

---

# **Molekulare Grundlagen von Proteinfehlfaltungskrankheiten: Methoden zur Molekulardynamiksimulation umgebungsgesteuerter Modellpeptide**

Martin Lingenheil

---



München 2009



---

**Molekulare Grundlagen von  
Proteinfehlfaltungskrankheiten:  
Methoden zur Molekulardynamiksimulation  
umgebungsgesteuerter Modellpeptide**

**Martin Lingenheil**

---

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

vorgelegt von  
Martin Lingenheil  
aus Kaufbeuren

München, im Juli 2009

Erstgutachter: Prof. Dr. Paul Tavan

Zweitgutachter: Prof. Dr. Martin Zacharias

Tag der mündlichen Prüfung: 2. Oktober 2009

# Inhaltsverzeichnis

<b>Zusammenfassung</b>	<b>vii</b>
<b>1 Einleitung</b>	<b>1</b>
1.1 Neurodegenerative Krankheiten . . . . .	1
1.2 Proteinfaltung und -fehlfaltung . . . . .	3
1.2.1 Proteinstruktur . . . . .	4
1.2.2 Faltungsprozess und Zeitskalen . . . . .	9
1.2.3 Partiiell gefaltete Zustände und Fehlfaltung . . . . .	11
1.2.4 Fehlfaltung des Prion-Proteins . . . . .	12
1.3 Techniken der Molekulardynamiksimulation . . . . .	14
1.3.1 Molekularmechanik . . . . .	14
1.3.2 Elektrostatische Wechselwirkung . . . . .	15
1.3.3 Temperaturkontrolle . . . . .	17
1.3.4 Das Abtastproblem . . . . .	18
1.3.5 <i>Replica-Exchange</i> -Techniken . . . . .	19
1.4 Aufbau und Gliederung der Arbeit . . . . .	23
<b>2 Minimalinvasive Temperaturkontrolle</b>	<b>27</b>
<b>3 Optimale Strategien für <i>Replica-Exchange</i>-Simulationen</b>	<b>43</b>
3.1 Optimale Temperaturleitern . . . . .	43
3.2 Effiziente Austauschschemata . . . . .	47
<b>4 Stabilität von H1 in Umgebungen unterschiedlicher Polarität</b>	<b>55</b>
<b>5 <i>Résumé</i> und Ausblick</b>	<b>81</b>
<b>Literaturverzeichnis</b>	<b>105</b>
<b>Danksagung</b>	<b>106</b>
<b>Lebenslauf</b>	<b>108</b>



# Zusammenfassung

Eine Doktorarbeit? Jetzt bin ich aber erleichtert. Ich dachte schon, es sei eine schreckliche Krankheit.

---

*Volzotan Smeik [1]*

Bei vielen neurodegenerativen Erkrankungen wie der Alzheimerschen, der Parkinsonschen und der Creutzfeld-Jakob-Krankheit ist die Fehlfaltung und Aggregation von bestimmten Proteinen und Peptiden ein wichtiger pathogener Faktor. Die Ursachen und Mechanismen der Fehlfaltung dieser Moleküle sind bis heute größtenteils unbekannt, insbesondere weil teilweise entfaltete Zustände der beteiligten Proteine und Peptide, die oftmals Startpunkte der Fehlfaltungsprozesse bilden, mit den gängigen experimentellen Techniken strukturell kaum aufzuklären sind. Im Gegensatz dazu lassen sich mit Hilfe von Molekulardynamik-(MD)-Simulationen die beteiligten Strukturen im Prinzip mit atomarer Auflösung charakterisieren. Ziel dieser Arbeit war es daher, Methoden zur MD-Simulation von Modellpeptiden zu entwickeln und anzuwenden, um Einsichten in die ersten Schritte der angesprochenen Fehlfaltungsprozesse zu gewinnen.

Da die Faltungseigenschaften von Peptiden stark von den Eigenschaften der Lösungsumgebung und hier insbesondere von der Temperatur abhängen, habe ich im ersten Teil meiner Arbeit eine Strategie zur Kontrolle der Temperatur bei MD-Simulationen von Makromolekülen in Lösung unter minimaler Störung der Dynamik entwickelt. Im Gegensatz zu gängigen Vorgehensweisen werden mit dieser neu entwickelten Strategie zur Temperaturkontrolle die Konformationsübergänge eines Peptids nicht verlangsamt, und das durch die Simulation abgetastete statistische Ensemble bleibt ungestört. Um das effiziente Abtasten des Konformationsraums mit der Fehlfaltung assoziierter Peptide zu gewährleisten, muss darüberhinaus auf sogenannte *replica-exchange*-Techniken zurückgegriffen werden, die durch einen regelmäßigen Konfigurationsaustausch zwischen parallelen Simulationen bei unterschiedlichen Temperaturen eine schnellere Konformationsdynamik des Peptids bewirken. Ein weiterer methodischer Teil meiner Arbeit beschäftigt sich daher mit Regeln zur optimalen Wahl der Temperaturleiter und des Austauschschemas für den Einsatz dieser Techniken. Insbesondere habe ich gezeigt, dass bisher bei der Ableitung entsprechender Regeln von falschen Voraussetzungen ausgegangen wurde, weshalb nur suboptimale Ergebnisse erzielt werden konnten. Aus der mathematischen Analyse des Problems und anhand eines Monte-Carlo-Modells habe ich eine tatsächlich optimale Strategie

entwickelt.

Schließlich habe ich diese Strategien auf die Simulation eines Modellpeptides angewandt, um einen Aspekt der Fehlfaltung des mit der Creutzfeld-Jakob-Krankheit assoziierten Prion-Proteins (PrP) näher zu untersuchen. Ziel einer Reihe von in diesem Rahmen durchgeführten *replica-exchange*-Simulationen war es, die Stabilität der ersten  $\alpha$ -Helix (H1) von PrP gegen ihre Entfaltung zu untersuchen. Hierzu wurde ein Modell-Peptid mit einer H1 entsprechenden Sequenz in unterschiedlich polaren Lösungsmitteln simuliert. Dabei zeigte sich, dass die entsprechende Peptidsequenz in Wasser mehrheitlich keine  $\alpha$ -helikale Faltung annimmt und vergleichsweise schnell entfaltet, während mit abnehmender Polarität des Lösungsmittels die Stabilität gegen eine Entfaltung deutlich zunimmt. Damit bestätigt sich eine Hypothese von Hirschberger *et al.* [*Biophys. J.* **90**, 3908-3918 (2006)], dass H1 für die Fehlfaltung von PrP keine Barriere darstellt, falls dieser Prozess, wie vermutet, als ersten Schritt den Wechsel von H1 von einer schwach in eine stark polare Umgebung beinhaltet.



# 1 Einleitung

Fangen wir weiter an!

---

*Bast I. Fonk*

Die Gesellschaften in den entwickelten Ländern stehen vor einem Umbruch. Die Lebenserwartung von 76,6 Jahren, die für einen heute in Deutschland geborenen Jungen berechnet wird, übersteigt die meiner Generation bei der Geburt prognostizierte durchschnittliche Lebensspanne um 7,1 Jahre und jene der Generation meines Doktorvaters sogar um 12,4 Jahre [2]. Schreibt man diese Entwicklung anhand der heutigen Bevölkerungsstruktur in die Zukunft fort, so wird sich in Deutschland die Zahl der Menschen über 80 Jahren von 3,5 Millionen im Jahr 2005 auf knapp über 10 Millionen im Jahr 2050 fast verdreifachen [3].

Was für sich genommen eine erfreuliche Entwicklung darstellt, rückt eine Reihe von neurodegenerativen Erkrankungen in den Fokus, deren Prävalenz oftmals mit zunehmendem Alter sprunghaft ansteigt [4]. So wird beispielsweise für die häufigste dieser Krankheiten, den Morbus Alzheimer, für das Jahr 2050 eine Zahl von weltweit rund 106 Millionen Erkrankten berechnet [5] (etwa 1,2% der prognostizierten Weltbevölkerung). Bereits heute werden die Kosten für die Pflege und die Behandlung von Alzheimer-Patienten in den USA bei vier Millionen Erkrankten auf jährlich 200 Millionen Dollar geschätzt [4]. Für die zweithäufigste neurodegenerative Erkrankung, den Morbus Parkinson, wird eine Verdopplung der Fälle innerhalb der nächsten 25 Jahre prognostiziert [6]. Eine intensive Erforschung der molekularen Grundlagen und Ursachen mit dem Ziel einer verbesserten und vor allem frühzeitigen Diagnose sowie einer ursächlichen Therapie ist also dringend geboten.

## 1.1 Neurodegenerative Krankheiten

Die Alzheimersche und die Parkinsonsche Krankheit gehören in eine Gruppe von Erkrankungen, die sich durch gemeinsame zelluläre und molekulare Mechanismen der Krankheitsentstehung auszeichnen [7]. Beobachtet wird das Absterben von Nervenzellen in bestimmten, für die jeweilige Krankheit typischen Hirnregionen und die Ablagerung von fehlerhaft gefalteten Proteinen oder deren Bruchstücken (Peptiden) in sogenannten amyloiden Fasern, linearen Aggregaten mit ähnlichen biophysikalischen und strukturellen Eigenschaften [8]. Dabei ist die

Fehlfaltung und Aggregation jeweils bestimmter Proteine und Peptide charakteristisch für verschiedene Krankheiten aus dieser Gruppe; etwa das A $\beta$ -Peptid und das  $\tau$ -Protein bei der Alzheimer-Krankheit oder das Protein  $\alpha$ -Synuclein bei der Parkinson-Krankheit. Es scheint heute wahrscheinlich, dass die Fehlfaltung und die frühen Schritte hin zu einer Aggregation dieser Proteine eine entscheidende Rolle spielen für das Absterben der Nervenzellen und damit auch für die Pathologie der entsprechenden Krankheiten [9–11].

Weitere, wesentlich seltener auftretende und nicht unbedingt altersassoziierte Erkrankungen aus dieser Gruppe sind zum Beispiel die amyotrophe Lateralsklerose, die Huntingtonsche Krankheit und die Creutzfeld-Jakob-Krankheit (CJD). Letztere gehört zu einer ungewöhnlichen Untergruppe der Proteinfehlfaltungskrankheiten, den transmissiblen spongiformen Enzephalopathien (TSE), die im Gegensatz zu den erstgenannten übertragbare Infektionskrankheiten sind. Der Hauptbestandteil der amyloiden Ablagerungen bei allen TSE ist, wie bei den anderen Amyloid-Krankheiten, eine abnorm gefaltete Form eines körpereigenen Proteins, in diesem Fall des Prion-Proteins PrP. Die normale (zelluläre) Form erhielt die Bezeichnung PrP<sup>C</sup> (engl. *prion protein cellular*), die abnormale Form wegen ihrer Verbindung zur Schafskrankheit Scrapie die Bezeichnung PrP<sup>Sc</sup> (engl. *prion protein scrapie*). Gemäß der heute weitgehend akzeptierten Prion-Hypothese ist PrP<sup>Sc</sup> gleichzeitig der Erreger der TSE, indem es seine ungewöhnliche Faltung im Rahmen eines autokatalytischen Prozesses auf PrP<sup>C</sup> überträgt [12–14]. Ein solches infektiöses Protein wird nach einem Vorschlag von Prusiner [13] als Prion bezeichnet.

Die bekannteste Vertreterin der TSE ist sicherlich eine Tierkrankheit: Die bovine spongiforme Enzephalopathie (BSE), welche als *Rinderwahnsinn* in den 1990iger Jahren in Großbritannien und schließlich um die Jahrtausendwende in Deutschland hohe Wellen schlug, Journalisten in die Kuhställe trieb, Verbraucher in Angst und Schrecken versetzte und durch den vorübergehenden Zusammenbruch des Rindfleischmarktes enorme wirtschaftliche Schäden verursachte. Auf eine Parlamentsanfrage aus dem Jahr 2001 nach den Kosten der BSE-Krise für den Steuerzahler im Vereinigten Königreich gab die Regierung eine Schätzung über 4,4 Milliarden Pfund Stirling bis zum Ende des Fiskaljahrs 2001/2002 zu Protokoll [15].

Die große Aufmerksamkeit, die die BSE sehr bald nach ihrer erstmaligen Beschreibung Mitte der 80iger Jahre des vergangenen Jahrhunderts und ihrer epidemischen Ausbreitung in der Rinderpopulation Großbritanniens erfuhr, rührte in erster Linie aus der Befürchtung her, die Krankheit könne möglicherweise durch den Verzehr von mit Prionen verseuchten Rinderprodukten auf den Menschen übertragen werden. Bei einer geschätzten Anzahl von 1-3 Millionen infizierter Tiere [16, 17], deren Fleisch bis Mitte der 90iger Jahre größtenteils in die menschliche Nahrung gelangt war, stellte eine solche Übertragung ein beträchtli-

ches Risiko für die allgemeine Gesundheit dar. Im Jahr 1996 wurde dann tatsächlich eine neue Variante der CJD (vCJD) beschrieben [18], die ungewöhnlich junge Menschen betraf, im Gegensatz zur eigentlich als Altersdemenz auftretenden CJD [19]. Bald gab es auch sehr stichhaltige Gründe anzunehmen, dass der Erreger der vCJD mit dem BSE-Erreger identisch ist [20, 21], und dass damit das Auftreten dieser neuen Variante am wahrscheinlichsten auf den Verzehr verseuchten Rindfleisch zurückzuführen ist. Glücklicherweise blieb das Auftreten der neuen Variante der CJD auf einige hundert Fälle beschränkt [22, 23], was vermutlich der geringen Infektiösität des BSE-Erregers für den Menschen zuzuschreiben ist [24].

Die BSE-Krise führte zu einer intensiven Erforschung der TSE in den vergangenen zwei Jahrzehnten und einigen Fortschritten. So wurde die detaillierte dreidimensionale Struktur von PrP<sup>C</sup> aufgeklärt [25], es wurden verschiedene Unterformen von PrP<sup>Sc</sup>, sogenannte Prion-Stämme, identifiziert, die sich vermutlich durch unterschiedliche dreidimensionale Strukturen im Detail unterscheiden und jeweils mit unterschiedlichen Erscheinungsformen der TSE assoziiert sind [26]. Die dreidimensionale Struktur von PrP<sup>Sc</sup> selbst ist jedoch noch immer ebenso ungeklärt wie die molekularen Mechanismen, die es PrP<sup>Sc</sup> ermöglichen, seine dreidimensionale Struktur auf PrP<sup>C</sup> zu übertragen. Gleiches gilt für die Einflussfaktoren und Mechanismen, die auf molekularer Ebene zur Fehlfaltung und Aggregation von A $\beta$ ,  $\tau$  und  $\alpha$ -Synuclein und damit zur Alzheimerschen bzw. zur Parkinsonschen Krankheit führen. Bis heute existiert für keine der Proteinefehlfaltungskrankheiten eine ursächliche Therapie – sie sind also bis dato durchgängig unheilbar. Eine Prophylaxe oder Therapie setzt am erfolgversprechendsten bei den Ursachen einer Krankheit an. Es ist daher dringend nötig aufzuklären, welche allgemeinen und spezifischen Faktoren die Faltung bzw. Fehlfaltung und die Aggregation der jeweiligen Proteine und Peptide bestimmen.

Im vergangenen Jahrzehnt reifte allmählich die Erkenntnis, dass die Neigung von Proteinen zur Fehlfaltung und Aggregation ein sehr verbreitetes Phänomen [27–33] von außerordentlicher, wegen seiner Verbindung zu den neurodegenerativen Krankheiten weit über die reine Grundlagenforschung hinausweisender Bedeutung ist. Das alte Paradigma der Proteinfaltung, wonach ein Protein nach seiner Synthese quasi spontan und mit traumwandlerischer Sicherheit einer festen dreidimensionalen Struktur zustrebt [34, 35], musste in den letzten Jahren zunehmend revidiert werden [36, 37].

## 1.2 Proteinfaltung und -fehlfaltung

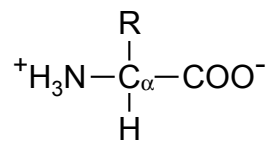
Proteine sind gleichermaßen Wunderwerke und Arbeitspferde des Lebens. Sie sind an jedem Vorgang in lebenden Organismen beteiligt: Sie transportieren Sauerstoff, katalysieren den Zuckerabbau, übermitteln als Rezeptorproteine Signale

oder kontrahieren als molekulare Motoren den Muskel. Die Liste ließe sich beliebig fortsetzen [38]. Proteine sind die entscheidenden Spieler in jedem der komplexen Netzwerke aus Verstärkung und Hemmung chemischer Reaktionen [39], welche das abstrakte Phänomen des Lebens in der materiellen Welt verankern.

Die Vielseitigkeit und Spezifität dieser Moleküle ist dabei eng verbunden mit ihrer erstaunlichen Fähigkeit zur steuerbaren, selbstorganisierten Faltung in bestimmte dreidimensionale Strukturen [40]. Mit dem Verlust seiner festen räumlichen Struktur verliert ein Protein üblicherweise auch seine Funktionstüchtigkeit [41–43]. Allerdings gibt es eine ganze Reihe von Proteinen, die, obwohl meist in der wässrigen Umgebung des Zytosols anzutreffen, dort überhaupt keine feste räumliche Struktur annehmen und erst durch die Wechselwirkung mit einer bestimmten Umgebung oder anderen Komponenten der Zelle an Struktur und damit ihre Funktion gewinnen [44, 45]. Zu dieser Gruppe gehört beispielsweise das  $\alpha$ -Synuclein, das erst auf einer Zellmembran „schwimmend“ eine definierte Struktur annimmt [46]. Auch das zelluläre Prion-Protein PrP<sup>C</sup> ist teilweise nicht zu einer festen dreidimensionalen Struktur gefaltet [25]. Interessanterweise sind in der Mehrzahl gerade diese ganz oder teilweise unstrukturierten Proteine Hauptakteure der verschiedenen Proteinfehlfaltungskrankheiten [37].

### 1.2.1 Proteinstruktur

Proteine entstehen als Makromoleküle durch die lineare Verkettung von ca. 100 bis ca. 1000 Monomeren<sup>1</sup>. Die einzelnen Bausteine entstammen einer Gruppe von 20 verschiedenen Aminosäuren, die daher als proteinogen bezeichnet werden [39]. Aminosäuren besitzen eine gemeinsame molekulare Grundstruktur<sup>2</sup>, die in Abb. 1.1 dargestellt ist.



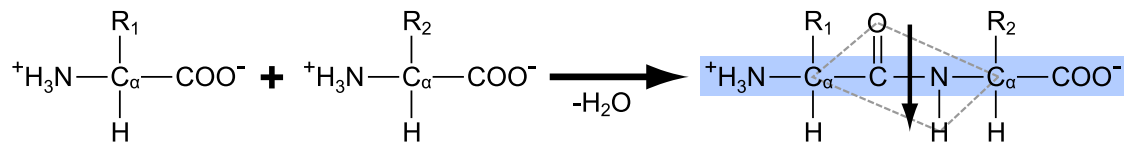
**Abbildung 1.1:** Strukturformel einer Aminosäure

Um ein zentrales Kohlenstoffatom ( $\text{C}_\alpha$ ) gruppieren sich neben einem inerten Wasserstoffatom eine Aminogruppe ( $\text{NH}_3^+$ ), eine Carboxylatgruppe ( $\text{COO}^-$ ) und eine Restgruppe (R), wobei sich die 20 proteinogenen Aminosäuren eben durch diese

---

<sup>1</sup>Riesenproteine wie etwa das Muskelprotein Titin [47] könne allerdings über 10 000 Monomere in sich vereinigen.

<sup>2</sup>Genau genommen trifft die in Abb. 1.1 gezeigte Struktur nur auf die sog.  $\alpha$ -Aminosäuren zu. Im Rahmen dieser Arbeit soll jedoch die Diskussion auf diese Gruppe beschränkt bleiben.



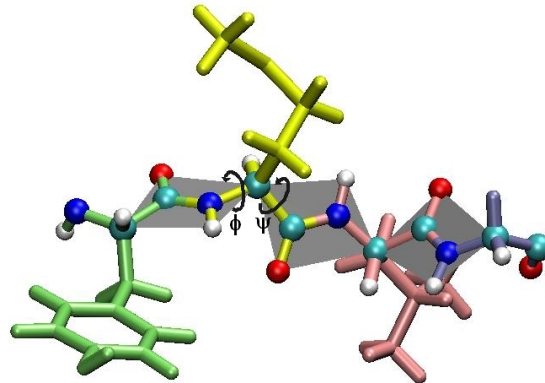
**Abbildung 1.2:** Bildung einer Peptidgruppe (gestrichelt umrandet) durch die Kondensationsreaktion zweier Aminosäuren: Da die Atome alle in einer Ebene liegen, bezeichnet man die Gruppe auch als Peptidplättchen. Die zentrale Bindung C-N ist bei Raumtemperatur torsionsstabil. Aufgrund der starken Elektronegativitätsunterschiede der Konstituenten trägt die Gruppe einen starken Dipol (durch den senkrechten Pfeil angedeutet). Die hellblau hinterlegte Kette von Atomen kann an ihren beiden Enden durch Reaktionen desselben Typs verlängert werden und bildet dann das sog. Peptidrückgrat.

Restgruppe unterscheiden. Der Charakter des Restes kann recht unterschiedlich sein. Das Spektrum reicht von kleinen, unpolaren Gruppen, beispielsweise einer Methylgruppe bei der Aminosäure Alanin, über aromatische Ringe (z.B. Phenylalanin) und polaren Gruppen (z.B. Serin) bis hin zu positiv und negativ geladenen Resten (z.B. Lysin und Aspartat). In neutraler, wässriger Lösung liegt eine solche Aminosäure, wie in Abb. 1.1 gezeigt ist, als Zwitterion vor.

Da das  $C_{\alpha}$ -Atom üblicherweise vier verschiedene Substituenten trägt – die einzige Ausnahme bildet die Aminosäure Glycin – handelt es sich bei den entsprechenden Aminosäuren um chirale Moleküle mit jeweils zwei Enantiomeren, die als L- und D-Aminosäuren bezeichnet werden. Interessanterweise sind ausschließlich die jeweiligen L-Formen proteinogen. Die tieferen Ursachen, die im Lauf der Evolution proteinbasierten Lebens zur Homochiralität führten, also zur Selektion ausschließlich der L-Aminosäuren als Bausteine der Proteine, sind bis heute nur andeutungsweise aufgeklärt [48, 49].

In einer in Abb. 1.2 skizzierten Reaktion bilden die Carboxylat- und Amino-Gruppen zweier Aminosäuren bei der Polymerisation eine sogenannte Peptidbindung. Die neugebildete Peptidgruppe (gestrichelter Kasten in Abb. 1.2) ist eine ebene Struktur mit einem starken Dipol (Pfeil in Abb. 1.2). Eine Torsion um die zentrale Bindungsachse (C-N) ist bei Raumtemperatur unmöglich, während Torsionen um die Bindungsachsen  $C_{\alpha}$ -C und N- $C_{\alpha}$  möglich sind. Diese Torsionen bilden bei weiterem Wachstum der Kette an den noch freien Amino- bzw. Carboxyl-Enden die Freiheitsgrade des Proteinrückgrates (in Abb. 1.2 blau hinterlegt).

Man kann sich ein Protein in einer vereinfachten Sicht also als Abfolge von drehbar miteinander verbundenen Peptidplättchen vorstellen. Diese Vorstellung ist in Abb. 1.3 – nun etwas weniger schematisch – visualisiert. Die Peptidplättchen entlang des Proteinrückgrates sind grau hinterlegt, um ihre flächige Struktur hervorzuheben. Jeweils zwischen zwei Plättchen befindet sich ein  $C_{\alpha}$ -Atom, an welchem sich stets eine Restgruppe vom Rückgrat abzweigt. Die beiden Torsionswinkel für die Drehung um die N- $C_{\alpha}$ - und die  $C_{\alpha}$ -C-Bindung werden der Konvention nach

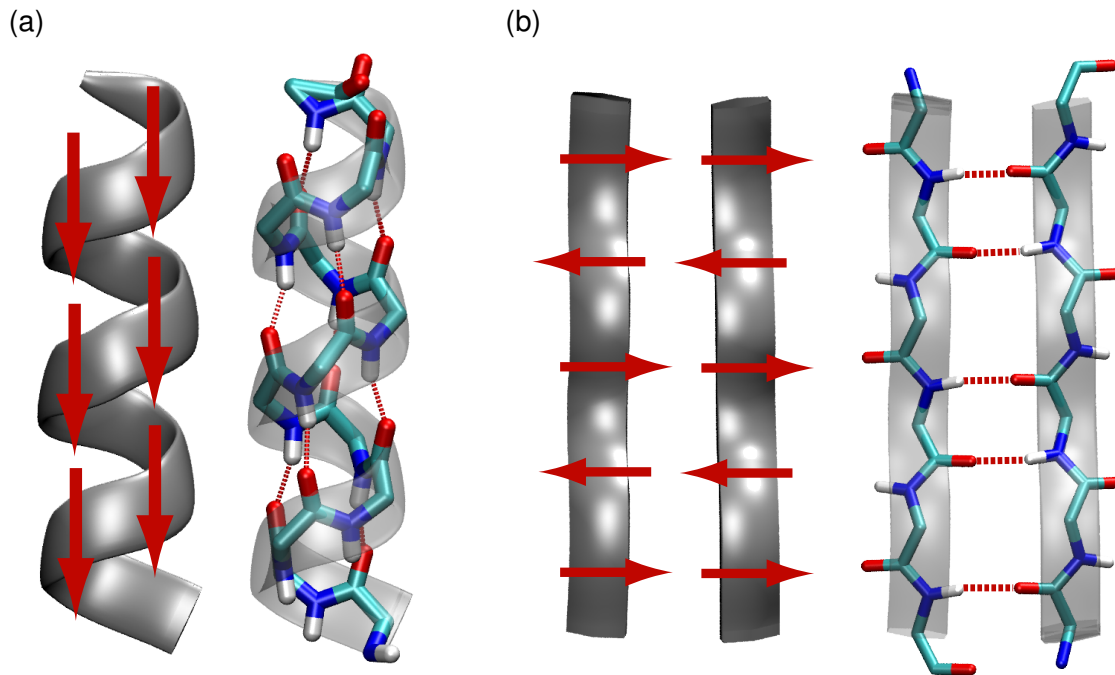


**Abbildung 1.3:** Ausschnitt aus einem Protein: Gezeigt ist die Abfolge der vier Aminosäuren Tyrosin (grün), Methionin (gelb), Leucin (rot) und Glycin (blau). Die Atome des Proteinerückgrats sind als Kugeln dargestellt (türkis: C, rot: O, blau: N, weiß: H), die Seitenketten im Stäbchenmodell. Die Torsionswinkel  $\phi$  und  $\psi$  zwischen den grau hinterlegten starren Peptidplättchen bilden die Freiheitsgrade des Proteinerückgrats. Die Abbildung ist der Diplomarbeit [50] von Christine Lutz entnommen.

mit  $\phi$  bzw.  $\psi$  bezeichnet.

Die Abfolge der Aminosäuren in der Kette, die sogenannte Sequenz oder Primärstruktur, bestimmt über die räumliche Anordnung, die die gesamte Kette in ihrer natürlichen Umgebung annimmt [34, 35]. Diese räumliche Anordnung des gesamten Moleküls wird als Tertiärstruktur bezeichnet. Neben ihrer individuellen Tertiärstruktur weisen viele Proteine trotz sehr unterschiedlicher Sequenzen stellenweise die gleiche lokale Ordnung des Peptidrückgrates zu sogenannten Sekundärstrukturelementen auf. Besonders wichtig für das Verständnis der strukturellen und funktionellen Eigenschaften von Proteinen sind in diesem Zusammenhang  $\alpha$ -Helix und  $\beta$ -Faltblatt, die 1951 von Pauling, Corey und Branson auf Basis einer Analyse der Geometrie der planaren Peptidgruppe und der damit möglichen Wasserstoffverbrückungen vorgeschlagen wurden [51, 52].

Bei der in Abb. 1.4a gezeigten  $\alpha$ -Helix verläuft das Peptidrückgrat helikal gewunden, wobei die Carbonyl-Gruppe (C=O) der Aminosäure  $i$  mit der Amidgruppe (N-H) der Aminosäure  $i + 4$  jeweils über eine Wasserstoffbrücke verbunden ist, die die Grundlage für die Stabilität dieser Art der Faltung bildet. Da sämtliche Dipole des Rückgrates in einer  $\alpha$ -Helix parallel stehen, addieren diese sich zu einem charakteristischen Makrodipol. Weitere stabilisierende Wechselwirkungen können über die Seitengruppen vermittelt werden, etwa über unterschiedlich geladene ionische Reste an den Positionen  $i$  und  $i + 4$  der Aminosäuresequenz. Bei helikaler Faltung des Peptidrückgrats können solche ionischen Reste aufgrund ihrer räumlichen Nähe eine sogenannte Salzbrücke ausbilden [53].



**Abbildung 1.4:** Die beiden wichtigsten Sekundärstrukturelemente in Proteinen jeweils als Schema (links) und als Stäbchendarstellung des Proteinrückgrats (rechts): Der Verlauf des Peptidrückgrats wird durch das silbergraue Band angedeutet; die Dipole der einzelnen Peptidplättchen werden durch rote Pfeile symbolisiert. In der Stäbchendarstellung werden zudem Wasserstoffbrücken durch eine gestrichelte rote Linie angedeutet. Die Farbkodierung der Atome in der Stäbchendarstellung entspricht jener in Abb. 1.3. Die in Teil (a) gezeigte  $\alpha$ -Helix ist gekennzeichnet durch eine schraubenförmig gewundene Konformation des Peptidrückgrats, bei der die Dipole der Peptidplättchen sämtlich parallel zueinander stehen. Die Carbonyl-Gruppen (C=O) in einer Windung der Helix sind jeweils mit den Amid-Gruppen (N-H) der folgenden Windung über Wasserstoffbrücken verbunden. Die Konformation des Peptidrückgrates im  $\beta$ -Faltblatt (b) ist linear gestreckt, wobei aufeinanderfolgende Peptid-Dipole antiparallel zueinander stehen. In einem  $\beta$ -Faltblatt verläuft ein zweiter Strang parallel oder antiparallel zum ersten, wobei die direkt miteinander interagierenden Dipole der beiden Stränge stets parallel übereinander angeordnet sind. Dies entspricht der in der Stäbchendarstellung gezeigten Wasserstoffverbrückung.

Im Gegensatz zur  $\alpha$ -Helix haben beim in Abb. 1.4b gezeigten  $\beta$ -Faltblatt die durch Wasserstoffbrücken verbundenen Aminosäuren keinen festen Abstand in der Sequenz eines Proteins. Stattdessen bilden hier die Peptidgruppen zweier unabhängiger Stränge stabilisierende Wasserstoffbrücken aus. Die Einzelstränge können dabei aus Teilen der Proteinkette gebildet werden, die in der Sequenz weit entfernt liegen; es sind sogar durchaus auch intermolekulare Faltblätter möglich und experimentell nachgewiesen [54].

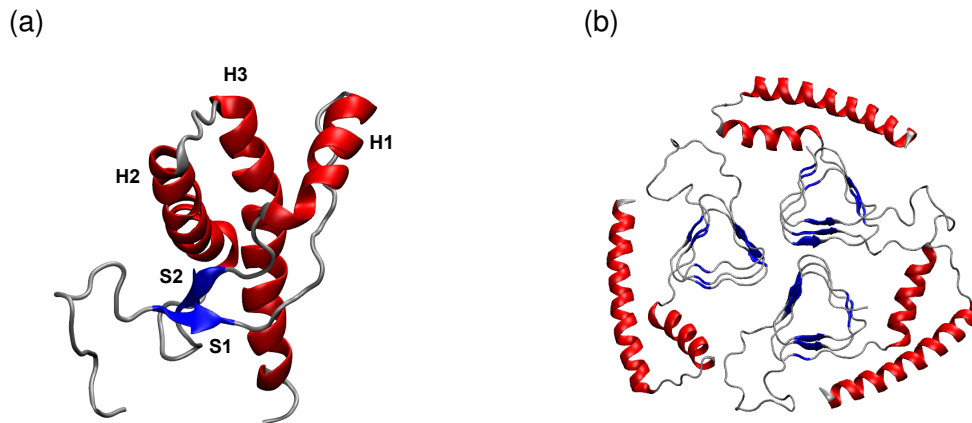
Tatsächlich sind es gerade solche intermolekularen  $\beta$ -Faltblätter, die den Kern jener amyloiden Fibrillen ausmachen [8], welche ein verbindendes Element der eingangs umrissenen Krankheiten bilden. Die einzelnen Stränge des Faltblatts

verlaufen dabei senkrecht zur Achse der amyloiden Fibrille und sind miteinander in Faserrichtung durch Wasserstoffbrücken verbunden. Eine besondere Rolle beim Wachstum der Fibrille kommt den nicht abgesättigten Amid- und Carbonyl-Gruppen an den beiden Längsseiten des  $\beta$ -Faltblatts zu (siehe Abb. 1.4b), an die sich sehr leicht ein weiterer Strang anlagern kann. Eine strukturelle Organisation wie die beschriebene wird als „gekreuzte  $\beta$ -Struktur“ (engl. *cross- $\beta$  structure*) bezeichnet und liefert im Röntgen- oder Elektronenbeugungsexperiment charakteristische Reflexe [55], die als Identifikationsmerkmal dienen [8]. Darüber hinaus sind jedoch die Details der strukturellen Organisation der meisten Amyloid-Strukturen schlecht oder gar nicht aufgeklärt, da sie sich den Methoden entziehen, die üblicherweise zur Strukturaufklärung in atomarer Auflösung angewandt werden: Weil sich aus den Fasern einerseits bisher keine dreidimensionalen Kristalle ausreichender Qualität züchten lassen, fällt die Röntgenkristallographie [56] als Methode aus; da sie andererseits unlöslich sind, kann auch die übliche Kernspinresonanzspektroskopie in Lösung [57] nicht zur Anwendung kommen. Die Unkenntnis des Ergebnisses der Proteinfehlfaltung erschwert naturgemäß die Erforschung des Fehlfaltungs- und Aggregationsprozesses und ist damit auch ein Haupthindernis bei der Erforschung der molekularen Grundlagen der Proteinfehlfaltungskrankheiten.

Den beiden in Abb. 1.4 gezeigten Sekundärstrukturelementen ist gemein, dass sie durch die starke Wasserstoffverbrückung eine relativ geringe interne Flexibilität aufweisen und daher üblicherweise das Grundgerüst jener Proteine bilden, die in ihrer natürlichen Umgebung eine definierte Faltung annehmen. Da diese beiden Sekundärstrukturelemente bei vielen verschiedenen Proteinen auftreten, die keine Ähnlichkeit bezüglich der Aminosäuresequenz aufweisen, ist die Neigung zur Sekundärstrukturbildung offensichtlich eine intrinsische Eigenschaft der Polypeptidkette, auf welche die Aminosäuresequenz nur modulierenden Einfluss ausübt; so etwa im Fall der  $\alpha$ -Helix beispielsweise über die oben bereits angesprochenen  $i, i + 4$ -Salzbrücken.

Neben diesen beiden Hauptmotiven der Sekundärstruktur existieren einige weitere, weniger häufige Elemente und schließlich auch Bereiche des Proteinrückgrats ohne feste Struktur. Die Tertiärstruktur eines Proteins, also die dreidimensionale Faltung einer ganzen Proteinkette, kann man als räumliche Anordnung der Sekundärstrukturelemente relativ zueinander auffassen. Als Beispiel zeigt Abb. 1.5a in einer Bänderdarstellung des Proteinrückgrats den gefalteten Teil des menschlichen zellulären Prion-Proteins PrP<sup>C</sup>. Es besteht aus drei  $\alpha$ -Helizes, üblicherweise mit H1, H2 und H3 bezeichnet und einem sehr kurzen, aus zwei Strängen S1 und S2 zusammengesetzten  $\beta$ -Faltblatt. Der in Abb. 1.5a gezeigte Teil von PrP<sup>C</sup> umfasst etwa 100 Aminosäuren im zweiten Teil der Proteinkette, die eine feste Tertiärstruktur besitzen, während ein erster Teil der Kette von ebenfalls etwa 100 Aminosäuren nativ ungefaltet vorliegt und in der Abbildung nicht sichtbar ist.





**Abbildung 1.5:** Tertiär- und Quartärstruktur am Beispiel des menschlichen Prion-Proteins PrP: Es handelt um eine Bänderdarstellung des Proteinrückgrates, wobei  $\alpha$ -Helizes rot und  $\beta$ -Faltblätter dunkelblau gekennzeichnet sind. Teil (a) zeigt die Tertiärstruktur des gefalteten Teils von PrP in seiner zellulären Form PrP<sup>C</sup>. Wesentliche Strukturmodule sind die  $\alpha$ -Helizes H1, H2 und H3, sowie ein kurzes  $\beta$ -Faltblatt bestehend aus den Strängen S1 und S2. Teil (b) zeigt als Beispiel für Quartärstruktur ein Modell für einen Trimer aus fehlgefaltetem Prion-Protein PrP<sup>Sc</sup> nach Stork *et al.* [58], der den Grundbaustein für die amyloiden Fibrillen bei den TSE bilden könnte.

Zur tertiären Ebene der Proteinstruktur gesellt sich im Fall von Proteinkomplexen die Quartärstruktur, das heißt die räumliche Konfiguration mehrerer Proteinketten, die den Proteinkomplex aufbauen. Ein Beispiel hierfür zeigt Abb. 1.5b. Es handelt sich um ein Modell nach Stork *et al.* [58] für den Grundbaustein der amyloiden Fasern bei den TSE: Dem Modell zufolge ein Trimer aus PrP<sup>Sc</sup>-Molekülen. Die Tertiärstruktur der einzelnen Moleküle in der modellierten PrP<sup>Sc</sup>-Faltung weicht deutlich von der PrP<sup>C</sup>-Struktur in Abb. 1.5a ab: Die kleinste  $\alpha$ -Helix H1 und ein Teil von H2 sind in Abb. 1.5b nicht mehr vorhanden. Dafür ist ein beträchtlicher Teil der Kette als sog.  $\beta$ -Helix organisiert. Das Proteinrückgrat verläuft dabei in Form einer Wendeltreppe mit dreieckiger Grundfläche, wobei die Dreiecksseiten aus  $\beta$ -Faltblättern bestehen. An die  $\beta$ -Helizes des in Abb. 1.5b gezeigten Trimers können oben und unten weitere Trimere angelagert werden, wodurch eine amyloide Fibrille mit der charakteristischen gekreuzten  $\beta$ -Struktur entsteht.

## 1.2.2 Faltungsprozess und Zeitskalen

Es ist lange bekannt [34, 35], dass die Primärstruktur, also die Abfolge der Aminosäuren in der Polypeptidkette, die Faltung eines Proteins bestimmt. So faltet sich eine große Zahl von meist kleineren Proteinen in wässriger Lösung spontan zu einer individuellen, festen dreidimensionalen Struktur. Diese Struktur wird als nativ bezeichnet. Die native Struktur stellt aus thermodynamischer Sicht ein

Struktur	Zeitskala (Untergrenze)	Referenz(en)
$\alpha$ -Helix	$10^{-7}$ s	[69, 70]
$\beta$ - <i>hairpin</i>	$10^{-6}$ s	[70, 71]
Kleinstproteine, Teildomänen	$10^{-5}$ s	[71–73]
komplexe Tertiärstruktur	$10^{-3}$ s	[68]

**Tabelle 1.1:** Zeitskalen für die Faltung einzelner Elemente der Proteinstruktur: Bei einem  $\beta$ -*hairpin* bilden zwei etwa gleich große Hälften eines Peptids ein zweisträngiges  $\beta$ -Faltblatt mit einer Kehre, die sich etwa in der Mitte der Peptidkette befindet. Die Faltung des Peptidrückgrates erinnert dabei an eine Haarnadel (engl. *hairpin*), was die Namensgebung für dieses einfachste  $\beta$ -Strukturmotiv erklärt.

wohldefiniertes Minimum der Gibbsschen freien Energie des Moleküls in seiner Umgebung dar [35, 59].

Neben den intramolekularen Wechselwirkungen des Proteins, etwa den im vorangegangenen Abschnitt angesprochenen Wasserstoffbrücken zwischen den Peptidgruppen des Rückgrats, beeinflusst insbesondere der Entropiebeitrag der Umgebung die freie Energie des nativen Zustands. Bei wasserlöslichen Proteinen mit definierter Struktur, sogenannten globulären Proteinen, beginnt die Faltung vermutlich mit einer Zusammenballung der unpolaren Teile der Kette in einem hydrophoben Kern. Dieser sogenannte hydrophobe Kollaps ist im wesentlichen durch ein starkes Anwachsen der Lösungsmittelentropie bei Minimierung der Kontaktflächen zwischen der polaren Wasserumgebung und dem unpolaren hydrophoben Kern getrieben. Der hydrophobe Kollaps wird begleitet von der Bildung von Sekundärstrukturelementen [37, 60–62]. Die entstandenen, noch vergleichsweise ungeordneten aber schon relativ kompakten Zustände reorganisieren sich sukzessive zur nativen Struktur, mit den entsprechenden tertiären Kontakten. Die Faltungsprozesse für Membranproteine sind wesentlich weniger weit erforscht. Klar ist, dass hier ein dem hydrophoben Kollaps entsprechender Prozess nicht vorkommt, da das Innere einer Zellmembran selbst eine sehr hydrophobe Umgebung darstellt.

Bei komplexeren Proteinen scheint der Prozess der Proteinfaltung oftmals über mehrere metastabile Zwischenzustände zu laufen [36, 37]. Diese Zwischenzustände liegen in einer trichterförmigen Energielandschaft, wobei die Faltung der Kette dabei relativ schnell auf den Zustand niedrigster freier Energie zugelenkt wird, das heißt den nativen Zustand. Es kann dabei jedoch auch Sackgassen und Fallen geben, die den Prozess behindern [36]. Viele Details dieses Prozesses sind jedoch noch unverstanden und werden seit Jahren intensiv theoretisch und experimentell untersucht [36, 37, 63–68]

Die Faltungsprozesse der einzelnen Strukturelemente eines Proteins laufen auf Zeitskalen ab, die je nach Komplexität variieren. In Tabelle 1.1 sind diese Zeitskalen mit entsprechenden Literaturziten in einer Übersicht zusammengefasst.

Elementare Prozesse sind relativ schnell: Die Bildung einer einzelnen  $\alpha$ -Helix kann auf der Zeitskala von  $10^{-7}$  s erfolgen [69, 70]; Faltungszeiten für eine einfache  $\beta$ -Faltblatt-Struktur sind mit mindestens  $10^{-6}$  s [70, 71] bereits um eine Größenordnung länger. Wiederum um eine Größenordnung darüber, bei  $10^{-5}$  s [71–73], sind die schnellsten Faltungszeiten für sehr kleine Proteine und Teildomänen größerer Proteine anzusiedeln. Dabei handelt es sich jedoch um Ausnahmen, da die Faltungszeit für die komplette native Tertiärstruktur der meisten Proteine eher im Zeitbereich oberhalb von  $10^{-3}$  s angesiedelt ist [68]. Insbesondere wenn kovalente Modifikationen des Proteins Teil des Faltungsprozesses sind, zum Beispiel die Ausbildung sogenannter Disulfidbrücken zwischen zwei Cystein-Seitengruppen, können auch durchaus Minuten oder gar Stunden für den Abschluss der Faltung vonnöten sein.

### 1.2.3 Partiiell gefaltete Zustände und Fehlfaltung

Keineswegs ist jedes Protein oder Peptid geneigt, in jeder Umgebung eine feste Faltungsstruktur anzunehmen. Wie eingangs bereits erwähnt wurde, hat sich in der jüngeren Vergangenheit herausgestellt, dass ein bedeutender Teil der Proteine nativ ungefaltet bleibt [44, 45]. Oft lassen diese Proteine sich jedoch durch Veränderung der Umgebungsbedingungen, beispielsweise durch ein Titrieren von Alkoholen zum ursprünglich wässrigen Lösungsmittel, in partiell oder komplett gefaltete Strukturen überführen [37, 74].

Umgekehrt lassen sich nativ gefaltete Moleküle durch eine Veränderung der physiko-chemischen Umgebungsbedingung zur Strukturänderung zwingen. Alle Proteine und Peptide verlieren bei steigenden Temperaturen ab einem bestimmten Punkt ihre native Faltung; man spricht von Hitzedenaturierung [39, 75]. Aber auch Kälte, veränderter pH-Wert, Druckerhöhung, bestimmte Lösungsmittel (beispielsweise Harnstoff) und weitere Umgebungsfaktoren können zur Denaturierung führen [76, 77]. Bei der Untersuchung der Faltungs- und Fehlfaltungseigenschaften von Proteinen und Peptiden im Reagenzglas (*in vitro*) ebenso wie auf dem Rechner (*in silico*) ist die Kontrolle dieser Faktoren daher von entscheidender Bedeutung. Kapitel 2 dieser Arbeit wird sich deshalb spezifisch mit der Frage der Temperaturkontrolle bei Simulationen von Makromolekülen in Lösung beschäftigen.

Partiell denaturierende Bedingungen führen bei einer Reihe von Proteinen *in vitro* zur Aggregation und zur Ausbildung amyloidartiger Fibrillen [27–33]. In ihrer nativen Faltungsstruktur sind Proteine relativ gut gegen Aggregationsprozesse geschützt, vermutlich da hydrophobe Bereiche der Kette, die potentiell die Zusammenlagerung mit anderen Proteinen befördern könnten, im Kern verpackt vorliegen und ebenso die Ränder nativer  $\beta$ -Faltblätter üblicherweise nicht exponiert liegen und also auch keinen potentiellen Anlagerungspunkt zur Verfügung

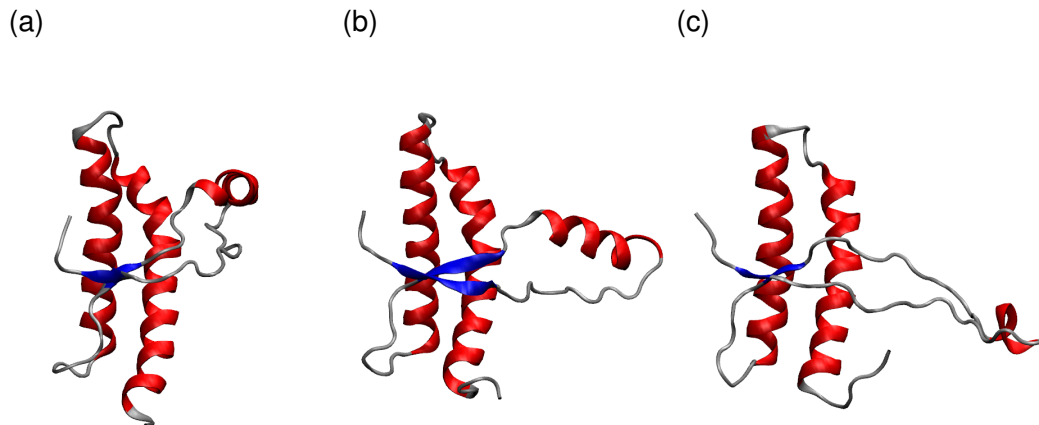
stellen [36]. Dagegen scheinen die teilweise gefalteten Zwischenzustände Ausgangspunkte für die mit der Faltung konkurrierende Aggregation zu bilden [37]. Teilgefaltete Zustände sind andererseits auch für eine Reihe nativ ungefalteter Moleküle Ausgangspunkt für die Bildung von Aggregaten [74].

Ein Beispiel bildet die lösungsmittelabhängige Faltung und *in-vitro*-Aggregation von  $\alpha$ -Synuclein [78]. Während hohe Konzentrationen fluoriertes Alkohole die für membrangebundenes  $\alpha$ -Synuclein typische,  $\alpha$ -helikale Konformation induzieren, erzeugen einfache Alkohole wie auch schwächer konzentrierte Fluoro-Alkohole einen partiell gefalteten Zwischenzustand, der den Ausgangspunkt für eine schnelle Aggregation zu typischen amyloiden Fasern bildet. Hierbei zeigte sich [78], dass insbesondere der Übergang in den aggregationsfähigen Zwischenzustand einfach mit der dielektrischen Konstante des Lösungsmittels korreliert ist. Diese Ergebnisse deuten darauf hin, dass die dielektrischen Eigenschaften der Umgebung eine wichtige Rolle bei der Faltung und Fehlfaltung von  $\alpha$ -Synuclein spielen.

### 1.2.4 Fehlfaltung des Prion-Proteins

Es ergeben sich hier nun interessante Parallelen zu Beobachtungen, die Hirschberger *et al.* [79] bei Molekulardynamiksimulationen von Mutanten des Prion-Proteins PrP<sup>C</sup> machten. Die Faltung des Proteins ist in Abb. 1.6 für drei Zeitpunkte aus einer dieser Simulationen gezeigt. Zu Beginn der Simulation (Abb. 1.6a) hat das Protein seine native Faltung. Nach einiger Zeit (Abb. 1.6b) entfernt sich die erste  $\alpha$ -Helix H1 vom hydrophoben Kern (H2 und H3) des Proteins (vgl. Abb. 1.5a). Nach dieser Trennung entfaltet H1 bis zum Ende der Simulationszeit von etwa 10 ns nahezu vollständig (Abb. 1.6c). Hirschberger *et al.* vermuteten, dass die Entfaltung der H1 ausgelöst wurde durch ihren Übergang von der schwach polaren Umgebung auf der Oberfläche des hydrophoben Kerns des Proteins in die hochpolare Umgebung des Volumenwassers. Ein solcher Prozess erscheint plausibel, da H1 wie jede  $\alpha$ -Helix durch die Wechselwirkung der Peptid-dipole (vgl. 1.4a) und darüberhinaus durch eine Reihe der bereits angesprochenen  $i, i + 4$ -Salzbrücken elektrostatisch stabilisiert ist [80]. Mit dem Übergang der Helix ins Volumenwasser ist eine Erhöhung der effektiven dielektrischen Konstante der Umgebung verbunden, was eine Schwächung dieser elektrostatischen Wechselwirkungen bedeutet.

Eine große Anzahl verschiedener Studien [81–85] deutet darauf hin, dass die von Hirschberger *et al.* [79] beobachtete Trennung einer Domäne, welche H1 und die beiden Stränge S1 und S2 des kurzen  $\beta$ -Faltblatts enthält (vgl. Abb. 1.5a), vom hydrophoben Kern des Proteins (H2 und H3) tatsächlich den ersten Schritt des Umfaltungsprozesses darstellt. Eine besondere Bedeutung kommt bei der Krankheitsentstehung möglicherweise der Oxidation des hydrophoben Aminosäurere-



**Abbildung 1.6:** Drei Schnappschüsse aus den Simulationen der Prion-Mutante M205R von Hirschberger *et al.* Die Simulation startet mit der in (a) gezeigten Struktur. Während der Simulation wurde ein Wegklappen der Helix H1 beobachtet (b), die anschließend fast vollständig entfaltete (c).

stes Methionin 205 (M205) zu [50], da diese Oxidation, ebenso wie die von Hirschberger *et al.* simulierte Punktmutation M205R, zu einer hydrophilen Störung der eigentlich hydrophoben Kontaktfläche zwischen den Domänen führt, was eine Erklärung für die Verbindung von oxidativem Stress und den TSE darstellen würde [50, 86–88]. Tatsächlich konnte kürzlich experimentell gezeigt werden, dass die Ersetzung aller Methionin-Reste im Prion-Protein durch isosterische und hydrophile Methoxinin-Reste ein Proteinmolekül mit hohem Anteil an  $\beta$ -Struktur und Aggregationsneigung entsteht [89].

Eine Bestätigung des von Hirschberger *et al.* skizzierten Mechanismus einer durch den Polaritätswechsel katalysierten Entfaltung von H1 kommt in diesem Zusammenhang besondere Bedeutung zu, da diese Helix vermutlich im Lauf der Umwandlung von PrP<sup>C</sup> zu PrP<sup>Sc</sup> entfalten muss [90, 91] und eine Rolle bei der Bindung von PrP<sup>C</sup> oder eines partiell entfalteten Zwischenzustandes an den katalytisch wirksamen PrP<sup>Sc</sup>-Komplex zu spielen scheint [92–97]. Im Zentrum der Diskussion steht dabei die Frage, ob die Entfaltung von H1 eine Barriere für den Umwandlungsprozess darstellt [83, 98]. Diese Frage wird in einer Simulationsstudie [99] behandelt, die Kapitel 4 dieser Arbeit bildet.

Die oben beschriebenen Umfaltungsprozesse werden üblicherweise experimentell mit Techniken beobachtet, die nur sehr grobe Strukturinformationen liefern, wie etwa der Zirkulardichroismus- [100] oder der Infrarot-Spektroskopie [101]. Eine Methode mit oftmals atomarer Genauigkeit bei der Strukturaufklärung stellt die Kernspinresonanz-Spektroskopie dar. Diese kann jedoch Faltungszustände mit Lebenszeiten unterhalb einer bestimmten Grenze nicht mehr detailliert auf-

lösen [101, 102], wie sie viele partiell gefaltete Zustände aufweisen. Eine theoretische Methode, die die Beobachtung solcher schnell wechselnder Zustände in atomarer Auflösung erlaubt, stellt die Molekulardynamiksimulation (MD-Simulation) dar. Ein Ziel meines Dissertationsprojektes bestand in der Anwendung von MD-Simulationen zur detaillierten Aufklärung der ersten Schritte von Proteinfaltungprozessen, die, wie oben skizziert, häufig von einer Veränderung der Umgebungsbedingungen induziert wird.

### 1.3 Techniken der Molekulardynamiksimulation

Die Wechselwirkungen und die Dynamik auf atomarer Ebene werden durch die Quantenmechanik beschrieben. Entsprechend müssen Prozesse, welche die Entstehung oder den Bruch einer kovalenten Bindung beinhalten, mit Hilfe quantenmechanischer Methoden beschrieben werden. Diese Methoden sind mit einem großen Rechenaufwand verbunden und daher auf Systemgrößen von höchstens etwa 100 Atomen beschränkt. Interessiert man sich aber in erster Linie für die langsamen, grobskalige Bewegungen von Makromolekülen in Lösung, die sog. Konformationsdynamik, so lässt sich ein solches System in guter Näherung im Rahmen der klassischen Mechanik beschreiben.

#### 1.3.1 Molekularmechanik

Bei einer molekularmechanischen (MM) MD-Simulation werden die  $N$  Atome eines Simulationssystems als klassische Punktteilchen mit den Koordinaten  $\{\mathbf{r}_i : i = 1, \dots, N\}$  behandelt, wobei man die einzelnen kartesischen Koordinaten  $r_i$  der Atome zu einem  $3N$ -dimensionalen Konfigurationsvektor  $\mathbf{R} \equiv (\mathbf{r}_1, \dots, \mathbf{r}_N)$  des Systems zusammenfassen kann. Die potentielle Energie des Systems wird durch eine analytische Energiefunktion  $E_{\text{MM}}(\mathbf{R})$  beschrieben, das sogenannte MM-Kraftfeld. Es existiert eine ganze Anzahl solcher MM-Kraftfelder; einen Überblick liefern Ponder und Case [103]. Die Kraft  $\mathbf{F}_i = -\nabla_i E_{\text{MM}}(\mathbf{R})$  auf Atom  $i$  ergibt sich als negativer Gradient bezüglich der Atomkoordinaten  $\mathbf{r}_i$  ebenfalls als analytische Funktion und ermöglicht die numerische Integration der Newtonschen Bewegungsgleichungen, beispielsweise mittels des Verlet-Algorithmus [104]. Um die schnellsten Schwingungsmoden des Systems ( $\sim 10$  fs) noch glatt abzutasten, werden dabei gewöhnlich Integrationsschritte von etwa 1 fs Dauer verwendet.

Üblicherweise setzt sich ein MM-Kraftfeld aus zwei Anteilen zusammen

$$E_{\text{MM}}(\mathbf{R}) = E_b(\mathbf{R}) + E_n(\mathbf{R}). \quad (1.1)$$

Der erste Anteil  $E_b(\mathbf{R})$  stellt eine klassische Näherung für die quantenmechanische Wechselwirkung kovalent verbundener Atome über einen Abstand von ma-

ximal drei kovalenten Bindungen dar und enthält Potentiale, die von internen Koordinaten des Moleküls wie Bindungslängen, Bindungswinkeln und Torsionswinkeln abhängt. Hierzu kommt bei der Verwendung des Kraftfeldes CHARMM22 [105] eventuell noch eine sogenannte CMAP-Korrektur [106, 107], d.h. eine Funktion, die von den beiden Torsionswinkeln  $\phi$  und  $\psi$  abhängt (vgl. Abb. 1.2 und Abb. 1.3) und einer verbesserten Beschreibung dieser beiden für die Faltung des Peptidrückgrats entscheidenden Freiheitsgrade dient [107].

Dazu kommen die Wechselwirkungen  $E_n(\mathbf{R})$ , die auch nicht chemisch gebundene Atome miteinander eingehen. Diese werden für alle Atompaare  $p = (i, j), j > i$  ausgewertet, zwischen welchen mindestens drei kovalente Bindungen liegen oder die überhaupt nicht Teil desselben Moleküls sind. Damit hat  $E_n(\mathbf{R})$  die Form

$$E_n(\mathbf{R}) = E_{\text{LJ}}(\mathbf{R}) + E_{\text{C}}(\mathbf{R}) = \sum_p 4\varepsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \sum_p \frac{q_i q_j}{r_{ij}}, \quad (1.2)$$

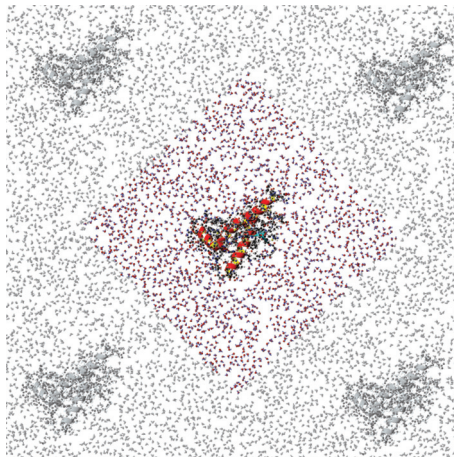
wobei  $r_{ij} = |\mathbf{r}_i - \mathbf{r}_j|$  den Abstand der Atome  $i$  und  $j$  angibt. In Gl. 1.2 beschreibt der erste Term  $E_{\text{LJ}}(\mathbf{R})$  die van-der-Waals-Wechselwirkung und die Pauli-Abstoßung zwischen den Atomen in der üblichen Form von Lennard-Jones-Potentialen. Das Lennard-Jones-Potential enthält als Parameter die Tiefe  $\varepsilon_{ij}$  des Potentialminimums und die Nullstelle des Potentials  $\sigma_{ij}$ . Der zweite Term  $E_{\text{C}}(\mathbf{R})$  in Gl. 1.2 beschreibt die elektrostatische Wechselwirkung der Atome  $i$  und  $j$  über ein Coulomb-Potential, in das die Partialladungen  $q_i$  und  $q_j$  der beiden Atome parametrisch eingehen.

Da die Anzahl der nichtgebundenen Paare  $p$  in einem Simulationssystem quadratisch mit der Anzahl  $N$  der Atome im Simulationssystem anwächst, stellt die Auswertung von  $E_n(\mathbf{R})$  für Simulationen von Makromolekülen mit expliziter Berücksichtigung der Lösungsumgebung einen enormen Rechenaufwand und gewöhnlich den entscheidenden Faktor für die erreichbaren Simulationszeiten dar. Das Lennard-Jones-Potential ist aufgrund seiner kurzreichweitigen Natur ( $\sim r_{ij}^{-6}$ ) für Atomabstände jenseits von etwa 10 Å für die Dynamik vernachlässigbar. Der energetische Effekt der vernachlässigten Wechselwirkungen kann im Rahmen einer Molekularfeldnäherung behandelt werden [108].

### 1.3.2 Elektrostatische Wechselwirkung

Die Coulomb-Wechselwirkung ( $\sim r_{ij}^{-1}$ ) ist dagegen langreichweitig und kann nicht ohne weiteres von einem bestimmten kleinen Abstand an vernachlässigt werden [109–111]. Es gibt zwei grundsätzlich unterschiedliche Ansätze dieses Problems Herr zu werden.

Die erste Lösungsmöglichkeit sind sogenannte Gittersummenmethoden. Diese nutzen aus, dass in MD-Simulationen von Molekülen in kondensierter Phase, wie



**Abbildung 1.7:** Beispiel für ein Simulationssystem mit einem Protein (hier PrP<sup>C</sup>), expliziter Wasserumgebung und periodischen Randbedingungen: Das zentrale Simulationsvolumen (in Farbe) ist von periodischen Kopien seiner selbst umgeben (grau). Die Abbildung ist der Dissertation von M. Stork [122] entnommen.

zum Beispiel eines Proteins in Lösung, üblicherweise periodische Randbedingungen angesetzt werden um Randeffekte zu vermeiden [111]. Ein Simulationssystem mit einigen tausend bis einigen zehntausend Atomen ist also von periodischen Kopien seiner selbst umgeben (siehe Abb. 1.7), und sobald ein Atom das zentrale Simulationsvolumen auf einer Seite verlässt tritt es auf der gegenüberliegenden Seite wieder ein. Für diesen Quasi-Kristall kann die Elektrostatik im Rahmen der numerischen Genauigkeit exakt ausgewertet werden, indem man sich der von P. P. Ewald [112] für die Berechnung der Elektrostatik von Kristallen entwickelten Summationstechnik oder optimierten, modernen Varianten wie der *particle-mesh*-Ewald-Technik (PME) [113] bedient. Der PME-Algorithmus skaliert  $N \log(N)$  mit der Systemgröße  $N$ . Diese Methode der Elektrostatikberechnung ist beispielsweise in der verbreiteten MD-Simulationssoftware GROMACS [114] implementiert. Mögliche Artefakte aufgrund der Näherung der Elektrostatik des realen, nicht-periodischen Systems durch jene des Quasi-Kristalls sind Gegenstand intensiver Diskussionen [115–121].

Als Alternative zu den Gittersummenmethoden bietet sich die Behandlung der Elektrostatik mittels schneller Multipolmethoden an, wie sie im von unserer Arbeitsgruppe entwickelten MD-Simulationsprogramm EGO-MMVI implementiert sind [110, 123–125]. Hierbei werden für die Berechnung der elektrostatischen Wechselwirkung über zunehmend größere Entfernungen die Punktladungen der Atome zu zunehmend größeren Ladungsverteilungen zusammengefasst, deren Potential durch ihre Multipolentwicklung genähert wird. Durch Taylorentwicklungen werden diese Potentiale an die Atomorte der einzelnen Wechselwirkungspartner extrapoliert, was einen linear mit der Systemgröße skalierenden Algorithmus ergibt. Dieser Algorithmus erlaubt es, mit vertretbarem Rechenaufwand alle



Wechselwirkungen innerhalb des sogenannten *minimum-image*-Abstandes [108] zu berechnen, der dem Radius der größten Kugel entspricht, die dem elementaren Simulationsvolumen (in Abb. 1.7 farbig) gerade noch einbeschrieben werden kann. Dadurch ist gewährleistet, dass jedes Atom im System mit maximal einem Bild jedes anderen Atoms wechselwirkt und Periodizitätsartefakte unterdrückt werden. Der Einfluss des Lösungsmittels jenseits des Abschneide-Radius wird in EGO-MMVI gewöhnlich durch eine Reaktionsfeldkorrektur berücksichtigt [110, 123, 126].<sup>3</sup>

Da die Faltungs-/Entfaltungs-Gleichgewichte einiger zur Aggregation neigender Peptide und Proteine teils sehr empfindlich auf Veränderungen der elektrostatischen Eigenschaften der Umgebung reagieren (vgl. Abschnitt 1.2.3), kommt einer genauen Beschreibung der Elektrostatik bei entsprechenden Simulationen eine große Bedeutung zu. Wie in Kapitel 4 anhand einer Simulationsstudie zur Stabilität eines PrP-H1-Peptids in Lösungsmitteln unterschiedlicher Polarität gezeigt wird, lässt sich diese Abhängigkeit bei entsprechend genauer Behandlung der Elektrostatik am Computermodell studieren. Weiterhin eignen sich entsprechend elektrostatisch empfindliche Modell-Peptide zur Untersuchung möglicher Artefakte im Zusammenhang mit den Gittersummenmethoden.

### 1.3.3 Temperaturkontrolle

Für die Besetzung der verschiedenen Konformationszustände eines Proteins oder Peptids ist die Temperatur von entscheidender Bedeutung, was sich an den oben bereits erwähnten Phänomenen der kalten und heißen Denaturierung von Proteinen [76] zeigt. Für  $\alpha$ -Synuclein wurde berichtet, dass Erwärmung *in vitro* partiell strukturinduzierend wirkt und zur Aggregation führt [127]. Für entsprechende Simulationen mit dem Ziel, Faltungs- und Entfaltungsprozesse für Makromoleküle zu beobachten und die entsprechenden Gleichgewichte abzutasten, ist es daher unbedingt nötig, die Temperatur des Moleküls genau kontrollieren zu können.

Jedes MD-Simulationsprogramm enthält eine Reihe von Näherungen, die dazu führen, dass die Energie im Simulationssystem nicht strikt erhalten ist, obwohl die zugrundeliegende Newtonsche Dynamik dies eigentlich gewährleisten sollte. In EGO handelt es sich dabei in erster Linie um die oben skizzierten Näherungen bei der Berechnung der Elektrostatik, insbesondere die Übergänge zwischen den einzelnen Ebenen der Multipolnäherung [110]. Dazu kommen Diskretisierungsfehler bei der Integration der Newton-Gleichungen [128]. Die Heizraten

---

<sup>3</sup>Die Implementierung der Elektrostatikberechnung in EGO wurde von mir dahingehend erweitert, dass nunmehr auch ungewöhnlich kleine Simulationssysteme (*minimum-image*-Radius kleiner als ca. 15 Å) mit periodischen Randbedingungen handhabbar sind.

in Peptid-Lösungsmittel-Simulationen können für die beiden Teile des Systems deutlich unterschiedlich sein.

Dieses algorithmische Heizen (und gelegentlich auch Kühlen) wird üblicherweise mit Hilfe sogenannter Thermostatenalgorithmen [129] kompensiert, welche allerdings sowohl die Dynamik als auch das statistische Ensemble des Makromoleküls mit unbekanntem Auswirkung modifizieren können. Kapitel 2 dieser Arbeit stellt ein Artikel [130] dar, der diese Störungen näher untersucht und Strategien zur ihrer Minimierung entwickelt. Besondere Bedeutung für meine Arbeit kommt diesen Untersuchungen auch deshalb zu, weil eine eventuelle Verzerrung des abgetasteten statistischen Ensembles durch den eingesetzten Thermostaten eine Verletzung der Grundannahmen der im folgenden noch zu behandelnden *replica-exchange*-Techniken darstellt.

### 1.3.4 Das Abtastproblem

MD-Simulationen haben einen Grundtakt: Die Dauer des Integrationsschritts bei der numerischen Integration der Newton-Gleichungen. In *all-atom*-Modellen ist dieser Grundtakt durch die schnellsten zu integrierenden Bewegungsmoden auf wenige Femtosekunden beschränkt. Möchte man daher beispielsweise Faltungsprozesse der Sekundärstruktur von Peptiden und Proteinen beobachten, wie sie für die Proteinfehlfaltung typisch sind, dann muss die Länge einer entsprechenden MD-Trajektorie jene Zeitskalen erreichen, die für solche Prozesse in Tab. 1.1 angegeben sind. Um die Faltung eines  $\alpha$ -helikalen Peptids zu beschreiben ( $\gtrsim 100$  ns), sind also mindestens etwa  $10^8$  Integrationsschritte nötig.

Die erste veröffentlichte MM-MD-Simulation eines Proteins [131] stammt aus dem Jahr 1977 und umfasste nur wenige tausend Integrationsschritte. Es konnte damit also nur die Dynamik des Proteins auf der Pikosekunden-Zeitskala beobachtet werden, zudem lediglich unter groben Näherungen der Wechselwirkungen und ohne die für die Proteinstruktur und -funktion so wichtige Lösungsumgebung. Die zunehmende Rechenleistung neuentwickelter Rechner erlaubte es van Gunsteren und Berendsen bereits im Jahr 1984 bei ähnlicher Simulationszeit erstmals auch einen kleinen Teil der Lösungsumgebung in die Simulation miteinzubeziehen [132].

Moderne Nachfahren dieser ersten biomolekularen Simulationen erreichen bei expliziter Berücksichtigung der Lösungsumgebung eine Dauer von bis zu  $10\ \mu\text{s}$  [133] und damit Zeitskalen, auf welchen sich tatsächlich sehr schnelle Proteinfaltungsprozesse abspielen (vgl. Tab. 1.1). Jedoch handelt es sich bei Trajektorien dieser Länge um „Einzelstücke“, die trotz ihrer Länge oftmals über den Großteil der Simulationszeit in einer bestimmten Konformation gefangen sind. Dies ist besonders dann von Nachteil, wenn die untersuchten Moleküle im thermodynamischen Gleichgewicht eine Vielzahl möglicher Konformationen anneh-

men können, wie es gerade für jene Peptide und Proteine typisch ist, die zur Fehlfaltung neigen (siehe Abschnitt 1.2.3 und [74]). Um eine ausreichende Statistik für die Besetzung der einzelnen Konformationen dieser Moleküle zu erhalten, ist daher stets eine ganze Anzahl von ausreichend langen MD-Simulationen mit unterschiedlichen Startbedingungen nötig.

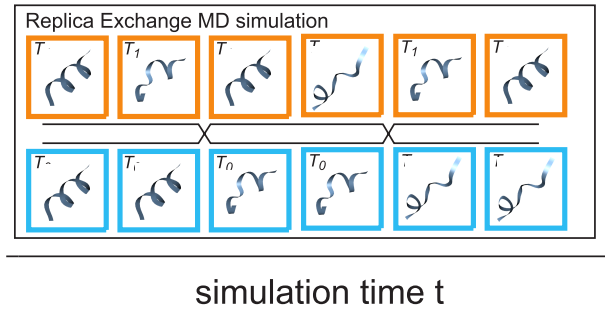
Der Grund für die seltenen Übergänge zwischen den einzelnen Konformationen liegt in oftmals hohen Energiebarrieren, welche die Konformationen voneinander trennen. Wenn diese Energiebarrieren enthalpischer Natur sind, dann erfolgt der Übergang über die Barriere bei erhöhter Temperatur deutlich schneller. Eine solche im wesentlichen enthalpische Barriere könnte beispielsweise die Entfaltung von  $\alpha$ -Helix 1 im Prion-Protein behindern, die sowohl durch die elektrostatischen Wechselwirkungen der Dipole des Rückgrates (vgl. 1.4a) wie auch durch zusätzliche elektrostatische Wechselwirkungen zwischen ionischen Seitengruppen (vgl. Abschnitt 1.2.4) stabilisiert wird.

Die statistischen Ensembles, die bei hohen Temperaturen erzeugt werden, sind jedoch nicht ohne weiteres auf niedrige Temperaturen übertragbar und daher schwer interpretierbar. In der Vergangenheit wurden deshalb eine ganze Reihe von Techniken zum beschleunigten Abtasten von Gleichgewichtsensembles (*enhanced-sampling*-Techniken, z.B. [134–140]) entwickelt, die meist erhöhte Temperaturen nutzen um die Konformationsübergänge zu beschleunigen, gleichzeitig aber das korrekte statistische Ensemble erzeugen.

### 1.3.5 Replica-Exchange-Techniken

Besonders verbreitet sind sogenannte *replica-exchange*-Techniken (RE), welche die parallele Simulation des gleichen Simulationssystem mit einer Nutzung der beschleunigten Konformationsdynamik bei hohen Temperaturen verbinden. Im einfachsten Fall der Temperatur-RE-Methode (TREM) [141–143] werden  $M$  Kopien oder Replikat des Simulationssystem bei  $M$  verschiedenen Temperaturen  $T_i, i = 1, \dots, M$  einer gewöhnlichen MD-Simulation unterworfen. In regelmäßigen Abständen wird versucht, die bei benachbarten Temperaturen erzeugten Konfigurationen auszutauschen. Das Prinzip ist in Abb. 1.8 für zwei Temperaturstufen  $T_1$  und  $T_2$  und ein helikales Peptid skizziert.

Wurde während eines MD-Intervalls bei  $T_2$  eine Barriere übersprungen und eine neue Peptid-Konformation erzeugt (in Abb. 1.8 z.B. im zweiten Intervall), so kann diese beim folgenden Austausch zur niedrigeren Temperatur  $T_1$  hin austauschen. Damit wird die neue Konformation in der bei  $T_1$  laufenden Teilsimulation sichtbar und trägt zur Stichprobe der Peptidkonformationen bei  $T_1$  bei. Von diesem Austausch an wird aber auch die noch nicht veränderte Konformation bei  $T_2$  weitersimuliert, was nun auch für diese die Wahrscheinlichkeit erhöht, eine neue Peptidkonformation zu erreichen. Auch diese neue Konfiguration kann dann in



**Abbildung 1.8:** Das Prinzip der *replica-exchange*-Methode, hier demonstriert an einem Beispiel mit zwei Temperaturstufen  $T_1$  (kalt und daher blau) und  $T_2$  (heiß und daher orange): Beide Simulationssysteme starten mit einer  $\alpha$ -helikalen Struktur. Jedes der Kästchen steht für ein MD-Intervall; jeweils zwischen den Intervallen wird ein Austausch versucht. Erfolgreiche Austauschversuche werden durch ein Kreuzen der Führungslinien zwischen den Kästchen angezeigt. Im zweiten Intervall findet bei der hohen Temperatur  $T_2$  eine Umfaltung in eine neue Peptidstruktur statt, die beim drauffolgenden Austausch zur niedrigen Temperatur  $T_1$  getauscht wird. Die nach  $T_2$  getauschte (ursprünglich kalte)  $\alpha$ -helikale Struktur entfaltet bald und generiert eine weitere Peptidkonformation, welche im folgenden nach unten propagiert wird. Die beschleunigte Konformationsdynamik bei  $T_2$  wirkt sich daher auf beide Trajektorien aus (vgl. die Führungslinien).

einem weiteren Austausch nach unten propagiert werden. Auf diese Art lässt sich die schnellere Konformationsdynamik bei  $T_2$  auch für die Stichprobe bei  $T_1$  nutzen.

Um sicherzustellen, dass das statistische Ensemble bei der Temperatur  $T_i$  nicht durch den Austausch von Strukturen mit dem fremden Ensemble bei  $T_j$  verzerrt werden, müssen diese Austausche einem Metropolis-Kriterium [144] gehorchen, welches garantiert, dass zwischen den verschiedenen Zuständen des generalisierten Ensembles, das sich aus den Einzelensembles bei den verschiedenen Temperaturen der RE-Leiter zusammensetzt, eine detaillierte Bilanz besteht [143]. Für TREM und zwei Temperaturstufen  $T_i$  und  $T_j$  wird ein Austausch unter diesem Kriterium [143] mit der Wahrscheinlichkeit

$$p_{ij} = \min \left[ 1, \exp(\Delta_{ij}) \right], \quad (1.3)$$

durchgeführt, wobei

$$\Delta_{ij} = (\beta_j - \beta_i) \left[ E(\mathbf{R}_j) - E(\mathbf{R}_i) \right]. \quad (1.4)$$

In Gl. 1.4 stehen  $\beta_j = 1/k_B T_j$  und  $\beta_i = 1/k_B T_i$  für die jeweiligen inversen Temperaturen ( $k_B$  ist die Boltzmann-Konstante);  $E(\mathbf{R}_j)$  steht für die potentielle Energie der Konfiguration  $\mathbf{R}_j$  des Simulationssystems auf der Temperaturstufe  $T_j$ , wie sie sich aus dem MM-Kraftfeld Gl. 1.1 ergibt, und  $E(\mathbf{R}_i)$  entsprechend für jene der Konfiguration  $\mathbf{R}_i$  auf der Temperaturstufe  $T_i$ .

Die mittlere Austauschwahrscheinlichkeit unter dem Austauschkriterium nach den Gleichungen 1.3 und 1.4 hängt davon ab, welchen Überlapp die Verteilun-

gen der potentiellen Energien des Simulationssystems bei  $T_i$  und  $T_j$  aufweisen [145]. Dieser Überlapp wiederum hängt von der Anzahl der Freiheitsgrade im Simulationssystem und vom Temperaturabstand  $T_j - T_i$  ab. Für sehr große Temperaturabstände oder Systeme mit sehr vielen Freiheitsgraden wird die mittlere Austauschwahrscheinlichkeit beliebig klein, was den Nutzen von TREM zunichte macht. RE-Strategien, die diesen Nutzen maximieren, zeichnen sich durch eine besonders schnelles Wandern der Replikate auf der Temperaturleiter aus [146]. Auf die Geschwindigkeit dieses Wanderns haben insbesondere die Konstruktion der Temperaturleiter [145, 147–149] und das Schema Einfluss, nach welchem die Paare von Replikaten ausgesucht werden, für die ein Austauschversuch gemacht wird [150, 151]. Diesen beiden Fragen widmen sich zwei Artikel [152, 153], welche die Abschnitte 3.1 und 3.2 dieser Arbeit bilden.

Die klassische TREM stößt wegen der vielen Freiheitsgrade des Simulationssystems bei Simulationen unter expliziter Berücksichtigung der Lösungsumgebung sehr schnell an Grenzen. Andererseits soll durch die Temperaturerhöhung eigentlich die Konformationsdynamik des Peptids beschleunigt werden und es würde daher genügen, allein die Temperatur des Peptids zu erhöhen [154]. Die Simulation eines heißen Peptids in einem kalten Lösungsmittel führt allerdings nicht zu einem wohldefinierten statistischen Ensemble und verletzt fundamentale Annahmen, die bei der Formulierung der TREM gemacht werden. Es existiert jedoch eine im Sinn der statistischen Physik strikte Umsetzung dieser Idee: Heißes Peptid in kaltem Lösungsmittel.

Dieses *replica exchange with solute tempering* (REST) genannte Verfahren [155] fußt auf der folgenden einfachen Überlegung: Im kanonischen Ensemble bei der Temperatur  $T_1$  kommt jeder Konfiguration  $\mathbf{R}$  mit einer potentiellen Energie  $E(\mathbf{R})$  (vgl. Gl. 1.2) ein Boltzmann-Gewicht  $\exp[-\beta_1 E(\mathbf{R})]$  zu. Erhöht man die Temperatur auf  $T_i$ , skaliert aber gleichzeitig die Funktion der potentiellen Energie  $E(\mathbf{R})$  mit dem Faktor  $\beta_1/\beta_i$ , so ergeben sich für sämtliche Konfigurationen  $\mathbf{R}$  wieder die gleichen Gewichte und also die gleiche Statistik wie bei der Temperatur  $T_1$ . Das System ist „statistisch kalt“ bei  $T_1$ , aber seiner mittleren kinetischen Energie nach trotzdem bei  $T_i$ . Daher findet kein Netto-Energiefluss statt, falls man ein solches System (in REST das Lösungsmittel) mit einem anderen System der Temperatur  $T_i$  (in REST das gelöste Molekül) koppelt – eine statische Nicht-Gleichgewichtssituation wird so vermieden.

Zur Umsetzung der Idee wird in REST das MM-Kraftfeld Gl. 1.1 in drei Teile geteilt; nämlich erstens in einen Anteil  $E^{\text{PP}}(\mathbf{R})$ , der die internen Wechselwirkungen des Moleküls (z.B. ein Peptid) beschreibt, dessen Konformationsdynamik beschleunigt werden soll, einen zweiten Anteil  $E^{\text{SP}}(\mathbf{R})$ , der die Wechselwirkungen des Moleküls mit der Lösungsumgebung beschreibt, und schließlich einen dritten Anteil  $E^{\text{SS}}(\mathbf{R})$ , der die Lösungsmittel-Lösungsmittel-Wechselwirkung beschreibt. Die Anteile  $E^{\text{SP}}$  und  $E^{\text{SS}}$  werden nun temperaturabhängig skaliert [99,

156], das heißt bei jeder Temperatur  $T_i$  wird ein anderes MM-Kraftfeld

$$E_i(\mathbf{R}) = E^{\text{PP}}(\mathbf{R}) + \sqrt{\beta_1/\beta_i} E^{\text{SP}}(\mathbf{R}) + (\beta_1/\beta_i) E^{\text{SS}}(\mathbf{R}) \quad (1.5)$$

verwendet, wobei  $\beta_1 = 1/k_B T_1$  die inverse Temperatur der niedrigsten Sprosse der Temperaturleiter darstellt. Das bedeutet insbesondere, dass nur diese Temperaturstufe mit einem unveränderten Kraftfeld simuliert wird und daher auch nur diese Temperaturstufe direkte physikalische Relevanz hat. Hier ist anzumerken, dass die von uns verwendete Form der Skalierung in Gl. 1.5 bezüglich des Skalierungsfaktors von  $E^{\text{SP}}$  von jener Form abweicht, wie sie von Liu *et al.* [155] für REST ursprünglich verwendet wurde. Da die Wahl des Skalierungsfaktors aber prinzipiell willkürlich ist, entschieden wir uns aus Gründen der leichteren Implementierung für die von Gl. 1.5 angegebene Form.

Ein entsprechend modifiziertes Metropolis-Austauschkriterium [155] garantiert wie bei TREM durch eine detaillierte Bilanz für das generalisierte REST-Ensemble, dass die Teilensembles bei den jeweiligen Temperaturen vom Austausch nicht berührt werden. Die Form der Austauschwahrscheinlichkeit Gl. 1.3 bleibt erhalten, allerdings lautet für REST der Exponent

$$\Delta_{ij} = (\beta_j - \beta_i) [E^{\text{PP}}(\mathbf{R}_j) - E^{\text{PP}}(\mathbf{R}_i)] + \left( \sqrt{\beta_1 \beta_j} - \sqrt{\beta_1 \beta_i} \right) [E^{\text{SP}}(\mathbf{R}_j) - E^{\text{SP}}(\mathbf{R}_i)] \quad (1.6)$$

etwas anders. Wichtig ist, dass in diesem Exponenten die in Gl. 1.5 voll skalierten internen Wechselwirkungen  $E^{\text{SS}}$  des Lösungsmittelsystems nun gar nicht mehr vorkommen. Für den Austausch ist also nunmehr ein System mit effektiv weniger Freiheitsgraden zu berücksichtigen, wodurch größere Abstände der einzelnen Temperaturstufen in der REST-Leiter möglich werden. Die skizzierte Idee des *solute tempering* wurde von uns kürzlich auf eine ähnliche *enhanced-sampling*-Technik übertragen, die im Vergleich mit RE noch größere Temperaturabstände zulässt. Ein entsprechendes Manuskript wurde kürzlich von uns zur Veröffentlichung eingereicht [156].

REST wie auch TREM wurden von R. Denschlag und mir für EGO in Form eines externen Shell-Skriptes implementiert. Diese Lösung bot sich an, da sich die notwendige Skalierung des MM-Kraftfelds nach Gl. 1.5 für die üblicherweise nur über Coulomb- und Lennard-Jones-Potentiale wechselwirkenden Lösungsmittelmoleküle [157–162] sehr einfach durch die Skalierung des Lennard-Jones-Parameters  $\epsilon_{ij}$  und der Partialladungen  $q_i$  der einzelnen Atome bewerkstelligen ließ (vgl. Gl. 1.2).

REST wurde in unserer Arbeitsgruppe erstmals von C. Lutz angewendet, um den in Abschnitt 1.2.4 skizzierten Zusammenhang zwischen der Oxidation des Aminosäure M205 im Prion-Protein PrP und den ersten Schritten zur Fehlfaltung des Proteins zu untersuchen [50]. Hier zeigte sich, dass bei großen Simulationssystemen der Nutzen von REST dadurch beschränkt ist, dass einige Replikate wegen

der kurzen erreichbaren Simulationszeiten kaum zur Statistik auf der niedrigsten Temperaturstufe (300 K) beitragen.

In der Folge wurde REST daher zur Untersuchung der Faltungs-/Entfaltungsgleichgewichte kleinerer Peptide angewandt, für die wesentlich längere Simulationszeiten zu erreichen sind. Anhand eines einfachen Monte-Carlo-Modells konnten wir zeigen, dass es für ein konvergentes Abtasten dieser Gleichgewichte nötig ist, eine ganze Reihe von Faltungs- *und* Entfaltungsereignissen während der Simulation zu beobachten [163]. Insbesondere zeigte sich auch, dass der Erfolg von RE-Techniken entscheidend davon abhängt, dass die kombinierte Zeit für die Faltung und die Entfaltung des Peptids durch eine Temperaturerhöhung beschleunigt wird.

R. Reichold konnte kürzlich unter Verwendung von REST erstmals das Faltungs-/Entfaltungsgleichgewicht eines  $\alpha$ -helikalen Peptids konvergent abtasten [107, 164]. Um diese Konvergenz zu demonstrieren, wurden zwei REST-Simulationen mit jeweils 10 Replikaten durchgeführt, die von vollständig gefalteten bzw. vollständig ungefalteten Startstrukturen ausgingen. Für das untersuchte Peptid war der Einfluss der extremalen Startstrukturen auf den Anteil helikaler Strukturen im Ensemble nach einer Simulationszeit von etwa 10 ns nicht mehr nachweisbar. Einen vergleichbaren Ansatz habe ich in der in Kapitel 4 vorgestellten Simulationsstudie verwendet.

## 1.4 Aufbau und Gliederung der Arbeit

Mein Dissertationsprojekt, welches zunächst in ein interdisziplinäres Projekt des Bayerischen Forschungsverbundes Prionen, später in den Sonderforschungsbereich 749 zur Erforschung der Dynamik und Intermediate molekularer Transformationen eingebunden war, hatte zum erklärten Ziel, Methoden zu entwickeln, um mittels MD-Simulationen Einblick in Proteinefehlfaltungsprozesse zu gewinnen. Da, wie in Abschnitt 1.2.3 diskutiert wurde, die Umgebungsbedingungen und hier insbesondere die Temperatur entscheidenden Einfluss auf die Besetzung der mit der Fehlfaltung assoziierten teilgefalteten Zustände haben, befasste ich mich vorderhand mit Strategien für eine optimale Temperaturkontrolle bei MD-Simulationen von Makromolekülen in Lösung. Hierzu führte ich zunächst eine beträchtliche Zahl relativ kurzer Simulationen an einem kleinen Modellpeptid durch, um den Einfluss gängiger Vorgehensweisen zur Temperaturkontrolle auf die Dynamik und das durch die Simulation abgetastete statistische Ensemble zu messen. Zusammen mit weiteren Simulationen, die von meinen Kollegen Robert Denschlag und Rudolf Reichold durchgeführt wurden, ließ sich daraus eine allgemeine Strategie zur Temperaturkontrolle ableiten, die die Konformationsdynamik des Makromoleküls, anders als übliche Methoden, nicht behindert und insbesondere auch für die spätere Anwendung in Verbindung mit REST geeignet ist.

Die Resultate dieser Arbeiten sind in einem Artikel [130] zusammengefasst, der Kapitel 2 dieser Dissertation bildet.

Da das Abtasten der vielen Konformationszustände von Peptiden und Proteinen wegen ihrer langsamen Konformationsdynamik mit heutigen MD-Methoden eine große Herausforderung darstellt und den Einsatz von verbesserten Abtastmethoden verlangt, war der nächste Schritt auf meinem Weg die Implementierung von REST für die Verwendung in unserer Arbeitsgruppe gemeinsam mit Robert Denschlag. Wie die in der Diplomarbeit [50] von Christine Lutz vorgestellten REST-Simulationen zeigen, ist ein erfolgreicher Einsatz von REST oder auch TREM stark von den Details des jeweils studierten Systems [163], insbesondere aber auch von Einzelheiten der gewählten Strategie abhängig. Aus systematischen Untersuchungen zur Verbesserung der mit RE-Techniken erreichbaren Ergebnisse ergaben sich mit Hilfe von Monte-Carlo-Modellen und mathematischen Analysen optimierte *replica-exchange*-Strategien, die in den beiden Artikeln [152, 153] vorgestellt werden, welche die Abschnitte 3.1 und 3.2 bilden.

Abschnitt 3.1 widmet sich dabei der Konstruktion einer optimalen RE-Temperaturleiter anhand eines einfachen Monte-Carlo-Modells, das die Zufallsbewegung eines Replikats auf der Leiter abbildet. Insbesondere zeigte sich anhand dieses Modells, dass frühere, aus theoretischen Überlegungen abgeleitete Regeln zur Konstruktion der Leiter [145, 147–149] für übliche RE-Strategien unerwarteterweise suboptimal sind. Anhand plausibler Überlegungen und mit Hilfe des Monte-Carlo-Modells kann empirisch eine tatsächlich optimale Regel angegeben werden. Der zweite Artikel [153] beschäftigt sich mit der Wahl eines optimalen Austauschschemas. Monte-Carlo-Simulationen zeigen, dass alternierende Austauschversuche nach oben und unten zu einer maximalen Wanderungsgeschwindigkeit der Replikate führen. Eine mathematische Analyse der Zufallsbewegungen unter Benutzung unterschiedlicher Austauschschemata erklärt diese Beobachtung ebenso wie die in Abschnitt 3.1 festgestellte Abweichung des Verhaltens realer RE-Systeme vom bisher angenommenen [145, 147–149].

Die Erfahrungen mit REST-Simulationen des Prion-Proteins [50] wie auch weitere Untersuchungen an einfachen Modellsystemen [163] zeigten, dass ein Abtasten der wichtigsten Konformationszustände ganzer Proteine wie etwa von  $\alpha$ -Synuclein auch mit REST nicht erreichbar sein würde. Vielversprechend erschien dagegen die Untersuchung der scheinbar ähnlich umgebungsabhängigen Helix 1 (H1) des Prion-Proteins mit Hilfe der neu implementierten und optimierten Methoden. Ich führte daher eine Reihe von REST-Simulationen an einem H1-Peptid in Lösungsmitteln unterschiedlicher Polarität durch, um die Frage zu klären, ob die Faltung der entsprechenden Sequenz tatsächlich durch die Polarität gesteuert wird und ob insbesondere die Entfaltung der Helix durch eine stark polare Umgebung erleichtert wird [79]. Indem ich für jedes Lösungsmittel jeweils zwei Simulationen mit vollständig helikalen bzw. vollständig entfalteten Startstrukt-



ren durchführte, erzielte ich Ober- und Untergrenzen für den jeweiligen Anteil helikaler Struktur im Gleichgewicht. Trotz einer teils deutlich eingeschränkten Konvergenz ließ sich durch Messung der Entfaltungsgeschwindigkeit der Helix in den unterschiedlichen Lösungsmitteln der vermutete, die Entfaltung von H1 katalysierende Einfluss hoch polarer Umgebungen nachweisen.



## 2 Minimalinvasive Temperaturkontrolle eines Makromoleküls in Lösung

Comme les lois inutiles  
affaiblissent les lois nécessaires  
[...]

---

*Charles-Louis de Secondat,  
Baron de Montesquieu [165]*

Wir wenden uns zunächst dem Problem der Temperaturkontrolle bei MD-Simulationen eines Makromoleküls in Lösung zu. Es handelt sich bei den folgenden Ausführungen um den Abdruck<sup>1</sup> des Artikels

Martin Lingenheil, Robert Denschlag, Rudolf Reichold, and Paul Tavan:  
The „Hot-Solvent/Cold-Solute“ Problem Revisited,  
*J. Chem. Theory Comput.* **4**, 1293-1306,

der von mir zusammen mit R. Denschlag, R. Reichold und P. Tavan verfasst wurde. Es wird gezeigt, dass die Kopplung eines inhomogenen Simulationssystems an einen globalen Thermostaten zu einem stationären Unterschied der Temperaturen von Makromolekül und Lösungsmittel, das heißt zu einem Nichtgleichgewichtszustand, führen kann. Nach einer Problemanalyse werden Strategien zur Beseitigung des Problems diskutiert, welche die Dynamik und Ensembleeigenschaften des Moleküls möglichst wenig stören.

---

<sup>1</sup>Reproduced with permission from The Journal of Chemical Theory and Computation, 4, 1293-1306, 2008. Copyright 2008 American Chemical Society.

## The “Hot-Solvent/Cold-Solute” Problem Revisited

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

*Lehrstuhl für Biomolekulare Optik, Ludwig-Maximilians-Universität München,  
Oettingenstrasse 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,<sup>1</sup> including processes of folding and refolding.<sup>2</sup> 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,<sup>3–5</sup> by exerting internal mechanical strain,<sup>6,7</sup> by introducing point mutations into the protein sequence,<sup>8,9</sup> or simply by elevating the temperature.<sup>9,10</sup>

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.<sup>11</sup> 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<sup>12</sup> or the viscosity.<sup>13</sup> 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.<sup>14</sup> 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,<sup>15</sup> the proton exchange and spin relaxation rates in nuclear magnetic resonance spectroscopy [see ref 16 and references therein], and the fluorescence depolarization rates<sup>17</sup> as well as the thermodynamical measures of protein stability.<sup>18</sup>

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

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<sup>19–21</sup> 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.<sup>14</sup> 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,<sup>22</sup> 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<sup>23</sup> 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 additionally the solute–solvent interaction energy is negligible 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<sup>24</sup> for studies of protein folding kinetics and is occasionally applied<sup>25,26</sup> 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<sup>27,28</sup> or by integrating the equations of motion with multiple-time-step algorithms.<sup>29</sup> Cooling may occur, for instance, if constraining bond lengths or angles with a too loose tolerance<sup>30</sup> or if neighbor lists for the calculation of the van der Waals interactions are not updated frequently enough.<sup>31</sup> 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,<sup>32</sup> a problem sometimes referred to as the “hot-solvent/cold-solute” problem.<sup>33</sup> 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.<sup>32,34,35</sup>

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.<sup>36</sup> 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:

- 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)?
- Ergodicity: Does the method generate the expected ensemble within the time typically covered by modern MD simulations?
- 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 thermostating 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<sup>37</sup> (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,<sup>40</sup> 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”.<sup>38,39</sup> 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.<sup>38</sup>

The more strongly coupling Nosé-Hoover thermostat<sup>41,42</sup> (NHT) is theoretically expected to generate the canonical distribution of microstates if certain conditions are obeyed thus conforming with the above question a).<sup>43</sup> 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.<sup>44</sup> Several studies<sup>42,44–47</sup> 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,<sup>48</sup> which has been successfully tested by Cheng and Merz<sup>33</sup> 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.<sup>49</sup> 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

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

Here, the acceleration  $\ddot{\mathbf{r}}_i(t)$  of atom  $i$  is caused not only by the forces  $\mathbf{F}_{i, \text{ff}}(t)$  derived from an MM force-field but also by a second term  $\mathbf{F}_{i, \text{therm}}(t) \equiv -m_i \gamma(t) \dot{\mathbf{r}}_i(t)$ , 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<sup>50</sup>  $T(t)$  by

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

with  $\tau$  denoting the coupling time and  $T_0$  the target temperature. For the Nosé-Hoover<sup>41,42</sup> thermostat,  $\gamma(t)$  is coupled on a time scale  $\tau_{\text{NHT}}$  to  $T(t)$  by the differential equation

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

**Perturbation of the Dynamics.** Every thermostat which is described by eq 1 perturbs the Newtonian dynamics generated by the forces  $\mathbf{F}_{i, \text{ff}}(t)$  through the admixture of additional thermostatic forces  $\mathbf{F}_{i, \text{therm}}(t)$ . 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

$$\xi_i^2 \equiv \langle \mathbf{F}_{i, \text{therm}}^2 \rangle_D / \langle \mathbf{F}_{i, \text{ff}}^2 \rangle_D \quad (4)$$

where the brackets  $\langle \dots \rangle_D$  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  $\xi_i$  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  $C_\alpha$ -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  $\xi_i^2$  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  $\langle \mathbf{F}_{i, \text{therm}}^2 \rangle_D$  required for the evaluation of the  $\xi_i^2$  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  $\mathbf{T}(t)$ . Hence, the correlation of  $\gamma^2$  and  $\dot{\mathbf{r}}_i^2$  vanishes and one obtains

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

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  $\hat{T} = \langle \mathbf{T} \rangle_D$  is the average temperature determined from the simulation. Equation 5 then becomes

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

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  $\rho(\mathbf{T})$  for the instantaneous kinetic temperature is a  $\chi^2$ -distribution<sup>51</sup>

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

where  $T = T_b$  is the expectation value of  $\mathbf{T}$ ,  $N_{\text{DoF}}$  is the number of degrees of freedom (DoF) of the system, and  $\Gamma(\dots)$  denotes the Euler  $\Gamma$ -function. Consequently, the variance  $\sigma_T^2$  of the temperature fluctuations is

$$\sigma_T^2 = \frac{2T^2}{N_{\text{DoF}}} \quad (8)$$

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  $N_{\text{DoF}} \rightarrow \infty$ , eq 7 becomes a normal distribution.

The size of  $\sigma_T^2$  together with the autocorrelation time<sup>50</sup>  $\tau_c$  of the temperature fluctuations critically influences the accuracy with which the equilibrium temperature  $T$  is determined by a given simulation. The variance  $\sigma_T^2$  of the

time averages  $\hat{T}$  obtained from a set of equilibrium simulations with durations  $D$  can be estimated<sup>50</sup> to be

$$\sigma_{\hat{T}}^2 = 2\sigma_T^2 \frac{\tau_c}{D} \quad (9)$$

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  $\sigma_{\hat{T}}^2$ ) with a correlation time  $\tau_c$  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  $T(t) = \langle \mathbf{T}(t) \rangle_{\text{ens}}$ . The ensemble average power exerted by the thermostat on a given atom  $i$  is

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

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

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

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  $\mathbf{T}(t)$  of the system, one obtains

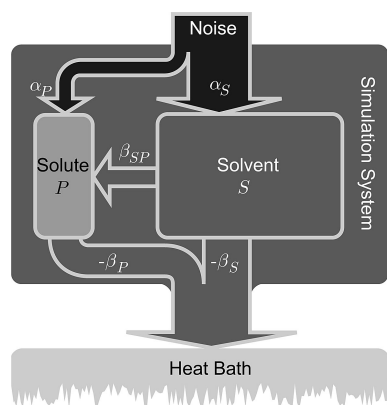
$$\beta_i(t) = \frac{3k_B T_i(t)}{2\tau} [T_0/T(t) - 1] \quad (12)$$

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  $\langle \dots \rangle_D$ . This allows to calculate for every subsystem  $\kappa$  from a simulation the (time) average thermostatic power

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

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

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.<sup>33</sup>



**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  $\alpha_\kappa$  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  $\kappa$ , 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  $\beta_\kappa(t)$  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

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

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

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

between  $S$  and  $P$ , which we assume to depend linearly on the temperature difference. Here, the time constant  $\tau_{SP}$  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.

$$\alpha_P + \beta_{SP} + \beta_P = 0 \quad (16)$$

$$\alpha_S - \beta_{SP} + \beta_S = 0 \quad (17)$$

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 thermostate a system exhibiting inhomogeneities with respect to the rates  $\alpha_\kappa$  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.<sup>36,47,52–57</sup>

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  $\tau_s$  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

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

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

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

where  $T_{0,P}$  denotes the target temperature, and  $\tau_P$  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  $\beta_P$  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.<sup>36,52–56,58</sup> For such small  $\tau_P$ , eq 19 is completely dominated by the thermostatic term. The reason for this dominance is that  $\tau_P$  is by at least 1 order of magnitude smaller than the solute–solvent coupling time  $\tau_{SP}$ , 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  $\alpha_P$ , this deviation is expected to be small because of the short time scale  $\tau_P$ .

**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  $\alpha_P$  is negligible on the time scale  $\tau_{SP}$  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  $\beta_P$  cancels the power  $\alpha_P$  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  $\tau_P$  by a proper choice of the target temperature  $T_{0,P}$ . In the limit  $\tau_P \rightarrow \infty$ , the thermostat variable  $\gamma$  in eqs 1 and 2 becomes a constant  $\gamma_P$ , and the thermostat scheme may actually be described by this single parameter. At large  $\tau_P$ , 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  $\alpha_P$ , “minimally invasive”.

To set up a CHF simulation as required in strategy **P.3**, the a priori unknown power  $\alpha_P$  of algorithmic heating has to be determined in order to specify the constant thermostat parameter  $\gamma_P$ , or, equivalently, the parameters  $T_{0,P}$  and  $\tau_P$  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  $\beta_P$  of the thermostat. The two heat balance eqs (19) of these tests then constitute a system of linear equations which determines the unknown parameters  $\beta_{SP}$  and  $\alpha_P$ . 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<sup>27</sup> and GROMACS<sup>59</sup> 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<sup>60,61</sup> with a moving-boundary reaction-field approach<sup>62</sup> and a multiple-time-step integration.<sup>29,63</sup> In the GROMACS simulations we used the PME method<sup>28</sup> 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<sup>30</sup> 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<sup>30</sup> 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  $\Delta t$  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 polyaniline octapeptide (8ALA) with charged termini described by the GROMOS96 force field<sup>64</sup> and embedded in a cubic box of 20 Å edge-length containing 236 simple point charge (SPC) water molecules.<sup>65</sup> 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 ( $\tau_P = \tau_S = 0.1$  ps,  $T_{0,S} = T_{0,P} = 300$  K). The second model system was an alanine dipeptide (ALDI) described by the CHARMM22 force field<sup>66</sup> in a cubic box of 21.3 Å edge length containing 324 water molecules modeled by the transferable three-point intermolecular potential (TIP3P).<sup>66,67</sup> 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<sup>a</sup>

name	software	protocol			thermostat parameters			
		C	$\Delta t/\text{fs}$	$D/\text{ns}$	$\tau_{\text{sys}}/\text{ps}$	$\tau_{\text{S}}/\text{ps}$	$\tau_{\text{P}}/\text{ps}$	$T_{0,\text{P}}/\text{K}$
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

<sup>a</sup> The simulation names code the varied parameters and temperature control scenarios. C specifies the type of bond length constraints,  $\Delta t$  the size of the basic integration time step, and  $D$  the duration of the simulation. The parameters  $\tau$  specify the coupling times of the BTs coupled to the whole system (sys), to the solvent (S), or to the solute (P).  $T_{0,\text{P}}$  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<sup>a</sup>

name	software	protocol			thermostat parameters			
		C	$\Delta t/\text{fs}$	$D/\text{ns}$	$\tau_{\text{sys}}/\text{ps}$	$\tau_{\text{S}}/\text{ps}$	$\tau_{\text{P}}/\text{ps}$	$T_{0,\text{P}}/\text{K}$
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

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

to the peptide ( $\tau_{\text{P}} = 500$  ps,  $T_{0,\text{P}} = 4800$  K,  $\hat{T}_{\text{P}} = 307.9$  K), the unknown parameters in eq 19 were determined as described in the section 2. We found the values  $\alpha_{\text{P}} = -2.04k_{\text{B}}$  K/ps for the algorithmic heating rate and  $\tau_{\text{SP}} = 1.61$  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  $T_{\text{P}} \approx 300.0$  K, these values were inserted into eq 19 yielding a “target temperature”  $T_{0,\text{P}} = 2340$  K. If the assumptions underlying our heat-flow model are correct, this choice should compensate through  $\beta_{\text{P}} = -\alpha_{\text{P}}$  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  $\mu\text{s}$  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  $T_{0,\text{P}}$  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 ( $T_{0,\text{P}} = 300$  K) or a slightly higher ( $T_{0,\text{P}} = 340$  K) target temperature as compared to  $T_{0,\text{S}}$ . 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  $\text{C}_{\alpha}$  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

**Table 3.** Simulation Parameters in Series #3<sup>a</sup>

peptide	software	protocol			thermostat parameters		
		C	$\Delta t/\text{fs}$	$D/\text{ns}$	$\tau_{\text{sys}}/\text{ps}$	$\tau_{\text{P}}/\text{ps}$	$T_{0,\text{P}}/\text{K}$
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

<sup>a</sup> For nomenclature see the caption to Table 1. In all simulations the solvent was coupled with  $\tau_{\text{S}} = 0.1$  ps to a Nosé-Hoover (NHT) or a Berendsen (BT) thermostat, respectively.

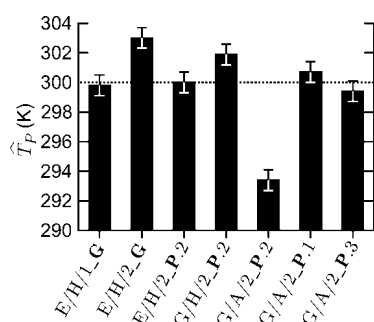
forces (eqs 5 and 6) and perturbation ratios (eq 4) in a third series of relatively short 250 ps simulations. Simulations were performed for 8ALA with varying coupling strengths and BTs and NHTs, respectively. Additionally, we determined the thermostatic forces and the perturbation ratio also for ALDI 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<sup>a</sup>

system		protocol			thermostat parameters		
peptide	environment	software	C	$\Delta t/\text{fs}$	$D/\text{ns}$	$\tau_P/\text{ps}$	$T_{0,P}/\text{K}$
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

<sup>a</sup> 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  $\Delta t$  was doubled to 2 fs. For our sample system, this doubling of  $\Delta t$

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  $\Delta t$  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  $\Delta t$  (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,<sup>32,34,68</sup> 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  $\beta_{SP}$  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,<sup>30</sup> 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

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

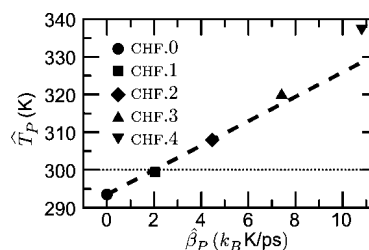
showing that the solute temperature  $T_P$  should depend linearly on the heating power  $\beta_P$  of the solute thermostat. To specify the unknown parameters  $\alpha_P$  and  $\tau_{SP}$  in eq 20, one needs measurements of  $T_S$  and  $T_P$  from two simulations employing different heating powers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

Estimates  $\hat{\beta}_P$  for the heating powers  $\beta_P$  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

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

which is, up to the use of different averages, identical to eq 18. Thus, at a constant coupling time  $\tau_P$ , the heating power  $\beta_P$  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  $\beta_P$  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

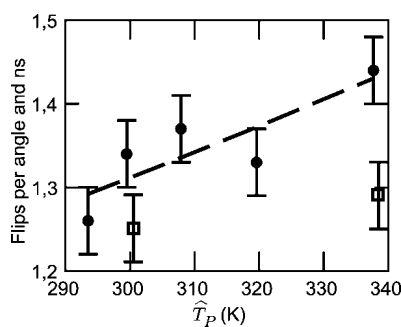


**Figure 3.** The average temperatures  $\hat{T}_P$  of the peptide ALA8 (in SPC water at  $\hat{T}_S = 300$  K) resulting from constant local heating with different powers  $\beta_P$  in the simulation sets CHF.0 to CHF.4 of series #2 (cf. Table 2). The prediction of linear heat flow model eq 20 is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

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  $\beta_P$  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.

The dashed line in Figure 3 expresses the linear relation 20 between  $\beta_P$  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  $\alpha_P$  and  $\tau_{SP}$  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  $\alpha_P$  and  $\tau_{SP}$  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  $\beta_P$  and  $T_P$ . In the case of 8ALA in explicit water, the assumption of a linear thermal coupling between solvent



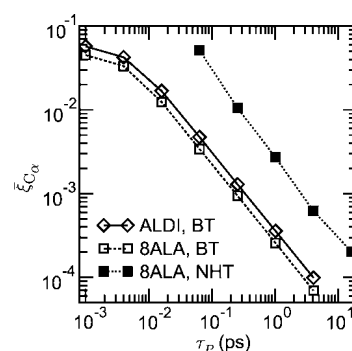
**Figure 4.** Temperature dependence of the peptide backbone dynamics of 8ALA. The graph shows the average number of transitions per angle and nanosecond of the  $\psi$ -dihedral angles between the  $\alpha$ -type region  $[-60^\circ, -30^\circ]$  and the  $\beta$ -type region  $[95^\circ, 145^\circ]$  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<sup>69</sup> 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<sup>69</sup> 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.<sup>70</sup>

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  $C_\alpha$  atoms of 8ALA and ALDI evaluated from simulation series #3 for the NHTs and BTs, respectively, for different coupling times  $\tau_P$  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  $C_\alpha$  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  $\tau_P$  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  $\tau_P$ . For the classical BT ( $\tau_P = 0.1$  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  $\tau_P$ , Nosé-Hoover coupling inflicts perturbations which are by 1 order of

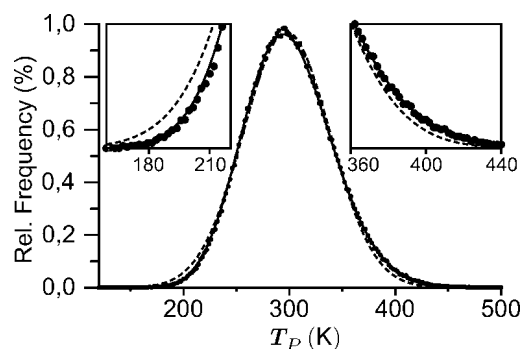
magnitude larger than in the Berendsen case (cf. Figure 5). For a Berendsen coupling of maximal strength ( $\tau_p = 0.001$  ps) the perturbation is comparable to that of a NHT with  $\tau_p$  as large as 0.064 ps. Furthermore, for Nosé-Hoover coupling  $\tau_p$  cannot be chosen larger than about 0.256 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<sup>44</sup> (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  $\tau_p$  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.

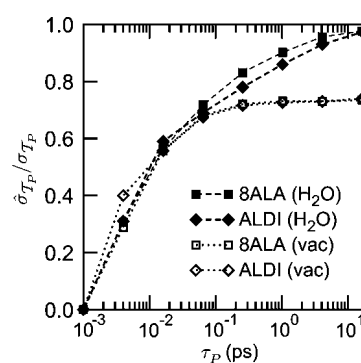
The above analysis was based on data for perturbation ratios derived through the approximate expression 6 and, therefore, depends on the validity of this equation. The first assumption made in the derivation (cf. section 2) of eq 6 was that the atomic velocities  $\dot{\mathbf{r}}_i(t)$  and the thermostat 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  $\chi^2$ -distribution (eq 7) for the instantaneous peptide temperature  $T_p(t)$  with results from simulation G/A/2\_P.3. For the 103 degrees of freedom of 8ALA, the  $\chi^2$ -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  $\chi^2$ -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



**Figure 6.** Distribution of the instantaneous temperature  $T_p(t)$  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  $\tau_p$  of a Berendsen solute thermostat. The model peptides are 8ALA (squares) and ALDI (diamonds). Simulations were performed in explicit water ( $H_2O$ , filled symbols) and vacuum (vac, empty symbols) for both peptides. Simulation parameters are given in Table 4.

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).

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  $\sigma_{T_p}$ , 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  $\tau_p$ . 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  $\tau_p$ . 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  $\tau_{SP}$  by a factor of 10. As a result, the separate BT is effectively disconnected from the peptide, the solute–solvent heat exchange term  $\beta_{SP}$  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  $\tau_p$  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  $\tau_p$  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  $\tau$  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 non-invasive strategy **P.2** can suffice for quite large solvent–solute systems. For instance, reinspecting a simulation<sup>8</sup> 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 environment  $S$  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  $T_{\text{sim}}$  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  $30\tau_{SP}$  for the respective subsystems to safely attune to  $T_{\text{sim}}$ .

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  $\tau_{SP}$ . Since  $\tau_{SP}$  is still unknown, an upper limit estimate (e.g.,  $\tau_{SP} \approx 20$  ps) should be used to determine the relaxation time  $t_{\text{relax}} \approx 10\tau_{SP}$ .

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  $|T_{P,1} - T_{\text{sim}}|$  is less than an acceptable tolerance  $\Delta T_P$ , 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  $\sigma_{\hat{T}_{P,1}}^2$  of the measured solute temperature  $\hat{T}_{P,1}$ , which forms an upper bound for the uncertainty  $\sigma_{T_P}^2$  in the prediction of the production run temperature  $T_P$ . If  $\Delta T_P$  is the accuracy required for the prediction, we should make sure that  $\sigma_{\hat{T}_{P,1}}^2 \leq \Delta T_P^2$ . By eq 9 the simulation time then is  $t_1 = 2\tau_c \sigma_{T_P}^2 / \Delta T_P^2$ , where  $\tau_c$  is the temperature autocorrelation time of the solute, and  $\sigma_{T_P}$  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  $\tau_{SP}$  (e.g., 1 ps), a reasonable choice for the target temperature is given by  $T_{0,P,2} = -\tau_P/\tau_{SP} \cdot |T_{P,1} - T_{\text{sim}}|$  (leading to 2-fold overcompensation if  $\tau_{SP}$  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

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

## 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 lichtschtaltbarer 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.



# 3 Optimale Strategien für *Replica-Exchange-Simulationen* mit explizitem Lösungsmittel

Citius, Altius, Fortius!

---

*Motto der neuzeitlichen  
olympischen Spiele*

Nachdem wir bisher Verfahren zur statistisch korrekten Temperaturkontrolle bei MD-Simulationen von Peptiden in Lösung eingehend erörtert haben, wenden wir uns nun dem Abtastproblem zu, das aus den notwendigerweise begrenzten Simulationszeiten erwächst. Anhand einfacher numerischer Modelle und mit Hilfe mathematischer Analyse wird gezeigt, wie sich die statistische Effizienz von *replica-exchange*-Simulationen von MD-Systemen mit explizitem Lösungsmittel optimieren lässt.

## 3.1 Optimale Temperaturleitern

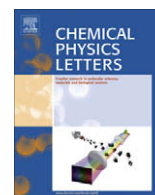
Wir beginnen mit einem Abdruck<sup>1</sup> des Artikels

Robert Denschlag, Martin Lingenheil, Paul Tavan:  
Optimal temperature ladders in replica exchange simulations  
*Chem. Phys. Lett.* **473**, 193-195 (2009),

den ich zusammen mit Robert Denschlag und Paul Tavan verfasst habe. Anhand von numerischen Simulationen unter Benutzung der *replica-exchange*-Standardstrategie wird gezeigt, dass frühere theoretische Modelle [145, 147–149] der Diffusion eines Replikats auf der Temperaturleiter falsche Voraussagen über eine entscheidende Kenngröße für die Effizienz von *replica exchange* liefern. Empirisch wird eine Regel für die Konstruktion einer Temperaturleiter abgeleitet, die tatsächlich diese Kenngröße maximiert.

---

<sup>1</sup>Mit freundlicher Genehmigung des Elsevier Verlags.



# Optimal temperature ladders in replica exchange simulations

Robert Denschlag, Martin Lingenheil, Paul Tavan \*

Lehrstuhl für Biomolekulare Optik, Ludwig-Maximilians-Universität, Oettingenstr. 67, 80538 München, Germany

## ARTICLE INFO

### Article history:

Received 19 January 2009

In final form 20 March 2009

Available online 25 March 2009

## ABSTRACT

In replica exchange simulations, a temperature ladder with  $N$  rungs spans a given temperature interval. Considering systems with heat capacities independent of the temperature, here we address the question of how large  $N$  should be chosen for an optimally fast diffusion of the replicas through the temperature space. Using a simple example we show that choosing average acceptance probabilities of about 45% and computing  $N$  accordingly maximizes the round trip rates  $r$  across the given temperature range. This result differs from previous analyses which suggested smaller average acceptance probabilities of about 23%. We show that the latter choice maximizes the ratio  $r/N$  instead of  $r$ .

© 2009 Published by Elsevier B.V.

## 1. Introduction

At given computer resources, the benefit of replica exchange [1–3] (RE) simulations crucially depends on the choice of certain parameters. Having chosen a temperature range  $[T_{\min}, T_{\max}]$ , which should be covered by the RE simulation, the optimal form of the temperature ladder ( $T_1 = T_{\min}, T_2, \dots, T_N = T_{\max}$ ) is an important issue [4–10]. Aiming at a minimal average round trip time of the replicas in the temperature space and assuming a constant heat capacity  $C$ , which should approximately apply to explicit solvent simulations [11], Nadler and Hansmann [9] have derived a formula

$$N \approx 1 + 0.594\sqrt{C} \ln(T_{\max}/T_{\min}) \quad (1)$$

for the number  $N$  of rungs in the temperature ladder. In Eq. (1) the (extensive) heat capacity  $C$  is given in units of the Boltzmann constant  $k_B$  and refers to the potential energy part of the total energy. As suggested by Okamoto et al. [12], from  $N$  one can determine the temperature rungs  $T_i, i = 1, \dots, N$ , in the ladder by

$$T_i = T_{\min}(T_{\max}/T_{\min})^{(i-1)/(N-1)}. \quad (2)$$

This choice is generally expected [3] to provide equal exchange probabilities  $p_{\text{acc}}(T_i, T_{i+1}) = p_{\text{acc}}(N)$  along the  $N$ -rung ladder.

Defining the function

$$\alpha(N) \equiv (T_{\max}/T_{\min})^{1/(N-1)}, \quad (3)$$

one immediately finds that the temperature rungs are given by the recursion

$$T_{i+1} = T_i \alpha(N). \quad (4)$$

Thus, for a given  $N$ , the ratio  $T_{i+1}/T_i$  is the constant  $\alpha(N)$ . For such a ladder and normally distributed potential energies, which is, along

with a constant heat capacity, typical for explicit solvent simulation systems, the average acceptance probabilities are very well approximated [6] by

$$p_{\text{acc}}(N) = \text{erfc} \left[ \sqrt{C} \frac{\alpha(N) - 1}{\alpha(N) + 1} \right], \quad (5)$$

where  $\text{erfc}(x) = 2/\sqrt{\pi} \int_x^\infty \exp(-x^2) dx$  is the complementary error function.

In summary, for a ladder spanning the temperature range  $[T_{\min}, T_{\max}]$  by the exponential spacing law Eq. (2), the temperature rungs  $T_i$  are uniquely given by  $N$ . Assuming a constant heat capacity and normally distributed potential energies, such a ladder then actually provides equal average acceptance probabilities [Eq. (5)]. Therefore, temperature ladders obeying Eqs. (2) and (5) are uniquely specified by choosing either a certain number  $N$  of rungs or a certain average acceptance probability  $p_{\text{acc}}$ .

## 2. Methods and simulation set-up

To check whether the formula given in Eq. (1) and suggested by Nadler and Hansmann [9] actually yields RE temperature ladders with minimal round trip times, we have designed simple test systems suited for computationally inexpensive RE Monte Carlo (REMC) simulations. The systems consist of  $d$  independent one-dimensional and harmonic oscillators in the canonical ensemble (we have chosen the same potential  $E = x^2$  for all oscillators). At each REMC step the coordinates of all  $d$  oscillators in a replica are randomly drawn from the associated normal distributions, the total energy  $E_i$  of the system at temperature  $T_i$  is calculated, and an exchange of systems at neighboring [3] temperatures is attempted with the Metropolis probability [13]  $p(i, i+1) = \min\{1, \exp[(1/k_B T_{i+1} - 1/k_B T_i)(E_{i+1} - E_i)]\}$ . We employed the standard exchange scheme [1], which alternately attempts exchanges between 'even' ( $T_{2i}, T_{2i+1}$ ) and 'odd' ( $T_{2i-1}, T_{2i}$ ) replica pairs. Below

\* Corresponding author. Fax: +49 89 2180 9220.

E-mail address: [tavan@physik.uni-muenchen.de](mailto:tavan@physik.uni-muenchen.de) (P. Tavan).

we call this RE scheme, which combines the standard exchange with the standard Metropolis criterion, the standard RE set-up.

Note the important fact that the heat capacity of our test system is independent of the temperature and is given by  $C = d/2$ . Therefore, it matches the conditions assumed in the derivation of Eq. (1). Note furthermore that the force constants of the harmonic oscillator potentials are of no concern because the Metropolis probability solely depends on the overlaps of the energy distributions.

First we consider the case  $d = 100$ . As extremal temperatures we choose  $T_{\min} = 300$  K and  $T_{\max} = 800$  K. With these parameters, Eqs. (1), (2), and (5) yield  $N = 5$ , the temperature ladder (300, 383.4, 389.9, 626.0, 800), and the average acceptance probability  $p_{\text{acc}} \approx 22\%$ , respectively. Previously also Kone and Kofke [6] and Rathore et al. [5] have suggested an acceptance probability of about 23% to be optimal. Thus, choosing the number of rungs through Eq. (1) seems to yield a reasonable acceptance probability.

The question as to whether the above choice actually entails minimal round trip times in REMC simulations can be addressed by comparing the set-up outlined above with alternatives defined by different choices of  $N$ . We tested ladders with  $N \in \{3, 4, \dots, 9, 10, 12, \dots, 18, 20\}$  each spanning the same temperature range [300 K, 800 K]. Every associated REMC simulation covered  $S = 500000$  MC steps. From each of these REMC simulations we determined the number of round trips  $M(N)$ . Here, a round trip was counted whenever a selected replica that started at  $T_{\min}$  subsequently reached  $T_{\max}$  and eventually returned to  $T_{\min}$ . Considering instead of the round trip time  $\tau(N)$  of a replica its inverse, the round trip rate  $r(N) \equiv M(N)/S$ , we asked which acceptance probability  $p_{\text{acc}}(N)$  [cf. Eq. (5)] belongs to the maximal rate  $r(N)$  measured in any of our simulations.

### 3. Results

Fig. 1 shows the measured round trip rate  $r$  as a function  $r(p_{\text{acc}})$  of the average acceptance probability  $p_{\text{acc}}$ . Two data points are additionally marked by the numbers of rungs in the associated ladders ( $N = 5, N = 7$ ). According to the graph the round trip rate  $r$  is maximal at  $p_{\text{acc}} \approx 0.42$  belonging to the  $N = 7$  rung ladder. This result differs from the expectation voiced above that  $r$  should be maximal at  $p_{\text{acc}} \approx 0.22$  or  $N = 5$ , respectively.

This surprising result raises the question why Eq. (1) yields a prediction for the optimal  $N$  (or  $p_{\text{acc}}$ ) differing from the measured one. Nadler and Hansmann [9] started the derivation of Eq. (1) by assuming for the round trip rate the plausible relation  $r(p_{\text{acc}}, N) = kp_{\text{acc}}/N(N - 1)$  with a certain constant  $k > 0$ . Using this assumption, one predicts that the rate  $r(0.22, 5) = k0.22/5(5 - 1) \approx 0.011k$  should be larger than the rate  $r(0.42, 7) = k0.42/7(7 - 1) \approx 0.010k$ , which is clearly at variance with the results of our simulation. Thus, the quoted relation does not yield

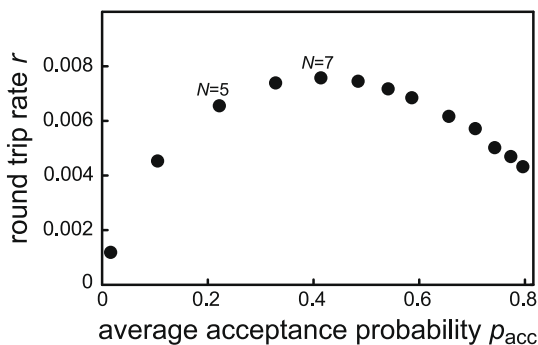


Fig. 1. Measured round trip rates  $r$  as a function of the average acceptance probability  $p_{\text{acc}}$ .

the correct round trip rate  $r(p_{\text{acc}})$  and, correspondingly, the choice of the ladder size  $N$  through Eq. (1) does not maximize  $r$ , if the standard RE set-up is used.

To understand how  $r$  depends on  $p_{\text{acc}}$  we introduce the average (and relative) temperature move

$$\kappa(T_i, T_{i+1}) \equiv p_{\text{acc}}(T_i, T_{i+1}) \frac{T_{i+1} - T_i}{T_i} \quad (6)$$

of a replica per exchange trial (here one MC step).  $\kappa$  measures the average velocity of the replicas in a properly scaled temperature space. This definition is motivated by our assumption that the average replica velocity  $\kappa$  should be proportional to the round trip rate, i.e. that  $\kappa = fr$  with a constant  $f > 0$ . Thus, we expect the largest round trip rates  $r$  for the largest velocities  $\kappa$ .

Using Eq. (4) and inverting Eq. (5) one finds for the average replica velocity

$$\kappa(p_{\text{acc}}) = p_{\text{acc}} \frac{2\text{erfc}^{-1}(p_{\text{acc}})}{\sqrt{C} - \text{erfc}^{-1}(p_{\text{acc}})}, \quad (7)$$

where  $\text{erfc}^{-1}$  denotes the inverse of the complementary error function. Thus, in contrast to the impression evoked by the definition in Eq. (6),  $\kappa$  is a constant within a given temperature ladder (because  $p_{\text{acc}}$  is a constant across each ladder).

The lines in Fig. 2 are the graphs of the function  $\kappa(p_{\text{acc}})$  given by Eq. (7) for systems with small and large heat capacities  $C = d/2$ . For  $d = 100$  (solid) the velocity  $\kappa$  becomes maximal at  $p_{\text{acc}} \approx 0.42$  (i.e. at  $N = 7$ ) and for  $d = 1000$  (dashed) at  $p_{\text{acc}} \approx 0.44$  (i.e. at  $N = 22$ ). Fig. 2 additionally displays scaled round trip rates  $fr$  measured for the small ( $f = 9.71$ ) and the large system ( $f = 24.6$ ), respectively. The good match of the scaled rates  $fr(p_{\text{acc}})$  (circles/squares) with the respective graphs  $\kappa(p_{\text{acc}})$  verifies our assumption that the average velocity  $\kappa(p_{\text{acc}})$  of the replicas in scaled temperature space is proportional to the round trip rate  $r(p_{\text{acc}})$ . Note that the value of the optimal average acceptance rate, at which the round trip rate  $r$  becomes maximal, depends only weakly on the system size  $d$ .

The noted weak dependence of the optimal average acceptance probability  $p_{\text{acc}}$  on the system size  $d = 2C$  can be understood by considering the limit of large systems ( $C \rightarrow \infty$ ). Using *Mathematica* [14] we have determined the derivative  $\kappa'(p_{\text{acc}}) \equiv d\kappa/dp_{\text{acc}}$  and its first order Taylor expansion  $\kappa'(p_{\text{acc}}) = a_0 + a_1 p_{\text{acc}} + \mathcal{O}(p_{\text{acc}}^2)$  at  $p_{\text{acc}} = 0.5$ . In linear approximation  $\kappa'$  vanishes at  $p_{\text{acc}}^0 = -a_0/a_1$  which is given by

$$p_{\text{acc}}^0 \approx 0.45 + g/\sqrt{C} \quad (8)$$

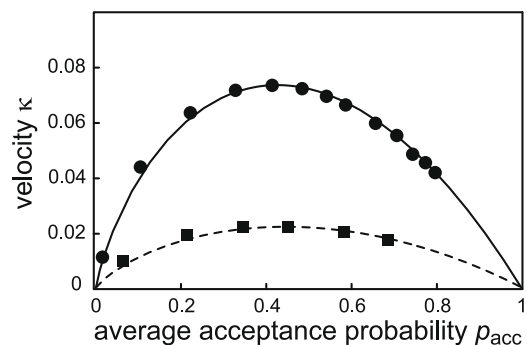
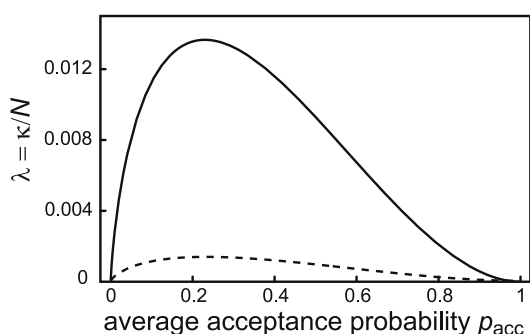


Fig. 2. Average replica velocity  $\kappa$  in scaled temperature space as a function of the average acceptance probability  $p_{\text{acc}}$ . The lines are graphs of  $\kappa(p_{\text{acc}})$  calculated from Eq. (7) for  $d = 100$  (solid) and  $d = 1000$  (dashed). The dots are the round trip rates  $r$  of Fig. 1 scaled by the factor  $f = 9.71$ . The squares are round trip rates scaled by  $f = 24.6$  and resulting from REMC simulations of  $d = 1000$  oscillators. Here, the rung numbers  $N \in \{10, 14, 18, 22, 30, 40\}$  increase from left to right.



**Fig. 3.**  $\lambda = \kappa/N$  as a function of the average acceptance probability  $p_{\text{acc}}$ . The lines are the graphs of Eq. (10) for  $d = 100$  (solid) and  $d = 1000$  (dashed).

with the constant  $g$  being of the order of  $-0.1$ . Thus for large systems the location  $p_{\text{acc}}^0$  of the maximum of  $\kappa$  approaches 0.45 from below. Note that we have also checked the limiting value 0.45 by numerically analyzing Eq. (7) for very large  $C$ .

The thus established limiting value of the optimal average acceptance probability leads to a new estimate

$$N \approx 1 + (\sqrt{C}/(2 \times 0.534) - 1/2) \ln(T_{\text{max}}/T_{\text{min}}) \quad (9)$$

for the optimal number of rungs in the associated temperature ladder. For computing Eq. (9) we have used Eq. (5),  $\text{erfc}^{-1}(0.45) \approx 0.534$ , and  $\ln[1 + 2 \times 0.534/(\sqrt{C} - 0.534)] \approx 2 \times 0.534/(\sqrt{C} - 0.534)$  for large  $C$ .

In summary, to maximize the round trip rate or minimize the round trip time, respectively, Eq. (9) has to be used instead of Eq. (1). From a practical point of view, however, instead of aiming at the maximal round trip rate, one may be content with a suboptimal rate if this choice is associated with a reduced computational effort. Instead of maximizing  $\kappa$  one may therefore consider the quantity

$$\lambda(p_{\text{acc}}) \equiv \kappa(p_{\text{acc}})/N(p_{\text{acc}}), \quad (10)$$

which exhibits a penalty linear in the number  $N$  of rungs. Fig. 3 shows the graphs of  $\lambda(p_{\text{acc}})$  for  $d = 100$  (solid line) and  $d = 1000$  (dashed line). In both cases  $\lambda$  is maximal at  $p_{\text{acc}} \approx 0.23$ . Applying once again the reasoning used in the derivation of Eq. (8) one finds that the optimal  $p_{\text{acc}}$  approaches 0.234 for  $C \rightarrow \infty$ . This result leads to the estimate

$$N \approx 1 + (0.594\sqrt{C} - 1/2) \ln(T_{\text{max}}/T_{\text{min}}) \quad (11)$$

for the number of rungs optimizing the specific compromise  $\lambda = \kappa/N$  between the round trip rate  $r$  and the number  $N$  of replicas.

Note here that Eqs. (11) and (1) become identical for large  $C$ . Thus, the number of rungs resulting from Eq. (1) effectively maximizes  $\lambda$  instead of  $\kappa$  (or, equivalently, a ladder size penalized round trip rate  $r/N$  instead of the round trip rate  $r$ ). That Nadler and Hansmann [9] have effectively maximized  $\lambda$  instead of  $r$  can be alternatively understood by inserting the definition Eq. (6) into (10). Using subsequently Eq. (4) and the approximation

$(T_{\text{max}}/T_{\text{min}})^{1/(N-1)} \approx 1 + \ln(T_{\text{max}}/T_{\text{min}})/(N-1)$ , which holds for large  $N$ , one finds  $\lambda \approx p_{\text{acc}} \ln(T_{\text{max}}/T_{\text{min}})/[N(N-1)]$  such that  $\lambda \propto p_{\text{acc}}/[N(N-1)]$ . Recall now that Nadler and Hansmann erroneously assumed a relation of this kind for the round trip rate  $r$  (i.e. for  $\kappa$ ).

We would like to remark that our results are transferable to simulated tempering [15,16] (ST) simulations. For ST, the average acceptance probability is given by

$$p_{\text{acc}}^{\text{ST}}(N) = \text{erfc}\left(\sqrt{C/2} \frac{\alpha(N) - 1}{\alpha(N) + 1}\right). \quad (12)$$

which is obtained from the RE expression Eq. (5) through replacing  $C$  by  $C/2$  [5,17]. In the limit of large systems also  $C/2$  becomes large and the optimal value of  $p_{\text{acc}}^{\text{ST}}$  is likewise at 45% or 23% depending on the optimized quantity. Similarly, the optimal  $N$  is given by Eqs. (9) or (11), respectively, through replacing  $C$  by  $C/2$ .

#### 4. Summary

For simulations employing the standard RE set-up, we have derived with Eqs. (9) and (11) two formulas for the optimal sizes  $N$  of temperature ladders obeying Eqs. (2) and (5). Here, optimal means that either the round trip rate  $r$  or the compromise  $r/N$  with computational effort is maximized. We have furthermore shown that the suggestion Eq. (1) of Nadler and Hansmann [9] maximizes the compromise  $r/N$  and not  $r$ , as claimed by the authors. An optimal  $r$  is obtained with average acceptance probabilities of about 45%, whereas an optimal  $r/N$  requires values of 23% matching earlier suggestions [6,5]. As a practical consequence of our study one sees that average acceptance probabilities chosen in the range from 20% to 45% are definitely ‘good’ choices featuring, however, slightly different merits.

#### Acknowledgement

This work was supported by the Deutsche Forschungsgemeinschaft (SFB 533/C1 and SFB 749/C4).

#### References

- [1] K. Hukushima, K. Nemoto, J. Phys. Soc. Jpn. 65 (1996) 1604.
- [2] U.H.E. Hansmann, Chem. Phys. Lett. 281 (1997) 140.
- [3] Y. Sugita, Y. Okamoto, Chem. Phys. Lett. 314 (1999) 141.
- [4] C. Predescu, M. Predescu, C.V. Ciobanu, J. Chem. Phys. 120 (2004) 4119.
- [5] N. Rathore, M. Chopra, J.J. de Pablo, J. Chem. Phys. 122 (2005) 024111.
- [6] A. Kone, D.A. Kofke, J. Chem. Phys. 122 (2005) 206101.
- [7] S. Trebst, M. Troyer, U.H.E. Hansmann, J. Chem. Phys. 124 (2006) 174903.
- [8] W. Nadler, U.H.E. Hansmann, Phys. Rev. E 75 (2007) 026109.
- [9] W. Nadler, U.H.E. Hansmann, J. Phys. Chem. B 112 (2008) 10386.
- [10] W. Nadler, J.H. Meinke, U.H.E. Hansmann, Phys. Rev. E 78 (2008) 061905.
- [11] B. Paschek, H. Nymeyer, A.E. Garcia, J. Struct. Biol. 157 (2007) 524.
- [12] Y. Okamoto, M. Fukugita, T. Nakazawa, H. Kawai, Protein Eng. 4 (1991) 639.
- [13] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, E. Teller, J. Chem. Phys. 21 (1953) 1087.
- [14] Wolfram Research, Inc., Mathematica Version 5.1, Champaign, IL, 2004.
- [15] A.P. Lyubartsev, A.A. Martinovski, S.V. Shevkunov, P.N. Vorontsov-Velyaminov, J. Chem. Phys. 96 (1992) 1776.
- [16] E. Marinari, G. Parisi, Europhys. Lett. 19 (1992) 451.
- [17] C. Zhang, J. Ma, J. Chem. Phys. 129 (2008) 134112.

## 3.2 Effiziente Austauschschemata

Die Diskussionen des vorangegangenen Abschnitts, die wesentlich auf Heuristik und Simulationsexperimenten beruhten, werden nun analytisch vertieft. Dazu dient der Abdruck des Artikels<sup>2</sup>

Martin Lingenheil, Robert Denschlag, Gerald Mathias, Paul Tavan  
„Efficiency of exchange schemes in replica exchange“,  
*Chem. Phys. Lett.* (2009). *Zur Veröffentlichung angenommen.*

den ich zusammen mit Robert Denschlag, Gerald Mathias und Paul Tavan verfasst habe. Hier wird beschrieben, dass das üblicherweise in *replica-exchange*-Simulationen verwendete Austauschschema tatsächlich eine optimale Wahl unter vier verglichenen Methoden darstellt. Durch die mathematische Analyse der aus diesem Austauschschema folgenden Zufallsbewegung wird gezeigt, dass dieses Schema wegen seines partiell deterministischen Charakters zu einer inhärent schnelleren Diffusion der Replikat auf der Temperaturleiter führt als eine vollständig stochastische Variante. Da frühere Modelle zur Diffusion der Replikat implizit von dieser vollständig stochastischen Variante ausgingen [145, 147–149], erklärt sich deren Inkonsistenz mit den im vorangegangenen Abschnitt 3.1 beschriebenen empirischen Beobachtungen.

---

<sup>2</sup>Mit freundlicher Genehmigung des Elsevier Verlags.

# Efficiency of exchange schemes in replica exchange

Martin Lingenheil, Robert Denschlag, Gerald Mathias, Paul Tavan\*

*Lehrstuhl für BioMolekulare Optik, Ludwig-Maximilians-Universität,  
Oettingenstr. 67, 80538 München, Germany*

---

## Abstract

In replica exchange simulations a fast diffusion of the replicas through the temperature space maximizes the efficiency of the statistical sampling. Here, we compare the diffusion speed as measured by the round trip rates for four exchange algorithms. We find different efficiency profiles with optimal average acceptance probabilities ranging from 8% to 41%. The best performance is determined by benchmark simulations for the most widely used algorithm, which alternately tries to exchange all even and all odd replica pairs. By analytical mathematics we show that the excellent performance of this exchange scheme is due to the high diffusivity of the underlying random walk.

---

## Introduction

The replica exchange (RE) method [1–3] has become a standard approach in molecular simulation to efficiently sample the rough energy landscapes of biomolecules in solution at a target temperature  $T_1$  (see e.g. Ref. [4]). In RE simulations,  $N$  simulation systems (replicas) are parallelly propagated in time using Monte Carlo (MC) or molecular dynamics (MD) algorithms. For the standard temperature RE method in particular, the replicas  $i \in \{1, \dots, N\}$  are identical with the exception of the respective simulation temperatures  $T_i$ . At a predefined temporal spacing an exchange between two replicas  $i$  and  $j$  is attempted and is accepted with the Metropolis [5] probability

$$p_{ij} = \min \left\{ 1, \exp \left[ (\beta_j - \beta_i)(E_j - E_i) \right] \right\}, \quad (1)$$

where  $E_i$  and  $E_j$  are the current potential energies of the replicas at the corresponding inverse temperatures  $\beta_i = 1/k_B T_i$  and  $\beta_j = 1/k_B T_j$ , respectively. Here,  $k_B$  denotes Boltzmann's constant. The exchange probability given by Eq. (1) satisfies the detailed balance condition and therefore guarantees that the ensembles sampled by the individual replicas remain undisturbed by the exchange.

Due to the exchanges, each replica performs a random walk through the temperature space  $[T_1, T_N]$ . During the high temperature phases of its trajectory, a replica crosses potential energy barriers more rapidly, leading in many cases (a relevant counter example has been given in Ref. [6]) to a faster convergence, compared to a straight forward simulation, of the statistical sampling at the lower temperatures. Here it is crucial for an optimal statistical sampling at the low temperatures that the replicas cycle between low and high temperatures as frequently as possible [7–11]. As we will demonstrate below, the sizes

of these round trip rates strongly depend on the detailed algorithm by which the replica pairs are selected for attempting an exchange.

A widely used exchange scheme [12–14] divides the set  $\mathcal{N} \equiv \{(i, i+1) | i = 1, \dots, N-1\}$  of next neighbors in the temperature ladder into the two subsets  $\mathcal{E} \subset \mathcal{N}$  and  $\mathcal{O} \subset \mathcal{N}$ , where  $\mathcal{E}$  contains all “even” pairs  $(2j, 2j+1) \in \mathcal{N}$  and  $\mathcal{O}$  contains the “odd” pairs  $(2j-1, 2j) \in \mathcal{N}$ . Exchanges are attempted alternatingly for the members of  $\mathcal{E}$  and  $\mathcal{O}$ . Because of the deterministic pattern of exchange trials, we will call this method the deterministic even/odd algorithm (DEO).

Besides the DEO scheme, we will also consider a randomized variant which, instead of alternatingly attempting exchanges among all even and all odd replica pairs, randomly chooses with equal probability one of the subsets  $\mathcal{E}$  and  $\mathcal{O}$ . Due to the stochastic selection of exchange sets, we call this method the stochastic even/odd algorithm (SEO). As we will show, the SEO scheme was implicitly assumed by a number of authors when theoretically deriving rules for optimal temperature ladders [7, 11]. Another reason for analyzing the SEO algorithm is that an exchange scheme equivalent to SEO is the straightforward choice when implementing simulated tempering [15].

Besides DEO also other exchange schemes have been discussed in the literature. The all-pair exchange (APE) method suggested by Brenner et al. [16] considers all possible exchange pairs including non-next neighbors. Finally, the very simple random next neighbor (RNN) algorithm [16, 17] chooses with equal probability at every exchange step a single pair from the set  $\mathcal{N}$  of next neighbors and attempts an exchange for this pair.

In this letter we will systematically check as to how the different exchange algorithms affect the diffusion of the replicas through the temperature space. This check will provide a rule for the optimal setup of RE simulations. For this purpose we will first introduce basic notions of the RE approach and a benchmark MC system. Using the benchmark system we will then compare the round trip rates obtained with the four differ-

---

\*Corresponding author. Fax: +49-89-2180-9220

Email address: tavan@physik.uni-muenchen.de (Paul Tavan)



ent exchange schemes. Because of the practical importance of the DEO algorithm, we will subsequently identify the reasons for its superior performance by analytical mathematics.

### Theoretical basics

It is general consensus that the distances of the  $N$  rungs  $T_i$  within the temperature ladder should be chosen to yield equal average acceptance probabilities  $\langle p_{i,i+1} \rangle = p_{\text{acc}}$  for the exchanges between neighboring replicas  $i$  and  $i + 1$  [3] provided that the simulated system does not undergo a phase transition within the range of the temperature ladder [8, 9]. If the system's heat capacity  $C$  is constant, which is approximately true for explicit solvent systems [18], then, following Okamoto et al. [19], the spacing law for equal average acceptance probabilities is

$$T_i = T_{\min} \cdot \alpha^{i-1}, \quad (2)$$

with the minimal temperature  $T_{\min}$  and a constant ratio  $\alpha = T_{i+1}/T_i$  of neighboring temperatures.

Given a certain temperature range  $[T_{\min}, T_{\max}]$  to be spanned by a simulation, the choice of the number  $N$  of replicas automatically determines the temperature ratio  $\alpha$  through

$$\alpha(N) = (T_{\max}/T_{\min})^{1/(N-1)}. \quad (3)$$

Next we assume Gaussian probability distributions

$$\rho(E_i) = \frac{1}{\sqrt{2\pi CT_i}} \exp\left[-\frac{(E_i - CT_i)^2}{2CT_i^2}\right] \quad (4)$$

for the potential energies  $E_i$  at the various temperatures  $T_i$  because these distributions are as typical for explicit solvent simulations as a constant heat capacity  $C$ . Note that, in Eq. (4),  $C$  denotes the (extensive) heat capacity in units of Boltzmann's constant  $k_B$  and refers to the potential energy part of the total energy.

With the potential energy distributions given by Eq. (4), the geometric temperature spacing by Eqs. (2) and (3), and the acceptance criterion by Eq. (1), the average acceptance probability according to Kone and Kofke [7] is

$$p_{\text{acc}} = \text{erfc}\left[\sqrt{C} \frac{\alpha(N) - 1}{\alpha(N) + 1}\right], \quad (5)$$

where  $\text{erfc}(x') = 2/\sqrt{\pi} \int_{x'}^{\infty} \exp(-x^2) dx$  is the complementary error function.

For a predefined temperature range  $[T_{\min}, T_{\max}]$ , Nadler and Hansmann [11] recently derived a formula to optimize an RE simulation setup with respect to the round trip rate  $r$ , i.e., to the average number of round trips a replica performs per unit time. In this optimal ladder spanning the interval  $[T_{\min}, T_{\max}]$ , the average acceptance probability  $p_{\text{acc}}$  is about 23% [20]. Consistently, Kone and Kofke [7] obtained the same value for  $p_{\text{acc}}$  when optimizing the diffusion of a replica on the temperature ladder. Most recently, however, we observed in sample simulations employing the DEO scheme that the allegedly optimal value  $p_{\text{acc}} \approx 23\%$  led to suboptimal round trip rates [20]. This surprising observation sparked our curiosity and led us to compare different exchange schemes using a very simple benchmark simulation system.

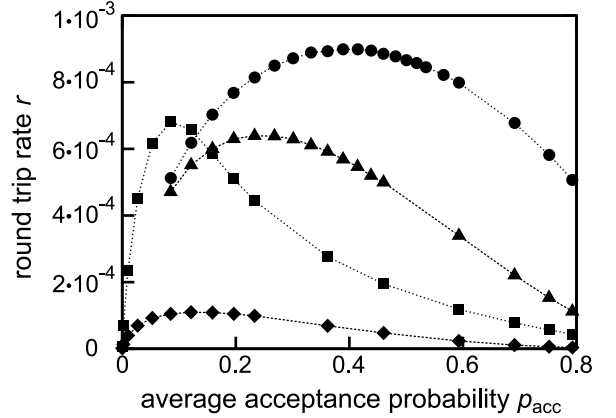


Figure 1: The round trip rates  $r$  measured for the four different exchange schemes and a system with  $C = 500$  as a function of the average acceptance probability  $p_{\text{acc}}$ : RNN (diamonds), APE (squares), SEO (triangles), DEO (circles). The dotted lines connecting the symbols are a guide for the eye.

### Benchmark simulations

For each of the four algorithms, DEO, SEO, APE, and RNN, we performed several RE Monte Carlo (REMC) benchmark simulations with differing numbers  $N$  of replicas but with a fixed temperature range  $T_{\min} = T_1 = 300$  K to  $T_{\max} = T_N = 800$  K and with the temperatures  $T_i$  spaced as given by Eq. (2). In these simulations we drew the potential energies  $E_i$  of the replicas at each REMC step from the distributions given by Eq. (4) choosing  $C = 500$  for the heat capacity. Then, one of the four algorithms was used to decide which exchanges should be considered, and the Metropolis criterion Eq. (1) was applied to evaluate the outcome of the exchange attempts. Every simulation comprised  $S = 10^7$  REMC steps. A round trip was counted if one of the replicas had traveled the complete way from  $T_1$  to  $T_N$  and back again. With the total number  $R$  of round trips counted during a simulation, the round trip rate is  $r \equiv R/NS$ .

Fig. 1 presents the measured round trip rates  $r$  as functions of  $p_{\text{acc}}(N)$ . The shown efficiency profiles  $r(p_{\text{acc}})$  of the various algorithms are markedly different. The simple RNN algorithm (diamonds in Fig. 1) shows by far the weakest performance and has its maximum round trip rate  $r_{\text{max}} \approx 10^{-4}$  at  $p_{\text{acc}} \approx 12\%$ . Because the RNN algorithm chooses only one pair from the set  $\mathcal{N}$  of next neighbors for an exchange trial and because this trial is successful with an average acceptance probability  $p_{\text{acc}}$ , the average number  $n_{\text{ex}}$  of actual exchanges per REMC step is equal to  $p_{\text{acc}}$ .

Compared with the RNN scheme, the more refined APE algorithm (squares in Fig. 1) yields much higher round trip rates with a maximal performance  $r_{\text{max}} \approx 6.8 \cdot 10^{-4}$  at  $p_{\text{acc}} \approx 9\%$ . At this value of  $p_{\text{acc}}$ , the average APE number of exchanges ( $n_{\text{ex}} = 0.55$ ) exceeds that of RNN ( $n_{\text{ex}} = 0.09$ ) by roughly a factor of 6. According to Fig. 1, here the APE round trip rate  $r$  is about 6.5 times higher than that of RNN. Hence, compared to RNN, the better performance of APE is mainly due to the larger value of  $n_{\text{ex}}$ , and the inclusion of non-next neighbor exchanges within APE seems to be of minor importance.

At  $p_{\text{acc}} \approx 23\%$  the round trip rate of the SEO algorithm

assumes its maximum value  $r_{\max} = 6.4 \cdot 10^{-4}$  (triangles in Fig. 1). At this point SEO exchanges nearly three times more pairs ( $n_{\text{ex}} = 1.45$ ) per REMC step than APE at its respective  $r_{\max}$ . Nevertheless, the maximal APE rate is higher than that of SEO. Interestingly, for SEO the position of  $r_{\max}$  in Fig. 1 perfectly agrees with the 23% acceptance probability predicted by the optimization formula of Nadler and Hansmann [11, 20] and with the point of maximal diffusivity predicted by Kone and Kofke [7].

As demonstrated by the circles in Fig. 1 the closely related DEO algorithm, performs everywhere better than SEO although both algorithms feature the same number  $n_{\text{ex}}$  of exchanges per REMC step for every choice of  $p_{\text{acc}}(N)$ . This improved performance of DEO is particularly pronounced at large  $p_{\text{acc}}$ , i.e. at large ladder sizes  $N$ . According to Fig. 1 the maximal DEO round trip rate  $r_{\max} = 9.0 \cdot 10^{-4}$  is found at  $p_{\text{acc}} \approx 41\%$ .

These findings suggest that the theory behind the optimizations performed by Kone and Kofke [7] as well as by Nadler and Hansmann [11] does apply to SEO but not to the established and widely used DEO scheme.

## Diffusive properties

*Elementary process.* To understand why DEO performs better than SEO, we analyzed the associated random walks performed by the replicas in the temperature space. A random walk is a sequence of statistically independent random experiments which we will call its elementary processes (EPs). After  $n$  EPs the displacement  $X \equiv \sum_{i=1}^n \Delta_i$  of the random walker is the sum of the displacements  $\Delta_i$  in the individual EPs. Since  $X$  is a sum of  $n$  identically distributed, statistically independent random variables  $\Delta_i$ , its variance  $\sigma^2(X)$  is given by  $n\sigma^2(\Delta)$  [21], where  $\sigma^2(\Delta)$  is the variance of the EP. If  $\langle d \rangle$  is the average duration of the random walk's EP, then its diffusivity, i.e. the gain in variance per unit time, is given by

$$D = \sigma^2(\Delta) / \langle d \rangle. \quad (6)$$

*Diffusivity of SEO.* In the SEO scheme, the EP may have the following three outcomes: (i) The replica moves one step upward ( $\Delta = +1$ ) on the temperature ladder with a total probability  $p_{\pm} p_{\text{acc}}$ , where  $p_{\pm} = 0.5$  is the probability of selecting the replica pair sets  $\mathcal{E}$  or  $\mathcal{O}$ , respectively. (ii) The replica moves one step downward ( $\Delta = -1$ ) on the ladder with the same probability. (iii) The replica does not move at all ( $\Delta = 0$ ) with the rejection probability  $1 - p_{\text{acc}}$ . Thus, the variance of the EP is

$$\sigma^2(\Delta) = p_{\pm} p_{\text{acc}} \left[ (+1)^2 + (-1)^2 \right] + (1 - p_{\text{acc}}) 0^2 = p_{\text{acc}} \quad (7)$$

and its average duration is  $\langle d \rangle = 1$ . According to Eq. (6), the diffusivity of the SEO algorithm is then

$$D_{\text{SEO}} = p_{\text{acc}}. \quad (8)$$

Thus, in SEO a replica performs a simple random walk on the temperature ladder as is usually assumed in theoretical analyses of the RE approach [7, 9, 11, 22].

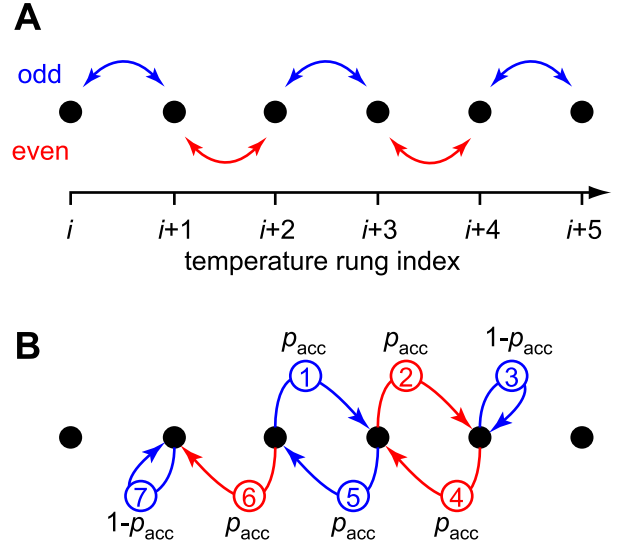


Figure 2: **A:** DEO exchanges within a temperatures ladder. The rungs  $i, i+1, \dots$  of the ladder are marked by black dots. The replica pairs which occupy the rungs joined by the blue double arrows are considered for exchange in odd REMC steps. Red double arrows refer to even REMC steps. **B:** An example for an EP. The upward march consists of  $k = 2$  successful steps, a failure in step 3, and of a downward march with  $k' = 3$  successful steps terminated by a failure in step 7. The probability of each step is indicated at the corresponding arrow.

*Diffusivity of DEO.* In contrast, the random walk imposed on a replica by the DEO algorithm is more complex. Let us first assume an infinite temperature ladder. For this situation, Fig. 2A sketches the logic of the exchanges. Assume that, in the odd REMC step  $j$ , the replica at rung  $i$  moves along the blue arrow upwards to rung  $i+1$ . Then, in the following even step  $j+1$ , it will be once again considered for an upward exchange toward rung  $i+2$  because DEO deterministically selects the red arrows for this exchange trial. As long as the exchange trials are successful, the replica keeps moving upward. This movement stops with the first exchange failure and, subsequently, the direction of the DEO replica movement is reversed. Thus, from the perspective of a single replica, the random walk consists of alternating upward and downward marches of variable lengths. Note that these lengths may also be zero if already the first exchange fails.

We define the EP of DEO as a pair consisting of an upward march and the following downward march. Let  $k \in \mathbb{N}_0$  denote the number of successful upward steps,  $k' \in \mathbb{N}_0$  the number of successful downward steps, and  $l \equiv k + k'$  the sum of both. Fig. 2B shows a typical example for such an EP. This EP causes a net displacement  $\Delta_{k,l} \equiv k - k' = 2k - l = -1$  and takes  $d_{k,l} \equiv k + k' + 2 = l + 2 = 7$  steps. The probability of the EP shown in Fig. 2B is the product of all indicated probabilities belonging to the terminated upward [ $p_{\text{acc}}^2(1-p_{\text{acc}})$ ] and downward marches [ $p_{\text{acc}}^3(1-p_{\text{acc}})$ ].

The general formula for the probability of  $k$  successful upward steps,  $k'$  successful downward steps, and two unsuccessful

exchange trials is

$$p_{k,l} \equiv p_{\text{acc}}^{k+k'} (1 - p_{\text{acc}})^2 = p_{\text{acc}}^l (1 - p_{\text{acc}})^2. \quad (9)$$

Thus, the variance of the EP is

$$\sigma^2(\Delta) = \sum_{l=0}^{\infty} \sum_{k=0}^l p_{k,l} \Delta_{k,l}^2 = \frac{2p_{\text{acc}}}{(1 - p_{\text{acc}})^2}, \quad (10)$$

and the average duration of the EP is

$$\langle d \rangle = \sum_{l=0}^{\infty} \sum_{k=0}^l p_{k,l} d_{k,l} = \frac{2}{1 - p_{\text{acc}}}. \quad (11)$$

From Eq. (6) we finally obtain for the DEO diffusivity

$$D_{\text{DEO}} = \frac{p_{\text{acc}}}{1 - p_{\text{acc}}}. \quad (12)$$

A comparison with Eq. (8) demonstrates that  $D_{\text{DEO}}$  is by a factor of  $1/(1 - p_{\text{acc}}) > 1$  larger than  $D_{\text{SEO}}$ .

*Round trip rates.* The above analysis has yielded diffusivities on infinite temperature ladders. However, what we want to know are round trip rates  $r$  on finite ladders. Such a rate  $r$  is the inverse of  $2\tau$ , where  $\tau$  is the mean first passage time required for a replica to diffusively cross the complete ladder. Therefore,  $r$  should be proportional [21] to the diffusivity  $D$  as long as the EPs of the underlying random walk are commensurate with the ladder size  $N$ . For the SEO scheme with its step sizes of 0 or  $\pm 1$  this requirement is automatically fulfilled. However, DEO features at decreasing probabilities also EPs of arbitrarily increasing step sizes. Therefore, the proportionality  $r_{\text{DEO}} \propto D_{\text{DEO}}$  is expected to hold only within a certain approximation.

Nadler and Hansmann [9] have calculated  $r$  assuming a SEO random walk directly from the corresponding master equation [21]. Inspection of Eq. (8) demonstrates that their result for the rate

$$r_{\text{SEO}} = \frac{p_{\text{acc}}}{2N(N-1)} \quad (13)$$

actually exhibits the expected proportionality  $r_{\text{SEO}} \propto D_{\text{SEO}} = p_{\text{acc}}$ . If we assume such a proportionality also for DEO we get  $r_{\text{DEO}} = r_{\text{SEO}} D_{\text{DEO}} / D_{\text{SEO}}$ . Hence with Eqs. (8), (12), and (13), we expect for DEO the round trip rate

$$r_{\text{DEO}} = \frac{p_{\text{acc}}}{(1 - p_{\text{acc}})2N(N-1)}. \quad (14)$$

*Comparison with measured rates.* In Fig. 3, we compare the round trip rates  $r(p_{\text{acc}})$  thus analytically calculated with rates measured in our test simulations (see Fig. 1). The dashed line in Fig. 3 is the SEO prediction of Eq. (13), where we have additionally used the unique relation between  $N$  and  $p_{\text{acc}}$  given by Eqs. (5) and (3). The agreement with the measured rates (triangles in Fig. 3) is perfect as expected. Also for DEO our expectations are met. As expected, the approximation Eq. (14) (solid line) deviates only a little from the measured rates (circles in Fig. 3).

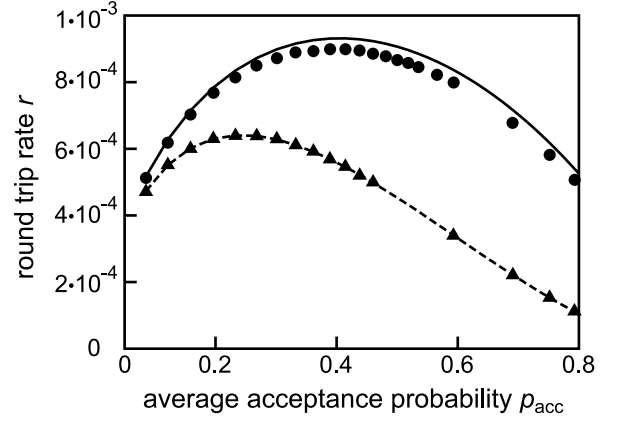


Figure 3: Analytically calculated round trip rates  $r_{\text{SEO}}$  [Eq. (13), dashed line] and  $r_{\text{DEO}}$  [Eq. (14), solid line] as functions of the average acceptance probability  $p_{\text{acc}}$ . Values measured in our test simulations are adopted from Fig. 1. The SEO values (triangles) perfectly match Eq. (13), whereas the DEO values (circles) slightly deviate from Eq. (14).

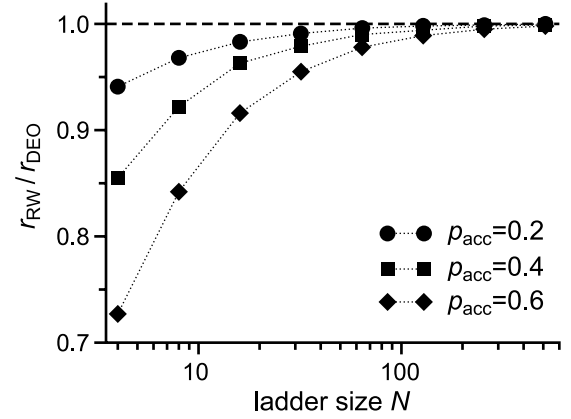


Figure 4: Ratio of a measured round trip rate  $r_{\text{RW}}$  (calculated with a fixed acceptance probability  $p_{\text{acc}}$  from extended DEO random walks on different ladder sizes  $N$ ) to the rate  $r_{\text{DEO}}$  given by Eq. (14). Ratios are shown for  $p_{\text{acc}} = 0.2$  (circles),  $p_{\text{acc}} = 0.4$  (squares),  $p_{\text{acc}} = 0.6$  (diamonds). The dotted lines connecting the symbols serve to guide the eye.

The deviation between Eq. (14) and the measurements noted for DEO must decrease with increasing ladder size, because for large ladders all those EPs which try to reach beyond the ladder boundary become less frequent. Fig. 4 illustrates this decrease for three different average acceptance probabilities. The figure compares a round trip rate  $r_{\text{RW}}$  measured in extended DEO random walks on ladders of increasing size  $N$  with the rate  $r_{\text{DEO}}$  from Eq. (14). Due to the decrease, the depicted ratio  $r_{\text{RW}}/r_{\text{DEO}}$  approaches the value one with growing  $N$  for all values of  $p_{\text{acc}}$ . Fig. 4 suggests that Eq. (14) is an upper limit for the actual round trip rate  $r_{\text{RW}}$  and that the overestimate of  $r_{\text{RW}}$  by  $r_{\text{DEO}}$  increases with  $p_{\text{acc}}$ . We note that at  $N = 10$  and  $p_{\text{acc}} = 0.4$  the deviation is less than 10%.

*Optimal temperature ladders.* The analytical approximation Eq. (14) for the DEO round trip rate  $r$  together with the unique correspondence between  $p_{\text{acc}}$  and  $N$  [Eqs. (3), (5)] enables us

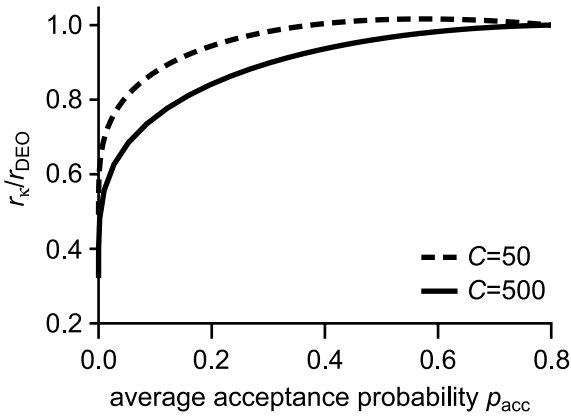


Figure 5: The ratio  $r_k/r_{\text{DEO}}$  as a function of the average acceptance probability  $p_{\text{acc}}$  for two systems with heat capacities  $C = 50$  (dashed line) and  $C = 500$  (solid line), respectively.

now to determine temperature ladders with a maximal  $r_{\text{DEO}}$ . Here we assume that a temperature range  $[T_{\text{min}}, T_{\text{max}}]$  for an RE simulation and a system with known heat capacity  $C$  are given. For our benchmark system, we find the maximum of  $r_{\text{DEO}}$  at the acceptance rate  $p_{\text{acc}} = 40.5\%$ . This prediction closely agrees with our MC measurements (Figs. 1, 3) which identified the maximal  $r$  at the ladder size  $N = 20$  corresponding to an average acceptance probability  $p_{\text{acc}} = 41.4\%$ .

In a previous study [20], we found empirically for the same benchmark system that  $r$  is quite well approximated by the expression

$$r_k \equiv c \kappa(p_{\text{acc}}) \approx c' p_{\text{acc}} / [N(p_{\text{acc}}) - 1] \quad (15)$$

with system dependent constants  $c$  and  $c'$ . From Eqs. (15), (3), and (5) one finds that the maximum of  $r_k$  is at  $p_{\text{acc}} = 44.4\%$ , which is also close to the MC value  $p_{\text{acc}} = 41.4\%$ .

As we have seen, the two apparently different analytical approximations Eqs. (14) and (15) yield quite similar predictions for the optimal  $p_{\text{acc}}$ . This puzzling result raises questions, which are resolved by Fig. 5. The figure shows the ratio  $r_k/r_{\text{DEO}} = c'N(p_{\text{acc}})(1 - p_{\text{acc}})$  as a function of  $p_{\text{acc}}$ . It demonstrates that for large  $p_{\text{acc}}$  the two expressions become equivalent and remain close at smaller  $p_{\text{acc}}$ . This similarity is more pronounced for very small systems (dashed line) than for larger ones (solid line). We would like to stress that the solid line is representative for all large systems with heat capacities  $C > 500$  (data not shown).

By Eq. (15),  $r_k$  is proportional to  $1/(N - 1)$ , which is characteristic for a directed motion with constant velocity, whereas  $r_{\text{DEO}}$  is by Eq. (14) proportional to  $1/[N(N - 1)]$ , which is typical for a diffusive motion. If the temperature ladders become so small that the EPs of the DEO random walk hit the ladder boundaries very frequently, then the approximation Eq. (14) is expected to become worse. For such small ladders one can expect that the constant velocity expression Eq. (15) becomes more accurate. To check this expectation, we calculated the  $p_{\text{acc}}$  of maximal  $r$  by Eqs. (14) and Eq. (15) for the small sys-

tem ( $C = 50$ ). For this system the constant velocity expression Eq. (15) predicts 42.6%, the diffusion expression Eq. (14) yields 43.5% whereas the target value obtained through MC is 41.5% (corresponding to a ladder size  $N = 7$ ). Thus, even for very small temperature ladders our analytical approximation  $r_{\text{DEO}}$  still predicts the location of the maximal round trip rate at nearly the same accuracy as the educated guess  $r_k$  put forward in Ref. [20], while it is somewhat better for larger systems.

## Summary

We have investigated the round trip rate performance of four different exchange algorithms. Our results demonstrate that the DEO algorithm yields the highest round rates over a wide range of average acceptance probabilities  $p_{\text{acc}}$ . Thus, using the DEO algorithm is not only a common but also a good practice in the application of replica exchange. This conclusion not only applies to replica exchange but also to simulated tempering [15]. Note, however, that according to our results (Fig. 1) the APE algorithm [16] might be an interesting alternative if one is interested in high round trip rates using minimally sized temperature ladders [20].

Examining the DEO random walk analytically we have shown that the reason for the higher DEO round trip rates compared with those of its randomized variant (SEO) is the intrinsically higher diffusivity of the corresponding random walk. Since the diffusivity advantage of DEO over SEO becomes larger with increasing  $p_{\text{acc}}$ , the acceptance probability maximizing the round trip rate is shifted from  $p_{\text{acc}} \approx 20\%$  (SEO) to  $p_{\text{acc}} \approx 40\%$  (DEO).

Finally, we have shown that the educated guess Eq. (15), which empirically had been determined to be a good optimization measure for the DEO round trip rates [20], yields results nearly equivalent to those of our analytical approximation Eq. (14).

Note that the DEO scheme does not permit a reverse move immediately after a successful replica swap. Therefore, it does not satisfy detailed balance and one may ask whether DEO can interfere with canonical sampling. However, Manousiouthakis and Deem [23] have shown that the equilibrium statistics, which is generated by the intermittent MD or MC simulation, is preserved if the sampling procedure satisfies a less strict "balance condition". This condition holds if each individual exchange trial satisfies local detailed balance, which is implied by the Metropolis criterion Eq. (1) for even and for odd exchanges. Thus, DEO represents a valid sampling strategy.

## Acknowledgement

This work was supported by the Deutsche Forschungsgemeinschaft (SFB 533/C1 and SFB 749/C4).

## References

- [1] K. Hukushima, K. Nemoto, J. Phys. Soc. Jpn. 65 (1996) 1604.
- [2] U. H. E. Hansmann, Chem. Phys. Lett. 281 (1997) 140.
- [3] Y. Sugita, Y. Okamoto, Chem. Phys. Lett. 314 (1999) 141.

- [4] D. J. Earl, M. W. Deem, *Phys. Chem. Chem. Phys.* 7 (2005) 3910.
- [5] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, E. Teller, *J. Chem. Phys.* 21 (1953) 1087.
- [6] R. Denschlag, M. Lingenheil, P. Tavan, *Chem. Phys. Lett.* 458 (2008) 244–248.
- [7] A. Kone, D. A. Kofke, *J. Chem. Phys.* 122 (2005) 206101.
- [8] S. Trebst, M. Troyer, U. H. E. Hansmann, *J. Chem. Phys.* 124 (2006) 174903.
- [9] W. Nadler, U. H. E. Hansmann, *Phys. Rev. E* 75 (2007) 026109.
- [10] W. Nadler, J. H. Meinke, U. H. E. Hansmann, *Phys. Rev. E* 78 (2008) 061905.
- [11] W. Nadler, U. H. E. Hansmann, *J. Phys. Chem. B* 112 (2008) 10386.
- [12] T. Okabe, M. Kawata, Y. Okamoto, M. Mikami, *Chem. Phys. Lett.* 335 (2001) 435.
- [13] M. J. Abraham, J. E. Gready, *J. Chem. Theory Comput.* 4 (2008) 1119.
- [14] D. van der Spoel, E. Lindahl, B. Hess, A. R. van Buuren, E. Apol, P. J. Meulenhoff, D. P. Tieleman, A. L. T. M. Sijbers, K. A. Feenstra, R. van Drunen, H. J. C. Berendsen, *Gromacs User Manual version 3.3*, [www.gromacs.org](http://www.gromacs.org) (2005).
- [15] R. Denschlag, M. Lingenheil, P. Tavan, G. Mathias, *J. Chem. Theory Comput.*, *submitted*.
- [16] P. Brenner, C. R. Sweet, D. VonHandorf, J. A. Izaguirre, *J. Chem. Phys.* 126 (2007) 074103.
- [17] F. Calvo, *J. Chem. Phys.* 123 (2005) 124106.
- [18] B. Paschek, H. Nymeyer, A. E. Garcia, *J. Struct. Biol.* 157 (2007) 524–533.
- [19] Y. Okamoto, M. Fukugita, T. Nakazawa, H. Kawai, *Protein Eng.* 4 (1991) 639–647.
- [20] R. Denschlag, M. Lingenheil, P. Tavan, *Chem. Phys. Lett.* 473 (2009) 244.
- [21] C. W. Gardiner, *Handbook of Stochastic Methods*, 2nd Edition, Springer, Berlin, 1985.
- [22] W. Nadler, U. H. E. Hansmann, *Phys. Rev. E* 76 (2007) 065701(R).
- [23] V. I. Manousiouthakis, M. W. Deem, *J. Chem. Phys.* 110 (1999) 2753.



## 4 Stabilität von H1 in Umgebungen unterschiedlicher Polarität

Time shall unfold what plighted  
cunning hides!

---

*Cordelia [166]*

Nach vielen, ausführlich diskutierten methodischen Entwicklungen können wir nun endlich auch eine relevante Anwendung in Angriff nehmen. Hierbei handelt es sich um eine Simulation des Faltungs-Entfaltungsgleichgewichts des Modell-peptids huH1, welches der Sequenz von Helix 1 des menschlichen Prion-Proteins entspricht. Dazu ist im folgenden der Artikel

Martin Lingenheil, Robert Denschlag, Paul Tavan

„Highly polar environments catalyze the unfolding of PrP<sup>C</sup> helix 1“,

abgedruckt, den ich zusammen mit Robert Denschlag und Paul Tavan verfasst und zur Veröffentlichung beim *European Biophysics Journal* eingereicht habe. Anhand von insgesamt acht REST-Simulationen wird die von Hirschberger *et al.* [79] aufgeworfene Frage untersucht, ob ein Wechsel der H1-Helix in eine Umgebung hoher Polarität ihre Entfaltung bewirken kann. In der Tat zeigt sich für huH1, dass das Peptid in Wasser mehrheitlich ungefaltet vorliegt und schnell entfaltet, während die Entfaltung in Umgebungen zunehmender Polarität zunehmend länger dauert.

# Highly polar environments catalyze the unfolding of PrP<sup>C</sup> helix 1.

M. Lingenheil · R. Denschlag · P. Tavan

Received: date / Accepted: date

**Abstract** The first  $\alpha$ -helix (H1) likely plays an important role in the conversion of the cellular prion protein (PrP<sup>C</sup>) into its pathogenic isoform (PrP<sup>Sc</sup>). In this conversion, H1 may either have to unfold or may represent a site of intermolecular contact. A recent molecular dynamics simulation suggested that H1 can unfold if it is detached from the protein core [Hirschberger et al. (2006), *Biophys. J.* 90:3908]. It has been hypothesized that the high dielectric constant  $\epsilon_S$  of the bulk water environment facilitates the unfolding of H1. To check this hypothesis, we performed a number of replica exchange molecular dynamics simulations of an H1 peptide in solvents of different  $\epsilon_S$ . We found that the equilibrium helix fraction in water is less than 40% in agreement with previous experimental findings and that the helix unfolds much faster in water than in less polar solvents. The kinetically stabilizing effect of the organic solvents is largely unspecific and correlates well with their dielectric constant  $\epsilon_S$ .

**Keywords** molecular dynamics · peptide studies · secondary structure · prion protein · helix 1 · dielectric constant

## Introduction

Transmissible spongiform encephalopathies (TSEs) or prion diseases, like the Creutzfeldt-Jakob disease, the Gerstmann-Sträussler-Scheinker syndrome, and the fatal familial insomnia in humans or the bovine spongi-

form encephalopathy, scrapie, and the chronic wasting disease in animals, are characterized by a refolding of the cellular prion protein PrP<sup>C</sup> into the pathological isoform PrP<sup>Sc</sup> (Prusiner 1998). In contrast to other known mammalian protein aggregation diseases like Alzheimer's and Parkinson's diseases, TSEs are infectious. According to the protein-only hypothesis (Griffith 1967; Prusiner 1982), this infectivity is due to the (auto)catalytic replication of PrP<sup>Sc</sup> by a template-guided misfolding of PrP<sup>C</sup> (Jarrett and Lansbury Jr. 1993; Cohen et al 1994; Eigen 1996).

The overall three-dimensional structure of the monomeric PrP<sup>C</sup> is conserved among mammalian species (Wüthrich and Riek 2001). The protein consists of a largely unstructured N-terminal region, which, in humans, extends to about residue 124, and of a globular C-terminal domain, which is composed of two short  $\beta$ -strands, S1 (residues 128-131) and S2 (residues 161-164), forming an antiparallel  $\beta$ -sheet and of three  $\alpha$ -helices, H1 (residues 144-154), H2 (residues 173-194), and H3 (residues 200-228), with H2 and H3 connected by a disulfide-bond between residues 179 and 214 (Zahn et al 2000). In contrast to the cellular isoform PrP<sup>C</sup>, the infectious isoform PrP<sup>Sc</sup> has hitherto proven unavailable to high-resolution methods of protein structure determination due to its propensity of forming fibrillar aggregates (Wille and Prusiner 1999). Yet, as determined by Fourier-transform infrared (FTIR) and by circular dichroism (CD) spectroscopy, the transition from PrP<sup>C</sup> to PrP<sup>Sc</sup> is marked by a distinct increase in  $\beta$ -structure and a moderate loss of  $\alpha$ -helix content (Pan et al 1993; Caughey et al 1991). These findings imply that part of the structural changes must comprise also the globular domain including at least a partial unfolding of H1, H2, or H3.

---

P. Tavan  
Department für Physik, LMU München  
Oettingenstraße 67, D-80538 München, Germany  
Tel.: +49-89-2180-9220  
Fax: +49-89-2180-9202  
E-mail: tavan@physik.uni-muenchen.de



A recent structure model by Cobb et al (2007) of *in vitro* generated PrP amyloid, which is based on data obtained by electron paramagnetic resonance spectroscopy and site-directed spin labeling as well as by hydrogen/deuterium exchange experiments (Lu et al 2007), ascribes the refolding to H2 and H3. The spiral model (DeMarco and Daggett 2004) for PrP<sup>Sc</sup> protofibrils, which is derived from an altered PrP conformation observed in a molecular dynamics (MD) simulation, excludes the region comprising the three helices completely from major rearrangements and localizes the structural differences between PrP<sup>C</sup> and PrP<sup>Sc</sup> in the N-terminal part of the globular domain, while yet another class of so-called  $\beta$ -helix models (Govaerts et al 2004; Stork et al 2005; Langedijk et al 2006) agrees with the spiral model in that H2 and H3 are largely conserved but, in contrast to the spiral model, require that H1 unfolds during the conversion. Indeed, antibody studies suggest that the H1 region becomes either buried or refolded upon formation of PrP<sup>Sc</sup> (Williamson et al 1998; Pan et al 2005). Subsequent studies with antibodies [see e.g. Pankiewicz et al (2006) and references therein] as well as several other lines of evidence (Norstrom and Mastrianni 2006; Watzlawik et al 2006; Solforosi et al 2007; Yu et al 2007; Hesp et al 2007) suggest that the H1 region is important for the specific binding of PrP<sup>C</sup> or a partially unfolded intermediate to the replicative PrP<sup>Sc</sup> complex.

There is strong evidence that partially unfolded intermediates feature in the conversion of PrP<sup>C</sup> to PrP<sup>Sc</sup> [for a review, see Surewicz et al (2006)]. In particular, it has been hypothesized that the initial step in the conversion of PrP<sup>C</sup> is characterized by the detachment of a subdomain (S1-H1-S2) comprising the two  $\beta$ -strands S1, S2 and the first  $\alpha$ -helix H1 from the relatively stable core (H2-H3) of the globular domain of the protein (Calzolari and Zahn 2003; Schwarzinger et al 2006; Hirschberger et al 2006; Eghiaian et al 2007; De Simone et al 2007). In fact, large scale rigid body motions and loss of tertiary contacts for the H1 region have been observed in a variety of equilibrium (Gspöner et al 2001; De Simone et al 2007) and non-equilibrium (Alonso et al 2002; Levy and Becker 2002; Colacino et al 2006; Pappalardo et al 2007) molecular dynamics (MD) simulations. In particular, in a recent MD simulation (Hirschberger et al 2006) of PrP<sup>C</sup> carrying the M205R mutation, which disturbs the hydrophobic packing of H1 to the H2-H3 core of the globular domain, H1 detached from the H2-H3 scaffold and subsequently unfolded quickly. This observation led Hirschberger et al (2006) to the question whether the two events are connected, i.e., whether the packing of the S1-H1-S2 do-

main to the hydrophobic core of the protein stabilizes H1.

Indeed, the  $\alpha$ -helical stability of H1 seems to be context-dependent (Fernández et al 2008): In pure water, nuclear magnetic resonance (NMR) spectroscopy of short peptides covering the H1 sequence indicates a significant helical propensity for the H1 residues (Sharman et al 1998; Liu et al 1999; Jamin et al 2002; Ziegler et al 2003). However, a well-formed  $\alpha$ -helical conformation which is stable on the NMR time scale would imply that typical patterns of medium-range nuclear Overhauser effects (NOEs) were observed. If mentioned at all, these NOEs were either absent (Sharman et al 1998; Ziegler et al 2003) or rare and weak (Liu et al 1999) in the cited studies. Taken together with CD and FTIR spectroscopy data, which indicate a helical content of less than 25% for various H1 peptides (Sharman et al 1998; Liu et al 1999; Thompson et al 2000; Jamin et al 2002; Tahiri-Alaoui et al 2003; Ziegler et al 2003), this suggests that, in aqueous solution, an  $\alpha$ -helical conformation exists in rapid equilibrium with essentially isoenergetic random coil conformations. Interestingly, for yet another H1 peptide in water, Kozin et al (2001) found no  $\alpha$ -helical structure at all but determined a well-defined  $\beta$ -hairpin motif using CD and NMR spectroscopy.

Megy et al (2004) found that, upon addition of 2,2,2-trifluoroethanol (TFE) as a cosolvent, the  $\beta$ -hairpin peptide studied earlier by Kozin et al (2001) could be refolded into a well-defined  $\alpha$ -helical conformation essentially identical to that of the corresponding sequence portion in PrP<sup>C</sup>. Similarly, significant concentrations of TFE and other organic cosolvents were shown to stabilize the nascent  $\alpha$ -helical conformations of the other H1 peptides (Sharman et al 1998; Liu et al 1999; Thompson et al 2000; Tahiri-Alaoui et al 2003; Ziegler et al 2003), which are largely unstructured in water. Remarkably, the effect of methanol (MeOH) as a cosolvent is comparable to that of TFE, which is well-known for its specific helix-stabilizing effect (Nelson and Kallenbach 1986). In view of all these data, Hirschberger et al (2006) hypothesized that the stabilization of H1 as an  $\alpha$ -helix by organic cosolvents is largely unspecific and probably due to the lower dielectric constant  $\epsilon_S$  of organic solvents. This view nicely agrees with earlier arguments of Morrissey and Shakhnovich (1999), who argued that electrostatic interactions are probably pivotal for the stability of the helical conformation of H1 due to the peptide's unusual abundance of charged residues. Since the dielectric screening of these charges weakens with decreasing  $\epsilon_S$ , the peptide-internal electrostatic interactions will become dominant in low- $\epsilon_S$  environments and so does an electrostatically stabilized  $\alpha$ -helical confor-

mation. If the stability of H1 should actually be dominated by electrostatic interactions, the detachment of H1 from the H2-H3 scaffold could indeed trigger its subsequent unfolding because the effective  $\epsilon_S$  on the surface of the protein's hydrophobic core is smaller than in the bulk solvent.

To check the hypothesis that the stability of the helical conformation of H1 against unfolding mainly depends on the solvent polarity, we performed MD simulations of a 13-residue polypeptide (huH1) covering the sequence portion 144-156 of the human prion protein (huPrP). Here, we used explicit solvent environments with experimental dielectric constants  $\epsilon_S$  ranging from 33 (MeOH) to 78 (water). Previous MD simulations of an H1 peptide in water indicated that the complete unfolding of the helix in water may take more than 90 ns at ambient temperature (Dima and Thirumalai 2004). In particular, the strong interactions of the charged amino acids in huH1 may lead to high energy barriers, which have to be crossed in the process of unfolding. Thus, to increase the number of conformational transitions, we applied replica exchange with solute tempering (Liu et al 2005) (REST), a method which has previously been shown to enhance the convergence of helix folding/unfolding equilibria (Huang et al 2007; Reichold et al 2009). However, even with REST it is notoriously difficult to acquire convergence of the folding/unfolding equilibrium (Denschlag et al 2008). Therefore, as a convergence test and in contrast to previous MD simulations of H1 peptides in water (Dima and Thirumalai 2004; Ji et al 2005), we start our simulations not only from  $\alpha$ -helical but also from unfolded states.

## Methods

### Model peptide

The peptide huH1 serving as a model for H1 in our simulations has the sequence Ac-DYEDRYRENMHR-NH<sub>2</sub>, which is equivalent to the sequence 144-156 of huPrP. The N-terminus is capped with an acetyl group (Ac) and the C-terminus with an amino group (NH<sub>2</sub>) to exclude electrostatic interactions of the termini with charged side chains because such interactions are absent when H1 is integrated into the protein. In the NMR structure of huPrP<sup>C</sup> determined at an acidic pH (Zahn et al 2000), H1 is two residues shorter (residues 144-154) than our huH1 peptide. We decided to simulate the longer sequence 144-156 identified for H1 in the NMR structure at neutral pH (Calzolari and Zahn 2003) because the positive charge of the C-terminal R156 will, in any case, influence the electrostatics of H1 and, in

particular, might contribute to the stability of the helix by favorable interaction with E152 or with the helix dipole (Munoz and Serrano 1994, 1995; Shoemaker et al 1987).

### Molecular mechanics force field

In our MD simulations, the peptide was described by the CHARMM22 force field (MacKerell et al 1998). We applied the CMAP correction (MacKerell et al 2004) to the potentials of the backbone dihedral angles  $\phi$  and  $\psi$  to obtain a better model for the  $\alpha$ -helical region of Ramachandran space than with standard CHARMM22. We simulated the huH1 peptide in three different explicit solvent environments: (i) pure water described by the TIP3P (Jorgensen et al 1983) water model modified as suggested by MacKerell et al (1998) for usage with the CHARMM22 force field, (ii) pure methanol described by the B3 model developed by Walser et al (2000), and (iii) an equimolar mixture of TIP3P water and B3 methanol.

To estimate to what extent the folding/unfolding kinetics of the huH1 peptide can be modified by the viscosity  $\eta$  of the respective solvent, we calculated  $\eta$  for each of the three solvent systems from sample simulations. These simulations were performed at 300 K and covered 2 ns each. All other simulation parameters were identical to those used in the simulations of the solvated peptides described further below. From the trajectories, we calculated the diffusion constants and estimated  $\eta$  using Stokes' law as described by Walser et al (1999). The resulting estimates are listed in Table 1 and show that the viscosities in the three solvent boxes are quite similar. The values for H<sub>2</sub>O and MeOH agree well with published data (Shen and Freed 2002; Walser et al 2000), which were obtained by different approaches. The viscosity  $\eta$  calculated for the mixed solvent interpolates the values of the pure solvents. Since the differences of the viscosities are small (see Tab. 1), a major influence on the folding/unfolding kinetics of the solute peptide can be excluded.

### MD simulation techniques

All simulations were carried out with the MD simulation program EGO-MMVI (Mathias et al 2003). The system geometry was given by a rhombic dodecahedron of  $R_i = 24 \text{ \AA}$  inscription radius and by periodic boundary conditions. We used a multiple time step integration (Eichinger et al 1997) of Newton's equation of motion with a basic time step of 1 fs. The long-range Coulomb interactions were treated by the

**Table 1** Solvent systems:  $N_S$  denotes the number of solvent molecules in the system;  $\eta$  denotes the dynamic viscosity;  $\epsilon_S$  denotes the dielectric constant assumed for the solvent continuum.

solvent	$N_S$	$\eta$ (Pa s)	$\epsilon_S$
H <sub>2</sub> O	2021	$4.9 \cdot 10^{-4}$	78
H <sub>2</sub> O/MeOH	1213	$4.7 \cdot 10^{-4}$	55
MeOH	944	$4.5 \cdot 10^{-4}$	33

combination of structure-adapted multipole expansions (Niedermeier and Tavan 1996) up to quadrupolar order augmented by a moving-boundary reaction-field approach (Mathias et al 2003). Here, the cutoff radius for the explicit evaluation of the electrostatic interactions was  $R_i$ . Beyond this radius, a dielectric continuum was assumed. We chose the dielectric constant  $\epsilon_S$  for this continuum using experimental values (Wohlfarth 2008; Mukherjee and Grunwald 1958) for the respective solvent (see column  $\epsilon_S$  in Table 1). In the case of the mixture H<sub>2</sub>O/MeOH, we assumed that  $\epsilon_S$  is the mean of the dielectric constants of the constituents. The van der Waals interactions were calculated explicitly for distances up to 10 Å; at larger distances, a mean-field approach (Allen and Tildesley 1987) was applied. Covalent bonds involving hydrogen atoms were kept fixed using the M-SHAKE algorithm (Kraeutler et al 2001).

#### Preparation of the simulation systems

A fully  $\alpha$ -helical starting structure of huH1 was generated using the Molden software (Schaftenaar and Noordik 2000) by setting the backbone dihedral angles to the values  $\phi_\alpha = -58^\circ$ ,  $\psi_\alpha = -47^\circ$  for an ideal  $\alpha$ -helix. This starting structure was placed into the centers of the simulation boxes, which were already filled with the respective solvent models. Any solvent molecule closer than 2 Å to the peptide was deleted. The resulting number  $N_S$  of solvent molecules in the respective boxes is given in the second column of Table 1. The generated simulation systems were subjected to an energy minimization followed by three equilibrations of 100 ps duration, in which two Berendsen thermostats (Berendsen et al 1984) (coupling times 0.1 ps) were separately heating the solvent and the peptide, respectively, to 100 K, 200 K, and finally 300 K. Additionally, a Berendsen barostat (Berendsen et al 1984) (coupling time 1 ps) equilibrated the system to ambient pressure. For the following production runs we deactivated the barostat. We also deactivated the peptide thermostat for the production simulations because such a thermostat can interfere with the conformational dynamics of the peptide (Lingenheil et al 2008).

During the equilibration, the C $_\alpha$ -atoms of the peptide were restrained to their initial positions by harmonic potentials with a force constant of 1.0 kcal/(mol/Å<sup>2</sup>). When these restraints were removed in the production runs, the initially folded peptides began to unfold. By visual inspection, we identified completely unfolded conformations in these trajectories which then served as starting structures for the simulations starting from the unfolded state. To refer to a specific simulation, we will use the name of the solvent (cf. Tab. 1) and the letters “f” for a fully folded ( $\alpha$ -helical) start or “u” for a fully unfolded (random coil) start, respectively. Thus, the simulation of the initially fully unfolded huH1 peptide in methanol would be called MeOH/u.

#### Replica exchange with solute tempering (REST)

For each of the prepared simulation systems, 12 copies (replicas) were simulated for 40 ns in parallel at different temperatures (300 K, 317 K, 335 K, 354 K, 374 K, 396 K, 420 K, 446 K, 474 K, 504 K, 536 K, 570 K). In contrast to the more common temperature replica exchange technique, within REST, the potential energies of a given system configuration  $\mathbf{x}$  at the temperatures  $T_i$  are different and are described by

$$E_i(\mathbf{x}) = E_{\text{pp}}(\mathbf{x}) + \sqrt{\frac{T_i}{T_0}} E_{\text{ps}}(\mathbf{x}) + \frac{T_i}{T_0} E_{\text{ss}}(\mathbf{x}), \quad (1)$$

where  $T_0 = 300$  K is the reference temperature, and  $E_{\text{pp}}$ ,  $E_{\text{ps}}$ , and  $E_{\text{ss}}$  are energies associated with the solute-solute, the solute-solvent, and the solvent-solvent interactions, respectively.

According to Eq. (1), only the replica at  $T_i = T_0$  has an unchanged potential energy, and, therefore, only this replica is of direct physical relevance. Note that the scaling of the solute-solvent energy  $E_{\text{ps}}$  with a factor  $\sqrt{T_i/T_0}$  in Eq. (1) deviates from the one which Liu et al (2005) employed in their original implementation of REST. With respect to the electrostatic interactions [including the direct Coulomb interactions and the indirect interactions mediated by the reaction field (Mathias et al 2003)], our choice of the scaling factors in Eq. (1) corresponds to a simple scaling of the solvent charges by a factor  $\sqrt{T_i/T_0}$ , while the form of the energy function (including the dielectric constant of the reaction field continuum) remains unchanged.

We expect that this scaling of the solvent’s partial charges as well as the differences in temperature will influence the dielectric response of a given solvent type. On the one hand, the scaling of the potential energy according to Eq. (1) is designed to make the solvent statistics independent of the temperature (Liu et al 2005).

REST thus conserves the orientational statistics of the solvent dipoles for the different temperatures. On the other hand, the strengths of the solvent dipoles increase with the temperature. We therefore expect that, within REST, the solvent’s dielectric constant at a temperature  $T_i$  is given by  $\epsilon_S(T_i) = \sqrt{T_i/T_0} \epsilon_S(T_0)$  implying that the solvent polarity increases with  $T_i$  along the temperature ladder. Furthermore, we can conclude that the ratio of the dielectric constants for two different solvent types is the same irrespective of the temperature.

During REST, all pairs of replicas which are simulated at neighboring temperatures  $T_i$  and  $T_j$  may exchange their temperatures every two picoseconds with a probability

$$P_{ij} = \min[1, \exp(-\Delta_{ij})] \quad (2)$$

with

$$\Delta_{ij} = (\beta_i - \beta_j) [E_{\text{pp}}(\mathbf{x}_j) - E_{\text{pp}}(\mathbf{x}_i)] + \left( \sqrt{\beta_0 \beta_i} - \sqrt{\beta_0 \beta_j} \right) [E_{\text{ps}}(\mathbf{x}_j) - E_{\text{ps}}(\mathbf{x}_i)]. \quad (3)$$

Here,  $\mathbf{x}_i$  and  $\mathbf{x}_j$  are the system configurations of the replicas  $i$  and  $j$  at the time point of an exchange trial and  $\beta = 1/k_B T$ , where  $k_B$  is Boltzmann’s constant.

The exchange prescribed by Eqs. (2, 3) preferentially transfers configurations of low potential energy toward the bottom rung of the temperature ladder (Huang et al 2007). This bias ensures that the canonical ensembles, which the replicas sample at their respective temperatures, remain undisturbed. Because the helical state of the huH1 peptide is stabilized by strong electrostatic interactions providing an enthalpic barrier to helix unfolding (Morrissey and Shakhnovich 1999; Ziegler et al 2003), the folded state is expected to preferentially occupy the low temperature rungs. Furthermore, because REST can substantially speed up the sampling of systems with large enthalpic barriers (Zuckerman and Lyman 2006; Denschlag et al 2008), the method should be capable of inducing a substantial number of unfolding events on the time scale of several tens of nanoseconds currently accessible to MD simulations with explicit solvent.

Due to the specific form of the REST potential energy functions [Eq. (1)], the potential energy of the solvent does not contribute to the exchange probability (2, 3). Therefore, in REST, only 12 replicas are needed to cover the temperature range between 300 K and 570 K with reasonable average acceptance probabilities for the exchange. We obtained the following average acceptance probabilities: 22% for H<sub>2</sub>O, 25% for H<sub>2</sub>O/MeOH, and 30% for MeOH. A more detailed analysis of the average exchange probabilities between the individual temperature rungs in the various simulation

is available through Fig. 10 and the accompanying text in the supporting information (SI). At comparable average acceptance probabilities, conventional replica exchange would require about 40 replicas to cover the given temperature range [see van der Spoel et al (2005); Nadler and Hansmann (2008); Denschlag et al (2009) for prescriptions on constructing replica exchange temperature ladders for explicit solvent systems].

Note that conventional replica exchange techniques require that the fluctuations of the total potential energy are canonical (Rosta et al 2009). Since the REST exchange probability depends only on the solute energies  $E_{\text{pp}}$  and  $E_{\text{sp}}$ , we solely have to guarantee that these energy terms exhibit canonical fluctuations. We have recently shown that the non-invasive strategy of temperature control applied in the present study leads to canonical fluctuations of the solute’s kinetic energy (Lingenheil et al 2008) suggesting that also the fluctuations of  $E_{\text{pp}}$  and  $E_{\text{sp}}$  are canonical.

#### Identification of helical structure

To identify the degree of helicity of a given peptide configuration, we use two different methods. The first one is based on the DSSP classifier by Kabsch and Sander (1983), which identifies helical structure according to typical H-bonding patterns. For a given peptide configuration, we define the DSSP helicity

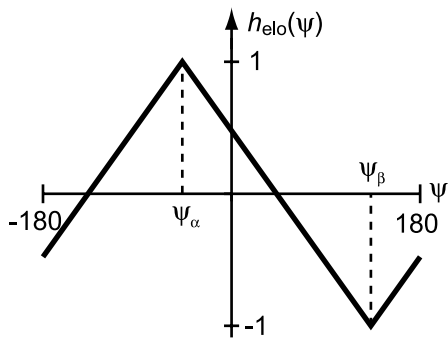
$$H_{\text{DSSP}} \equiv h/h_{\text{max}}, \quad (4)$$

where  $h$  is the number of residues involved in helical structure ( $3_{10}$ -helix,  $\alpha$ -helix, or  $\pi$ -helix) according to DSSP and  $h_{\text{max}}$  is the number of helical residues for an ideal  $\alpha$ -helix. In the case of huH1,  $h_{\text{max}} = 11$ .

A second helicity measure is based on the  $\psi$ -dihedral angles. Residues involved in a helical structure typically have  $\psi$ -angles near  $\psi_\alpha = -47^\circ$ , whereas in an extended structure, the  $\psi$ -angle is typically around  $\psi_\beta = \psi_\alpha + 180^\circ = 133^\circ$ . Fig. 1 shows a simple dihedral  $\alpha$ - $\beta$ -scoring function  $h_{\text{elo}}(\psi)$ , which linearly decreases from 1 to  $-1$  as  $\psi$  changes from  $\psi_\alpha$  to  $\psi_\beta$ . To measure the helicity in a sequence portion  $i \rightarrow j$  of the huH1 peptide comprising the residues from  $i$  to  $j$ , we use the average  $\alpha$ - $\beta$ -score

$$H_{\text{elo}}(i \rightarrow j) = \frac{\sum_{n=i}^j h_{\text{elo}}(\psi_n)}{j - i + 1} \quad (5)$$

of the  $\psi$ -dihedral angles ( $\psi_i, \dots, \psi_j$ ) within the region  $i \rightarrow j$ . Here, well-formed  $\alpha$ -helices are indicated by  $H_{\text{elo}} > 0.8$ , whereas random coil structures are indicated by values of  $H_{\text{elo}}$  near zero. To analyze the general behavior of huH1 in terms of helicity, we will use



**Fig. 1** The dihedral helicity-elongation scoring function  $h_{\text{elo}}(\psi)$ .

$H_{\text{elo}}^{\text{core}} \equiv H_{\text{elo}}(146 \rightarrow 153)$ , which measures the helicity of the eight core residues corresponding to the two central rungs of the helical peptide. To get a structurally more resolved insight, we will furthermore analyze our trajectories in terms of the N-terminal helicity-elongation score  $H_{\text{elo}}^{\text{N}} \equiv H_{\text{elo}}(146 \rightarrow 149)$  and by the C-terminal score  $H_{\text{elo}}^{\text{C}} \equiv H_{\text{elo}}(150 \rightarrow 153)$ . Thus,  $H_{\text{elo}}^{\text{core}}$  is the average value of  $H_{\text{elo}}^{\text{C}}$  and  $H_{\text{elo}}^{\text{N}}$ .

## Results and discussion

We will now first present our REST simulations of huH1 in pure water and scrutinize the statistical reliability of the results. Subsequently, we will report our observations on the peptide’s quite different behavior in the remaining solvents, and, finally, we will examine the contribution of the electrostatic interactions to the differences thus identified.

HuH1 in water is flexible

Our REST simulations H<sub>2</sub>O/f and H<sub>2</sub>O/u started from completely folded and completely unfolded ensembles, respectively, and thus with grossly different initial conditions. One expects that these initial conditions introduce progressively decaying biases into the conformational ensembles sampled during the simulations. Specifically, we expect the H<sub>2</sub>O/f simulation (initially  $H_{\text{elo}}^{\text{core}} = 1.0$ ,  $H_{\text{DSSP}} = 1.0$ ) to be biased toward helical structures and the H<sub>2</sub>O/u simulation (initially  $H_{\text{elo}}^{\text{core}} = -0.57$ ,  $H_{\text{DSSP}} = 0.0$ ) toward extended structures. With increasing duration  $t$  of the simulations, the biasing influence of the initial conditions will decrease, and we correspondingly expect that the average helicities in the two simulations will (in the long run) both converge toward a common value, measuring the equilibrium helicity of our huH1 computer model.

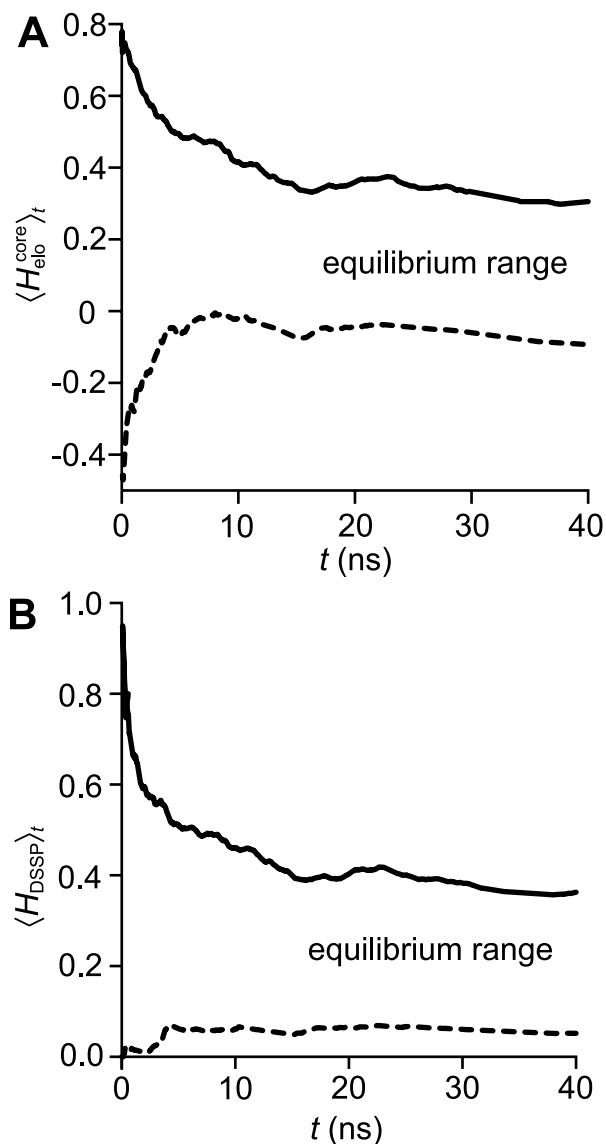
The progressive decay of the initial helicity bias is illustrated by Fig. 2. Part A of the figure shows the cu-

mulative average  $\langle H_{\text{elo}}^{\text{core}} \rangle_t$  of the core helicity-elongation at 300 K as a function of the simulation time  $t$ . Immediately after the equilibration ( $t = 0$ ), all replicas are still close to their respective starting structures. Correspondingly, the H<sub>2</sub>O/f (solid line) and H<sub>2</sub>O/u (dashed line) curves in Fig. 2A start out at  $\langle H_{\text{elo}}^{\text{core}} \rangle_0 \approx 0.8$  and  $\langle H_{\text{elo}}^{\text{core}} \rangle_0 \approx -0.4$ , respectively. During the first 12 ns, the initially folded structures of simulation H<sub>2</sub>O/f continuously loose their  $\alpha$ -helical character, which is reflected by the marked decrease of the solid line in Fig. 2A. For the remainder of the simulation time, the figure indicates a nearly stationary average helicity-elongation score  $\langle H_{\text{elo}}^{\text{core}} \rangle_t \approx 0.35$ . Similarly, the H<sub>2</sub>O/u structures loose part of their bias toward extended conformations early in the simulation, as indicated by the increase of the dashed line in Fig. 2A up to the random coil value of  $\langle H_{\text{elo}}^{\text{core}} \rangle_t \approx 0.0$  during the first 5 ns. For the remaining 35 ns of simulation time, the average helicity-elongation score  $\langle H_{\text{elo}}^{\text{core}} \rangle_t$  remains stationary. The fact that the helical structures in simulation H<sub>2</sub>O/f decay more slowly than the extended structures in simulation H<sub>2</sub>O/u reflects the enthalpic stability provided to the helices by the typical backbone hydrogen bonds and, possibly, additional helix stabilizing mechanisms.

Fig. 2B shows a similar analysis of the water results at 300 K in terms of the average helicity  $\langle H_{\text{DSSP}} \rangle_t$  as measured by DSSP. In agreement with Fig. 2A, also Fig. 2B indicates that the initially  $\alpha$ -helical structures of the H<sub>2</sub>O/f simulation (solid line) progressively decay for about 15 ns followed by a phase of stationarity. Since the DSSP classifier does not distinguish between extended and random coil structures, the H<sub>2</sub>O/u curve (dashed line) does not exhibit a pronounced relaxation phase. Nonetheless, the distinct jump at  $t = 4.5$  ns shows that at least one event of helix folding must have occurred early in the simulation.

The considerable gaps between the solid and dashed curves in Figs. 2A and 2B which remain after 40 ns simulation time are a clear sign of lacking convergence. For a converged sampling of the peptide’s folding/unfolding equilibrium, in Figs. 2A and 2B, the average helicities for H<sub>2</sub>O/f and H<sub>2</sub>O/u would have to be equal and, thus, independent of the starting ensemble. However, since the initial helicity biases of the simulations H<sub>2</sub>O/f and H<sub>2</sub>O/u are extremal, the associated average helicities shown in Fig. 2 are upper and lower bounds for the equilibrium values. Therefore, the equilibrium helicities should be localized somewhere in between the two curves.

According to Fig. 2, the average helicity-elongation of the core residues of huH1 in water is thus between 0.0 and 0.35. These values indicate that the equilibrium ensemble comprises a major random coil component



**Fig. 2** Average helicity at 300 K plotted as a function of the simulation time  $t$  for simulation H<sub>2</sub>O/f (solid lines) and for simulation H<sub>2</sub>O/u (dashed lines). **A:** Cumulative average helicity-elongation score  $\langle H_{\text{elo}}^{\text{core}} \rangle_t$ . **B:** Cumulative average helicity  $\langle H_{\text{DSSP}} \rangle_t$  according to DSSP.

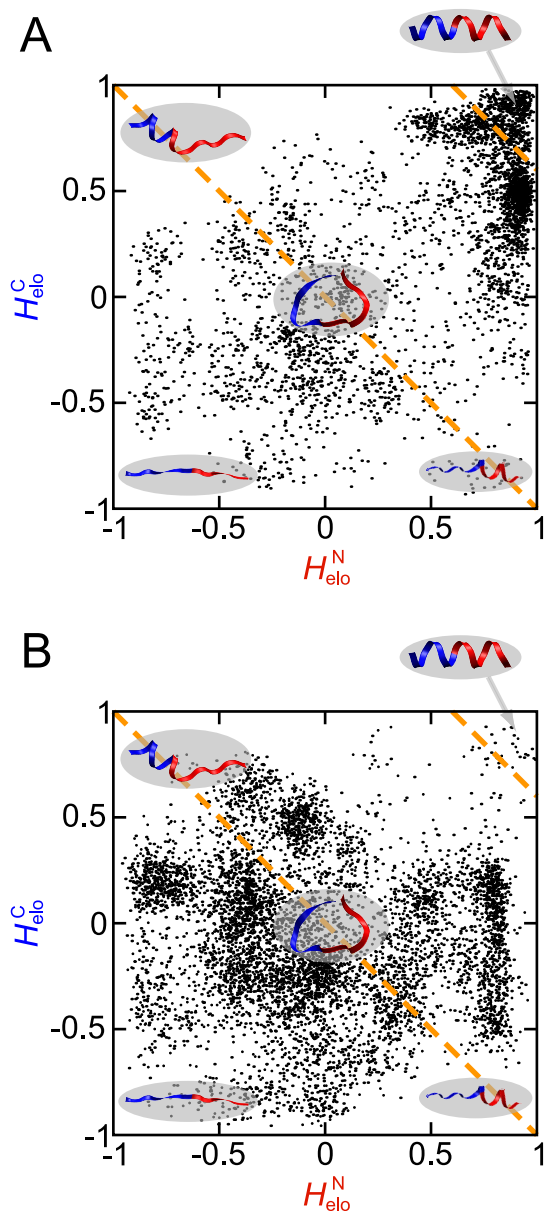
and a minor but probably non-vanishing helical contribution. Similarly, the average equilibrium helicity as classified by DSSP is expected in a range of minimally 7% and maximally 40%. This result nicely agrees with experimental estimates of 7%-30% (Ziegler et al 2003; Fernández et al 2008) for the total helix content of similar H1 peptides in aqueous solution determined by CD spectroscopy. It furthermore agrees with the 34%-42% helical propensities obtained by NMR spectroscopy for the core residues of huH1 in water (Liu et al 1999).

On the other hand, the REST results displayed by Fig. 2 are at variance with MD results of Dima and Thirumalai (2004) on a peptide covering the sequence portion 144-153 of mouse PrP. These authors analyzed their 90 ns MD trajectories by a helicity measure (Klimov and Thirumalai 2003) comparable to our  $H_{\text{elo}}$  score and predicted a helical fraction of about 60%. However, since Dima and Thirumalai (2004) started their simulations with a completely helical peptide and employed standard MD at room temperature, one can expect that their simulation remained even more biased toward the initial helical conformation despite a simulation time as long as 90 ns [see Reichold et al (2009) for an analysis of slow convergence in extended MD simulations of  $\alpha$ -helical peptides].

Encouraged by the apparent agreement of our REST simulations with the available experimental knowledge, we now will use the complete “virtual reality” provided by these simulations to derive a structurally resolved insight into the folding properties of the H1 sequence. To this end, Fig. 3A displays for the 300 K replica of the H<sub>2</sub>O/f simulation the distribution of sampled peptide structures in the plane spanned by the N- and C-terminal helicity-elongation scores  $H_{\text{elo}}^{\text{N}}$  and  $H_{\text{elo}}^{\text{C}}$ , respectively. Each sampled structure is represented by a dot. Additionally, the figure shows prototypical peptide conformations characterizing the various regions of the  $H_{\text{elo}}$ -plane.

The  $\alpha$ -helical starting structure of simulation H<sub>2</sub>O/f is located in the upper right corner of Fig. 3A, which, therefore, is highly populated. In addition to the folded starting structure, the 300 K replica sampled a variety of partially unfolded conformers. Here, the dominant conformation is  $\alpha$ -helical at the N-terminus ( $H_{\text{elo}}^{\text{N}} > 0.8$ ) with a partially unfolded C-terminus ( $H_{\text{elo}}^{\text{C}} \approx 0.4$ ). Furthermore, as Fig. 3A shows, strictly helical values  $H_{\text{elo}}^{\text{N}} > 0.8$  for the N-terminal part of the peptide are frequently combined even with negative values of the C-terminal elongation score  $H_{\text{elo}}^{\text{C}}$ , i.e. with extended C-terminal structures. The inverse situation did not occur in a single instance during the simulation H<sub>2</sub>O/f. Thus, the unfolding of the C-terminal end of the helix does not destabilize the N-terminal part of the helix as much as vice versa. Apart from the helical and partially helical conformations neighboring the starting structure in Fig. 3A, the H<sub>2</sub>O/f simulation sampled a wide variety of random coil and extended structures, which is illustrated by a considerable and widely scattered population of the central part of the  $H_{\text{elo}}$ -plane.

Fig. 3B presents an equivalent diagram for the simulation H<sub>2</sub>O/u. The figure is dominated by a large and widely scattered population occupying the center of the  $H_{\text{elo}}$ -plane and indicating random coil conformations.



**Fig. 3** Distribution of peptide structures in the plane spanned by the N-terminal helicity  $H_{\text{elo}}^{\text{N}}$  (red) and by the C-terminal helicity  $H_{\text{elo}}^{\text{C}}$  (blue) of huH1. Each dot represents a snapshot taken every 60 ps from the respective 300 K replica. The five schematic structures of the peptide backbone are prototypical for the various regions of the  $H_{\text{elo}}$ -plane. The orange dashed lines indicate borders between peptide conformations. The upper right line separates perfect  $\alpha$ -helices ( $H_{\text{elo}}^{\text{core}} > 0.8$ ) from other conformations. The diagonal line separates definitely non-helical structures ( $H_{\text{elo}}^{\text{core}} < 0.0$ ) from conformations with residual helicities. **A:** REST simulation  $\text{H}_2\text{O}/\text{f}$ . **B:** REST simulation  $\text{H}_2\text{O}/\text{u}$ .

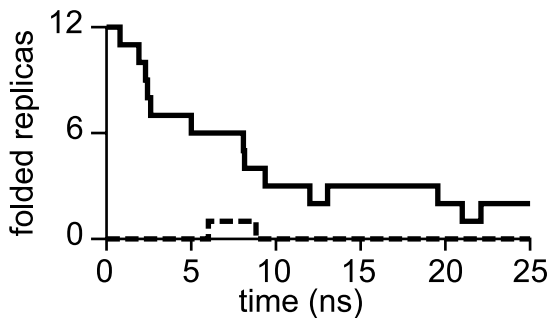
As in Fig. 3A, it is difficult to discern individual clusters of points in this region. Thus, the random coil “conformations” comprise a large number of quickly communicating isoenergetic subconformations. In contrast to the  $\text{H}_2\text{O}/\text{f}$  simulation in Fig. 3A, the immediate neighborhood of the starting structure (lower left corner) is not preferentially populated compared with the random coil conformations in the center of the  $H_{\text{elo}}$ -plane. Thus, the initially extended structure rapidly decays in contrast to the initial  $\alpha$ -helix of Fig. 3A.

The most striking difference between the two simulations  $\text{H}_2\text{O}/\text{f}$  and  $\text{H}_2\text{O}/\text{u}$  is, however, that the  $\alpha$ -helical region of the  $H_{\text{elo}}$ -plane in Fig. 3B is sparsely populated quite in contrast to Fig. 3A. Note furthermore that, in both simulations, we frequently observed structures with a strongly helical N-terminus ( $H_{\text{elo}}^{\text{N}} > 0.8$ ) and an imperfectly folded C-terminal half ( $H_{\text{elo}}^{\text{C}} < 0.4$ ). Within simulation  $\text{H}_2\text{O}/\text{u}$  this intermediate conformation eventually lead in two cases to a fully helical structure. The frequent occurrence of this intermediate is consistent with earlier observations (Liu et al 1999; Alonso et al 2002; Dima and Thirumalai 2004; Hirschberger et al 2006) that the  $\alpha$ -helical propensity of the C-terminal part of H1 is somewhat lower than that of the N-terminal part and suggests that the N-terminal residues form a nucleation site for  $\alpha$ -helix folding.

#### Statistical validity and convergence behavior

The populations of the  $H_{\text{elo}}$ -planes in Figures 3A and 3B differ and show clear signatures of the respective initial conditions. Thus, the comparison of Figs. 3A and 3B once more illustrates that the REST samplings  $\text{H}_2\text{O}/\text{f}$  and  $\text{H}_2\text{O}/\text{u}$  did not reach convergence within the given time spans of 40 ns. We have observed the lacking convergence already in Fig. 2. Here the helicity averages after 40 ns showed a clear memory of the respective initial conditions. Without a comparison of simulations employing different initial conditions, the fact that the helicity averages in Fig. 2 seem to be nearly stationary for the last 25 ns of the simulations could easily be mistaken as a sign of convergence. However, the observed quasi-stationarity is a typical artifact resulting from an incomplete sampling of a folding/unfolding equilibrium using replica exchange techniques (Denschlag et al 2008).

We recently introduced a method to assess the statistical validity of RE simulations of peptide folding/unfolding equilibria by counting the number of folding and unfolding events because convergence requires the (multiple) occurrence of such events (Denschlag et al 2008). Our method is based on the following general observation: In replica exchange simulations, each exchange of



**Fig. 4** The number of folded TE replicas as a function of the simulation time for the simulations H<sub>2</sub>O/f (solid line) and H<sub>2</sub>O/u (dashed line).

configurations introduces a discontinuity into the trajectory of a replica  $i$  which is simulated at the temperature  $T_i$  and whose configuration is modified by the exchange. Therefore, one cannot identify real folding processes by analyzing the trajectories of the replicas at the various temperatures  $T_i$ . Instead one has to consider the continuous trajectories, which move through the temperature space upon exchange and which, for this reason, are termed temperature exchange (TE) replicas. In the given case of the huH1 peptide, we count a folding event if the trajectory of a TE replica makes the complete way from the fully unfolded region ( $H_{\text{elo}}^{\text{core}} < 0.0$ ) to the fully helical region ( $H_{\text{elo}}^{\text{core}} > 0.8$ ) in the  $H_{\text{elo}}$ -plane. The borders of these two domains are indicated by the dashed orange lines in Fig. 3. For an unfolding process to be identified, the TE replica has to make the reverse way from the helical to the non-helical domain.

Fig. 4 presents the number of folded members defined by this criterion within the ensemble of the 12 TE replicas for the simulations H<sub>2</sub>O/f and H<sub>2</sub>O/u. In the case of H<sub>2</sub>O/f (solid line in Fig. 4), most of the initially  $\alpha$ -helical structures decay quickly, leaving only three of the initially 12 helices after 10 ns. Due to laws of statistics and the sorting of the enthalpically favorable folded TE replicas to low temperatures (cf. Methods), the unfolding of the next TE replica takes the more time the less folded TE replicas are left (Denschlag et al 2008). During the 40 ns, the simulation H<sub>2</sub>O/f (solid line) showed only three folding events. One TE replica did not unfold at all during the 40 ns simulation and is responsible for the high population of the helical corner in Fig. 3A. In the case of H<sub>2</sub>O/u (dashed line in Fig. 4), we observed two folding events. Here, a well-formed  $\alpha$ -helix existed for about 3 ns but eventually unfolded again.

Taken together, Fig. 4 indicates that, although neither the H<sub>2</sub>O/f nor the H<sub>2</sub>O/u simulation have sufficiently sampled the folding/unfolding equilibrium to

derive the exact weights of folded and unfolded conformations in the equilibrium ensemble of huH1, our simulations were long enough to observe several folding events. We thus were able to derive upper (40%) and lower (7%) bounds for the helicity of our huH1 model in TIP3P water at 300 K.

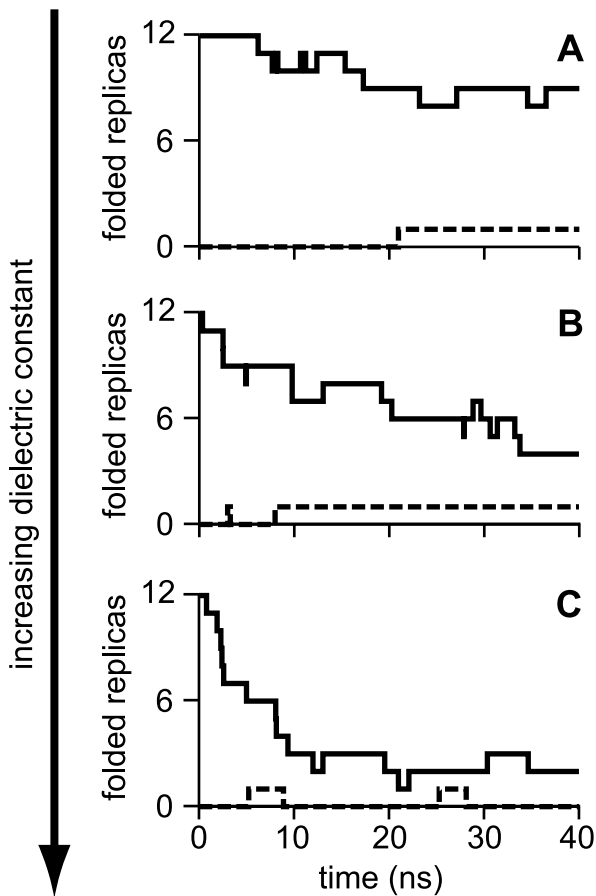
To determine the equilibrium helicity of huH1 in water with greater accuracy, we would have to observe several more folding events. Assuming that the folding time of our computer model of huH1 in the REST setup is somewhere near the 40 ns time span of our simulations, which is plausible given the 2-3 folding events observed in the H<sub>2</sub>O simulations, we would have to simulate well above 100 ns to accurately calculate the equilibrium helicity of huH1 in water. Since such an extensive computational effort would not have been feasible for a number of different explicit solvents environments, and because we were mainly interested in the relative kinetic stabilization of the huH1 helix by environments of different polarity, we abstained from further simulating in H<sub>2</sub>O for the benefit of three other solvent systems.

The huH1 helix is kinetically stabilized by weakly polar solvents

Fig. 5 analyzes the folding/unfolding events of huH1 in different solvents. This analysis is equivalent to the one displayed by Fig. 4 for huH1 in H<sub>2</sub>O. For a simplified comparison, the latter data are repeated in panel C. The panels A, B show the results of the REST simulations with the solvents MeOH and the mixture H<sub>2</sub>O/MeOH, respectively. Thus, the dielectric constant  $\epsilon_S$  (cf. Table 1) of the solvent increases from panel A to C.

As the solid line in panel A indicates, 9 helices remained in the MeOH/f simulation after 40 ns of REST. Thus, for the least polar solvent MeOH, the helical conformation is essentially stable on the time scale of our simulation even at temperatures as high as 500 K. Upon increasing the solvent polarity by an admixture of water (solid line, panel B) considerably less helices (i.e. 4) survived the unfolding simulation. Finally, and in sharp contrast with the MeOH/f results of panel A, in the highly polar water most helices were unfolded after 40 ns with only 2 remaining (solid line, panel C). Thus, in our unfolding simulations, the number of unfolded peptides found after 40 ns of REST simulation monotonously increased with increasing solvent polarity. This result suggests that the unfolding kinetics is accelerated by environments of increasing polarity. The fact that the two helices observed in the H<sub>2</sub>O/u simulation (dashed line in Fig. 5C) unfolded very quickly after





**Fig. 5** The number of folded TE replicas as a function of the simulation time. The line styles distinguish initial conditions: folded (solid) and unfolded (dashed). **A:** MeOH. **B:** H<sub>2</sub>O/MeOH. **C:** H<sub>2</sub>O.

their generation whereas the two newly formed helices in simulations H<sub>2</sub>O/MeOH/u and MeOH/u persisted until the end of the respective simulation further supports our suggestion that helix unfolding seems to be much faster in water than in the other solvents.

With respect to the folding times, the simulations starting from the unfolded ensembles (dashed lines) cannot reveal any significant solvent dependence because the folding of a helix from an unfolded conformation is a rare event on the given 40 ns time scale for all solvents (1 event for MeOH, 2 events for H<sub>2</sub>O/MeOH and H<sub>2</sub>O each). The apparently quite small folding rates are the main reason for the poor convergence of our REST simulations. Convergence would be guaranteed if these rates could be reliably measured from the simulation data (Denschlag et al 2008). However, particularly for the weakly polar solvents, such a measurement would require significantly longer simulation times, which are inaccessible to us. The convergence proved even slower in simulations of huH1 in TFE ( $\epsilon_s = 27$ , data not shown). This is the reason why we have abstained from

a detailed discussion of the results of these simulation, although these results confirm the general trend of increased kinetic helix stability with decreasing  $\epsilon_s$ .

Given the poor convergence particularly for the weakly polar solvents (cf. Fig. 5), one may ask the question as to whether there is anything to be learned from such simulations? Clearly, instead of accurate estimates for thermodynamic quantities such as the ratio between folded and unfolded peptides in the equilibrium ensemble, we can only give more or less extended ranges. Nevertheless, these simulations lead to two additional and clear-cut conclusions.

The first conclusion pertains to the kinetics of huH1 helix unfolding. The discussion of Fig. 5 has suggested that the unfolding of huH1 is accelerated by solvents of increasing polarity. To quantify this acceleration, we calculated the apparent unfolding times  $\tau_u = \langle \tau_i \rangle$  for huH1 in the different solvents. To this end, we determined for every unfolding event  $i$  shown in Fig. 5 the time  $\tau_i$  which the respective TE replica had been folded before. The expected error in  $\tau_u$  was calculated as  $E(\tau_u) = \sigma(\tau_u)/\sqrt{M}$ , where  $M$  is the number of measurements  $\tau_i$  and  $\sigma(\tau_u)$  is their standard deviation. For all replicas found folded at the end of the REST simulations, a lower limit of  $\tau_i = 40 \text{ ns} - t_i$  for the life time was assumed, where  $t_i$  denotes the time of the preceding folding event. Thus, the resulting values for  $\tau_u$  underestimate the actual time needed to unfold a folded helix within our REST setup. This underestimation is the more severe the more replicas are folded at the end of the simulation, which particularly pertains to MeOH (cf. Fig. 5A). Note here that one cannot determine the mean unfolding times at 300 K from the apparent unfolding times  $\tau_u$  observed in our REST simulations because the latter times refer to a temperature-activated process and, therefore, are probably much shorter than the former. However, the ratios of the REST unfolding times in the different solvents should be measures for the corresponding ratios at 300 K.

Table 2 shows the REST unfolding times  $\tau_u$  of huH1 obtained for the various solvents. According to the data,  $\tau_u$  significantly decreases with increasing solvent polarity. For instance, for MeOH the REST unfolding time  $\tau_u$  is at least 1.7 times larger than for water, showing that the unfolding energy barrier is significantly larger in MeOH. It is plausible that the delayed unfolding of the helical conformation of the H1 sequence, i.e. the larger kinetic stability, is also responsible for the enhanced thermodynamic stability of the helix conformation observed experimentally for H1 peptides in organic solvents (Sharman et al 1998; Liu et al 1999; Thompson et al 2000; Tahiri-Alaoui et al 2003; Ziegler et al 2003; Megy et al 2004).

**Table 2** Mean REST unfolding times  $\tau_u$  of huH1 in the different solvents.  $E(\tau_u)$  denotes the estimated error.  $M$  denotes size of the samples entering the calculations.

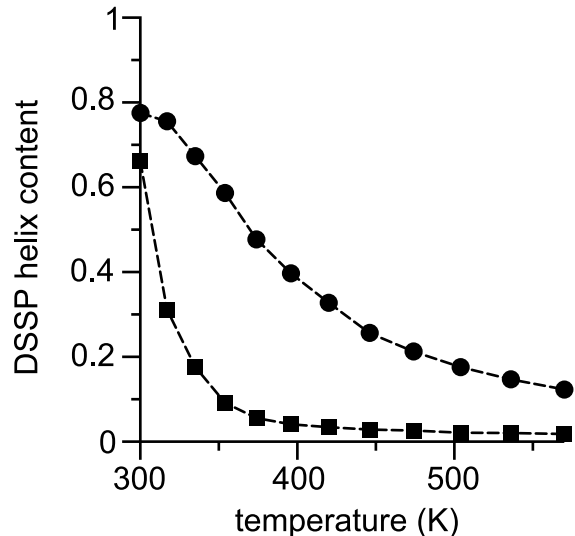
solvent	$\tau_u$ (ns)	$E(\tau_u)$ (ns)	$M$
MeOH	22.7	3.7	18
H <sub>2</sub> O/MeOH	15.0	3.1	19
H <sub>2</sub> O	8.1	2.8	17

As outlined in the Introduction, the main objective of this study was to answer the question whether the  $\alpha$ -helical conformation of the H1 sequence as found in PrP<sup>C</sup> is more likely to unfold, once the S1-H1-S2 domain becomes separated from the hydrophobic core of the protein thereby bringing H1 from an environment with a low dielectric constant into one with a high dielectric constant. The fact that the REST unfolding rates in Tab. 2 strictly increase with increasing solvent polarity indicates that this is indeed the case, if the ratios of the REST unfolding times  $\tau_u$  (cf. Tab. 2) qualitatively represent the corresponding ratios at 300 K.

Helical huH1 has a specifically low potential energy

The second conclusion which can be derived from our REST simulations pertains to the energetics of the huH1 peptide. This energetics is revealed by the distribution of helical conformations along the temperature ladder. As an example, Fig. 6 shows the average DSSP helix fraction observed at the 12 rungs of the temperature ladder during the last 30 ns of the simulations H<sub>2</sub>O/MeOH/f (circles) and H<sub>2</sub>O/MeOH/u (squares). Similar comparisons for the other solvents are presented as in Fig. 11 of the SI. If convergence were achieved, the helicities would have to be independent of the initial conditions. Conversely, the size of the gap between the two curves in Fig. 6 indicates how far our simulations deviate from convergence at the various temperatures. At the 300 K target temperature of the REST sampling, the two curves for H<sub>2</sub>O/MeOH/f and H<sub>2</sub>O/MeOH/u in Fig. 6 are very close to each other indicating a helix fraction between 67% and 79%. Furthermore, in simulation H<sub>2</sub>O/MeOH/u (squares), the helical content of the REST ensemble is essentially confined to the three lowest temperature rungs. Recall that this simulation showed only one folding event leading to a stable helix (cf. the dashed curve in Fig. 5B).

Fig. 7 illustrates the distributions of folded and unfolded states observed during the REST simulation H<sub>2</sub>O/MeOH/u. Fig. 7A pertains to the single TE replica which underwent the folding into a stable helix. The figure shows the random walk of this replica on the temperature ladder and demonstrates that, after fold-



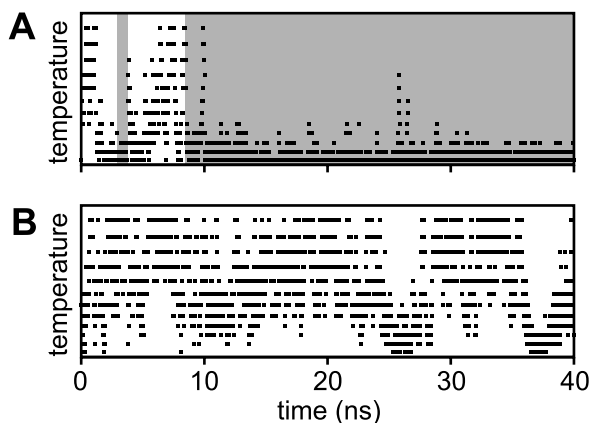
**Fig. 6** Average helical content according to DSSP observed during the final 30 ns of the simulation H<sub>2</sub>O/MeOH/f (circles) and H<sub>2</sub>O/MeOH/u (squares) as a function of the temperature.

ing (gray shaded area), the replica remained confined to the lowest temperature rungs. In contrast, the random walk of an unfolded replica (Fig. 7B) rarely visits these rungs, while moving freely along the remainder of the ladder.

Now recall that the Metropolis exchange criterion given by Eqs. (2, 3) preferentially moves structures with low potential energies to low temperatures. Hence, the pronounced sorting of the helical replica in simulation H<sub>2</sub>O/MeOH/u clearly demonstrates that its helical conformation is characterized by a very low potential energy compared with the unfolded conformations of the other replicas (Huang et al 2007). The strong sorting of helical structures to low temperatures is also evident from the monotonous decrease of the helicity profiles shown for H<sub>2</sub>O/MeOH in Fig. 6 and for the other solvents in Fig. 11 of the SI. Thus, the average potential energy of the  $\alpha$ -helical states must be much lower than those of the unfolded states.

### The role of electrostatic interactions

Various studies have previously indicated that the stability of H1 in PrP and H1 peptides is dominantly caused by electrostatic interactions (Morrissey and Shakhnovich 1999; Ziegler et al 2003; Fernández et al 2008; Megy et al 2004; Dima and Thirumalai 2004). Thus, these interactions are the likely cause for the kinetic stabilization of the huH1 helix. To check this issue, we



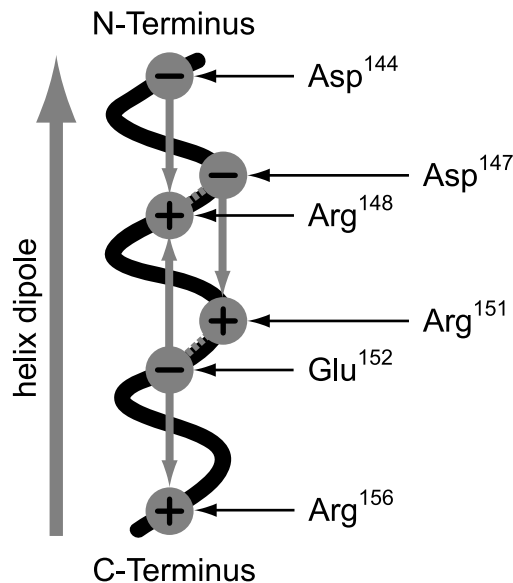
**Fig. 7** The random walks performed by two replicas in the  $\text{H}_2\text{O}/\text{MeOH}/u$  simulation. The position of the replica on the temperature ladder is indicated by a square in intervals of 200 ps. **A:** A TE replica that undergoes folding transitions. The life time of the helix is indicated by a gray background. **B:** A TE replica that remains unfolded during the simulation.

now consider the electrostatics of the folded state in more detail.

Like every  $\alpha$ -helix, also that of huH1 is stabilized by hydrogen bonds within the backbone. However, as schematically shown by Fig. 8, the H1 sequence enables four favorable  $i, i + 4$  salt bridges ( $\text{Asp}^{144}\text{-Arg}^{148}$ ,  $\text{Asp}^{147}\text{-Arg}^{151}$ ,  $\text{Arg}^{148}\text{-Glu}^{152}$ , and  $\text{Glu}^{152}\text{-Arg}^{156}$ ) (Morrissey and Shakhnovich 1999; Ziegler et al 2003; Fernández et al 2008; Megy et al 2004; Dima and Thirumalai 2004; Speare et al 2003; De Simone et al 2007). The total dipole moment generated by the ionic groups in the huH1 helix is directed antiparallel to the macrodipole generated by the the backbone peptide groups and, thus, lowers the electrostatic energy.

The core of the helix is dominated by a quadrupolar motif (Fernández et al 2008) composed of  $\text{Asp}^{147}$ ,  $\text{Arg}^{148}$ ,  $\text{Arg}^{151}$ , and  $\text{Glu}^{152}$ . The associated quadrupole field is short-ranged and, therefore, the solvent's dielectric properties cannot strongly alter the potential energy of the quadrupolar charge group. However, during unfolding, the charges making up the quadrupole must become separated and solvated, which is energetically favored in highly polar solvents. Thus, in weakly polar solvents, the enthalpic barrier towards unfolding should have a large contribution associated with the breaking of the  $i, i + 4$  salt bridges. As a result, the apparent life time of the folded state in our REST simulations (cf. Tab. 2) should correlate with the frequency  $f_{\text{sb}}$  of salt bridges observed in this state.

To check whether such a correlation exists, we analyzed whether the four potential  $i, i + 4$  salt bridges in huH1 were realized whenever the peptide backbone connecting the two residues was classified as  $\alpha$ -helical by

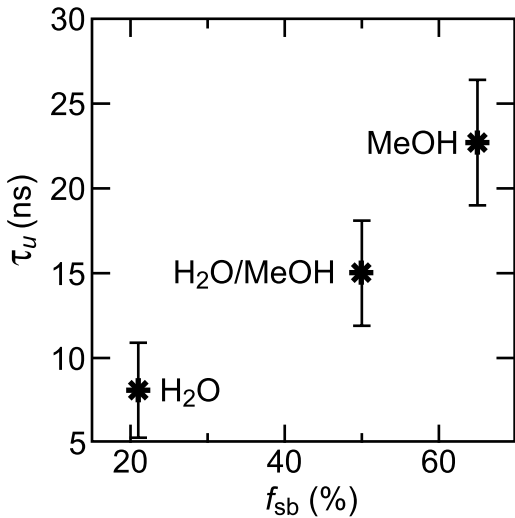


**Fig. 8** Schematic representation of the electrostatic network in the  $\alpha$ -helical conformation of huH1.

DSSP, i.e. if the typical backbone hydrogen bonds were present. The existence of a salt bridge was assumed if the distance  $d$  of the C-atoms in the guanidinium and the carboxylate ionic groups of the respective residues was less than  $5 \text{ \AA}$ . The analysis was performed for every solvent using the 300 K replicas of the simulations starting with folded ensembles.

Fig. 9 demonstrates the expected correlation between the mean REST unfolding time  $\tau_u$  of the helix and the relative frequency  $f_{\text{sb}}$  of the salt bridges in the helical state. In water, the ions are solvated most of the time ( $f_{\text{sb}} = 21\%$ ). Hence, the unfolding of the helix does not additionally require the breaking of salt bridges and the REST unfolding rate is correspondingly large. The solvation of the ions becomes less and less favorable with decreasing solvent polarity. Then, the salt bridges become stable and provide an additional energy barrier to the unfolding of the connecting helical turn. Hence, the sketched mechanism certainly contributes to the observed delayed unfolding of the helix in weakly polar solvents. We conclude that the free energy required for the unfolding transition must be dominated by the enthalpy for the breakup of the intricate network of electrostatic interactions shown in Fig. 8.

To obtain a microscopic insight into the interactions of the peptide with the solvent molecules and, thus, into the reason for the different kinetics, we investigated the solvation of the charged side chains for the different solvents and peptide conformations. Fig. 12 of the SI shows the radial distribution of the solvent around the side chain anions. The figure demonstrates that the sol-



**Fig. 9** The mean REST unfolding time  $\tau_u$  plotted over the relative frequency  $f_{sb}$  of  $i, i + 4$  salt bridges in the helical state.

vation of the anionic side chains in H<sub>2</sub>O is independent of the folding state of the peptide. In MeOH, however, the solvation of the side chains is by about 20% smaller for the unfolded states than for the helix suggesting that additional *intra*-peptide salt bridges replace some solute-solvent interactions upon unfolding.

To check whether the formation of salt bridges is actually more common in the unfolded than in the folded state, we calculated for H<sub>2</sub>O and MeOH and the two conformational states the average numbers of salt bridges considering all possible (i.e. 12) combinations of oppositely charged residues. Further above we had seen that in MeOH and in the helical state the probability for the formation of the typical  $i, i + 4$  salt bridges is very high (65%, cf. Fig. 9). Our counting has now shown that the non-helical configurations in MeOH exhibit an average of 3.5 salt bridges whereas in the helical state only 2.6 salt bridges are formed. In contrast, for H<sub>2</sub>O, the average number of salt bridges is always small with 0.7 (non-helical) and 0.8 (helical), respectively. Consequently, in MeOH the non-helical configurations are even more strongly stabilized by salt bridges than the helical ones. The prevalence of many enthalpically stabilized unfolded conformations is a likely reason for the slow folding/unfolding dynamics in MeOH and thus for the slow convergence toward equilibrium.

#### Temperature mixing

The strong sorting of folded replicas toward the lowest rungs of the REST temperature ladder illustrated by

**Table 3** Average number  $N_r$  of round trips per TE replica observed for the various simulations.

folded		unfolded	
solvent	$N_r$	solvent	$N_r$
H <sub>2</sub> O	21	H <sub>2</sub> O	23
H <sub>2</sub> O/MeOH	5	H <sub>2</sub> O/MeOH	9
MeOH	3	MeOH	5

Fig. 7 may have reduced the efficiency of our REST simulations. For REST to be optimally efficient, all replicas should cycle as rapidly as possible from low to high temperatures and back again (Katzgraber et al 2006; Huang et al 2007; Nadler et al 2008). To gain some clarity with respect to this point, we checked the mixing behavior of our simulations by calculating the average number  $N_r$  of round trips performed by the TE replicas for the various simulations. The results of these calculations are given in Table 3 and demonstrate that the mixing is strongly impeded in the weakly polar solvents compared with water. Also,  $N_r$  is always larger for the simulation starting with an unfolded ensemble than when starting with a folded ensemble.

It is important to note that the weak mixing in H<sub>2</sub>O/MeOH and MeOH does not become evident from a simple inspection of the average exchange probabilities  $p_{acc}$  for the different simulations shown in Fig. 10 in the SI. Although  $p_{acc}$  varies somewhat over the temperature ladder for all solvents, it is not possible to identify any “bottleneck” regions on the ladders with a greatly reduced  $p_{acc}$ .

However, the fact that  $p_{acc}$  is approximately constant and sizeable is not sufficient to ensure good temperature mixing (Katzgraber et al 2006; Nadler et al 2008). Although  $p_{acc}$  in MeOH exceeds the value in H<sub>2</sub>O by nearly a factor of 1.5, the random walks in MeOH still are strongly confined (see Tab. 3) probably due to a large potential energy gap between the peptide’s different conformations (Huang et al 2007). The enhanced formation of salt bridges with decreasing solvent polarity demonstrated above is the most likely reason for this effect, which may explain the weak convergence particularly of our MeOH simulations.

As pointed out by Huang et al (2007), using conventional replica exchange may be superior to REST in such a situation. To still allow for the efficient REST simulation of huH1 and other systems exhibiting such a temperature mixing problem, one should consider the application of feedback-optimized temperature ladders (Katzgraber et al 2006; Nadler et al 2008) or of a modified REST scheme which includes some of the water molecules into the unscaled part of the simulation system (Huang et al 2007).

## Summary

The role of helix 1 in the conversion of the prion protein from the cellular isoform PrP<sup>C</sup> to the scrapie isoform PrP<sup>Sc</sup> has been discussed ever since the NMR structure of PrP<sup>C</sup> has been published. H1 has been suggested to unfold during the conversion of PrP<sup>C</sup> to PrP<sup>Sc</sup> (Gov-aerts et al 2004; Stork et al 2005; Langedijk et al 2006). If an unfolding of H1 is necessary, the stability of the helical conformation of H1 is a key to the question whether H1 constitutes a barrier (Ziegler et al 2003) for the conversion of PrP<sup>C</sup> to PrP<sup>Sc</sup> or rather a weak point (Hirschberger et al 2006) at which this conversion starts. Furthermore, the evidence (Norstrom and Mastrianni 2006; Watzlawik et al 2006; Solforosi et al 2007; Yu et al 2007; Hesp et al 2007) for the H1 region to be involved in the binding of PrP<sup>C</sup> to PrP<sup>Sc</sup> immediately leads to the question whether, during this process, the helix remains firmly folded and, thus, is a passive player or whether it plays an active part due to conformational flexibility.

In our simulations of the huH1 peptide, we observed that the isolated H1 sequence is dominantly non-helical in water. A DSSP analysis of the sampled peptide structures indicates a helical content in a range of 7 – 40% consistent with experimental values for H1 peptides in water (Ziegler et al 2003; Fernández et al 2008). Furthermore, we observed that the helical conformation of huH1 is strongly stabilized against unfolding by weakly polar solvents (such as methanol or TFE), which is also consistent with experimental data (Megy et al 2004; Sharman et al 1998; Liu et al 1999; Thompson et al 2000; Tahiri-Alaoui et al 2003; Ziegler et al 2003). In our REST simulations, an increasing kinetic stabilization of the helical conformation of huH1 was strongly correlated with a decreasing polarity of the environment as measured by the dielectric constant.

The specific distribution of charged residues in the sequence of H1 leads to a very low electrostatic energy of the  $\alpha$ -helical state in weakly polar environments. If the conversion of PrP<sup>C</sup> to PrP<sup>Sc</sup> requires the unfolding of H1, the stability of this helix will indeed represent a barrier as long as it is confined to a low-dielectric environment. However, in a highly polar environment such as bulk water, the efficient screening of the electrostatic interactions reduces the electrostatic stabilization of the helix thereby lowering the unfolding barrier. The unusual electrostatics of the H1 sequence can also explain why H1 peptides are extraordinarily sensitive environmental switches which assume a helical conformation in weakly polar environments and are predominantly non-helical in bulk water (Megy et al 2004; Sharman et al 1998; Liu et al 1999; Thompson et al 2000; Tahiri-

Alaoui et al 2003; Ziegler et al 2003; Fernández et al 2008).

For an H1 sequence embedded into the prion protein the kinetic stability of the helix will also be steered by the polarity of the environment. Thus, any conformational fluctuation which brings H1 from the low-dielectric protein surface into the bulk solvent will catalyze its unfolding. There is a number of reasons to assume that H1 may become separated from the hydrophobic core of PrP<sup>C</sup> early in the conversion process to PrP<sup>Sc</sup> (Calzolari and Zahn 2003; Schwarzinger et al 2006; Hirschberger et al 2006; Eghiaian et al 2007; De Simone et al 2007; Gsponer et al 2001; Alonso et al 2002; Levy and Becker 2002; Colacino et al 2006; Pappalardo et al 2007). Then the helix will face a considerably more polar environment and its unfolding will become much more likely.

In a subsequent step, the unfolded H1 region may take part in a specific interaction with a binding partner (Norstrom and Mastrianni 2006; Watzlawik et al 2006; Solforosi et al 2007; Yu et al 2007; Hesp et al 2007). Here, a potential binding partner would have to offer a peculiar electrostatic environment which can efficiently compete with the  $i, i + 4$  salt bridges potentially promoting the formation of a H1 helix. In this respect, it is interesting to note that the exact distribution of positive and negative charges within H1 seems to be quite important for the conversion efficiency of PrP<sup>C</sup> to the pathogenic PrP<sup>Sc</sup> (Norstrom and Mastrianni 2006).

**Acknowledgements** This work was supported by the Bavarian joint research project ForPrion (LMU02) and by Deutsche Forschungsgemeinschaft (SFB533/C1, SFB749/C4). Access to the resources of the Leibniz-Rechenzentrum HPC Linux Cluster (project UH408) is gratefully acknowledged.

## Electronic supplementary material

Below is the link to the electronic supplementary information (SI) which contains three figures and associated discussions presenting the average acceptance probabilities along the rungs of the temperature ladders during our REST simulations (Fig. 10), temperature resolved helix contents (Fig. 11), and solvent distribution functions near the peptide (Fig. 12).

## References

- Allen MP, Tildesley DJ (1987) Computer Simulations of Liquids. Oxford University Press, Oxford
- Alonso DOV, Cham A, Daggett V (2002) Simulations of biomolecules: characterization of the early steps in the pH-induced conformational conversion of the hamster, bovine

- and human forms of the prion protein. *Phil Trans R Soc Lond A* 360:1165–1178
- Berendsen HJC, Postma JPM, van Gunsteren WF, Di Nola A, Haak JR (1984) Molecular dynamics with coupling to an external bath. *J Chem Phys* 81(8):3684–3690
- Calzolari L, Zahn R (2003) Influence of pH on nmr structure and stability of the human prion protein globular domain. *J Biol Chem* 278(37):35,592–35,596
- Caughey BW, Dong A, Bhat KS, Ernst D, Hayes SF, Caughey WS (1991) Secondary structure-analysis of the scrapie-associated protein prp 27-30 in water by infrared-spectroscopy. *Biochemistry* 30(31):7672–7680
- Cobb NJ, Sönnichsen FD, Mchaourab H, Surewicz WK (2007) Molecular architecture of human prion protein amyloid: A parallel, in-register  $\beta$ -structure. *Proc Natl Acad Sci USA* 104(48):18,946–18,951
- Cohen FE, Pan KM, Huang Z, Baldwin M, Fletterick RJ, Prusiner SB (1994) Structural clues to prion replication. *Science* 264(5158):530–531
- Colacino S, Tiana G, Broglia RA, Colombo G (2006) The determinants of stability in the human prion protein: Insights into folding and misfolding from the analysis of the change in the stabilization energy distribution in different conditions. *Prot Struct Funct Bioinf* 62:698–707
- De Simone A, Zagari A, Derreumaux P (2007) Structural and hydration properties of the partially unfolded states of the prion protein. *Biophys J* 93(4):1284–1292
- DeMarco ML, Daggett V (2004) From conversion to aggregation: Protofibril formation of the prion protein. *Proc Natl Acad Sci USA* 101(8):2293–2298
- Denschlag R, Lingenheil M, Tavan P (2008) Efficiency reduction and pseudo-convergence in replica exchange sampling of peptide folding-unfolding equilibria. *Chem Phys Lett* 458:244–248
- Denschlag R, Lingenheil M, Tavan P (2009) Optimal temperature ladders on replica exchange. *Chem Phys Lett* 473:193–195
- Dima RI, Thirumalai D (2004) Probing the instabilities in the dynamics of helical fragments from mouse PrP<sup>C</sup>. *Proc Natl Acad Sci USA* 101(43):15,335–15,340
- Eghiaian F, Daubenfeld T, Quenet Y, van Audenhaege M, Bouin AP, van der Rest G, Grosclaude J, Rezaei H (2007) Diversity in prion protein oligomerization pathways results from domain expansion as revealed by hydrogen/deuterium exchange and disulfide linkage. *Proc Natl Acad Sci USA* 104(18):7414–7419
- Eichinger M, Grubmüller H, Heller H, P T (1997) Famusamm: an algorithm for rapid evaluation of electrostatic interactions in molecular dynamics simulations. *J Comput Chem* 18:1729–1749
- Eigen M (1996) Prionics or the kinetic basis of prion diseases. *Biophys Chem* 63:A1–A18
- Fernández A, Crespo A, Maddipati S, Scott R (2008) Bottom-up engineering of peptide cell translocators based on environmentally modulated quadrupole switches. *ACS nano* 2(1):61–68
- Govaerts C, Wille H, Prusiner SB, Cohen FE (2004) Evidence for assembly of prions with left-handed  $\beta$ -helices into trimers. *Proc Natl Acad Sci USA* 101:8342–8347
- Griffith JS (1967) Self-replication and scrapie. *Nature* 215(5105):1043–1044
- Gsponer J, Ferrara P, Calfisch A (2001) Flexibility of the murine prion protein and its asp178asn mutant investigated by molecular dynamics simulations. *J Mol Graphics Modell* 20:169–182
- Hesp JR, Raven NDH, Sutton JM (2007) A role for his155 in binding of human prion peptide 144–167 to immobilised prion protein. *Biochem Biophys Res Commun* 362:695–699
- Hirschberger T, Stork M, Schropp B, Winklhofer KF, Tatzelt J, Tavan P (2006) Structural instability of the prion protein upon M205S/R mutations revealed by molecular dynamics simulations. *Biophys J* 90:3908–3918
- Huang XH, Hagen M, Kim B, Friesner RA, Zhou RH, Berne BJ (2007) Replica exchange with solute tempering: Efficiency in large scale systems. *J Phys Chem B* 111(19):5405–5410
- Jamin N, Coic YM, Landon C, Ovtracht L, Baleux F, Neumann JM, Sanson A (2002) Most of the structural elements of the globular domain of murine prion protein form fibrils with predominant beta-sheet structure. *FEBS Lett* 529(2-3):256–260
- Jarrett JT, Lansbury Jr PT (1993) Seeding “one-dimensional crystallization” of amyloid: A pathogenic mechanism in alzheimer’s disease and scrapie? *Cell* 73(6):1055–1058
- Ji HF, Zhang HY, Shen LA (2005) The role of electrostatic interaction in triggering the unraveling of stable helix 1 in normal prion protein. A molecular dynamics simulation investigation. *J Biomol Struct Dyn* 22(5):563–570
- Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. *J Chem Phys* 79(2):926–935
- Kabsch W, Sander C (1983) Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 22(12):2577–2637
- Katzgraber HG, Trebst S, Huse DA, Troyer M (2006) Feedback-optimized parallel tempering monte carlo. *J Stat Mech* P03018:1742–5468
- Klimov DK, Thirumalai D (2003) Dissecting the assembly of a beta(16-22) amyloid peptides into antiparallel beta sheets. *Structure* 11(3):295–307
- Kozin SA, Bertho G, Mazur AK, Rabesona H, Girault JP, Haertle T, Takahashi M, Debey P, Hoa GHB (2001) Sheep prion protein synthetic peptide spanning helix 1 and beta-strand 2 (residues 142-166) shows beta-hairpin structure in solution. *Journal of Biological Chemistry* 276(49):46,364–46,370
- Kraeutler V, van Gunsteren WF, Hünenberger PH (2001) A fast SHAKE algorithm to solve distance constraint equations for small molecules in molecular dynamics simulations. *J Comput Chem* 22(5):501–508
- Langedijk JPM, Fuentes G, Boshuizen R, Bonvin AMJJ (2006) Two-rung model of a left-handed  $\beta$ -helix for prions explains species barrier and strain variation in transmissible spongiform encephalopathies. *J Mol Biol* 360:907–920
- Levy Y, Becker OM (2002) Conformational polymorphism of wild-type and mutant prion proteins: Energy landscape analysis. *Prot Struct Funct Genet* 47(4):458–468
- Lingenheil M, Denschlag R, Reichold R, Tavan P (2008) The “hot-solvent/cold-solute” problem revisited. *J Chem Theory Comput* 4:1293–1306
- Liu AZ, Riek P, Zahn R, Hornemann S, Glockshuber R, Wuthrich K (1999) Peptides and proteins in neurodegenerative disease: Helix propensity of a polypeptide containing helix 1 of the mouse prion protein studied by nmr and cd spectroscopy. *Biopolymers* 51(2):145–152
- Liu P, Kim B, Friesner RA, Berne BJ (2005) Replica exchange with solute tempering: A method for sampling biological systems in explicit water. *Proc Natl Acad Sci USA* 102(39):13,749–13,754
- Lu X, Wintrodde PL, Surewicz WK (2007)  $\beta$ -sheet core of human prion protein amyloid fibrils as determined by hydrogen/deuterium exchange. *Proc Natl Acad Sci USA* 104(5):1510–1515
- MacKerell AD Jr, Feig M, Brooks CL III (2004) Extending the treatment of backbone energetics in protein force fields: Lim-

- itations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J Comput Chem* 25(11):1400–1415
- MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiorkiewicz-Kuczera J, Yin D, Karplus M (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J Phys Chem B* 102:3586–3616
- Mathias G, Egwolf B, Nonella M, Tavan P (2003) A fast multipole method combined with a reaction field for long-range electrostatics in molecular dynamics simulations: The effects of truncation on the properties of water. *J Chem Phys* 118:10,847–10,860
- Megy S, Bertho G, Kozin SA, Debey P, Hoa GHB, Girault JP (2004) Possible role of region 152-156 in the structural duality of a peptide fragment from sheep prion protein. *Protein Sci* 13(12):3151–3160
- Morrissey MP, Shakhnovich EI (1999) Evidence for the role of PrP<sup>C</sup> helix 1 in the hydrophilic seeding of prion aggregates. *Proc Natl Acad Sci USA* 96(20):11,293–11,298
- Mukherjee LM, Grunwald E (1958) Physical properties and hydrogen bonding in the system Ethanol-2,2,2-Trifluoroethanol. *J Phys Chem* 62(10):1311–1314
- Munoz V, Serrano L (1994) Elucidating the folding problem of helical peptides using empirical parameters. *Nature Struct Biol* 1:399–409
- Munoz V, Serrano L (1995) Elucidating the folding problem of helical peptides using empirical parameters. II. helix macrodipole effects and rational modification of the helical content of natural peptides. *J Mol Biol* 245:275–296
- Nadler W, Hansmann UHE (2008) Optimized explicit-solvent replica exchange molecular dynamics from scratch. *J Phys Chem B* 112:10,386–10,387
- Nadler W, Meinke JH, Hansmann UHE (2008) Folding proteins by first-passage-times-optimized replica exchange. *Phys Rev E* 78:061,905
- Nelson JW, Kallenbach NR (1986) Stabilization of the ribonuclease s-peptide  $\alpha$ -helix by trifluoroethanol. *Prot Struct Funct Genet* 1:211–217
- Niedermeier C, Tavan P (1996) Fast version of the structure adapted multipole method-efficient calculation of electrostatic forces in protein dynamics. *Molecular Simulation* 17(1):57–66
- Norstrom EM, Mastrianni JA (2006) The charge structure of helix 1 in the prion protein regulates conversion to pathogenic PrP<sup>Sc</sup>. *J Virol* 80(17):8521–8529
- Pan KM, Baldwin M, Nguyen J, Gasset M, Serban A, Groth D, Mehlhorn I, Huang ZW, Fletterick RJ, Cohen FE, Prusiner SB (1993) Conversion of  $\alpha$ -helices into  $\beta$ -sheets features in the formation of the scrapie prion proteins. *Proc Natl Acad Sci USA* 90:10,962–10,966
- Pan T, Chang BG, Wong P, Li CY, Li RL, Kang SC, Robinson JD, Thompsett AR, Tein P, Yin SM, Barnard G, McConnell I, Brown DR, Wisniewski T, Sy MS (2005) An aggregation-specific enzyme-linked immunosorbent assay: Detection of conformational differences between recombinant prp protein dimers and PrP<sup>Sc</sup> aggregates. *J Virol* 79(19):12,355–12,364
- Pankiewicz J, Prelli F, Sy MS, Kascsak RJ, Kascsak RB, Spinner DS, Carp RI, Meeker HC, Sadowski M, Wisniewski T (2006) Clearance and prevention of prion infection in cell culture by anti-prp antibodies. *Eur J Neurosci* 23(10):2635–2647
- Pappalardo M, Milardi D, Grasso D, La Rosa C (2007) Steered molecular dynamics studies reveal different unfolding pathways of prions from mammalian and non-mammalian species. *New J Chem* 31(6):901–905
- Prusiner SB (1982) Novel proteinaceous infectious particles cause scrapie. *Science* 216(4542):136–144
- Prusiner SB (1998) Prions. *Proc Natl Acad Sci USA* 95:13,363–13,383
- Reichold R, Fierz B, Kiefhaber T, Tavan P (2009) Toward md descriptions of  $\alpha$ -helix folding. *J Chem Theory Comput Submitted*
- Rosta E, Buchete NV, Hummer G (2009) Thermostat artifacts in replica exchange molecular dynamics simulations. *J Chem Theory Comput* 5:1393–1399
- Schaftenaar G, Noordik J (2000) Molden: a pre- and post-processing program for molecular and electronic structures. *J Comput Aid Mol Des* 14:123–134
- Schwarzinger S, Horn AHC, Ziegler J, Sticht H (2006) Rare large scale subdomain motions in prion protein can initiate aggregation. *J Biomol Struct Dyn* 23(6):581–590
- Sharman GJ, Kenward N, Williams HE, Landon M, Mayer RJ, Searle MS (1998) Prion protein fragments spanning helix 1 and both strands of beta sheet (residues 125-170) show evidence for predominantly helical propensity by cd and nmr. *Fold Des* 3(5):313–320
- Shen M, Freed KF (2002) Long time dynamics of Met-Enkephalin: Comparison of explicit and implicit solvent models. *Biophys J* 82:1791–1808
- Shoemaker KR, Kim PS, York EJ, Stewart JM, Baldwin RL (1987) Tests of the helix dipole model for stabilization of  $\alpha$ -helices. *Nature* 326(6113):563–567
- Solfrosi L, Bellon A, Schaller M, Cruite JT, Abalos GC, Williamson RA (2007) Toward molecular dissection of prp-prpc interaction. *J Biol Chem* 282(10):7465–7471
- Speare JO, Rush TS, Bloom ME, Caughey B (2003) The role of helix 1 aspartates and salt bridges in the stability and conversion of prion protein. *J Biol Chem* 278(14):12,522–12,529
- Stork M, Giese A, Kretzschmar HA, Tavan P (2005) Molecular dynamics simulations indicate a possible role of parallel  $\beta$ -helices in seeded aggregation of poly-gln. *Biophys J* 88:2442–2451
- Surewicz WK, Jones EM, Apetri AC (2006) The emerging principles of mammalian prion propagation and transmissibility barriers: Insight from studies in vitro. *Acc Chem Res* 39(9):654–662
- Tahiri-Alaoui A, Bouchard M, Zurdo J, James W (2003) Competing intrachain interactions regulate the formation of beta-sheet fibrils in bovine PrP peptides. *Protein Sci* 12(3):600–608
- Thompson A, White AR, McLean C, Masters CL, Cappai R, Barrow CJ (2000) Amyloidogenicity and neurotoxicity of peptides corresponding to the helical regions of PrP<sup>C</sup>. *J Neurosci Res* 62(2):293–301
- van der Spoel D, Lindahl E, Hess B, van Buuren AR, Apol E, Meulenhoff PJ, Tieleman DP, Sijbers ALTM, Feenstra KA, van Drunen R, Berendsen HJC (2005) Gromacs User Manual version 3.3. [www.gromacs.org](http://www.gromacs.org)
- Walser R, Mark AE, van Gunsteren WF (1999) On the validity of Stoke's law at the molecular level. *Chem Phys Lett* 303:583–586
- Walser R, Mark AE, van Gunsteren WF (2000) The effect of force-field parameters on properties of liquids: Parametrization of a simple three-site model for methanol. *J Chem Phys* 112(23):10,450–10,459
- Watzlawik J, Skora L, Frense D, Griesinger C, Zweckstetter M, Schulz-Schaeffer WJ, Kramer ML (2006) Prion protein helix 1 promotes aggregation but is not converted into beta-sheet. *J Biol Chem* 281(40):30,242–30,250

- 
- Wille H, Prusiner SB (1999) Ultrastructural studies on scrapie prion protein crystals obtained from reverse micellar solutions. *Biophys J* 76:1048–1062
- Williamson RA, Peretz D, Pinilla C, Ball H, Bastidas RB, Rozen-shteyn R, Houghten RA, Prusiner SB, Burton DR (1998) Mapping the prion protein using recombinant antibodies. *J Virol* 72(11):9413–9418
- Wohlfarth C (2008) *Static Dielectric Constants of Pure Liquids and Binary Liquid Mixtures*, Landolt-Börnstein: Numerical Data and Functional Relationships in Science and Technology, vol 17. Springer
- Wüthrich K, Riek R (2001) Three-dimensional structures of prion proteins. *Adv Protein Chem* 57:55–82
- Yu S, Yin S, Li C, Wong P, Chang B, Xiao F, Kang SC, Yan H, Xiao G, Tien P, Sy MS (2007) Aggregation of prion protein with insertion mutations is proportional to the number of inserts. *Biochem J* 403:343–351
- Zahn R, Liu A, Lührs T, Riek R, von Schroetter C, López-García F, Billeter M, Calzolari L, Wider G, Wüthrich K (2000) Nmr solution structure of the human prion protein. *Proc Natl Acad Sci USA* 97(1):145–150
- Ziegler J, Sticht H, Marx UC, Muller W, Rosch P, Schwarzinger S (2003) Cd and nmr studies of prion protein (prp) helix 1 - novel implications for its role in the PrP<sup>C</sup> → PrP<sup>Sc</sup> conversion process. *J Biol Chem* 278(50):50,175–50,181
- Zuckerman DM, Lyman E (2006) A second look at canonical sampling of biomolecules using replica exchange simulation. *J Chem Theory Comput* 2:1200–1202



Supporting information to the manuscript

Highly polar environments  
catalyze the unfolding of PrP<sup>C</sup> helix 1

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

*Theoretische Biophysik, Lehrstuhl für BioMolekulare Optik,  
Ludwig-Maximilians-Universität München*

---

\*Corresponding author. Address: Department für Physik, LMU München, Oettingenstraße 67, D-80538 München, Germany, Tel.: +49-89-2180-9220, Fax: +49-89-2180-9202, e-mail: tavan@physik.uni-muenchen.de

## Average exchange probabilities

To check whether the temperature ladder chosen for the REST simulations enabled sufficiently frequent exchanges between the individual rungs of the temperature ladder, we calculated the average exchange probabilities  $p_{\text{acc}}$  between neighboring temperature rungs for all simulations performed in the present study. Fig. 10A shows  $p_{\text{acc}}$  for the two H<sub>2</sub>O simulations with folded (black circles) and unfolded (red squares) start ensembles. The curves are similar and do not exhibit strong variations along the temperature ladder. In both cases, the smallest  $p_{\text{acc}}$  is observed for the exchange between the lowest two temperature rungs.

A very similar effect can be observed in panel B of Fig. 10 which shows  $p_{\text{acc}}$  for the simulations in the mixed solvent H<sub>2</sub>O/MeOH. Also in this case,  $p_{\text{acc}}$  is fairly homogeneous along the temperature ladder. However, particularly the simulation starting with an unfolded ensemble (red squares) has a distinctly low exchange probability between the lowest two temperature rungs. As demonstrated in the main article (cf. Fig. 7), during this simulation, a single replica had folded and afterwards occupied the lowest temperature rung almost exclusively. The unusually low potential energy of the helical conformation prevents exchanges of this conformation to higher temperature rungs (if these are occupied by non-helical conformations) and thus leads to a reduced  $p_{\text{acc}}$ .

Finally, panel C of Fig. 10 presents the  $p_{\text{acc}}$  values for the simulations in MeOH. Also here, only a moderate variation of  $p_{\text{acc}}$  along the temperature ladder can be seen. For both simulations MeOH/f and MeOH/u, the curves look quite similar apart from a reduced  $p_{\text{acc}}$  for simulation MeOH/f at about 450 K. Interestingly, for simulation MeOH/f this temperature range harbors the interface between temperature rungs mainly occupied by helical configurations and those mainly occupied by non-helical ones (cf. Fig. 11B).

Thus, in summary, it appears that, although the interfaces between the “mainly helical” and “mainly non-helical parts” of the temperature ladder have a somewhat reduced average exchange probability  $p_{\text{acc}}$ , this reduction is not strong enough as to completely

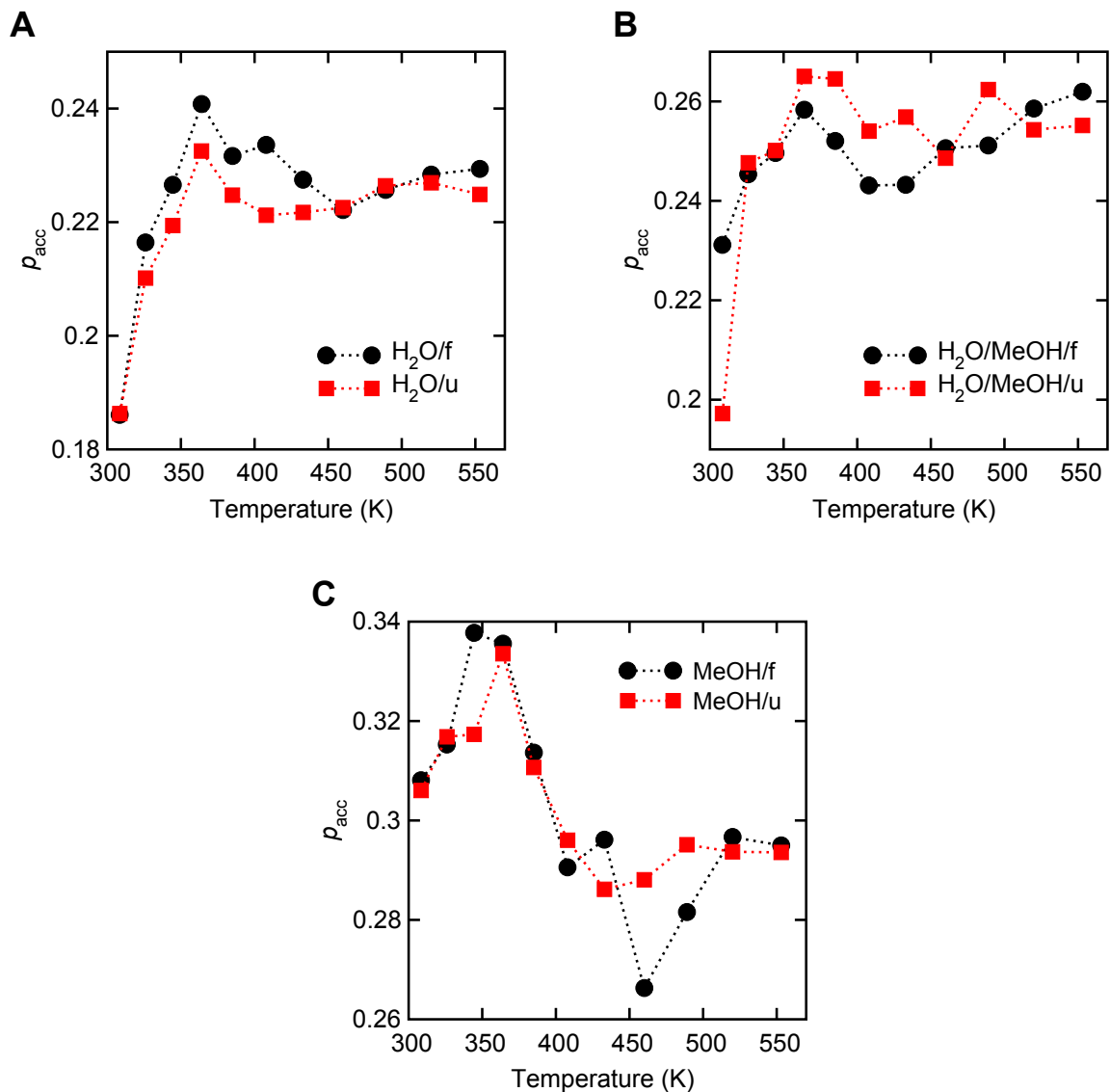


Figure 10: The average REST exchange probabilities  $p_{acc}$  for the individual temperature rungs observed in different solvents: H<sub>2</sub>O (panel A), H<sub>2</sub>O/MeOH (panel B), and MeOH (panel C). Black circles indicate the values for helical start configurations, red squares indicate the values for non-helical start configurations. Although the values vary somewhat, there is no clear indication of any “bottleneck” regions where the exchange probability is low enough to prevent the passing of replicas beyond the bottleneck.

prevent the passing of replicas over the interface region.

## REST helicities for H<sub>2</sub>O, MeOH

Figure 11 presents the average helicities in the REST ensembles for the solvents H<sub>2</sub>O (panel A), and MeOH (panel B). The presentation is equivalent to that in Fig. 6 in the manuscript. The large differences between the curves for an folded start ensemble (circles) and an unfolded start ensemble (squares) show that the simulations are not yet converged. A sorting of helical structures towards low temperatures is observed for the unfolding simulations for H<sub>2</sub>O and MeOH (circles in Figure 11A and B) and for the folding simulation in MeOH (panel B, squares). Since no stable helix was ever observed during the folding simulation H<sub>2</sub>O/u (see Fig. 5, panel C), the helical content is similarly small at all temperatures (squares in Fig. 11A).

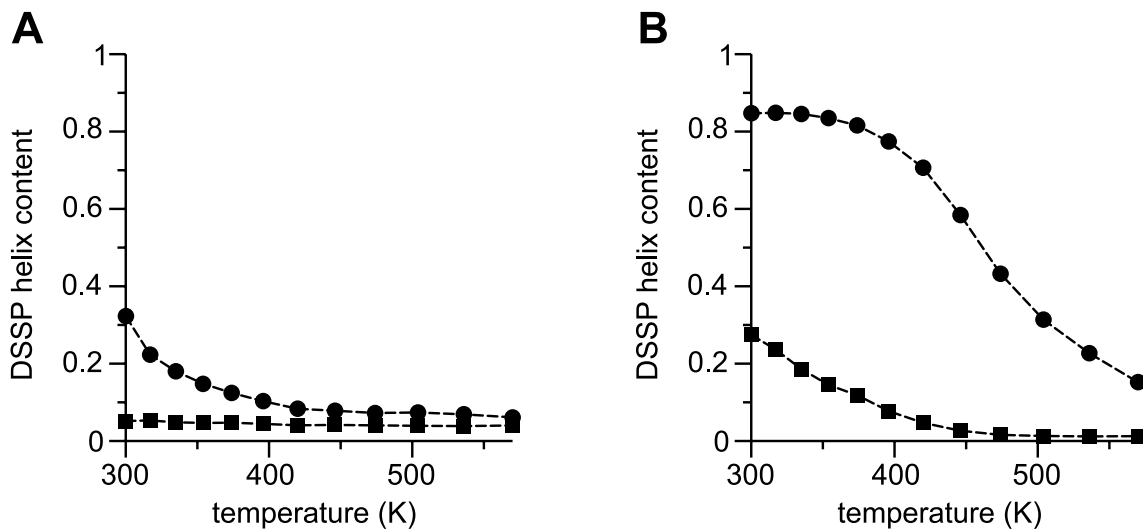


Figure 11: Average helical content according to DSSP observed during the final 30 ns for the unfolding (circles) and the folding simulation (squares) with H<sub>2</sub>O (panel A) and MeOH (panel B).

## Solvation of the backbone and charged side chains

To investigate the reasons for the different behavior of the peptide in the different solvents, we analyzed the solvation of the backbone and of the charged side chains in the different solvents depending on the conformation of the peptide. To this end, we identified for every solvent in the 300 K ensembles of the two simulations (f and u) all  $\alpha$ -helical ( $H_{\text{DSSP}} > 0.8$ ) and all non-helical ( $H_{\text{DSSP}} = 0$ ) configurations. From these sets of configurations, we then calculated the radial density  $g(r)$  of the solvent oxygen atoms around (i) the backbone carbonyl C atoms and (ii) the carboxylate C atoms of the anionic side chains.

Figure 12A compares  $g(r)$  observed in water for the backbone carbonyl C atoms of the helical (black solid line) and non-helical (red dashed line) set of configurations. For both sets of configurations, a first shell of solvent atoms is found at about 3 Å, a second one at 5 Å, and a third one at 7.5 Å. Beyond the third shell, the solvent density is already close to the bulk density. Note that the total solvent density within the range of the first three solvent shells around the backbone carbonyl C atoms is about 30% lower if the peptide is folded into a  $\alpha$ -helix with intact backbone hydrogen bonds than for non-helical configurations. This decreased solvent density expresses the shielding of the backbone from the solvent that is caused by the helical structure.

Figure 12B shows an equivalent plot illustrating the solvation of the backbone dipoles in MeOH. The differences between the solvent structures around the helical (black solid line) and non-helical (red dashed line) configurations are more pronounced for MeOH (panel B) than for H<sub>2</sub>O (panel A). The ordering into shells is stronger for MeOH especially with exposed backbone dipoles (the dashed line in panel B even exhibits a fourth discernable maximum) than for H<sub>2</sub>O (panel A).

The second line of plots in Fig. 12 presents the radial density  $g(r)$  of solvent oxygen atoms around the carboxylate C atoms, thus illustrating the solvation of the anionic side chains. Panel C shows the data for H<sub>2</sub>O as solvent, panel D shows the respective data for MeOH. In both cases, the curves for helical and non-helical configurations are similar. The

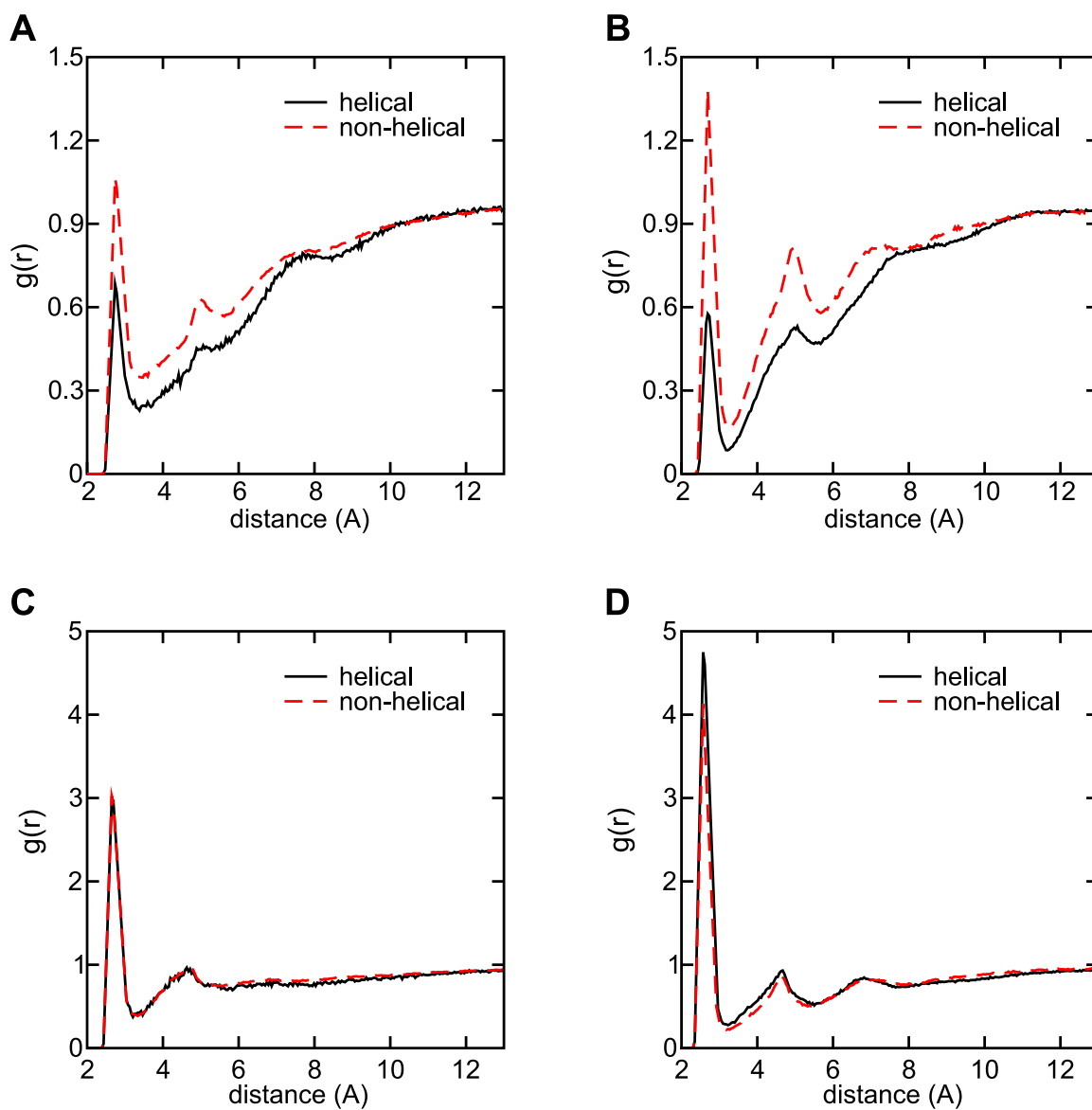


Figure 12: The radial density  $g(r)$  of solvent oxygen atoms as a function of the distance  $r$ . All panels show  $g(r)$  for helical (black solid line) and non-helical (red dashed line) peptide configurations. In the first line of panels, A shows  $g(r)$  for the backbone carbonyl carbon atoms for  $H_2O$ , panel B the corresponding functions for MeOH. The second line of panels shows  $r(g)$  for the carboxylate carbon atoms of the anionic side chains in  $H_2O$  (panel C) and MeOH (panel D).

curves for water in panel C show only two maxima in contrast to the corresponding curves for MeOH in panel D, which exhibit a distinct third maximum. Thus, once again, the solvent structure of MeOH in the near range appears more rigid than the H<sub>2</sub>O structure.

The radial density  $g(r)$  in H<sub>2</sub>O is practically the same for helical (black solid line, panel C) and non-helical (red dashed line, panel C) configurations. In contrast, the data for MeOH in panel D indicate that, in the helical state, the density of solvent molecules in the first shell ( $\sim 3$  Å) around a side chain anion is by about 20% larger than in the non-helical states. This finding suggests that in this less polar solvent the non-helical states feature more intra-peptide salt bridges than the helical state.





## 5 *Résumé* und Ausblick

There are no happy endings,  
because nothing ends.

---

*Schmendrick the Magician [167]*

Am Beginn meines Dissertationsprojekts stand die Idee, durch MD-Simulationen Einblicke in die Ursachen und Mechanismen von Proteinefehlfaltungsprozessen zu gewinnen, deren molekulare Details sich experimentellen Methoden entziehen. Trotz der stetig zunehmenden Leistung moderner Rechner und fortgeschrittener Algorithmen stellt die Simulation solcher Prozesse noch immer eine enorme Herausforderung dar (siehe Abschnitt 1.3).

Zunächst müssen die Umgebungsbedingungen in entsprechenden Simulationen ausreichend genau modelliert werden, da gerade Veränderungen dieser Bedingungen Fehlfaltungsprozesse induzieren können (siehe Abschnitt 1.2.3). In der Einleitung wurde bereits skizziert, dass hierbei insbesondere die Temperatur von Bedeutung ist.

### Temperaturkontrolle

Kapitel 2 meiner Dissertation befasste sich daher mit Strategien zur Temperaturkontrolle bei MD-Simulationen von Makromolekülen in Lösung. Dabei sind zwei Grundprobleme zu lösen. Zum einen führt die Verletzung der Energieerhaltung insbesondere durch die näherungsweise Behandlung der elektrostatischen Wechselwirkungen ohne Gegenmaßnahmen zu einer konstanten Drift der Gesamtenergie des Simulationssystems. Zum anderen können bei inhomogenen Systemen aufgrund der inhomogenen Fehlerverteilung stationäre Temperaturgradienten, das heißt artifizielle Nichtgleichgewichtszustände, entstehen. Demgegenüber sollte das durch die Simulation abgetastete statistische Ensemble möglichst dem in Experimenten vermessenen Ensemble entsprechen. Es sind in der Literatur eine Reihe von sogenannten Thermostatenalgorithmen für MD-Simulationen vorgeschlagen worden, welche die angesprochenen Probleme lösen sollen (einen Überblick gibt Hünenberger [129]).

Die beiden verbreitetsten Thermostatenalgorithmen, der Berendsen-Thermostat (BT) und der Nosé-Hoover-Thermostat (NHT), bildeten Gegenstand der Untersuchungen in Kapitel 2. Problematisch an allen Thermostaten ist, dass sie in die „natürliche“ Newtonsche Dynamik des Systems eingreifen, mit generell schwer

absehbaren Folgen. Dieser Eingriff in die Dynamik des Systems lässt sich anhand des in Kapitel 2 eingeführten Störquotienten quantifizieren. Es zeigte sich, dass die durch einen Nosé-Hoover-Thermostaten verursachte Störung bei gleicher Kopplungsstärke etwa eine Größenordnung über der von einem BT verursachten Störung der Dynamik liegt.

Andererseits führt der Einsatz eines BT im Gegensatz zum NHT zu einer Unterdrückung jenes Teils der kanonischen Energiefluktuationen, die auf längeren Zeitskalen ablaufen, als durch den Kopplungsparameter des Thermostaten vorgegeben ist. Dies kann, wie in Kapitel 2 anhand des Beispiels des Octaalanin-Peptids gezeigt wurde, zu einer verlangsamten Konformationsdynamik führen. Eine solche Behinderung der Konformationsdynamik konnten wir für den NHT nicht feststellen. Daher lässt sich vermuten, dass die durch den BT unterdrückten kanonischen Energiefluktuationen das Überqueren von Energiebarrieren erleichtern.

Wie bereits angedeutet wurde verkompliziert sich bei der Simulation inhomogener Systeme das Problem zusätzlich dadurch, dass die einzelnen Subsysteme jenen algorithmischen Fehlern unterschiedlich stark unterliegen, welche zu einer Energiedrift führen. So erwartet man, dass sich Ungenauigkeiten bei der Behandlung elektrostatischer Wechselwirkungen für die stark polare Wasserumgebung stärker bemerkbar machen, als für ein relativ schwach polares Peptid – etwa Poly-Alanin. Koppelt man ein solches Peptid-Wasser-System mit unterschiedlichen lokalen Heizraten an einen einzigen, globalen Thermostaten, so führt dies zu dem aus der Literatur bekannten *hot-solvent/cold-solute* Problem. Dabei liegen dauerhaft unterschiedliche lokale Temperaturen von Peptid und Umgebung und ein ständiger Wärmefluss zwischen beiden vor. Die übliche Lösung für dieses Problem, ein separater Thermostat für Peptid und Wassersystem, hat zur Folge, dass sich die Störung der Dynamik, im Fall des BT auch die Verzerrung des statistischen Ensembles, nunmehr direkt auf das Peptid auswirken. Die durch einen BT bewirkte verzögerte Konformationsdynamik des Peptids würde insbesondere das effiziente Abtasten der Konformationszustände fehlfaltender Peptide erschweren, die ja gerade das zentrale Ziel dieser Arbeit ist.

Da, wie in Kapitel 2 gezeigt wurde, bereits ein relativ kleines Lösungsmittelsystem ausreicht, um als Wärmebad für das Peptid zu fungieren und dort die kanonischen Energiefluktuationen zu induzieren, ist es bei expliziter Behandlung der Lösungsumgebung nicht notwendig, diese Fluktuationen künstlich mittels eines Thermostatalgorithmus zu erzeugen. Die einzige Aufgabe des Peptidthermostaten ist es daher, die konstante algorithmische Heizrate zu kompensieren. Aus dieser Erkenntnis heraus und nach einer Analyse der Wärmeflüsse in Peptid-Wasser-Systemen formulierten wir zwei Strategien zur Temperaturkontrolle von Makromolekülen in Lösung, die einen kompensierenden Wärmefluss unter minimaler Störung der Dynamik gewährleisten. Ist die thermische Kopplung zwi-

---

schen den Subsystemen stark genug, so genügt es, lediglich das Lösungsmittelsystem an einen Thermostaten zu koppeln und die Kontrolle der Temperatur des Makromoleküls dem thermischen Kontakt mit dem Lösungsmittel zu überlassen. Diese nichtinvasive Strategie führte beispielsweise in der in Kapitel 4 vorgestellten Simulationsstudie zum gewünschten Erfolg. Andernfalls kann eine minimalinvasive Strategie verfolgt werden. Dazu muss die Heizrate für das Makromolekül über Testsimulationen wie in Kapitel 2 bestimmt werden. Ein eigener Thermostat für das Makromolekül kann dann so eingestellt werden, dass er lediglich diese konstante Heizrate kompensiert und ansonsten nur sehr schwach an die Dynamik koppelt.

Bei Benutzung der in Kapitel 2 skizzierten Thermostatenstrategien lässt sich also sowohl die gewünschte Temperatur homogen im gesamten System einstellen, wie auch das korrekte statistische Ensemble für das Makromolekül erzeugen, wobei die dynamische Störung des Moleküls auf ein Minimum beschränkt bleibt. Da bei der Ableitung des Austauschkriteriums (Gleichungen 1.3 und 1.6) für die REST-Technik angenommen wird [155], dass im Zuge der zwischen zwei Austauschen ablaufenden MD-Simulationen das kanonische Ensemble des Makromoleküls bei der jeweiligen Temperatur abgetastet wird, kommt der Erzeugung dieses Ensembles für den Einsatz von REST besondere Bedeutung zu.

Die in Kapitel 2 vorgestellten Strategien sind bisher ausschließlich zur Temperaturkontrolle kleinerer Peptide in Lösung angewendet worden. Hier ist meist die nichtinvasive Strategie ausreichend. Ein systematischer Test der beiden Strategien für die Temperaturkontrolle größerer Systeme, beispielsweise eines ganzen Proteins, steht noch aus. Wünschenswert ist auch eine Weiterentwicklung des Protokolls zur Bestimmung der Heizrate des Makromoleküls für die minimalinvasiven Strategie, da diese bis dato relativ umständlich ist. Obwohl die minimalinvasive Strategie prinzipiell auf beliebig viele Subsysteme mit unterschiedlichen lokalen Heizraten erweiterbar ist, etwa auf Membran-Protein-Wasser-Systeme, wird die Bestimmung der lokalen Heizraten zunehmend komplex, da in der bisherigen Formulierung stets auch die Stärke der thermischen Kopplung zwischen den einzelnen Subsystemen ermittelt werden muss. Daher müsste für dieses Einsatzgebiet zunächst eine Möglichkeit gefunden werden, die lokale Heizrate ohne aufwendige Testsimulationen zu bestimmen.

## Optimierung von *Replica-Exchange*-Strategien

Eine zweite Aufgabe, die es im Lauf meiner Arbeit zu lösen galt, bestand darin, den Einsatz von *replica-exchange*-Techniken (RE-Techniken) für das Abtasten der Konformationszustände fehlfaltender Peptide zu optimieren. Die den RE-Techniken zugrundeliegende Idee ist es, die bei hohen Temperaturen beschleunigte Konformationsdynamik zu nutzen, um in einer MD-Trajektorie schnell neue

Konformationen zu erzeugen. Durch das Wandern der Replikatate können diese Konformationen dann auch in der Stichprobe bei der niedrigsten Temperatur mit dem ihnen im entsprechenden Ensemble zukommenden Gewicht auftauchen (vgl. Abschnitt 1.3.5). Daher hängt der Nutzen von RE-Techniken ganz wesentlich davon ab, dass die einzelnen Replikatate möglichst schnell durch den Temperaturraum wandern [145–149]. Die Effizienz dieses Wanderns wurde in Kapitel 3 anhand der sogenannten *round-trip*-Rate  $r$  gemessen, also der Anzahl an Runden, die ein Replikat pro Zeiteinheit zwischen der niedrigsten und der höchsten Temperatur der RE-Temperaturleiter dreht.

In Abschnitt 3.1 wurde zunächst der Einfluss der Temperaturleiter auf  $r$  untersucht. Dabei beschränkten wir uns in unseren Untersuchungen auf Systeme mit expliziter Berücksichtigung des Lösungsmittels. Ist hier das von der Leiter zu überdeckende Temperaturintervall vorgegeben, so ist die Temperaturleiter eindeutig bestimmt, wenn entweder eine bestimmten Anzahl  $N$  von Sprossen oder eine mittlere Austauschwahrscheinlichkeit  $p_{\text{acc}}$  zwischen den Sprossen vorgegeben wird. Wie in Abschnitt 3.1 nachgewiesen wurde, sind die von Kone und Kofke [145] sowie von Nadler und Hansmann [149] vorgeschlagenen Konstruktionsregeln für Temperaturleitern äquivalent: Diese Regeln sagen eine optimale *round-trip*-Rate  $r$  bei einer mittleren Austauschwahrscheinlichkeit von  $p_{\text{acc}} \approx 23\%$  voraus.

Unter Verwendung eines einfachen Monte-Carlo-Modells, das die Bewegung der Replikatate auf der Temperaturleiter simuliert, stellten wir fest, dass sich eine maximale *round-trip*-Rate  $r_{\text{max}}$  nicht etwa bei  $p_{\text{acc}} \approx 23\%$ , sondern vielmehr bei  $p_{\text{acc}} \approx 45\%$  ergibt. Dieses Ergebnis war auch für uns selbst überraschend, da die bei der Konstruktion ihrer Regeln von Kone und Kofke [145] bzw. von Nadler und Hansmann [149] gemachten Annahmen durchaus plausibel erschienen. Des Rätsels Lösung ergab sich erst später aus der mathematischen Analyse der Zufallsbewegungen unter verschiedenen Austauschschemata, die in Abschnitt 3.2 behandelt wurde. Zunächst erlaubte uns das Monte-Carlo-Modell jedoch bereits die Einführung der empirischen Größe  $\kappa$  (etwa die mittlere auf der Temperaturleiter zurückgelegte Strecke je Zeitschritt), die sich in sehr guter Näherung als linear zu den gemessenen *round-trip*-Raten  $r$  erwies. Die Optimierung von  $\kappa$  optimiert also die automatisch  $r$ .

Die weitere Untersuchung der Eigenschaften von  $\kappa$  als Funktion der mittleren Austauschwahrscheinlichkeit  $p_{\text{acc}}$  ergab, dass  $\kappa(p_{\text{acc}})$  für alle praktisch relevanten Systemgrößen sein Maximum bei  $p_{\text{acc}} \approx 45\%$  annimmt. Für eine bezüglich  $r$  optimale Temperaturleiter müssen also die Abstände der Sprossen in der Temperaturleiter so gewählt werden, dass sich für alle benachbarten Sprossenpaare diese mittlere Austauschwahrscheinlichkeit ergibt. Weiterhin konnten wir zeigen, dass die Leitern nach Kone und Kofke [145] bzw. Nadler und Hansmann [149] näherungsweise  $r/N$  maximieren. Je nach Zielsetzung und verfügbaren Rechen-

---

ressourcen kann die Gewichtung des höheren Rechenaufwands durch eine vergrößerte Anzahl  $N$  von Leitersprossen, die in diesem Optimierungsmaß enthalten ist, durchaus sinnvoll sein.

Die Analysen in Abschnitt 3.1 beschränkten sich auf Simulationssysteme, bei welchen die Lösungsumgebung explizit berücksichtigt wird, da hier üblicherweise keine Phasenübergänge zu erwarten sind [147]. Bei solchen Systemen ist erstens die Wärmekapazität des Systems in guter Näherung konstant [168]. Daher ist es sehr einfach, eine Temperaturleiter zu konstruieren, die zu den gleichen mittleren Austauschwahrscheinlichkeiten  $p_{\text{acc}}$  zwischen den jeweils benachbarten Temperaturstufen führen [169]. Zweitens weisen solche Systeme eine sehr kleine Autokorrelationszeit der potentiellen Energien auf, welche über Gleichung 1.4 in die Akzeptanzwahrscheinlichkeit für die Austausch eingehen. Das führt dazu, dass aufeinanderfolgende Austauschversuche als statistisch unabhängige Zufallsexperimente betrachtet werden können.

Für Simulationssysteme ohne explizites Lösungsmittel oder REST-Systeme brechen diese Annahmen oft zusammen, da einzelne langlebige Peptidkonformationen (vgl. Tab. 1.1) stark unterschiedliche mittlere potentielle Energien aufweisen können. Ein typisches Beispiel bildet hier das in Kapitel 4 behandelte H1-Peptid. Sind die Austausche keine statistisch unabhängigen Ereignisse mehr, dann ist auch das in Abschnitt 3.1 verwendete Monte-Carlo-Modell nicht mehr anwendbar, da es eine Korrelation zwischen aufeinanderfolgenden Austauschen nicht berücksichtigt. Es existieren Vorschläge [170, 171] zur Konstruktion optimaler Temperaturleitern auch für solche Systeme. Diese Vorschläge einer ähnlichen Prüfung wie in Abschnitt 3.1 zu unterziehen, erscheint gerade im Hinblick auf die Anwendung von REST sehr wichtig. Ein erster Schritt hierzu wird in der Konstruktion eines einfachen Testsystems bestehen, das die Korrelation aufeinanderfolgender Austauschversuche beinhaltet.

In Abschnitt 3.2 wurde dann der Einfluss des Austauschschemas auf die *round-trip*-Rate  $r$  untersucht. Üblicherweise wird bei RE für eine bestimmte Temperaturstufe  $i$  abwechselnd versucht, das gerade auf der Stufe  $i$  befindliche Replikat nach oben bzw. nach unten auszutauschen (DEO<sup>1</sup>-Schema). Eine vollständig stochastische Variante dieses Schemas (SEO<sup>2</sup>) ergibt sich, wenn die Richtung des Austauschversuchs nicht alternierend sondern zufällig gewählt wird. Ein anderes, algorithmisch wesentlich komplizierteres Schema (APE<sup>3</sup>) [151] versucht die Diffusion der Replikate auf der Leiter zu verbessern, indem nicht nur Austausche zwischen benachbarten Sprossen der Temperaturleiter sondern alle denkbaren Paarungen in den Austauschversuch miteinbezogen werden. Schließlich wurde im Rahmen des Vergleichs in Abschnitt 3.1 noch ein weiteres vollständig stocha-

---

<sup>1</sup>engl. *deterministic even/odd*

<sup>2</sup>engl. *stochastic even/odd*

<sup>3</sup>engl. *all pair exchange*

stisches Schema (RNN<sup>4</sup>) geprüft.

Erneut benutzten wir Monte-Carlo-Modelle, um die jeweilige Zufallsbewegung unter diesen unterschiedlichen Austausch schemata zu untersuchen. Das DEO-Schema schnitt hierbei deutlich am besten ab. Im Einklang mit unseren Ergebnissen aus Abschnitt 3.1 fanden wir (bei etwas besserer Auflösung) eine maximale *round-trip*-Rate bei einer mittleren Austauschwahrscheinlichkeit von  $p_{\text{acc}} \approx 40\%$ . Die maximale Rate für das SEO-Schema ergab sich bei  $p_{\text{acc}} \approx 23\%$ , also genau bei jener mittleren Austauschwahrscheinlichkeit, die von Kone und Kofke [145] bzw. Nadler und Hansmann [149] als optimal vorausgesagt worden waren.

Wie die mathematische Analyse der Zufallsbewegungen von SEO und DEO in Abschnitt 3.2 gezeigt hat, führt SEO tatsächlich zu einer einfachen, vollständig zufälligen Bewegung eines Replikates auf der Temperaturleiter, wie in den früheren theoretischen Analysen implizit auch für DEO [145, 149] angenommen worden war. Im Gegensatz dazu führt das teilweise deterministische Schema DEO zu einer Zufallsbewegung mit einer inhärent schelleren Diffusion der Replikate auf der Temperaturleiter, welche auch die größere *round-trip*-Rate  $r$  von DEO gegenüber SEO bei ansonsten gleichen Bedingungen erklärt. DEO ist SEO also in der Praxis stets vorzuziehen.

Aus der Analyse der Zufallsbewegungen ließ sich weiterhin ein analytischer Ausdruck für  $r$  als Funktion der mittleren Austauschwahrscheinlichkeit  $p_{\text{acc}}$  herleiten. Die Optimierung von  $r(p_{\text{acc}})$  bezüglich  $p_{\text{acc}}$  ergab schließlich eine optimale DEO-Austauschwahrscheinlichkeit von etwa 40% für alle praktisch relevanten Systemgrößen. Die weitere Analyse der für den DEO-Algorithmus hergeleiteten Funktion  $r(p_{\text{acc}})$  zeigte schließlich auch deren Verwandtschaft zu dem in Abschnitt 3.2 eingeführten Optimierungsmaß  $\kappa(p_{\text{acc}})$ .

Um die Effizienz von RE sicherzustellen, sollten alle Replikate oft genug Gelegenheit gehabt haben, bei hohen Temperaturen neue Peptidkonformationen zu erzeugen und diese im Rahmen ihrer Wanderung in die Stichprobe der Konformationen bei der niedrigsten Temperatur einfließen zu lassen. Vor Beginn jeder RE-Simulation sollte daher geprüft werden [146], ob für das zu simulierende System, die geplante Temperaturleiter und die vorgesehene Simulationszeit überhaupt erwartet werden kann, dass jedes Replikat eine ausreichende Anzahl von Wanderungen zwischen der höchsten und der niedrigsten Temperatur des Systems vollzieht. Genau diese Vorausabschätzung einer wichtigen Kenngröße für die Effizienz von RE ist nunmehr durch den in Abschnitt 3.2 abgeleiteten Ausdruck  $r(p_{\text{acc}})$  für Systeme mit expliziter Berücksichtigung des Lösungsmittels möglich geworden: Eine einfache Multiplikation von  $r$  mit der geplanten Simulationszeit ergibt die voraussichtliche Anzahl an Temperaturzyklen.

Weiterhin zeigen unsere Ergebnisse in Abschnitt 3.2, dass das von Brenner *et*

---

<sup>4</sup>engl. *random next neighbor*

---

al. [151] vorgeschlagene APE Austauschschema für sehr niedrige mittlere Austauschwahrscheinlichkeiten sehr gute Ergebnisse liefert. In der Tat übertrifft APE bei Austauschwahrscheinlichkeiten  $p_{acc} < 10\%$  sogar das DEO-Schema bezüglich  $r$ . Stehen für eine RE-Simulation nur geringe Rechenressourcen zur Verfügung, so lässt sich bei Verwendung des APE-Schemas auch mit einer ausgedünnten Temperaturleiter noch eine zufriedenstellende *round-trip*-Rate erzielen.

Der in Abschnitt 3.2 festgestellte relativ schlechte Erfolg von APE bei größeren Temperaturleitern geht wesentlich darauf zurück, dass dieses Schema seiner Konstruktion nach zwischen zwei MD-Intervallen maximal ein Paar von Replikaten unter allen für den Austausch geprüften auch tatsächlich austauscht. Da APE in seiner ursprünglichen Formulierung stets alle denkbaren Paare von Replikaten einer Temperaturleiter in die Prüfung miteinbezieht, wird auch bei großen Temperaturleitern stets nur höchstens ein Paar von Replikaten ausgetauscht. Diese Beschränkung lässt sich durchbrechen, indem die Temperaturleiter in Abschnitte unterteilt und auf jeden Abschnitt separat ein APE-Algorithmus angewandt wird. Ob ein solches Verfahren bezüglich  $r$  möglicherweise sogar besser abschneidet als das DEO-Schema, lässt sich anhand des von uns in Abschnitt 3.2 vorgestellten Monte-Carlo-Modells prüfen. Erste Versuche in diese Richtung zeigten allerdings, dass es schwierig ist, die Größe der Abschnitte passend zu wählen, um das DEO-Schema tatsächlich übertreffen zu können.

## Abtasten der Faltungszustände eines Modellpeptids

Nach der Implementierung von REST und den oben beschriebenen umfangreichen methodischen Arbeiten, bestand nun der nächste Schritt darin, die verbesserten Methoden bei der Untersuchung von Fehlfaltungsprozessen zur Anwendung zu bringen. Mit Hilfe der verbesserten Simulationstechniken erschien es möglich, die in Abschnitt 1.2.4 bereits angerissene Frage nach den Eigenschaften der Helix 1 (H1) des Prion-Proteins (PrP) und ihrer möglichen Rolle bei der Fehlfaltung von PrP zu beantworten. Zur Untersuchung dieser Frage konnten wir uns in Kapitel 4 auf die Betrachtung eines Modellpeptids (huH1) mit 13 Aminosäureresten beschränken, welches der Sequenz nach H1 im menschlichen Prion-Protein entspricht. Die im Vergleich zum Gesamtprotein viel kleinere Systemgröße von H1 ermöglichte eine Gesamtsimulationszeit von knapp 4  $\mu$ s.

Im wesentlichen sollten mit der in Kapitel 4 vorgestellten Peptidstudie zwei eng verwandte Fragen geklärt werden: Eine Reihe experimenteller Untersuchungen [98, 172–178] hatte ergeben, dass H1-Peptide in Wasser größtenteils nichthelikal vorliegen, dass aber bei Zugabe von Alkoholen zum Lösungsmittel, ähnlich wie bei  $\alpha$ -Synuclein (vgl. Abschnitt 1.2.3), der Anteil gefalteter Peptide deutlich ansteigt. Ein erstes Ziel der in Kapitel 4 durchgeführten Simulationen bestand deshalb darin, anhand des Faltungs/-Entfaltungs-Gleichgewichts von huH1 prototy-

pisch zu prüfen, ob sich die durch eine Veränderung der Lösungsmittelpolarität induzierte Verschiebung der Gewichte unterschiedlicher Peptidkonformationen im Gleichgewichtsensemble mit Hilfe der entwickelten MD-Methoden abbilden lässt.

Zum anderen galt es, die von Hirschberger *et al.* [79] aufgestellte Hypothese zu prüfen, dass H1 durch den Übergang von einer schwach polaren in eine hoch polare Umgebung zur Entfaltung gebracht werden kann. Daher wurde huH1 in vier verschiedenen Lösungsmitteln mit stark unterschiedlicher Polarität simuliert sowie die entsprechende Entfaltungsgeschwindigkeit der helikalen Konformation gemessen. Obwohl die hierbei gemessenen Entfaltungsgeschwindigkeiten wegen der Anwendung von REST und der damit verbundenen Temperaturerhöhung und der Kraftfeldskalierung nicht den Entfaltungsgeschwindigkeiten bei Raumtemperatur entsprechen, kann dennoch davon ausgegangen werden, dass das Verhältnis der REST-Entfaltungsgeschwindigkeiten in den unterschiedlichen Lösungsmitteln zumindest qualitativ dem Verhältnis bei nativen Bedingungen entspricht.

Die in Kapitel 4 präsentierten Simulationsergebnisse zeigen, dass die Entfaltungsgeschwindigkeit von huH1 in reinem Wasser maximal ist und mit abnehmender Polarität der Lösungsumgebung stark abnimmt. Damit bestätigte sich die Hypothese [79], dass die Entfaltung der helikalen Konformation von H1 durch einen Wechsel der Umgebungspolarität katalysiert werden kann. Daraus kann im Bezug auf den Fehlfaltungsprozess von PrP nun der folgende Schluss gezogen werden: Erstens stellt die Entfaltung von H1 nicht notwendigerweise eine Barriere für den Umfaltungsprozess [83, 98] dar, da die kinetische Stabilität der Helix von der Umgebung gesteuert wird. Daher hängt die Antwort auf diese Frage davon ab, ob H1 im Verlauf des Umfaltungsprozesses zeitweise in eine hoch polare Umgebung wechselt. Da außer den MD-Simulationen von Hirschberger *et al.* [79] auch eine Reihe weiterer Beobachtungen die Trennung von H1 vom Rest des Proteins und damit einen solchen Wechsel in der Polarität der Umgebung als ersten Schritt der Fehlfaltung von PrP nahelegen [81–85], ergibt sich aus den Ergebnissen in Kapitel 4 die schnelle Entfaltung von H1 als logische Folge dieses Ereignisses. Die Entfaltung von H1 stellt daher zweitens vermutlich ein sehr frühes Ereignis im Umfaltungsprozess von PrP dar.

Die Ursachen für die kinetische Stabilisierung von huH1 in Lösungsmitteln niedriger Polarität wurden näher untersucht. Ein grober Vergleich mit dem Verhalten eines anderen Peptids deutet darauf hin, dass ein Teil des Effektes spezifisch für die Sequenz von H1 ist. Eine nähere Untersuchung zeigte, dass hier insbesondere die bereits mehrfach erwähnten  $i, i + 4$ -Salzbrücken eine wichtige Rolle spielen, von welchen huH1 gleich vier anzubieten hat. Wie gezeigt werden konnte, werden diese Salzbrücken umso seltener gebildet, je höher die Polarität der Umgebung ist, da vermutlich die ionischen Aminosäurereste, die in einer schwach po-



---

laren Umgebung an der Salzbrücke teilhaben, in einer hoch polaren Umgebung stattdessen effizient solvatisiert werden können.

Die in Kapitel 4 festgestellte Abhängigkeit des Verhaltens von huH1 von den dielektrischen Eigenschaften der Umgebung macht dieses Peptid als Testsystem interessant, um mögliche Artefakte [115–121] durch die Behandlung der elektrostatischen Wechselwirkungen durch Gittersummenmethoden zu untersuchen (siehe Abschnitt 1.3.2). Hierzu wurden von mir bereits eine Reihe von Simulationen durchgeführt, welche die in Kapitel 4 vorgestellten Verfahren einsetzen. Erste Ergebnisse deuten darauf hin, dass bei der Verwendung von Gittersummenmethoden und sehr kleinen Simulationsvolumina die Entfaltung der helikalen Konformation von huH1 behindert wird. Aufgrund der in Kapitel 4 gemachten Beobachtung, dass die kinetische Stabilität von der Polarität der Umgebung gesteuert wird, liegt es nahe zu vermuten, dass die effektive Polarität einer Lösungsumgebung bei der Verwendung von Gittersummenmethoden stark von der Größe des Simulationsvolumens abhängt.

Bezüglich der zweiten Frage, dem Gewicht der helikal gefalteten Strukturen im Gleichgewichtsensemble von huH1, sind die in Kapitel 4 präsentierten Ergebnisse weniger eindeutig. Die Konvergenz der Faltungs-/Entfaltungs-Gleichgewichte in unpolaren Lösungsmitteln ist zum Teil deutlich eingeschränkt. Indem wir jedoch für jedes Lösungsmittel zwei REST-Simulationen durchführten, die mit vollständig  $\alpha$ -helikalen bzw. vollständig ungefalteten Strukturen starteten, ließen sich Ober- und Untergrenzen für den Helix-Gehalt bei 300 K ermitteln. Tatsächlich entsprechen die so ermittelten, relativ gut konvergierten Grenzen (7% – 40%) für den Anteil helikaler Strukturen in Wasser den experimentellen Ergebnissen für vergleichbare H1-Peptide. Weiterhin zeigten unsere Simulationen, dass huH1 in einer Mischung aus Wasser und Methanol zu mindestens 63% in gefalteter Form vorliegt. Nachdem Reichold *et al.* [107, 164] bereits gezeigt hatten, dass es unter Verwendung von REST möglich ist, das Faltungs-/Entfaltungs-Gleichgewicht eines helikalen Peptides mit Methoden der Molekulardynamik konvergent abzutasten, wurde hier nun erstmals nachgewiesen, dass sich auch der Einfluss des Lösungsmittels auf dieses Gleichgewicht zumindest qualitativ abbilden lässt. Die schlechte Konvergenz der Grenzen des Helix-Anteils für das Peptid in den anderen beiden Lösungsmitteln erlaubt freilich keine Aussage darüber, ob allgemein der Anteil helikaler Strukturen am Gleichgewichtsensemble mit sinkender Lösungsmittelpolarität zunimmt, wie sich dies aufgrund der festgestellten kinetischen Stabilisierung vermuten lässt.

Anhand der in Kapitel 4 vorgestellten Simulationsstudie erweist sich, dass trotz des Einsatzes von REST und optimierter Temperaturleitern das Abtasten der verschiedenen Konformationszustände fehlfaltender Peptide und Proteine noch immer eine bedeutende Hürde darstellt. Für Systeme von der Größe von huH1 kann vermutlich durch die zusätzliche Leistung moderner Rechner und den Einsatz

neuer, weiter verbesserter Abtast-Techniken, beispielsweise des von uns kürzlich entwickelten, von REST abgeleiteten SST<sup>5</sup>-Technik [156], ein ausreichend gutes Abtasten der Faltungs-/Entfaltungsgleichgewichte erzielt werden, um routinemäßig quantitative Vergleiche mit experimentellen Observablen zuzulassen.

War es für das relativ kleine huH1-Peptid mit den mir zur Verfügung stehenden Mitteln möglich, ausreichend lange Simulationszeiten zu erreichen, um den Einfluss des Lösungsmittels auf dieses Gleichgewicht in einem Fall nachzuweisen, so ist dies für ganze Proteine, etwa für  $\alpha$ -Synuclein, auf absehbare Zeit undenkbar, solange die Lösungsumgebung explizit berücksichtigt werden muss. Abhilfe versprechen hier neue Ansätze zur impliziten Behandlung der Umgebung als dielektrisches Kontinuum [179–181]. Die im Rahmen dieser Arbeit durchgeführten und in Kapitel 4 vorgestellten Simulationen eignen sich hervorragend als Referenz für den Test eines solchen Kontinuumsmodells, da eine Reihe verschiedener Lösungsmittel mit sehr unterschiedlichen dielektrischen Eigenschaften simuliert wurden und außerdem das huH1-Peptid offensichtlich auch sensibel auf diese Eigenschaften reagiert. Auch hier mag der Einsatz von SST dazu beitragen, die statistische Qualität der von mir bisher generierten Referenzdaten noch einmal deutlich zu verbessern.

---

<sup>5</sup>engl. *simulated-solute-tempering*

# Literaturverzeichnis

Eine neue Doktorarbeit ist  
immer auch eine Orgie von  
alten Doktorarbeiten.

---

Dr. Oztafan Kolibril [1]

- [1] Moers, W. *Rumo und die Wunder im Dunkeln*. Piper, München (2003). [vii](#), [91](#)
- [2] Statistische Jahrbücher für die Bundesrepublik Deutschland und das Ausland. Statistisches Bundesamt, Wiesbaden (1950-2008).  
URL [www.destatis.de](http://www.destatis.de) [1](#)
- [3] 11. Koordinierte Bevölkerungsvorausberechnung – Annahmen und Ergebnisse. Statistisches Bundesamt, Wiesbaden (2006).  
URL [www.destatis.de](http://www.destatis.de) [1](#)
- [4] Prusiner, S. B. Neurodegenerative Diseases and Prions. *New Engl. J. Med.* **344**, 1516–1526 (2001). [1](#)
- [5] Brookmeyer, R., E. Johnson, K. Ziegler-Graham und M. Arrighi. Forecasting the Global Burden of Alzheimer's Disease. *Alzheimer's and Dementia* **3.3**, 186–191 (2007). [1](#)
- [6] Dorsey, E. R., R. Constantinescu, J. P. Thompson, K. M. Biglan, R. G. Holloway *et al.* Projected number of people with Parkinson's disease in the most populous nations, 2005 through 2030. *Neurology* **68**, 384–386 (2007). [1](#)
- [7] Ross, C. A. und M. A. Poirier. Protein aggregation and neurodegenerative disease. *Nat. Med.* **10 Suppl.**, S10–S17 (2004). [1](#)
- [8] Fändrich, M. On the structural definition of amyloid fibrils and other polypeptide aggregates. *Cell. Mol. Life Sci.* **64**, 2066–2078 (2007). [1](#), [7](#), [8](#)
- [9] Bence, N. F., R. M. Sampat und R. R. Kopito. Impairment of the Ubiquitin-Proteasom System by Protein Aggregation. *Science* **292**, 1552–1555 (2001). [2](#)
- [10] Bucciantini, M., E. Giannoni, F. Chiti, F. Baroni, L. Formigli *et al.* Inherent toxicity of aggregates implies a common mechanism for protein misfolding diseases. *Nature* **416**, 507–511 (2002). [2](#)

- [11] Ross, C. A. und M. A. Poirier. What is the role of protein aggregation in neurodegeneration? *Nat. Rev. Mol. Cell Biol.* **6**, 891–898 (2005). [2](#)
- [12] Griffith, J. S. Self-replication and Scrapie. *Nature* **215**, 1043–1044 (1967). [2](#)
- [13] Prusiner, S. B. Novel Proteinaceous Infectious Particles Cause Scrapie. *Science* **216**, 136–144 (1982). [2](#)
- [14] Cohen, F. E., K.-M. Pan, Z. Huang, M. Baldwin, R. J. Fletterick und S. B. Prusiner. Structural Clues to Prion Replication. *Science* **264**, 530–531 (1994). [2](#)
- [15] Hayman, H. und R. Fletcher-Vane. BSE: Cost. *UK Hansard* **624** (2001). [2](#)
- [16] Anderson, R. M., C. A. Donnelly, N. M. Ferguson, M. E. J. Woolhouse, C. J. Watt *et al.* Transmission dynamics and epidemiology of BSE in British cattle. *Nature* **328**, 779–788 (1996). [2](#)
- [17] Donnelly, C. A., N. M. Ferguson, A. C. Ghani und R. M. Anderson. Implications of BSE infection screening data for the scale of the British BSE epidemic and current European infection levels. *Proc. R. Soc. London B* **269**, 2179–2190 (2002). [2](#)
- [18] Will, R. G., J. W. Ironside, M. Zeidler, S. N. Cousens, K. Estibeiro *et al.* A new variant of Creutzfeldt-Jakob disease in the UK. *Lancet* **347**, 921–925 (1996). [3](#)
- [19] Valleron, A. J., P. Y. Boelle, R. Will und J. Y. Cesbron. Estimation of epidemic size and incubation time based on age characteristics of vCJD in the United Kingdom. *Science* **294**, 1726–1728 (2001). [3](#)
- [20] Bruce, M. E., R. G. Will, J. W. Ironside, I. McConnell, D. Drummond *et al.* Transmissions to mice indicate that 'new variant' CJD is caused by the BSE agent. *Nature* **389**, 498–501 (1997). [3](#)
- [21] Hill, A. F., M. Desbruslais, S. Joiner, K. C. L. Sidle, I. Gowland, J. Collinge, L. J. Doey und P. Lantos. The same prion strain causes vCJD and BSE. *Nature* **389**, 448–450 (1997). [3](#)
- [22] Chadeau-Hyam, M. und A. Alperovitch. Risk of variant Creutzfeldt-Jakob disease in France. *Int. J. Epidemiol.* **34**, 46–52 (2005). [3](#)
- [23] Ghani, A. C., C. A. Donnelly, N. M. Ferguson und R. M. Anderson. Updated projections of future vCJD deaths in the UK. *BMC Infect. Dis.* **3**, 4(1–8) (2003). [3](#)

- [24] Ferguson, N. M. und C. A. Donnelly. Assessment of the risk posed by bovine spongiform encephalopathy in cattle in Great Britain and the impact of potential changes to current control measures. *Proc. R. Soc. London B* **270**, 1579–1584 (2003). [3](#)
- [25] Wüthrich, K. und R. Riek. Three-dimensional structures of prion proteins. *Adv. Prot. Chem.* **57**, 55–82 (2001). [3](#), [4](#)
- [26] Collinge, J. Prion diseases of humans and animals: Their causes and molecular basis. *Annu. Rev. Neurosci.* **24**, 519–550 (2001). [3](#)
- [27] Booth, D. R., M. Sunde, V. Bellotti, C. V. Robinson, W. L. Hutchinson *et al.* Instability, unfolding and aggregation of human lysozym variants underlying amyloid fibrillogenesis. *Nature* **358**, 787–793 (1997). [3](#), [11](#)
- [28] Guijarro, J. I., M. Sunde, J. A. Jones, I. D. Campbell und C. M. Dobson. Amyloid fibril formation by an SH3 domain. *Proc. Natl. Acad. Sci. USA* **95**, 4224–4228 (1998). [3](#), [11](#)
- [29] Serpell, L. C., C. C. F. Blake und P. E. Fraser. Molecular structure of fibrillar Alzheimer's A $\beta$  fragment. *Biochemistry* **39**, 13269–13275 (2000). [3](#), [11](#)
- [30] El-Agnaf, O. M. A. und G. B. Irvine. Formation and properties of amyloid-like fibrils derived from  $\alpha$ -synuclein and related proteins. *J. Struct. Biol.* **130**, 300–309 (2000). [3](#), [11](#)
- [31] Bellotti, V., P. Mangione und G. Merlini. Immunoglobulin light chain amyloidosis – The archetype of structural and pathogenic variability. *J. Struct. Biol.* **130**, 280–289 (2000). [3](#), [11](#)
- [32] McParland, V. J., N. M. Kad, A. P. Kalverda, A. Brown, P. Kirwin-Jones, M. G. Hunter, M. Sunde und S. E. Radford. Partially unfolded state of  $\beta$ -(2)-microglobulin and amyloid formation in vitro. *Biochemistry* **39**, 8735–8746 (2000). [3](#), [11](#)
- [33] Lim, A., A. M. Makhov, J. Bond, H. Inouye, L. H. Connors, J. D. Griffith, B. W. Erickson, D. A. Kirschner und C. E. Costello. Betabellins 15D and 16D, *de novo* designed  $\beta$ -sandwich proteins that have amyloidogenic properties. *J. Struct. Biol.* **130**, 363–370 (2000). [3](#), [11](#)
- [34] 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. USA* **47**, 1309–1314 (1961). [3](#), [6](#), [9](#)
- [35] Anfinsen, C. B. Principles that govern the folding of protein chains. *Science* **181**, 223–224 (1973). [3](#), [6](#), [9](#), [10](#)

- [36] Dobson, C. M. Protein folding and misfolding. *Nature* **426**, 884–890 (2003). [3](#), [10](#), [12](#)
- [37] Uversky, V. N. Protein folding revisited. A polypeptide chain at the folding-misfolding-nonfolding cross-roads: Which way to go? *Cell. Mol. Life Sci.* **60**, 1852–1871 (2003). [3](#), [4](#), [10](#), [11](#), [12](#)
- [38] Alberts, B., D. Bray, J. Lewis, M. Raff, K. Roberts und J. D. Watson. *Molecular biology of the cell*. Garland Publishing Inc., New York, 1. Auflage (1983). [4](#)
- [39] Stryer, L. *Biochemie*. Viehweg, Braunschweig, 4. Auflage (1987). [4](#), [11](#)
- [40] Vendruscolo, M., J. Zurdo, C. E. MacPhee und C. M. Dobson. Protein folding and misfolding: A paradigm of self-assembly and regulation in complex biological system. *Phil. Trans. R. Soc. Lond. A* **361**, 1205–1222 (2003). [4](#)
- [41] Wu, H. Studies on denaturation of proteins. XIII. A theory of denaturation. *Chin. J. Physiol.* **5**, 321–344 (1931). [4](#)
- [42] Anson, M. L. und A. E. Mirsky. The equilibrium between active native Trypsin and inactive denatured Trypsin. *J. Gen. Physiol.* **17**, 393–408 (1934). [4](#)
- [43] Lumry, R. und H. Eyring. Conformation changes of proteins. *J. Phys. Chem.* **58**, 110–120 (1954). [4](#)
- [44] Uversky, V. N. Natively unfolded proteins: A point where biology waits for physics. *Protein Sci.* **11**, 739–756 (2002). [4](#), [11](#)
- [45] Fink, A. L. Natively unfolded proteins. *Curr. Opin. Struct. Biol.* **15**, 35–41 (2005). [4](#), [11](#)
- [46] Ulmer, T. S., A. Bax, N. B. Cole und R. L. Nussbaum. Structure and dynamics of micelle-bound human  $\alpha$ -synuclein. *JBC* **280**, 9595–9603 (2005). [4](#)
- [47] Labeit, S. und B. Kolmerer. Titins: Giant Proteins in Charge of Muscle Ultrastructure and Elasticity. *Science* **270**, 293–296 (1995). [4](#)
- [48] Blackmond, D. G. und M. Klussmann. Spoilt for choice: Assessing phase behavior models for the evolution of homochirality. *Chem. Commun.* **39**, 3990–3996 (2007). [5](#)
- [49] Podlech, J. Origin of organic molecules and biomolecular homochirality. *Cell. Mol. Life Sci.* **58**, 44–60 (2001). [5](#)

- [50] Lutz, C. *Strukturelle Stabilität des zellulären Prionproteins unter oxidativem Stress: Eine Simulationsstudie*. Diplomarbeit, Fakultät für Physik, Ludwig-Maximilians-Universität, München (2007). [6](#), [13](#), [22](#), [24](#)
- [51] 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. USA* **37**, 235–240 (1951). [6](#)
- [52] Pauling, L. und R. B. Corey. The pleated sheet, a new layer configuration of polypeptide chains. *Proc. Natl. Acad. Sci. USA* **37**, 251–256 (1951). [6](#)
- [53] Chakrabartty, A. und B. R. L. Stability of  $\alpha$ -helices. *Adv. Prot. Chem.* **46**, 141–176 (1995). [6](#)
- [54] Haire, L. F., S. M. Whyte, N. Vasisht, A. C. Gill, C. Verma, E. J. Dodson, G. G. Dodson und P. M. Bayley. The crystal structure of the globular domain of sheep prion protein. *J. Mol. Biol.* **336**, 1175–1183 (2004). [7](#)
- [55] Sunde, M., L. C. Serpell, M. Bartlam, P. E. Fraser, M. B. Pepys und C. C. F. Blake. Common Core Structure of Amyloid Fibrils by Synchrotron X-ray Diffraction. *J. Mol. Biol.* **273**, 729–739 (1997). [8](#)
- [56] Blundell, T. L. und L. N. Johnson. *Protein Crystallography*. Academic Press, London (1976). [8](#)
- [57] Canet, D. *NMR – Konzepte und Methoden*. Springer, Berlin (1994). [8](#)
- [58] Stork, M., A. Giese, H. A. Kretzschmar und P. Tavan. Molecular Dynamics Simulations Indicate a Possible Role of Parallel  $\beta$ -Helices in Seeded Aggregation of Poly-Gln. *Biophys. J.* **88**, 2442–2451 (2005). [9](#)
- [59] 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). [10](#)
- [60] Dill, K. A. Theory for the Folding and Stability of Globular Proteins. *Biochemistry* **24**, 1501–1509 (1985). [10](#)
- [61] Chan, H. S. und K. A. Dill. Origins of structure in globular proteins. *Proc. Natl. Acad. Sci. USA* **87**, 6388–6392 (1990). [10](#)
- [62] Gregoret, L. M. und F. E. Cohen. Protein Folding Effect of Packing Density on Chain Conformation. *J. Mol. Biol.* **219**, 109–122 (1991). [10](#)
- [63] Baldwin, R. L. The nature of protein-folding pathways – the classical versus the new view. *J. Biomol. NMR* **5**, 103–109 (1995). [10](#)
- [64] Dill, K. A. und H. S. Chan. From Levinthal to pathways to funnels. *Nat. Struct. Biol.* **4**, 10–19 (1997). [10](#)

- [65] Baldwin, R. L. und G. D. Rose. Is protein folding hierarchic? I. Local structure and peptide folding. *Trends Biochem. Sci.* **24**, 26–33 (1999). [10](#)
- [66] Baldwin, R. L. und G. D. Rose. Is protein folding hierarchic? II. Folding intermediates and transition states. *Trends Biochem. Sci.* **24**, 77–83 (1999). [10](#)
- [67] Frauenfelder, H., P. G. Wolynes und R. H. Austin. Biological Physics. *Rev. Mod. Phys.* **71**, 419–430 (1999). [10](#)
- [68] Buchner, J., (Hrsg.) und T. Kiefhaber (Hrsg.). *Protein Folding Handbook*. Wiley-VHC, Weinheim (2005). [10](#), [11](#)
- [69] Gilmanshin, R., S. Williams, R. H. Callender, W. H. Woodruff und R. B. Dyer. Fast events in protein folding: Relaxation dynamics of secondary and tertiary structure in native apomyoglobin. *Proc. Natl. Acad. Sci. USA* **94**, 3709–3713 (1997). [10](#), [11](#)
- [70] Eaton, W. A., V. Muñoz, P. A. Thompson, E. R. Henry und J. Hofrichter. Kinetics and dynamics of loops,  $\alpha$ -helices,  $\beta$ -hairpins, and fast-folding proteins. *Acc. Chem. Res.* **31**, 745–753 (1998). [10](#), [11](#)
- [71] Snow, C. D., H. Nguyen, V. S. Pande und M. Gruebele. Absolute comparison of simulated and experimental protein-folding dynamics. *Nature* **420**, 102–106 (2002). [10](#), [11](#)
- [72] Mayor, U., C. M. Johnson, V. Daggett und A. R. Fersht. Protein folding and unfolding in microseconds to nanoseconds by experiment and simulation. *Proc. Natl. Acad. Sci. USA* **97**, 13518–13522 (2000). [10](#), [11](#)
- [73] Kubelka, J., J. Hofrichter und W. A. Eaton. The protein folding 'speed limit'. *Curr. Opin. Struct. Biol.* **14**, 76–88 (2004). [10](#), [11](#)
- [74] Uversky, V. N., C. J. Oldfield und A. K. Dunker. Intrinsically Disordered Proteins in Human Diseases: Introducing the D<sup>2</sup> Concept. *Annu. Rev. Biophys.* **37**, 215–246 (2008). [11](#), [12](#), [19](#)
- [75] Fitter, J. A measure of conformational entropy change during thermal protein unfolding using neutron spectroscopy. *Biophys. J.* **84**, 3924–3930 (2003). [11](#)
- [76] Privalov, P. L. Physical basis of the stability of the folded conformations of Proteins. In T. E. Creighton (Hrsg.), *Protein Folding*, Kapitel 3, Seiten 83–126. W. H. Freeman, San Francisco (1992). [11](#), [17](#)
- [77] Royer, C. A. Revisiting volume changes in pressure-induced protein unfolding. *Biochim. Biophys. Acta* **1595**, 201–209 (2002). [11](#)



- [78] Munishkina, L. A., C. Phelan, V. N. Uversky und A. L. Fink. Conformational Behavior and Aggregation of  $\alpha$ -Synuclein in Organic Solvents: Modeling the Effects of Membranes. *Biochemistry* **42**, 2720–2730 (2003). [12](#)
- [79] 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). [12](#), [24](#), [55](#), [88](#)
- [80] Morrissey, M. P. und E. I. Shakhnovich. Evidence for the role of PrP<sup>C</sup> helix 1 in the hydrophilic seeding of prion aggregates. *Proc. Natl. Acad. Sci. USA* **96**, 11293–11298 (1999). [12](#)
- [81] Surewicz, W. K., E. M. Jones und A. C. Apetri. The emerging principles of mammalian prion propagation and transmissibility barriers: Insight from studies in vitro. *Acc. Chem. Res.* **39**, 654–662 (2006). [12](#), [88](#)
- [82] Calzolari, L. und R. Zahn. Influence of pH on NMR structure and stability of the human prion protein globular domain. *J. Biol. Chem.* **278**, 35592–35596 (2003). [12](#), [88](#)
- [83] Schwarzinger, S., A. H. C. Horn, J. Ziegler und H. Sticht. Rare Large Scale Subdomain Motions in Prion Protein can Initiate Aggregation. *J. Biomol. Struct. Dyn.* **23**, 581–590 (2006). [12](#), [13](#), [88](#)
- [84] Eghiaian, F., T. Daubenfeld, Y. Quenet, M. van Audenhaege, A.-P. Bouin, G. van der Rest, J. Grosclaude und H. Rezaei. Diversity in prion protein oligomerization pathways results from domain expansion as revealed by hydrogen/deuterium exchange and disulfide linkage. *Proc. Natl. Acad. Sci. USA* **104**, 7414–7419 (2007). [12](#), [88](#)
- [85] De Simone, A., A. Zagari und P. Derreumaux. Structural and hydration properties of the partially unfolded states of the prion protein. *Biophys. J.* **93**, 1284–1292 (2007). [12](#), [88](#)
- [86] Guentchev, M., T. Voigtländer, C. Haberler, M. H. Groschup und H. Budka. Evidence for oxidative stress in experimental prion disease. *Neurobiol. Dis.* **7**, 270–273 (2000). [13](#)
- [87] Requena, J. R., M. N. Dimitrova, G. Legname, S. Teijeira, S. B. Prusiner und R. L. Levine. Oxidation of methionine residues in the prion protein by hydrogen peroxide. *Arch. Biochem. Biophys.* **432**, 188–195 (2004). [13](#)
- [88] Breydo, L., O. V. Bocharova, N. Makarava, V. V. Salnikov, M. Anderson und I. V. Baskakov. Methionine Oxidation interferes with Conversion of the Prion Protein into the fibrillar Proteinase K-Resistant conformation. *Biochemistry* **44**, 15534–15543 (2005). [13](#)

- [89] Wolschner, C., A. Giese, H. A. Kretschmar, R. Huber, L. Moroder und N. Budisaa. Design of anti- and pro-aggregation variants to assess the effects of methionine oxidation in human prion protein. *Proc. Natl. Acad. Sci. USA* (2009). *Early Edition*, DOI: 10.1073/pnas.0902688106. 13
- [90] Pan, T., B. G. Chang, P. Wong, C. Y. Li, R. L. Li *et al.* An aggregation-specific enzyme-linked immunosorbent assay: Detection of conformational differences between recombinant PrP protein dimers and PrP<sup>Sc</sup> aggregates. *J. Virol.* **79**, 12355–12364 (2005). 13
- [91] Williamson, R. A., D. Peretz, C. Pinilla, H. Ball, R. B. Bastidas, R. Rozenshteyn, R. A. Houghten, S. B. Prusiner und D. R. Burton. Mapping the prion protein using recombinant antibodies. *J. Virol.* **72**, 9413–9418 (1998). 13
- [92] Pankiewicz, J., F. Prelli, M. S. Sy, R. J. Kascsak, R. B. Kascsak *et al.* Clearance and prevention of prion infection in cell culture by anti-PrP antibodies. *Eur. J. Neurosci.* **23**, 2635–2647 (2006). 13
- [93] Norstrom, E. M. und J. A. Mastrianni. The Charge Structure of Helix 1 in the Prion Protein Regulates Conversion to Pathogenic PrP<sup>Sc</sup>. *J. Virol.* **80**, 8521–8529 (2006). 13
- [94] Hesp, J. R., N. D. H. Raven und c. J. M. Sutton. A role for His155 in binding of human prion peptide 144–167 to immobilised prion protein. *Biochem. Biophys. Res. Commun.* **362**, 695–699 (2007). 13
- [95] Watzlawik, J., L. Skora, D. Frense, C. Griesinger, M. Zweckstetter, W. J. Schulz-Schaeffer und M. L. Kramer. Prion protein helix 1 promotes aggregation but is not converted into  $\beta$ -sheet. *J. Biol. Chem.* **281**, 30242–30250 (2006). 13
- [96] Solfrosi, L., A. Bellon, M. Schaller, J. T. Cruite, G. C. Abalos und R. A. Williamson. Toward molecular dissection of PrP<sup>C</sup>-PrP<sup>Sc</sup> interaction. *J. Biol. Chem.* **282**, 7465–7471 (2007). 13
- [97] Yu, S., S. Yin, C. Li, P. Wong, B. Chang *et al.* Aggregation of prion protein with insertion mutations is proportional to the number of inserts. *Biochem. J.* **403**, 343–351 (2007). 13
- [98] Ziegler, J., H. Sticht, U. C. Marx, W. Müller, P. Rosch und S. Schwarzinger. CD and NMR studies of prion protein (PrP) helix 1 - Novel implications for its role in the PrP<sup>C</sup>  $\rightarrow$  PrP<sup>Sc</sup> conversion process. *J. Biol. Chem.* **278**, 50175–50181 (2003). 13, 87, 88
- [99] Lingenheil, M., R. Denschlag und P. Tavan. Highly polar environments catalyze the unfolding of PrP<sup>C</sup> helix 1. *Eur. Biophys. J.* (2009). *Zur Veröffentlichung eingereicht.* 13, 21

- [100] Johnson Jr., W. C. Secondary structure of proteins through circular dichroism spectroscopy. *Ann. Rev. Biophys. Biophys. Chem.* **17**, 145–166 (1988). [13](#)
- [101] Hamm, P., J. Helbing und J. Bredenbeck. Two-Dimensional Infrared Spectroscopy of Photoswitchable Peptides. *Annu. Rev. Phys. Chem.* **59**, 291–317 (2008). [13](#), [14](#)
- [102] Bieri, O. und T. Kiefhaber. Elementary steps in protein folding. *Biol. Chem.* **380**, 923–929 (1999). [14](#)
- [103] Ponder, J. W. und D. A. Case. Force Fields for Protein Simulation. *Adv. Prot. Chem.* **66**, 27–85 (2003). [14](#)
- [104] Verlet, L. Computer “Experiments” on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Phys. Rev.* **159**, 98–103 (1967). [14](#)
- [105] MacKerell, A. D., D. Bashford, M. Bellott, R. L. Dunbrack, 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). [15](#)
- [106] 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). [15](#)
- [107] Reichold, R. *Rechnergestützte Beschreibung der Struktur und Dynamik von Peptiden und ihren Bausteinen*. Dissertation, Fakultät für Physik, Ludwig-Maximilians-Universität, München (2009). [15](#), [23](#), [89](#)
- [108] Allen, M. P. und D. J. Tildesley. *Computer Simulations of Liquids*. Oxford University Press, Oxford (1987). [15](#), [17](#)
- [109] 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). [15](#)
- [110] Mathias, G., B. Egwolf, M. Nonella und P. Tavan. A fast multipole method combined with a reaction field for long-range electrostatics in molecular dynamics simulations: The effects of truncation on the properties of water. *J. Chem. Phys.* **118**, 10847–10860 (2003). [15](#), [16](#), [17](#)
- [111] Tavan, P., H. Carstens und G. Mathias. Molecular dynamics simulations of proteins and peptides: Problems, achievements, and perspectives. In *Protein Folding Handbook*, Band 1, Seiten 1170–1195. Wiley-VHC, Weinheim (2005). [15](#), [16](#)

- [112] Ewald, P. P. Die Berechnung optischer und elektrostatischer Gitterpotentiale. *Ann. Phys.* **369**, 253–287 (1921). [16](#)
- [113] Darden, T., D. York und L. Pedersen. Particle mesh Ewald: An  $N \cdot \log(N)$  method for Ewald sums in large systems. *J. Chem. Phys.* **98**, 10089 (1993). [16](#)
- [114] van der Spoel, D., E. Lindahl, B. Hess, G. Groenhof, A. E. Mark und H. J. C. Berendsen. GROMACS: fast, flexible and free. *J. Comput. Chem.* **26**, 1701–1718 (2005). [16](#)
- [115] Hünenberger, P. H. und McCammon. Effect of artificial periodicity in simulations of biomolecules under Ewald boundary conditions: A continuum electrostatics study. *Biophys. Chem.* **78**, 69–88 (1999). [16](#), [89](#)
- [116] Weber, W., P. H. Hünenberger und J. A. McCammon. Molecular Dynamics Simulations of a Polyalanine Octapeptide under Ewald Boundary Conditions: Influence of Artificial Periodicity on Peptide Conformation. *J. Phys. Chem. B* **104**, 3668–3675 (2000). [16](#), [89](#)
- [117] Kastenholz, M. A. und P. H. Hünenberger. Influence of Artificial Periodicity and Ionic Strength in Molecular Dynamics Simulations of Charged Biomolecules Employing Lattice-Sum Methods. *J. Phys. Chem. B* **108**, 774–788 (2004). [16](#), [89](#)
- [118] Villareal, M. A. und G. G. Montich. On the Ewald Artifacts in Computer Simulations. The Test-Case of the Octaalanine Peptide with Charged Termini. *J. Biomol. Struct. Dyn.* **23**, 135–142 (2005). [16](#), [89](#)
- [119] Baumketner, A. und J.-E. Shea. The Influence of Different Treatments of Electrostatic Interactions on the Thermodynamics of Folding of Peptides. *J. Phys. Chem. B* **109**, 21322–21328 (2005). [16](#), [89](#)
- [120] Lins, R. D. und U. Röthlisberger. Influence of Long-Range Electrostatic Treatments on the Folding of the N-Terminal H4 Histone Tail Peptide. *J. Chem. Theory Comput.* **2**, 246–250 (2006). [16](#), [89](#)
- [121] Reif, M., V. Kräutler, M. A. Kastenholz, X. Daura und P. H. Hünenberger. Molecular Dynamics Simulations of a Reversibly Folding  $\beta$ -Heptapeptide in Methanol: Influence of the Treatment of Long-Range Electrostatic Interactions. *J. Phys. Chem. B* **113**, 3112–3128 (2009). [16](#), [89](#)
- [122] Stork, M. *Molekulardynamik-Simulationen von amyloidogenen Proteinen in Lösung: Stabilitätsuntersuchungen und Weiterentwicklung einer Kontinuumsmethode*. Dissertation, Fakultät für Physik, Ludwig-Maximilians-Universität (2006). [16](#)

- 
- [123] Mathias, G. *Elektrostatische Wechselwirkungen in komplexen Flüssigkeiten und ihre Beschreibung mit Molekulardynamiksimulationen*. Dissertation, Fakultät für Physik, Ludwig-Maximilians-Universität, München (2004). [16](#), [17](#)
- [124] Niedermeier, C. und P. Tavan. Fast version of the structure adapted multipole method – Efficient calculation of electrostatic forces in protein dynamics. *Mol. Simul.* **17**, 57–66 (1996). [16](#)
- [125] Niedermeier, C. und P. Tavan. A structure adapted multipole method for electrostatic interactions in protein dynamics. *J. Chem. Phys.* **101**, 734–748 (1994). [16](#)
- [126] Trioni, I. G., R. Sperb, P. E. Smith und W. F. van Gunsteren. A generalized reaction field method for molecular-dynamics simulations. *J. Chem. Phys.* **102**, 5451–5459 (1995). [17](#)
- [127] Uversky, V. N., J. Li und A. L. Fink. Evidence for a Partially Folded Intermediate in  $\alpha$ -Synuclein Fibril Formation. *J. Biol. Chem.* **276**, 10737–10744 (2001). [17](#)
- [128] Eichinger, M., H. Grubmüller, H. Heller und T. P. FAMUSAMM: An algorithm for rapid evaluation of electrostatic interactions in molecular dynamics simulations. *J. Comput. Chem.* **18**, 1729–1749 (1997). [17](#)
- [129] Hünenberger, P. H. Thermostat Algorithms for Molecular Dynamics Simulations. *Adv. Polym. Sci.* **173**, 105–149 (2005). [18](#), [81](#)
- [130] Lingenheil, M., R. Denschlag, R. Reichold und P. Tavan. The “hot-solvent/cold-solute” problem revisited. *J. Chem. Theory Comput.* **4**, 1293–1306 (2008). [18](#), [24](#)
- [131] McCammon, J. A., B. R. Gelin und M. Karplus. Dynamics of folded proteins. *Nature* **267**, 585–590 (1977). [18](#)
- [132] 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). [18](#)
- [133] Freddolino, P. L., L. F., M. Gruebele und K. . Schulten. Ten-microsecond molecular dynamics simulation of a fast-folding WW domain. *Biophys. J.* **94**, L75–L77 (2008). [18](#)
- [134] Mitsutake, A., Y. Sugita und Y. Okamoto. Generalized-ensemble algorithms for molecular simulations of biopolymers. *Biopolymers* **60**, 96–123 (2001). [19](#)

- [135] Sugita, Y., A. Kitao und Y. Okamoto. Multidimensional replica-exchange method for free-energy calculations. *J. Chem. Phys.* **113**, 6042–6051 (2000). [19](#)
- [136] Fukunishi, H., O. Watanabe und S. Takada. On the Hamiltonian replica exchange method for efficient sampling of biomolecular systems: Application to protein structure prediction. *J. Chem. Phys.* **116**, 9058–9067 (2002). [19](#)
- [137] Affentranger, R. und I. Tavernelli. A novel Hamiltonian replica exchange MD protocol to enhance protein conformational space sampling. *J. Chem. Theory Comput.* **2**, 217–228 (2006). [19](#)
- [138] Nakajima, N., H. Nakamura und A. Kidera. Multicanonical ensemble generated by molecular dynamics simulation for enhanced conformational sampling of peptides. *J. Phys. Chem. B* **101**, 817–824 (1997). [19](#)
- [139] Berg, B. A. und T. Neuhaus. Multicanonical Ensemble – a New Approach to Simulate 1st-Order Phase-Transitions. *Phys. Rev. Lett.* **68**, 9–12 (1992). [19](#)
- [140] Berg, B. A. und T. Neuhaus. Multicanonical Algorithms for 1st Order Phase-Transitions. *Phys. Lett. B* **267**, 249–253 (1991). [19](#)
- [141] Hukushima, K. und K. Nemoto. Exchange Monte Carlo method and application to spin glass simulations. *J. Phys. Soc. Jpn.* **65**, 1604 (1996). [19](#)
- [142] Hansmann, U. H. E. Parallel tempering algorithm for conformational studies of biological molecules. *Chem. Phys. Lett.* **281**, 140 (1997). [19](#)
- [143] Sugita, Y. und Y. Okamoto. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* **314**, 141 (1999). [19](#), [20](#)
- [144] Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller und E. Teller. Equation of State Calculations by Fast Computing Machines. *J. Chem. Phys.* **21**, 1087 (1953). [20](#)
- [145] Kone, A. und D. A. Kofke. Selection of temperature intervals for parallel-tempering simulations. *J. Chem. Phys.* **122**, 206101(1–2) (2005). [21](#), [24](#), [43](#), [47](#), [84](#), [86](#)
- [146] Abraham, M. J. und J. E. Gready. Ensuring mixing efficiency of replica-exchange molecular dynamics simulations. *J. Chem. Theory Comput.* **4**, 1119–1128 (2008). [21](#), [84](#), [86](#)
- [147] Nadler, W. und U. H. E. Hansmann. Generalized ensemble and tempering simulations: A unified view. *Phys. Rev. E* **75**, 026109 (2007). [21](#), [24](#), [43](#), [47](#), [84](#), [85](#)

- [148] Nadler, W. und U. H. E. Hansmann. Dynamics and optimal number of replicas in parallel tempering simulations. *Phys. Rev. E* **76**, 065701(R) (2007). [21](#), [24](#), [43](#), [47](#), [84](#)
- [149] Nadler, W. und U. H. E. Hansmann. Optimized Explicit-Solvent Replica Exchange Molecular Dynamics from Scratch. *J. Phys. Chem. B* **112**, 10386–10387 (2008). [21](#), [24](#), [43](#), [47](#), [84](#), [86](#)
- [150] Calvo, F. All-exchanges parallel tempering. *J. Chem. Phys.* **123**, 124106 (2005). [21](#)
- [151] Brenner, P., C. R. Sweet, D. VonHandorf und J. A. Izaguirre. Accelerating the replica exchange method through an efficient all-pairs exchange. *J. Chem. Phys.* **126**, 074103 (2007). [21](#), [85](#), [87](#)
- [152] Denschlag, R., M. Lingenheil und P. Tavan. Optimal temperature ladders in replica exchange simulations. *Chem. Phys. Lett.* **473**, 193–195 (2009). [21](#), [24](#)
- [153] Lingenheil, M., R. Denschlag, G. Mathias und P. Tavan. Efficiency of exchange schemes in replica exchange. *Chem. Phys. Lett.* (2009). *Zur Veröffentlichung angenommen.* [21](#), [24](#)
- [154] Kubitzki, M. B. und B. L. de Groot. Molecular dynamics simulations using temperature-enhanced essential dynamics replica exchange. *Biophys. J.* **92**, 4262–4270 (2007). [21](#)
- [155] 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. USA* **102**, 13749–13754 (2005). [21](#), [22](#), [83](#)
- [156] Denschlag, R., M. Lingenheil, P. Tavan und G. Mathias. Simulated solute tempering. *J. Chem. Theory Comput.* (2009). *Zur Veröffentlichung eingereicht.* [22](#), [90](#)
- [157] Berendsen, H. J. C., J. P. M. Postma, W. F. van Gunsteren und J. Hermans. Interaction Models for Water in Relation to Protein Hydration. In B. Pullmann (Hrsg.), *Intermolecular Forces*, Seiten 331–342. Reidel, Dordrecht (1981). [22](#)
- [158] Jorgensen, W. L., J. Chandrasekhar, J. D. Madura, R. W. Impey und M. L. Klein. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **79**, 926–935 (1983). [22](#)
- [159] Berendsen, H. J. C., J. R. Grigera und T. P. Straatsma. The missing term in effective pair potentials. *J. Phys. Chem.* **91**, 6269–6271 (1987). [22](#)

- [160] Berweger, C. D., W. F. van Gunsteren und F. Müller-Plathe. Force-Field Parametrization by Weak-Coupling – Reengineering SPC Water. *Chem. Phys. Lett.* **232**, 429 (1995). [22](#)
- [161] Walser, R., A. E. Mark und W. F. van Gunsteren. The effect of force-field parameters on properties of liquids: Parametrization of a simple three-site model for methanol. *J. Chem. Phys.* **112**, 10450–10459 (2000). [22](#)
- [162] Bordat, P., J. Sacristan, D. Reith, S. Girard, A. Glattli und F. Müller-Plathe. An improved dimethyl sulfoxide force field for molecular dynamics simulations. *Chem. Phys. Lett.* **374**, 201–205 (2003). [22](#)
- [163] 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). [23](#), [24](#)
- [164] Reichold, R., B. Fierz, T. Kiefhaber und P. Tavan. Toward MD descriptions of  $\alpha$ -helix folding. *J. Chem. Theory Comput.* (2009). *Zur Veröffentlichung eingereicht.* [23](#), [89](#)
- [165] de Montesquieu, C.-L. *De l'esprit des lois*. Barillot et fils, Genf (1749). [27](#)
- [166] Shakespeare, W. *The Tragedy of King Lear*. Project Gutenberg, Champaign (1997).  
URL [www.gutenberg.org/etext/1128](http://www.gutenberg.org/etext/1128) [55](#)
- [167] Beagle, P. S., J. Bass und A. Rankin, Jr. The Last Unicorn. Rankin/Bass Productions, Incorporated Television Company (1982). *Animationsfilm nach dem gleichnamigen Roman von P. S. Beagle.* [81](#)
- [168] Paschek, B., H. Nymeyer und A. E. Garcia. Replica exchange simulation of reversible folding/unfolding of the Trp-cage miniprotein in explicit solvent: On the structure and possible role of internal water. *J. Struct. Biol.* **157**, 524–533 (2007). [85](#)
- [169] Okamoto, Y., M. Fukugita, T. Nakazawa und H. Kawai.  $\alpha$ -Helix Folding by Monte-Carlo Simulated Annealing in Isolated C-Peptide of Ribonuclease-A. *Protein Eng.* **4**, 639–647 (1991). [85](#)
- [170] Nadler, W., J. H. Meinke und U. H. E. Hansmann. Folding proteins by first-passage-times-optimized replica exchange. *Phys. Rev. E* **78**, 061905 (2008). [85](#)
- [171] Katzgraber, H. G., S. Trebst, D. A. Huse und M. Troyer. Feedback-optimized parallel tempering Monte Carlo. *J. Stat. Mech.* Seite P03018 (2006). [85](#)



- [172] Sharman, G. J., N. Kenward, H. E. Williams, M. Landon, R. J. Mayer und M. S. Searle. Prion protein fragments spanning helix 1 and both strands of  $\beta$ -sheet (residues 125-170) show evidence for predominantly helical propensity by CD and NMR. *Fold. Des.* **3**, 313–320 (1998). [87](#)
- [173] Liu, A. Z., P. Riek, R. Zahn, S. Hornemann, R. Glockshuber und K. Wüthrich. Peptides and proteins in neurodegenerative disease: Helix propensity of a polypeptide containing helix 1 of the mouse prion protein studied by NMR and CD spectroscopy. *Biopolymers* **51**, 145–152 (1999). [87](#)
- [174] Jamin, N., Y. M. Coic, C. Landon, L. Ovtracht, F. Baleux, J. M. Neumann und A. Sanson. Most of the structural elements of the globular domain of murine prion protein form fibrils with predominant  $\beta$ -sheet structure. *FEBS Lett.* **529**, 256–260 (2002). [87](#)
- [175] Thompson, A., A. R. White, C. McLean, C. L. Masters, R. Cappai und C. J. Barrow. Amyloidogenicity and neurotoxicity of peptides corresponding to the helical regions of PrP<sup>C</sup>. *J. Neurosci. Res.* **62**, 293–301 (2000). [87](#)
- [176] Tahiri-Alaoui, A., M. Bouchard, J. Zurdo und W. James. Competing intrachain interactions regulate the formation of  $\beta$ -sheet fibrils in bovine PrP peptides. *Protein Sci.* **12**, 600–608 (2003). [87](#)
- [177] Kozin, S. A., G. Bertho, A. K. Mazur, H. Rabesona, J. P. Girault, T. Haertle, M. Takahashi, P. Debey und G. H. B. Hoa. Sheep prion protein synthetic peptide spanning helix 1 and  $\beta$ -strand 2 (residues 142-166) shows  $\beta$ -hairpin structure in solution. *JBC* **276**, 46364–46370 (2001). [87](#)
- [178] Megy, S., G. Bertho, S. A. Kozin, P. Debey, G. H. B. Hoa und J. P. Girault. Possible role of region 152-156 in the structural duality of a peptide fragment from sheep prion protein. *Protein Sci.* **13**, 3151–3160 (2004). [87](#)
- [179] Egwolf, B. und P. Tavan. Continuum description of solvent dielectrics in molecular-dynamics simulations of proteins. *JCP* **118**, 2039–2056 (2003). [90](#)
- [180] 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). [90](#)
- [181] 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). [90](#)
- [182] Grimm, J. und W. Grimm. Hänsel und Gretel. In *Die Kinder- und Hausmärchen der Brüder Grimm*, Seiten 94–105. Gondrom, Bindlach (1996). [106](#)

# Danksagung

Mein Märchen ist aus, dort läuft  
eine Maus, wer sie fängt darf  
sich eine große Pelzkappe  
daraus machen.

---

*Anonymer Märchenerzähler  
wiedergegeben durch die  
Brüder Grimm [182]*

Die letzten Seiten einer solchen Arbeit gehören jenen, die auf sehr vielfältige Weise zu ihrer Entstehung beigetragen haben.

An erster Stelle steht mein Dank an Paul Tavan, meinen Doktorvater. So eigentümlich altertümlich diese Bezeichnung daherkommt, so treffsicher beschreibt sie Pauls Rolle für meine Promotion. Dank gebührt ihm für den großen persönlichen Einsatz, mit dem er dafür sorgte, dass das Kind stets zu essen hatte, sowohl was die Finanzierung wie auch die wissenschaftliche Begleitung betrifft. Sein Interesse galt nicht allein dem Erfolg des Projekts, sondern ebenso dem persönlichen Wohlergehen seines Schützlings und dessen Persönlichkeitsbildung. Mich von der Besserwisserei zu therapieren ist zweifelsohne kein Leichtes gewesen. Zuletzt möchte ich ihm für die Möglichkeit danken, mir meine Hörner in der Lehre abzustoßen – auch dies ein Projekt angewandter Persönlichkeitsbildung.

Ist Paul Tavan mein Doktorvater, so ist Robert Denschlag zweifelsohne mit Recht als mein Doktorbruder zu bezeichnen. Die unzähligen Diskussionen, die wir mit Verve und Freude führten, manches Mal auch mit der puren Lust an der Provokation, hatten wohl zuweilen den Charakter geschwisterlicher Zwistigkeiten. Ein guter Teil der Erkenntnisse, die in dieser Arbeit dokumentiert sind, hat seine Wurzel in diesen Diskussionen. Ein typisches Unigewächs wie mich durch einen Pragmatiker vom Schlage eines Robert Denschlags zu ergänzen, war sicherlich ein genialer Schachzug von Pauls Seite, ob nun absichtsvoll oder nicht. Auch für die gute Kooperation in der Lehre sei Robert gedankt. Schließlich will ich ihm dafür danken, dass er trotz meiner triumphal besseren Körperfettwerte die Zusammenarbeit nicht aufgekündigt hat.

Martina Stork und Thomas Hirschberger sei für ihre Geduld gedankt, mit der sie mich bei der Hand nahmen und mir die Orientierung im Labyrinth der Forschung im Bereich der neurodegenerativen Krankheiten erleichterten. Christine Lutz danke ich für ihren Willen, sich auf das Abenteuer einer „höchst experimentellen“ Diplomarbeit aus diesem Bereich einzulassen und dabei meine Hilfe anzunehmen, da doch kaum etwas mehr schmeichelt als ernsthaft um Rat gefragt

---

zu werden. Rudolf Reichold möchte ich danken für die gute Zusammenarbeit in der Peptid-Gruppe, seinen klaren Blick und sein kritisches Auge. Gerald Mathias danke ich für seine konzentrierte und pragmatische Art, die die Zusammenarbeit mit ihm zu einer sehr angenehmen Erfahrung machte, die gerne hätte länger wären dürfen als jene zwei Monate, in denen sich unser Dasein in der Arbeitsgruppe überschritt. Sebastian Bauer gilt mein Dank für seine Bereitschaft, meine Simulationen zu den Gittersummen-Periodizitätsartefakten fortzuführen – auf die Resultate bin ich gespannt. Daneben sei beiden, Sebastian und Gerald, für das Korrekturlesen der Arbeit gedankt.

Galina Babitzki, Benjamin Rieff, Bernhard Schropp und Verena Schultheis haben mich neben den bereits genannten als Doktorandenkollegen auf meinem Weg lange Zeit begleitet. Ihnen sei gedankt für ihre Fragen, die Bereitschaft, sich der meinen anzunehmen, sowie für die kollegiale Atmosphäre in der Arbeitsgruppe. Kürzer aber als nicht weniger angenehme Zeit- und sonstige Genossen erlebte ich Matthias Schmitz, Thomas Zauner, Benedikt Breitenfeld und Stefan Maciej. Für die beiden letzteren gilt im besonderen auch das Christine oben erteilte Lob für den steten Strom ihrer Fragen.

Für ihren unschätzbaren Einsatz für mein Werkzeug während der vergangenen vier Jahre danke ich Sebastian Bauer, Rudolf Reichold, Thomas Hirschberger, Robert Denschlag und Karl-Heinz Mantel. Weiterhin danke ich Alexandra Michaelis, Marianne Widmann-Diermeier, Barbara Podolski, Ines Haame und Birgit Carell für ihre Unterstützung bei administrativen Problemen und schließlich Rosi Lederer für administrative Unterstützung und ihr Organisationstalent im Rahmen des bayerischen Forschungsverbunds Prionen.

Ein Schaf, das mit den Wölfen heult, wurde ich dankenswerterweise nach dem Ableben einer regelmäßigen Theoretiker-Kaffeepausen-Gruppe von der AG Huber adoptiert – besonders des Herrn Biedermanns Bindekräfte seien an diesem Punkt lobend erwähnt. Gesa Palte reiht sich ein unter die bereits zweimal zu Ehren gekommenen ewig fragenden Diplomanden. Die Gruppe der anonymen Tee-Trinker sei bedankt für die freundliche Aufnahme eines Quotentheoretikers, ihre Funktion als Jammersenke und die Gelegenheit für ein wenig Clownerie. Tobias Schrader ist ein Unikum, dem ich für seine Bereitschaft danken darf, uns Theoretiker niemals aufzugeben. Den Teilnehmern unserer täglichen Essensausflüge in die Mensa sei gedankt für den bunten Blumenstrauß an wissenschaftlichen und nichtwissenschaftlichen Gesprächen.

Für ein Leben außerhalb der Universität ist zu danken: Der Genossenschaft, die in den letzten Monaten den vorübergehenden Totalausfall eines Genossen verkraften mußte; der Dienstagsrunde, die die geistigen Geburtswehen eines Doktors anhand von Cosymo erleben durfte; schließlich meiner Familie, ohne deren helfende Hand ich nicht hier säße und diese Zeilen tippen würde.

# Lebenslauf

Name: Martin Lingenheil

Geburtsdatum: 12.01.1979

Geburtsort: Kaufbeuren

seit 05/2009      Wissenschaftlicher Mitarbeiter  
am Helmholtz-Zentrum München

03/2005–05/2009      Wissenschaftlicher Mitarbeiter an der  
Ludwig-Maximilians-Universität München

12/2004      Diplom in Allgemeiner Physik

10/1999–12/2004      Studium der Physik, Technische Universität München

08/1998–09/1999      Zivildienst bei der Stadt Kaufbeuren

06/1998      Allgemeine Hochschulreife

09/1989–06/1998      Staatliches Gymnasium Kaufbeuren