulation is also relevant for the application of generalized ensemble techniques like replica exchange molecular dynamics[19–21] which has recently become very popular in order to enhance the sampling efficiency. These techniques rely on the assumption that the applied MD method samples the canonical ensemble at the respective temperature.

When simulating macromolecules in solution, the solvent environment, which is essential for the properties of the solute, can either be treated implicitly using continuum approximations or explicitly by including part of the solvent into the simulation system.[14] The following discussions exclusively deal with the latter case and are devoted to the task of controlling the temperature of a solute macromolecule in explicit solvent. This task can comprise additional challenges if nonequilibrium relaxation processes are studied. Here, frequently, energy is released in one part of the system and then dissipated into the rest of the simulation box, e.g. from a solute molecule to the surrounding solvent. Since the kinetics of energy relaxation and heat transport can influence the dynamical properties of the solute,[22] any applied temperature control method should make sure that the natural energy relaxation processes are unimpaired.

Generally, the ideal temperature control scheme for solute−solvent systems would be to simulate the complete simulation system microcanonically, i.e. at constant total energy in the *NVE* ensemble. One can show[23] that, in this situation, an arbitrary subset of degrees of freedom in thermal contact with the rest of the system (e.g., the solute's kinetic degrees of freedom) will sample the canonical ensemble if the energy fluctuations of the subsystem are insignificant compared with the total energy in the rest of the system. Furthermore, one can show that the subsystem will sample the isothermal−isobaric ensemble if also the subsystem's volume fluctuations are negligible compared with the volume of the rest of the system. Finally, one expects that all those configurational degrees of freedom of the solute which directly interact with the solvent system will sample the canonical or isothermal−isobaric ensemble, respectively, if additionaly the solute−solvent interaction energy is neglibible compared with the solvent-internal interaction energy. In MD simulations systems, the latter condition is fulfilled if the solvent atoms by far outnumber those of the solute. Concurrently, by using the *NVE* approach, the solute's Newtonian dynamics are left completely undisturbed. The *NVE* strategy has been recommended[24] for studies of protein folding kinetics and is occasionally applied[25,26] to eliminate a putative influence of thermostat algorithms on the simulated dynamics.

Unfortunately, the simple *NVE* strategy is not easily applied to extended MD simulations. Numerical inaccuracies associated with approximation schemes serving to speed up the computations generally lead to a violation of energy conservation. For example, heating may be caused by certain approximate treatments of long-range electrostatic interactions[27,28] or by integrating the equations of motion with multiple-time-step algorithms.[29] Cooling may occur, for instance, if constraining bond lengths or angles with a too loose tolerance[30] or if neighbor lists for the calculation of the van der Waals interactions are not updated frequently enough.[31] The defect of energy conservation could, in principle, be compensated by using an ergostat algorithm which would just scale the velocities of all atoms at every time step by an appropriate factor to keep the total energy exactly constant. However, the rates of algorithmic energy drift can vary among the constituents of an inhomogeneous simulation system leading to unphysical steady state temperature gradients,[32] a problem sometimes referred to as the "hot-solvent/cold-solute" problem.[33] For example, such a gradient can result from an approximate treatment of the electrostatic interactions, which may render a mildly polar solute less affected by algorithmic noise than a strongly polar aqueous solvent.[32,34,35]

Thus, specifically for equilibrium simulations of macromolecules in solution, the applied temperature control has to fulfill an important requirement: The temperature distribution has to be homogeneous throughout the inhomogeneous simulation system. As a strategy guaranteeing such a homogeneous temperature distribution it has been suggested to couple the subsystems independently to separate thermostats.[36] Further below we will check this strategy among others because it is the central aim of this work to determine an optimal strategy for generating a homogeneous temperature distribution in solute−solvent simulation systems.

From a general point of view, the appropriateness of a given temperature control method involves the following three aspects:

a) Thermodynamics: Does the method generate the expected thermodynamical ensemble in principle (i.e., with simulations of infinite length and in the absence of numerical errors)?

b) Ergodicity: Does the method generate the expected ensemble within the time typically covered by modern MD simulations?

c) Dynamics: Is the time dependence and spatial distribution of the thermostatic forces realistic? For a solute in solution, for example, one would prefer to have no such forces at all beyond the thermostatting Newtonian interactions with the solvent.

A number of different algorithms has been proposed as realizations of the required thermostats (for a review see ref 36). Each of these algorithms has its specific merits and drawbacks. A critical discussion of these issues is another goal of our study.

For example, the widely used Berendsen thermostat[37] (BT) has the advantage to couple only weakly to the dynamics of the controlled system (see the original paper ref 37 for this issue). On the other hand, it is clear from theoretical considerations that the BT does not create a canonical distribution of microstates,[40] i.e. it introduces artifacts of type a). Furthermore, the BT violates energy equipartition by redistributing energy from high to low frequency modes, which leads to the so-called "flying-ice-cube effect".[38,39] It is unclear whether this effect is specific to the Berendsen method and closely related methods or it can occur with any thermostat belonging to the more general class of velocity rescaling algorithms.[38]

The more strongly coupling Nosé-Hoover thermostat[41,42] (NHT) is theoretically expected to generate the canonical distribution of microstates if certain conditions are obeyed thus conforming with the above question a).[43] However, within the time covered by a typical MD simulation, amplitudes of temperature fluctuations were observed which were by 1 order of magnitude larger than those expected for a canonical ensemble.[44] Several studies[42,44–47] have shown that Nosé-Hoover coupled systems do not necessarily acquire ergodicity in a reasonable time [cf. question b) above] if these systems are small, stiff, or at low temperatures. Additionally, by its very construction as a velocity rescaling algorithm, also the NHT could show the flying-ice-cube artifact (although we are not aware of any reports on a corresponding example).

As a reaction to these problems, modifications to both the Berendsen and Nosé-Hoover schemes have been proposed. The most frequently employed variant of the Nosé-Hoover thermostat is the so-called Nosé-Hoover chain,[48] which has been successfully tested by Cheng and Merz[33] as a remedy to the hot-solvent/cold-solute problem. No artifacts or deviations from the canonical ensemble have been reported so far. Only recently, Bussi et al.[49] suggested a modification of the Berendsen scheme in order to reliably generate a canonical distribution for systems that otherwise would sample the microcanonical ensemble. Both, the Nosé-Hoover chain and the modified Berendsen thermostat induce temperature fluctuations of the correct size by artificially scaling the atomic velocities. For systems, however, which anyway sample the canonical ensemble, such a thermostat introduces an unnecessary perturbation of the dynamics, i.e. artifacts of type c). The generic example for such a system is a solute molecule in a sufficiently large explicit solvent system, which, as discussed above, always samples a canonical ensemble although possibly at the wrong temperature because of algorithmic inaccuracies.

Concerning temperature control of macromolecules in solution, we want to show how one can (i) generate the appropriate ensemble for the solute molecule in adequate time, (ii) leave invariant the time scales of energy relaxation and of equilibrium fluctuations, and (iii) guarantee a homogeneous temperature distribution in equilibrium simulations with (iv) minimal perturbation of the solute's Newtonian dynamics.

For this purpose we will scrutinize in section 2 the existing temperature control scenarios for MD simulations of solvent–solute systems by partially recollecting and partially developing associated theoretical concepts. These considerations will lead to the definition of strategies for a minimally invasive control of a solute temperature. In section 3 we will sketch the methods which we employed in a series of quite extended test simulations on peptides in aqueous solution. As explained in section 4, these simulations were specifically designed to estimate the extent to which the theoretically expected effects of temperature control do actually modify the properties of a solute peptide. Section 5 discusses the results and suggests a practical procedure ensuring a minimally invasive temperature control.

2. Theory

Thermostats. The most widely used class of thermostat algorithms is based on the rescaling of atomic velocities. The equation of motion for an atom which belongs to a system under the rule of such a thermostat is

$$m_i\ddot{\mathbf{r}}_i(t) = \mathbf{F}_{i,\mathrm{ff}}(t) - m_i\gamma(t)\dot{\mathbf{r}}_i(t) \tag{1}$$

Here, the acceleration $\ddot{\mathbf{r}}_i(t)$ of atom i is caused not only by the forces $\mathbf{F}_{i,\,\mathrm{ff}}(t)$ derived from an MM force-field but also by a second term $\mathbf{F}_{i,\,\mathrm{therm}}(t) \equiv -m_i\gamma(t)\dot{\mathbf{r}}_i(t)$, which is proportional to the atom's velocity $\dot{\mathbf{r}}_i(t)$ and to a generally time dependent thermostat parameter $\gamma(t)$.

In the Berendsen scheme, $\gamma(t)$ is directly given in terms of the instantaneous kinetic temperature[50] $\boldsymbol{T}(t)$ by

$$\gamma(t) = \frac{1}{2\tau}\left[1 - \frac{T_0}{\boldsymbol{T}(t)}\right] \tag{2}$$

with τ denoting the coupling time and T_0 the target temperature. For the Nosé-Hoover[41,42] thermostat, $\gamma(t)$ is coupled on a time scale τ_{NHT} to $\boldsymbol{T}(t)$ by the differential equation

$$\frac{d\gamma}{dt} = \frac{1}{\tau_{\mathrm{NHT}}^2}\left[\frac{\boldsymbol{T}(t)}{T_0} - 1\right] \tag{3}$$

Perturbation of the Dynamics. Every thermostat which is described by eq 1 perturbs the Newtonian dynamics generated by the forces $\mathbf{F}_{i,\,\mathrm{ff}}(t)$ through the admixture of additional thermostatic forces $\mathbf{F}_{i,\,\mathrm{therm}}(t)$. For a solute–solvent system, these thermostatic forces introduce artifacts of type c) concerning the dynamics (cf. section 1). The resulting perturbation can be measured for a selected atom i by the quotient

$$\xi_i^2 \equiv \langle \mathbf{F}_{i,\mathrm{therm}}^2\rangle_D / \langle \mathbf{F}_{i,\mathrm{ff}}^2\rangle_D \tag{4}$$

where the brackets $\langle ...\rangle_D$ denote temporal averages over a simulation of a given duration D. The perturbation quotients (4) will depend on the system size and on the particular thermostat, i.e. on the form of $\gamma(t)$, as well as on the coupling time.

The perturbation quotients ξ_i are strictly local measures for the influence of a thermostat on a simulated dynamics. However, one may also consider the local perturbation inflicted on a certain group G of atoms within a simulation system, e.g. on the C_α-atoms of a solute peptide embedded in a solvent environment. Then the root mean quotient $\bar{\xi}_G \equiv \sqrt{\langle \xi_i^2\rangle_G}$ over the ξ_i^2 belonging to G can be used to compare how the dynamics of a solute is perturbed in different solute–solvent systems.

Instead of calculating the averages $\langle \mathbf{F}_{i,\,\mathrm{therm}}^2\rangle_D$ required for the evaluation of the ξ_i^2 directly from a simulation, one can also give a simple estimate for these average square forces.

Assuming a sufficiently large simulation system, the velocities of the individual atoms will negligibly contribute to the temperature $\boldsymbol{T}(t)$. Hence, the correlation of γ^2 and $\dot{\mathbf{r}}_i^2$ vanishes and one obtains

$$\langle \mathbf{F}_{i,\text{therm}}^2\rangle_D = \langle m_i^2\gamma^2\dot{\mathbf{r}}_i^2\rangle_D = m_i^2\langle\gamma^2\rangle_D\langle\dot{\mathbf{r}}_i^2\rangle_D \tag{5}$$

Assuming furthermore that the system is in equilibrium, that the atomic velocity distributions are undisturbed by the thermostat, and that the system is free of internal constraints (such as fixed bond lengths), the mean square velocity of atom i is expected to be $\langle\dot{\mathbf{r}}_i^2\rangle_D \approx 3k_B\hat{T}/m_i$, where $\hat{T} = \langle\boldsymbol{T}\rangle_D$ is the average temperature determined from the simulation. Equation 5 then becomes

$$\langle \mathbf{F}_{i,\text{therm}}^2\rangle_D \approx 3m_i k_B\hat{T}\langle\gamma^2\rangle_D \tag{6}$$

We will check this estimate by sample simulations and show that it already holds for relatively small systems.

Inserting the estimate 6 into eq 4, one can recognize that the perturbation quotients of a given system which is simulated with different thermostatic strategies solely differ with respect to $\langle\gamma^2\rangle_D$. Thus, in this case, comparisons of the mean square scaling activities $\langle\gamma^2\rangle_D$ suffice for the evaluation of different thermostatic strategies concerning the size of local perturbations of the dynamics. However, thermostats do not only cause local perturbations of the Newtonian dynamics but may also interfere with ensemble properties like, for example, size and time scales of the temperature fluctuations.

Temperature Fluctuations. In an MD simulation, the statistics of the temperature fluctuations provides a probe for artifacts of type a) and b) pertaining the generation of the desired ensemble (section 1). For a system in contact with a heat bath of temperature T_b, the distribution of microstates is either given by the canonical or by the isothermal−isobaric ensemble. However, with respect to the temperature fluctuations, both ensembles are equal. The associated probability density $\rho(\boldsymbol{T})$ for the instantaneous kinetic temperature is a χ^2-distribution[51]

$$\rho(\boldsymbol{T}) = \frac{(N_\text{DoF}\boldsymbol{T}/2T)^{N_\text{DoF}/2}}{\Gamma(N_\text{DoF}/2)\boldsymbol{T}}\ \exp\left[-\frac{N_\text{DoF}\boldsymbol{T}}{2T}\right] \tag{7}$$

where $T = T_b$ is the expectation value of $\boldsymbol{T}$, N_DoF is the number of degrees of freedom (DoF) of the system, and $\Gamma(...)$ denotes the Euler Γ-function. Consequently, the variance σ_T^2 of the temperature fluctuations is

$$\sigma_T^2 = \frac{2T^2}{N_\text{DoF}} \tag{8}$$

Under the influence of a thermostat, the statistics can deviate from what is expected for a canonical ensemble. This deviation constitutes a measure for the global influence of the thermostat and for how close a simulation is to sampling the canonical ensemble. In the limit $N_\text{DoF}\to\infty$, eq 7 becomes a normal distribution.

The size of σ_T^2 together with the autocorrelation time[50] τ_c of the temperature fluctuations critically influences the accuracy with which the equilibrium temperature T is determined by a given simulation. The variance $\sigma_{\hat{T}}^2$ of the time averages $\hat{T}$ obtained from a set of equilibrium simulations with durations D can be estimated[50] to be

$$\sigma_{\hat{T}}^2 = 2\sigma_T^2\frac{\tau_c}{D} \tag{9}$$

In order to judge whether a particular strategy is suited to correctly tune the temperature T, one has to perform a test simulation which is long enough to determine T with sufficient accuracy. For a small solute (large σ_T^2) with a correlation time τ_c longer than 10 ps, an accuracy of 1 K may require simulation times of up to 10 ns.

Power of a Thermostat. By means of the observables introduced above, one can judge to what extent a thermostat can perturb the dynamical and equilibrium properties of a solute in solute−solvent simulations. Such perturbations can, of course, be avoided by using no thermostat at all. However, as outlined in section 1, this approach is generally not feasible because algorithmic inaccuracies, which are inevitable in large scale simulations using efficient MD codes, represent heat drains or sources that have to be compensated.

To properly tune this compensation, we consider the work performed by the thermostatic forces $\mathbf{F}_{i,\text{therm}}(t)$ on the atoms i for an ensemble of simulation systems with the temperature $T(t) = \langle\boldsymbol{T}(t)\rangle_\text{ens}$. The ensemble average power exerted by the thermostat on a given atom i is

$$\beta_i(t) = \langle\mathbf{F}_{i,\text{therm}}(t)\cdot\dot{\mathbf{r}}_i(t)\rangle_\text{ens} \tag{10}$$

Using the definition of $\mathbf{F}_{i,\text{therm}}$ [see eq 1] and the Berendsen expression 2 for γ leads to

$$\beta_i(t) = \frac{1}{\tau}\langle\varepsilon_{i,\text{kin}}(t)[T_0/\boldsymbol{T}(t) - 1]\rangle_\text{ens} \tag{11}$$

with the usual definition for the kinetic energy $\varepsilon_{i,\text{kin}}(t)$ of atom i. Employing once more the assumption of a negligible correlation between the velocity [and, thus, the kinetic energy $\varepsilon_{i,\text{kin}}(t)$] of a single atom and the kinetic temperature $\boldsymbol{T}(t)$ of the system, one obtains

$$\beta_i(t) = \frac{3k_BT_i(t)}{2\tau}[T_0/T(t) - 1] \tag{12}$$

where k_B is the Boltzmann constant, and $T_i(t) \equiv 2/3k_B\langle\varepsilon_{i,\text{kin}}(t)\rangle_\text{ens}$ is the ensemble average temperature of atom i. For equilibrated systems the ensemble averages employed in eq 12 can be replaced by temporal averages $\langle...\rangle_D$. This allows to calculate for every subsystem κ from a simulation the (time) average thermostatic power

$$\hat{\beta}_\kappa = \frac{k_B\hat{T}_\kappa}{2\tau}[T_0/\hat{T} - 1] \tag{13}$$

per degree of freedom using the average temperature $\hat{T}_\kappa \equiv \langle\mathbf{T}_\kappa\rangle_D$ of the subsystem κ, the corresponding average $\hat{T} \equiv \langle\mathbf{T}\rangle_D$ of the temperature $\boldsymbol{T}(t)$ controlled by the BT, and the thermostat parameters T_0 and τ.

Further below we will use eq 13 to determine the thermostatic power exerted by a BT on a solute peptide from sample simulations. These data will be used to check the validity of a heat conduction model which we will now introduce to analyze the hot-solvent/cold-solute problem occasionally hampering MD simulations of inhomogeneous systems.[33]

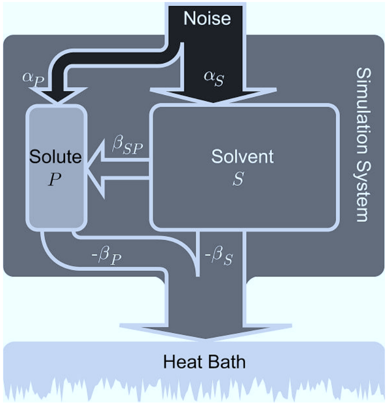


Figure 1. Heat flow model representing specifically the “hot-solvent/cold-solute” case for an inhomogeneous system consisting of two subsystems with different heating rates. The simulation system is coupled to a single thermostat, representing an external heat bath. Bright and dark colors code low and high temperatures, respectively. Heat flows driven by temperature gradients and heat sources are marked by arrows. A detailed discussion is given in the text.

Heat Flow Model. In simulations of solute−solvent systems, the algorithmic heat drains or sources may be inhomogeneously distributed and, thus, the temperature may likewise be inhomogeneous. According to requirement (iii) stated at the bottom of the Introduction, such inhomogeneous temperature distributions should be avoided. Figure 1 sketches a heat flow model from which one can derive strategies for the reliable control of the solute temperature. As drawn, the model refers to a particular strategy employing a single thermostat for the whole system. For further reference we denote this strategy by **G**.

The model depicted in Figure 1 consists of two subsystems $\kappa \in \{P, S\}$ with P denoting the solute and S denoting the solvent. The powers α_κ of algorithmic heating per DoF are assumed to be constant and homogeneous within the subsystems. Furthermore, the temperature is assumed to be homogeneous within each subsystem κ, i.e. for the atomic temperatures T_i we have $T_i = T_\kappa$ for all $i \in \kappa$. According to eq 12, the ensemble average work $\beta_i(t)$ exerted on atom i per unit time by the global thermostat then only depends on whether i is part of P or S, respectively. Thus, for the subsystems $\kappa \in \{P, S\}$ the respective thermostatic powers per DoF are given by

$$\beta_\kappa(t) = \frac{k_B T_\kappa(t)}{2\tau}[T_0/T(t) - 1] \tag{14}$$

If the local temperatures T_P and T_S differ, as is assumed in Figure 1, there will be a net heat flow

$$\beta_{SP}(t) = \frac{k_B[T_S(t) - T_P(t)]}{2\tau_{SP}} \tag{15}$$

between S and P, which we assume to depend linearly on the temperature difference. Here, the time constant τ_{SP} characterizes the thermal coupling of the subsystems.

The heat flowchart shown in Figure 1 immediately suggests stationarity conditions. In the steady state, the net heat flow must individually vanish for each of the two subsystems, i.e.

$$\alpha_P + \beta_{SP} + \beta_P = 0 \tag{16}$$

$$\alpha_S - \beta_{SP} + \beta_S = 0 \tag{17}$$

Now suppose for a moment that the temperature distribution is homogeneous throughout the system, i.e. $T_P(t) = T_S(t) = T(t)$. According to eq 14 the thermostatic powers $\beta_P(t)$ and $\beta_S(t)$ exerted on the subsystems are then equal, and, by eq 15, the heat flow $\beta_{SP}(t)$ between S and P vanishes. Equations 16 and 17 then immediately require as the stationarity condition that $\alpha_P = \alpha_S$, i.e. that the heat sources in the subsystems work at equal powers.

If this is not the case ($\alpha_P \neq \alpha_S$), the temperature cannot be homogeneously distributed in the stationary state, and, by eq 15, there will be a nonvanishing heat exchange $\beta_{SP} \neq 0$ between the subsystems. As a result, a steady state temperature difference is inevitable whenever, upon applying scenario **G**, a single global thermostat is used to thermostatize a system exhibiting inhomogeneities with respect to the rates α_κ of algorithmic heating. This is the origin of the hot-solvent/cold-solute problem as described e.g. in ref 34.

Separate Thermostats. To avoid temperature inhomogeneities, it has become a standard in simulations of macromolecules to couple *separate* thermostats to the subsystems.[36,47,52–57] We will denote this temperature control scenario by **P**. In the following discussion of scenario **P**, we will concentrate on the temperature control of the solute P, assuming that the temperature of the solvent S is reliably controlled. Such a reliable control can be achieved by a solvent thermostat combining a coupling time τ_S on the subpicosecond time scale (e.g., $\tau_S = 0.1$ ps) with a target temperature $T_{0,S}$ equal to the intended temperature. This choice of thermostat tuning actually is the standard (see e.g. refs 33, 36, 37, 47, 52–56), and, thus, we call it the classical setup.

For a scenario P, in which a separate BT is coupled to a (thermally homogeneous) solute P, the controlled temperature $T(t)$ is the solute temperature $T_P(t)$. Thus, we obtain from eq 14 the simplified expression

$$\beta_P(t) = \frac{k_B}{2\tau_P}[T_0 - T_P(t)] \tag{18}$$

for the power of the thermostat acting on P. With eqs 18 and 15, the solute’s stationarity condition 16 may be rewritten as

$$\alpha_P + \frac{k_B(T_S - T_P)}{2\tau_{SP}} + \frac{k_B(T_{0,P} - T_P)}{2\tau_P} = 0 \tag{19}$$

where $T_{0,P}$ denotes the target temperature, and τ_P denotes the coupling time of the solute thermostat. The first term characterizes the algorithmic heating within P, the second term characterizes the heat flow between P and S, and the third term characterizes the power β_P of the thermostat separately coupled to P.

Equation 19 is the quintessence of our linear heat flow model and may be used to predict the effects of three different thermostatic strategies within scenario **P**. In all these strategies, S is coupled to a classical BT and P is decoupled from this thermostat. The three strategies are as follows:

P.1 The solute P is coupled to a classical thermostat. Here, the use of a correspondingly small coupling time $\tau_P \approx 0.1$

ps is the standard.[36,52–56,58] For such small τ_P, eq 19 is completely dominated by the thermostatic term. The reason for this dominance is that τ_P is by at least 1 order of magnitude smaller than the solute–solvent coupling time τ_{SP}, which is typically larger than 1 ps (see further below). Neglecting the heat flow contribution, the deviation $T_P - T_{0,P}$ from the target temperature is given by $2\alpha_P\tau_P/k_B$. For moderate algorithmic heating rates α_P, this deviation is expected to be small because of the short time scale τ_P.

P.2 No separate thermostat is coupled to the solute P, i.e. $\tau_P \rightarrow \infty$, and solely the thermostatted solvent S acts as a heat bath. We call this strategy "noninvasive" because it does neither alter the Newtonian dynamics nor the energy relaxation properties of P. The expected temperature difference $T_P - T_S = 2\alpha_P\tau_{SP}/k_B$ will be small if the local heating α_P is negligible on the time scale τ_{SP} of the thermal coupling between the subsystems.

P.3 The solute is coupled to a thermostat with a very large coupling time $\tau_P \gg \tau_{SP}$ to realize a "constant heat flow" (CHF) approach. As suggested by the heat balance eq 19, a homogeneous ($T_P = T_S$) and stationary temperature distribution only requires that the thermostatic power β_P cancels the power α_P of algorithmic heating, i.e., $\alpha_P = k_B(T_P - T_{0,P})/2\tau_P$. This condition can be satisfied for an arbitrarily large coupling time τ_P by a proper choice of the target temperature $T_{0,P}$. In the limit $\tau_P \rightarrow \infty$, the thermostat variable γ in eqs 1 and 2 becomes a constant γ_P, and the thermostat scheme may actually be descibed by this single parameter. At large τ_P, the thermostat works in a heating/cooling limit as a constant heat source/drain, and this activity solely serves to maintain the energy balance. Because of eq 2, the perturbation of the Newtonian dynamics [cf. eq 1] inflicted by such a CHF thermostat can be made very small. Therefore, we call the CHF approach to the solute's temperature control, which is applicable to non-negligible local heating rates α_P, "minimally invasive".

To set up a CHF simulation as required in strategy **P**.3, the a priori unknown power α_P of algorithmic heating has to be determined in order to specify the constant thermostat parameter γ_P, or, equivalently, the paramteres $T_{0,P}$ and τ_P if a traditional Berendsen thermostat is used in the heating/cooling limit. To this end, the solute temperature T_P has to be measured in two test simulations with different heating powers β_P of the thermostat. The two heat balance eqs (19) of these tests then constitute a system of linear equations which determines the unknown parameters β_{SP} and α_P. A detailed description of the setup protocol is given in Appendix A.

In the following we will examine the temperature control strategies **G** and **P**.1 −**P**.3 introduced above by test simulations. Based on these results, we formulate guidelines for a temperature control satisfying the four conditions summarized at the bottom of the Introduction.

3. Methods

MD Simulation Techniques. The software packages EGO-MMII[27] and GROMACS[59] were used in several series of MD simulations. Besides EGO we also applied GROMACS because it provides an NHT in addition to a BT, because it is computationally efficient for very small systems, and because it can provide data for a crosscheck of results. In EGO the electrostatic interactions are treated combining structure-adapted multipole expansions[60,61] with a moving-boundary reaction-field approach[62] and a multiple-time-step integration.[29,63] In the GROMACS simulations we used the PME method[28] with a 10 Å cutoff for the real space contribution, with a grid spacing of 0.5 Å, and with a sixth order interpolation of the charges to the grid. For both EGO and GROMACS simulations, the van der Waals interactions were truncated at 10 Å. If not stated otherwise, the simulations were carried out with explicit solvent using periodic boundary conditions and with a BT ($\tau_S = 0.1$ ps, $T_{0,S} = 300$ K) rapidly coupled to the solvent to guarantee that the solvent was closely kept at the target value. Bond lengths were constrained using the M-SHAKE algorithm[30] with relative tolerances of 10^{-4} when using GROMACS, which is the recommended default value, and 10^{-6} when using EGO, which is hard-coded in the source code in this case.

We applied different simulation protocols to vary the heating properties within the simulation systems. Here, the first parameter was the software used for simulation, which we denote by E for EGO or G for GROMACS. Since the M-SHAKE algorithm is known[30] to have a cooling effect, we varied the number of constraints by either constraining no bonds at all (N), only bonds involving a hydrogen atom (H), or all bonds (A). The last parameter which presumably influences the heating in the system is the length Δt of the basic integration time step, which we simply denote by its value in femtoseconds. Thus, a standard EGO simulation (constraints on bonds involving hydrogen atoms and $\Delta t = 1$ fs) would be denoted by E/H/1.

Model Systems. The first model system was a polyalanine octapeptide (8ALA) with charged termini described by the GROMOS96 force field[64] and embedded in a cubic box of 20 Å edge-length containing 236 simple point charge (SPC) water molecules.[65] The number of DoF for the peptide then is 153/143/103 for N-/H-/A-constraining, respectively. The starting conformation was always fully extended. The system was equilibrated for 300 ps during which solute and solvent were coupled to separate BTs ($\tau_P = \tau_S = 0.1_{\text{ps}}$, $T_{0,S} = T_{0,P} = 300$ K). The second model system was an alanine dipeptide (ALDI) described by the CHARMM22 force field[66] in a cubic box of 21.3 Å edge length containing 324 water molecules modeled by the transferable three-point intermolecular potential (TIP3P).[66,67] Here, the number of peptide DoF is 66/54/45 for N-/H-/A-constraining, respectively. The system was prepared as described for 8ALA, except that the equilibration time was only 100 ps.

MD Simulations. A first series of seven MD simulations of 8ALA in SPC water served to study the various situations encountered in the temperature control of inhomogeneous systems. Table 1 associates acronyms to these simulations and lists the employed parameters. In particular, in the last simulation G/A/2_**P**.3, the CHF approach was applied to the peptide. Using the data from the preceding simulation G/A/2_**P**.2 ($\tau_P = \infty$, $\hat{T}_P = 293.4$ K), in which only S was coupled to a classical BT and the data from an independent 10 ns test simulation with an additional CHF thermostat coupled

Table 1. Simulation Names and Associated Parameters in Series #1[a]

	protocol				thermostat parameters			
name	software	C	Δt/fs	D/ns	τ_{sys}/ps	τ_S/ps	τ_P/ps	$T_{0,P}$/K
E/H/1_**G**	EGO	H	1	20	0.1	–	–	–
E/H/2_**G**	EGO	H	2	20	0.1	–	–	–
E/H/2_***P***.2	EGO	H	2	20	–	0.1	–	–
G/H/2_***P***.2	GROMACS	H	2	20	–	0.1	–	–
G/A/2_***P***.2	GROMACS	A	2	20	–	0.1	–	–
G/A/2_***P***.1	GROMACS	A	2	20	–	0.1	0.1	300
G/A/2_***P***.3	GROMACS	A	2	20	–	0.1	500	2340

[a] The simulation names code the varied parameters and temperature control scenarios. C specifies the type of bond length constraints, Δt the size of the basic integration time step, and D the duration of the simulation. The parameters τ specify the coupling times of the BTs coupled to the whole system (sys), to the solvent (S), or to the solute (P). $T_{0,P}$ is the target temperature of a thermostat coupled to the solute. The solute peptide was 8ALA in SPC water. See the text for further information.

Table 2. Simulation Parameters in Series #2[a]

	protocol				thermostat parameters			
name	software	C	Δt/fs	D/ns	τ_{sys}/ps	τ_S/ps	τ_P/ps	$T_{0,P}$/K
CHF.0	GROMACS	A	2	200 × 2	–	0.1	–	–
CHF.1	GROMACS	A	2	200 × 2	–	0.1	500	2340
CHF.2	GROMACS	A	2	200 × 2	–	0.1	500	4800
CHF.3	GROMACS	A	2	200 × 2	–	0.1	500	7700
CHF.4	GROMACS	A	2	200 × 2	–	0.1	500	11100
CLS.1	GROMACS	A	2	200 × 2	–	0.1	0.1	300
CLS.2	GROMACS	A	2	200 × 2	–	0.1	0.1	340

[a] The model peptide was 8ALA in SPC water at $\hat{T}_S = 300$ K. Except for the simulation set CHF.0, in which only T_S was controlled, separate BTs were applied to S and P. See the caption to Table 1 for further information.

Table 3. Simulation Parameters in Series #3[a]

	protocol				thermostat parameters		
peptide	software	C	Δt/fs	D/ns	τ_{sys}/ps	τ_P/ps	$T_{0,P}$/K
8ALA	GROMACS	N	1	0.25	NHT	0.064	300
8ALA	GROMACS	N	1	0.25	NHT	0.256	300
8ALA	GROMACS	N	1	0.25	NHT	1.024	300
8ALA	GROMACS	N	1	0.25	NHT	4.096	300
8ALA	GROMACS	N	1	0.25	NHT	16.384	300
8ALA/ALDI	EGO	N	1	0.25	BT	0.001	300
8ALA/ALDI	EGO	N	1	0.25	BT	0.004	300
8ALA/ALDI	EGO	N	1	0.25	BT	0.016	300
8ALA/ALDI	EGO	N	1	0.25	BT	0.064	300
8ALA/ALDI	EGO	N	1	0.25	BT	0.256	300
8ALA/ALDI	EGO	N	1	0.25	BT	1.024	300
8ALA/ALDI	EGO	N	1	0.25	BT	4.096	300

[a] For nomenclature see the caption to Table 1. In all simulations the solvent was coupled with $\tau_S = 0.1$ ps to a Nosé-Hoover (NHT) or a Berendsen (BT) thermostat, respectively.

to the peptide ($\tau_P = 500$ ps, $T_{0,P} = 4800$ K, $\hat{T}_P = 307.9$ K), the unknown parameters in eq 19 were determined as described in the section 2. We found the values $\alpha_P = -2.04k_B$ K/ps for the algorithmic heating rate and $\tau_{SP} = 1.61$ ps for the solute–solvent coupling time, which actually is in the picosecond time range as claimed further above. To realize a CHF thermostat maintaining the peptide at $T_P \approx 300.0$ K, these values were inserted into eq 19 yielding a “target temperature” $T_{0,P} = 2340$ K. If the assumptions underlying our heat-flow model are correct, this choice should compensate through $\beta_P = -\alpha_P$ the algorithmic energy drift in the G/A/2_**P**.3 simulation.

The setup of a second series of simulations was chosen such that the effects of the local temperature and of a thermostat on the dynamics of 8ALA can be studied. We performed seven sets of 200 simulations each. Every single simulation had a duration of 2 ns, amounting to 400 ns per set and a total of 2.8 μs of simulation time. The simulation parameters are summarized in Table 2. All simulations were performed with the G/A/2 protocol. In the first set (CHF.0), no thermostat was coupled to the peptide, while in the following four sets (CHF.1 to CHF.4) a BT targeting at increasingly large temperatures $T_{0,P}$ was coupled in an extremely slow fashion to the peptide. In the last two sets (CLS.1 and CLS.2), a separate classical BT was coupled to the peptide using either the same ($T_{0,P} = 300$ K) or a slightly higher ($T_{0,P} = 340$ K) target temperature as compared to $T_{0,S}$. The 200 initial conditions were obtained by taking snapshots every 20 ps from a 2 ns preparatory simulation at 300 K, with the peptide’s C_α atoms harmonically coupled to their initial coordinates of an extended conformation.

To compare the different thermostatic strategies discussed in section 2, we determined the corresponding thermostatic forces (eqs 5 and 6) and perturbation ratios (eq 4) in a third series of relatively short 250 ps simulations. Simulations were performed for 8ALA with varying coupling strengths and BTs and NHTs, respectively. Additionally, we determined the thermostatic forces and the perturbation ratio also for ALDI and Berendsen coupling again varying the coupling strength. The simulation parameters of the third series are given in Table 3. As these simulations served to compare thermostatic and force-field forces, no bond lengths were constrained thus eliminating constraint forces.

Finally, a fourth series of slightly more extended simulations (500 ps) was designed to examine how the solute’s variance of temperature fluctuations (cf. the corresponding paragraph in section 2) is affected by the coupling times of a BT. We studied 8ALA and ALDI in water and in vacuum

Table 4. Simulation Parameters in Series #4[a]

system		protocol				thermostat parameters	
peptide	environment	software	C	Δt/fs	D/ns	τ_P/ps	$T_{0,P}$/K
8ALA	water/vac	EGO	A	2	0.5	0.001	300
8ALA	water/vac	EGO	A	2	0.5	0.004	300
8ALA	water/vac	EGO	A	2	0.5	0.016	300
8ALA	water/vac	EGO	A	2	0.5	0.064	300
8ALA	water/vac	EGO	A	2	0.5	0.256	300
8ALA	water/vac	EGO	A	2	0.5	1.024	300
8ALA	water/vac	EGO	A	2	0.5	4.096	300
ALDI	water/vac	EGO	H	2	0.5	0.001	300
ALDI	water/vac	EGO	H	2	0.5	0.004	300
ALDI	water/vac	EGO	H	2	0.5	0.016	300
ALDI	water/vac	EGO	H	2	0.5	0.064	300
ALDI	water/vac	EGO	H	2	0.5	0.256	300
ALDI	water/vac	EGO	H	2	0.5	1.024	300
ALDI	water/vac	EGO	H	2	0.5	4.096	300

[a] For nomenclature see the caption to Table 1. BTs were used for solvent and solute. The solvent was coupled with a coupling time of 0.1 ps.

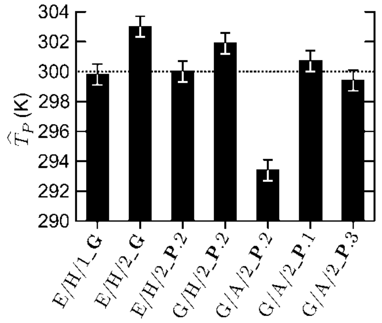


Figure 2. Average peptide temperature $\hat{T}_P$ observed in the first series of simulations on 8ALA in SPC water. The associated acronyms and parameters characterizing the members of the series are given in Table 1.

by E/H/2 simulations using the same set of coupling times as in series #3. Table 4 summarizes the simulations of the last series.

4. Results and Discussion

Temperature Control Scenarios. As outlined above, the series of equilibrium simulations on the model peptide 8ALA in explicit water as characterized by Table 1 served to exemplify the problems connected with the temperature control of inhomogeneous systems. Figure 2 shows the average peptide temperatures obtained in these sample simulations. Using eq 9, the remaining uncertainty of these average temperatures was estimated to $\sigma_{\hat{T}_P} < 0.7$ K. The solvent temperatures were 300.0 K where not mentioned explicitly.

In simulation E/H/1_**G**, we used the standard simulation protocol for EGO (see section 3 for details), which includes a classical BT coupled to the whole simulation system and, thus, represents an example for scenario **G** outlined in section 2. Neither the resulting temperatures of the peptide (cf. Figure 2) nor of the solvent showed any statistically significant deviations from the 300 K target value suggesting that in E/H/1_**G** the algorithmic noise was weak.

Figure 2 indicates that this behavior was lost in simulation E/H/2_**G**, in which the basic integration time step Δt was doubled to 2 fs. For our sample system, this doubling of Δt led to a 3.0 K increase of the peptide temperature, indicating that the modified simulation setup has caused certain algorithmic inaccuracies. When using EGO, the choice of a larger Δt is expected to reduce the accuracy of the integration algorithm because the employed highly efficient multiple-time-step algorithm does not exactly guarantee energy conservation and because the corresponding violation increases with the size of Δt (see refs 63 and 29 for a discussion). According to Figure 2, the combination of a global Berendsen thermostat with a reduced accuracy of integration in simulation E/H/2_**G** apparently led to a moderately elevated temperature for the peptide and to a slightly (0.3 K) cooler temperature for the larger solvent system. Nevertheless, the temperature of the total system was accurately kept at 300.0 K by the thermostat. Apart from changed signs (hot solute in cold solvent), this result is an example for the classical problem reported in the literature,[32,34,68] which can arise in scenario **G** from indiscriminately coupling a thermostat to all parts of an inhomogeneous system and which is described by the heat flow model sketched in Figure 1.

However, as demonstrated by the average peptide temperature displayed in Figure 2 for simulation E/H/2_**P**.2, this temperature control problem was eliminated by simply decoupling the peptide from the thermostat, i.e. by realizing scenario **P**.2. This observation suggests that in the E/H/2 simulations the solvent experiences a considerable cooling, whereas the level of algorithmic noise within the peptide is very low. According to our experience, such a decoupling of the solute is a proper solution for most temperature control problems which can occur in simulations of inhomogeneous systems using either EGO or GROMACS.

The fact that the application of scenario **P**.2 cannot always remove such problems is demonstrated by the results of simulation G/H/2_**P**.2, which was carried out with GROMACS using the same settings as in the EGO simulation E/H/2_**P**.2. According to Figure 2, in the G/H/2_**P**.2 simulation the peptide was by about 2 K too hot, indicating that the rate β_{SP} of heat transport from the peptide P into the solvent S was too slow to compensate the algorithmic heating $\alpha_P > 0$ of the solute occurring in this case.

It may be expected that introducing additional M-SHAKE constraints into the peptide system leads to a local cooling,[30] which might compensate the observed algorithmic heating of *P*. This is the reason why we carried out simulation G/A/2_**P**.2, which differs from G/H/2_**P**.2 only in the number of constraints (50 vs 10) within the peptide. In fact, Figure 2 displays for simulation G/A/2_**P**.2 a peptide temperature which is by 6.6 K cooler than that of the solvent, implying that the original heating has been overcompensated by the local cooling. A deviation of this size is unacceptable in simulations serving to probe the equilibrium properties of the solute. Thus, the simulation setup G/A/2 is a typical case in which one of the two remaining temperature control strategies **P**.1 and **P**.3 described in the section 2 should be applied.

Hence, in simulation G/A/2_**P**.1 we utilized a separate classical BT for temperature control of the peptide, while in simulation G/A/2_**P**.3 we applied a CHF thermostat. Figure 2 shows that in both cases there is no significant deviation of the observed peptide temperatures from the solvent temperature. Both methods are capable of correctly thermostatting the solute. For the CHF thermostat we conclude that the choice of parameters (cf. section 3) was correct and that the underlying heat flow model describes the situation in this case. This success has motivated us to further scrutinize the validity of this model.

Validity of the Heat Flow Model. The second quite extended series of simulations (see Table 2 for the parameters) can serve to assess the validity of eq 19, which expresses the contents of the model. With eq 18 the model 19 can be equivalently reformulated as

$$T_P = T_S + \frac{2\tau_{SP}}{k_B}(\beta_P + \alpha_P) \tag{20}$$

showing that the solute temperature T_P should depend linearly on the heating power β_P of the solute thermostat. To specify the unknown parameters α_P and τ_{SP} in eq 20, one needs measurements of T_S and T_P from two simulations employing different heating powers β_P.

Estimates $\hat{\beta}_P$ for the heating powers β_P can be determined from simulations by evaluating eq 13 specifically for the case of a solute thermostat, i.e. for $\kappa = P$, $T_0 = T_{0,P}$, $\tau = \tau_P$, and $\hat{T} = \hat{T}_P$. One obtains

$$\hat{\beta}_P = \frac{k_B}{2\tau_P}[T_{0,P} - \hat{T}_P] \tag{21}$$

which is, up to the use of different averages, identical to eq 18. Thus, at a constant coupling time τ_P, the heating power β_P is steered by the choice of the target temperature $T_{0,P}$ and measured through the average peptide temperature $\hat{T}_P$. Therefore, the linear relationship 20 between T_P and β_P can be checked by comparing with data points ($\hat{T}_P$,$\hat{\beta}_P$) obtained from simulations employing different target temperatures $T_{0,P}$.

An inspection of the first five simulation sets in series #2 listed in Table 2 shows that this set qualifies both for the evaluation of the unknown parameters in eq 20 and for the check of this linear equation. In all simulation sets of series #2, the simulation protocol was G/A/2 just like in the simulation G/A/2.**P**.2 of the first series. However, the temperature control scenario **P**.2 (no separate thermostat for the peptide) was employed only in simulation CHF.0. In the remaining CHF simulations a BT was coupled to *P* using an extremely slow coupling time $\tau_P = 500$ ps combined with a large and increasing target temperature (cf. Table 2). According to eq 21 this choice leads to a heating power β_P of this thermostat, which increases from simulation CHF.0 ($\beta_P = 0$) to simulation CHF.4. Figure 3 shows the observed stationary peptide temperatures $\hat{T}_P$ as a function of the observed heating power $\hat{\beta}_P$. In the case of the simulation set CHF.0 (black dot) the result of simulation G/A/2_**P**.2 (cf. Figure 2) is closely recovered because the same temperature control setting **P**.2 was applied, i.e. $\hat{T}_P$ was by 6.5 K smaller than the solvent temperature of $\hat{T}_S = 300$ K. With nonzero and successively growing $\hat{\beta}_P$ the peptide temperature $\hat{T}_P$ is seen to increase.

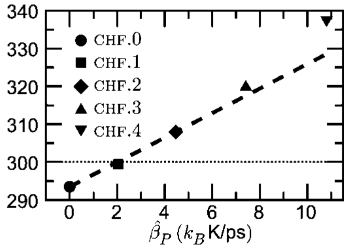


Figure 3. The average temperatures $\hat{T}_P$ of the peptide ALA8 (in SPC water at $\hat{T}_S = 300$ K) resulting from constant local heating with different powers β_P in the simulation sets CHF.0 to CHF.4 of series #2 (cf. Table 2). The prediction of linear heat flow model eq 20 is drawn as a dashed line, and the solvent temperature $\hat{T}_S$ is indicated as a dotted line (see the text for explanation).

The dashed line in Figure 3 expresses the linear relation 20 between β_P and T_P. The required parameters were determined as $\alpha_P = -2.02k_B$ K/ps and $\tau_{SP} = 1.60$ ps from the simulation sets CHF.0 and CHF.2. Therefore, the dashed line linearly interpolates between the data points ($\hat{\beta}_P$,$\hat{T}_P$) of these two simulation sets. The above values of the parameters α_P and τ_{SP} closely agree with those calculated earlier (see Methods) for setting up the CHF thermostat used in simulation G/A/2.**P**.3. This result is expected because in both cases the parameters α_P and τ_{SP} were computed from simulations employing the same parameters.

In simulation set CHF.1, the peptide temperature was nearly identical to $\hat{T}_S$ with $\hat{T}_P = 299.5$ K (black square in Figure 3) because here the thermostat parameters were chosen equal to those of the simulation G/A/2.**P**.3 (series #1), which realizes the **P**.3 strategy. The temperature T_P predicted for CHF.1 by the dashed line deviates by only 0.5 K from the observed average. This deviation is probably significant because the temperature averages shown in the figure are extremely well converged ($\sigma_{\hat{T}_P} < 0.1$ K) due to the extended statistics. If a similar interpolation would be constructed using the data from the simulation sets CHF.3 or CHF.4 instead of CHF.2, the error in the prediction for CHF.1 would increase to 1.1 K or 2.2 K, respectively, with increasing violation of the approximate linear relation 20 between β_P and T_P. In the case of 8ALA in explicit water, the assumption of a linear thermal coupling between solvent

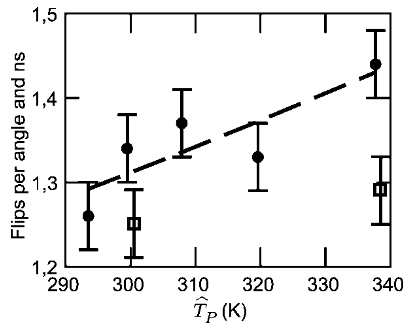


Figure 4. Temperature dependence of the peptide backbone dynamics of 8ALA. The graph shows the average number of transitions per angle and nanosecond of the ψ-dihedral angles between the α-type region [−60°,−30°] and the β-type region [95°, 145°] for the five CHF simulation sets (filled circles) and the two CLS sets (empty squares) over the observed average peptide temperature $\hat{T}_P$. The error bars give the range of plus/minus one standard deviation. Additionally, an Arrhenius[69] model (dashed line) fitted to the CHF data is plotted. The simulation parameters are summarized in Table 2.

and solute (eq 15), thus, obviously breaks down if $\hat{T}_P$ deviates by more than about 10 K from $\hat{T}_S$, which is probably also true for related simulation systems. In test simulations serving to set up a CHF thermostat through eq 19, the deviation $|\hat{T}_P - \hat{T}_S|$ should, thus, be smaller than about 10 K if one wants to guarantee an accurate tuning of T_P in applications of strategy **P**.3.

Backbone Dynamics. As we have seen further above, the use of an inappropriate strategy for temperature control can lead to peptide temperatures considerably deviating from that of the solvent. It seems likely that such a deviation can entail an altered conformational dynamics of the peptide. To check this expectation, we analyzed the second simulation series also in this respect. Due to the extremely slow thermostat coupling employed in CHF.0 to CHF.4, here, the dynamics should be exclusively affected by differences in the peptide temperatures.

Figure 4 shows how the kinetics of conformational transitions in 8ALA is modified by T_P in CHF.0−4 (black dots). This kinetics is measured by local flip rates of backbone torsional angles (see the figure caption). As expected, the flip rates increase with the temperature. A simple Arrhenius model[69] fitted to the CHF data is drawn as a dashed line. This model yields an energy barrier of $434k_B$ K for the backbone flips. This value is well in the range of typical barrier heights reported for biomolecules in the literature.[70]

Having estimated the influence of the temperature on the conformational dynamics of our sample peptide 8ALA in SPC water, it seems appropriate to check whether a separate classical BT (as frequently applied in strategy **P**.1) changes the dynamics. Here, particularly a slowing down seems possible because a rapidly coupled thermostat can interfere with long-lasting energy fluctuations within the peptide, which are caused by random in- and outflow of energy from the solvent. For the purpose of such a check, we carried out the simulation sets CLS.1−2 listed in Table 2, in which a classical BT separately coupled to P enforced temperatures $\hat{T}_P$ of about 300 K and 340 K, respectively.

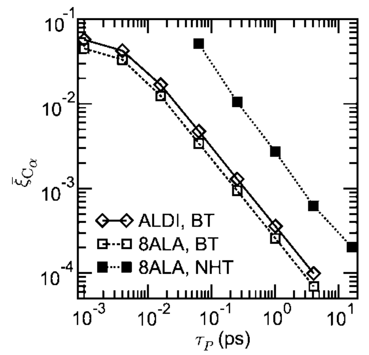


Figure 5. Root mean perturbation quotients $\bar{\xi}_{C\alpha}$ at the C_α atoms of 8ALA and ALDI evaluated from simulation series #3 for the NHTs and BTs, respectively, for different coupling times τ_P of the peptide thermostats.

Figure 4 compares the flip rates observed when using a classical Berendsen thermostat (open squares) with the data for the CHF thermostat (filled circles) and demonstrates that our expectation is actually met. Thus, if one wants to sample the equilibrium fluctuations of a peptide in solution by MD as rapidly as possible, or if one wants to gain access to the kinetics of nonequilibrium relaxation processes, the separate coupling of a classical BT to a small peptide seems counterproductive.

We interpret the above result by the following physical picture: A rapidly coupled BT likewise dampens fluctuations to higher and lower energies, thus leading to the correct average temperature. However, barrier crossings are enabled by rare accidental accumulations of a critical amount of energy in the respective collective coordinates. Particularly by dampening the higher energy fluctuations of the peptide, a classical BT makes such accumulations and, thus, barrier crossings less likely. Note that we have additionally checked the performance of a NHT in the same setting. We found no reduction of flip rates (data not shown) as could be expected for a thermostat maintaining the canonical energy fluctuations.

Local Perturbations of the Dynamics. The flips of backbone dihedral angles are collective movements and, therefore, are not directly related to the perturbation which a thermostat inflicts on the dynamics of individual atoms. To check the latter, we collected from simulation series #3 (cf. Table 3) all those forces acting on the C_α atoms of 8ALA which are required for the evaluation of the perturbation quotients (4). We carried out this data collection for BTs and NHTs with coupling times τ_P covering 4 orders of magnitude. In the case of the smaller ALDI model, we concentrated on the Berendsen approach.

Figure 5 shows the resulting perturbation ratios (4) evaluated using the approximate expression 6. As demonstrated by the squares marking the 8ALA results, the perturbations $\bar{\xi}_{C\alpha}$ are small for both thermostats and decrease over a wide range linearly with the inverse of τ_P. For the classical BT (τ_P = 0.1 ps) the $\bar{\xi}_{C\alpha}$ are only about 0.5%. Furthermore, the smaller ALDI model exhibits slightly larger $\bar{\xi}_{C\alpha}$ (open diamonds) than 8ALA (open squares). However, this size-induced difference is much smaller than that between the NHTs and BTs. At a given τ_P, Nosé-Hoover coupling inflicts perturbations which are by 1 order of magnitude larger than in the Berendsen case (cf. Figure 5). For a Berendsen coupling of maximal strength ($\tau_P = 0.001$ ps) the perturbation is comparable to that of a NHT with τ_P as large as 0.064 ps. Furthermore, for Nosé-Hoover coupling τ_P cannot be chosen larger than about 0.256 ps where $\bar{\xi}_{C\alpha}$ is about 1% and, thus, not particularly small. In the given case of 8ALA, one otherwise observes long-lasting and artificial temperature oscillations, i.e. the so-called Toda daemon[44] (data not shown).

One can compare the perturbations shown in Figure 5 to those which are inflicted by a CHF thermostat as employed in strategy **P**.3. In simulation G/A/2_**P**.3, the peptide 8ALA was kept at 300 K with a perturbation ratio of $\bar{\xi}_{C\alpha} \approx 10^{-4}$. As can be seen from Figure 5, this ratio corresponds to a Berendsen coupling time larger than 1 ps in the classical thermostat setup. However, a classical BT with $T_{0,P} = T_S$ and $\tau_p \geq 1$ ps cannot properly control the temperature because then τ_p is in the range of solvent−solute coupling time ($\tau_{SP} = 1.6$ ps), i.e. is too slow (cf. section 2). On the other hand, a more strongly coupled thermostat with $\tau_P = 0.1$ ps does the job, but then the perturbation is more than ten times stronger than for a CHF thermostat.

The above analysis was based on data for perturbation ratios derived through the approximate expression 6 and, therefore, depends on the validity of this equation. The first assumption made in the derivation (cf. section 2) of eq 6 was that the atomic velocities $\dot{\mathbf{r}}_i(t)$ and the thermostat variable $\gamma(t)$ are uncorrelated. We have checked this assumption for simulation series #3 by evaluating eq 5 with and without taking the correlation into account; the relative difference was less than 10^{-2} for both 8ALA and ALDI (data not shown).

The second assumption was that the individual atomic velocities $\dot{\mathbf{r}}_i(t)$ are drawn from an undisturbed Maxwell distribution and can be checked by comparing results of the exact expression 5 with results of the approximate expression 6. We evaluated these expressions for the trajectories of series #3 and determined the root-mean-square deviations. In the worst case of a BT at the maximum coupling strength ($\tau_P = 0.001$ ps), we found root-mean-square deviations amounting to 8.3% of the mean thermostatic force for 8ALA and to 14% for ALDI. In view of the moderate statistics provided by the 250 ps simulations employed in series #3, the estimate 6 is fairly reliable. Thus, eq 6 is adequate if one wants to estimate thermostatic forces.

Temperature Fluctuations. In our suggestion of the minimally invasive CHF thermostat characterizing strategy **P**.3 we were guided by the notion that a properly thermostatted explicit solvent system is a canonical heat bath for an uncontrolled solute. To check this assumption, we compare in Figure 6 the canonical χ^2-distribution (eq 7) for the instantaneous peptide temperature $\boldsymbol{T}_P(t)$ with results from simulation G/A/2_**P**.3. For the 103 degrees of freedom of 8ALA, the χ^2-distribution (solid line) resembles a Gaussian (dashed line), which is expected for very large systems. Remarkably, the MD results (circles) closely reproduce the slight asymmetry of the χ^2-distribution. This agreement strongly indicates that the peptide has sampled the canonical ensemble in the simulation G/A/2_**P**.3. We have verified this result for a series of further CHF simulations. It did not change for larger solvent systems and was independent of the coupling time for the solvent thermostat provided that the solvent temperature remained well-tuned (data not shown).

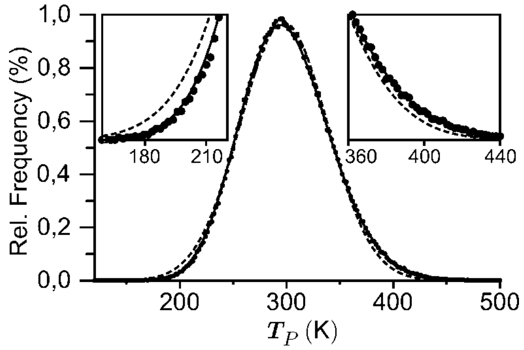


Figure 6. Distribution of the instantaneous temperature $\boldsymbol{T}_P(t)$ of 8ALA (in SPC water at 300 K) during the 20 ns MD simulation G/A/2_**P**.3 (dots). The dashed line is a Gaussian fit to the data. The canonical distribution (eq 7) is drawn as a solid line.

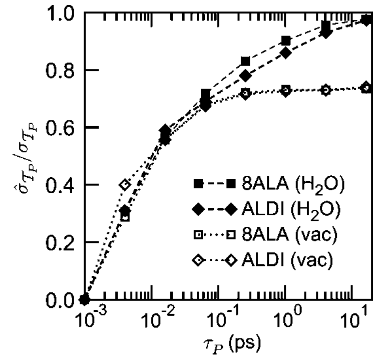


Figure 7. Ratio $\hat{\sigma}_{T_P}/\sigma_{T_P}$ of measured and canonical temperature fluctuations for various coupling times τ_P of a Berendsen solute thermostat. The model peptides are 8ALA (squares) and ALDI (diamonds). Simulations were performed in explicit water (H$_2$O, filled symbols) and vacuum (vac, empty symbols) for both peptides. Simulation parameters are given in Table 4.

To estimate how a classical BT separately coupled to a peptide (strategy **P**.1) affects its global statistical properties, we determined the temperature fluctuations of the peptides 8ALA and ALDI, respectively, as measured by the standard deviation $\hat{\sigma}_{T_P}$ in a fourth series of simulations (for details see Table 4). Figure 7 shows the ratio of $\hat{\sigma}_{T_P}$ and σ_{T_P}, which is the value theoretically expected for a canonical ensemble and is given by eq 8. For peptides in explicit solvent the figure shows that $\hat{\sigma}_{T_P}/\sigma_{T_P}$ is always smaller than one and approaches that limit for large τ_P. Thus, in the classical setting ($\tau_P = 10^{-2}$ ps) a BT strongly suppresses the canonical temperature fluctuations. These fluctuations successively become restored with increasing τ_P. The full range of canonical fluctuations is reached at coupling times $\tau_P > 10$ ps, i.e. at values exceeding the solvent-peptide heat coupling time τ_{SP} by a factor of 10. As a result, the separate BT is effectively disconnected from the peptide, the solute−solvent heat exchange term β_{SP} dominates the heat balance eq 19, and strategy **P**.1 reduces to the noninvasive strategy **P**.2.

Figure 7 not only reveals the general suppression of temperature fluctuations within a peptide by a classical BT but also demonstrates through a comparison with vacuum simulation data that these fluctuations are caused (i) by a fast exchange of kinetic and potential energy within a peptide and (ii) by a slower energy exchange with the solvent.

In vacuum simulations, the exchange of kinetic and potential energy within the peptide is the only cause of temperature fluctuations. As shown by the data, a rapidly coupled Berendsen thermostat ($\tau_P < 0.1$ ps) suppresses these microcanonical fluctuations in the same way as it suppresses the canonical temperature fluctuations of a solvated peptide. However, at slower coupling times τ_P the thermostat is seen to no longer affect the microcanonical fluctuations. The clear saturation of $\hat{\sigma}_{T_P}/\sigma_{T_P}$ at $\tau_P > 0.1$ ps demonstrates that the microcanonical fluctuations occur on time scales below 0.1 ps. In contrast, additional fluctuations of a solvated peptide are still suppressed by the thermostat with even slower coupling. Thus, as claimed above, they occur on longer time scales.

In order to retain the correct statistics for the solute, it is important to choose the coupling time τ_P for the thermostat longer than the typical time scale of the canonical fluctuations, which, in our case, is in the range of 10 ps, as can be seen from Figure 7. However, this time may even be longer for more weakly coupling solvents or larger solutes.

5. Conclusions

Every thermostat changes the dynamics of the controlled system to a larger or lesser extent. Measured on a microscopic scale, these changes are by about 1 order of magnitude smaller for BTs than for NHTs (cf. the data on the perturbation quotients displayed in Figure 5). On the other hand, NHTs, in contrast to BTs, guarantee the canonical ensemble. For instance, as shown by the results on the temperature fluctuations (Figure 7), BTs suppress all those canonical energy fluctuations which are slower than the time scale τ at which the BT is coupled to the system.

Whether such changes can modify the specific observables to be extracted from a simulation and to be compared with experimental data is *a priori* unclear in many cases. Even if one suspects that a given thermostat could possibly introduce an artifact into the computation of a certain observable, one may have to spend an enormous computational effort for a statistically clear proof. In fact, to prove a suspected dampening of peptide flip rates by a standard BT, we had to spend about 400 ns of simulation time on each of the data points to get the statistical certainty shown in Figure 4.

Especially if the popular strategy **P**.1 is applied to a solute−solvent system, the specific drawbacks of the various thermostat algorithms may directly affect the properties of the solute. The **P**.1 strategy with a BT is expected to cause artifacts of type a), i.e. artifacts resulting from an incorrect thermodynamical ensemble. In fact, as we have shown for a sample peptide, the dampening of the canonical energy fluctuations due to the BT can lead to reduced peptide flip rates. Furthermore, one expects that the combination of **P**.1 with the NHT will render the solute vulnerable to artifacts of type b), i.e. lacking ergodicity. Using the **P**.1 strategy with other thermostats which suffer neither from type a) nor type b) drawbacks (e.g., the Nosé-Hoover chain) still perturbs the dynamics much more strongly than necessary, i.e. such a strategy is prone to introduce artifacts of type c) (dynamics).

Given the need for some sort of temperature control in large scale MD simulations of complex systems, the optimal strategies to avoid artifacts of types a), b), and c) are **P**.2 or **P**.3, respectively. Here, the minimally invasive strategy **P**.3, which employs a constant heat flow to compensate the algorithmic heat production in the solute, has to be applied only if the noninvasive strategy **P**.2 turns out to be ineffective in a sufficiently extended test simulation. Strategy **P**.3 reduces the perturbation of the solute's dynamics to a minimum while keeping it nevertheless properly tempered. The precise protocol to set up a **P**.3 scheme is given in the Appendix.

The preservation of the canonical ensemble within the solute through strategies **P**.2 and **P**.3 (despite the use of a standard BT for the solvent which strongly perturbs the temperature fluctuations in this part of the system) is the most important result of this paper and proves our hypothesis that an explicitly simulated solvent of the correct temperature T_S represents the optimal thermostat for a solute. Admittedly, our quantitative analysis of the applicability of strategies **P**.2 and **P**.3 is restricted to relatively small peptides because an extended statistics is required for reliable results. Already for the small peptides with their short temperature autocorrelation times of 15 ps, it takes more than 10 ns to determine the average temperature with an accuracy of 1 K. For larger systems, the temperature autocorrelation times increase and so do the simulation times required for accurate temperature measurements. Too short simulations can easily lead to the false impression that the solute temperature sizably differs from the solvent temperature. To our experience, the noninvasive strategy **P**.2 can suffice for quite large solvent−solute systems. For instance, reinspecting a simulation^8 of the C-terminal domain of the human prion protein (residues 125−228), which employed a global thermostat coupling (strategy **G**), we found that the protein temperature deviated by more than 10 K from that of the solvent. Subsequent simulations of a slightly larger fragment (residues 114−228), which employed strategy **P**.2 but otherwise the same simulation setup, showed no significant temperature difference. In the few cases in which one observes a seemingly intolerable temperature difference between solute and solvent, one can still use the solvent as the heat bath by applying the minimally invasive strategy **P**.3 to keep the solute well tempered.

It should be noted that our heat flow model and the associated setup protocol for the constant heat flow strategy **P**.3 are restricted to two subsystems with homogeneous local algorithmic heating rates. For simulations of more complex systems such as protein-DNA assemblies in solution, for which one expects more than two different heating rates, a constant heat flow strategy can be analogously designed. However, it will become increasingly difficult to determine the local heating rates of the various subsystems which have to be compensated.

Acknowledgment. This work was supported by the Bavarian joint research project ForPrion (LMU02), by the VolkswagenStiftung (I/79 884), and by Deutsche Forschungsgemeinschaft (SFB533/C1, SFB749/C4).

Appendix: Setting up Strategies P.2 and P.3

Here, we give a detailed description of the steps needed in order to set up a simulation system containing a macromolecule P in thermal equilibrium with an explicit solvent enviroment S according to the strategies **P**.2 and **P**.3, respectively, using the standard Berendsen algorithm. After preparation (e.g., removal of close solvent−solute contacts by energy minimization), the following steps are necessary:

a) Heating phase: The subsystems are heated using two separate classical BTs (e.g., $\tau_S = \tau_P = 0.1$ ps) to the temperature T_{sim} desired in the production simulation. Depending on the initial deviations of the solute temperature T_P and solvent temperature T_S, it may take a simulation time of up to $30\tau_{S/P}$ for the respective subsystems to safely attune to T_{sim}.

b) Relaxation phase I: The solute is decoupled from its thermostat ($\tau_P = \infty$) and relaxes to its new steady state temperature $T_{P,1}$. The time constant for the relaxation to the steady state is the solvent−solute coupling time τ_{SP}. Since τ_{SP} is still unknown, an upper limit estimate (e.g., $\tau_{SP} \approx 20$ ps) should be used to determine the relaxation time $t_{\mathrm{relax}} \approx 10\tau_{SP}$.

c) Test simulation I: Here, the solute remains decoupled from its thermostat and the simulation serves to determine its average temperature $T_{P,1}$. If the deviation from equilibrium measured by $|T_{P,1} - T_{\mathrm{sim}}|$ is less than an acceptable tolerance ΔT_P, then the noninvasive strategy **P**.2 is applicable, and one may directly continue the simulation for data production f).

The necessary simulation time t_1 for the test depends on the tolerable uncertainty $\sigma^2_{\hat{T}_{P,1}}$ of the measured solute temperature $\hat{T}_{P,1}$, which forms an upper bound for the uncertainty $\sigma^2_{T_P}$ in the prediction of the production run temperature T_P. If ΔT_P is the accuracy required for the prediction, we should make sure that $\sigma^2_{\hat{T}_{P,1}} \leq \Delta T_P^2$. By eq 9 the simulation time then is $t_1 = 2\tau_c\sigma^2_{T_P}/\Delta T_P^2$, where τ_c is the temperature autocorrelation time of the solute, and σ_{T_P} is the standard deviation of its temperature fluctuations, which were observed during the test run. One typically obtains simulation times of several nanoseconds.

d) Relaxation phase II: The solute is coupled to a separate thermostat with a coupling time $\tau_P \geq 500$ ps intended for the **P**.3 production run. Using an estimate for τ_{SP} (e.g., 1 ps), a reasonable choice for the target temperature is given by $T_{0,P,2} = -\tau_P/\tau_{SP}\cdot|T_{P,1} - T_{\mathrm{sim}}|$ (leading to 2-fold overcompensation if τ_{SP} was exact). The duration of this relaxation phase is the same as in step b).

e) Test simulation II: The average temperature $T_{P,2}$ is determined. The simulation time t_2 should be equal to t_1 in step c).

f) Production simulation: If strategy **P**.2 turned out to be applicable in step c), the settings in this simulation are chosen identically (in fact, one may regard the test run as the initial part of the production simulation). Otherwise, the target temperature $T_{0,P}$ for a **P**.3 simulation is determined from the two test simulations by

$$T_{0,P} = T_{\mathrm{sim}} + \frac{T_{0,P,2} - T_{P,2}}{T_{P,2} - T_{P,1}}(T_{\mathrm{sim}} - T_{P,1}) \qquad (22)$$

References

(1) Karplus, M.; McCammon, J. A. *Nat. Struct. Biol.* **2002**, *9*, 646–652.

(2) Gnanakaran, S.; Nymeyer, H.; Portman, J.; Sanbonmatsu, K. Y.; Garca, A. E. *Curr. Opin. Struct. Biol.* **2003**, *13*, 168–174.

(3) Grubmüller, H.; Heymann, B.; Tavan, P. *Science* **1996**, *271*, 997–999.

(4) Rief, M.; Oesterhelt, F.; Heymann, B.; Gaub, H. E. *Science* **1997**, *275*, 1295–1297.

(5) Brockwell, D. J.; Paci, E.; Zinober, R. C.; Beddard, G. S.; Olmsted, P. D.; Smith, D. A.; Perham, R. N.; Radford, S. E. *Nat. Struct. Biol.* **2003**, *10*, 731–737.

(6) Kucera, K.; Lambry, J. C.; Martin, J. L.; Karplus, M. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 5805–5807.

(7) Spörlein, S.; Carstens, H.; Satzger, H.; Renner, C.; Behrendt, R.; Moroder, L.; Tavan, P.; Zinth, W.; Wachtveitl, J. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 7998–8002.

(8) Hirschberger, T.; Stork, M.; Schropp, B.; Winklhofer, K. F.; Tatzelt, J.; Tavan, P. *Biophys. J.* **2006**, *90*, 3908–3918.

(9) Mayor, U.; Guydosh, N. R.; Johnson, C. M.; Grossmann, J. G.; Sato, S.; Jas, G. S.; Freund, S. M. V.; Alonso, D. O. V.; Daggett, V.; Fersht, A. R. *Nature* **2003**, *421*, 863–867.

(10) Shea, J.-E.; Brooks, C. L. I. *Annu. Rev. Phys. Chem.* **2001**, *52*, 499–535.

(11) Privalov, P. L. Physical basis of the stability of the folded conformations of Proteins. In *Protein Folding*; Creighton, T. E., Ed.; W. H. Freeman: San Francisco, 1992.

(12) Munishkina, L. A.; Phelan, C.; Uversky, V. N.; Fink, A. L. *Biochemistry* **2003**, *42*, 2720–2730.

(13) Vitkup, D.; Ringe, D.; Petsko, G. A.; Karplus, M. *Nat. Struct. Biol.* **2000**, *7*, 34–38.

(14) Tavan, P.; Carstens, H.; Mathias, G. Molecular Dynamics Simulations of Proteins and Peptides: Problems, Achievements, and Perspectives. In *Protein Folding Handbook. Part I*; Buchner, J., Kiefhaber, T., Eds.; Wiley-VCH: 2005.

(15) Frauenfelder, H.; Petsko, G. A.; Tsernoglou, D. *Nature* **1979**, *280*, 558–563.

(16) Ishima, R.; Torchia, D. A. *Nat. Struct. Biol.* **2000**, *7*, 740–743.

(17) Ichiye, T.; Karplus, M. *Biochemistry* **1983**, *22*, 2884–2893.

(18) Lazaridis, T.; Karplus, M. *Biophys. Chem.* **2003**, *100*, 367–395.

(19) Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **1999**, *314*, 141–151.

(20) Hansmann, U. H. E. *Chem. Phys. Lett.* **1997**, *281*, 140–150.

(21) Liu, P.; Kim, B.; Friesner, R. A.; Berne, B. J. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 13749–13754.

(22) Carstens, H. Konformationsdynamik lichtschaltbarer Peptide: Molekulardynamiksimulationen und datengetriebene Modell-

bildung, Dissertation, Fakultät für Physik, Ludwig-Maximilians-Universität München, 2004.

(23) Koper, G. J. M.; Reiss, H. *J. Phys. Chem.* **1996**, *100*, 422–432.

(24) Swope, W. C.; Pitera, J. W.; Suits, F. *J. Phys. Chem. B* **2004**, *108*, 6571–6581.

(25) Swope, W. C.; Pitera, J. W.; Suits, F.; Pitman, M.; Eleftheriou, M.; Fitch, B. G.; Germain, R. S.; Rayshubski, A.; Ward, T. J. C.; Zhestkov, Y.; Zhou, R. *J. Phys. Chem. B* **2004**, *108*, 6582–6594.

(26) Nutt, D. R.; Smith, J. C. *J. Chem. Theory Comput.* **2007**, *3*, 1550–1560.

(27) Mathias, G.; Egwolf, B.; Nonella, M.; Tavan, P. *J. Chem. Phys.* **2003**, *118*, 10847–10860.

(28) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089−10092.

(29) Grubmüller, H.; Tavan, P. *J. Comput. Chem.* **1998**, *19*, 1534–1552.

(30) Kraeutler, V.; van Gunsteren, W. F.; Hünenberger, P. H. *J. Comput. Chem.* **2001**, *22*, 501–508.

(31) Sagui, C.; Darden, T. A. *Annu. Rev. Biophys. Biomol. Struct.* **1999**, *28*, 155–179.

(32) Oda, K.; Miyagawa, H.; Kitamura, K. *Mol. Simul.* **1996**, *16*, 167–177.

(33) Cheng, A.; Merz, K. M. *J. Phys. Chem.* **1996**, *100*, 1927–1937.

(34) Guenot, J.; Kollman, P. A. *Protein Sci.* **1992**, *1*, 1185–1205.

(35) Arnold, G. E.; Ornstein, R. L. *Proteins* **1994**, *18*, 19–33.

(36) Hünenberger, P. H. *Adv. Polym. Sci.* **2005**, *173*, 105–149.

(37) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Di Nola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684–3690.

(38) Harvey, S. C.; Tan, R. K.-Z.; Cheatham, T. E. *J. Comput. Chem.* **1998**, *19*, 726–740.

(39) Chui, S.-W.; Clark, M.; Subramaniam, S.; Jakobsson, E. *J. Comput. Chem.* **2000**, *21*, 121–131.

(40) Morishita, T. *J. Chem. Phys.* **2000**, *113*, 2976–2982.

(41) Nosé, S. *J. Chem. Phys.* **1984**, *81*, 511–519.

(42) Hoover, W. G. *Phys. Rev. A* **1985**, *31*, 1695–1697.

(43) Tuckerman, M. E.; Liu, Y.; Ciccotti, G.; Martyna, G. J. *J. Chem. Phys.* **2001**, *115*, 1678–1702.

(44) Holian, B. L.; Voter, A. F.; Ravelo, R. *Phys. Rev. E* **1995**, *52*, 2338–2347.

(45) Posch, H. A.; Hoover, W. G.; Vesely, F. J. *Phys. Rev. A* **1986**, *33*, 4253–4265.

(46) D'Alessandro, M.; Tenenbaum, A.; Amadei, A. *J. Phys. Chem. B* **2002**, *106*, 5050–5057.

(47) Tobias, D. J.; Martyna, G. J.; Klein, M. L. *J. Phys. Chem.* **1993**, *97*, 12959–12966.

(48) Martyna, G. J.; Klein, M. L.; Tuckerman, M. *J. Chem. Phys.* **1992**, *97*, 2635–2643.

(49) Bussi, G.; Donadio, D.; Parrinello, M. *J. Chem. Phys.* **2007**, *126*, 014101(1−7)

(50) Allen, M. P.; Tildesley, D. J. *Computer Simulations of Liquids*; Oxford University Press: Oxford, 1987.

(51) Stange, K. *Angewandte Statistik*; Springer: Heidelberg, 1970.

(52) Weber, W.; Hünenberger, P. H.; McCammon, J. A. *J. Phys. Chem. B* **2000**, *104*, 3668–3675.

(53) Daura, X.; Gademann, K.; Schäfer, H.; Jaun, B.; Seebach, D.; van Gunsteren, W. F. *J. Am. Chem. Soc.* **2001**, *123*, 2393–2404.

(54) Villareal, M. A.; Montich, G. G. *J. Biomol. Struct. Dyn.* **2005**, *23*, 135–142.

(55) van den Bosch, M.; Swart, M.; Snijders, J. G.; Berendsen, H. J.; Mark, A. E.; Oostenbrink, C.; van Gunsteren, W. F.; Canters, G. W. *ChemBioChem* **2005**, *6*, 738–746.

(56) Monticelli, L.; Simoes, C.; Belvisi, L.; Colombo, G. *J. Phys.: Condens. Matter* **2006**, *18*, S329–S345.

(57) Fox, T.; Kollman, P. A. *Proteins* **1996**, *25*, 315–334.

(58) Kong, Y.; Ma, J.; Karplus, M.; Lipscomb, W. N. *J. Mol. Biol.* **2006**, *356*, 237–247.

(59) Lindahl, E.; Hess, B.; van der Spoel, D. *J. Mol. Model.* **2001**, *7*, 306–317.

(60) Niedermeier, C.; Tavan, P. *J. Chem. Phys.* **1994**, *101*, 734–748.

(61) Niedermeier, C.; Tavan, P. *Mol. Simul.* **1996**, *17*, 57–66.

(62) Mathias, G.; Tavan, P. *J. Chem. Phys.* **2004**, *120*, 4393–4403.

(63) Eichinger, M.; Grubmüller, H.; Heller, H.; Tavan, P. *J. Comput. Chem.* **1997**, *18*, 1729–1749.

(64) Scott, W. R. P.; Hünenberger, P. H.; Tironi, I. G.; Mark, A. E.; Billeter, S. R.; Fennen, J.; Torda, A. E.; Huber, T.; Krüger, P.; van Gunsteren, W. F. *J. Phys. Chem. A* **1999**, *103*, 3596–3607.

(65) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Hermans, J. *Interaction Models for Water in Relation to Protein Hydration. In Intermolecular Forces*; Pullman, B., Ed.; D. Reidel Publishing Company: Dordrecht, 1981.

(66) MacKerell, A. D.; et al. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.

(67) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.

(68) Guenot, J.; Kollman, P. A. *J. Comput. Chem.* **1993**, *14*, 295–311.

(69) Gardiner, C. W. *Handbook of Stochastic Methods*, 2nd ed.; Springer: Berlin, 1985.

(70) Zuckerman, D. M.; Lyman, E. *J. Chem. Theory Comput.* **2006**, *2*, 1200–1202.

CT8000365

5 Zusammenfassung und Ausblick

In dieser Arbeit wurden die Eigenschaften biomolekularer Systeme unterschiedlicher Größe mit an die Systemgröße angepassten theoretischen Methoden untersucht. Diese Methoden wurden in Kapitel 1 eingeführt. Kapitel 2 enthielt dann den Abdruck einer Publikation [14], in der die Dichtefunktionaltheorie eingesetzt wurde um aufzuklären, welche Faktoren die Konformationseigenschaften des 4-Mercaptoprolins und verwandter Prolinderivate bestimmen. In Kapitel 3 fand der Übergang zu größeren Systemen statt, zu deren Beschreibung auf die Methode der molekularmechanischen Molekulardynamik-Simulation zurückgegriffen wurde. In diesem Kapitel wurde ein Manuskript [33] präsentiert, in dem die Eigenschaften helikaler Peptide untersucht wurden. In der in Kapitel 4 abgedruckten Publikation [145] wurde schließlich eine methodische Fragestellung zum Thema der Temperaturkontrolle in MD-Simulationen diskutiert.

Bevor ich unmittelbar anschließend die Hauptergebnisse der drei letzten Kapitel noch einmal zusammenfasse, möchte ich anmerken, dass ich noch an einer weiteren Publikation als Koautor beteiligt war. Es handelt sich um einen im *Journal of Physical Chemistry B* (Band 112, S. 12217-12230) im Jahr 2008 erschienenen Artikel mit dem Titel „A Polarizable Force Field for Computing the Infrared Spectra of the Polypeptide Backbone“, für den meine Doktorandenkollegen Verena Schultheis und Bernhard Schropp, ich selbst, sowie der Leiter unserer Arbeitsgruppe Paul Tavan als Autoren verantwortlich zeichnen. Zu dieser Arbeit habe ich die Ergebnisse einer Reihe von quantenmechanischen Rechnungen beigesteuert. Mit diesen Rechnungen wurden die Schwingungsspektren des Moleküls N-Methyl-Acetamid, das ein Modell der Amidgruppen darstellt, genauer charakterisiert. Mit diesen Resultaten habe ich wichtiges Datenmaterial für die Entwicklung eines neuartigen polarisierbaren Kraftfelds für Amidgruppen geliefert.

Mercaptoprolin

In der Natur kommt die Aminosäure Prolin oft in modifizierter Form vor. In Kapitel 1 wurde das Protein Kollagen angesprochen, das viele Prolinreste enthält, in denen eines der an den Pyrrolidinring gebundenen Wasserstoffatome (H) in einem posttranslationalen Prozess durch eine Hydroxylgruppe (OH) ersetzt wird. Diese Prolinreste werden dann als Hydroxyprolin (Hyp) bezeichnet. Dessen häufigste Variante ist das (4R)-Hydroxyprolin, in dem die Hydroxylgruppe eines der beiden Wasserstoffatome am C^{γ}-Atom des Prolins ersetzt. Wird das andere Wasserstoffatom ersetzt, erhält man das (4S)-Hydroxyprolin (vgl. Abbildung 1 in Kapitel 2). Es ist seit längerem bekannt, dass diese beiden Epimere unterschiedliche Konformationen des Pyrrolidinrings bevorzugen. Während (4R)-Hyp den C^{γ}-*exo*-Ringknick („*exo* pucker“) bevorzugt, bevorzugt (4S)-Hyp den C^{γ}-*endo*-Ringknick („*endo* pucker“). Die gleichen Präferenzen zeigen die $4R$- und $4S$-Epimere des Fluoroprolins (Flp), das statt der Hydrodroxylgruppe des Hydroxyprolins ein Fluoratom enthält.

In Kapitel 2 wurden nun erstmals die Struktureigenschaften der $4R$- und $4S$-Epimere des Mercaptoprolins (Mpc) untersucht, das statt der Hydroxyl- eine Thiolgruppe (SH) enthält. Durch NMR-Spektroskopie wurde gezeigt, dass dieser Austausch der Hydroxylgruppe durch die weniger elektronegative Thiolgruppe zu drastisch veränderten Konformationseigenschaften führt: Das $4R$-Epimer des Mpc bevorzugt den C^{γ}-*endo*-Ringknick, das $4S$-Epimer den C^{γ}-*exo*-Ringknick. Verglichen mit Hyp und Flp zeigen die beiden Epimere des Mpc also genau die entgegengesetzten Präferenzen. Diese Beobachtung ist wichtig für das Design von Peptiden und Proteinen, in denen die Thiolgruppe zur Thioether- oder Disulfidverknüpfung innerhalb oder zwischen Peptidketten genutzt werden soll.

Um herauszufinden, wodurch diese Unterschiede in den Konformationseigenschaften erzeugt werden, wurden die Modellmoleküle, die auch experimentell betrachtet worden waren, mit Hilfe der Dichtefunktionaltheorie analysiert. In diesen Rechnungen konnte insbesondere die unerwartete, aber experimentell beobachtete Neigung des ($4R$)-Mpc zum C^{γ}-*endo*-Ringknick und des ($4S$)-Mpc zum C^{γ}-*exo*-Ringknick reproduziert werden. Die DFT-Rechnungen belegen außerdem, dass die Bevorzugung eines Knick-Typs für eine gegebene 4-Substitution hauptsächlich durch die Wechselwirkung des Bindungsdipols an der Substitutionsstelle mit dem Dipol der vorhergehenden Amidbindung bestimmt wird. Die Kombination von ($4R$)-Hyp oder ($4R$)-Flp mit einem C^{γ}-*endo*-Ringknick führt zu einer ungünstigen antiparallelen Ausrichtung der Dipole, wohingegen der C^{γ}-*exo*-Ringknick fast orthogonale Dipole aufweist, deren Wechselwirkung kaum zur Gesamtenergie beiträgt. Daher ist der C^{γ}-*exo*-Ringknick für ($4R$)-Hyp- und ($4R$)-Flp-Derivate günstiger. Ganz anders ist die Situation im Mpc. Dort ist das Dipolmoment an der Substitutionsstelle sehr schwach und bewirkt deshalb für die $4R$-*endo*-Kombination kaum energetische Nachteile. Die beim Mpc beobachtete Bevorzugung der $4R$-*endo*-Geometrie impliziert dann, dass die Summe aller Energiebeiträge (Elektrostatik, Van-der-Waals, kovalente Geometrie usw.) ohne die besprochene Dipol-Dipol-Wechselwirkung insgesamt leicht zugunsten des C^{γ}-*endo*-Ringknicks für das untersuchte ($4R$)-Mpc-Derivat ausfällt. Der C^{γ}-*exo*-Ringknick von ($4R$)-Hyp und ($4R$)-Flp resultiert somit eher aus einer weniger ungünstigen Dipol-Dipol-Orientierung als aus einer spezifischen günstigen Wechselwirkung. Für andere $4R$- oder $4S$-Substitutionen mit kleinem Bindungsdipol würde man, analog zu Mpc, deshalb auch die $4R$-*endo*- und die $4S$-*exo*-Kombination erwarten, obwohl diese bisher als ungewöhnlich und ungünstig galten.

Die in Kapitel 2 vorgestellte Arbeit hat einen klaren Zusammenhang zwischen 4-

Substitutionen des Prolins und den bevorzugten Geometrien des Prolinrings aufgezeigt. Die Möglichkeit, die Prolingeometrie durch Wahl geeigneter Analoga gezielt einzustellen, zusammen mit der wichtigen Rolle des Prolinrests in Peptid- und Proteinstrukturen, kann ein sehr wirkungsvolles Werkzeug zum Design und für Faltungsstudien von Peptiden sein.

Helikale Peptide

Das Ziel des in Kapitel 3 vorgestellten Manuskripts war die Untersuchung des lichtinduzierten Faltungsprozesses eines helikalen Peptids. Dazu sollte ein Peptid entwickelt werden, dessen Faltungsprozess sowohl mit experimentellen Methoden als auch in Molekulardynamik-Simulationen untersucht werden kann. Als die größten Hindernisse auf dem Weg zu realistischen Simulationen eines derartigen Prozesses wurden in Kapitel 3 zwei Probleme identifiziert: Zum einen ist es trotz der Fortschritte auf dem Gebiet der Computertechnik nach wie vor schwierig, den Konfigurationsraum selbst eines kleinen Peptids hinreichend umfassend abzutasten, um z.B. Aussagen über seine Konformationseigenschaften treffen zu können. Dieser Sachverhalt wird oft als *Sampling*-Problem bezeichnet. Dieses Problem wird verschärft, wenn, wie in den Simulationen in Kapitel 3, zur Beschreibung der Lösungsmittelumgebung des Peptids ein explizites Lösungsmittelmodell eingesetzt wird, wodurch die Systemgröße und damit der Rechenaufwand stark steigen. Auf den Einsatz eines expliziten Lösungsmittelmodells kann derzeit, wie in Kapitel 3 erläutert wurde, jedoch wegen der eingeschränkten Anwendbarkeit der bislang vorgeschlagenen impliziten Lösungsmittelmodelle nicht verzichtet werden. Zum anderen ist in jeder Simulation, in der ein molekularmechanisches Kraftfeld verwendet wird, unklar, wie sich die dem Kraftfeld inhärenten Näherungen auf die Eigenschaften des in der Simulation betrachteten molekularen Systems auswirken. In Kapitel 3 wurde nun diskutiert, wie sich dieses Kraftfeld-Problem und das oben erwähnte Sampling-Problem auf die Beschreibung des Faltungs/Entfaltungs-Gleichgewichts von helikalen Peptiden auswirken.

Anhand einiger Beispiele wurde gezeigt, dass zur Charakterisierung derartiger Gleichgewichtsensembles bei einer Temperatur von 300 K konventionelle MD-Simulationen auf einer Zeitskala von 100 ns nicht ausreichen. Vielmehr muss eine der sogenannten *enhanced sampling*-Techniken eingesetzt werden. Unsere Wahl fiel mit REST auf eine bisher selten eingesetzte aber speziell für Simulationen von Peptiden in Lösung entwickelte Methode aus der Familie der *replica exchange*-Methoden [144]. Wir konnten zeigen, dass sich mit REST der Konfigurationsraum helikaler Peptide effizient abtasten lässt. Ausserdem konnten wir zeigen, dass diese Technik eine einfache näherungsweise Bestimmung der thermischen Schmelzkurven helikaler Peptide ermöglicht. In diesem Zusammenhang wurde von uns durch die Kombination der herkömmlichen temperaturbasierten *replica exchange*-Methode (TREM) mit REST ein neues Verfahren (TREM/REST) entwickelt, das die jeweiligen Vorzüge der beiden Methoden vereint.

Um das Konvergenzverhalten der REST-Simulationen zu untersuchen, wurden für zwei unserer Modellpeptide ($\tilde{\mathrm{P}}$ und $\mathrm{P}_{\mathrm{pHP}}$, vgl. Kapitel 3) je zwei REST-Simulationen durchgeführt. Jeweils eine Simulation ging von gefalteten Startstrukturen aus, die jeweils andere von entfalteten Strukturen. Durch einen Vergleich der in diesen REST-Simulationen betrachteten erweiterten Ensembles konnte dann die Konvergenz der Simulationen beurteilt werden. Um die Ensembles miteinander vergleichen zu können, muss eine geeignete Observable gewählt werden. Da wir uns für die Faltungseigenschaften unserer Peptide interessieren, fiel unsere Wahl auf den Helixgehalt als Observable. Für die Konvergenzuntersuchungen wurde dann schließlich der Helixgehalt der erweiterten Ensembles miteinander verglichen sowie der der Replikate bei 300 K. Aus diesem Vergleich wurde für beide Peptide eine anfängliche Relaxationsphase von 10 ns Dauer identifiziert. Nach dieser Zeit hing der beobachtete Helixgehalt nicht mehr von der Wahl der Startstrukturen ab.

Zumindest für die von uns betrachteten Peptide ermöglicht die REST-Methode also die Erzeugung von Gleichgewichtsensembles auf einer Zeitskala von einigen 10 ns und löst somit das Sampling-Problem. Dabei ist jedoch zu beachten, dass sich auf diese Weise nur die Gleichgewichtseigenschaften eines molekularen Systems, nicht jedoch dessen Dynamik untersuchen lassen. Zur Untersuchung des lichtinduzierten Faltungsprozesses ist die REST-Methode demnach nicht geeignet. Dagegen können die in den REST-Simulationen erzeugten Gleichgewichtsensembles dazu verwendet werden, das Kraftfeld-Problem zu analysieren.

Schnell hatte sich nämlich gezeigt, dass auch unsere Peptide, vor allem das Modellpeptid P_{E} (vgl. Tabelle 1 in Kapitel 3), anfällig für ein Artefakt des CHARMM22-Kraftfelds (C22) sind. In mehreren konventionellen MD-Simulationen von P_{E} bei 300 K fand während der ersten Nanosekunden eine Umfaltung der ursprünglichen α-Helizes in sehr stabile π-Helizes statt. Somit konnten wir bestätigen, dass das Kraftfeld C22 im Allgemeinen nicht zur Untersuchung α-helikaler Peptide geeignet ist. Als Konsequenz wurde die CMAP-Erweiterung zu C22 (vgl. Abschnitt 1.3.2 und Anhang A) in unserem MD-Simulationsprogramm EGO-MMVI implementiert. In den REST-Simulationen unserer Peptide wurde dann dieses erweiterte Kraftfeld C22/CMAP eingesetzt. Die Analyse der in diesen Simulationen generierten Gleichgewichtsensembles zeigte schließlich, dass dieses Kraftfeld für die von uns betrachteten Peptide nun tatsächlich α-helikale Strukturen erzeugt. Für einige dieser Peptide konnte der Helixgehalt auch durch CD-Messungen bestimmt werden. Der Vergleich dieser experimentell ermittelten Werte mit den entsprechenden Werten aus unseren Simulationen zeigte jedoch klar, dass das Kraftfeld C22/CMAP α-helikale Peptidstrukturen deutlich überstabilisiert. Dass C22/CMAP zu einer Überstabilisierung von α-Helizes neigen könnte, war zwar im Zusammenhang mit Untersuchungen einiger Dipeptide und Peptide in impliziter Lösungsmittelumgebung schon vermutet worden [146, 147], konnte für helikale Peptide in expliziter Lösungsmittelumgebung von uns jedoch zum ersten Mal systematisch nachgewiesen werden. Wir konnten also zeigen, dass C22/CMAP gegenüber C22 zwar einen eindeutigen qualitativen

Fortschritt für die Beschreibung des Faltungs/Entfaltungs-Gleichgewichts α-helikaler Peptide in MD-Simulationen darstellt, gleichzeitig aber noch immer erhebliche quantitative Defizite aufweist.

Unabhängig von den bisher diskutierten Aspekten bilden die im Rahmen dieser Arbeit erzeugten Simulationsdaten einen wertvollen Datensatz, der zur Validierung neuer Methoden herangezogen werden kann. So ist es z.B. bei der Entwicklung von Kontinuumsmethoden [148–151] zur impliziten Beschreibung eines Lösungsmittels wichtig, auf statistisch valide Referenzdaten zurückgreifen zu können. Die in den REST-Simulationen erzeugten Gleichgewichtsensembles sollten sich dazu gut eignen.

Das ursprüngliche Ziel dieser Arbeit, die Untersuchung des lichtinduzierten Faltungsprozesses eines helikalen Peptids, wurde allerdings nicht ganz erreicht. Dies hatte im Wesentlichen zwei Gründe. Erstens stellten sich die ursprünglich für diese Untersuchung vorgesehenen Peptide für Faltungsexperimente als ungeeignet heraus. In CD-Messungen zeigten sie nur einen vernachlässigbaren Helixanteil. Daraufhin wurde eine neue Peptidsequenz vorgeschlagen, die im Experiment einen deutlich höheren Helixgehalt aufwies. Dieses Peptid war bislang allerdings zu groß, um mit unseren Mitteln in MD-Simulationen eingehend untersucht werden zu können. Zweitens mussten, bevor wir uns dem Faltungsprozess zuwenden konnten, zuerst das Sampling- und das Kraftfeld-Problem behandelt werden.

Neben der detaillierten Analyse dieser beiden Probleme wurden in dieser Arbeit immerhin zwei wichtige Schritte auf dem Weg zur Untersuchung des Faltungsprozesses unternommen. So konnten wir in Faltungsstudien eines unserer Modellpeptide ($\tilde{\mathrm{P}}$) zeigen, dass die Faltung des C22/CMAP-Modells dieses Peptids auf einer Zeitskala von einigen Nanosekunden stattfindet. Diese Zeitskala ist in MD-Simulationen bereits mit heutigen Mitteln gut zugänglich. Der zweite wichtige Schritt war die Entwicklung eines molekularmechanischen Kraftfelds zur Beschreibung der lichtspaltbaren Verbrückung unserer photosensitiven Peptide, wodurch die Voraussetzungen für MM-MD-Simulationen dieser Peptide geschaffen wurden.

Unsere bisherigen Aktivitäten beschränkten sich auf den offenen Zustand der lichtschaltbaren Peptide. Als nächster Schritt könnte nun die Anwendung der in Kapitel 3 validierten Methoden auf die geschlossenen Peptide folgen, um deren Gleichgewichtseigenschaften zu analysieren. Auf diese Weise würde auch ein umfangreiches Ensemble von Startstrukturen generiert, das als Ausgangspunkt für repräsentative Faltungssimulationen der lichtschaltbaren Peptide dienen könnte.

Temperaturkontrolle während der MD-Simulation

Prozesse wie die im letzten Abschnitt behandelte lichtinduzierte Faltung eines helikalen Peptids werden stark von der Temperatur des Systems beeinflusst. Deshalb kommt der Temperaturkontrolle in MD-Simulationen eine große Bedeutung zu. In

Kapitel 4 wurde eine Publikation vorgestellt, die sich mit dieser Thematik beschäftigt.

Bislang wurde eine ganze Reihe von Algorithmen zur Temperaturkontrolle in MD-Simulationen vorgeschlagen. Diese sogenannten Thermostaten kontrollieren die Temperatur eines Simulationssystems, indem sie in die Dynamik des Systems eingreifen. Somit ist der Einsatz eines Thermostaten immer mit einer Störung der natürlichen Newtonschen Mechanik des jeweiligen Systems verbunden. In Kapitel 4 wurde mit dem *Störungsquotienten* nun eine Größe eingeführt, die es erlaubt, diese Störung zu quantifizieren. Dadurch können unterschiedliche Thermostaten bezüglich der Stärke ihres Eingriffs in die Dynamik verglichen werden.

Es wurden zwei Klassen von Thermostaten untersucht: der Berendsen- (BT) und der Nosé-Hoover-Thermostat (NHT). Wir konnten zeigen, dass die Stärke der Eingriffe des NHT in der Regel eine Größenordnung über der des BT liegt. Allerdings konnte auch gezeigt werden, dass ein mit der Zeitkonstante τ an das System gekoppelter BT diejenigen kanonischen Energiefluktuationen unterdrückt, die auf einer Zeitskala länger als τ stattfinden. Generell unklar ist jedoch, wie sich die durch einen Thermostaten verursachten Störungen überhaupt auf die Daten auswirken, die aus einer Simulation extrahiert werden. Am Beispiel des Alanin-Dipeptids zeigte sich, dass sich eine Beeinflussung von dessen Konformationsdynamik durch einen Thermostaten nur in ausgesprochen ausgedehnten Simulationen verlässlich nachweisen lässt.

Bei der Temperaturkontrolle heterogener Systeme, wie z.B. der in Kapitel 3 betrachteten Peptide in wässriger Lösung, ist ein zusätzlicher Aspekt zu beachten. Oft unterscheiden sich die durch algorithmisches Rauschen verursachten Heizraten der Subsysteme. In den Peptid/Wasser-Systemen aus Kapitel 3 liegt z.B. die Heizrate des Wassers deutlich über der des Peptids. Werden diese beiden Subsysteme nun an einen gemeinsamen Thermostaten gekoppelt, führt die Differenz ihrer Heizraten zu einer Differenz ihrer Temperaturen. In unserem Peptid/Wasser-System läge also die Temperatur des Peptids unter der des Lösungsmittels. Dieses bekannte Phänomen wird auch als „Hot-Solvent/Cold-Solute“-Problem bezeichnet. Dieses Problem wurde in Kapitel 4 anhand eines Modells des Wärmeflusses zwischen den Subsystemen und dem Thermostaten detailliert analysiert. Oft werden zur Lösung dieses Problems die beiden Subsysteme an zwei separate Thermostaten gekoppelt. Dies hat jedoch zur Folge, dass die jeweiligen Nachteile des eingesetzten Thermostaten direkt auf die Eigenschaften des Peptids durchschlagen, wodurch beim Einsatz eines BT insbesondere dessen thermodynamisches Ensemble erheblich gestört wird.

Diese Erkenntnis hat uns dazu veranlasst, zwei neue Strategien zur Temperaturkontrolle in MD-Simulationen heterogener Systeme vorzuschlagen: die minimalinvasive und die noninvasive Kopplung. Beide Strategien basieren auf der Idee, das Lösungsmittel als Wärmebad für das Peptid zu betrachten. Bei der noninvasiven Kopplung wird lediglich die Temperatur des Lösungsmittels durch einen Thermostaten kontrolliert. Das Peptid ist an keinen Thermostaten gekoppelt. Seine Temperatur wird nur durch den thermischen Kontakt zu seiner Umgebung, dem Lösungsmittel, kontrolliert. Reicht dieser Kontakt nicht aus, um die Temperatur des Peptids auf dem gewünschten Wert zu halten, muss die Strategie der minimalinvasiven Kopplung angewendet werden. Hier wird das Peptid an einen separaten Thermostaten gekoppelt, der lediglich einen konstanten Wärmefluss zur Verfügung stellt, der die algorithmische Wärmeentwicklung des Peptids kompensiert.

Der entscheidende Vorteil dieser beiden von uns entwickelten Strategien ist, dass das thermodynamische Ensemble des Peptids erhalten bleibt. In den meisten Fällen reicht der thermische Kontakt zwischen Peptid und Lösungsmittel aus, um die noninvasive Kopplung anwenden zu können. Dies war auch bei den in Kapitel 3 betrachteten Peptid/Wasser-Systemen der Fall. Unsere Analyse der in Kapitel 4 vorgeschlagenen Strategien zur Temperaturkontrolle in heterogenen Systemen beschränkte sich bisher auf die Untersuchung kleiner Peptide in Lösung. Ob diese Strategien auch auf größere Proteinsysteme anwendbar sind, muss sich noch zeigen. Ebenso wurde die Strategie der minimalinvasiven Kopplung bislang nur für Systeme diskutiert, die sich aus zwei Subsystemen mit jeweils einheitlichen algorithmischen Heizraten zusammensetzen. Die Erweiterung auf kompliziertere Systeme, wie z.B. Protein/DNA-Komplexe in Lösung, sollte jedoch ohne prinzipielle Schwierigkeiten möglich sein.

Wesentliche Erkenntnisse

Im Rahmen dieser Arbeit wurden wichtige Beiträge zur Entwicklung neuer Methoden auf dem Gebiet der numerischen Simulation biomolekularer Systeme geleistet. Das zu Beginn des Kapitels erwähnte neuartige polarisierbare Kraftfeld ist ein vielversprechendes Konzept zur effizienten Berechnung der Schwingungsspektren von Proteinen. Die von uns vorgeschlagenen Strategien zur Temperaturkontrolle in Molekulardynamik-Simulationen heterogener Systeme stellen einen deutlichen Fortschritt in diesem Bereich dar.

Neben diesen methodischen Beiträgen sind als wesentliche Erkenntnisse dieser Arbeit zwei Punkte zu nennen. Zum einen wurde mit der Wechselwirkung dipolarer Bindungen eine schlüssige und gleichzeitig einfache Erklärung für die Ursachen der Konformationspräferenzen verschiedener Prolinderivate vorgestellt. Zum anderen konnte erstmals gezeigt werden, dass und wie sich das Faltungs/Entfaltungs-Gleichgewicht α-helikaler Peptide in Abhängigkeit der Temperatur in Molekulardynamik-Simulationen effizient beschreiben lässt.

A CMAP — der neue Energieterm in CHARMM

In diesem Anhang soll das Funktionsprinzip des in dem molekularmechanischen Kraftfeld CHARMM22/CMAP neu eingeführten Energieterms [101] beschrieben und erklärt werden. Außerdem soll gezeigt werden, wie dieses neue Kraftfeld in unserem MD-Simulationsprogramm EGO-MMVI implementiert wurde.

Ziel der Einführung der CMAP war der Wunsch der CHARMM-Entwickler, die Beschreibung der Energetik des Proteinrückgrats durch ihr Kraftfeld zu verbessern [101]. In CHARMM22 (C22) wird die Energie des Proteinrückgrats durch die Diederenergie der Winkel ϕ und ψ dominiert. Diese wird jeweils durch einen einfachen Kosinus-Term (vgl. Gl. 1.7) beschrieben [100].

Die hierbei verwendeten Parameter wurden so gewählt, dass die relativen Energien ausgewählter Konformationen des Alanin- bzw. Glycin-Dipeptids möglichst gut getroffen wurden [100]. Die Optimierung der Parameter des Proteinrückgrats beschränkte sich also ursprünglich auf einige wenige Punkte in der von den Diederwinkeln ϕ und ψ aufgespannten Ebene. Diese Beschränkung wurde bei der Entwicklung von C22/CMAP aufgehoben, Zielfunktion der Parameteroptimierung war jetzt die komplette ϕ/ψ-Ebene [101]. Hierzu wurden für das Alanin-, Glycin-, und Prolin-Dipeptid mit quantenmechanischen Methoden die Energiewerte in dieser Ebene auf einem quadratischen Gitter (Gitterkonstante 15°) bestimmt, d.h.

$$E_{\mathrm{QM}}(\phi_i, \psi_j) \text{ mit } \phi_i, \psi_j = -180°, -165°, \ldots, 180°. \tag{A.1}$$

Nun musste eine Erweiterung des C22-Kraftfelds gefunden werden, die dafür sorgt, dass das erweiterte Kraftfeld diese quantenmechanisch berechnete Energiefläche hinreichend genau approximiert.

Der Versuch, dies ohne eine Erweiterung der analytischen Form des Kraftfelds (vgl. Abschnitt 1.3.2) zu erreichen, scheiterte. Selbst eine Parametrisierung der beiden Diederwinkel ϕ und ψ durch je sechs Kosinusterme (gleichbedeutend mit 24 freien Parametern) erwies sich als ungenügend. Auch eine Erweiterung des Kraftfelds um einen zusätzlichen Energieterm

$$E_{\mathrm{cross}}(\phi, \psi) = \sum_{n,m=1}^{4} K_{n,m}[1 + \cos(n\phi + m\psi - \delta_{n,m})] \tag{A.2}$$

brachte nicht den gewünschten Erfolg [101].

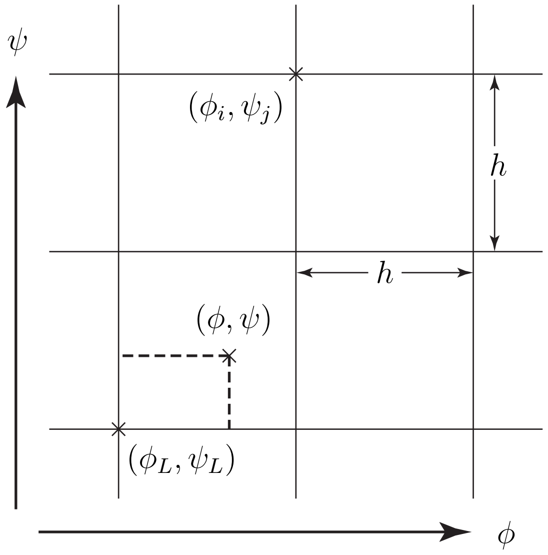


Abbildung A.1: Für die bei der CMAP-Korrektur eingesetzte bikubische Interpolation wird über die ϕ/ψ-Ebene ein quadratisches Gitter gelegt mit den Gitterpunkten (ϕ_i, ψ_j) und der Gitterkonstanten $h = 15°$. Zur Berechnung der CMAP-Energie $E_{\mathrm{CMAP}}(\phi, \psi)$ nach Gl. (A.3) muss als erstes bestimmt werden, in welcher Gitterzelle der Punkt (ϕ, ψ) liegt. (ϕ_L, ψ_L) sind dann die Koordinaten der linken unteren Ecke dieser Zelle.

So wurde mit der CMAP (*enery Correction MAP*) schließlich ein völlig neues Konzept in das Kraftfeld C22 eingeführt.

A.1 Das Prinzip

Diese Energiekorrektur beruht auf einem Gitter-basierten Interpolationsverfahren, das eine fast vollständige Reproduktion der QM-Energiefläche durch das MM-Kraftfeld erlaubt. Hierzu wird durch zweidimensionale bikubische Interpolation eine zweimal stetig differenzierbare Funktion

$$E_{\mathrm{CMAP}}(\phi, \psi) = \sum_{i,j=0}^{3} c_{i,j} \left(\frac{\phi - \phi_L}{h}\right)^i \left(\frac{\psi - \psi_L}{h}\right)^j \tag{A.3}$$

mit der Eigenschaft

$$E_{\mathrm{MM}}(\phi_i, \psi_j) = E_{\mathrm{C22}}(\phi_i, \psi_j) + E_{\mathrm{CMAP}}(\phi_i, \psi_j) = E_{\mathrm{QM}}(\phi_i, \psi_j) \tag{A.4}$$

konstruiert. Zur Auswertung der Funktion $E_{\mathrm{CMAP}}(\phi, \psi)$ muss zuerst bestimmt werden, in welcher Gitterzelle das Koordinatenpaar (ϕ, ψ) liegt. (ϕ_L, ψ_L) sind dann die

Koordinaten der linken unteren Ecke dieser Zelle (vgl. Abbildung A.1). Zu jeder Gitterzelle gehören 16 Koeffizienten $c_{i,j}$ $(i, j = 0, 1, 2, 3)$. Diese müssen nun so gewählt werden, dass E_{CMAP} zum einen die Interpolationsbedingung

$$E_{\text{CMAP}}(\phi_i, \psi_j) = E_{\text{QM}}(\phi_i, \psi_j) - E_{\text{C22}}(\phi_i, \psi_j) \tag{A.5}$$

erfüllt und zum anderen zweimal stetig differenzierbar ist.

Eine Möglichkeit hierzu ist die Verwendung der Werte der vier Größen

$$E_{\text{CMAP}}, \ \frac{\partial E_{\text{CMAP}}}{\partial \phi}, \ \frac{\partial E_{\text{CMAP}}}{\partial \psi}, \ \frac{\partial^2 E_{\text{CMAP}}}{\partial \phi \partial \psi} \tag{A.6}$$

an den Punkten (ϕ_i, ψ_j). Sind diese Werte bekannt, können die Koeffizienten $c_{i,j}$ für jede Gitterzelle durch eine lineare Transformation aus diesen Werten an den vier Ecken der Zelle bestimmt werden [152]. $E_{\text{CMAP}}(\phi_i, \psi_j)$ wird durch die Interpolationsbedingung (Gl. A.5) festgelegt, Werte für die drei anderen Terme in (A.6) müssen jedoch noch bestimmt werden. Während durch diese Wahl von $E_{\text{CMAP}}(\phi_i, \psi_j)$ sichergestellt wird, dass die resultierende Funktion die gewünschten Interpolationseigenschaften bestitzt, lässt sich über die anderen drei Werte der Verlauf der Energiefläche zwischen den Gitterpunkten steuern. Somit ist eine geschickte Wahl dieser Werte Ausschlag gebend für die Qualität des Energieterms E_{CMAP}. Zur Bestimmung dieser Werte schlagen die Entwickler der CMAP die Verwendung einer kubischen Spline-Interpolation [152] vor [101].

Da eine Kenntnis dieses Interpolationsverfahrens das Verständnis der relevanten Teile des Programmcodes unseres MD-Pakets EGO-MMVI stark vereinfacht, soll es im folgenden Abschnitt vorgestellt werden. Die Darstellung ist bewusst kompakt gehalten und orientiert sich an [153].

A.2 Spline-Interpolation

Seien $x_0 < x_1 < \ldots < x_n$ Stützstellen mit zugehörigen Stützwerten $y_0, y_1, \ldots, y_n$, $x_i, y_i \in \mathbb{R}$ $(i = 0, 1, \ldots, n)$. Dann versteht man unter einem kubischen Spline eine Funktion $s \in \mathcal{C}^{(2)}([x_0, x_n])$, wobei s eingeschränkt auf das Intervall $[x_{i-1}, x_i]$ $(i = 1, 2, \ldots, n)$ ein Polynom höchstens dritten Grades ist [153]. Die Interpolationsbedingungen sind $s(x_i) = y_i$ $(i = 0, 1, \ldots, n)$.

Unter Ausnutzung der Interpolations- und Stetigkeitsbedingungen lässt sich zeigen, dass

$$s(x) = -s''_{i-1}\tfrac{(x-x_i)^3}{6h_i} + s''_i\tfrac{(x-x_{i-1})^3}{6h_i} + \left(\tfrac{y_i}{h_i} - s''_i\tfrac{h_i}{6}\right)(x - x_{i-1}) - \left(\tfrac{y_{i-1}}{h_i} - s''_{i-1}\tfrac{h_i}{6}\right)(x - x_i) \tag{A.7}$$

und

$$s'(x) = -s''_{i-1}\frac{(x-x_i)^2}{2h_i} + s''_i\frac{(x-x_{i-1})^2}{2h_i} + \frac{y_i - y_{i-1}}{h_i} - (s''_i - s''_{i-1})\frac{h_i}{6} \tag{A.8}$$

für $x \in [x_{i-1}, x_i]$. Dabei gilt $s''_i = s''(x_i)$ und $h_i = x_i - x_{i-1}$ $(i = 0, 1, \ldots, n)$.

Die Forderung der Stetigkeit von $s'(x)$ und periodischer Randbedingungen ($s'_0 = s'_n$ und $s''_0 = s''_n$) führt nun zu folgendem linearen Gleichungssystem

$$\begin{pmatrix} 2 & \lambda_1 & 0 & \cdots & 0 & \mu_1 \\ \mu_2 & 2 & \lambda_2 & & & 0 \\ 0 & \mu_3 & 2 & \lambda_3 & & \vdots \\ \vdots & & \ddots & \ddots & \ddots & 0 \\ 0 & & & \mu_{n-1} & 2 & \lambda_{n-1} \\ \lambda_n & 0 & \cdots & 0 & \mu_n & 2 \end{pmatrix} \begin{pmatrix} s''_1 \\ s''_2 \\ \\ \vdots \\ \\ s''_n \end{pmatrix} = \begin{pmatrix} d_1 \\ d_2 \\ \\ \vdots \\ \\ d_n \end{pmatrix} \tag{A.9}$$

mit

$$\lambda_i := \frac{h_{i+1}}{h_i + h_{i+1}}, \tag{A.10a}$$

$$\mu_i := 1 - \lambda_i = \frac{h_i}{h_i + h_{i+1}} \tag{A.10b}$$

und

$$d_i := \frac{6}{h_i + h_{i+1}}\left(\frac{y_{i+1} - y_i}{h_{i+1}} - \frac{y_i - y_{i-1}}{h_i}\right) \qquad (i = 1, 2, \ldots, n-1), \tag{A.10c}$$

sowie

$$\lambda_n := \frac{h_1}{h_n + h_1}, \tag{A.10d}$$

$$\mu_n := 1 - \lambda_n = \frac{h_n}{h_n + h_1}, \tag{A.10e}$$

$$d_n := \frac{6}{h_n + h_1}\left(\frac{y_1 - y_n}{h_1} - \frac{y_n - y_{n-1}}{h_n}\right). \tag{A.10f}$$

Für die CMAP-Korrektur gilt $n = 24$ und $h_i = h$ $(i = 1, 2, \ldots, 24)$. Eingesetzt in Gl. (A.9-A.10) erhalten wir somit zur Bestimmung der s''_i $(i = 1, 2, \ldots, 24)$ das Gleichungssystem

$$\begin{pmatrix} 2 & 1/2 & 0 & \cdots & 0 & 1/2 \\ 1/2 & 2 & 1/2 & & & 0 \\ 0 & 1/2 & 2 & 1/2 & & \vdots \\ \vdots & & \ddots & \ddots & \ddots & 0 \\ 0 & & & 1/2 & 2 & 1/2 \\ 1/2 & 0 & \cdots & 0 & 1/2 & 2 \end{pmatrix} \begin{pmatrix} s''_1 \\ s''_2 \\ \\ \vdots \\ \\ s''_{24} \end{pmatrix} = \frac{3}{h^2} \begin{pmatrix} y_{24} - 2y_1 + y_2 \\ y_1 - 2y_2 + y_3 \\ \\ \vdots \\ \\ y_{23} - 2y_{24} + y_1 \end{pmatrix}. \tag{A.11}$$

Setzt man nun die Lösung dieses Gleichungssystems in Gl. (A.8) ein, ergibt sich folgender Ausdruck für die Ableitung des Splines an einer der Stützstellen[1]

$$s'(x_i) = a_{12} \sum_{j=1}^{11} a_j (y_{(i+j) \bmod 24} - y_{(i-j) \bmod 24}) \tag{A.12}$$

mit $a_1 = 564719, a_2 = -151316, a_3 = 40545, a_4 = -10864, a_5 = 2911, a_6 = -780, a_7 = 209, a_8 = -56, a_9 = 15, a_{10} = -4, a_{11} = 1, a_{12} = 1/(702520h)$.

Mit Gleichung (A.12) lassen sich jetzt aus den Werten $E_{\mathrm{CMAP}}(\phi_i, \psi_j)$ Werte für die Ableitungen in Gl. (A.6) an den Gitterpunkten (ϕ_i, ψ_j) bestimmen:

$$\frac{\partial E_{\mathrm{CMAP}}}{\partial \phi}(\phi_i, \psi_j) = a_{12} \sum_{k=1}^{11} a_k [E_{\mathrm{CMAP}}(\phi_{(i+k) \bmod 24}, \psi_j) - E_{\mathrm{CMAP}}(\phi_{(i-k) \bmod 24}, \psi_j)] \tag{A.13a}$$

$$\frac{\partial E_{\mathrm{CMAP}}}{\partial \psi}(\phi_i, \psi_j) = a_{12} \sum_{k=1}^{11} a_k [E_{\mathrm{CMAP}}(\phi_i, \psi_{(j+k) \bmod 24}) - E_{\mathrm{CMAP}}(\phi_i, \psi_{(j-k) \bmod 24})] \tag{A.13b}$$

und daraus auch

$$\begin{aligned} \frac{\partial^2 E_{\mathrm{CMAP}}}{\partial \phi \partial \psi}(\phi_i, \psi_j) = a_{12}^2 \sum_{k,l=1}^{11} a_k a_l [& E_{\mathrm{CMAP}}(\phi_{(i+k) \bmod 24}, \psi_{(j+l) \bmod 24}) \\ & - E_{\mathrm{CMAP}}(\phi_{(i-k) \bmod 24}, \psi_{(j+l) \bmod 24}) \\ & - E_{\mathrm{CMAP}}(\phi_{(i+k) \bmod 24}, \psi_{(j-l) \bmod 24}) \\ & + E_{\mathrm{CMAP}}(\phi_{(i-k) \bmod 24}, \psi_{(j-l) \bmod 24})]. \end{aligned} \tag{A.13c}$$

A.3 Berechnung der CMAP-Kräfte

Mit den aus einer kubischen Spline-Interpolation unter Berücksichtigung periodischer Randbedingungen hergeleiteten Gleichungen (A.13) lassen sich nun also für jeden Gitterpunkt (ϕ_i, ψ_j) $(\phi_i, \psi_j = -180°, -165°, \ldots, 180°)$ neben den aus Gl. (A.5) bekannten Größen $E_{\mathrm{CMAP}}(\phi_i, \psi_j)$ Werte für die restlichen drei in Gl. (A.6) aufgeführten Ausdrücke bestimmen. Aus diesen Werten können dann, wie bereits erwähnt, durch eine lineare Transformation [152] Koeffizienten $c_{i,j}$ bestimmt werden, die, eingesetzt in Gl. (A.3), dafür sorgen, dass E_{CMAP} die gewünschten Eigenschaften hat (vgl. A.1).

---

[1] Hier bezeichnet "mod" die Modulodivision.

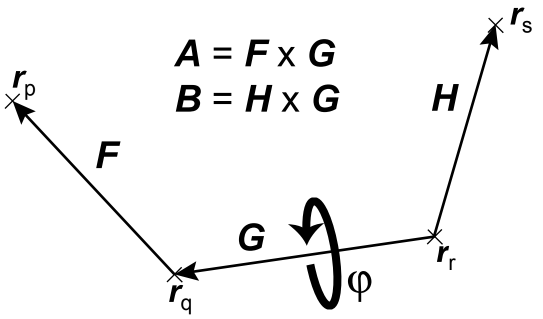


Abbildung A.2: Der Diederwinkel φ beschreibt den Winkel zwischen den beiden Ebenen, die von den Atomen $(\boldsymbol{r}_p, \boldsymbol{r}_q, \boldsymbol{r}_r)$ bzw. $(\boldsymbol{r}_q, \boldsymbol{r}_r, \boldsymbol{r}_s)$ aufgespannt werden. Die Ableitung dieses Winkels nach den Koordinaten der Atome kann aus den eingezeichneten Vektoren berechnet werden [154].

Der Einsatz des Kraftfeldbeitrags E_{CMAP} in einer MD-Simulation führt nun zu zusätzlichen Kräften. Wie diese aus E_{CMAP} berechnet werden, soll im folgenden erklärt werden. Da $E_{\text{CMAP}}(\phi, \psi)$ mit ϕ und ψ von zwei Diederwinkeln abhängt, deren Wert jeweils durch die Positionen von vier Atomen bestimmt ist, ist E_{CMAP} eine Funktion der Koordinaten von acht Atomen. Sind $\boldsymbol{r}_n$ $(n = a, \ldots, g)$ die Orte dieser Atome, gilt also[1]

$$E = E(\phi, \psi) = E(\phi(\boldsymbol{r}_a, \boldsymbol{r}_b, \boldsymbol{r}_c, \boldsymbol{r}_d), \psi(\boldsymbol{r}_e, \boldsymbol{r}_f, \boldsymbol{r}_g, \boldsymbol{r}_h)). \tag{A.14}$$

Daraus folgt für die Komponente i der CMAP-Kraft $\boldsymbol{F}$ auf das Atom n der Ausdruck

$$F_n^i = -\frac{\partial E}{\partial r_n^i} = -\frac{\partial E}{\partial \phi}\frac{\partial \phi}{\partial r_n^i} - \frac{\partial E}{\partial \psi}\frac{\partial \psi}{\partial r_n^i} \tag{A.15}$$

für alle $n \in \{a, b, c, d, e, f, g, h\}$ und $i \in \{x, y, z\}$. Die Werte $\frac{\partial E}{\partial \phi}$ und $\frac{\partial E}{\partial \psi}$ lassen sich aus Gl. (A.3) zu

$$\frac{\partial E}{\partial \phi} = \sum_{i,j=0}^{3} i\, c_{i,j} \left(\frac{\phi - \phi_L}{h}\right)^{i-1} \left(\frac{\psi - \psi_L}{h}\right)^{j} \tag{A.16a}$$

und

$$\frac{\partial E}{\partial \psi} = \sum_{i,j=0}^{3} j\, c_{i,j} \left(\frac{\phi - \phi_L}{h}\right)^{i} \left(\frac{\psi - \psi_L}{h}\right)^{j-1} \tag{A.16b}$$

bestimmen. Für die Ableitungen $\frac{\partial \phi}{\partial r_n^i}$ und $\frac{\partial \psi}{\partial r_n^i}$ der Diederwinkel nach den Atomkoordinaten leiten Blondel und Karplus folgende Beziehungen her [154]:

$$\frac{\partial \varphi}{\partial \boldsymbol{r}_p} = -\frac{|\boldsymbol{G}|}{\boldsymbol{A}^2}\boldsymbol{A} \tag{A.17a}$$

---

[1] Um die Übersichtlichkeit zu erhöhen wird E_{CMAP} bis auf weiteres E genannt.

$$\frac{\partial \varphi}{\partial \boldsymbol{r}_q} = \frac{|\boldsymbol{G}|}{\boldsymbol{A}^2}\boldsymbol{A} + \frac{\boldsymbol{F}\cdot\boldsymbol{G}}{\boldsymbol{A}^2|\boldsymbol{G}|}\boldsymbol{A} - \frac{\boldsymbol{H}\cdot\boldsymbol{G}}{\boldsymbol{B}^2|\boldsymbol{G}|}\boldsymbol{B} \tag{A.17b}$$

$$\frac{\partial \varphi}{\partial \boldsymbol{r}_r} = \frac{\boldsymbol{H}\cdot\boldsymbol{G}}{\boldsymbol{B}^2|\boldsymbol{G}|}\boldsymbol{B} - \frac{\boldsymbol{F}\cdot\boldsymbol{G}}{\boldsymbol{A}^2|\boldsymbol{G}|}\boldsymbol{A} - \frac{|\boldsymbol{G}|}{\boldsymbol{B}^2}\boldsymbol{B} \tag{A.17c}$$

$$\frac{\partial \varphi}{\partial \boldsymbol{r}_s} = \frac{|\boldsymbol{G}|}{\boldsymbol{B}^2}\boldsymbol{B}. \tag{A.17d}$$

Hierbei sind, wie in Abbildung A.2 dargestellt, $\boldsymbol{F}$, $\boldsymbol{G}$ und $\boldsymbol{H}$ Verbindungsvektoren der den Diederwinkel aufspannenden Atome und $\boldsymbol{A}$ sowie $\boldsymbol{B}$ zwei daraus abgeleitete Vektoren. Mit den Rechenvorschriften der Gleichungen (A.16) und (A.17) können nun in einer MD-Simulation über Gleichung (A.15) die durch die CMAP verursachten Kräfte bestimmt werden. Genau so wurden die CMAP-Kräfte auch in unserem MD-Simulationsprogramm EGO-MMVI implementiert. Hinweise zur Benutzung der CMAP in EGO-MMVI sind im `Wiki` [155] unserer Arbeitsgruppe hinterlegt.

A.4 Die Energiefläche

In diesem letzten Abschnitt soll schließlich veranschaulicht werden, wie die Einführung der CMAP die Energielandschaft des CHARMM-Kraftfelds in der ϕ/ψ-Ebene verändert hat. In C22 werden zur Beschreibung dieser Energiefläche für alle Aminosäuren außer Prolin die gleichen Parameter verwendet [100]. Für die CMAP-Korrektur gibt es drei Parametersätze: einen für Glycin, einen für Prolin und einen für alle anderen kanonischen Aminosäuren [101]. Bei unserem Vergleich der Kraftfelder C22 und C22/CMAP beschränken wir uns auf den letzten Parametersatz.

Abbildung A.3 zeigt eine Konturdarstellung der Energie der Diederwinkel ϕ und ψ im Kraftfeld C22. Für diese Energie gilt (vgl. Gl. 1.7)

$$E(\phi,\psi) = E(\phi) + E(\psi) = \{0.2\,[1+\cos(\phi+180)] + 0.6\,[1+\cos(\psi)]\}\,\frac{\mathrm{kcal}}{\mathrm{mol}}. \tag{A.18}$$

Analog dazu zeigt Abbildung A.4 die Energie des Kraftfelds C22/CMAP, die sich aus der C22-Energie (Gl. A.18) und der CMAP-Energie zusammensetzt:

$$E(\phi,\psi) = E(\phi) + E(\psi) + E_{\mathrm{CMAP}}(\phi,\psi). \tag{A.19}$$

Ein Vergleich der beiden Abbildungen macht deutlich, dass die Einführung der CMAP eine viel differenziertere Modellierung der Energielandschaft in der ϕ/ψ-Ebene erlaubt.

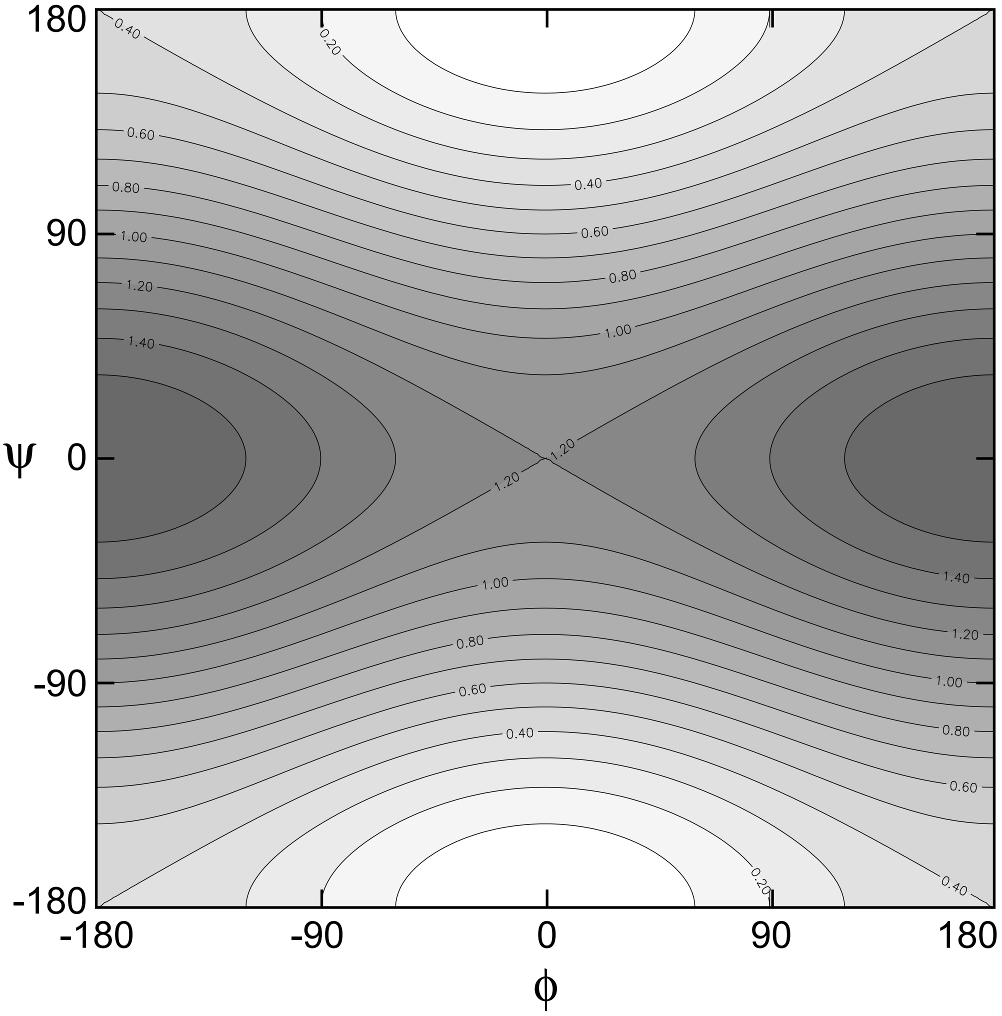


Abbildung A.3: Konturdarstellung der Diederenergie $E(\phi, \psi)$ im Kraftfeld C22. In C22 gilt $E(\phi, \psi) = E(\phi) + E(\psi)$ mit der in Gl. (A.18) angegebenen Parametrisierung. Der Abstand der Konturlinien beträgt 0.1 kcal/mol.

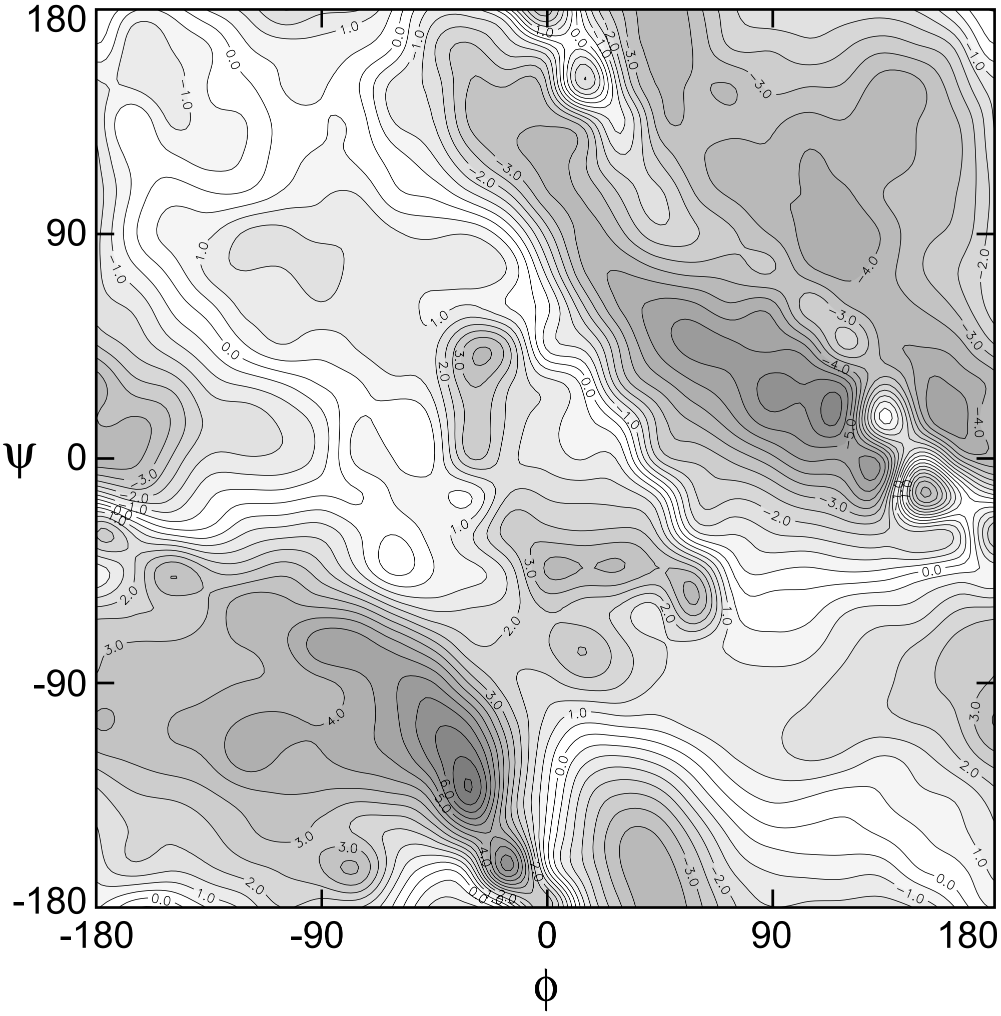


Abbildung A.4: Konturdarstellung der im Kraftfeld C22/CMAP mit den Diederwinkeln ϕ und ψ assoziierten Energie $E(\phi,\psi)$. In C22/CMAP gilt $E(\phi,\psi) = E(\phi)+E(\psi)+E_{\mathrm{CMAP}}(\phi,\psi)$. $E(\phi)$ und $E(\psi)$ sind die aus C22 übernommenen Kosinusterme, E_{CMAP} dagegen wird durch bikubische Interpolation (vgl. Gl. A.3) bestimmt. Der Abstand der Konturlinien beträgt 0.5 kcal/mol.

Literaturverzeichnis

[1] Stryer, L. *Biochemistry.* W. H. Freeman and Company, New York, 3. Auflage (1988).

[2] Holmes, K. C., D. Popp, W. Gebhard und W. Kabsch. Atomic model of the actin filament. *Nature* **347**, 44–49 (1990).

[3] Mizuno, N., S. Toba, M. Edamatsu, J. Watai-Nishii, N. Hirokawa, Y. Y. Toyoshima und M. Kikkawa. Dynein and kinesin share an overlapping microtubule-binding site. *EMBO J.* **23**, 2459–2467 (2004).

[4] Oesterhelt, D. und W. Stoeckenius. Functions of a new photoreceptor membrane. *Proc. Natl. Acad. Sci. U.S.A.* **70**, 2853–2857 (1973).

[5] Lehninger, A. L. *Biochemistry.* Worth Publishers, New York, 2. Auflage (1975).

[6] Wu, H. Studies on denaturation of proteins. XIII. A theory of denaturation. *Chin. J. Physiol.* **5**, 321–344 (1931).

[7] Anson, M. L. und A. E. Mirsky. The equilibrium between active native Trypsin and inactive denatured Trypsin. *J. Gen. Physiol.* **17**, 393–408 (1934).

[8] Lumry, R. und H. Eyring. Conformation changes of proteins. *J. Phys. Chem.* **58**, 110–120 (1954).

[9] Prusiner, S. B. Prions. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 13363–13383 (1998).

[10] Stork, M. *Molekulardynamik-Simulationen von amyloidogenen Proteinen in Lösung: Stabilitätsuntersuchungen und Weiterentwicklung einer Kontinuumsmethode*. Dissertation, Ludwig-Maximilians-Universität, München (2007).

[11] Holmgren, S. K., L. E. Bretscher, K. M. Taylor und R. T. Raines. A hyperstable collagen mimic. *Chem. Biol.* **6**, 63–70 (1999).

[12] Improta, R., C. Benzi und V. Barone. Understanding the role of stereoelectronic effects in determining collagen stability. 1. A quantum mechanical study of proline, hydroxyproline, and fluoroproline dipeptide analogues in aqueous solution. *J. Am. Chem. Soc.* **123**, 12568–12577 (2001).

[13] Fietzek, P. P. und K. Kühn. Information contained in the amino acid sequence of the $\alpha 1$(I)-chain of collagen and its consequences upon the formation of the triple helix, of fibrils and crosslinks. *Mol. Cell. Biochem.* **8**, 141–157 (1975).

[14] Cadamuro, S. A., R. Reichold, U. Kusebauch, H.-J. Musiol, C. Renner, P. Tavan und L. Moroder. Konformationseigenschaften des 4-Mercaptoprolins und verwandter Derivate. *Angew. Chem.* **120**, 2174–2177 (2008).

[15] Anfinsen, C. B., E. Haber, M. Sela und F. H. White, Jr. The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. *Proc. Natl. Acad. Sci. U.S.A.* **47**, 1309–1314 (1961).

[16] Epstein, C. J., R. F. Goldberger und C. B. Anfinsen. The genetic control of tertiary protein structure: studies with model systems. *Cold Spring Harbor Symp. Quant. Biol.* **28**, 439–449 (1963).

[17] Anfinsen, C. B. Principles that govern the folding of protein chains. *Science* **181**, 223–230 (1973).

[18] Baldwin, R. L. The nature of protein-folding pathways – the classical versus the new view. *J. Biomol. NMR* **5**, 103–109 (1995).

[19] Dill, K. und H. Chan. From Levinthal to pathways to funnels. *Nat. Struct. Biol.* **4**, 10–19 (1997).

[20] Baldwin, R. L. und G. D. Rose. Is protein folding hierarchic? I. Local structure and peptide folding. *Trends Biochem. Sci.* **24**, 26–33 (1999).

[21] Baldwin, R. L. und G. D. Rose. Is protein folding hierarchic? II. Folding intermediates and transition states. *Trends Biochem. Sci.* **24**, 77–83 (1999).

[22] Frauenfelder, H., P. G. Wolynes und R. H. Austin. Biological Physics. *Rev. Mod. Phys.* **71**, 419–430 (1999).

[23] Buchner, J. und T. Kiefhaber(Herausgeber), *Protein Folding Handbook.* Wiley-VCH, Weinheim (2005).

[24] Pauling, L., R. B. Corey und H. R. Branson. The structure of proteins: two hydrogen-bonded helical configurations of the polypeptide chain. *Proc. Natl. Acad. Sci. U.S.A.* **37**, 205–211 (1951).

[25] Barlow, D. J. und J. M. Thornton. Helix geometry in proteins. *J. Mol. Biol.* **201**, 601–619 (1988).

[26] Weaver, T. M. The π-helix translates structure into function. *Protein Sci.* **9**, 201–206 (2000).

[27] Fodje, M. N. und S. Al-Karadaghi. Occurrence, conformational features and amino acid propensities for the π-helix. *Protein Eng.* **15**, 353–358 (2002).

[28] Ramachandran, G. N., C. Ramakrishnan und V. Sasisekharan. Stereochemistry of polypeptide chain configurations. *J. Mol. Biol.* **7**, 95–99 (1963).

[29] Levinthal, C. Are there pathways for protein folding? *J. Chim. Phys.* **65**, 44–45 (1968).

[30] Jackson, S. E. How do small single-domain proteins fold? *Fold. Des.* **3**, R81–R91 (1998).

[31] Karplus, M. und D. L. Weaver. Protein folding dynamics: The diffusion-collision model and experimental data. *Protein Sci.* **3**, 650–668 (1994).

[32] Waltho, J. P., V. A. Feher, G. Merutka, H. J. Dyson und P. E. Wright. Peptide models of protein folding initiation sites. 1. Secondary structure formation by peptides corresponding to the G- and H-helices of myoglobin. *Biochemistry* **32**, 6337–6347 (1993).

[33] Reichold, R., B. Fierz, T. Kiefhaber und P. Tavan. Toward MD descriptions of α-helix folding. *Manuskript* .

[34] Johnson, W. C., Jr. Secondary structure of proteins through circular dichroism spectroscopy. *Ann. Rev. Biophys. Biophys. Chem.* **17**, 145–166 (1988).

[35] Creighton, T. E. *Proteins: structures and molecular properties.* W. H. Freeman and Company, New York, 2. Auflage (1993).

[36] Brahms, S. und J. Brahms. Determination of protein secondary structure in solution by vacuum ultraviolet circular dichroism. *J. Mol. Biol.* **138**, 149–178 (1980).

[37] Greenfield, N. und G. D. Fasman. Computed circular dichroism spectra for the evaluation of protein conformation. *Biochemistry* **8**, 4108–4116 (1969).

[38] Chen, Y.-H., J. T. Yang und H. M. Martinez. Determination of the secondary structures of proteins by circular dichroism and optical rotatory dispersion. *Biochemistry* **11**, 4120–4131 (1972).

[39] Chen, Y.-H., J. T. Yang und K. H. Chau. Determination of the helix and β form of proteins in aqueous solution by circular dichroism. *Biochemistry* **13**, 3350–3359 (1974).

[40] Whitmore, L. und B. A. Wallace. Protein secondary structure analyses from circular dichroism spectroscopy: Methods and reference databases. *Biopolymers* **89**, 392–400 (2008).

[41] Hirst, J. D. und C. L. Brooks, III. Helicity, circular dichroism and molecular dynamics of proteins. *J. Mol. Biol.* **243**, 173–178 (1994).

[42] Woody, R. W. und I. Tinoco, Jr. Optical rotation of oriented helices. III. Calculation of the rotatory dispersion and circular dichroism of the alpha- and 3_{10}-helix. *J. Chem. Phys.* **46**, 4927–4945 (1967).

[43] Besley, N. A. und J. D. Hirst. Theoretical studies toward quantitative protein circular dichroism calculations. *J. Am. Chem. Soc.* **121**, 9636–9644 (1999).

[44] Hirst, J. D., S. Bhattacharjee und A. V. Onufriev. Theoretical studies of time-resolved spectroscopy of protein folding. *Faraday Discuss.* **122**, 253–267 (2002).

[45] Oakley, M. T., B. M. Bulheller und J. D. Hirst. First-principles calculations of protein circular dichroism in the far-ultraviolet and beyond. *Chirality* **18**, 340–347 (2006).

[46] Bulheller, B. M., A. Rodgerb und J. D. Hirst. Circular and linear dichroism of proteins. *Phys. Chem. Chem. Phys.* **9**, 2020–2035 (2007).

[47] Greenfield, N. J. Using circular dichroism spectra to estimate protein secondary structure. *Nature Protocols* **1**, 2876–2890 (2007).

[48] Canet, D. *NMR – Konzepte und Methoden.* Springer, Berlin (1994).

[49] Levine, S. G. A short history of the chemical shift. *J. Chem. Ed.* **78**, 133 (2001).

[50] Ernst, R. R. *Nobel Lecture.* http://nobelprize.org/nobel_prizes/chemistry/laureates /1991/ernst-lecture.html, Stand: 05.11.2008 (1991).

[51] Karplus, M. Contact electron-spin coupling of nuclear magnetic moments. *J. Chem. Phys.* **30**, 11–15 (1959).

[52] Karplus, M. Vicinal proton coupling in nuclear magnetic resonance. *J. Am. Chem. Soc.* **85**, 2870–2871 (1963).

[53] Cantor, C. R. und P. R. Schimmel. *Biophysical chemistry. Part II: Techniques for the study of biological structure and function.* W. H. Freeman and Company, New York (1980).

[54] Anet, F. A. L. und A. J. R. Bourn. Nuclear magnetic resonance spectral assignments from Nuclear Overhauser Effects. *J. Am. Chem. Soc.* **87**, 5250–5251 (1965).

[55] Neuhaus, D. und M. Williamson. *The nuclear Overhauser effect in structural and conformational analysis.* VCH Verlagsgesellschaft, Weinheim (1989).

[56] Brünger, A. T., R. L. Campbell, G. M. Clore, A. M. Gronenborn, M. Karplus, G. A. Petsko und M. M. Teeter. Solution of a protein crystal structure with a model obtained from NMR interproton distance restraints. *Science* **235**, 1049–1053 (1987).

[57] Brünger, A. T. *X-Plor (Version 3.1) A System for X-ray Crystallography and NMR.* Yale University Press, New Haven (1992).

[58] Parr, R. G. und W. Yang. *Density-functional theory of atoms and molecules.* Oxford University Press, New York (1989).

[59] Thomas, L. H. The calculation of atomic fields. *Proc. Camb. Phil. Soc.* **23**, 542–548 (1927).

[60] Fermi, E. Un metodo statistico per la determinazione di alcune proprieta dell' atomo. *Rend. R. Acc. Naz. dei Lincei* **6**, 602–607 (1927).

[61] Fermi, E. Eine statistische Methode zur Bestimmung einiger Eigenschaften des Atoms und ihre Anwendung auf die Theorie des periodischen Systems der Elemente. *Z. Phys.* **48**, 73–79 (1928).

[62] Hohenberg, P. und W. Kohn. Inhomogeneous electron gas. *Phys. Rev.* **136**, B864–B871 (1964).

[63] Kohn, W. und L. J. Sham. Self-consistent equations including exchange and correlation effects. *Phys. Rev.* **140**, A1133–A1138 (1965).

[64] Haken, H. und H. C. Wolf. *Molekülphysik und Quantenchemie: Einführung in die experimentellen und theoretischen Grundlagen.* Springer-Verlag, Berlin Heidelberg New York, 2. Auflage (1994).

[65] Becke, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A* **38**, 3098–3100 (1988).

[66] Perdew, J. P. Density-functional approximation for the correlation energy of the inhomogeneous electron gas. *Phys. Rev. B* **33**, 8822–8824 (1986).

[67] Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **98**, 5648–5652 (1993).

[68] Lee, C., W. Yang und R. C. Parr. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* **37**, 785–789 (1988).

[69] Frisch, M. J., G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb *et al.* Gaussian 98 (Revision A.7), Gaussian Inc., Pittsburgh PA (1998).

[70] Hutter, J., A. Alavi, T. Deutsch, M. Bernasconi, S. Goedecker, D. Marx, M. Tuckerman und M. Parinello. *CPMD: Car-Parinello Molecular Dynamics, version 3.3a.* MPI für Festkörperforschung and IBM Zurich Research Laboratory 1995-96 (1999).

[71] Ahlrichs, R., M. Bär, M. Häser, H. Horn und C. Kölmel. Electronic structure calculations on workstation computers: The program system turbomole. *Chem. Phys. Lett.* **162**, 165–169 (1989).

[72] Treutler, O. und R. Ahlrichs. Efficient molecular numerical integration schemes. *J. Chem. Phys.* **102**, 346–354 (1995).

[73] Neese, F., U. Becker, D. Ganiouchine, S. Koßmann, T. Petrenko, C. Riplinger und F. Wennmohs. Orca – an ab initio, DFT and semiempirical SCF-MO package, http://www.thch.uni-bonn.de/tc/orca/, Stand: 05.11.2008.

[74] Schmidt, M., K. Baldridge, J. Boatz, S. Elbert, M. Gordon *et al.* General atomic and molecular electronic structure system. *J. Comput. Chem.* **14**, 1347–1363 (1993).

[75] Gordon, M. und M. Schmidt. Advances in electronic structure theory: GAMESS a decade later. In C. Dykstra, G. Frenking, K. Kim und G. Scuseria (Herausgeber), *Theory and applications of computational chemistry: the first forty years*, Seiten 1167–1189. Elsevier, Amsterdam (2005).

[76] VandeVondele, J., M. Krack, F. Mohamed, M. Parrinello, T. Chassaing und J. Hutter. Quickstep: fast and accurate density functional calculations using a mixed Gaussian and plane waves approach. *Comp. Phys. Comm.* **167**, 103–128 (2005).

[77] Lippert, G., J. Hutter und M. Parrinello. A hybrid Gaussian and plane wave density functional scheme. *Mol. Phys.* **92**, 477–487 (1997).

[78] Lippert, G., J. Hutter und M. Parrinello. The Gaussian and augmented-plane-wave density functional method for ab initio molecular dynamics simulations. *Theor. Chem. Acc.* **103**, 124–140 (1999).

[79] Schäfer, A., C. Huber und A. Reinhart. Fully optimized contracted Gaussian basis sets of triple zeta valence quality for atoms Li to Kr. *J. Chem. Phys.* **100**, 5829–5835 (1994).

[80] Schultheis, V., R. Reichold, B. Schropp und P. Tavan. A polarizable force field for computing the infrared spectra of the polypeptide backbone. *J. Phys. Chem. B* **112**, 12217–12230 (2008).

[81] Ponder, J. W. und D. A. Case. Force Fields for Protein Simulations. *Adv. Prot. Chem.* **66**, 27–85 (2003).

[82] MacKerell, A. D., Jr. Empirical force fields for biological macromolecules: overview and issues. *J. Comput. Chem.* **25**, 1584–1604 (2004).

[83] Weiner, P. K. und P. A. Kollman. AMBER – assisted model-building with energy refinement – a general program for modeling molecules and their interactions. *J. Comput. Chem.* **2**, 287–303 (1981).

[84] Weiner, S. J., P. A. Kollman, D. A. Case, U. C. Singh, C. Ghio, G. Alagona, S. Profeta und P. Weiner. A new force-field for molecular mechanical simulation of nucleic-acids and proteins. *J. Am. Chem. Soc.* **106**, 765–784 (1984).

[85] Weiner, S. J., P. A. Kollman, D. T. Nguyen und D. A. Case. An all atom force-field for simulations of proteins and nucleic-acids. *J. Comput. Chem.* **7**, 230–252 (1986).

[86] Cornell, W. D., P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, Jr. *et al.* A second generation force field for the simulation of proteins and nucleic acids. *J. Am. Chem. Soc.* **117**, 5179–5197 (1995).

[87] Kollman, P. A. Advances and continuing challenges in achieving realistic and predictive simulations of the properties of organic and biological molecules. *Acc. Chem. Res.* **29**, 461–469 (1996).

[88] Wang, J., P. Cieplak und P. A. Kollman. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem.* **21**, 1049–1074 (2000).

[89] Duan, Y., C. Wu, S. Chowdhury, M. C. Lee, G. Xiong *et al.* A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.* **24**, 1999–2012 (2003).

[90] Garcia, A. E. und K. Y. Sanbonmatsu. α-Helical stabilization by side chain shielding of backbone hydrogen bonds. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 2782–2787 (2002).

[91] Nymeyer, H. und A. E. Garcia. Simulation of the folding equilibrium of α-helical peptides: A comparison of the generalized Born approximation with explicit solvent. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 13934–13939 (2003).

[92] Sorin, E. J. und V. S. Pande. Exploring the helix-coil transition via all-atom equilibrium ensemble simulations. *Biophys. J.* **88**, 2472–2493 (2005).

[93] Hornak, V., R. Abel, A. Okur, B. Strockbine, A. Roitberg und C. Simmerling. Comparison of multiple amber force fields and development of improved protein backbone parameters. *Proteins* **65**, 712–725 (2006).

[94] van Gunsteren, W. F. und H. J. C. Berendsen. *Groningen Molecular Simulation (GROMOS) Library Manual.* BIOMOS b.v., University of Groningen, Groningen (1987).

[95] Scott, W. R. P., P. H. Hünenberger, I. G. Tironi, A. E. Mark, S. R. Billeter *et al.* The GROMOS biomolecular simulation program package. *J. Phys. Chem. A* **103**, 3596–3607 (1999).

[96] Jorgensen, W. L. und J. Tirado-Rives. The OPLS potential functions for proteins. Energy minimizations for crystals of cyclic peptides and crambin. *J. Am. Chem. Soc.* **110**, 1657–1666 (1988).

[97] Jorgensen, W. L., D. S. Maxwell und J. Tirado-Rives. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* **118**, 11225–11236 (1996).

[98] Kaminski, G. A., R. A. Friesner, J. Tirado-Rives und W. L. Jorgensen. Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *J. Phys. Chem. B* **105**, 6474–6487 (2001).

[99] Brooks, B. R., R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan und M. Karplus. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **4**, 187–217 (1983).

[100] MacKerell, A. D., Jr., D. Bashford, M. Bellott, R. L. Dunbrack, Jr., J. D. Evanseck *et al.* All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **102**, 3586–3616 (1998).

[101] MacKerell, A. D., Jr., M. Feig und C. L. Brooks, III. Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.* **25**, 1400–1415 (2004).

[102] Feig, M., A. D. MacKerell, Jr. und C. L. Brooks, III. Force field influence on the observation of π-helical protein structures in molecular dynamics simulations. *J. Phys. Chem. B* **107**, 2831–2836 (2003).

[103] MacKerell, A. D., Jr., M. Feig und C. L. Brooks, III. Improved treatment of the protein backbone in empirical force fields. *J. Am. Chem. Soc.* **126**, 698–699 (2004).

[104] Buck, M., S. Bouguet-Bonnet, R. W. Pastor und A. D. MacKerell, Jr. Importance of the CMAP correction to the CHARMM22 protein force field: Dynamics of hen lysozyme. *Biophys. J.* **90**, L36–L38 (2006).

[105] Carstens, H. *Konformationsdynamik lichtschaltbarer Peptide: Molekulardynamiksimulationen und datengetriebene Modellbildung.* Dissertation, Ludwig-Maximilians-Universität, München (2004).

[106] Guillot, B. A reappraisal of what we have learnt during three decades of computer simulations on water. *J. Mol. Liq.* **101**, 219–260 (2002).

[107] Tavan, P., H. Carstens und G. Mathias. Molecular dynamics simulations of proteins and peptides: Problems, achievements, and perspectives. In J. Buchner und T. Kiefhaber (Herausgeber), *Protein Folding Handbook. Part 1.*, Seiten 1170–1195. Wiley-VCH, Weinheim (2005).

[108] Ren, P. und J. Ponder. Polarizable atomic multipole water model for molecular mechanics simulation. *J. Phys. Chem. B* **107**, 5933–5947 (2003).

[109] Patel, S., A. D. MacKerell, Jr. und C. L. Brooks, III. CHARMM fluctuating charge force field for proteins: II protein/solvent properties from molecular dynamics simulations using a nonadditive electrostatic model. *J. Comput. Chem.* **25**, 1504–1514 (2004).

[110] Maple, J. R., Y. Cao, W. Damm, T. A. Halgren, G. A. Kaminski, L. Y. Zhang und R. A. Friesner. A polarizable force field and continuum solvation methodology for modeling of protein-ligand interactions. *J. Chem. Theory Comput.* **1**, 694–715 (2005).

[111] Wang, Z.-X., W. Zhang, C. Wu, H. Lei, P. Cieplak und Y. Duan. Strike a balance: optimization of backbone torsion parameters of AMBER polarizable force field for simulations of proteins and peptides. *J. Comput. Chem.* **27**, 781–790 (2006).

[112] Wang, Z.-X., W. Zhang, C. Wu, H. Lei, P. Cieplak und Y. Duan. Strike a balance: optimization of backbone torsion parameters of AMBER polarizable force field for simulations of proteins and peptides (Erratum). *J. Comput. Chem.* **27**, 994 (2006).

[113] Schropp, B. und P. Tavan. The polarizability of point-polarizable water models: density functional theory/molecular mechanics results. *J. Phys. Chem. B* **112**, 6233–6240 (2008).

[114] Schropp, B. *Entwicklung und Analyse polarisierbarer Wassermodelle für Molekulardynamiksimulationen.* Dissertation, Ludwig-Maximilians-Universität, München. In Vorbereitung.

[115] Allen, M. P. und D. J. Tildesley. *Computer simulation of liquids.* Clarendon Press, Oxford (1987).

[116] van Gunsteren, W. F. und H. J. C. Berendsen. Computer simulation of molecular dynamics: Methodology, applications, and perspectives in chemistry. *Angew. Chem. Int. Ed.* **29**, 992–1023 (1990).

[117] Becker, O. M., A. D. MacKerell, Jr., B. Roux und M. Watanabe(Herausgeber), *Computational biochemistry and biophysics.* Marcel Dekker, Inc., New York (2001).

[118] Frenkel, D. und B. Smit. *Understanding molecular simulation: From algorithms to applications.* Academic Press, San Diego (2002).

[119] van Gunsteren, W. F., D. Bakowies, R. Baron, I. Chandrasekhar, M. Christen *et al.* Biomolecular modeling: Goals, problems, perspectives. *Angew. Chem. Int. Ed.* **45**, 4064–4092 (2006).

[120] Verlet, L. Computer “experiments” on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules. *Phys. Rev.* **165**, 201–214 (1967).

[121] Grubmüller, H. *Molekulardynamik von Proteinen auf langen Zeitskalen.* Dissertation, Technische Universität München, München (1994).

[122] Eichinger, M., H. Grubmüller, H. Heller und P. Tavan. FAMUSAMM: An algorithm for rapid evaluation of electrostatic interactions in molecular dynamics simulations. *J. Comput. Chem.* **18**, 1729–1749 (1997).

[123] Grubmüller, H. und P. Tavan. Multiple time step algorithms for molecular dynamics simulations of proteins: How good are they? *J. Comput. Chem.* **19**, 1534–1552 (1998).

[124] Brooks III, C. L., M. Karplus und B. M. Pettitt. *Proteins: A theoretical perspective of dynamics, structure, and thermodynamics.* John Wiley & Sons, New York (1988).

[125] McCammon, J. A., B. R. Gelin und M. Karplus. Dynamics of folded proteins. *Nature* **267**, 585–590 (1977).

[126] Levitt, M. Molecular dynamics of macromolecules in water. *Chem. Scr.* **29A**, 197–203 (1989).

[127] van Gunsteren, W. F. und H. J. C. Berendsen. Computer simulation as a tool for tracing the conformational differences between proteins in solution and in the crystalline state. *J. Mol. Biol.* **176**, 559–564 (1984).

[128] Duan, Y. und P. A. Kollman. Pathways to a protein folding intermediate observed in a 1 microsecond simulation in aqueous solution. *Science* **282**, 740–744 (1998).

[129] Freddolino, P. L., F. Liu, M. Gruebele und K. Schulten. Ten-microsecond molecular dynamics simulation of a fast-folding WW domain. *Biophys. J.* **94**, L75–L77 (2008).

[130] Berendsen, H. J. C. Bio-molecular dynamics comes of age. *Science* **271**, 954–955 (1996).

[131] Grubmüller, H., B. Heymann und P. Tavan. Ligand binding: Molecular mechanics calculation of the streptavidin-biotin rupture force. *Science* **271**, 997–999 (1996).

[132] Spörlein, S., H. Carstens, H. Satzger, C. Renner, R. Behrendt, L. Moroder, P. Tavan, W. Zinth und J. Wachtveitl. Ultrafast spectroscopy reveals subnanosecond peptide conformational dynamics and validates molecular dynamics simulation. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 7998–8002 (2002).

[133] Lavery, R., A. Leburn, J.-F. Allemand, D. Bensimon und V. Croquette. Structure and mechanics of single biomolecules: Experiment and simulation. *J. Phys. - Condens. Mat.* **14**, R383–R414 (2002).

[134] Karplus, M. und J. A. McCammon. Molecular dynamics simulations of biomolecules. *Nature Struct. Biol.* **9**, 646–652 (2002).

[135] Carstens, H., C. Renner, A. G. Milbradt, L. Moroder und P. Tavan. Multiple loop conformations of peptides predicted by molecular dynamics simulations are compatible with nuclear magnetic resonance. *Biochemistry* **44**, 4829–4840 (2005).

[136] Stork, M., A. Giese, H. A. Kretzschmar und P. Tavan. Molecular dynamics simulations indicate a possible role of parallel β-helices in seeded aggregation of poly-Gln. *Biophys. J.* **88**, 2442–2451 (2005).

[137] Hirschberger, T., M. Stork, B. Schropp, K. F. Winklhofer, J. Tatzelt, und P. Tavan. Structural instability of the prion protein upon M205S/R mutations revealed by molecular dynamics simulations. *Biophys. J.* **90**, 3908–3918 (2006).

[138] Hukushima, K. und K. Nemoto. Exchange Monte Carlo method and application to spin glass simulations. *J. Phys. Soc. Jpn* **65**, 1604–1608 (1996).

[139] Hansmann, U. H. E. Parallel tempering algorithm for conformational studies of biological molecules. *Chem. Phys. Lett.* **281**, 140–150 (1997).

[140] Sugita, Y. und Y. Okamoto. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* **314**, 141–151 (1999).

[141] Denschlag, R., M. Lingenheil und P. Tavan. Efficiency reduction and pseudo-convergence in replica exchange sampling of peptide folding-unfolding equilibria. *Chem. Phys. Lett.* **458**, 244–248 (2008).

[142] Denschlag, R. *Effizientes Abtasten der Gleichgewichtsensembles von Peptiden in Lösung: Methoden und Anwendungen.* Dissertation, Ludwig-Maximilians-Universität, München. In Vorbereitung.

[143] Nymeyer, H. How efficient is replica exchange molecular dynamics? An analytic approach. *J. Chem. Theory Comput.* **4**, 626–636 (2008).

[144] Liu, P., B. Kim, R. A. Friesner und B. J. Berne. Replica exchange with solute tempering: A method for sampling biological systems in explicit water. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 13749–13754 (2005).

[145] Lingenheil, M., R. Denschlag, R. Reichold und P. Tavan. The "hot-solvent/cold-solute" problem revisited. *J. Chem. Theory Comput.* **4**, 1293–1306 (2008).

[146] Tanizaki, S., J. Clifford, B. D. Connelly und M. Feig. Conformational sampling of peptides in cellular environments. *Biophys. J.* **94**, 747–759 (2008).

[147] Feig, M. Is alanine dipeptide a good model for representing the torsional preferences of protein backbones? *J. Chem. Theory Comput.* **4**, 1555–1564 (2008).

[148] Egwolf, B. und P. Tavan. Continuum description of solvent dielectrics in molecular-dynamics simulations of proteins. *J. Chem. Phys.* **118**, 2039–2056 (2003).

[149] Stork, M. und P. Tavan. Electrostatics of proteins in dielectric solvent continua. I. Newton's third law marries qE forces. *J. Chem. Phys.* **126**, 165105 (2007).

[150] Stork, M. und P. Tavan. Electrostatics of proteins in dielectric solvent continua. II. First applications in molecular dynamics simulations. *J. Chem. Phys.* **126**, 165106 (2007).

[151] Bauer, S. Dissertation, Ludwig-Maximilians-Universität, München. In Vorbereitung.

[152] Press, W. H., S. A. Teukolsky, W. T. Vetterling und B. P. Flannery. *Numerical Recipes in C*. Cambridge University Press, Cambridge, UK, 2. Auflage (1992).

[153] Puls, J. und S. Stintzing. *Numerik für Physiker*. Ludwig-Maximilians-Universität München – Sektion Physik, München, 2. Auflage (2004).

[154] Blondel, A. und M. Karplus. New formulation for derivatives of torsion angles and improper torsion angles in molecular mechanics: Elimination of singularities. *J. Comput. Chem.* **17**, 1132–1141 (1996).

[155] http://iwww.bmo.physik.uni-muenchen.de/internes/BMOwiki/index.php/CMAP, Stand: 14.11.2008.

Danksagung

Meinem Betreuer Paul Tavan möchte ich für das Vertrauen danken, das er mir während der letzten Jahre entgegengebracht hat und für die große Freiheit bei der Ausgestaltung meiner Projekte. Besonders dankbar bin ich dafür, dass ich mit dem Mercaptoprolin-Projekt betraut wurde, dessen Bearbeitung mir großen Spaß gemacht hat. Auch für meine Finanzierung über ein von der VolkswagenStiftung gefördertes Projekt bin ich ihm sehr dankbar. Von seiner Begeisterung für das Präsentieren und Vortragen sowie seinem Bestreben, diese an seine Mitarbeiter weiterzugeben, habe ich bestimmt profitiert. Schließlich möchte ich ihm für unser alljährliches Skiseminar in Leogang danken. Immerhin war dies für mich der Anlass, nach mehrjähriger Pause meinen Spaß am Skifahren wiederzuentdecken.

Zwei der in dieser Arbeit vorgestellten Projekte waren Gemeinschaftsprojekte mit anderen Arbeitsgruppen. Für die gute Zusammenarbeit bei diesen Projekten möchte ich Luis Moroder und seinen Mitarbeitern sowie Beat Fierz und Thomas Kiefhaber danken. Ohne ihre Beiträge wäre diese Arbeit nicht zustande gekommen.

Für die Unterstützung bei allen Verwaltungsvorgängen bedanke ich mich bei Frau Michaelis, Frau Widmann-Diermeier und Frau Podolski, den Sekretärinnen unseres Lehrstuhls. Karl-Heinz Mantel und Rudi Schwarz danke ich für die Organisation diverser Veranstaltungen sozialer Art wie z.B. unserer Wandertage.

Am meisten wurden die Jahre meiner Promotion jedoch durch jene Personen geprägt, mit denen ich tagein, tagaus zu tun hatte: meine Doktorandenkollegen. Martin Lingenheil und Robert Denschlag waren in ihrer Begeisterung und ihrem Engagement für ihre Arbeit immer vorbildlich. Dafür, dass sie sich immer die Zeit genommen haben, sich meiner Probleme und Fragen anzunehmen und sich ihnen in der ihnen eigenen Diskussionwut zu widmen, bin ich sehr dankbar. Ohne ihre Unterstützung wäre meine Arbeit in dieser Form nicht möglich gewesen. Nicht weniger hilfsbereit waren die restlichen Mitglieder unserer Arbeitsgruppe, Thomas Hirschberger, Verena Schultheis, Galina Babizki, Bernhard Schropp, Christine Lutz, Sebastian Bauer und Benjamin Rieff. Bei Thomas Hirschberger möchte ich mich für die geduldige Einführung in die Verwaltung unserer Rechner bedanken, bei Sebastian Bauer für die bereitwillige Übernahme meiner Pflichten in diesem Bereich. Ihm danke ich auch für das Korrekturlesen von Teilen meiner Doktorarbeit. Bernhard Schropp und Robert Denschlag danke ich für Unterstützung bei der Wartung unserer Rechencluster.

Ganz besonderen Dank hat jedoch meine langjährige Zimmerkollegin Martina Stork verdient, die mir immer mit Rat und Tat zur Seite stand. Besonders wertvoll waren ihre EGO-Kenntnisse, die mir bei der CMAP-Implementierung sehr geholfen haben. Und auch für das Korrekturlesen von Teilen meiner Doktorarbeit möchte ich ihr noch einmal ausdrücklich danken.

Lebenslauf

Name:	Rudolf Reichold
Geburtsdatum:	27.02.1978
Geburtsort:	München

07/2004 – 08/2008	Wissenschaftlicher Mitarbeiter am Lehrstuhl für BioMolekulare Optik, Theoretische Biophysik, LMU München
12/2003	Diplom in Physik
11/1998 – 12/2003	Studium der Physik, LMU München
08/1997 – 08/1998	Zivildienst, Kreiskrankenhaus Starnberg
06/1997	Abitur
1988 – 1997	Otto-von-Taube-Gymnasium Gauting