Molekulardynamik-Simulationen von amyloidogenen Proteinen in Lösung: Stabilitätsuntersuchungen und Weiterentwicklung einer Kontinuumsmethode

Martina Stork

München 2006

Molekulardynamik-Simulationen von amyloidogenen Proteinen in Lösung: Stabilitätsuntersuchungen und Weiterentwicklung einer Kontinuumsmethode

Martina Stork

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

vorgelegt von
Martina Stork
aus Krumbach

München, im Dezember 2006

Erstgutachter: Prof. Paul Tavan
Zweitgutachter: Prof. Regina de Vivie-Riedle
Tag der mündlichen Prüfung: 07.02.2007

Zusammenfassung

Viele neurodegenerative Erkrankungen, wie die transmissiblen spongiformen Enzephalopathien (TSE), die Alzheimer- und die Huntington-Krankheit, sind durch charakteristische Ablagerungen im Gehirn, sogenannte Amyloide, gekennzeichnet. Amyloide sind oftmals fibrilläre Aggregate von normalerweise löslichen Proteinen, deren dreidimensionale Strukturen sich bei der Aggregation verändern. Bedauerlicherweise waren hochauflösende Methoden biophysikalischer Strukturaufklärung bislang auf Amyloide nicht anwendbar. Dagegen können Molekulardynamik (MD)-Simulationen amyloidogener Proteine und Peptide in ihrer Lösungsmittelumgebung dazu beitragen, die Mechanismen der auftretenden Konformationsänderungen zu verstehen und die Strukturen amyloider Fasern aufzuklären. Die korrekte und effiziente Beschreibung der Lösungsmittelumgebung spielt dabei eine entscheidende Rolle.

Im ersten Teil dieser Arbeit wird die Konformationsdynamik Amyloid bildender Peptide und Proteine in expliziter wässriger Umgebung untersucht. In MD-Simulationen des zellulären Prion Proteins (PrP^C) werden durch Einführung der Punktmutationen M205S und M205R entscheidende Faktoren für die korrekte Faltung und strukturelle Stabilität des Proteins identifiziert. Ferner wird für die Grundstruktur der bei TSE auftretenden pathogenen Isoform PrP^{Sc} ein Modell basierend auf dem Strukturmotiv einer parallelen β-Helix entwickelt. Analog dazu werden Peptide aus poly-Glutamin, die den mutmaßlichen Aggregationskeim bei der Huntington-Krankheit darstellen, als parallele β-Helizes unterschiedlicher Formen und Größen modelliert. In MD-Simulationen ermitteln wir aus diesen Strukturen thermodynamisch stabile monomere und dimere Aggregationskeime.

Da die erreichbaren Simulationszeiten in expliziten Lösungsmitteln verglichen mit den Zeitskalen der Proteindynamik zu kurz sind, wird im zweiten Teil dieser Arbeit eine effiziente Kontinuumsmethode für Proteine in polaren Lösungsmitteln weiterentwickelt. In dieser Methode wird das durch die Polarisation des Lösungsmittels hervorgerufene Reaktionsfeld (RF) durch normalverteilte RF-Dipoldichten an den Orten der Proteinatome beschrieben. Die sich daraus ergebenden RF-Kräfte auf die Proteinatome berücksichtigen aber nicht den Druck an den dielektrischen Grenzflächen, der vom Kontinuum auf das Protein ausgeübt wird, und verletzen damit das 3. Newtonsche Gesetz. Dies führt in MD-Simulationen zu erheblichen Artefakten. In dieser Arbeit wird diese Kontinuumsmethode so umformuliert und erweitert, dass die resultierenden RF-Kräfte dem Prinzip *Actio=Reactio* gehorchen. Die modifizierte Kontinuumsmethode wird in ein MD-Programm implementiert und an Hand geeigneter Systeme parametrisiert. In ausgedehnten MD-Simulationen des Alanin-Dipeptids wird die Korrektheit und Effizienz der Methode demonstriert.

Inhaltsverzeichnis

Abkürzungen

BSE	bovine spongiforme Enzephalopathie; schwammartige Hirndegeneration beim Rind, 2
CJK	Creutzfeldt-Jakob-Krankheit, 1
EM	Elektronenmikroskop, 9
fCJK	vererbte (familiäre) Form der CJK, 1
GB	Generalisierte Born-Methode, 18
MD	Molekulardynamik, 11
MM	Molekularmechanik, 12
NMR	*nuclear magnetic resonance*; Kernspinresonanz, 7
PB	Poisson-Boltzmann, 17
PrP	Prion-Protein, 6
PrP^{C}	*cellular prion protein*; zelluläres Prion-Protein, 4
PrP^{Sc}	Prion-Protein Scrapie, 4
RF	Reaktionsfeld, 17
sCJK	sporadische Form der CJK, 1
TSE	transmissible spongiforme Enzephalopathie; übertragbare schwammartige Erkrankung des Gehirns bei Mensch und Tier, 2
vCJK	neue Variante der Creutzfeldt-Jakob-Krankheit, 3
ZNS	zentrales Nervensystem, 1

1 Einführung

1.1 Spongiforme Enzephalopathien

Der Neuropathologe Alfons M. Jakob und der Neurologe Hans G. Creutzfeldt beschrieben 1920 [1] bzw. 1921 [2] eine Erkrankung des Zentralnervensystems (ZNS) beim Menschen, die bald darauf Creutzfeldt-Jakob-Krankheit (CJK) genannt wurde. Anfänglich war diese neurodegenerative Erkrankung klinisch nur vage und nicht allgemeingültig definiert und meist nicht von anderen neurodegenerativen Erkrankungen unterscheidbar. Wie die Alzheimer-Krankheit tritt CJK vor allem bei älteren Menschen ab 60 Jahren auf, führt zu Bewegungsstörungen, fortschreitender Demenz und schließlich zum Tod [3].

Untersucht man das Gehirn eines Verstorbenen unter dem Lichtmikroskop, so erhält man jedoch ein für CJK charakteristisches Bild. Abbildung 1.1 zeigt einen solchen Hirnschnitt. Das zentrale Nervensystem weist eine Vielzahl von Löchern (Vakuolen) auf und erhält dadurch eine schwammartige Struktur. Die CJK wird deshalb auch als eine *spongiforme* (schwammartige) *Enzephalopathie* (Degeneration des Zentralnervensystems) bezeichnet und unterscheidet sich darin von anderen neurodegenerativen Erkrankungen. Bei einer Nachuntersuchung der Hirnproben, die von Jakob und Creutzfeldt entnommen worden waren, stellte sich heraus, dass die meisten nicht die für CJK typischen Merkmale aufweisen. Deshalb würde man kurioserweise heute nur wenige der von Jakob und keinen einzigen der von Creutzfeldt beschriebenen Fälle der Creutzfeldt-Jakob-Krankheit, sondern anderen neurodegenerativen Erkrankungen zuordnen [4].

Heutzutage weiß man, dass CJK bei etwa einem von einer Million Menschen *sporadisch* auftritt, d.h. ohne erkennbare statistische Häufungen [6]. Man bezeichnet diese sporadische Form von CJK auch als sCJK. Die sCJK ist somit im Vergleich zur Alzheimer-Krankheit mit 189 Fällen pro einer Million Menschen (Jahr 2001, USA) [7] eher selten. Neben der sporadischen Erscheinungsform tritt CJK in bestimmten Familien aber auch gehäuft auf und wird über die Generationen vererbt [8]. Diese Variante wird deshalb auch *familiäre* CJK (fCJK) genannt.

Spongiforme Enzephalopathien gibt es nicht nur beim Menschen, sondern auch bei Tieren. Mitte des 18. Jahrhunderts wurde bei Schafen die Krankheit *Scrapie* oder auf deutsch Traberkrankheit beschrieben. Wie aus den Benennungen schon hervorgeht, fallen erkrankte Tiere durch ihr ungewöhnliches Verhalten wie auffälliges Kratzen und eine eigenartige Gangart auf. Scrapie trat in großen Epidemien auf und wurde als eine *übertragbare* (transmissible) Krankheit erkannt [9]. Auf welchem Wege

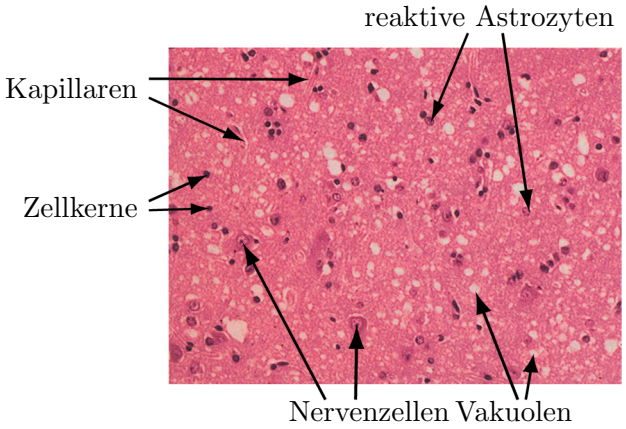

Abbildung 1.1: Lichtmikroskopisches Bild eines eingefärbten Gewebeschnitts aus dem zerebralen Kortex eines an CJK verstorbenen Patienten. Deutlich zu sehen sind die krankhaften Vakuolen (weiß) von 5-20 μm im Durchmesser (Abbildung entnommen aus [5] und beschriftet mit Hilfe von Armin Giese).

die Krankheit übertragen wird, ist noch nicht abschließend geklärt, allerdings geht man überwiegend von einer oralen Infektion aus, aber auch eine Übertragung über die Haut wird für möglich gehalten [10].

1959 erkannte William Hadlow die Ähnlichkeit zwischen der Schafkrankheit Scrapie und der Menschenkrankheit Kuru, als er unter dem Lichtmikroskop ZNS-Material von erkrankten Tieren und Menschen betrachtete und ähnliche Bilder wie in Abbildung 1.1 erhielt [11]. Die Krankheit Kuru trat bei Aborigines in Neu-Guinea auf, die die Gehirne ihrer Verstorbenen verzehrten [12]. Sie entstand somit wie beim Schaf durch den Kontakt mit infektiösem Material der gleichen Spezies. Derartige übertragbare Erkrankungen des ZNS mit einem charakteristischen Bild wie in Abbildung 1.1 werden unter dem Oberbegriff der *transmissiblen spongiformen Enzephalopathien* (TSE) zusammengefasst. Später wurde ebenso die Ähnlichkeit von Kuru und CJK erkannt [13]. Diese Ähnlichkeit wurde experimentell bekräftigt, als es gelang, Kuru [14] und sogar die vererbte Form von CJK auf Schimpansen zu übertragen [15]. Dies war in der Medizingeschichte etwas völlig Neues, da CJK die erste Krankheit war, die sowohl als vererbbar als auch infektiös erkannt wurde [16].

1987 wurde in Großbritannien zum ersten Mal eine ähnliche Erkrankung bei Rindern festgestellt, die *Bovine Spongiforme Enzephalopathie* (BSE). Seit Ausbruch der Epidemie waren bis Ende 2005 in Großbritannien annähernd 185 000 Rinder erkrankt, während in der restlichen Welt bis dahin ca. 5500 Fälle von BSE [17] entdeckt worden waren. In Deutschland gab es bis zu diesem Zeitpunkt 390 bestätigte Fälle von BSE [18]. Da die meisten Tiere allerdings vermutlich schon vor dem Ausbruch der Krankheit geschlachtet worden waren, geht man von insgesamt einer Million infizierter Tiere aus [19]. Die Ursache dieser Tierseuche wurde in der geänderten Herstellungsweise von Tiermehl ab den späten 1970er Jahren und der anschließenden Verfütterung an Rinder gesehen. Meist wird vermutet, dass dabei eine Übertragung von Schaf zu Rind stattgefunden hat, die eine Überschreitung

Krankheit	Spezies	Ursache/Übertragungswege
sCJK (sporadische CJK)	Mensch	Spontane Entstehung
fCJK (familiäre CJK)	Mensch	Vererbung
GSS (Gerstmann-Sträussler-Scheinker Syndrom)	Mensch	Vererbung
Kuru	Mensch	Verzehr von infektiösem menschlichen Hirnmaterial
iCJK (iatrogene CJK)	Mensch	Infektion durch ärztliche Eingriffe, z.B. Injektion kontaminierter Stoffe
vCJK (neue Variante der CJK)	Mensch	Verzehr von infektiösem Fleisch
Traberkrankeit/Scrapie	Schaf	Infektion, vermutlich oral
BSE (bovine spongiforme Enzephalopathie)	Rind	Verfütterung von infektiösem Fleisch- und Knochenmehl
CWD (chronic wasting disease)	Hirsch, Elch	unbekannt

Tabelle 1.1: Übertragungswege einiger Varianten von Spongiformen Enzephalopathien bei Menschen und Tieren.

der Artenbarriere bedeutet [20]. Auf Grund des gewaltigen Ausmaßes der Seuche und der vermuteten Überschreitung der Artenbarriere wurden Befürchtungen laut, dass sich auch der Mensch durch den Verzehr infizierter Rinder mit der Krankheit anstecken könnte. Tatsächlich wurde knapp 10 Jahre nach der ersten BSE-Krise eine neue Variante der Creutzfeldt-Jakob-Krankheit (vCJK) beim Menschen entdeckt [21], die, anders als CJK, vor allem bei jüngeren Menschen auftritt [22]. Inzwischen wird vCJK tatsächlich auf eine orale Ansteckung mit dem BSE-Erreger zurückgeführt [22]. Bis Juni 2006 sind in Großbritannien mindestens 111 Menschen an vCJK gestorben [23]. Auf Grund der unbekannten Inkubationszeit kann die Anzahl der zukünftigen vCJK-Fälle allerdings schwer prognostiziert werden.

Nach der Übertragung der Krankheit auf den Menschen hat sich die Suche nach dem Erreger, sowie nach möglichen Impfstoffen oder Heilmitteln intensiviert. In Tabelle 1.1 sind verschiedene TSE-Erkrankungen bei Mensch und Tier aufgelistet, für die ein möglicher gemeinsamer Erreger verantwortlich ist. Dieser Erreger muss sowohl sporadische Varianten wie sCJK, ererbte Varianten wie fCJK oder GSS als auch transmissible Varianten wie Scrapie, BSE und vCJK erklären können.

Lange Zeit hielt man einen Virus mit einer sehr langen Inkubationszeit für den Erreger von TSE-Erkrankungen. Im Jahr 1982 wies Stanley Prusiner dann allerdings nach, dass der Erreger überwiegend aus Eiweiß besteht, und führte für diese unbekannte Substanz die Bezeichnung *Prion* ein [24]. Wie eine proteinartige Substanz anderes Gewebe infizieren und sich dabei ohne Erbinformationen vermehren konnte, war aber zunächst rätselhaft. Im Hirn-Material von Tieren, die an Scrapie erkrankt waren, fand man schließlich ein körpereigenes Protein, das in einer ungewöhnlichen räumlichen Struktur vorlag, und bezeichnete es als Prion-Protein Scrapie, abgekürzt

PrP^{Sc} [24]. Dieses Protein fand man auch im Hirn-Material gesunder Tiere, allerdings in einer gänzlich anderen räumlichen Struktur [25]. Man nannte das Protein in dieser Form zelluläres Prion-Protein, abgekürzt PrP^{C}. Prusiner stellte die These auf, dass das Prion-Protein alleine durch Änderung seiner Konformation und Aggregation die TSE-Erkrankungen auslöst [20, 26], ein Mechanismus, der schon früher in einem anderen Kontext als *protein-only-Hypothese* vorgeschlagen worden war [27]. Zum besseren Verständnis dieser Hypothese möchte ich im fogenden Abschnitt auf den Aufbau und die räumliche Struktur von Proteinen eingehen.

1.2 Aufbau von Proteinen

Abbildung 1.2a zeigt die Zusammensetzung von Proteinen aus ihren Grundbausteinen, den Aminosäuren. Jede Aminosäure besteht aus einem zentralen Kohlenstoffatom C_α, an das eine Aminogruppe NH_2, eine Carboxylgruppe COOH, ein Wasserstoffatom H und eine Seitenkette R gebunden sind [28, 29]. Es gibt 20 natürlich vorkommende Aminosäuren, die sich jeweils in ihrer Seitenkette unterscheiden. Bei der Synthese zu einem Protein verbindet sich die Aminogruppe einer Aminosäure mit der Carboxylgruppe einer anderen Aminosäure unter Abspaltung eines Wassermoleküls. Diese Bindung wird Peptidbindung genannt. Wenige durch Peptidbindungen verbundene Aminosäuren bezeichnet man als Peptid. Den Aminosäurerest in einem Protein oder Peptid nennt man Residuum. Die Residuen werden beginnend am N-Terminus, dem Ende mit einer Aminogruppe, an das kein weiteres Residuum gebunden ist, bis zum C-Terminus, dem Ende mit einer Carboxylgruppe, an das kein weiteres Residuum gebunden ist, durchnummeriert.

Die Abfolge der Aminosäuren nennt man die *Primärstruktur* des Proteins. Als Proteinrückgrat oder auch Hauptkette wird die Abfolge der Stickstoffe N, der Kohlenstoffe C_α und der Kohlenstoffe C bezeichnet (in Abbildung 1.2a grau hinterlegt). Durch die Verbindung der Aminosäuren entsteht ein starres Peptidplättchen, bei dem das C_α-Atom und die CO-Gruppe eines Residuums, sowie die NH-Gruppe und das C_α-Atom des nächsten Residuums in einer Ebene liegen (in Abbildung 1.2a gestrichelt umrandet) [30, 31]. Das Peptidplättchen besitzt ein ausgeprägtes Dipolmoment, da die stark elektronegativen O- und N-Atome die Elektronenladungswolken zu sich ziehen. Die Starrheit und das Dipolmoment des Peptidplättchens schränken die Möglichkeiten für eine räumliche Anordnung des Proteinrückgrats stark ein.

Als *Sekundärstruktur* eines Proteins bezeichnet man dessen lokale Strukturmotive, die durch Wechselwirkungen im Proteinrückgrat stabilisiert werden. Dabei geht das partiell negativ geladene O-Atom der CO-Gruppe eines Residuums eine starke anziehende elektrostatische Wechselwirkung (*Wasserstoffbrückenbindung*) mit dem partiell positiv geladenen H-Atom der NH-Gruppe eines anderen Residuums ein. Äquivalent dazu kann man die Wasserstoffbrückenbindung auch als anziehende Wechselwirkung zwischen den elektrischen Dipolen der Peptidplättchen auffassen.

Linus Pauling, Robert Corey und Herman Branson entdeckten 1951 die we-

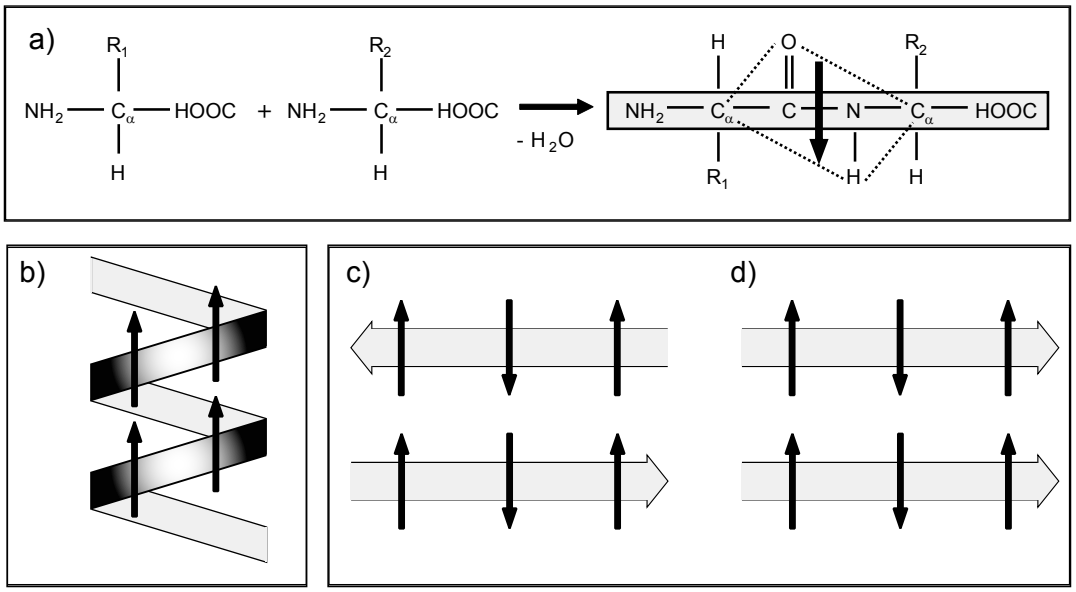

Abbildung 1.2: Aufbau von Proteinen. a) Zwei Aminosäuren bilden unter Abspaltung eines Wassermoleküls eine Peptidbindung aus. Dabei entsteht ein starres Peptidplättchen (gepunktet umrahmt) mit einem ausgeprägten Dipolmoment (Pfeil). Das Proteinrückgrat ist grau hinterlegt. Verlauf des Proteinrückgrats bei b) einer α-Helix, c) einem antiparallelen β-Faltblatt und d) einem parallelen β-Faltblatt. Diese wichtigen Sekundärstrukturmotive von Proteinen werden durch die Wechselwirkungen zwischen den Dipolen der Peptidplättchen (schwarze Pfeile), oder äquivalent dazu durch Wasserstoffbrückenbindungen, stabilisiert. Die Abbildungen basieren auf einer Abbildung in [32].

sentlichen Sekundärstrukturmotive von Proteinen [30], die durch eine regelmäßige Abfolge der Wasserstoffbrückenbindungen innerhalb des Proteinrückgrats charakterisiert werden. Abbildung 1.2b zeigt schematisch das Sekundärstrukturmotiv einer α-Helix. Hier ist das Proteinrückgrat spiralförmig gewunden und durch Wasserstoffbrückenbindungen zwischen einem Residuum n und dem Residuum n+4 stabilisiert. Die Abbildungen 1.2c und d zeigen schematisch das zweite wichtige Sekundärstrukturmotiv, das β-Faltblatt (engl. *β-sheet*), in den Varianten eines *antiparallelen* bzw. eines *parallelen* β-Faltblatts. Ein β-Faltblatt setzt sich aus zwei Abschnitten, den Strängen (engl. *strands*), zusammen. Je nachdem, ob beide Stränge eines Faltblatts in die gleiche Richtung laufen, d.h. beide vom N- zum C-Terminus oder in entgegengesetzte Richtungen, entsteht ein paralleles oder antiparalleles β-Faltblatt. Im Gegensatz zur α-Helix kann ein β-Faltblatt auch Abschitte aus nicht benachbarten Teilen des Proteins umfassen.

Die dreidimensionale Anordnung des Proteins, d.h. auch die Anordnung der Sekundärstrukturmotive zueinander, wird als *Tertiärstruktur* bezeichnet. Die native Sekundär- und Tertiärstruktur eines Proteins ist nach einem zentralen Dogma der

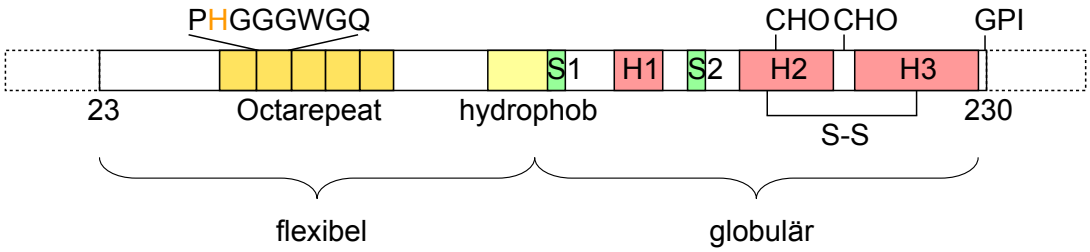

Abbildung 1.3: Wichtige Abschnitte der Sequenz des (humanen) PrP. Zwei Signalsequenzen (Res 1-22) und (Res 231-253) werden nach der Synthese des PrP entfernt. Merkmale des maturierten PrP (Res 23-230) sind eine kupferbindende Octarepeat-Region (Res 51-91), eine hydrophobe Domäne (Res 113-128), eine Schwefelbrücke (Cys-179, Cys-214) und zwei Bindungsstellen für Zuckerketten (Asn-181, Asn-197). Das zelluläre Prion-Protein PrP^C besteht aus einem flexiblen (Res 23-124) und einem globulären, strukturierten Abschnitt (Res 125-228). Letzterer enthält drei α-Helizes (H1, H2, H3) und ein β-Faltblatt, bestehend aus zwei Strängen (S1, S2). Die Abbildung basiert auf einem Entwurf von T. Hirschberger.

Proteinfaltung [33] schon durch seine Primärstruktur festgelegt und entspricht dem globalen Minimum der freien Energie. Im Falle des Prion-Proteins sind jedoch, wie oben erwähnt, zwei stabile dreidimensionale Strukturen bekannt.

1.3 Das Prion-Protein

Das Prion-Protein (PrP) kommt bei einer Vielzahl von Spezies vor und tritt in großen Mengen an den synaptischen Endplättchen von Nervenzellen auf. Über die physiologische Funktion des PrP ist bislang noch wenig bekannt. Da PrP Kupfer bindet, wird das Protein oft mit der Regulierung des Kupferhaushalts in Verbindung gebracht [34]. Neuere Forschungsergebnisse geben Hinweise darauf, dass es auch eine Rolle bei der Neuroprotektion von Nervenzellen spielen könnte [35].

Da die Primärstrukturen fast aller PrP-Varianten dem menschlichen PrP stark ähneln [36], beziehen sich im Folgenden die Angaben zu Residuen auf die menschliche Variante und können leicht auf andere Spezies übertragen werden. Abbildung 1.3 zeigt die wichtigsten Abschnitte der Sequenz. PrP besitzt nach seiner Synthese im Ribosom 253 Residuen. Im Endoplasmatischen Retikulum wird die N-terminale Signalsequenz entfernt (Res 1-22), Zuckerketten (CHO) angehängt, sowie eine Schwefelbrücke ausgebildet. Nach der Entfernung der C-terminalen Signalsequenz (Res 231-253) wird ein GPI (glycosylphosphatidylinositol)-Anker hinzugefügt, mit dem das zelluläre PrP^C schließlich an der Zellmembran verankert wird [37]. Weitere Merkmale von PrP sind eine kupferbindende Octarepeat Region (Res 51-91) sowie eine hydrophobe Domäne (Res 113-128). Obwohl sich die Primärstrukturen der beiden Isoforme des Prion-Proteins, PrP^C und PrP^{Sc}, nicht unterscheiden, weisen die

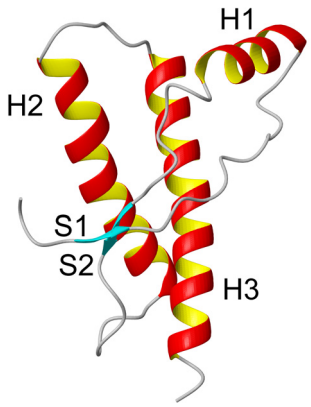

Abbildung 1.4: NMR-Struktur des globulären Teils des humanen Prion-Proteins [39]. Zu sehen sind drei spiralförmig gewundene α-Helizes und ein β-Faltblatt, zusammengesezt aus zwei Strängen (Pfeile). Die Koordinaten für diese Abbildung sind der Protein Data Bank [40] entnommen, Eintrag 1QM2. Die Abbildung wurde mit dem Grafikprogramm MOLMOL [41] erstellt.

zugehörigen dreidimensionalen Strukturen große Unterschiede auf.

In Kernspinresonanz-Messungen (NMR, vom engl. *nuclear magnetic resonance*) zeigte sich, dass PrP^C aus einem N-terminalen flexiblen (Res 23-124) und einem C-terminalen globulären, strukturierten Abschnitt (Res 125-228) besteht [38, 39]. Im globulären Teil des PrP^C befinden sich drei α-Helizes (H1, H2, H3) (rot in Abbildung 1.3) und ein antiparalleles β-Faltblatt, bestehend aus den zwei Strängen S1 und S2 (grün in Abbildung 1.3). Die dreidimensionale Struktur des globulären Teils des humanen PrP^C ist in Abbildung 1.4 dargestellt. Zu sehen ist hier nur das Proteinrückgrat mit den Sekundärstrukturmotiven der drei spiralförmig gewundenen α-Helizes (rot) und das β-Faltblatt, dessen Stränge (grün) als Bänder mit Pfeilen dargestellt sind, die in Richtung des C-Terminus der Aminosäuresequenz zeigen.

Die genaue dreidimensionale Struktur des fehlgefalteten PrP^{Sc} ist im Gegensatz zu der von PrP^C bis jetzt noch nicht aufgeklärt. Dies liegt daran, dass sich hier die atomare Strukturaufklärung durch experimentelle Verfahren als schwierig erweist, da die PrP^{Sc}-Fibrillen zu wenig geordnet für Röntgenbeugungsmessungen und zu wenig löslich für die NMR-Spektroskopie sind. Es konnte jedoch gezeigt werden [42, 43], dass PrP^{Sc} einen großen Anteil an β-Faltblattstrukturen aufweist. In Modellen für die dreidimensionale Struktur von PrP^{Sc} [44–46] wird dies berücksichtigt.

Man geht davon aus, dass sich PrP^C nach einem autokatalytischen Mechanismus in PrP^{Sc} umfaltet und aggregiert [20, 47–49]. Dabei lagert sich PrP^C an ein spontan umgefaltetes PrP^{Sc} an und faltet sich daraufhin ebenfalls zu PrP^{Sc} um. Das neu entstandene PrP^{Sc} dient wiederum als eine Art Schablone für die Umfaltung von PrP^C in PrP^{Sc}. Der Krankheitsverlauf lässt darauf schließen, dass erst nach der Bildung eines sogenannten Aggregationskeims, der ein stabiles Oligomer aus umgefalteten

Proteinen ist, ein schnelles Wachstum von Ablagerungen stattfindet [47].

Ablagerungen, wie sie in den Gehirnen TSE-Kranker auftreten, nennt man Amyloid-Plaques. Die aggregierenden Proteine, wie bei TSE das Prion-Protein, bezeichnet man als amyloidogen. Entnimmt man aus dem Gehirn Proben mit PrP^{Sc}-Material und gibt ein bestimmtes Enzym hinzu, so erhält man fadenförmige Amyloid-Fibrillen, die in ähnlicher Form auch bei anderen neurodegenerativen Erkrankungen auftreten [42, 50]. Die charakteristischen spongiformen Veränderungen im Gehirn (vgl. Abbildung 1.1) sind mit dem Auftreten von PrP^{Sc} korreliert [5], jedoch ist noch nicht bekannt, wie sie genau entstehen. Ebenso ist nicht geklärt, ob die bei TSE auftretenden Symptome auf einer verminderten Menge an PrP^{C} und dem damit einhergehenden Verlust einer möglichen Funktion von PrP^{C} (*loss-of-function*) beruhen, oder darauf, dass das aggregierte PrP^{Sc} direkt toxische Wirkung besitzt (*gain-of function*). Es gibt allerdings Anzeichen dafür, dass die toxische Wirkung eher von kleineren oligomeren Aggregaten als von größeren Amyloid-Plaques herrührt.

Die Konzepte zur Vorbeugung und Therapie der TSE-Erkrankungen sind vielfältig. Man könnte das Prion-Protein in seiner zellulären Form stabilisieren und damit die Entstehung von PrP^{Sc} von vornherein verhindern. Dazu müsste man die Schwachstellen von PrP^{C} kennen, die seine Umfaltung in die PrP^{Sc}-Isoform begünstigen. Am anderen Ende des Krankheitsmechanismus ansetzend, könnte man versuchen, vorhandene Amyloid-Plaques aufzulösen. Hierzu müssten wiederum die Schwachstellen von PrP^{Sc} bekannt sein. Bei diesem Ansatz besteht fatalerweise die Gefahr, dass aus den eventuell ungiftigen Plaques kleinere toxische Aggregate abgespalten werden, die den Krankheitsverlauf noch beschleunigen könnten. Weiterhin könnte man versuchen, den autokatalytischen Prozess der Umfaltung zu stören, indem man die Anlagerung von PrP^{C} an vorhandene PrP^{Sc}-Aggregate verhindert [51]. Für alle drei Angriffspunkte wäre es von großem Vorteil, mehr über die dreidimensionale Struktur von PrP^{Sc} zu wissen. Seit geraumer Zeit wird daher versucht, die Struktur von PrP^{Sc}-Aggregaten experimentell aufzuklären.

1.4 Strukturaufklärung von Amyloid-Fibrillen

Auch bei anderen neurodegenerativen Erkrankungen wie Alzheimer, Parkinson und Huntington kommt es zur Bildung von Amyloid-Plaques und -Fibrillen. Im Unterschied zu den TSE-Erkrankungen sind diese Erkrankungen aber nicht ansteckend [52]. In den auftretenden Ablagerungen im Gehirn befindet sich ein für die jeweilige Erkrankung charakteristisches Protein oder Peptid. Bei TSE ist dies, wie schon in Kapitel 1.3 beschrieben wurde, das Prion-Protein, bei der Alzheimer-Erkrankung das Aβ-Peptid, bei Parkinson das Protein α-Synuklein und bei der Huntingtonschen Krankheit das Exon-1 Peptid des Huntingtin. Obwohl diese Proteine und Peptide eine sehr unterschiedliche Abfolge von Aminosäuren besitzen, weisen die Ablagerungen große Gemeinsamkeiten auf. Einiges deutet darauf hin, dass allen Ablagerungen ein ähnliches Strukturmotiv zu Grunde liegt, das wenig von den spezifischen Seiten-

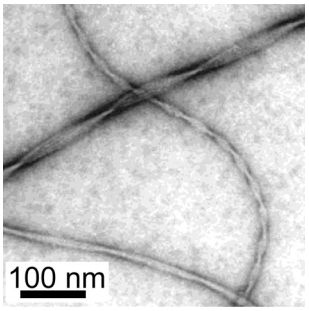

Abbildung 1.5: Elektronenmikroskopische Aufnahme von Amyloid-Fibrillen, die aus einem Fragment des mit Alzheimer assoziierten Aβ-Peptids bestehen. Abbildung entnommen aus [55].

ketten der Proteine abhängt [53–55].

In *elektronenmikroskopischen* (EM) Aufnahmen vieler Amyloid-Fibrillen erkennt man gleichmäßige, unverzweigte Fasern, die zwischen 0.1 und 10 μm lang sind [56]. Die Durchmesser vieler Fasern liegen im Bereich von 100 Å [50, 56], wovon es jedoch auch größere Abweichungen gibt. In Abbildung 1.5 sind EM-Aufnahmen von Fibrillen aus dem mit Alzheimer assoziierten Aβ-Peptid zu sehen. Neben dem typischen Erscheinungsbild unter dem Elektronenmikroskop charakterisiert man Amyloid durch die Bindung des Farbstoffs Kongo-Rot und die dabei zu beobachtende grüne Doppelbrechung unter polarisiertem Licht [5]. Diese Eigenschaft ist ein Indiz für einen hohen Anteil an β-Faltblatt-Struktur [5].

Weitere Gemeinsamkeiten in der räumlichen Struktur von Amyloid-Fibrillen zeigen sich in *Röntgenbeugungsmessungen* [56, 57]. Röntgenlicht (Wellenlänge 0.1 Å bis 100 Å) wird hierbei senkrecht zur Längsachse der Fibrillen eingestrahlt und das auftretende Beugungsmuster des ausgehenden Lichtes analysiert. Abbildung 1.6 zeigt ein für viele Amyloid-Strukturen charakteristisches Röntgenbeugungsmuster [53]. Aus dem *meridionalen* Reflex kann man auf die Struktur in Richtung der Fibrillenachse und aus dem *äquatorialen* Reflex auf die Struktur senkrecht zur Fibrillenachse schließen. Der scharfe meridionale Reflex bei 4,75 Å ist charakteristisch für den Abstand zwischen den Hauptketten zweier Stränge eines β-Faltblatts. Dieser Abstand wird von der festen Länge der Wasserstoffbrückenbindungen zwischen den O- und H-Atomen in einem β-Faltblatt (vgl. Abbildungen 1.2c und d) bestimmt. Ein derartiger meridionaler Reflex deutet auf eine sogenannte cross-β-Struktur hin, in der die Stränge der β-Faltblätter senkrecht zur Faserachse und die Wasserstoffbrückenbindungen parallel zur Faserachse verlaufen [50].

Abbildung 1.6b zeigt eine entsprechende Struktur. Der äquatoriale Reflex bei 10 Å entspricht einer parallelen Anordnung von Hauptketten, bei der die zugehörigen Seitenketten den Zwischenraum ausfüllen. Solch ein Beugungsmuster entsteht z.B., wenn mehrere β-Faltblätter hintereinander angeordnet sind. Abbildung 1.6c zeigt die Sicht entlang der Faserachse der in Abbildung 1.6b gezeigten Faser, die mit dem äquatorialen Reflex vereinbar ist. Bei einigen Amyloid-Fibrillen, wie z.B. bei

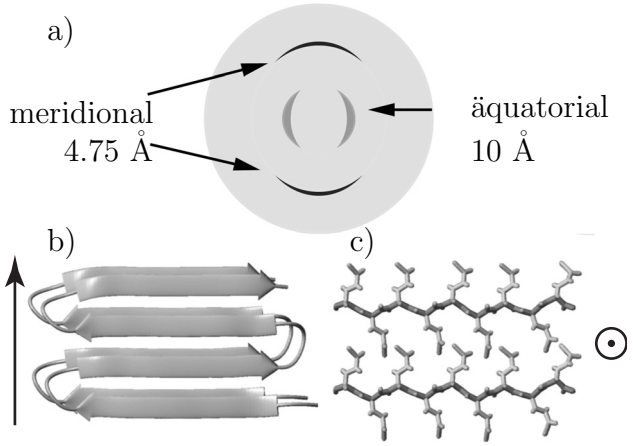

Abbildung 1.6: a) Röntgenbeugunsmuster vieler Amyloid-Fibrillen in stark vereinfachter Darstellung. Der *meridionale* Reflex beschreibt die Struktur in Richtung der Faserachse, der *äquatoriale* Reflex die Struktur senkrecht zur Faserachse. Eine Struktur, die mit beiden Reflexen vereinbar ist, wird hier mit der Blickrichtung b) senkrecht zur und c) entlang der Fibrillenachse gezeigt. b) Der meridionale Reflex ist charakteristisch für den Abstand zweier Stränge eines β-Faltblatts und weist auf eine regelmäßige Anordnung von β-Faltblättern senkrecht zur Faserachse hin. c) Der äquatoriale Abstand ist charakteristisch für hintereinader angeordneten Hauptketten (dunkelgrau) bei denen die Seitenketten (hellgrau) den Zwischenraum einnehmen.

Fibrillen aus dem exon-1 Peptid des Huntingtin, fehlt allerdings der äquatoriale 10 Å-Reflex im Röntgenbeugungsmuster [58, 59]. Dies deutet auf eine Anordnung von β-Faltblättern hin, die sich grundlegend von der in Abbildung 1.6c gezeigten unterscheidet.

Die Röntgenbeugung an Fibrillen gibt zwar Hinweise auf grundsätzliche Strukturmotive, wie die Ausrichtung von β-Faltblättern, erlaubt aber keine atomare Auflösung. Mit der *Röntgenstrukturanalyse* [60] an Protein-Kristallen konnten hingegen schon viele Protein-Strukturen bestimmt werden. Dazu muss das Protein kristallisiert werden, was bei Amyloid-Fibrillen noch nicht in für eine Messung ausreichender Form gelang. Jedoch konnte die Struktur eines Enzyms, das zu einem großen Teil aus β-Faltblattstrukturen besteht [61] und möglicherweise Ähnlichkeiten mit Bausteinen von Amyloid-Fibrillen besitzt [44, 45], aufgeklärt werden.

Eine weitere bewährte Technik zur Strukturaufklärung von Proteinen ist die *NMR-Spektroskopie in Lösung* [62], mit der z.B. die dreidimensionale Struktur des Prion-Proteins (Abbildung 1.4) aufgeklärt werden konnte [39]. Die Strukturaufklärung mittels NMR beruht auf der Dipol-Dipol-Kopplung zwischen Kernspins [28]. Dadurch kann der Abstand zweier Wasserstoffatome ermittelt werden, sofern sie weniger als 5 Å voneinander entfernt sind. Die gewonnenen Messwerte müssen dann den einzelnen Atompaaren zugeordnet werden. Anschließend wird mittels Computersimulation in Verfahren des *Structure Refinement* [63, 64] versucht, diejenige

dreidimensionale Struktur zu finden, die möglichst gut alle vorgegebenen Abstände erfüllt. Die Koordinaten der mittels NMR-Messungen und Röntgenstrukturanalyse bestimmten Strukturen sind unter einem Buchstabencode (1QM2 für das PrP^C in Abbildung 1.4) in der Protein Data Bank abgelegt [40] und im Internet frei verfügbar. Amyloid-Fibrillen sind jedoch für eine NMR-Messung in Lösung auf Grund ihrer geringen Löslichkeit nicht geeignet.

Größeren Erfolg verspricht hier das Verfahren der *Festkörper-NMR* [56, 65, 66], bei dem die Probe zur Messung nicht verdünnt werden muss. Dies bringt jedoch den Nachteil mit sich, dass sich die Probe nicht schnell und isotrop wie in einer Flüssigkeit bewegt, und resultiert in der schlechteren Auflösung von ca. 6 Å für atomare Abstände [56]. In den letzten Jahren wurden zunehmend Modelle für Amyloid-Fibrillen entwickelt, die auf Festkörper-NMR-Messungen beruhen. So konnte z.B. für das bei Alzheimer auftretende Aβ-Peptid gezeigt werden, dass sich je nach Länge des beteiligten Peptid-Fragments entweder parallele oder antiparallele β-Faltblätter bilden [65, 66].

Ein weiteres Verfahren zur Strukturbestimmung ist die *paramagnetische Elektronen-Spin-Resonanz (EPR)-Spektroskopie*, die auf der Dipol-Dipol-Kopplung von Elektronen beruht. Mit ihrer Hilfe konnten ebenfalls Aussagen über die Anordnung von β-Faltblatt-Strukturen in Fibrillen aus Fragmenten des Aβ-Peptids [67], sowie in Fibrillen aus α-Synuklein getroffen werden [68].

Besitzt man erste Anhaltspunkte für die Struktur von Amyloid-Fibrillen oder kleineren Aggregaten, bietet die Methode der *Molekulardynamik (MD)-Simulation* [69–72] geeignete Verfahren [63, 64], Strukturvorschläge zu überprüfen und zu verfeinern. Für Peptide aus poly-Gln, die den möglichen Aggregationskeim für die Fibrillen des exon-1 Peptids des Huntingtin bilden, wurde von Perutz et al. [58] eine Hypothese für die Grundstruktur der Fibrillen aufgestellt. In dieser Arbeit werde ich in MD-Simulationen auf der Grundlage des vorgeschlagenen Modells eine alternative Struktur entwickeln, die sich, im Gegensatz zum ursprünglich vorgeschlagenen Modell, in MD-Simulationen als stabil erweist. Ebenso werde ich für die Struktur von PrP^{Sc}-Fibrillen, basierend auf den Ergebnissen von EM-Messungen an zweidimensionalen Kristallstrukturen [44], ein detailliertes Modell für deren Grundstruktur entwickeln.

Ist die Struktur eines Aggregats oder des zugehörigen löslichen Proteins sogar in atomarer Auflösung verfügbar [40], kann in MD-Simulationen dessen Konformationsdynamik untersucht werden, um so strukturelle Schwachstellen und damit potentielle Angriffspunkte für eine Therapie zu identifizieren. In dieser Arbeit werde ich in MD-Simulationen die Konformationsdynamik des PrP^C und zweier Varianten untersuchen, um so Faktoren für die Strukturstabilität von PrP^C zu ermitteln. Im Folgenden werde ich daher die Methode der MD-Simulation vorstellen.

1.5 Grundlagen der Molekulardynamik-Simulation

In MD-Simulationen lässt sich die thermische Bewegung von Molekülen in ihrer Lösungsmittelumgebung in atomarer Auflösung betrachten. Die Funktionsweise und Struktur von Biomolekülen wie Proteinen oder Peptiden ist stark von ihrer Umgebung abhängig. In einer MD-Simulation muss deshalb die Lösungsmittelumgebung möglichst gut die native Umgebung des Biomoleküls widerspiegeln. Diese kann oft als wässrige, zum Teil auch ionenhaltige Lösung bei physiologischen Werten für Druck und Temperatur beschrieben werden [46, 73]. Wasser geht als sehr polares Lösungsmittel starke elektrostatische Wechselwirkungen mit den dipolaren Peptidgruppen im Proteinrückgrat und geladenen oder polaren Seitenketten eines Proteins ein. Die korrekte Beschreibung langreichweitiger elektrostatischer Wechselwirkungen ist deshalb in MD-Simulationen von zentraler Bedeutung [74]. Abhängig von dem untersuchten Prozess sind unterschiedliche physikalische Modelle für die Beschreibung des Proteins und seiner Umgebung sinnvoll.

Will man chemische Reaktionen beobachten, also Prozesse, die das Bilden und Brechen chemischer Bindungen beinhalten, muss das System *quantenmechanisch* beschrieben werden. Hier werden die Elektronen durch ihre Wellenfunktion dargestellt. Auf Grund des großen Rechenaufwandes ist dabei aber die Größe des Systems auf wenige 100 Atome beschränkt. Interessiert man sich nur für die Konformationsdynamik von Molekülen, lässt sich das Simulationssystem gut im Rahmen der klassischen Mechanik beschreiben. Bei solchen *molekularmechanischen* (MM) MD-Simulationen wird die Wechselwirkung der Atome von einer analytisch vorgegebenen und geeignet parametrisierten Energiefunktion, dem sogenannten Kraftfeld, bestimmt [70]. Da die mikroskopische Beschreibung des Lösungsmittels einen Großteil der Rechenzeit in Anspruch nimmt, gibt es vielfältige Ansätze (z.B. [75, 76]), das Lösungsmittel makroskopisch als dielektrisches Kontinuum zu betrachten [77–81].

1.5.1 Molekularmechanik

In einer MM-MD-Simulation von Molekülen wird die Bewegung der Atome durch eine Energiefunktion $E(\mathbf{R})$ bestimmt, deren genaue Form in einem *Kraftfeld* [82, 83] angegeben ist. Ein Überblick über verschiedene Kraftfelder findet sich in [84]. Durch Gradientenbildung bezüglich der Atomorte lässt sich aus $E(\mathbf{R})$ zu jeder Konfiguration $\mathbf{R} \equiv (\mathbf{r}_1, \ldots, \mathbf{r}_N)$ eines Simulationssystems, wobei $(\mathbf{r}_1, \ldots, \mathbf{r}_N)$ die Positionen aller N Atome sind, die resultierende Kraft auf jedes Atom ermitteln. Die Bewegung der Atome ergibt sich dann aus der numerischen Integration der Newtonschen Bewegungsgleichungen, beispielsweise mittels des Verlet-Algorithmus [85]. Die Integrationsschrittweite bei der numerischen Integration ist durch die Bewegung der schnellsten Freiheitsgrade im System nach oben beschränkt und liegt im Bereich von einer Femtosekunde.

In dieser Arbeit wurde das Kraftfeld CHARMM22 [82] verwendet[1] , dessen Ener-

[1] In einem von mir betreuten Projekt wurde der Konvertierer GROMACS2LIS entwickelt, der es

giefunktion für ein System von N Atomen die folgende Form besitzt (vgl. auch [64]):

$$\begin{aligned} E(\mathbf{R}) = & \sum_{\text{Bindungen}} K_b(b-b_0)^2 + \sum_{\text{Winkel}} K_\theta(\theta-\theta_0)^2 \\ & + \sum_{\substack{\text{Improper}\\ \text{Dihedralwinkel}}} K_\xi(\xi-\xi_0)^2 + \sum_{\substack{\text{Proper}\\ \text{Dihedralwinkel}}} \sum_{i=1}^{m} K_{\phi_i}[1+cos(n_i\phi_i-\phi_{i,0})] \\ & + \sum_{\substack{\text{nicht}-\text{geb.}\\ \text{Atompaare } (i,j)}} \left\{ 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] + \frac{q_i q_j}{\varepsilon_s r_{ij}} \right\}. \end{aligned} \tag{1.1}$$

Die ersten beiden Terme der rechten Seite von Gleichung (1.1) geben die Energien chemischer Bindungen zwischen zwei Atomen sowie die Energien von Winkeln zwischen drei Atomen in Form von harmonischen Potentialen an. Kovalente Bindungen der Länge b sowie Winkel θ führen harmonische Schwingungen um ihre Gleichgewichtswerte b_0 bzw. θ_0 aus, bestimmt durch die zugehörigen Kraftkonstanten K_b bzw. K_θ. Die Torsion um eine Bindung wird durch Dihedralwinkel beschrieben, deren Energien je nach Art des Winkels über harmonische Potentiale oder Kosinusfunktionen gegeben sind [dritter und vierter Term in Gleichung (1.1)]. Durch *improper* Dihedralwinkel werden Atomgruppen, die nur wenig um ihren Gleichgewichts-Dihedralwinkel ξ_0 ausgelenkt werden, mittels harmonischer Potentiale mit den Kraftkonstanten K_ξ annähernd planar gehalten. Sogenannte *proper* Dihedralwinkel besitzen mehrere Energieminima und sind durch eine Kosinusfunktion oder durch die Summe von m (*Multiplizität*) Kosinusfunktionen parametrisiert. Hier bestimmt die Periodizität n_i die Anzahl der Energieminima eines Summanden und der Phasenfaktor $\phi_{i,0}$ die Lage der Minima.

Atome, die nur indirekt miteinander verbunden sind und zwischen denen mindestens drei kovalente Bindungen liegen, sowie Atome, die verschiedenen Molekülen angehören, wechselwirken über Lennard-Jones- und Coulombpotentiale [5. und 6. Term in Gleichung (1.1)], deren Stärken von ihrem Abstand r_{ij} abhängen.[1] Die Beschränkung auf nicht-gebundene Atome berücksichtigt, dass die entsprechenden Wechselwirkungen zwischen gebundenen Atomen schon in die Bindungs- und Winkelenergien eingehen. Im Lennard-Jones-Potential wird das bindende Potential ($\sim r_{ij}^{-6}$), welches aus der Dispersionswechselwirkung zwischen induzierten atomaren Dipolen resultiert, gemeinsam mit einem antibindenden Potential ($\sim r_{ij}^{-12}$) parametrisiert.

ermöglicht, im MD-Programm EGO-MMII neben CHARMM22 auch andere Kraftfelder zu verwenden.

[1] Die elektrostatische Wechselwirkung zweier Atome, die genau durch drei kovalente Bindungen verbunden sind, wird in einigen Kraftfeldern [83, 86] durch einen Skalierungsfaktor abgeschwächt. In andere Kraftfelder [82, 87], wie auch in CHARMM22, geht sie unskaliert ein.

Dieses stark abstoßende Potential gibt die Pauli-Abstoßung zwischen den Elektronenwolken zweier chemisch nicht gebundener Atome wieder. Meist, und im Folgenden auch hier, wird die Summe beider Anteile als van-der-Waals (vdW)-Energie bezeichnet. Die vdW-Energie wird durch ihre Potentialtiefe ϵ_{ij} und ihre Nullstelle bei σ_{ij} parametrisiert. Die Werte ϵ_{ij} und σ_{ij} ergeben sich aus der Kombination der Parameter $\epsilon_{ii} \equiv \epsilon_i$ und $\sigma_{ii} \equiv \sigma_i$ für die Wechselwirkung zwischen zwei identischen Atomtypen zu $\epsilon_{ij} = \sqrt{\epsilon_i \epsilon_j}$ und $\sigma_{ij} = (\sigma_i + \sigma_j)/2$. Aus der Breite σ_i lässt sich der vdW- oder Hartkugelradius R_i^{vdW} eines Atoms i berechnen, dessen Zusammenhang mit σ_i in der Literatur jedoch nicht einheitlich definiert ist (vgl. z.B. die unterschiedlichen Definitionen in [64, 86]). In dieser Arbeit wird als vdW-Radius eines Atoms die Hälfte des Abstandes zweier Atome identischen Typs bezeichnet, bei dem das Lennard-Jones Potential sein Minimum ϵ_i annimmt. Er berechnet sich aus Gleichung (1.1) zu

$$R_i^{vdW} = \sigma_i \sqrt[6]{2}/2. \tag{1.2}$$

Die elektrostatische Energie zwischen zwei Atomen i und j ergibt sich aus der Coulombwechselwirkung ihrer Punktladungen q_i und q_j, abgeschirmt durch die Dielektrizitätskonstante ϵ_s des Moleküls, die meist zu $\epsilon_s \equiv 1$ gesetzt wird. Bei der Parametrisierung elektrostatischer Eigenschaften eines Atoms durch eine Punktladung am Atomort bleibt unberücksichtigt, dass sich die Elektronenladungswolken im elektrischen Feld anderer Atome verschieben. Um diesen Effekt zu berücksichtigen, werden sogenannte polarisierbare Kraftfelder [72, 84] entwickelt. So wird in unserer Arbeitsgruppe die Polarisierbarkeit durch veränderliche normalverteilte Dipoldichten $\mathbf{P}_i$ an den Orten der Atome berücksichtigt [88, 89]. Die Dipole werden bei einer spezifischen Konfiguration $\mathbf{R}$ des Moleküls vom elektrischen Feld aller anderen Dipole und Punktladungen induziert und ergeben sich, unter Berücksichtigung Atom-spezifischer Polarisierbarkeiten α_i, z.B. durch eine selbstkonsistente Berechnung [89, 90].

Die Coulomb- und vdW-Wechselwirkungen sind Paarwechselwirkungen zwischen allen nicht-gebundenen Atomen sowohl des gelösten Moleküls als auch der Lösungsmittelmoleküle. Durch die Vielzahl dieser Wechselwirkungen ergibt sich ein sehr großer Rechenaufwand. Da man nicht ein Protein in einer unnatürlich eng begrenzten Wasserumgebung beschreiben möchte, setzt man das Simulationssystem periodisch in alle Raumrichtungen fort. Die vdW-Wechselwirkungen sind auf Grund des r^{-6}-Terms kurzreichweitig und werden deshalb nur zwischen Atomen explizit berechnet, deren Abstand geringer als einige (~ 10) Å ist. Der mittlere Einfluss von weiter entfernten Atomen wird durch einen Korrekturterm berücksichtigt [69]. Die elektrostatischen Wechselwirkungen dagegen sind auf Grund des r^{-1}-Terms langreichweitig und können nicht schon bei kleineren Abständen vernachlässigt werden [70, 91]. Gittersummationsmethoden wie die Ewald-Methode [92, 93] nutzen die Periodizität des Systems geschickt bei der Berechnung der elektrostatischen Kräfte aus. Alternativ können in schnellen Multipolmethoden die Punktladungen zu Ladungsverteilungen zusammengefasst und das Potential in größerer Entfernung durch eine Multipolent-

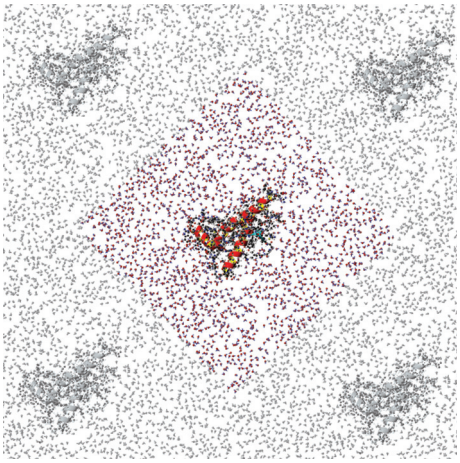

Abbildung 1.7: MM-MD-Simulationssystem eines Proteins in expliziter wässriger Lösung. Das Grundsystem (in Farbe) besteht aus einem Protein (Mitte), hier PrP^c aus Abbildung 1.4, umgeben von Wassermolekülen. Es wird periodisch in alle Raumrichtungen fortgesetzt (grau).

wicklung genähert werden, wie es auch in unserem Simulationsprogramm EGO-MMII geschieht [91, 94, 95]. Die Kräfte auf die einzelnen Atome werden vermittels einer Taylorentwicklung aus dieser Multipolentwicklung berechnet, so dass sich ein linear mit der Systemgröße skalierender Algorithmus ergibt. Um Artefakte auf Grund der künstlichen Periodizität zu vermeiden, werden hier nur Wechselwirkungen zwischen Atomen berücksichtigt, die sich innerhalb einer Kugel mit einem vorgegebenen Abschneideradius befinden, der kleiner ist als der Radius einer dem Grundsystem einbeschriebenen Kugel. Der mittlere Einfluss des dielektrischen Lösungsmittels außerhalb der Abschneidekugel wird durch eine Kirkwood-Reaktionsfeldkorrektur berücksichtigt [91].

Rechenzeitbegrenzung bei MD-Simulationen

Ein System, das in dieser Arbeit in MM-MD-Simulationen untersucht wird (Kapitel 2), ist das Prion-Protein in Lösung. Abbildung 1.7 zeigt das Grundsystem, das sich aus 1694 Proteinatomen, 25775 Wassermolekülen und 153 (Na^+, Cl^-) Ionen zusammensetzt. Trotz der geschickten Behandlung langreichweitiger Wechselwirkungen lässt es sich selbst auf modernen Rechenclustern nur wenige Nanosekunden lang simulieren. Wird die Rechnung parallel auf sechs (1.6 GHz) Prozessoren durchgeführt, benötigt man für 10 ns simulierte Zeit ca. 7 Wochen Rechenzeit. Wie ich später zeigen werde, war dieser Zeitraum in diesem Fall lang genug, um Instabilitäten in der Proteinstruktur identifizieren zu können [73], aber nicht ausreichend lang, um Faltungsprozesse, die sich auf Zeitskalen von μs (β-helikale Peptide) bis Minuten (Proteine) abspielen [96], beobachten zu können. Ebenso reichte, wie in Kapitel 3

gezeigt werden wird, in einer MM-MD-Simulation von Peptiden aus poly-Gln in wässriger Lösung [46] eine Simulationszeit von 10 ns aus, um thermodynamisch *instabile* Peptidstrukturen identifizieren zu können. Für amyloidogene Peptide, die auf dieser Zeitskala *stabil* sind, möchten wir aber wesentlich längere Simulationen durchführen. Erst dann wird sich herausstellen, ob diese Strukturen lange genug stabil sind, um Aggregationskeime für amyloide Fasern bilden zu können, und ob sich auch tatsächlich andere Peptide an sie anlagern.

Die Simulationszeiten könnten deutlich ausgedehnt werden, wenn die aufwändige Berechnung elektrostatischer Wechselwirkungen zwischen dem Protein oder Peptid und den mikroskopischen Wasserkonfigurationen entfallen würde. Wie ich schon weiter oben erklärt habe, ist das Wasser für die Konformationsdynamik des Biomoleküls extrem wichtig und kann deshalb nicht einfach weggelassen werden. Jedoch kann man versuchen, die Wasserumgebung eines Proteins als makroskopisches Kontinuum zu beschreiben. Da sich gängige Kontinuumsmethoden für den Einsatz in MD-Simulationen als ungeeignet erweisen, werde ich eine vielversprechende Kontinuumsmethode weiterentwickeln. Im Folgenden stelle ich das Konzept der Kontinuumsmethoden vor, gehe auf die Nachteile gängiger Methoden ein und beschreibe in Grundzügen unseren Ansatz für die makroskopische Betrachtung polarer Lösungsmittel.

1.5.2 Kontinuumsmethoden

Die Reduzierung der Rechenzeit bei Kontinuumsmethoden beruht auf zwei Effekten. Zum einen ist die Rechenzeit für einen Integrationsschritt bei manchen [75], allerdings nicht allen [97] Kontinuumsmethoden erheblich kürzer, als bei einem expliziten Lösungsmittel. Zum anderen beschreiben Kontinuumsmethoden den mittleren Einfluss des Lösungsmittels auf eine bestimmte Proteinkonfiguration [78]. Für explizites Lösungsmittel müsste dafür bei einer Proteinkonfiguration eine aufwändige Mittelung über mikroskopische Wasserkonfigurationen durchgeführt werden. Ein Maß für den mittleren Einfluss des Lösungsmittels ist die freie Solvatationsenergie. Darunter versteht man die Differenz der freien Energie eines Moleküls im Vakuum und im Lösungsmittel. In polaren Lösungsmitteln bilden die elektrostatischen Wechselwirkungen den Hauptbeitrag zur freien Solvatationsenergie. Diese Wechselwirkungen werden deshalb im Folgenden näher betrachtet.

Abbildung 1.8a zeigt die entsprechende elektrostatische Beschreibung. Das gelöste Molekül, hier das Prion-Protein aus Abbildung 1.7, wird durch ein Volumen $\mathcal{V}_s$ mit einer niedrigen Dielektrizitätskonstante ε_s repräsentiert, welches von einem dielektrischen Kontinuum $\mathcal{V}_c$ mit einer hohen Dielektrizitätskonstante ε_c umgeben ist ($\varepsilon_c \approx 78$ für Wasser [99]). Die Wechselwirkungen innerhalb des Proteins werden unverändert durch das gegebene Kraftfeld (1.1) beschrieben und entsprechend wird für ε_s der Wert 1 gewählt.

Die Wechselwirkung zwischen Protein und Wasser beruht auf der Polarisation des dielektrischen Kontinuums durch die Ladungsverteilung $\rho(\mathbf{r}) = \sum_i q_i \delta(\mathbf{r} - \mathbf{r}_i)$

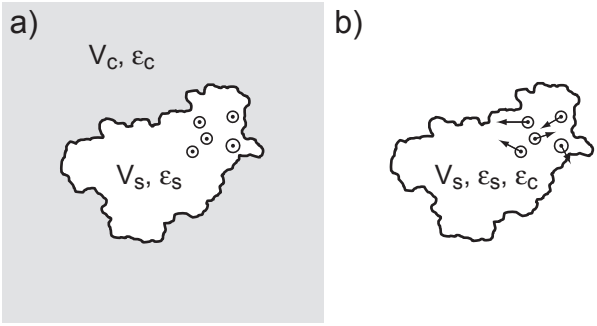

Abbildung 1.8: a) Konzept eines Kontinuummodells für ein Protein im polaren Lösungsmittel. Das Protein wird durch ein Volumen $\mathcal{V}_s$ (weiß) mit einer niedrigen Dielektrizitätskonstante ε_s beschrieben. Es ist umgeben von einem dielektrischen Kontinumm $\mathcal{V}_c$ (grau) mit einer hohen Dielektrizitätskonstante ε_c. Einige Proteinatome sind beispielhaft als Kugeln mit punktförmigen Partialladungen dargestellt. b) Konzept der Kontinuumsmethode von Egwolf und Tavan [98]. Die Polarisation des Dielektrikums, die durch die Ladungen im Protein hervorgerufen wird und ein elektrisches Feld am Ort des Proteins hervorruft, wird durch RF-Dipole (Pfeile) an den Atomorten repräsentiert.

im Protein. Das polarisierte Kontinuum erzeugt daraufhin ein elektrisches Feld am Ort des Proteins. Da dieses Feld die Reaktion des Lösungsmittels auf das Protein darstellt, wird es *Reaktionsfeld* (RF) genannt. Das gesamte elektrostatische Potential $\Phi(\mathbf{r})$ an einem Ort $\mathbf{r}$ ergibt sich aus der Lösung der *Poisson-Gleichung* [100]

$$\boldsymbol{\nabla}_\mathbf{r} \cdot \left[\varepsilon(\mathbf{r})\boldsymbol{\nabla}_\mathbf{r}\Phi(\mathbf{r})\right] = -4\pi\rho(\mathbf{r}), \tag{1.3}$$

in der die ortsabhängige Dielektrizitätsfunktion $\varepsilon(\mathbf{r})$ im Proteinvolumen $\mathcal{V}_s$ den Wert $\varepsilon(\mathbf{r}) \equiv \varepsilon_s$, außerhalb, in $\mathcal{V}_c$, den Wert $\varepsilon(\mathbf{r}) \equiv \varepsilon_c$ annimmt. Das Potential $\Phi(\mathbf{r})$ lässt sich darstellen als Summe aus dem mit ε_s abgeschirmten Coulombpotential der Partialladungen im Protein und dem RF-Potential, aus dem sich durch Gradientenbildung das Reaktionsfeld ergibt. Befinden sich Ionen im Lösungsmittel, tritt eine zusätzliche Abschirmung des Potentials auf und es muss statt (1.3) die Poisson-Boltzmann-Gleichung gelöst werden. Im Folgenden wird nur der Fall eines polaren Lösungsmittels betrachtet, eine ausführliche Darstellung für ionische Lösungsmittel findet sich in [101].

Analytische Lösungen der Poisson-Gleichung (1.3) sind nur für einfache geometrische Objekte gegeben, wie sie z.B. in EGO-MMII bei der Berechnung des Kirkwood-Reaktionsfeldes im Fall von Kugeln genutzt werden [91]. Für komplizierte Volumina, wie das Proteinvolumen $\mathcal{V}_s$ aus Abbildung 1.8a, können die Gleichungen nicht analytisch gelöst werden. Kontinuumsmethoden gehen dieses Problem grundsätzlich auf zwei verschiedenen Wegen an und lassen sich diesbezüglich in zwei Kategorien einteilen: Während *PB-Methoden* versuchen, die Poisson- oder Poisson-Boltzmann-Gleichung numerisch zu lösen, verwenden sogenannte *Generalisierte Born* (GB)-

Methoden einen analytischen Ausdruck für eine Reaktionsfeldenergie, der nur in wenigen ausgesuchten Grenzfällen den PB-Lösungen entspricht.

Bei PB-Methoden werden das Proteinvolumen $\mathcal{V}_s$ und angrenzende Bereiche aus $\mathcal{V}_c$ [102–104] oder nur die Oberfläche des Proteins [105] diskretisiert. Während bei manchen Methoden (finite Differenzen) [102, 103] die Diskretisierung des Systems starr ist, passen *adaptive* Methoden (finite Elemente [104] und Randelemente [105]) die Diskretisierung an die jeweilige Proteinstruktur dynamisch an. Besonders mit adaptiven PB-Methoden lassen sich sehr gute Ergebnisse bei der Berechnung der freien Solvatationsenergie statischer Proteinkonfigurationen erzielen [106], doch sind sie zumeist zu zeitaufwändig für den Einsatz in MD-Simulationen [81, 107]. Verschiedene Ansätze [108–110] zielen darauf ab, die Effizienz der PB-Methoden zu steigern.

GB-Methoden [75] dagegen sind speziell auf den Einsatz in MD-Programmen zugeschnitten. Hier wird eine Reaktionsfeldenergie ΔG^B aus einer paarweisen, abgeschirmten Wechselwirkung zwischen den Ladungen im Protein an Hand der empirischen Formel [75, 80]

$$\Delta G^B = -\frac{1}{2}\left(\frac{1}{\varepsilon_s} - \frac{1}{\varepsilon_c}\right) \sum_{i,j} \frac{q_i q_j}{\sqrt{r_{ij}^2 + \alpha_i \alpha_j \exp(-r_{ij}^2 / F \alpha_i \alpha_j)}} \tag{1.4}$$

berechnet, in der den sogenannten Born-Radien α_i der Atome eine entscheidende Rolle zukommt. F ist ein empirischer Faktor, der zumeist auf 4 gesetzt wird [75]. Im Fall zweier einander sehr naher Ladungen ($r_{ij} < 0.1\ \alpha_i \alpha_j$) ergibt sich aus (1.4) näherungsweise die Onsager RF-Energie [111] eines Dipols in einer sphärischen Kavität [75]. Im Grenzfall weit voneinander entfernter Ladungen ($r_{ij} > 2.5\ \alpha_i \alpha_j$) reduziert sich die Summe aus ΔG^B (1.4) und der Coulomb-Energie der Ladungen im Protein (letzter Term in (1.1)) auf die Summe aus den Born-Energien von Ionen im Dielektrikum [112]

$$W^B = -\frac{1}{2}\left(\frac{1}{\varepsilon_s} - \frac{1}{\varepsilon_c}\right) \sum_i \frac{q_i^2}{\alpha_i} \tag{1.5}$$

und den Coulomb-Energien

$$W^C = \frac{1}{2\varepsilon_c} \sum_{i,j} \frac{q_i q_j}{r_{ij}} \tag{1.6}$$

von Punktladungen im Kontinuum $\mathcal{V}_c$, die nun mit ε_c, statt mit ε_s wie in (1.1), abgeschirmt sind. Die verallgemeinerte Born-Energie (1.4) ist stark abhängig von der Parametrisierung der Born-Radien, deren Berechnung die größte Schwierigkeit in GB-Methoden darstellt [79, 80]. Zwar konnten mit GB-Methoden schon MD-Simulationen durchgeführt werden, die sich über mehrere hundert Mikrosekunden erstrecken [113], jedoch unterscheiden sich die resultierenden Konformationslandschaften von Peptiden zum Teil stark von denen in expliziten Lösungsmitteln [114, 115]. Dies schließt GB-Methoden für die Untersuchung der Konformationsdynamik unserer amyloidogenen Peptide aus.

Eine äußerst vielversprechende Kontinuumsmethode wurde von Egwolf und Tavan (ET) entwickelt [98, 101, 116]. Da sie auf der Lösung der Poisson-Gleichung beruht, ist von ihr, im Gegensatz zu den GB-Methoden, eine korrekte Beschreibung des Reaktionsfelds zu erwarten. Sie unterscheidet sich jedoch von gängigen PB-Methoden durch ihre sehr viel größere Effizienz und scheint deshalb für den Einsatz in MD-Simulationen geradezu prädestiniert zu sein.

In der ET-Methode wurde die Poisson-Gleichung (1.3) so umformuliert, dass das Reaktionsfeld und das RF-Potential durch RF-Dipoldichten an den Orten der Proteinatome exakt repräsentiert werden. Nähert man die Dipoldichten und weitere Funktionen zur Darstellung der Atomvolumina, die dann das Proteinvolumen $\mathcal{V}_s$ ergeben, durch Gauss-Funktionen, so ergeben sich gekoppelte Gleichungen für atomare RF-Dipole, die iterativ numerisch gelöst werden können. Abbildung 1.8b zeigt das Prinzip dieser Methode: RF-Dipole (Pfeile) an den Orten der Punktladungen im Proteinvolumen $\mathcal{V}_s$ repräsentieren die Polarisation des Kontinuums (in Abbildung 1.8a grau). Das Reaktionsfeld $\mathbf{E}^\varepsilon$ und die RF-Energie W^ε lassen sich dann näherungsweise aus analytischen Formeln berechnen. Die sich ergebenden RF-Kräfte $\mathbf{F}_i^\varepsilon = q_i \mathbf{E}^\varepsilon(\mathbf{r}_i)$ auf atomare Ladungen q_i im Volumen $\mathcal{V}_s$ stimmen sehr gut mit denen expliziter Lösungsmittel überein. Die RF-Energieen W^ε sind vergleichbar mit den Ergebnissen aus PB-Methoden [116].

Trotz der guten Ergebnisse bei statischen Proteinkonfigurationen ist diese Kontinuumsmethode aber noch nicht für den Einsatz in MD-Simulationen geeignet. Dies liegt daran, dass die RF-Kräfte $\mathbf{F}_i^\varepsilon = q_i \mathbf{E}^\varepsilon(\mathbf{r}_i)$ nur eine reine *Actio*-Kraft des Kontinuums, repräsentiert durch die RF-Dipole, auf die Punktladungen im Protein darstellen. Sie enthalten jedoch keine *Reactio*-Kräfte der Punktladungen auf die RF-Dipole, die das 3. Newtonsche Gesetz fordert und die dem dielektrischen Druck an Grenzflächen [103] entsprechen. Vernachlässigt man das Prinzip von *Actio=Reactio*, so ergibt sich eine große Gesamtkraft und ein Drehmoment auf das Protein, verbunden mit einer starken Aufheizung des Systems. Methoden zur Korrektur der Translations- und Rotationsbewegung des Proteins und der Temperatur führen zu großen Artefakten in der Dynamik des Proteins. In dieser Doktorarbeit wird geprüft, wie die ET-Methode erweitert werden kann, so dass sie dem 3. Newtonschen Gesetz gehorcht und zu stabilen und effizienten MD-Simulationen führt.

1.6 Ziele und Gliederung dieser Arbeit

Die vorliegende Doktorarbeit entstand im Rahmen eines interdisziplinären Projektes des Bayerischen Forschungsverbundes Prionen. Ein erstes Ziel dieser Arbeit war es, in Molekulardynamik (MD)-Simulationen die Konformationsdynamik des zellulären Prion-Proteins PrP^C in seiner natürlichen Lösungsmittelumgebung zu beschreiben. Dazu habe ich ein Simulationssystem bestehend aus Protein, Wasserumgebung und Ionen vorbereitet, Prozeduren zur Simulation von Proteinen entwickelt und dieses System mit dem MD-Programm EGO-MMII einige Nanosekunden lang simuliert. Die

Simulationsdaten wurden anschließend mit Methoden zur Mustererkennung und der statistischen Datenauswertung, wie z.B. einem Verfahren zur hierarchischen Klassifikation von Daten, analysiert. Gemeinsam mit Thomas Hirschberger und seinem damaligen Werkstudenten Bernhard Schropp wurden die Simulationen zu PrP^C fortgeführt und zusätzlich zwei Punktmutationen (M205S/R) des Prion-Proteins untersucht. Die Ergebnisse dieser Simulationsstudie sind in einem Artikel [73] veröffentlicht und als Abschnitt 2 in diese Arbeit eingebunden.

Neben Simulationsstudien an PrP^C war es ein zweites Ziel des Projektes, Modelle für die dreidimensionale Struktur des krankheitsassoziierten PrP^{Sc} zu bilden. Zusammen mit Dr. Armin Giese vom Zentrum für Neuropathologie und Prionforschung in München habe ich zwei Vorschläge für die Grundstruktur von PrP^{Sc} entwickelt. Aus diesen Modellen habe ich dreidimensionale Computerstrukturen konstruiert und im Detail unter anderem mit Verfahren zur Strukturbestimmung von Proteinen, wie z.B. dem *Simulated Annealing* [117], ausgearbeitet. Analog dazu habe ich Modelle für die dreidimensionale Struktur von Fasern aus poly-Glutamin (Gln) entwickelt, welche die Grundbausteine von Fibrillen bilden, die bei der Huntingtonschen Krankheit in den Gehirnen erkrankter Menschen entstehen. Mittels MD-Simulationen sollten thermodynamisch instabile Strukturen verworfen, sowie die stabilste Grundstruktur identifiziert werden. Die resultierenden Modelle für Fibrillen aus poly-Gln und PrP^{Sc}, sowie die Ergebnisse der Simulationsstudien an den poly-Gln-Strukturen, wurden in einer Veröffentlichung [46] zusammengefasst, die als Abschnitt 3.1 in diese Arbeit eingebunden ist. In Abschnitt 3.2 werden die Randbedingungen und Verfahren bei der Modellbildung von PrP^{Sc} detaillierter beschrieben und die PrP^{Sc}-Fibrillen in ihrer dreidimensionalen Struktur abgebildet.

Bei den Simulationen des Prion-Proteins und der poly-Gln-Peptide wurde schnell deutlich, dass die derzeit technisch möglichen Simulationszeiten zu kurz sind, um abschließende Aussagen zur Langzeitstabilität der betrachteten Moleküle machen zu können. Da die mikroskopische Beschreibung des Lösungsmittels die Hauptursache für die begrenzte Simulationszeit (maximal 10–100 ns) ist, wäre es von großem Vorteil, sie durch eine effiziente Kontinuumsbeschreibung zu ersetzen. Ein vielversprechender Ansatz bot sich in den Ergebnissen der Doktorarbeit von B. Egwolf [98], in der eine Kontinuumstheorie für Proteine in wässriger Umgebung entwickelt worden war. Erste Schritte zur Implementierung dieser Methode in unser MD-Programm EGO-MMII unter Benutzung schon existierender schneller Berechnungsverfahren für die langreichweitige Elektrostatik wurden in einer Diplomarbeit unternommen. Meine Aufgabe war es nun, diese Implementierung so zu testen, zu korrigieren und gegebenenfalls zu erweitern, dass damit MD-Simulationen an Peptiden durchgeführt werden konnten. In Kapitel 4 werden die Ergebnisse dieses Projektes zusammengefasst. Nach einer umfangreichen Fehlerbehebung und Vervollständigung der Implementierung wurde in Simulationen an kleinen Peptiden deutlich, dass die Kontinuumsmethode durch die Verletzung des 3. Newtonschen Gesetzes (*Actio=Reactio*) in der bestehenden Form noch nicht für den Einsatz in MD-Simulationen geeignet war. Dies machte eine gründliche Einarbeitung in die Theorie der Kontinuumsmethode

notwendig. Die Ergebnisse der daraus entstandenen, zusammen mit Paul Tavan erarbeiteten Umformulierung der Kontinuumsmethode und deren Erweiterung auf *Reactio*-Kräfte sind in einem Artikel [118] veröffentlicht, der hier in Abschnitt 4.1 abgedruckt ist.

Die um die *Reactio*-Kräfte erweiterte Theorie, sowie die Möglichkeit zur Berechnung der Reaktionsfeld-Energie W^{ε}, wurde in EGO-MMII von mir implementiert. An einem geeigneten Testsystem habe ich die Kontinuumsmethode parametrisiert und anschließend die aus umfangreichen MD-Simulationen gewonnenen freien Energielandschaften des Alanin-Dipeptids mit denen aus Simulationen mit einer expliziten Lösungsmittelumgebung verglichen. Dabei wurden verschiedene Methoden zur Berechnung der RF-Dipole (selbstkonsistente Berechnung und *Extended Lagrange*-Methode) untersucht. Die Ergebnisse aus diesen Simulationen sind als Folgeartikel [119] direkt im Anschluss an [118] veröffentlicht. In die vorliegende Arbeit ist dieser Artikel als Abschnitt 4.2 eingebunden.

In Kapitel 5 fasse ich die Resultate meiner Arbeit ausführlich zusammen und gebe einen Ausblick auf mögliche Ansatzpunkte zukünftiger Projekte.

2 Mutationsstudie am Prion-Protein

Zunächst sollen die MD-Simulationen am zellulären Prion-Protein PrP^C und an zwei Varianten mit den Punktmutationen M205S und M205R vorgestellt werden. Der folgende Abschnitt ist ein Abdruck[1] des Artikels

> Thomas Hirschberger, Martina Stork, Bernhard Schropp, Konstanze F. Winklhofer, Jörg Tatzelt and Paul Tavan: „Structural Instability of the Prion Protein upon M205S/R Mutations Revealed by Molecular Dynamics Simulations.“ *Biophysical Journal* 90, 3908-3918 (2006),

den ich gemeinsam mit Thomas Hirschberger, Bernhard Schropp und Paul Tavan aus unserer Arbeitsgruppe, sowie mit Konstanze F. Winklhofer und Jörg Tatzelt vom Institut für Neurobiochemie der LMU München verfasst habe. Zu dieser Veröffentlichung findet sich auf der Internetseite das Biophysical Journal [2] Zusatzmaterial in Form einer ergänzenden Grafik, die im Anschluss an den Artikel hier ebenfalls abgedruckt ist, sowie mpg-Filme, die die Trajektorien der simulierten Moleküle zeigen.

Meine Beiträge zu dieser Arbeit bestanden vor allem in der Vorbereitung der Simulationssysteme und der Prozeduren zur Simulation dieser Proteine, sowie in der Auswertung der Simulationsdaten zusammen mit Thomas Hirschberger und Bernhard Schropp.

[1] Mit freundlicher Genehmigung des Biophysical Journal.
[2] http://www.biophysj.org

Structural Instability of the Prion Protein upon M205S/R Mutations Revealed by Molecular Dynamics Simulations

Thomas Hirschberger,* Martina Stork,* Bernhard Schropp,* Konstanze F. Winklhofer,† Jörg Tatzelt,† and Paul Tavan*

*Theoretische Biophysik, Lehrstuhl für BioMolekulare Optik, Ludwig-Maximilians-Universität, Munich, Germany; and †Neurobiochemie, Adolf-Butenandt-Institut, Ludwig-Maximilians-Universität, Munich, Germany

ABSTRACT The point mutations M205S and M205R have been demonstrated to severely disturb the folding and maturation process of the cellular prion protein (PrP^C). These disturbances have been interpreted as consequences of mutation-induced structural changes in PrP, which are suggested to involve helix 1 and its attachment to helix 3, because the mutated residue M205 of helix 3 is located at the interface of these two helices. Furthermore, current models of the prion protein scrapie (PrP^{Sc}), which is the pathogenic isoform of PrP^C in prion diseases, imply that helix 1 disappears during refolding of PrP^C into PrP^{Sc}. Based on molecular-dynamics simulations of wild-type and mutant PrP^C in aqueous solution, we show here that the native PrP^C structure becomes strongly distorted within a few nanoseconds, once the point mutations M205S and M205R have been applied. In the case of M205R, this distortion is characterized by a motion of helix 1 away from the hydrophobic core into the aqueous environment and a subsequent structural decay. Together with experimental evidence on model peptides, this decay suggests that the hydrophobic attachment of helix 1 to helix 3 at M205 is required for its correct folding into its stable native structure.

INTRODUCTION

In prion diseases, like the Creutzfeldt-Jakob disease, scrapie, or the bovine spongiform encephalopathy, the central pathogenic process is the refolding of cellular prion proteins (PrP^C) into the pathological and β-sheet-rich isoform PrP^{Sc}, which aggregates into amyloid fibers (1). The occurrence of such a template-guided refolding process indicates that the native solution structure of PrP^C cannot be very stable. According to nuclear magnetic resonance (NMR) spectroscopy (2,3), in this solution structure only the C-terminal domain (residues 125–28) exhibits a well-defined tertiary folding pattern, whereas the remaining N-terminal part is randomly coiled. Moreover, the major result of a refined NMR analysis was the precise structural definition of a large fraction of side chains, showing that the globular domain of PrP^C contains a tightly packed hydrophobic core (4). Due to the marginal stability of the PrP^C structure, small perturbations such as point mutations of single residues in the globular C-terminal domain are likely to cause large-scale structural changes.

A candidate for such a perturbing mutation is methionine 205. The position of M205 within the C-terminal domain of human PrP^C (3) is shown in Fig. 1. M205 is part of helix 3 and its hydrophobic side chain is buried in the region of contact with helix 1. Therefore, a replacement of M205 by a hydrophilic residue could weaken this hydrophobic contact and thereby destabilize the hydrophobic core. Indeed, a previous study in cell culture (5) revealed that the two mutations M205S and M205R significantly interfere with folding and maturation of PrP^C in the secretory pathway of neuronal cells. In contrast to wild-type PrP^C, both mutants adopt a misfolded and partially protease-resistant conformation, lack the glycosylphosphatidylinositol anchor, and are not complex glycosylated. Interestingly, PrP-ΔH1, a different PrP mutant in which helix 1 was completely deleted, shows the same phenotype (5).

The question as to whether helix 1 is stable or not has been a central topic in a series of further studies, which applied circular dichroism (CD) (6) and NMR spectroscopy (7–10), bioinformatics tools of secondary-structure prediction (7,11), molecular modeling (12,13), or molecular-dynamics (MD) simulations (11) to peptides covering the helix 1 sequence. Some of these studies came to the conclusion that helix 1 is stable (9), or even remarkably stable (7), whereas others came to the opposite conclusion (8,10). Its sequence is characterized by an unusual abundance of charged residues (13), which are all exposed to the solvent. The region of contact with helix 3, however, exhibits mildly polar (Y149, Y150, N153) or nonpolar (M154) residues, such that helix 1 acquires an amphiphilic character.

The interest in the stability of helix 1 has been driven by the little available knowledge on the structure of the pathogenic PrP^{Sc} isoform of PrP^C, according to which PrP^{Sc} is largely formed by β-sheets and has a reduced α-helical content (14,15). Therefore, certain parts of the PrP sequence, which in PrP^C form α-helices, must refold into β-sheets during the conformational transition from PrP^C to PrP^{Sc}. The

Submitted September 29, 2005, and accepted for publication February 13, 2006.

Address reprint requests to Paul Tavan, Lehrstuhl für BioMolekulare Optik, Ludwig-Maximilians-Universität, Oettingenstr. 67, D-80538 München, Germany. Tel.: 49-89-2180-9220; Fax: 49-89-2180-9202; E-mail: paul.tavan@physik.uni-muenchen.de.

0006-3495/06/06/3908/11 $2.00

doi: 10.1529/biophysj.105.075341

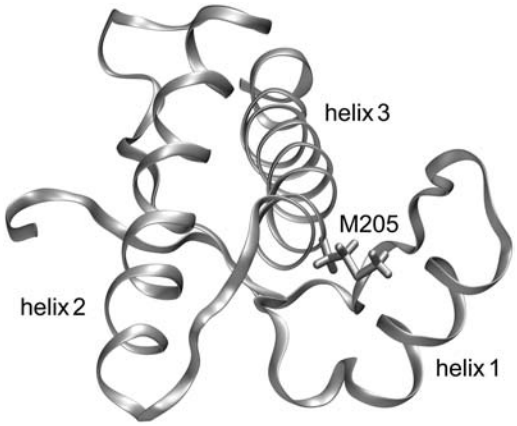

FIGURE 1 Backbone folding pattern of the human prion protein in the C-terminal domain (residues 125–228) according to the NMR structure in Zahn et al. (3). Besides a two-stranded antiparallel β-sheet, the structure comprises three α-helices, helix 1 (144–154), helix 2 (173–194), and helix 3 (200–228). Also drawn is the hydrophobic side chain of methionine 205, which is part of helix 3. According to the figure, this side chain is buried in the contact region between helix 3 and helix 1. This structural arrangement suggests that the replacement of M205 by a hydrophilic residue like serine or arginine could induce an intrusion of water molecules into the contact region and, thus, loosen the attachment of the two helices.

question, then, is which parts do refold, and, in particular, do they cover helix 1?

Unfortunately, PrP^{Sc} is currently inaccessible to high-resolution techniques of structural analysis: conventional solution NMR fails, because the PrP^{Sc} amyloid fibers are too large, and x-ray diffraction cannot be applied, because nobody has succeeded in arranging these fibers into three-dimensional crystals. Therefore, one has to resort to necessarily somewhat speculative modeling considerations when addressing the structure of PrP^{Sc}.

According to recent modeling suggestions, helices 2 and 3 are preserved in the PrP^{C}-to-PrP^{Sc} transition, whereas the role of helix 1 is controversial (16–19). According to one of these PrP^{Sc} models (16), which was designed to match electron microscopy data on two-dimensional crystals of PrP^{Sc} building blocks, helix 1 refolds and becomes a part of a large, left-handed, and triangular β-helix. Because such a structure cannot tolerate charged residues within its hydrophobic core, other alignments of parts of the PrP sequence onto such a β-helix were recently developed (17,18). In contrast, helix 1 remains stable according to a simulation-based model for a PrP^{Sc} protofibril (19).

Motivated by these discussions and by the specific results on the putative significance of M205 for the stabilization of helix 1 (5), we decided to carry out computer experiments aimed at checking whether the point mutations M205S and M205R can actually induce a destabilization of helix 1. As our testing scenario, we have chosen MD simulations of the C-terminal domain of human PrP^{C} as well as of two suitably modeled mutants in pure water at room temperature and ambient pressure. Arguments explaining the thermodynamic and statistical background of the applied testing scenario can be found in Stork et al. (18).

METHODS

Simulation system

As the starting structure for our simulation of the C-terminal domain (residues 125–228) of human PrP^{C}, which we call wtPrP, we chose the NMR structure (entry 1QM2 of the Protein Data Bank (20)) determined by Zahn et al. (3). In addition, we created simulation models for the two variants PrP-M205S and PrP-M205R by introducing the respective point mutations M205S and M205R into the structure of wtPrP using Swiss-PdbViewer (21). Charged N- and C-terminal groups were chosen for all three models. A rhombic dodecahedron just covering an enclosed sphere with a radius of 52 Å has been chosen to define the geometry of a periodic simulation system. Because the maximum extension of the PrP is ~51 Å along helix 3, the simulation cell is larger than twice the maximum extension of PrP in any direction. Initially this unit cell was filled with molecular mechanics (MM) models of water molecules. The transferable three-point interaction potential originally suggested in Jorgensen et al. (22) and modified in MacKerell et al. (23) was chosen as the MM force field for these water models. The force-field parameters of the proteins were adopted from CHARMM22 (23).

All simulations were carried out using the NPT ensemble with the MD program EGO-MMII (24). The temperature T and the pressure p were controlled by a thermostat ($\tau = 0.1$ ps) and a barostat ($\tau = 1.0$ ps, $\beta = 5.0$ Pa) (25), respectively. Covalent bonds involving hydrogen atoms were kept fixed by the M-SHAKE procedure (26). A basic time step of 2 fs was chosen for the multiple time-step integration (27) of Newton's equations of motion employed by EGO-MMII. The long-range Coulomb interactions were treated by the combination of structure-adapted multipole expansions (28,29) with a moving-boundary reaction-field approach explained and tested by Mathias and co-workers (24,30). Here, the dielectric and ionic continuum surrounding each atom in the system at a distance of ~52 Å were described by a dielectric constant $\varepsilon = 79.0$ and a Debye-Hückel parameter $\kappa = 0.13$ Å^{-1}, respectively. That value of κ corresponds to a 165-mM NaCl concentration. Van der Waals interactions were calculated explicitly up to distances of 10.5 Å; at larger distances, a mean-field approach (31) was applied.

The water system was initially equilibrated for 1 ns at $T = 300$ K and $p = 1.013 \times 10^5$ Pa. For solvation the proteins were positioned at the center of the equilibrated water box and all water molecules closer than 2.0 Å to a protein atom were removed, resulting in a total amount of ~25,800 water molecules surrounding a given protein. This corresponds to a 2.15-mM protein concentration. In addition, ~150 of the water molecules were randomly selected and replaced by Na^+ and Cl^- ions, yielding a neutral simulation system and the 165-mM NaCl concentration used above for the characterization of the Debye-Hückel continuum. Thus, the simulation system covered ~79,200 atoms. For equilibration, the proteins were initially kept fixed, whereas the surrounding solvent molecules were thermally moving for several hundred picoseconds at $T = 500$ K and $T = 300$ K. Next, the rigid constraints were removed and solely the positions of the protein C_α atoms were constrained by harmonic potentials (force constant 2.1×10^2 kJ/(mol Å^2)). These systems were cooled by energy minimization within 1 ps to $T < 0.1$ K and subsequently heated within 120 ps to $T = 300$ K. Within another 300 ps, the constraining force constants were slowly reduced to zero until the proteins were free to move within the solvent. This procedure served to adjust the modeled protein structures to the MM force field, or, equivalently, to partially remove the prejudices imposed onto the structures by the modeling. The simulation systems thus obtained were the starting points for the following unconstrained 10-ns simulations at $T = 300$ K and $p = 1.013 \times 10^5$ Pa. Coordinates were saved every picosecond. Note, here, that 10-ns simulations of systems with ~80,000 atoms are computationally quite expensive. Using six processors (1.6 GHz), the total computation time for all three simulations was ~20 weeks.

Measures of overall stability

To obtain an overall measure for the structural stability of the three PrP variants during MD simulation, the three trajectories were separately analyzed by computing root mean-square deviations from average initial structures. Restricting the analysis to the backbone as given by the coordinates of the C_α atoms, the average initial structures were calculated from the first 100 ps of the free simulations. Each backbone structure sampled by the unconstrained simulation was subsequently fitted onto the respective initial structure by minimization of the root mean-square deviation, and the resulting time series of minimal values $d_{\mathrm{rms}}(t)$ was saved for plotting.

Classification of backbone trajectories and determination of protein conformations by hierarchical cluster analysis

For a detailed analysis of the backbone conformations sampled by the MD trajectories of the three PrP variants, the ϕ/ψ dihedral angles along the protein backbones were collected resulting in time series of 206-dimensional feature vectors $\boldsymbol{\alpha} \equiv (\psi_{125}, \phi_{126}, \psi_{126}, \ldots, \phi_{227}, \psi_{227}, \phi_{228}) \in [-\pi, \pi]^{206}$. The corresponding data sets $\{\boldsymbol{\alpha}(t) \mid t = 1, \ldots, 10{,}000\}$ represent point densities in the angular configuration space $[-\pi, \pi]^{206}$, for which we derived smooth, parametric, and analytical maximum-likelihood (ML) density models $p(\boldsymbol{\alpha} \mid \sigma_{\mathrm{ML}})$ in the form of mixtures of $R = 100$ univariate normal distributions with identical statistical weights $1/R$ and widths σ_{ML}. The parameters of these mixture models, that is the centers $\boldsymbol{w}_r \in [-\pi, \pi]^{206}$, $r = 1, \ldots, 100$ and the common width σ_{ML} of the Gaussians, were optimized by a safely converging algorithm maximizing the likelihood of the density estimates provided by the mixture models $p(\boldsymbol{\alpha} \mid \sigma_{\mathrm{ML}})$ (32–34).

As is well known (see, e.g., Carstens et al. (34) for further references and a discussion), each conformation of a peptide whose dynamics is sampled by a MD trajectory corresponds to a local maximum of the point density in the reduced configuration space $[-\pi, \pi]^{206}$, because such a maximum marks a local minimum of a corresponding free energy landscape. However, the distinction of a local density maximum from a statistical density fluctuation requires a smoothening of the density at multiple scales. Therefore, to identify the hierarchies of conformations and subconformations, we constructed scale-space representations $p(\boldsymbol{\alpha} \mid \sigma) = p(\boldsymbol{\alpha} \mid \sigma_{\mathrm{ML}}) \times g(\boldsymbol{\alpha} \mid \sigma_{\mathrm{conv}})$ of the ML models by convolution with a Gaussian kernel $g(\boldsymbol{\alpha} \mid \sigma_{\mathrm{conv}})$ of varying width σ_{conv}. The resulting models $p(\boldsymbol{\alpha} \mid \sigma)$ are mixtures of normal distributions at fixed centers $\boldsymbol{w}_r$ and of variable width σ. For these smoothened models $p(\boldsymbol{\alpha} \mid \sigma)$, we employed gradient ascents to detect all of their local maxima and, thus, the number $N(\sigma)$ and prototypical geometries $\boldsymbol{\alpha}_k(\sigma)$ of all conformations $k = 1, \ldots, N(\sigma)$ at the given spatial resolution σ.

By considering plots of $N(\sigma)$ vs. σ obtained for the three protein models we selected a common resolution $\sigma_c = 1.85$ rendering $N(\sigma_c) = 4$, 3, and 7 conformational states for wtPrP, PrP-M205S, and PrP-M205R, respectively (cf. the figure in the Supplementary Material). The resulting prototypical states k of the three PrP variants are denoted by wt1-wt4, S1-S3, and R1-R7, respectively. For graphical illustration of these prototypical structures, we picked those snapshots from the MD trajectories whose feature vectors $\boldsymbol{\alpha}(t)$ are closest to the density maxima at $\boldsymbol{\alpha}_k(\sigma_c)$ in dihedral space $[-\pi, \pi]^{206}$. Finally, at the selected resolution σ_c every protein geometry $\boldsymbol{\alpha}(t)$ contained in the time series was classified by gradient ascent as belonging to one of the conformations k.

We checked by comparison with other possible choices that the selected resolution $\sigma_c = 1.85$ represents a reasonable compromise between a sufficiently simplified but still detailed representation of the backbone fluctuations and relaxations sampled by the trajectories. For instance, an increase of the resolution by reducing σ from 1.85 to 1.7 slightly increases the number $N(\sigma)$ of conformations from four to five for wtPrP, leaves that number invariant at the value of three for PrP-M205S, and strongly increases it from 4 to 11 for PrP-M205R. The latter increase indicates that many major conformational transitions must have been sampled by the corresponding trajectory. This conjecture will be substantiated further below.

Classification of trajectories in terms of secondary-structure elements

For an analysis of local secondary-structure motifs we applied the software tool DSSP (Database of Secondary Structure in Proteins) by Kabsch and Sander (35), which employs H-bonding patterns and various other geometrical features to assign secondary-structure labels to the residues of a protein. DSSP classifies each residue in every snapshot as belonging to one of the eight classes ''α-helix'', ''isolated β-bridge'', ''extended strand'', ''3-10-helix'', ''π-helix'', ''H-bonded turn'', ''bend'', or ''other''. Snapshots taken every 50 ps from the trajectories served as input for DSSP. As a result, one obtains trajectories of secondary-structure labels. Using suitable color coding, one can represent the local secondary-structure dynamics sampled by a trajectory as a graph covering as many lines as there are residues in the simulated protein model.

RESULTS

To allow a most vivid insight into the processes described by our three 10-ns simulations of the wild-type and the two mutant PrP models, we have provided three mpg-movies in the online supplement to this article (http://www.biophys.org). The movies show the backbone fluctuations of the three proteins in a ribbon representation. The M205 residue and its mutated variants are highlighted by an all-atom representation. Snapshots taken every 20 ps were collected to generate these movies.

Movies cannot be printed, nor does their format permit a quantitative comparison between the three simulations. For this reason, the observation and classification tools described above have to be used.

PrP-M205S and PrP-M205R are less stable than wtPrP

Fig. 2 shows the time series $d_{\mathrm{rms}}(t)$ of the root mean-square deviations from the initial backbone structures (cf. Methods) for the three PrP variants. For wtPrP (*light shaded*), the values of $d_{\mathrm{rms}}(t)$ are smaller than those for the two mutants throughout the whole simulation time of 10 ns, indicating

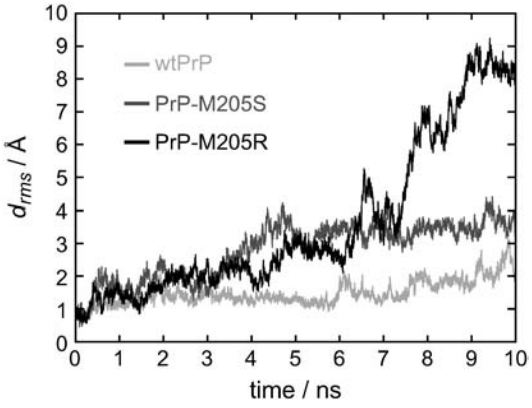

FIGURE 2 Time series of the minimal root mean-square deviations $d_{\mathrm{rms}}(t)$ of the PrP structures from their respective initial structures.

that the MD simulation preserves the starting structure of wtPrP much better than those of the mutants. Among the mutants, PrP-M205R (*solid*) is seen to deviate much more strongly from the starting structure than PrP-M205S (*dark shaded*).

According to Fig. 2, for wtPrP the values of $d_{\mathrm{rms}}(t)$ fluctuate around 1.5 Å for the first 6 ns. In the remaining time span the deviation $d_{\mathrm{rms}}(t)$ increases to values of ~2.5 Å, indicating that a conformational change occurred at ~6 ns.

For PrP-M205S, the values of $d_{\mathrm{rms}}(t)$ initially exhibit a rapid increase. Within the following first 3 ns, they fluctuate around 2 Å before they subsequently rise to a second plateau characterized by values near 3.5 Å. These data indicate that the starting conformation of PrP-M205S, which had been chosen close to the native structure of PrP^{C}, is changed by two major conformational transitions. Correspondingly, the data suggest that each of the two plateaus represents a meta-stable conformation, in which PrP-M205S happens to be temporarily caught during the 10-ns computer experiment.

In the case of PrP-M205R, the values of $d_{\mathrm{rms}}(t)$ moderately increase in two consecutive stages during the first 6 ns until they cross the $d_{\mathrm{rms}}(t)$-trajectory of PrP-M205S. After 6 ns, the $d_{\mathrm{rms}}(t)$ trajectory of PrP-M205R steeply increases to reach a peak of 5 Å at ~6.5 ns. After a short decline at ~7 ns, this trajectory exhibits further jumps toward larger values until it seems to reach a plateau at ~8.5 Å within the last nanosecond of the simulation. This behavior suggests that the R-mutant undergoes substantial conformational changes by passing through several short-lived intermediate states.

For the two mutants, the above analysis of the $d_{\mathrm{rms}}(t)$ time series has shown that the assumed PrP^{C}-like starting structures do not represent stable conformations. In contrast, for wtPrP, significantly smaller deviations were observed. For all three PrP variants time points of conformational transitions were identified. However, the inspection of the $d_{\mathrm{rms}}(t)$ trajectory did not provide any insights into the structural changes described by the 10-ns computer simulations. Clearly, viewing the movies of the trajectories gives detailed impressions of the conformational changes sampled by the simulations. A coarse-grained and printable version of these movies, which catches in a sketchy manner their essential contents, is obtained by applying our hierarchical classification scheme to the three trajectories. Here the backbone configurations sampled by the trajectories are represented in terms of coarse-grained prototypical conformations as described in Methods. The results of that conformational analysis will now be individually presented for each of the three simulation trajectories, starting with wtPrP.

Conformational equilibrium fluctuations of wtPrP

Fig. 3 *A* shows the prototypical structures wt1–wt4, representing the coarse-grained states identified by our classification scheme in the MD trajectory of wtPrP, and

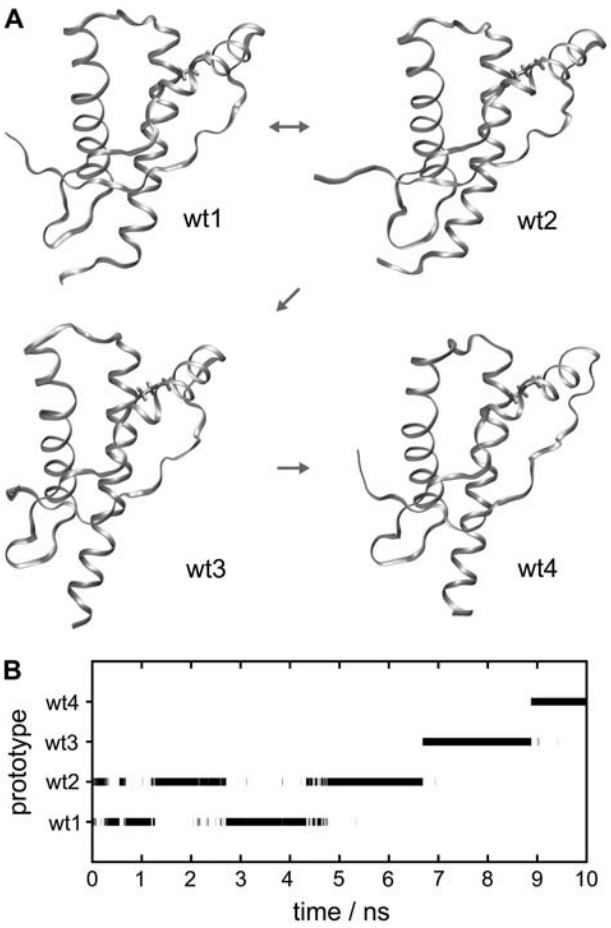

FIGURE 3 (*A* and *B*) Conformations of wtPrP and time series of classifications. The structures of wt1, wt2, wt3, and wt4 are related to local density maxima in dihedral space. Thus, they represent classes of structurally similar configurations within the simulation time series of wtPrP. (*A*) The small arrows between the conformations indicate the chronological order in which they were occupied during the simulation. (*B*) The protein configurations sampled by the simulation of wtPrP are assigned to the classes defined by the prototypical configurations wt1–wt4.

Fig. 3 *B* shows the corresponding classification of the trajectory (cf. Methods). Within the first 6.7 ns, the wtPrP model repeatedly fluctuates between the backbone conformations wt1 and wt2, before it subsequently changes to conformation wt3 and finally to wt4. By comparing the drawings in Fig. 3 *A* we can now identify the structural contents of these transitions.

For instance, one of the features of the conformational transition at ~6–7 ns, whose occurrence was already indicated by the $d_{\mathrm{rms}}(t)$ trajectory (cf. the discussion of Fig. 2), and which now has been identified as the wt2 → wt3 transition, appears to be a spontaneous refolding of the C-terminal part of helix 3. This part of helix 3 loses its original NMR shape during the first few picoseconds of the simulation and thus is disordered in wt1 and wt2. It regains its α-helical structure in wt3 and wt4. Inspection of the movie of the wtPrP trajectory (see Supplementary Material)

clearly shows these processes. The marginal stability of the C-terminal part of helix 3 thus described by our MD simulation agrees with the NMR data presented by Zahn et al. (3), who found an enhanced accessibility to amide proton exchange at the corresponding residues. A second feature of the wt2 → wt3 transition seems to be a slight loosening of the extended polypeptide chain between the disordered N-terminal tail and helix 1. According to NMR (3), this sequence portion partially contributes to a rather rigid, small β-sheet found in the PrP structure. The wt3 → wt4 transition observed at 9 ns is characterized by a bending of the C-terminal part of helix 2, a straightening of helix 3, and a change of the connecting loop.

The trajectory in Fig. 3 *B* indicates an equilibrium between the conformations wt1 and wt2, which mainly differ in the orientations of the flexible N- and C-termini. Unfortunately, our 10-ns trajectory is too short to provide corresponding evidence also for the transitions to conformations wt3 and wt4: a return to the wt1 or wt2 conformation has not (yet) been observed.

In PrP-M205S, helix 2 is decaying

Fig. 4 *A* shows the three conformations S1, S2, and S3 identified by our classifier for PrP-M205S. Conformation S1 closely resembles the PrP^C-like initial structure. The main difference is a slight increase of the gap between helix 1 and helix 3 near position 205 (as compared to the native structure in Fig. 1 or to the nearly native conformations wt1–wt4 in Fig. 3 *A*) and indicates a partial solvation of the Ser-205 by water molecules (not shown) entering the interfacial region between the two helices. As is apparent from the classification of the trajectory in Fig. 4 *B*, conformation S1 persists for the first 4.7 ns of our computer experiment. According to Fig. 2, at about this time the $d_{\mathrm{rms}}(t)$ trajectory reaches a second plateau, indicating a major conformational transition.

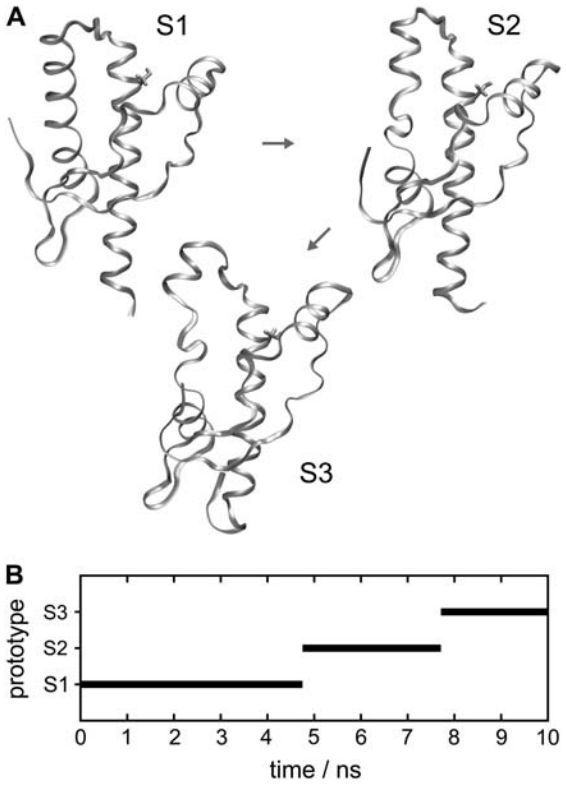

FIGURE 4 Conformations of PrP-M205S (*A*) and time series of occupancies (*B*).

Fig. 4 *B* identifies the new conformation of PrP-M205S as S2, and the drawing in Fig. 4 *A* reveals the structural nature of the S1 → S2 transition. Apart from small fluctuations at the C- and N-termini it mainly consists in an unfolding of ~1.5 coils of helix 2 covering the region between residues 181 and 186. As revealed by Fig. 4 *B*, conformation S2 lives for ~3 ns, until it further decays to conformation S3, which is the conformation persisting until the end of our simulation. Fig. 4 *A* demonstrates that the S2 → S3 transition is characterized by a continued unfolding of helix 2 such that the unfolded region now covers residues 181–188. This continued unfolding of helix 2 is accompanied by an unfolding of the C-terminal part of helix 3 by 1.5 coils.

According to the NMR data on amide hydrogen exchange in native PrP^C (3), helix 2 is quite rigid in the region between 181 and 186, whereas the part from residues 187–194 is supposed to exhibit a considerable conformational flexibility. Earlier MD studies on PrP fragments spanning a part (residues 180–193) of helix 2 (36) or helices 2 and 3 (11) essentially agreed with these NMR data, although the latter study indicated a decreased helical order already for residues 182–186. The latter finding thus agrees with our result that the mutation-induced unfolding of helix 2 starts in the supposedly rigid sequence range 181–186. These data suggest that the 181–186 part of helix 2 is stabilized into its natively rigid structure by the unperturbed structural ensemble characteristic for the native PrP sequence.

Note here that the large-scale processes of conformational decay exhibited by Fig. 4 are hardly reflected by the $d_{\mathrm{rms}}(t)$ time series in Fig. 2, underlining the limited value of this simple observable. To monitor the further fate of the observed decay, which most probably will go on after 10 ns, more extended simulations will be required.

However, independent of that fate, already the observed processes appear to mark a fast and unidirectional decay of the PrP^C-like initial structure for PrP-M205S. Because our wtPrP model of a natively stable structure did not show any comparable signs of instability, particularly concerning the 181–186 sequence of helix 2, we conclude that this rigid part of the native PrP structure is destabilized by the M205S mutation. Furthermore we conclude from the observed instability of the PrP-M205S model that this protein, if it happened to assume a PrP^C-like structure, would leave this conformation rapidly. Conversely, the M205S mutant will most likely never acquire the native PrP^C structure, because

it apparently represents an unstable state of very high free energy.

One would now expect that helix 2 decays also in our mutant model PrP-M205R, because in this mutation, too, the hydrophobic M205 is replaced by a polar residue. In fact, it is likely that such a decay will be observed in corresponding computer experiments. However, our single computer experiment on PrP-M205R happened to reveal a further and different property of the native PrP^C structure.

In PrP-M205R, helix 1 is decaying

In the case of the experiment on PrP-M205R, our statistical tool for classification of backbone conformations identifies a set of seven prototypical structures, R1–R7 (see Fig. 5 *A*). Fig. 5 *B* shows that the PrP-M205R model remained in conformation R1 during the first 6.2 ns. According to Fig. 5 *A*, the main feature of conformation R1 is an opening of a gap between helices 1 and 3 (cf. conformations wt1–wt4 in Fig. 3 *A*), which, as in the S1 conformation discussed above, serves to allow a solvation of the polar residue replacing M205 by water molecules entering the interfacial region (further data concerning the water are not shown). Because the charged arginine residue is larger and more polar than the small serine, the gap between the two helices is larger in R1 than in S1. Thus the M205R mutation induces a stronger perturbation into the native PrP^C structure than M205S.

Fig. 5 *B* shows that shortly after 6.2 ns, the PrP-M205R conformation switches to states R2 and R3, returns to R1, and through the very short-lived intermediates R4–R6 jumps to R7, where it happens to remain for the last 2 ns of our simulation. According to Fig. 5 *A*, the initial fluctuations within the set of states R1–R3 mainly consist of reversible and strong fluctuations of helix 1 into the aqueous phase, strongly increasing the gap and corresponding angle between helices 1 and 3. However, at ~7.5 ns the water-exposed helix 1 starts to unfold in several stages. In the R1 → R4 transition, a first coil of helix 1 unfolds. The R4 → R5/R6 transition is characterized by the partial unfolding of a second coil resulting in a complete unfolding of the amphiphilic part of helix 1. Finally, in the transition to R7, residues D144 and Y145 also leave the original α-helical structure. In R7 only the four residues 146–149 remain in an α-helical structure. Note here that the above decay process is accompanied by an unfolding of 1.5 coils in the C-terminal part of helix 3, whereas the remaining parts of the PrP^C structure remain essentially unchanged. Also in this case the fate of the ongoing decay process is unclear.

As in the case of M205S, the instability observed for the M205R model indicates that the M205R protein will most likely never acquire the native PrP^C structure. Furthermore, our finding supports the suggestions by Winklhofer et al. (5) that the point mutation M205R destabilizes helix 1 and that M205R differs structurally from wtPrP.

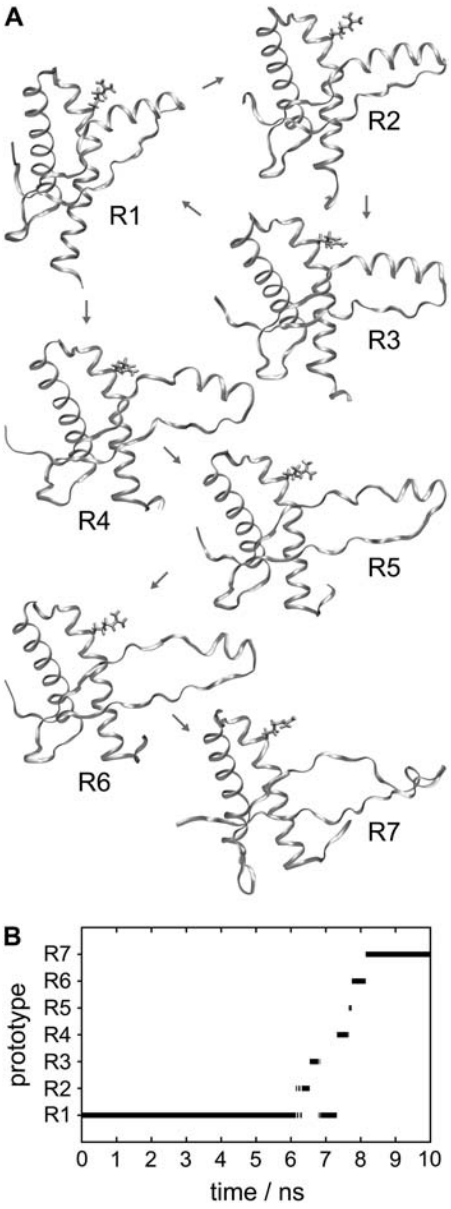

FIGURE 5 Conformations of PrP-M205R (*A*) and time series of occupancies (*B*).

Trajectories of secondary-structure features

Fig. 6 shows the classification of the three trajectories in terms of secondary-structure elements obtained by the software tool DSSP (cf. Methods). The DSSP plots enable a local structural analysis complementing the above characterization of the backbone conformational dynamics.

Before examining the time series, consider the DSSP classification of the NMR structure of PrP^C by Zahn et al. (3), which is depicted in Fig. 6 at the left and right margins of the DSSP time series. Generally, the DSSP classification agrees quite well with the secondary-structure assignments given by Zahn et al. (3) (cf. legend to Fig. 1). DSSP,

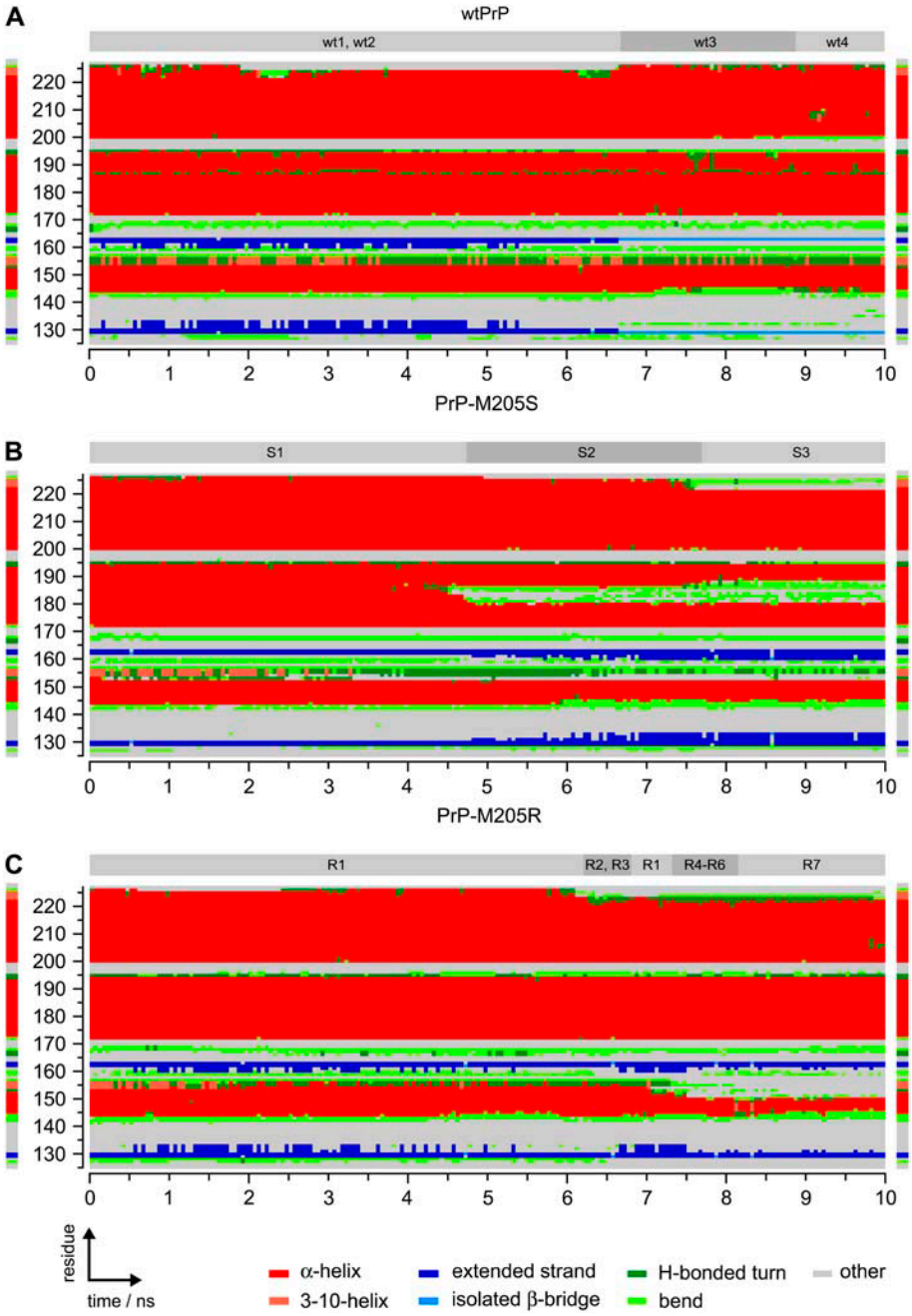

FIGURE 6 Classification of trajectories in terms of secondary-structure elements by DSSP (35). The residue-wise DSSP classifications are shown for the trajectories for wtPrP (*A*), PrP-M205S (*B*), and PrP-M205R (*C*). As a reference, the DSSP classification of the NMR structure (3) is displayed at the left and right margins. For comparison, the classification in terms of prototypical backbone structures from Figs. 3 *B*, 4 *B*, and 5 *B* is indicated at the top of the DSSP plots.

however, classifies the C-terminal residues 223–226 of helix 3 as forming 3-10-helical (*light red*) or bend (*light green*) structures and classifies residues 227 and 228 as ''other'' (*gray*). Furthermore, DSSP marks only residues 129 and 130, as well as 162 and 163, as ''extended strands'' (*blue*), whereas according to Zahn et al. (3) the small β-sheet covers residues 128–131 and 161–164.

When one examines the DSSP time series for wtPrP shown in Fig. 6 *A*, one gains complementary insights into the conformational dynamics. During the first 6.7 ns covering the wt1 $\leftrightarrow$ wt2 equilibrium, the C-terminal residues of helix 3 unfold and show an enhanced flexibility (*green* or *gray* instead of *red* at the top lines of the graph). Correspondingly, in the conformations wt1 and wt2 (cf. Fig. 3 *A*), the terminal part of helix 3 is unfolded. Likewise, DSSP confirms its subsequent refolding during the lifetime of states wt3 and wt4. As we have pointed out above, in the discussion of the wtPrP conformations, this C-terminal flexibility agrees with NMR (3).

A detail that cannot be detected by visual inspection of wt1–wt4 in Fig. 3 *A* but is revealed by Fig. 6 *A* is the growth of the two short β-strands (*blue*) toward helix 1 during the first half of the simulation time, as well as their subsequent shortening toward a single pair of isolated β-bridges present

during the lifetimes of wt3 and wt4 (*light blue*). This shortening can only be partially guessed from the loosening of the polypeptide chain between the flexible N-terminal tail and helix 1 displayed by the backbone conformations in Fig. 3 *A*. However, this loosening also covers the loop region between β-strand 1 and helix 1 (132–143) and this feature is not reflected in the DSSP plot. The corresponding residues are simply classified as ''other'' or ''bend''.

The shortening of the β-strands is accompanied by an enhanced flexibility at the two N-terminal residues 144 and 145 of helix 1, which temporarily switch to bend (*light green*) or turn (*green*) structures in the period between 7.2 ns and 9.7 ns. The fact that these two residues rejoin helix 1 at 9.7 ns indicates the equilibrium character of these fluctuations. Such a conformational equilibrium has been suggested earlier by Zahn et al., who found an increased amide proton exchange at these residues (3). These authors additionally reported enhanced proton exchange for residue 155 (3) at the C-terminus, which is consistent with the fluctuations of residues 154–156 between 3-10-helical (*light red*) and turn (*green*) structures observed in the DSSP time series for wtPrP.

Further details revealed by the DSSP time series for wtPrP and overlooked by our previous coarse-grained conformational analysis are certain fluctuations in helix 2. Here, residues 187 and 188 persistently switch and residues 189–194 occasionally fluctuate between α-helical (*red*) and turn structures (*green*). Also this flexibility agrees with an enhanced proton exchange (3), as noted above in our discussion of PrP-M205S.

In part, the DSSP time series for PrP-M205S and PrP-M205R in Fig. 6, *B* and *C*, show features similar to the one just discussed for wtPrP. For instance, they also show large-scale fluctuations from α-helical to bend or turn structures at the C-terminal residues 221–228 of helix 3. Because these particular fluctuations were reversible in the case of wtPrP and were shown to be in agreement with NMR data on amide proton exchange (3), they are equilibrium fluctuations. However, in the DSSP trajectories for PrP-M205S and PrP-M205R these fluctuations happen to occur at much larger timescales. They are observed only in the second parts of the two simulations, and a return to the integral helical structure does not occur until the end of the simulations. However, this does not mean that helix 3 cannot be restored later. It simply shows that certain fluctuations can cover vastly different time spans and that the simulation time of 10 ns is too short to cover all equilibrium processes although it is near the limits of computational feasibility for such a large system.

According to the DSSP analysis, the small β-sheet (*blue*) is stable for the first 5 ns and subsequently grows in the case of PrP-M205S, whereas it persistently fluctuates between a larger and normal size in PrP-M205R. A shortening, which happened to occur in the wtPrP simulation, is missing. For the stated reason of limited sampling, these findings do not imply that such a shortening will never occur in the mutants nor that a restoration to the original size cannot occur in the wtPrP model. Instead, the tendency of β-sheet growth found in parts of all three simulations when taken together with the stability of the β-sheet revealed by NMR appears to indicate that the employed CHARMM22 force field can satisfactorily describe this type of structural element.

Besides the discussed fluctuations, which seem to be equilibrium processes, the DSSP plots of the mutant models show two remarkable features, which do not even remotely find any correspondence in the DSSP plot of wtPrP. Starting at 4 ns, the DSSP time series for PrP-M205S in Fig. 6 *B* exhibits a large green and gray zone in the red band that serves to mark helix 2. Its onset approximately coincides with the S1 $\rightarrow$ S2 transition, which we have assigned to a beginning unfolding of helix 2 starting at residues 181–186 and continuing toward residues 187 and 188 during S3. Similarly, the DSSP time series for PrP-M205R in Fig. 6 *C* clearly indicates the partial decay of helix 1 noted further above: after 7 ns, the red band marking helix 1 becomes sizably smaller, covering solely residues 146–150 at 10 ns (instead of 144–154, as in the NMR structure of native PrP^C). However, the reason for and structural nature of this helix 1 decay are not revealed by the DSSP analysis. In this respect, our backbone conformational classification has proven its strengths. It has clearly revealed that helix 1, before it starts to decay, bends away from the hydrophobic core of the protein and concomitantly intrudes into the aqueous phase. Thus, only this classification suggests that helix 1 is mainly stabilized by the attachment to the hydrophobic core found in the native structure.

DISCUSSION AND CONCLUSIONS

Despite the progress of computer technology, the MD simulations presented here, each covering the motion of 80,000 atoms over a time span of 10 ns, are still quite costly. For very stable proteins the corresponding limitation to the 10-ns timescale does not represent a severe restriction, because this timescale still enables at least a partial sampling of the equilibrium conformations. However, for simulations of marginally stable proteins like PrP^C and even worse for nonequilibrium simulations, which start at an unstable initial structure (like our M205S- and M205R-PrP models), this limitation becomes severe.

In the case of a marginally stable protein, one may happen to sample a rare fluctuation for a few nanoseconds, which nevertheless has a very small statistical weight in the equilibrium ensemble. The shortening of the small β-sheet observed during the last 3 ns of our wtPrP simulation may be an example of such bad luck. This conjecture is motivated by the fact that a different MD study (37) starting at the same NMR structure of wild-type human PrP and covering the same 10-ns time span did not show any such shortening of the small β-sheet. It revealed only elongating fluctuations like those observed during the first 7 ns of our wtPrP

simulation. Furthermore, the hydrogen exchange data of Zahn et al. (3) argue for a quite stable β-sheet structure. Nevertheless, our computer experiment on wtPrP has illustrated once again that this protein exhibits large regions of conformational flexibility, which not only cover the various loops connecting the secondary structural elements, but also comprise helical regions. In agreement with NMR (3), these are the two termini of helix 1 and the C-termini of helices 2 and 3. The large-scale fluctuations revealed by our 10-ns simulation fit with the notion that native PrP is only a marginally stable protein.

In the case of nonequilibrium decay processes induced by point mutations, computational limitations restrict our simulation approach to single attempts and preclude frequent repetitions, which would allow us to gain at least some limited statistics on possible decay paths. Each of these single attempts has, then, the character of a random experiment sampling only one of many possible decay paths. It would be another instance of bad luck, of course, if the particular random experiment happened to choose a highly unlikely instead of a highly probable decay path. But independent of the likeliness of the chosen path, the outcome of such a random experiment clearly indicates the existence of a particular decay path and, therefore, the existence of a structural weak point. Such an experiment on an unstable mutant model additionally shows how the native protein is stabilized and which of its elements are susceptible to weak perturbations.

Our simulations of mutant PrP models have thus identified two elements of the native PrP^C structure, which are susceptible to weak perturbations. A first weak point is helix 2 in the sequence range 181–188, and a second weak point is the amphiphilic helix 1, which is destabilized as soon as it becomes completely exposed to the aqueous phase. As will be discussed now, the former result agrees with previous findings, whereas the latter is at variance with certain claims found in the literature, but agrees with many others and with the available data.

In the presentation of our results, we had already mentioned that the instability of helix 2 identified by us agrees with conclusions derived from earlier but more restricted MD studies (11,36). It is interesting to note that it also agrees with previous suggestions derived by application of secondary-structure prediction tools. For instance, according to Kallberg et al. the sequence 179–191 of PrP should adopt a β-strand structure (38), because this sequence portion mainly consists of the three bulky branched amino acids threonine (T), valine (V), and isoleucine (I), which are known to have a high β-strand propensity (39). In fact, corresponding peptides have been shown to form amyloid fibers and to exhibit a β-sheet structure (40,41). Similarly, a peptide covering the residues 173–195 from helix 2 was found to prefer α-helical over extended β-type conformations only by the small free energy difference of 5–8 kJ mol^{-1} (42).

In contrast, our identification of the amphiphilic helix 1 as a second weak point of the PrP^C structure disagrees with the interpretations by Ziegler et al. (7) of NMR and CD data on various short peptides spanning helix 1. According to these authors, helix 1 allegedly is remarkably stable, with its stability creating a barrier for the conversion of PrP^C to PrP^{Sc}. Other authors (8,9), however, who also studied helix 1 peptides by NMR and CD, came to less stringent conclusions concerning the stability of helix 1, ranging from the cautious statement that ''it would be no surprise if ... helix 1 were preserved during the conformational transition from PrP^C to PrP^{Sc}'' (9) to the conclusion that ''despite the propensity of the individual residues to preferentially populate helical space there is no well-formed helical conformation'' (8). Even more remarkably, Kozin et al. (10) determined by NMR for yet another helix 1 peptide in aqueous solution a well-defined β-hairpin structure and no α-helix content at all.

If one looks more closely at the experimental data gained for those peptides, which actually showed a certain helical propensity for the residues belonging to helix 1 (7–9), one notices close similarities despite the strongly different interpretations. At room temperature in aqueous solution, several of these peptides showed no medium-range NOEs whose presence would be indicative of a helical structure that is stable at the NMR timescale (7,8), whereas yet another peptide gave rise only to a few very weak signals of this type (9). Even Ziegler et al. (7), who strongly advocated the ''remarkable stability'' of helix 1 in their conclusion, had to admit in their results section that the absence of tertiary NOE peaks ''suggests high conformational flexibility'' of the helix 1 peptides.

The CD data presented in Ziegler et al. (7), Sharman et al. (8), and Liu et al. (9) for these peptides (dilute aqueous solution, room temperature) consistently indicate α-helical content in the range of 10–25% (cf., e.g., Fig. 4 in Liu et al. (9) or Fig. 2 *b* in Sharman et al. (8)). Together with the conformational flexibility following from NMR, these data indicate that the helix 1 residues are in rapid equilibrium of helical and random coil conformations. At the stated conditions, these conformations are thus essentially isoenergetic and are connected by small barriers. In the case of the particular peptide studied by Kozin et al. (10), a β-hairpin structure even marks a pronounced free energy minimum in conformational space.

For the other peptides, the NMR data on chemical shifts of the H^α-protons or the $^{13}C^\alpha$-atoms of the helix 1 residues indicate a somewhat larger helix propensity than the percentage given by the CD helix content. Concerning the NMR chemical shifts, the peptide data presented by Liu et al. (9) allow a direct comparison with corresponding data for native PrP^C (3). Taking the conformation-dependent chemical shift differences $\Delta\delta(^{13}C^\alpha)$ as a measure (compare Figs. 2 *a* in Liu et al. (9) and 3 *b* in Zahn et al. (3)), one finds that the helical propensity of the helix 1 residues is less pronounced in the isolated peptide than in PrP^C by a factor of 2.

The thus emerging high conformational flexibility of the helix 1 sequence at room temperature in aqueous solution

fully agrees with an extended MD study on a correspondingly solvated helix 1 peptide, in which only the initial coil (residues 144–148) remained in the initial α-helical conformation (11). Note that this remaining structure roughly agrees with the conformation of helix 1 observed by us after 10 ns in our PrPM205R simulation. On the other hand, the apparent stability of helix 1 in the native PrP^C structure, which follows from the NMR data of Zahn et al. (3), also agrees with our MD simulations, according to which helix 1 remained stable as long as it remained in the vicinity of the hydrophobic core of the protein made up of helices 2 and 3 (wtPrP, PrP-M205S and initial phase of PrP-M205R). Thus, it seems that in PrP^C helix 1 is stabilized by the nearby hydrophobic core, and the question arises, by which physical mechanism is this stabilization effected?

Concerning the answer to this question the quoted peptide studies, and a recent investigation by Megy et al. (43) provide important clues. It has been shown that the addition of large amounts of organic solvents can strongly stabilize α-helical structures (7–9) even in the case of the peptide (43), which folds into a β-hairpin in aqueous solution (10). Notably the effect of trifluoroethanol as cosolvent, which is well-known for its helix-stabilizing effect, was comparable to that of methanol. Therefore, the stabilization is largely nonspecific and, thus, should be due to the corresponding reduction of the dielectric constant ε_s (cf. Megy et al. (43) and Munishkina et al. (44)). If this conjecture holds, our simulation data concerning the behavior of helix 1 find a neat explanation.

According to our simulations, the amphiphilic helix 1 is stable as long as it is attached with its nonpolar face to the nonpolar surface (e.g., M205 of helix 3) of the hydrophobic core of the protein. At the surface of such a hydrophobic core, ε_s is smaller than within the bulk aqueous solvent, arguing that in wild-type PrP^C the amphiphilic helix 1 is stabilized by this low-dielectric environment. As observed in our M205R simulation, helix 1 becomes destabilized upon exposure to an aqueous environment, i.e., to an environment of larger ε_s. Thus, the increasing α-helical content of the helix 1 peptides at increasing concentrations of organic cosolvents determined by NMR and CD (7–9,43) agrees with and explains our M205R result.

Conversely, the suggested ε_s-dependence of the helix 1 stability sheds light on the functional role of residue M205. When this residue is present, it provides a hydrophobic attachment site to the few nonpolar residues contained in helix 1, which, during folding, can pull this highly charged part of the PrP sequence from the aqueous phase into the low-dielectric environment near the protein surface. By providing an environment of low ε_s, the hydrophobic core of PrP can thus act as an intramolecular chaperone supporting the folding of helix 1 into a stable structure. If that attachment site is eliminated by mutation into a polar or charged residue, as in M205S or M205R, this chaperone function is eliminated. The highly charged residues of helix 1 will stay solvated in a high ε_s environment, away from the hydrophobic core, and their folding into the rigid α-helical structure characteristic for native PrP^C will be prevented. This hampered folding explains why the two mutations have an effect identical to that of the deletion of helix 1 on the maturation of PrP^C (5). Thus, the functional role of M205 appears to be mainly to ensure the correct folding of helix 1.

Our study has thus provided evidence that the formation of the tightly packed hydrophobic core plays a major role in the folding of native PrP^C. It is interesting that this core can also be impaired by mutations linked to inherited prion diseases in humans. In a recent study, it was shown that two pathogenic PrP mutations within the hydrophobic core, T183A and F198S, show the same biochemical behavior as M205S/R (45), which, according to our results, is explained by a destabilization of the tertiary structure.

SUPPLEMENTARY MATERIAL

An online supplement to this article can be found by visiting BJ Online at http://www.biophys.org.

We are grateful to Matthias Schmitz and Armin Giese for stimulating and useful discussions.

This work was supported by the Bundesministerium für Bildung and Forschung (01KO0108) and by the Bavarian joint research project ForPrion (LMU02).

REFERENCES

1. Prusiner, S. B. 1998. Prions. *Proc. Natl. Acad. Sci. USA*. 95:13363–13383.
2. Riek, R., S. Hornemann, G. Wider, M. Billeter, R. Glockshuber, and K. Wuthrich. 1996. NMR structure of the mouse prion protein domain PrP(121–231). *Nature*. 382:180–182.
3. Zahn, R., A. Z. Liu, T. Luhrs, R. Riek, C. von Schroetter, F. L. Garcia, M. Billeter, L. Calzolai, G. Wider, and K. Wuthrich. 2000. NMR solution structure of the human prion protein. *Proc. Natl. Acad. Sci. USA*. 97:145–150.
4. Riek, R., G. Wider, M. Billeter, S. Hornemann, R. Glockshuber, and K. Wuthrich. 1998. Prion protein NMR structure and familial human spongiform encephalopathies. *Proc. Natl. Acad. Sci. USA*. 95:11667–11672.
5. Winklhofer, K. F., J. Heske, U. Heller, A. Reintjes, W. Muranyi, I. Moarefi, and J. Tatzelt. 2003. Determinants of the in vivo folding of the prion protein: a bipartite function of helix 1 in folding and aggregation. *J. Biol. Chem.* 278:14961–14970.
6. Speare, J. O., T. S. Rush, M. E. Bloom, and B. Caughey. 2003. The role of helix 1 aspartates and salt bridges in the stability and conversion of prion protein. *J. Biol. Chem.* 278:12522–12529.
7. Ziegler, J., H. Sticht, U. C. Marx, W. Muller, P. Rosch, and S. Schwarzinger. 2003. CD and NMR studies of prion protein (PrP) helix 1: novel implications for its role in the PrPC → PrP^{Sc} conversion process. *J. Biol. Chem.* 278:50175–50181.
8. Sharman, G. J., N. Kenward, H. E. Williams, M. Landon, R. J. Mayer, and M. S. Searle. 1998. Prion protein fragments spanning helix 1 and both strands of β sheet (residues 125–170) show evidence for predominantly helical propensity by CD and NMR. *Fold. Des.* 3:313–320.
9. Liu, A. Z., P. Riek, R. Zahn, S. Hornemann, R. Glockshuber, and K. Wuthrich. 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: 145–152.

10. Kozin, S. A., G. Bertho, A. K. Mazur, H. Rabesona, J. P. Girault, T. Haertle, M. Takahashi, P. Debey, and G. H. B. Hoa. 2001. Sheep prion protein synthetic peptide spanning helix 1 and β-strand 2 (residues 142–166) shows β-hairpin structure in solution. *J. Biol. Chem.* 276: 46364–46370.
11. Dima, R. I., and D. Thirumalai. 2004. Probing the instabilities in the dynamics of helical fragments from mouse PrPc. *Proc. Natl. Acad. Sci. USA.* 101:15335–15340.
12. Derreumaux, P. 2001. Evidence that the 127–164 region of prion proteins has two equi-energetic conformations with β or α features. *Biophys. J.* 81:1657–1665.
13. Morrissey, M. P., and E. I. Shakhnovich. 1999. Evidence for the role of PrPC helix 1 in the hydrophilic seeding of prion aggregates. *Proc. Natl. Acad. Sci. USA.* 96:11293–11298.
14. Caughey, B. W., A. Dong, K. S. Bhat, D. Ernst, S. F. Hayes, and W. S. Caughey. 1991. Secondary structure analysis of the scrapie-associated protein Prp 27–30 in water by infrared spectroscopy. *Biochemistry.* 30:7672–7680.
15. Pan, K. M., M. Baldwin, J. Nguyen, M. Gasset, A. Serban, D. Groth, I. Mehlhorn, Z. W. Huang, R. J. Fletterick, F. E. Cohen, and S. B. Prusiner. 1993. Conversion of α-helices into β-sheets features in the formation of the scrapie prion proteins. *Proc. Natl. Acad. Sci. USA.* 90:10962–10966.
16. Wille, H., M. D. Michelitsch, V. Guenebaut, S. Supattapone, A. Serban, F. E. Cohen, D. A. Agard, and S. B. Prusiner. 2002. Structural studies of the scrapie prion protein by electron crystallography. *Proc. Natl. Acad. Sci. USA.* 99:3563–3568.
17. Govaerts, C., H. Wille, S. B. Prusiner, and F. E. Cohen. 2004. Evidence for assembly of prions with left-handed β-helices into trimers. *Proc. Natl. Acad. Sci. USA.* 101:8342–8347.
18. Stork, M., A. Giese, H. A. Kretzschmar, and P. Tavan. 2005. Molecular dynamics simulations indicate a possible role of parallel β-helices in seeded aggregation of poly-Gln. *Biophys. J.* 88:2442–2451.
19. DeMarco, M. L., and V. Daggett. 2004. From conversion to aggregation: protofibril formation of the prion protein. *Proc. Natl. Acad. Sci. USA.* 101:2293–2298.
20. Berman, H. M., J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne. 2000. The Protein Data Bank. *Nucleic Acids Res.* 28:235–242.
21. Guex, N., and M. C. Peitsch. 1997. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis.* 18:2714–2723.
22. Jorgensen, W. L., J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein. 1983. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79:926–935.
23. MacKerell, A. D., D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, and M. Karplus. 1998. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B.* 102:3586–3616.
24. Mathias, G., B. Egwolf, M. Nonella, and P. Tavan. 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.* 118:10847–10860.
25. Berendsen, H. J. C., J. P. M. Postma, W. F. van Gunsteren, A. Dinola, and J. R. Haak. 1984. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* 81:3684–3690.
26. Krautler, V., W. F. van Gunsteren, and P. H. Hunenberger. 2001. A fast SHAKE: algorithm to solve distance constraint equations for small molecules in molecular dynamics simulations. *J. Comput. Chem.* 22: 501–508.
27. Eichinger, M., H. Grubmuller, H. Heller, and P. Tavan. 1997. FAMUSAMM: an algorithm for rapid evaluation of electrostatic interactions in molecular dynamics simulations. *J. Comput. Chem.* 18: 1729–1749.
28. Niedermeier, C., and P. Tavan. 1994. A structure-adapted multipole method for electrostatic interactions in protein dynamics. *J. Chem. Phys.* 101:734–748.
29. Niedermeier, C., and P. Tavan. 1996. Fast version of the structure adapted multipole method-efficient calculation of electrostatic forces in protein dynamics. *Mol. Simulat.* 17:57–66.
30. Mathias, G., and P. Tavan. 2004. Angular resolution and range of dipole-dipole correlations in water. *J. Chem. Phys.* 120:4393–4403.
31. Allen, M. P., and D. J. Tildesley. 1987. Computer Simulation of Liquids. Clarendon, Oxford, UK.
32. Kloppenburg, M., and P. Tavan. 1997. Deterministic annealing for density estimation by multivariate normal mixtures. *Phys. Rev. E.* 55: R2089–R2092.
33. Albrecht, S., J. Busch, M. Kloppenburg, F. Metze, and P. Tavan. 2000. Generalized radial basis function networks for classification and novelty detection: self-organization of optimal Bayesian decision. *Neural Netw.* 13:1075–1093.
34. Carstens, H., C. Renner, A. G. Milbradt, L. Moroder, and P. Tavan. 2005. Multiple loop conformations of peptides predicted by molecular dynamics simulations are compatible with nuclear magnetic resonance. *Biochemistry.* 44:4829–4840.
35. Kabsch, W., and C. Sander. 1983. Dictionary of protein secondary structure: pattern-recognition of hydrogen-bonded and geometrical features. *Biopolymers.* 22:2577–2637.
36. Pappalardo, M., D. Milardi, C. La Rosa, C. Zannoni, E. Rizzarelli, and D. Grasso. 2004. A molecular dynamics study on the conformational stability of PrP 180–193 helix II prion fragment. *Chem. Phys. Lett.* 390:511–516.
37. Sekijima, M., C. Motono, S. Yamasaki, K. Kaneko, and Y. Akiyama. 2003. Molecular dynamics simulation of dimeric and monomeric forms of human prion protein: insight into dynamics and properties. *Biophys. J.* 85:1176–1185.
38. Kallberg, Y., M. Gustafsson, B. Persson, J. Thyberg, and J. Johansson. 2001. Prediction of amyloid fibril-forming proteins. *J. Biol. Chem.* 276:12945–12950.
39. Chou, P. Y., and G. D. Fasman. 1974. Prediction of protein conformation. *Biochemistry.* 13:222–245.
40. Gasset, M., M. A. Baldwin, D. H. Lloyd, J. M. Gabriel, D. M. Holtzman, F. Cohen, R. Fletterick, and S. B. Prusiner. 1992. Predicted α-helical regions of the prion protein when synthesized as peptides form amyloid. *Proc. Natl. Acad. Sci. USA.* 89:10940–10944.
41. Thompson, A., A. R. White, C. McLean, C. L. Masters, R. Cappai, and C. J. Barrow. 2000. Amyloidgenicity and neurotoxicity of peptides corresponding to the helical regions of PrPC. *J. Neurosci. Res.* 62: 293–301.
42. Tizzano, B., P. Palladino, A. De Capua, D. Marasco, F. Rossi, E. Benedetti, C. Pedone, R. Ragone, and M. Ruvo. 2005. The human prion protein α2 helix: a thermodynamic study of its conformational preferences. *Proteins.* 59:72–79.
43. Megy, S., G. Bertho, S. A. Kozin, P. Debey, G. H. B. Hoa, and J. P. Girault. 2004. Possible role of region 152–156 in the structural duality of a peptide fragment from sheep prion protein. *Protein Sci.* 13:3151–3160.
44. Munishkina, L. A., C. Phelan, V. N. Uversky, and A. L. Fink. 2003. Conformational behavior and aggregation of α-synuclein in organic solvents: modeling the effects of membranes. *Biochemistry.* 42:2720–2730.
45. Kiachopoulos, S., A. Bracher, K. F. Winklhofer, and J. Tatzelt. 2005. Pathogenic mutations located in the hydrophobic core of the prion protein interfere with folding and attachment of the glycosylphosphatidylinositol anchor. *J. Biol. Chem.* 280:9320–9329.

Supplemental Material

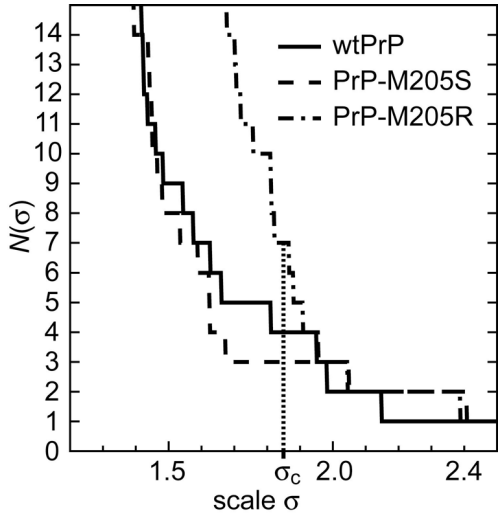

Figure: Numbers $N(\sigma)$ of prototypes for the three PrP-variants as functions of the spatial scale σ. Characteristically, the numbers of prototypes $N(\sigma)$ decrease stepwise and monotonously with increasing σ, i.e. with decreasing resolution. At the selected common scale $\sigma_c = 1.85$ the trajectory of wtPrP exhibits the four prototypical conformations wt1-wt4, PrP-M205S shows the three conformations S1-S3, and PrP-M205R is described by the states R1-R7.

3 β-helikale Modelle für Amyloid-Strukturen

Nun soll die Entwicklung von β-helikalen Modellen für die Strukturen amyloider Fibrillen aus Poly-Glutamin und für PrP^{Sc} beschrieben werden. Während sich Abschnitt 3.1 hauptsächtlich mit den Grundbausteinen dieser Fibrillen beschäftigt, einschließlich der Verifizierung entsprechender Modellvorschläge durch MD-Simulationen, veranschaulicht Abschnitt 3.2 am Beispiel von PrP^{Sc} die Zusammensetzung dieser Grundstrukturen zu Amyloid-Fibrillen.

3.1 Grundstrukturen für Fibrillen aus poly-Gln und für PrP^{Sc}

Der folgende Abschnitt ist ein Abdruck[1] des Artikels

> Martina Stork, Armin Giese, Hans A. Kretzschmar and Paul Tavan:
> „Molecular dynamics simulations indicate a possible role of parallel β-helices in seeded aggregation of poly-Gln.“
> *Biophysical Journal* 88, 2442-2451 (2005),

den ich zusammen mit Paul Tavan, sowie mit Armin Giese und Hans Kretzschmar vom Zentrum für Neuropathologie und Prionforschung der LMU München im Biophysical Journal veröffentlicht habe. Zu dieser Veröffentlichung findet sich auf der Internetseite das Biophysical Journal[2] Zusatzmaterial, in dem wir unter anderem den Einfluss der Temperatur auf die Stabilität der poly-Gln-Peptide demonstrieren. Dieses Material ist hier im Anschluss an den Artikel ebenfalls abgedruckt.

[1] Mit freundlicher Genehmigung des Biophysical Journal.
[2] http://www.biophysj.org

Molecular Dynamics Simulations Indicate a Possible Role of Parallel β-Helices in Seeded Aggregation of Poly-Gln

Martina Stork,* Armin Giese,† Hans A. Kretzschmar,† and Paul Tavan*
*Theoretische Biophysik, Lehrstuhl für BioMolekulare Optik, Ludwig-Maximilians-Universität, D–80538 Munich, Germany; and †Zentrum für Neuropathologie und Prionforschung, Ludwig-Maximilians-Universität, Munich, Germany

ABSTRACT The molecular structures of amyloid fibers characterizing neurodegenerative diseases such as Huntington's or transmissible spongiform encephalopathies are unknown. Recently, x-ray diffraction patterns of poly-Gln fibers and electron microscopy images of two-dimensional crystals formed from building blocks of prion rods have suggested that the corresponding amyloid fibers are generated by the aggregation of parallel β-helices. To explore this intriguing concept, we study the stability of small β-helices in aqueous solution by molecular dynamics simulations. In particular, for the Huntington aggregation nucleus, which is thought to be formed of poly-Gln polymers, we show that three-coiled β-helices are unstable at the suggested circular geometries and stable at a triangular shape with 18 residues per coil. Moreover, we demonstrate that individually unstable two-coiled triangular poly-Gln β-helices become stabilized upon dimerization, suggesting that seeded aggregation of Huntington amyloids requires dimers of at least 36 Gln repeats (or monomers of ~54 Gln) for the formation of sufficiently stable aggregation nuclei. An analysis of our results and of sequences occurring in native β-helices leads us to the proposal of a revised model for the PrP^{Sc} aggregation nucleus.

INTRODUCTION

Amyloid fibers are β-sheet-rich, large, and insoluble aggregates of peptides or proteins. Deposition of such aggregates in the brain is the pathological hallmark of common neurodegenerative diseases. The deposits and smaller aggregation intermediates appear to be involved in the pathogenesis of neuronal cell death (Caughey and Lansbury, 2003). Because the fibers are large and hard to crystallize, high-resolution structural analysis is severely hampered. In addition, it is notoriously difficult to stabilize aggregation intermediates for this purpose.

Although the peptides forming amyloid fibers exhibit little or no sequence similarity, the fibers are assumed to have common structural motifs (Sunde et al., 1997). The aggregation is supposed to proceed through a structural conversion and oligomerization leading to an ordered aggregation nucleus, which initiates fiber growth (Rochet and Lansbury, 2000). In x-ray diffraction many fibers show the so-called cross-β pattern, which is characterized by a sharp 4.75-Å meridional and a broader 10-Å equatorial reflection (Sunde and Blake, 1998). Fig. 1, *a* and *b*, depicts a sample structure that fits to such a pattern. Here, stacks of antiparallel β-sheets form a hydrogen-bonded aggregate along the fiber axis. The meridional 4.75-Å reflection is due to the spacing between the main chains of the hydrogen-bonded β-strands (Fig. 1 *a*). For the given example, the equatorial 10-Å reflection is due to the parallel stacking of the backbone in the cross section of the fiber visible in Fig. 1 *b*. Note that the cross-β pattern is also compatible with double-layered β-strands aggregating through in-register parallel β-sheet formation along the fiber axis as suggested by Petkova et al. (2002) for $A\beta_{1-40}$ on the basis of solid-state NMR-data.

However, the 10-Å reflection is absent or very weak in x-ray diffraction patterns of amyloid fibers associated with Huntington's disease and with particular fragments of the Alzheimer Aβ peptide (Perutz et al., 2002a). Perutz et al. (2002a) have focused on the structure of these fibers. Also here, a strong meridional reflection at 4.75 Å indicates that β-sheets must be oriented perpendicular to the fiber axis. On the other hand, the absence of the 10-Å reflections shows that the backbone cannot exhibit a parallel stacking. Instead, a circular β-helix (Fig. 1 *c*) as well as other shapes (Jenkins and Pickersgill, 2001) of parallel β-helices (like the one in Fig. 1 *e*) are compatible with such a diffraction pattern.

As a particular example, Perutz et al. (2002a) have interpreted the x-ray diffraction patterns of fibers aggregated from short poly-L-Gln peptides $D_2Q_{15}K_2$ in terms of a model, according to which the fibers have a diameter of ~3 nm and are made up of circular β-sheets. This example is important, because Huntington's disease, like other related neurodegenerative diseases, is caused by large (≥35) Gln repeats (Ross et al., 2003). Furthermore, fibers formed by an exon-1 protein of huntingtin, which contains a 51-Gln repeat, exhibit similar x-ray diffraction patterns (Perutz et al., 2002b). These and other facts have led Perutz et al. (2002a) to speculate that the poly-Gln segments directly promote the growth of the respective amyloid fibers and provide the required building blocks in the form of at least two coils of circular parallel β-helices comprising 18 or 20 residues each (Perutz et al., 2002a). (Adopting the nomenclature suggested by Jenkins

Submitted September 8, 2004, and accepted for publication January 14, 2005.

Address reprint requests to Paul Tavan, Theoretische Biophysik, Lehrstuhl für BioMolekulare Optik, LMU, Oettingenstrasse 67, D–80538 Munich, Germany. Tel.: 49-89-2180-9220; Fax: 49-89-2180-9202; E-mail: paul.tavan@physik.uni-muenchen.de.

0006-3495/05/04/2442/10 $2.00

doi: 10.1529/biophysj.104.052415

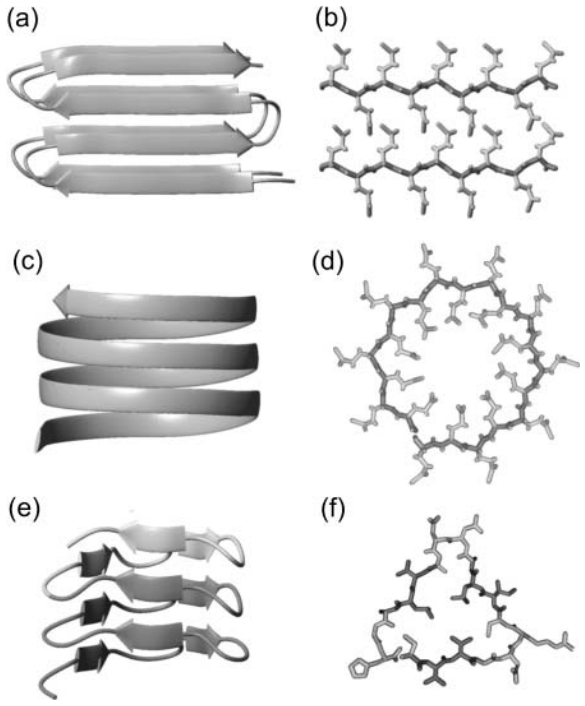

FIGURE 1 Model of stacked antiparallel β-sheets corresponding to a cross-β x-ray pattern (*a* and *b*), as well as structures of a circular (*c* and *d*) and a triangular (*e* and *f*) parallel β-helix with 20 and 18 residues per helical coil, respectively. The β-helices are compatible with the absence of the 10-Å reflection. The circular model visualizes a suggestion by Perutz et al. (2002a) for the building blocks of poly-Gln amyloid fibers. The triangular structure is a three-coiled fragment of a native left-handed β-helix in the enzyme LpxA (Raetz and Roderick, 1995) (residues 125–179); dark shaded marks a two-coiled fragment. (*a*, *c*, and *e*) Side views. (*b*, *d*, and *f*) Cross sections; in the cross-section *f* covering the residues 127–144, three β-strands (*dark shaded*) are separated by turns (*light shaded*).

and Pickersgill (2001), we call a complete turn of a parallel β-helix a *coil*.) Fig. 1, *c* and *d*, shows a three-coiled sample structure constructed according to the suggestion of Perutz et al. (2002a) for 55 and 61 poly-Gln repeats. In these circular parallel β-helices the side chains point alternatingly into and out of the fiber axis. Note that the cylindrical structures, which were suggested by Perutz et al. (2002a) to be formed from these β-helical building blocks and to represent the core motif for huntingtin aggregation, do not exactly match any of the native parallel β-helical structures classified by Jenkins and Pickersgill (2001).

At the same time, a parallel β-helix motif has also been proposed by Wille et al. (2002) for scrapie prion. According to these authors, among the known secondary structural motifs only parallel β-helices appeared to be compatible with electron microscopy images of two-dimensional crystals formed from building blocks of prion rods.

This unusual secondary structure motif is known from several protein families in which left- and right-handed β-helices are found (Jenkins and Pickersgill, 2001). Many left-handed β-helices resemble an equilateral prism as in the case of UDP-*N*-acetylglucosamine acyltransferase (LpxA) (Raetz and Roderick, 1995), contain 18 residues per helical coil, and occasionally exhibit interruptions of the helical coils by loop structures. Fig. 1 *e* depicts a three-coiled β-helical fragment of LpxA and Fig. 1 *f* one of its helical coils. It consists of three β-strands linked by turns. Six of the 18 side groups are oriented toward the core of the helix, which is densely packed with hydrophobic or weakly polar residues (Raetz and Roderick, 1995). In contrast, the sequences and shapes of right-handed β-helices are less regular, and the sizes of their helical coils are larger (Jenkins and Pickersgill, 2001). Whereas all these proteins avoid aggregation by capping the β-helices (Richardson and Richardson, 2002), an isolated β-helix domain of the P22 tailspike protein quickly aggregates into amyloid-like fibers by linear polymerization (Schuler et al., 1999). To explain the corresponding data the parallel β-helix has been speculated to provide the structural motif of fiber growth (Schuler et al., 1999).

One may ask whether, and how, one can substantiate or reject the quoted speculations on the structure of amyloid fibers. Here, we claim that molecular dynamics (MD) simulations (van Gunsteren and Berendsen, 1990) can provide additional evidence, because they can provide measures for the stability of suggested building blocks.

Processes of protein folding and aggregation sensitively depend on the thermodynamic conditions and on a subtle balance of the electrostatic interactions within the protein-solvent system (Warshel and Russell, 1984) which, therefore, have to be adequately represented in MD simulations (Tavan et al., 2005). To generate a well-defined thermodynamic ensemble in MD, one has to enclose the protein and a small (3- to 8-nm) neighborhood of solvent molecules by periodic boundaries (Allen and Tildesley, 1987). For an adequate treatment of the electrostatic interactions one has to apply either Ewald (Frenkel and Smit, 2002) or moving-boundary reaction-field techniques (Mathias et al., 2003; Mathias and Tavan, 2005).

MD simulations of protein-solvent systems comprising 10^4–10^5 atoms are time-consuming. Therefore, they are currently limited to small peptides (like those shown in Fig. 1) in solution and to processes occurring within $\sim$10 ns. Nevertheless, the accessible timescale should be long enough to investigate the stability of selected and predefined β-helical peptide structures. If a small model structure, like the one depicted in Fig. 1 *c*, should dissolve within nanoseconds upon simulation of its thermal motion in aqueous solution, then the current MD simulation techniques (Mathias et al., 2003, Lindahl et al., 2001) should be reliable enough (Ponder and Case, 2003) to allow the prediction that this structure is thermodynamically unstable. In processes of amyloid fiber formation such a lack of short-time stability is important, because a correspondingly flexible monomeric structure cannot serve as a template and nucleus for aggregation. Conversely, if a suggested structure does not decay

within 10 ns of MD simulation, this may be taken as an indication of its metastability (because it cannot be excluded that it will decay at longer timescales) and, therefore, of its being a candidate nucleus for aggregation. The latter evidence can then be strengthened, if the metastability becomes more pronounced upon a rise of temperature or upon oligomerization. We hold this optimistic view on the capabilities of MD simulations, despite the serious deficiencies of the available molecular mechanics (MM) force fields (particularly concerning the neglect of the electronic polarizability (Ponder and Case, 2003)). Of course, the working hypothesis outlined above has to be validated by first considering β-helical sample structures, whose stabilities can be a priori estimated from general arguments.

Adopting this MD-based testing scenario, we want to check the suggestion of Perutz et al. (2002a) according to which the poly-Gln aggregation nuclei are circular β-helices (cf. Fig. 1 *c*) aggregating into water-filled nanotubes. Here we will exclude from consideration the short $D_2Q_{15}K_2$ model peptides, which according to Perutz et al. form circles aggregating into cylinders, because the extremely slow kinetics of their seeded aggregation (Chen et al., 2001) indicates the necessity of forming large oligomers for nucleation in this case. Instead we will consider larger Gln polymers comprising at least 37 residues, because for such peptides the aggregation kinetics becomes much faster (Chen et al., 2001). Our decision to study Gln polymers is additionally motivated by the fact that they pose no difficulties in matching a sequence onto a structural model. We have selected left-handed β-helices, because they exhibit a smaller variability of shape, size, and sequence than the right-handed β-helices and occur at smaller sizes of the helical coils. These sizes are compatible with the 3-nm diameter determined by Perutz et al. (2002a) for their poly-Gln fibers. Therefore, we will model the circular β-helices suggested by these authors as left-handed.

A related MD study also aiming at a stability assessment of amyloid structural motifs has been presented by Ma and Nussinov (2002a,b). These authors have studied stacks of β-sheets like those depicted in Fig. 1, *a* and *b*, for alanine-rich peptides and sequence portions of the Alzheimer Aβ-peptide.

METHODS

To construct a supposedly stable β-helix model, we selected two and three regularly shaped coils from LpxA (entry 1LXA (Berman et al., 2000) of the Protein Data Bank) as depicted in Fig. 1 *e* and characterized in the figure caption. These β-helical fragments start and terminate at locations within the turns, because many native β-helices start, end, or are interrupted by loops at turn positions. By acetylating (Ac) the N- and amidating (NH_2) the C-terminus, we avoid terminal charges and emulate the fragments as parts of the native β-helix.

Triangular three-coiled models were built by matching poly-Ala, poly-Gln, poly-Ser, and poly-Ile sequences onto the structure of the native LpxA fragment. For poly-Gln an additional two-coiled triangular structure was modeled. A dimer was formed by arranging two of these structures into an overall four-coiled β-helix. To achieve a close fit, here, the two-coiled monomers were restricted to 36 glutamines by omitting the N-terminal residue.

Circular poly-Gln tubes with 18 and 20 residues per coil (cf. Fig. 1 *c*) were built with InsightII (Accelrys, San Diego, CA) in the suggested dimensions (Perutz et al., 2002a). Starting from planar poly-Gln β-sheets with 55 and 61 residues, respectively, the peptides were bent into circles by modifying the angles between the C_α atoms. Stretching the resulting circular β-helix into the axial direction rendered the 4.75 Å helical pitch.

MD simulations

As our simulation system we chose a periodic rhombic dodecahedron large enough to enclose a sphere with a radius of 34 Å. Initially, this system was filled with 7813 water molecules modeled by the CHARMM variant of the TIP3P potential (Jorgensen et al., 1983; MacKerell et al., 1998) because the parameters of the peptide force field were also adopted from CHARMM22 (MacKerell et al., 1998).

All simulations were carried out in the *NpT* ensemble with the program package EGO-MMII (Mathias et al., 2003). The temperature T and the pressure p were controlled by a thermostat ($\tau = 0.1$ ps) and a barostat ($\tau = 1.0$ ps, $\beta = 5.0$ Pa) (Berendsen et al., 1984). Covalent bonds of hydrogen atoms were kept fixed using M-SHAKE (Kräutler et al., 2001). The integration time step was 2 fs.

The long-range Coulomb interactions were treated by a particular combination of structure-adapted multipole expansions (Niedermeier and Tavan, 1994) with a moving-boundary reaction-field approach and a multiple time-step integrator (Mathias et al., 2003, Mathias and Tavan, 2004). Van der Waals interactions were calculated explicitly up to 10.5 Å; at larger distances a mean field approach (Allen and Tildesley, 1987) was applied.

For solvation, the peptides were positioned at the center of the equilibrated (1 ns, $T = 300$ K, $p = 1.013 \times 10^5$ Pa) water box and all water molecules <2.0 Å from a peptide atom were removed. For equilibration the peptides initially were kept fixed, whereas the surrounding solvent molecules were thermally moving for several hundred picoseconds at $T = 500$ K and $T = 300$ K. Next, the rigid constraints were removed and solely the positions of the peptide C_α atoms were constrained by harmonic potentials (force constant 2.1×10^2 kJ/(mol·Å^2)). By energy minimization these systems were cooled within 1 ps to below $T = 0.1$ K and subsequently heated within 120 ps to $T = 300$ K. Within another 300 ps the constraining force constants were slowly reduced to zero until the peptides were free to move within the solvent. This procedure served to adjust the modeled peptide structures to the MM force field or, equivalently, to partially remove the prejudices imposed on the structures by the modeling. The simulation systems thus obtained were the starting points for the following unconstrained 2- to 10-ns simulations at $T = 300$ K and $p = 1.013 \times 10^5$ Pa (coordinates saved every 0.2 ps).

Measures for structural stability

Initial structures of the various coils j, $j = 1, \ldots, M$, within an M-coiled β-helix, were obtained as averages over the time interval from 100 ps to 400 ps of the unconstrained simulations. The subsequently sampled MD structures were matched onto the respective initial structures by least-square fits using the positions of the C_α atoms. The resulting root mean-square deviations $d^j_{\mathrm{rms}}(t)$ were taken as measures of the structural stabilities of the coils j. To judge the overall stability of a β-helical peptide containing M coils, the average value $\langle d_{\mathrm{rms}}\rangle(t) = (1/M)\sum_{j=1}^{M} d^j_{\mathrm{rms}}(t)$ of the coil-specific deviations $d^j_{\mathrm{rms}}(t)$ was calculated.

A huge number of further observations covering empirical measures for secondary-structure content, hydrogen bonding, simulated (Φ, Ψ) dihedral-angle distributions, and deviations of these dihedral angles from their initial values has also been calculated for cross-checking conclusions derived from

the coil-specific and coil-average Cartesian root mean-square deviations. However, in this article we will not include this huge amount of data, which can be extracted from the virtual reality of MD simulations, because we want to save space and avoid an overstretching of the reader's patience. Images of the peptides were created with the molecular graphics program MOLMOL (Koradi et al., 1996).

RESULTS AND DISCUSSION

The well-known deficiencies of the current nonpolarizable MM force fields (Ponder and Case, 2003), which are shared by the force field employed by us (MacKerell et al., 1998), generally preclude reliable predictions of peptide structures by MD simulation. Nevertheless, one can expect that natively stable peptide secondary structures, such as α-helices or β-sheets, at least represent local minima of the free-energy landscape generated by such a force field. If this is the case, the corresponding peptide conformations represent metastable states and in short-time simulations their metastability shows up as a conformational invariance for a certain time span. The average size of this time span is determined by the minimal height of the free energy barriers surrounding the given local minimum or, equivalently, by the depth of that minimum.

If one starts an MD simulation with a peptide structure, which is close to a metastable conformation in configuration space, one expects only slight structural rearrangements of the given starting structure, because it will rapidly relax toward that metastable state. On the other hand, if a starting structure is far away from a metastable conformation, it will become rapidly and strongly distorted.

In this study, we apply the method of MD simulation to a secondary structure motif, the left-handed parallel β-helix, which, to our best knowledge, has not yet been treated by MD simulation before. Therefore, to judge the stabilities of our peptide structures in solution, we first have to construct a stability scale. To this aim we will use simulation results on structures, for which we a priori expect grossly different stabilities from general arguments.

Stability scale

According to the above arguments, natively stable β-helices should show signs of structural stability also in our nanosecond MD simulations. To check whether this is actually the case and how stability is monitored by MD, we first consider the two- (*dark shaded*) and three-coiled fragments (*dark and light shaded*) of the LpxA helix depicted in Fig. 1 *e*.

In solution the stability of such a fragment will decrease, if it has charged N- and C termini, because the charged ends will be drawn into the solvent. Thus, such a fragment will be artificially destabilized as compared to the native situation. To check as to whether this effect shows up in our short-time simulations, we have simulated the LpxA fragments with charged and with neutralized ends, respectively, at 300 K for 2 ns starting from the structures shown in Fig. 1 *e*.

Fig. 2, *a* and *b*, compares the two-coiled structures observed after 2-ns simulation. Fig. 2 *a* demonstrates that the solvation of the charged termini is complete after this short time span and has broken up the β-sheet hydrogen bonds in a zipper-like fashion. In contrast, the 2-ns structure of the fragment with neutral termini shown in Fig. 2 *b* still resembles the starting structure (cf. Fig. 1 *e*, *dark shaded*), if one disregards a small conformational transition at one of the turns within the N-terminal coil 1.

Thus, the simple two-coiled example of a native β-helical fragment has shown that the destabilizing solvation effect induced by charged ends is strong enough to be observable within a short (2-ns) MD simulation. On the other hand, one may also expect that the depth of the local free-energy minimum corresponding to a fragment of natively stable structure will increase with the size of that fragment. Therefore, the question arises whether short-time simulations can catch also this stabilizing effect counteracting the destabilization induced by the solvation tendency of the charged ends.

For an answer consider Fig. 2, *c* and *d*, which compares the 2-ns structures of the three-coiled models obtained for charged and neutral ends, respectively. Apart from very small conformational changes near the termini, the starting structure is preserved in both models. Thus, the stabilization induced by the additional coil sufficiently deepens the local free energy minimum of the native conformation and prevents a thermally activated escape within 2 ns independently of the destabilizing charge solvation effect.

Because we want to evaluate the stability of small β-helical structures undisturbed by these solvation effects we have generally modeled the fragments of the native LpxA

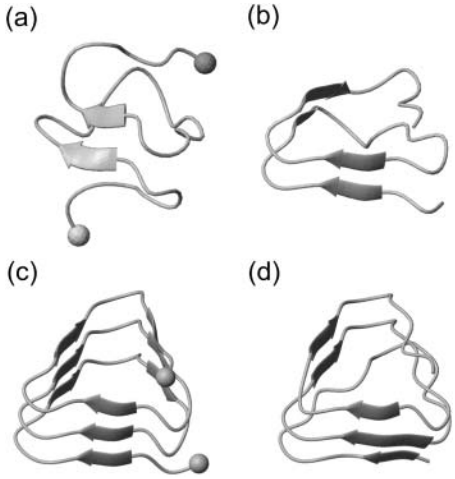

FIGURE 2 Snapshots of LpxA fragments after 2 ns of MD simulation: two coils (*a*) with charged N- and C-termini marked by spheres, and (*b*) with neutral termini; three coils (*c*) with charged termini (*spheres*), and (*d*) with neutral termini.

as well as the other β-helical peptides described in Methods with neutral, i.e., acetylated and amidated, termini. To check whether the asserted stability of the two-coiled LpxA fragment (neutral ends) pertains to longer time spans than the 2 ns discussed above, we have simulated its dynamics for another 2 ns.

According to Fig. 3 *a*, the root mean-square deviations $d^j_{\mathrm{rms}}(t)$ increase at both coils, $j = 1, 2$, after $\sim$1 ns (see Methods for the definition of $d^j_{\mathrm{rms}}(t)$). This increase reflects the conformational change at one of the turns, which we have already mentioned in connection with Fig. 2 *b*. The additional strong increase of $d^1_{\mathrm{rms}}(t)$ after 3 ns is caused by a loosening of the N-terminus, which is apparent in the 4-ns structural snapshot depicted in Fig. 3 *b*.

As a result, the interaction between two coils is apparently too weak to keep a hydrated two-coiled fragment of a native β-helix stable within a 4 ns MD simulation. Therefore we checked, whether adding a third coil suffices to stabilize the native structure for 4 ns or longer.

Fig. 3 *c* shows the 10-ns evolution of the $d^j_{\mathrm{rms}}(t)$ values for the three LpxA coils (cf. Fig. 1 *e*). At the coils $j = 2, 3$ the $d^j_{\mathrm{rms}}(t)$ values remain small for $\sim$6.5 ns, indicating that these coils keep their respective initial conformations. Visual inspection of the structures of the N-terminal coil 1 reveals that the increase of $d^1_{\mathrm{rms}}(t)$ at $\sim$1 ns is caused by a small conformational change at one of the turns, which is similar to the one noted above for the two-coiled structure (data not shown). As demonstrated by the average structures of the coils during the first 6.5 ns (drawn in *black* in Fig. 3 *d*), the fragment remains essentially stable within this time span. At 6.5 ns the concomitant rise of all three $d^j_{\mathrm{rms}}(t)$ indicates a conformational change whose nature becomes apparent by comparison of the light shaded (averages over the last 3.5 ns) and black structures in Fig. 3 *d*. Accordingly, the conformational change is confined to the top and bottom coils of the fragment, whereas the central coil remains nearly unchanged.

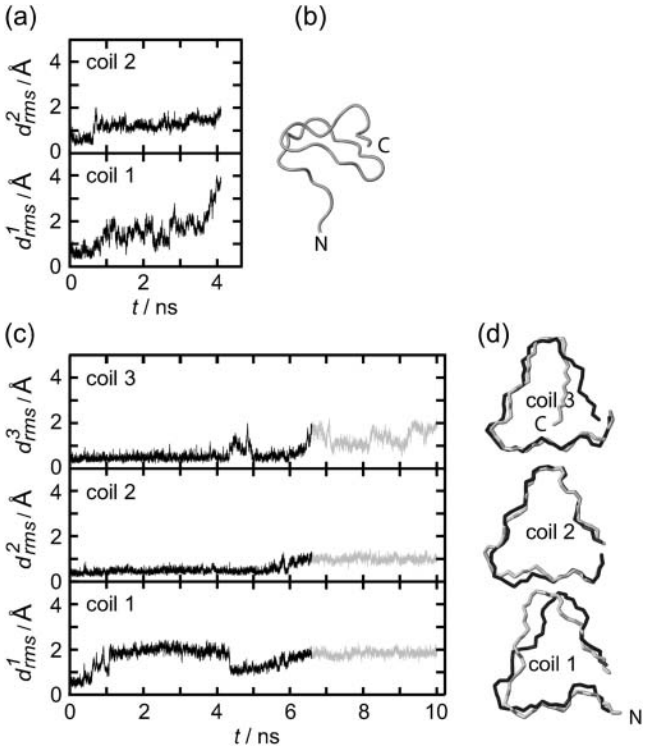

FIGURE 3 Temporal evolution of the root mean-square deviations $d^j_{\mathrm{rms}}(t)$ of the various coils j in MD simulations of two- (*a*) and three-coiled (*c*) LpxA fragments. Coils are numbered starting at the N-terminus. (*b*) Snapshot of the two-coiled fragment after 4 ns. (*c*) Two simulation periods marking different conformations (*black* and *light shaded*) of the three-coiled fragment are distinguished in the plots of $d^j_{\mathrm{rms}}(t)$. (*d*) Average structures of the three coils for the black and light shaded periods.

As a result, the three-coiled fragment of LpxA is much more stable than the two-coiled fragment, and this enhanced stability is particularly pronounced at the central coil, which is sandwiched by the two partially water-exposed coils and, therefore, experiences an approximately native molecular environment. We conclude 1), that a natively stable β-helical coil remains stable for 10 ns in our MD simulations if it is in a native environment, and 2), that the stability of hydrated fragments of a native β-helix increases with the number of coils. Furthermore, the kind of structural stability, which is characterized by the $d^j_{\mathrm{rms}}(t)$ curves depicted in Fig. 3 *c*, defines a stability scale for our MD approach:

Three-coiled β-helix peptides with similar $d^j_{\mathrm{rms}}(t)$ curves will be called ''stable'', whereas peptides exhibiting a stronger increase of the $d^j_{\mathrm{rms}}(t)$ values will be classified as ''unstable''. This approach relies on the assumption that the MM force field can identify local minima in the free energy landscape, can account for basic physicochemical effects deepening or flattening these minima, and does not require the accuracy necessary for quantitative predictions of thermodynamic stability or of transition rates between metastable states.

To check the validity of the stability scale defined above, we additionally analyzed a triangular three-coiled poly-Ala β-helix. We a priori expect that this structure is unstable, because poly-Ala peptides preferentially form α-helices (Chakrabartty et al., 1991; Blondelle et al., 1997). Fig. 4 *a* depicts the evolution of the average deviation $\langle d_{\mathrm{rms}} \rangle(t)$ of the corresponding coil structures from the initial values (see

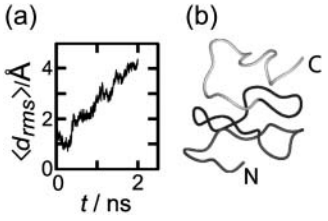

FIGURE 4 (*a*) Average root mean-square deviation of the β-helical coils (cf. Methods) for a three-coiled triangular poly-Ala β-helix during 2 ns of simulation. (*b*) Snapshot of structure at 2 ns.

Methods for the definition of $\langle d_{\mathrm{rms}}\rangle(t)$). The deviation $\langle d_{\mathrm{rms}}\rangle(t)$ rapidly increases within 2 ns to reach a value of 4.4 Å. This value is larger than the corresponding value of 2.8 Å reached for the unstable two-coiled LpxA fragment after 4 ns, and much larger than the 10-ns value of 1.6 Å for the stable three-coiled LpxA fragment. Thus, our stability scale classifies the three-coiled poly-Ala β-helix as ''very unstable''. The 2-ns structural snapshot in Fig. 4 *b* confirms that the poly-Ala peptide is actually decaying toward a random coil.

This result verifies that our MD-based testing can provide reasonable estimates for the stabilities of small β-helical peptides. Despite the inaccuracies of MM-force fields (Ponder and Case, 2003) it should allow us to reliably judge the stabilities of related β-helical models.

Poly-Gln models

We first address the question of whether the three-coiled circular poly-Gln β-helices containing 18 and 20 residues per coil suggested by Perutz et al. (2002a) are stable. Fig. 5, *a* and *b*, shows the corresponding $\langle d_{\mathrm{rms}}\rangle(t)$ curves obtained from 10-ns simulations. Immediately after the start of the unconstrained simulations, the $\langle d_{\mathrm{rms}}\rangle(t)$ values steeply increase and reach a value of ~2 Å at 2 ns, in both cases indicating that the β-helices quickly deviate from their initial circular shapes. The quoted 2-ns value is two times larger than the corresponding 2-ns value of 1.0 Å observed for the stable three-coiled LpxA fragment, and therefore, both circular poly-Gln models are classified as unstable.

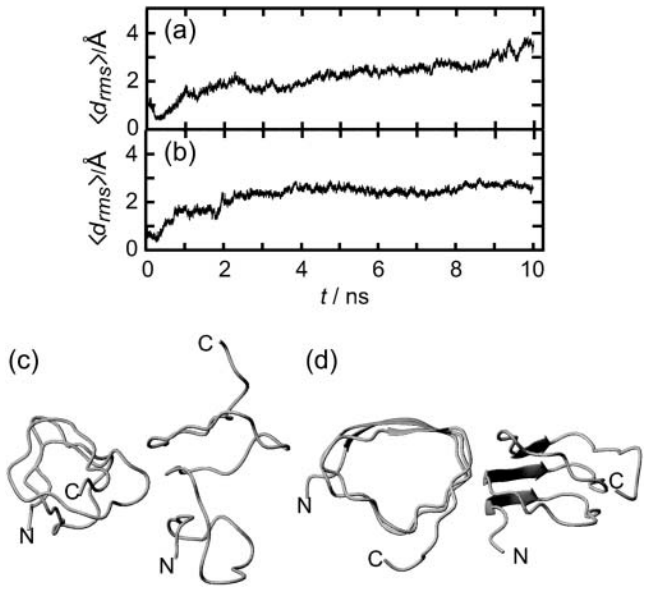

FIGURE 5 Temporal evolution of the $\langle d_{rms}\rangle(t)$ values of circular poly-Gln β-helices with 18 (*a*) and 20 residues (*b*) per coil. (*c* and *d*) Snapshots of the respective structures after 10 ns.

In fact, an analysis of the structures acquired after 2 ns reveals that the circular β-strand coils, in which the Gln residues alternatingly pointed toward and away from the center of the helix, have decayed by partial or complete outward side-chain flips into coil structures now exhibiting partial or complete turns at the respective flip positions. Concomitantly with these fast conformational flips several water molecules initially filling the cores of the circular models were sqeezed out into the aqueous environment or into the region between adjacent coils (data not shown). Correspondingly, the initial conformational relaxation of the two circular models may be described as a beginning hydrophobic collapse and as the breaking of the circular symmetry by introduction of turns and of undistorted β-strands. Beyond 2 ns, the $\langle d_{\mathrm{rms}}\rangle(t)$ curves of the two circular β-helices start to differ:

For the 18-member coils (cf. Fig. 5 *a*) the average deviation $\langle d_{\mathrm{rms}}\rangle(t)$ continues to increase until it reaches a value of 3.7 Å at 10 ns, which is nearly as large as the 2-ns value (4.4 Å) of the very unstable poly-Ala model and by 2.1 Å larger than the 10 ns value of the stable three-coiled LpxA fragment. Therefore, the considered circular poly-Gln β-helix must be very unstable. This judgment is corroborated by the 10-ns snapshot shown in Fig. 5 *c*. Instead of the initial circular shape a nearly random coil is seen.

For the 20-member coils, in contrast, the $\langle d_{\mathrm{rms}}\rangle(t)$ curve in Fig. 5 *d* does not show a likewise continuous and large increase after the first 2 ns. Although the slight further increase signifies ongoing conformational changes, the near constancy of the curve between 4 ns and 10 ns seems to indicate that the peptide is approaching a new stable state. The 10-ns snapshot in Fig. 5 *d* shows that the initial circular shape is lost. Interestingly, the coils now exhibit several turns while retaining an overall structure, which remotely resembles a triangular β-helix. This finding has inspired us to speculate that triangular poly-Gln β-helices might be stable, and therefore we carried out simulations of such poly-Gln structures.

Fig. 6 *a* shows the 10-ns temporal evolution of the $d^{\mathrm{j}}_{\mathrm{rms}}(t)$ values for the three coils of the triangular poly-Gln β-helix (cf. Methods). Surprisingly, all three $d^{\mathrm{j}}_{\mathrm{rms}}(t)$ curves behave just like those of the native LpxA fragment in Fig. 3 *c*. After 10 ns, the value of $d^{2}_{\mathrm{rms}}(t)$ belonging to the central coil 2 is $<$0.7 Å, whereas the partially water-exposed coils 1 and 3 exhibit higher values (1.5 Å and 2.0 Å, respectively). As a result, the three-coiled triangular poly-Gln β-helix is apparently as stable as the corresponding native LpxA fragment. In Fig. 6 *b*, the black and light shaded structures represent averages over the first 2.5 ns and the last 3 ns, respectively. Their good match visualizes the asserted structural stability. In particular, the central coil 2 is seen to be perfectly preserved.

To check this surprising result we have carried out additional 8-ns stability tests at an elevated temperature of 330 K both for the triangular three-coiled poly-Gln model

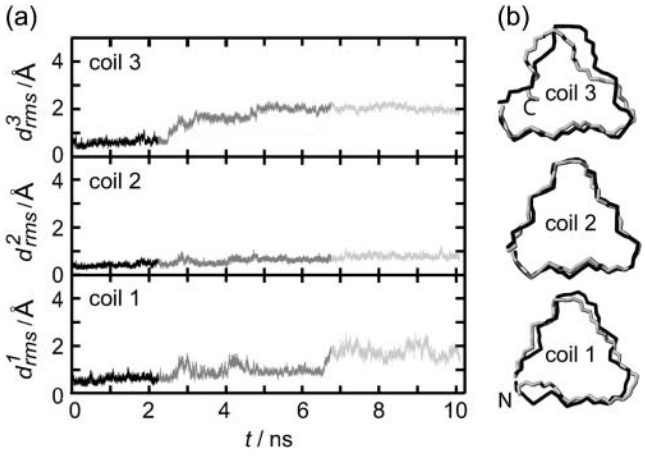

FIGURE 6 (*a*) $d^j_{rms}(t)$ curves for the various coils j of a three-coiled triangular poly-Gln β-helix. The gray scale (*black*, *shaded*, and *light shaded*) distinguishes three simulation periods. (*b*) Average structures of the three coils for the black and light shaded simulation periods.

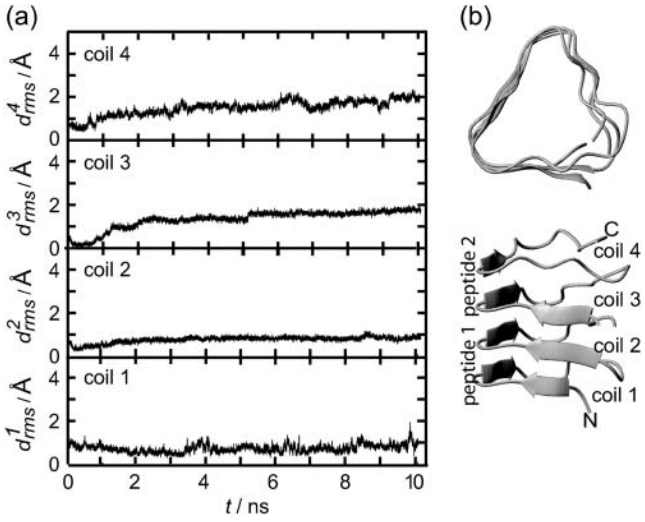

FIGURE 7 (*a*) Coil-specific root mean-square deviations for a dimer of two-coiled triangular poly-Gln β-helices. (*b*) Average structure (10 ns, *top* and *side views*).

and for the corresponding fragment of the native LpxA β-helix. The results of these tests are documented in the first figure of the Supplementary Material and demonstrate that the poly-Gln model becomes additionally stabilized by raising the temperature, which is a clear signature of hydrophobic stabilization. In contrast, the LpxA fragment appears to have lost some of its stability at these conditions.

In summary, three-coiled triangular poly-Gln β-helices are calculated to be at least as stable as fragments of a natively stable β-helix. Therefore, they should be sufficiently stable as to qualify as building blocks for amyloid fiber aggregation.

Our finding that the stability of β-helical fragments increases with the number of coils, which we derived from the comparison of two- and three-coiled LpxA fragments, now suggests that a two-coiled triangular poly-Gln β-helix might be unstable. As demonstrated by the second figure in the Supplementary Material, this is actually the case: The $d^j_{rms}(t)$ curves and 4-ns snapshot are similar to those of the two-coiled LpxA fragment depicted in Fig. 3, *a* and *b*).

Now the important question arises as to whether the aggregation of two individually unstable two-coiled structures can stabilize the resulting dimer. To clarify this issue Fig. 7 *a* shows the 10-ns $d^j_{rms}(t)$ curves for the four coils of the poly-Gln dimer. After 10 ns the average value $\langle d_{rms}\rangle(t)$ is 1.4 Å, which is of the same order of magnitude as the corresponding results for the three-coiled LxpA (1.6 Å) and poly-Gln (1.4 Å) models. Accordingly, the dimer is classified as stable. The 10-ns average structures depicted in Fig. 7 *b* verify this conclusion. Even the fast (2 ns) increase of $d^3_{rms}(t)$ to a value of 2 Å, exhibited in Fig. 7 *a*, does not change this result, because visual analysis of the structures reveals that the indicated change of coil 3 is confined to the N-terminus of peptide 2 and does not modify the overall shape of the aggregate. In summary, our data suggest that the dimerization of poly-Gln peptides, which comprise at least 36 residues, can create stable nuclei for seeded aggregation of triangular β-helices.

To find an explanation of why certain β-helices are stable and others preferentially decay, consider the space-filling models of three-coiled β-helices in Fig. 8 *a*. In the triangular LpxA, poly-Gln, and poly-Ile structures the centers of the helices are seen to be densely packed by the van der Waals spheres surrounding the atoms of the bulky nonpolar or weakly polar side groups. However, in some of the triangular structures (△), that is, for the poly-Ala and poly-Ser models, and in the two circular poly-Gln structures (○) with 18 and 20 residues suggested by Perutz et al. (2002a), the peptide atoms leave sufficient space to open sizable channels near the centers of the β-helices. Upon hydration these channels

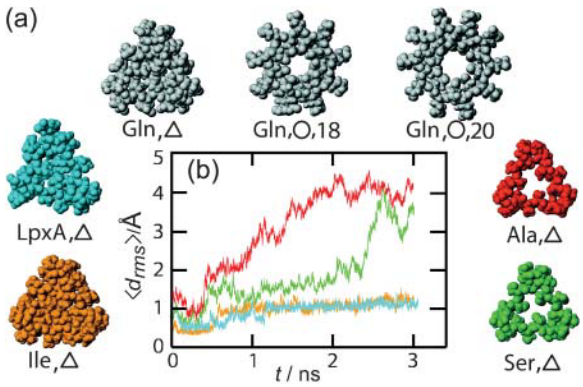

FIGURE 8 (*a*) Space-filling models of various three-coiled β-helices viewed along their axes. (*b*) Average root mean-square deviations for the models colored in *a* from 3-ns simulations; color coding as in *a*.

become filled with 5–6 (Gln/○,18; Ala/△; Ser/△), and even 10 (Gln/○,20) water molecules per helical coil.

The simulation results for some of these structures, presented above, now argue that β-helices are destabilized by water channels and are stabilized by hydrophobic interactions, if they have a sufficiently nonpolar and densely packed core. This argument implies that the triangular poly-Ser/△ structure depicted in Fig. 8 *a* should decay, whereas poly-Ile/△ should be stable.

Fig. 8 *b* compares the $\langle d_{\mathrm{rms}}\rangle(t)$ curves obtained by 3-ns simulations for poly-Ile/△ (*orange*) and poly-Ser/△ (*green*), with the reference curves of the stable LpxA (*blue*) and unstable poly-Ala (*red*) models, respectively. It shows that the predictions of the poly-Ser/△ decay and of the poly-Ile/△ stability are confirmed. Visual inspections of the final structures corroborate this result (data not shown). As a corollary we conclude that if poly-Gln β-helices exist, they are by no means water-filled nanotubes, in contrast to the expectation of Perutz et al. (2002a).

DISCUSSION, BIOMEDICAL IMPLICATIONS, AND A MODEL FOR PRP^{SC}

Originally, our MD-based stability study of parallel β-helices had been designed to check the suggestion of Perutz et al. (2002a) that the poly-Gln amyloid fibers involved in Huntington's disease are composed of circular β-helices. For the proposed 18 and 20 residues per coil, the circular structures turned out to be unstable. However, in the 20-residue case, after 10 ns of simulation, the structure remotely resembled a triangular β-helix.

This observation inspired us to examine left-handed triangular poly-Gln β-helices with 18 residues per coil as is common in native structures (Jenkins and Pickersgill, 2001). We found that a triangular three-coiled poly-Gln model is at least as stable as a native fragment of LpxA of the same size. A similar stability was determined for a dimeric aggregate of two-coiled poly-Gln β-helices, which, when considered individually, were found unstable.

Our results suggest that the dimerization of poly-Gln peptides with at least 36 residues can form stable nuclei for the aggregation of amyloid protofibrils shaped as triangular β-helices. However, in the case of poly-Gln peptides with $\sim$54 residues, a dimerization should not be required anymore. These conclusions agree well with the repeat length-dependence of disease risk (Ross et al., 2003) and of nucleation kinetics observed for poly-Gln peptides in vitro (Chen et al., 2001). Here, only Gln repeats with at least 37 residues were found to efficiently promote amyloid formation and Huntington's disease was attributed by Ross et al. (2003) to repeats exceeding 35–45 residues.

Our conclusions also agree with an in vitro mutational analysis of the structural organization of poly-Gln aggregates (Thakur and Wetzel, 2002). In this study the aggregation kinetics of various poly-Gln peptides with occasional Pro-Gly insertions has been investigated, because Pro-Gly inserts are known to be compatible with β-turn formation and incompatible with β-extended chain. Interestingly, the spontaneous aggregation kinetics of a K_2-$(Q_{10}$-PG$)_3$-Q_{10}-K_2 peptide was found to be as fast as that of a pure poly-Gln peptide (K_2-Q_{45}-K_2) (Thakur and Wetzel, 2002). The 12 residues of a (Q_{10}-PG) repeat exactly cover two thirds of a coil in a triangular left-handed parallel β-helix with the PG residues located at the second turn. Thus, this type of repeat unit is extremely compatible with the suggested secondary structure. Other repeat units, e.g., (Q_7-PG) or (Q_8-PG), in peptides of similar size were found to aggregate much less readily (Thakur and Wetzel, 2002) and are incompatible with the structure of a triangular parallel β-helix. A somewhat slower aggregation than for the pure poly-Gln peptide was found for (Q_9-PG) repeats (Thakur and Wetzel, 2002). When matching such sequences to a triangular parallel β-helix one would have to require that the corresponding helical coils contain short-cut turns rendering coils with $<$18 residues. Note here that such short-cut turns and coils with $<$18 residues also occur in native parallel β-helices (Jenkins and Pickersgill, 2001). Therefore, also, the result on the (Q_9-PG) aggregation is still compatible with the suggested secondary-structure motif. Thus, our results and the quoted in vitro data support the speculation that the core structure of the poly-Gln-induced amyloid fibers in Huntington's disease is that of a triangular β-helix with 18 residues per coil.

Based on our findings the question arises whether parallel β-helices may also feature in other diseases, such as prion diseases, that are linked to seeded aggregation and amyloid formation. Wille et al. (2002) suggested triangular β-helices as the decisive structural motif promoting PrP^{Sc} amyloid formation and provided a model by matching the hamster PrP sequence from about residue 90 to about residue 176 onto a regular triangular β-helix covering five layers arranged in a planar hexamer. However, all possible alignments of the hamster PrP sequence with this structure imply that some charged residues are located in the interior of the β-helix. In contrast, in native triangular β-helices charged residues exclusively point out of the helix or are found within loops, which generally protrude from the turns (Jenkins and Pickersgill, 2001). Moreover, no native hexamer of β-helices has been published, whereas several trimeric structures are known (Jenkins and Pickersgill, 2001).

Because of these objections we considered a trimer model (see Fig. 9 *a*), which fits the size constraints imposed by electron microscopy analysis (Wille et al., 2002) and can easily accommodate loops. Based on alternative alignments of the sequence by allowing loops to protrude from the outside corners of identical four-layered beta-helices in a trimeric arrangement, we selected two refined models that both avoid charged residues in the interior of the β-helix and steric interference at the trimer interface, as depicted in Fig. 9 *b*. Because our above findings suggest that Gln side chains in the interior of a triangular β-helix can stabilize such

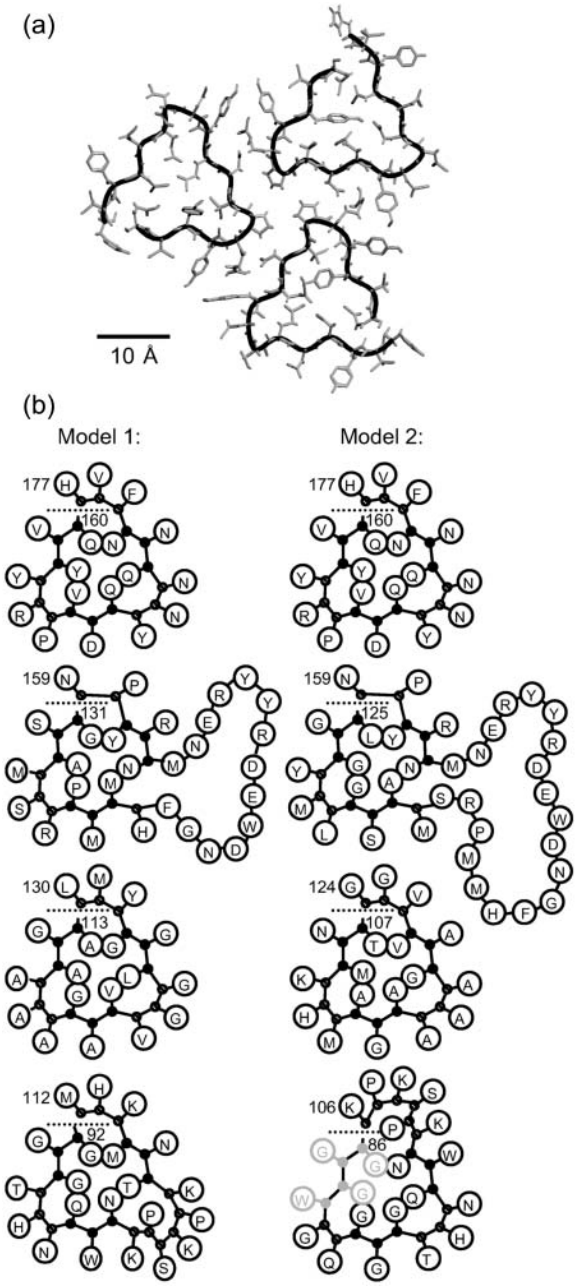

FIGURE 9 (*a*) Trimer model of PrP^{Sc} constructed from triangular β-helices; shown are residues 160–177. (*b*) Two triangular β-helix models for PrP^{Sc} with four coils. Model 1 comprises residues 92–177, and model 2 residues 86–177; residues 86–89 may not be part of the helix anymore. The orientation of the single coils in the trimeric arrangement corresponds to the upper right coil in *a*.

a secondary structure, we here consider Gln side chains as an adequate filling material, although Gln does not occur inside the native left-handed trimeric β-helices listed in Jenkins and Pickersgill (2001). Correspondingly, both models exhibit several Gln side chains in the core of the β-helix. The models comprise, apart from their otherwise regular β-helical structure depicted in Fig. 1, *e* and *f*, one small loop (residues 101–105), one bigger loop (residues 140–153 in model 1, residues 134–153 in model 2), and one tight corner (residues 158, 159). In model 1 both loops protrude from the same corner, as is found in all natively trimeric β-helical proteins listed in Jenkins and Pickersgill (2001). In contrast, model 2 exhibits loops at different corners and, therefore, represents an unusual structure. Besides, both models present a Tyr-Tyr-Arg motif in an outward loop, which has been implied as a PrP^{Sc}-specific epitope (Paramithiotis et al., 2003).

Note that very recently, Govaerts et al. (2004) published an alternative refined version of the model of Wille et al. (2002), also using a trimeric instead of a hexameric arrangement, but with a sequence alignment differing from our suggestions. As in our model 2, in the model of Govaerts et al. (2004) the loops protrude from different corners, and in one case, a loop starts even within a side of the β-helix. As a result the β-helical coil at the top layer is neither complete nor regular. Such defects could hamper fiber growth which is supposed to proceed through aggregation of an unfolded soluble monomer to a structural template provided by a complete β-helical coil of the existing oligomer. Note furthermore that a different trimeric model using stacked β-sheets, recently published by DeMarco and Daggett (2004), would fit into the electron microscopy images.

Testing models other than poly-Gln by MD simulations is hampered by the uncertainties of sequence alignment (including loop structures) and by the sheer sizes of the presumed aggregation nuclei. Whereas in the poly-Gln case these nuclei may actually be specifically folded monomers (Chen et al., 2002) larger complexes appear to be required with other polypeptides such as prion proteins (Eigen, 1996). Nevertheless, corresponding models may be valuable, because they entail predictions regarding the effect of amino acid exchanges on the stabilities of monomeric and oligomeric structures, which can be tested experimentally.

SUPPLEMENTARY MATERIAL

An online supplement to this article can be found by visiting BJ Online at http://www.biophysj.org

This project has been funded by the Bayerischer Forschungsverbund Prionen (project LMU02).

REFERENCES

Allen, M. P., and D. J. Tildesley. 1987. Computer Simulation of Liquids. Clarendon, Oxford, UK.

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

Berman, H. M., J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne. 2000. The protein data bank. *Nucleic Acids Res.* 28:235–242.

Blondelle, S. E., B. Forood, R. A. Houghten, and E. Pérez-Payá. 1997. Polyalanine-based peptides as models for self-associated β-pleated-sheet complexes. *Biochemistry*. 36:8393–8400.

Caughey, B., and P. T. Lansbury, Jr. 2003. Protofibrils, pores, fibrils, and neurodegeneration: Separating the responsible protein aggregates from the innocent bystanders. *Annu. Rev. Neurosci*. 26:267–298.

Chakrabartty, A., J. Schellman, and R. Baldwin. 1991. Large differences in the helix propensities of alanine and glycine. *Nature*. 351:586–588.

Chen, S., V. Berthelier, W. Yang, and R. Wetzel. 2001. Polyglutamine aggregation behavior *in vitro* supports a recruitment mechanism of cytotoxicity. *J. Mol. Biol*. 311:173–182.

Chen, S., F. A. Ferrone, and R. Wetzel. 2002. Huntington's disease age-of-onset linked to polyglutamine aggregation nucleation. *Proc. Natl. Acad. Sci. USA*. 99:11884–11889.

DeMarco, M. L., and V. Daggett. 2004. From conversion to aggregation: protofibril formation of the prion protein. *Proc. Natl. Acad. Sci. USA*. 101:2293–2298.

Eigen, M. 1996. Prionics or the kinetic basis of prion diseases. *Biophys. Chem.* 63:A1–A18.

Frenkel, D., and B. Smit. 2002. Understanding Molecular Simulation: From Algorithms to Applications. Academic Press, San Diego.

Govaerts, C., H. Wille, S. B. Prusiner, and F. E. Cohen. 2004. Evidence for assembly of prions with left-handed β-helices into trimers. *Proc. Natl. Acad. Sci. USA*. 101:8342–8347.

Jenkins, J., and R. Pickersgill. 2001. The architecture of parallel β-helices and related folds. *Prog. Biophys. Mol. Biol*. 77:111–175.

Jorgensen, W. L., J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein. 1983. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys*. 79:926–935.

Koradi, R., M. Billeter, and K. Wüthrich. 1996. MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graph*. 14: 51–55.

Kräutler, V., W. F. van Gunsteren, and P. H. Hünenberger. 2001. A fast SHAKE algorithm to solve distance constraint equations for small molecules in molecular dynamics simulations. *J. Comput. Chem*. 22: 501–508.

Lindahl, E., B. Hess, and D. van der Spoel. 2001. GROMACS 3.0: A package for molecular simulation and trajectory analysis. *J. Mol. Model*. 7:306–317.

Ma, B., and R. Nussinov. 2002a. Molecular dynamics simulations of alanine rich β-sheet oligomers: Insight into amyloid formation. *Protein Sci*. 11:2335–2350.

Ma, B., and R. Nussinov. 2002b. Stabilities and conformations of Alzheimer's β-amyloid peptide oligomers ($A\beta_{16-22}$, $A\beta_{16-35}$, $A\beta_{10-35}$): Sequence effects. *Proc. Natl. Acad. Sci. USA*. 99:14126–14131.

MacKerell, A. D., Jr., D. Bashford, M. Bellott, R. L. Dunbrack, Jr., J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher III, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin, and M. Karplus. 1998. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B*. 102:3586–3616.

Mathias, G., B. Egwolf, M. Nonella, and P. Tavan. 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*. 118:10847–10860.

Mathias, G., and P. Tavan. 2004. Angular resolution and range of dipole-dipole correlations in water. *J. Chem. Phys*. 120:4393–4403.

Niedermeier, C., and P. Tavan. 1994. A structure adapted multipole method or electrostatic interactions in protein dynamics. *J. Chem. Phys*. 101:734–748.

Paramithiotis, E., M. Pinard, T. Lawton, S. LaBoissiere, V. L. Leathers, W.-Q. Zou, L. A. Estey, J. Lamontagne, M. T. Lehto, L. H. Kondejewski, G. P. Francoeur, M. Papadopoulos, A. Haghighat, S. J. Spatz, M. Head, R. Will, J. Ironside, K. O'Rourke, Q. Tonelli, H. C. Ledebur, A. Chakrabartty, and N. R. Cashman. 2003. A prion protein epitope selective for the pathologically misfolded conformation. *Nat. Med*. 9: 893–899.

Perutz, M. F., J. T. Finch, J. Berriman, and A. Lesk. 2002a. Amyloid fibers are water-filled nanotubes. *Proc. Natl. Acad. Sci. USA*. 99:5591–5595.

Perutz, M. F., B. J. Pope, D. Owen, E. E. Wanker, and E. Scherzinger. 2002b. Aggregation of proteins with expanded glutamine and alanine repeats of the glutamine-rich and asparagine-rich domains of Sup35 and of the amyloid β-peptide of amyloid plaques. *Proc. Natl. Acad. Sci. USA*. 99:5596–5600.

Petkova, A. T., Y. Ishii, J. J. Balbach, O. N. Antzutkin, R. D. Leapman, F. Delaglio, and R. Tycko. 2002. A structural model for Alzheimer's β-amyloid fibrils based on experimental constraints from solid state NMR. *Proc. Natl. Acad. Sci. USA*. 99:16742–16747.

Ponder, J. W., and D. A. Case. 2003. Force fields for protein simulation. *Adv. Protein Chem.* 66:27–85.

Raetz, C. R. H., and S. L. Roderick. 1995. A left-handed parallel β helix in the structure of UDP-*N*-acetylglucosamine acyltransferase. *Science*. 270: 997–1000.

Richardson, J. S., and D. C. Richardson. 2002. Natural β-sheet proteins use negative design to avoid edge-to-edge aggregation. *Proc. Natl. Acad. Sci. USA*. 99:2754–2759.

Rochet, J.-C., and P. T. Lansbury, Jr. 2000. Amyloid fibrillogenesis: themes and variations. *Curr. Opin. Struct. Biol*. 10:60–68.

Ross, C. A., M. A. Poirier, E. E. Wanker, and M. Amzel. 2003. Polyglutamine fibrillogenesis: The pathway unfolds. *Proc. Natl. Acad. Sci. USA*. 100:1–3.

Schuler, B., R. Rachel, and R. Seckler. 1999. Formation of fibrous aggregates from a non-native intermediate: isolated P22 tailspike β-helix domain. *J. Biol. Chem.* 274:18589–18596.

Sunde, M., and C. C. F. Blake. 1998. From the globular to the fibrous state: protein structure and structural conversion in amyloid formation. *Q. Rev. Biophys*. 31:1–39.

Sunde, M., L. C. Serpell, M. Bartlam, P. E. Fraser, M. B. Pepys, and C. C. F. Blake. 1997. Common core structure of amyloid fibrils by synchrotron X-ray diffraction. *J. Mol. Biol*. 273:729–739.

Tavan, P., H. Carstens, and G. Mathias. 2005. Molecular dynamics simulations of proteins and peptides: problems, achievements, and perspectives. *In* Protein Folding Handbook, Part I. J. Buchner and T. Kiefhaber, editors. Wiley-VCH, Weinheim, Germany. 1170–1195.

Thakur, A., and R. Wetzel. 2002. Mutational analysis of the structural organization of polyglutamine aggregates. *Proc. Natl. Acad. Sci. USA*. 99:17014–17019.

van Gunsteren, W. F., and H. J. C. Berendsen. 1990. Computer simulation of molecular dynamics: methodology, applications, and perspectives in chemistry. *Angew. Chem. Int. Ed. Engl*. 29:992–1023.

Warshel, A., and S. T. Russell. 1984. Calculations of electrostatic interactions in biological systems and in solutions. *Q. Rev. Biophys*. 17: 283–422.

Wille, H., M. D. Michelitsch, V. Guénebaut, S. Supattapone, A. Serban, F. E. Cohen, D. A. Agard, and S. B. Prusiner. 2002. Structural studies of the scrapie prion protein by electron crystallography. *Proc. Natl. Acad. Sci. USA*. 99:3563–3568.

Supplemental Material

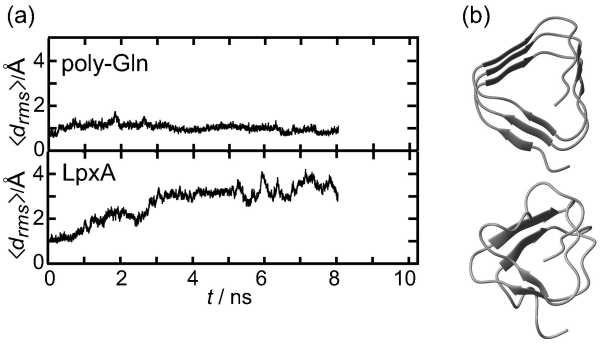

Figure 1: (a) Temporal evolution of the $\langle d_{rms} \rangle(t)$-values of the triangular poly-Gln β-helix and the three-coiled LpxA-fragment at an elevated temperature of 330 K. (b) Structures of these three-coiled β-helices after 8 ns.

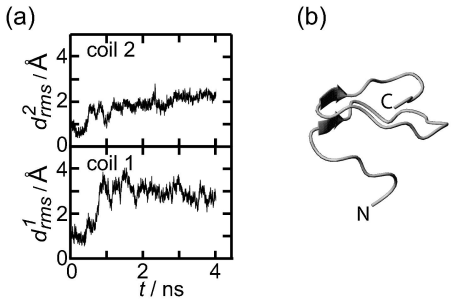

Figure 2: (a) $d^j_{rms}(t)$-curves for the various coils j of a two-coiled *triangular* poly-Gln β-helix. (b) Snapshot after 4 ns.

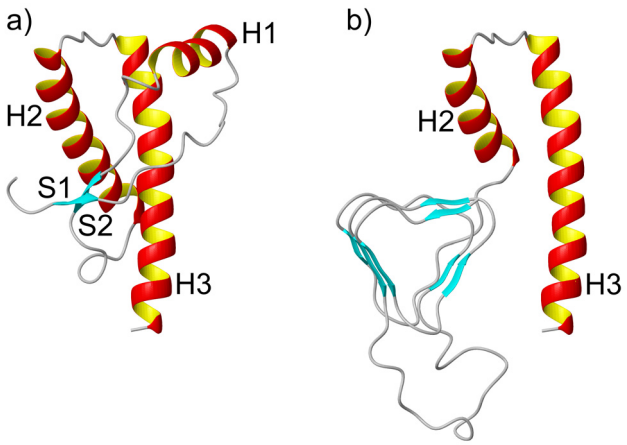

Abbildung 3.1: Dreidimensionale Strukturen des Prion-Proteins des Hamsters. a) NMR-Struktur des PrP^{C}, Residuen 125-228. b) Modell für PrP^{Sc}, Residuen 106-228. Die Koordinaten für a) wurden der Protein Data Bank [40] entnommen, Eintrag 1B10 [120]. Die Abbildungen wurden mit dem Grafikprogramm MOLMOL [41] erstellt.

3.2 Struktur einer PrP^{Sc}-Fibrille

Die im vorhergehenden Abschnitt entwickelten Modelle für die Grundstrukturen von Fibrillen aus poly-Gln und PrP^{Sc} besitzen einen Durchmesser von etwa 30 Å. Dieser Durchmesser ist kleiner als die in EM-Messungen ermittelten Durchmesser amyloider Fibrillen, die oft im Bereich von 100 Å liegen [50], wie etwa die in Abbildung 1.5 gezeigten Fibrillen aus Teilen des Aβ-Peptids [56]. Diese Grundstrukturen werden deshalb als *Protofibrillen* bezeichnet, die erst durch ihre Zusammensetzung eine Fibrille ergeben [58]. Im folgenden Abschnitt wird die Entwicklung von dreidimensionalen Modellen für Fibrillen aus PrP^{Sc} beschrieben und deren Übereinstimmung mit Daten aus experimentellen Verfahren der Strukturaufklärung (vgl. Abschnitt 1.4) demonstriert.

Das bestimmende Strukturmotiv der PrP^{Sc}-Modelle ist das einer triangulären parallelen β-Helix. Bei der Modellierung wurde die PrP-Sequenz des syrischen Hamsters verwendet, die sich nur wenig von der des Menschen (Abbildung 1.3) unterscheidet. Abbildung 3.1a zeigt die dreidimensionale Struktur des strukturierten, globulären Teils des Hamster-PrP^{C}, der aus den Residuen 125-228 besteht. Im Vergleich dazu zeigt Abbildung 3.1b eines unserer Modelle[1] für den strukturierten Teil des PrP^{Sc}, der einen etwas längeren Abschnitt der Sequenz, die Residuen 106-228, umfasst. Wie schon in Abschnitt 1.3 beschrieben wurde, besteht der strukturierte Teil

[1]Modell 2 aus Abschnitt 3.1, welches hier um die erste Windung der β-Helix gekürzt wurde. In analoger Weise ergeben sich dreidimensionale Modelle für das Modell 2 in seiner vollen Länge sowie für Modell 1.

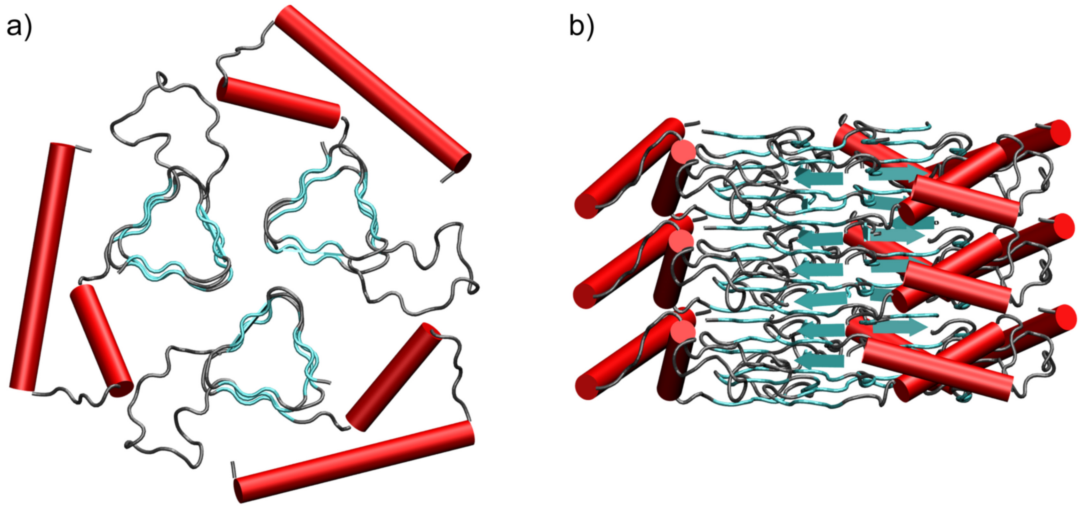

Abbildung 3.2: Modell einer PrP^{Sc}-Fibrille. a) Die Grundbausteine aus Abbildung 3.1b bilden ein Trimer. b) Durch Stapelbildung der parallelen β-Helizes können sich weitere Grundbausteine an beide Enden eines vorhandenen Aggregats anlagern. Die Abbildungen wurden mit dem Grafikprogramm VMD [124] erstellt.

des PrP^{C} aus drei α-Helizes und einem β-Faltblatt mit zwei Strängen. In unserem Modell für PrP^{Sc} ist Helix 3 unverändert geblieben, während Helix 2 in leicht gekürzter Form vorliegt. Helix 1 hat sich dagegen komplett umgefaltet und ist jetzt Bestandteil einer flexiblen Loopstruktur, die von einer Ecke einer parallelen β-Helix ausgeht. Das β-Faltblatt, sowie ein in PrP^{C} unstrukturierter Abschnitt (Res 106-124), der eine hydrophobe Domäne beinhaltet (vgl. Abbildung 1.3), sind jetzt Teile der parallelen β-Helix. Dieses Modell für PrP^{Sc} erfüllt viele Vorgaben, die sich aus Messungen an derartigen Fibrillen ergeben haben. Dazu zählt unter anderem, dass der Anteil an α-Helix zurückgeht, während der Anteil an β-Faltblatt zunimmt [42, 43] und dass die hydrophobe Domäne in PrP^{Sc} eine β-Faltblatt-Struktur annimmt [121].

Das dreidimensionale Modell für PrP^{Sc} entstand durch eine Anordnung verschiedener Sequenzbereiche (vgl. Abbildung 9 in [46]) auf die Strukturen nativer triangulärer β-Helizes [61, 122] mit dem Programm SwissPDB [123]. In analoger Weise wurden Koordinaten für die Loopstrukturen (Res 134-153 und Res 101-105) erzeugt. Anschließend wurden diese Einzelteile mit den verbleibenden α-Helizes aus PrP^{C} zu einer Struktur zusammengefügt. Mit Verfahren des *Structure Refinement* [63, 64], unter anderem durch Energieminimierung der β-Helix und ein *Simulated Annealing* der Loopstrukturen, wurden vorhandene Verspannungen im Modell beseitigt.

Inspiriert von den Dimensionsvorgaben elektronenmikroskopischer Aufnahmen an zweidimensionalen kristallinen Aggregaten aus PrP^{Sc}-Fragmenten [44], sowie von der Anordnung nativer triangulärer β-Helizes in Kristallstrukturen [61], wurde eine

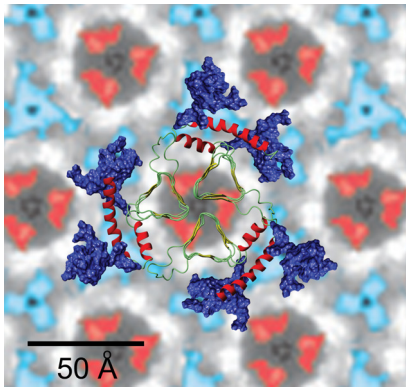

Abbildung 3.3: Modell einer PrP^{Sc}-Fibrille mit Zuckerseitenketten (dunkelblau), unterlegt mit einer elektronenmikroskopischen Aufnahme an zweidimensionalen Aggregaten aus PrP^{Sc}-Fragmenten [44]. Die Abbildung der PrP^{Sc}-Struktur wurde mit dem Grafikprogramm MOLMOL [41] erstellt.

Fibrille als Trimer dieser Grundstrukturen modelliert. Abbildung 3.2a zeigt eine Aufsicht auf die entstandene Trimerstruktur. Solch eine Fibrille kann in beide Richtungen durch Anlagerung weiterer Grundbausteine wachsen. Dabei stapeln sich die triangulären, β-helikalen Strukturen aufeinander und bilden intermolekulare Wasserstoffbrückenbindungen aus. Abbildung 3.2b zeigt eine Seitenansicht einer so entstandenen Fibrille, bestehend aus drei Trimeren. Auch in dieser Struktur wurden mit Verfahren des *Structure Refinement* Verspannungen gelöst.

Ist dieses Modell plausibel, so muss es auch genügend Platz für die an die Residuen Asn-181 und Asn-197 gebundenen Zuckerseitenketten (vgl. Abbildung 1.3) bieten. Analog zu [125] wurden Zuckerseitenketten mit dem Programm InsightII (Accelrys, San Diego, CA) modelliert und an die Bindungsstellen im PrP^{Sc}-Modell gesetzt. Abbildung 3.3 zeigt eine Aufsicht auf eine PrP^{Sc}-Fibrille, in der die Zuckerseitenketten als dunkelblaue Strukturen erscheinen. Wie zu sehen ist, zeigen die Zuckerseitenketten nach außen und fügen sich somit räumlich gut in das vorhandene Modell ein. Unterlegt ist die Struktur mit einem Differenzenbild [45] aus EM-Messungen an zweidimensionalen Aggregaten aus verschiedenen PrP^{Sc}-Fragmenten [44, 45]. Die hellblau eingefärbten Strukturen weisen auf die Anordnung der Zuckerseitenketten hin, die gut mit der Anordnung in unserem Modell übereinstimmt. Die rot eingefärbten Strukturen lassen auf die Lage negativ geladener Seitenketten im Bereich der Residuen 141-176 schließen, die sich in unserem Modell in der Loopstruktur befinden. Durch eine etwas andere Ausrichtung der flexiblen Loopstrukturen könnte auch diese Randbedingung erfüllt werden. Insgesamt passen die Dimensionen uneres Modells hervorragend zu den EM-Messungen.

Zwei weitere Modelle, die kürzlich von anderen Gruppen publiziert wurden [45, 126], wobei das erste große Ähnlichkeiten mit unserem Modell besitzt, sind ebenfalls mit den EM-Aufnahmen vereinbar. In weiteren Messungen an PrP^{Sc}-Fibrillen könnte überprüft werden, welches der Modelle zusätzliche Randbedingungen erfüllt.

Beispielsweise kann man mit Verfahren des *Crosslinking* [127] räumlich benachbarte Abschnitte in PrP^{Sc} kovalent verbinden. Durch eine anschließende Aufspaltung der Aggregate und mit Hilfe einer massenspektroskopischen Analyse der erhaltenen Fragmente kann man ermitteln, welche Fragmente durch einen Crosslink verbunden wurden und daraus folgern, dass diese Abschnitte in den PrP^{Sc}-Aggregaten räumlich benachbart sind [128]. Hiermit erhält man nicht nur Informationen über die räumliche Anordnung eines Grundbausteins (vgl. Abbildung 3.1), sondern auch über die Lage der Grundbausteine zueinander (vgl. Abbildung 3.2). Die PrP^{Sc}-Modelle könnten mit derartigen Messergebnissen verglichen und auf diese Weise weiter verfeinert werden. Ist ein Modell mit den räumlichen Vorgaben gänzlich unvereinbar, könnte es auf Grund dieser Daten verworfen werden.

4 Weiterentwicklung einer Kontinuumsmethode

In diesem Kapitel wird die in Abschnitt 1.5.2 beschriebene Kontinuumsmethode für polare Lösungsmittel von Egwolf und Tavan [98] für den Einsatz in MD-Simulationen weiterentwickelt. In Abschnitt 4.1 wird die Methode so umformuliert und erweitert, dass sie dem Prinzip *Actio=Reactio* gehorcht. In Abschnitt 4.2 wird die erweiterte Methode in ein MD-Programm implementiert, geeignet parametrisiert und an einem einfachen System getestet.

4.1 Umformulierung und Erweiterung der Kontinuumsmethode

Der folgende Abschnitt ist ein Abdruck[1] des Artikels

> Martina Stork and Paul Tavan:
> „Electrostatics of proteins in dielectric solvent continua:
> I. Newton's third law marries qE forces."
> *Journal of Chemical Physics* 126, 165105/1-15 (2007),

in dem ich die Kontinuumsmethode zusammen mit Paul Tavan umformuliert und erweitert habe. Neben den theoretischen Betrachtungen zu Reaktionsfeld-Kräften und -Energien wird die modifizierte Methode mit der analytischen Kirkwood-Lösung für ein Ion in einer sphärischen Kavität verglichen, sowie ein System aus zwei Ionen im Kontinuum betrachtet. Das Verhalten der Methode in Grenzfällen, wie z.B. im Limes für isolierte Atome, liefert wichtige Hinweise auf weitere Möglichkeiten ihrer Optimierung.

[1] Mit freundlicher Genehmigung des American Institute of Physics.

Electrostatics of proteins in dielectric solvent continua. I. Newton's third law marries *qE* forces

Martina Stork and Paul Tavan[a)]
Theoretische Biophysik, Lehrstuhl für BioMolekulare Optik, Ludwig-Maximilians-Universität München, Oettingenstrasse 67, D-80538 München, Germany

(Received 13 November 2006; accepted 2 March 2007; published online 27 April 2007)

The authors reformulate and revise an electrostatic theory treating proteins surrounded by dielectric solvent continua [B. Egwolf and P. Tavan, J. Chem. Phys. **118**, 2039 (2003)] to make the resulting reaction field (RF) forces compatible with Newton's third law. Such a compatibility is required for their use in molecular dynamics (MD) simulations, in which the proteins are modeled by all-atom molecular mechanics force fields. According to the original theory the RF forces, which are due to the electric field generated by the solvent polarization and act on the partial charges of a protein, i.e., the so-called *qE* forces, can be quite accurately computed from Gaussian RF dipoles localized at the protein atoms. Using a slightly different approximation scheme also the RF energies of given protein configurations are obtained. However, because the *qE* forces do not account for the dielectric boundary pressure exerted by the solvent continuum on the protein, they do not obey the principle that *actio* equals *reactio* as required by Newton's third law. Therefore, their use in MD simulations is severely hampered. An analysis of the original theory has led the authors now to a reformulation removing the main difficulties. By considering the RF energy, which represents the dominant electrostatic contribution to the free energy of solvation for a given protein configuration, they show that its negative configurational gradient yields mean RF forces obeying the *reactio* principle. Because the evaluation of these mean forces is computationally much more demanding than that of the *qE* forces, they derive a suggestion how the *qE* forces can be modified to obey Newton's third law. Various properties of the thus established theory, particularly issues of accuracy and of computational efficiency, are discussed. A sample application to a MD simulation of a peptide in solution is described in the following paper [M. Stork and P. Tavan, J. Chem. Phys., **126**, 165106 (2007). © *2007 American Institute of Physics*. [DOI: 10.1063/1.2720387]

I. INTRODUCTION

The structures and functional dynamics of soluble proteins and peptides are governed by electrostatic interactions among the dipolar peptide groups making up the backbone, the amino acid side groups of charged and polar residues occurring in the sequence, and the strong dipoles and ionic charges within the aqueous environment.[1] Therefore, attempts to quantitatively describe the dynamical processes of protein folding and function by molecular dynamics (MD) simulations based on all-atom molecular mechanics (MM) force fields must include a proper treatment of the electrostatics.[2–5]

Usually, in such MD simulations finite boxes are filled with MM models of the solvent and solute molecules, periodic boundary conditions are applied, and the long-range electrostatic interactions are evaluated by Ewald lattice summation methods.[6,7] Alternatively, the long-range electrostatics may be treated by combining fast hierarchical multipole expansions[8,9] with toroidal boundary conditions and a reaction field (RF) correction.[10] In both cases the number of solvent atoms has to exceed the number of protein atoms by at least a factor of 10, if one wants to avoid computational artifacts arising from the long-range nature of the electrostatics and from the finite size of the periodic simulation systems.[11,12] As a result one obtains simulation systems containing $N=10^4-10^5$ atoms, is forced to simulate an almost pure aqueous solution, which is only slightly polluted by the biological macromolecule of interest, and has to spend most of the computer time on the description of the thermal motion of the water molecules.

The enormous computational effort thus posed by protein-solvent simulations currently limits the time scales accessible to MD to a few tens or at most hundreds of nanoseconds. There are only rare cases, in which the conformational dynamics mediating the folding or function of peptides or proteins actually proceeds on such short time scales (for typical examples, see, e.g., Refs. 13–15). Generally, however, these processes proceed on time scales, which are by many orders of magnitude longer.

The apparent time scale gap has inspired many efforts, which try to get rid of the time-consuming explicit description of the solvent electrostatics by applying so-called implicit solvent models. Here, one tries to replace the microscopic modeling of the solvent by a mean field approximation of its interactions with the solute protein. Because these interactions are dominated by the electrostatic free energy of solvation, one has to find solutions to the

[a)]Author to whom correspondence should be addressed. Electronic mail: tavan@physik.uni-muenchen.de

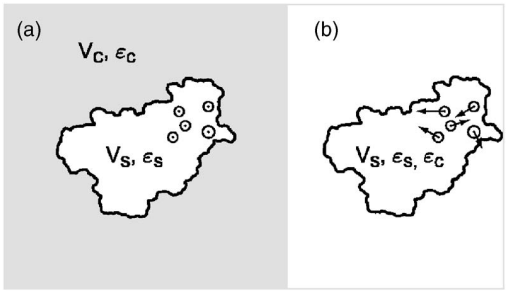

FIG. 1. (a) Concept of usual continuum models for proteins in polar solvents: A protein (white) is represented as an irregularly shaped cavity $\mathcal{V}_s$ of low dielectric constant ε_s and the polar solvent by a surrounding dielectric continuum $\mathcal{V}_c$ of high dielectric constant ε_c (gray). Protein atoms are indicated by circles and their (partial) charges by dots. (b) Concept of the continuum method by Egwolf and Tavan (Ref. [16]): The polarization of the dielectric environment, which is caused by the protein charges and generates an electric field exerting forces on these charges, is represented by so-called RF dipoles (arrows) localized at the protein atoms.

electrostatics problem of an irregularly shaped cavity $\mathcal{V}_s$ filled with point charges (the solute) and embedded in a dielectric and possibly ionic continuum $\mathcal{V}_c$ (the solvent). Figure 1(a) sketches this scenario for the purely dielectric case. When applied to MD simulations of soluble proteins, this scenario requires that one has to solve a partial differential equation at each MD time step. However, standard numerical methods to solve the corresponding Poisson or Poisson-Boltzmann equations are much too expensive to be used in MD simulations and have many other drawbacks (see Ref. 16 for a review and discussion).

To circumvent the problem of actually having to solve a partial differential equation at each MD time step, the so-called generalized Born (GB) methods[17] have been suggested (for a review, see Ref. 18). GB methods introduce local screening functions, which are supposed to describe the solvent-induced shielding of the electrostatic interactions within the solute and are empirically parametrized using sets of sample molecules. The GB approach enables the computation of impressive MD trajectories covering several hundreds of microseconds.[19] Unfortunately, however, the resulting free energy landscapes drastically differ from those obtained with explicit solvent.[20,21] Thus, the GB methods apparently oversimplify the complicated electrostatics problem for the sake of computational efficiency and, therefore, fail in structure prediction.

A quite accurate method to compute the free energy of solvation for a solute molecule and, in particular, the RF forces acting on the charges of its atoms has recently been developed by Egwolf and Tavan for the Poisson (Ref. 16) and the linearized Poisson-Boltzmann (Ref. 22) cases. Because the latter case is mathematically somewhat more complicated and because these complications could obscure the principles, we will restrict the following discussion to the Poisson case of a purely dielectric continuum. An extension to additionally include ion concentrations is then in principle straightforward (though tedious in detail). Correspondingly, Fig. 1(b) sketches the basic concept of the Egwolf-Tavan (ET) method toward the solution of the Poisson equation for proteins in polar solvents.[16] According to this analytical approach the electrostatic field $\mathbf{E}^{\varepsilon}(\mathbf{r})$ and potential $\Phi^{\varepsilon}(\mathbf{r})$, which are generated at positions $\mathbf{r}$ within the volume $\mathcal{V}_s$ of the solute protein by the polarization of the surrounding solvent continuum $\mathcal{V}_c$, i.e., the RF and its potential, are *exactly* represented by atom-centered dipole densities within the solute. The use of a Gaussian approximation for these dipole densities and for atomic shape functions collectively defining $\mathcal{V}_s$ leads to a set of coupled equations for associated atomic dipoles, the so-called RF dipoles, which can be solved numerically in an iterative fashion. The RF $\mathbf{E}^{\varepsilon}(\mathbf{r})$ within $\mathcal{V}_s$ and the solvation free energy W^{ε} of the solute molecule are then calculated from these dipoles.

The computational effort required by the ET method is comparable to that of introducing the electronic polarizability through inducible atomic dipoles into a MM force field. This effort is by several orders of magnitude smaller than that of standard numerical methods for the solution of the partial differential Poisson equation. The resulting RF forces $\mathbf{F}_i^{\varepsilon}=q_i\mathbf{E}^{\varepsilon}(\mathbf{r}_i)$ acting on atomic charges q_i at positions $\mathbf{r}_i \in \mathcal{V}_s$ agree very well with those calculated from analytical solutions for a spherical protein model and from explicit solvent simulations of a realistic protein model. Furthermore, as shown for a sample set of small molecules, the solvation free energies W^{ε} closely match those obtained by standard numerical methods.[16]

However, the progress sketched above represents only a first step towards MD simulations with implicit solvent, because the RF forces $\mathbf{F}_i^{\varepsilon}=q_i\mathbf{E}^{\varepsilon}(\mathbf{r}_i)$ solely cover the *actio* of the polarized continuum on the protein charges. As clearly explained in Ref. 23, such *qE* forces neglect the *reactio* forces, which are required by Newton's third law and give rise to the so-called dielectric boundary pressure.[24] Unfortunately, existing models for the inclusion of that pressure are neither sufficiently accurate nor computationally efficient.[23] Note that such boundary pressure models are superfluous in a recent method, which instead of numerically solving the Poisson equation treats a polarization density free energy functional within an extended Lagrangian dynamics scheme.[25] Because of the use of a free energy functional this method automatically accounts for Newton's third law. As a major drawback the resulting MD approach to the dynamics of peptides in solution cannot beat the explicit inclusion of large numbers of solvent molecules into MD simulations because of inferior computational efficiency.

If one chooses instead the approach through the Poisson equation and neglects the *actio=reactio* principle in a MD simulation of a protein embedded in a solvent continuum (implying that one simply adds the *actio qE* forces $\mathbf{F}_i^{\varepsilon}$ to the intramolecular forces acting on the atoms), one obtains a large total force and torque on the protein as well as a deposition of heat into the system, which have to be corrected at each MD integration step. However, corresponding methods for translation and rotation correction as well as for temperature control introduce large additional forces and, therefore, will cause severe dynamical artifacts.

In this paper we address the question, to what extent such artifacts can be avoided in the framework of the continuum theory outlined above, i.e., whether the ET theory can be adapted to MD simulations with implicit solvent with-

out destroying its merits by the introduction of large dynamical artifacts. For this purpose we first summarize and partially reformulate in Sec. II the key concepts and mathematics of the ET theory. An analysis of the expression for the solvation free energy W^{ε} in Secs. II B and II C will then show that the theory covers the desired *reactio* forces representing the dielectric boundary pressure at least "in principle," although this insight does not immediately lead to a computationally efficient algorithm. It will, however, provide clues on how the *actio qE* forces can be corrected by *reactio* forces, which eliminate the total force. As discussed in Sec. III these forces can be efficiently calculated in a linearly scaling fashion by integrating the RF dipoles into the structure adapted multipole method,[8,9] previously developed for the computationally efficient treatment of the long-range electrostatics in conventional MM-MD simulations.[10] A discussion of further aspects of the theory, of possible extensions, and of remaining tasks will conclude the paper.

II. THEORY

A. Exact formulation of reaction field and potential

As mentioned in the Introduction, Fig. 1(a) sketches the scenario of conventional continuum approaches towards proteins in polar solvents. Here, the solute protein is represented as an arbitrarily shaped cavity $\mathcal{V}_s$ of low dielectric constant ε_s and the polar solvent by an infinite surrounding continuum $\mathcal{V}_c$ of high dielectric constant ε_c. Adopting the concepts commonly used in MM-MD models of protein-solvent systems,[26,27] the electrostatic properties of the protein atoms are described by partial point charges and ε_s has the vacuum value of 1. Correspondingly, in continuum approaches one employs for ε_c the dielectric constant of the solvent at the given thermodynamic conditions (e.g., $\varepsilon_c=78$ for pure water at room temperature and ambient pressure). Because the charges of the protein atoms polarize the solvent continuum $\mathcal{V}_c$, this polarization generates an electrostatic reaction field $\mathbf{E}^{\varepsilon}$ and potential Φ^{ε} at positions $\mathbf{r}$ within the volume $\mathcal{V}_s$ occupied by the solute protein. Continuum theories aim at the computation of these functions.

The ET theory starts with an exact reformulation of the continuum electrostatics of proteins in polar solvents.[16] To enable a discussion of that reformulation we will now briefly summarize its key ingredients and the resulting equations.

1. Atomization of the protein volume

Any continuum theory requires a partitioning of the whole space into the mutually exclusive volumes $\mathcal{V}_s$ and $\mathcal{V}_c$ covering the protein and the surrounding dielectric continuum, respectively. If one defines $\mathcal{V}_s$ by the characteristic function

$$\Theta(\mathbf{r}) \equiv \left\{ \begin{array}{ll} 1 & \text{if } \mathbf{r} \in \mathcal{V}_s \\ 0 & \text{if } \mathbf{r} \in \mathcal{V}_c, \end{array} \right\} \tag{1}$$

one obtains the dielectric function

$$\varepsilon(\mathbf{r}) \equiv \varepsilon_c - (\varepsilon_c - \varepsilon_s)\Theta(\mathbf{r}), \tag{2}$$

which assumes the values ε_s and ε_c in the corresponding parts of the system.

The protein volume $\mathcal{V}_s$ can now be *exactly* decomposed into partially overlapping atomic volumes ϑ_i, if one models the shapes of the protein atoms i by normalized Gaussians

$$G(|\mathbf{r}-\mathbf{r}_i|;\sigma_i) \equiv (2\pi\sigma_i^2)^{-3/2} \exp\left[-\frac{(\mathbf{r}-\mathbf{r}_i)^2}{2\sigma_i^2} \right] \tag{3}$$

of width σ_i and centered around the positions $\mathbf{r}_i$ of the atoms. Multiplying the normalized atomic shape functions (3) with effective atomic volumes v_i enables the definition of partition functions

$$a_i(\mathbf{r}) \equiv v_i G(|\mathbf{r}-\mathbf{r}_i|;\sigma_i) \left[\sum_j v_j G(|\mathbf{r}-\mathbf{r}_j|;\sigma_j) \right]^{-1}, \tag{4}$$

which provide an exact fuzzy partition[28] of $\mathcal{R}^3$ with the property

$$\sum_i a_i(\mathbf{r}) = 1 \quad \text{for all } \mathbf{r} \in \mathcal{R}^3. \tag{5}$$

Note here that the partition functions a_i depend on $\mathbf{r}$ solely through the distances $|\mathbf{r}-\mathbf{r}_j|$ from the atoms at $\mathbf{r}_j$. Restricting the partition defined by Eqs. (4) and (5) to the protein volume $\mathcal{V}_s$ by

$$\vartheta_i(\mathbf{r}) \equiv a_i(\mathbf{r})\Theta(\mathbf{r}) \tag{6}$$

then provides the announced fuzzy decomposition of the characteristic function Θ into atomic contributions ϑ_i, because

$$\sum_i \vartheta_i(\mathbf{r}) = \Theta(\mathbf{r}). \tag{7}$$

By the partition (7) the space $\mathcal{V}_s$ occupied by the solute protein is thus exactly decomposed into partially overlapping atomic volumes ϑ_i. This *atomization* provides an irregular but natural fuzzy discretization of $\mathcal{V}_s$ suitable to calculate the electrostatic field and potential, and is the first key ingredient to the ET theory.

2. Continuum electrostatics of a solute protein

For a protein-solvent system characterized by the dielectric function (2) and by a charge distribution ρ associated with the atoms in $\mathcal{V}_s$, the electric field $\mathbf{E}$ obeys Gauss's law, which in Gaussian units[29] is given by

$$\nabla_{\mathbf{r}} \cdot [\varepsilon(\mathbf{r})\mathbf{E}(\mathbf{r})] = 4\pi\rho(\mathbf{r}). \tag{8}$$

Here, the field has to obey the condition

$$\nabla_{\mathbf{r}} \times \mathbf{E}(\mathbf{r}) = \mathbf{0}, \tag{9}$$

which is fulfilled, if $\mathbf{E}$ derives from an electrostatic potential Φ by

$$\mathbf{E}(\mathbf{r}) = -\nabla_{\mathbf{r}}\Phi(\mathbf{r}). \tag{10}$$

Insertion into Gauss's law (8) yields the Poisson equation

$$\nabla_{\mathbf{r}} \cdot [\varepsilon(\mathbf{r})\nabla_{\mathbf{r}}\Phi(\mathbf{r})] = -4\pi\rho(\mathbf{r}), \tag{11}$$

which is the partial differential equation to be solved.

3. Alternative formulations

According to the ET approach the Poisson equation (11), which identifies the charge distribution ρ within the protein as the only source of the electrostatic potential, can be reformulated in two equivalent ways, which both reveal certain dipole densities within $\mathcal{V}_s$ as additional sources of the potential. Having defined the constant

$$c_\varepsilon \equiv \frac{\varepsilon_c - \varepsilon_s}{4\pi}, \tag{12}$$

one can, for instance, introduce a first dipole density

$$\mathbf{P}^{\mathbf{E}}(\mathbf{r}) \equiv c_\varepsilon \Theta(\mathbf{r}) \boldsymbol{\nabla}_{\mathbf{r}} \Phi(\mathbf{r}), \tag{13}$$

which is confined by the step function Θ to the protein volume $\mathcal{V}_s$. Using the expression (2) for the dielectric function $\varepsilon(\mathbf{r})$ then enables to rewrite Eq. (11) in the form[16]

$$\Delta_{\mathbf{r}} \Phi(\mathbf{r}) = -\frac{4\pi}{\varepsilon_c}[\rho(\mathbf{r}) - \boldsymbol{\nabla}_{\mathbf{r}} \cdot \mathbf{P}^{\mathbf{E}}(\mathbf{r})]. \tag{14}$$

This is the Poisson equation of a system with the dielectric constant ε_c, whose electrostatic potential Φ is generated by the charge density ρ and the dipole density $\mathbf{P}^{\mathbf{E}}$. The latter density represents the polarization of the surrounding solvent continuum by a polarization within the protein volume $\mathcal{V}_s$. In the ET theory, Eq. (14) is the starting point for the calculation of the electric field.

For the calculation of the potential a different route was chosen, which is based on the fact that the Poisson equation (11) equivalently may be rearranged to read

$$\Delta_{\mathbf{r}}[\varepsilon(\mathbf{r})\Phi(\mathbf{r})] = -4\pi[\rho(\mathbf{r}) - \boldsymbol{\nabla}_{\mathbf{r}} \cdot \mathbf{P}^{\Phi}(\mathbf{r})], \tag{15}$$

with a second dipole density $\mathbf{P}^{\Phi}(\mathbf{r})$, which in contrast to the volume density (13) is a surface density. It is defined by

$$\mathbf{P}^{\Phi}(\mathbf{r}) \equiv -c_\varepsilon[\boldsymbol{\nabla}_{\mathbf{r}} \Theta(\mathbf{r})]\Phi(\mathbf{r}). \tag{16}$$

Whereas the first alternative formulation (14) of the Poisson equation provides a good starting point for accurate numerical computations of the field, the second formulation (15) is better suited for computations of the potential.[16]

4. Atomization of the densities

Taking advantage of the fuzzy partition defined by Eqs. (6) and (7), in the ET approach the two dipole densities introduced above are decomposed by

$$\mathbf{P}^{\mathbf{E}}(\mathbf{r}) = \sum_i \mathbf{P}_i^{\mathbf{E}}(\mathbf{r}) \quad \text{and} \quad \mathbf{P}^{\Phi}(\mathbf{r}) = \sum_i \mathbf{P}_i^{\Phi}(\mathbf{r}) \tag{17}$$

into partial atomic densities

$$\mathbf{P}_i^{\mathbf{E}}(\mathbf{r}) \equiv c_\varepsilon \vartheta_i(\mathbf{r}) \boldsymbol{\nabla}_{\mathbf{r}} \Phi(\mathbf{r}) \tag{18}$$

and

$$\mathbf{P}_i^{\Phi}(\mathbf{r}) \equiv -c_\varepsilon[\boldsymbol{\nabla}_{\mathbf{r}} \vartheta_i(\mathbf{r})]\Phi(\mathbf{r}), \tag{19}$$

respectively. Like the partition functions ϑ_i also the atomic dipole densities $\mathbf{P}_i^{\Phi}$ and $\mathbf{P}_i^{\mathbf{E}}$ are centered around the positions $\mathbf{r}_i$ of the associated atoms. Because of Eqs. (4)–(6) their values at a point $\mathbf{r}$ depend on the distance vectors $\mathbf{r} - \mathbf{r}_i$.

According to Eqs. (10), (14), and (15), the RF contributions to the total electric field and potential are generated by these atomic dipole densities. In Fig. 1(b), which sketches this modified scenario, they are jointly represented by arrows. Note that in this drawing the surrounding dielectric continuum $\mathcal{V}_c$, which was still indicated in Fig. 1(a), has been omitted. Instead a mapping of the polarization originally located in $\mathcal{V}_c$ now appears within the protein volume $\mathcal{V}_s$ in the form of the atomic dipole densities marked by the arrows.

In principle, also the charge density ρ within $\mathcal{V}_s$ could be discretized by the fuzzy partitioning applied above. However, Egwolf and Tavan[16] preferred to choose a different approach by adopting the concepts commonly used in MM force fields for proteins. Here, ρ is discretized according to

$$\rho(\mathbf{r}) = \sum_i q_i \delta(\mathbf{r} - \mathbf{r}_i) \tag{20}$$

by point charges q_i localized at the centers $\mathbf{r}_i$ of the atoms.

5. The electric field and potential

By the above atomizations of the charge and dipole densities within $\mathcal{V}_s$ the electric field and potential become sums over atomic contributions [cf. Eqs. (31) and (34) in Ref. 16]. As one can easily show the resulting expressions can be rewritten such that the total electric field

$$\mathbf{E}(\mathbf{r}) \equiv \mathbf{E}^C(\mathbf{r}) + \mathbf{E}^{\varepsilon}(\mathbf{r}) \tag{21}$$

becomes a superposition of the Coulomb field $\mathbf{E}^C$ and the RF field $\mathbf{E}^{\varepsilon}$. The former field

$$\mathbf{E}^C(\mathbf{r}) \equiv \frac{1}{\varepsilon(\mathbf{r})} \sum_i [-\boldsymbol{\nabla}_{\mathbf{r}} \phi^{q_i}(\mathbf{r})] \tag{22}$$

is obtained in the usual way from the unshielded Coulomb potentials

$$\phi^{q_i}(\mathbf{r}) \equiv \frac{q_i}{|\mathbf{r} - \mathbf{r}_i|} \tag{23}$$

of the partial charges q_i at the positions $\mathbf{r}_i$. The RF

$$\mathbf{E}^{\varepsilon}(\mathbf{r}) \equiv \frac{1}{\varepsilon(\mathbf{r})} \sum_i \{-\boldsymbol{\nabla}_{\mathbf{r}} \phi^{\mathbf{P}_i^{\mathbf{E}}}(\mathbf{r}) + 4\pi \mathbf{P}_i^{\mathbf{E}}(\mathbf{r})\} \tag{24}$$

derives from the atomic dipole densities $\mathbf{P}_i^{\mathbf{E}}$ on the one hand directly through the associated polarizations $4\pi\mathbf{P}_i^{\mathbf{E}}$ of the atomic partial volumes ϑ_i and on the other hand indirectly through the unshielded electrostatic potentials

$$\phi^{\mathbf{P}_i^{\mathbf{E}}}(\mathbf{r}) \equiv -\boldsymbol{\nabla}_{\mathbf{r}} \cdot \int_{\mathbf{r}' \in \mathcal{R}^3} \frac{\mathbf{P}_i^{\mathbf{E}}(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} dV' \tag{25}$$

generated by the $\mathbf{P}_i^{\mathbf{E}}$. Whereas in the case of the Coulomb potentials (23) the functional dependence on the distance vectors $\mathbf{r} - \mathbf{r}_i$ between the point $\mathbf{r}$ and the locations $\mathbf{r}_i$ of the atoms is obvious, and the corresponding functional dependence of the potentials (25) becomes clear only after reconsideration of our remarks following Eq. (19). The dipole densities $\mathbf{P}_i^{\mathbf{E}}$ are the solutions of the coupled system of integrodifferential equations

$$\nabla_{\mathbf{r}} \times \mathbf{P}_i^{\mathbf{E}}(\mathbf{r}) = \frac{c_\varepsilon}{\varepsilon_c} \nabla_{\mathbf{r}} \times \vartheta_i(\mathbf{r}) \sum_j \nabla_{\mathbf{r}} [\phi^{q_j}(\mathbf{r}) + \phi^{\mathbf{P}_j^{\mathbf{E}}}(\mathbf{r})], \quad (26)$$

which follows from the requirement (9) that the curl of the electric field should vanish everywhere.[16]

Like the electric field (21) also the potential is a superposition

$$\Phi(\mathbf{r}) \equiv \Phi^C(\mathbf{r}) + \Phi^\varepsilon(\mathbf{r}) \quad (27)$$

of Coulombic and RF contributions. With the atomic point charges (20) the Coulomb part is given by the sum

$$\Phi^C(\mathbf{r}) = \frac{1}{\varepsilon(\mathbf{r})} \sum_i \phi^{q_i}(\mathbf{r}) \quad (28)$$

over the unshielded Coulomb potentials (23). Due to the atomization (17) the RF part becomes the sum

$$\Phi^\varepsilon(\mathbf{r}) = \frac{1}{\varepsilon(\mathbf{r})} \sum_i \phi^{\mathbf{P}_i^\Phi}(\mathbf{r}) \quad (29)$$

over the unshielded potentials

$$\phi^{\mathbf{P}_i^\Phi}(\mathbf{r}) = -\nabla_{\mathbf{r}} \cdot \int_{\mathbf{r}' \in \mathcal{R}^3} \frac{\mathbf{P}_i^\Phi(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} dV' \quad (30)$$

generated by the second type of atomic dipole densities [cf. Eq. (19)]. These densities follow directly from the Coulomb potentials (23) and from the solutions $\mathbf{P}_j^{\mathbf{E}}$ of the equation system (26) by

$$\mathbf{P}_i^\Phi(\mathbf{r}) = -\frac{c_\varepsilon}{\varepsilon_c} [\nabla_{\mathbf{r}} \vartheta_i(\mathbf{r})] \sum_j [\phi^{q_j}(\mathbf{r}) + \phi^{\mathbf{P}_j^{\mathbf{E}}}(\mathbf{r})]. \quad (31)$$

This relation guarantees that the electric field (21) is the negative gradient of the potential (27), i.e., that the total field and potential obey Eq. (10). Within the volume $\mathcal{V}_s$ occupied by the protein, where the dielectric function $\varepsilon(\mathbf{r})$ assumes the constant value ε_s, Eq. (10) obviously holds for the Coulomb contributions $\mathbf{E}^C$ and Φ^C to the total field and potential, respectively. Therefore, within $\mathcal{V}_s$ also the RF individually obeys $\mathbf{E}^\varepsilon(\mathbf{r}) = -\nabla_{\mathbf{r}} \Phi^\varepsilon(\mathbf{r})$.

B. Discussion of the exact theory

If one somehow manages to solve the integrodifferential equations (26) for the dipole densities $\mathbf{P}_i^{\mathbf{E}}$, then the *qE* force exerted by the RF (24) on a protein atom j with point charge q_j is

$$\mathbf{F}_j^\varepsilon = q_j \mathbf{E}^\varepsilon(\mathbf{r}_j). \quad (32)$$

In fact, Egwolf and Tavan[16] suggested to calculate the RF forces in MD simulations from this equation apparently being unaware of the difficulties, which are caused by the associated neglect of Newton's third law (cf. the Introduction). On the other hand, because their theory is exact, it also provides clues for avoiding these difficulties.

To gain access to these clues we consider the electrostatic energy W of the charge distribution ρ in $\mathcal{V}_s$, which in *linear* media (like the solvent continuum assumed by us), is given by[29]

$$W = \frac{1}{2} \int_{\mathbf{r} \in \mathcal{R}^3} \rho(\mathbf{r}) \Phi(\mathbf{r}) dV. \quad (33)$$

As immediately follows from Eq. (27), the energy is made up of Coulomb and RF contributions W^C and W^ε, respectively. Inserting the atomization (20) of the charge distribution and the expression (28) for Φ^C into Eq. (33) yields by evaluation of the integral for the Coulomb contribution the usual expression

$$W^C = \frac{1}{2\varepsilon_s} \sum_{k, i \neq k} q_k \phi^{q_i}(r_{ki}), \quad (34)$$

with $r_{ki} \equiv |\mathbf{r}_k - \mathbf{r}_i|$. Here, we have used the fact that the Coulomb potentials (23) depend on $\mathbf{r}$ only through the distances $|\mathbf{r} - \mathbf{r}_i|$. Similarly one finds for the RF part

$$W^\varepsilon = \frac{1}{2\varepsilon_s} \left[\sum_{k, i \neq k} q_k \phi^{\mathbf{P}_i^\Phi}(\mathbf{r}_k) \right] \left[+ \sum_k q_k \int_{\mathbf{r}' \in \mathcal{R}^3} \frac{\mathbf{P}_k^\Phi(\mathbf{r}')(\mathbf{r}_k - \mathbf{r}')}{|\mathbf{r}_k - \mathbf{r}'|^3} dV' \right], \quad (35)$$

showing that the potentials $\phi^{\mathbf{P}_i^\Phi}$ generated by the dipole densities $\mathbf{P}_i^\Phi$ at other atoms i and the polarization of the atom k itself contribute to the RF energy of each charge q_k.

The RF energy W^ε covers the main contribution to the free energy of solvation for a molecule in a given atomic configuration $(\mathbf{r}_1, \ldots, \mathbf{r}_N)$.[16] Therefore, with respect to *changes* of the atomic configuration the RF energy $W^\varepsilon \equiv W^\varepsilon(\mathbf{r}_1, \ldots, \mathbf{r}_N)$ represents a *potential of mean force* and the *mean RF forces* exerted by the solvent on the atoms j of a solute protein are given by

$$\mathbf{F}_j^{W^\epsilon} = -\nabla_{\mathbf{r}_j} W^\varepsilon(\mathbf{r}_1, \ldots, \mathbf{r}_N). \quad (36)$$

Now suppose that a given atom j in a protein is neutral ($q_j = 0$). Then Eqs. (36) and (35) predict that a generally nonvanishing mean RF force $\mathbf{F}_j^{W^\epsilon}$ acts on that atom, because also neutral atoms carry dipole densities $\mathbf{P}_j^\Phi$ and because W^ε depends on the coordinate $\mathbf{r}_j$ of that atom. This "mean RF force" is quite obviously ignored by Eq. (32) representing the *qE* force and is one of the missing *reactio* forces, which are supposed to account for the dielectric boundary pressure.

The problem of integrating continuum electrostatics into MD simulations can thus be formulated as follows: A computational scheme has to be constructed by which one can rapidly and accurately obtain the mean RF forces (36). As its basic ingredient a computationally efficient, approximate solution $\tilde{\mathbf{P}}_i^{\mathbf{E}}(\mathbf{r})$ of the basic equation system (26) is needed, because its solutions, the dipole densities $\mathbf{P}_i^{\mathbf{E}}$, determine through Eq. (24) the field, through Eq. (31) the densities $\mathbf{P}_i^\Phi$, through the latter densities by Eq. (29) the potential (27), and, thus, through Eq. (33) the free energy of solvation W^ε as given by Eq. (35).

C. Approximate theory reformulated

In addition to the exact theory discussed so far Egwolf and Tavan[16] have also suggested a set of approximations

capable of yielding the desired solution for Eq. (26). We will now sketch the key concepts of these approximations and concurrently reformulate the mathematics in such a way that the problems posed by the original approach can be more easily tackled.

1. Protein volume

The first approximation pertains to the definition (1) of the cavity $\mathcal{V}_s$ filled by the protein. Here, the step function Θ and its exact decomposition (7) were replaced by the superposition

$$\widetilde{\Theta}(\mathbf{r}) \equiv \sum_i \widetilde{\vartheta}_i(\mathbf{r}) \tag{37}$$

of approximate atomic partition functions $\widetilde{\vartheta}_i$, which are given by the Gaussian shape functions (3) weighted by the effective volumes v_i, i.e.,

$$\widetilde{\vartheta}_i(\mathbf{r}) \equiv v_i G(|\mathbf{r}-\mathbf{r}_i|;\sigma_i). \tag{38}$$

Due to the normalization of the Gaussians, the total volume V_s of the protein is simply the sum of the effective atomic volumes v_i. Imposing the condition that $\Theta(\mathbf{r}_i)=\widetilde{\Theta}(\mathbf{r}_i)$ at all positions $\mathbf{r}_i$ of the protein atoms yields conditions for the effective volumes v_i, which can be met by iterating the equation system

$$v_i^{(n)} = v_i^{(n-1)}\left[\sum_j v_j^{(n-1)} G(r_{ij};\sigma_j)\right]^{-1}. \tag{39}$$

As a result one obtains an approximate characteristic function $\widetilde{\Theta}(\mathbf{r})$, which exhibits "bumps" at the atomic sites and "dips" in between. As opposed to the exact step function $\Theta(\mathbf{r})$, the approximation $\widetilde{\Theta}(\mathbf{r})$ smoothly decays to zero on an atomic length scale at the boundary between $\mathcal{V}_s$ and $\mathcal{V}_c$.

2. Gaussian dipole densities

The second approximation pertains to the shapes of the atomic dipole densities $\mathbf{P}_i^{\mathbf{E}}$ defined by Eq. (18). Using once again the Gaussian atomic shape functions (3) they were modeled by

$$\widetilde{\mathbf{P}}_i^{\mathbf{E}}(\mathbf{r}) \equiv \mathbf{p}_i G(|\mathbf{r}-\mathbf{r}_i|;\sigma_i), \tag{40}$$

with the point dipoles $\mathbf{p}_i$. These dipoles are the spatial integrals over the model densities $\widetilde{\mathbf{P}}_i^{\mathbf{E}}$ and will be called "RF dipoles" from now on.

3. RF dipole equations reformulated

The spatial integration of the key equation system (26) for the dipole densities $\mathbf{P}_i^{\mathbf{E}}$ yields then[16] [with the approximations (38) and (40) as well with a Taylor expansion of the integral in Eq. (25) around the center $\mathbf{r}_i$ of the model density $\widetilde{\mathbf{P}}_i^{\mathbf{E}} \approx \mathbf{P}_i^{\mathbf{E}}$] a linear equation system [Eq. (52) in Ref. 16] for the RF dipoles $\mathbf{p}_i$, which can be solved by iteration. As one can show by simple rearrangements, this equation system may be equivalently written as

$$\mathbf{p}_i = \frac{\alpha_i}{\varepsilon_s}\nabla_{\mathbf{r}_i}\sum_{j\neq i}\left[\phi_G^{q_j}(\mathbf{r}_i;\sigma_i) + \phi_G^{\mathbf{p}_j}(\mathbf{r}_i;\sigma_{ij})\right]. \tag{41}$$

Here, the quantity

$$\phi_G^{q_j}(\mathbf{r}_i;\sigma_i) \equiv \frac{q_j}{r_{ij}}\,\mathrm{erf}\left(\frac{r_{ij}}{\sqrt{2}\sigma_i}\right) \tag{42}$$

is the unshielded potential generated by the point charge q_j and averaged over the normal distribution at atom i. It is given in terms of the error function $\mathrm{erf}(x)$. Here, like in the following, we characterize potentials or fields obtained from Gaussian distributions by the subscript G. Thus, the corresponding and properly shielded field is

$$\mathbf{E}_G^{q_j}(\mathbf{r}_i;\sigma_i) \equiv -\frac{1}{\varepsilon_s}\nabla_{\mathbf{r}_i}\phi_G^{q_j}(\mathbf{r}_i;\sigma_i), \tag{43}$$

for which an explicit expression is given in Appendix A. Furthermore, the quantity

$$\phi_G^{\mathbf{p}_j}(\mathbf{r}_i;\sigma_{ij}) = -\nabla_{\mathbf{r}_i}\cdot\left[\mathbf{p}_j\frac{1}{r_{ij}}\,\mathrm{erf}\left(\frac{r_{ij}}{\sqrt{2}\sigma_{ij}}\right)\right] \tag{44}$$

is the unshielded potential generated by the dipole density $\widetilde{\mathbf{P}}_j^{\mathbf{E}}$, $i\neq j$, and averaged over the normal distribution at atom i. Due to the underlying integration over both densities, the spatial scale σ_{ij} appearing in the error function contains contributions from both Gaussian widths, i.e., $\sigma_{ij}^2=\sigma_i^2+\sigma_j^2$. Also here we define the corresponding field by

$$\mathbf{E}_G^{\mathbf{p}_j}(\mathbf{r}_i;\sigma_{ij}) \equiv -\frac{1}{\varepsilon_s}\nabla_{\mathbf{r}_i}\phi_G^{\mathbf{p}_j}(\mathbf{r}_i;\sigma_{ij}) \tag{45}$$

and have provided an explicit expression in Appendix A. Finally, the quantity

$$\alpha_i \equiv \frac{3}{2}\sqrt{\frac{\pi}{2}}\varepsilon_s\sigma_i^3 S(\nu_i^\epsilon) \tag{46}$$

with the *screening function*

$$S(x) \equiv \frac{x}{1+(1-x/(2\sqrt{2}))/2} \tag{47}$$

and the dielectrically weighted *volume fraction*

$$\nu_i^\epsilon \equiv \frac{v_i}{(2\pi\sigma_i^2)^{3/2}}\left(1-\frac{\varepsilon_s}{\varepsilon_c}\right) \tag{48}$$

of the effective and Gaussian atomic volumes v_i and $(2\pi\sigma_i^2)^{3/2}$, respectively, is the RF *polarizability* of atom i. The volume iteration (39) guarantees that the volume fractions are in the range

$$0 \leqslant \nu_i^\epsilon < 1. \tag{49}$$

According to the definition (48) the ν_i^ϵ vanish for $\varepsilon_s=\varepsilon_c$. In this case also $S(x)$ and α_i vanish and there are no RF dipoles as one expects. Otherwise both quantities are positive and, in particular, α_i has the dimension of a volume as required for a polarizability [cf. Eq. (46)].

If one introduces with Eqs. (43) and (45) the electric field

$$\mathbf{E}_{\mathrm{pol}}^{q,\mathbf{p}}(\mathbf{r}_i) \equiv \sum_{j\neq i} [\mathbf{E}_G^{q_j}(\mathbf{r}_i;\sigma_i) + \mathbf{E}_G^{\mathbf{p}_j}(\mathbf{r}_i;\sigma_{ij})] \tag{50}$$

generated by all other charges q_j and RF dipoles $\mathbf{p}_j$ within the protein and polarizing a given Gaussian atom $i \neq j$, then the linear equation system (41) assumes the simple form

$$\mathbf{p}_i = -\alpha_i \mathbf{E}_{\mathrm{pol}}^{q,\mathbf{p}}(\mathbf{r}_i) \tag{51}$$

showing that the RF dipoles $\mathbf{p}_i$, up to an essential sign, behave just like ordinary inducible point dipoles in a polarizable force field, if these dipoles derive from isotropic atomic polarizabilities α_i (see, e.g., Ref. 30). In fact, if one uses Gaussian instead of point dipoles in such a polarizable force field, the two approaches are, up to the sign, formally identical.

The noted formal equivalence of the ET approach to a polarizable force field with Gaussian atomic dipoles was obscured by the original formulation [cf. Eq. (52) in Ref. 16] of the equation system (41) for the RF dipoles, because it contained a contribution of the RF dipole $\mathbf{p}_i$ on the right hand side expressing a kind of a self-polarization of an atom, which is absent in polarizable force fields. In our reformulation (41) this nasty self-polarization has been absorbed into the polarizability (46) of the atom i, which is given in terms of the screening function (47) and the dielectrically weighted volume fraction (48).

4. Electric field reformulated

Now also the resulting approximation for the total field (21) at an atomic position $\mathbf{r}_i$ can be reformulated. For this purpose the dipole densities $\mathbf{P}_i^{\mathbf{E}}$ in Eq. (24) are replaced by their models (40), which enables the evaluation of the integral in Eq. (25) for the dipole potential. The use of Eqs. (41)–(48) and (50) yields after some calculation

$$\widetilde{\mathbf{E}}(\mathbf{r}_i) = \mathbf{E}^{q,\mathbf{p}}(\mathbf{r}_i) - S(\nu_i^{\epsilon})\mathbf{E}_{\mathrm{pol}}^{q,\mathbf{p}}(\mathbf{r}_i) + 4\pi\frac{1}{\varepsilon_s}\sum_{j\neq i}\widetilde{\mathbf{P}}_j^{\mathbf{E}}(\mathbf{r}_i), \tag{52}$$

where the last term is due to the overlaps of the Gaussian polarizations (40) of neighboring atoms j at the position $\mathbf{r}_i$ of the given atom $i \neq j$. The first term derives from the Coulomb potentials $\phi^{q_j}(\mathbf{r}_i)$ of the point charges q_j defined by Eq. (23) and from the potentials $\phi_G^{\mathbf{p}_j}(\mathbf{r}_i;\sigma_j)$ of the Gaussian RF dipoles $\mathbf{p}_j$, which are obtained from Eq. (44) upon replacing σ_{ij} by σ_j, i.e.,

$$\mathbf{E}^{q,\mathbf{p}}(\mathbf{r}_i) = -\frac{1}{\varepsilon_s}\boldsymbol{\nabla}_{\mathbf{r}_i}\Phi^{q,\mathbf{p}}(\mathbf{r}_i) \tag{53}$$

with

$$\Phi^{q,\mathbf{p}}(\mathbf{r}_i) \equiv \sum_{j\neq i} [\phi^{q_j}(\mathbf{r}_i) + \phi_G^{\mathbf{p}_j}(\mathbf{r}_i;\sigma_j)]. \tag{54}$$

The second term in Eq. (52) is the polarizing field (50) multiplied with the screening function $S(\nu_i^{\epsilon})$. Note that this term was absent in the corresponding and strictly equivalent expressions for the electric field and the associated qE reaction field forces, which instead contained the RF dipole $\mathbf{p}_i$ implying that this dipole exerts a force on a charge q_i at the same atom [cf. Eqs. (53) and (54) in Ref. 16]. This self-interaction has now been removed by observing that according to the linear response Eq. (51) the RF dipole $\mathbf{p}_i$ is induced by $\mathbf{E}_{\mathrm{pol}}^{q,\mathbf{p}}$, which is generated by charges and RF dipoles at other atoms $j \neq i$.

The removal of this self-interaction is important, because it allows us to state the following properties:

(i) All ingredients to the electric field (52) are functions of the interatomic distances r_{ij}, i.e., $\widetilde{\mathbf{P}}_j^{\mathbf{E}}(\mathbf{r}_i) \equiv \widetilde{\mathbf{P}}_j^{\mathbf{E}}(r_{ij})$ and $\phi^{q_j}(\mathbf{r}_i) \equiv \phi^{q_j}(r_{ij})$, or distance vectors $\mathbf{r}_i - \mathbf{r}_j$, i.e., $\phi_G^{\mathbf{p}_j}(\mathbf{r}_i) \equiv \phi_G^{\mathbf{p}_j}(\mathbf{r}_i - \mathbf{r}_j)$ as one immediately sees from the corresponding definitions (40), (23), and (44). But whereas the first two contributions to the field are the negative gradients of distance dependent pair potentials, the last term apparently does not derive from such potentials. This observation indicates that the condition (9) does not hold anymore exactly for the field (52) because of the applied approximations.

(ii) In the limit of large interatomic distances $r_{ij} \gg \sigma_i, \sigma_j$ the polarizing field (50) and the field (53) become identical, because in this limit $\mathrm{erf}(r/\sqrt{2}\sigma) \to 1$, which implies that the potentials Eqs. (42) and (44) associated with Gaussian distributions become the usual potentials of point charges and dipoles, respectively. Because in this limit $G(r;\sigma) \to \delta(r)$, the third contribution to the field (52) additionally vanishes. Thus, the far field limit of Eq. (52) is

$$\widetilde{\mathbf{E}}_{\mathrm{far}}(\mathbf{r}_i) = [1 - S(\nu_i^{\epsilon})]\mathbf{E}_{\mathrm{far}}^{q,\mathbf{p}}(\mathbf{r}_i) \tag{55}$$

with the field

$$\mathbf{E}_{\mathrm{far}}^{q,\mathbf{p}}(\mathbf{r}_i) = -\frac{1}{\varepsilon_s}\boldsymbol{\nabla}_{\mathbf{r}_i}\Phi_{\mathrm{far}}^{q,\mathbf{p}}(\mathbf{r}_i) \tag{56}$$

derived from the potential

$$\Phi_{\mathrm{far}}^{q,\mathbf{p}}(\mathbf{r}_i) \equiv \sum_{j\neq i} [\phi^{q_j}(\mathbf{r}_i) + \phi^{\mathbf{p}_j}(\mathbf{r}_i)] \tag{57}$$

of point charges and dipoles and where the latter potentials are given by

$$\phi^{\mathbf{p}_j}(\mathbf{r}_i) = -\boldsymbol{\nabla}_{\mathbf{r}_i}\cdot\frac{\mathbf{p}_j}{r_{ij}}. \tag{58}$$

The far field limit (55) explains why we have called $S(\nu_i^{\epsilon})$ a *screening function*: Apart from the factor $1 - S(\nu_i^{\epsilon})$, the limit (55) is an ordinary electric field in a continuum of dielectric constant ε_s, which is sampled at a point $\mathbf{r}_i$ and is generated by point charges and point dipoles at distant positions $\mathbf{r}_j$. Because the effective volumes ν_i^{ϵ} are in the range given by the inequality (49), the factor $1 - S(\nu_i^{\epsilon})$ is in the range

$$1 \geqslant 1 - S(\nu_i^{\epsilon}) > 0.24, \tag{59}$$

showing that the dielectric solvent continuum surrounding the protein induces at large interatomic distances a simple shielding of the electrostatic interactions among the charges and RF dipoles within the protein, which depends on the dielectrically weighted volume fraction ν_i^{ϵ} of the target atom. For $\varepsilon_s = \varepsilon_c$ the

factor becomes 1, the RF dipoles vanish, and one regains the Coulomb field generated by the point charges, which is the usual model in MM-MD simulations.

5. Energy reformulated

As follows from Eq. (35) the computation of the electrostatic energy (33) requires an expression for the second type of dipole densities $\mathbf{P}_i^{\Phi}(\mathbf{r})$ defined by Eq. (19). Because Eq. (31) relates these densities to the dipole densities $\mathbf{P}_i^{\mathbf{E}}(\mathbf{r})$ used for the computation of the field, the corresponding approximations (40) and (38) naturally induce a suitable approximation also for the $\mathbf{P}_i^{\Phi}(\mathbf{r})$, which enables the evaluation of the integrals in Eqs. (35) and (30). Following the derivation in Ref. 16 one obtains after some regrouping of the results [and by using Eq. (54) for the potential generated by the point charges and RF dipoles as well as by employing Eq. (48) for the dielectrically weighted volume fractions ν_i^{ϵ}] the revised expression

$$\tilde{W} = \frac{1}{2\varepsilon_s}\sum_i q_i\{\Phi^{q,\mathbf{p}}(\mathbf{r}_i) - \nu_i^{\epsilon}[\Phi_{G,\Phi}^{q,\mathbf{p}}(\mathbf{r}_i) + B_i] - \left(1-\frac{\varepsilon_s}{\varepsilon_c}\right)\sum_{j\neq i}\vartheta_j(\mathbf{r}_i)[\Phi_{G,\Phi}^{q,\mathbf{p}}(\mathbf{r}_j) + B_j] - 4\pi\sum_{j\neq i}\mathbf{p}_j\mathbf{r}_{ij}G(r_{ij};\sigma_j) \tag{60}$$

for the total electrostatic energy. Here, a new potential

$$\Phi_{G,\Phi}^{q,\mathbf{p}}(\mathbf{r}_i) \equiv \sum_{k\neq i}[\phi_{G,\Phi}^{q_k}(\mathbf{r}_i) + \phi_{G,\Phi}^{\mathbf{p}_k}(\mathbf{r}_i)] \tag{61}$$

has been introduced, which is composed of the screened potentials

$$\phi_{G,\Phi}^{q_k}(\mathbf{r}_i) \equiv \frac{q_k}{r_{ik}} f^q(r_{ik};c_\phi\sigma_i) \tag{62}$$

of the charges and

$$\phi_{G,\Phi}^{\mathbf{p}_k}(\mathbf{r}_i) \equiv \frac{\mathbf{p}_k\mathbf{r}_{ik}}{r_{ik}^3} f^{\mathbf{p}}(r_{ik};c_\phi\sigma_i;\sigma_k) \tag{63}$$

of the RF dipoles. The short-range screening functions f^q and $f^{\mathbf{p}}$ are given by Eqs. (B1) and (B2) in Appendix B. Furthermore, the constants

$$B_i \equiv \sqrt{\frac{\pi}{2}}\frac{q_i}{c_\phi\sigma_i} \tag{64}$$

appearing in Eq. (60) allow us to adjust the total energy through choice of the factor c_ϕ to the Born energy[31] of a solute protein. In Ref. 16 the value

$$c_\phi = 2.774 \tag{65}$$

has been suggested for this parameter. Finally, the RF contribution $\tilde{W}^{\epsilon}$ to the total electrostatic energy (60) is obtained by subtracting the Coulomb contribution (34).

As opposed to the strictly equivalent expression [Eq. (67)] in Ref. 16, the reformulated representation (60) of the electrostatic energy $\tilde{W}$ does not contain any diagonal ($i=j$) terms in the multiple summations required for its evaluation. Their role is taken over by the atomic contributions $q_i\nu_i^{\epsilon}B_i$ to the Born energy of the protein and by the energies of the charges q_i in the potentials $\Phi_{G,\Phi}^{q,\mathbf{p}}(\mathbf{r}_i)$, which are generated by the charges and RF dipoles at the other atoms $k\neq i$ and are weakened by the local screening factors ν_i^{ϵ}. Recall here that in the reformulated Eqs. (41) for the RF dipoles and Eq. (52) for the electric field the removal of self-contributions was similarly achieved by introducing a screening function $S(\nu_i^{\epsilon})$.

Together with the definitions (54) and (61) of the various potentials the reformulated representation (60) of $\tilde{W}$ now reveals that this total electrostatic energy and its RF part $\tilde{W}^{\epsilon}$ derive from distant dependent pair interactions among the atoms and from atomic Born contributions. Therefore, the approximate forces

$$\tilde{\mathbf{F}}_j^W \equiv -\boldsymbol{\nabla}_{\mathbf{r}_j}\tilde{W}, \tag{66}$$

which include the Coulomb and the mean RF forces [cf. Eq. (36)], will obey Newton's third law, if the configurational derivatives of the Born contributions $q_i\nu_i^{\epsilon}B_i$ vanish. This is actually the case, because the q_i and B_i are constants and because the volume fractions ν_i^{ϵ} are proportional to effective volumes v_i determined by a self-consistency iteration [cf. Eq. (39)].

However, the representation (60) of $\tilde{W}$ also reveals why one cannot expect to gain a very efficient MD algorithm on this route of reasoning. For instance, the contribution containing the partition functions ϑ_j requires the time-consuming evaluation of a triple sum over the atoms. In contrast, the representation (52) of the electric field, from which one can obtain the qE forces using Eq. (32), solely contains double summations. Therefore we asked the question, whether an efficient algorithm for the approximate computation of the mean RF forces can be constructed by a suitable modification of the qE forces.

D. *Reactio* forces

If one wants to compute the mean forces $-\boldsymbol{\nabla}_{\mathbf{r}_i}W$ by complementing the *actio* qE forces with suitable *reactio* forces and has as starting points solely approximate expressions for $\mathbf{E}$ and W [implying, in particular, that the field (52) is not exactly the negative gradient of the potential, from which the energy (60) is derived],[16] one has to make sure that the two approaches toward force computation approximately match in certain limiting cases. For this purpose the far distance limit, in which the field assumes the simple form (55), appears to be particularly suited. Therefore, we now consider the far distance limit also for the representation (60) of the total electrostatic energy and for the associated approximate forces (66).

1. Large distance limit of $\tilde{W}$

For interatomic distances r large compared to the Gaussian atomic radii σ the two screening functions appearing in Eqs. (62) and (63) and entering $\tilde{W}$ through the potential (61) rapidly converge to one, i.e., $f^q, f^{\mathbf{p}} \to 1$. In this limit the

Gaussians as well as the partition functions $\vartheta_j(\mathbf{r}_i)$ in Eq. (60) vanish (because the corresponding summations cover only terms with $i \neq j$). Therefore, the far distance contribution $\tilde{W}_{\text{far}}$ to the energy is

$$\tilde{W}_{\text{far}} = \frac{1}{2\varepsilon_s} \sum_i q_i \{[1 - \nu_i^\epsilon] \Phi_{\text{far}}^{q,\mathbf{p}}(\mathbf{r}_i) - \nu_i^\epsilon B_i\} \tag{67}$$

with the atomic Born contributions (64) and with the potential $\Phi_{\text{far}}^{q,\mathbf{p}}$ of point charges and dipoles given by Eq. (57). Thus, $\tilde{W}_{\text{far}}$ derives from exactly the same potential $\Phi_{\text{far}}^{q,\mathbf{p}}$ as the far distance contribution (55) to the field (52) [cf. Eq. (56)]. Note, however, that in Eq. (67) the potential is scaled by factors $1-\nu_i^\epsilon$ whereas in Eq. (55) the field is scaled by factors $1-S(\nu_i^\epsilon)$. This difference is due to the different routes chosen for the approximate computation of the potential and field, respectively. Keeping this difference in mind and comparing the far distance contributions to the mean forces (66) obtained from $\tilde{W}_{\text{far}}$ with the far distance qE forces resulting from Eq. (56) provide now the clues for the desired modification of the *actio* qE forces associated with the field (52).

For this purpose we first calculate the negative gradient

$$-\nabla_{\mathbf{r}_i} \tilde{W}_{\text{far}} = \frac{1}{2}(1 - \nu_i^\epsilon) q_i \mathbf{E}_{\text{far}}^{q,\mathbf{p}}(\mathbf{r}_i) + \frac{1}{2} \sum_{j \neq i} (1 - \nu_j^\epsilon) \{q_i \mathbf{E}^{q_j}(\mathbf{r}_i) + [\nabla_{\mathbf{r}_i} \otimes \mathbf{E}^{q_j}(\mathbf{r}_i)] \mathbf{p}_i\} \tag{68}$$

of $\tilde{W}_{\text{far}}$, where $\otimes$ denotes a tensor product, where the field $\mathbf{E}_{\text{far}}^{q,\mathbf{p}}$ is given by Eq. (56) and where the fields

$$\mathbf{E}^{q_j}(\mathbf{r}) \equiv -\frac{1}{\varepsilon_s} \nabla_{\mathbf{r}} \phi^{q_j}(\mathbf{r}) \tag{69}$$

of the point charges q_j are derived from the unscreened Coulomb potentials (23). The corresponding limit for the *actio* qE forces associated with the field (55) is, apart from a factor of 1/2 and a different screening factor $1-S(\nu_i^\epsilon)$ replacing $1-\nu_i^\epsilon$, identical to the first term in Eq. (68). It represents the *actio* forces. But the second term is new. It reveals—in addition to forces acting on the charge q_i—also forces acting on the dipole $\mathbf{p}_i$, which both are generated by the properly screened Coulomb fields $(1-\nu_j^\epsilon)\mathbf{E}^{q_j}$ of the other charges q_j. These two additional force contributions serve to guarantee Newton's third law because they introduce symmetry into the interactions among the charges and introduce *reactio* forces into the interactions between dipoles and charges.

2. Reactio forces compensating the qE forces

The above consideration of the far distance limit suggests how one should modify the qE forces associated with the electric field (52) in order to guarantee the *actio=reactio* principle. Like in Eq. (68) we split the force

$$\mathbf{F}_i = \mathbf{F}_i^a + \mathbf{F}_i^r \tag{70}$$

on an atom at $\mathbf{r}_i$ into an *actio* contribution

$$\mathbf{F}_i^a = \frac{1}{2} q_i \tilde{\mathbf{E}}(\mathbf{r}_i), \tag{71}$$

which is half the qE force generated by the electric field (52), and a *reactio* contribution

$$\mathbf{F}_i^r = \mathbf{F}_i^{r,1} + \mathbf{F}_i^{r,2} + \mathbf{F}_i^{r,3}, \tag{72}$$

whose three components are associated with corresponding components of the field (52).

The first component

$$\mathbf{F}_i^{r,1} = \frac{1}{2} \sum_{j \neq i} \{q_i \mathbf{E}^{q_j}(\mathbf{r}_i) + [\nabla_{\mathbf{r}_i} \otimes \mathbf{E}_G^{q_j}(\mathbf{r}_i; \sigma_i)] \mathbf{p}_i\} \tag{73}$$

derives from the field component $\mathbf{E}^{q,\mathbf{p}}$ given by Eq. (53), contains the fields (69) and (43), and covers the *reactio* forces $\mathbf{F}_{ij}^r$, which are required to compensate the qE *actio* $\mathbf{F}_{ji}^a$ of the charge q_i and RF dipole $\mathbf{p}_i$ at atom i on the charges q_j at the other atoms $j \neq i$. It has the same general form as the second term in Eq. (68), which is also composed of Coulomb fields $\mathbf{E}^{q_j}$ acting on the charge q_i and gradients $\nabla_{\mathbf{r}_i} \otimes \mathbf{E}^{q_j}$ of such fields acting on the RF dipole $\mathbf{p}_i$. However, because the *Gaussian* dipole densities $\tilde{\mathbf{P}}_i^{\mathbf{E}}$ contribute pair forces

$$\mathbf{F}_{ji}^{a,q_j,\mathbf{p}_i} = -\frac{1}{2\varepsilon_s} q_j \nabla_{\mathbf{r}_j} \phi_G^{\mathbf{p}_i}(\mathbf{r}_j; \sigma_i)$$

to the *actio* qE force $\mathbf{F}_j^{a,1} \equiv q_j \mathbf{E}^{q,\mathbf{p}}$ on q_j also the *reactio* forces

$$\mathbf{F}_{ij}^{r,\mathbf{p}_i,q_j} = [\nabla_{\mathbf{r}_i} \otimes \mathbf{E}_G^{q_j}(\mathbf{r}_i; \sigma_i)] \mathbf{p}_i$$

generated by the charges q_j and acting on the density $\tilde{\mathbf{P}}_i^{\mathbf{E}}$ carry a Gaussian signature as indicated by the parameter σ_i and by the subscript G at the field $\mathbf{E}_G^{q_j}$. With the indicated choice of the parameter σ_i this field is given by Eqs. (42) and (43). An explicit expression for the field gradient $[\nabla_{\mathbf{r}_i} \otimes \mathbf{E}_G^{q_j}(\mathbf{r}_i; \sigma)]$ is provided in Appendix A.

The second component

$$\mathbf{F}_i^{r,2} = -\frac{1}{2} \sum_{j \neq i} S(\nu_j^\epsilon) \{q_i \mathbf{E}_G^{q_j}(\mathbf{r}_i; \sigma_j) + [\nabla_{\mathbf{r}_i} \otimes \mathbf{E}_G^{q_j}(\mathbf{r}_i; \sigma_{ij})] \mathbf{p}_i\} \tag{74}$$

of the *reactio* force $\mathbf{F}_i^r$ belongs to the polarizing field defined by Eq. (50), whose contribution to the total electric field (52) at atom i is steered by the value of the screening function $S(\nu_i^\epsilon)$. The construction principle of the second *reactio* force component is identical to that of the first component (73); its ingredients are defined by Eqs. (43) and (45).

In the far distance limit $r \gg \sigma$ all partial fields contributing to the first two *reactio* force components become ordinary fields $\mathbf{E}^{q_j}$ of point charges q_j. Thus, their sum becomes identical to the *reactio* force derived from $\tilde{W}_{\text{far}}$, i.e., to the second term at the right hand side (rhs) of Eq. (68), if one replaces the screening function $S(\nu_j^\epsilon)$ by ν_j^ϵ. Because all partial fields contributing to the first two *reactio* force components as well as to the associated qE *actio* force components derive from potentials [cf. Eqs. (43) and (45)], one can easily write down an energy function $\hat{W}$, from which one gets these forces by taking configurational derivatives like in Eq. (66).

Unfortunately, this possibility ceases to exist for the third component

$$\mathbf{F}_i^{r,3} = -\frac{1}{2\varepsilon_s}\sum_{j\neq i} q_j 4\pi \tilde{\mathbf{P}}_i^{\mathbf{E}}(\mathbf{r}_j). \tag{75}$$

This *reactio* force component is caused by the overlaps of the Gaussian dipole densities $\tilde{\mathbf{P}}_j^{\mathbf{E}}$ with the positions $\mathbf{r}_i$ of neighboring $(i \neq j)$ point charges q_i. The overlaps generate a field and a corresponding *qE actio* on atom i [cf. Eq. (52)], which is compensated by $\mathbf{F}_i^{r,3}$. When combined with the other force components, Newton's third law is guaranteed. Thus, the sum over all forces (70) vanishes and no net force acts on the system.

Nearest neighbor contributions, which are related to the forces $\mathbf{F}_i^{r,3}$ and likewise vanish at distances $r \gg \sigma$, are also present in the approximate expression for the potential of mean force $\tilde{W}$ [cf. the last two terms at the rhs of Eq. (60)]. However, the nearest neighbor forces $\mathbf{F}_i^{r,3}$ (as well as their *actio* counterparts $\mathbf{F}_i^{a,3}$) derived from our approximation (52) of the field are not negative gradients (36) of the nearest neighbor contributions to $\tilde{W}$, because of the different approximation scheme used in the respective derivations. Therefore, the $\mathbf{F}_i^{r,3}$ cannot guarantee that also the total torque, which is exerted by the forces (70) on the system, is exactly zero. In MD simulations based on the forces (70)–(75) one will therefore have to compensate that residual torque by applying a rotation correction at each integration step of the dynamics.

III. SUMMARY AND DISCUSSION

Already the discussion of the exact ET theory in Sec. II B has demonstrated that *reactio* forces start to show up, if one considers the mean RF forces (36) instead of the *qE* forces (32). One of the reasons why this interesting discrepancy has been previously overlooked may be the apparent complexity of the ET theory. Despite its conceptual simplicity, the details of this theory are truly complicated.[16] Part of these complexities are due to the various self-polarizations and -interactions appearing in the original equations. In our reformulation we have removed these self-terms by introducing (i) screening functions $S(\nu_i^{\epsilon})$ into the equations [Eq. (41)] for the RF dipoles and [Eq. (52)] for the electric field, (ii) screening factors ν_i^{ϵ}, and (iii) explicit atomic Born contributions into the equation [Eq. (60)] for the electrostatic energy. By considering the far distance limits of the forces derived from the field and from the energy, respectively, the revised expressions then enabled an identification of those terms, by which the *qE* forces have to be complemented, if Newton's third law is supposed to hold. Before turning to an outlook concerning the use of the suggested method in MD simulations of biomolecules, we will briefly discuss the viability of and physical insights provided by our reformulation of the ET theory.

A. Equivalence with the original theory

Figure 2 demonstrates that the reformulated approach is strictly equivalent to the original one and illustrates that the approximation (52) of the field is actually better than that [Eq. (60)] of the energy. For this purpose the figure compares the *qE* force components $F_x(x)$, which were obtained from the analytical Kirkwood solution[32] for a unit charge $+e$ in a spherical cavity (line), with those (dots) obtained by Eq. (52) for a regularly discretized model of that cavity. Here, the cavity is centered at the origin of a Cartesian coordinate system, the charge moves along the x axis from the center toward the surface of the sphere, and the discretized model sphere is constructed from a hexagonal lattice of identical atoms. Figure 3 shows a cross section through that sphere together with further details, which will be discussed further below in a different context. Apart from the fact that the lattice constant ξ has been chosen two times smaller (resulting in a better discretization), the comparison in Fig. 2 is

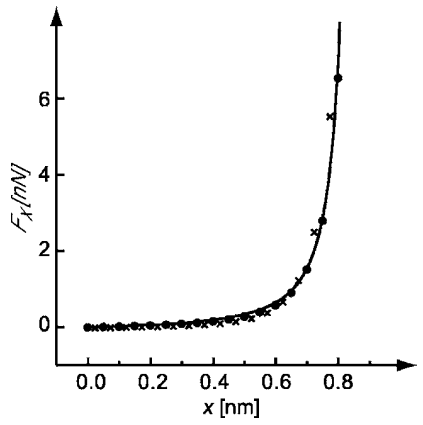

FIG. 2. The x component of the qE force $\mathbf{F}_i = q_i\mathbf{E}(\mathbf{r}_i)$ acting on a unit charge $q_i = e$ in an otherwise uncharged spherical protein centered around the origin of a Cartesian coordinate system as a function of its position on the x axis. Line: Kirkwood solution for the charge in a spherical cavity (with a radius of 0.89 nm) (Ref. [32]). Dots: reformulated ET method using the expression (52) for the field and the discretized atomic model of a sphere depicted in Fig. 3. Crosses: negative difference quotients $\hat{F}_{x,i}^{W}(i\xi+\xi/2) \equiv -\{\tilde{W}^{\epsilon}[(i+1)\xi] - \tilde{W}^{\epsilon}(i\xi)\}/\xi$ of the energy (60) for $i=0,\ldots,15$ and $\xi=0.05$ nm. For further explanations see the text and the caption to Fig. 3.

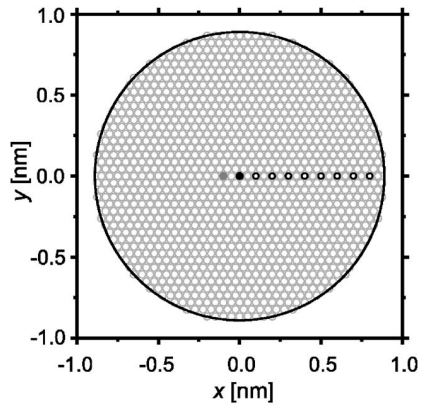

FIG. 3. Cross section through a discrete model of a Kirkwood sphere (large black circle) composed of 33 477 Gaussian atoms densly packed on a hexagonal lattice. The sizes of the atoms are indicated through their Gaussian radii $\sigma=0.022$ nm (small gray circles). An elementary charge (black dot) hops along lattice sites x_c (small black circles) on the x axis away from an uncharged atom (gray dot) at x_u just to the left of the origin. The RF dipole at the uncharged atom causes a *reactio* force, as shown in Fig. 4. For explanation see the text.

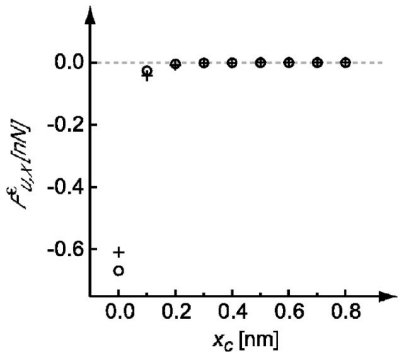

FIG. 4. x components $F^{\epsilon}_{u,x}$ of RF forces on an uncharged atom (gray dot) at x_u responding to a charged atom (black dot) in the spherical model protein depicted by Fig. 3. Open circles: numerical configurational derivatives $-d\tilde{W}/dx_u$ for different x positions x_c of the charge. Crosses: forces calculated by Eqs. (70)–(75).

identical to the one, which was given in Fig. 4 of Ref. 16 and which originally served to illustrate the remarkable accuracy of the approximations employed for the computation of the electric field. In the present context this first comparison (line versus dots) serves to prove the equivalence of the revised expression (52) for the field with the original one. The second comparison (line versus crosses) exhibited by the figure pertains to forces $F_x=-dW/dx$ obtained by taking negative derivatives of energy expressions. According to the Kirkwood solution the negative derivative $-dW^{\mathrm{Kirk}}/dx$ of the exact RF energy W^{Kirk} of the charge is identical to the corresponding qE force (line). Evaluation of the energy expression (60) for the charge hopping within the discretized model sphere along the atoms i centered at the x axis and approximating $-d\tilde{W}/dx$ by the finite differences $\hat{F}^{W}_{x,i}(x)$ yields the crosses at the positions $x=i\xi+\xi/2$. As claimed above these crosses show larger deviations from the exact solution than the dots. Besides the issues of computational efficiency this accuracy difference between the approximations employed for the computation of the energy and of the field, respectively, is another reason why one should preferentially calculate atomic forces from the properly extended qE expressions (70)–(75) instead through Eq. (66).

B. RF dipoles are real

Clearly, our reformulation would have been superfluous, if it would not have led to new insights. Here, as a first result, the removal of the self-polarizations in the original equations for the RF dipoles [Eq. (52) in Ref. 16] has revealed their formal equivalence with inducible Gaussian dipoles in a usual polarizable force field. This equivalence has become apparent by inspection of the linear response equation (51), in which the unusual negative sign reminds us that the atomic RF dipoles are image objects representing the polarization of the surrounding continuum.

This first insight is not only a question of esthetics but guides the attention to the usual theories of polarizable force fields (cf., e.g., Refs. 30 and 33), in which the forces on the atoms are obtained by configurational derivatives of electrostatic energies. According to these theories electrostatic forces on the atoms act not only through their partial charges but also through their induced dipoles. In the same way the inspection of the configurational derivative (36) of the exact potential of mean force has shown that the RF dipoles contribute to the mean forces acting on the atoms.

This second insight then guides to a further reading of the relevant chapters in the well-known book of Jackson,[29] which clearly state that the *reactio* forces exerted by a dielectric continuum on the polarizing charges are obtained by a procedure, which is essentially identical to the configurational derivative (36). Therefore the forces, which according to the exact expression (36) act through the atomic RF dipole densities on the atoms, represent the continuum *reactio* forces, i.e., the so-called dielectric boundary pressure. Because within the approximation (66) these forces are mediated by the RF dipoles, we arrive at the conclusion that the RF dipoles should be conceived as real physical entities and not as pure auxiliary variables.

C. Mean RF forces for the Kirkwood sphere

Although our investigation has shown that the approach towards the mean RF forces through the configurational derivatives of W^{ε} [cf. Eqs. (36) and (66)] and, thus, towards forces obeying Newton's third law is precluded by its computational inefficiency and its inferior accuracy, its principle viability can nevertheless be demonstrated. For this purpose we consider once more the discretized Kirkwood sphere depicted in Fig. 3, which we already used for the force comparisons in Fig. 2. But now we compute the *reactio* RF forces $\mathbf{F}^{\varepsilon}_{u}(x_c)$ on an uncharged atom (gray dot) at a position x_u on the x axis, which are caused by a unit charge moving along positions x_c (also on the x axis) towards the boundary of the sphere. By symmetry it suffices to consider the x components $F^{\varepsilon}_{u,x}(x_c)$ of these forces, which we can either evaluate from Eqs. (70)–(75) or by numerically computing the configurational derivative of the approximate energy $\tilde{W}$. This numerical derivative is obtained by the following procedure: The charge is positioned at one of the indicated positions x_c (black circles and black dot) within the discretized spherical model protein. Then the position x_u of the selected uncharged atom is varied by small shifts of ± 0.0001 and ± 0.0002 nm along the x axis. For each shifted position the effective volumes (39) and RF dipoles (41) are self-consistently calculated and the energy Eq. (60) is evaluated. The configurational derivative $d\tilde{W}/dx_u$ is then obtained through a numerical symmetric four-point derivative[34] for each x_c.

For the *reactio* RF forces on the uncharged atom marked in Fig. 3, Fig. 4 compares the numerical configurational derivatives $-d\tilde{W}/dx_u$ (open circles) with data (crosses) computed by Eqs. (70)–(75). The results of the two different computational approaches exhibit a quite close agreement and both indicate that the uncharged atom is effectively repelled by the charge. The agreement demonstrates that our construction (70)–(75) of *reactio* forces is essentially correct.

The close agreement exhibited by Fig. 4 is surprising (i) because different approximations are employed on the respective routes of computation and (ii) because the numerical derivatives $-d\tilde{W}/dx_u$ involve configurations of the discretized sphere, in which the original regularity of the atomic

arrangement is locally perturbed. From a fundamental property of Jacobi's classical theta function[35] one knows that superpositions (37) of identically weighted Gaussians on a regular lattice result in extremely smooth characteristic functions $\tilde{\Theta}(\mathbf{r})$, which, however, become strongly perturbed upon the slightest deviations from regularity. Correspondingly, our approach toward the approximate solution of the Poisson equation should perform particularly well for regular lattices and should exhibit much larger numerical inaccuracies for perturbed lattices. Therefore, numerical differentiations of the kind employed above, apart from involving differences of large numbers obtained from iterative self-consistency procedures, cannot be very accurate. In fact, we have checked that the relative deviations of the forces computed by the two approaches are moderate (1%) as long as the forces are large (at $x_c=0$) and become sizable (8%) only for very small forces (at $x_c=8$ nm).

D. Integration into a fast multipole method

The far distance limits [Eq. (55)] of the electric field and [Eq. (67)] of the electrostatic energy are not only significant, because they have provided clues on how to introduce *reactio* forces, but also because they show that at distances $r \gg \sigma$ the interactions involving Gaussian dipole distributions reduce to those of point dipoles. Therefore the computation of these interactions becomes much simpler at larger distances and nicely fits into a computational scheme like the structure adapted multipole method (SAMM),[8,9] which applies distance classes and associated hierarchical multipole expansions to the computation of long-range electrostatic interactions in proteins. SAMM reduces the algorithmic complexity of the classical electrostatic many body problem to linear scaling. Therefore, we employ this scheme for the computation of the polarizing field (50), of the forces (70)–(75) and energies (60) in our implementation of the ET approach into the parallelized molecular dynamics program package EGO-MMII.[10] Concerning the details of this implementation we merely note that separate data structures are required for fields and potentials associated with RF dipoles, because only the charges contribute to the forces on the RF dipoles [cf. Eqs. (73)–(75)], whereas the charges *and* RF dipoles contribute to the polarizing field (50). Moreover, a separate bookkeeping of the RF data is favorable, because it allows us to individually measure the RF contributions to the forces and energies. Further details of our implementation and particularly all issues of computational efficiency will be discussed in connection with forthcoming sample applications serving to evaluate and optimize the simulation approach.

E. Extension to Gaussian charge distributions

As pointed out further above, the formalism of the original ET theory is quite complex. Although our reformulation has diminished some of these complexities, it did not yet reach the aim of removing them entirely. Interestingly, the reformulated theory becomes formally much simpler, if one employs Gaussian charge distributions of widths σ_i instead of point charges to model the charge distribution within a protein. As a result the Coulomb potentials q/r of the point charges are replaced by their Gaussian counterparts $(q/r)\mathrm{erf}(r/\sqrt{2}\sigma)$. As one can easily show, such an extension eliminates the somewhat complicated and confusing fact that in Eqs. (41), (50), (54), (73), and (74) some fields or potentials depend on σ_{ij}, others on σ_i or σ_j, and still others contain no reference to Gaussian widths at all, because they are associated with point charges. With Gaussian charge distributions all these interaction expressions exclusively depend on the mixed widths σ_{ij} (where $\sigma_{ij}^2=\sigma_i^2+\sigma_j^2$) and all interactions among dipoles and charges at atoms i and j are screened at short distances on the same scale σ_{ij}. This scale can also be used in Eq. (39) for the computation of the effective atomic volumes v_i, if the arguments employed for its derivation are properly adjusted. Apart from a slight modification of the screening function (47), which then simplifies to $S(x)=x/[1+(1-x)/2]$, the computation of the qE forces and of the associated *reactio* forces remain unchanged. For the Kirkwood sphere model we have checked that such an extension yields qE forces essentially identical to those depicted in Fig. 2. Solely the computation of the energy (60) requires some additional analytics, because a certain integral [i.e., the one appearing in Eq. (62) of Ref. 16] must still be determined for its evaluation. However, because conventional MM force fields for proteins employ partial point charges and because a transition to Gaussian charge distributions would cost a certain reparametrization effort, we have postponed further investigations of this intriguing concept.

F. Isolated atom limits

Despite the announced postponement of the Gaussian extension we would like to illustrate the differences between point charges and Gaussian charge distributions already now. As examples we use the most simple limiting cases of (a) two ions and (b) an ion and an uncharged atom in a solvent continuum. Although these examples are actually outside the intended range of applicability of the ET theory, which has been originally designed for single macromolecules in solution, they provide insights into the way how RF effects are covered.

Consider first two ions with charges $+e$ and $-e$ separated by a distance r in continuous aqueous solution ($\varepsilon_c=80$). The assumption that the charges have identical van der Waals radii $R_{\mathrm{vdW}}=1.7$ Å defines then the contact distance $r_0/R_{\mathrm{vdW}}=2$. Furthermore, the common ion radius R_{vdW} can be used to define a common Gaussian width σ in such a way that the volumes of a hard sphere with radius R_{vdW} and of a Gaussian soft sphere with width σ are identical. One immediately gets the ratio $\sigma/R_{\mathrm{vdW}}=0.64$. Setting the dielectric constant within the ions to $\varepsilon_s=1$ fixes the last parameter required for the computation of the attractive electrostatic force $\mathbf{F}(r)=F(r)\mathbf{r}/r$ acting according to Eqs. (70)–(75) on one of the ions.

Figure 5 compares the amplitudes $F_1(r)$ of the electrostatic forces on ion 1 calculated by our RF approach for point (solid curve) and Gaussian charge models (black dashed curve). At the contact distance $r_0/R_{\mathrm{vdW}}=2$, which is indicated by a vertical bar, the two models predict nearly the

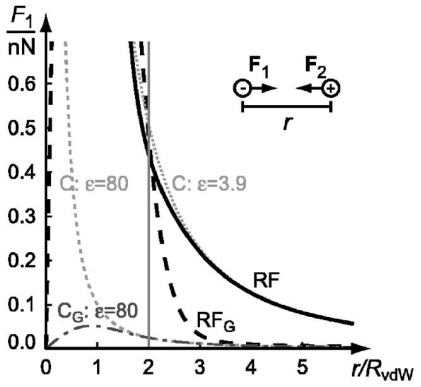

FIG. 5. Absolute value of the electrostatic force acting between two oppositely charged ions ($\pm e$) with identical van der Waals radii $R_{vdW}=1.7$ Å in an aqueous continuum ($\varepsilon_c=80$). Line styles differentiate approaches. Solid: RF approach with point charges. Black dashes: extension to Gaussian charge distributions ($\sigma/R_{vdW}=0.64$). Light gray: Coulomb interaction of point charges screened by $\varepsilon_c=80$ (dashes) and $\varepsilon_c=3.9$ (dots). Dark gray dash-dot: Coulomb interaction of Gaussian charge distributions screened by $\varepsilon_c=80$. For discussion see the text.

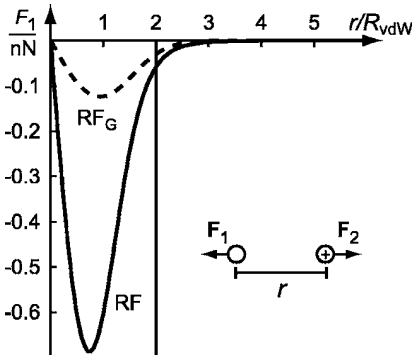

FIG. 6. Amplitude of the RF force acting on an uncharged atom in an aqueous continuum ($\varepsilon_c=80$), which is polarized by an ion ($+e$). Solid line: RF approach with point charges. Dashed line: extension to Gaussian charge distributions ($\sigma/R_{vdW}=0.64$). For discussion see the text.

same force. The differences at smaller distances, notably the fact that the electrostatic force vanishes at $r=0$ for the Gaussian and diverges for the point charge model, do not matter, because the ions can never reach this range of distances. At larger distances the force decays much more rapidly for the Gaussian than for the point charge model. Beginning at the contact distance the asymptotics of the slow decay (point charge model) is apparently described by the dotted curve (light gray), which is the Coulomb force acting between two actual point charges surrounded *everywhere* by a mildly polar continuum of dielectric constant $\varepsilon_c=3.9$. In contrast, the rapid force decay (Gaussian model, black dashed curve) appears to approach the Coulomb forces acting between actual point (light gray dashed) or Gaussian charges (dark gray dashed-dotted) entirely embedded in a strongly polar continuum of dielectric constant $\varepsilon_c=80$. A closer inspection of the mathematics reveals that the asymptotic decay of the Gaussian charge model (black dashed curve) is actually governed by a slightly smaller dielectric constant $\varepsilon_c=53.3$, which nevertheless is large enough to mimic a strongly polar solvent. In this context we would like to remark that there are widespread MM models for water [e.g., the simple point charge (SPC) model[36]] with similar [$\varepsilon_c(\mathrm{SPC})\approx 65$] dielectric constants.

The above example shows that the reformulated ET approach provides reasonable descriptions of electrostatic interactions even for scenarios (such as the one exemplified above), which originally were thought to be outside its range of applicability. If one envisages future applications to problems of docking or peptide-peptide interactions, the Gaussian charge model appears to be superior, because it provides a much stronger shielding of the electrostatic interactions between separate molecules than the point charge model. Furthermore, the close similarity of the Gaussian and point charge forces at contact distance appears to indicate that no substantial reparametrizations of force field parameters are required in a transition from point to Gaussian charges (if the Gaussian widths are sufficiently small). On the other hand, the very strong shielding ($\varepsilon_c=53.3$) associated with Gaussian charges is acquired already at very small distances from the van der Waals surface and it remains to be checked by comparisons with simulations using explicit water models, to what extent this behavior is correct. In this context it should be noted that the strong decay of the electrostatic forces beyond the contact distance observed for the Gaussian model closely resembles (even quantitatively, data not shown) a screened Coulomb potential employed in generalized Born[17] (GB) or other[37] implicit solvent models. However, such models cannot account for the RF forces repelling an uncharged atom in aqueous solution from a nearby ion (and *vice versa)*.

Figure 6 illustrates how this "lipophobic interaction," which in our approach is caused by a repulsion between the RF dipole induced at the uncharged atom and the charge of the ion, comes out for point (solid curve) and Gaussian charge models (black dashed curve). Like in the first example the van der Waals radii of the atom and ion were chosen identical $R_{vdW}=1.7$ Å, their distance is denoted by r, and they are surrounded by a dielectric continuum with $\varepsilon_c=80$. At contact distance the RF force predicted by the point charge model is slightly larger than that of the Gaussian charge model. Here, the lipohobic force is by about a factor of 10 smaller than the attractive force acting between the two ions. At larger distances both charge models predict similarly rapid decays of the lipophobic interaction. The lipophobic repulsion becomes of course much larger, if the ion approaches a whole bunch of uncharged atoms instead only a single one, because then the set of RF dipoles induced within the uncharged cluster can combine their efforts to repel the charge. Thus, although the ET theory originally was not designed for a description of isolated atoms or molecules in solution, the shown examples demonstrate that even such cases are reasonably covered.

G. Outlook

In the above manuscript we have outlined an approach toward the efficient and quite accurate computation of RF forces, which obey Newton's third law and thus account for the dielectric surface pressure. Although the revised theory looks quite promising already at its present stage, there are still a few shortcomings, which should be addressed in future

analytical work. The main isssue here is the fact that the "third component" of the RF force [cf. Eq. (75)] as opposed to the remaining components cannot be represented as the derivative of a potential function. This is the reason why neither the total energy nor the angular momentum will be exactly conserved in RF/MD simulations although the violations of these conservation laws are expected to be small. Furthermore several practical hurdles are still to be passed before the present theory can be safely applied to MD simulations of biomolecules embedded in a dielectric continuum.

The first hurdle pertains to the choice of the Gaussian widths σ_i, which characterize the sizes of the various atoms and determine by Eq. (39) their effective volumes v_i. The parameters σ_i and v_i are highly important, because they steer through Eq. (46) the atomic RF polarizabilities, through the linear response (51) the sizes of the RF dipoles, and through Eqs. (70)–(75) the sizes of the RF forces acting on the atoms. Egwolf and Tavan had suggested to choose the σ_i proportional to the atomic van der Waals Radii R_i given in common MM force fields.[16] However, it is *a priori* not clear whether the R_i, which are usually chosen to yield suitable descriptions of van der Waals interactions, can also render acceptable atomic RF polarizabilities. Thus, it must be studied how a given choice of the Gaussian widths σ_i affects the resulting description of the RF forces and solvation free energies. Such a study requires extended comparisons of simulations with and without explicit solvent as well as investigations of the associated free energy landscapes.

The second hurdle pertains to the choice of a suitable dynamics for the RF dipoles. Clearly, one could compute the RF dipoles self-consistently at each integration step of a MD simulation. However, such an attempt would be conceptually incompatible with the notion that the RF energy represents a potential of *mean* force and with the fact that dielectric relaxation times in liquids such as water are in the range between a few hundreds of femtoseconds up to several picoseconds whereas the atomic motions sampled by MD within a biomolecule are about two orders of magnitude faster. Thus, a self-consistent computation of the RF dipoles within MD would render them as much too rapidly fluctuating variables. As a result, the modeling of the RF dipole dynamics must account for the relatively slow dielectric relaxation.

The last hurdle pertains to the choice of techniques, by which one can control the expectedly small net torques exerted by the RF forces on the given biomolecule. As repeatedly indicated in the paper such torques arise because certain contributions, i.e., those given by Eq. (75), to the RF forces apparently do not represent negative gradients of a potential function. Thus, a required torque control will represent a source of algorithmic noise, whose size and dynamical consequences have to be estimated.

In the follow-up manuscript we will start to tackle this hurdle using the most simple biomolecule, which is alanine dipeptide, as our first example, because its free energy landscape has been extensively studied using various different MD approaches.[26,37–40]

ACKNOWLEDGMENTS

The authors would like to thank Bernhard Egwolf and Gerald Mathias for useful discussions. The project has been funded by the Bayerischer Forschungsverbund Prionen (Project No. LMU02) and by the Deutsche Forschungsgemeinschaft (SFB 533/C1).

APPENDIX A: ELECTRIC FIELDS AND FIELD GRADIENTS: EXPLICIT EXPRESSIONS

Actual computations of the RF dipoles, e.g., by iterating the linear response equation system (41), require explicit expressions for the electric fields $\mathbf{E}_G^{q_j}(\mathbf{r}_i;\sigma)$ and $\mathbf{E}_G^{\mathbf{p}_j}(\mathbf{r}_i;\sigma)$ defined by Eqs. (43) and (45), respectively. For the field at $\mathbf{r}_i$ generated by a point charge q_j at $\mathbf{r}_j$ one obtains

$$\mathbf{E}_G^{q_j}(\mathbf{r}_i;\sigma)=\frac{1}{\varepsilon_s}\left[q_j\frac{\mathbf{r}_{ij}}{r_{ij}^3}\,\mathrm{erf}\left(\frac{r_{ij}}{\sqrt{2}\sigma}\right)-\sqrt{\frac{2}{\pi}}q_j\frac{\mathbf{r}_{ij}}{r_{ij}^2}\frac{1}{\sigma}\exp\left(-\frac{r_{ij}^2}{2\sigma^2}\right)\right], \tag{A1}$$

which in the large distance limit $r_{ij}\gg\sigma$ reduces to the point charge expression

$$\mathbf{E}_G^{q_j}(\mathbf{r}_i;\sigma)\to\frac{q_j}{\varepsilon_s}\frac{\mathbf{r}_{ij}}{r_{ij}^3}. \tag{A2}$$

This far field limit is important for the implementation of the approach into the SAMM algorithm (cf. Sec. III D), because it safely applies to almost all SAMM distance classes.[8,10] Solely for interactions among atoms within the innermost SAMM distance class ($r_{ij}<1$ nm) one has to employ the more complicated expression (A1). Similarly the field generated by a dipole density $\tilde{\mathbf{P}}_j^{\mathbf{E}}$ and acting on atom $i\neq j$ is given by

$$\mathbf{E}_G^{\mathbf{p}_j}(\mathbf{r}_i;\sigma)=\frac{1}{\varepsilon_s}\left\{\left[3\frac{\mathbf{p}_j^{\varepsilon}\mathbf{r}_{ij}}{r_{ij}^5}\mathbf{r}_{ij}-\frac{\mathbf{p}_j^{\varepsilon}}{r_{ij}^3}\right]\mathrm{erf}\left(\frac{r_{ij}}{\sqrt{2}\sigma}\right)-\sqrt{\frac{2}{\pi}}\frac{1}{\sigma^3}\left(\frac{\mathbf{p}_j^{\varepsilon}\mathbf{r}_{ij}}{r_{ij}^2}\mathbf{r}_{ij}+3\frac{\mathbf{p}_j^{\varepsilon}\mathbf{r}_{ij}}{r_{ij}^4}\mathbf{r}_{ij}\sigma^2-\frac{\mathbf{p}_j^{\varepsilon}}{r_{ij}^2}\sigma^2\right)\times\exp\left(-\frac{r_{ij}^2}{2\sigma^2}\right)\right\} \tag{A3}$$

and, in the large distance limit, reduces to the field of point dipoles $\mathbf{p}_j^{\varepsilon}$, i.e.,

$$\mathbf{E}_G^{\mathbf{p}_j}(\mathbf{r}_i;\sigma)\to\frac{1}{\varepsilon_s}\left[3\frac{\mathbf{p}_j^{\varepsilon}\mathbf{r}_{ij}}{r_{ij}^5}\mathbf{r}_{ij}-\frac{\mathbf{p}_j^{\varepsilon}}{r_{ij}^3}\right]. \tag{A4}$$

To compute the *reactio* forces acting on the RF dipoles from Eqs. (73) and (74), one additionally has to specify the field gradients

$$[\nabla_{\mathbf{r}_i} \otimes \mathbf{E}_G^{q_j}(\mathbf{r}_i;\sigma)] = -\frac{q_j}{\varepsilon_s}\left\{\left[3\frac{\mathbf{r}_{ij}\otimes\mathbf{r}_{ij}}{r_{ij}^5} - \frac{\mathbb{1}}{r_{ij}^3}\right]\mathrm{erf}\left(\frac{r_{ij}}{\sqrt{2}\sigma}\right) - \sqrt{\frac{2}{\pi}}\frac{1}{\sigma^3}\left[\frac{\mathbf{r}_{ij}\otimes\mathbf{r}_{ij}}{r_{ij}^2} + 3\frac{\mathbf{r}_{ij}\otimes\mathbf{r}_{ij}}{r_{ij}^4}\sigma^2 - \frac{\mathbb{1}}{r_{ij}^2}\sigma^2\right] \times\exp\left(-\frac{r_{ij}^2}{2\sigma^2}\right)\right\}, \tag{A5}$$

where $\mathbb{1}$ denotes the identity matrix. The associated far field limit is

$$[\nabla_{\mathbf{r}_i} \otimes \mathbf{E}_G^{q_j}(\mathbf{r}_i;\sigma)] \to -\frac{q_j}{\varepsilon_s}\left[3\frac{\mathbf{r}_{ij}\otimes\mathbf{r}_{ij}}{r_{ij}^5} - \frac{\mathbb{1}}{r_{ij}^3}\right]. \tag{A6}$$

APPENDIX B: EXPLICIT EXPRESSIONS FOR SCREENING FUNCTIONS

The screening functions $f^q(r;\sigma)$ and $f^{\mathbf{p}}(r;\sigma;\rho)$ occurring in Eqs. (62) and (63), respectively, are given by

$$f^q(r;\sigma) = 1 - \exp\left(-\frac{r^2}{2\sigma^2}\right) + r\sqrt{\frac{\pi}{2\sigma^2}}\left[1 - \mathrm{erf}\left(\frac{r}{\sqrt{2}\sigma}\right)\right] \tag{B1}$$

and by

$$f^{\mathbf{p}}(r;\sigma;\rho) = \mathrm{erf}\left(\frac{r}{\sqrt{2}\rho}\right) - \sqrt{1+\frac{\rho^2}{\sigma^2}}\,\mathrm{erf}\left(\frac{r}{\sqrt{2}\rho}\sqrt{\frac{\sigma^2}{\sigma^2+\rho^2}}\right) \times\exp\left(-\frac{r^2}{2(\sigma^2+\rho^2)}\right). \tag{B2}$$

[1] A. Warshel and S. T. Russell, Q. Rev. Biophys. **17**, 283 (1984).
[2] W. F. van Gunsteren and H. J. C. Berendsen, Angew. Chem., Int. Ed. Engl. **29**, 992 (1990).
[3] M. Karplus and J. A. McCammon, Nat. Struct. Biol. **9**, 646 (2002).
[4] J. W. Ponder and D. A. Case, Adv. Protein Chem. **66**, 27 (2003).
[5] P. Tavan, H. Carstens, and G. Mathias, in *Protein Folding Handbook*, edited by J. Buchner and T. Kiefhaber (Wiley-VCH, Weinheim, 2005), Pt. 1, pp. 1170–1195.
[6] T. A. Darden, D. York, and L. Pedersen, J. Chem. Phys. **98**, 10089 (1993).
[7] B. A. Luty, I. G. Tironi, and W. F. van Gunsteren, J. Chem. Phys. **103**, 3014 (1995).
[8] C. Niedermeier and P. Tavan, J. Chem. Phys. **101**, 734 (1994).
[9] C. Niedermeier and P. Tavan, Mol. Simul. **17**, 57 (1996).
[10] G. Mathias, B. Egwolf, M. Nonella, and P. Tavan, J. Chem. Phys. **118**, 10847 (2003).
[11] W. Weber, P. H. Hünenberger, and J. A. McCammon, J. Phys. Chem. B **104**, 3668 (2000).
[12] G. Mathias and P. Tavan, J. Chem. Phys. **120**, 4393 (2004).
[13] O. Bieri, J. Wirz, B. Hellrung, M. Schutkowski, M. Drewello, and T. Kiefhaber, Proc. Natl. Acad. Sci. U.S.A. **96**, 9597 (1999).
[14] S. Spörlein, H. Carstens, H. Satzger, C. Renner, R. Behrendt, L. Moroder, P. Tavan, W. Zinth, and J. Wachtveitl, Proc. Natl. Acad. Sci. U.S.A. **99**, 7998 (2002).
[15] F. Krieger, B. Fierz, O. Bieri, M. Drewello, and T. Kiefhaber, J. Mol. Biol. **332**, 265 (2003).
[16] B. Egwolf and P. Tavan, J. Chem. Phys. **118**, 2039 (2003).
[17] W. C. Still, A. Tempczyk, R. C. Hawley, and T. Hendrickson, J. Am. Chem. Soc. **112**, 6127 (1990).
[18] D. Bashford and A. Case, Annu. Rev. Phys. Chem. **51**, 129 (2000).
[19] C. D. Snow, N. Nguyen, V. S. Pande, and M. Gruebele, Nature (London) **420**, 102 (2002).
[20] R. Zhou and B. J. Berne, Proc. Natl. Acad. Sci. U.S.A. **99**, 12777 (2002).
[21] H. Nymeyer and A. E. García, Proc. Natl. Acad. Sci. U.S.A. **100**, 13934 (2003).
[22] B. Egwolf and P. Tavan, J. Chem. Phys. **120**, 2056 (2004).
[23] M. K. Gilson, M. E. Davis, B. A. Luty, and J. A. McCammon, J. Phys. Chem. **97**, 3591 (1993).
[24] A clear discussion of the dielectric boundary pressure and of its derivation either through Maxwell's stress tensor within liquid dielectrics or through the variation of the free energy is given in L. D. Landau and E. M. Lifschitz, *Elektrodynamik der Kontinua* (Akademie-Verlag, Berlin, 1985), Chap. 15.
[25] N. Levy, D. Borgis, and M. Marchi, Comput. Phys. Commun. **169**, 69 (2005).
[26] A. D. MacKerell, Jr., D. Bashford, M. Bellott *et al.*, J. Phys. Chem. B **102**, 3586 (1998).
[27] W. F. van Gunsteren, S. R. Billeter, A. A. Eising, P. H. Hünenberger, P. Krüger, A. E. Mark, W. R. P. Scott, and I. G. Tironi, *Biomolecular Simulation: The GROMOS96 Manual and User Guide* (Vdf Hochschulverlag AG an der ETH Zürich, Zürich, 1996).
[28] M. Kloppenburg and P. Tavan, Phys. Rev. E **55**, 2089 (1997).
[29] J. D. Jackson, *Classical Electrodynamics* (Wiley, New York, 1975).
[30] P. Ahlström, A. Wallquist, S. Engström, and B. Jönsson, Mol. Phys. **68**, 563 (1989).
[31] M. Born, Z. Phys. **1**, 45 (1920).
[32] J. G. Kirkwood, J. Chem. Phys. **2**, 351 (1934).
[33] T. M. Nymand and P. Linse, J. Chem. Phys. **112**, 6152 (2000).
[34] I. Bronstein and K. Semendjajew, *Taschenbuch der Mathematik* (Harri Deutsch, Frankfurt, 1991), p. 768.
[35] R. Remmert, *Funktionentheorie I* (Springer, Berlin, 1989).
[36] P. Höchtl, S. Boresch, W. Bitomsky, and O. Steinhauser, J. Chem. Phys. **109**, 4927 (1998).
[37] S. A. Hassan and E. L. Mehler, Proteins **47**, 45 (2002).
[38] H. Hu, M. Elstner, and J. Hermans, Proteins **50**, 451 (2003).
[39] M. Marchi, D. Borgis, N. Levy, and P. Ballone, J. Chem. Phys. **114**, 4377 (2001).
[40] F. Fogolari, G. Esposito, P. Viglino, and H. Molinari, J. Comput. Chem. **22**, 1 (2001).

4.2 MD-Simulationen mit der modifizierten Kontinuumsmethode

Es folgt nun ein Abdruck[1] des Artikels

> Martina Stork and Paul Tavan:
> „Electrostatics of proteins in dielectric solvent continua:
> II. First applications in molecular dynamics simulations."
> *Journal of Chemical Physics* 126, 165106/1-13 (2007),

der den Folgeartikel zu der in Abschnitt 4.1 abgedruckten Veröffentlichung darstellt. Hier wird der Einsatz der Kontinuumsmethode in MD-Simulationen mit dem Programm EGO-MMII beschrieben. Nach einer Parametrisierung mit Hilfe des Prion-Protein-Systems aus Abschnitt 2 wird die Methode am Alanin-Dipeptid getestet. Dabei wird das effiziente *Extended Lagrange*-Verfahren zur Berechnung der RF-Dipole eingesetzt.

[1] Mit freundlicher Genehmigung des American Institute of Physics.

Electrostatics of proteins in dielectric solvent continua. II. First applications in molecular dynamics simulations

Martina Stork and Paul Tavan[a)]
Theoretische Biophysik, Lehrstuhl für BioMolekulare Optik, Ludwig-Maximilians-Universität München, Oettingenstrasse 67, D-80538 München, Germany

(Received 13 November 2006; accepted 2 March 2007; published online 27 April 2007)

In the preceding paper by Stork and Tavan, [J. Chem. Phys. **126**, 165105 (2007)], the authors have reformulated an electrostatic theory which treats proteins surrounded by dielectric solvent continua and approximately solves the associated Poisson equation [B. Egwolf and P. Tavan, J. Chem. Phys. **118**, 2039 (2003)]. The resulting solution comprises analytical expressions for the electrostatic reaction field (RF) and potential, which are generated within the protein by the polarization of the surrounding continuum. Here the field and potential are represented in terms of Gaussian RF dipole densities localized at the protein atoms. Quite like in a polarizable force field, also the RF dipole at a given protein atom is induced by the partial charges and RF dipoles at the other atoms. Based on the reformulated theory, the authors have suggested expressions for the RF forces, which obey Newton's third law. Previous continuum approaches, which were also built on solutions of the Poisson equation, used to violate the *reactio* principle required by this law, and thus were inapplicable to molecular dynamics (MD) simulations. In this paper, the authors suggest a set of techniques by which one can surmount the few remaining hurdles still hampering the application of the theory to MD simulations of soluble proteins and peptides. These techniques comprise the treatment of the RF dipoles within an extended Lagrangian approach and the optimization of the atomic RF polarizabilities. Using the well-studied conformational dynamics of alanine dipeptide as the simplest example, the authors demonstrate the remarkable accuracy and efficiency of the resulting RF-MD approach. © *2007 American Institute of Physics*. [DOI: 10.1063/1.2720389]

I. INTRODUCTION

As outlined in the Introduction of the preceding paper, which we will refer from now on as "Paper I,"[1] much longer time scales could be covered by molecular dynamics (MD) simulations of biological macromolecules in solution if accurate and computationally tractable implicit solvent models were available. In such hybrid approaches, only the solute molecule is treated at atomic resolution, whereas the solvent is represented by a dielectric and possibly also ionic continuum. Within corresponding MD simulations, one has to solve a partial differential equation (PDE), like, e.g., the Poisson equation or the linearized Poisson-Boltzmann equation, on the fly with the integration of the protein dynamics. If one tries to avoid this complicated task by resorting to the simple "generalized Born" approximation,[2] one can cover impressive time scales[3] with MD but one gets free energy landscapes, which drastically differ from those obtained with explicit solvent.[4,5] If one tries to tackle this task by standard numerical methods for solving PDEs, one is confronted with an intractable computational effort.[6]

In that respect it was fortunate that Egwolf and Tavan (ET) succeeded in deriving, by means of analytical mathematics, a conceptually different approach for solving the Poisson and linearized Poisson-Boltzmann equations,[6,7] which led to highly efficient numerical algorithms. For a given protein configuration $\mathbf{R} \equiv (\mathbf{r}_1, \ldots, \mathbf{r}_N)$, the method yields the electric field $\mathbf{E}(\mathbf{r}_i)$ at the positions $\mathbf{r}_i$ of the protein atoms i and the total electrostatic energy $W(\mathbf{R})$ of the system. In addition to the Coulomb interactions among the protein charges, the total electrostatic field $\mathbf{E}$ and energy W also cover the so-called reaction field (RF) contributions $\mathbf{E}^{\varepsilon}$ and W^{ε}, respectively, which describe the effect of the continuum polarization on the protein charge distribution. As is well known, for each protein configuration $\mathbf{R}$, the RF contribution $W^{\varepsilon}(\mathbf{R})$ to the total electrostatic energy $W(\mathbf{R})$ is, up to surface terms associated with the hydrophobic effect and with van der Waals interactions, the free energy of solvation.[6] With respect to changes of that configuration, $W^{\varepsilon}(\mathbf{R})$ is thus a potential of mean force.[1]

According to the ET theory[6,7] and its reformulation in Paper I,[1] which, for the sake of clarity, was restricted to the formally somewhat simpler Poisson case, the polarization of the dielectric continuum surrounding a protein can be *exactly* represented by dipole densities $\mathbf{P}_i^{\mathbf{E}}(\mathbf{r})$ localized at the protein atoms. *Approximating* these densities by Gaussian models with amplitudes $\mathbf{p}_i$ and widths σ_i leads to a "linear response" equation system for the amplitudes

$$\mathbf{p}_i = -\alpha_i \mathbf{E}_{\mathrm{pol}}^{q,\mathbf{p}}(\mathbf{r}_i), \tag{1}$$

which has to be self-consistently solved, because the electric field $\mathbf{E}_{\mathrm{pol}}^{q,\mathbf{p}}$ polarizing the protein atoms through their RF polarizabilities α_i depends not only on the charges q_j but also on the amplitudes $\mathbf{p}_j$ of the Gausssian dipole densities at the

[a)]Author to whom correspondence should be addressed. Electronic mail: tavan@physik.uni-muenchen.de

other atoms $j \neq i$ [cf. Eqs. (50) and (51) of Paper I]. As explained in Paper I, the RF polarizabilities α_i are functions of the dielectric constants ε_s and ε_c chosen for the protein volume and for the surrounding dielectric continuum, respectively (e.g., with the choices $\varepsilon_s=1$ and $\varepsilon_c=80$ for a protein in water at room temperature and ambient pressure). The α_i furthermore depend on the Gaussian widths σ_i through so-called effective atomic volumes v_i, which also have to be determined by a self-consistency iteration. This volume iteration is defined by

$$v_i^{(n)} = v_i^{(n-1)} \left[\sum_j v_j^{(n-1)} G(r_{ij};\sigma_j) \right]^{-1}, \tag{2}$$

where G denotes a normal distribution and r_{ij} the distance between the protein atoms i and j. Here, the Gaussian "overlaps" $G(r_{ij};\sigma_j)$ measure the values of normal distributions centered around positions $\mathbf{r}_j$ at the position of atom i.

Inspection of Eq. (1) immediately shows that the amplitudes $\mathbf{p}_i$, which are called "RF dipoles," follow from the electric field $\mathbf{E}_{\mathrm{pol}}^{q,\mathbf{p}}$ nearly in the same way as induced dipoles in a polarizable force field (see, e.g., Ref. 8). The important difference is the negative sign, which reflects the fact that the atomic RF dipoles are images of the polarization within the surrounding continuum. Once the effective volumes v_i and the RF dipoles $\mathbf{p}_i$ are known, the RF contributions $\mathbf{E}^{\varepsilon}$ to the field and the solvation free energy W^{ε} can be approximately calculated from analytical formulas [cf. Eqs. (52) and (60) of Paper I].

Because the resulting approximate solvation free energy $\tilde{W}^{\varepsilon}$ is composed of pair interactions among the atomic charges and RF dipoles depending on the distance vectors $\mathbf{r}_{ij}$, the corresponding mean RF forces $-\nabla_{\mathbf{r}_i}\tilde{W}^{\varepsilon}$ act both on charged and uncharged protein atoms and obey Newton's third law. However, this approach to the force computation leads neither to an efficient nor to a very accurate algorithm. In contrast, the resulting approximation $\tilde{\mathbf{E}}^{\varepsilon}$ for the electric field is quite accurate, can be evaluated much more efficiently, and can be used for a construction of RF forces compatible with Newton's third law.[1] Such a construction is necessary, because the simple idea of obtaining the mean RF forces by multiplying the atomic charges q_i with the local fields $\tilde{\mathbf{E}}^{\varepsilon}(\mathbf{r}_i)$ yields instead the so-called qE forces, which solely act on the charged atoms and violate Newton's *reactio* principle.[9] According to that principle, the surrounding continuum exerts a "dielectric surface pressure" on the protein counterbalancing the qE forces. Thus, our construction of approximate mean RF forces [cf. Eqs. (70)–(75) of Paper I], which obey Newton's third law and act both on charged and uncharged atoms, automatically accounts for the required counterbalance.

With the thus accomplished "marriage" between Newton's third law and RF forces obtained from an approximate but numerically efficient solution of the Poisson equation, the main hurdle toward MD simulations with implicit solvent appears to be surmounted. However, as outlined in Sec. III F of Paper I, a few technical points still need to be addressed before the viability of the approach can be evaluated. These points cover the following questions: (i) How should one include the required self-consistent evaluations of the effective atomic volumes v_i [Eq. (2)] and of the RF dipoles $\mathbf{p}_i$ [Eq. (1)] into the integration of the Newtonian dynamics of atomic motion? (ii) How should one choose the widths σ_i of the Gaussian atomic dipole densities $\tilde{\mathbf{P}}_i^{\mathbf{E}}(\mathbf{r})$ for the various "atom types" occurring in a protein? Here, as is common in molecular mechanics (MM) force fields, the notion of atom type refers to an atom of given chemical identity covalently linked to specific other atoms in a well-defined stereochemical arrangement. This question is important, because the RF polarizabilities α_i of the atoms depend on the chosen σ_i through the v_i [cf. Eqs. (46)–(48) of Paper I]. Considering the local character of the interactions among the v_i revealed by Eq. (2) leads to the conclusion that the RF polarizabilities α_i are shaped for given σ_i by the respective local covalent bonding patterns. Therefore, the σ_i represent atom type parameters of the method and one has to provide procedures for an optimal choice.

In the following, we will first address the algorithmic question (i) and will subsequently characterize in Sec. III the methods employed for the optimization of σ_i. Because this optimization will rely on comparisons with MD simulations of proteins and peptides embedded in explicit water, we will also sketch the corresponding simulation systems and techniques. Additionally, this section will explain the procedures employed in our testing scenario for the implicit solvent MD approach, which mainly deals with the free energy landscape of alanine dipeptide in solution.

II. EXTENDED LAGRANGE METHOD FOR UPDATING RF DIPOLES

Using Eq. (2) the effective atomic volumes v_i entering the ET approach have to be self-consistently determined for each protein configuration $\mathbf{R}$. Because this configuration changes at each MD integration step T ($T=1,2,\ldots$), the v_i have to be continuously updated. Experience has shown that the volume iteration [Eq. (2)] exhibits a rapid convergence and that the v_i vary only a little with the change from $\mathbf{R}(T)$ to $\mathbf{R}(T+1)$. Therefore, only one iteration step suffices to keep the v_i very close to their self-consistent values, if one uses the $v_i(T)$ resulting at integration step T as starting values for the volume computations at step $T+1$. Furthermore, Eq. (2) shows that the v_i are determined by Gaussian overlaps between covalently bound or neighboring atoms. Due to the local character of the Gaussians, only a small set of neighboring atoms has to be considered for the computation of each v_i and this algorithmic step can be easily integrated into distance class schemes for the calculation of electrostatic forces (like the ones employed in particle-mesh Ewald[10] or in structure-adapted multipole[11] methods). As a result, the computational effort required for the calculation of the v_i can be kept very small and the atomic RF polarizabilities α_i are readily determined from the v_i through Eqs. (46)–(48) of Paper I.

In contrast, the iteration of the RF dipoles $\mathbf{p}_i$, which is required by the self-consistent field (SCF) condition (1) for each configuration $\mathbf{R}(T)$, exhibits a much slower convergence. Moreover, the differences $|\mathbf{p}_i(T+1)-\mathbf{p}_i(T)|$ are not al-

ways small. Therefore, at each MD integration step, many iterations may be necessary to reach convergence, even if the results $\mathbf{p}_i(T)$ of the preceding MD integration step are employed for the computation of $\mathbf{p}_i(T+1)$. The attempt of keeping the RF dipoles at their SCF values is, thus, computationally quite expensive.

More importantly, however, the attempt of keeping the RF dipoles at their SCF values is hardly compatible with the physical meaning of these variables: According to the ET theory, the RF dipoles are images of the orientational polarization induced by the protein charges within a surrounding polar solvent. As is well known, the dielectric relaxation times of polar liquids range from a few hundred femtoseconds up to several picoseconds.[12] These times are much longer than the integration time steps of about 1 fs, which are usually employed in MD simulations for a smooth sampling of the very fast bond vibrations. Correspondingly, the orientational polarization of the liquid represents an average over the structures sampled by a protein on a femtosecond time scale and can respond only to much slower conformational changes. RF dipoles, which by the SCF requirement are forced to follow the femtosecond protein motions, exhibit correspondingly fast fluctuations and, therefore, do not adequately model the more slowly fluctuating orientational solvent polarization.

For all these reasons, we have additionally selected a second approach to dynamically adjust the dipoles within MD, which is the extended Lagrangian (EL) method.[13] By adopting this method, the RF dipoles can be treated as dynamical variables whose speed of adjustment can be steered by assigning suitable "masses." The EL method is frequently used in polarizable force fields to update the variables describing the polarization.[14–19] If these variables are inducible atomic dipoles (see, e.g., Ref. 16), the dipoles are allowed to fluctuate within a potential U, whose minimum is marked by their SCF values.

In analogy to the latter case, we thus define the potential function

$$U = \sum_i \left\{ \frac{\mathbf{p}_i^2}{2\alpha_i} + \mathbf{p}_i \mathbf{E}_{\mathrm{pol}}^{q,\mathbf{p}}(\mathbf{r}_i) \right\}, \tag{3}$$

whose gradients are given by

$$\frac{\partial U}{\partial \mathbf{p}_i} = \frac{\mathbf{p}_i}{\alpha_i} + \mathbf{E}_{\mathrm{pol}}^{q,\mathbf{p}}(\mathbf{r}_i). \tag{4}$$

Whenever the gradients vanish, the SCF condition (1) is fulfilled. The equations of motions for the dipoles are then

$$\hat{m}_i \frac{d^2\mathbf{p}_i}{dt^2} = -\left[\frac{\mathbf{p}_i}{\alpha_i} + \mathbf{E}_{\mathrm{pol}}^{q,\mathbf{p}}(\mathbf{r}_i) \right], \tag{5}$$

where the $\hat{m}_i$ are the above mentioned masses of the reaction field dipoles. The $\hat{m}_i$ have the dimension amu/e^2. Because these masses merely determine how quickly the dipoles react to a change of the coordinates, they can be replaced by the global relaxation time $\tau_p = \sqrt{\hat{m}_i \alpha_i}$, which should be chosen such that it matches the average dielectric relaxation time of the modeled solvent at the given thermodynamic conditions.

The equations of motion (5) can now be numerically integrated by the Verlet algorithm[20]

$$\mathbf{p}_i(t+\Delta t) = 2\mathbf{p}_i(t) - \mathbf{p}_i(t-\Delta t) - [\mathbf{p}_i(t) + \alpha_i \mathbf{E}_{\mathrm{pol}}^{q,\mathbf{p}}(\mathbf{r}_i,t)] \frac{\Delta t^2}{\tau_p^2} \tag{6}$$

in parallel with the integration of the atomic motion. The overall movement of the dipoles can be monitored by defining a dipole temperature T_p. For this purpose, we define by

$$E_{\mathrm{kin}} = \frac{\tau_p^2}{2} \sum_i \frac{1}{\alpha_i} \left(\frac{\mathbf{p}_i(t) - \mathbf{p}_i(t-\Delta t)}{\Delta t} \right)^2 \tag{7}$$

the kinetic energy of the dipoles. Making use of the equipartition theorem

$$E_{\mathrm{kin}} = \frac{f_p}{2} N k_B T_p, \tag{8}$$

where $f_p = 3$ is the number of degrees of freedom per dipole and k_B is the Boltzmann constant, we obtain the dipole temperature

$$T_p = \frac{\tau_p^2}{3Nk_B} \sum_i \frac{1}{\alpha_i} \left(\frac{\mathbf{p}_i(t) - \mathbf{p}_i(t-\Delta t)}{\Delta t} \right)^2. \tag{9}$$

Due to the coupling of the atomic and dipolar motions, the associated dynamical subsystems will have, on the average, the same temperature. For temperature control, one can, thus, either couple the atoms or the dipoles or both to a heat bath using, e.g., a Berendsen thermostat.[21] If one decides to couple also the dipoles to the heat bath, one may choose a coupling time constant $\tau_{p,T}$ different from the one selected for the atoms (τ_T) to keep the system near the joint target temperature T_0.

The SCF and EL approaches outlined above have been implemented in our parallelized MD program EGO-MMII.[22] The Appendix sketches the flow of computational steps characterizing this implementation. With such an implementation at hand, a MM/RF-MD simulation approach is, in principle, available. In practice, however, the question as to how one should choose the widths σ_i characterizing the Gaussian RF dipole density remains to be answered [cf. question (ii) further above].

III. METHODS

In their original paper, Egwolf and Tavan had suggested choosing the Gaussian widths σ_i of the RF dipole densities proportional to the so-called van der Waals radii R_i of the atoms, because the choice of the σ_i should be guided by the "sizes" of the atoms.[6] Common MM force fields such as CHARMM (Refs. 23 and 24) provide such radii R_i as parameters for the Lennard-Jones potentials modeling the van der Waals interactions. Egwolf and Tavan apparently assumed that these parameters measure something like the effective sizes of the atoms within molecules or, more specifically, within particular chemical binding motifs. However, a close inspection, e.g., of the CHARMM22 force field[24] reveals that this assumption is misleading. For instance, radii as small as 0.2 Å are assigned to the so-called polar hydrogen atoms. In

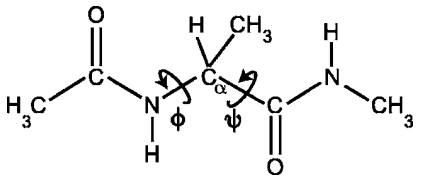

FIG. 1. Chemical structure of alanine dipeptide (AlDi). The angles ϕ and ψ at the central C_α atom are the two important conformational degrees of freedom of this minimal peptide model.

optimizations of empirical force fields such counterintuitive values can result, because they happened to enable a satisfactory description of the protein building blocks containing polar hydrogens. But because of the complex interdependencies among the parameters, the results of force field optimizations are by no means unique. For instance, the older CHARMM19 force field employed larger radii of 0.8 Å for the polar hydrogens.[23] Hence, the parameters R_i listed in a given force field do not necessarily represent good measures of "atom sizes."

A. Scenario for optimization and testing

Because atom sizes are not easily defined for atoms in molecules, we chose to consider the Gaussian widths σ_i as free parameters of the ET theory and, correspondingly, set up an optimization procedure for their choice. Here, we decided that comparisons between the RF forces and the *mean forces* exerted in MD simulations by an explicitly modeled solvent on the atoms of a solute protein should be a relevant measure.

In addition to the construction of a corresponding optimization setup, a testing procedure had to be designed, which can answer the question of whether the optimized atomic parameters σ_i are *transferable* to other systems. Note in this context that one generally has to group the atoms into classes whenever one wants to determine transferable atomic parameters for a MM force field. These classes are called atom types and comprise atoms in well-defined chemical bonding motifs (such as, e.g., hydrogens in CH_3 groups). To what extent the atom types chosen in a given force field for specific purposes (e.g., to enable the definition of transferable parameters for the so-called bonded interactions) suitably classify the atoms also for the computation of RF forces is *a priori* unclear and remains to be checked.

For our optimization setup, we selected the C-terminal part of the cellular prion protein (PrP^C) in aqueous solution as a sample protein because a PrP^C model was at hand from a previous MD study,[25] it contains many of the CHARMM22 atom types in reasonable amounts, and it is a typical representative for small, globular, and soluble proteins. To evaluate the transferability of the parameters and the quality of the associated MM/RF-MD method, we selected alanine dipeptide (AlDi) as our example because this small peptide, whose chemical structure is shown in Fig. 1, has only two essential degrees of freedom, its free energy landscape has been frequently studied,[24,26–31] and it is accessible to MD simulations with explicit solvent at a moderate computational effort. AlDi should be a particularly hard testing case for RF-MD, because its few atoms provide a very coarse discretization for the RF dipole density and the ET approach is expected to work better for larger molecules.[6]

B. Simulation setups

PrP^C covering residues 125–228 was taken from entry 1qm2 (Ref. 32) of the Protein Data Bank. AlDi was constructed by the program MOLDEN.[33] Water molecules were modeled by the CHARMM22 variant of the TIP3P potential.[24,34] For simulations with explicit solvent, the respective solute was placed into a periodic rhombic dodecahedron filled with TIP3P water molecules (PrP^C: 25 775 water molecules, AlDi: 1011 water molecules). The size of the system was chosen such that it encloses a sphere with a radius R, which is larger than the diameter of the solute (PrP^C: $R=51.4$ Å, AlDi: $R=18.9$ Å). All simulations were carried out with the program package EGO-MMII, in which the Coulomb interactions are treated by a particular combination of structure-adapted multipole expansions[11] with a moving boundary reaction field approach and a multiple time-step integrator.[22,35] van der Waals interactions were calculated explicitly up to 10 Å; at larger distances, a mean field approach[36] was applied. The covalent bonds of the hydrogen atoms were always kept rigid by the M-SHAKE procedure (with a relative tolerance of 10^{-6}).[37] The integration time step was set to 1 fs.

PrP^C in TIP3P water was equilibrated for several nanoseconds in the NpT ensemble. Here, the temperature T and pressure p were controlled by Berendsen thermo- and barostats,[21] respectively, using the relaxation times $\tau_T=0.5$ ps and $\tau_b=1.0$ ps as well as the value $\beta=5.0$ Pa^{-1} for the isothermal compressibility of water. With these parameters, the target values $T_0=300$ K and $P_0=1.0\times10^5$ Pa were closely and stably assumed. After equilibration we switched to the NVT ensemble for data acquisition.

In the case of AlDi in TIP3P water, the same procedure was carried out for an initial equilibration at 300 K and normal pressure. Using the TIP3P thermal expansion coefficient[38] of 1.4×10^{-3}/K, we calculated the volume V_{500}, which TIP3P water assumes at $T=500$ K if the liquid state is maintained. The system was then expanded to $V=V_{500}$ and concurrently heated to $T=500$ K. At these parameters, it was equilibrated for several nanoseconds by an NVT simulation.

During RF-MD simulations of AlDi ($\varepsilon_s=1$, $\varepsilon_c=80$) and during MD simulations of this peptide in vacuum ($\varepsilon_s=\varepsilon_c=1$), the atoms were weakly coupled ($\tau_T=1$ ps) by a Berendsen thermostat to a heat bath at $T_0=500$ K. When using the EL technique for the dipole dynamics in RF-MD simulations, such a thermostat was additionally applied to the dipole temperature using the same parameters. In these EL simulations, the global relaxation time τ_p of the dipoles was chosen as 2.5 ps although smaller values in the range between 0.1 and 1 ps might be more realistic estimates for the dielectric relaxation time of water at the given conditions.

At the fixed starting configuration of the peptide, initial values of the RF dipoles and effective volumes were generated by iterating the volume and dipole equations (2) and (1), respectively, until self-consistency was reached (starting at

$v_i=1$ and $p_i=0$ for all atoms i). As convergence measures, we used relative changes from iteration step n to step $n+1$, which for the v_i are given by the average

$$\Delta v = \frac{1}{N}\sum_i \frac{|v_i^n - v_i^{n+1}|}{v_i^{n+1}}, \tag{10}$$

and for the RF dipoles by

$$\Delta p = \max_i \frac{|\mathbf{p}_i^n - \mathbf{p}_i^{n+1}|}{|\mathbf{p}_i^{n+1}|}. \tag{11}$$

As soon as these relative changes fell below predefined thresholds η_v and η_p, respectively, the iterations were stopped. During the start-up phase of a simulation, the thresholds were set to 10^{-6}, and it turned out that for the volumes, convergence was reached after about 250 iteration steps, whereas for the dipoles, 1000 steps were required. The results were employed as starting values $v_i(0)$ and $\mathbf{p}_i(0)$ for the EL- and SCF-type simulations. During these simulations, the threshold η_v was relaxed to a value of 10^{-3} and the converged volumes $v_i(T)$ of a preceding dynamics step T were employed as initial values for the volume iteration (2) in a subsequent dynamics step $T+1$. Because of the rapid convergence of Eq. (2), only one iteration step had to be carried out to maintain $\Delta v < \eta_v$. To speed up the SCF-type simulations, the RF dipole threshold was relaxed to an even larger value $\eta_p=10^{-2}$, and by using the dipoles $\mathbf{p}_i(T)$ as initial values, only at most five iterations (1) were required to reach the corresponding convergence. For the simplest implementation, the number of iterations was therefore set to the constant value of 5. In all RF and vacuum simulations, the peptide was equilibrated at $T=500$ K for several nanoseconds before data acquisition.

C. Parameter optimization

To realize the optimization scenario outlined in Sec. III A for the parameters σ_i of the ET theory, the mean forces exerted by the TIP3P water molecules on the atoms of PrP^C and the corresponding RF forces had to be calculated. For this purpose, we arbitrarily selected a single PrP^C configuration $\mathbf{R}_0$, namely, the one found after equilibration. All PrP^C atoms were restrained to their positions in $\mathbf{R}_0$ by applying harmonic potentials with a force constant of 20 kcal/(mol Å^2). From a subsequent 400 ps simulation of the constrained protein surrounded by water molecules moving freely in the *NVT* ensemble ($T=300$ K), we stored every 4 ps snapshots of the PrP^C configuration $\mathbf{R}_t$ and of the electrostatic forces $\mathbf{F}_i^e(t)$ exerted by the water molecules on the PrP^C atoms i. Then the ensemble averages $\langle \mathbf{F}_i^e \rangle$ over the 100 forces $\mathbf{F}_i^e(t)$ were calculated for each i. For the configuration $\mathbf{R}_0$, for each specific choice of the parameter set $\Sigma \equiv \{\sigma_i; i=1,\ldots,N\}$, and for each particular grouping $\mathcal{G}$ of the atoms into atom types, we calculated the corresponding forces $\mathbf{F}_i^i(\mathbf{R}_0;\Sigma,\mathcal{G})$ by our RF approach.[39]

As a quality measure for the parameter set Σ and for the atom grouping $\mathcal{G}$, we chose the average square difference

$$\Delta F(\Sigma,\mathcal{G}) \equiv \frac{1}{N}\sum_{i=1}^{N} [\langle \mathbf{F}_i^e \rangle - \mathbf{F}_i^i(\mathbf{R}_0;\Sigma,\mathcal{G})]^2 \tag{12}$$

between the average forces $\langle \mathbf{F}_i^e \rangle$ and the RF forces $\mathbf{F}_i^i(\mathbf{R}_0;\Sigma,\mathcal{G})$ on the PrP^C atoms i. In the computation of the latter forces, we used the values $\varepsilon_s=1$ and $\varepsilon_c=80$ for the dielectric constants within the protein and solvent, respectively, although one might argue that setting ε_c to the somewhat larger dielectric constant of TIP3P water at the given conditions[40] ($\varepsilon_c \approx 97$) would be more consistent.

The optimal parameter set Σ^{opt} and grouping $\mathcal{G}^{\mathrm{opt}}$ jointly mark the global minimum of the functional $\Delta F(\Sigma,\mathcal{G})$. In view of the highly nonlinear character of this functional and of the complicated mutual dependencies among the parameters, this optimal solution is not easily found. To obtain a first guess, we decided to choose a simple gradient descent procedure based on consecutive line searches. For a given grouping $\mathcal{G}$ and starting at an initial choice $\sigma_{i,0}=1.06$ Å for all atom types occurring in PrP^C, the σ_i were varied for a first atom group in steps of 0.03 Å covering a small range of ±0.06 Å around $\sigma_{i,0}$. The value $\sigma_{i,\min}$ with the smallest ΔF was chosen as the new parameter σ_i for the given atom group. This procedure was carried out for all other groups and was iterated over the set of the atom groups until ΔF changed by less than 0.005 [kcal/(mol Å)]^2 from one optimization cycle to the next. Very large and very small σ_i were avoided by constraining the values to the range between 0.6 and 1.6 Å. To check to what extent the thus obtained σ_i are independent of the serial order of the line searches, we carried out six independent optimizations with different serial orders of the atom groups. The values of ΔF found for different sequential line searches generally turned out to be quite similar, indicating that our gradient descent does not strongly depend on the chosen sequence. Correlation graphs (cf. Figs. 2 and 3 for examples) comparing the average forces $\langle \mathbf{F}_i^e \rangle$ with the RF forces $\mathbf{F}_i^i(\mathbf{R}_0;\Sigma,\mathcal{G})$ were additionally used to monitor the progress of the optimization.

The initially chosen grouping $\mathcal{G}^0$ had to differ from the CHARMM22 prescription $\mathcal{G}^{c22}$, because not all of the CHARMM22 atom types occur in statistically sufficient amounts in the PrP^C structure. For atoms of a given chemical identity and of rare atom type, we therefore combined the corresponding small groups into residual classes denoted as HR_{RF}, CR_{RF}, NR_{RF}, and SR_{RF}. The progress of the optimization showed that we had to split a few large groups into smaller groups to achieve better or more reasonable results. As our final choice, we selected the grouping $\hat{\mathcal{G}}^{\mathrm{opt}}$, which rendered the smallest ΔF among the few investigated groupings. Table I relates the thus determined grouping $\hat{\mathcal{G}}^{\mathrm{opt}}$ to the CHARMM22 choice $\mathcal{G}^{c22}$. The corresponding "optimal" parameter set $\hat{\Sigma}^{\mathrm{opt}}$ was obtained as an average over the six independent gradient descents carried out for $\hat{\mathcal{G}}^{\mathrm{opt}}$.

D. Transferability check

Similar to our above choice of an arbitrary PrP^C configuration as a training set for parameter optimization, we furthermore decided to take a single configuration of a very

TABLE I. Optimized atomic grouping $\hat{\mathcal{G}}^{\mathrm{opt}}$ and Gaussian widths $\hat{\sigma}^{\mathrm{opt}}$ compared with the grouping $\mathcal{G}^{\mathrm{c22}}$ and the widths $\sigma^{c22}=0.8R_i^{c22}$ taken directly from the CHARMM22 force field.

$\hat{\mathcal{G}}^{\mathrm{opt}}$	$\hat{\sigma}^{\mathrm{opt}}$ (Å)	σ^{c22} (Å)	$\mathcal{G}^{\mathrm{c22}}$
O_{RF}	1.25	1.36	O
OC_{RF}	0.84	1.36	OC
$OH1_{RF}$	0.87	1.42	OH1
H_{RF}	0.93	0.18	H
HP_{RF}	1.51	1.09	HP
HA_{RF}	1.41	1.06	HA, non-methyl
HM_{RF}	0.85	1.06	HA, methyl
HB_{RF}	0.87	1.06	H
HR_{RF}	1.61	0.72	HR1
		0.56	HR3
HC_{RF}	1.00	0.18	HC
CA_{RF}	0.71	1.60	CA
C_{RF}	1.22	1.61	C
$CT1_{RF}$	0.77	1.83	CT1
$CT2_{RF}$	0.64	1.75	CT2
$CT3_{RF}$	1.38	1.66	CT3
CC_{RF}	0.64	1.61	CC
CR_{RF}	1.16	1.83	CP1
		1.75	CP2, CP3
		1.45	CPH1, CPH2
$NH2_{RF}$	0.84	1.49	NH2
$NH1_{RF}$	1.06	1.49	NH1
$NC2_{RF}$	0.64	1.49	NC2
NR_{RF}	0.64	1.49	NH3, N, NR1, NR2
SR_{RF}	1.61	1.61	S
		1.59	SM

different system, i.e., of AlDi, for a first transferability check. Also here we chose an arbitrary configuration resulting from the equilibration of this small peptide in TIP3P water at 300 K, fixed it by applying harmonic constraints, and evaluated the average forces $\langle \mathbf{F}_i^e \rangle$ exerted by the TIP3P water molecules on the peptide atoms i using the procedure described further above. Furthermore, we evaluated for the given peptide configuration the RF forces $\mathbf{F}_i^i(\mathbf{R}_0; \hat{\Sigma}^{\mathrm{opt}}, \hat{\mathcal{G}}^{\mathrm{opt}})$. Also here correlation graphs were used as tools to study the transferability of the parameters from PrP^C to AlDi.

E. Generation of energy maps

Free energy landscapes $\Delta G(\phi, \psi)$ of AlDi were calculated by various methods (for the definition of the dihedral angles ϕ and ψ characterizing the conformations of this peptide, see Fig. 1). A first such landscape was obtained from a 50 ns simulation of the peptide in TIP3P water at $T=500$ K, during which the dihedral angles were stored every 500 fs. A histogram density estimate $p(\phi, \psi)$ was obtained from the resulting data set of (ϕ, ψ) angles by subdividing the (ϕ, ψ) space into 40^2 square bins. The free energy then is

$$\Delta G(\phi, \psi) = -k_B T \ln p(\phi, \psi) - G_0, \tag{13}$$

where k_B is the Boltzmann constant and G_0 is a constant serving to set the minimum of ΔG to zero. Because empty bins in the histogram density estimate $[p(\phi, \psi)=0]$ would lead to infinite free energies, an upper energy cutoff $\Delta G_{\max} = -k_B T \ln(1/M_{\max})$ was introduced, where $M_{\max}$ is the maximum bin count found in the histogram, and ΔG was set to $\Delta G_{\max}$ for all empty bins. In the same way, free energy maps were obtained from 500 ns RF-MD simulations of AlDi at 500 K using the EL and SCF algorithms and from a likewise extended MD simulation of AlDi in vacuum.

The availability of an analytical expression for the total electrostatic energy $W(\mathbf{R})$ of a protein in a continuum solvent [cf. Eq. (60) of Paper I] offers, in the simple case of AlDi, an alternative approach to $\Delta G(\phi, \psi)$. For this purpose, one simply has to replace in the MM force field $E(\mathbf{R})$ the Coulomb electrostatics $W^c(\mathbf{R})$ of the peptide by $W(\mathbf{R})$ and one obtains a free energy function $G^\varepsilon(\mathbf{R})$. Because the remaining degrees of freedom of the peptide are either very stiff (bond lengths, bond angles, and the dihedral angle around the central C–N bond) or very labile (torsions of the methyl groups), the free energy landscape $\Delta G(\phi, \psi)$ is, to a very good approximation, given by $G^\varepsilon(\phi, \psi)$ if the remaining degrees of freedom are chosen at their equilibrium values. This reasoning is the so-called saddle point approximation for the Boltzmann ensemble of the remaining degrees of freedom. Thus, $\Delta G(\phi, \psi)$ can be obtained by simply scanning the (ϕ, ψ) space in steps of 6° and by minimizing $G^\varepsilon(\mathbf{R})$ with respect to the remaining degrees of freedom. Technically this minimization was carried out by the following procedure: The peptide structure was constrained to a point (ϕ, ψ) by additional dihedral potentials, whose equilibrium values select this point and whose force constants were stepwise increased from 1 to 100 kcal/(mol rad^2). During this increase the peptide was allowed to move thermally in a 200 K heat bath for 40 ps using the force field extended by the RF forces [Eqs. (70)–(75) of Paper I]. Subsequently, the force constants were set to the very large values of 5000 kcal/(mol rad^2) and $G^\varepsilon(\mathbf{R})$ was minimized for 10 ps by velocity scaling. The same procedure was also applied to scan the potential energy surface $E(\phi, \psi)$ of AlDi, to which $G^\varepsilon(\phi, \psi)$ reduces for $\varepsilon_s = \varepsilon_c = 1$, because then $W^c = W$.

IV. RESULTS AND DISCUSSION

Before one can apply the revised ET approach to MM/RF-MD simulations of soluble proteins and peptides, the parameters σ_i of the theory have to be specified. Adopting the strategy commonly used in the construction of transferable protein force fields, the choice of these parameters requires a classification of the atoms occurring in a protein into atom types, which are defined in terms of local covalent bonding motifs. Such a strategy appears to be natural for the choice of the σ_i, because these parameters enter the theory mainly through the overlaps of the atomic Gaussian distributions [cf., e.g., Eq. (2)] and because sizable overlaps occur only for covalently bonded atoms (if the σ_i are of atomic size).

A. Choice of the parameters

Following the suggestion in Ref. 6, we initially tried to simply adopt the atom classification $\mathcal{G}^{\mathrm{c22}}$, which is defined in the CHARMM22 force field[24] employed by us, and to choose the parameters σ_i according to

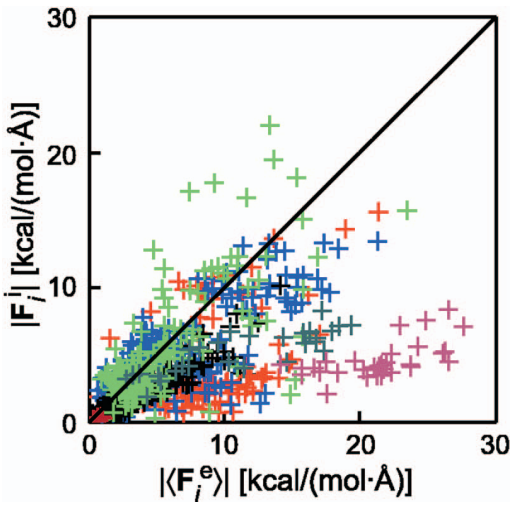

FIG. 2. (Color) Correlation of the absolute values of the mean electrostatic forces $\langle \mathbf{F}_i^e \rangle$ exerted by TIP3P water on the PrP^C atoms with the absolute values of the associated RF forces $\mathbf{F}_i^i(c_R=0.8)$ parametrized using Eq. (14), the CHARMM22 van der Waals radii R_i^{c22}, and atom grouping $\mathcal{G}^{c22}$. The color code generally refers to the chemical type of the respective atom: H, blue; O, red; C, black; N, green; and S, brown. Additionally, the O and C atoms occurring in carboxyl groups (CHARMM22 atom types OC and CC) are distinguished: OC, gray; CC, red-gray.

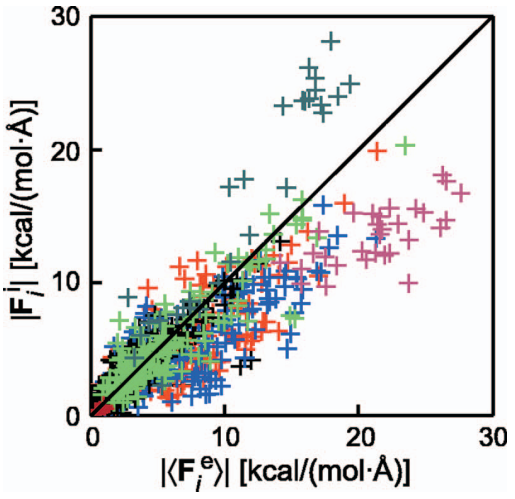

FIG. 3. (Color) Force correlations for the RF forces $\mathbf{F}_i^i(\hat{\Sigma}^{opt}, \hat{\mathcal{G}}^{opt})$ on the PrP^C atoms obtained with the optimized parameters listed in Table I. For further explanation, see the caption to the analogous Fig. 2.

$$\sigma_i = c_R R_i^{c22} \tag{14}$$

proportional to the van der Waals radii R_i^{c22} specified in this force field for various atom types. The quality functional ΔF defined in Eq. (12) then solely depends on the proportionality factor c_R, and one can determine that factor by minimizing ΔF with respect to c_R for the PrP^C force comparison (cf. Sec. III C). A line search of $\Delta F(c_R)$ yielded a minimum at $c_R=0.80$. This value shows that the optimal CHARMM22 parameters σ_i^{c22} obtained through Eq. (14) from this force field are actually of atomic size.

Figure 2 correlates for the PrP^C data set the resulting absolute values $|\mathbf{F}_i^i(c_R=0.8, \mathcal{G}^{c22})|$ of the RF forces with the absolute values $|\langle \mathbf{F}_i^e \rangle|$ of the corresponding average forces calculated in explicit TIP3P water at 300 K. For each PrP^C atom, the graph contains a single cross representing a pair of force values. One observes a close matching for some but large deviations for other RF forces. The deviations are particularly large for the atoms of the CHARMM22 types carboxyl carbon (CC, red-gray) and carboxyl oxygen (OC, gray), which belong to negatively charged carboxyl groups and, therefore, experience sizable average forces. These average solvent forces are underestimated by the RF forces at the given parametrization.

The force correlation displayed by Fig. 2 is quite obviously not very convincing and one may ask to what extent it can be improved by a more diligent choice of the parameters σ_i and of the grouping $\mathcal{G}$. To explore this issue, we applied the quite simple and preliminary optimization procedure described in Sec. III C using the PrP^C data set of average electrostatic solvent forces $\langle \mathbf{F}_i^e \rangle$ as our reference. The choice of this quite limited data set was guided by the practical reasons outlined in Sec. III A. More importantly, however, it was motivated by the following consideration: Only if already an optimization using such a limited data set can render substantially improved and transferable force correlations, a thorough parametrization using much larger training sets and more sophisticated optimization procedures will make sense in the future.

Table I shows the final grouping $\hat{\mathcal{G}}^{opt}$, its relation to $\mathcal{G}^{c22}$, and the final parameter set $\hat{\Sigma}^{opt}$ obtained by our optimization. For $\hat{\mathcal{G}}^{opt}$ the gradient descent procedure (cf. Sec. III C) converged to values of the target functional ΔF, which were about three times smaller than the values found for the original CHARMM22 choice, indicating that, in fact, more diligent choices of $\mathcal{G}$ and Σ can yield substantially improved force correlations.

If one compares the optimized parameters $\hat{\Sigma}^{opt}$ in Table I with the CHARMM22-derived choice Σ^{c22}, sizable changes become apparent. For instance, the very small widths $\sigma_H^{c22}=0.18$ Å of polar hydrogens have become larger ($\sigma_H^{opt}=0.93$ Å) upon optimization. Intuitively this increased width seems to be a more pleasing measure of atom size. On the other hand, some of the optimized values for other hydrogen atoms have become counterintuitively large (types HP, HA, and HR), raising the question of whether our simple optimization procedure is really adequate.

A comparison of the force correlations in Fig. 3 with the correlations in Fig. 2 visualizes the substantial improvement, which was (despite the above caveats) generated by our optimization. The force pairs of most nitrogen, carbon, and hydrogen atoms neatly aligned along the diagonal. Solely the carbon (gray) and oxygen (red-gray) atoms of the carboxyl groups sizably deviated from that line. With the optimized parameters, the RF approach now overestimates the solvent forces on the carboxyl carbons and still moderately underestimates the solvent forces on the carboxyl oxygens (implying that the total force on a carboxyl group may be better modeled than the forces on its individual atoms). Nevertheless, already our first and limited optimization attempt has shown that, in principle, substantial improvements can be achieved by more appropriate choices of the σ_i.

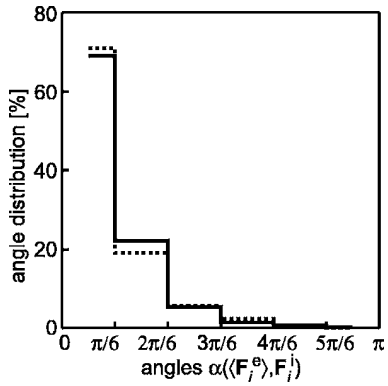

FIG. 4. Distributions of the angles α between the electrostatic solvent forces $\langle \mathbf{F}_i^e \rangle$ and $\mathbf{F}_i^i$ on the PrP^C atoms. Solid line: optimized parameters as in Fig. 3; dotted line: CHARMM22-derived parameters as in Fig. 2.

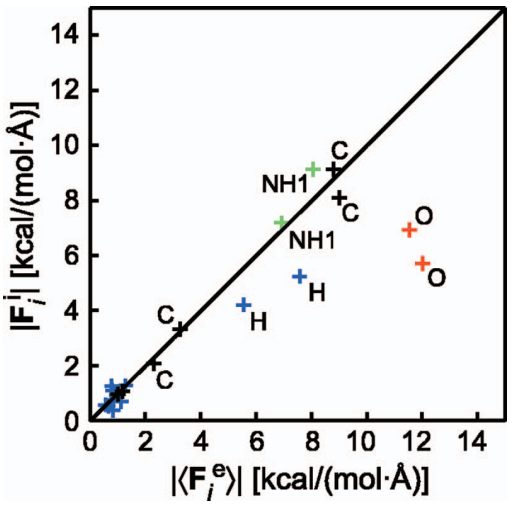

FIG. 5. (Color) Force correlations for the absolute values of the RF forces $|\mathbf{F}_i^i(\hat{\Sigma}^{\text{opt}}, \hat{\mathcal{G}}^{\text{opt}})|$ on the atoms of AlDi in an arbitrarily selected α-helical conformation. Atoms are labeled by their CHARMM22 atom types. For further explanation, see the caption to Fig. 2, and for discussion, see text.

If one searches for a possible cause why the strong underestimates of the forces on the atoms of carboxyl groups, which were found with the CHARMM22-derived parameter set, turned into moderate overestimates (carbons) or underestimates (oxygens), one should consult Table I. According to the table, the widths of the Gaussians attached to the CCs became smaller upon optimization by a factor of 2.5 and those of the OCs became smaller by only a factor of 1.6. It thus seems that shrinking parameters σ are associated with increasing RF forces and that the optimization rendered the CC parameters too small while leaving the OC parameters too large. In view of the highly nonlinear way that the Gaussian widths of covalently connected atoms enter the RF approach, such a suboptimal result of a gradient descent optimization had to be expected. More extended and detailed studies on the mutual dependencies of the parameters focusing on specific chemical motifs, the application of more sophisticated optimization procedures, and more extended sets of training data will be required for a further improvement of the RF description.

Concerning the latter point, the inclusion of further solvents into the training database will be of particular importance, because our optimization was solely based on water. In addition to strongly shielding the electric field, this highly protic solvent can form specifically ordered hydrogen bonds to solvent-exposed donors and acceptors of hydrogen bonds. Solvent-exposed carboxyl groups are prominent examples for corresponding structural components of proteins. If such ordered hydrogen bonds are present, then the average forces measured for an explicit water model are expected to be larger than continuum RF forces. If an optimization is supposed to mainly aim at a model for the dielectric properties of the solvent, nonprotic polar solvents such as dimethylsulfoxide or simple alcohols should be included into a reference data set.

The above comparisons solely referred to the absolute sizes of the forces exerted by the solvent on the atoms of a solute protein. Thus, it still could be that the directions of the RF forces show strong deviations. To check this question, we show in Fig. 4 histograms for the distributions of the angles between the average forces $\langle \mathbf{F}_i^e \rangle$ exerted by the TIP3P water on the PrP^C and the RF forces $\mathbf{F}_i^i$ for the optimized and CHARMM22-derived parameter sets. The two histograms are quite similar and show a pronounced peak at small angles, indicating that the RF solvent forces point actually in similar directions as the forces derived from the molecular modeling of the surrounding water. In a future parameter optimization, a more detailed analysis of such data will be required, distinguishing, e.g., very small forces, for which large angular deviations can be tolerated, and large forces, whose directions should closely match. At the present stage of methods development, the histogram indicates a satisfactory agreement of force directions.

B. Transferability

Having shown that substantial improvements of force correlations can be achieved by a parameter optimization like the one employed by us, the question of transferability must be addressed next. For the reasons explained in Sec. III A, we chose the small peptide AlDi depicted in Fig. 1 as our test case. Like in the optimization, we had to select a specific peptide configuration for the transferability check. The random choice of a configuration from the 300 K equilibration trajectory of AlDi in TIP3P water happened to be of α-helical type.

Figure 5 compares for this α-helical configuration the absolute values $|\langle \mathbf{F}_i^e \rangle|$ of the mean forces exerted by TIP3P water at 300 K on the peptide atoms with the absolute values $|\mathbf{F}_i^i(\hat{\Sigma}^{\text{opt}}, \hat{\mathcal{G}}^{\text{opt}})|$ of the RF forces. The shown correlation is of similar quality as the optimized correlation in Fig. 3, indicating that the optimized parameters $\hat{\Sigma}^{\text{opt}}$ are transferable (note the different scales used in Figs. 3 and 5).

If one looks at some of the details exhibited by Fig. 5, one immediately recognizes that the RF forces underestimate the average forces $\langle \mathbf{F}_i^e \rangle$ on the amide hydrogens H and carbonyl oxygens O of the two peptide groups. To what extent these underestimates of the average forces on the donors (H) and acceptors (O) of hydrogen bonds with the aqueous environment are caused by suboptimal parameters, are a specific

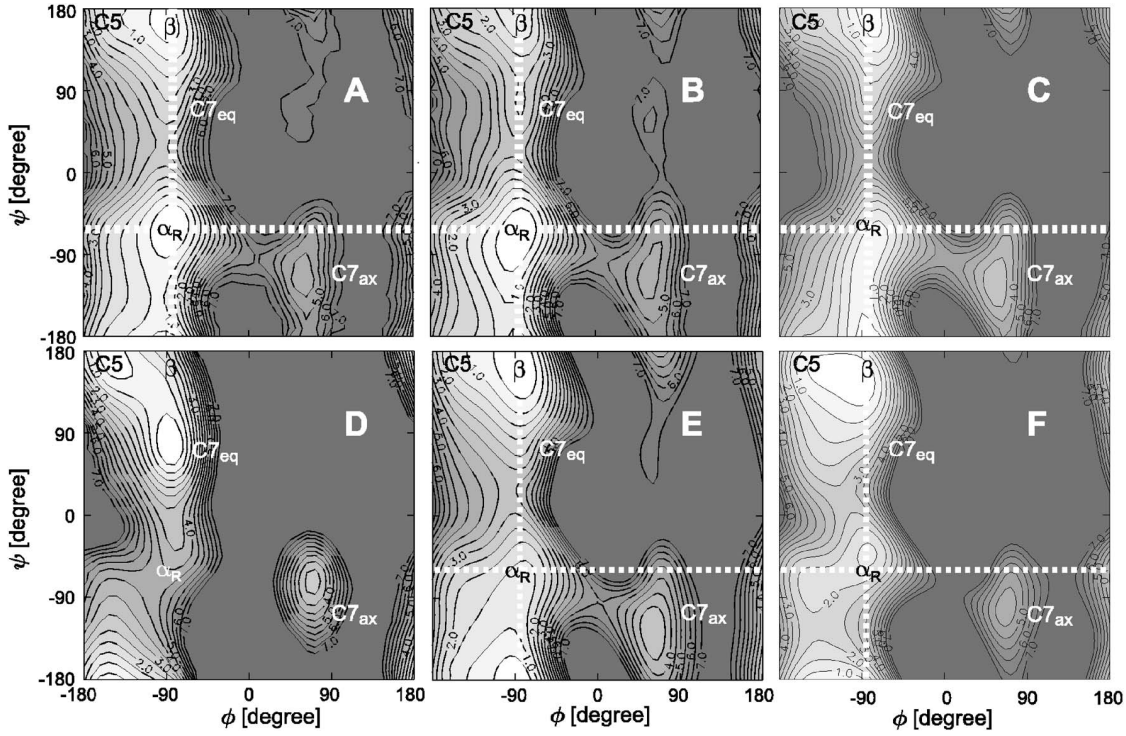

FIG. 6. Free energy surfaces $\Delta G(\phi,\psi)$ of a CHARMM22 model for AlDi calculated by different methods. (A) 50 ns MD simulation in TIP3P water at 500 K. (B) 500 ns EL/RF-MD at 500 K with parameters $\hat{\Sigma}^{\mathrm{opt}}$ and $\hat{\mathcal{G}}^{\mathrm{opt}}$. (C) 500 ns EL/RF-MD at 500 K with parameters Σ^{c22} and $\mathcal{G}^{\mathrm{c22}}$. (D) 500 ns MD in vacuum. (E) Free energy scan $\Delta G^{e}(\phi,\psi|\hat{\Sigma}^{\mathrm{opt}},\hat{\mathcal{G}}^{\mathrm{opt}})$. (F) Free energy scan $\Delta G^{e}(\phi,\psi|\Sigma^{\mathrm{c22}},\mathcal{G}^{\mathrm{c22}})$. The white dashed bars in (A)–(C) and (E) and (F) indicate the locations of the vertical and horizontal slices shown in Fig. 7.

property of the chosen AlDi configuration, or signify fundamental limitations of a continuum modeling (cf. the discussion further above) cannot be decided at the current stage of methods development, but will become clear from an extended parametrization study to be carried out in the future. Presently it remains to be checked how strongly the different force values affect the free energy landscapes of peptides. For this test we chose AlDi as the simplest example because of the reasons listed in Sec. III A.

C. Energy landscapes of alanine dipeptide

The top panel (A–C) of Fig. 6 compares the free energy landscapes $\Delta G(\phi,\psi)$ of AlDi obtained (A) from MD simulations using TIP3P water, (B) from RF-MD simulations of EL type with the optimized and (C) CHARMM22-derived parameter sets. For a more extended sampling of the (ϕ,ψ) plane, an elevated temperature of 500 K was used. The contour levels represent steps of 0.5 kcal/mol and the color coding distinguishes energy levels from $\Delta G=0$ (white) to the respective energy cutoff $\Delta G_{\max}$ (dark gray) [see Sec. III E for details].

Figure 6(A) reproduces the well-known CHARMM22 free energy surface of AlDi in TIP3P water, which has been calculated many times before (see, e.g., Ref. 26). Its global minimum is denoted as α_R and indicates that AlDi preferentially assumes a right-handed α-helical conformation. The extended β-strand conformation represents a second nearly isoenergetic state, which at $T=300$ K and ambient pressure is found[26] only about 0.3 kcal/mol above α_R. At these conditions, the free energy surface exhibits a third state that is 5.0 kcal/mol above α_R, which is metastable and is commonly denoted as $\mathrm{C7}_{\mathrm{ax}}$. The (ϕ,ψ) positions and ΔG values of these conformational states are nearly invariant upon heating to $T=500$ K. To guide the eyes, these (ϕ,ψ) positions are marked by the symbols α_R, β, and $\mathrm{C7}_{\mathrm{ax}}$, which are identically placed in the remaining plots B, C, E, and F.

The visual comparison of the free energy surfaces in Figs. 6(A) (reference) and 6(B) (EL/RF-MD, $\hat{\Sigma}^{\mathrm{opt}}$) immediately reveals a striking similarity. This impression is quantitatively supported by the inspection of Fig. 7(a), which shows vertical (left) and horizontal (right) surface cuts, whose locations are indicated by the white dashed lines in Fig. 6(A). In Fig. 7(a) the cuts through the reference surface of Fig. 6(A) are drawn as dotted lines and those through the EL/RF-MD surface of Fig. 6(B) are drawn as solid lines. The lines are seen to agree very closely except for the region near the extended β-strand conformation. Apparently the EL/RF-MD simulation predicts the β state at a free energy, which is about 0.5 kcal/mol above the value found by MD in TIP3P water. In view of the coarse $9^\circ\times 9^\circ$ square bins employed for the statistics and of the inevitable statistical fluctuations, one may summarize this comparison by the statement that the EL/RF-MD simulation, which is based on the parameters $\hat{\Sigma}^{\mathrm{opt}}$, quite accurately reproduces the free energy landscape of AlDi in explicit solvent.

The effect of the parameter optimization, which was based on force comparisons for $\mathrm{PrP}^{\mathrm{C}}$, on the free energy

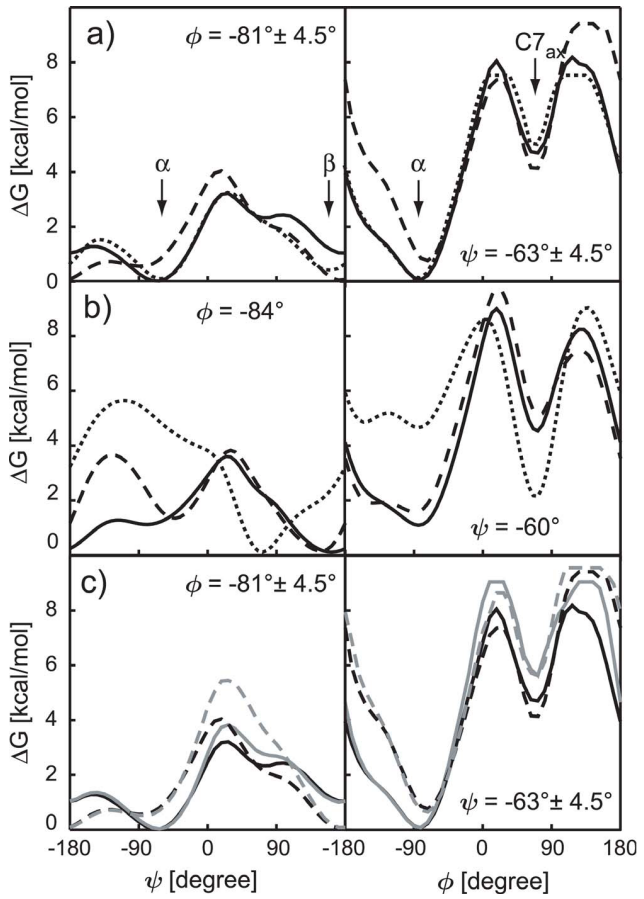

FIG. 7. Vertical (left) and horizontal (right) slices of the free energy surface $\Delta G(\phi,\psi)$ for a CHARMM22 model of AlDi calculated by different methods. (a) Cuts through the surfaces shown in the top panel of Fig. 6(A) (dotted), 6(B) (solid), and 6(C) (dashed); see the corresponding caption for details. (b) Cuts through surfaces generated by (ϕ,ψ) scanning: potential energy (dotted), cuts through the surfaces in Figs. 6(E) (solid) and 6(F) (dashed). (c) Comparison of ΔG slices from EL-type (black) and SCF-type (gray) RF-MD simulations; the line style distinguishes parameter sets: $\hat{\Sigma}^{\mathrm{opt}}$, $\hat{\mathcal{G}}^{\mathrm{opt}}$ (solid) and Σ^{c22}, $\mathcal{G}^{\mathrm{c22}}$ (dashed).

surface $\Delta G(\phi,\psi)$ of AlDi can be recognized by comparing Fig. 6(C), which was obtained from an EL/RF-MD simulation with the original CHARMM22-derived parameter set, with Figs. 6(A) and 6(B). With the strongly suboptimal parameter set $(\Sigma^{\mathrm{c22}},\mathcal{G}^{\mathrm{c22}})$, the results deteriorate, as is witnessed by the fact that the global minimum at α_R has turned into a saddle point. The quantitative aspects of this deterioration can be better judged by comparing the dashed curves in Fig. 7(a), which represent cuts through Fig. 6(C), with the other curves. With the suboptimal parameters, the α_R conformation is energetically elevated by about 0.8 kcal/mol and the β conformation becomes the global minimum. Moreover, the barrier between the α_R and β states depicted at the center of the left graph in Fig. 7(a) increases by about 1 kcal/mol. Nevertheless, the overall structure of $\Delta G(\phi,\psi)$ in Fig. 6(C) does not dramatically deteriorate upon use of $(\Sigma^{\mathrm{c22}},\mathcal{G}^{\mathrm{c22}})$, as one can deduce from the nearly invariant locations and depths of the minima at β and $\mathrm{C7_{ax}}$.

The above observations contain two messages: (i) The free energy landscapes deduced from EL/RF-MD simulations depend to a certain degree, but not dramatically, on the choice of the Gaussian widths σ_i. Because these parameters determine how the volume of a protein is defined and partitioned into atomic contributions, and because the sets $\hat{\Sigma}^{\mathrm{opt}}$ and Σ^{c22} strongly differ (cf. Table I), the EL/RF-MD approach seems to have a certain intrinsic stability with respect to parameter variations. (ii) An optimization based on force comparisons (cf. Sec. IV A) can bring the EL/RF-MD quite close to MD simulations with explicit water.

The bottom panel (D–F) of Fig. 6 serves to show that (a) the free energy surface $\Delta G(\phi,\psi)$ of AlDi is shaped by the RF forces, that (b) parameters Σ suitable for the description of *RF forces* are not necessarily optimal for the evaluation of the *RF energy* expression (60) of Paper I, and that (c) the procedure chosen for parameter optimization needs further improvement.

In support of claim (a), Fig. 6(D) displays $\Delta G(\phi,\psi)$ for a vacuum simulation, according to which there are new minima C5 and $\mathrm{C7_{eq}}$, while the minima α_R and β are absent. Furthermore, the conformation $\mathrm{C7_{ax}}$ has become stabilized at a shifted position [cf. Figs. 6(A)–6(D)]. The key effects of the solvent or, equivalently, of the RF forces on the AlDi solution structure are thus the stabilization of the α-helical and β-strand conformations as well as the destabilization of $\mathrm{C7_{ax}}$, proving that $\Delta G(\phi,\psi)$ is shaped by the RF forces.

The doubtful statement (b) concerning the joint suitability of parameters for force *and* energy computations is explained by the free energy scans in Figs. 6(E) and 6(F). These scans were calculated from Eq. (60) of Paper I with the optimized [Fig. 6(E)] and CHARMM22-derived [Fig. 6(F)] parameter sets. Interestingly, the CHARMM22-derived parameters seem to perform quite well in the direct computation of the configurational free energy through Eq. (60) of Paper I, as becomes apparent by comparing Fig. 6(F) with the reference surface in Fig. 6(A). The three minima α_R, β, and $\mathrm{C7_{ax}}$ are found at nearly identical locations in the (ϕ,ψ) plane and the overall structures of the two surfaces are quite similar. The inspection of the corresponding surface cuts [compare the dashed curves in Fig. 7(b) with the dotted curves in Fig. 7(a)] shows, however, that the free energy scanning predicts the β conformation as the global minimum and the α_R state to be about 1.2 kcal/mol above that minimum [dashed curve, Fig. 7(b), left]. However, despite these differences the initial impression that the scanning procedure provides for the CHARMM22-derived parameters an acceptable description of the reference surface is maintained. In contrast, the free energy surface scan in Fig. 6(E), which was calculated with the force-optimized parameters, performs much worse. As evidenced by the corresponding cuts [solid curves in Fig. 7(b)], the minimum at α_R has nearly turned into a saddle point upon optimization.

The shape of the force-optimized free energy scan would be even worse than the one depicted in Fig. 6(E) if we did not adjust a further parameter for better comparison. The addressed parameter is the factor c_ϕ, which according to Eq. (65) of Paper I should have the value 2.774 [see Paper I and Ref. 6 for details concerning the role of this parameter in scaling the Gaussian widths in direct free energy calculations]. Whereas we have maintained the original choice for the computation of Fig. 6(F), we had to diminish c_ϕ to the

value 1.4 in order to bring the surface [Fig. 6(E)] computed with $\hat{\Sigma}^{\mathrm{opt}}$ as close as possible to the reference surface [Fig. 6(A)]. But even with the reduced choice, the resulting surface [Fig. 6(E)] became at most similar to suboptimal the EL/RF-MD result [Fig. 6(C)] calculated with the parameters Σ^{c22}. Thus, as claimed by statement (b), parameters improving the description of RF forces may actually deteriorate the evaluation of RF energies.

The basic reason for this contrary behavior of force and direct free energy computations can be traced back to the use of different routes of approximation and to the fact that the approximate RF forces are not exactly the negative derivatives of the RF free energies in our approach (cf. Paper I). For the choice of a parameter optimization procedure, this contrary behavior implies that one should try to maintain an acceptable description of the free energies while optimizing the RF forces and try to modify the target functional (12) accordingly. This insight is equivalent to the above statement (c) that the applied optimization procedure needs further improvement. In view of the relative stability of the RF-MD free energy surfaces with respect to parameter variations, which we noted further above, it seems likely that an improved procedure can be found.

If we presently neglect the deterioration of the direct free energy computation, which was caused by our simplistic parameter optimization, the high quality of the resulting EL/RF-MD simulation approach remains a key result of this preliminary study. Recall in this context that we have argued for an EL-type MD procedure not only for reasons of enhanced computational efficiency but also because the EL approach allows us to model the relatively slow dielectric relaxation of the polar solvent through the choice of the relaxation time τ_p (cf. Sec. II). Our rough guess, which we employed in the simulations, was $\tau_p=2.5$ ps. Thus, it remains to be checked how a much faster relaxation of the RF dipoles, which is modeled by a SCF-type procedure and, hence, proceeds on a time scale of about 1 fs, affects the simulation results.

Figure 7(c) compares cuts through the EL/RF-MD free energy surfaces [Figs. 6(B) (black solid line) and 6(C) (black dashed line)] with cuts through corresponding SFC/RF-MD results (gray solid and dashed lines). While one can see an overall agreement between the black and gray curves for the two parameter sets, there are also certain deviations which may be due to the faster dielectric relaxation modeled by SCF or to other reasons associated with different levels of algorithmic noise and the different methods applied for noise control:

Firstly, what we call a SCF-type procedure is not a SCF method in a strict sense, because the applied convergence criteria are extraordinarily crude. Instead, our SCF-type procedure is an inaccurate approximation to a strict SCF method and, therefore, generates a considerable amount of algorithmic noise. Secondly, as pointed out repeatedly in Paper I, the RF forces, while conserving the linear momentum, do not exactly conserve the angular momentum. Therefore, at present, we have to apply rotation corrections, which generate additional algorithmic noise both in SCF- and EL-type simulations. All these sources of algorithmic noise heat up the peptide during a SCF/RF-MD simulation such that a weak coupling Berendsen thermostat is required to approximately maintain the desired target temperature of 500 K. However, our simulations have shown that the noise-induced heat production is small enough to be safely controlled by a thermostat with a coupling time as slow as $\tau_T=1$ ps. For the EL/RF-MD simulations, we applied for the same reasons such weak coupling thermostats to the atoms and to the dipoles.

As a final remark, we would like to discuss simulation times, i.e., the efficiency of the method. On a single processor of a modern personal computer, one MD integration step for AlDi in TIP3P water involving the update of 3055 atomic coordinates took 2.8×10^{-1} s. In contrast, one EL/RF-MD integration step updating the 22 atomic positions and dipoles required only 4.6×10^{-4} s. Thus, the speedup factor from replacing the 1011 TIP3P water molecules by 22 RF dipoles is 609. With our current and not yet fully optimized implementation of the EL/RF-MD approach, one integration step took a factor of 3.5 longer than the integration step of a vacuum simulation.

D. Summary and outlook

The above results provide a "proof of principle" that a highly efficient and accurate RF-MD simulation approach for proteins in polar solvents can be constructed by combining the electrostatic force expressions (70)–(75) derived in Paper I with an EL approach and with optimized parameters for the Gaussian widths of the atomic RF dipole densities. Here, the measures of efficiency and accuracy are the free energy landscapes of solute peptides resulting from MD simulations with explicit solvent.

We have used the simplest example (AlDi) as a check of efficiency and accuracy. This example has shown that already a parametrization based on a single conformation of a specific soluble protein ($\mathrm{PrP}^{\mathrm{C}}$) can lead to RF forces, which in an EL/RF-MD simulation of our small peptide yield a free energy landscape closely resembling the one obtained by an explicit solvent simulation. This close match was achieved, although the detailed force comparisons shown in Fig. 5 revealed certain differences between the RF forces and the mean forces exerted by an explicit solvent model on the peptide atoms. Thus it seems that peptide free energy landscapes resulting from the EL/RF-MD approach are quite insensitive to these details or, conversely, that the free energy landscapes of AlDi are not very much influenced by specific effects of hydrogen bonding between the TIP3P water and the amide groups. Similarly it turned out that the EL/RF-MD landscapes are quite stable with respect to substantial changes of the parameters Σ. On the other hand, the example has also shown that the target functional (12) employed for the optimization, which exclusively aims at an improvement of forces, is actually too simplistic, because it disturbs the computation of a much more sensitive quantity, which is also delivered by our approach, i.e., the configurational free energy $G^{\varepsilon}(\mathbf{R})$ of the peptide.

As a result, our investigation has also identified a number of tasks which have to be tackled in an envisaged trans-

formation of the current RF-MD approach into a generally applicable, reliable, and easy-to-use method. To conclude the two parts of the current investigation, we will now sketch these tasks:

(i) As explained in Secs. III E and III F of Paper I, the use of Gaussian charge distributions for the atomic partial charges of a MM force field can be easily integrated into the RF approach, does not seem to require substantial force field reparametrizations, and offers distinct advantages concerning the "isolated atom limit," which is relevant for a proper treatment of protein-protein or protein-substrate interactions in RF-MD docking simulations. Therefore, the transition to Gaussian charges is the first imminent step.

(ii) The use of Gaussian charges requires that these distributions are essentially confined to the interior of the van der Waals spheres of the atoms.[1] This requirement sets an upper limit of about $\sigma_{ij} \leqslant 0.8 R_{ij}$ to the Gaussian widths σ_i of the charge and RF dipole distributions, where $\sigma_{ij}^2 = \sigma_i^2 + \sigma_j^2$ and R_{ij} is half the van der Waals contact distance between two nonbonded atoms i and j. The upper limit imposes a constraint to any parameter optimization, which was not yet included into the current one.

(iii) The target functional for parameter optimization should be extended by observables, which can guarantee that the computation of the configurational free energies $G^{\varepsilon}(\mathbf{R})$ does not deteriorate too much at the expense of a more accurate force computation. In view of the finding that for AlDi reasonable $G^{\varepsilon}(\phi, \psi)$ were obtained with the CHARMM22-derived parameter set, this task seems to be tractable.

(iv) Extended training data sets including different proteins in various solvents and more sophisticated optimization procedures should be applied to compute transferable parameters.

(v) The *reactio* force contribution given by Eq. (75) of Paper I and its *actio* counterpart are the sources of residual total torques acting on the peptide. Therefore, they should be modified such that they exactly conserve also the angular momentum (without deteriorating the accuracy of the current force computation). This task appears to be quite straightforward, but a corresponding revision requires extended accuracy checks. As a result, one will get a Hamiltonian dynamics. Then one can steer the peptide dynamics at will by adding variable amounts of noise and friction. Currently a corresponding coupling to a heat bath is required to maintain the target temperature.

(vi) Finally, the viability of the resulting method has to be verified for more complex peptides than AlDi. To this aim, reference data sets have to be generated from extended MD simulations of these peptides in explicit solvent. In view of the efficiency of the RF-MD approach, the generation of corresponding test data will be much less demanding.

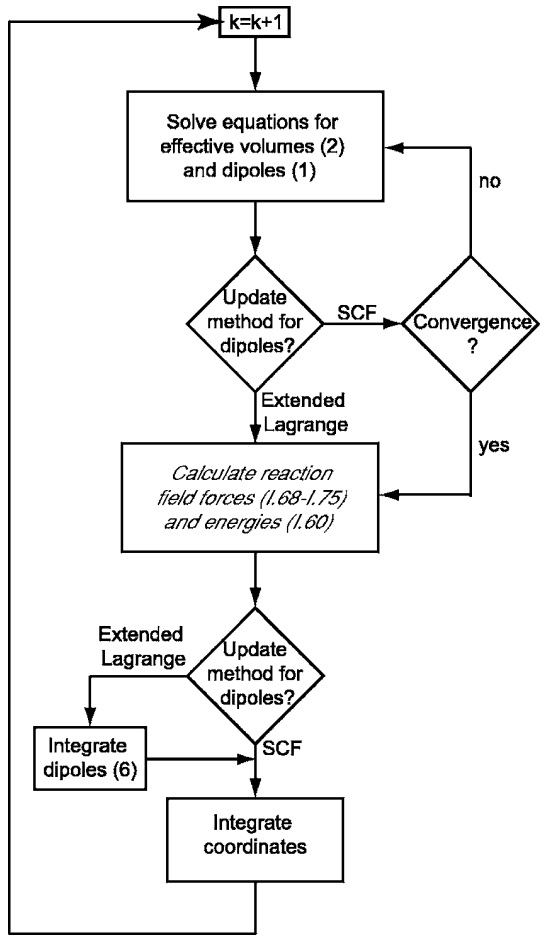

FIG. 8. Flow chart for the calculation of RF dipoles, forces, and energies in the MD program EGO-MMII.

ACKNOWLEDGMENTS

This project was funded by the Bayerischer Forschungsverbund Prionen (Project No. LMU02) and by the Deutsche Forschungsgemeinschaft (SFB 533/C1).

APPENDIX: IMPLEMENTATION

Figure 8 displays a flow chart sketching the two alternative algorithms introduced in Sec. II for the calculation of the RF forces. These algorithms, the SCF and EL approaches, differ concerning the update of the RF dipoles. At a given integration step k, the atomic coordinates $\mathbf{r}_i$, volumes v_i, and dipoles $\mathbf{p}_i$ are given as results of the preceding step $k-1$. Both methods start by iterating once, first, the volume [Eq. (2)] and, subsequently, the dipole [Eq. (1)] using the current values v_i and $\mathbf{p}_i$ as initial values $v_{i,0}$ and $\mathbf{p}_{i,0}$. In the case of the SCF algorithm, the updated values $v_{i,1}$ and $\mathbf{p}_{i,1}$ are chosen as inputs for iterating Eqs. (2) and (1) until convergence, as measured by predefined accuracy criteria, has been reached. The SFC dipoles yield then the RF forces and energies, and the atomic coordinates are integrated by a multiple time step algorithm[41] from the total forces. In the case of the EL method, the volume equation is iterated only once and the value "$\alpha_i \mathbf{E}_{\mathrm{pol}}^{q,\mathbf{p}}(\mathbf{r}_i)$" [cf. dipole equation (1)] is calculated. Then the RF forces and energies are calculated using

the old values for the dipoles. By using Eq. (6), the RF dipoles are integrated parallel to the propagation of the atomic coordinates.

[1] M. Stork and P. Tavan, J. Chem. Phys. **126**, 165105 (2006), preceding paper.
[2] D. Bashford and A. Case, Annu. Rev. Phys. Chem. **51**, 129 (2000).
[3] C. D. Snow, N. Nguyen, V. S. Pande, and M. Gruebele, Nature (London) **420**, 102 (2002).
[4] R. Zhou and B. J. Berne, Proc. Natl. Acad. Sci. U.S.A. **99**, 12777 (2002).
[5] H. Nymeyer and A. E. García, Proc. Natl. Acad. Sci. U.S.A. **100**, 13934 (2003).
[6] B. Egwolf and P. Tavan, J. Chem. Phys. **118**, 2039 (2003).
[7] B. Egwolf and P. Tavan, J. Chem. Phys. **120**, 2056 (2004).
[8] P. Ahlström, A. Wallquist, S. Engström, and B. Jönsson, Mol. Phys. **68**, 563 (1989).
[9] M. K. Gilson, M. E. Davis, B. A. Luty, and J. A. McCammon, J. Phys. Chem. **97**, 3591 (1993).
[10] T. A. Darden, D. York, and L. Pedersen, J. Chem. Phys. **98**, 10089 (1993).
[11] C. Niedermeier and P. Tavan, J. Chem. Phys. **101**, 734 (1994).
[12] M. L. Horng, J. A. Gardecki, A. Papazyan, and M. Maroncelli, J. Phys. Chem. **99**, 17311 (1995).
[13] R. Car and M. Parrinello, Phys. Rev. Lett. **55**, 2471 (1985).
[14] M. Sprik and M. L. Klein, J. Chem. Phys. **89**, 7556 (1988).
[15] M. Sprik, J. Phys. Chem. **95**, 2283 (1991).
[16] D. van Belle and S. J. Wodak, Comput. Phys. Commun. **91**, 253 (1995).
[17] V. M. Anisimov, G. Lamoureux, I. V. Vorobyov, N. Huang, B. Roux, and A. D. MacKerell, Jr., J. Chem. Theory Comput. **1**, 153 (2005).
[18] G. A. Kaminski, R. A. Friesner, and R. Zhou, J. Comput. Chem. **24**, 267 (2003).
[19] E. Harder, B. Kim, R. A. Friesner, and B. J. Berne, J. Chem. Theory Comput. **1**, 169 (2005).
[20] L. Verlet, Phys. Rev. **159**, 98 (1967).
[21] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak, J. Chem. Phys. **81**, 3684 (1984).
[22] G. Mathias, B. Egwolf, M. Nonella, and P. Tavan, J. Chem. Phys. **118**, 10847 (2003).
[23] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, J. Comput. Chem. **4**, 187 (1983).
[24] A. D. MacKerell, Jr., D. Bashford, M. Bellott *et al.*, J. Phys. Chem. B **102**, 3586 (1998).
[25] T. Hirschberger, M. Stork, B. Schropp, K. F. Winklhofer, J. Tatzelt, and P. Tavan, Biophys. J. **90**, 3908 (2006).
[26] P. E. Smith, J. Chem. Phys. **111**, 5568 (1999).
[27] F. Fogolari, G. Esposito, P. Viglino, and H. Molinari, J. Comput. Chem. **22**, 1 (2001).
[28] S. A. Hassan and E. L. Mehler, Proteins **47**, 45 (2002).
[29] H. Hu, M. Elstner, and J. Hermans, Proteins **50**, 451 (2003).
[30] H. Carstens, Ph.D. thesis, Ludwig-Maximilians-Universität, 2004.
[31] N. Levy, D. Borgis, and M. Marchi, Comput. Phys. Commun. **169**, 69 (2005).
[32] R. Zahn, A. Liu, T. Lührs, R. Riek, C. von Schroetter, F. L. Garcia, M. Billeter, L. Calzolai, G. Wider, and K. Wüthrich, Proc. Natl. Acad. Sci. U.S.A. **97**, 145 (2000).
[33] G. Schaftenaar and J. H. Noordik, J. Comput.-Aided Mol. Des. **14**, 123 (2000).
[34] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, J. Chem. Phys. **79**, 926 (1983).
[35] G. Mathias and P. Tavan, J. Chem. Phys. **120**, 4393 (2004).
[36] M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids* (Clarendon, Oxford, 1987).
[37] V. Kräutler, W. F. van Gunsteren, and P. H. Hünenberger, J. Comput. Chem. **22**, 501 (2001).
[38] W. L. Jorgensen and C. Jenson, J. Comput. Chem. **19**, 1179 (1998).
[39] For a more accurate comparison, one should calculate for each of the 100 configurations $\mathbf{R}_t$ the forces $\mathbf{F}_i^t(t)$ and the ensemble averages $\langle \mathbf{F}_i^t \rangle$. However, such a procedure would increase the computing time required for the optimization of the σ_i by a factor of 100. In a test calculation, we verified that the reaction field forces $\mathbf{F}_i^t(0)$ at the configuration $\mathbf{R}_0$ are very good approximations for the average forces $\langle \mathbf{F}_i^t \rangle$ (data not shown).
[40] P. Höchtl, S. Boresch, W. Bitomsky, and O. Steinhauser, J. Chem. Phys. **109**, 4927 (1998).
[41] M. Eichinger, H. Grubmüller, H. Heller, and P. Tavan, J. Comput. Chem. **18**, 1729 (1997).

5 Zusammenfassung und Ausblick

In dieser Arbeit wurde in Molekulardynamik-Simulationen die Stabilität amyloidogener Proteine und Peptide in wässriger Lösung untersucht, sowie mit der Weiterentwicklung einer Kontinuumsmethode die Grundlage für zeitlich ausgedehntere Simulationen geschaffen.

Wie in der Einleitung bereits ausführlich beschrieben wurde, kommt es bei vielen neurodegenerativen Erkrankungen zu amyloiden Ablagerungen im Gehirn. Diese Amyloide bestehen aus Proteinen oder Peptiden, die, beispielsweise bei der Creutzfeldt-Jakob-Krankheit, in zwei Isoformen, d.h. in zwei verschiedenen stabilen dreidimensionalen Strukturen, vorliegen können. Während das entsprechende Protein in seiner nativen Form löslich ist, bildet es in der pathogenen Isoform Aggregate, so z.B. das Prion-Protein PrP bei CJK. In MD-Simulationen an der löslichen Form PrP^{C} konnten in dieser Arbeit strukterelle Schwachstellen identifiziert werden, von denen die Umfaltung des Proteins ausgehen kann. Für die pathogene Isoform PrP^{Sc} und für Peptide aus poly-Gln wurden mögliche Aggregationskeime für die Bildung amyloider Fasern ermittelt.

Da bei einer mikroskopischen Beschreibung des Lösungsmittels nur kurze Zeiträume simuliert werden können und so nur begrenzt Aussagen über die Konformationsdynamik von Proteinen und Peptiden möglich sind, wurde eine Kontinuumsmethode für den Einsatz in MD-Simulationen weiterentwickelt. Damit wurden die Grundlagen für ausgedehntere Simulationen geschaffen, die es ermöglichen sollen, Vorhersagen der zeitlich begrenzten Simulationen statistisch zu überprüfen und detailliertere Aussagen über die Mechanismen der Umfaltung und Aggregation amyloidogener Proteine zu machen.

Im Folgenden fasse ich die Ergebnisse aus den einzelnen Teilprojekten dieser Arbeit zusammen und gebe jeweils einen kurzen Ausblick auf mögliche Ansatzpunkte zukünftiger Projekte.

Das zelluläre Prion-Protein

In Kapitel 2 dieser Arbeit wurde in MD-Simulationen die Konformationsdynamik des zellulären Prion-Proteins PrP^{C} des Menschen sowie die zweier Varianten des Proteins untersucht. Den Anlass für diese Simulationen gab eine Zellkulturstudie [129], in der gezeigt worden war, dass die Punktmutationen M205S und M205R die Faltung und Maturierung des PrP^{C} erheblich stören. Mittels MD-Simulationen sollte

überprüft werden, ob diese Störungen dadurch erklärt werden können, dass die Varianten eine andere räumliche Struktur besitzen als das native PrP^C. Dazu wurden in die Struktur des nativen PrP^C die Mutationen eingebaut und die Stabilität dieser Strukturen mittels MD-Simulationen untersucht.

Das Residuum Methionin 205 (M205) ist Teil der Helix 3 des PrP^C (vgl. Abbildung 1.4) und steht in engem räumlichen Kontakt mit hydrophoben und leicht polaren Seitenketten aus Helix 1. Durch Einführung der Punktmutationen M205R bzw. M205S, in denen das hydrophobe Methionin in einem Fall durch das geladene Arginin (R), im anderen Fall durch das polare Serin (S) ersetzt wird, wurde dieser hydrophobe Kontakt unterbunden. Mittels MD-Simulationen wurde die thermische Bewegung jeder der drei Varianten des Proteins in wässriger Lösung über einen Zeitraum von 10 ns verfolgt.

Während bei nativem PrP^C lediglich Fluktuationen um die NMR-Struktur auftraten, die im Einklag mit Daten aus NMR-Messungen [39] standen, waren bei den beiden Mutanten die Änderungen der dreidimensionalen Strukturen erheblich. Bei M205S betrafen die gravierendsten strukturellen Änderungen Helix 2. Dies weist auf eine Schwachstelle von PrP^C in diesem Bereich hin. Bei der Variante M205R löste sich Helix 1 vom Rest des Prion-Proteins ab, wurde solvatisiert und löste sich daraufhin nahezu vollständig auf. Dieses Ergebnis weist darauf hin, dass Residuum M205 Helix 1 durch hydrophobe Wechselwirkungen an den hydrophoben Kern von PrP^C bindet und dadurch die korrekte Faltung von Helix 1 ermöglicht. Damit ist die funktionelle Rolle des M205-Residuums in der PrP^C-Sequenz geklärt. Dieses Residuum fungiert als intramolekulares Chaperon für die Faltung von Helix 1. Wir schließen ferner aus den in solch kurzen Simulationszeiten beobachteten großen Instabilitäten, dass die Punktmutanten M205S und M205R des PrP^C nicht in der nativen Struktur vorliegen und bestätigen so die Vermutung der Zellkulturstudie.

In der soeben skizzierten Arbeit wurden eine Reihe von Fragen aufgeworfen, die erst durch wesentlich längere Simulationen beantwortet werden können. Bei der Überprüfung entsprechender Vermutungen könnte der Einsatz von Kontinuumsmethoden helfen. Einiges deutet darauf hin, dass die Faltung von Helix 1 im Wesentlichen von der Dielektrizitätskonstante der Umgebung bestimmt wird. In ihrer gefalteten Form befindet sich Helix 1 in nahem Kontakt mit dem Rest des Proteins, der ein Volumen mit einer niedrigen Dielektrizitätskonstante ϵ (≈ 1) darstellt. Löst sie sich vom Protein, so befindet sie sich in einer wässrigen Umgebung mit einer hohen Dielektrizitätskonstante ϵ (≈ 78) und entfaltet sich dort. Unsere Vermutung, dass die Dielektrizitätskonstante die Struktur der Helix 1 bestimmt, könnte in einer Kontinuumssimulation, bei der das Lösungsmittel allein durch seine Dielektrizitätskonstante charakterisiert wird, leicht durch eine Variation von ϵ überprüft werden.

Eine längere MD-Simulation am nativen PrP^C könnte schließlich klären, ob die kleineren Konformationsänderungen, die in der MD-Simulation auftraten, und die sich während der Simulationszeit von 10 ns nicht wieder zurückbildeten, reversible Fluktuationen der Proteinstruktur darstellen.

Das Prion-Protein Scrapie

Da die erreichbaren Simulationszeiten in MD-Simulationen deutlich zu kurz sind, um eine Umfaltung von PrP^{C} in seine infektiöse Isoform PrP^{Sc} beobachten zu können, haben wir auch am anderen Ende des Umfaltungsprozesses angesetzt und, wie in Kapitel 3 beschrieben wurde, Modelle für einen Aggregationskeim des PrP^{Sc} entwickelt. Auf der Grundlage von EM-Messungen an zweidimensionalen Kristallstrukturen aus PrP^{Sc}-Fragmenten entwickelten wir ein Modell für den PrP^{Sc}-Aggregationskeim, das, wie von Wille et al. [44] vorgeschlagen, vom Strukturmotiv einer parallelen β-Helix bestimmt wird, sich aber zudem an die Randbedingungen hält, die sich aus einer Analyse nativer β-Helizes ergeben.

Da im Inneren nativer β-Helizes keine geladenen Seitenketten zu finden sind, wurde dieser Befund als zwingende Randbedingung bei der Enwicklung unseres Modells aufgenommen. Ihre Erfüllung war nur durch die Einführung zweier Loopstrukturen, die von den Ecken der β-Helizes ausgehen, zu erreichen. Durch unterschiedliche Anordnungen der PrP-Sequenz auf die parallele β-Helix und die Loopstrukturen entstanden zwei alternative Modelle. Da sich native β-Helizes in Kristallen oft zu Trimerstrukturen zusammensetzen, wurde diese Anordnung auch für die Grundbausteine im PrP^{Sc}-Modell übernommen. In den erhaltenen Strukturen ließen sich sowohl die weiterhin vorhandene α-Helix 3 des PrP^{C} und der erhaltene Teil der α-Helix 2, sowie die an das Protein gebundenen Zuckerseitenketten gut unterbringen. Das Wachstum an beiden Enden einer solchen Fibrille lässt sich durch eine Stapelung der β-Helizes erklären. Die Dimensionen unseres Modells für einen PrP^{Sc}-Aggregationskeim stimmen insgesamt sehr gut mit den EM-Messungen überein.

Die Frage stellt sich, ob wir im Rahmen unserer Möglichkeiten das PrP^{Sc}-Modell noch weiter überprüfen oder auch verfeinern können. Für längere MD-Simulationen in expliziter Lösungsmittelumgebung ist der entstandene Aggregationskeim (Durchmesser ≈ 80 Å) zu groß. Zudem sind die Möglichkeiten zur Ausrichtung der Seitenketten und für die Anordnung der Sequenz auf die β-helikale Struktur und auf die Loopstrukturen so vielfältig, dass auch Ergebnisse von MD-Simulationen in impliziter Lösungsmittelumgebung nur sehr spekulativ sein könnten.

Unser Modell könnte jedoch mit weiteren Daten aus experimentellen Methoden der Strukturaufklärung, z.B. mittels Crosslinking-Experimenten [127, 128], verfeinert werden. Zudem macht unser Modell aber auch jetzt schon Vorhersagen, die in Experimenten überprüft werden könnten. So müsste eine Ersetzung einer der ungeladenen Seitenketten, die bei uns ins Innere der β-Helix zeigen, durch eine geladene Seitenkette destabilisierend auf die Struktur wirken. In experimentellen Messungen sollte sich dies durch ein verändertes Aussehen der PrP^{Sc}-Aggregate oder durch eine Verlangsamung des Aggregationsverlaufs bemerkbar machen.

Ein Aggregationskeim für amyloide Fasern aus poly-Gln

Ein Aggregationskeim amyloider Fasern, der sich auf Grund seiner geringen Größe gut für die Untersuchung in MD-Simulationen eignet und ebenfalls in Kapitel 3 betrachtet wurde, sind Peptide aus poly-Glutamin (Gln). Diese Peptide bilden vermutlich die Grundbausteine von Fibrillen, die bei der Huntingtonschen Krankheit in den Gehirnen erkrankter Menschen auftreten. Wie für PrP^{Sc} wurde auch für die Grundstruktur dieser Aggregate das Strukturmotiv einer parallelen β-Helix vorgeschlagen [58]. Im Gegensatz zur Modellierung der PrP^{Sc}-Fibrillen, bei der es vielfältige Möglichkeiten für die Anordnung der Sequenz auf die Struktur der β-Helix gibt, stellt sich dieses Problem bei der poly-Gln-Sequenz nicht.

Da das Strukturmotiv der parallelen β-Helix in MD-Simulationen bislang noch nicht behandelt worden war, musste zuerst überprüft werden, ob unser Kraftfeld [82] überhaupt dazu geeignet ist, Vorhersagen über die Stabilitäten paralleler β-Helizes zu treffen. In MD-Simulationen in wässriger Lösung erwies sich ein Fragment einer nativ stabilen β-Helix-Struktur (LpxA) [61] als stabil, solange es aus mindestens drei β-helikalen Windungen bestand. Im Gegensatz dazu erwies sich eine gleich große parallele β-Helix aus poly-Alanin (Ala), das in Wirklichkeit keine parallele β-Helix sondern eine α-Helix bildet, als instabil. Daraus konnten wir schließen, dass unser Kraftfeld die Stabilitäten β-helikaler Peptide korrekt wiedergeben kann und es uns erlaubt, auch Vorhersagen für die Stabilitäten β-helikaler Peptide zu treffen.

Die poly-Gln-Peptide wurden anfänglich nach einem Vorschlag von Perutz et al. [58] als β-Helizes mit einer zirkulären Grundstruktur modelliert. Dabei wurden zwei Modelle mit jeweils drei Windungen untersucht, die sich in der Zahl ihrer Residuen pro β-helikaler Windung unterschieden. Das Modell mit 18 Residuen pro Windung löste sich nach kurzer Zeit größtenteils auf. Auch das Modell mit 20 Residuen pro Windung behielt seine Anfangsstruktur während der Simulation nicht. Hier ähnelte die ursprünglich zirkuläre Grundstruktur durch die Herausbildung mehrerer Ecken nach 10 ns der triangulären Grundstruktur des β-helikalen LpxA-Fragments. Dies veranlasste uns dazu, das poly-Gln-Peptid ebenfalls als trianguläre parallele β-Helix mit drei Windungen zu modellieren. Poly-Gln in dieser Struktur erwies sich in MD-Simulationen während einer Simulationszeit von 10 ns als mindestens so stabil wie das Fragment einer nativen β-Helix.

Das trianguläre β-helikale Modell mit drei Windungen (insgesamt $\approx$ 54 Residuen) scheint deshalb ein geeigneter Kandidat für einen Aggregationskeim bei der Bildung amyloider Fasern zu sein. Nach dem Mechanismus der *Seeded Aggregation* [47, 48] lagern sich an einen solchen Keim weitere Peptide an. Dabei stellt die Struktur des Aggregationskeims eine Art Schablone für die Änderung der dreidimensionalen Struktur der Peptide dar. Wie wir in einer weiteren Simulation zeigen konnten, besitzt nicht nur ein Peptid mit drei Windungen genügend Kurzzeitstabilität, um als Aggregationskeim in Frage zu kommen, sondern auch ein Dimer bestehend aus zwei Peptiden mit je zwei Windungen. Jedes dieser Peptide setzt sich aus 36 Glutaminen zusammen. Die Stabilität dieses Dimers kann den experimentellen Befund [130]

erklären, wonach die Huntington-Krankheit bei Menschen auftritt, in deren exon-1 Peptid des Huntingtin mindestens 35-46 Glutamine in Folge vorkommen.

Am Ende dieses Projektes stellte sich die Frage, warum sich die Strukturen mancher β-helikaler Peptide (poly-Gln mit zirkulärer Grundstruktur und poly-Ala) auflösen, während die anderer β-helikaler Peptide (poly-Gln mit triangulärer Grundstruktur und LpxA) erhalten bleiben. Durch die Modellierung und MD-Simulation zweier weiterer Peptide aus poly-Isoleucin (Ile) bzw. poly-Serin (Ser) in der Struktur triangulärer β-Helizes konnten wir eine einfache und schlüssige Antwort auf diese Frage finden. Alle Strukturen, in deren Zentren hydrophobe oder schwach polare Seitenketten dicht gepackt waren (poly-Gln mit triangulärer Grundstruktur, LpxA, poly-Ile) blieben in den Simulationen stabil, während alle Strukturen, in deren Zentren sich offene Kanäle befanden (poly-Gln mit zirkulären Grundstrukturen, poly-Ala, poly-Ser), die vom wässrigen Lösungsmittel gefüllt werden konnten, zerfielen. Daraus konnten wir die Schlussfolgerung ziehen, dass das Strukturmotiv einer parallelen β-Helix durch einen hydrophoben Kern stabilisiert wird und dass poly-Gln-Peptide, falls sie die Struktur einer parallelen β-Helix annehmen, keinesfalls wassergefüllte Nanoröhren darstellen, wie es von Perutz et al. [58] vorgeschlagen worden war.

Auch hier lassen sich weitere Fragen erst in ausgedehnteren Simulationen klären. Um unseren Vorschlägen für die möglichen Aggregationskeime für Fasern aus poly-Gln mehr Substanz zu geben, müsste gezeigt werden, dass die in Frage kommenden Peptide in den modellierten β-helikalen Strukturen auch länger stabil bleiben. Weiterhin wäre es interessant zu untersuchen, ob sich aus einem ungefalteten poly-Gln-Peptid auch spontan eine β-helikale Struktur bilden kann und ob bei der Simulation mehrerer Peptide in einem System eine spontane Aggregation auftritt. Für all diese Fragestellungen sind Simulationen in mikroskopischen Lösungsmitteln zu zeitaufwändig.

Eine Kontinuumsmethode für MD-Simulationen

Viele der oben angesprochenen offenen Fragen zur Konformationsdynamik und Umfaltung amyloidogener Proteine und Peptide lassen sich erst durch ausgedehntere MD-Simulationen klären. Durch den Einsatz einer effizienten Kontinuumsmethode zur Beschreibung der Wasserumgebung könnten relevante Zeitskalen zugänglich werden. Deshalb wurde in dieser Arbeit eine von Egwolf und Tavan (ET) entwickelte Kontinuumsmethode [98, 116] so umformuliert und modifiziert, dass sie in MD-Simulationen eingesetzt werden kann. Dies wurde detailliert in Kapitel 4, Abschnitt 4.1 dargestellt.

Wie schon in der Einleitung in Abschnitt 1.5.2 beschrieben wurde, wird in der ET-Methode das durch die Polarisation des Kontinuums erzeugte Reaktionsfeld näherungsweise durch normalverteilte RF-Dipoldichten an den Orten der Proteinatome repräsentiert. Atomare RF-Dipole $\mathbf{p}_i$ lassen sich in selbstkonsistenter Weise aus dem Feld aller Ladungen q_i und RF-Dipole $\mathbf{p}_i$ bestimmen. Das Reaktionsfeld $\mathbf{E}^{\varepsilon}(\mathbf{r}_i)$ und

das RF-Potential ergeben sich dann analytisch aus den RF-Dipolen. Durch die einfache Beziehung

$$\mathbf{F}_i^{\varepsilon} = q_i \mathbf{E}^{\varepsilon}(\mathbf{r}_i)$$

übt das Reaktionsfeld Kräfte $\mathbf{F}_i^{\varepsilon}$ auf die geladenen Atome aus. Da diese Kräfte aus einer Wechselwirkung zwischen den RF-Dipolen und den Punktladungen der Atome resultieren, müssten nach dem 3. Newtonschen Gesetz gleich große, entgegengesetzte Kräfte auch auf die RF-Dipole wirken. In der ET-Theorie wurde dieses *Actio=Reactio*-Prinzip vernachlässigt, was zur Entstehung einer Gesamtkraft und eines Gesamtdrehmoments auf das gelöste Molekül führt. In MD-Simulationen ergeben sich daraus große Artefakte.

Durch theoretische Überlegungen konnten wir zeigen, dass sich aus der ET-Methode im Prinzip auch Ausdrücke für *Reactio*-Kräfte ableiten lassen. Die RF-Energie stellt ein Potential der mittleren Kraft dar, aus dem man durch Bildung der negativen Gradienten bezüglich der Atomorte mittlere RF-Kräfte erhält. Solche aus der RF-Energie gewonnenen RF-Kräfte erfüllen das *Actio=Reactio*-Prinzip und wirken im Gegensatz zu den $q\mathbf{E}$-Kräften auch auf ungeladene Atome. Da die direkte Berechnung von RF-Kräften aus der RF-Energie sehr kompliziert und rechenaufwändig ist, haben wir nach einem alternativen Weg gesucht, die bestehenden RF-Kräfte so zu modifizieren, dass sie die *Reactio*-Kräfte beinhalten.

Als ersten Schritt dahin haben wir einen störenden Selbstanteil entfernt, der sowohl in der Berechnung der RF-Dipole als auch bei der Berechnung des Reaktionsfeldes vorhanden war und eine Kraft eines RF-Dipols auf die Ladung des gleichen Atoms hervorrief. Die entsprechenden Gleichungen wurden so umformuliert, dass nur die Felder, die von den Atomen und RF-Dipolen fremder Ladungen erzeugt wurden, zur Entstehung eines RF-Dipols beitrugen. Es ergab sich der einfache Zusammenhang

$$\mathbf{p}_i = -\,\alpha_i\, \mathbf{E}_{\mathrm{pol}}^{q,\mathbf{p}}(\mathbf{r}_i)$$

zwischen dem RF-Dipol $\mathbf{p}_i$ eines Atoms und dem polarisierenden Feld $\mathbf{E}_{\mathrm{pol}}^{q,\mathbf{p}}(\mathbf{r}_i)$ aller anderen Ladungen und Dipole, der über eine RF-Polarisierbarkeit α_i vermittelt wird. In die RF-Polarisierbarkeit α_i gehen die Breiten σ_i der normalverteilten Dipoldichten, die Dielektrizitätskonstanten ε_{s} des Proteinvolumens und ε_{c} der Lösungsmittelumgebung und effektive atomare Volumina v_i ein. Das negative Vorzeichen in der Gleichung drückt aus, dass die Dipole die Polarisation des Kontinuums in das Proteinvolumen „hineinspiegeln“. Mit Hilfe dieser Gleichung für die RF-Dipole konnte auch die Berechnungsvorschrift für die RF-Kräfte so umformuliert werden, dass die störende Selbstkraft entfiel.

Als weiteren Schritt zur Einführung von *Reactio*-Kräften verglichen wir die mittleren RF-Kräfte, die sich aus den negativen Gradienten der RF-Energie bezüglich der Atomorte ergeben, mit den oben umformulierten RF-Kräften der ET-Theorie für den Grenzfall weit entfernter Atome. Aus der großen Ähnlichkeit einzelner Terme konnten wir einen Vorschlag für die Einführung geeigneter *Reactio*-Kräfte entwickeln. Dass die so modifizierten RF-Kräfte sinnvoll sind, konnten wir an Hand der analyti-

schen Kirkwood-Lösung für eine Punktladung in einer sphärischen Kavität und für die einfachen Fälle zweier Ionen im Kontinuum sowie eines Ions und eines ungeladenen Atoms im Kontinuum zeigen.

MD-Simulationen mit der Kontinuumsmethode

Die Erweiterung der RF-Kräfte um einen *Reactio*-Anteil schuf die Grundlage für die Durchführung stabiler MD-Simulationen mit der Kontinuumsmethode. Die erweiterte Methode sowie die Möglichkeiten zur Berechnung der RF-Energie W^{ε} wurden von mir in das MD-Programm EGO-MMII implementiert und an Hand eines einfachen Systems parametrisiert und getestet, wie in Abschnitt 4.1 beschrieben wurde.

Die RF-Dipole und somit auch die RF-Kräfte und -Energien sind abhängig von den Breiten σ_i der normalverteilten Dipoldichten. Diese Parameter wurden so optimiert, dass die resultierenden RF-Kräfte möglichst gut den elektrostatischen Kräften entsprechen, die von einer expliziten Wasserumgebung im Mittel auf die Atome eines Proteins ausgeübt werden. Die Parametrisierung wurde für eine Konfiguration des zellulären Prion-Proteins durchgeführt. Die so erhaltenen optimierten Breiten lassen sich jedoch auch auf andere Peptide übertragen, wie am Beispiel des kleinen Alanin-Dipeptids gezeigt wurde. Hier ergab sich für den Vergleich zwischen den RF-Kräften und den mittleren Kräften einer expliziten Wasserumgebung auf das Peptid eine ähnlich gute Übereinstimmung wie für das Prion-Protein.

Zudem wurde die Effizienz der Methode erheblich gesteigert, indem die zeitaufwändige, selbstkonsistente Berechnung der RF-Dipole durch eine Berechnung mit der *Extended Lagrange*-Methode ersetzt wurde. In dieser Methode ergeben sich in einer MD-Simulation die RF-Dipole analog zu den Atomkoordinaten aus einer numerischen Integration geeigneter Bewegungsgleichungen. Mit diesem Ansatz kann zudem, im Gegensatz zur selbstkonsistenten Berechnung der RF-Dipole, berücksichtigt werden, dass die dielektrische Relaxationszeit des Wassers im Bereich von einigen hundert Femtosekunden bis einigen Pikosekunden liegt, weshalb sich auch die RF-Dipole ähnlich langsam an eine veränderte Proteinkonfiguration anpassen sollten.

Die freien Energielandschaften in den relevanten Freiheitsgraden des Alanin-Dipeptids, den Dihedralwinkeln ϕ und ψ, wurden bei verschiedenen Parametrisierungen und für verschiedene Methoden der Dipolberechnung analysiert. Vergleicht man die sich ergebenden Landschaften der freien Energie mit denen aus einer Simulation des Alanin-Dipeptids in einer expliziten Wasserumgebung, zeigt sich, dass deren wesentliche Eigenschaften mit der Kontinuumsbeschreibung reproduziert werden können. Zudem erwies sich die Extended-Lagrange-Methode als effizientes und geeignetes Verfahren für die Berechnung der RF-Dipole.

Die Optimierung der Breiten wirkte sich zwar sehr positiv auf die in den MD-Simulationen erhaltenen Energielandschaften aus, verschlechterte aber die Ergebnisse für die Energielandschaften, die sich aus einer direkten Berechnung der Reaktionsfeld-Energie W^{ε} ergeben. Diese Verschlechterung beruht darauf, dass bei der Para-

metrisierung der Breiten nur die RF-Kräfte, nicht aber die RF-Energien berücksichtigt wurden.

In diesen ersten MD-Simulationen mit der umformulierten und erweiterten ET-Methode konnte gezeigt werden, dass sie grundsätzlich stabile MD-Simulationen ermöglicht und dass die modifizierten RF-Kräfte den elektrostatischen Einfluss der Wasserumgebung auf ein Peptid beschreiben können. Aus diesen Simulationen ergaben sich allerdings auch Ansatzpunkte dafür, wie die Methode weiter optimiert werden kann. Abschließend möchte ich einige noch zu lösende Aufgaben aufzeigen.

Durch die Einführung von *Reactio*-Kräften konnte zwar die Gesamtkraft auf ein Protein, jedoch nicht das Gesamtdrehmoment vollständig beseitigt werden. Deshalb müssen in jedem Integrationsschritt geeignete Korrekturen des Gesamtdrehmoments durchgeführt werden, die zu Artefakten führen können. Ein Gesamtdrehmoment ergibt sich daraus, dass die RF-Kräfte durch die Näherungen der ET-Theorie nicht direkt aus einem Potential abgeleitet sind. Weitere theoretische Überlegungen und ein Vergleich mit polarisierbaren Kraftfeldern könnten Wege für eine Modifizierung der RF-Kräfte aufzeigen, so dass das Gesamtdrehmoment verschwindet.

Weiterhin ist, wie oben aufgezeigt wurde, die Parametrisierung der Breiten der Dipoldichten in der modifizierten RF-Methode noch nicht optimal. Das Verfahren zur Optimierung der Breiten müsste so geändert werden, dass es sich nicht nur an den RF-Kräften orientiert, sondern auch die RF-Energien in geeigneter Weise berücksichtigt.

Ein weiterer vielversprechender Ansatzpunkt ist die Beschreibung der atomaren Ladungen durch normalverteilte Ladungsdichten, statt wie bisher durch Punktladungen. Wie wir in Abschnitt 4.1 dieser Arbeit gezeigt haben, vereinfachen und vereinheitlichen sich dadurch nicht nur die Gleichungen für die RF-Dipole und RF-Kräfte, sondern es ergibt sich auch ein physikalisch stimmigeres Bild bei der Beschreibung isolierter Atome. In unserem MD-Programm EGO-MMII ist es schon jetzt möglich, die RF-Dipole und RF-Kräfte bei normalverteilten Ladungsdichten zu berechnen. Inwieweit dies auch für die Berechnung der Coulomb-Kräfte sinnvoll ist, ohne dass eine aufwändige Reparametrisierung des Kraftfelds durchgeführt werden muss, ist noch zu untersuchen.

Wenn die Kontinuumsmethode auch für die Beschreibung voneinander isolierter Moleküle gute Ergebnisse liefert, könnte sie auch in Verfahren des *Docking* eingesetzt werden, bei denen untersucht wird, wie und ob sich ein Molekül an ein anderes Molekül anlagert. Dies ist besonders im Bereich amyloidogener Proteine und Peptide von großem Interesse. Beispielsweise könnte man hier untersuchen, wie PrP^{Sc}-Proteine oder poly-Gln-Peptide zu amyloiden Fasern aggregieren, und ob sich die Aggregation durch eine Veränderung der Kontaktbereiche, wie z.B. durch das Anbringen geeigneter Liganden, unterbinden lässt.

Literaturverzeichnis

[1] Creutzfeldt, H. G. Über eine eigenartige herdförmige Erkrankung des Zentralnervensystems. *Z. Gesamte Neurol. Psychiatrie* **57**, 1–18 (1920). 1

[2] Jakob, A. Über eigenartige Erkrankungen des Zentralnervensystems mit bemerkenswertem anatomischem Befunde (spastische Pseudosklerose-Encephalomyelopathie mit disseminierten Degenerationsherden). *Z. Gesamte Neurol. Psychiatrie* **64**, 147–228 (1921). 1

[3] Roos, R., D. C. Gajdusek und J. G. Clarence, Jr. The Clinical Characteristics of Transmissible Creutzfeldt-Jakob disease. *Brain* **96**, 1–20 (1973). 1

[4] Katscher, F. It's Jakob's disease, not Creutzfeldt's. *Nature* **393**, 11 (1998). 1

[5] DeArmond, S. J., J. W. Ironside, E. Bouzamondo-Bernstein und D. Peretz. Neuropathology of Prion Diseases. In S. B. Prusiner (Herausgeber), *Prion Biology and Diseases*, Seiten 777–856. Cold Spring Harbor Laboratory Press, USA (2004). 2, 8, 9

[6] Holman, R. C., A. S. Khan, E. D. Belay und L. B. Schonberger. Creutzfeldt-Jakob Disease in the United States, 1979-1994: Using National Mortality Data to Assess the Possible Occurrence of Variant Cases. *Emerging Infectious Diseases* **2**, 333–337 (1996). 1

[7] Arias, E., R. N. Anderson, H.-C. Kung, S. L. Murphy und K. D. Kochanek. Deaths: Final Data for 2001. *National Vital Statistics Reports* **52**, 1–116 (2003). 1

[8] Kirschbaum, W. Zwei eigenartige Erkrankungen des Zentralnervensystems nach Art der spastischen Pseudosklerose (Jakob). *Z. Gesamte Neurol. Psychiatrie* **92**, 1–25 (1924). 1

[9] Cuillé, J. und P. L. Chelle. La maladie dite tremblante du mouton est-elle inoculable? *C. R. Seances Acad. Sci* **203**, 1552–1554 (1936). 1

[10] Prusiner, S. B., E. Williams, J.-L. Laplanche und M. Shinagwa. Scrapie, Chronic Wasting Disease, and Transmissible Mink Encephalopathy. In S. B. Prusiner (Herausgeber), *Prion Biology and Diseases*, Seiten 545–594. Cold Spring Harbor Laboratory Press, USA (2004). 2

[11] Hadlow, W. J. Scrapie and Kuru. *Lancet* **2**, 289–290 (1959). 2

[12] Gajdusek, D. C. und V. Zigas. Degenerative Disease of the Central Nervous System in New-Guinea – The Endemic Occurence of Kuru in the Native Population. *N. Engl. J. Med.* **257**, 974–978 (1957). 2

[13] Klatzo, I., D. C. Gajdusek und V. Zigas. Pathology of Kuru. *Lab. Invest.* **8**, 799–847 (1959). 2

[14] Gajdusek, D. C., C. J. Gibbs und M. Alpers. Experimental Transmission of a Kuru-Like Syndrome to Chimpanzees. *Nature* **209**, 794–795 (1966). 2

[15] Gibbs, C. J., D. C. Gajdusek, D. M. Asher, M. P. Alpers, E. Beck, P. M. Daniel und W. B. Matthews. Creutzfeldt-Jakob Disease (Spongiform Encephalopathy) – Transmission to Chimpanzee. *Science* **161**, 388–389 (1968). 2

[16] Kong, Q., W. K. Surewicz, R. B. Petersen, W. Zou, S. G. Chen und P. Gambetti. Neuropathology of Prion Diseases. In S. B. Prusiner (Herausgeber), *Prion Biology and Diseases*, Seiten 673–775. Cold Spring Harbor Laboratory Press, USA (2004). 2

[17] World organisation for animal health, Internetseite vom 22.06.06, http://www.oie.int/. 2

[18] Bundesministerium für Ernährung, Landwirtschaft und Verbraucherschutz, Internetseite vom 22.06.06, http://www.bmelv.de. 2

[19] 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* **382**, 779–788 (1996). 2

[20] Prusiner, S. B. Prions. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 13363–13383 (1998). 3, 4, 7

[21] 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

[22] Prusiner, S. B., E. Williams, J.-L. Laplanche und M. Shinagwa. Scrapie, Chronic Wasting Disease, and Transmissible Mink Encephalopathy. In S. B. Prusiner (Herausgeber), *Prion Biology and Diseases*, Seiten 545–594. Cold Spring Harbor Laboratory Press, USA (2004). 3

[23] Department of Health, Internetseite vom 22.06.06, http://www.doh.gov.uk. 3

[24] Prusiner, S. B. Novel proteinaceous infectious particles cause scrapie. *Science* **216**, 136–144 (1982). 3, 4

[25] Basler, K., B. Oesch, M. Scott, D. Westaway, M. Walchli, D. F. Groth, M. P. McKinley, S. B. Prusiner und C. Weissmann. Scrapie and cellular PrP isoforms are encoded by the same chromosomal gene. *Cell* **46**, 417–428 (1986). 4

[26] Prusiner, S. B. Molecular-biology of prion diseases. *Science* **252**, 1515–1522 (1991). 4

[27] Griffith, J. S. Self-replication and scrapie. *Nature* **215**, 1043–1044 (1967). 4

[28] Creighton, T. E. *Proteins. Structures and Molecular Properties*. W. H. Freeman and Company, New York, 2. Auflage (1993). 4, 10

[29] Branden, C. und J. Tooze. *Introduction to Protein Structure*. Garland Publishing, Inc., New York, 2. Auflage (1999). 4

[30] Pauling, L., R. B. Corey und H. R. Branson. The structure of proteins - 2 hydrogen-bonded helical configurations of the polypeptide chain. *Proc. Natl. Acad. Sci. U.S.A.* **37**, 205–211 (1951). 4, 5

[31] Eisenberg, D. The discovery of the alpha-helix and beta-sheet, the principal structural features of proteins. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 11207–11210 (2003). 4

[32] Schmitz, M. *Entwicklung, Anwendung und Vergleich von Methoden zur Berechnung von Infrarotspektren einzelner Moleküle in polaren Lösungsmitteln*. Dissertation, Ludwig-Maximilians-Universität, München (2004). 5

[33] Anfinsen, C. Principles that govern folding of protein chains. *Science* **181**, 223–230 (1973). 6

[34] Brown, D. R., K. Qin, J. W. Herms, A. Madlung, J. Manson *et al*. The cellular prion protein binds copper in vivo. *Nature* **390**, 684–687 (1997). 6

[35] Mitteregger, G., M. Vosko, B. Krebs, W. Xiang, J. Burk, G. F. Hamann und H. A. Kretzschmar. The role of the octarepeat region and its histidines in neuroprotective function of PRPC. *Brain Pathology* **16**, 133 (2006). 6

[36] Schätzl, H. M., M. Da Costa, L. Taylor, F. E. Cohen und S. B. Prusiner. Prion Protein gene variation among primates. *J. Mol. Biol.* **245**, 362–374 (1995). 6

[37] Tatzelt, J. und K. F. Winklhofer. Folding and misfolding of the prion protein in the secretory pathway. *Amyloid* **11**, 162–172 (2004). 6

[38] Riek, R., S. Hornemann, G. Wider, M. Billeter, R. Glockshuber und K. Wuthrich. NMR structure of the mouse prion protein domain PrP(121-231). *Nature* **382**, 180–182 (1996). 7

[39] Zahn, R., A. Liu, T. Luhrs, R. Riek, C. V. Schroetter *et al*. NMR Solution Structure of the Human Prion Protein. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 145 (2000). 7, 10, 90

[40] Berman, H. M., J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov und P. E. Bourne. The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242 (2000). 7, 11, 51

[41] Koradi, R., M. Billeter und K. Wüthrich. MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graphics* **14**, 51–55 (1996). 7, 51, 53

[42] Pan, K. M., M. Baldwin, J. Nguyen, M. Gasset, A. Serban *et al*. Conversion of alpha-helices into beta-sheets features in the formation of the scrapie prion proteins. *Proc. Natl. Acad. Sci. U.S.A.* **90**, 10962–10966 (1993). 7, 8, 52

[43] Caughey, B. W., A. Dong, K. S. Bhat, D. Ernst, S. F. Hayes und W. S. Caughey. Secondary structure analysis of the scrapie-associated protein PrP 27-30 in water by infrared spectroscopy. *Biochemistry* **30**, 7672–7680 (1991). 7, 52

[44] Wille, H., M. D. Michelitsch, V. Guenebaut, S. Supattapone, A. Serban, F. E. Cohen, D. A. Agard und S. B. Prusiner. Structural studies of the scrapie prion protein by electron crystallography. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 3563–3568 (2002). 7, 10, 11, 52, 53, 91

[45] Govaerts, C., H. Wille, S. B. Prusiner und F. E. Cohen. Evidence for assembly of prions with left-handed β-helices into trimers. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 8342–8347 (2004). 7, 10, 53

[46] Stork, M., A. Giese, H. A. Kretzschmar und P. Tavan. Molecular Dynamics Simulations Indicate a Possible Role of Parallel β-Helices in Seeded Aggregation of Poly-Gln. *Biophys. Journal* **88**, 2442–2451 (2005). 7, 12, 16, 20, 52

[47] Jarrett, J. T. und P. T. Lansbury, Jr. Seeding „One-Dimensional Crystallization“ of Amyloid: A Pathogenic Mechanism in Alzheimer's Disease and Scrapie? *Cell* **73**, 1055–1058 (1993). 7, 8, 92

[48] Eigen, M. Prionics or The kinetic basis of prion diseases. *Biophys. Chem.* **63**, A1–A18 (1996). 7, 92

[49] Bieschke, J., P. Weber, N. Sarafoff, M. Beekes, A. Giese und H. Kretzschmar. Autocatalytic self-propagation of misfolded prion protein. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 12207–12211 (2004). 7

[50] Sunde, M. und C. C. F. Blake. From the globular to the fibrous state: protein structure and structural conversion in amyloid formation. *Q. Rev. Biophys.* **31**, 1–39 (1998). 8, 9, 51

[51] Bertsch, U., K. F. Winklhofer, T. Hirschberger, J. Bieschke, P. Weber *et al*. Systematic identification of anti-prion drugs by high-throughput screening based on scanning for intensely fluorescent targets (SIFT). *J. of Virology* **79**, 7785–7791 (2005). 8

[52] Weissmann, C. The state of the prion. *Nat. Rev. Microbiol.* **2**, 861–871 (2004). 8

[53] 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). 9

[54] Fändrich, M. und C. M. Dobson. The behaviour of polyamino acids reveals an inverse side chain effect in amyloid structure formation. *The EMBO Journal* **21**, 5682–5690 (2002). 9

[55] Tycko, R. Insights into the Amyloid Folding Problem from Solid-State NMR. *Biochemistry* **42**, 3151–3159 (2003). 9

[56] Tycko, R. Progress towards a molecular-level structural understanding of amyloid fibrils. *Curr. Opin. Struc. Biol.* **14**, 96–103 (2004). 9, 11, 51

[57] Cantor, C. R. und P. R. Schimmel. *Biophysical Chemistry. Part II: Techniques for the study of biological structure and function*. W. H. Freeman and Company, New York (1980). 9

[58] Perutz, M. F., J. T. Finch, J. Berriman und A. Lesk. Amyloid fibers are water-filled nanotubes. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 5591–5595 (2002). 10, 11, 51, 92, 93

[59] Perutz, M. F., B. J. Pope, D. Owen, E. E. Wanker und E. Scherzinger. Aggregation of proteins with expanded glutamine and alanine repeats of the glutamine-rich and asparagine-rich domains of Sup35 and of the amyloid β-peptide of amyloid plaques. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 5596–5600 (2002). 10

[60] Blundell, T. L. und L. N. Johnson. *Protein Crystallography*. Academic Press, London (1976). 10

[61] Raetz, C. R. H. und S. L. Roderick. A Left-Handed Parallel β Helix in the Structure of UDP-N-Acetylglucosamine Acyltransferase. *Science* **270**, 997–1000 (1995). 10, 52, 92

[62] Canet, D. *NMR – Konzepte und Methoden*. Springer, Berlin (1994). 10

[63] Brunger, A. T., R. L. Campbell, G. M. Clore, A. M. Gronenborn, M. Karplus, G. A. Petsko und M. M. Teeter. Solution of a protein crystal-structure with a model obtained from NMR interproton distance restraints. *Science* **235**, 1049–1053 (1987). 10, 11, 52

[64] Brünger, A. T. *X-PLOR, version 3.1. A system for X-ray crystallography and NMR*. Yale University Press, New Haven and London (1992). 10, 11, 13, 14, 52

[65] Lansbury, P. T., P. R. Costa, J. M. Griffiths, E. J. Simon, M. Auger *et al*. Structural model for the β-amyloid fibril based on interstrand alignment of an antiparallel-sheet comprising a C-terminal peptide. *Nat. Struct. Biol.* **2**, 990–998 (1997). 11

[66] Petkova, A. T., Y. Ishii, J. J. Balbach, O. N. Antzutkin, R. D. Leapman, F. Delaglio und R. Tycko. A structural model for Alzheimer's β-amyloid fibrils based on experimental constraints from solid state NMR. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 16742–16747 (2002). 11

[67] Török, M., S. Milton, R. Kayed, P. Wu, T. McIntire, C. G. Glabe und R. Langen. Structural and Dynamic Features of Alzheimer's Aβ Peptide in Amyloid Fibrils Studied by Site-directed Spin Labeling. *J. Biol. Chem.* **277**, 40810–40815 (2002). 11

[68] Der-Sarkissian, A., C. C. Jao, J. Chen und R. Langen. Structural Organization of α-Synuclein Fibrils Studied by Site-directed Spin Labeling. *J. Biol. Chem.* **278**, 37530–3753 (2003). 11

[69] Allen, M. P. und D. J. Tildesley. *Computer Simulation of Liquids*. Clarendon, Oxford (1987). 11, 14

[70] van Gunsteren, W. F. und H. J. C. Berendsen. Computer Simulation of Molecular Dynamics: Methodology, Applications, and Perspectives in Chemistry. *Angew. Chem. Int. Ed. Engl.* **29**, 992–1023 (1990). 11, 12, 14

[71] Frenkel, D. und B. Smit. *Understanding Molecular Simulation: From Algorithms to Applications*. Academic Press, San Diego (2002). 11

[72] Tavan, P., H. Carstens und G. Mathias. Molecular dynamics simulations of proteins and peptides: Problems, achievements, and perspectives. In J. Buchner und T. Kiefhaber (Herausgeber), *Protein Folding Handbook. Part 1.*, Seiten 1170–1195. Wiley-VCH, Weinheim (2005). 11, 14

[73] 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. Journal* **90**, 3908–3918 (2006). 12, 15, 20

[74] Warshel, A. und S. T. Russell. Calculations of electrostatic interactions in biological systems and in solutions. *Q. Rev. Biophys.* **17**, 283–422 (1984). 12

[75] Still, W. C., A. Tempczyk, R. C. Hawley und T. Hendrickson. Semianalytical Treatment of Solvation for Molecular Mechanics and Dynamics. *J. Am. Chem. Soc.* **112**, 6127–6129 (1990). 12, 16, 18

[76] Fogolari, F., G. Esposito, P. Viglino und H. Molinari. Molecular Mechanics and Dynamics of Biomolecules Using a Solvent Continuum Model. *J. Comp. Chem.* **22**, 1–13 (2001). 12

[77] Honig, B. und A. Nicholls. Classical Electrostatics in Biology and Chemistry. *Science* **268**, 1144–1149 (1995). 12

[78] Roux, B. und T. Simonson. Implicit solvent models. *Biophys. Chem.* **78**, 1–20 (1999). 12, 16

[79] Bashford, D. und A. Case. Generalized Born models for macromolecular solvation effects. *Ann. Rev. Phys. Chem.* **51**, 129–152 (2000). 12, 18

[80] Feig, M. und C. L. Brooks III. Recent advances in the development and application of implicit solvent models in biomolecule simulations. *Curr. Opin. Struc. Biol.* **14**, 217–224 (2004). 12, 18

[81] Baker, N. A. Improving implicit solvent simulations: a Poisson-centric view. *Curr. Opin. Struc. Biol.* **15**, 137–143 (2005). 12, 18

[82] MacKerell, A. D., Jr, D. Bashford, M. Bellott, R. L. Dunbrack, Jr, J. D. Evanseck *et al*. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **102**, 3586–3616 (1998). 12, 13, 92

[83] Jorgensen, W. L. und J. Tirado-Rives. The OPLS Potential Functions for Proteins. Energy Minimizations for Crystals of Cyclic Peptides and Crambin. *J. Am. Chem. Soc.* **110**, 1657–1666 (1988). 12, 13

[84] Ponder, J. W. und D. A. Case. Force Fields for Protein Simulation. *Adv. Protein Chem.* **66**, 27–85 (2003). 12, 14

[85] Verlet, L. Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard–Jones Molecules. *Phys. Rev.* **159**, 98–103 (1967). 12

[86] Brooks, B. R., R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan und M. Karplus. CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *J. Comp. Chem.* **4**, 187–217 (1983). 13, 14

[87] van Gunsteren, W., S. Billeter, A. Eising, P. Hünenberger, P. Krüger, A. Mark, W. Scott und I. Tironi. *Biomolecular Simulation: The GROMOS96 Manual and User Guide*. Vdf Hochschulverlag AG, ETH Zürich, Zürich, Switzerland (1996). 13

[88] Schropp, B. *Ableitung polarisierbarer Kraftfelder aus der Dichtefunktionaltheorie: Eine Pilotstudie*. Diplomarbeit, Ludwig-Maximilians-Universität München, Lehrstuhl für BioMolekulare Optik, AG Theoretische Biophysik (2005). 14

[89] Rieff, B. *Erweiterung und Anwendung der strukturadaptierten Multipolmethode für polarisierbare Kraftfelder am Beispiel reinen Wassers*. Diplomarbeit, Ludwig-Maximilians-Universität München, Lehrstuhl für BioMolekulare Optik, AG Theoretische Biophysik (2006). 14

[90] Ahlström, P., A. Wallquist, S. Engström und B. Jönsson. A molecular dynamics study of polarizable water. *Mol. Phys.* **68**, 563–581 (1989). 14

[91] 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). 14, 15, 17

[92] Darden, T. A., D. York und L. Pedersen. Particle mesh Ewald: An N log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **98**, 10089–10092 (1993). 14

[93] Luty, B. A., I. G. Tironi und W. F. van Gunsteren. Lattice-sum methods for calculating electrostatic interactions in molecular simulations. *J. Chem. Phys.* **103**, 3014–3021 (1995). 14

[94] Niedermeier, C. und P. Tavan. A structure adapted multipole method for electrostatic interactions in protein dynamics. *J. Chem, Phys.* **101**, 734–748 (1994). 15

[95] 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). 15

[96] Eaton, W. A., V. Munoz, S. J. Hagen, G. S. Jas, L. J. Lapidus, E. R. Henry und J. Hofrichter. Fast kinetics and mechanisms in protein folding. *Annu. Rev. Biophys. Biomol. Struct.* **29**, 327–359 (2000). 15

[97] Levy, N., D. Borgis und M. Marchi. A dielectric continuum model of solvation for complex solutes. *Comput. Phys. Commun.* **169**, 69–74 (2005). 16

[98] Egwolf, B. und P. Tavan. Continuum description of solvent dielectrics in molecular-dynamics simulations of proteins. *J. Chem. Phys.* **120**, 2056–2068 (2004). 17, 19, 20, 55, 93

[99] Barthel, J., K. Bachhuber, R. Buchner und H. Hetzenauer. Dielectric spectra of some common solvents in the microwave region. Water and lower alcohols. *Chem. Phys. Lett.* **165**, 369–373 (1990). 16

[100] Jackson, J. D. *Classical Electrodynamics*. John Wiley & Sons, Inc., New York, London, Sydney, Toronto (1975). 17

[101] Egwolf, B. *Proteine in wässriger Umgebung: Kontinuumstheorie der Lösungsmittelelektrostatik und ihre effiziente Berechnung*. Dissertation, Ludwig-Maximilians-Universität München, Germany (2004). 17, 19

[102] Davis, M. E., J. D. Madura, B. A. Luty und J. A. McCammon. Electrostatics and diffusion of molecules in solution: Simulations with the University of Houston Brownian dynamics program. *Comput. Phys. Commun.* **62**, 187–197 (1991). 18

[103] Gilson, M. K., M. E. Davis, B. A. Luty und J. A. McCammon. Computation of Electrostatic Forces on Solvated Molecules Using the Poisson-Boltzmann Equation. *J. Phys. Chem.* **97**, 3591–3600 (1993). 18, 19

[104] Holst, M., N. Baker und F. Wang. Adaptive Multilevel Finite Element Solution of the Poisson–Boltzmann Equation I. Algorithms and Examples. *J. Comp. Chem.* **21**, 1319–1342 (2000). 18

[105] Bordner, A. J. und G. A. Huber. Boundary Element Solution of the Linear Poisson–Boltzmann Equation and a Multipole Method for the Rapid Calculation of Forces on Macromolecules in Solution. *J. Comp. Chem.* **24**, 353–367 (2003). 18

[106] Kollman, P. A., I. Massova, C. R. a B. Kuhn, S. Huo, L. Chong *et al.* Calculating structures and free energies of complex molecules: Combining molecular mechanics and continuum models. *Acc. Chem. Res.* **33**, 889–897 (2000). 18

[107] Feig, M., A. Onufriev, M. S. Lee, W. Case und C. L. Brooks III. Performance comparison of generalized Born and Poisson methods in the calculation of electrostatic solvation energies for protein structures. *J. Comp. Chem.* **25**, 265–284 (2004). 18

[108] Fogolari, F., A. Brigo und H. Molinari. Protocol for MM/PBSA Molecular Dynamics Simulations of Proteins. *Biophys. Journal* **85**, 159–166 (2003). 18

[109] Prabhu, N. V., P. Zhu und K. A. Sharp. Implementation and Testing of Stable, Fast Implicit Solvation in Molecular Dynamics using the Smooth-Permittivity Finite Difference Poisson Boltzmann Method. *J. Comp. Chem.* **25**, 2049–2064 (2004). 18

[110] Lu, Q. und R. Luo. A Poisson Boltzmann dynamics method with nonperiodic boundary condition. *J. Chem. Phys.* **119**, 11035–11047 (2003). 18

[111] Onsager, L. Electric Moments of Molecules in Liquids. *J. Am. Chem. Soc.* **58**, 1486–1493 (1936). 18

[112] Born, M. Volumen und Hydratationswärme der Ionen. *Zeitschrift für Physik* **1**, 45–48 (1920). 18

[113] Snow, C. D., N. Nguyen, V. S. Pande und M. Gruebele. Absolute comparison of simulated and experimental protein-folding dynamics. *Nature* **420**, 102–106 (2002). 18

[114] Zhou, R. und B. J. Berne. Can a continuum solvent model reproduce the free energy landscape of a β-hairpin folding in water? *Proc. Natl. Acad. Sci. U.S.A.* **99**, 12777–12782 (2002). 18

[115] Nymeyer, H. und A. E. García. Simulation of the folding equilibrium of a-helical peptides: A comparison of the generalized Born approximation with explicit solvent. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 13934–13939 (2003). 18

[116] Egwolf, B. und P. Tavan. Continuum description of solvent dielectrics in molecular-dynamics simulations of proteins. *J. Chem. Phys.* **118**, 2039–2056 (2003). 19, 93

[117] Bruenger, A. T., A. Krukowski und J. Erikson. Slow-Cooling Protocols for Crystallographic Refinement by Simulated Annealing. *Acta Cryst. A* **46**, 558–593 (1990). 20

[118] Stork, M. und P. Tavan. Electrostatics of proteins in dielectric solvent continua. I. Newton's third law marries qE forces. *J. Chem. Phys.* **126**, 165105/1–15 (2007). 21

[119] Stork, M. und P. Tavan. Electrostatics of proteins in dielectric solvent continua. II. First applications in molecular dynamics simulations. *J. Chem. Phys.* **126**, 165106/1–13 (2007). 21

[120] James, T. L., H. Liu, N. B. Ulyanov, S. Farr-Jones, H. Zhang *et al.* Solution structure of a 142-residue recombinant prion protein corresponding to the infectious fragment of the scrapie isoform. *Proc. Natl. Acad. Sci. U.S.A.* **94**, 10086–10091 (1997). 51

[121] Gasset, M., M. Baldwin, D. Lloyd, J.-M. Gabriel, D. Holtzmann, F. Cohen, R. Fletterick und S. Prusiner. Predicted α-helical regions of the prion protein when synthesized as peptides form amyloid. *Proc. Natl. Acad. Sci. U.S.A.* **89**, 10940–10944 (1992). 52

[122] Jenkins, J. und R. Pickersgill. The architecture of parallel β-helices and related folds. *Prog. Biophys. Mol. Bio.* **77**, 111–175 (2001). 52

[123] Guex, N. und M. Peitsch. SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. *Electrophoresis* **18**, 2714–2723 (1997). 52

[124] Humphrey, W., A. Dalke und K. Schulten. VMD – Visual Molecular Dynamics. *J. Mol. Graphics* **14**, 33–38 (1996). 52

[125] Zuegg, J. und J. E. Gready. Molecular dynamics simulation of human prion protein including both N-linked oligosaccharides and the GPI anchor. *Glycobiology* **10**, 959–974 (2000). 53

[126] DeMarco, M. L. und V. Daggett. From conversion to aggregation: Protofibril formation of the prion protein. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 2293–2298 (2004). 53

[127] Piening, N., P. Weber, T. Hogen, M. Beekes, H. Kretzschmar und A. Giese. Photoinduced crosslinking of prion protein oligomers and prions. *Amyloid* **13**, 67–77 (2006). 54, 91

[128] Onisko, B., E. G. Fernandez, M. L. Freire, A. Schwarz, M. Baier, F. Camina, J. R. Garcia, V. S. Rodriguez-Segade und J. R. Requena. Probing Prpsc Structure Using Chemical Cross-Linking and Mass Spectrometry: Evidence of the Proximity of Gly90 Amino Termini in the PrP 27-30 Aggregate. *Biochemistry* **44**, 10100–10109 (2005). 54, 91

[129] Winklhofer, K. F., J. Heske, U. Heller, A. Reintjes, W. Muranyi, I. Moarefi und J. Tatzelt. Determinants of the in vivo folding of the prion protein: a bipartite function of helix 1 in folding and aggregation. *J. Biol. Chem.* **278**, 14961–14970 (2003). 89

[130] Ross, C. A., M. A. Poirier, E. E. Wanker und M. Amzel. Polyglutamine fibrillogenesis: The pathway unfolds. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 1–3 (2003). 92

Danksagung

An dieser Stelle möchte ich allen danken, die zum Gelingen meiner Doktorarbeit beigetragen haben:

Mein Doktorvater Paul Tavan hat mir dieses spannende und vielseitige Projekt ermöglicht und die Arbeit durch viele wertvolle Anregungen und durch seinen oftmals inspirierenden Blickwinkel bereichert. Für seinen großen persönlichen Einsatz und die großzügige Finanzierung meiner Doktorarbeit möchte ich ihm herzlich danken.

Armin Giese und Thomas Hirschberger möchte ich für die erfolgreiche und sehr anregende Zusammenarbeit bei den Prion-Projekten danken. Gerald Mathias konnte mir oft mit wertvollen Ratschlägen weiterhelfen und brachte mir in zahlreichen Diskussionen, zusammen mit Bernhard Egwolf, viele Fragestellungen der Elektrostatik näher.

Mein Zimmerkollege Rudolf Reichold stand mir bei allen kleineren und größeren Problemen unermüdlich mit Rat und Tat zur Seite. Für das kritische und sorgfältige Korrekturlesen meiner Doktorarbeit möchte ich mich ganz besonders bei ihm und bei Verena Schultheis bedanken.

Allen derzeitigen und ehemaligen Mitgliedern unserer Arbeitsgruppe danke ich für die stete Diskussions- und Hilfsbereitschaft. Dies waren, neben den bereits oben genannten, Galina Babizki, Sebastian Bauer, Heiko Carstens, Tassilo Christ, Robert Denschlag, Martin Lingenheil, Prof. Marco Nonella, Benjamin Rieff, Matthias Schmitz, Bernhard Schropp, Paul Strodel, Andreas Weiss sowie die Diplomanden und Werkstudenten der Arbeitsgruppe.

Mitglieder der AG Zinth und AG Riedle gewährten mir während unserer gemeinsamen Mensabesuche Einblicke in die Welt der experimentellen biophysikalischen Forschung und haben das sehr angenehme Arbeitsklima am Lehrstuhl für BioMolekulare Optik mit geprägt.

Die Mitarbeiter der Werkstatt unter Leitung von Rudi Schwarz standen mir bei allen handwerklichen Arbeiten bereitwillig und unkompliziert zur Seite. Die Sekretärinnen des Lehrstuhls, Frau Podolski, Frau Michaelis, Frau Klemradt und Frau Widmann-Diermeier, waren mir bei allen organisatorischen Tätigkeiten gerne behilflich. Frau Dr. Lederer und dem Bayerischen Forschungsverbund Prionen möchte ich für die Finanzierung meiner Doktorarbeit danken.

Freunde haben dafür gesorgt, dass ich mich in meiner zuweilen spärlichen Freizeit auch einmal von der Arbeit lösen konnte. Besonders bei meinen Freunden aus dem Schwedischen Chor München sowie bei Luzie Schmiz-Rölz und Martin Rölz möchte ich mich herzlich bedanken. Marcus Jautze stand mir lange zur Seite und hat mich stets unterstützt und ermutigt. Für die Fürsorge möchte ich ihm und vor allem auch meinen Eltern danken.