Entwicklung einer Kontinuumselektrostatik Methode für Molekulardynamik Simulationen

Sebastian Bauer

München 2014

für Rita & Lorenz

Entwicklung einer Kontinuumselektrostatik Methode für Molekulardynamik Simulationen

Sebastian Bauer

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

vorgelegt von
Sebastian Bauer
aus München

München, Juni 2014

Erstgutachter: Prof. Dr. Paul Tavan
Zweitgutachter: Prof. Dr. Martin Zacharias
Tag der mündlichen Prüfung: 26. November 2014

Zusammenfassung

Die Struktur und Dynamik von Proteinen wird wesentlich von elektrostatischen Wechselwirkungen mit dem umgebenden Lösungsmittel bestimmt. Bei Simulationen der Molekulardynamik (MD), in denen die auf die Proteinatome wirkenden Kräfte aus einem atomaren Kraftfeld berechnet werden, wird das Lösungsmittel üblicherweise ebenfalls atomar aufgelöst dargestellt. Alternative, reduzierte und damit potenziell effizientere Beschreibungen verzichten dagegen auf eine solche explizite Beschreibung des Lösungsmittels und betten das betrachtete Protein stattdessen in ein dielektrisches Kontinuum ein. Zur Berechnung der elektrostatischen Wechselwirkungen muss dann die dielektrische Poisson Gleichung (PG) in jedem Integrationsschritt einer MD-Simulation gelöst werden.

Da die PG nur für wenige und sehr einfache Geometrien analytisch gelöst werden kann, wurden diverse numerische Näherungsverfahren etabliert, die jedoch die Reaktionskräfte, die das Kontinuum auf die Proteinatome ausübt, vernachlässigen. Deshalb und wegen ihrer mangelnden Recheneffizienz sind diese Methoden für MD-Simulationen ungeeignet. Sogenannte *Generalized Born* (GB) Methoden erlauben auf der anderen Seite zwar effiziente und dynamisch korrekte Berechnungen elektrostatischer Kräfte, liefern jedoch keine Lösung der PG.

Anknüpfend an Ergebnisse von Egwolf und Tavan (ET, *J. Chem. Phys.* **118**, 2039–2056, 2003) wurde in dieser Arbeit eine neue Darstellung und eine approximative Lösung der PG für Proteine im dielektrischen Kontinuum entwickelt, welche auch die benötigten Reaktionskräfte liefert und deren Effizienz für MD-Simulationen ausreicht. Diese Resultate sind in drei Publikationen dokumentiert /5-7/, die das Kernstück der vorliegenden Dissertation darstellen.

Ausgehend von einer akribischen Analyse des ET-Ansatzes wird in /5/ die PG in eine exakte und atomar aufgelöste Reaktionsfeld-(RF-)Darstellung umformuliert. Einfache Gaußsche Näherungen ermöglichen, wie gezeigt wird, eine effiziente Berechnung der elektrostatischen Wechselwirkungen. Vergleiche mit analytischen und numerischen Lösungen der PG führen dann zu Einsichten in die Qualität der entwickelten Näherungen und die konzeptionellen Schwächen von GB Methoden.

In /6/ wird daraus das als „HADES“ bezeichnete (von engl. *Hamiltonian dielectric solvent*) MD-Verfahren entwickelt. Es wird gezeigt, wie aus den Ergebnissen von /5/ analytische Ausdrücke für approximative RF-Kräfte abgeleitet werden können, welche die Reaktionskräfte einschließen und daher Hamilton'sche, d.h. energie- und impulserhaltende MD-Simulationen ermöglichen. Es wird eine effiziente Implementierung von HADES vorgestellt.

Die letzte Arbeit /7/ demonstriert die Effizienz von HADES-MD anhand von Replika-Austausch-Simulationen zum Schmelzen eines α-helikalen Peptids und skizziert die noch nötigen Schritte zur Optimierung der Parameter von HADES.

Verzeichnis der im Rahmen dieser Arbeit entstandenen Publikationen

/1/ B. Rieff, G. Mathias, S. Bauer, and P. Tavan. Density Functional Theory Combined with Molecular Mechanics: The Infrared Spectra of Flavin in Solution. *Photochem. Photobiol.* **87**, 511–523 (2011).

/2/ B. Rieff, S. Bauer, G. Mathias, and P. Tavan. IR Spectra of Flavins in Solution: DFT/MM Description of Redox Effects. *J. Phys. Chem. B* **115**, 2117–2123 (2011).

/3/ B. Rieff, S. Bauer, G. Mathias, and P. Tavan. DFT/MM Description of Flavin IR Spectra in BLUF domains. *J. Phys. Chem. B* **115**, 11239–11253 (2011).

/4/ M. Schwörer, B. Breitenfeld, P. Tröster, S. Bauer, K. Lorenzen, P. Tavan, and G. Mathias. Coupling DFT to Polarizable Force Fields for Efficient and Accurate Hamiltonian Molecular Dynamics Simulations. *J. Chem. Phys.* **138**, 244103 (2013).

/5/ **S. Bauer, G. Mathias, and P. Tavan. Electrostatics of Proteins in Dielectric Solvent Continua. I. An Accurate and Efficient Reaction Field Description. *J. Chem. Phys.* 140, 104102 (2014)**

/6/ **S. Bauer, P. Tavan, and G. Mathias. Electrostatics of Proteins in Dielectric Solvent Continua. II. Hamiltonian Reaction Field Dynamics. *J. Chem. Phys.* 140, 104103 (2014)**

/7/ **S. Bauer, P. Tavan, and G. Mathias. Exploring Hamiltonian Dielectric Solvent Molecular Dynamics. *Chem. Phys. Lett*, 612, 20 (2014)**

Die durch **Fettdruck** hervorgehobenen Arbeiten sind in den Text der Dissertation eingearbeitet und dort nachgedruckt.

Inhaltsverzeichnis

1 Einleitung

In the beginning the universe was created.
This has made a lot of people very angry
and been widely regarded as a bad move.

Douglas Adams,
The Restaurant at the End of the Universe

Proteine bilden elementare Bestandteile einer jeden lebenden Zelle. Die etwa 10^{10} bis 10^{12} natürlich vorkommenden Varianten der Proteine erfüllen dabei äußerst spezialisierte Aufgaben. Unter anderem steuern sie als enzymatische Katalysatoren chemische Reaktionen in Zellen, sie dienen zum Transport und zur Speicherung von Nährstoffen und erlauben gerichtete Bewegungen von Muskeln sowie die Erzeugung und Weiterleitung von Nervenpulsen. Ferner sind sie an der Immunabwehr beteiligt und tragen als Bestandteile der Haut und der Knochen zur Stabilisierung des Körpers bei [1–4].

All diese hoch spezialisierten Aufgaben können Proteine jedoch nur in ihrer nativen dreidimensionalen Struktur erfüllen. Fehlfaltungen nach der Proteinsynthese führen häufig zur Fehlfunktion oder gar zum Zelltod. Es sind eine Reihe von schwerwiegenden Krankheiten bekannt, die durch solche Fehlfaltungen entstehen [5], wie etwa Alzheimer [6], Creutzfeldt-Jakob [7] oder Parkinson [8, 9]. Trotz großer Anstrengungen sowohl der experimentellen als auch der theoretischen Forschung ist es jedoch noch nicht gelungen, alle Aspekte der Proteinfunktion zu erklären.

1.1 Struktur und Dynamik von Proteinen

Proteine werden in den Ribosomen einer Zelle aus einer festen in der DNA kodierten Sequenz der 20 natürlich vorkommenden Aminosäuren polymerisiert [1]. Die Länge dieser Sequenz variiert dabei von etwa 100 bis zu einigen 1000 Aminosäuren – Titin, das größte bekannte menschliche Protein besteht aus über 30.000 Aminosäuren [10].

Abbildung 1.1 zeigt die chemische Struktur einer sogenannten α-Aminosäure. Sie besteht aus einem zentralen C_α Atom an das neben einer Carboxygruppe (COOH), eine Aminogruppe (NH_2), ein einzelnes Wasserstoffatom (H) und ein für die Aminosäure charakteristisches Residuum (R) gebunden sind. Bei der Peptidsynthese bilden, wie in Abb. 1.1 dargestellt, zwei Aminosäuren unter Wasserabspaltung eine sogenannte Peptidbindung. Das C_α-Atom und die CO-Gruppe der einen und die NH-Gruppe und das C_α-Atom der nächsten Aminogruppe liegen dabei in einer Ebene und bilden zusammen das sogenannte Peptidplättchen (grau hinterlegt). Auf Grund der hohen Elektronegativität der beiden Atome O und N weist das Peptidplättchen ein starkes Dipolmoment auf, das als scharzer Pfeil in Abb. 1.1 dargestellt ist. Durch

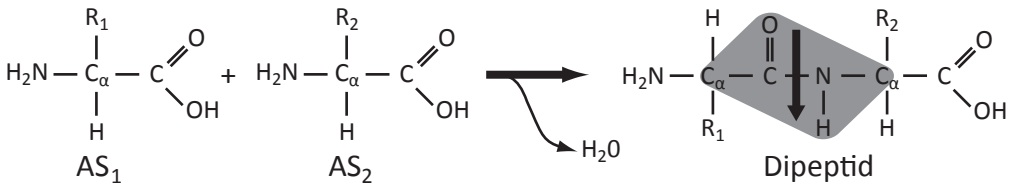

Abbildung 1.1: Synthese von Proteinen und Peptiden aus Aminosäuren: Zwei Animosäuren AS_1 und AS_2, die sich lediglich durch ihre Residuen R_1 und R_2 unterscheiden verbinden sich unter Wasserabspaltung zu einem Dipeptid. Das torsionsstabile Peptidplättchen ist grau dargestellt, die Richtung seines starken Dipolmoments durch einen Pfeil.

stetige Wiederholung dieser chemischen Reaktion in den Ribosomen einer Zelle entstehen bei der Peptidsynthese Proteine [1].

Bei der Faltung eines Proteins bilden sich zunächst in Teilen der Aminosäurekette lokale Strukturmotive, die sogenannte Sekundärstruktur, aus. Diese meist α-helikalen oder β-faltblattartigen Bereiche, die durch elektrostatische Wechselwirkungen der Atome des Proteinrückgrats (engl. *Backbone*) stabilisiert werden, lagern sich bereits während der Synthese oder innerhalb einiger Mikrosekunden bis Sekunden danach zu einer für jedes Protein zumeist wohldefinierten endgültigen räumlichen Struktur, der Tertiärstrukur zusammen [11]. Abbildung 1.2 zeigt schematisch die Tertiärstruktur eines AppA-BLUF Dimers, das aus zwei AppA-BLUF Proteinen besteht. Die α-helikalen Bereiche sind hier violett, die β-faltblattartigen Bereiche gelb dargestellt. Die ungefalteten flexiblen Bereiche des Moleküls (hellgrau und cyan) erlauben Veränderungen der Tertiärstruktur.

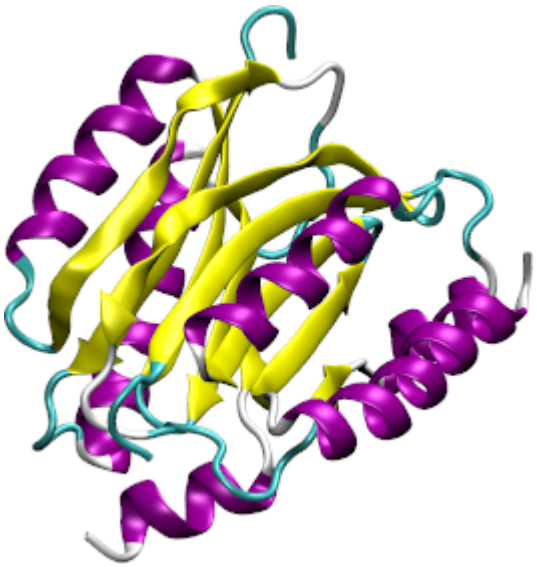

Abbildung 1.2: Darstellung des Proteins AppA-BLUF Dimers. Das Monomer des Proteins besteht aus drei α-helikalen Regionen (violett) und einigen β-faltblattartigen Bereichen (gelb). Cyan und hellgrau sind die ungefalteten flexiblen Teile des Proteinrückgrats dargestellt. Zur Visualisierung wurde das Programmpaket VMD verwendet [12].

Zwar ist die Sequenz der Aminosäuren eines Proteins und somit die Primärstruktur in der DNA festgelegt, ihre Kenntnis allein erlaubt aber noch keine Aussage über die räumliche Struktur und damit die Funktion des Proteins [13, 14]. Obwohl die Torsionsstabilität des Peptidplättchens die Zahl der möglichen räumlichen Anordnungen eines Protein bereits deutlich einschränkt, bleibt eine große Zahl lokaler Minima im Raum möglicher Konformationen [15]. Tatsächlich wurde gezeigt, dass bereits eine reduzierte, gitterbasierte Beschreibung der Proteinfaltung ein NP-vollständiges Problem darstellt [16]. Seit Anfinsen 1973 zeigte, dass Proteine eine eindeutige, wohldefinierte räumliche Struktur aufweisen [17] gilt das sogenannte Faltungsproblem, also die Vorhersage der Tertiärstruktur eines Proteins aus seiner Aminosäuresequenz, daher als eine der wichtigsten wissenschaftlichen Fragestellungen unseres Jahrtausends [18].

Erst die Kenntnis der Tertiärstruktur erlaubt es, die Funktionsweise eines Proteins zu erklären. Daher wurden eine Reihe von biophysikalischen Methoden zur Strukturaufklärung, wie die Röntgenstrukturanalyse [19, 20], die Kernspinresonanz (NMR) [21, 22] oder die zeitaufgelöste Spektroskopie [23] entwickelt. Jedoch hat jede dieser Methoden entscheidende Nachteile, die eine genaue Bestimmung der dreidimensionalen Struktur von Biomolekülen erschweren.

Die Röntgenstrukturanalyse bestimmt die atomare Struktur eines Proteins aus dem Beugungsmuster von Röntgenstrahlen an einem Proteinkristall. Dazu muss das Protein kristallisiert und bei tiefen Temperaturen untersucht werden [24, 25]. Es ist also nicht klar, ob auf Grund der nicht-physiologsichen Umgebung die nativ-funktionale Struktur des Proteins bestimmt wird. Ferner reicht die Auflösung der Methode nicht aus, um auch die Position von Wasserstoffatomen zu bestimmen [26–29].

Die Strukturaufklärung mit NMR basiert auf der Messung paarweiser Atomabstände anhand der Dipol-Dipol-Kopplung atomarer Kernspins [21, 22, 30]. Dabei ist es im Gegensatz zur Röntgenstrukturanalyse auch möglich, die Orte von Wasserstoffatomen zu ermitteln. Die Bestimmung der vollständigen dreidimensionalen Struktur aus den Atomabständen ist jedoch nicht eindeutig [31–33] und wird mit zunehmender Größe des Moleküls schwieriger, so dass ein weiterer empirischer Strukturverfeinerungsprozess notwendig ist [34, 35].

Spektroskopische Methoden dienen häufig eher der Untersuchung chemischer Reaktionen oder Anregungen in Biomolekülen. Indirekt liefern aber auch sie Informationen über die Struktur eines Moleküls und seiner Umgebung, die in den jeweiligen Spektren kodiert sind. So ergeben etwa die angesprochenen Sekundärstrukturelemente α-Helix und β-Faltblatt unterschiedliche Spektren im Infrarotbereich. Auf Grund der großen Zahl möglicher Schwingungsmoden eines Proteins ist die Extraktion dieser verschlüsselten Strukturinformation jedoch schwierig.

1.2 Molekulardynamik Simulationen

Einen direkten Zugang zur Untersuchung von Struktur und Dynamik von Proteinen stellen Molekulardynamik (MD) Simulationen dar [36, 37]. Dabei ist es möglich, die thermischen Bewegungen der Proteinatome in ihrer nativen Lösungsmittelumgebung mit atomarer Auf-

lösung zu untersuchen. Eine quantenmechanische Beschreibung der Kräfte in einem Peptid–Lösungsmittelsystem ist, auf Grund des damit verbundenen Rechenaufwands, selbst mit modernen Computern nur für Systeme mit einigen 100 bis wenigen 1000 Atomen möglich und beschränkt sich typischerweise auf Zeitskalen von wenigen Nanosekunden.

Liegt der Fokus der Untersuchung nicht auf chemischen Reaktionen sondern lediglich auf der Konformationsdynamik, also zum Beispiel der Faltung oder Umfaltung eines Proteins, ist häufig eine einfachere und daher effizientere Beschreibung der Wechselwirkungen ausreichend. Dabei werden die Atome als klassische Teilchen behandelt, die sich in einem empirischen Potential, dem sogenannten molekülmechanischen (MM) Kraftfeld, $U(\mathbf{R} \,|\, \mathbf{A})$ bewegen. Dieses hängt nur von den Kernkoordinaten $\mathbf{R} = \{\mathbf{r}_1, \ldots, \mathbf{r}_N)$ der N Atome ab und approximiert mit einfachen analytischen Funktionen die Wechselwirkungen der Atome. Die Parameter $\mathbf{A}$ eines solchen *all-atom* Kraftfeldes $U(\mathbf{R} \,|\, \mathbf{A})$ werden durch quantenmechanische Rechnungen und durch Vergleich mit experimentellen Daten anhand einfacher Beispielmoleküle bestimmt [38–40]. Einen Überblick über verschiedene Kraftfelder, die in MM/MD-Simulationen von Proteinen verwendet werden, findet man beispielsweise in [41]. Aus dem Potential $U(\mathbf{R} \,|\, \mathbf{A})$ können nun die Newton'schen Bewegungsgleichungen

$$m_i \ddot{\mathbf{r}}_i = -\nabla_i U(\mathbf{R} \,|\, \mathbf{A}) \tag{1.1}$$

für ein Teilchen i mit Masse m_i berechnet und numerisch, z.B. mit dem Verlet-Algorithmus [42], integriert werden, so dass man eine zeitlich diskretisierte Trajektorie der Proteindynamik erhält.

Typische MM Kraftfelder unterteilen die Wechselwirkungsenergie

$$U(\mathbf{R} \,|\, \mathbf{A}) = U_{\mathrm{B}}(\mathbf{R}) + U_{\mathrm{NB}}(\mathbf{R}) \tag{1.2}$$

in die sogenannten gebundenen oder *bonded* Anteile $U_{\mathrm{B}}(\mathbf{R})$, die etwa Schwingungen kovalent gebundener Atome sowie Winkelschwingungen und die Torsion um die Dihedralwinkel umfassen, und die langreichweitigen nicht gebundenen oder *non bonded* Anteile $U_{\mathrm{NB}}(\mathbf{R})$. Der Aufwand für die Auswertung der kurzreichweitigen *bonded* Anteile skaliert linear mit der Zahl N der Atome, da jedes Atom eine geringe und feste Zahl von Wechselwirkungspartnern besitzt und ist somit für die Effizienz einer MM/MD-Simulation unkritisch.

Die langreichweitigen Anteile der Wechselwirkungsenergie

$$U_{\mathrm{NB}}(\mathbf{R}) = U_{\mathrm{vdW}}(\mathbf{R}) + U_{\mathrm{elec}}(\mathbf{R}) \tag{1.3}$$

setzen sich aus der van der Waals (vdW) Wechselwirkung $U_{\mathrm{vdW}}(\mathbf{R})$ und der Coulomb Wechselwirkung $U_{\mathrm{C}}(\mathbf{R})$ zusammen. Dabei wird $U_{\mathrm{vdW}}(\mathbf{R})$ häufig durch das Lennard Jones (LJ) Potential

$$U_{\mathrm{LJ}}(\mathbf{R}) = \sum_{i<j} 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right], \tag{1.4}$$

approximiert, das nur vom Abstand $r_{ij} = |\mathbf{r}_i - \mathbf{r}_j|$ zweier Atome i und j abhängt. Es beschreibt die bindende, zu r_{ij}^{-6} proportionale Dispersionswechselwirkung und die abstoßende, zu r_{ij}^{-12} proportionale Pauli-Repulsion. Die Parameter ϵ_{ij} und σ_{ij} bestimmen die Tiefe und den Nulldurchgang der jeweiligen Paarwechselwirkung von $U_{\mathrm{LJ}}(\mathbf{R})$.

Auf Grund der kurzen Reichweite des r_{ij}^{-6} Wechselwirkungsterms wird das LJ Potential in MM/MD-Simulationen häufig ab einem bestimmten Abstand R_c vernachlässigt. Dieses Abschneiden liefert dann für jedes Atom eine konstante Zahl von Wechselwirkungspartnern, so dass auch hier ein lineares Skalierungsverhalten des Rechenaufwands mit der Zahl N der Atome erreicht wird. Andererseits kann das LJ Potential auch durch Taylor-Entwicklung in eine schnelle Multipolmethode eingebunden werden, die dann ebenfalls linear mit N skaliert und keine Abschneideartefakte zeigt [43].

Die elektrostatische Wechselwirkungsenergie

$$U_{\text{elec}} = \frac{1}{2} \sum_i q_i \Phi(r_i) \tag{1.5}$$

ergibt sich aus der Wechselwirkung der Partialladungen q_i mit dem elektrostatischen Potential $\Phi(\mathbf{r})$. Für eine Beschreibung der atomaren Partialladungen als Punktladungen, wie sie in den meisten MM Kraftfeldern verwendet wird, ist das elektrostatische Potential am Ort $\mathbf{r}$ gegeben durch das Coulomb Potential

$$\Phi^{\text{C}}(\mathbf{r}) = \sum_j \frac{q_j}{|\mathbf{r} - \mathbf{r}_j|} \tag{1.6}$$

das von den Ladungen q_j aller anderen Atome $j \neq i$ erzeugt wird. Im Gegensatz zum vdW Potential fällt die elektrostatische Wechselwirkung deutlich langsamer mit r^{-1} ab und es ist daher nicht möglich, diese Wechselwirkung abzuschneiden. Bei einer naiven Auswertung von Gleichung (1.5) als Summe von Paarwechselwirkungen skaliert der Rechenaufwand für die Berechnung der elektrostatischen Energie mit $\mathcal{O}(N^2)$. Zwar existieren Methoden, wie etwa die Ewald-Summation [44] oder schnelle Multipol-Methoden [43, 45–48], um den mit der Berechnung der Coulomb Wechselwirkung verbundenen Rechenaufwand auf $\mathcal{O}(N \log N)$ beziehungsweise $\mathcal{O}(N)$ zu reduzieren. Die Berechnung der elektrostatischen Wechselwirkungen bleibt aber auf Grund der großen Zahl an Wechselwirkungspartnern, die bei MM/MD-Simulationen von Protein/Lösungsmittel-Systemen zu berücksichtigen sind, der die Geschwindigkeit bestimmende Faktor.

Durch die Verwendung eines MM Kraftfelds konnten McCammon, Gelin und Karplus bereits 1977 eine erste MM/MD-Simulation des Trypsin-Inhibitors aus der Bauchspeicheldrüse von Rindern (BPTI, von engl. *Bovine Pancreatic Trypsin Inhibitor*) durchführen [49]. Bei dieser Simulation, die eine Zeitspanne von etwa 9 ps abdeckte, wurde BPTI jedoch im Vakuum simuliert, d.h. alle Einflüsse des umgebenden Lösungsmittels wurden vernachlässigt. Nichtsdestoweniger ermöglichte diese Pionierarbeit bereits Einsichten in die Dynamik von Proteinen [49].

Die physiologische Umgebung des BPTI, wie auch der meisten anderen Proteine, ist jedoch eine im wesentlichen wässrige Lösung. Diese hochpolare Umgebung steuert die elektrostatischen Wechselwirkungen und damit die Bildung und Stabilisierung der Sekundär- und Tertiärstruktur. Daher darf sie bei der Beschreibung von Faltungsprozessen keinesfalls vernachlässigt werden. In MM/MD-Simulationen wird das Protein deshalb meist von einer großen Zahl von Lösungsmittelmolekülen umgeben.

Abbildung 1.3 zeigt beispielhaft ein typisches Simulationssystem. Das bereits in Abb. 1.2 dargestellte AppA-BLUF Dimer ist dort von etwa 10000 TIP3P [50, 51] Wassermodellen

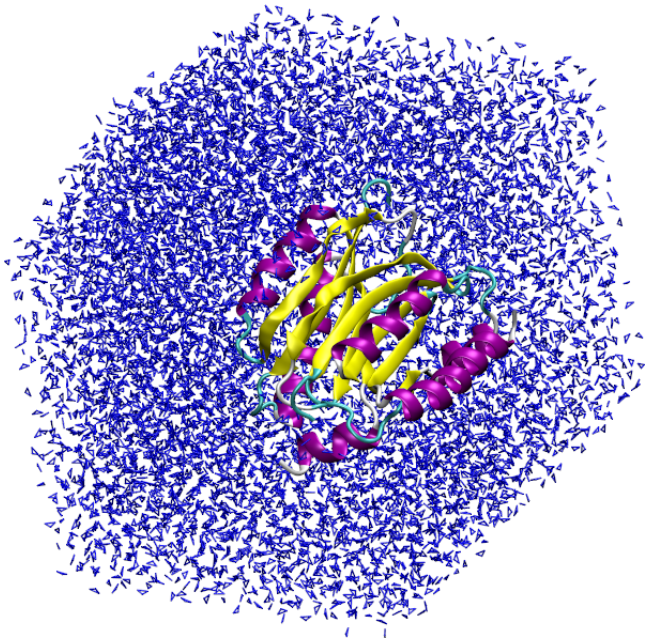

Abbildung 1.3: Ein typisches MM/MD-Simulationssystem. Das AppA-Dimer aus Abb. 1.2 ist von 10000 Wassermolekülen umgeben. Die Zahl der Wasseratome übersteigt die der Proteinatome um etwa einen Faktor zehn. Zur Visualisierung wurd das Programmpaket VMD verwendet [12].

(blau dargestellt) umgeben. Um physiologische Proteinkonzentrationen zu erreichen, muss die Zahl der Lösungsmittelatome die Zahl der Proteinatome, wie in Abb. 1.3, mindestens um eine Größenordnung übertreffen [52–54], so dass im Wesentlichen leicht verunreinigtes lauwarmes Wasser simuliert wird.

Um das Faltungsproblem zu lösen, d.h. um mit MD-Simulationen die Tertiärstruktur eines Proteins aus seiner Primärstruktur vorher zu sagen, müssen diese bis in den Bereich von Mikro- oder gar Millisekunden ausgedehnt werden. Die an der Proteinfaltung beteiligten dynamischen Prozesse laufen jedoch auch auf wesentlich kürzeren Zeitskalen ab [23, 55, 56].

In einer MD-Simulation ist der Zeitschritt Δt der numerischen Integration durch die schnellsten Freiheitsgrade des Systems nach oben beschränkt und liegt auf Grund der schnellen Bindungs- und Winkelschwingungen bei etwa einer Femtosekunde. Die aufwändige Berechnung der langreichweitigen elektrostatischen Wechselwirkungen zusammen mit der großen Zahl von Lösungsmittelatomen begrenzen deshalb die in MM/MD-Simulationen erreichbaren Zeitskalen auf einige 100 ns. Faltungs- und Umfaltungsprozesse dauern allerdings typischerweise Mikrosekunden bis Sekunden, so dass die atomar aufgelöste MM/MD-Simulation der Faltung eines Proteins in Lösung noch nicht in greifbarer Nähe ist.

1.3 Kontinuumselektrostatik

In MM/MD-Simulationen ist die thermische Bewegung der Lösungsmittelatome häufig nicht von Interesse. Wegen der großen Polarität von Wasser, das den Hauptanteil des Lösungsmit-

tels von Proteinen ausmacht, müssen jedoch zumindest die mittleren elektrostatischen Eigenschaften der Lösungsmittelumgebung berücksichtigt werden. Dazu kann die explizite, atomar aufgelöste Darstellung des Lösungsmittels durch eine Molekularfeld-Näherung ersetzt und das Protein in ein dielektrisches Kontinuum eingebettet werden [57–59].

Abbildung 1.4 zeigt eine solche Beschreibung schematisch am Beispiel des AppA-BLUF Dimers. Die Wassermoleküle aus Abbildung 1.3 wurden hier durch ein grau dargestelltes dielektrisches Kontinuum ersetzt. Dadurch wurde die Zahl der Atome im Simulationssystem drastisch reduziert.

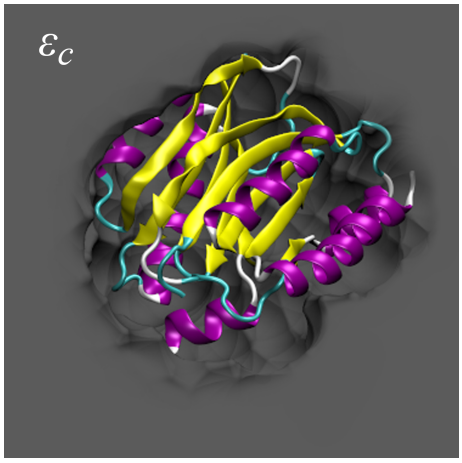

Abbildung 1.4: Das AppA-BLUF Dimer umgeben von einem dielektrischen Kontinuum mit der Dielektrizitätskonstante ε_c (grau). Die Darstellung erlaubt einen Blick in das Proteinvolumen $\mathfrak{V}_s$. Der komplizierte Rand des Volumens $\mathfrak{V}_s$ ist angedeutet. Zur Visualisierung wurd das Programmpaket VMD verwendet [12].

Zur Auswertung der elektrostatischen Energie (1.5) eines solchen Systems ist es nötig, das elektrostatische Potential $\Phi(\mathbf{r})$ als Lösung der dielektrischen Poisson-Gleichung (PG)

$$\nabla \cdot [\varepsilon(\mathbf{r}) \nabla \Phi(\mathbf{r})] = -4\pi\rho(\mathbf{r}) \tag{1.7}$$

zu berechnen. Dabei setzt sich die Ladungsverteilung

$$\rho(\mathbf{r}) = \sum_i q_i \delta(\mathbf{r} - \mathbf{r}_i) \tag{1.8}$$

in üblichen MM/MD Kraftfeldern aus den Punktladungen q_i der Atome i zusammen und die dielektrische Funktion

$$\varepsilon(\mathbf{r}) = \varepsilon_c - (\varepsilon_c - \varepsilon_s)\Theta(\mathbf{r}) \tag{1.9}$$

wird durch die charakteristische Funktion

$$\Theta(\mathbf{r}) = \begin{cases} 1 & \text{falls} \quad \mathbf{r} \in \mathfrak{V}_s \\ 0 & \text{sonst} \end{cases} \tag{1.10}$$

des Proteinvolumens $\mathfrak{V}_s$ beschrieben. Sie nimmt innerhalb von $\mathfrak{V}_s$ den Wert ε_s und außerhalb den Wert ε_c an.

Die Lösung $\Phi(\mathbf{r})$ der PG (1.7) lässt sich gemäß

$$\Phi(\mathbf{r}) = \Phi^{\mathrm{C}}(\mathbf{r}) + \Phi^{\mathrm{RF}}(\mathbf{r}) \tag{1.11}$$

als Summe des Coulomb-Potentials $\Phi^{\mathrm{C}}(\mathbf{r})$ und des Reaktionsfeld-(RF-)Potentials $\Phi^{\mathrm{RF}}(\mathbf{r})$ darstellen, wobei letzteres als Differenz der Lösungen $\Phi(\mathbf{r} \,|\, \varepsilon_s, \varepsilon_c)$ der PG zu den Werten $\varepsilon_c \neq \varepsilon_s$ und $\varepsilon_c = \varepsilon_s$ definiert ist. Das RF-Potential erfasst damit den elektrostatischen Einfluss des umgebenden dielektrischen Kontinuums auf die Proteinatome.

Die Verwendung der Kontinuumselektrostatik bei MD-Simulationen vernachlässigt sowohl die granulare molekulare Struktur des Lösungsmittels, die insbesondere an der Oberfläche des gelösten Proteins einen wichtigen Einfluss haben sollte, als auch die endliche Dauer der dielektrischen Relaxation [60], die im Bereich von Femtosekunden bis Pikosekunden abläuft. Während die Bedeutung der Granularität des Lösungsmittels nur durch die Verfügbarkeit einer verlässlichen und effizienten Kontinuumsmethode und den Vergleich mit expliziten Darstellungen des Lösungsmittels abschätzbar ist, sollte die vernachlässigte dielektrische Relaxation für die Beschreibung der Konformationsdynamik von Proteinen von geringer Bedeutung sein. Diese Dynamik läuft nämlich typischerweise auf Zeitskalen weit oberhalb von Pikosekunden ab und weist im Zeitbereich dieser Relaxation lediglich Strukturfluktuationen um atomare Gleichgewichtslagen auf. Bei diesen Fluktuationen sollten sich die durch die Vernachlässigung der dielektrischen Relaxation entstandenen Fehler herausmitteln.

Für sehr wenige Beispiele mit einfachen Geometrien des Volumens $\mathfrak{V}_s$ ist die Lösung der PG analytisch möglich. Das wohl berühmteste Beispiele hierfür ist die Lösung von Max Born aus dem Jahr 1920 für ein isoliertes sphärisches Ion [61]. Ein erstes Modell für die Elektrostatik von Proteinen in dielektrischen Medien lieferte Kirkwood im Jahr 1934 [62]. Er fand die Lösung der PG für eine beliebige Ladungsverteilung in einem sphärischen Volumen.

Abbildung 1.5 zeigt für diesen Fall die Lösungsenergie einer Einheitsladung e im Abstand r vom Mittelpunkt einer Kugel mit Radius $R = 0.89\,\mathrm{nm}$. Die Abbildung zeigt, dass sich die Energie für eine Ladung nahe dem Rand der Kugel am stärksten ändert. Als Modell eines Proteins ist die Kirkwood-Kugel jedoch ungeeignet, da die hydrophilen Seitenketten eines Proteins in das Lösungsmittel eindringen und die Oberfläche eines Proteins deshalb nicht glatt sphärisch sondern rau ist. Abbildung 1.4 zeigt diesen Umstand am Beispiel des AppA-BLUF Dimers und macht deutlich, dass die Geometrien von Proteinvolumina $\mathfrak{V}_s$ kompliziert sind. Für solche Geometrien existieren aber keine analytischen Lösungen der PG.

Es gibt aber eine Vielzahl von numerischen Methoden, um das elektrostatische Potential $\Phi(\mathbf{r})$ für beliebige Geometrien zu berechnen. Diese Methoden lassen sich in zwei Kategorien einteilen. Gittermethoden diskretisieren das Proteinvolumen $\mathfrak{V}_s$ und einen Teil der angrenzenden Lösungsmittelumgebung durch ein reguläres Gitter. Die PG (1.7) kann dann lokal approximiert werden, sodass das elektrostatische Potential $\Phi(\mathbf{r})$ auf den Gitterpunkten als Lösung eines linearen Gleichungssystems bestimmt werden kann. Diese Lösung kann entweder direkt oder durch Iteration gefunden werden [63–69].

Randelementmethoden (BEM, von engl. *boundary element methods*) verwenden eine Integraldarstellung der PG (1.7) und eine Diskretisierung der Oberfläche des Proteins zur Berechnung des Potentials [70–76]. Dabei werden auf den Elementen der Oberfläche von $\mathfrak{V}_s$

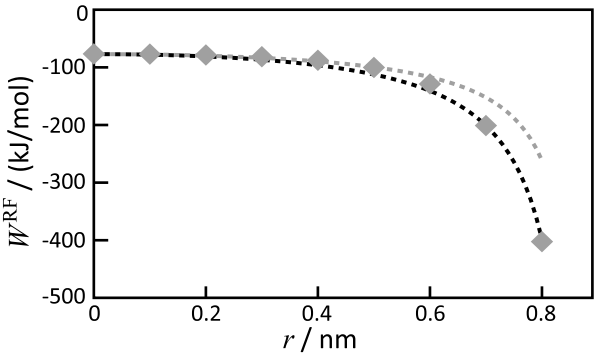

Abbildung 1.5: Energie einer Einheitsladung e im Abstand r vom Kugelmittelpunkt einer Kugel mit Radius R, die in einem Dielektrikum ε_c eingebettet ist. Die Abbildung zeigt die analytische Lösung von Kirkwood (schwarz gestrichelt), sowie die Approximation von Egwolf und Tavan (graue Vierecke) und das Ergebnis der Generalized Born Methode (grau gestrichelt).

geeignete Oberflächenladungen bestimmt, die das RF des Lösungmittelkontinuums beschreiben.

Beide Verfahren liefern sehr genaue Ergebnisse bei der Bestimmung der elektrostatischen Energie einer festen Konfiguration $\mathbf{R}$ der N Atome eines Proteins und sind dann jedoch mit großem Rechenaufwand verbunden. Weitere Probleme sind numerische Ungenauigkeiten bei der Berechnung von Feldern und Kräften und die Schwierigkeit den dielektrischen Druck, also die Reaktio-Kräfte auf die Atome, zu berücksichtigen. Daher führen solche Methoden nicht zu Hamilton'schen MD-Simulationen [77–79].

Einen alternativen Ansatz zu den diskutierten numerischen Kontinuumsmethoden liefern sogenannte generalized Born (GB) Methoden [80–91], welche speziell für den Einsatz in MD-Simulationen konzipiert sind und beeindruckende Simulationsdauern erlauben [92]. GB-Methoden verwenden eine heuristische Verallgemeinerung der Lösung der PG für isolierte Ionen [61] in einem Dielektrikum auf beliebige Molekülgeometrien. Sie bieten folglich keine Lösung der PG.

Stattdessen bestimmen sie die RF-Energie

$$\Delta G^{\mathrm{B}}(\mathbf{R}) = -\frac{1}{2}\left(\frac{1}{\varepsilon_s} - \frac{1}{\varepsilon_c}\right)\sum_{i,j}\frac{q_i q_j}{\sqrt{r_{ij}^2 + R_i R_j \exp(r_{ij}^2/4R_iR_j)}}. \tag{1.12}$$

aus paarweise abgeschirmten Wechselwirkungen. Eine zentrale Rolle spielen dabei die effektiven Born-Radien R_i, deren Wahl die Qualität der Näherung (1.12) wesentlich bestimmt [90, 93–95]. Es wurden eine Reihe von Methoden zur Berechnung der R_i vorgeschlagen [80–91]. Dennoch liefert Gleichung (1.12) auch bei Verwendung sogenannter perfekter Born-Radien keine exakte Lösung der PG [96, 97].

Es ist daher nicht verwunderlich, dass sich die aus GB-MD-Simulationen gewonnenen freien Energielandschaften stark von denen aus Simulationen mit expliziten Lösungsmitteln unterscheiden [93–95, 98–101]. Abbildung 1.5 zeigt schon für das einfache, aber für theoretische

Untersuchungen wichtige, Beispiel der Kirkwood-Kugel die großen Abweichungen der Vorhersagen einer GB-Methode [85], welche die Coulomb-Feld-Approximation verwendet [82], von der analytischen Lösung.

Ein vielversprechender Ansatz zur approximativen Lösung der PG (1.7) wurde aufbauend auf Konzepten von Sklenar [102] von Egwolf und Tavan (ET) [103] entwickelt und von Stork und Tavan (ST) [104, 105] erweitert. Dieser Ansatz beruht auf der exakten Umformulierung der PG (1.7) in die Form

$$\Delta\Phi(\mathbf{r}) = -\frac{4\pi}{\varepsilon_c}\left[\rho(\mathbf{r}) - \nabla\cdot\mathbf{P}(\mathbf{r})\right], \tag{1.13}$$

wobei die Dipoldichte $\mathbf{P}(\mathbf{r})$ die Selbstkonsistenzbedingung

$$\mathbf{P}(\mathbf{r}) = \chi_e\Theta(\mathbf{r})\nabla\Phi(\mathbf{r}) \tag{1.14}$$

mit der Suszeptibilität

$$\chi_e = (\varepsilon_c - \varepsilon_s)/4\pi \tag{1.15}$$

erfüllen muss. Gleichung (1.14) zeigt unmittelbar, dass es sich bei dieser Dipoldichte um eine Antipolarisation handelt, die durch den Faktor $\Theta(\mathbf{r})$ auf das Proteinvolumen $\mathfrak{V}_s$ eingeschränkt ist.

Nach der PG (1.13) kann der Einfluss eines Dielektrikums im Außenraum von $\mathfrak{V}_s$ also vollständig durch die Antipolarisationsdichte $\mathbf{P}(\mathbf{r})$ dargestellt werden. Außerdem ist es damit möglich, die natürliche Diskretisierung des Volumens $\mathfrak{V}_s$ durch die Atome zu verwenden, um das Potential $\Phi(\mathbf{r})$ durch die Potentiale atomarer Partialladungen q_i und atomarer Antipolarisationsdichten $\mathbf{P}_i(\mathbf{r})$ darzustellen [103].

Die von ET eingeführten Näherungen der atomaren Antipolarisationsdichten durch Gauß'sche Dipolverteilungen

$$\tilde{\mathbf{P}}_i(\mathbf{r}) = \tilde{\mathbf{p}}_i G(\mathbf{r}\,|\,\mathbf{r}_i,\sigma_i) \tag{1.16}$$

mit den normierten Gaußfunktionen

$$G(\mathbf{r}\,|\,\mathbf{r}_i,\sigma_i) = \frac{1}{(2\pi\sigma^2)^{3/2}}\exp\left(-\frac{(\mathbf{r}-\mathbf{r}_i)^2}{2\sigma_i}\right) \tag{1.17}$$

lieferten, wie Abb. 1.5 zeigt, für ein Modell der Kirkwood-Kugel bereits hervorragende Ergebnisse in Bezug auf die RF-Energien. Leider wurden durch die Verwendung der sogenannten $q\mathbf{E}$-Kräfte jedoch die Reaktio-Kräfte und damit der dielektrische Druck vernachlässigt, so dass mit dem ET Ansatz keine Hamilton'schen RF-MD-Simulationen möglich waren [65].

Die Erweiterung der Methode durch ST konnte diesen Fehler teilweise beheben. Hier wurde zunächst durch eine Umformulierung der ET Methode die konzeptionelle Ähnlichkeit des Ansatzes zu gewöhnlichen polarisierbaren Kraftfeldern gezeigt. Dies ermöglichte die wesentliche Erkenntnis, dass die atomaren RF-Dipoldichten nicht nur als Quellen des RF-Potentials, sondern als reale wechselwirkende Dipole anzusehen sind. Aus der Betrachtung des Grenzfalls isolierter Atome konstruierten ST Kräfte, die zwar Newtons Reaktionsprinzip genügen, die jedoch immer noch nicht direkt durch

$$\mathbf{f}_i = -\nabla_i W^{\mathrm{RF}}(\mathbf{R}) \tag{1.18}$$

aus der RF-Energie abgeleitet waren. Dadurch waren zwar erste RF-MD-Simualtionen an dem einfachen Dipeptid Ac-Ala-NHMe möglich, die bereits vielversprechende Ergebnisse lieferten, die aber keine stabile energieerhaltende Dynamik erlaubten. Trotz dieses vielversprechenden Ansatzes war es also nicht möglich, eine Methode zur Berechnung von atomaren RF-Kräften für Hamilton'sche MD-Simulationen abzuleiten.

1.4 Ziele und Überblick

Diese Arbeit ist motiviert von dem Wunsch die Konformationsdynamik von Proteinen am Computer zu simulieren. Daraus ergibt sich die Notwendigkeit eine effiziente und dennoch genaue Kontinuumsmethode für MD-Simulationen zu entwickeln. Die angesprochenen Schwierigkeiten der vorhandenen und etablierten Methoden auf diesem Gebiet zeigen, wie schwierig es ist, einen akzeptablen Kompromiss zwischen Genauigkeit und Effizienz zu finden.

Genaue Gittermethoden scheiden als Ausgangspunkt für eine solche Methode auf Grund ihres hohen Rechenaufwands und des Fehlens analytischer Kraftausdrücke, welche auch die Reaktionskräfte einschließen, aus. Die unter Umständen effizienten GB-Methoden erlauben zwar energieerhaltende MD-Simulationen, liefern aber bei der Berechnung der Lösungsmittelenergien nicht die nötige Genauigkeit, da sie die PG (1.7) nicht lösen.

Die vielversprechenden Ergebnisse von ET und ST und die konzeptionelle Nähe ihrer Methode zu polarisierbaren Kraftfeldern lassen sie als sinnvollen Ausgangspunkt erscheinen. Ziel dieser Arbeit war es daher, aufbauend auf den Konzepten von ET/ST eine Kontinuumsmethode zu entwickeln, die nicht nur die effiziente und genaue Berechnung von RF-Energien, sondern darüber hinaus auch analytisch darstellbare atomare Kräfte liefert, welche die Reaktionskräfte einschließen und somit Hamilton'sche RF-MD-Simulationen von Peptiden und Proteinen erlauben. Dieses Ziel erzwang eine detaillierte Untersuchung der beiden Vorgängermethoden sowie eine gründliche Umformulierung der Theorie, die in Kapitel 2 beschrieben wird.

Die in Abschnitt 2.1 abgedruckte Publikation [97] erläutert diese Neuformulierung im Detail und leitet aus der exakten ET Theorie eine exakte atomare RF-Darstellung der PG (1.7) her. In dieser Darstellung wird das RF-Potential durch atomare radiale Abschirmladungsverteilungen und atomare antipolarisierbare Dipoldichten beschrieben. Es wird gezeigt, welchen Bedingungen diese Ladungsverteilungen und Dipoldichten genügen müssen, um zu gewährleisten, dass das elektrostatische Potential die PG (1.7) exakt löst. Zudem wird dargestellt, wie diese atomaren Dichten geeignet approximiert werden können, um zu einer effizienten Methode zur Berechnung des RF-Potentials $\Phi^{\mathrm{RF}}(\mathbf{r})$ für Peptide und Proteine zu gelangen. Die Genauigkeit der Methode wird anhand von Vergleichen mit analytischen und numerischen Ergebnissen für einige Beispiele, unter anderem für die bereits in Abschnitt 1.3 angesprochene Kirkwood-Kugel und das Dipeptid Ac-Ala-NHMe, gezeigt. Die komplementären Konstruktionsfehler der ursprünglichen Formulierung von ET und der etablierten GB-Methoden werden diskutiert.

Abschnitt 2.3 zeigt im Anschluss die nötigen theoretischen und mathematischen Konzepte, die

es ermöglichen aus der Approximation der RF-Energie atomare Kräfte analytisch zu berechnen. Wegen der Selbstkonsistenzbedingungen, denen unter anderem die anti-polarisierbaren RF-Dipole genügen müssen, ist es nicht möglich, die Kräfte direkt aus Gleichung (1.18) zu berechnen. Stattdessen werden die durch Iteration bestimmten atomaren Größen als freie Variablen behandelt, die jenen Zwangsbedingungen unterliegen, welche durch die Selbstkonsistenzbedingungen gegeben sind. Die sich daraus ergebenden Lagrange'schen Bewegungsgleichungen ermöglichen so die Berechnung von Lagrange-Multiplikatoren und damit die Berechnung der atomaren Kräfte. Außerdem ist es so möglich, eine alternative Formulierung $\breve{W}^{\mathrm{RF}}$ für die RF-Energie W^{RF} anzugeben, die es erlaubt, die RF-Kräfte direkt aus Gleichung (1.18) zu berechnen.

Anhand diverser MD-Simualtionen wird die energieerhaltende Struktur der neuen HADES (von engl. *Hamiltonian dielectric solvent*, Hamilton'sches dielektrisches Lösungsmittel) Methode nachgewiesen. Es wird geprüft, welche Konvergenzkriterien nötig sind, um energieerhaltenende Simulationen durchführen zu können. Des weiteren zeigt diese Arbeit, wie HADES recheneffizient in das bestehende MD Softwarepacket IPHIGENIE [48, 106] integriert werden konnte. Die dabei verwendeten Methoden wurden in IPHIGENIE auch zur Beschleunigung der Berechnung polarisierbarer Kraftfelder zweckentfremdet. Ein Vergleich der freien Energielandschaft von Ac-Ala-NHMe aus einer HADES-MD-Simulation mit der aus einer expliziten Lösungsmittelsimulation zeigt das Potential der entwickelten Methode.

Die in Kapitel 3 eingebundene Veröffentlichung stellt eine Anwendung der HADES Methode auf ein α-helikales Modelpeptid aus 10 Aminosäuren (150 Atomen) vor. Schon für die in [97] entwickelte und nicht optimierte Parametrisierung von HADES korrelieren die mit HADES und der Gittermethode DelPhi berechneten RF-Energien recht gut. Replica Exchange Simulationen, die sowohl mit HADES als auch mit explizitem Lösungsmittel durchgeführt wurden, zeigen, dass schon die „erratene" Parametrisierung vergleichbare Schmelzkurven liefert. Es wird allerdings auch klar, dass eine Verbesserung der Parametrisierung wichtig und möglich ist.

Kapitel 4 fasst die wichtigsten Ergebnisse dieser Arbeit zusammen, gibt Aufschluss über die Schritte, die zur Optimierung der vorgestellten Methode noch nötig sind, und skizziert ihre Erweiterung auf den Fall der linearisierten Poisson-Boltzmann-Gleichung. Letztere bietet sich beispielweise für MD-Simulationen von hochgeladenen Molekülen, wie etwa der DNA, an, da dann auch eine Darstellung eines ionenhaltigen Lösungsmittels möglich wird [107].

2 Theorie

"Begin at the beginning," the King said, gravely, "and go on till you come to an end; then stop."

Lewis Carroll,
Alice in Wonderland

Zunächst sollen hier die beiden Arbeiten vorgestellt werden, die zeigen, wie die PG von Peptiden und Proteinen, welche in dielektrische Kontinua eingebettet sind, exakt in eine atomare RF-Darstellung umformuliert werden kann, aus der unmittelbar eine einfach berechenbare Näherungslösung folgt, (Abschnitte 2.1 und 2.2) und wie diese approximative Lösung effiziente und Hamilton'sche MD Simulationen ermöglicht (Abschnitt 2.3).

2.1 Die Reaktionsfelddarstellung der Kontinuumselektrostatik von Proteinen

Die nachfolgende Publikation[1]

„Electrostatics of proteins in dielectric sovent continua.
I. An accurate and efficient reaction field description"

Sebastian Bauer, Gerald Mathias, and Paul Tavan
J. Chem. Phys. **140**, 104102 (2014)

die ich zusammen mit Gerald Mathias und Paul Tavan verfasst habe, entwickelt für Proteine in dielektrischen Kontinua eine neue Darstellung des RF-Potentials, die einen früheren Ansatz von Egwolf und Tavan (ET) [103] aufgreift, entscheidend erweitert und dadurch korrigiert. Dabei wird zunächst eine exakte atomare RF-Darstellung der dielektrischen PG und ihrer Lösung hergeleitet. Einfache, analytisch behandelbare Beispiele motivieren diese Darstellung, welche die Stärken und Schwächen des ET-Ansatzes sowie der weit verbreiteten *Generalized Born* (GB) Methoden [80, 82] erklärt.

In Anlehnung an das ET-Verfahren wird anschließend eine approximative Darstellung entwickelt, die eine effiziente Berechnung der RF-Energien von Proteinen und Peptiden erlaubt. Anhand geeigneter Beispiele werden Resultate dieser neuen Näherungsmethode mit analytischen und numerischer Beschreibungen verglichen, so dass ihre günstigen Eigenschaften offen zu Tage treten. Gleichzeitig illustrieren diese Vergleiche die Konstruktionsfehler der ET und GB Ansätze.

[1] Mit freundlicher Genehmigung der Verlags

THE JOURNAL OF CHEMICAL PHYSICS **140**, 104102 (2014)

Electrostatics of proteins in dielectric solvent continua. I. An accurate and efficient reaction field description

Sebastian Bauer, Gerald Mathias, and Paul Tavan[a)]
Lehrstuhl für BioMolekulare Optik, Ludwig–Maximilians Universität München, Oettingenstr. 67, 80538 München, Germany

(Received 10 October 2013; accepted 17 February 2014; published online 10 March 2014)

We present a reaction field (RF) method which accurately solves the Poisson equation for proteins embedded in dielectric solvent continua at a computational effort comparable to that of an electrostatics calculation with polarizable molecular mechanics (MM) force fields. The method combines an approach originally suggested by Egwolf and Tavan [J. Chem. Phys. **118**, 2039 (2003)] with concepts generalizing the Born solution [Z. Phys. **1**, 45 (1920)] for a solvated ion. First, we derive an exact representation according to which the sources of the RF potential and energy are inducible atomic anti-polarization densities and atomic shielding charge distributions. Modeling these atomic densities by Gaussians leads to an approximate representation. Here, the strengths of the Gaussian shielding charge distributions are directly given in terms of the static partial charges as defined, e.g., by standard MM force fields for the various atom types, whereas the strengths of the Gaussian anti-polarization densities are calculated by a self-consistency iteration. The atomic volumes are also described by Gaussians. To account for covalently overlapping atoms, their effective volumes are calculated by another self-consistency procedure, which guarantees that the dielectric function $\varepsilon(\mathbf{r})$ is close to one everywhere inside the protein. The Gaussian widths σ_i of the atoms i are parameters of the RF approximation. The remarkable accuracy of the method is demonstrated by comparison with Kirkwood's analytical solution for a spherical protein [J. Chem. Phys. **2**, 351 (1934)] and with computationally expensive grid-based numerical solutions for simple model systems in dielectric continua including a di-peptide (Ac-Ala-NHMe) as modeled by a standard MM force field. The latter example shows how weakly the RF conformational free energy landscape depends on the parameters σ_i. A summarizing discussion highlights the achievements of the new theory and of its approximate solution particularly by comparison with so-called generalized Born methods. A follow-up paper describes how the method enables Hamiltonian, efficient, and accurate MM molecular dynamics simulations of proteins in dielectric solvent continua. © *2014 AIP Publishing LLC*. [http://dx.doi.org/10.1063/1.4867280]

I. INTRODUCTION

The structure and the functional dynamics of soluble proteins are dominated by electrostatic interactions with the surrounding aqueous solvent.[1] Therefore, attempts of theoretical descriptions must properly account for these interactions.[2–5] Atomistic molecular dynamics (MD) simulations, in which protein-solvent systems are described by standard molecular mechanics (MM) force fields such as CHARMM22,[6] AMBER,[7] or GROMOS[8] typically employ periodic boundary conditions, take advantage of this periodicity by computing the long-range parts of the electrostatics through Ewald summations (see, e.g., Ref. 9), and model the aqueous solvent by simple non-polarizable three-point potentials like TIP3P[10] or SPC/E.[11] To guarantee sufficiently low protein concentrations in the simulation system and to limit periodicity artifacts, the number of solvent atoms should exceed the number of protein atoms by a factor larger than 10.[5,12,13] Although the dynamics of the solute material, e.g., its folding, unfolding, or functional rearrangement, commonly is in the focus of the scientific interest, such MD trajectories thus mainly cover the dynamics of liquid water, which is slightly polluted by protein material.

As of today, the enormous computational effort required by atomistic simulations of protein-solvent systems still poses a challenge. A large part of this effort could be saved, if the surrounding solvent could be replaced by a computationally inexpensive continuum model, which nevertheless accurately describes the key dielectric protein-solvent interactions. A corresponding scenario for MD simulations is schematically represented in Figure 1(a).

There have been many attempts[14–39] to construct a continuum approach, which meets the above criteria. However, in our view, which will be further substantiated in this work, all these attempts essentially represent failures (see also corresponding discussions in Refs. 5, 38, 40–43). In fact, it is the complexity of the underlying mathematical task, which so far has prevented accurate and computationally efficient solutions. Already in the most simple case of a purely polar solvent, this task demands the solution of a partial differential equation (PDE) on the fly with the integration of the protein dynamics.

[a)]Electronic mail: paul.tavan@physik.uni-muenchen.de

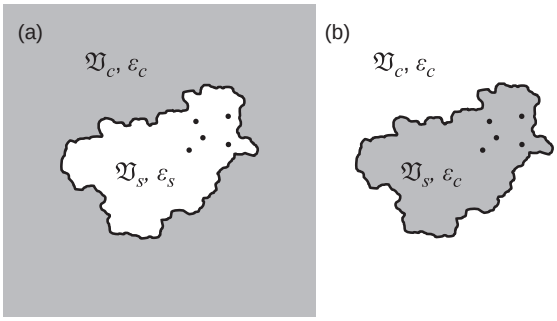

FIG. 1. (a) Concept of continuum solvent models: A solute molecule (white) is represented as a distribution of point charges (black dots) in a cavity $\mathfrak{V}_s$ with the dielectric constant ε_s. It is surrounded by a continuum $\mathfrak{V}_c$ of dielectric constant $\varepsilon_c > \varepsilon_s$. The charges generate a polarization (gray) in $\mathfrak{V}_c$. b) The approach by Egwolf and Tavan[14] removes the polarization in $\mathfrak{V}_c$ (now white), replaces it by an anti-polarization (gray) strictly confined to $\mathfrak{V}_s$, and assumes that ε_c additionally applies to $\mathfrak{V}_s$, i.e., that the solvent continuum covers the whole space.

The PDE to be solved is, using Gaussian units, the dielectric Poisson equation (PE)

$$\nabla \cdot [\varepsilon(\mathbf{r})\nabla\Phi(\mathbf{r})] = -4\pi\rho(\mathbf{r}) \tag{1}$$

for the electrostatic potential $\Phi(\mathbf{r})$, which is supposed to be generated by a distribution

$$\rho(\mathbf{r}) = \sum_i q_i \delta(\mathbf{r} - \mathbf{r}_i) \tag{2}$$

of N point charges q_i located at the positions $\mathbf{r}_i$ of the atoms i, if standard MM force fields are applied to the solute protein $\mathcal{P}$. The dielectric function

$$\varepsilon(\mathbf{r}) = \varepsilon_c - (\varepsilon_c - \varepsilon_s)\Theta(\mathbf{r}) \tag{3}$$

appearing in the PE is given in terms of the characteristic function

$$\Theta(\mathbf{r}) = \begin{cases} 1, & \text{if } \mathbf{r} \in \mathfrak{V}_s \\ 0, & \text{if } \mathbf{r} \in \mathfrak{V}_c. \end{cases} \tag{4}$$

separating the volume $\mathfrak{V}_s$ occupied by $\mathcal{P}$ from the solvent surroundings $\mathfrak{V}_c$ [cf. Fig. 1(a)].

There are, of course, numerical methods employing grid-discretizations[29–33] of $\mathfrak{V}_s$ and of a good portion of the surrounding $\mathfrak{V}_c$ or boundary element methods,[34–37] which can solve Eq. (1) for the potential $\Phi(\mathbf{r})$ generated by $\mathcal{P}$ at a fixed configuration $\mathbf{R} = (\mathbf{r}_1, \ldots, \mathbf{r}_N)^T \in \mathbb{R}^{3N}$ of the N atoms and allow accurate calculations of the electrostatic energy, which in linear media[44] is given by

$$W(\mathbf{R}) = \frac{1}{2}\int \rho(\mathbf{r}\,|\,\mathbf{R})\Phi(\mathbf{r}\,|\,\mathbf{R})\,\mathrm{d}V. \tag{5}$$

However, these approaches are computationally very expensive and suffer from discretization effects, which exclude sufficiently accurate computations of the electrostatic fields and forces at the positions $\mathbf{r}_i$ of the atoms. Moreover, these methods account for the reaction forces, which are generated by the dielectric boundary pressure,[27] only through heuristic approximations,[45–47] which occasionally entail severe artifacts.[48] For all these reasons, these methods do not preserve the Hamiltonian structure of the dynamics and cannot be used for microcanonical MM-MD simulations. Instead they have to apply some sort of stochastic dynamics, which can partially compensate[39] the various dynamical artifacts arising from the inevitable violations of Newton's reaction principle.

Instead of solving the PE at each integration step, which is the task that has to be tackled, generalized Born (GB) models[15–26] introduce local screening functions for the electrostatic interactions, which supposedly generalize the Born solution for the solvation energy of an isolated ion.[49] Because GB methods offer simple analytical expressions for the atomic interaction energies and forces, they enable MD simulations over impressive timescales.[50]

Grycuk[20] has demonstrated that the GB approach does not solve the PE (1) even after its approximate correction by and beyond the so-called Coulomb field approximation (CFA). Here, he used the well-known Kirkwood solution[51] for the solvated "spherical protein," which we denote by $\mathcal{K}$, as his reference, because $\mathcal{K}$ is one of the few non-trivial cases, in which the PE can be analytically solved. For instance, if the low-dielectric interior of $\mathcal{K}$ harbors a single charge at a distance r from the center, then GB/CFA underestimates the change $\Delta G(r) \equiv |G(r) - G(0)|$ of the electrostatic contribution to the free energy by a factor up to two.[20,25] Note that Kharkats *et al.*[40] had shown already in 1976 for a dipolar pair of point charges that there is no enclosing cavity, for which the CFA is correct.

Thus, it is not particularly surprising that free energy landscapes resulting from extended GB-MD simulations, which moreover strongly depend on the applied parameter sets, substantially differ from landscapes derived by explicit solvent simulations.[41,42] Based on our exact results to be presented further below, we will analyze the GB approach and its systematic shortcomings more deeply at several locations of this paper.

A. The Egwolf-Tavan approach

In 2003, Egwolf and Tavan[14] (ET) proposed a new approach to the continuum electrostatics of proteins in solution. Partially adopting earlier arguments by Sklenar *et al.*,[52] they first noted that the PE (1) can be exactly reformulated as

$$\Delta\Phi(\mathbf{r}) = -\frac{4\pi}{\varepsilon_c}\left[\rho(\mathbf{r}) - \nabla\cdot\mathbf{P}(\mathbf{r})\right]. \tag{6}$$

Here, ET introduced a new type of dipole density by

$$\mathbf{P}(\mathbf{r}) \equiv \frac{\varepsilon_c - \varepsilon_s}{4\pi}\Theta(\mathbf{r})\nabla\Phi(\mathbf{r}) \tag{7}$$

as the gradient of the unknown electrostatic potential $\Phi(\mathbf{r})$. This novel dipole density $\mathbf{P}(\mathbf{r})$ is non-zero solely in the region $\mathfrak{V}_s$ occupied by $\mathcal{P}$ and vanishes outside because of the multiplicative factor $\Theta(\mathbf{r})$ [cf. Eq. (4)].

The partial differential equations (1) and (6) have identical solutions $\Phi(\mathbf{r})$, if the ET dipole density $\mathbf{P}(\mathbf{r})$ self-consistently fulfills Eq. (7). This density $\mathbf{P}(\mathbf{r})$ is actually an *anti-polarization*,[38] as one sees with the definitions

$$\chi_e \equiv (\varepsilon_c - \varepsilon_s)/4\pi > 0 \tag{8}$$

of the electric susceptibility and $\mathbf{E}(\mathbf{r}) = -\nabla\Phi(\mathbf{r})$ of the electric field. Then Eq. (7) becomes

$$\mathbf{P}(\mathbf{r}) = -\chi_e \mathbf{E}(\mathbf{r})\Theta(\mathbf{r}), \tag{9}$$

which differs from the usual expression for polarization densities by the essential minus sign and by the factor $\Theta(\mathbf{r})$ restricting the anti-polarization density $\mathbf{P}(\mathbf{r})$ to the spatial region $\mathfrak{V}_s$ occupied by the solute protein. In contrast, conventional polarization densities are proportional to the electric field and are confined to the surroundings $\mathfrak{V}_c$ of $\mathfrak{V}_s$.

At this point, it is important to note that the boundary conditions, which, in the context of the PE (1), are imposed at the surface of $\mathfrak{V}_s$, are replaced in the ET representation (6) of continuum electrostatics by the self-consistency condition (7) [or, equivalently, (9)] for $\mathbf{P}(\mathbf{r})$. Hence, the solution of the PE (1) amounts within the ET theory to the self-consistent computation of the anti-polarization density $\mathbf{P}(\mathbf{r})$.[14]

Figure 1(b) illustrates the ET representation (6) of the continuum electrostatics for a solvated protein. Here, the source of $\Phi(\mathbf{r})$ is not only the charge density $\rho(\mathbf{r})$ but also the anti-polarization density $\mathbf{P}(\mathbf{r})$ within $\mathfrak{V}_s$. $\mathbf{P}(\mathbf{r})$ effectively replaces the polarization density, which in the conventional view covers the surroundings $\mathfrak{V}_c$ [cf. Fig. 1(a)]. In the ET representation, the two sources are everywhere strongly shielded by ε_c implying that the solvent continuum fills the whole space. Within $\mathfrak{V}_s$ the shielding of $\rho(\mathbf{r})$ is much too strong and, therefore, has to be compensated by $\mathbf{P}(\mathbf{r})$.

For general protein geometries, Eqs. (6) and (7) cannot be solved analytically. To derive numerically tractable approximate solutions, one can exploit the fact that the protein atoms, which are located at the positions $\mathbf{r}_i \in \mathfrak{V}_s$, define a natural discretization of the space $\mathfrak{V}_s$. This discretization can be mathematically formulated by assigning to every atom i a normalized Gaussian distribution

$$G(\mathbf{r}\,|\,\mathbf{r}_i, \sigma_i) = \frac{1}{\left(2\pi\sigma_i^2\right)^{3/2}} \exp\left[-\frac{(\mathbf{r}-\mathbf{r}_i)^2}{2\sigma_i^2}\right], \tag{10}$$

which is centered at $\mathbf{r}_i$ and has the width σ_i. Then these functions define by

$$\vartheta_i(\mathbf{r}) = \Theta(\mathbf{r}) \frac{G(\mathbf{r}\,|\,\mathbf{r}_i, \sigma_i)}{\sum_j G(\mathbf{r}\,|\,\mathbf{r}_j, \sigma_j)} \tag{11}$$

an exact fuzzy partition[53] of $\mathfrak{V}_s$ into fuzzy atomic sets $\mathfrak{v}_i$ characterized by the $\vartheta_i(\mathbf{r})$, which have the property

$$\sum_i \vartheta_i(\mathbf{r}) = \Theta(\mathbf{r}) \tag{12}$$

for all $\mathbf{r}$ and reduce to a Voronoi tessellation of $\mathfrak{V}_s$ in the limit $\sigma_i \to 0$. Defining the volumes v_i of the fuzzy atomic regions $\mathfrak{v}_i$ by the integrals

$$v_i \equiv \int \vartheta_i(\mathbf{r})\,\mathrm{d}V \tag{13}$$

and the protein volume V_s by a corresponding integral over $\Theta(\mathbf{r})$, one has

$$\sum_i v_i = V_s. \tag{14}$$

Inserting now Eq. (12) into Eq. (7) yields with the definition

$$\mathbf{P}_i(\mathbf{r}) \equiv \chi_e \vartheta_i(\mathbf{r}) \nabla\Phi(\mathbf{r}), \tag{15}$$

of atomic anti-polarization densities the decomposition $\mathbf{P}(\mathbf{r}) = \sum_i \mathbf{P}_i(\mathbf{r})$ of the total anti-polarization density $\mathbf{P}(\mathbf{r})$ into atomic contributions. Together with the decomposition (2) of the charge distribution into atomic partial charges, the PE (6) finally reads

$$\Delta\Phi(\mathbf{r}) = -\frac{4\pi}{\varepsilon_c} \sum_i [q_i \delta(\mathbf{r}-\mathbf{r}_i) - \nabla\cdot\mathbf{P}_i(\mathbf{r})]. \tag{16}$$

According to the form (16) of the PE (1), the electrostatic potential is generated by certain atomic anti-polarization densities $\mathbf{P}_i(\mathbf{r})$, which have to self-consistently fulfill Eq. (15), and by the atomic partial charges q_i. All these atomic sources of the exact solution $\Phi(\mathbf{r})$ of the PE (16) are everywhere strongly shielded by ε_c [cf. Fig. 1(b)]. If one wants to solve Eq. (16), one needs to determine the atomic characteristic functions $\vartheta_i(\mathbf{r})$, which generate an exact fuzzy partition of $\mathfrak{V}_s$, because the $\vartheta_i(\mathbf{r})$ define by Eq. (15) the atomic anti-polarization densities $\mathbf{P}_i(\mathbf{r})$. Generally there will be neither an analytical nor a simple numerical solution to this problem and, therefore, one has to resort to analytical approximations.

With the aim of deriving accurate and computationally inexpensive approximations and partially adopting an earlier suggestion by Schaefer and Karplus,[16] ET replaced[14] the exact atomic functions $\vartheta_i(\mathbf{r})$ and $\mathbf{P}_i(\mathbf{r})$ by the Gaussian models

$$\tilde{\vartheta}_i(\mathbf{r}\,|\,\mathbf{r}_i, \tilde{v}_i, \sigma_i) = \tilde{v}_i G(\mathbf{r}\,|\,\mathbf{r}_i, \sigma_i) \tag{17}$$

and

$$\tilde{\mathbf{P}}_i(\mathbf{r}\,|\,\mathbf{r}_i, \tilde{\mathbf{p}}_i, \sigma_i) = \tilde{\mathbf{p}}_i G(\mathbf{r}\,|\,\mathbf{r}_i, \sigma_i), \tag{18}$$

which are constructed from the atomic Gaussian distributions (10). Here, the Gaussian widths σ_i may be derived, e.g., from the van der Waals (vdW) radii available in standard MM force fields. Because the Gaussians (10) are normalized, the parameters $\tilde{v}_i$ and $\tilde{\mathbf{p}}_i$ of the models (17) and (18) are the respective spatial integrals and, therefore, are the atomic volumes and the reaction field (RF) dipoles. The approximate characteristic function

$$\tilde{\Theta}(\mathbf{r}) = \sum_i \tilde{\vartheta}_i(\mathbf{r}\,|\,\mathbf{r}_i, \tilde{v}_i, \sigma_i) \tag{19}$$

and approximate anti-polarization density

$$\tilde{\mathbf{P}}(\mathbf{r}) = \sum_i \tilde{\mathbf{P}}_i(\mathbf{r}\,|\,\mathbf{r}_i, \tilde{\mathbf{p}}_i, \sigma_i) \tag{20}$$

are then simply sums of the atomic Gaussian models, whose parameters $\tilde{v}_i$ and $\tilde{\mathbf{p}}_i$, respectively, can be calculated by self-consistency iterations.[14]

The self-consistency conditions for the volumes $\tilde{v}_i$ of the Gaussian atoms defined by Eq. (17) derive from the requirement that $\tilde{\Theta}(\mathbf{r}_i) = \Theta(\mathbf{r}_i) = 1$ for all $\mathbf{r}_i$ and can be met by the fixed-point iteration

$$\tilde{v}_i^{(n)} = \tilde{v}_i^{(n-1)} \left[\sum_j \tilde{v}_j^{(n-1)} G(\mathbf{r}_i\,|\,\mathbf{r}_j, \sigma_j)\right]^{-1}. \tag{21}$$

The conditions for the atomic RF dipoles $\tilde{\mathbf{p}}_i$ result from extended and complicated analytical calculations and are given by Eq. (52) in Ref. 14.

The approximate ET approach allowed rapid and accurate computations of solvation energies for polar and charged molecules. Moreover, it was shown to provide an accurate solution of the PE (16) in the case of the Kirkwood sphere $\mathcal{K}$.[51] However, electrostatic RF forces on atomic charges $q_i \in \mathfrak{V}_s$ were not derived from the negative gradient of the approximate electrostatic energy $\tilde{W}$. Instead they were calculated from an approximate expression for the electric field $\tilde{\mathbf{E}}(\mathbf{r})$ as so-called $q\mathbf{E}$-forces thereby neglecting the reaction forces[38] associated with the dielectric boundary pressure.[27] The resulting violation of Newton's reaction principle excluded applications to RF/MM-MD simulations.

B. The Stork-Tavan revision

A careful revision and reformulation of the ET approach by Stork and Tavan[38,39] (ST) uncovered its close similarity to polarizable force fields, which employ inducible atomic dipoles.[54–56] As shown by ST, the atomic RF dipoles

$$\tilde{\mathbf{p}}_i = -\alpha_i \mathbf{E}_{\mathrm{pol}}^{q,\tilde{\mathbf{p}}}(\mathbf{r}_i), \tag{22}$$

which represent the strengths of the Gaussian anti-polarization densities (18), follow by a negative linear response from a polarizing field $\mathbf{E}_{\mathrm{pol}}^{q,\tilde{\mathbf{p}}}(\mathbf{r})$, which is generated by the partial charges q_j and RF dipoles $\tilde{\mathbf{p}}_j$ of all other atoms $j \neq i$, through atomic polarizabilities $\alpha_i > 0$ [cf. Eqs. (41)–(51) in Ref. 38]. Furthermore, ST considered the RF dipoles not only as sources of the electrostatic field but also as targets of electrostatic forces, which therefore can comply with Newton's reaction principle.

However, the electrostatic forces (which include the RF contributions) were not obtained as configurational derivatives

$$\mathbf{f}_i = -\nabla_{\mathbf{r}_i} \tilde{W}(\mathbf{R}) \tag{23}$$

of the ET approximation $\tilde{W}(\mathbf{R})$ to the electrostatic energy (5) but were constructed by considering approximate expressions of the electric field in the large distance limit.[38] Therefore, the forces did neither preserve the energy nor the angular momentum. This non-Hamiltonian character strongly hampered RF/MM-MD simulations.

C. The solution

It is the aim of this and of a follow-up[57] paper to present a solution for the problems left unsolved by the ET approach and its ST revision. In the current paper, we first transform the ET representation of continuum electrostatics into a new but equally exact form, which we call the "RF representation". This transformation contains a few seemingly strange but decisive steps, whose origin and motivation will be subsequently explained by scrutinizing, within the framework of the ET approach to the continuum electrostatics of solute-solvent systems, the analytical solution[14,44,49] for the field-exposed Born ion. Together with the Gaussian approximations (17) and (18) employed earlier[14,38] this simple example suggests improved approximations for the atomic dipole densities $\mathbf{P}_i(\mathbf{r})$. The quality of the resulting expression for the approximate electrostatic potential $\tilde{\Phi}(\mathbf{r})$ is discussed by comparisons with exact (where available) and numerical solutions of the PE for the Kirkwood sphere $\mathcal{K}$, for pair interactions of solvated ions and atoms, and for the solvated N-Acetylalanine methylamide (Ac-Ala-NHMe) in various conformations.

The follow-up paper derives by Eq. (5) from the approximate electrostatic potential $\tilde{\Phi}(\mathbf{r})$ expressions for the associated electrostatic energy $\tilde{W}(\mathbf{R})$, from which the electrostatic forces are calculated by Eq. (23) as derivatives. It is demonstrated that these forces enable Hamiltonian RF/MM-MD simulations.[57]

II. EXACT THEORY

We will now show how the continuum electrostatics of protein-solvent systems can be exactly transformed from the ET representation, which is specified by Eqs. (11), (15), and (16), into the RF representation. For the sake of a most compact derivation, the underlying motivations and guidelines are initially left aside. They will be discussed afterwards using the exact solution[14,44,49] of the field-exposed Born ion as an instructive example.

A. The RF representation of protein continuum electrostatics

First, we decompose the exact atomic dipole densities (15) by

$$\mathbf{P}_i(\mathbf{r}) = \mathbf{P}_i^{\mathrm{rad}}(\mathbf{r}) + (\varepsilon_c/\varepsilon_s)\hat{\mathbf{P}}_i(\mathbf{r}) \tag{24}$$

into radial parts

$$\mathbf{P}_i^{\mathrm{rad}}(\mathbf{r}) = (\mathbf{r} - \mathbf{r}_i) f_i(|\mathbf{r} - \mathbf{r}_i|) \tag{25}$$

with the scalar functions

$$f_i(r) \equiv \frac{1}{4\pi r} \int \frac{\mathbf{r}}{r} \cdot \mathbf{P}_i(\mathbf{r} + \mathbf{r}_i)\, \mathrm{d}\Omega \tag{26}$$

and remaining non-radial parts $\hat{\mathbf{P}}_i(\mathbf{r})$, where the integration is performed over the complete solid angle.

Inspecting the ET representation (16) of the PE suggests that one can consider the sources $-\nabla \cdot \mathbf{P}_i^{\mathrm{rad}}(\mathbf{r})$ of the potential $\Phi(\mathbf{r})$ as dipolar charge distributions

$$\rho_i^{\mathrm{rad}}(\mathbf{r}) \equiv -\nabla \cdot \mathbf{P}_i^{\mathrm{rad}}(\mathbf{r}). \tag{27}$$

As shown in Appendix A, these dipolar atomic charge densities $\rho_i^{\mathrm{rad}}(\mathbf{r})$ depend only on the distance $|\mathbf{r} - \mathbf{r}_i|$, i.e., are radially symmetric functions. The total spatial integrals

$$\int \rho_i^{\mathrm{rad}}(\mathbf{r})\, \mathrm{d}V = 0 \tag{28}$$

over these atomic charge densities vanish, as one concludes by inserting Eq. (27) into Eq. (28), by applying Gauss's theorem,[58] and by noting that $\mathbf{P}_i^{\mathrm{rad}}(\mathbf{r})$ vanishes outside $\mathfrak{V}_s$ because of Eqs. (15), (11), and (4).

Next, the dipolar charge densities $\rho_i^{\mathrm{rad}}(\mathbf{r})$ are decomposed into two radial and oppositely charged distributions. We

choose the combinations

$$\rho_i^{\mathrm{rad}}(\mathbf{r}) \equiv (\varepsilon_c/\varepsilon_s)[-\hat{q}_i\delta(\mathbf{r}-\mathbf{r}_i)+\hat{\rho}_i(\mathbf{r})] \tag{29}$$

of point charges $-\hat{q}_i$ at $\mathbf{r}_i$, whose strengths are defined by

$$\hat{q}_i \equiv -q_i(1-\varepsilon_s/\varepsilon_c), \tag{30}$$

with radial charge distributions $\hat{\rho}_i(\mathbf{r})$ of the opposite strengths

$$\int \hat{\rho}_i(\mathbf{r})\,\mathrm{d}V = \hat{q}_i, \tag{31}$$

which are obtained by integrating $\hat{\rho}_i(\mathbf{r})$ over all space and, thus, guarantee that Eq. (28) holds. Because of Eq. (15), each radial charge distribution $\hat{\rho}_i(\mathbf{r})$ is localized within the corresponding atomic fuzzy region $\mathfrak{v}_i$ defined by the characteristic function $\vartheta_i(\mathbf{r})$. Outside $\mathfrak{v}_i$, each radial charge distribution $\hat{\rho}_i(\mathbf{r})$ completely shields the associated point charge $-\hat{q}_i$.

If we now insert the decomposition (24) of the exact atomic dipole densities $\mathbf{P}_i(\mathbf{r})$ into the PE (16) and apply the equivalent replacement (27) of the radial parts $-\nabla\cdot\mathbf{P}_i^{\mathrm{rad}}(\mathbf{r})$ by the dipolar charge densities $\rho_i^{\mathrm{rad}}(\mathbf{r})$, which become decomposed according to Eq. (29), we finally obtain, after a few lines of algebra, the RF representation

$$\Delta\Phi(\mathbf{r}) = -\frac{4\pi}{\varepsilon_s}\sum_i[q_i\delta(\mathbf{r}-\mathbf{r}_i)+\hat{\rho}_i(\mathbf{r})-\nabla\cdot\hat{\mathbf{P}}_i(\mathbf{r})] \tag{32}$$

for the PE of a solute protein $\mathcal{P}$, which is strictly equivalent to Eqs. (1) and (6).

According to the form (32) of the PE, the electrostatic potential $\Phi(\mathbf{r})$ is generated by the partial charges q_i, by the oppositely charged radial atomic shielding charge distributions $\hat{\rho}_i(\mathbf{r})$ of total strengths $\hat{q}_i$, and by the non-radial and diminished contributions $\hat{\mathbf{P}}_i(\mathbf{r})$ to the atomic densities (15). As is characteristic for a RF representation, all sources of the potential $\Phi(\mathbf{r})$ are weakly shielded by ε_s. Correspondingly the Coulomb contribution

$$\Phi^{\mathrm{C}}(\mathbf{r}) = \frac{1}{\varepsilon_s}\sum_i \phi(\mathbf{r}\,|\,\mathbf{r}_i, q_i) \tag{33}$$

to $\Phi(\mathbf{r})$ is a weakly shielded superposition of the Coulomb potentials $\phi(\mathbf{r}\,|\,\mathbf{r}_i, q_i)$ explicitly specified through Eq. (B12) in Appendix B,[59] which are generated by the atomic point charges q_i. Furthermore the electrostatic action of the continuum ε_c has been completely absorbed into the RF potential

$$\Phi^{\mathrm{RF}}(\mathbf{r}) = \frac{1}{\varepsilon_s}\sum_i[\phi(\mathbf{r}\,|\,\hat{\rho}_i)+\phi(\mathbf{r}\,|\,\hat{\mathbf{P}}_i)], \tag{34}$$

which is a sum of the unknown potentials generated by the equally unknown densities $\hat{\rho}_i(\mathbf{r})$ and $\hat{\mathbf{P}}_i(\mathbf{r})$.

The main advantage of the RF representation (32) of the PE and of its solution

$$\Phi(\mathbf{r}) = \Phi^{\mathrm{C}}(\mathbf{r})+\Phi^{\mathrm{RF}}(\mathbf{r}) \tag{35}$$

is the clear separation of the Coulomb and RF contributions, which is automatically given in, e.g., boundary integral methods, but was absent in the original ET formulation. Correspondingly, the separate accessibility of the RF potential enables us to compute the solvation contribution $W^{\mathrm{RF}}(\mathbf{R})$ to the total electrostatic energy (5). If one evaluates $\Phi^{\mathrm{RF}}(\mathbf{r})$ as given by Eq. (34) at the atomic positions $\mathbf{r}_i$, the RF energy of $\mathcal{P}$ is

$$W^{\mathrm{RF}}(\mathbf{R}) = \frac{1}{2}\sum_i q_i\Phi^{\mathrm{RF}}(\mathbf{r}_i\,|\,\mathbf{R}), \tag{36}$$

whose configurational derivatives $\nabla_{\mathbf{r}_i}W^{\mathrm{RF}}(\mathbf{R})$ in principle provide access to the mean forces exerted by the solvent continuum on the protein atoms.[57]

Because of Eqs. (15) and (9), the dipole densities $\mathbf{P}_i(\mathbf{r})$ have to be self-consistently calculated from the electric field $\mathbf{E}(\mathbf{r})$ within the atomic regions $\mathfrak{v}_i$. With the decomposition (24), this condition reads

$$\mathbf{P}_i^{\mathrm{rad}}(\mathbf{r}) + \frac{\varepsilon_c}{\varepsilon_s}\hat{\mathbf{P}}_i(\mathbf{r}) = -\chi_e\vartheta_i(\mathbf{r})\mathbf{E}(\mathbf{r}), \tag{37}$$

where the atomic characteristic functions $\vartheta_i(\mathbf{r})$ belong to the exact decomposition (11) of the region $\mathfrak{V}_s$ and should not be confused with the Gaussian models $\tilde{\vartheta}_i(\mathbf{r})$ of the approximate ET approach. For arbitrarily shaped proteins, these conditions cannot be analytically solved for the non-radial and radial contributions to the atomic anti-polarization densities $\hat{\mathbf{P}}_i(\mathbf{r})$ and $\mathbf{P}_i^{\mathrm{rad}}(\mathbf{r})$, respectively, or for the radial shielding charge distributions $\hat{\rho}_i(\mathbf{r})$ equivalently replacing the latter. One can, however, easily derive exact conditions for the atomic spatial integrals of these densities, which should be fulfilled, of course, by any approximation to $\hat{\mathbf{P}}_i(\mathbf{r})$ and $\hat{\rho}_i(\mathbf{r})$.

With Eqs. (30) and (31), the desired integral conditions on the radial shielding charge distribution $\hat{\rho}_i(\mathbf{r})$ have already been specified. Thus, the integrals over approximate distributions $\tilde{\rho}_i(\mathbf{r})$ should also yield the exact values $\hat{q}_i$ given by Eq. (30). For the non-radial contributions $\hat{\mathbf{P}}_i(\mathbf{r})$ to the atomic anti-polarization densities suitable conditions follow by considering the spatial integrals

$$\int \mathbf{P}_i^{\mathrm{rad}}(\mathbf{r})\,\mathrm{d}V + \frac{\varepsilon_c}{\varepsilon_s}\hat{\mathbf{p}}_i = -v_i\chi_e\langle\mathbf{E}\rangle_{\mathfrak{v}_i} \tag{38}$$

of Eq. (37). Here, χ_e is given by Eq. (8) and we have used the definitions

$$\hat{\mathbf{p}}_i \equiv \int \hat{\mathbf{P}}_i(\mathbf{r})\,\mathrm{d}V$$

for the strengths $\hat{\mathbf{p}}_i$ of the non-radial atomic anti-polarization densities $\hat{\mathbf{P}}_i(\mathbf{r})$ and

$$\langle\mathbf{E}\rangle_{\mathfrak{v}_i} \equiv (1/v_i)\int \mathbf{E}(\mathbf{r})\vartheta_i(\mathbf{r})\,\mathrm{d}V \tag{39}$$

for the average electric fields within the atomic regions $\mathfrak{v}_i$, whose volumes v_i are given by Eq. (13). Because the spatial integrals over purely radial vector fields $\mathbf{P}_i^{\mathrm{rad}}(\mathbf{r})$ as defined by Eqs. (25) and (26) vanish by symmetry, i.e.,

$$\int \mathbf{P}_i^{\mathrm{rad}}(\mathbf{r})\,\mathrm{d}V = 0, \tag{40}$$

the condition (38) becomes

$$\hat{\mathbf{p}}_i = -v_i\chi_e(\varepsilon_s/\varepsilon_c)\langle\mathbf{E}\rangle_{\mathfrak{v}_i}. \tag{41}$$

Hence, the strengths $\hat{\mathbf{p}}_i$ of the non-radial parts of the atomic anti-polarization densities $\hat{\mathbf{P}}_i(\mathbf{r})$ should be anti-parallel to the

electric field averaged over the atomic regions $\mathfrak{v}_i$ and diminished by a factor $\varepsilon_s/\varepsilon_c$, which one can try to achieve by some sort of self-consistency iteration.

In summary, the RF theory of the protein continuum electrostatics outlined above yields the two conditions (30) and (41) for strengths $\hat{q}_i$ and $\hat{\mathbf{p}}_i$ of the atomic sources $\hat{\rho}_i(\mathbf{r})$ and $\hat{\mathbf{P}}_i(\mathbf{r})$, respectively, of the RF potential $\Phi^{\mathrm{RF}}(\mathbf{r})$ given by (34). Furthermore, it provides a direct access to the RF energy $W^{\mathrm{RF}}(\mathbf{R})$, which is a potential of mean force for the influence of the solvent continuum on the solute protein $\mathcal{P}$.

B. The most simple and analytically solvable example

The following discussion of the most simple and analytically tractable example[14] for a solute "molecule" serves to clarify the ET approach and its RF extension. It will explain the motivation for the decompositions (24) of the atomic anti-polarizations $\mathbf{P}_i(\mathbf{r})$ into non-radial and radial parts and (29) of the dipolar charge densities $\rho_i^{\mathrm{rad}}(\mathbf{r})$ into point charges $-\hat{q}_i$ and surrounding shielding charge distributions $\hat{\rho}_i(\mathbf{r})$. This example considers a single solvated ion, which is exposed to a homogeneous external electric field $\mathbf{E}^{\mathrm{ext}}$. The latter is supposed to model the field, which acts in a protein on a given atom and is generated by the other atoms.

For this analysis, we choose a stepwise fashion, because the solvated ion exposed to $\mathbf{E}^{\mathrm{ext}}$ is, as we show in Sec. S1 of the supplementary material,[60] the superposition of an empty solvated cavity $\mathcal{C}$ in a homogeneous field[44] with the solvated Born ion $\mathcal{B}$.[49] Correspondingly, we will denote the solvated ion exposed to $\mathbf{E}^{\mathrm{ext}}$ from now on as $\mathcal{B} \cup \mathcal{C}$.

The ion is assumed to be a spherical cavity with the radius R and volume $v = 4\pi R^3/3$, has the charge q at its center, has the interior dielectric constant ε_s, and is surrounded by a solvent continuum of dielectric constant ε_c. The ion occupies the spatial region $\mathfrak{v}$ defined by the characteristic function

$$\vartheta(\mathbf{r}) = \begin{cases} 1, & \text{if } |\mathbf{r}| \le R, \\ 0, & \text{else.} \end{cases} \tag{42}$$

Because the system contains in the given case only a single atom, the associated atomic region $\mathfrak{v}$ is identical to the region $\mathfrak{V}_s$, i.e., $\vartheta(\mathbf{r}) = \Theta(\mathbf{r})$.

1. The field-exposed cavity $\mathcal{C}$

According to Chap. 4.4 in the textbook of Jackson[44] and Eq. (S98) of the supplementary material,[60] the potential of $\mathcal{C}$ is given by

$$\Phi^{\mathcal{C}}(\mathbf{r}) = -\mathbf{r}\cdot\mathbf{E}^{\mathrm{ext}} + \begin{cases} \mathbf{r}\cdot\mathbf{p}/R^3\varepsilon_c, & \text{if } \mathbf{r} \in \mathfrak{v}, \\ (1/\varepsilon_c)\phi(\mathbf{r}\,|\,\mathbf{p}), & \text{else,} \end{cases} \tag{43}$$

where $\phi(\mathbf{r}\,|\,\mathbf{p})$ is the potential[59] Eq. (B13) of the RF point dipole

$$\mathbf{p} \equiv -\alpha\mathbf{E}^{\mathrm{ext}}. \tag{44}$$

This RF dipole is located at the center of $\mathcal{C}$, is oriented anti-parallel to $\mathbf{E}^{\mathrm{ext}}$, and is induced by $\mathbf{E}^{\mathrm{ext}}$ through the RF polarizability

$$\alpha = \frac{\varepsilon_c - \varepsilon_s}{2 + \varepsilon_s/\varepsilon_c} R^3 \tag{45}$$

of $\mathcal{C}$. Note that the potential (43) obeys the asymptotic boundary condition $\Phi^{\mathcal{C}}(\mathbf{r}) = -\mathbf{r}\cdot\mathbf{E}^{\mathrm{ext}}$ for $|\mathbf{r}| \to \infty$.

Inserting the potential (43) into the definition (15) of the atomic anti-polarization densities yields, after a little algebra using Eqs. (44), (45), and (8), the homogeneous dipole density

$$\mathbf{P}^{\mathcal{C}}(\mathbf{r}) = \mathbf{p}\,\vartheta(\mathbf{r})/v, \tag{46}$$

which is confined to $\mathfrak{v}$.

The potential $\phi(\mathbf{r}\,|\,\mathbf{P}^{\mathcal{C}})$, which is generated by the anti-polarization density (46), is calculated in Appendix C. The resulting expressions (C4) for $\mathbf{r} \notin \mathfrak{v}$ and (C5) for $\mathbf{r} \in \mathfrak{v}$ immediately show that, everywhere in space, the potential (43) can be more compactly written as

$$\Phi^{\mathcal{C}}(\mathbf{r}) = -\mathbf{r}\cdot\mathbf{E}^{\mathrm{ext}} + (1/\varepsilon_c)\phi(\mathbf{r}\,|\,\mathbf{P}^{\mathcal{C}}). \tag{47}$$

By the construction presented in Appendix C, $\Phi^{\mathcal{C}}(\mathbf{r})$ solves the ET representation

$$\Delta\Phi^{\mathcal{C}}(\mathbf{r}) = (4\pi/\varepsilon_c)\nabla\cdot\mathbf{P}^{\mathcal{C}}(\mathbf{r}) \tag{48}$$

of the PE [cf. Eqs. (6) and (16)] with the given asymptotic boundary condition.

Because the anti-polarization density (46) was derived above through Eq. (15) from the cavity potential (43), and because this potential $\Phi^{\mathcal{C}}(\mathbf{r})$ can be equivalently expressed in terms of $\mathbf{P}^{\mathcal{C}}(\mathbf{r})$ through Eq. (47), the anti-polarization density $\mathbf{P}^{\mathcal{C}}(\mathbf{r})$ obviously fulfills the self-consistency condition (15), that is,

$$\mathbf{P}^{\mathcal{C}}(\mathbf{r}) = \chi_e\vartheta(\mathbf{r})\nabla\Phi^{\mathcal{C}}(\mathbf{r}), \tag{49}$$

which guarantees the fulfillment of the boundary conditions at the surface of $\mathfrak{v}$.

The RF representation of the ET solution (47) for the PE (48) is readily obtained, because the homogeneous anti-polarization density $\mathbf{P}^{\mathcal{C}}(\mathbf{r})$ is purely non-radial [cf. Eq. (46)] such that all radial contributions vanish. Hence, the decomposition (24) may be rewritten as

$$\hat{\mathbf{P}}^{\mathcal{C}}(\mathbf{r}) = (\varepsilon_s/\varepsilon_c)\,\mathbf{P}^{\mathcal{C}}(\mathbf{r}), \tag{50}$$

and the cavity's PE (48) becomes

$$\Delta\Phi^{\mathcal{C}}(\mathbf{r}) = (4\pi/\varepsilon_s)\nabla\cdot\hat{\mathbf{P}}^{\mathcal{C}}(\mathbf{r}), \tag{51}$$

which is the RF representation (32) of the PE for $\mathcal{C}$. The RF solution $\Phi^{\mathcal{C}}(\mathbf{r})$, whose general form is given by (34), then immediately follows from the ET solution (47) as

$$\Phi^{\mathcal{C}}(\mathbf{r}) = -\mathbf{r}\cdot\mathbf{E}^{\mathrm{ext}} + (1/\varepsilon_s)\phi(\mathbf{r}\,|\,\hat{\mathbf{P}}^{\mathcal{C}}). \tag{52}$$

The corresponding RF potential (34) of $\mathcal{C}$ is given by

$$\Phi^{\mathrm{RF},\mathcal{C}}(\mathbf{r}) = (1/\varepsilon_s)\phi(\mathbf{r}\,|\,\hat{\mathbf{P}}^{\mathcal{C}}). \tag{53}$$

Explicit expressions for $\phi(\mathbf{r}\,|\,\hat{\mathbf{P}}^{\mathcal{C}})$ are provided by Eqs. (C4) and (C5) in Appendix C, if one replaces the original ET dipole $\mathbf{p}$ [cf. Eq. (44)] by the diminished dipole $\hat{\mathbf{p}} = (\varepsilon_s/\varepsilon_c)\mathbf{p}$.

Figure 2 compares the electrostatic scenarios employed in the ET (a) and RF (b) representations of the field-exposed

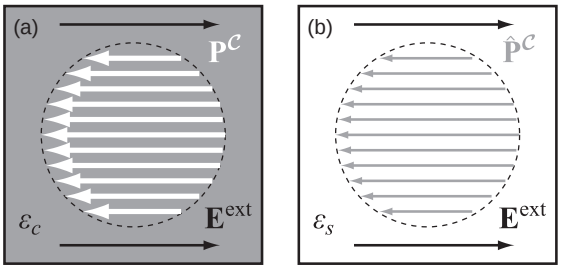

FIG. 2. Two representations of $\mathcal{C}$: (a) ET scenario as formulated by Eqs. (47) and (48); (b) RF approach as expressed by Eqs. (51) and (52); for explanation, see the text.

cavity $\mathcal{C}$. The dark gray background in (a) serves to indicate that the solvent continuum fills the whole space (including the interior of the cavity). Correspondingly, the potential $\phi(\mathbf{r}\,|\,\mathbf{P}^{\mathcal{C}})$, which is generated by the anti-polarization density $\mathbf{P}^{\mathcal{C}}(\mathbf{r})$ [fat white arrows in Fig. 2(a)] in $\mathfrak{v}$, is strongly shielded by ε_c. The white background in (b), in contrast, indicates for the RF scenario a weak or vanishing shielding of the potential $\phi(\mathbf{r}\,|\,\hat{\mathbf{P}}^{\mathcal{C}})$, which is generated by the strongly diminished anti-polarization density $\hat{\mathbf{P}}^{\mathcal{C}}(\mathbf{r})$ [cf. Eq. (50)] (thin gray arrows). According to Eq. (34), this weakly shielded potential $(1/\varepsilon_s)\phi(\mathbf{r}\,|\,\hat{\mathbf{P}}^{\mathcal{C}})$ is the RF potential $\Phi^{\text{RF},\mathcal{C}}(\mathbf{r})$ generated by the field-exposed cavity $\mathcal{C}$.

2. The Born ion

Next we treat the classical case[49] of the Born ion $\mathcal{B}$ in the framework of the ET approach and of its RF extension. As is also documented by Eq. (S97) in the supplementary material,[60] its potential is given by

$$\Phi^{\mathcal{B}}(\mathbf{r}) = \begin{cases} (1/\varepsilon_s)\phi(\mathbf{r}\,|\,q) + \hat{q}/R\varepsilon_s, & \text{if } \mathbf{r} \in \mathfrak{v}, \\ (1/\varepsilon_c)\phi(\mathbf{r}\,|\,q), & \text{else.} \end{cases} \tag{54}$$

Here we have used the expression[59] (B12) for the potential $\phi(\mathbf{r}\,|\,q)$ of a point charge q and the definition (30) of the shielding charge $\hat{q}$. With $r \equiv |\mathbf{r}|$, the Born potential (54) and the definitions (8) and (15) yield the anti-polarization density

$$\mathbf{P}^{\mathcal{B}}(\mathbf{r}) = \frac{1}{4\pi}\frac{\varepsilon_c}{\varepsilon_s}\hat{q}\frac{\mathbf{r}}{r^3}\vartheta(\mathbf{r}) \tag{55}$$

associated to $\Phi^{\mathcal{B}}$, which is a radial function confined to $\mathfrak{v}$. The potential $\phi(\mathbf{r}\,|\,\mathbf{P}^{\mathcal{B}})$ generated by the anti-polarization density (55) is calculated in Appendix D and is given by

$$\frac{1}{\varepsilon_c}\phi(\mathbf{r}\,|\,\mathbf{P}^{\mathcal{B}}) = \begin{cases} \left(\frac{1}{\varepsilon_s} - \frac{1}{\varepsilon_c}\right)\phi(\mathbf{r}\,|\,q) + \frac{1}{\varepsilon_s}\frac{\hat{q}}{R}, & \text{if } \mathbf{r} \in \mathfrak{v}, \\ 0, & \text{else.} \end{cases} \tag{56}$$

A comparison with the Born expression (54) for the electrostatic potential of $\mathcal{B}$ shows that it can be more generally expressed as

$$\Phi^{\mathcal{B}}(\mathbf{r}) = (1/\varepsilon_c)[\phi(\mathbf{r}\,|\,q) + \phi(\mathbf{r}\,|\,\mathbf{P}^{\mathcal{B}})]. \tag{57}$$

As follows from the construction in Appendix D, $\Phi^{\mathcal{B}}(\mathbf{r})$ solves the ET representation (16) of the PE, i.e.,

$$\Delta\Phi^{\mathcal{B}}(\mathbf{r}) = -(4\pi/\varepsilon_c)[q\delta(\mathbf{r}) - \nabla\cdot\mathbf{P}^{\mathcal{B}}(\mathbf{r})]. \tag{58}$$

Because $\mathbf{P}^{\mathcal{B}}(\mathbf{r})$ was derived above through Eq. (15) from the Born potential $\Phi^{\mathcal{B}}(\mathbf{r})$ as given by Eq. (54), the consistency of this potential and of the anti-polarization density (55) with the basic Eqs. (16) and (15) of the exact ET theory has been demonstrated.

To derive the RF representation (34) of $\Phi^{\mathcal{B}}(\mathbf{r})$ we first note that, in this case, the non-radial contribution $\hat{\mathbf{P}}^{\mathcal{B}}(\mathbf{r})$ to the decomposition (24) of the anti-polarization density $\mathbf{P}^{\mathcal{B}}(\mathbf{r})$ vanishes such that $\mathbf{P}^{\text{rad},\mathcal{B}} = \mathbf{P}^{\mathcal{B}}$. The dipolar charge distribution $\rho^{\text{rad},\mathcal{B}} = -\nabla\cdot\mathbf{P}^{\mathcal{B}}(\mathbf{r})$, which by Eq. (27) belongs to $\mathbf{P}^{\text{rad},\mathcal{B}}$, is readily calculated by using the explicit expression (55) for $\mathbf{P}^{\mathcal{B}}(\mathbf{r})$. As is shown in Appendix E, one recovers the decomposition (29) of the dipolar charge distribution $\rho^{\text{rad},\mathcal{B}}$, if one uses the expression (30) for the shielding charge $\hat{q}$ and defines the shielding charge distribution $\hat{\rho}(\mathbf{r})$ occurring in Eq. (29) for the Born case as

$$\hat{\rho}^{\mathcal{B}}(\mathbf{r}) = \frac{\hat{q}}{4\pi R^2}\delta(r - R). \tag{59}$$

This shielding charge distribution is constant on the spherical surface of $\mathfrak{v}$ and obviously fulfills the condition (31) on its total strength. Therefore, it completely shields the point charge $-\hat{q}$, which, according to the decomposition (29) of $\rho^{\text{rad},\mathcal{B}}$, is located at the center of $\mathfrak{v}$. Consequently, the total potential $\phi(\mathbf{r}\,|\,\rho^{\text{rad},\mathcal{B}}) = \phi(\mathbf{r}\,|\,\mathbf{P}^{\mathcal{B}})$ vanishes outside $\mathfrak{v}$ [cf. Eq. (56)].

Thus, $\hat{\rho}^{\mathcal{B}}(\mathbf{r})$ allows us to bring the PE of the Born ion into the RF representation (32) which then reads

$$\Delta\Phi^{\mathcal{B}}(\mathbf{r}) = -(4\pi/\varepsilon_s)[q\delta(\mathbf{r}) + \hat{\rho}^{\mathcal{B}}(\mathbf{r})] \tag{60}$$

and which is solved by the potential

$$\Phi^{\mathcal{B}}(\mathbf{r}) = (1/\varepsilon_s)[\phi(\mathbf{r}\,|\,q) + \phi(\mathbf{r}\,|\,\hat{\rho}^{\mathcal{B}})]. \tag{61}$$

The RF potential (34) of $\mathcal{B}$ is the weakly shielded contribution

$$\Phi^{\text{RF},\mathcal{B}}(\mathbf{r}) = (1/\varepsilon_s)\phi(\mathbf{r}\,|\,\hat{\rho}^{\mathcal{B}}) \tag{62}$$

to the total potential (61), which is generated by the shielding charge distribution (59) of the Born ion. An explicit expression for this potential is calculated in Appendix F and is specified by Eq. (F2).

Figure 3 compares the ET (a) and RF (b) scenarios for the Born case. Just like in Fig. 2(a), also in Fig. 3(a) the gray background indicates the intrusion of the solvent continuum into $\mathfrak{v}$. The fat white arrows in $\mathfrak{v}$ represent the radial anti-polarization density $\mathbf{P}^{\mathcal{B}}(\mathbf{r})$ defined by Eq. (55), which serves to effectively expel the intruded solvent continuum from $\mathfrak{v}$.

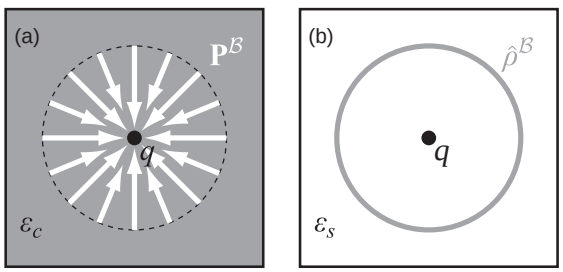

FIG. 3. Two representations of $\mathcal{B}$: (a) ET scenario as formulated by Eqs. (57) and (58); (b) RF approach as expressed by Eqs. (61) and (60); for explanation see the text.

According to the RF scenario sketched by Fig. 2(b), the solvent continuum has been eliminated through the introduction [cf. Eq. (29)] of the shielding charge distribution $\hat{\rho}^{\mathcal{B}}(\mathbf{r})$, which for $\mathcal{B}$ is confined to the surface of $\mathfrak{v}$ (gray circle).

3. The RF potential $\Phi^{RF,\mathcal{B}\cup\mathcal{C}}(\mathbf{r})$ of $\mathcal{B}\cup\mathcal{C}$

As shown in Sec. S1 of the supplementary material,[60] the potential of the field exposed Born ion is simply the superposition

$$\Phi^{\mathcal{B}\cup\mathcal{C}}(\mathbf{r}) = \Phi^{\mathcal{B}}(\mathbf{r}) + \Phi^{\mathcal{C}}(\mathbf{r}) \tag{63}$$

of the potentials of $\mathcal{B}$ and $\mathcal{C}$. In the ET representation, they are given by Eqs. (57) and (47), whereas Eqs. (61) and (52) pertain to the RF representation.

The definition (15) of the ET anti-polarization densities then immediately demonstrates that the density

$$\mathbf{P}^{\mathcal{B}\cup\mathcal{C}}(\mathbf{r}) = \mathbf{P}^{\mathcal{B}}(\mathbf{r}) + \mathbf{P}^{\mathcal{C}}(\mathbf{r}) \tag{64}$$

of $\mathcal{B}\cup\mathcal{C}$ is the sum of the completely radial density $\mathbf{P}^{\mathcal{B}}(\mathbf{r})$ given by Eq. (55) and the completely non-radial density $\mathbf{P}^{\mathcal{C}}(\mathbf{r})$ specified by Eq. (46). If one now considers the case $\mathcal{B}\cup\mathcal{C}$ as a most simple paradigm for any atom of a protein in solution (cf. the beginning of Sec. II B), then the decomposition (64) of $\mathbf{P}^{\mathcal{B}\cup\mathcal{C}}(\mathbf{r})$ into radial and non-radial parts *a posteriori* justifies the corresponding decomposition (24) of the exact atomic anti-polarization densities $\mathbf{P}_i(\mathbf{r})$ into $\mathbf{P}_i^{\mathrm{rad}}(\mathbf{r})$ and $\hat{\mathbf{P}}_i(\mathbf{r})$.

As soon as the necessity of separating the radial and non-radial contributions has become clear, the re-interpretation of $-\nabla\cdot\mathbf{P}_i^{\mathrm{rad}}(\mathbf{r})$ as an atomic dipolar charge distribution $\rho_i^{\mathrm{rad}}(\mathbf{r})$ is obvious and its further decomposition (29) into point- and shielding charge distributions can be deduced from case $\mathcal{B}$. Then one arrives at the new representation (34) of the RF potential, which for $\mathcal{B}\cup\mathcal{C}$ is

$$\Phi^{\mathrm{RF},\mathcal{B}\cup\mathcal{C}}(\mathbf{r}) = (1/\varepsilon_s)[\phi(\mathbf{r}\,|\,\hat{\rho}^{\mathcal{B}}) + \phi(\mathbf{r}\,|\,\hat{\mathbf{P}}^{\mathcal{C}})]. \tag{65}$$

The RF energy (36) of $\mathcal{B}\cup\mathcal{C}$ follows from evaluating the RF potential $\Phi^{\mathrm{RF},\mathcal{B}\cup\mathcal{C}}(\mathbf{r})$ at the position $\mathbf{r}=0$ of the ion, i.e.,

$$W^{\mathrm{RF},\mathcal{B}\cup\mathcal{C}} = (q/2\varepsilon_s)[\phi(0\,|\,\hat{\rho}^{\mathcal{B}}) + \phi(0\,|\,\hat{\mathbf{P}}^{\mathcal{C}})]. \tag{66}$$

Taking the diminished dipole $\hat{\mathbf{p}}$ as the source of the unshielded potential $\phi(\mathbf{r}\,|\,\hat{\mathbf{P}}^{\mathcal{C}})$, one finds $\phi(0\,|\,\hat{\mathbf{P}}^{\mathcal{C}})=0$ from the explicit expression (C5) implying that the anti-polarization density $\hat{\mathbf{P}}^{\mathcal{C}}(\mathbf{r})$ does not contribute to the solvation energy of $\mathcal{B}\cup\mathcal{C}$, that is, $W^{\mathrm{RF},\mathcal{B}\cup\mathcal{C}} = W^{\mathrm{RF},\mathcal{B}}$. Next, Eq. (F2) yields

$$\phi(0\,|\,\hat{\rho}^{\mathcal{B}}) = \hat{q}/R \tag{67}$$

for $\phi(\mathbf{r}\,|\,\hat{\rho}^{\mathcal{B}})$ at $\mathbf{r}=0$ and Eq. (66) reduces to

$$W^{\mathrm{RF},\mathcal{B}} = q\hat{q}/2\varepsilon_s R, \tag{68}$$

which is with the definition (30) of $\hat{q}$ the famous Born solvation energy of the ion.[49]

4. Discussion of $\mathcal{B}\cup\mathcal{C}$

If we compare the analytical solution (65) for the RF potential $\Phi^{\mathrm{RF},\mathcal{B}\cup\mathcal{C}}(\mathbf{r})$ of the field-exposed Born ion $\mathcal{B}\cup\mathcal{C}$ with the exact expression (34) for the RF potential $\Phi^{\mathrm{RF}}(\mathbf{r})$ of a protein $\mathcal{P}$, then we immediately see that the RF theory describes a protein as something like a collection of field exposed Born ions, because it combines Born contributions $\phi(\mathbf{r}\,|\,\hat{\rho}_i)$ with cavity terms $\phi(\mathbf{r}\,|\,\hat{\mathbf{P}}_i)$.

Considering the basic concepts of GB models[15,17,20,21] in the light of the exact RF expression (34) for $\Phi^{\mathrm{RF}}(\mathbf{r})$, one recognizes that GB models solely account for the Born contributions $\phi(\mathbf{r}\,|\,\hat{\rho}_i)$. This follows from the fact that the cavity terms $\phi(\mathbf{r}\,|\,\hat{\mathbf{P}}_i)$ are not included into this modeling concept: GB solely considers Coulomb interactions, which are modified by the shielding charge distributions; there is no explicit description of interactions between empty cavities and ions.

The approximate ET approach,[14] on the other hand, happened to completely neglect the Born contributions $\phi(\mathbf{r}\,|\,\hat{\rho}_i)$ and solely included cavity terms $\phi(\mathbf{r}\,|\,\hat{\mathbf{P}}_i)$. The Born contributions, which arise from the radial contributions $\mathbf{P}_i^{\mathrm{rad}}(\mathbf{r})$ to the atomic anti-polarization densities $\mathbf{P}_i(\mathbf{r})$, got lost during the construction of the approximate approach, because it was based on spatial integrations over the atomic volumes. Equation (40) reveals that the integrals over the $\mathbf{P}_i^{\mathrm{rad}}(\mathbf{r})$ vanish by symmetry such that only the $\hat{\mathbf{P}}_i(\mathbf{r})$ survived the integration. Thus, the $\hat{\mathbf{P}}_i(\mathbf{r})$ were erroneously considered as the sole contributions to the $\mathbf{P}_i(\mathbf{r})$.

A detection of the radial parts $\mathbf{P}_i^{\mathrm{rad}}(\mathbf{r})$ would have required to scrutinize the Born case within the framework of the ET approach as we have done above. Unfortunately, this consistency check, which would have revealed the completely radial character of $\mathbf{P}^{\mathcal{B}}(\mathbf{r})$ through Eq. (55), was omitted.

As a result, the GB model of protein continuum electrostatics and the approximate ET approach toward the same problem are characterized by fundamental and complementary shortcomings. We will take up this issue repeatedly further below.

III. APPROXIMATE THEORY

According to Sec. II A, the exact RF potential (34) of a protein $\mathcal{P}$ is generated by generally unknown atomic charge and anti-polarization densities $\hat{\rho}_i(\mathbf{r})$ and $\hat{\mathbf{P}}_i(\mathbf{r})$. Thus, an approximate theory requires suitable guesses $\tilde{\rho}_i(\mathbf{r})$ and $\tilde{\mathbf{P}}_i(\mathbf{r})$ for these functions. For the design of such guesses the corresponding analytically known densities $\hat{\rho}^{\mathcal{B}}(\mathbf{r})$ and $\hat{\mathbf{P}}^{\mathcal{C}}(\mathbf{r})$ of a field-exposed Born ion $\mathcal{B}\cup\mathcal{C}$, which are specified by Eqs. (59) and (50), respectively, and generate the exact RF potential (65), may serve as a guideline. Furthermore, the densities $\hat{\rho}_i(\mathbf{r})$ and $\hat{\mathbf{P}}_i(\mathbf{r})$ inherit from their parent anti-polarization densities $\mathbf{P}_i(\mathbf{r})$ through the unknown characteristic functions $\vartheta_i(\mathbf{r})$ appearing in Eq. (15) the confinement to the atomic fuzzy volumes $\mathfrak{v}_i$. Hence, choosing a certain approximate specification $\tilde{\vartheta}_i(\mathbf{r})$ of these spatial regions can give another hint, how one should choose the approximations $\tilde{\rho}_i(\mathbf{r})$ and $\tilde{\mathbf{P}}_i(\mathbf{r})$.

A. Gaussian approximation for $\vartheta_i(\mathbf{r})$

Following the ET suggestion, we will choose the Gaussian models $\tilde{\vartheta}_i(\mathbf{r}\,|\,\mathbf{r}_i, \tilde{v}_i, \sigma_i)$ defined by Eq. (17) for the characteristic functions $\vartheta_i(\mathbf{r})$. The parameters $\tilde{v}_i$ of these models measure the volumes of the Gaussian atoms and are adjusted to the respective configuration $\mathbf{R}$ of $\mathcal{P}$ by the self-consistency

iteration (21). The Gaussian widths σ_i are fixed parameters, which are typical for atoms i in given covalent binding motifs and can be chosen to optimize the accuracy of the approximation.

For the case $\mathcal{B} \cup \mathcal{C}$ of a single field-exposed ion, ET suggested the choice

$$\sigma = [(2/\pi)^{1/2}/3]^{1/3} R, \tag{69}$$

because then the Gaussian sphere $\tilde{\vartheta}(\mathbf{r} \,|\, 0, v, \sigma)$ defined by Eqs. (10), (17), and (21) has the same volume $4\pi R^3/3$ as the Born sphere $\mathfrak{v}$ [cf. Eq. (38) in Ref. 14].

B. Gaussian approximation for the $\hat{\rho}_i(\mathbf{r})$

In Sec. II B 2, the shielding charge distribution $\hat{\rho}^{\mathcal{B}}(\mathbf{r})$ of $\mathcal{B}$ turned out to be constant on the spherical surface of $\mathfrak{v}$ [cf. Eq. (59)] with the overall strength $\hat{q}$ given by the exact Eq. (30) in terms of the ion's charge q. For atoms i of Gaussian shape $\tilde{\vartheta}_i(\mathbf{r} \,|\, \mathbf{r}_i, \tilde{v}_i, \sigma_i)$ it seems, thus, reasonable to replace the exact shielding charge distributions $\hat{\rho}_i(\mathbf{r})$ by the Gaussian models

$$\tilde{\rho}_i(\mathbf{r} \,|\, \mathbf{r}_i, \hat{q}_i, \hat{\sigma}_i) = \hat{q}_i G(\mathbf{r} \,|\, \mathbf{r}_i, \hat{\sigma}_i) \tag{70}$$

and to derive the strengths $\hat{q}_i$ by Eq. (30) from the atomic partial charges q_i. Because $\hat{\rho}_i(\mathbf{r})$ originates from the divergence $-\nabla \cdot \mathbf{P}_i^{\mathrm{rad}}(\mathbf{r})$ of a radial anti-polarization density confined to $\mathfrak{v}_i$, its Gaussian width $\hat{\sigma}_i$ should be somewhat larger than the width σ_i of the confining function $\tilde{\vartheta}_i(\mathbf{r} \,|\, \mathbf{r}_i, \tilde{v}_i, \sigma_i)$.

If one applies the Gaussian model (70) also to the isolated ion $\mathcal{B}$, for which σ is fixed by Eq. (69), then the Gaussian width $\hat{\sigma}$ of the model shielding charge distribution $\tilde{\rho}^{\mathcal{B}}(\mathbf{r} \,|\, 0, \hat{q}, \hat{\sigma})$ can be determined by the requirement that the Gaussian model should lead to a Born energy $\tilde{W}^{\mathrm{RF},\mathcal{B}}$, which is equal to the exact Born energy $W^{\mathrm{RF},\mathcal{B}}$ given by Eq. (68).

To evaluate $\hat{\sigma}$, we first calculate the approximate RF potential

$$\tilde{\Phi}^{\mathrm{RF},\mathcal{B}}(\mathbf{r}) = (1/\varepsilon_s)\, \phi(\mathbf{r} \,|\, \tilde{\rho}^{\mathcal{B}}), \tag{71}$$

which is generated by the Gaussian shielding charge distribution $\tilde{\rho}^{\mathcal{B}}(\mathbf{r} \,|\, 0, \hat{q}, \hat{\sigma})$ of the analytically determined strength $\hat{q}$. It is given by the unshielded potential $\phi(\mathbf{r} \,|\, \tilde{\rho}^{\mathcal{B}})$ $\equiv \phi(\mathbf{r} \,|\, 0, \hat{q}, \hat{\sigma})$ specified[59] by Eq. (B1). At $\mathbf{r} = 0$ it has the value

$$\phi(0 \,|\, 0, \hat{q}, \hat{\sigma}) = \sqrt{2/\pi}\, (\hat{q}/\hat{\sigma}), \tag{72}$$

which is equal to the value (67) of the exact potential $\phi(\mathbf{r} \,|\, \hat{\rho}^{\mathcal{B}})$ at $\mathbf{r} = 0$, if the Gaussian width $\hat{\sigma}$ of the shielding charge distribution $\tilde{\rho}^{\mathcal{B}}(\mathbf{r} \,|\, 0, \hat{q}, \hat{\sigma})$ is chosen as

$$\hat{\sigma} = \sqrt{2/\pi}\, R. \tag{73}$$

Figure 4 compares the radial dependences of the approximate (71) and exact (62) Born potentials. By construction the potentials agree at $r = 0$, such that the Gaussian model yields the exact Born energy (68). Moreover, also the first and all higher derivatives vanish for both potentials at $r = 0$.[14] Because their sources are shielding charges of equal strengths, they also become identical in the limit $r \to \infty$. Thus, in both limits ($r \to 0$ and $r \to \infty$) the approximate Born potential $\tilde{\Phi}^{\mathrm{RF},\mathcal{B}}(\mathbf{r})$ shows the correct asymptotic behavior. Close

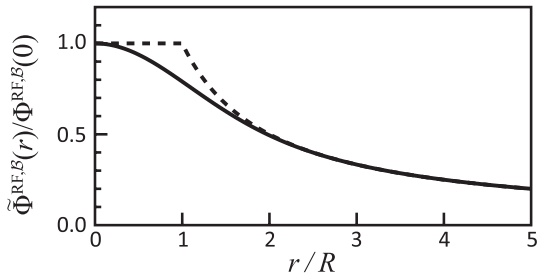

FIG. 4. Born ion: Radial decay of the approximate (solid) and exact (dashed) RF potentials $\tilde{\Phi}^{\mathrm{RF},\mathcal{B}}(r)$ and $\Phi^{\mathrm{RF},\mathcal{B}}(r)$, respectively, measured with respect to the common value $\Phi^{\mathrm{RF},\mathcal{B}}(0)$.

matches are visible in the regions up to $r/R \lesssim 0.5$ and beyond $r/R \gtrsim 1.5$. Near the surface of $\mathfrak{v}$, the approximate RF potential $\tilde{\Phi}^{\mathrm{RF},\mathcal{B}}(\mathbf{r})$ is a smoothed version of $\Phi^{\mathrm{RF},\mathcal{B}}(\mathbf{r})$.

For the range $0.5 < r/R < 1.5$, Figure 4 seems to indicate that $\tilde{\Phi}^{\mathrm{RF},\mathcal{B}}(r)$ represents a poor approximation to $\Phi^{\mathrm{RF},\mathcal{B}}(r)$. However, there is no physical scenario, in which the sizable deviations near $r/R = 1$ can be probed. Such probing would require that a point-like test charge approaches the ion up to the surface of its Born sphere, whose radius R is generally smaller than its vdW radius R_{vdW}.[61] But there are no point-like test charges in MM protein models. Possible test charges are partially or integrally charged atoms, which have comparable radii R as the ion considered in Fig. 4. In the absence of chemical bonds they could probe the potential at distances $r/R \geq 2$. However, such ions contribute an anti-polarizable cavity and a polarizing charge to the electrostatics problem and, therefore, cannot be discussed by solely considering the single ion case underlying Fig. 4. Interactions of an ion with a cavity or with another ion will be discussed further below.

Near $r/R = 1$, the Gaussian shielding potential $\tilde{\Phi}^{\mathrm{RF},\mathcal{B}}(r)$ shown in Figure 4 and specified by Eq. (72) does not represent an astonishingly good approximation to the potential $\Phi^{\mathrm{RF},\mathcal{B}}(r)$. Nevertheless, we have depicted this potential, because it will serve through insertion into Eq. (34) as a model for the atom-centered basis functions $\phi(\mathbf{r} \,|\, \hat{\rho}_i)$. The superposition of these approximate basis functions will have the task to model the RF potential of a whole protein $\mathcal{P}$, which is embedded in a dielectric continuum. In this context, the correct asymptotic behavior of the basis function $\phi(\mathbf{r} \,|\, 0, \hat{q}, \hat{\sigma})$ for $r \to 0$ and $r \to \infty$, which was noted in the discussion of Fig. 4, is of key importance. A suitable model for the second type $\phi(\mathbf{r} \,|\, \hat{\mathbf{P}}_i)$ of atom-centered basis functions, which also enter the superposition (34), will be now suggested by considering the case $\mathcal{C}$.

C. Gaussian approximation for $\hat{\mathbf{P}}_i(\mathbf{r})$

The homogeneous anti-polarization density $\mathbf{P}^{\mathcal{C}}(\mathbf{r})$, which is induced by the homogeneous external field $\mathbf{E}^{\mathrm{ext}}$ in the space $\mathfrak{v}$ occupied by the cavity $\mathcal{C}$, has the strength $\mathbf{p}$ given by the linear response Eq. (44). Hence, if one considers $\hat{\mathbf{P}}^{\mathcal{C}}(\mathbf{r})$ [cf. Eq. (50)] as a model for the cavity contributions $\hat{\mathbf{P}}_i(\mathbf{r})$ to the atomic anti-polarization densities $\mathbf{P}_i(\mathbf{r})$ and if one assumes that the external field, which is generated by the other atoms j of $\mathcal{P}$, is nearly constant within the atomic region $\mathfrak{v}_i$

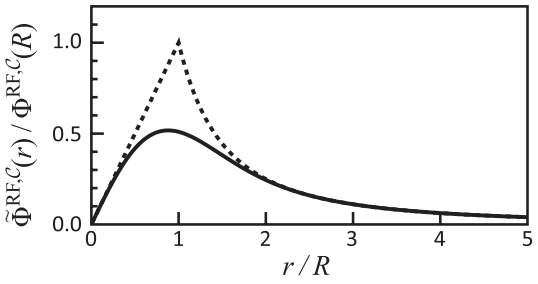

FIG. 5. Field-exposed cavity: One-dimensional sections of the approximate (solid) and exact (dashed) RF potentials $\tilde{\Phi}^{\mathrm{RF},\mathcal{C}}(r)$ and $\Phi^{\mathrm{RF},\mathcal{C}}(r)$, respectively, in the direction of $\mathbf{E}^{\mathrm{ext}}$. Both potentials are given in units of $\Phi^{\mathrm{RF},\mathcal{C}}(R)$.

characterized by $\vartheta_i(\mathbf{r})$ and occupied by atom i, then also $\hat{\mathbf{P}}_i(\mathbf{r})$ should be almost constant within $\mathfrak{v}_i$.

Interestingly, Gaussian approximations $\tilde{\mathbf{P}}_i(\mathbf{r}\,|\,\mathbf{r}_i, \tilde{\mathbf{p}}_i, \sigma_i)$ [as defined by Eq. (18)] for the anti-polarization densities $\hat{\mathbf{P}}_i(\mathbf{r})$ are constant within the Gaussian atoms defined by $\tilde{\vartheta}_i(\mathbf{r}\,|\,\tilde{v}_i, \mathbf{r}_i, \sigma_i)$ as is witnessed by the constant ratio $\tilde{\mathbf{P}}_i(\mathbf{r}\,|\,\mathbf{r}_i, \tilde{\mathbf{p}}_i, \sigma_i)/\tilde{\vartheta}_i(\mathbf{r}\,|\,\tilde{v}_i, \mathbf{r}_i, \sigma_i) = \tilde{\mathbf{p}}_i/\tilde{v}_i$. Therefore, the ET suggestion[14] to approximate almost constant anti-polarization densities like $\hat{\mathbf{P}}_i(\mathbf{r})$ by Gaussian dipole distributions $\tilde{\mathbf{P}}_i(\mathbf{r}\,|\,\mathbf{r}_i, \tilde{\mathbf{p}}_i, \sigma_i)$ should be reasonable.

If one applies this Gaussian modeling also to the field-exposed cavity $\mathcal{C}$, then the corresponding density $\tilde{\mathbf{P}}^{\mathcal{C}}(\mathbf{r}\,|\,0, \hat{\mathbf{p}}, \sigma)$, whose strength $\hat{\mathbf{p}} = (\varepsilon_s/\varepsilon_c)\mathbf{p}$ derives from the analytical result (44) and whose width σ is given by Eq. (69), generates the approximate RF potential

$$\tilde{\Phi}^{\mathrm{RF},\mathcal{C}}(\mathbf{r}) = (1/\varepsilon_s)\phi(\mathbf{r}\,|\tilde{\mathbf{P}}^{\mathcal{C}}).$$

The unshielded potential $\phi(\mathbf{r}\,|\tilde{\mathbf{P}}^{\mathcal{C}}) \equiv \phi(\mathbf{r}\,|\,0, \hat{\mathbf{p}}, \sigma)$, which is specified by Eq. (B5), is now supposed to approximate the exact potential $\phi(\mathbf{r}\,|\,\mathbf{P}^{\mathcal{C}})$, which is given by Eqs. (C4) and (C5).

Figure 5 illustrates the quality, by which the approximate RF potential $\tilde{\Phi}^{\mathrm{RF},\mathcal{C}}(r)$ models the exact solution $\Phi^{\mathrm{RF},\mathcal{C}}(r)$ for $\mathcal{C}$ in the direction of $\mathbf{E}^{\mathrm{ext}}$. Most importantly for the use of $\tilde{\Phi}^{\mathrm{RF},\mathcal{C}}(\mathbf{r})$ as a model for the atom-centered basis functions $\phi(\mathbf{r}\,|\,\hat{\mathbf{P}}_i)$ in the RF potential (34) of a protein $\mathcal{P}$, the two potentials become identical for $r \to 0$ and $r \to \infty$. Also for $\mathcal{C}$, the approximate RF potential $\tilde{\Phi}^{\mathrm{RF},\mathcal{C}}(r)$ is a smoothed version of the cavity's exact potential $\Phi^{\mathrm{RF},\mathcal{C}}(r)$ in the vicinity of the surface of $\mathfrak{v}$. Here, $\tilde{\Phi}^{\mathrm{RF},\mathcal{C}}(r)$ has the largest deviations from $\Phi^{\mathrm{RF},\mathcal{C}}(r)$, which, however, cannot be physically probed and, therefore, are irrelevant (see the discussion of Fig. 4).

GB models completely ignore the cavity-type contributions $\phi(\mathbf{r}\,|\,\hat{\mathbf{P}}_i)$ to the RF potential of a protein $\mathcal{P}$, for which suitable models $\tilde{\Phi}^{\mathrm{RF},\mathcal{C}}(\mathbf{r})$ are shown in Fig. 5. Similarly, the approximate ET approach erroneously did not include the Born-type contributions $\phi(\mathbf{r}\,|\,\hat{\rho}_i)$, which can be modeled by basis functions derived from the atom-centered approximation $\tilde{\Phi}^{\mathrm{RF},\mathcal{B}}(r)$ to $\Phi^{\mathrm{RF},\mathcal{B}}(r)$ shown in Fig. 4.

D. Approximate RF approach for solute proteins $\mathcal{P}$

The above comparisons suggest that the Gaussian models $\tilde{\rho}_i(\mathbf{r}\,|\,\mathbf{r}_i, \hat{q}_i, \hat{\sigma}_i)$ and $\tilde{\mathbf{P}}_i(\mathbf{r}\,|\,\mathbf{r}_i, \tilde{\mathbf{p}}_i, \sigma_i)$ for the shielding charge distributions $\hat{\rho}_i(\mathbf{r})$ and for the anti-polarization densities $\hat{\mathbf{P}}_i(\mathbf{r})$ should provide quite accurate approximations also for the atoms i of a solvated protein, if the interactions with the other protein atoms j are properly taken into account.

The corresponding approximate description is obtained from the exact representation (34) of the RF potential $\Phi^{\mathrm{RF}}(\mathbf{r})$ of $\mathcal{P}$, if one replaces the exact densities by the Gaussian approximations $\tilde{\rho}_i(\mathbf{r}\,|\,\mathbf{r}_i, \hat{q}_i, \hat{\sigma}_i)$ and $\tilde{\mathbf{P}}_i(\mathbf{r}\,|\,\mathbf{r}_i, \tilde{\mathbf{p}}_i, \sigma_i)$ as well as the exact atomic characteristic functions $\vartheta_i(\mathbf{r})$ by the models $\tilde{\vartheta}_i(\mathbf{r}\,|\,\mathbf{r}_i, \tilde{v}_i, \sigma_i)$. Note that $\tilde{\rho}_i(\mathbf{r}\,|\,\mathbf{r}_i, \hat{q}_i, \hat{\sigma}_i)$ has the same strength $\hat{q}_i$ [cf. Eq. (30)] as its exact relative $\hat{\rho}_i(\mathbf{r})$, because we impose the condition (31) also on $\tilde{\rho}_i(\mathbf{r}\,|\,\mathbf{r}_i, \hat{q}_i, \hat{\sigma}_i)$. The exact RF potential (34) of a solute protein $\mathcal{P}$ then becomes the approximate RF potential

$$\tilde{\Phi}^{\mathrm{RF}}(\mathbf{r}) = \frac{1}{\varepsilon_s}\sum_i[\phi(\mathbf{r}\,|\,\tilde{\rho}_i) + \phi(\mathbf{r}\,|\,\tilde{\mathbf{P}}_i)], \tag{74}$$

which is generated by the $\tilde{\rho}_i(\mathbf{r}\,|\,\mathbf{r}_i, \hat{q}_i, \hat{\sigma}_i)$ and $\tilde{\mathbf{P}}_i(\mathbf{r}\,|\,\mathbf{r}_i, \tilde{\mathbf{p}}_i, \sigma_i)$ and which yields the total approximate potential $\tilde{\Phi}(\mathbf{r})$ by replacing $\Phi^{\mathrm{RF}}(\mathbf{r})$ in Eq. (35) with $\tilde{\Phi}^{\mathrm{RF}}(\mathbf{r})$. The atom-centered basis functions $\phi(\mathbf{r}\,|\,\tilde{\rho}_i) \equiv \phi(\mathbf{r}\,|\,\mathbf{r}_i, \hat{q}_i, \hat{\sigma}_i)$ and $\phi(\mathbf{r}\,|\,\tilde{\mathbf{P}}_i) \equiv \phi(\mathbf{r}\,|\,\mathbf{r}_i, \tilde{\mathbf{p}}_i, \sigma_i)$ of a protein's RF potential are specified by Eqs. (B1) and (B5).

Because the Gaussian models $\tilde{\mathbf{P}}_i(\mathbf{r}\,|\,\mathbf{r}_i, \tilde{\mathbf{p}}_i, \sigma_i)$ are supposed to represent the exact atomic anti-polarization densities $\hat{\mathbf{P}}_i(\mathbf{r})$, their total strengths $\tilde{\mathbf{p}}_i$ have to obey a condition, which is analogous to Eq. (41) for the exact quantities $\hat{\mathbf{p}}_i$. Here, the exact electrostatic field $\mathbf{E}(\mathbf{r})$ has to be replaced by the approximate field $\tilde{\mathbf{E}}(\mathbf{r}) = -\nabla\tilde{\Phi}(\mathbf{r})$ and the exact atomic field-average (39) by the average

$$\langle\tilde{\mathbf{E}}(\mathbf{r}_i)\rangle_{\mathfrak{v}_i} = \frac{1}{\tilde{v}_i}\int \tilde{\mathbf{E}}(\mathbf{r})\,\tilde{\vartheta}_i(\mathbf{r}\,|\,\mathbf{r}_i, \tilde{v}_i, \sigma_i)\,\mathrm{d}V \tag{75}$$

over the atomic regions $\mathfrak{v}_i$. Within the chosen ET approach, the volumes $\tilde{v}_i$ of these regions are self-consistently calculated from Eq. (21). By the definition (17) of $\tilde{\vartheta}_i(\mathbf{r}\,|\,\mathbf{r}_i, \tilde{v}_i, \sigma_i)$, the average (75) is equivalent to the average $\langle\tilde{\mathbf{E}}(\mathbf{r}_i)\rangle_{\sigma_i}$ over the Gaussian shape function (10) defined by

$$\langle f(\mathbf{r}_i)\rangle_{\sigma_i} \equiv \int f(\mathbf{r})G(\mathbf{r}\,|\,\mathbf{r}_i, \sigma_i)\,\mathrm{d}V. \tag{76}$$

Following the arguments in Sec. 3 of Ref. 38, which are elaborated in Sec. S2 of the supplementary material,[60] the defining equation (41) of the atomic RF dipoles $\hat{\mathbf{p}}_i$ can be rearranged into the self-consistency condition

$$\tilde{\mathbf{p}}_i = -\alpha_i\langle\tilde{\mathbf{E}}(\mathbf{r}_i)\rangle_{\sigma_i}, \tag{77}$$

for anti-polarizable RF dipoles $\tilde{\mathbf{p}}_i$, which is analogous to the ST condition (22). The polarizing field is

$$\langle\tilde{\mathbf{E}}(\mathbf{r}_i)\rangle_{\sigma_i} = \frac{1}{\varepsilon_s}\sum_{j\neq i}[\langle\mathbf{E}(\mathbf{r}_i\,|\,\mathbf{r}_j, q_j)\rangle_{\sigma_i} + \langle\mathbf{E}(\mathbf{r}_i\,|\,\tilde{\mathbf{P}}_j)\rangle_{\sigma_i} + \langle\mathbf{E}(\mathbf{r}_i\,|\,\tilde{\rho}_j)\rangle_{\sigma_i}], \tag{78}$$

where the fields of the point charges q_j and Gaussian dipole or charge distributions $\tilde{\mathbf{P}}_j(\mathbf{r})$ or $\tilde{\rho}_j(\mathbf{r})$, respectively, are specified in Appendix B. Note that the Gaussian average $\langle\ldots\rangle_{\sigma_i}$ defined by Eq. (76) leads for any of the above potentials and fields to formally identical expressions, if the original widths σ_j are replaced by the larger widths $\sigma_{ij} \equiv (\sigma_i^2 + \sigma_j^2)^{1/2}$.

The atomic RF polarizabilities

$$\alpha_i = \left(3\,\varepsilon_s\,\sigma_i^3/2\right)\sqrt{\pi/2}\;S\left(v_i^{\varepsilon}\right) \tag{79}$$

appearing in Eq. (77) are identical to the ones given by ST with the screening function

$$S(x) = \frac{x}{1+(1-x/(2\sqrt{2}))/2} \tag{80}$$

and with the dielectrically weighted volume fractions

$$v_i^{\varepsilon} = \frac{\tilde{v}_i^{\gamma}}{\left(2\pi\sigma_i^2\right)^{3/2}}\left(1-\varepsilon_s/\varepsilon_c\right). \tag{81}$$

As is revealed by a careful comparison with Eq. (48) in Ref. 38, Eq. (81) contains the scaled atomic volumes

$$\tilde{v}_i^{\gamma} = \gamma\tilde{v}_i, \tag{82}$$

instead of the volumes $\tilde{v}_i$ self-consistently obtained through Eq. (21). The scaling by a factor $\gamma \gtrsim 1$ is necessary, because the self-consistency iteration (21) guarantees $\tilde{\Theta}(\mathbf{r}) = 1$ only at the atomic positions $\mathbf{r}_i$. Elsewhere, $\tilde{\Theta}(\mathbf{r})$ is on the average smaller than one (Fig. S13 in the supplementary material[60] provides an illustrative example). Therefore, the approximate anti-polarization density $\tilde{\mathbf{P}}(\mathbf{r})$, which belongs through Eq. (7) to the approximate characteristic function $\tilde{\Theta}(\mathbf{r})$ will be on the average smaller than the exact density $\mathbf{P}(\mathbf{r})$. Correspondingly, the RF dipoles $\tilde{\mathbf{p}}_i$ will be slightly too small. Because the atomic RF polarizabilities (79) monotonously increase with the scaled volumes $\tilde{v}_i^{\gamma}$, the factor $\gamma \gtrsim 1$ can empirically correct these minor underestimates of the $\tilde{\mathbf{p}}_i$.

1. Parameters of the approximate RF potential $\tilde{\Phi}^{RF}(\mathbf{r})$

Within the framework introduced above, the description of a macromolecule $\mathcal{P}$ embedded in a solvent continuum requires the specification of the parameter set

$$\Lambda = \{\gamma, \zeta, \sigma_i \,|\, i = 1, \ldots, N\}. \tag{83}$$

Here, γ is the minor correction (82) of the atomic volumes $\tilde{v}_i$, $\zeta \geq 1$ scales the Gaussian widths σ_i of the characteristic functions $\tilde{\vartheta}_i(\mathbf{r}\,|\,\mathbf{r}_i, \tilde{v}_i, \sigma_i)$, which by Eqs. (15) and (18) are also the widths of the atomic dipole densities $\tilde{\mathbf{P}}_i(\mathbf{r}\,|\,\mathbf{r}_i, \tilde{\mathbf{p}}_i, \sigma_i)$, into somewhat larger widths

$$\hat{\sigma}_i = \zeta\sigma_i \tag{84}$$

of the radial atomic shielding charge distributions $\tilde{\rho}_i(\mathbf{r}\,|\,\mathbf{r}_i, \hat{q}_i, \hat{\sigma}_i)$.

In the case $\mathcal{B}$, the scaling factor $\zeta^{\mathcal{B}}$ has been implicitly determined as $\zeta^{\mathcal{B}} = \sqrt[3]{6/\pi}$ by the requirement that $\tilde{\rho}^{\mathcal{B}}(\mathbf{r}\,|\,0, \hat{q}, \hat{\sigma})$ should yield the correct Born energy $W^{\mathrm{RF},\mathcal{B}}$ [cf. the remark following Eq. (73)], because here $\hat{\sigma}$ was expressed by the cavity radius R and because, for the isolated cavity, also σ had been given in terms of R [cf. Eq. (69)].

For $\mathcal{P}$ a suitable choice for ζ will be determined below by comparison with the exact RF energy of Kirkwood's spherical protein model.[51] Estimates for the widths σ_i of the Gaussian atoms i can be derived by Eq. (69) from atomic hard sphere radii R_i, which may be determined[14] from atomic vdW parameters listed in standard MM force fields.[6–8] On the other hand, the σ_i also can be optimized by correlating the atomic RF forces from explicit and continuum MM-MD descriptions[39] or by matching the reaction potential matrices[62] obtained from the approximate RF potential (74) and a numerical solution of the PE (1).

2. RF energy $\tilde{W}^{RF}(\mathbf{R})$

As soon as a reasonable parameter set Λ for a protein $\mathcal{P}$ of structure $\mathbf{R}$ is known, the approximate RF potential $\tilde{\Phi}^{\mathrm{RF}}(\mathbf{r})$ is calculated by the following procedure: First, the strengths $\hat{q}_i$ of the Gaussian shielding charge distributions $\tilde{\rho}_i(\mathbf{r}\,|\,\mathbf{r}_i, \hat{q}_i, \hat{\sigma}_i)$ are calculated by Eq. (30) from the partial charges q_i of the protein atoms. Next, a joint iteration of Eqs. (21) and (77) serves to self-consistently determine the effective atomic volumes $\tilde{v}_i$ and RF dipoles $\tilde{\mathbf{p}}_i$. The self-consistency procedures are terminated as soon as the maximum deviation $\max_i |\tilde{\Theta}(\mathbf{r}_i) - 1| < \chi_v$ of the approximate characteristic function from its target value of 1 and the maximum absolute change $\max_{i,u} |\tilde{p}_{i,u}^{(n)} - \tilde{p}_{i,u}^{(n-1)}| < \chi_p$ of the components $u \in \{x, y, z\}$ of the RF dipoles $\tilde{\mathbf{p}}_i$ between cycle (n) and $(n-1)$ are below certain thresholds χ_v and χ_p, respectively. In all RF calculations, the thresholds were chosen to be $\chi_v = 10^{-14}$ and $\chi_p = 10^{-14}$ D.

If we now evaluate $\tilde{\Phi}^{\mathrm{RF}}(\mathbf{r})$ as given by Eq. (74) at the atomic positions $\mathbf{r}_i$, the approximate RF energy of $\mathcal{P}$ is

$$\tilde{W}^{\mathrm{RF}}(\mathbf{R}) = \frac{1}{2}\sum_i q_i\tilde{\Phi}^{\mathrm{RF}}(\mathbf{r}_i\,|\,\mathbf{R}). \tag{85}$$

The RF forces required for a Hamiltonian MD simulation can be calculated by taking the configurational derivative of $\tilde{W}^{\mathrm{RF}}(\mathbf{R})$ subject to the constraints of Eqs. (21) and (77). Details will be described in Part II of this work.[57]

IV. COMPARISONS WITH EXACT AND NUMERICAL SOLUTIONS

In the construction of the approximate RF potential (74) for a solute protein $\mathcal{P}$ we have exploited our analytical results on the shielding charge distribution $\hat{\rho}^{\mathcal{B}}(\mathbf{r})$ of the isolated Born ion $\mathcal{B}$ and on the anti-polarization density $\hat{\mathbf{P}}^{\mathcal{C}}(\mathbf{r})$ of the field-exposed cavity, which generate the exact RF potential (65) of the field-exposed Born ion $\mathcal{B} \cup \mathcal{C}$. Beyond these most simple cases an analytical solution is solely available for Kirkwood's "spherical protein".[51]

A. Kirkwood's spherical protein

In 1934, Kirkwood[51] succeeded to find the analytical solution for the electrostatics of a spherical protein $\mathcal{K}$ embedded in a dielectric continuum ε_c. $\mathcal{K}$ is a spherical cavity $\mathfrak{V}_{\mathcal{K}}$ filled with point charges. Because of the superposition principle, which is applicable to the PE (1) as long as one considers a fixed cavity $\mathfrak{V}_{\mathcal{K}}$, it suffices to consider a unit charge e at an arbitrary position $\mathbf{r}$ within $\mathfrak{V}_{\mathcal{K}}$. The RF potential of an arbitrary charge distribution within $\mathfrak{V}_{\mathcal{K}}$ is then the sum of the RF potentials, which are generated by the individual charges residing at the same positions in $\mathfrak{V}_{\mathcal{K}}$.

As a result, $\mathcal{K}$ filled with a single charge is the most relevant reference model for the continuum electrostatics of a solute protein and, therefore, has been frequently utilized to judge the quality of approximate methods.[14,20,23–26,38,63] Therefore also we considered $\mathcal{K}$ with the aim of checking our approximate RF approach.

Following a corresponding suggestion in Ref. 14, a granular version $\tilde{\mathcal{K}}$ of $\mathcal{K}$ was constructed by placing 4187 Gaussian atoms i of identical widths $\sigma_i = 0.04212$ nm on a hexagonal close packing grid (lattice constant 0.1 nm) and by removing all grid points outside the exact Kirkwood sphere $\mathfrak{V}_\mathcal{K}$ of radius $R^\mathcal{K} = 0.89$ nm. In the model $\tilde{\mathcal{K}}$, the charge e must occupy a grid point k, of course, i.e., $q_i = e\delta_{ik}$ and $\mathbf{r} = \mathbf{r}_k$.

The coarse grained model $\tilde{\mathcal{K}}$ served us to fix the parameters γ and ζ for protein models. Here, the atomic volume scaling factor γ [cf. Eq. (82)] was determined by visually comparing the scaled characteristic function $\gamma\tilde{\Theta}^\mathcal{K}(\mathbf{r})$ of $\tilde{\mathcal{K}}$ with its exact counterpart $\Theta^\mathcal{K}(\mathbf{r})$. Figure S13 in the supplementary material[60] and the accompanying text explain our choice $\gamma = 1.03$ and illustrate the quality by which the scaled superposition $\gamma\tilde{\Theta}^\mathcal{K}(\mathbf{r})$ of Gaussians describes $\mathfrak{V}_\mathcal{K}$.

Furthermore, the scaling factor ζ, which relates the σ_i by Eq. (84) to the widths $\hat{\sigma}_i$ of the Gaussian shielding charge distributions $\tilde{\rho}_i(\mathbf{r}\,|\,\mathbf{r}_i, \hat{q}_i, \hat{\sigma}_i)$, was chosen by considering the solvation energy $W^{\mathrm{RF},\mathcal{B}}$ of the Born ion with radius $R^\mathcal{K}$ [cf. Eq. (68)], which is obtained by placing q_k into the center $\mathbf{r}_0 = 0$ of $\mathcal{K}$. With $\zeta = 1.5439$, we found that our coarse grained RF model of this large Born ion has the correct Born energy, i.e., $\tilde{W}^{\mathrm{RF},\mathcal{K}}(\mathbf{r}_0) = W^{\mathrm{RF},\mathcal{B}}$. Note that this value of ζ is by 24 % larger than the value $\zeta^\mathcal{B} = \sqrt[3]{6/\pi}$ applicable to an isolated ion in solution. Hence, within our RF approach the shielding charge distributions of atoms within proteins are more extended than those of isolated ions. We assume that the thus determined values of γ and ζ should be transferable to more realistic protein models and note that all parameters $\Lambda^\mathcal{K}$ [cf. Eq. (83)], which are required for the RF treatment of $\tilde{\mathcal{K}}$, are now specified.

Figure 6 illustrates the thus defined approximate RF treatment of $\tilde{\mathcal{K}}$ for a cross section covering the x-y plane and a case, in which q_k (white dot) is located at one of the Gaussian atoms (small circles) on the x axis close to the surface of $\mathcal{K}$ (large circle). Only the charged atom at $\mathbf{r}_k$ features a nonzero Gaussian shielding charge distribution $\tilde{\rho}_k(\mathbf{r})$ (larger filled gray circle), while all atoms carry mutually induced Gaussian anti-polarizations $\tilde{\mathbf{P}}_i(\mathbf{r})$ with the strengths $\tilde{\mathbf{p}}_i$ (black arrows). In the vicinity of q_k, the $\tilde{\mathbf{p}}_i$ point toward q_k. At larger distances the orientations of the $\tilde{\mathbf{p}}_i$ change. Here the dipoles become very small (note the logarithmic scaling of drawn dipole sizes). The thus determined dipole and shielding charge densities immediately yield by Eq. (74) the approximate RF potential $\tilde{\Phi}^{\mathrm{RF},\mathcal{K}}(\mathbf{r})$ and with $\tilde{W}^{\mathrm{RF},\mathcal{K}}(\mathbf{r}_k) = (q_k/2)\tilde{\Phi}^{\mathrm{RF},\mathcal{K}}(\mathbf{r}_k)$ the approximate RF energy.

Figure 7 shows $\tilde{W}^{\mathrm{RF},\mathcal{K}}(\mathbf{r}_k)$ (black triangles) for all lattice points $\mathbf{r}_k = k\,\hat{\mathbf{e}}_x$, $k = 0, 1, \ldots, 8$, on the positive x axis ($\hat{\mathbf{e}}_x$ is the associated basis vector). These results for $\tilde{\mathcal{K}}$ are compared with the exact Kirkwood[51] energies $W^{\mathrm{RF},\mathcal{K}}(x)$ (black dashed) calculated for $\mathcal{K}$. The triangles are seen to closely match the dashed line demonstrating the quality of our RF description. As is demonstrated by Figure S15 in the supplementary material, the RF description is even better than grid-based numerical solutions obtained with the well-known program DelPhi.[30,32]

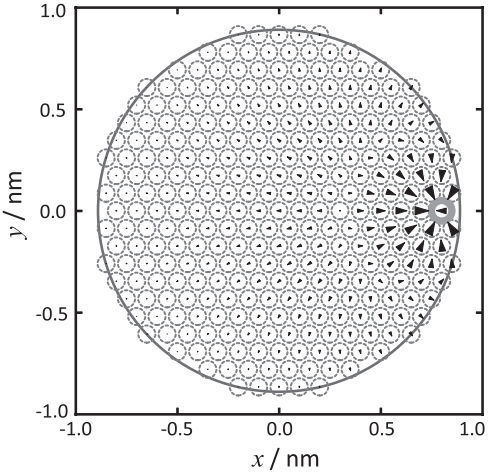

FIG. 6. Cross section through the discrete model $\tilde{\mathcal{K}}$ for the Kirkwood sphere $\mathcal{K}$ (large circle). The sizes of the Gaussian atoms i, which occupy a hexagonal grid, are indicated through their Gaussian radii σ (small circles). The charge q_k (white dot) at $\mathbf{r}_k$ induces a Gaussian shielding charge distribution $\tilde{\rho}_k(\mathbf{r})$ (gray filled circle) and RF dipoles $\tilde{\mathbf{p}}_i$ (black arrows), which are the sources of the RF potential $\tilde{\Phi}^{\mathrm{RF},\mathcal{K}}(\mathbf{r})$. The sizes of the arrows are proportional to $\ln|\tilde{\mathbf{p}}_i|$.

For further comparisons, Fig. 7 additionally displays the GB/CFA description[20,25] (gray dashed), which applies to the exact sphere $\mathcal{K}$, and the approximate ET result[14] (gray diamonds) for the coarse grained model $\tilde{\mathcal{K}}$. Whereas GB/CFA systematically and sizably underestimates $|W^{\mathrm{RF},\mathcal{K}}(x)|$ (cf. our corresponding remark in Sec. I), the ET description is seen to be of a comparably high quality as our RF approach (black triangles) despite the neglected radial contributions $\mathbf{P}_i^{\mathrm{rad}}(\mathbf{r})$ to the $\mathbf{P}_i(\mathbf{r})$. Hence, the question arises whether our RF approach can improve the description of protein electrostatics previously achieved by the ET approach at all.

In the search of an answer, we first refer to Fig. S14 in Sec. S3 B of the supplementary material,[60] which demonstrates for the granular Kirkwood sphere $\tilde{\mathcal{K}}$ that the ET models $\tilde{\mathbf{P}}_i^{\mathrm{ET}}(\mathbf{r})$ of the atomic dipole densities $\mathbf{P}_i(\mathbf{r})$ actually

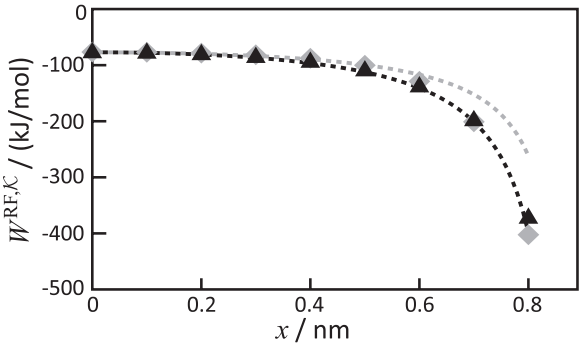

FIG. 7. Kirkwood's analytical solution[51] (black dashed) for the RF energy $W^{\mathrm{RF},\mathcal{K}}(x)$ of a unit charge e moving in a spherical cavity $\mathfrak{V}_\mathcal{K}$ of radius $R_\mathcal{K} = 0.89$ nm along the x axis of the hexagonal grid shown in Fig. 6 is compared for $0 < x \leq 0.8$ nm with our approximate RF prediction (black triangles) and with the approximate ET result[14] (gray diamonds). The gray dashed line is the corresponding GB/CFA result.[20,25]

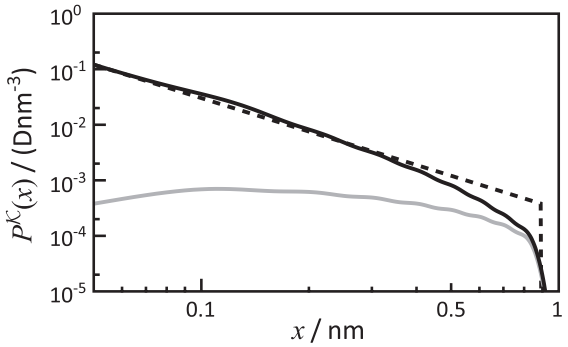

FIG. 8. Double logarithmic profiles of the total dipole densities $|\mathbf{P}^{\mathcal{K}}(x)|$ (exact, $\mathcal{K}$, black dashed), $|\tilde{\mathbf{P}}^{\mathcal{K}}(x)|$ (RF, $\tilde{\mathcal{K}}$, black solid), and $|\tilde{\mathbf{P}}^{\mathrm{ET},\mathcal{K}}(x)|$ (ET, $\tilde{\mathcal{K}}$, gray solid) along the x-axis of $\mathcal{K}$ for e at $\mathbf{r}_0 = 0$.

differ from the RF models $\tilde{\mathbf{P}}_i(\mathbf{r})$ concerning orientations and sizes.

Next we consider the exact anti-polarization density $\mathbf{P}^{\mathcal{K}}(\mathbf{r})$ of the whole Kirkwood sphere $\mathcal{K}$, which is defined by Eq. (7). For q_k at $\mathbf{r}_k = 0$ it is already known, because in this case $\mathcal{K}$ becomes a Born ion $\mathcal{B}$ with the radius $R^{\mathcal{K}}$, whose exact anti-polarization density $\mathbf{P}^{\mathcal{B}}(\mathbf{r}) \equiv \mathbf{P}^{\mathcal{K}}(\mathbf{r})$ is given by Eq. (55). In the ET scenario, the approximate counterpart $\tilde{\mathbf{P}}^{\mathrm{ET},\mathcal{K}}(\mathbf{r})$ is given by Eq. (20). In our RF approach, the computation of $\tilde{\mathbf{P}}^{\mathcal{K}}(\mathbf{r})$ is not quite as simple, because of the decomposition Eq. (24) of the $\mathbf{P}_i(\mathbf{r})$ into radial and non-radial parts. However, in Appendix G we derive an analytical expression for $\tilde{\mathbf{P}}^{\mathcal{K}}(\mathbf{r})$ [Eqs. (G1) and (G2)], which enables for $\tilde{\mathcal{K}}$ a comparison of $\tilde{\mathbf{P}}^{\mathcal{K}}(\mathbf{r})$ and of $\tilde{\mathbf{P}}^{\mathrm{ET},\mathcal{K}}(\mathbf{r})$ with $\mathbf{P}^{\mathcal{K}}(\mathbf{r})$.

Figure 8 compares $\mathbf{P}^{\mathcal{K}}(\mathbf{r})$ (black dashed) with the approximations $\tilde{\mathbf{P}}^{\mathcal{K}}(\mathbf{r})$ (black solid) and $\tilde{\mathbf{P}}^{\mathrm{ET},\mathcal{K}}(\mathbf{r})$ (gray, solid) along the positive x axis ($\mathbf{r} = x\,\hat{\mathbf{e}}_x$, $x > 0$). The RF approximation $\tilde{\mathbf{P}}^{\mathcal{K}}(\mathbf{r})$ is seen to closely agree with $\mathbf{P}^{\mathcal{K}}(x)$ for $x \in [0, 0.5]$ nm. Here, both curves represent a $1/x^2$ behavior. Solely close to the surface of $\mathcal{K}$, where $\mathbf{P}^{\mathcal{K}}(x)$ discontinuously jumps to zero, the RF approximation shows a smoothed decay, which is caused by the rough surface of the granular model $\tilde{\mathcal{K}}$, by the corresponding sizable deviations from the exact spherical shape of $\mathcal{K}$, and by the Gaussian shapes of the atoms. The ET approximation $\tilde{\mathbf{P}}^{\mathrm{ET},\mathcal{K}}(\mathbf{r})$, in contrast, is constant near the origin and by orders of magnitude too small. Magically, however, this deficiency does not hamper the description of $W^{\mathrm{RF},\mathcal{K}}(x)$ as we have noted above. But there are, as we will show now, situations, in which the incomplete ET anti-polarization density actually leads to strongly erroneous RF energies.

B. Ion–cavity interaction

For systems other than $\mathcal{B}$, $\mathcal{C}$, and $\mathcal{K}$, analytical solutions of the PE (1) do not exist. However, if such systems are sufficiently small, accurate numerical approximations can be calculated at a manageable effort by grid based PE solvers.[30,32] We will now discuss a simple example, namely, the interaction of two spherical objects of radius $R = 0.17$ nm, one representing a neutral atom, which is modeled as an empty cavity ($\mathcal{C}$), and the other an ion ($\mathcal{B}$) of unit charge $q = +e$. Both objects have the interior permittivity $\varepsilon_s = 1$, are embedded in an aqueous continuum ($\varepsilon_c = 80$), and are separated by the center-to-center distance r, which is varied in the range [0, 2] nm with steps $\delta r = 0.01$ nm. We will denote this solvated two-atomic molecule by "$\mathcal{BC}$" and calculate its RF energy $W^{\mathrm{RF},\mathcal{BC}}(r) = (1/2)q\,\Phi^{\mathrm{RF}}(\mathbf{r}_q \,|\, r)$, where $\mathbf{r}_q$ denotes the position of q, using different approximations.

Here, we applied the grid-based numerical approximation offered by DelPhi[30,32] using its focusing algorithm. Each of the hierarchically nested cubic grids, whose lattice constants decreased in the sequence 0.4 nm, 0.1 nm, 0.025 nm, and 0.01 nm, covered 301^3 points. Thus, the coarsest and finest grids represented boxes of 120^3 nm^3 and 3^3 nm^3, respectively.

Furthermore, we have calculated an approximate RF energy $\tilde{W}^{\mathrm{RF},\mathcal{BC}}(r)$ by a GB/CFA approach, in which the effective distance-dependent Born radius $R_{\mathcal{BC}}(r)$ of $\mathcal{BC}$ was most accurately calculated by numerically evaluating the volume integral in Eq. (14) of Ref. 17 on a grid covering as many as 4.9×10^9 points. Like in DelPhi, the ion $\mathcal{B}$ and the cavity $\mathcal{C}$ were modeled as spheres.

Finally, we have calculated $\tilde{W}^{\mathrm{RF},\mathcal{BC}}(r)$ using our RF and the ET[14] approach, which both model the interacting objects $\mathcal{B}$ and $\mathcal{C}$ as Gaussian spheres. Their widths σ were chosen by the volume criterion Eq. (69) applied above already to the isolated objects, the scaling factor ζ defined in Eq. (84) was set to $\zeta^{\mathcal{B}}$, and the factor γ from Eq. (82) to one. The CPU times needed for DelPhi[30,32] and GB/CFA, respectively, were 10^7 and 10^4 times longer than the times consumed by our RF or the ET approach.

Choosing the solvation energy $W^{\mathrm{RF},\mathcal{B}}$ of the isolated Born ion [cf. Eq. (68)] as the reference, Figure 9 compares the changes

$$\Delta\tilde{W}^{\mathrm{RF},\mathcal{BC}}(r) \equiv \tilde{W}^{\mathrm{RF},\mathcal{BC}}(r) - W^{\mathrm{RF},\mathcal{B}} \tag{86}$$

of the various approximate RF energies $\tilde{W}^{\mathrm{RF},\mathcal{BC}}(r)$ with the distance r of $\mathcal{B}$ and $\mathcal{C}$. Considering the difference between the DelPhi (black dashed) and all other displayed results one first notices that, for DelPhi, $\tilde{W}^{\mathrm{RF},\mathcal{BC}}(r)$ does not converge to the Born limit $W^{\mathrm{RF},\mathcal{B}}$, which should be reached for $r \to \infty$. This offset is a finite grid-size artifact.

GB/CFA (gray dashed), in contrast, reaches the correct limit for $r \to \infty$. For $r/R < 2$ it closely resembles the DelPhi result. Hence, GB/CFA performs for $\mathcal{BC}$ much better than in

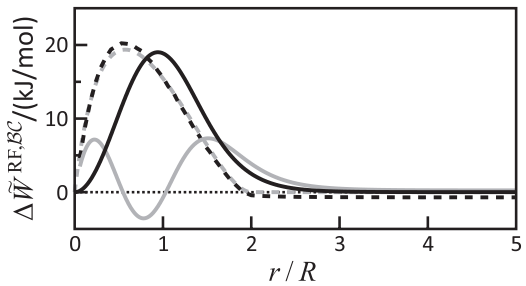

FIG. 9. Approximate RF energy changes (86) calculated by our RF approach (black solid) and by DelPhi[30,32] (black dashed) for a solute ion $\mathcal{B}$ and an empty cavity $\mathcal{C}$ as functions of their distance r/R. For comparison, ET (gray solid) and GB/CFA (gray dashed) results are also given. For discussion see the text.

the case of the spherical protein $\mathcal{K}$ (cf. Fig. 7). Note that a naive GB approach, which employs a constant Born radius and abstains from the volume estimate effected by the CFA, would predict $\Delta\tilde{W}^{\mathrm{RF},\mathcal{BC}}(r) = 0$ for all r.

Turning now to our RF description of $\Delta\tilde{W}^{\mathrm{RF},\mathcal{BC}}(r)$ (black solid) we recognize that it represents a slightly smoothed version of the DelPhi curve and predicts non-zero deviations from $W^{\mathrm{RF},\mathcal{B}}$ also for distances $r/R > 2$. The reason is that the Gaussian spheres of our RF model start to overlap already here, whereas the hard sphere overlap of the DelPhi and GB/CFA models starts at $r/R \leq 2$. Apart of this smoothing, the DelPhi and RF energy curves are similar. In particular, comparably large values are predicted for the energy barrier, which has to be surmounted upon shifting the cavity into the ion or vice versa.

Whereas the DelPhi solution (black dashed) in Fig. 9 supports our RF approximation, it is at variance with the ET energy curve (gray solid). Upon decreasing r this curve initially suggests the existence of an energy barrier. However, as soon as the Gaussian atoms show a considerable overlap the ET curve becomes grossly wrong. This failure results from the ET neglect of the radial contributions $\mathbf{P}_i^{\mathrm{rad}}(\mathbf{r})$ to the dipole densities in the interiors of the Gaussian atoms. On the other hand, the ET energetics is correct at large distances and not too bad at near contact. This finding explains the good performance of the ET approach on the coarse grained version of $\mathcal{K}$, because here the shortest distances between the single ion and the surrounding neutral atoms are $r \approx 1.5\,R$.

In the given case of the two atom system $\mathcal{BC}$, the suboptimal description delivered by our RF approach is caused by the rather poor approximation of the volume $\mathfrak{V}_s$ occupied by $\mathcal{BC}$ through two overlapping Gaussian spheres and by the fact that the approximate RF potential (74) is represented by only three basis functions, which are the potentials of two atomic anti-polarization densities $\tilde{\mathbf{P}}_i(\mathbf{r})$ and one shielding charge distribution $\tilde{\rho}_k(\mathbf{r})$. As one can see from the case $\tilde{\mathcal{K}}$ discussed above, in which a single Gaussian ion is surrounded by 4186 Gaussian cavities, the errors of our approximate RF approach become rapidly smaller with an increasing number of atoms. Thus $\mathcal{BC}$ represents a worst-case scenario for this approach.

On the other hand the $\mathcal{BC}$ example demonstrates that GB/CFA may also work quite well in certain cases and does not have to be as faulty as it turned out to be in the case $\mathcal{K}$ documented by Fig. 7. It has been shown that the quality of GB/CFA descriptions depends on the shape of the solute molecule and is hard to predict for a given molecule.[23–26] Thus, for GB/CFA there seems to be no systematic way to ensure the quality of the description, whereas our RF approach is granted to become better for large systems.

Note here that the supplementary material[60] compares in Sec. S4 further energy curves $\tilde{W}^{\mathrm{RF},\mathcal{BC}}(r)$ obtained by our RF approach and by DelPhi, respectively, for ions $\mathcal{B}$ and cavities $\mathcal{C}$ featuring unequal radii. This section additionally explains why only distances $r > \max(R_{\mathcal{C}}, R_{\mathcal{B}})$ are relevant when one envisages applications to molecular structures. For such distances, the deviations of our RF energies from the DelPhi results turn out to be small in all considered examples. Section S5 furthermore considers the cases of two oppositely and equally charged ions of equal radii and compares the RF, ET, and GB approximations with the numerical DelPhi solutions.

C. Free energy landscape of a model dipeptide

For a polar molecule, which is embedded in a polar solvent and features a given atomic configuration $\mathbf{R}$, the RF energy $\tilde{W}^{\mathrm{RF}}(\mathbf{R})$ covers the dominant electrostatic contribution to the free energy of solvation and represents a potential of mean force with respect to changes of $\mathbf{R}$. When added to a MM energy function $E^{\mathrm{MM}}(\mathbf{R})$, the shape of the resulting free energy function $G(\mathbf{R}) \approx E^{\mathrm{MM}}(\mathbf{R}) + \tilde{W}^{\mathrm{RF}}(\mathbf{R})$ maps the conformational landscape of the molecule.

Because simple model dipeptides such as Ac-Ala-NHMe possess only the two essential conformational degrees of freedom φ and ψ, which Fig. 10 highlights within Ac-Ala-NHMe's chemical structure, their conformational landscapes can be represented by graphs $\Delta G(\varphi, \psi) \equiv G(\varphi, \psi) - G_0$ resembling the well-known Ramachandran plots [the constant G_0 sets the minimum of $\Delta G(\varphi, \psi)$ to zero].

We have calculated free energy landscapes $\Delta G(\varphi, \psi)$ of Ac-Ala-NHMe in a dielectric continuum $\varepsilon_c = 80$ by DelPhi and by our RF method. Here, the parameters (γ, ζ) were set to the protein standard values (1.03, 1.5439). For the required MM energy function $E^{\mathrm{MM}}(\mathbf{R})$, we chose CHARMM22[6] combined with its CMAP[64] correction as implemented in the MD simulation program IPHIGENIE,[65] which additionally implements our RF approach for MD simulations.[57] The Ac-Ala-NHMe configurations $\mathbf{R}(\varphi, \psi)$ were adopted from Ref. 39, where the (φ, ψ)-plane was sampled in steps of 6° while all remaining configurational degrees of freedom were relaxed using the ST free energy function. In addition to the molecular configurations $\mathbf{R}(\varphi, \psi)$ both methods need values for the radii R_i of the atomic spheres (DelPhi) or σ_i of the atomic Gaussians (RF).

In our DelPhi description of Ac-Ala-NHMe, the atomic radii $R_{\tau(i)} = 2^{1/6}\sigma_{\tau(i)}^{\mathrm{vdW}}/2$ were derived from the vdW equilibrium distances $2^{1/6}\sigma_{\tau}^{\mathrm{vdW}}$ defined by CHARMM22 for atoms i of type τ. Figure S22 in the supplementary material[60] explains the atom types τ occurring in Ac-Ala-NHMe and Table S2 lists the associated radii R_τ. Furthermore, we chose 301^3 grid points, a three step focusing such that 1%, 70%, and 90% of the associated grid volumes were filled with Ac-Ala-NHMe.

For the RF calculation, we employed a simple definition of atom types $X(i)$, which solely distinguishes chemical elements. Table I lists our initial guesses σ_X^{ini} for the atomic Gaussian widths, which were motivated by choosing the associated spherical radii R_X^{ini} equal to typical C$-X$ bond lengths.

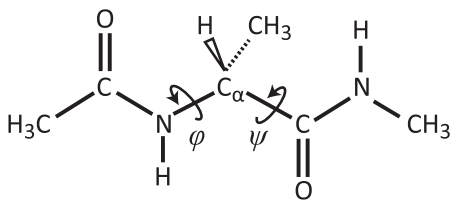

FIG. 10. Structure of Ac-Ala-NHMe. The conformational degrees of freedom are the two dihedral angles φ and ψ.

TABLE I. Atom types X, Gaussian widths σ_X, and associated [Eq. (69)] spherical radii R_X employed in the RF calculation of $\Delta G(\varphi, \psi)$ for Ac-Ala-NHMe.

X	C	N	O	H
R_X^{ini} [Å]	1.56	1.41	1.25	0.94
σ_X^{ini} [Å]	1.00	0.90	0.80	0.60
R_X [Å]	1.35	1.23	1.09	0.82
σ_X [Å]	0.87	0.78	0.69	0.52

The σ_X^{ini} were scaled by the factor $\lambda = 0.87$ to match the RF and DelPhi landscapes $W^{\text{RF}}[\mathbf{R}(\varphi, \psi)]$ at a single configuration. Here, we chose the α-helical configuration with $(\varphi_\alpha, \psi_\alpha) = (-66^\circ, -42^\circ)$, which marks the global minimum G_0 of the DelPhi landscape $G[\mathbf{R}(\varphi, \psi)]$. The scaled radii R_X and widths σ_X are listed in the last two lines of Table I.

The left and central graphs displayed by Figure 11 compare the free energy landscapes $\Delta G[\mathbf{R}(\varphi, \psi)]$ of Ac-Ala-NHMe computed by DelPhi and our RF approach, respectively, and immediately reveal the close similarity of the results. There are also minor differences. The global minimum of the RF description, for instance, is not at the same α-helical configuration $\mathbf{R}(\varphi_\alpha, \psi_\alpha)$ as the global minimum of DelPhi, but at an extended configuration $\mathbf{R}(\varphi_\beta, \psi_\beta)$ characterized by $(\varphi_\beta, \psi_\beta) = (-60^\circ, 144^\circ)$. According to DelPhi, $\mathbf{R}(\varphi_\beta, \psi_\beta)$ is by 1.6 kJ/mol above $\mathbf{R}(\varphi_\alpha, \psi_\alpha)$, whereas our RF approach predicts it by 0.4 kJ/mol below $\mathbf{R}(\varphi_\alpha, \psi_\alpha)$.

Clearly, these and other slight differences could be strongly diminished by a more elaborate choice of the Gaussian atomic widths σ_X, but this is not the objective here. Instead, the comparison in the left part of Fig. 11 serves to show that our RF approach is quite insensitive to the choice of the parameters σ_X, because already an almost arbitrary (but reasonable) choice brings the cheaply gained RF results close to those of the costly DelPhi calculations (with DelPhi the CPU time to compute the landscape $\Delta G[\mathbf{R}(\varphi, \psi)]$ was 10^5 times longer than with our RF approach using one core of a current PC processor). The insensitivity of our RF approach concerning the choice of the parameters σ_X is demonstrated in a little more depth by Sec. S7 of the supplementary material.[60] Here it is shown that the method is solely sensitive to the overall size of the σ_X, but hardly to the detailed differences among the σ_X as long as they are within the range of sizes given by Table I.

The right graph of Fig. 11 illustrates the importance of the RF contribution $W^{\text{RF}}[\mathbf{R}(\varphi, \psi)]$ to the free energy landscape $\Delta G[\mathbf{R}(\varphi, \psi)]$ of Ac-Ala-NHMe by showing the potential energy landscape $E^{\text{MM}}[\mathbf{R}(\varphi, \psi)]$ predicted by the CHARMM22/CMAP force field for Ac-Ala-NHMe isolated in the vacuum, which radically differs from the corresponding DelPhi and RF landscapes. Here the global minimum of ΔG is located at $(-156^\circ, 162^\circ)$. The local minimum in the α-helical region near $(-60^\circ, -60^\circ)$ is absent and a new minimum shows up at $(-78^\circ, 66^\circ)$. Furthermore, one immediately recognizes that the isolated Ac-Ala-NHMe is essentially constrained to the extended configurations in the upper left corner of the (φ, ψ)-plane, whereas in a polar liquid it may switch from extended to α-helical configurations.

From the related results presented by ST in Ref. 39, one now expects that free energy landscapes $\Delta G(\varphi, \psi)$ obtained by MD simulations of Ac-Ala-NHMe, which is embedded in a dielectric continuum as described by our RF approach, should be close to the free energy scans shown in Fig. 11. This is one of the questions, which will be answered in Part II of this work.[57]

V. SUMMARY, DISCUSSION, AND OUTLOOK

Motivated by consistent treatment of the Born ion $\mathcal{B}$ within the ET representation[14] (15) and (16) of continuum electrostatics, we have identified the radial contributions (25) to the atomic anti-polarization densities $\mathbf{P}_i(\mathbf{r})$ as the key contributions, which were missing in the original ET approach. Thus, the somewhat tedious and formal analysis of the cases $\mathcal{C}$ and $\mathcal{B}$ in Secs. II B 1 and II B 2, respectively, originally had provided the guideline for the formulation of the exact representation (34) of the RF potential $\Phi^{\text{RF}}(\mathbf{r})$ for solvated proteins in terms of atomic potentials generated by $\mathcal{B}$-type shielding charge densities $\hat{\rho}_i(\mathbf{r})$ of strengths $\hat{q}_i$, which are given by the atomic partial charges q_i through Eq. (30), and by $\mathcal{C}$-type anti-polarization densities $\hat{\mathbf{P}}_i(\mathbf{r})$, whose strengths $\hat{\mathbf{p}}_i$ must be self-consistently calculated from the locally averaged fields $\langle \mathbf{E} \rangle_{\mathfrak{v}_i}$.

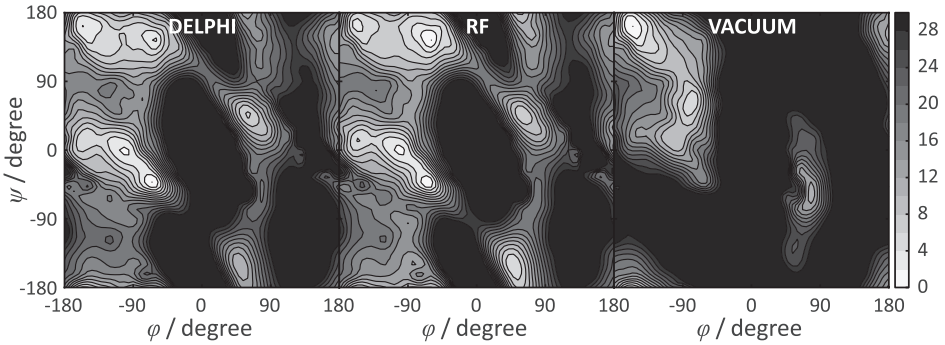

FIG. 11. Free energy landscapes $\Delta G[\mathbf{R}(\varphi, \psi)]$ of Ac-Ala-NHMe computed with the CHARMM22/CMAP force field[6,64] for identical configurations $\mathbf{R}(\varphi, \psi)$ describing the dielectric solvent continuum $\varepsilon_c = 80$ by DelPhi (left) and by our RF approach (center). The energy landscape $E^{\text{MM}}[\mathbf{R}(\varphi, \psi)]$ isolated molecule ($\varepsilon_c = 1$) is shown at the right. The contour levels represent steps of 2 kJ/mol. The color bar indicates the ΔG scale in kJ/mol. Energies above 30 kJ/mol were identified for a better resolution of the low-energy landscape.

Because Gaussian models for the densities $\hat{\rho}_i(\mathbf{r})$ and $\hat{\mathbf{P}}_i(\mathbf{r})$ delivered reasonable approximations for the electrostatic potentials in the analytically accessible cases $\mathcal{B}$ and $\mathcal{C}$ (cf. Figs. 4 and 5), an approximate expression $\tilde{\Phi}^{\mathrm{RF}}(\mathbf{r})$ for the RF potential was readily derived as the sum (74) of simple analytical functions, whose variables $\tilde{\mathbf{p}}_i$ and v_i are computed by the self-consistency iterations (21) and (77). Beside the scaling parameters (γ, ζ), for which the Kirkwood sphere $\mathcal{K}$ suggested the standard values (1.03, 1.5439), the parameters of the thus established approximate RF approach are the widths σ_i of the Gaussian atoms, which can be estimated from atomic vdW parameters available in common MM force fields. Results for the free energy landscape of the small model dipeptide Ac-Ala-NHMe indicated a weak dependence on the detailed choice of the σ_i.

A. Consequences for GB models

Because the representation (34) of the RF potential $\Phi^{\mathrm{RF}}(\mathbf{r})$ is exact, it enables an analysis of GB approaches.[15,17] For a most simple discussion, we assume that all partially charged atoms j have the same radius, i.e., $R_j = R$ for all j. According to Still *et al.*,[15] one can then write the GB model of the RF potential at the position $\mathbf{r}_i$ of the ion i as

$$\Phi^{\mathrm{RF,GB}}(\mathbf{r}_i) = \frac{1}{\varepsilon_s}\frac{\hat{q}_i}{R} + \sum_{j\neq i} \Phi^{\mathrm{SC,GB}}(r_{ij} \mid \hat{q}_j, R), \tag{87}$$

where $r_{ij} \equiv |\mathbf{r}_i - \mathbf{r}_j|$. The first term in Eq. (87) is the RF potential $\Phi^{\mathrm{RF},\mathcal{B}}(0)$ of a Born ion i at $\mathbf{r} = \mathbf{r}_i = 0$ [cf. Eqs. (62) and (67)], which is generated by its screening charge distribution $\hat{\rho}_i^{\mathcal{B}}(\mathbf{r})$ [cf. Eq. (59)], and the other terms are the potentials

$$\Phi^{\mathrm{SC,GB}}(r_{ij} \mid \hat{q}_j, R) \equiv \frac{1}{\varepsilon_s}\frac{\hat{q}_j}{\sqrt{r_{ij}^2 + R^2 \exp\left(-r_{ij}^2/4R^2\right)}} \tag{88}$$

generated by the other screening charges $\hat{q}_j$ [cf. Eq. (30)].

According to Eq. (34), the exact RF potential $\Phi^{\mathrm{RF}}(\mathbf{r})$ contains a sum over potentials $\phi(\mathbf{r} \mid \hat{\rho}_j)$, which are generated by the atomic shielding charge distributions $\hat{\rho}_j(\mathbf{r})$ of overall strengths $\hat{q}_j$ [cf. Eq. (31)]. In this respect, the GB approach resembles the exact potential. However, $\Phi^{\mathrm{RF}}(\mathbf{r})$ additionally contains the potentials $\phi(\mathbf{r} \mid \hat{\mathbf{P}}_j)$ generated by the atomic dipole densities $\hat{\mathbf{P}}_j(\mathbf{r})$, for which the GB expression (87) offers no correspondence.

In the simple scenario of identically sized ions, we can try a quantitative comparison. For this purpose, we bring the approximate counterpart (74) of Eq. (34) into a form comparable to the GB expression (87). With the translations (69) and (73) of the Gaussian widths σ_j and $\hat{\sigma}_j$ into the common atomic radius R and with the potentials $\phi(\mathbf{r} \mid \mathbf{r}_j, \hat{q}_j, \hat{\sigma}_j)$ of the Gaussian shielding charge distributions $\tilde{\rho}_j(\mathbf{r} \mid \mathbf{r}_j, \hat{q}_j, \hat{\sigma}_j)$, which are specified by Eqs. (B1) and (B4), one finds for the approximate RF potential at $\mathbf{r}_i$,

$$\tilde{\Phi}^{\mathrm{RF}}(\mathbf{r}_i) = \frac{1}{\varepsilon_s}\frac{\hat{q}_i}{R} + \frac{1}{\varepsilon_s}\sum_{j\neq i} \phi(\mathbf{r}_i \mid \mathbf{r}_j, \hat{q}_j, R\sqrt{2/\pi}) + \frac{1}{\varepsilon_s}\sum_{j\neq i} \phi(\mathbf{r}_i \mid \mathbf{r}_j, \tilde{p}_j, R[\sqrt{2/\pi}/3]^{1/3}). \tag{89}$$

Like in the GB expression (87), the first term is the potential (62) of the Born ion i at its center as specified by Eq. (67) and the second term a sum of modified Coulomb potentials

$$\Phi^{\mathrm{SC,RF}}(r_{ij} \mid \hat{q}_j, R) \equiv (1/\varepsilon_s)\,\phi\,(\mathbf{r}_i \mid \mathbf{r}_j, \hat{q}_j, \sqrt{2/\pi}\,R), \tag{90}$$

which are generated by the Gaussian screening charge distributions of strengths $\hat{q}_j$ at the other ions. As pointed out above, the third term has no correspondence in the GB expression (87).

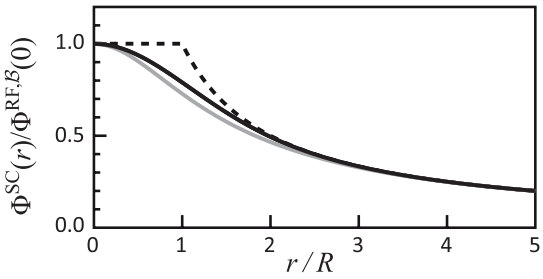

FIG. 12. The RF (black) and GB (gray) expressions for the modified Coulomb potentials $\Phi^{\mathrm{SC}}(r_{ij} \mid \hat{q}_j, R)$ of a screening charge $\hat{q}_j$ are compared with each other and with the RF potential $\Phi^{\mathrm{RF},\mathcal{B}}(r)$ (dashed) of a Born ion [cf. Eq. (62)] as functions of the distance $r \equiv r_{ij}$. All potentials are given in units of the RF potential $\Phi^{\mathrm{RF},\mathcal{B}}(0)$, which is generated by the shielding charge distribution $\hat{\rho}_j^{\mathcal{B}}(\mathbf{r})$ at the center $\mathbf{r}_j = 0$ of the Born ion j.

Figure 12 demonstrates the modified Coulomb potentials $\Phi^{\mathrm{SC,GB}}(r_{ij} \mid \hat{q}_j, R)$ (gray) and $\Phi^{\mathrm{SC,RF}}(r_{ij} \mid \hat{q}_j, R)$ (black) defined by Eqs. (88) and (90), respectively, are actually very similar. Thus, one could equally well exchange the GB-type expression[15] for the potentials of the $\hat{q}_j$ by the potential of a corresponding Gaussian charge distributions and still get a GB model.

Note furthermore that the black solid and dashed curves in Fig. 12 are simply copies of the corresponding curves in Fig. 4 and that the dashed curve represents the exact solution for the potential of an isolated Born ion. Hence, if one considers the GB and RF expressions (88) and (90) as approximations to the Born potential $\Phi^{\mathrm{RF},\mathcal{B}}(r)$, then our Gaussian RF approximation performs even slightly better than GB (see, however, our discussion following Fig. 4 on the physical interpretation of such a comparison).

Summarizing, we may state now that our approximate RF description (74) reduces to a GB model, if the nonradial contributions $\hat{\mathbf{P}}_i(\mathbf{r})$ are neglected and if the screening charge densities $\hat{\rho}_i(\mathbf{r})$, which originate from the radial contributions $\mathbf{P}_i^{\mathrm{rad}}(\mathbf{r})$ to the dipole densities $\mathbf{P}_i(\mathbf{r})$, are taken as the sole sources of $\Phi^{\mathrm{RF}}(\mathbf{r})$. An approximate anti-polarization density exclusively constructed from the $\mathbf{P}_i^{\mathrm{rad}}(\mathbf{r})$, however, cannot fulfill the boundary conditions at the surface of $\mathfrak{V}_s$, because this task requires self-consistently determined contributions $\hat{\mathbf{P}}_i(\mathbf{r})$ at all protein atoms. With the aim of repairing

the thus generated deficiency, GB approaches try to approximately fulfill the boundary conditions by applying the CFA or other corrections.[15–26] On the other hand, the approximate ET approach[14] while perfectly fulfilling the boundary conditions unfortunately overlooked the GB contributions to the RF potential.

As a result, the extension of the ET approach by the radial dipole densities $\mathbf{P}_i^{\mathrm{rad}}(\mathbf{r})$ and its transformation into the RF scenario can now also be viewed as a marriage of ET and GB concepts, which removes complementary shortcomings of both approaches. Concerning GB it renders the complicated attempts[15–26] of estimating effective configuration-dependent atomic radii superfluous. With respect to ET it removes basic inconsistencies implying erroneous small distance limits. Particularly, the newly derived representation (85) of the RF energy is key for the development of applications to MD in a continuum electrostatics setting.

B. Computational aspects

The formulas Eq. (74) for the approximate RF potential $\tilde{\Phi}^{\mathrm{RF}}(\mathbf{r})$, Eq. (77) and Eq. (21) for the self-consistency iterations of the $\tilde{\mathbf{p}}_i$ and v_i, and Eq. (85) for the RF energy $\tilde{W}^{\mathrm{RF}}$ are quite simple and easily programmable, as one sees by inspecting the explicit formulas listed in Appendix B. Implementations should exploit the fact that, with increasing distance, the potentials of Gaussian charges and dipoles rapidly approach the potentials of their point-like relatives.

By integrating the longer range parts of the RF electrostatics into the hierarchical fast multipole method, which is called SAMM_4 and is part of the MD program package IPHIGENIE,[65] our existing RF implementation features a linearly scaling computational effort for solute molecules of increasing size. Because IPHIGENIE offers the treatment of polarizable MM force fields,[55,56] whose polarizable degrees of freedom are modeled as inducible Gaussian dipoles located at the heavy atoms, the programming of our RF approach, which partially [Eq. (77)] has the form of a corresponding anti-polarizable force field, was considerably simplified. Correspondingly, IPHIGENIE also offered a suitable platform for the development of a Hamiltonian RF-MD method, which will be described[57] in Part II of this work. As a result, the cost of RF energy and force calculations is comparable to that of a polarizable MM force field, which also requires self-consistency iterations of atomic dipoles.

C. Outlook

The theory developed in this paper is confined to the continuum electrostatics of proteins in dielectric media. The presence of ions, which cause an increased shielding of electrostatic interactions, is important particularly for highly charged macromolecules like DNA. It may be described within the Debye-Hückel approximation by the linearized Poisson-Boltzmann equation (LPBE).[66]

In principle, the inclusion of an ionic continuum into our theory should pose no problem, because a corresponding template is available with Ref. 67, which extends the ET theory[14] to the case of the LPBE. In fact, a closer look at this extension reveals that it can be easily integrated into our RF approach. As far as an efficient implementation, which tries to take advantage of a fast multipole method like SAMM_4, is concerned, there are, however, technical difficulties, because one has to replace the expansions of ordinary multipole potentials by expansions of exponentially screened multipole potentials. But with sufficient patience also these difficulties can be surmounted.

The second important task, which has still to be tackled, is the fine-tuning of the atomic Gaussian widths σ_i in such a way that solvation energies $\tilde{W}^{\mathrm{RF}}(\mathbf{R})$ calculated for a large number of different peptides and proteins represent reasonable approximations to experimental findings or to well-tuned calculations with DelPhi or related methods. At the current stage, experience suggests that relative changes of $\tilde{W}^{\mathrm{RF}}(\mathbf{R})$ induced by changes of the molecular configuration are only weakly affected by different choices of the σ_i as long as an overall scaling factor, which brings the solvation energy for a given configuration of a solute molecule close to a known reference value, is correspondingly chosen (see the Ac-Ala-NHMe example discussed above).

ACKNOWLEDGMENTS

This work was supported by the Deutsche Forschungsgemeinschaft (SFB749/C4).

APPENDIX A: TRANSFORMING RADIAL ANTI-POLARIZATION INTO RADIAL CHARGE DENSITIES

To show that the charge distribution (29) corresponding to the radial anti-polarization density (25) depends only on the distance $|\mathbf{r}-\mathbf{r}_i|$, we first assume $\mathbf{r}_i = 0$. Calculating the charge distribution from Eq. (27) and using spherical coordinates, we get

$$\nabla \cdot [f(r)\mathbf{r}] = \frac{1}{r^2}\frac{\partial}{\partial r} r^2 f(r) r,$$

where Eq. (25) has been employed for the dipole distribution. One immediately finds

$$\nabla \cdot [f(r)\mathbf{r}] = 3f(r) + rf'(r)$$

with $f'(r)$ denoting the derivative of $f(r)$ as given by Eq. (26). Shifting $\mathbf{r}_i$ back to its original position one gets

$$\rho_i^{\mathrm{rad}}(\mathbf{r}) = 3f(|\mathbf{r}-\mathbf{r}_i|) + |\mathbf{r}-\mathbf{r}_i| f'(|\mathbf{r}-\mathbf{r}_i|), \qquad \text{(A1)}$$

which is the desired radial charge distribution.

APPENDIX B: ELECTROSTATIC POTENTIALS AND FIELDS

The approximate RF potential (74) and the associated polarizing field (78) are given in terms of the potentials and fields of Gaussian charges $\tilde{\rho}(\mathbf{r}\,|\,\mathbf{r}_j, q_j, \sigma_j)$ and dipoles $\tilde{\mathbf{P}}(\mathbf{r}\,|\,\mathbf{r}_j, \tilde{\mathbf{p}}_j, \sigma_j)$ which are specified below (including the special case of point charges and dipoles, i.e., $\sigma_j = 0$).

At the position $\mathbf{r}_i$ of an atom i, the electrostatic potential

$$\phi(\mathbf{r}_i \,|\, \mathbf{r}_j, q_j, \sigma_j) = q_j\, \mu(r_{ij}, \sigma_j) \tag{B1}$$

generated by a Gaussian charge distribution $\tilde{\rho}(\mathbf{r} \,|\, \mathbf{r}_j, q_j, \sigma_j)$ is given in terms of the function

$$\mu(r, \sigma) = (1/r)\,\mathrm{erf}(r/\sqrt{2}\sigma) \tag{B2}$$

and of the inter-atomic distance $r_{ij} \equiv |\mathbf{r}_{ij}| \equiv |\mathbf{r}_i - \mathbf{r}_j|$. In the limit of $r \to 0$, we get from

$$\lim_{r\to 0} \mu(r, \sigma) = \sqrt{2}/\sqrt{\pi}\sigma\,, \tag{B3}$$

the potential of a Gaussian charge distribution at its center

$$\phi(\mathbf{r}_j \,|\, \mathbf{r}_j, q_j, \sigma_j) = \sqrt{2/\pi}\; q_j/\sigma_j\,. \tag{B4}$$

Similarly, the potential of a Gaussian dipole density $\tilde{\mathbf{P}}(\mathbf{r} \,|\, \mathbf{r}_j, \tilde{\mathbf{p}}_j, \sigma_j)$ is

$$\phi(\mathbf{r}_i \,|\, \mathbf{r}_j, \mathbf{p}_j, \sigma_j) = \left(\mathbf{p}_j \cdot \mathbf{r}_{ij}/r_{ij}^2\right)\, \kappa(r_{ij}, \sigma_j) \tag{B5}$$

with

$$\kappa(r, \sigma) = \mu(r, \sigma) - \sigma^2 \eta(r, \sigma) \tag{B6}$$

and

$$\eta(r, \sigma) = \sqrt{2/\pi}(1/\sigma^3)\exp(-r^2/2\sigma^2). \tag{B7}$$

For $\mathbf{r}_i \to \mathbf{r}_j$, the potential $\phi(\mathbf{r}_i \,|\, \mathbf{r}_j, \mathbf{p}_j, \sigma_j)$ vanishes as one concludes from Eqs. (9), (49), and Appendix C in Ref. 14.

The corresponding electric fields

$$\mathbf{E}(\mathbf{r}_i \,|\, \mathbf{r}_j, q_j, \sigma_j) = q_j \left(\mathbf{r}_{ij}/r_{ij}^2\right) \kappa(r_{ij}, \sigma_j) \tag{B8}$$

and

$$\mathbf{E}(\mathbf{r}_j \,|\, \mathbf{r}_j, \mathbf{p}_j, \sigma_j) = [\kappa(r_{ij}, \sigma_j)\, \mathbf{G}(\mathbf{r}_{ij}) - \eta(r_{ij}, \sigma_j)\, \mathbf{A}(\mathbf{r}_{ij})]\mathbf{p}_j \tag{B9}$$

are the negative gradients of the potentials (B1) and (B5), respectively, where we used the tensorial expressions (see, e.g., Ref. 65)

$$\mathbf{A}(\mathbf{r}) = (\mathbf{r} \otimes \mathbf{r})\,/r^2 \tag{B10}$$

and

$$\mathbf{G}(\mathbf{r}) = (3\mathbf{A} - \mathbf{I})\,/r^2. \tag{B11}$$

In the limit of vanishing Gaussian widths σ_j, the potentials (B1) and (B5) reduce for $i \neq j$ to the well-known potentials

$$\phi(\mathbf{r}_i \,|\, \mathbf{r}_j, q_j) = q_j/r_{ij} \tag{B12}$$

of point charges and

$$\phi(\mathbf{r}_i \,|\, \mathbf{r}_j, \mathbf{p}_j) = \mathbf{p}_j \cdot \mathbf{r}_{ij}/r_{ij}^3 \tag{B13}$$

of point dipoles.

APPENDIX C: POTENTIAL OF THE DIPOLE DENSITY $\mathbf{P}^{\mathcal{C}}$

The electrostatic potential of an arbitrary dipole density $\mathbf{P}(\mathbf{r})$ is given by[44]

$$\phi(\mathbf{r} \,|\, \mathbf{P}) = -\nabla \cdot \int \frac{\mathbf{P}(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}\, \mathrm{d}V'. \tag{C1}$$

Inserting the homogeneous anti-polarization density (46), which is generated by $\mathbf{E}^{\mathrm{ext}}$ in the spherical cavity $\mathfrak{v}$, using the definition (42) of the characteristic function ϑ, and orienting $\mathbf{r}$ along the z-axis of the Cartesian coordinate system gives

$$\phi(\mathbf{r} \,|\, \mathbf{P}^{\mathcal{C}}) = -\frac{\mathbf{p}}{v} \cdot \nabla I^{\mathcal{C}}(\mathbf{r}) \tag{C2}$$

with the integral

$$I^{\mathcal{C}}(\mathbf{r}) = \int_{\mathfrak{v}} (r^2 - 2rr'\cos\theta' + r'^2)^{-1/2}\, \mathrm{d}V'$$

expressed by spherical coordinates. Integration over φ' and θ' yields[58]

$$I^{\mathcal{C}}(\mathbf{r}) = 2\pi \int_0^R \frac{r'}{r}\left(r + r' - \sqrt{(r - r')^2}\right)\, \mathrm{d}r'. \tag{C3}$$

For $\mathbf{r} \notin \mathfrak{v}$, we have $r > r'$. Therefore, the integral is

$$I^{\mathcal{C}}(\mathbf{r}) = 2\pi \int_0^R \frac{2r'^2}{r} = \frac{4\pi}{3}\,\frac{R^3}{r}.$$

Insertion into Eq. (C2) gives

$$\phi(\mathbf{r} \,|\, \mathbf{P}^{\mathcal{C}}) = \frac{\mathbf{p} \cdot \mathbf{r}}{r^3}, \qquad \mathbf{r} \notin \mathfrak{v}. \tag{C4}$$

A comparison with Eq. (B13) demonstrates that this is the electrostatic potential $\phi(\mathbf{r} \,|\, \mathbf{p})$ of a point dipole $\mathbf{p}$ located at the origin. For $\mathbf{r} \in \mathfrak{v}$, the integral (C3) splits into contributions with $r > r'$ and $r < r'$, respectively,

$$I^{\mathcal{C}}(\mathbf{r}) = 4\pi \int_0^r \frac{r'^2}{r}\, \mathrm{d}r' + 4\pi \int_r^R r'\, \mathrm{d}r',$$

such that the integral is

$$I^{\mathcal{C}}(\mathbf{r}) = -4\pi \left(\frac{1}{3}r^2 - R^2\right).$$

With $v = 4\pi R^3/3$, the potential Eq. (C2) of the homogeneous anti-polarization density $\mathbf{P}^{\mathcal{C}}(\mathbf{r})$ defined by Eq. (49) now is

$$\phi(\mathbf{r} \,|\, \mathbf{P}^{\mathcal{C}}) = \frac{\mathbf{p} \cdot \mathbf{r}}{R^3}, \qquad \mathbf{r} \in \mathfrak{v}. \tag{C5}$$

APPENDIX D: POTENTIAL OF THE ANTI-POLARIZATION DENSITY $\mathbf{P}^{\mathcal{B}}$

Instead of using Eq. (C1) for the potential of a dipole density $\mathbf{P}(\mathbf{r})$, we start with the equivalent expression[68]

$$\phi(\mathbf{r} \,|\, \mathbf{P}) = \int \frac{\mathbf{P}(\mathbf{r}') \cdot (\mathbf{r} - \mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|^3}\, \mathrm{d}V'. \tag{D1}$$

Inserting for $\mathbf{P}$ the anti-polarization density $\mathbf{P}^{\mathcal{B}}$ of the Born ion [Eq. (55)], we obtain

$$\phi(\mathbf{r} \,|\, \mathbf{P}^{\mathcal{B}}) = \frac{1}{4\pi}\,\frac{\varepsilon_c}{\varepsilon_s}\,\hat{q}\, I^{\mathcal{B}}(\mathbf{r}) \tag{D2}$$

with the integral

$$I^{\mathcal{B}}(\mathbf{r}) = \int_{\mathfrak{v}} \frac{r'\cos\theta - r}{r'^2\sqrt{r^2 - rr'\cos\theta + r'^2}}\, \mathrm{d}V'$$

written with spherical coordinates. Integrating over φ' and θ', we find[58]

$$I^{\mathcal{B}}(\mathbf{r}) = -2\pi \int_0^R \frac{1}{r'^2}\left[1 - \frac{r - r'}{\sqrt{(r-r')^2}}\right] \mathrm{d}r'. \qquad \text{(D3)}$$

For $r > r'$, the integrand vanishes. For $\mathbf{r} \notin \mathfrak{v}$, we have $r > r'$. Therefore, the potential

$$\phi(\mathbf{r} \mid \mathbf{P}^{\mathcal{B}}) = 0, \qquad \mathbf{r} \notin \mathfrak{v}, \qquad \text{(D4)}$$

vanishes outside the ion. For $\mathbf{r} \in \mathfrak{v}$, the integral (D3) is readily evaluated yielding

$$I^{\mathcal{B}}(\mathbf{r}) = 4\pi(1/R - 1/r).$$

With Eq. (D2), the potential of the anti-polarization density $\mathbf{P}^{\mathcal{B}}(\mathbf{r})$ is

$$\phi(\mathbf{r} \mid \mathbf{P}^{\mathcal{B}}) = \frac{\varepsilon_c}{\varepsilon_s}\hat{q}\left(\frac{1}{R} - \frac{1}{r}\right), \qquad \mathbf{r} \in \mathfrak{v}, \qquad \text{(D5)}$$

in the interior of the ion. Using the definition (30) for the shielding charge $\hat{q}$ in the $1/r$-term and the expression[59] (B12) for the potential $\phi(\mathbf{r} \mid q)$ of a point charge, one finally gets from Eqs. (D4) and (D5) the expression (56) for the potential $\phi(\mathbf{r} \mid \mathbf{P}^{\mathcal{B}})$.

APPENDIX E: THE CHARGE DENSITY $\rho^{\mathrm{rad},\mathcal{B}}$

With the definition (55) of the anti-polarization density $\mathbf{P}^{\mathcal{B}}(\mathbf{r})$ and Eq. (27), the equivalent charge distribution is

$$\rho^{\mathrm{rad},\mathcal{B}}(\mathbf{r}) = -\frac{1}{4\pi}\frac{\varepsilon_c}{\varepsilon_s}\hat{q}\nabla \cdot \left[\frac{\mathbf{r}}{r^3}\vartheta(\mathbf{r})\right]. \qquad \text{(E1)}$$

With the identity $\mathbf{r}/r^3 = -\nabla 1/r$, the divergence applied to the first term in the bracket is $\nabla \cdot \nabla 1/r = \Delta 1/r$, which gives[44] $\Delta 1/r = -4\pi\delta(\mathbf{r})$. Furthermore, the divergence of the spherical characteristic function is[14]

$$\nabla\vartheta(\mathbf{r}) = -\delta(r - R)\mathbf{r}/r,$$

such that Eq. (E1) becomes

$$\rho^{\mathrm{rad},\mathcal{B}}(\mathbf{r}) = -\frac{1}{4\pi}\frac{\varepsilon_c}{\varepsilon_s}\hat{q}\left[4\pi\delta(\mathbf{r}) - \frac{1}{r^2}\delta(r - R)\right].$$

The second term in brackets is nonzero only at $r = R$ and we can write

$$\rho^{\mathrm{rad},\mathcal{B}}(\mathbf{r}) = \frac{\varepsilon_c}{\varepsilon_s}\hat{q}\left[-\delta(\mathbf{r}) + \frac{1}{4\pi R^2}\delta(r - R)\right]. \qquad \text{(E2)}$$

With the definition (59) of the surface charge distribution $\hat{\rho}^{\mathcal{B}}$, Eq. (E2) immediately results in the decomposition (29) for the charge distribution $\rho^{\mathrm{rad},\mathcal{B}}$.

APPENDIX F: THE BORN POTENTIAL $\Phi^{\mathcal{B}}$ EXPRESSED IN TERMS OF $\hat{\rho}^{\mathcal{B}}$

Because the anti-polarization density $\mathbf{P}^{\mathcal{B}}$ and the charge distribution $\rho^{\mathrm{rad},\mathcal{B}}$ generate the same electrostatic potential, the Born potential (57), which is given in terms of $\phi(\mathbf{r} \mid \mathbf{P}^{\mathcal{B}})$, can be expressed equally well in terms of $\phi(\mathbf{r} \mid \rho^{\mathrm{rad},\mathcal{B}})$. According to Eq. (29), $\rho^{\mathrm{rad},\mathcal{B}}$ consists of the point charge $-\hat{q}$ and the shielding charge distribution $\hat{\rho}^{\mathcal{B}}$, which are both magnified by the factor $\varepsilon_c/\varepsilon_s$. Thus, the potential

$$\phi(\mathbf{r} \mid \rho^{\mathrm{rad},\mathcal{B}}) = -\frac{\varepsilon_c}{\varepsilon_s}\phi(\mathbf{r} \mid \hat{q}) + \frac{\varepsilon_c}{\varepsilon_s}\phi(\mathbf{r} \mid \hat{\rho}^{\mathcal{B}})$$

can be split into the associated scaled potentials. Inserting the definition (30) of $\hat{q}$ into the potential $\phi(\mathbf{r} \mid \hat{q})$ yields

$$\phi(\mathbf{r} \mid \rho^{\mathrm{rad},\mathcal{B}}) = \left(\frac{\varepsilon_c}{\varepsilon_s} - 1\right)\phi(\mathbf{r} \mid q) + \frac{\varepsilon_c}{\varepsilon_s}\phi(\mathbf{r} \mid \hat{\rho}^{\mathcal{B}}). \qquad \text{(F1)}$$

Replacing in the expression (57) for $\Phi^{\mathcal{B}}$ the potential $\phi(\mathbf{r} \mid \mathbf{P}^{\mathcal{B}})$ by $\phi(\mathbf{r} \mid \rho^{\mathrm{rad},\mathcal{B}})$ as given now by Eq. (F1) yields Eq. (61), which is completely specified once an explicit expression for the potential $\phi(\mathbf{r} \mid \hat{\rho}^{\mathcal{B}})$ of the shielding charge distribution $\hat{\rho}^{\mathcal{B}}$ is available.

For its calculation, we insert the specifications Eq. (F1) of $\phi(\mathbf{r} \mid \rho^{\mathrm{rad},\mathcal{B}})$ and Eqs. (D4) and (D5) of $\phi(\mathbf{r} \mid \mathbf{P}^{\mathcal{B}})$ into the equivalence of these potentials. We obtain

$$\phi(\mathbf{r} \mid \hat{\rho}^{\mathcal{B}}) = \begin{cases} \hat{q}/R & \text{if } \mathbf{r} \in \mathfrak{v}, \\ \phi(\mathbf{r} \mid \hat{q}) & \text{if } \mathbf{r} \notin \mathfrak{v}, \end{cases} \qquad \text{(F2)}$$

where we have used the definition (B12) for the potential of a point charge.

APPENDIX G: TOTAL ANTI-POLARIZATION DENSITY $\tilde{\mathbf{P}}(\mathbf{r})$

The RF approximation

$$\tilde{\mathbf{P}}(\mathbf{r}) = \sum_i \left[\tilde{\mathbf{P}}_i^{\mathrm{rad}}(\mathbf{r}) + (\varepsilon_c/\varepsilon_s)\tilde{\mathbf{P}}_i(\mathbf{r})\right] \qquad \text{(G1)}$$

of the total anti-polarization density (7) can be calculated, once the radial contributions $\tilde{\mathbf{P}}_i^{\mathrm{rad}}$ are reconstructed from the Gaussian shielding charge densities $\tilde{\rho}_i(\mathbf{r} \mid \mathbf{r}_i, \hat{q}_i, \hat{\sigma}_i)$. For this purpose, one must first construct by Eq. (29) associated dipolar charge densities $\tilde{\rho}_i^{\mathrm{rad}}$ and subsequently solve the equation

$$-\nabla \cdot \tilde{\mathbf{P}}_i^{\mathrm{rad}}(\mathbf{r}) = \frac{\varepsilon_c}{\varepsilon_s}\hat{q}_i\left[-\delta(\mathbf{r}) + G(\mathbf{r} \mid 0, \hat{\sigma}_i)\right],$$

for $\tilde{\mathbf{P}}_i^{\mathrm{rad}}$, where we have chosen $\mathbf{r}_i = 0$ for simplicity. By symmetry, the solution must have the form $\tilde{\mathbf{P}}_i^{\mathrm{rad}}(\mathbf{r}) = p(r)\hat{\mathbf{e}}_{\mathbf{r}}$, where $\hat{\mathbf{e}}_{\mathbf{r}}$ is the radial unit vector. Expressing the divergence in spherical coordinates leads to

$$\frac{1}{r^2}\frac{\partial}{\partial r}r^2 p(r) = \frac{\varepsilon_c}{\varepsilon_s}\hat{q}_i\left[-\delta(\mathbf{r}) + G(\mathbf{r} \mid 0, \hat{\sigma}_i)\right].$$

Direct integration yields[58]

$$p(r) = \frac{\varepsilon_c}{\varepsilon_s}\frac{\hat{q}_i}{4\pi}\frac{1}{r}\left[\frac{1}{r} - \kappa(r, \hat{\sigma}_i)\right]$$

with κ given by Eq. (B6). Thus, the desired RF approximation is

$$\tilde{\mathbf{P}}_i^{\mathrm{rad}}(\mathbf{r}) = \frac{\hat{q}_i\varepsilon_c}{4\pi\varepsilon_s}\frac{\mathbf{r} - \mathbf{r}_i}{|\mathbf{r} - \mathbf{r}_i|^2}\left[\frac{1}{|\mathbf{r} - \mathbf{r}_i|} - \kappa(|\mathbf{r} - \mathbf{r}_i|, \hat{\sigma}_i)\right]. \qquad \text{(G2)}$$

According to Eq. (G2), $\tilde{\mathbf{P}}_i^{\mathrm{rad}}$ is a superposition of the electric fields of the point charge $\hat{q}_i$ and of a Gaussian charge distribution with total charge $-\hat{q}_i$, which are both localized at $\mathbf{r}_i$.

[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: Part 1*, edited by J. Buchner and T. Kiefhaber (Wiley-VCH, Weinheim, 2005), pp. 1170–1195.
[6]A. D. MacKerell, 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, J. Phys. Chem. B **102**, 3586 (1998).
[7]D. Pearlman, D. Case, J. Caldwell, W. Ross, T. Cheatham III, S. DeBolt, D. Ferguson, G. Seibel, and P. Kollman, Comput. Phys. Commun. **91**, 1 (1995).
[8]M. Christen, P. H. Hünenberger, D. Bakowies, R. Baron, R. Bürgi, D. P. Geerke, T. N. Heinz, M. A. Kastenholz, V. Kräutler, C. Oostenbrink, C. Peter, D. Trzesniak, and W. F. van Gunsteren, J. Comput. Chem. **26**, 1719 (2005).
[9]T. A. Darden, D. York, and L. Pedersen, J. Chem. Phys. **98**, 10089 (1993).
[10]W. L. Jorgensen, J. Chandrasekhar, and J. D. Madura, J. Chem. Phys. **79**, 926 (1983).
[11]H. Berendsen, J. Grigera, and T. Straatsma, J. Phys. Chem. **91**, 6269 (1987).
[12]W. Weber, P. H. Hünenberger, and J. A. McCammon, J. Phys. Chem. B **104**, 3668 (2000).
[13]G. Mathias, B. Egwolf, M. Nonella, and P. Tavan, J. Chem. Phys. **118**, 10847 (2003).
[14]B. Egwolf and P. Tavan, J. Chem. Phys. **118**, 2039 (2003).
[15]W. C. Still, A. Tempczyk, R. C. Hawley, and T. Hendrickson, J. Am. Chem. Soc. **112**, 6127 (1990).
[16]M. Schaefer and M. Karplus, J. Phys. Chem. **100**, 1578 (1996).
[17]D. Bashford and A. Case, Annu. Rev. Phys. Chem. **51**, 129 (2000).
[18]M. S. Lee, F. R. Salsbury, Jr., and C. L. Brooks III, J. Chem. Phys. **116**, 10606 (2002).
[19]W. Im, M. S. Lee, and C. L. Brooks III, J. Comput. Chem. **24**, 1691 (2003).
[20]T. Grycuk, J. Chem. Phys. **119**, 4817 (2003).
[21]M. Wojciechowski and B. Lesyng, J. Phys. Chem. B **108**, 18368 (2004).
[22]E. Gallicchio and R. M. Levy, J. Comput. Chem. **25**, 479 (2004).
[23]G. Sigalov, P. Scheffel, and A. Onufriev, J. Chem. Phys. **122**, 094511 (2005).
[24]G. Sigalov, A. Fenley, and A. Onufriev, J. Chem. Phys. **124**, 124902 (2006).
[25]J. Mongan, W. A. Svrcek-Seiler, and A. Onufriev, J. Chem. Phys. **127**, 185101 (2007).
[26]A. Fenley, J. C. Gordon, and A. Onufriev, J. Chem. Phys. **129**, 075101 (2008).
[27]M. K. Gilson, M. E. Davis, B. A. Luty, and J. A. McCammon, J. Phys. Chem. **97**, 3591 (1993).
[28]M. Gilson, J. A. McCammon, and J. D. Madura, J. Comput. Chem. **16**, 1081 (1995).
[29]M. E. Davis, J. D. Madura, B. A. Luty, and J. A. McCammon, Comput. Phys. Commun. **62**, 187 (1991).
[30]W. Rocchia, E. Alexov, and B. Honig, J. Phys. Chem. B **105**, 6507 (2001).
[31]R. Luo, L. David, and M. K. Gilson, J. Comput. Chem. **23**, 1244 (2002).
[32]A. Nicholls and B. Honig, J. Comput. Chem. **12**, 435 (1991).
[33]W. Geng, S. Yu, and G. Wei, J. Chem. Phys. **127**, 114106 (2007).
[34]A. H. Juffer, E. F. Botta, B. A. van Keulen, A. van der Ploeg, and H. J. Berendsen, J. Comput. Phys. **97**, 144 (1991).
[35]J. Liang and S. Subramaniam, Biophys. J. **73**, 1830 (1997).
[36]M. D. Altman, J. P. Bardhan, J. K. White, and B. Tidor, J. Comput. Chem. **30**, 132 (2009).
[37]R. Yokota, J. P. Bardhan, M. G. Knepley, L. Barba, and T. Hamada, Comput. Phys. Commun. **182**, 1272 (2011).
[38]M. Stork and P. Tavan, J. Chem. Phys. **126**, 165105 (2007).
[39]M. Stork and P. Tavan, J. Chem. Phys. **126**, 165106 (2007).
[40]Y. I. Kharkats, A. A. Kornyshev, and M. A. Vorotyntsev, J. Chem. Soc., Faraday Trans. 2 **72**, 361 (1976).
[41]R. Zhou and B. J. Berne, Proc. Natl. Acad. Sci. U.S.A. **99**, 12777 (2002).
[42]H. Nymeyer and A. E. García, Proc. Natl. Acad. Sci. U.S.A. **100**, 13934 (2003).
[43]M. Feig, A. Onufriev, M. S. Lee, W. Im, D. A. Case, and C. L. Brooks III, J. Comput. Chem. **25**, 265 (2004).
[44]J. D. Jackson, *Classical Electrodynamics* (John Wiley & Sons, Inc., New York/London/Sydney/Toronto, 1975).
[45]W. Im, D. Beglov, and B. Roux, Comput. Phys. Commun. **111**, 59 (1997).
[46]Q. Lu and R. Luo, J. Chem. Phys. **119**, 11035 (2003).
[47]W. Geng and G. W. Wei, J. Comput. Phys. **230**, 435 (2011).
[48]One grid-based approach[47] to dielectric boundary forces yields, for instance, for a system of two identical Born ions separated by the distance r in the limit $r \to 0$ a large nonzero force acting on one of the ions, instead of the physically required vanishing force correctly predicted by Ref. 45; see Figs. 5(b) and 5(d) in the quoted paper and compare also with Fig. 1 in Ref. 57.
[49]M. Born, Z. Phys. **1**, 45 (1920).
[50]C. D. Snow, N. Nguyen, V. S. Pande, and M. Gruebele, Nature (London) **420**, 102 (2002).
[51]J. G. Kirkwood, J. Chem. Phys. **2**, 351 (1934).
[52]H. Sklenar, F. Eisenhaber, M. Poncin, and R. Lavery, in *Theoretical Biochemistry and Molecular Biophysics 2: Proteins*, edited by D. L. Beveridge and R. Lavery (Adenine Press, New York, 1991), pp. 317–335.
[53]M. Kloppenburg and P. Tavan, Phys. Rev. E **55**, 2089 (1997).
[54]P. Ahlström, A. Wallquist, S. Engström, and B. Jönsson, Mol. Phys. **68**, 563 (1989).
[55]M. Schwörer, B. Breitenfeld, P. Tröster, K. Lorenzen, P. Tavan, and G. Mathias, J. Chem. Phys. **138**, 244103 (2013).
[56]P. Tröster, K. Loenzen, M. Schwörer, and P. Tavan, J. Phys. Chem. B **117**, 9486 (2013).
[57]S. Bauer, P. Tavan, and G. Mathias, J. Chem. Phys. **140**, 104103 (2014).
[58]I. N. Bronstein, K. A. Semendjajew, G. Musiol, and H. Mühlig, *Taschenbuch der Mathematik* (Verlag Harri Deutsch, Frankfurt am Main, 1999).
[59]Appendix B contains explicit expressions for the potentials and fields of point or Gaussian charges and dipoles.
[60]See the supplementary material at http://dx.doi.org/10.1063/1.4867280 which provides on 16 pages in 7 sections a total of 12 figures (S13– S24), 13 equations (S91)– (S103), and two tables (S2, S3) as additional material to the derivation of the approximate RF approach (III) and to the comparisons (IV) with numerical results.
[61]As one concludes from Table I and S2,[60] the Born radii R_X of the heavy atoms of type X, which occur in the di-peptide Ac-Ala-NHMe, are by about a factor of 1.5 smaller than their vdW radii $R_{\mathrm{vdW}}(X)$.
[62]J. P. Bardhan, J. Chem. Phys. **129**, 144105 (2008).
[63]D. Borgis, N. Levy, and M. Marchi, J. Chem. Phys. **119**, 3516 (2003).
[64]A. D. MacKerell, Jr., M. Feig, and C. L. Brooks, III, J. Comput. Chem. **25**, 1400 (2004).
[65]K. Lorenzen, M. Schwörer, P. Tröster, S. Mates, and P. Tavan, J. Chem. Theory Comput. **8**, 3628 (2012).
[66]P. Debye and E. Hückel, Physikalische Zeitschrift **24**, 185 (1923).
[67]B. Egwolf and P. Tavan, J. Chem. Phys. **120**, 2056 (2004).
[68]D. J. Griffiths, *Introduction to Electrodynamics* (Prentice Hall International, Inc., 1999).

2.2 Zusatzinformationen zur Berechnung der Reaktionsfeld-Energie von Proteinen

Der folgende Abdruck[2]

Supporting material to:

Electrostatics of proteins in dielectric sovent continua.
I. An accurate and efficient reaction field description“

Sebastian Bauer, Gerald Mathias, and Paul Tavan
J. Chem. Phys. **140**, 104102 (2014)

enthält zusätzliche Informationen zur approximativen Berechnung von RF-Energien, die nicht im Haupttext der oben abgedruckten Veröffentlichung enthalten sind. Auf 16 Seiten, bestehend aus sieben Abschnitten, mit 13 Gleichungen, 12 Abbildungen und zwei Tabellen werden (i) weitere Unterschiede und Gemeinsamkeiten zwischen der neuen Methode und ihrem ET Vorgänger [103], (ii) Vergleiche mit den Resultaten numerischer Verfahren und (iii) die Abhängigkeit der Methode von ihren Parametern diskutiert und erläutert.

[2] Mit freundlicher Genehmigung des Verlags. Kostenlos einsehbar unter DOI: 10.1063/1.4867280

Supplementary material to:

Electrostatics of proteins in dielectric solvent continua. I. An accurate and efficient reaction field description

Sebastian Bauer, Gerald Mathias, and Paul Tavan[a)]

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

[a)]Electronic mail: paul.tavan@physik.uni-muenchen.de

S1. THE FIELD-EXPOSED BORN ION: CONVENTIONAL TREATMENT

Here we solve the PE (1) for a field-exposed Born ion, which is analytically treated within the ET theory in Sec. II B, in the conventional setting by imposing the usual boundary conditions at the surface of the spherical cavity $\mathfrak{v}$. We assume that the homogeneous external field $\mathbf{E}^{\text{ext}} = E_0\hat{\mathbf{e}}_z$ points into the z-direction. Thus, the problem has a cylindrical symmetry implying that the potential $\Phi(\mathbf{r})$ should depend in spherical coordinates solely on r and θ but not on φ, i.e. $\Phi(\mathbf{r}) = \Phi(r,\theta)$.

If one inserts the definitions (3) of the dielectric function $\varepsilon(\mathbf{r})$ and (42) of the characteristic function $\vartheta(\mathbf{r}) = \Theta(\mathbf{r})$ into the PE (1) one sees that the potential $\Phi(\mathbf{r})$ has to obey the PE $\Delta\Phi(\mathbf{r}) = -(4\pi/\varepsilon_s)q\delta(\mathbf{r})$ for $\mathbf{r} \in \mathfrak{v}$. Thus, $\Phi(r,\theta)$ can be expressed for $r \leq R$ as the expansion[2]

$$\Phi_{\text{in}}(r,\theta) = \frac{q}{\varepsilon_s r} + \sum_{n=0}^{\infty} A_n r^n P_n(\cos\theta), \tag{S91}$$

where the functions $P_n(\cos\theta)$ are the Legendre polynomials. For $r > R$ the potential $\Phi(\mathbf{r})$ can be expanded as[3]

$$\Phi_{\text{out}}(r,\theta) = \sum_{n=0}^{\infty} \left[B_n r^n + C_n r^{-(n+1)}\right] P_n(\cos\theta). \tag{S92}$$

At the surface of $\mathfrak{v}$, i.e. at $r = R$, the potential has to be continuous, i.e.

$$\Phi_{\text{in}}\big|_{r=R} = \Phi_{\text{out}}\big|_{r=R}. \tag{S93}$$

Its derivatives, i.e. the fields, have to be continuous

$$\frac{\partial\Phi_{\text{in}}}{\partial\theta}\Big|_{r=R} = \frac{\partial\Phi_{\text{out}}}{\partial\theta}\Big|_{r=R} \tag{S94}$$

for the component tangential to the surface of the spherical region $\mathfrak{v}$ and discontinuous according to

$$\varepsilon_s\frac{\partial\Phi_{\text{in}}}{\partial r}\Big|_{r=R} = \varepsilon_c\frac{\partial\Phi_{\text{out}}}{\partial r}\Big|_{r=R} \tag{S95}$$

for the normal component. Furthermore, $\Phi_{\text{ext}}(r,\theta)$ has to obey the asymptotic boundary condition $\Phi_{\text{ext}}(\mathbf{r}) = -\mathbf{r}\cdot\mathbf{E}^{\text{ext}}$, i.e. $\Phi_{\text{ext}}(r,\theta) = -rE_0\cos\theta$ for $r \to \infty$.

Eqs. (S93), (S94) and (S95) have to be fulfilled for all values of θ and must therefore separately hold for each component of the expansions (S91) and (S92). Solving the resulting

system of linear equations and using the definition (30) of the shielding charge $\hat{q}$ yields the final expression

$$\Phi(\mathbf{r}) = -\mathbf{E}^{\text{ext}} \cdot \mathbf{r} + \begin{cases} \dfrac{q}{\varepsilon_s r} + \dfrac{\hat{q}}{\varepsilon_s R} - \dfrac{1-\varepsilon_s/\varepsilon_c}{2+\varepsilon_s/\varepsilon_c}\,\mathbf{E}^{\text{ext}} \cdot \mathbf{r}, & \text{for } r \le R \\ \dfrac{q}{\varepsilon_c r} - \dfrac{R^3}{r^3}\dfrac{1-\varepsilon_s/\varepsilon_c}{2+\varepsilon_s/\varepsilon_c}\,\mathbf{E}^{\text{ext}} \cdot \mathbf{r}, & \text{else} \end{cases} \tag{S96}$$

for the potential of the Born ion in a homogeneous external field. Note that this expression has been previously derived in the framework of the ET theory (see Appendix B2 of Ref. 1).

We now demonstrate that the potential (S96) is a superposition of the potentials $\Phi^{\mathcal{C}}(\mathbf{r})$ of the empty field exposed cavity and $\Phi^{\mathcal{B}}(\mathbf{r})$ of the isolated Born ion. $\Phi^{\mathcal{B}}(\mathbf{r})$ has been calculated by Born[4] and is given by

$$\Phi^{\mathcal{B}}(\mathbf{r}) = \begin{cases} \dfrac{q}{\varepsilon_s r} + \dfrac{\hat{q}}{\varepsilon_s R}, & \text{for } r \le R \\ \dfrac{q}{\varepsilon_c r}, & \text{else.} \end{cases} \tag{S97}$$

Within $\mathfrak{v}$ this is the Coulomb potential of the central point charge q weakly shielded by ε_s, which is superimposed by the likewise weakly shielded constant potential $\hat{q}/\varepsilon_s R$. In the surroundings of $\mathfrak{v}$, $\Phi^{\mathcal{B}}(\mathbf{r})$ is the Coulomb potential of the central point charge strongly shielded by ε_c.

The calculation of $\Phi^{\mathcal{C}}(\mathbf{r})$, on the other hand, is one of the examples in the textbook of Jackson[3] yielding

$$\Phi^{\mathcal{C}}(\mathbf{r}) = -\mathbf{r} \cdot \mathbf{E}^{\text{ext}} - \begin{cases} \dfrac{1-\varepsilon_s/\varepsilon_c}{2+\varepsilon_s/\varepsilon_c}\,\mathbf{E}^{\text{ext}} \cdot \mathbf{r}, & \text{for } r \le R, \\ \dfrac{R^3}{r^3}\dfrac{1-\varepsilon_s/\varepsilon_c}{2+\varepsilon_s/\varepsilon_c}\,\mathbf{E}^{\text{ext}} \cdot \mathbf{r}, & \text{else.} \end{cases} \tag{S98}$$

$\Phi^{\mathcal{C}}(\mathbf{r})$ consists everywhere of the potential $-\mathbf{r} \cdot \mathbf{E}^{\text{ext}}$ of $\mathbf{E}^{\text{ext}}$. Inside $\mathfrak{v}$ this field is simply elevated, whereas its radial component suddenly drops at the surface of $\mathfrak{v}$ and rapidly ($\sim r^{-2}$) approaches the limiting field $\mathbf{E}^{\text{ext}}$ as $r \to \infty$.

A visual comparison of Eqs. (S98) and (S97) with Eq. (S96) now immediately shows that the potential (S96) of the field exposed Born ion is the superposition $\Phi^{\mathcal{B}}(\mathbf{r}) + \Phi^{\mathcal{C}}(\mathbf{r})$. Therefore we denote the case of the field-exposed Born ion from now on as $\mathcal{B} \cup \mathcal{C}$ and its potential correspondingly as $\Phi^{\mathcal{B}\cup\mathcal{C}}(\mathbf{r})$.

Note that the conditions under which the superposition principle also holds in continuum electrostatics immediately follow from the PE (1). If two such electrostatics problems, differing e.g. by the charge distributions or polarizing fields, refer to identical characteristic functions $\Theta(\mathbf{r})$, i.e. if the low-dielectric volumes $\mathfrak{V}_s$ of the problems to be superimposed are identical, then the dielectric function is invariant, the differential operator $\nabla\cdot\varepsilon(\mathbf{r})\nabla$ is linear, and the potentials obey the superposition principle.

S2. DERIVATION OF RF POLARIZABILITIES

The atomic polarizabilities α_i, which are defined by Eq. (79) for our RF approach, are identical to the ones given by ST.[5] For a proof, which has been omitted in Ref. 5, we consider the approximate counterpart to Eq. (41), according to which the atomic RF dipoles are proportional to the atomic average fields $\langle\mathbf{E}\rangle_{\mathfrak{v}_i}$. With the approximate field $\tilde{\mathbf{E}}(\mathbf{r}) = -\nabla\tilde{\Phi}(\mathbf{r})$, which is the sum $\tilde{\Phi}(\mathbf{r}) = \Phi^{\mathrm{C}}(\mathbf{r}) + \tilde{\Phi}^{\mathrm{RF}}(\mathbf{r})$ of Coulombic and RF contributions, one gets

$$\tilde{\mathbf{p}}_i = \frac{v_i\varepsilon_s}{4\pi}\left(1-\frac{\varepsilon_s}{\varepsilon_c}\right)\left\langle\nabla\left[\Phi^{\mathrm{C}}(\mathbf{r}) + \tilde{\Phi}^{\mathrm{RF}}(\mathbf{r})\right]\right\rangle_{\sigma_i}. \tag{S99}$$

According to Eqs. (33) and (74), the potentials $\Phi^{\mathrm{C}}(\mathbf{r})$ and $\tilde{\Phi}^{\mathrm{RF}}(\mathbf{r})$ are sums over contributions $\phi(\mathbf{r}\,|\,\mathbf{r}_j, q_j)$, $\phi(\mathbf{r}\,|\,\mathbf{r}_j, \hat{q}_j, \hat{\sigma}_j)$, and $\phi(\mathbf{r}\,|\,\mathbf{r}_j, \tilde{\mathbf{p}}_j, \sigma_j)$ originating from all atoms j. While the local volume averages $\langle\ldots\rangle_{\sigma_i}$ [cf. Eq. (76)] of the radial fields generated by the point charges q_j and the Gaussian shielding charges $\hat{q}_j$ vanish for $j = i$, the Gaussian anti-polarization densities (18) with the strengths $\tilde{\mathbf{p}}_i$ generally contribute the non-vanishing local field averages[1]

$$\langle-\nabla\phi(\mathbf{r}\,|\,\mathbf{r}_i, \tilde{\mathbf{p}}_i, \sigma_i)\rangle_{\sigma_i} = -\sqrt{\frac{2}{\pi}}\frac{1}{3\sigma_{ii}}\tilde{\mathbf{p}}_i, \tag{S100}$$

where $\sigma_{ii} = \sqrt{2}\sigma_i$. With the definition (78) of the polarizing field $\langle\tilde{\mathbf{E}}(\mathbf{r})\rangle_{\sigma_i}$, which solely contains contributions $j \neq i$, Eq. (S99) thus becomes

$$\tilde{\mathbf{p}}_i = \frac{v_i\varepsilon_s}{4\pi}\left(1-\frac{\varepsilon_s}{\varepsilon_c}\right)\langle\tilde{\mathbf{E}}(\mathbf{r}_i)\rangle_{\sigma_i} + \frac{v_i}{12(\sqrt{\pi}\sigma_i)^3}\left(1-\frac{\varepsilon_s}{\varepsilon_c}\right)\tilde{\mathbf{p}}_i. \tag{S101}$$

Solving this equation for $\tilde{\mathbf{p}}_i$ and using the definitions (80), (81) and (82) with $\gamma = 1$ finally yields the self-consistency condition (77) with the polarizability (79).

S3. GRANULAR KIRKWOOD SPHERE

A. Scaled characteristic function

If the characteristic function $\Theta(\mathbf{r})$, which defines according to Eq. (4) the space $\mathfrak{V}_s$ of filled by a protein, is approximated according to Eq. (19) as a superposition $\tilde{\Theta}(\mathbf{r})$ of Gaussian atoms (17) and if the volumes $\tilde{v}_i$ are iteratively determined by Eq. (21), then the correct value $\Theta(\mathbf{r}) = 1$ ($\mathbf{r} \in \mathfrak{V}_s$) is ensured by $\tilde{\Theta}(\mathbf{r})$ only at the positions $\mathbf{r} = \mathbf{r}_i$ of the atoms. In the case of the coarse grained model $\tilde{\mathcal{K}}$ of the Kirkwood sphere $\mathcal{K}$ introduced in Sec. IV A this leads, on average, to an underestimate of $\Theta^{\mathcal{K}}(\mathbf{r})$ by $\tilde{\Theta}^{\mathcal{K}}(\mathbf{r})$. This underestimate had been apparent already in Figure 3 of Ref. 1, which compared the exact step function $\Theta^{\mathcal{K}}(\mathbf{r})$ in a section along the x-axis of the hexagonal grid with the Gaussian superposition $\tilde{\Theta}^{\mathcal{K}}(\mathbf{r})$.

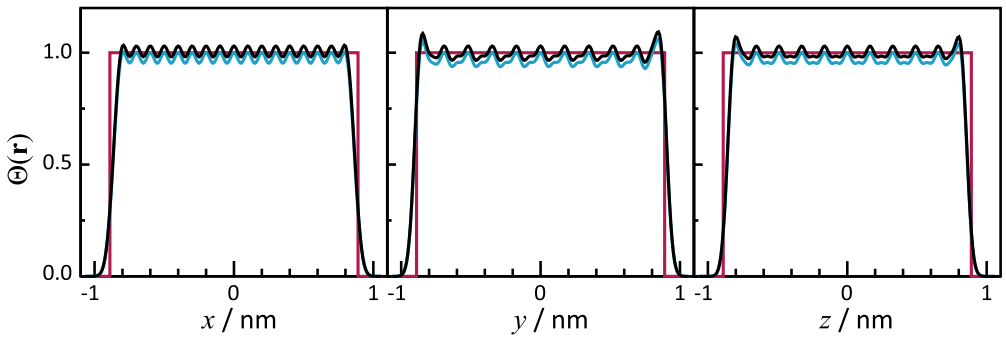

FIG. S13: The characteristic function $\Theta^{\mathcal{K}}(\mathbf{r})$ (red line) of an exact Kirkwood sphere $\mathcal{K}$ of radius $R^{\mathcal{K}} = 0.89$ nm is compared along the three Cartesian axes with the unscaled and scaled approximations $\tilde{\Theta}^{\mathcal{K}}(\mathbf{r})$ (ET, blue) and $\gamma\tilde{\Theta}^{\mathcal{K}}(\mathbf{r})$ (RF, $\gamma = 1.03$, black).

By displaying $\Theta^{\mathcal{K}}(x)$ (red) and $\tilde{\Theta}^{\mathcal{K}}(x)$ (blue) the left graph in Fig. S13 repeats the quoted comparison. In addition the graph shows an upscaled version $\gamma\tilde{\Theta}^{\mathcal{K}}(x)$ (black) of the approximate characteristic function with the choice $\gamma = 1.03$. Analogous data are shown in the center and right graphs of the figure for cuts of the various functions along y and z axes, respectively.

Quite apparently, the indicated choice of γ shifts the scaled function $\gamma\tilde{\Theta}^{\mathcal{K}}(\mathbf{r})$ for all three cuts in such a way to larger values that the average underestimate of $\Theta^{\mathcal{K}}(\mathbf{r})$ disappears. On average the scaled model $\gamma\tilde{\Theta}^{\mathcal{K}}(\mathbf{r})$ is very close to $\Theta^{\mathcal{K}}(\mathbf{r})$. We would like to stress at this point that our choice is by no means unique and that similar choices would have almost the same effects. There is necessarily a certain degree of arbitrariness associated with any choice of γ, because there is no strict measure, which enables a comparison of the fuzzy and crisp volume descriptions $\tilde{\Theta}^{\mathcal{K}}(\mathbf{r})$ and $\Theta^{\mathcal{K}}(\mathbf{r})$, respectively.

On the other hand, the case of the Kirkwood sphere and of its coarse grained Gaussian superposition model demonstrates that the volume model $\tilde{\Theta}(\mathbf{r})$ originally suggested in Ref. 1 most likely represents for all densely packed macromolecules an underestimate. Furthermore it enables an estimate for its approximate correction. Thus, the parameter γ is from now on fixed to the value 1.03.

B. RF and ET dipoles

Figure S14 compares the RF dipoles $\tilde{\mathbf{p}}_i$, which are assigned by the new RF and original ET approaches to the Gaussian atoms (green circles) of the coarse grained Kirkwood sphere $\tilde{\mathcal{K}}$ and are induced by a single charged atom k (yellow dot). The associated shielding charge distribution $\tilde{\rho}_k(\mathbf{r})$ of the new RF approach is indicated by the filled blue circle and the shapes of the exact spheres $\mathcal{K}$ by the large gray circles. For the drawings the dipole sizes were logarithmically scaled according to the prescription $c \ln(|\tilde{\mathbf{p}}_i|/|\tilde{\mathbf{p}}_{\min}|)$, with c chosen such that the largest dipole just fits in one of the atomic circles. Note that the largest RF dipole of the ET approach is by about a factor of three larger than the largest RF dipole of the revised method, although the largest dipoles of the two results are drawn at equal sizes.

Fig. S14 demonstrates that the $|\tilde{\mathbf{p}}_i|$ assigned by the revised method to the uncharged atoms surrounding the charge q_k decrease much more rapidly with increasing distance than

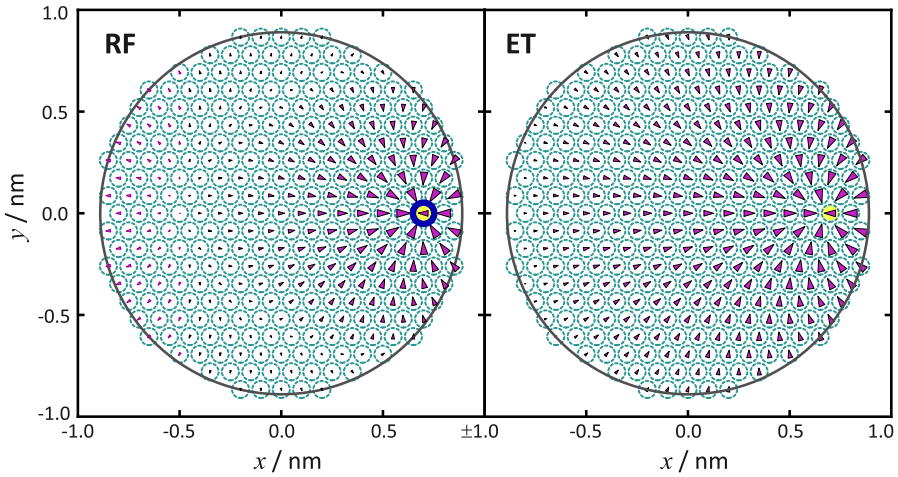

FIG. S14: The RF dipoles $\tilde{\mathbf{p}}_i$ (magenta arrows) of the Kirkwood sphere $\tilde{\mathcal{K}}$ (cf. Fig. 6 in the main text) in a cross section covering the x-y plane as predicted by the RF and ET approaches, respectively. The inducing charge q_k (yellow dot) is located on the x axis at $x = 0.7$ nm. In the RF setting it is surrounded by a Gaussian shielding charge distribution $\tilde{\rho}_k(\mathbf{r})$ (filled blue circle).

those calculated by the original ET approach. While all ET dipoles within $\tilde{\mathcal{K}}$ point toward q_k, a reversal of dipole orientations can be detect in the graph labeled by RF beyond a certain distance ($\approx 1\,\mathrm{nm}$). Hence, the revised theory renders even qualitatively different dipole densities.

C. Numerical solution

Figure S15 compares our RF solution for the coarse grained model $\tilde{\mathcal{K}}$ of the Kirkwood sphere $\mathcal{K}$ (radius $R^{\mathcal{K}} = 0.89$ nm), which is embedded in a dielectric continuum $\varepsilon_c = 80$ and is introduced in Sec. IV A, with a numerical solution for a similar model $\hat{\mathcal{K}}$ of $\mathcal{K}$, in which the Gaussian atoms (widths $\sigma = 0.0421\,\mathrm{nm}$) occupying the hexagonal lattice are replaced by spherical cavities. Their radii $r_c = [3/\sqrt{2\pi}]^{1/3}\sigma$ are chosen such that the volumes of the Gaussian and spherical atoms are identical. The associated diameters $2r_c = 0.13\,\mathrm{nm}$ are larger than the lattice constant of 0.1 nm such that neighboring spheres considerably overlap.

The numerical solution was calculated with DELPHI[6,7] using a cubic grid with 301^3 points and a hierarchical focusing spanning the lattice constants 0.390 nm, 0.078 nm, 0.020 nm, 0.007 nm. Thus, the largest grid had an inner radius of 58.5 nm and the smallest grid an inner radius of 0.976 nm. In addition, the exact Kirkwood solution[2] is indicated by the black dashed line in Fig. S15.

As is apparently generic for numerical solutions, the solvation energy $|W^{\mathrm{RF},\mathcal{K}}(0)|$ of the

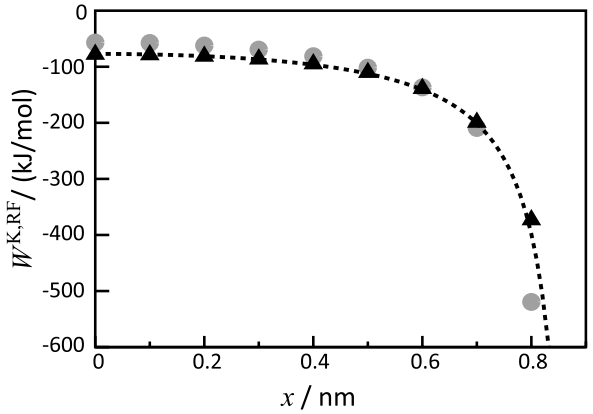

FIG. S15: Exact (dashed) RF energy $W^{\mathrm{RF},\mathcal{K}}(x)$ of a unit charge in a Kirkwood sphere $\mathcal{K}$ of radius $R^{\mathcal{K}} = 0.89$ nm at a distance x from its center. The black triangles represent our RF result for a coarse grained model $\tilde{\mathcal{K}}$ of $\mathcal{K}$ made up from a hexagonal lattice of Gaussian spheres. The gray dots indicate the numerical solution for the coarse grained model $\hat{\mathcal{K}}$, in which the lattice is occupied by spherical cavities.

associated large Born ion of radius $R^{\mathcal{K}} = 0.89$ nm is underestimated by the numerical approach. For increasing shifts of the unit charge along the x-axis, the numerical treatment (gray dots) predicts a steeper decrease of the RF energy $W^{\mathrm{RF},\mathcal{K}}(x)$ than the exact (dashed line) or our approximate treatment (black triangles).

S4. ION–CAVITY INTERACTIONS FOR DIFFERENT RADII

In the main text we have demonstrated the quality of our RF approximation for the basic example $\mathcal{BC}$, i.e. the interaction of an ion with a neutral atom of the same radius $R = 0.17\,\mathrm{nm}$. Figure S16 checks the performance of the RF approximation for objects of different size. It shows $\tilde{W}^{\mathrm{RF},\mathcal{BC}}(r)$ calculated for the same ion and cavities with either smaller radii (Fig. S16a) of $0.8R$ (black solid) and $0.9R$ (red solid) or larger radii (Fig. S16b) of $1.1R$ (green solid) and $1.2R$ (blue solid). The corresponding DELPHI results are marked by the same colors but are distinguished by the use of dashed lines. The DELPHI setup is described in the main text.

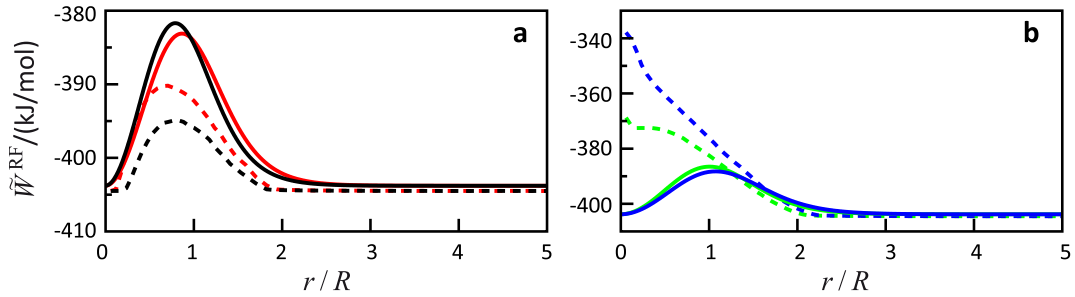

FIG. S16: Distance-dependent ion–cavity RF energies for differently sized spherical cavities. The ion has the radius $R = 0.17\,\mathrm{nm}$. The radii of the cavities were chosen either smaller (a), i.e. $0.9R$ (red) and $0.8R$ (black), or larger (b), i.e. $1.1R$ (green) and $1.2R$ (blue). The DELPHI results for strict spheres are marked by dashed lines, our RF approximation for Gaussian spheres by solid lines.

If smaller cavities interact with the ion (Fig. S16a), our RF approach qualitatively reproduces the DELPHI results. DELPHI, as always, underestimates the Born solvation energy ($r \to \infty$), which is correct in RF. With RF, the energy barriers are still predicted at the correct (DELPHI) positions, but their heights are substantially overestimated. Decreasing the cavity radius from $0.9R$ (red) to $0.8R$ (black) leads according to RF (solid) to an increasing barrier height, whereas DELPHI predicts the opposite behavior. No such inversion is found at distances $r/R > 1$, i.e. beyond the positions of the barriers. Here, the RF curves (solid) semi-quantitatively show the same behavior as the DELPHI curves.

Also if larger cavities interact with the ion, the RF energies closely approximate the DELHPI results at distances $r/R > 1$. Disregarding the small DELPHI error at $r \to \infty$, the RF results essentially reproduce those of DELPHI in the indicated range. Again, the absolute height of the barrier is predicted by RF to increase for a decreasing cavity radius, whereas DELPHI predicts the opposite behavior at distances $r/R < 1$. Here the larger cavity swallows the ion and we end up with a different situation, i.e. with a single ion of increased size. According to Eq. (68) it has a smaller solvation energy $|W^{\mathrm{RF}}|$ than the swallowed ion (cf. Fig. S16b for $r \to 0$).

Our RF approximation does not properly account for this situation, because the width $\hat{\sigma} = \sqrt{2/\pi}R$ of the Gaussian shielding charge distribution $\tilde{\rho}(\mathbf{r}\,|\,\mathbf{r} = 0, \hat{q}, \hat{\sigma})$ belongs to the original ion and, therefore, keeps its small size, although it is now embedded in a large cavity. One could, of course, get rid of this drawback by the use of widths $\hat{\sigma}$, which change upon swallowing events.

However, for molecular structures the atomic radii can always be chosen in such a way that the volume is properly described and swallowing is excluded. Swallowing could only occur for covalently bound atoms with strongly different radii $R_i > R_j$ and small bond length $r_{ij} < R_i$. Such cases are avoided, however, if the chosen atomic RF radii R_i obey $\max(R_i, R_j) < r_{ij}$. Then, distances as small as $r/R = 1$ in Fig. S16 represent the worst case scenario, while the usual cases (see Table I for typical radii) are found at distances $r/R > 1.5$, where our RF approximation performs very well.

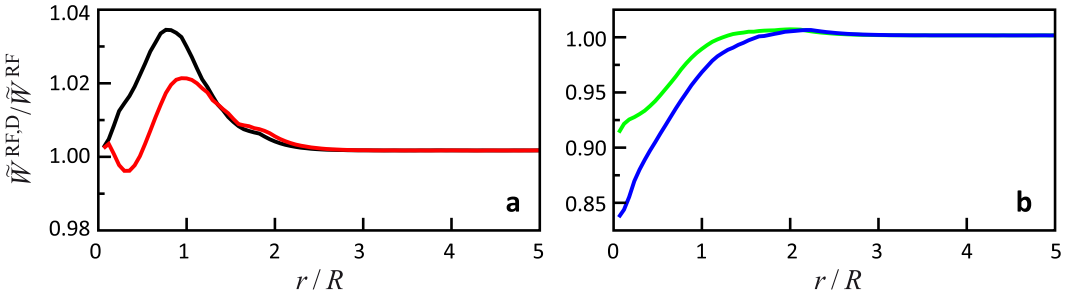

FIG. S17: Relative deviation $\tilde{W}^{\mathrm{RF,D}}(r)/\tilde{W}^{\mathrm{RF}}(r)$ of the DELPHI and RF results for a smaller (a), i.e. $0.8R$ (black) and $0.9R$ (red), and a larger (b), i.e. $1.1R$ (green) and $1.2R$ (blue), cavity interacting with an ion of radius R.

The quality of the RF approach is underlined by the relative deviation $\tilde{W}^{\mathrm{RF,D}}(r)/\tilde{W}^{\mathrm{RF}}(r)$ of the DELPHI and RF results shown in Figure S17. At the generic distances $r/R > 1.5$ the deviation is smaller than 2%.

As demonstrated by Figure S18, which shows the energy differences $\tilde{W}^{\mathrm{RF}}(r) - \tilde{W}^{\mathrm{RF,D}}(r)$,

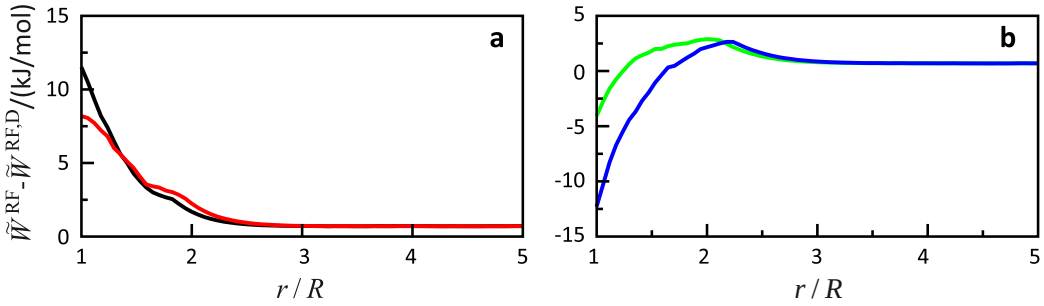

FIG. S18: Difference $\tilde{W}^{\mathrm{RF}}(r) - \tilde{W}^{\mathrm{RF,D}}(r)$ of the RF and DELPHI results for a smaller (a), i.e. $0.8R$ (black) and $0.9R$ (red), and a larger (b), i.e. $1.1R$ (green) and $1.2R$ (blue), cavity interacting with an ion of radius R.

the absolute deviations are smaller than 5 kJ/mol for $r > 1.5R$.

S5. ION–ION INTERACTIONS

In standard MM force fields essentially all atoms carry partial charges. Therefore, the interaction of two ions $\mathcal{B} \cup \mathcal{C}$, which are exposed to their mutually polarizing fields, is the most frequently encountered case in such systems. To illustrate our RF description and its quality for such cases, we have calculated the changes

$$\Delta\tilde{W}^{\mathrm{RF}}(r \,|\, e, \pm e) \equiv \tilde{W}^{\mathrm{RF}}(r \,|\, e, \pm e) - 2W^{\mathrm{RF},\mathcal{B}} \tag{S102}$$

of the approximate RF energies $\tilde{W}^{\mathrm{RF}}(r \,|\, e, \pm e)$ and of their DelPhi counterparts as functions of the center-to-center distance r. Figure S19 shows the results for oppositely (black) and identically (gray) charged ions, respectively.

According to Fig. S19, at large distances ($r/R \gg 1$) the RF energy

$$\Delta\tilde{W}^{\mathrm{RF}}(r \,|\, e, \pm e) \approx -\left(1 - \frac{\varepsilon_s}{\varepsilon_c}\right) \frac{1}{\varepsilon_s} \left(\mp \frac{e^2}{r}\right)$$

is repulsive (upper sign) for opposite charges (black curves) and attractive (lower sign) for identical charges (gray curves). Within the accuracy of the drawing our RF energies (solid lines) are almost indistinguishable from the DelPhi results (dashed lines) in the relevant distance range $r/R > 1$ (Ref. 8 presents a related comparison for GB/CFA). But also at smaller distances the two descriptions are quite similar.

It is important to note that the RF energies calculated by our RF approach and by DelPhi are not exactly symmetric for the two cases, i.e. the black and gray curves in Fig. S19 are

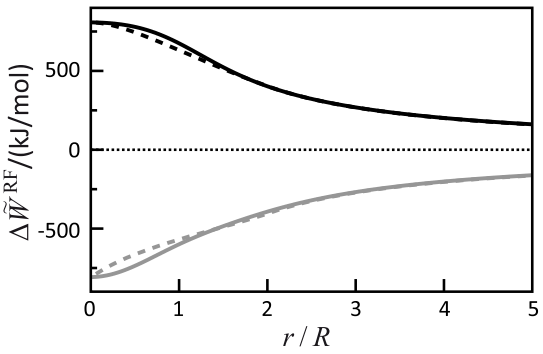

FIG. S19: Distance dependent RF energy changes $\Delta\tilde{W}^{\mathrm{RF}}(r\,|\,e,\pm e)$ as defined by Eq. (S102) and as calculated by our RF approach (solid) and by DelPhi (dashed), respectively, for two oppositely $[(e,-e)$, black] and two equally $[(e,e)$, gray] charged ions of identical size $R = 0.17\,\mathrm{nm}$.

not exact mirror images of each other with respect to the dotted line. The deviations from mirror symmetry can be measured by the sum

$$\Delta W^{\mathrm{RF,S}}(r) \equiv \Delta W^{\mathrm{RF}}(r\,|\,e,e) + \Delta W^{\mathrm{RF}}(r\,|\,e,-e), \tag{S103}$$

of the changes $\Delta W^{\mathrm{RF}}(r\,|\,e,e)$ and $\Delta W^{\mathrm{RF}}(r\,|\,e,-e)$ for two equally and two oppositely charged charged ions interacting at the distance r.

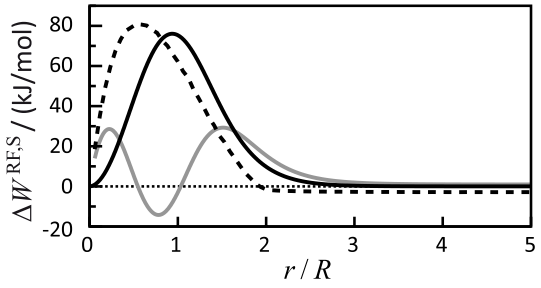

FIG. S20: Deviations $\Delta W^{\mathrm{RF,S}}(r)$ from mirror symmetry of the RF energies shown in Figure S19: RF approach (black solid), DelPhi (black dashed). As a reference, the deviation from symmetry obtained with the ET approach (gray solid) is also given.

Figure S20 compares the deviations $\Delta W^{\mathrm{RF,S}}(r)$ from mirror symmetry obtained by our RF approach (black solid), by DelPhi (black dashed), and by the ET approach (gray solid), which demonstrate that the cases of equally and oppositely charged ions are non-symmetric with respect to changing the sign of one of the charges.

A comparison with Fig. 9 immediately reveals that for all three approaches the deviations $\Delta W^{\mathrm{RF,S}}(r)$ are simply four times the RF energies $\Delta\tilde{W}^{\mathrm{RF},\mathcal{BC}}(r)$, which result from merging an ion $\mathcal{B}$ with a cavity $\mathcal{C}$ and are depicted in Fig. 9. An inspection of the numbers demonstrates

that the equality $\Delta W^{\mathrm{RF,S}}(r) = 4\Delta\tilde{W}^{\mathrm{RF},\mathcal{BC}}(r)$ holds to numerical accuracy: Consequently, the merging of the four ions appearing on the r.h.s. of Eq. (S103) with associated cavities upon reduction of r is the only source of the symmetry breaking. Each of these processes of merging the volume of an ion with an equally sized cavity adds the same RF energy $\Delta\tilde{W}^{\mathrm{RF},\mathcal{BC}}(r)$.

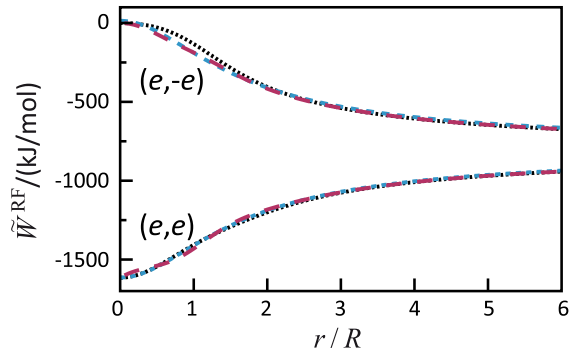

FIG. S21: RF energies $\tilde{W}^{\mathrm{RF}}(r\,|\,e,\pm e)$ calculated by our RF approach (black), by ET (red), and by GB (blue) respectively, for two oppositely $(e,-e)$ and two equally (e,e) charged ions of identical size $R = 0.17\,\mathrm{nm}$ interacting at the distance r.

Figure S21 now documents the RF energies $\tilde{W}^{\mathrm{RF}}(r\,|\,e,\pm e)$ obtained by the ET approach (red) underlying the gray curve in Fig. S20. For reference our RF results already shown in Fig. S19 are repeated (black). One clearly recognizes identical descriptions at $r/R > 1.5$ and differences for smaller r. For the interested reader we have added the GB result (blue) to Fig. S21, which shows no deviation from mirror symmetry for $r/R < 2$.

S6. PARAMETERS FOR THE DELPHI DESCRIPTION OF ALDI

Figure S22 depicts the chemical structure and the CHARMM22 atom types of AlDi, which were used in the free energy calculation with DELPHI. The necessary atomic radii were adapted from the CHARMM22 force field as described in Ref. 5 and are summarized in table S2.

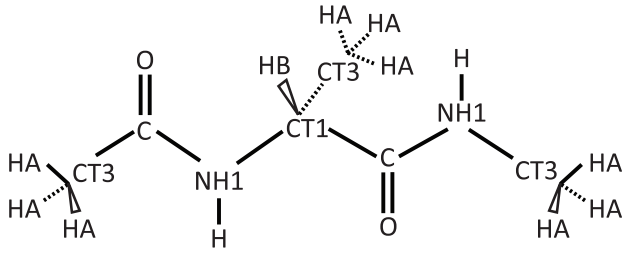

FIG. S22: Chemical structure of the model dipeptide Ace-Ala-NHMe with atoms i labeled by their CHARMM22[9] atom types $\tau(i)$.

atom type τ	C	CT1	CT3	H	HA/HB	NH1	O
R_τ [Å]	2.01	2.29	2.08	0.23	1.33	1.86	1.70

TABLE S2: Atomic radii R_τ employed in the DELPHI calculation of $\Delta G(\varphi, \psi)$ for AlDi.

S7. FREE ENERGY LANDSCAPES $\Delta G[\mathbf{R}(\varphi, \psi)]$

To illustrate the dependence of our RF approach on the choice of the parameters, i.e. the Gaussian widths σ_X of the various atom types, we have further simplified the already simple choice specified by Table I into the one given by Table S3, which distinguishes only heavy (C, N, O) from small (H) atoms. The value of σ_X chosen for the heavy atoms is the frequency-weighted average of the values given in Table I, where "frequency" means the frequency of occurrence in the structure of AlDi.

X	C, N, O	H
σ_X [Å]	0.82	0.52

TABLE S3: Atom types X and Gaussian widths σ_X employed in an alternative RF calculation of $\Delta G(\varphi, \psi)$ for AlDi.

With the most simple parameter set of Table S3 we calculated the free energy landscape $\Delta G[\mathbf{R}(\varphi, \psi)]$. Figure S23 compares the result, which is labeled as RF′, with the energy

landscape RF presented already in Fig. 11 of the main text and calculated with the somewhat more sophisticated choice of Table I.

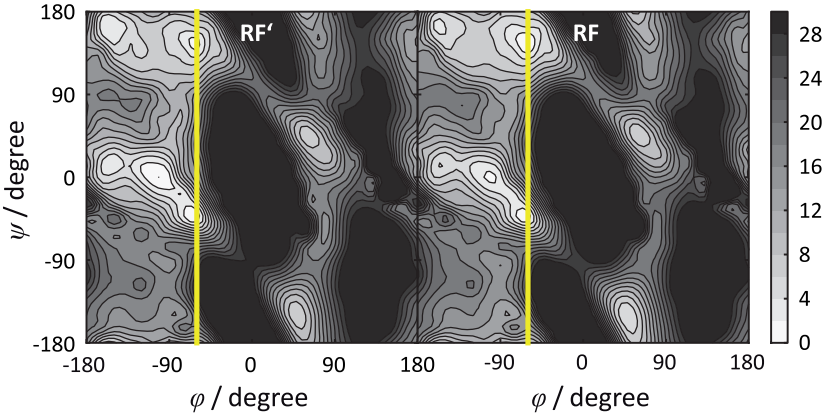

FIG. S23: Free energy landscapes $\Delta G[\varphi, \psi)]$ of AlDi calculated by our RF approach. RF′: most simple parameters of Table S3, RF: simple parameters of Table I. The yellow lines indicate the locations of the one-dimensional cuts $\Delta G[-60°, \psi)]$ displayed by Fig. S24.

Quite obviously the overall structure of the free energy landscape is invariant under the change of the parameters. Only a very careful inspection indicates minor changes. For a more quantitative insight into these changes we have extracted the cross-sections $\Delta G[\mathbf{R}(-60°, \psi)]$ through the two landscapes as indicated by the yellow lines in Fig. S23.

For better comparability we have shifted all cross-sections (see Fig. S24) such that $\Delta G[\mathbf{R}(-60°, -42°)] = 0$, i.e. that the location of the α-helical minimum marks in all three cases the zero of the energy scale. Correspondingly, we had to shift the DELPHI cross-section relative to the RF reference by -0.2 kJ/mol and the alternative RF′ cross-section by -24.1 kJ/mol to lower energies. The close match of the DELPHI and RF absolute free energies $G[\mathbf{R}(-60°, -42°)]$ at the α-helical minimum results from the match of the DELPHI and RF Born energies $G[\mathbf{R}(-66°, -42°)]$ achieved by the global scaling of the RF Gaussian widths described in Sec. IV C. The -24.1 kJ/mol difference between the Born energies of RF and RF′ at the AlDi configuration $\mathbf{R}(-60°, -42°)$ thus indicates that the effective volume of AlDi in the RF′ model is slightly overestimated as compared to the RF model.

Comparing now the cross-sections $\Delta G[\mathbf{R}(-60°, \psi)]$ of RF (black), simplified RF′ (gray), and DELPHI (black dashed) one sees that the minima and maxima are localized at the same angles ψ by all three methods, that the energies of the second minimum at $(-60°, 144)$ are predicted within ± 1.1 kJ/mol, and that the maxima are calculated with a larger spread of ± 4 kJ/mol. A glance at Fig. S23 demonstrates that the maxima are

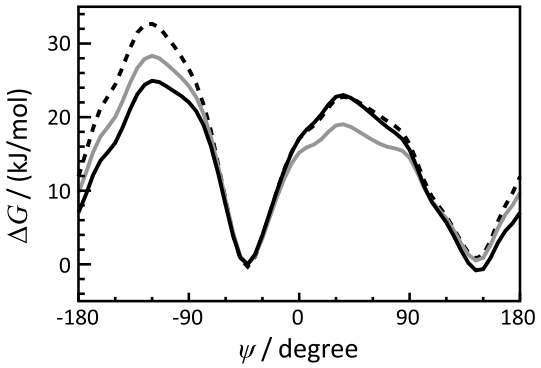

FIG. S24: Cross-sections $\Delta G[\mathbf{R}(-60°, \psi)]$ through the free energy landscape of AlDi. Black: RF with the simple parameters of Table I, gray: RF with the most simple parameters of Table S3, black dashed: DELPHI.

in regions of $\Delta G[\mathbf{R}(\varphi, \psi)]$, in which the free energy landscape features large gradients $|\partial \Delta G[\mathbf{R}(\varphi, \psi)]/\partial \varphi|$. Thus, small errors are magnified here.

In summary, the above study has demonstrated that the locations of the maxima and minima, the relative energies, and the sizes of local gradients are quite insensitive to the choice of the Gaussian widths σ_X and that solely the overall energetic location G_0 of the free-energy surface, which can be tuned by an overall scaling of the σ_X, is plagued by considerable uncertainties. To what extent these uncertainties can be minimized by more sophisticated choices of the Gaussian widths σ_X, which then should be calibrated using large sets of sample molecules with known solvation energies, remains to be seen. It must be stressed, however, that all alternative conformations of a given molecule, i.e. the conformational dependence of $\Delta G(\mathbf{R})$, should be accurately predicted by our RF approach once the Born solvation energy $G_0 \equiv G(\mathbf{R}_0)$ of a particular configuration $\mathbf{R}_0$ is known.

REFERENCES

[1]B. Egwolf and P. Tavan, J. Chem. Phys. **118**, 2039 (2003).

[2]J. G. Kirkwood, J. Chem. Phys. **2**, 351 (1934).

[3]J. D. Jackson, *Classical Electrodynamics* (John Wiley & Sons, Inc., New York, London, Sydney, Toronto, 1975).

[4]M. Born, Zeitschrift für Physik **1**, 45 (1920).

[5]M. Stork and P. Tavan, J. Chem. Phys. **126**, 165105 (2007).

[6]W. Rocchia, E. Alexov, and B. Honig, J. Phys. Chem. B **105**, 6507 (2001).

[7]A. Nicholls and B. Honig, J. Comp. Chem. **12**, 435 (2004).

[8]D. Borgis, N. Levy, and M. Marchi, J. Chem. Phys. **119**, 3516 (2003).

[9]A. D. MacKerell, 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, J. Phys. Chem. B **102**, 3586 (1998).

2.3 Hamilton'sche Dynamik für MD Simulationen in dielektrischen Kontinua

Der folgende Abdruck[3]

Electrostatics of proteins in dielectric sovent continua.
II. Hamiltonian reaction field dynamics"
Sebastian Bauer, Paul Tavan, and Gerald Mathias
J. Chem. Phys. **140**, 104103 (2014)

erklärt die Ableitung von analytischen Ausdrücken für die auf die Proteinatome wirkenden RF-Kräfte aus der oben abgeleiteten Darstellung der RF-Energie eines Proteins, erläutert die Implementierung der Kraftberechnung in das Simulationsprogramm IPHIGENIE [48, 106], und demonstriert an Simulationen einfacher Beispiele den Hamilton'schen Charakter und die Genauigkeit von HADES-MD.

[3] Mit freundlicher Genehmigung des Verlags.

THE JOURNAL OF CHEMICAL PHYSICS **140**, 104103 (2014)

Electrostatics of proteins in dielectric solvent continua. II. Hamiltonian reaction field dynamics

Sebastian Bauer, Paul Tavan, and Gerald Mathias[a)]
Lehrstuhl für BioMolekulare Optik, Ludig-Maximilians Universität München, Oettingenstr. 67, 80538 München, Germany

(Received 11 October 2013; accepted 1 January 2014; published online 10 March 2014)

In Paper I of this work [S. Bauer, G. Mathias, and P. Tavan, J. Chem. Phys. **140**, 104102 (2014)] we have presented a reaction field (RF) method, which accurately solves the Poisson equation for proteins embedded in dielectric solvent continua at a computational effort comparable to that of polarizable molecular mechanics (MM) force fields. Building upon these results, here we suggest a method for linearly scaling Hamiltonian RF/MM molecular dynamics (MD) simulations, which we call "Hamiltonian dielectric solvent" (HADES). First, we derive analytical expressions for the RF forces acting on the solute atoms. These forces properly account for all those conditions, which have to be self-consistently fulfilled by RF quantities introduced in Paper I. Next we provide details on the implementation, i.e., we show how our RF approach is combined with a fast multipole method and how the self-consistency iterations are accelerated by the use of the so-called direct inversion in the iterative subspace. Finally we demonstrate that the method and its implementation enable Hamiltonian, i.e., energy and momentum conserving HADES-MD, and compare in a sample application on Ac-Ala-NHMe the HADES-MD free energy landscape at 300 K with that obtained in Paper I by scanning of configurations and with one obtained from an explicit solvent simulation. [http://dx.doi.org/10.1063/1.4867281]

I. INTRODUCTION

Atomistic molecular dynamics (MD) simulations of soluble proteins, in which the respective solute-solvent system is described by a standard molecular mechanics (MM) force field such as CHARMM22,[1] AMBER,[2] or GROMOS,[3] have to properly account for the dominant electrostatic interactions.[4–8] In such simulations the solvent is usually[7] represented by simple three-point water potentials.[9,10] For a proper representation of the electrostatic solute-solvent interactions the number of solvent atoms must exceed that of the solute by at least a factor of ten.[8,11,12] Therefore, most of the computational effort must be spent on simulating the thermal motions within the solvent, which limits, particularly for large solutes, the time-scales accessible by such MM-MD simulations.

To reduce this effort so-called implicit solvent models of the generalized Born (GB) type were applied.[13–21] However, the free energy landscapes obtained by GB/MM-MD simulations turned out to be unreliable.[22,23] Furthermore and as explained in more detail in Paper I of this work,[12] GB approaches generally do not solve the dielectric Poisson equation (PE)

$$\nabla \cdot [\varepsilon(\mathbf{r})\nabla\Phi(\mathbf{r})] = -4\pi \sum_i q_i \delta(\mathbf{r}-\mathbf{r}_i) \qquad (1)$$

for the electrostatic potential $\Phi(\mathbf{r})$, which is generated by the protein's partial charges q_i and is the quantity of interest, if one wants to apply a continuum approach to the solvent. Here, the symbol

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

denotes the dielectric function, which assigns the value ε_s to the interior of the protein $\mathcal{P}$ and the usually much larger value ε_c to the solvent continuum by means of the characteristic function $\Theta(\mathbf{r})$, which has the value of 1 inside and 0 outside of $\mathcal{P}$.[12,24] The PE (1) can be numerically solved, of course, through the use of regular grids[25–27] or boundary elements.[28] However, these methods are computationally very expensive[12,29,30] and, moreover, do not directly yield the electrostatic forces acting on the atoms. Therefore, they cannot be employed for MD simulations. A similar critique[30] applies to the free energy functional approach suggested in Ref. 31.

Aiming at the use in MD simulations we have derived in Paper I of this work[12] an alternative route toward solving the PE. Extending earlier attempts by Egwolf and Tavan (ET)[24] as well as Stork and Tavan (ST)[30] this approach employs the natural discretization of the protein volume, which is given by its N atoms i occupying the positions $\mathbf{r}_i$. Through an exact fuzzy decomposition of the characteristic function $\Theta(\mathbf{r})$ into atomic volumes it yields with Eqs. (32)–(35) in Ref. 12 an exact atomistic reaction field (RF) representation of the PE (1) and of its solution. Additionally and following the general concepts of ET and ST accurate approximations are derived by modeling the atoms as Gaussian volumes.

The corresponding approximate characteristic function

$$\tilde{\Theta}(\mathbf{r}) = \sum_i \tilde{\vartheta}_i(\mathbf{r}) \qquad (3)$$

a)Electronic mail: gerald.mathias@physik.uni-muenchen.de

is a superposition of atomic functions

$$\tilde{\vartheta}_i(\mathbf{r}) = \tilde{v}_i G(\mathbf{r} \mid \mathbf{r}_i, \sigma_i), \tag{4}$$

which are given by atomic volumes $\tilde{v}_i$ and by atom-centered Gaussian shape functions

$$G(\mathbf{r} \mid \mathbf{r}_i, \sigma_i) = \frac{1}{v^{\mathrm{G}}(\sigma_i)} \exp\left[-\frac{(\mathbf{r}-\mathbf{r}_i)^2}{2\sigma_i^2}\right] \tag{5}$$

of width σ_i, which are normalized by $v^{\mathrm{G}}(\sigma_i) \equiv (2\pi\sigma_i^2)^{3/2}$. If $\tilde{\Theta}(\mathbf{r})$ is supposed to represent a reasonable approximation to the exact characteristic function $\Theta(\mathbf{r})$, the volumes $\tilde{v}_i$ should be determined such that $\tilde{\Theta}(\mathbf{r}_k) = 1$ at each $\mathbf{r}_k$. Starting from, e.g., $\tilde{v}_k^{(0)} = v^{\mathrm{G}}(\sigma_k)$ these N conditions can be solved by the self-consistency iteration

$$\tilde{v}_k^{(n+1)} = \tilde{v}_k^{(n)}\left[\sum_i \tilde{\vartheta}_i^{(n)}(\mathbf{r}_k)\right]^{-1} \tag{6}$$

that assigns new volumes $\tilde{v}_k^{(n+1)}$ from the volumes $\tilde{v}_k^{(n)}$ of the nth iteration. Convergence can be assumed as soon as the dimensionless measure

$$\delta_v^{(n)} = \max_k \left|1 - \sum_i \tilde{v}_i^{(n)} G(\mathbf{r} \mid \mathbf{r}_i, \sigma_i)\right| \tag{7}$$

drops below a given threshold χ_v.

Based on a meticulous analysis and correction of the previous suggestions of ET and ST this discretization led to the approximate reaction field (RF) representation

$$\Delta\tilde{\Phi}(\mathbf{r}) = -\frac{4\pi}{\varepsilon_s}\sum_i [q_i\delta(\mathbf{r}-\mathbf{r}_i) + \tilde{\rho}_i(\mathbf{r}) - \nabla\cdot\tilde{\mathbf{P}}_i(\mathbf{r})] \tag{8}$$

of the PE (1), in which approximate quantities are characterized by the tildes. Equation (8) distinguishes three different types of atomic quantities as sources of the approximate electrostatic potential $\tilde{\Phi}(\mathbf{r})$. These are the partial charges q_i, which give rise to the superposition

$$\Phi^{\mathrm{C}}(\mathbf{r}) = \frac{1}{\varepsilon_s}\sum_i \phi(\mathbf{r} \mid \mathbf{r}_i, q_i) \tag{9}$$

of the associated Coulomb potentials[32] $\phi(\mathbf{r} \mid \mathbf{r}_i, q_i)$. Further sources are the Gaussian shielding charge distributions

$$\tilde{\rho}_i(\mathbf{r} \mid \mathbf{r}_i, \hat{q}_i, \hat{\sigma}_i) = \hat{q}_i G(\mathbf{r} \mid \mathbf{r}_i, \hat{\sigma}_i), \tag{10}$$

which have the total charges

$$\hat{q}_i = -(1-\varepsilon_s/\varepsilon_c)q_i \tag{11}$$

and the widths $\hat{\sigma}_i = \zeta\sigma_i$. Here, the parameter $\zeta > 1$ broadens the $\hat{\sigma}_i$ and ranges from $\zeta_B \equiv (6/\pi)^{1/3} \approx 1.24$ ideal for a single ion to $\zeta_{\mathcal{P}} \equiv 1.5439$ for densely packed atoms.[12] The last types of sources are the Gaussian RF dipole densities

$$\tilde{\mathbf{P}}_i(\mathbf{r} \mid \mathbf{r}_i, \tilde{\mathbf{p}}_i, \sigma_i) = \tilde{\mathbf{p}}_i G(\mathbf{r} \mid \mathbf{r}_i, \sigma_i) \tag{12}$$

of total strengths $\tilde{\mathbf{p}}_i$ and widths σ_i.

The Gaussian shielding charge and dipole distributions $\tilde{\rho}_i(\mathbf{r} \mid \mathbf{r}_i, \hat{q}_i, \hat{\sigma}_i)$ and $\tilde{\mathbf{P}}_i(\mathbf{r} \mid \mathbf{r}_i, \tilde{\mathbf{p}}_i, \sigma_i)$, respectively, generate the RF contribution[32]

$$\tilde{\Phi}^{\mathrm{RF}}(\mathbf{r}) = \frac{1}{\varepsilon_s}\sum_i \phi(\mathbf{r} \mid \mathbf{r}_i, \hat{q}_i, \hat{\sigma}_i) + \phi(\mathbf{r} \mid \mathbf{r}_i, \tilde{\mathbf{p}}_i, \sigma_i) \tag{13}$$

to the approximate electrostatic potential

$$\tilde{\Phi}(\mathbf{r}) = \Phi^{\mathrm{C}}(\mathbf{r}) + \tilde{\Phi}^{\mathrm{RF}}(\mathbf{r}), \tag{14}$$

if and only if[12] the RF dipoles

$$\tilde{\mathbf{p}}_k = -\alpha_k \langle\tilde{\mathbf{E}}(\mathbf{r}_k)\rangle_{\sigma_k} \tag{15}$$

are calculated self-consistently from the polarizing field[32]

$$\tilde{\mathbf{E}}(\mathbf{r}_k) = \frac{1}{\varepsilon_s}\sum_{i\neq k}[\mathbf{E}(\mathbf{r}_k \mid \mathbf{r}_i, q_i) + \mathbf{E}(\mathbf{r}_k \mid \mathbf{r}_i, \hat{q}_i, \hat{\sigma}_i) + \mathbf{E}(\mathbf{r}_k \mid \mathbf{r}_i, \tilde{\mathbf{p}}_i, \sigma_i)], \tag{16}$$

which is the derivative $\tilde{\mathbf{E}}(\mathbf{r}) = -\nabla\tilde{\Phi}(\mathbf{r})$ of the approximate potential (14). The brackets $\langle\ldots\rangle$ in Eq. (15) denote the Gaussian averages

$$\langle f(\mathbf{r}_k)\rangle_{\sigma_k} = \int f(\mathbf{r})G(\mathbf{r} \mid \mathbf{r}_k, \sigma_k)\mathrm{d}V, \tag{17}$$

which are easily evaluated for all components of $\tilde{\mathbf{E}}$. Furthermore, the atomic anti-polarizabilities α_k are given by

$$\alpha_k = (3/2)\sqrt{\pi/2}\,\varepsilon_s\sigma_k^3 S(\nu_k^\varepsilon) \tag{18}$$

and depend through

$$S(\nu_k^\varepsilon) = 8\,\nu_k^\varepsilon/(12 - \sqrt{2}\nu_k^\varepsilon) \tag{19}$$

and

$$\nu_k^\varepsilon = (1-\varepsilon_s/\varepsilon_c)\gamma\tilde{v}_k/v^{\mathrm{G}}(\sigma_k) \tag{20}$$

on the atomic volumes $\tilde{v}_k$.[12,30] Here, the parameter $\gamma \gtrsim 1$ introduces a minor empirical correction resulting from the representation (3) of the protein volume. In Paper I, γ has been chosen as 1.0 for isolated spherical objects and as 1.03 for proteins.[12]

The self-consistency iteration (15) of the RF dipoles $\tilde{\mathbf{p}}_k$ is assumed to be converged, if in the nth iteration the largest change

$$\delta_p^{(n)} = \max_{k,u}\left|\tilde{p}_{k,u}^{(n)} - \tilde{p}_{k,u}^{(n-1)}\right| \tag{21}$$

of the Cartesian components $\tilde{p}_{k,u}^{(n)}$, $u \in \{x, y, z\}$, is smaller than a threshold χ_p.[12] Because $\delta_p^{(n)}$ depends via the α_k on the atomic volumes $\tilde{v}_k^{(n)}$, the convergence of the $\tilde{\mathbf{p}}_k$ can be ascertained only after convergence of the volume iteration (6) has been reached.

In linear media, the RF energy[33]

$$\tilde{W}^{\mathrm{RF}}(\mathbf{R}) = \frac{1}{2}\sum_i q_i\tilde{\Phi}^{\mathrm{RF}}(\mathbf{r}_i \mid \mathbf{R}), \tag{22}$$

which belongs to the approximate RF potential (13) and depends on the configuration $\mathbf{R} \equiv (\mathbf{r}_1, \ldots, \mathbf{r}_N)$ of the protein atoms, approximates the dominant electrostatic contribution to the free energy of solvation. For the protein atoms $\tilde{W}^{\mathrm{RF}}(\mathbf{R})$ represents a potential of mean force, which should be suitable for MD simulations.

By comparing with available exact solutions and with numerical results calculated by means of the grid method DelPhi[26,27] for several simple model systems, including the minimal peptide model Ac-Ala-NHMe, we have shown in Paper I that our approximate RF approach is highly accurate.

Moreover, for the chosen sample systems it turned out to be by many orders of magnitude faster than DelPhi.

However, before our RF method can be applied to MD simulations, analytical expressions for the RF forces $\mathbf{f}_k^{\mathrm{RF}} = -\nabla_k \tilde{W}^{\mathrm{RF}}$ have to be given and this is the purpose of the current contribution. In the corresponding derivation we will avoid all those errors, which plagued the earlier ST approach.[30,34] Beside many other nasty details, here the key issue was the erroneous assumption[30] that the derivative $-\nabla_k \tilde{W}^{\mathrm{RF}}$ does not have to include the configurational dependence of the atomic volumes $\tilde{v}_k$ and the RF dipoles $\tilde{\mathbf{p}}_k$, which are calculated by self-consistency iterations.

To properly account for the specified configurational dependence we will employ Lagrangian multipliers[35] and, thereby, derive expressions for the RF forces, which preserve the total energy together with the linear and angular momenta during a MD simulation and, thus, yield a Hamiltonian dynamics for a macromolecule in a dielectric continuum. Correspondingly, we will call the resulting RF approach to MD "Hamiltonian dielectric solvent" (HADES).

The HADES forces turn out to have many contributions, whose relative strengths will be studied taking an ion pair as a relevant but most simple example. Furthermore, by comparing HADES with a conventional polarizable force field (PFF) we will derive an energy expression $\check{W}^{\mathrm{RF}}(\mathbf{R})$, which directly renders the HADES forces through $\mathbf{f}_k^{\mathrm{RF}} = -\nabla_k \check{W}^{\mathrm{RF}}$. Next, we will describe an efficient implementation in the MM-MD program IPHIGENIE[36,37] and demonstrate the Hamiltonian character of HADES-MD simulations. In this connection further computational aspects of HADES-MD are discussed. Finally, the free energy landscape of Ac-Ala-NHMe derived by HADES-MD is addressed before the paper is concluded by a short summary and outlook.

II. THEORY

In HADES the atomic volumes $\tilde{v}_i$ and RF dipoles $\tilde{\mathbf{p}}_i$ depend through Eqs. (6) and (15), respectively, on the atomic configuration $\mathbf{R}$. Unfortunately, the derivative

$$\mathbf{f}_k^{\mathrm{RF}} = -\nabla_k \tilde{W}^{\mathrm{RF}} - \sum_i \left(\frac{\partial \tilde{\mathbf{p}}_i}{\partial \mathbf{r}_k} + \frac{\partial \tilde{\mathbf{p}}_i}{\partial \tilde{v}_i} \frac{\partial \tilde{v}_i}{\partial \mathbf{r}_k} \right) \frac{\partial \tilde{W}^{\mathrm{RF}}}{\partial \tilde{\mathbf{p}}_i}, \tag{23}$$

cannot be evaluated, because the functional forms, by which the atomic volumes $\tilde{v}_i = \tilde{v}_i(\mathbf{R})$ depend on the coordinates $\mathbf{r}_k$ and by which the RF dipoles $\tilde{\mathbf{p}}_i = \tilde{\mathbf{p}}_i(\mathbf{R}, \tilde{v}_i)$ depend on the $\mathbf{r}_k$ and $\tilde{v}_k$, are unknown.

This problem can be circumvented by the use of Lagrangian multipliers.[35] For this purpose we consider the volumes $\tilde{v}_i$ and RF dipoles $\tilde{\mathbf{p}}_i$ as independent variables and define the generalized coordinates

$$\mathbf{q} = (\mathbf{R}, \tilde{\mathbf{p}}_1, \ldots, \tilde{\mathbf{p}}_N, \tilde{v}_1, \ldots, \tilde{v}_N).$$

Then the self-consistency conditions (6) for the $\tilde{v}_i$ give rise to N holonomic constraints $h_i^{\tilde{v}}(\mathbf{q}) = 0$ with

$$h_i^{\tilde{v}}(\mathbf{q}) = 1 - \sum_j \tilde{v}_j G(\mathbf{r}_i \mid \mathbf{r}_j, \sigma_j). \tag{24}$$

Similarly, the conditions (15) for the $\tilde{\mathbf{p}}_i$ generate $3N$ holonomic constraints $\mathbf{h}_i^{\tilde{\mathbf{p}}}(\mathbf{q}) = \mathbf{0}$ with

$$\mathbf{h}_i^{\tilde{\mathbf{p}}}(\mathbf{q}) = \tilde{\mathbf{p}}_i + \alpha_i(\tilde{v}_i)\langle \tilde{\mathbf{E}}(\mathbf{r}_i)\rangle_{\sigma_i}, \tag{25}$$

where the N vector-valued functions $\mathbf{h}_i^{\tilde{\mathbf{p}}}$ collect the three constraints for the components of $\tilde{\mathbf{p}}_i$. Because the additional degrees of freedom introduced by the $\tilde{v}_i$ and $\tilde{\mathbf{p}}_i$ are all balanced by constraints, the effective number of degrees of freedom is preserved.

The kinetic and potential energies $T(\dot{\mathbf{q}}) = \sum_i m_i \dot{\mathbf{r}}_i^2/2$ and $U(\mathbf{q}) = U(\mathbf{R})$, respectively, where $U(\mathbf{R})$ is given by a protein force field, remain unchanged, whereas the RF energy

$$\tilde{W}^{\mathrm{RF}}(\mathbf{q}) = \tilde{W}^{\mathrm{RF}}(\mathbf{R}, \tilde{\mathbf{p}}_1, \ldots, \tilde{\mathbf{p}}_N) \tag{26}$$

now explicitly depends on the RF dipoles $\tilde{\mathbf{p}}_i$. The atomic volumes $\tilde{v}_i$, however, do not explicitly appear in $\tilde{W}^{\mathrm{RF}}$ but influence it only via the constraints (25). The Lagrangian

$$L(\mathbf{q}, \dot{\mathbf{q}}) = T - (U + \tilde{W}^{\mathrm{RF}}) \tag{27}$$

of the extended system then yields, together with the constraints (24) and (25), the Euler-Lagrange equations[35]

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{q}_\mu} = \frac{\partial L}{\partial q_\mu} + \sum_i \frac{\partial \mathbf{h}_i^{\tilde{\mathbf{p}}}}{\partial q_\mu} \cdot \boldsymbol{\lambda}_i^{\tilde{\mathbf{p}}} + \sum_i \frac{\partial h_i^{\tilde{v}}}{\partial q_\mu} \lambda_i^{\tilde{v}}. \tag{28}$$

Here, the $\lambda_i^{\tilde{v}}$ and the vectors $\boldsymbol{\lambda}_i^{\tilde{\mathbf{p}}}$ are the Lagrangian multipliers coupling the volume and dipole constraints to the dynamics.

A. Lagrangian dipoles

Replacing in Eq. (28) the general variables q_μ by the RF dipoles $\tilde{\mathbf{p}}_k$ yields

$$\mathbf{0} = -\frac{\partial \tilde{W}^{\mathrm{RF}}}{\partial \tilde{\mathbf{p}}_k} + \sum_i \frac{\partial \mathbf{h}_i^{\tilde{\mathbf{p}}}}{\partial \tilde{\mathbf{p}}_k} \boldsymbol{\lambda}_i^{\tilde{\mathbf{p}}} \tag{29}$$

because the Lagrangian (27) does not depend on the dipole velocities $\dot{\tilde{\mathbf{p}}}_k$. Inserting now the RF energy (22) and the dipole constraints (25) together with the polarizing field (16) into Eq. (29) one finds

$$\mathbf{0} = \frac{1}{2}\sum_{i\neq k} \langle \mathbf{E}(\mathbf{r}_k \mid \mathbf{r}_i, q_i)\rangle_{\sigma_k} + \boldsymbol{\lambda}_k^{\tilde{\mathbf{p}}} + \sum_{i\neq k} \alpha_i \left\langle \frac{\partial \mathbf{E}(\mathbf{r}_i \mid \mathbf{r}_k, \tilde{\mathbf{p}}_k, \sigma_k)}{\partial \tilde{\mathbf{p}}_k} \right\rangle_{\sigma_i} \boldsymbol{\lambda}_i^{\tilde{\mathbf{p}}} \tag{30}$$

demonstrating that the multipliers $\boldsymbol{\lambda}_k^{\tilde{\mathbf{p}}}$ have the dimension of an electric field. Replacing these fields in Eq. (30) by the "Lagrangian dipoles"

$$\check{\mathbf{p}}_k \equiv 2\alpha_k \boldsymbol{\lambda}_k^{\tilde{\mathbf{p}}}/\varepsilon_s \tag{31}$$

leads after a few lines of algebra to the new self-consistency conditions

$$\check{\mathbf{p}}_k = -\alpha_k \langle \check{\mathbf{E}}(\mathbf{r}_k)\rangle_{\sigma_k} \tag{32}$$

for the Lagrangian dipoles $\check{\mathbf{p}}_k$, where the anti-polarizing field

$$\check{\mathbf{E}}(\mathbf{r}_k) \equiv \frac{1}{\varepsilon_s}\sum_{i\neq k} [\mathbf{E}(\mathbf{r}_k \mid \mathbf{r}_i, q_i) + \mathbf{E}(\mathbf{r}_k \mid \mathbf{r}_i, \check{\mathbf{p}}_i, \sigma_i)] \tag{33}$$

is generated by the $\check{\mathbf{p}}_i$ and by the partial charges q_i of the other atoms i. Thus, the Lagrangian multipliers $\boldsymbol{\lambda}_k^{\tilde{\mathbf{p}}}$ for the constraint (25) are self-consistently calculated by iterating equation (32).

Note that neither the RF dipoles $\tilde{\mathbf{p}}_i$ nor the shielding charges $\hat{q}_i$ contribute to the new anti-polarizing field $\check{\mathbf{E}}(\mathbf{r}_k)$ defined by Eq. (33). Furthermore, according to this definition the $\check{\mathbf{p}}_i$ are, just like the RF dipoles $\tilde{\mathbf{p}}_i$, strengths of Gaussian dipole distributions with widths σ_i. Finally, in a computation of HADES forces the $\check{\mathbf{p}}_i$ and the $\tilde{\mathbf{p}}_i$ are concomitantly iterated until the convergence criterion defined in connection with Eq. (21) is met for both quantities.

B. Volume multipliers

Next we turn to the Euler-Lagrange equations

$$0 = \sum_i \lambda_i^{\tilde{v}} \frac{\partial h_i^{\tilde{v}}}{\partial v_k} + \boldsymbol{\lambda}_k^{\tilde{\mathbf{p}}} \cdot \frac{\partial \mathbf{h}_k^{\tilde{\mathbf{p}}}}{\partial v_k} \tag{34}$$

for the atomic volumes $\tilde{v}_k$, which follow from Eq. (28) by restricting the generalized coordinates q_μ to the $\tilde{v}_k$. Inserting the definitions (24) and (25) for $h_i^{\tilde{v}}$ and $\mathbf{h}_k^{\tilde{\mathbf{p}}}$, respectively, into Eq. (34) yields

$$0 = -\sum_i \lambda_i^{\tilde{v}} G(\mathbf{r}_i \,|\, \mathbf{r}_k, \sigma_k) + \frac{\partial \alpha_k}{\partial v_k} \boldsymbol{\lambda}_k^{\tilde{\mathbf{p}}} \cdot \langle \mathbf{E}(\mathbf{r}_k) \rangle_{\sigma_k}. \tag{35}$$

To simplify the notation we define the scaled multiplier $\lambda_i \equiv 2\varepsilon_s \lambda_i^{\tilde{v}}$. With the definitions (5) of the normalized Gaussian $G(\mathbf{r}_i \,|\, \mathbf{r}_k,\ \sigma_k)$, (18) of the atomic polarizability α_k, and (31) of the Lagrangian dipole $\check{\mathbf{p}}_k$, and with the self-consistency condition (15) of the RF dipoles $\tilde{\mathbf{p}}_k$ one gets the conditions

$$\lambda_k = -4\pi \frac{\check{\mathbf{p}}_k \cdot \tilde{\mathbf{p}}_k}{\tilde{v}_k v_k^{\varepsilon}} - \sum_{i \neq k} \lambda_i \exp\left(-r_{ik}^2 / 2\sigma_k^2\right) \tag{36}$$

for the scaled multipliers λ_k, which have to be self-consistently fulfilled. Convergence is reached as soon as

$$\delta_\lambda^{(n)} = \max_k \left| \lambda_k^{(n)} - \lambda_k^{(n-1)} \right| \tag{37}$$

drops below a given threshold χ_λ, which has the dimension of an energy. Since the atomic Lagrangian energies λ_k depend on the $\tilde{v}_k$, $\tilde{\mathbf{p}}_k$, and $\check{\mathbf{p}}_k$, the latter quantities have to be converged before the self-consistency iteration (36) can be stopped.

C. Equations of motion

Restricting finally the Euler-Lagrange equations (28) to the atomic coordinates $\mathbf{r}_k$ yields the equations of motion

$$m_k \ddot{\mathbf{r}}_k = -\nabla_k \left[U + \tilde{W}^{\mathrm{RF}} - \sum_i \frac{1}{2\alpha_i} \left(\mathbf{h}_i^{\tilde{\mathbf{p}}} \cdot \check{\mathbf{p}}_i \right) - \sum_i \frac{\lambda_i}{2\varepsilon_s} h_i^{\tilde{v}} \right], \tag{38}$$

which can be applied, as soon as the self-consistency iterations of the quantities $\tilde{v}_k$, $\tilde{\mathbf{p}}_k$, $\check{\mathbf{p}}_k$, and λ_k are converged.

The forces appearing on the right hand side of Eq. (38) consist of the MM forces $-\nabla_k U$, which include Coulomb, van der Waals, and the so-called bonded interactions. The second term

$$\begin{aligned} -\nabla_k \tilde{W}^{\mathrm{RF}} = {} & \frac{q_k}{2\varepsilon_s} \sum_{i \neq k} \left[\mathbf{E}(\mathbf{r}_k \,|\, \mathbf{r}_i, \hat{q}_i, \hat{\sigma}_i) + \mathbf{E}(\mathbf{r}_k \,|\, \mathbf{r}_i, \tilde{\mathbf{p}}_i, \sigma_i) \right] \\ & + \frac{\hat{q}_k}{2\varepsilon_s} \sum_{i \neq k} \langle \mathbf{E}(\mathbf{r}_k \,|\, \mathbf{r}_i, q_i) \rangle_{\hat{\sigma}_k} \\ & + \frac{1}{2\varepsilon_s} \sum_{i \neq k} \langle \nabla \mathbf{E}(\mathbf{r}_k \,|\, \mathbf{r}_i, q_i) \rangle_{\sigma_k} \tilde{\mathbf{p}}_k \end{aligned} \tag{39}$$

is composed of the *actio* force, which arises from the action of the shielding charges $\hat{q}_i$ and RF dipoles $\tilde{\mathbf{p}}_i$ at the other atoms i on the partial charge q_k of the target atom k, and of the *reaction force*, which belongs to the action of the shielding charge $\hat{q}_k$ and RF dipole $\tilde{\mathbf{p}}_k$ of the target atom k on the partial charges of all other atoms i. Here, we have introduced the short hand notation $\nabla \mathbf{E} \equiv \partial \mathbf{E} / \partial \mathbf{r}$ for the field gradient tensor.[32] Note that the force (39) obeys Newton's third law as follows from the identities

$$q_k \mathbf{E}(\mathbf{r}_k \,|\, \mathbf{r}_i, \hat{q}_i, \hat{\sigma}_i) = -\hat{q}_i \langle \mathbf{E}(\mathbf{r}_i \,|\, \mathbf{r}_k, q_k) \rangle_{\hat{\sigma}_i} \tag{40}$$

and

$$q_k \mathbf{E}(\mathbf{r}_k \,|\, \mathbf{r}_i, \tilde{\mathbf{p}}_i, \sigma_i) = -\langle \nabla \mathbf{E}(\mathbf{r}_i \,|\, \mathbf{r}_k, q_k) \rangle_{\sigma_i} \tilde{\mathbf{p}}_i, \tag{41}$$

which can be easily checked with the help of the explicit expression given for the various fields in the Appendix.

The third term on the r.h.s. of Eq. (38) becomes with the definitions (25) of $\mathbf{h}_i^{\tilde{\mathbf{p}}}$ and (16) of the polarizing field $\tilde{\mathbf{E}}(\mathbf{r}_k)$ the force

$$\begin{aligned} & \nabla_k \sum_i \frac{1}{2\alpha_i} \left(\mathbf{h}_i^{\tilde{\mathbf{p}}} \cdot \check{\mathbf{p}}_i \right) \\ & = \frac{q_k}{2\varepsilon_s} \sum_{i \neq k} \mathbf{E}(\mathbf{r}_k \,|\, \mathbf{r}_i, \check{\mathbf{p}}_i, \sigma_i) + \frac{\hat{q}_k}{2\varepsilon_s} \sum_{i \neq k} \langle \mathbf{E}(\mathbf{r}_k \,|\, \mathbf{r}_i, \check{\mathbf{p}}_i, \sigma_i) \rangle_{\hat{\sigma}_k} \\ & \quad + \frac{1}{2\varepsilon_s} \sum_{i \neq k} \langle \nabla \mathbf{E}(\mathbf{r}_k \,|\, \mathbf{r}_i, \check{\mathbf{p}}_i, \sigma_i) \rangle_{\sigma_k} \tilde{\mathbf{p}}_k + \frac{1}{2} \langle \nabla \tilde{\mathbf{E}}(\mathbf{r}_k) \rangle_{\sigma_k} \check{\mathbf{p}}_k, \end{aligned} \tag{42}$$

which expresses through the first three terms the actions of the Lagrangian dipoles $\check{\mathbf{p}}_i$ at all other atoms i on the partial charge q_k, the shielding charge $\hat{q}_k$, and the RF dipole $\tilde{\mathbf{p}}_k$ of atom k. The last term in Eq. (42) is the force, which is exerted by the q_i, $\hat{q}_i$, and $\tilde{\mathbf{p}}_i$ of all other atoms i through the gradient of the polarizing field (16) on the Lagrangian dipole $\check{\mathbf{p}}_k$ of atom k. The constraining force (42) accounts for the changes $\partial \tilde{W}^{\mathrm{RF}} / \partial \tilde{\mathbf{p}}_i$ of the RF energy, which are according to Eq. (23) caused by the variations $\partial \tilde{\mathbf{p}}_i / \partial \mathbf{r}_k$ of the RF dipoles upon the motion of atom k. Also this force complies with Newton's reaction principle because of the identity (41) and the identities

$$\hat{q}_k \langle \mathbf{E}(\mathbf{r}_k \,|\, \mathbf{r}_i, \tilde{\mathbf{p}}_i, \sigma_i) \rangle_{\hat{\sigma}_k} = -\langle \nabla \mathbf{E}(\mathbf{r}_i \,|\, \mathbf{r}_k, \hat{q}_k, \hat{\sigma}_k) \rangle_{\sigma_i} \tilde{\mathbf{p}}_i \tag{43}$$

and

$$\langle \nabla \mathbf{E}(\mathbf{r}_k \,|\, \mathbf{r}_i, \check{\mathbf{p}}_i, \sigma_i) \rangle_{\sigma_k} \tilde{\mathbf{p}}_k = -\langle \nabla \mathbf{E}(\mathbf{r}_i \,|\, \mathbf{r}_k, \tilde{\mathbf{p}}_k, \sigma_k) \rangle_{\sigma_i} \check{\mathbf{p}}_i. \tag{44}$$

The last contribution to the HADES force results from the fourth term on the r.h.s. of Eq. (38) by inserting the definition

(24) for the constraints $h_i^{\tilde{v}}$ and is given by

$$\sum_i \frac{\lambda_i}{2\varepsilon_s}\frac{\partial h_i^{\tilde{v}}}{\partial \mathbf{r}_k} = \frac{1}{2\varepsilon_s}\sum_{i\neq k}[\lambda_i \nabla_i \tilde{\vartheta}_k(\mathbf{r}_i) - \lambda_k \nabla_k \tilde{\vartheta}_i(\mathbf{r}_k)] \quad (45)$$

in terms of gradients of the Gaussian atomic functions (4). They obviously obey Newton's third law. The constraining force (45) expresses the changes $\partial \tilde{W}^{\mathrm{RF}}/\partial \tilde{\mathbf{p}}_i$ of the RF energy with the variations $\partial \tilde{\mathbf{p}}_i/\partial \tilde{v}_i$ of the RF dipoles $\tilde{\mathbf{p}}_i$, which arise, if the volumes $\tilde{v}_i$ of their atoms change ($\partial \tilde{v}_i/\partial \mathbf{r}_k \neq 0$) with the location of atom k.

D. HADES pair forces

As a result we may state that the use of Lagrangian multipliers has eventually enabled us to derive explicit expressions for the HADES forces $\mathbf{f}_k^{\mathrm{RF}}$ on the atoms $k = 1, \ldots, N$, of a protein embedded in a dielectric continuum ε_c. Originally these forces had been defined by Eq. (23) as gradients of our RF energy function $\tilde{W}^{\mathrm{RF}}$. Therefore, they conserve the total energy as well as the linear and angular momenta, if all self-consistency conditions are strictly met. Further below, the conservation laws will be checked for a simple HADES-MD example, which will serve to verify our implementation and the chosen convergence thresholds χ_v, χ_p, and χ_λ.

Inspecting the contributions (39), (42), and (45) to the HADES forces $\mathbf{f}_k^{\mathrm{RF}}$ we recognize that they can be written as the sums

$$\mathbf{f}_k^{\mathrm{RF}} = \frac{1}{\varepsilon_s}\sum_{i\neq k}\left(\mathbf{f}_{ki}^{q} + \mathbf{f}_{ki}^{\check{\mathbf{p}}} + \mathbf{f}_{ki}^{v}\right), \quad (46)$$

of atomic pair forces. Here the $\mathbf{f}_{ki}^{q}$ express the RF forces (39), whereas the $\mathbf{f}_{ki}^{\check{\mathbf{p}}}$ and the $\mathbf{f}_{ki}^{v}$ represent the constraining forces (42) and (45), respectively. All these contributions $\mathbf{f}_{ki}^{x}$, $x \in \{q,\ \check{\mathbf{p}},\ v\}$, to $\mathbf{f}_k^{\mathrm{RF}}$ will now be characterized step by step.

For this purpose we use as a most simple example two ions k and i with the opposite charges $+e$ and $-e$, respectively, and with identical radii $R = 1.7$ Å, which are separated by the distance $r_{ki} \equiv |\mathbf{r}_k - \mathbf{r}_i|$. The dielectric constants are $\varepsilon_s = 1$ in the interior of the ionic cavities and $\varepsilon_c = 80$ in the surrounding continuum, the Gaussian widths of the RF dipoles are $\sigma \approx 1.1$ Å, and the two scaling factors are chosen as $\zeta = \zeta_B$ and $\gamma = 1.0$. The choice of these parameters is described in Sec. IV B of Paper I of this work, whereas the energetics of this ion pair is thoroughly discussed in Sec. S4 of the corresponding supplementary material.[12] In the computations of the HADES pair forces $\mathbf{f}_{ki}^{x}$ we chose for the self-consistency iterations the thresholds $\chi_v = 10^{-14}$, $\chi_p = 10^{-14}$ Debye, and $\chi_\lambda = 10^{-14}$ kcal/mol. In the given example all pair forces $\mathbf{f}_{ki}^{x}$ are parallel to the connection $\mathbf{r}_{ki}$. Therefore, they can be measured by the scalar projections $f_{ki}^{x} = \mathbf{f}_{ki}^{x} \cdot \mathbf{r}_{ki}/r_{ki}$.

Figure 1 compares the projections f_{ki}^{x} with the corresponding projection f_{ki}^{RF} of the total HADES force, which is represented by the black line in Fig. 1(a). f_{ki}^{RF} is everywhere repulsive and has a maximum of 183 kJ mol^{-1}Å^{-1} at $r_{ki}/R = 1.25$. Of similar shape and magnitude is the projection f_{ki}^{q}

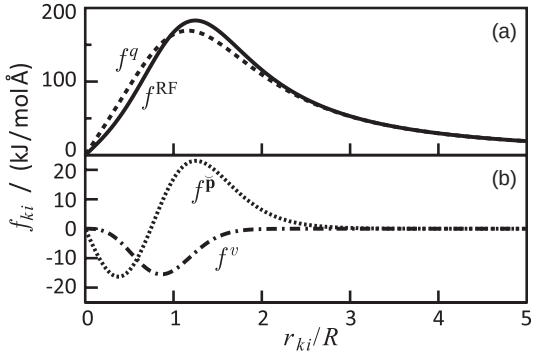

FIG. 1. Projections f_{ki}^{x} of the total HADES force $x = \mathrm{RF}$ and of its three contributions $x = q,\ \check{\mathbf{p}},\ v$ for two oppositely charged ions ($\pm e$) with radii R at a distance r_{ki}/R. (a) f_{ki}^{RF} (solid) and its main contribution f_{ki}^{q} (dashed). (b) The constraining forces $f_{ki}^{\check{\mathbf{p}}}$ (dotted) and f_{ki}^{v} (dashed-dotted). Note the different scales in (a) and (b).

(dashed line in Fig. 1(a)) of the first contribution

$$\mathbf{f}_{ki}^{q} = \frac{1}{2}[q_k \mathbf{E}(\mathbf{r}_k \mid \mathbf{r}_i, \hat{q}_i, \hat{\sigma}_i) + \hat{q}_k \langle \mathbf{E}(\mathbf{r}_k \mid \mathbf{r}_i, q_i)\rangle_{\hat{\sigma}_k} + q_k \mathbf{E}(\mathbf{r}_k \mid \mathbf{r}_i, \tilde{\mathbf{p}}_i, \sigma_i) + \langle \nabla \mathbf{E}(\mathbf{r}_k \mid \mathbf{r}_i, q_i)\rangle_{\sigma_k}\tilde{\mathbf{p}}_k] \quad (47)$$

to $\mathbf{f}_k^{\mathrm{RF}}$, which describes the interactions of the partial charges with the Gaussian shielding charges and RF dipoles. According to Figure 1(a), f_{ki}^{q} visibly differs from f_{ki}^{RF} only in the overlap region $r/R < 2.5$ of the Gaussians associated to the two ions.

This small difference is compensated by the constraining forces $\mathbf{f}_{ki}^{\check{\mathbf{p}}}$ and $\mathbf{f}_{ki}^{v}$, whose projections are depicted as dotted and dashed-dotted lines, respectively, in Fig. 1(b). Here, the projection of the constraining force

$$\mathbf{f}_{ki}^{\check{\mathbf{p}}} = \frac{1}{2}[q_k \mathbf{E}(\mathbf{r}_k \mid \mathbf{r}_i, \check{\mathbf{p}}_i, \sigma_i) + \langle \nabla \mathbf{E}(\mathbf{r}_k \mid \mathbf{r}_i, q_i)\rangle_{\sigma_k}\check{\mathbf{p}}_k + \hat{q}_k \langle \mathbf{E}(\mathbf{r}_k \mid \mathbf{r}_i, \check{\mathbf{p}}_i, \sigma_i)\rangle_{\hat{\sigma}_k} + \langle \nabla \mathbf{E}(\mathbf{r}_k \mid \mathbf{r}_i, \hat{q}_i, \hat{\sigma}_i)\rangle_{\sigma_k}\check{\mathbf{p}}_k + \langle \nabla \mathbf{E}(\mathbf{r}_k \mid \mathbf{r}_i, \check{\mathbf{p}}_i, \sigma_i)\rangle_{\sigma_k}\tilde{\mathbf{p}}_k + \langle \nabla \mathbf{E}(\mathbf{r}_k \mid \mathbf{r}_i, \tilde{\mathbf{p}}_i, \sigma_i)\rangle_{\sigma_k}\check{\mathbf{p}}_k], \quad (48)$$

which derives from the self-consistency condition of the RF dipoles $\tilde{\mathbf{p}}_i$ and describes the interactions of the Lagrangian dipoles with the partial charges, shielding charges, and RF dipoles, is repulsive at large distances, has a maximum at $r_{ik}/R \approx 1.25$, and changes sign at $r_{ik}/R \approx 0.75$. In contrast, the projection (dashed-dotted line) of the other constraining force

$$\mathbf{f}_{ki}^{v} = \frac{1}{2}[\lambda_i \nabla_i \tilde{\vartheta}_k(\mathbf{r}_i) - \lambda_k \nabla_k \tilde{\vartheta}_i(\mathbf{r}_k)], \quad (49)$$

which belongs to the self-consistency condition for the atomic volumes $\tilde{v}_i$, is attractive for strongly overlapping ions.

Recall here that the constraining forces $\mathbf{f}_{ki}^{\check{\mathbf{p}}}$ and $\mathbf{f}_{ki}^{v}$ derive according to Eq. (23) from partial derivatives $\partial \tilde{\mathbf{p}}_i/\partial \mathbf{r}_k$ and $\partial \tilde{v}_i/\partial \mathbf{r}_k$. As is demonstrated by the distance dependences of $\tilde{p}_i$ and $\tilde{v}_i$ displayed in Fig. 2, these quantities show pronounced changes in the ranges $0 < r_{ki}/R < 2.5$ and $0 < r_{ki}/R < 2$, respectively. A glance at Fig. 1(b) demonstrates that the shapes and ranges of the associated pair forces $\mathbf{f}_{ki}^{\check{\mathbf{p}}}$ and $\mathbf{f}_{ki}^{v}$ nicely fit to the distance-derivatives of the two curves shown in Fig. 2.

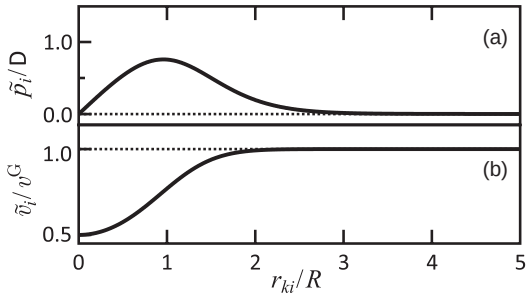

FIG. 2. Self-consistent RF dipole $|\tilde{\mathbf{p}}_i|$ (a) and volume $\tilde{v}_i$ (b) of the negative ion i as functions of the ion-ion distance r_{ki}.

E. Comparison with polarizable force fields

The relation of the ET approach with a polarizable force field (PFF)[38] was first recognized by ST.[30] This relation is preserved by HADES, because it also employs anti-polarizable RF dipoles [cf. Eq. (15)]. The HADES forces (47)–(49), however, seem to be much more complicated than those of a PFF. For a better understanding it is instructive to derive also the forces of a PFF by Lagrangian multipliers.

The Lagrangian $L^{\mathrm{PFF}} = T - (U + W^{\mathrm{PFF}})$ of a PFF simulation features the linear response polarization energy[33]

$$W^{\mathrm{PFF}} = 1/2 \sum_{i,j\neq i} q_i \phi(\mathbf{r}_i \,|\boldsymbol{\mu}_j), \tag{50}$$

of the point charges q_i in the potential of the atomic point dipoles $\boldsymbol{\mu}_k$ in close analogy to the HADES Lagrangian (27) and the RF energy (22). The $\boldsymbol{\mu}_k$ are determined by polarizabilities α_k and the self-consistency conditions

$$\boldsymbol{\mu}_k = \alpha_k \sum_{i\neq k} [\mathbf{E}(\mathbf{r}_k \,|\, \mathbf{r}_i, q_i) + \mathbf{E}(\mathbf{r}_k \,|\, \mathbf{r}_i, \boldsymbol{\mu}_i)], \tag{51}$$

which can be represented by holonomic constraints similar to the constraints (25) of the RF dipoles, i.e., as $\mathbf{h}_k^{\boldsymbol{\mu}} = 0$. The potential and field of the point dipoles $\boldsymbol{\mu}_j$ in Eqs. (50) and (51), respectively, are obtained from Eqs. (A3) and (A7) in the limit $\sigma \to 0$ of vanishing Gaussian widths. In contrast to HADES, the PFF polarizabilities α_k are constant and, therefore, volume constraints are unnecessary. Furthermore, a PFF does not contain shielding charges.

Substituting $\tilde{W}^{\mathrm{RF}}$ in the Lagrangian (27) by W^{PFF} and replacing the Lagrangian multipliers $\boldsymbol{\lambda}_k^{\boldsymbol{\mu}}$, which appear in the resulting analogue of the Euler-Lagrange equation (29), by the PFF-type Lagrangian dipoles $\check{\boldsymbol{\mu}}_k = 2\alpha_k \boldsymbol{\lambda}_k^{\boldsymbol{\mu}}$ one finds, as explained in the text accompanying Eq. (30), also for the new Lagrangian dipoles $\check{\boldsymbol{\mu}}_k$ self-consistency conditions

$$\check{\boldsymbol{\mu}}_k = \alpha_k \sum_{i\neq k} [\mathbf{E}(\mathbf{r}_k \,|\, \mathbf{r}_i, q_i) + \mathbf{E}(\mathbf{r}_i \,|\, \mathbf{r}_k, \check{\boldsymbol{\mu}}_i)] \tag{52}$$

analogous to those in Eq. (32). They are formally identical to the conditions (51) for the PFF dipoles $\boldsymbol{\mu}_k$. With the choice $\check{\boldsymbol{\mu}}_k \equiv \boldsymbol{\mu}_k$ the conditions (52) are met, if Eq. (51) holds. Thus, the PFF dipoles are their own Lagrangian multipliers and only one set of inducible dipoles has to be iterated.

The PFF forces appearing in the equations of motion analogous to Eq. (38) are then negative gradients of the effective[35] potential energy

$$\check{W}^{\mathrm{PFF}} = W^{\mathrm{PFF}} - \sum_i \frac{\boldsymbol{\mu}_i}{2\alpha_i} \cdot \mathbf{h}_i^{\boldsymbol{\mu}}. \tag{53}$$

Note that the values of $\check{W}^{\mathrm{PFF}}$ and W^{PFF} are identical, if the self-consistency conditions (51) are fulfilled, because then $\mathbf{h}_i^{\boldsymbol{\mu}} = 0$ for all i. Inserting explicit expressions for W^{PFF} and $\mathbf{h}_i^{\boldsymbol{\mu}}$ into Eq. (53) yields after a few rearrangements

$$\check{W}^{\mathrm{PFF}} = \sum_{i,k\neq i} q_i \phi(\mathbf{r}_i \,|\boldsymbol{\mu}_k) - \frac{1}{2} \sum_{i,k\neq i} \boldsymbol{\mu}_i \cdot \mathbf{E}(\mathbf{r}_i \,|\boldsymbol{\mu}_k) + \sum_i \frac{\boldsymbol{\mu}_i^2}{2\alpha_i}, \tag{54}$$

which is the PFF energy expression given by Ahlström *et al.*[38] It covers the total energy of the partial charges in the potential of the dipoles, the dipole–dipole interaction energy, and the self-energy of the polarization. One can readily show that among all possible polarizations, the self-consistent dipoles *minimize* $\check{W}^{\mathrm{PFF}}$.[38]

Just like in Eq. (53) for the PFF, we can also define for HADES the effective RF energy

$$\check{W}^{\mathrm{RF}} = \tilde{W}^{\mathrm{RF}} - \frac{1}{2} \sum_i \check{\mathbf{p}}_i \langle \tilde{\mathbf{E}}(\mathbf{r}_i) \rangle_{\sigma_i} - \sum_i \frac{\check{\mathbf{p}}_i \cdot \tilde{\mathbf{p}}_i}{2\alpha_i} - \sum_i \frac{\lambda_i}{2\varepsilon_s} h_i^{\tilde{v}}, \tag{55}$$

from which the forces were calculated in Sec. II C as negative gradients with respect to the atomic positions $\mathbf{r}_k$. The third term resembles the self-energy in $\check{W}^{\mathrm{PFF}}$. In HADES, however, the RF dipoles $\tilde{\mathbf{p}}_k$ cannot be identified with their Langrangian siblings $\check{\mathbf{p}}_k$, because the shielding charges $\hat{q}_i$ polarize only the $\tilde{\mathbf{p}}_k$ but not the $\check{\mathbf{p}}_k$. Only in the limit of a large separation between all atoms, i.e., $r_{jk} \gg \sigma_{jk}$ for $j \neq k$, we find the relation $\check{\mathbf{p}}_k = (\varepsilon_c/\varepsilon_s)\tilde{\mathbf{p}}_k$. Furthermore, in this limit one can show that the $\check{\mathbf{p}}_k$ and $\tilde{\mathbf{p}}_k$ *maximize* $\check{W}^{\mathrm{RF}}$ due to the anti-polarizable nature of the HADES RF. The last term in Eq. (55), which contributes no energy but yields the forces $\mathbf{f}_{ki}^{v}$ arising from the volume constraints $h_i^{\tilde{v}}$ as defined by Eq. (24), has no PFF correspondence.

III. IMPLEMENTATION

The HADES forces, which are specified by Eqs. (46)–(49), turned out to be quite complicated. Now one may fear that the efficiency of their evaluation cannot suffice for HADES-MD simulations, because the atomic volumes $\tilde{v}_k$, the RF dipoles $\tilde{\mathbf{p}}_k$, the Lagrangian dipoles $\check{\mathbf{p}}_k$, and the multipliers λ_k have to be self-consistently calculated to fulfill the conditions (6), (15), (32), and (36), respectively.

We will now show that these difficulties can be surmounted, if one takes advantage of fast multipole methods[39–41] (FMM) for the efficient treatment of long-range interactions and of polynomial extrapolations[42,43] combined with direct inversions in the iterative subspace[44–46] (DIIS) for the speed-up of the self-consistency iterations.

A. FMM for HADES

FMMs are particularly useful, because the HADES forces contain many fields of Gaussian charge and dipole distributions, which become fields of point charges and dipoles

already at quite small inter-atomic distances, and because FMMs anyway distinguish the computation of nearby interactions from more distant ones. Similarly, this distinction can be exploited to neglect the atomic characteristic functions (4) in the iterative calculation Eq. (6) of the atomic volumes for large inter-atomic distances r_{ki}, because these Gaussian functions are extremely short-ranged.[47]

Our implementation of HADES-MD takes advantage of an existent FMM program called IPHIGENIE,[36] which realizes the structure adapted multipole method[36,48–50] ($SAMM_4$) for the evaluation of long range electrostatic interactions. Meanwhile $SAMM_4$ has been extended to the van der Waals interactions as modeled by Lennard Jones potentials (Lorenzen, ongoing dissertation). IPHIGENIE covers PFFs and, in particular, efficiently handles Gaussian charge and dipole distributions[37,51] as required by HADES.

The computation of the electrostatic HADES pair interactions has been integrated into the $SAMM_4$ scheme, which applies a hierarchical decomposition of the simulated system into nested clusters of decreasing size and eventually resolves the system into individual atoms for the treatment of nearest neighbor interactions. At larger distances the long-range interactions are treated by fourth order fast multipole expansions of point-like charge and dipole distributions.[36,37] Because the HADES electrostatics is generated at large distances by point-like charges (q_i, $\hat{q}_i$) and dipoles ($\tilde{\mathbf{p}}_i$, $\check{\mathbf{p}}_i$), it nicely fits into the computational scheme of $SAMM_4$.

As is illustrated by Table I, the computation of the HADES force $\mathbf{f}_k^{\mathrm{RF}}$ on an atom k requires access to 15 different contributions of other atoms i to the electrostatic potential, field, and field gradient at $\mathbf{r}_k$. These contributions are generated by the partial charges q_i, shielding charges $\hat{q}_i$, RF dipoles $\tilde{\mathbf{p}}_i$, and Lagrangian dipoles $\check{\mathbf{p}}_i$ of the atoms i and affect beside the properties $\tilde{\mathbf{p}}_k$ and $\check{\mathbf{p}}_k$ of atom k also pair quantities, i.e., the RF energies

$$\tilde{W}_{ki}^{\mathrm{RF}} \equiv \frac{q_k}{2\varepsilon_s}[\phi(\mathbf{r}_k|\mathbf{r}_i,\hat{q}_i,\hat{\sigma}_i)+\phi(\mathbf{r}_k|\mathbf{r}_i,\tilde{\mathbf{p}}_i,\sigma_i)],$$

the RF forces $\mathbf{f}_{ki}^{q}$, and the constraint forces $\mathbf{f}_{ki}^{\check{\mathbf{p}}}$.

Note that the quantities in columns two and three must be calculated only once at the beginning of the self-consistency iterations and are stored separately, because the partial charges q_i and the shielding charges $\hat{q}_i$ are constant. Similarly, also the fields generated by the inducible dipoles $\tilde{\mathbf{p}}_i$ and $\check{\mathbf{p}}_i$ are kept separately. They are updated in each iteration step until convergence is reached. The quantities listed in the last two rows of Table I are necessary for the calculation of the HADES pair forces and include the computationally expensive field gradients of $\tilde{\mathbf{p}}_i$ and $\check{\mathbf{p}}_i$. They must be calculated only once immediately after convergence.

Note furthermore that the Gaussian averages, which are indicated by the brackets $\langle\ldots\rangle_\sigma$ in Table I and defined by Eq. (17), can be neglected at sufficiently large distances. This is the case, in particular, for the $SAMM_4$ computation of the long-range parts of the electrostatics. Then, for instance, a distant source $\check{\mathbf{p}}_i$ requires only the calculation of $\mathbf{E}$ instead of the three field variants listed in the last column of Table I, which considerably reduces the complexity and effort of the computation.

TABLE I. Respective potentials, fields, and field gradients and their Gaussian averaged counterparts of the electrostatic sources q_i, $\hat{q}_i$, $\tilde{\mathbf{p}}_i$, and $\check{\mathbf{p}}_i$ required to calculate the HADES properties of atom k listed in the first column. The sets $\mathcal{E}^q \equiv \{\langle\mathbf{E}\rangle_{\hat{\sigma}_k}, \langle\nabla\mathbf{E}\rangle_{\sigma_k}\}$ and $\mathcal{E}^{\check{\mathbf{p}}} \equiv \{\mathbf{E}, \langle\mathbf{E}\rangle_{\hat{\sigma}_k}, \langle\nabla\mathbf{E}\rangle_{\sigma_k}\}$ collect fields and field gradients generated by the partial charges q_i and Lagrangian dipoles $\check{\mathbf{p}}_i$, respectively.

	q_i	$\hat{q}_i$	$\tilde{\mathbf{p}}_i$	$\check{\mathbf{p}}_i$
$\tilde{\mathbf{p}}_k$	$\langle\mathbf{E}\rangle_{\sigma_k}$	$\langle\mathbf{E}\rangle_{\sigma_k}$	$\langle\mathbf{E}\rangle_{\sigma_k}$	…
$\check{\mathbf{p}}_k$	$\langle\mathbf{E}\rangle_{\sigma_k}$	…	…	$\langle\mathbf{E}\rangle_{\sigma_k}$
$\tilde{W}_{ki}^{\mathrm{RF}}$	…	ϕ	ϕ	…
$\mathbf{f}_{ki}^{q}$	$\mathcal{E}^q$	$\mathbf{E}$	$\mathbf{E}$	…
$\mathbf{f}_{ki}^{\check{\mathbf{p}}}$	$\langle\nabla\mathbf{E}\rangle_{\sigma_k}$	$\langle\nabla\mathbf{E}\rangle_{\sigma_k}$	$\langle\nabla\mathbf{E}\rangle_{\sigma_k}$	$\mathcal{E}^{\check{\mathbf{p}}}$

B. DIIS for HADES

The self-consistency iterations of the atomic volumes $\tilde{v}_k$, RF dipoles $\tilde{\mathbf{p}}_k$, Lagrangian dipoles $\check{\mathbf{p}}_k$, and multipliers λ_k show a quite slow convergence. DIIS[44–46] can reduce the required number of iterations by orders of magnitude. Instead of calculating in an iteration step a quantity $a^{(n+1)}$ from its predecessor $a^{(n)}$, DIIS employs an estimated predecessor $\hat{a}^{(n)} = \sum_{j=0}^{l-1} c_{n-j} a^{(n-j)}$. This estimate is calculated from a history covering the l values $\{a^{(n-l+1)}, \ldots, a^{(n)}\}$ and from expansion coefficients c_{n-j}, which are solutions of a $(l+1)$-dimensional system of linear equations. This system is constructed from the shifts $\{\delta a^{(n-l+1)}, \ldots, \delta a^{(n)}\}$, which are defined by $\delta a^{(n)} = a^{(n)} - \hat{a}^{(n-1)}$. If self-consistency is required for N atomic variables, the quantities $a^{(n)}$ addressed above are actually N-dimensional vectors collecting these variables. Because the optimal length l of a DIIS extrapolation depends on the optimization problem, in HADES the lengths l_v, l_p, and l_λ of the DIIS extrapolations for the $\tilde{v}_k$, $\tilde{\mathbf{p}}_k$ and $\check{\mathbf{p}}_k$, and λ_k, respectively, can be individually chosen.

C. Extrapolation of initial values

When executing one step of the numerical integration of the equations of motion (38) with a small time step Δt, the atomic coordinates change only a little. Correspondingly all those HADES quantities, which have to meet self-consistency conditions, should experience likewise small changes. Employing Lagrangian polynomials[42,43] to extrapolate suitable starting values

$$a(t+\Delta t) = \sum_{i=0}^{m-1} c_i a(t - i\Delta t) \tag{56}$$

for the self-consistency iterations from histories $\{a[t-(m-1)\Delta t], \ldots, a(t)\}$ of lengths m using the coefficients

$$c_i = \prod_{\substack{0\le j<m \\ j\neq i}} \frac{1+j}{i-j}, \tag{57}$$

which solely depend on m. In our implementation the parameters m_v, m_p, and m_λ steering the histories of the HADES quantities $\tilde{v}_k$, $\tilde{\mathbf{p}}_k$ and $\check{\mathbf{p}}_k$, and λ_k, respectively, can be individually chosen.

Note that the extrapolation schemes described above and in Sec. III B have also been implemented for PFFs, which express atomic polarizations by Gaussian inducible dipoles, thereby increasing the timescales and system sizes accessible to PFF-MD simulations with IPHIGENIE (Ref. 51, e.g., provides a sample application).

IV. METHODS

We studied the properties of our HADES implementation by MD simulations of the simple model di-peptide Ac-Ala-NHMe, whose chemical structure is depicted in Fig. 10 of Paper I.[12] The parameters required for the MM and HADES descriptions of this model peptide are described in Sec. IV D of Paper I. Beyond the CHARMM22/CMAP force field,[1,52] the MD simulations usually applied the velocity Verlet[53,54] algorithm with a time step $\Delta t = 1$ fs for the integration of the equations of motion (38). The lengths of covalent bonds involving H atoms were kept fixed by the M-SHAKE[55] algorithm.

As a check how the tightness of the convergence parameters χ_v, χ_p, and χ_λ affects energy conservation, we simulated Ac-Ala-NHMe several times for time spans $\tau = 10$ ns by HADES-MD ($\varepsilon_c = 80$) and once by a reference MD ($\varepsilon_c = 1$). The HADES-MD simulations differed in the choice of the convergence parameters and are denoted by $\mathcal{S}(\chi_v, \chi_p, \chi_\lambda)$. All simulations started at the same state belonging to the average temperature $\langle T\rangle_\tau = 280$ K. Energies were saved every 100 fs. Starting from the tight initial guesses $\chi_v^{\mathrm{ini}} = 10^{-14}$, $\chi_p^{\mathrm{ini}} = 10^{-14}$ D, and $\chi_\lambda^{\mathrm{ini}} = 10^{-14}$ kcal/mol the convergence parameters were successively loosened until beyond certain values $\chi_v^{\max}$, $\chi_p^{\max}$, and $\chi_\lambda^{\max}$ violations of energy conservation became apparent. Here, the parameters $\tilde{v}_i$, $\tilde{p}_i$, $\check{\mathbf{p}}_i$, and λ_i were brought to self-consistency in the given sequence to enable separate studies of the associated convergence thresholds. In HADES-MD simulations all these parameters are simultaneously iterated.

To estimate a possible energy drift, each trajectory was divided into 20 segments i of equal duration $\tau(i) = \tau/20$. Then average total energies

$$\langle E\rangle_{\tau(i)} = \langle K + U + \tilde{W}^{\mathrm{RF}}\rangle_{\tau(i)}, \tag{58}$$

where K denotes the kinetic energy, were calculated. The drift $\dot{Q}$ and its statistical uncertainty were then estimated from these data by linear regression.

Applying the optimal convergence parameters $\chi_v^{\max}$, $\chi_p^{\max}$, and $\chi_\lambda^{\max}$, we carried out six independent 100 ns replica exchange[56–58] (RE) HADES-MD simulations of Ac-Ala-NHMe sampling the dihedral angles φ and ψ at the central C_α every 0.5 ps. With the optimized[59] temperature ladder 300 K, 360 K, 432 K, 519 K, 624 K, and 750 K replica exchanges were attempted every ps using the "deterministic even-odd" (DEO) scheme.[60] Employing the second order integration algorithm specified in Ref. 61 (with $\gamma = 1/\mathrm{ps}$) the equations of motion were extended toward a Langevin dynamics. For each temperature T of the ladder we calculated a 60 × 60 bin histogram $p(\varphi, \psi)$ from the sampled angles (φ, ψ) to obtain the free energy landscape

$$\Delta G(\varphi, \psi) = -k_{\mathrm{B}} T \ln p(\varphi, \psi) - G_0 \tag{59}$$

of Ac-Ala-NHMe at T, where k_{B} is the Boltzmann constant and G_0 serves to set the minimum of ΔG to zero. For an improved low-energy resolution sparsely populated bins were identified through the upper energy cutoff $\Delta G_{\max} = 30$ kJ/mol.

For a direct comparison to explicit solvent data we carried out NVT simulations of a Ac-Ala-NHMe molecule surrounded by 1363 TIP3P water molecules[9] within a periodic rhombic dodecahedron of volume 40.3 nm^3. Long-range electrostatic interactions were treated with the SAMM$_4$ algorithm[36] employing minimum image truncation and a Kirkwood type reaction field correction with a dielectric constant $\varepsilon = 80$ for the surrounding medium.[11] Enhanced sampling was achieved by the "simulated solute tempering" (SST) method,[62] which effectively tempers only the Ac-Ala-NHMe solute and keeps the solvent at the reference temperature 300 K by scaling the TIP3P interaction potentials. In the SST/MD simulations we chose a temperature ladder with five rungs at 300 K, 356 K, 424 K, 504 K, and 600 K, for which short preparatory simulations yielded the required dimensionless weights of 0.0, 6.5, 11.6, 15.3, and 17.9.[62] During SST/MD we conducted every 100 integration steps one DEO exchange attempt. A canonical temperature distribution was obtained by a Bussi thermostat with a coupling time of $\tau = 0.1$ ps.[63] The production trajectories started from seven independent initial conditions and accumulated a total simulation length of 105 ns, from which we sampled Ac-Ala-NHMe coordinates every 1 ps. During SST/MD the system resided about 20% of the simulation time at each of the rungs resulting in a total of 21 045 configurations at 300 K, from which we obtained a free energy landscape by the same procedure as described for the HADES-RE/MD simulations above.

V. RESULTS

A numerical integration of a Hamiltonian dynamics with the velocity Verlet algorithm does not exactly conserve the energy. Instead, for a given time step Δt the energy is[64]

$$E(t \mid \Delta t) = \langle E(t \mid \Delta t)\rangle_\tau + f(t)\Delta t^2 + \mathcal{O}(\Delta t^4), \tag{60}$$

where $\langle E(t \mid \Delta t)\rangle_\tau$ is a temporal average over a reasonably large time span τ, which converges in the analytical limit $\Delta t \to 0$ to the exact energy E_0, and where the shape function $f(t)$ of the fluctuations has the property $\lim_{\tau\to\infty}\langle f(t)\rangle_\tau = 0$. Then the fluctuations of $E(t \mid \Delta t)$ around the average $\langle E(t \mid \Delta t)\rangle_\tau$ are

$$\Delta E(t \mid \Delta t) = f(t)\Delta t^2 + \mathcal{O}(\Delta t^4). \tag{61}$$

If one integrates the Hamiltonian dynamics starting at identical initial conditions but using different time steps (e.g., Δt_0 and $\Delta t > \Delta t_0$), then the scaled energy fluctuations

$$\delta E(t \mid \Delta t', \Delta t) = (\Delta t/\Delta t')^2 \Delta E(t \mid \Delta t') \tag{62}$$

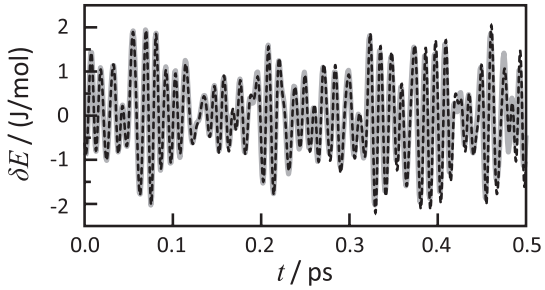

FIG. 3. HADES energy fluctuations $\Delta E(t \mid \Delta t)$ of Ac-Ala-NHMe calculated with $\Delta t = 0.1$ fs (black, dotted) are compared with the scaled fluctuations $\delta E(t \mid \Delta t', \Delta t)$ calculated with $\Delta t' = 10\Delta t$ (gray).

have according to Eq. (61) the almost identical shapes

$$\delta E(t \mid \Delta t', \Delta t) = f(t)\Delta t^2 + \mathcal{O}'(\Delta t^4). \tag{63}$$

Conversely one concludes that the dynamics integrated by the velocity Verlet algorithm is Hamiltonian, if for $\Delta t' > \Delta t$ the fluctuations (61) closely match the scaled fluctuations (63).

Figure 3 shows such a comparison for the HADES dynamics of Ac-Ala-NHMe for the time steps $\Delta t = 0.1$ fs and $\Delta t' = 10\Delta t$ combined with the tight convergence criteria. The almost perfect match of the two curves proves that HADES-MD is actually energy conserving.

Moreover, the HADES-MD simulation $\mathcal{S}(\chi_v^{\text{ini}}, \chi_p^{\text{ini}}, \chi_\lambda^{\text{ini}})$ of Ac-Ala-NHMe yielded through the linear regression for the average total energies $\langle E \rangle_{\tau(i)}$ defined by Eq. (58) an energy drift $\dot{Q}$ of -0.012 ± 0.006 kJ/(mol ns), while the vacuum MD simulation of Ac-Ala-NHMe provided the reference value $\dot{Q} = +0.046 \pm 0.013$ kJ/(mol ns). Hence, HADES-MD conserves the energy as well as plain MD as long as the convergence parameters are extremely tight.

Note here that GB methods like GB using molecular volume (GBMV)[16] or GB with simple switching (GBSW)[17,65] which employ complex estimates of the effective Born radii for improved[29] approximations, do not render a Hamiltonian dynamics. Instead they are quite noisy and add considerable amounts of heat to the simulated system.[65]

A. Choice of convergence parameters

The noted energy conservation within the limits of numerical accuracy prevails for a while, even if the convergence parameters are successively loosened. Because a loosening of these parameters considerably enhances the speed of HADES-MD, we have checked which maximal values are still compatible with a numerically accurate energy conservation. The parameters were loosened in the sequence χ_v, χ_p, and χ_λ. Figure 4 shows the energy drifts $\dot{Q}$ resulting from the associated simulations $\mathcal{S}(\chi_v, \chi_p, \chi_\lambda)$.

According to Fig. 4(a) the energy drift $\dot{Q}$ stays near the typical values signifying numerical accuracy as long as the parameter χ_v steering the convergence of the volume iteration (6) is below $\chi_v^{\max} = 10^{-6}$. It becomes abruptly large above 10^{-5}, where it signifies considerable algorithmic noise heating the molecule. Similar jumps are visible in Figs. 4(b) and

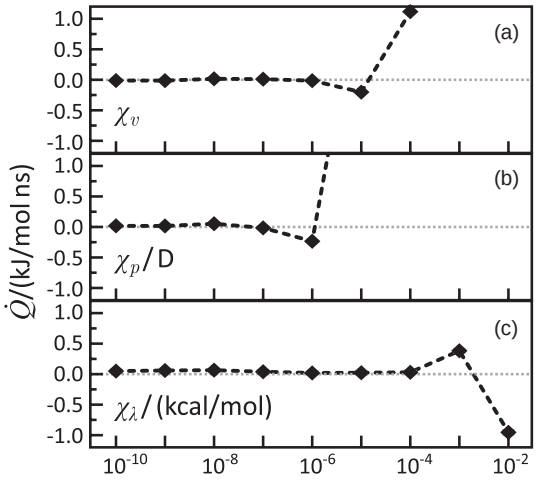

FIG. 4. HADES-MD energy drifts $\dot{Q}$ calculated for Ac-Ala-NHMe through linear regressions from the average energies (58); the estimated statistical uncertainties are too small to be visible on the given scale. (a) $\mathcal{S}(\chi_v, \chi_p^{\text{ini}}, \chi_\lambda^{\text{ini}})$, (b) $\mathcal{S}(\chi_v^{\max}, \chi_p, \chi_\lambda^{\text{ini}})$, and (c) $\mathcal{S}(\chi_v^{\max}, \chi_p^{\max}, \chi_\lambda)$.

4(c) above $\chi_p^{\max} = 10^{-7}$ D and $\chi_\lambda^{\max} = 10^{-4}$ kcal/mol. Hence, for Ac-Ala-NHMe the maximal values of the convergence parameters are well-defined and we assume that they can be transferred to other molecules.

B. Efficiency issues

The loosening of the convergence criteria from χ^{ini} to $\chi^{\max}$ entails for HADES-MD of Ac-Ala-NHMe a speed-up by a factor four. If one additionally applies the DIIS algorithm[44–46] to the self-consistency iterations the speed-up becomes seven. Finally, employing also the extrapolation of initial values described in Sec. III C enhances the speed-up factor to 12. Furthermore, the implementation shows even for such a small system like Ac-Ala-NHMe, which has only 22 atoms, a very good MPI (message passing interface) parallelization behavior. Using four cores of a single CPU (central processing unit) instead of only one yields a speed-up factor 25. Then HADES-MD of Ac-Ala-NHMe is by a factor of 100 faster than the explicit solvent MD simulation of Ac-Ala-NHMe in TIP3P water (using four cores also in this case). If instead a much more complex and accurate polarizable water model like TL5P[51] is employed as the solvent, then the HADES-MD speed-up is a factor of 600. On the other end, a complete neglect of the solvent, i.e., a vacuum MD simulation of Ac-Ala-NHMe is about a factor of ten faster than a HADES-MD simulation of this molecule.

C. Free energy landscape of Ac-Ala-NHMe

The HADES-RE/MD setup described in Sec. IV was applied to compute free energy landscapes $\Delta G^{\text{H}}(\varphi, \psi)$ as defined by Eq. (59) at the temperatures T covered by the RE temperature ladder. This RE setup resulted in a mean exchange rate of 36% between replicas at neighboring rungs of the ladder. Within the 600 ns of HADES-RE/MD each replica completed on the average 2376 so-called round trips[59] up and

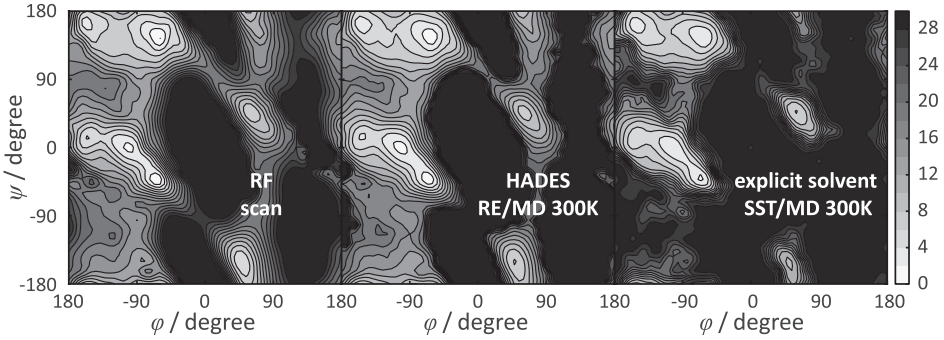

FIG. 5. The free energy landscape $\Delta G^{\mathrm{H}}(\varphi, \psi)$ of Ac-Ala-NHMe resulting at $T = 300$ K from HADES-RE/MD (center) is compared with the RF energy scan $\Delta G^{\mathrm{RF}}[\mathbf{R}(\varphi, \psi)]$ (left) presented in Paper I[12] and the explicit solvent MD free energy landscape $\Delta G^{\mathrm{E}}(\varphi, \psi)$ (right). The contour levels represent steps of 2 kJ/mol. The color bar indicates the ΔG scale in kJ/mol.

down the entire ladder. Thus, each replica took quite often advantage of the accelerated dynamics at 750 K. The 105 ns explicit solvent SST/MD simulation achieved a larger mean exchange rate of 55% due to the enhanced exchange probability of SST.[59,62]

Because Ac-Ala-NHMe is a quite stiff molecule featuring only two essential degrees of conformational freedom (φ, ψ), the free energy landscape $\Delta G^{\mathrm{H}}(\varphi, \psi)$ obtained by HADES-RE/MD at $T = 300$ K should be very similar to the one calculated in Paper I by our RF method for selected minimum energy configurations $\mathbf{R}(\varphi, \psi)$ embedded in a dielectric continuum $\varepsilon_c = 80$. With the MM energy function $U(\mathbf{R})$ and the reaction field energy $\tilde{W}^{\mathrm{RF}}(\mathbf{R})$ defined by Eq. (22) the corresponding free energy expression $\Delta G^{\mathrm{RF}}(\varphi, \psi)$ follows from $G^{\mathrm{RF}}(\mathbf{R}) \equiv \tilde{W}^{\mathrm{RF}}(\mathbf{R}) + U(\mathbf{R})$ by subtracting the smallest value $G_{\min}$ encountered in the set $G^{\mathrm{RF}}[\mathbf{R}(\varphi, \psi)]$ of energy values.

Figure 5 compares $\Delta G^{\mathrm{RF}}[\mathbf{R}(\varphi, \psi)]$ (left) with the 300 K HADES-RE/MD result $\Delta G^{\mathrm{H}}(\varphi, \psi)$ (center). As expected, the two landscapes are very similar. They feature local minima of almost identical shapes at the same locations within the (φ, ψ) plane. Solely some of the high energy barriers connecting minima are better resolved in the free energy scan $\Delta G^{\mathrm{RF}}[\mathbf{R}(\varphi, \psi)]$ than in $\Delta G^{\mathrm{H}}(\varphi, \psi)$. In contrast, landscapes $\Delta G^{\mathrm{H}}(\varphi, \psi)$ obtained from replicas at higher temperatures (data not shown) resolve also these barriers while preserving the structures and locations of the various local minima. Thus, they almost perfectly match $\Delta G^{\mathrm{RF}}[\mathbf{R}(\varphi, \psi)]$.

Also in terms of energy values the differences between the 300 K landscape $\Delta G^{\mathrm{H}}(\varphi, \psi)$ and $\Delta G^{\mathrm{RF}}[\mathbf{R}(\varphi, \psi)]$ are very small. For instance, the global minimum of $\Delta G^{\mathrm{H}}(\varphi, \psi)$ is located at $(\varphi_\alpha, \psi_\alpha) = (-66^\circ, -42^\circ)$ in the α-helical region while the lowest local minimum of the extended β-sheet region at $(\varphi_\beta, \psi_\beta) = (-60^\circ, 144^\circ)$ is found slightly above at 0.20 kJ/mol. For $\Delta G^{\mathrm{RF}}[\mathbf{R}(\varphi, \psi)]$ the global minimum is at $(\varphi_\beta, \psi_\beta)$ and the α-helical conformation $(\varphi_\alpha, \psi_\alpha)$ is located 0.4 kJ/mol above.[12] Note that the energetic locations of the two local minima found in $\Delta G^{\mathrm{H}}(\varphi, \psi)$ happen to be closer to the results of the landscape calculated in Paper I[12] by numerically solving the Poisson equation on a hierarchy of nested grids.[26,27]

The close match of $\Delta G^{\mathrm{RF}}[\mathbf{R}(\varphi, \psi)]$ and $\Delta G^{\mathrm{H}}(\varphi, \psi)$ apparent in Fig. 5 highlights the key progress achieved by our thoroughly revised continuum approach and the associated HADES-MD, because a similar but much less convincing comparison has been presented before with Figs. 6(B) and 6(E) in Ref. 34. While the quoted comparison served the authors (ST) as a "proof of principle" it concurrently demonstrated the incompatibility of the applied free energy function with the approximate force expression used in MD.

Fig. 6 of Ref. 34 additionally demonstrates the large parameter dependence of the free energy surfaces obtained with the predecessor methods either by scanning the configurations $\mathbf{R}(\varphi, \psi)$ or by MD simulation (the parameters mentioned here are the widths σ_i of the Gaussian atoms). This claim follows by visual comparison of Figs. 6(B) and 6(C). Our revised approach, in contrast, hardly depends on these parameters as has been demonstrated by Figs. S23 and S24 in the supplementary material of Paper I.[12]

Finally, the explicit solvent MD free energy landscape $\Delta G^{\mathrm{E}}(\varphi, \psi)$, which represents a target reference for our HADES-MD, is depicted in Fig. 5 on the right hand side. The free energy wells of the α-helical and extended β-sheet region display similar overall shapes and feature minima at the same positions as in ΔG^{H}, respectively. Also in ΔG^{E} the global minimum is found in the α-helical region. Its extended β minimum lies 0.7 kJ/mol above. In the region of α_{L}-like structures a third minimum is found at $\Delta G^{\mathrm{E}}(60^\circ, 48^\circ) = 4.4$ kJ/mol for the explicit solvent MD and at $\Delta G^{\mathrm{H}}(60^\circ, 48^\circ) = 6.6$ kJ/mol for HADES-MD. Thus, their difference is larger than for the two main minima but is still within an quite acceptable error range of $k_{\mathrm{B}}T \approx 2.5$ kJ/mol. The minima in the fourth well of the free energy landscape near $(48^\circ, -150^\circ)$ at $\Delta G^{\mathrm{H}}(48^\circ, -150^\circ) = 5.5$ kJ/mol and $\Delta G^{\mathrm{E}}(48^\circ, -156^\circ) = 7.9$ kJ/mol have an equally small difference. The transition regions between the minima appear slightly more rugged in ΔG^{E} and the associated barriers appear to be larger that in ΔG^{H}. However, also the statistical error of ΔG^{E} is larger due to the about 30 fold smaller data set.

With educated guesses for the atomic widths σ_i and hardly any further parameter optimization[12] the key features of the Ac-Ala-NHMe free energy landscape obtained by the HADES simulations match those of the explicit solvent simulations within an error range of $k_{\mathrm{B}}T$. We expect that optimized σ_i will bring these free energy surfaces even closer together.

VI. SUMMARY AND OUTLOOK

With Eq. (55) we have derived an effective RF energy function $\check{W}^{\mathrm{RF}}(\mathbf{R})$, which parametrically depends on the atomic volumes $\tilde{v}_i$, RF dipoles $\tilde{\mathbf{p}}_k$, Lagrangian dipoles $\check{\mathbf{p}}_k$, and Lagrangian multipliers λ_k. These parameters are self-consistently calculated by iterating Eqs. (6), (15), (32), and (36), respectively. The HADES forces (46) are then obtained by $\mathbf{f}_k^{\mathrm{RF}} = -\nabla_k \check{W}^{\mathrm{RF}}(\mathbf{R})$ as negative gradients of this RF energy function. As is illustrated in Sec. V, the associated equations of motion (38) therefore represent a Hamiltonian dynamics.

With a suitable implementation, like the one sketched in Sec. III, this dynamics can be employed for computationally quite efficient HADES-MD simulations. For the solvated model di-peptide Ac-Ala-NHMe, for instance, the computation of free energy landscapes with HADES-MD is by about two orders of magnitude faster than with a simple explicit solvent model like TIP3P. For more accurate and, hence, more expensive solvent models the computational gain is correspondingly larger.

The efficiency of HADES is currently determined by the number of iterations required to reach self-consistency. Even employing extrapolation and DIIS schemes the HADES-MD of Ac-Ala-NHMe required on average 8.9 SCF iterations to converge. If one could adapt algorithms that require the equivalence of a single iteration, such as the extended Lagrange technique[66] that has already been applied to polarizable force fields[67,68] or the "always stable predictor corrector" (ASPC) method,[69] HADES-MD could be sped up by almost another order of magnitude. Because of the four sets of variables, i.e., the $\tilde{v}_i$, $\tilde{\mathbf{p}}_i$, $\check{\mathbf{p}}_i$, and λ_i, which have to be jointly brought to self-consistency for HADES, such methods will require additional considerations. For instance it may be advantageous for the stability and efficiency of such algorithms to apply them only to the RF dipoles $\tilde{\mathbf{p}}_i$ and Lagrange dipoles $\check{\mathbf{p}}_i$, which depend on computationally expensive long range interactions, and determine the computationally cheap $\tilde{v}_i$ and λ_i still self-consistently. Note, furthermore, that an extended Lagrange scheme may additionally serve to model a delayed dielectric relaxation,[34,70] which ranges up to several picoseconds for water, by suitably chosen mass-parameters for the dipoles.

On the other hand HADES-MD, just like other electrostatic continuum methods, neglects van der Waals contributions to the solvent-solute interaction energy and contributions of the solvent entropy to the free energy of solvation, which are both included in explicit solvent simulations. Therefore, one will have to complement HADES with surface and volume dependent free energy contributions, see, e.g., Refs. 71 and 72 and references therein. Here, the Gaussian volume representation (3) of the solute molecule could prove advantageous for the computation of the necessary solvent accessible surface area for each atom, if one adapts the ideas of Refs. 73 and 74.

As was demonstrated in Paper I of this work, our RF approach and, therefore, also its HADES-MD version remedy key conceptual and practical problems hampering the popular[21] GB descriptions of solute peptides and proteins. Moreover, HADES-MD enables for peptides embedded in dielectric continua Hamiltonian simulations, which are excluded in GB simulations employing variable Born radii.[65]

However, due to the limited experience with applications of HADES-MD to solvated peptides, its practical merits (and possible limitations) cannot yet be comprehensively judged. Clearer insights will require extended test simulations and refined choices of the free parameters, i.e., the widths σ_i of the Gaussian atoms employed for the model construction. Quite clearly, however, it is the first method, which enables accurate solutions of the PE for peptides embedded in dielectric continua leading directly to Hamiltonian MD simulations.

ACKNOWLEDGMENTS

This work was supported by the Deutsche Forschungsgemeinschaft (SFB749/C4).

APPENDIX: ELECTROSTATIC POTENTIALS AND FIELDS

The computations of the RF energies and dipoles require analytical expressions for the electrostatic potentials, fields, and field gradients of Gaussian charges and dipoles. The electrostatic potential at the position $\mathbf{r}_i$ of an atom i, which is generated by a Gaussian charge distribution of width σ_j, strength q_j, and centered at $\mathbf{r}_j$ is

$$\phi(\mathbf{r}_i \mid \mathbf{r}_j, q_j, \sigma_j) = q_j\, \mu(r_{ij}, \sigma_j) \tag{A1}$$

with the distance dependent function

$$\mu(r, \sigma) \equiv (1/r)\,\mathrm{erf}(r/\sqrt{2}\sigma). \tag{A2}$$

The potential generated by a Gaussian dipole density of strength $\mathbf{p}_j$ and width σ_j, which is localized at $\mathbf{r}_j$, is

$$\phi(\mathbf{r}_i \mid \mathbf{r}_j, \mathbf{p}_j, \sigma_j) = (\mathbf{p}_j \cdot \mathbf{r}_{ij})\,\kappa(r_{ij}, \sigma_j)/r_{ij}^2 \tag{A3}$$

with

$$\kappa(r, \sigma) \equiv \mu(r, \sigma) - \sigma^2 \eta(r, \sigma) \tag{A4}$$

and

$$\eta(r, \sigma) \equiv \sqrt{\frac{2}{\pi}}\frac{1}{\sigma^3}\exp\left(-\frac{r^2}{2\sigma^2}\right). \tag{A5}$$

The corresponding fields are

$$\mathbf{E}(\mathbf{r}_i \mid \mathbf{r}_j, q_j, \sigma_j) = q_j \mathbf{r}_{ij}\,\kappa(r_{ij}, \sigma_j)/r_{ij}^2 \tag{A6}$$

and

$$\mathbf{E}(\mathbf{r}_i \mid \mathbf{r}_j, \mathbf{p}_j, \sigma_j) = [\kappa(r_{ij}, \sigma_j)\,\mathbf{G}(\mathbf{r}_{ij}) - \eta(r_{ij}, \sigma_j)\,\mathbf{A}(\mathbf{r}_{ij})]\,\mathbf{p}_j \tag{A7}$$

with the tensorial functions

$$\mathbf{A}(\mathbf{r}) = (\mathbf{r} \otimes \mathbf{r})/r^2 \tag{A8}$$

and

$$\mathbf{G}(\mathbf{r}) = (3\mathbf{A} - \mathbf{I})/r^2. \tag{A9}$$

The field gradient of a Gaussian charge is then given by

$$\nabla\mathbf{E}(\mathbf{r}_i \mid \mathbf{r}_j, q_j, \sigma_j) = -q_j[\kappa(r_{ij}, \sigma_j)\mathbf{G}(\mathbf{r}_{ij}) - \eta(r_{ij}, \sigma_j)\mathbf{A}(\mathbf{r}_{ij})]. \tag{A10}$$

Finally, with

$$\mathbf{B}(\mathbf{r},\mathbf{p}) \equiv \mathbf{A}(\mathbf{r})\,(\mathbf{p}\cdot\mathbf{r}) \tag{A11}$$

and

$$\mathbf{H}(\mathbf{r},\mathbf{p}) \equiv \frac{1}{r}[5\mathbf{A}(\mathbf{r},\mathbf{p}) - \mathbf{p}\otimes\mathbf{r} - \mathbf{r}\otimes\mathbf{p} - (\mathbf{p}\cdot\mathbf{r})\,\mathbf{I}] \tag{A12}$$

the field gradient of the Gaussian dipole is

$$\begin{aligned}\nabla\mathbf{E}(\mathbf{r}_i \mid \mathbf{r}_j,\mathbf{p}_j,\sigma_j) &= -\{\left[3\kappa(r_{ij},\sigma_j)/r_{ij}^2 - \eta(r_{ij},\sigma_j)\right]\mathbf{H}(\mathbf{p}_{ij},\mathbf{r}_{ij}) \\ &\quad + \eta(r_{ij},\sigma_j)\mathbf{B}(\mathbf{r}_{ij},\mathbf{p}_{ij})/\sigma_j^2\}. \end{aligned}\tag{A13}$$

[1] A. D. MacKerell, 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, J. Phys. Chem. B **102**, 3586 (1998).
[2] D. Pearlman, D. Case, J. Caldwell, W. Ross, T. Cheatham III, S. DeBolt, D. Ferguson, G. Seibel, and P. Kollman, Comput. Phys. Commun. **91**, 1 (1995).
[3] M. Christen, P. H. Hünenberger, D. Bakowies, R. Baron, R. Bürgi, D. P. Geerke, T. N. Heinz, M. A. Kastenholz, V. Kräutler, C. Oostenbrink, C. Peter, D. Trzesniak, and W. F. van Gunsteren, J. Comput. Chem. **26**, 1719 (2005).
[4] A. Warshel and S. T. Russell, Q. Rev. Biophys. **17**, 283 (1984).
[5] W. F. van Gunsteren and H. J. C. Berendsen, Angew. Chem. Int. Ed. Engl. **29**, 992 (1990).
[6] M. Karplus and J. A. McCammon, Nat. Struct. Biol. **9**, 646 (2002).
[7] J. W. Ponder and D. A. Case, Adv. Protein Chem. **66**, 27 (2003).
[8] P. Tavan, H. Carstens, and G. Mathias, in *Protein Folding Handbook. Part 1*, edited by J. Buchner, and T. Kiefhaber (Wiley-VCH, Weinheim, 2005), pp. 1170–1195.
[9] W. L. Jorgensen, J. Chandrasekhar, and J. D. Madura, J. Chem. Phys. **79**, 926 (1983).
[10] H. Berendsen, J. Grigera, and T. Straatsma, J. Phys. Chem. **91**, 6269 (1987).
[11] G. Mathias, B. Egwolf, M. Nonella, and P. Tavan, J. Chem. Phys. **118**, 10847 (2003).
[12] S. Bauer, G. Mathias, and P. Tavan, J. Chem. Phys. **140**, 104102 (2014).
[13] W. C. Still, A. Tempczyk, R. C. Hawley, and T. Hendrickson, J. Am. Chem. Soc. **112**, 6127 (1990).
[14] M. Schaefer and M. Karplus, J. Phys. Chem. **100**, 1578 (1996).
[15] D. Bashford and A. Case, Ann. Rev. Phys. Chem. **51**, 129 (2000).
[16] M. S. Lee, F. R. Salsbury Jr., and C. L. Brooks III, J. Chem. Phys. **116**, 10606 (2002).
[17] W. Im, M. S. Lee, and C. L. Brooks III, J. Comput. Chem. **24**, 1691 (2003).
[18] T. Grycuk, J. Chem. Phys. **119**, 4817 (2003).
[19] M. Wojciechowski and B. Lesyng, J. Phys. Chem. B **108**, 18368 (2004).
[20] E. Gallicchio and R. M. Levy, J. Comput. Chem. **25**, 479 (2004).
[21] C. D. Snow, N. Nguyen, V. S. Pande, and M. Gruebele, Nature (London) **420**, 102 (2002).
[22] R. Zhou and B. J. Berne, Proc. Natl. Acad. Sci. U.S.A. **99**, 12777 (2002).
[23] H. Nymeyer and A. E. García, Proc. Natl. Acad. Sci. U.S.A. **100**, 13934 (2003).
[24] B. Egwolf and P. Tavan, J. Chem. Phys. **118**, 2039 (2003).
[25] M. E. Davis, J. D. Madura, B. A. Luty, and J. A. McCammon, Comput. Phys. Commun. **62**, 187 (1991).
[26] W. Rocchia, E. Alexov, and B. Honig, J. Phys. Chem. B **105**, 6507 (2001).
[27] A. Nicholls and B. Honig, J. Comp. Chem. **12**, 435 (1991).
[28] R. Yokota, J. P. Bardhan, M. G. Knepley, L. Barba, and T. Hamada, Comput. Phys. Commun. **182**, 1272 (2011).
[29] M. Feig, A. Onufriev, M. S. Lee, W. Im, D. A. Case, and C. L. Brooks III, J. Comput. Chem. **25**, 265 (2004).
[30] M. Stork and P. Tavan, J. Chem. Phys. **126**, 165105 (2007).
[31] N. Levy, D. Borgis, and M. Marchi, Comput. Phys. Commun. **169**, 69 (2005).
[32] Appendix contains explicit expressions for the potentials, fields, and field gradients of point or Gaussian charges and dipoles.
[33] J. D. Jackson, *Classical Electrodynamics* (John Wiley & Sons, Inc., New York, 1975).
[34] M. Stork and P. Tavan, J. Chem. Phys. **126**, 165106 (2007).
[35] J. V. José and E. J. Saletan, *Classical Dynamics: A Contemporary Approach* (Cambridge University Press, Cambridge, 1998).
[36] K. Lorenzen, M. Schwörer, P. Tröster, S. Mates, and P. Tavan, J. Chem. Theory Comput. **8**, 3628 (2012).
[37] M. Schwörer, B. Breitenfeld, P. Tröster, K. Lorenzen, P. Tavan, and G. Mathias, J. Chem. Phys. **138**, 244103 (2013).
[38] P. Ahlström, A. Wallquist, S. Engström, and B. Jönsson, Mol. Phys. **68**, 563 (1989).
[39] A. A. Appel, SIAM J. Sci. Stat. Comput. **6**, 85 (1985).
[40] J. Barnes and P. Hut, Nature (London) **324**, 446 (1986).
[41] L. Greengard and V. Rokhlin, J. Comput. Phys. **73**, 325 (1987).
[42] I. N. Bronstein, K. A. Semendjajew, G. Musiol, and H. Mühlig, *Taschenbuch der Mathematik* (Verlag Harri Deutsch, Frankfurt am Main, 1999).
[43] B. I. Kvasov, *Methods of Shape Preserving Spline Approximation* (World Scientific, Singapore, 2000).
[44] P. Pulay, Chem. Phys. Lett. **73**, 393 (1980).
[45] C. Császár and P. Pulay, J. Mol. Struct. **114**, 31 (1984).
[46] D. Marx and J. Hutter, *Ab Initio Molecular Dynamics* (Cambridge University Press, Cambridge, 2009), pp. 116–117.
[47] Nearest neighbor lists are efficiently generated in the FMM computation of the long-range interactions by a downward movement along a hierarchical tree decomposition of the simulation system; for HADES they are extended by interaction partners, which in MM force fields are usually excluded by the so-called 1-3 criterion identifying covalently bound atoms; furthermore they are split by a Gaussian distance cutoff into nearby Gaussian and more distant point-like interaction partners.
[48] C. Niedermeier and P. Tavan, J. Chem. Phys. **101**, 734 (1994).
[49] C. Niedermeier and P. Tavan, Mol. Simul. **17**, 57 (1996).
[50] M. Eichinger, H. Grubmüller, H. Heller, and P. Tavan, J. Comput. Chem. **18**, 1729 (1997).
[51] P. Tröster, K. Lorenzen, M. Schwörer, and P. Tavan, J. Phys. Chem. B **117**, 9486 (2013).
[52] A. D. MacKerell, Jr., M. Feig, and C. L. Brooks III, J. Comput. Chem. **25**, 1400 (2004).
[53] L. Verlet, Phys. Rev. **159**, 98 (1967).
[54] W. C. Swope, H. C. Andersen, P. H. Berens, and K. R. Wilson, J. Chem. Phys. **76**, 637 (1982).
[55] V. Kräutler, W. F. van Gunsteren, and P. H. Hünenberger, J. Comput. Chem. **22**, 501 (2001).
[56] K. Hukushima and K. Nemoto, J. Phys. Soc. Japan **65**, 1604 (1996).
[57] U. H. E. Hansmann, Chem. Phys. Lett. **281**, 140 (1997).
[58] Y. Sugita and Y. Okamoto, Chem. Phys. Lett. **314**, 141 (1999).
[59] R. Denschlag, M. Lingenheil, and P. Tavan, Chem. Phys. Lett. **473**, 193 (2009).
[60] M. Lingenheil, R. Denschlag, G. Mathias, and P. Tavan, Chem. Phys. Lett. **478**, 80 (2009).
[61] E. Vanden-Eijnden and G. Ciccotti, Chem. Phys. Lett. **429**, 310 (2006).
[62] R. Denschlag, M. Lingenheil, P. Tavan, and G. Mathias, J. Chem. Theory Comput. **5**, 2847 (2009).
[63] G. Bussi and M. Parrinello, Comput. Phys. Commun. **179**, 26 (2008).
[64] M. E. Tuckerman, *Statistical Mechanics: Theory and Molecular Simulation*, 1st ed. (Oxford University Press, New York, 2010), Chap. 3.13, pp. 121–124.
[65] J. Chocholousova and M. Feig, J. Comput. Chem. **27**, 719 (2006).
[66] R. Car and M. Parrinello, Phys. Rev. Lett. **55**, 2471 (1985).
[67] M. Sprik and M. L. Klein, J. Chem. Phys. **89**, 7556 (1988).
[68] S. W. Rick, S. J. Stuart, and B. J. Berne, J. Chem. Phys. **101**, 6141 (1994).
[69] J. Kolafa, J. Comput. Chem. **25**, 335 (2004).
[70] I. G. Tironi, B. A. Luty, and W. F. van Gunsteren, J. Chem. Phys. **106**, 6068 (1997).
[71] J. R. Allison, K. Boguslawski, F. Fraternali, and W. F. van Gunsteren, J. Phys. Chem. B **115**, 4547 (2011).
[72] J. Kleinjung, W. R. P. Scott, J. R. Allison, W. F. van Gunsteren, and F. Fraternali, J. Chem. Theory Comput. **8**, 2391 (2012).
[73] J. A. Grant and B. T. Pickup, J. Phys. Chem. **99**, 3503 (1995).
[74] J. Weiser, P. S. Shenkin, and W. C. Still, J. Comput. Chem. **20**, 688 (1999).

3 Anwendung

If you try and take a cat apart to see how it works, the first thing you have on your hands is a nonworking cat.

Douglas Adams,
The Salmon of Doubt

In Kapitel 2 wurde mit HADES eine Methode entwickelt, die MD Simulationen von Peptiden und Proteinen in dielektrischen Kontinua erlaubt. Die in Abschnitt 2.3 abgedruckte Arbeit beweist die Hamilton'sche Natur der Methode und zeigt anhand des Ace-Ala-NHMe, eines Minimalmodells für Peptide, die Effizienz von HADES. Der dabei verwendete Parametersatz, der die Größen von Atomen misst, wurde einfach geschätzt und nicht für Peptidsimulationen optimiert.

3.1 Erste Versuche mit HADES-MD

In der Anschlussarbeit

„Exploring Hamiltonian Dielectric Solvent Molecular Dynamics"
Sebastian Bauer, Paul Tavan, and Gerald Mathias
Chem. Phys. Lett, **612**, 20 (2014)

die ich zusammen mit Paul Tavan und Gerald Mathias verfasst habe, wird gezeigt, dass die in Abschnitt 2.3 beschriebene Effizienz von HADES auch bei etwas größeren Systemen bestehen bleibt. Als Testsystem dient uns dabei ein α-helikales, 150 Atome umfassendes Dekapeptid, dessen Schmelzkurve schon früher von Reichold in seiner Dissertation [108] mit *Replica Exchange with Solute Tempering* MD-Simulationen unter Verwendung eines expliziten Modells für das Lösungsmittel berechnet worden war und das sich daher zum Vergleich mit HADES Replika-Austausch-MD-Simulationen anbot. Hier zeigte sich eine um etwa den Faktor 20 höhere Abtastleistung von HADES-MD. Ferner lieferte auch der geschätzte Parametersatz schon eine recht ähnliche Schmelzkurve. HADES erwies sich auch, im Gegensatz zu GB Methoden, als deutlich weniger sensibel gegenüber der Parameterwahl. Andererseits zeigt die Arbeit aber auch, dass eine sorgfältige Optimierung der Parameter den zwingend erforderlichen nächsten Schritt darstellt.

Chemical Physics Letters 612 (2014) 20–24

Contents lists available at ScienceDirect

Chemical Physics Letters

journal homepage: www.elsevier.com/locate/cplett

Exploring Hamiltonian dielectric solvent molecular dynamics

Sebastian Bauer, Paul Tavan, Gerald Mathias *

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

ARTICLE INFO

Article history:
Received 30 April 2014
In final form 28 July 2014
Available online 4 August 2014

ABSTRACT

Hamiltonian dielectric solvent (HADES) is a recent method [7,25], which enables Hamiltonian molecular dynamics (MD) simulations of peptides and proteins in dielectric continua. Sample simulations of an α-helical decapeptide with and without explicit solvent demonstrate the high efficiency of HADES-MD. Addressing the folding of this peptide by replica exchange MD we study the properties of HADES by comparing melting curves, secondary structure motifs and salt bridges with explicit solvent results. Despite the unoptimized *ad hoc* parametrization of HADES, calculated reaction field energies correlate well with numerical grid solutions of the dielectric Poisson equation.

1. Introduction

A major factor, which limits the conformational sampling of peptides and proteins by molecular dynamics (MD) simulations, is the explicit inclusion of the structure determining aqueous solvent. If one describes such solute–solvent systems by standard all-atom molecular mechanics (MM) force fields [1–3], then the solvent atoms typically outnumber the solute atoms by at least an order of magnitude [4–6]. Therefore, instead of focusing on the solute–solute and solute–solvent interactions, one has to spend most of the computational effort on the calculation of the interactions between the water molecules.

A large part of this effort would be saved, if the surrounding water could be replaced by a continuum model that needs to be computationally inexpensive and physically correct. This task demands the solution of the dielectric Poisson equation (PE) on the fly with the numerical integration of the protein dynamics [7]. A corresponding approach neglects, of course, the dielectric relaxation [8] (femtoseconds to picoseconds) of the water and its structure near a protein surface. Whereas the former approximation may be of minor importance for proteins, because their conformational dynamics proceeds on much slower time scales (>nanoseconds), the significance of the latter is still unclear [9] and can be assessed only if a continuum approach, which meets the above criteria, is available. In this context, complementary information from hybrid approaches, which interpolate between an explicit and a continuum description by using a few layers of explicit solvent, may be helpful [10].

There have been many attempts to construct such a continuum approach for MD simulations. However, as demonstrated in Ref. [7], all these attempts essentially represent failures (see also corresponding discussions in Refs. [11–15]). Atomic forces derived from numerical solutions of the PE [16,17], for instance, do not comply with Newton's reaction principle, because they do not yield exact atomic reaction forces, and, therefore, violate energy conservation. On the other hand, a free energy functional approach [18], which actually yields a Hamiltonian dynamics, turned out to be slower [14] than explicit solvent simulations. Finally, the popular generalized Born methods (see e.g. Refs. [19–21]) do not solve the PE [7,20,22].

Starting with the reformulation of the PE [15,23,24], which replaces the polarization of the surrounding continuum by an anti-polarization within the solute protein, we recently succeeded [7,25] to construct a continuum approach for MM-MD simulations, which actually meets the above requirements and is called 'Hamiltonian dielectric solvent' (HADES).

Like all continuum methods also the HADES reaction field (RF) energies and forces, which summarize the interaction of the protein charges with the continuum, depend on the description of the effective atomic volumes v_i [9,26], which collectively define the space $\mathfrak{V}_S$ occupied by the protein and, hence, separate the interior region characterized by a small dielectric constant ε_S from the exterior region characterized by the large dielectric constant ε_C of the solvent continuum. Therefore the parameters σ_i, which, in HADES, steer the atomic sizes [7], have to be carefully chosen, if one wants to achieve realistic descriptions. In the preceding works [7,25], however, only an *ad hoc* parametrization has been provided. Moreover, the contended computational efficiency has solely been demonstrated by HADES-MD simulations on a small dipeptide.

* Corresponding author.
E-mail address: gerald.mathias@physik.uni-muenchen.de (G. Mathias).

http://dx.doi.org/10.1016/j.cplett.2014.07.070
0009-2614/

With this contribution we want to demonstrate that the alleged computational efficiency pertains also to larger peptides. Furthermore, we want to check to what extent the preliminary *ad hoc* choice [7] Λ of the atomic size parameters σ_i affects the conformational landscape of such a model system.

As a sample system we chose the α-helical decapeptide Ace-AAAKEAAAKK-NH_2, which we call $\mathcal{P}$ from now on, because the melting curve of its CHARMM22/ CMAP [1,27] model has been previously studied [28] by replica exchange [29,30] (RE) and by replica exchange with solute tempering [31,32] (REST) MD simulations in explicit water, which was described by the 'three point transferable intermolecular potential' [33] (TIP3P). Note that the chosen sequence contains four charged residues that can form internal salt bridges, which may cause difficulties in continuum approaches featuring improper descriptions of atomic volumes [9]. Hence, comparisons of HADES-RE-MD on $\mathcal{P}$ with the previous explicit solvent descriptions [28] can indicate to what extent our first choice Λ of the volume parameters σ_i was reasonable.

Whereas standard non-polarizable MM force fields are known to exhibit distinct conformational preferences [6,34], HADES should be impartial in this respect. Thus the choice of an α-helical model peptide was solely dictated by the much larger computational effort required for a statistically meaningful explicit solvent sampling of β-hairpin folding-unfolding equilibria.

2. HADES continuum electrostatics

In HADES the protein volume $\mathfrak{V}_s$ is described by the characteristic function

$$\Theta(\mathbf{r}) = \begin{cases} 1, & \text{if } \mathbf{r} \in \mathfrak{V}_s \\ 0, & \text{else} \end{cases} \tag{1}$$

and is decomposed [24] by local characteristic functions $\vartheta_i(\mathbf{r})$ into atomic volumes $\mathfrak{v}_i$. The atomic functions $\vartheta_i(\mathbf{r})$ are centered around the positions $\mathbf{r}_i$ of the atoms i and obey $\sum_i \vartheta_i(\mathbf{r}) = \Theta(\mathbf{r})$ for all $\mathbf{r}$. Using this partition, the PE has been exactly transformed [7] into the new representation

$$\Delta\Phi(\mathbf{r}) = -\frac{4\pi}{\varepsilon_s}\sum_i \left[q_i\delta(\mathbf{r}-\mathbf{r}_i) + \hat{\rho}_i(\mathbf{r}) - \nabla\cdot\hat{\mathbf{P}}_i(\mathbf{r})\right]. \tag{2}$$

Here, the atomic partial charges q_i, which generate the Coulomb potential $\Phi^{\mathrm{C}}(\mathbf{r})$, are clearly separated from the sources $\hat{\rho}_i(\mathbf{r})$ and $\hat{\mathbf{P}}_i(\mathbf{r})$ of the RF potential $\Phi^{\mathrm{RF}}(\mathbf{r})$, which is the difference between the solutions $\Phi(\mathbf{r}|\varepsilon_s, \varepsilon_c)$ of the PE for $\varepsilon_s \neq \varepsilon_c$ and $\varepsilon_s = \varepsilon_c$. The sources of $\Phi^{\mathrm{RF}}(\mathbf{r})$ are the atomic shielding charge distributions $\hat{\rho}_i(\mathbf{r})$ and antipolarization densities $\hat{\mathbf{P}}_i(\mathbf{r})$, which are both confined to the atomic volumes $\mathfrak{v}_i$. In the representation (2) the usual boundary conditions imposed to $\Phi(\mathbf{r})$ are replaced by the conditions

$$\hat{q}_i = -q_i(1-\varepsilon_s/\varepsilon_c) \tag{3}$$

for the strengths (volume integrals) $\hat{q}_i$ of the $\hat{\rho}_i(\mathbf{r})$ and by the self-consistency conditions

$$\hat{\mathbf{p}}_i = -\frac{v_i\varepsilon_s}{4\pi}\left(1-\frac{\varepsilon_s}{\varepsilon_c}\right)\langle\mathbf{E}\rangle_{\mathfrak{v}_i} \tag{4}$$

for the strengths $\hat{\mathbf{p}}_i$ of the $\hat{\mathbf{P}}_i(\mathbf{r})$, where $v_i = \int \vartheta_i(\mathbf{r})d^3r$ denotes the atomic volumes and $\langle\mathbf{E}\rangle_{\mathfrak{v}_i} = (1/v_i)\int \vartheta_i(\mathbf{r})[-\nabla\Phi(\mathbf{r})]d^3r$ the atomic field averages.

The exact representation (2) of the PE is not very useful by itself for computations, because the exact characteristic functions $\vartheta_i(\mathbf{r})$ are difficult to handle analytically or numerically. However, it provides an excellent starting point to derive simple and computationally tractable approximations. Following a previous suggestion [24] we introduced the Gaussian models $\tilde{\vartheta}_i(\mathbf{r}\,|\,\mathbf{r}_i, \tilde{v}_i, \sigma_i)$ for the exact atomic shape functions $\vartheta_i(\mathbf{r})$. These atomic models are specified by the approximate volumes $\tilde{v}_i$ and by the Gaussian widths σ_i. Whereas the $\tilde{v}_i$ are fixed by a self-consistency condition [24], the σ_i are the parameters of HADES. The introduction of the $\tilde{\vartheta}_i(\mathbf{r}\,|\,\mathbf{r}_i, \tilde{v}_i, \sigma_i)$ immediately leads to Gaussian approximations $\tilde{\rho}_i(\mathbf{r}\,|\mathbf{r}_i, \hat{q}_i, \hat{\sigma}_i)$ and $\tilde{\mathbf{P}}_i(\mathbf{r}\,|\,\mathbf{r}_i, \tilde{\mathbf{p}}_i, \sigma_i)$ also for the $\hat{\rho}_i(\mathbf{r})$ and $\hat{\mathbf{P}}_i(\mathbf{r})$. Here, the strengths $\tilde{\mathbf{p}}_i$ of the Gaussian models $\tilde{\mathbf{P}}_i(\mathbf{r}\,|\,\mathbf{r}_i, \tilde{\mathbf{p}}_i, \sigma_i)$ have to fulfill a self-consistency condition analogous to Eq. (4), whereas the exact condition (3) identically applies to the strengths $\hat{q}_i$ of the models $\tilde{\rho}_i(\mathbf{r}\,|\mathbf{r}_i, \hat{q}_i, \hat{\sigma}_i)$, whose widths $\hat{\sigma}_i$ differ by a factor in the range $1.2 \le \zeta \le 1.6$ from the σ_i. This factor is another HADES parameter [7].

As a result of the Gaussian approximations, the self-consistent HADES computation of the approximate RF potential $\tilde{\Phi}^{\mathrm{RF}}(\mathbf{r}_i)$ is essentially analogous to that of a potential, which is generated by induced dipoles of a polarizable force field [7]. One immediately obtains an analytically tractable approximation

$$\tilde{W}^{\mathrm{RF}}(\mathbf{R}) = \sum_i (q_i/2)\,\tilde{\Phi}^{\mathrm{RF}}(\mathbf{r}_i) \tag{5}$$

for the electrostatic contribution to the solvation free energy, which enables the derivation [25] of explicit expressions for the atomic RF forces. Note that these forces obey Newton's third law and enable Hamiltonian MD simulations [25].

3. Simulation setups

Issues of computational efficiency were addressed by comparing for the α-helical decapeptide $\mathcal{P}$ introduced above the timings of five different MD simulation setups. Three of them employed our MD package IPHIGENIE [25,35–37]. The explicit solvent setup I consisted of the 150 atom peptide $\mathcal{P}$ solvated by 1809 TIP3P water models adding up to a total of $N = 5577$ atoms enclosed in a periodic orthorhombic dodecahedron of inner radius $R_i = 21.1$ Å. This carefully equilibrated system (temperature $T = 300$ K, volume $V = 53.1\ \mathrm{nm}^3$) was adopted from Reichold [28] and is, together with his REST-MD and RE-MD simulations on explicitly solvated $\mathcal{P}$, described in Sections S1 and S2 of the Supporting Information (SI). The setups II and III applied HADES- and vacuum-MD, respectively, to the thermal motion of $\mathcal{P}$'s 150 atoms. The setups IV and V used the NAMD program [38], which offers [39] a popular GB method [40,41]. They applied GB- (IV) and vacuum-MD (V) to $\mathcal{P}$. All timing simulations were carried out on a single core of a 4×16 core 2.5 GHz AMD Opteron 6274 workstation.

In all simulations $\mathcal{P}$ was described by the CHARMM22/CMAP force field [1,27]. Bond lengths involving hydrogen atoms were constrained by the MSHAKE algorithm [42]. In setup I the dynamics was integrated in the *NVT* ensemble by the velocity Verlet algorithm [43,44] with a time step $\Delta t = 1$ fs and the solvent was coupled to a Berendsen thermostat (coupling constant $\tau = 0.5$ ps, target temperature $T_0 = 300$ K). The long-range electrostatic and dispersive interactions were treated by the combination of the fast structure-adapted multipole method with a RF correction, which is called $\mathrm{SAMM}_{4,1}$/RF [5,35,37]. In the continuum setups II and IV the dielectric constants were chosen as $\varepsilon_c = 80$ and $\varepsilon_s = 1$. In the setups II and III a Langevin dynamics was applied using a second order integration algorithm [45] (with $T_0 = 300$ K, $\gamma = 1$/ps, $\Delta t = 1$ fs). Related temperature control methods were applied in the NAMD setups IV and V.

HADES-MD simulations were carried out with the two *ad hoc* choices Λ and Λ' for the Gaussian atomic widths σ_i listed in Table S2 of the SI. Λ distinguishes [7] only the four atom types $X \in \{\mathrm{H}, \mathrm{O}, \mathrm{N}, \mathrm{C}\}$, to which it assigns values σ_X in the range 0.52–0.87 Å. In Λ' the widths σ_X of the carboxylate oxygens in E and of the amino hydrogens in K are reduced by the factor 0.8 to generate an enhanced solubility. All other HADES parameters and convergence criteria

were adopted from Ref. [25]. We have checked the energy conservation of these setting for HADES on a 10 ns *NVE* trajectory employing the procedures described in Ref. [25] and find a negligible average heat production of $\dot{Q} = 0.17 \pm 0.67\,\mathrm{kJ\,mol^{-1}\,ns^{-1}}$.

4. Efficiency of HADES

As compared to the time consumed by the vacuum-MD for a given time span in setup III, HADES-MD (setup II) was by a factor of 20 and explicit solvent MD (setup I) by a factor of 460 slower, such that the sampling with HADES is 23 times faster than with explicit solvent. Similarly, GB-MD of setup IV was 5 times slower than vacuum-MD of setup V. Hence, for $\mathcal{P}$ the sampling speed of GB is by the factor 4 larger than that of HADES. This disadvantage of HADES-MD can be partially alleviated by favorable MPI parallelization properties. Despite the small size of $\mathcal{P}$ one gets, e.g. speedups by the factors 1.9 or 3.2, respectively, if one employs 2 and 4 cores instead of only one.

Note furthermore that GB truncates the electrostatic and dispersive interactions at finite distances (16–20 Å). HADES does not need such a cutoff, because it employs $\mathrm{SAMM}_{4,1}$, which scales linearly with the number of atoms [37]. Therefore, the 20 times larger computational effort of HADES as compared to vacuum-MD should be independent of peptide size. A similar size-invariance should apply to its 20-fold computational advantage over explicit solvent MD. In contrast, as stated by Tanner et al. [39], for 'large systems and long cutoffs the computational expense' of GB-MD 'can exceed that of explicit solvent simulations' [39].

We are currently preparing a release of our MD code IPHIGENIE, which will be freely available at https://sourceforge.net/projects/iphigenie/ under the GNU General Public License. In a concomitant publication we will demonstrate the linear scaling and the excellent parallelization properties of the code and describe further improvements of our HADES implementation.

5. α-Helix melting with HADES

The apparent efficiency of HADES-MD and the available RE- and REST-MD prediction [28] on the helix melting profile of $\mathcal{P}$ in TIP3P water (see Sections S1 and S2 of the SI) inspired us to recalculate this profile for $\mathcal{P}$ from three HADES-RE-MD simulations, spanning the range from 200 K to 540 K with an optimal [46] temperature ladder of 16 rungs. Just like Reichold's REST-MD simulations (cf. Section S1.4 in the SI) also here the first two 25 ns simulations started with sets of replicas, which were either all completely folded (simulation $\mathcal{P}$.H.α) or all completely unfolded ($\mathcal{P}$.H.u).

Section S3 of the SI demonstrates that the conformational sampling provided by these HADES-RE-MD simulations rapidly converged. This conclusion is proven by comparing the average α-helix fractions $\langle H_\alpha\rangle_\tau(T)$ of $\mathcal{P}$, which were extracted from the trajectories $\mathcal{P}$.H.α and $\mathcal{P}$.H.u at the RE temperature rungs T. The fractions represent averages over time spans $\tau = 5$ ns and were measured by the dictionary of protein secondary structure [47] (DSSP) using the procedures specified in Section S1.5 of the SI. After 10 ns the two ensembles showed similar helix fractions, which were independent of the different initial conditions.

For the subsequent production trajectory $\mathcal{P}$.H.α/u comprising 30 ns of HADES-RE-MD the initial conditions were randomly selected from the converged ensembles of the first two simulations and the α-helix fraction profile $\langle H_\alpha\rangle_\tau(T)$ of $\mathcal{P}$ was averaged over the last $\tau = 20$ ns of this trajectory. Figure 1 compares the resulting profile (diamonds) with the corresponding profile [28] (circles) from REST-MD with explicit TIP3P solvent. Note that the REST-MD profile $\langle H_\alpha\rangle_\tau(T)$ (cirlces) represents a very good approximation to a physical melting curve, despite the scaling of the solvent–solvent and solvent–solute interactions employed for efficiency reasons in the REST technique [31,32]. The high quality of this approximation is demonstrated in Section S2.2 of the SI by comparison with RE-MD simulations. The REST-MD melting curve (circles) predicts for the CHARMM/CMAP model of $\mathcal{P}$ in TIP3P water an α-helix melting point at about 366 K and a 90% helix content at 300 K.

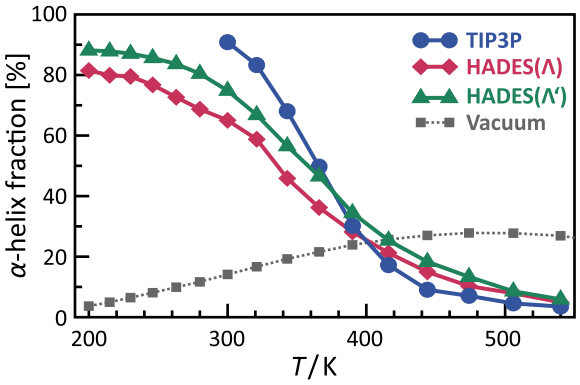

Fig. 1. Diamonds: HADES-RE-MD temperature dependence of $\mathcal{P}$'s α-helix fraction $\langle H_\alpha\rangle_\tau(T)$ averaged over the last 20 ns of the simulation $\mathcal{P}$.H.α/u executed with the speculative [7] parameter set Λ; triangles: modified parameter set Λ'; circles: explicit TIP3P solvent REST-MD result from Reichold's [28] thesis; squares: vacuum RE-MD.

Also the α-helix fraction profile (diamonds) obtained from HADES-RE-MD with the same model for $\mathcal{P}$ features the monotonous sigmoid decline toward higher temperatures T and, therefore, signifies α-helix melting. However, HADES-RE-MD assigns the melting point to 335 K and reaches a helix content of at most 82% at 200 K. The fact that one gets a melting curve at all is a first encouraging sign as is demonstrated by the dotted α-helix fraction profile, which was obtained by averaging over the last $\tau = 50$ ns of a 100 ns vacuum-RE-MD spanning the range $T \in [200, 908]$ K by 24 replicas (squares). Due to the existence of stable salt bridges, which are incompatible with the formation of an α-helix (e.g. E_5–K_{10}), the accumulation of a substantial helix fraction (28%) requires for the isolated $\mathcal{P}$ a lot of heat.

6. Salt bridges in HADES

As is well known from GB models of peptide folding [9] salt bridges can be destabilized by reducing the Born radii of the ions. Now one may ask, whether a similar overestimate of ion-pair stability, which has severely hampered GB simulations of peptide folding–unfolding equilibria [9], might be the cause for the difference between the HADES-RE-MD (diamonds) and explicit solvent REST-MD (circles) melting curves in Figure 1. Therefore, we have repeated the 30 ns simulation $\mathcal{P}$.H.α/u with the slightly modified parameter set Λ'. This set features 20% reduced Gaussian widths σ_i at the carboxyl oxygens of E and amino hydrogens of K (cf. Table S2 in the SI), which should entail an enhanced solubility of these ionic residues. We obtained the helix melting profile depicted in Figure 1 by triangles, which actually exhibits a higher melting temperature of 360 K and an 84% helix fraction already at 263 K. Thus, the reduction of Gaussian widths σ_i at E and at the Ks actually moves the structural ensemble predicted by HADES-RE-MD closer to the TIP3P REST-MD ensemble.

Now the question arises, to what extent the three salt bridges E_5–K_9, E_5–K_{10}, E_5–K_4, which may form in $\mathcal{P}$, are actually affected by the parameter change $\Lambda \to \Lambda'$. For a quantitative estimate we have calculated for all these salt bridges from the HADES-RE-MD and REST-MD-trajectories free energy profiles $\Delta G(d)$ at $T = 300$ K as functions of the distance d between the carboxylate carbon C_δ of E_5 and the amino nitrogens N_ζ of the respective K_i. These functions

Table 1
Frequency[a] of salt bridges at $T = 300$ K

Solvent model	E_5–K_9	E_5–K_{10}	E_5–K_4
TIP3P	11.1%	1.5%	1.8%
HADES/Λ	9.5%	6.0%	4.0%
HADES/Λ'	6.7%	2.6%	2.3%

[a] A salt bridge is formed, if $d[C_\delta(E_5), N_\zeta(K_i)] < 4.5$ Å.

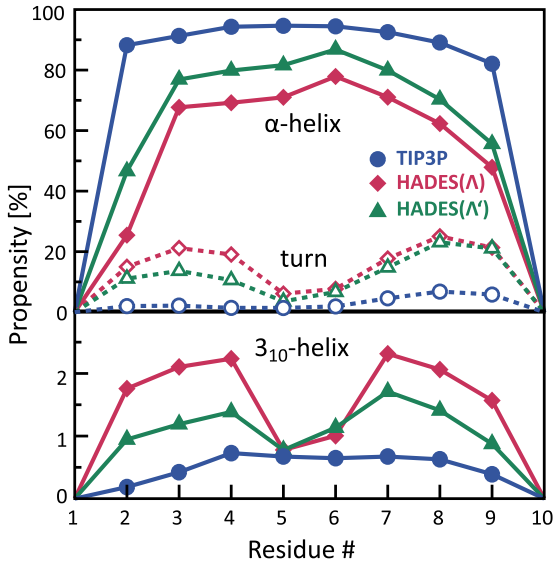

Fig. 2. Propensities at 300 K for local secondary structure motifs as measured by DSSP [47] at each residue and as predicted for $\mathcal{P}$ in TIP3P solvent by REST-MD (circles) or by HADES-RE-MD with the parameter sets Λ (diamonds) and Λ' (triangles); top: α-helix (solid symbols) and turn (open symbols); bottom: 3_{10}-helix.

are depicted in Figure S8 and discussed in Section S4 of the SI. At distances d larger than equilibrium distance $d_S \equiv 3.8$ Å of such a salt bridge, the continuum profiles $\Delta G(d)$ markedly differ from the TIP3P results, which clearly exhibit the molecular granularity of this solvent model through local maxima at about 4.5 Å. Assuming that distances $d \leq 4.5$ Å signify the formation of a salt bridge, we have calculated at 300 K relative frequencies for the three possibilities enumerated above.

The salt bridge frequencies listed in Table 1 show that the reduction of the Gaussian widths σ_i at the ionic groups, which accompanies the $\Lambda \to \Lambda'$ transition actually reduces the frequencies of all three salt bridges. As compared to the explicit TIP3P solvent case, this is favorable for the less frequent salt bridges E_5–K_{10} and E_5–K_4 and less favorable for E_5–K_9, which is the energetically preferred salt bridge in TIP3P solvent (see Figure S8 in the SI). Hence, we see that the stability of salt bridges can be actually steered through the Gaussian widths of atoms in ionic groups. Furthermore, we see that the initial crude guess Λ of the σ_i was not too bad either in this respect. In contrast to GB methods, HADES does not show a similarly extreme sensitivity to the choice of ion size parameters (cf. e.g. Figure 10 in Ref. [9]). As a result it is still unclear, whether an overestimated salt-bridge stability is the real reason for the improvement of the α-helix melting curve upon exchange of Λ by Λ'.

But whatever the reason may be, Figure 2 demonstrates that the parameter change $\Lambda \to \Lambda'$ significantly changes at 300 K all local propensities for secondary structure motives, which can be measured for each residue by DSSP [47]. For all motifs and at all residues the $\Lambda \to \Lambda'$ transition is seen to bring the various HADES-RE-MD propensities (transition from diamonds to triangles) closer to the explicit solvent results (circles). These changes indicate that there is ample room for further improvements of HADES by a

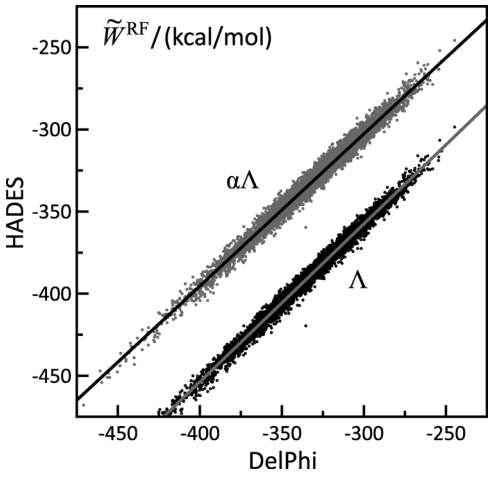

Fig. 3. Correlation of approximate RF energies (electrostatic solvation free energies) $\tilde{W}^{RF}(\mathbf{R}_n)$ (5) calculated for 112 000 structures $\mathbf{R}_n$ of $\mathcal{P}$ by DelPhi and HADES with the parameters Λ (black dots, gray regression line) and $\alpha\Lambda$ (gray dots, black line).

systematic and diligent optimization of the parameters Λ, which should proceed along the lines suggested by Stork and Tavan [48] for the predecessor method.

7. RF Energies from HADES

The quality of the first and *ad hoc* guess for the HADES parameters Λ can be seen by a different testing scenario, which solely addresses the energetics of the structures $\mathbf{R}_n$ sampled by the HADES-RE-MD simulations on $\mathcal{P}$ at all considered replicas T. For a corresponding demonstration we have collected a total of 112 000 such configurations $\mathbf{R}_n$ and have recomputed the approximate RF energy $\tilde{W}^{RF}(\mathbf{R}_n)$ with the program DelPhi [49,50], which represents the electrostatic potential on a hierarchy of cubic grids and solves the PE numerically. Here, the atomic van der Waals radii provided by the CHARMM22 force field [1] served to define the molecular volume $\mathfrak{V}_s$ of $\mathcal{P}$ (for further computational details see Section S5 of the SI).

Note here that HADES and DelPhi employ for $\mathcal{P}$ strongly different volume descriptions: HADES models $\mathfrak{V}_s$ by small overlapping Gaussians whereas DelPhi estimates $\mathfrak{V}_s$ through generally large overlapping hard spheres. Nevertheless, there should be a certain amount of correlation between the approximate RF energies $\tilde{W}^{RF}(\mathbf{R}_n)$ obtained by the two methods for the given set of structures.

Figure 3 shows the resulting correlation between the approximate RF energies $\tilde{W}^{RF}(\mathbf{R}_n)$ obtained by HADES/Λ and DelPhi (black dots). The two data sets are seen to be very well correlated as expressed by the correlation coefficient $r = 0.995$ and by the slope $b = 0.962$ of the indicated linear regression (gray line; for the definition of the quantities r and b see Eqs. (S3) and (S4) in Section S5.1 of the SI). This excellent match confirms once again that the approximations underlying HADES provide a quite accurate solution of the PE (2). The only difference to DelPhi is that HADES/Λ predicts the electrostatic contributions $W^{RF}(\mathbf{R}_n)$ to the free energies of solvation on average 56.3 kcal/mol below those calculated by DelPhi.

As is proven by the correlation labeled by $\alpha\Lambda$ in Figure 3 (gray dots, black line), this difference can be largely removed by globally scaling the parameter set Λ with the constant $\alpha = 1.22$, i.e. by homogeneously increasing the Gaussian widths σ_i of all atoms. Then the revised HADES RF energies overestimate their DelPhi counterparts on average by only 0.2 kcal/mol (a difference that could be brought essentially to zero by further optimization of α), whereas the

correlations deteriorate only very little as is witnessed by the values $r = 0.991$ and $b = 0.925$.

The above comparison of the DelPhi/HADES correlations measured with Λ and with its upscaled version $\alpha\Lambda$ has demonstrated that the changes of the RF energies, which are caused by altered configurations of a given peptide, hardly depend on the overall size of the Gaussian widths σ_i. Therefore, also the dynamically important RF forces should hardly depend on the overall size of the σ_i. Different choices of the scaling parameter α will lead to essentially the same RF dynamics.

8. Outlook

In contrast to the presumably weak parameter dependence of the RF forces, the total energies of solvation will sensitively depend on α and on all detailed choices σ_i. The parameter set Λ' with its enhanced solubility of $\mathcal{P}$'s ionic side groups, for instance, shifts the average solvation energy down by further 66.8 kcal/mol. As a result, a careful and diligent parameterization of HADES is the next task, which has to be tackled, to further develop this computationally efficient, scalable, and dynamically stable approach to the continuum dynamics of peptides and proteins into a method, which can accurately predict energies of solvation together with conformational landscapes.

Acknowledgement

This work has been supported by the Deutsche Forschungsgemeinschaft (SFB 749/C4). The authors thank Rudolf Reichold for providing the explicit solvent simulation data.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.cplett.2014.07.070.

References

[1] A.D. MacKerell Jr., et al., J. Phys. Chem. B 102 (1998) 3586.
[2] D. Pearlman, et al., Comput. Phys. Commun. 91 (1995) 1.
[3] M. Christen, et al., J. Comput. Chem. 26 (2005) 1719.
[4] W. Weber, P.H. Hünenberger, J.A. McCammon, J. Phys. Chem. B 104 (2000) 3668.
[5] G. Mathias, B. Egwolf, M. Nonella, P. Tavan, J. Chem. Phys. 118 (2003) 10847.
[6] P. Tavan, H. Carstens, G. Mathias, in: J. Buchner, T. Kiefhaber (Eds.), Protein Folding Handbook. Part 1, Wiley-VCH, Weinheim, 2005, p. 1170.
[7] S. Bauer, G. Mathias, P. Tavan, J. Chem. Phys. 140 (2014) 104102.
[8] R. Jimenez, G. Fleming, P. Kumar, M. Maroncelli, Nature 369 (1994) 471.
[9] R. Geney, M. Layten, R. Gomperts, V. Hornak, C. Simmerling, J. Chem. Theory Comput. 2 (2006) 115.
[10] G. Mancini, G. Brancato, V. Barone, J. Chem. Theory Comput. 10 (2014) 1150.
[11] R. Zhou, B.J. Berne, Proc. Natl. Acad. Sci. U.S.A. 99 (2002) 12777.
[12] H. Nymeyer, A.E. García, Proc. Natl. Acad. Sci. U.S.A. 100 (2003) 13934.
[13] M. Feig, A. Onufriev, M.S. Lee, W. Im, D.A. Case, J. Comput. Chem. 25 (2003) 265.
[14] N. Levy, D. Borgis, M. Marchi, Comput. Phys. Commun. 169 (2005) 69.
[15] M. Stork, P. Tavan, J. Chem. Phys. 126 (2007) 165105.
[16] W. Im, D. Beglov, B. Roux, Comput. Phys. Commun. 111 (1997) 59.
[17] W. Geng, G.W. Wei, J. Comput. Phys. 230 (2011) 435.
[18] M. Marchi, D. Borgis, N. Levy, P. Ballone, J. Chem. Phys. 114 (2001) 4377.
[19] D. Bashford, A. Case, Ann. Rev. Phys. Chem. 51 (2000) 129.
[20] T. Grycuk, J. Chem. Phys. 119 (2003) 4817.
[21] A. Fenley, J.C. Gordon, A. Onufriev, J. Chem. Phys. 129 (2008) 075101.
[22] A. Onufriev, D.A. Case, D. Bashford, J. Comput. Chem. 23 (2002) 1297.
[23] H. Sklenar, F. Eisenhaber, M. Poncin, R. Lavery, in: D.L. Beveridge, R. Lavery (Eds.), Theoretical Biochemistry & Molecular Biophysics, 2. Proteins, Adenine Press, New York, 1991, p. 317.
[24] B. Egwolf, P. Tavan, J. Chem. Phys. 118 (2003) 2039.
[25] S. Bauer, P. Tavan, G. Mathias, J. Chem. Phys. 140 (2014) 104103.
[26] M. Born, Zeit. Phys. 1 (1920) 45.
[27] A.D. MacKerell Jr., M. Feig, C.L. Brooks III, J. Comput. Chem. 25 (2004) 1400.
[28] R. Reichold, Rechnergestützte Beschreibung der Struktur und Dynamik von Peptiden und ihren Bausteinen, Fakultät für Physik, Ludwig-Maximilians-Universität, 2009 (Dissertation).
[29] U.H.E. Hansmann, Chem. Phys. Lett. 281 (1997) 140.
[30] Y. Sugita, Y. Okamoto, Chem. Phys. Lett. 314 (1999) 141.
[31] P. Liu, B. Kim, R.A. Friesner, B.J. Berne, Proc. Natl. Acad. Sci. U.S.A. 102 (2005) 13749.
[32] R. Denschlag, M. Lingenheil, P. Tavan, G. Mathias, J. Chem. Theory Comput. 5 (2009) 2847.
[33] W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, M.L. Klein, J. Chem. Phys. 79 (1983) 926.
[34] J.W. Ponder, D.A. Case, Adv. Protein Chem. 66 (2003) 27.
[35] K. Lorenzen, M. Schwörer, P. Tröster, S. Mates, P. Tavan, J. Chem. Theory Comput. 8 (2012) 3628.
[36] M. Schwörer, B. Breitenfeld, P. Tröster, K. Lorenzen, P. Tavan, G. Mathias, J. Chem. Phys. 138 (2013) 244103.
[37] K. Lorenzen, P. Tavan, J. Chem. Theory Comput. (2014), http://dx.doi.org/10.1021/ct500319a.
[38] NAMD was developed by the Theoretical and Computational Biophysics Group in the Beckman Institute for Advanced Science and Technology at the University of Illinois at Urbana-Champaign; J.C. Phillips et al., J. Comp. Chem. 26 (2005) 1781.
[39] D.E. Tanner, K.-Y. Chan, J.C. Phillips, K. Schulten, J. Chem. Theory Comput. 7 (2011) 3635.
[40] A. Onufriev, D. Bashford, D.A. Case, J. Phys. Chem. B 104 (2000) 3712.
[41] A. Onufriev, D. Bashford, D.A. Case, Proteins: Struct. Funct. Bioinf. 55 (2004) 383.
[42] V. Kräutler, W.F. van Gunsteren, P.H. Hünenberger, J. Comp. Chem. 22 (2001) 501.
[43] L. Verlet, Phys. Rev. 159 (1967) 98.
[44] W.C. Swope, H.C. Andersen, P.H. Berens, K.R. Wilson, J. Chem. Phys. 76 (1982) 637.
[45] E. Vanden-Eijnden, G. Ciccotti, Chem. Phys. Lett. 429 (2006) 310.
[46] R. Denschlag, M. Lingenheil, P. Tavan, Chem. Phys. Lett. 473 (2009) 193.
[47] W. Kabsch, C. Sander, Biopolymers 22 (1983) 2577.
[48] M. Stork, P. Tavan, J. Chem. Phys. 126 (2007) 165106.
[49] W. Rocchia, E. Alexov, B. Honig, J. Phys. Chem. B 105 (2001) 6507.
[50] A. Nicholls, B. Honig, J. Comput. Chem. 12 (1991) 435.

3.2 Zusatzinformationen zur Peptidfaltung mit HADES

Das Manuskript

„Supporting Information to:
Exploring Hamiltonian Dielectric Solvent Molecular Dynamics“
Sebastian Bauer, Paul Tavan, and Gerlad Mathias
Chem. Phys. Lett, **612**, 20 (2014)

enthält zusäztliche Informationen zur Peptidfaltung mit dem HADES Modell, die nicht im Haupttext enthalten sind. Dabei werden zum einen die im Haupttext verwendeten Methoden und Ergebnisse der Dissertation von Reichold [108] zusammengefasst, zum anderen werden Fragen der Konvergenz der HADES-MD Simulationen und der Salzbrückenstabilität im HADES Modell behandelt und die für HADES verwendeten Parameter aufgelistet.

Supporting Information: Exploring Hamiltonian Dielectric Solvent Molecular Dynamics

Sebastian Bauer, Paul Tavan, Gerald Mathias*

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

S1. Methods employed by Reichold [1]

Here we sketch the setup of the MD simulation systems used by Reichold [1] to characterize the conformational ensemble of the peptide $\mathcal{P}$ in aqueous solution at various temperatures. After a specification of the MD methods we provide the equilibration procedures preceding the simulations serving for data acquisition. Next, the various types and details of these data production runs are explained and listed. Finally we outline the observables used in the statistical data analysis for the identification of secondary structures.

S1.1. Setup of the Simulation System

An ideal α-helical model of $\mathcal{P}$ was built with the program MOLDEN 4.0 [2]. This α-helix was defined by choosing the values $(-58°, -47°)$ for the pair of dihedral angles (ϕ, ψ) which characterize the backbone geometry at the C_α atoms. To enable MD simulations of $\mathcal{P}$ in explicit solvent, a periodic box shaped as orthorhombic dodecahedron was filled with water molecules modeled by the CHARMM22 variant [3, 4] of Jorgensen's [5] three point transferable interaction potential (TIP3P). The periodic water box was equilibrated by a MD simulation in the NpT ensemble controlling the temperature T and pressure p by tightly coupled Berendsen [6] thermo- and barostats (targets: $T_0 = 300$ K, $p_0 = 1.0$ bar). Now, the peptide model was transferred into the center of the box, and water molecules overlapping with $\mathcal{P}$ were removed. Subsequently, the system was equilibrated toward T_0 and p_0 in a series of MD simulations during which the initial restraints on $\mathcal{P}$ were gently and slowly relaxed, such that $\mathcal{P}$ essentially retained its initial α-helical structure. As a result [1] the inner radius of the box was 21.1 Å and contained 1809 water models surrounding the 150 atoms of the α-helical peptide $\mathcal{P}$.

*Corresponding author.
Email address: gerald.mathias@physik.uni-muenchen.de (Gerald Mathias)

S1.2. Simulation Methods

The MD simulations were performed with the program package EGO [7, 8], which is the predecessor of IPHIGENIE [9–12], $\mathcal{P}$ was described by the CHARMM22/CMAP [4, 13] force field. The long-range electrostatic interactions were treated by a combination of fast structure-adapted multipole expansions [14, 15] with a moving boundary reaction-field approach and a multiple time step integrator [8]. Here, the default values [8] were chosen and the basic time step was set to $\Delta t = 1$ fs. The van der Waals interactions were calculated explicitly up to 10 Å while at larger distances a mean field approach [16] was applied. Covalent bonds involving hydrogen atoms were kept fixed by the M-SHAKE algorithm [17].

Whenever a different, statistically independent, and α-helical starting structure of $\mathcal{P}$ was required, a 100 ps NVT simulation was added at $T = 300$ K. In all these and subsequent simulations the Berendsen thermostat was exclusively coupled to the solvent following the suggestion of Ref. [18] for temperature control.

S1.3. Enhanced Sampling

To gain access to the α-helix folding and unfolding equilibria of $\mathcal{P}$ the conventional "temperature" replica exchange (RE) method [19–21] and the "replica exchange with solute tempering (REST)" technique [22] as reformulated by Denschlag et al. [23] were applied. For the given system, which contains only a few peptide degrees of freedom surrounded by large numbers of water molecules, REST can span a given temperature range (e.g. from 300 K to 500 K) with much fewer replicas than RE and, nevertheless, can guarantee a high exchange probability.

In contrast to the case of RE, the higher temperature REST replicas sample ensembles, which are associated with an energy function describing scaled solvent-solvent and solvent-solute interactions. Therefore these ensembles cannot accurately describe the physical behavior of the solvated peptide at elevated temperatures. Instead they describe something like a hot peptide (at T_i) in a cool solvent (at T_0).

To avoid the latter drawback of REST and, concurrently, to take advantage of its efficient sampling, a new

Table S1: Simulations

simulation[a]	sampling method
$\mathcal{P}.\alpha$	REST
$\mathcal{P}$.u	REST
$\mathcal{P}.\alpha$/u	RE/REST

[a]Names $\mathcal{P}.\eta$ of extended ensemble simulations, whose postfixes η characterize the conformations of starting structures: (α) α-helical, (u) unfolded, (α/u) mixture.

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

S1.4. Simulations

Table S1 specifies the simulations carried out by Reichold [1]. The table assigns to each simulation a unique name. Furthermore it characterizes the force field employed in the respective simulation for the peptide/water system, the sampling method applied, the durations and multiplicities of the trajectories collected under the given name, and finally the employed starting structures which have been mostly α-helical but sometimes also unfolded or a mixture of the two.

The setup of the 25 ns REST and RE/REST simulations listed in S1 deserves a more detailed characterization. The REST simulations employed an optimal [24] 10 rung temperature ladder spanning the range [300, 540] K. Exchanges between neighboring replicas were attempted every 10 ps. This setup resulted in exchange probabilities in the range between 30% and 40%.

For $\mathcal{P}$ in TIP3P we have carried out the two REST simulations $\mathcal{P}.\alpha$ and $\mathcal{P}$.u starting either with 10 independent α-helical starting structures or with 10 independent unfolded structures. By this setup we wanted to enable a check of the REST sampling convergence [25]. The required unfolded starting structures were randomly selected from the 540 K replicas of the simulation $\mathcal{P}.\alpha$.

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

S1.5. Secondary Structure Analysis and Helix Fraction

The local secondary structures of the peptides were measured by the dictionary of protein secondary structure (DSSP) [26], which classifies each residue into one of eight classes, i.e. as locally α- or π-helical, β-strand, etc. If N_α is the DSSP number of α-helical and N the total number of residues in a given structure then one may define the *helix fraction* by

$$H_\alpha = \frac{N_\alpha}{N-2}, \tag{S1}$$

which is normalized by $N-2$, because DSSP does not classify the first and last residue in a sequence. Evaluating H_α at every MD time step t gives a helix fraction trajectory $H_\alpha(t)$, whose average over a time span τ is denoted by $\langle H_\alpha \rangle_\tau$.

S2. Results of Reichold's simulations[1]

After demonstrating the safe convergence of the REST simulations $\mathcal{P}.\alpha$ and $\mathcal{P}$.u, it is shown that the REST results $\langle H_\alpha \rangle_\tau(T)$ are melting curves despite the applied scaling of the solvent-solute and solvent-solvent interactions.

S2.1. Convergence of REST Simulations

The convergence of REST-MD was checked by comparing relevant observables extracted from the two REST melting curves $\langle H_\alpha \rangle_\tau(T)$, which were obtained for $\mathcal{P}$ from the simulations $\mathcal{P}.\alpha$ and $\mathcal{P}$.u. The approach toward the folding-unfolding equilibrium was monitored by the mutual approach of the two helix contents $\langle H_\alpha \rangle_\tau(T_0)$ at the target temperature $T_0 = 300$ K of the REST sampling and by the average helix content $\langle\langle H_\alpha \rangle_\tau\rangle_T$ over the whole generalized REST ensemble. As soon as they match, the equilibrium has been reached and its sampling is sufficient. Remaining differences can indicate the degree of convergence and the quality of the statistics.

Figure S4 demonstrates that the REST simulations $\mathcal{P}.\alpha$ and $\mathcal{P}$.u are actually well converged both at the target temperature of 300 K (gray curves) and in the generalized ensemble (black curves). The convergence is proven by the fact that, after 25 ns, the helical content has become independent of the initial conditions in both ensembles. Apparently, the convergence to the respective equilibria has already been reached after about 10 ns. Therefore, averages over the last $\tau \equiv 15$ ns of our 25 ns REST simulations should represent equilibrium properties. Such averages are, of course, still subject to certain statistical fluctuations which could be diminished only upon a substantial

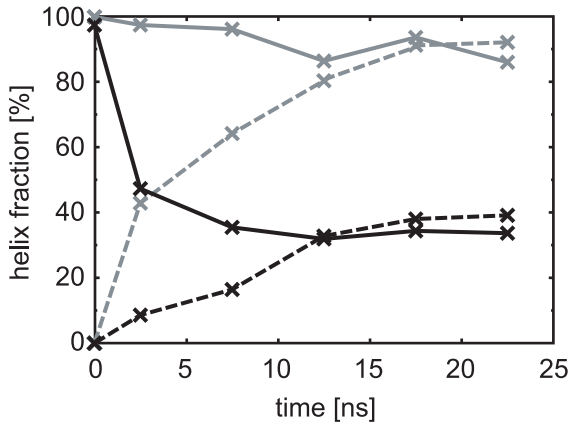

Figure S4: DSSP helix fractions $\langle H_\alpha \rangle_\tau$ averaged over periods τ = 5 ns within the two 25 ns REST simulations $\mathcal{P}.\alpha$ (solid curves) and $\mathcal{P}$.u (dashed curves) and over the residues of $\mathcal{P}$. The gray curves refer to $\langle H_\alpha \rangle_\tau(T_0)$ measured for the replica at T_0 = 300 K. The black curves are averages $\langle \langle H_\alpha \rangle_\tau \rangle_T$ over the generalized REST ensembles made up of all replicas T.

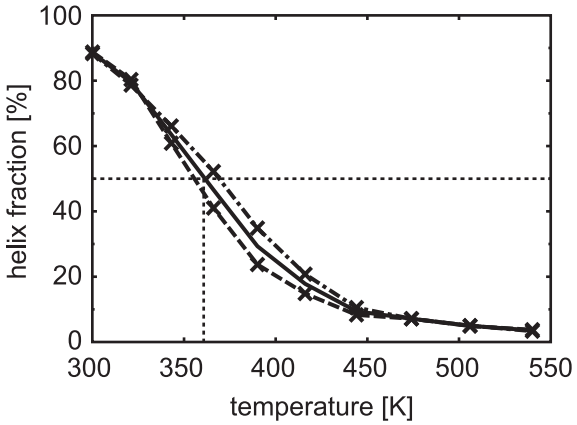

Figure S5: Equilibrium DSSP helix fractions $\langle H_\alpha \rangle_\tau(T)$ averaged over the last τ = 15 ns of the two 25 ns REST simulations $\mathcal{P}.\alpha$ (dashed curve) and $\mathcal{P}$.u (dot-dashed curve) and over the residues of $\mathcal{P}$. Also given is the average of the two results (solid curve).

extension of the sampling period starting after the initial 10 ns period of relaxation.

Figure S5 shows the equilibrium temperature dependences $\langle H_\alpha \rangle_\tau(T)$ of the α-helix fractions calculated from the simulations $\mathcal{P}.\alpha$ and $\mathcal{P}$.u together with the associated average. Interestingly, the dashed curve, which belongs to the simulation $\mathcal{P}.\alpha$ featuring α-helical initial structures, is found nearly everywhere below the dot-dashed curve belonging to the unfolded starting structures. The remaining small deviations thus indicate the considerable quality of the underlying statistics. The average curve $\langle H_\alpha \rangle_\tau(T)$ should be associated with an even smaller statistical uncertainty.

S2.2. Does REST Yield Melting Curves?

The temperature dependences $\langle H_\alpha \rangle_\tau(T)$ depicted in Figure S5 resemble a melting curve of an α-helical peptide. In fact, if a sufficiently extended RE sampling had been used instead of REST, the $\langle H_\alpha \rangle_\tau(T)$ would actually represent melting curves.

However, we did not employ RE but instead the more efficient REST sampling for the computation of $\langle H_\alpha \rangle(T)$ raising the question, to what extent the curves in Fig. S5 approximates the melting curve despite the REST energy scaling. For an answer we simply have to check whether a RE simulation yields a similar dependence. If this should turn out to be the case, the REST curves $\langle H_\alpha \rangle(T)$ would represent approximate melting curves.

For a corresponding check we performed simulation $\mathcal{P}.\alpha$/u with the RE/REST method introduced in section S1.3. Because peptide $\mathcal{P}$ had shown a rapid convergence toward equilibrium in the REST setting, we did not execute two simulations with opposite initial conditions in the related RE/REST approach. Instead, as starting structures we used peptide/water configurations drawn from the ensembles generated by the REST simulation $\mathcal{P}.\alpha$ (for details see Section S1.4).

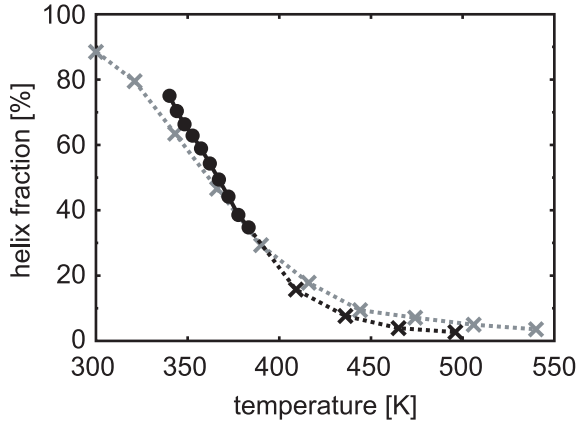

Figure S6: Temperature dependence of the equilibrium DSSP helix fractions $\langle H_\alpha \rangle_\tau(T)$ resulting from the RE/REST simulation $\mathcal{P}.\alpha$/u (black) and the REST simulations $\mathcal{P}.\alpha/\mathcal{P}$.u (gray). The values for RE replicas are indicated by dots, those for the REST replicas by crosses.

Figure S6 compares the RE/REST melting curve from simulation $\mathcal{P}.\alpha$/u with the REST result obtained from $\mathcal{P}.\alpha/\mathcal{P}$.u (Figure S5). Despite the scaling of solvent-solvent and peptide-solvent interactions applied in REST, the curves almost match. This finding indicates that the REST energy scaling has only a quite small influence on the helical content of $\mathcal{P}$.

Note here that simulations $\mathcal{P}.\alpha/\mathcal{P}$.u and $\mathcal{P}.\alpha$/u nicely illustrate the efficiency gain of REST vs. RE. In $\mathcal{P}.\alpha/\mathcal{P}$.u a temperature range of 240 K is covered by 10 replicas whereas in the RE part of $\mathcal{P}.\alpha$/u the same number of replicas covers only a range of 43 K. Because both setups operate with comparable exchange probabilities, the

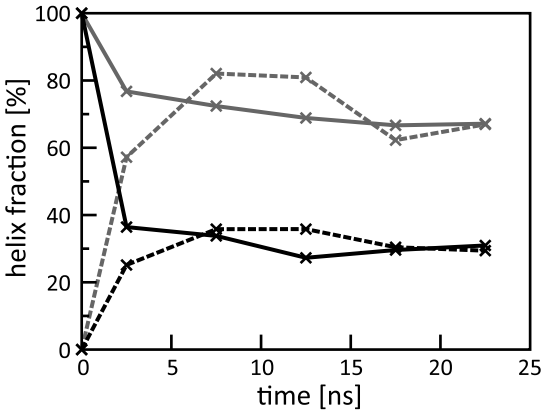

Figure S7: DSSP helix fractions $\langle H_\alpha\rangle_\tau$ of $\mathcal{P}$ averaged over periods $\tau = 5$ ns within the two 25 ns HADES-RE-MD simulations $\mathcal{P}$.H.α (solid curves) and $\mathcal{P}$.H.u (dashed curves) and over the residues of $\mathcal{P}$. The gray curves refer to $\langle H_\alpha\rangle_\tau(T_0)$ measured for the replica at $T_0 = 300$ K. The black curves are averages $\langle\langle H_\alpha\rangle_\tau\rangle_T$ over all replicas in the HADES-RE-MD ensembles.

REST sampling is nearly six times more efficient than RE for the given system.

S3. Convergence of HADES-RE-MD

The convergence of HADES-RE-MD, which had been speculatively parameterized by the set [27] Λ (see Table S2), was checked by the same procedure, which was employed by Reichold [1] and is described in Section S2.1 above. Hence, also we measure the helical content of $\mathcal{P}$ by averaging the DSSP [26] helix-fraction $H_\alpha(t)$ as defined by Eq. (S1) over time spans τ of HADES-RE-MD and over the $(N-2)$ accessible residues of $\mathcal{P}$, and also we compare the two HADES-RE-MD melting curves $\langle H_\alpha\rangle_\tau(T)$ obtained for $\mathcal{P}$ from the corresponding simulations $\mathcal{P}$.H.α and $\mathcal{P}$.H.u.

The HADES-RE-MD-Figure S7 closely resembles the corresponding REST-MD-Figure S4. After about 10 ns the helix contents have become independent of the initial conditions in both cases and, hence, the respective ensembles have reached their equilibria. This convergence is apparently a little faster for HADES-RE-MD than for REST-MD. Furthermore, the helix content $\langle H_\alpha\rangle_\tau(T_0)$ predicted for $\mathcal{P}$ at $T_0 = 300$ K is according to HADES-RE-MD only 67 % whereas according to REST-MD it is 90 %.

Note here, that both predictions on the helix content are most likely much larger than the true helix content of $\mathcal{P}$ in real water at 300 K, because the CHARMM22/CMAP force field apparently [1] yields gross overestimates of α-helical stability. From CD measurements reported for related peptides in Reichold's thesis, for which additional simulations were presented, and from the calculated melting curves one may conclude that the true 300 K helix content of $\mathcal{P}$ should be below 20 %.

Nevertheless, the differences of the HADES-RE-MD and the REST-MD descriptions signify that the speculative and most simple parameter set Λ listed in Table S2 may be of very low quality, if it is suggested to be a model for an explicit TIP3P solvent. Therefore this table additionally contains a second guess featuring by 20 % diminished Gaussian widths at the glutamate oxygens ("OC") and at the lysine amino hydrogens ("HC"), which should enhance the solubility of the Es and the single K appearing in the sequence of $\mathcal{P}$. As is shown in the main text of the paper this modification can reduce the difference between the HADES-RE-MD and REST-MD predictions of $\langle H_\alpha\rangle_\tau(T_0)$.

Table S2: Speculative Gaussian Widths σ_X of Atom Types X for HADES

σ_X	Λ	Λ'
σ_{H}	0.5192	0.5192
σ_{HC}	0.5192	0.4000
σ_{N}	0.7788	0.7788
σ_{C}	0.8654	0.8654
σ_{O}	0.6923	0.6923
σ_{OC}	0.6923	0.5500

S4. Salt Bridge propensities

Figure S8 shows free energy profiles $\Delta G(d)$ of $\mathcal{P}$ for the three possible salt bridges E_5-K_9, E_5-K_{10}, E_5-K_4 as functions of the distance d between the carboxylate carbon C_δ of E_5 and the amino nitrogens N_ζ of the respective K_i. The free energy profiles

$$\Delta G(d) = -k_{\mathrm{B}}T \ln p(d) - G_0 \qquad \text{(S2)}$$

were obtained at $T = 300$ K from explicit solvent REST-MD simulations [1] (TIP3P) and from HADES-RE-MD with the parameter sets Λ and Λ', respectively, through normalized histograms $p(d)$ using a bin size of 0.1 Å. The respective values G_0 were determined as $\min[-k_{\mathrm{B}}T \ln p(d)]$.

With explicit TIP3P solvent (blue curves) all three salt bridges are clearly indicated by local minima of $\Delta G(d)$ near $d_{\mathrm{S}} \equiv 3.8$ Å. Beyond this salt bridge distance d_{S}, $\Delta G(d)$ is seen to increase up to about 4.5 Å, which appears to be the minimal distance compatible with a TIP3P water molecule partially intruding into the breaking salt bridge. This distance is indicated by the dotted line in the figure and has been employed as a cutoff to compute the salt bridge frequencies listed in the paper.

The salt bridge E_5-K_9, which links two residues separate by just one α-helical turn, is apparently much more stable than the other two. The corresponding free energy profile $\Delta G(d)$ (blue curve in the topmost graph) exhibits a second minimum, which apparently corresponds to a hydrogen bonded structure linking E_5 and K_9. A similar remnant of the granular structure of the TIP3P solvent is seen in the profile $\Delta G(d)$ of E_5-K_4. The key difference between the profile of E_5-K_9 and those of the other two

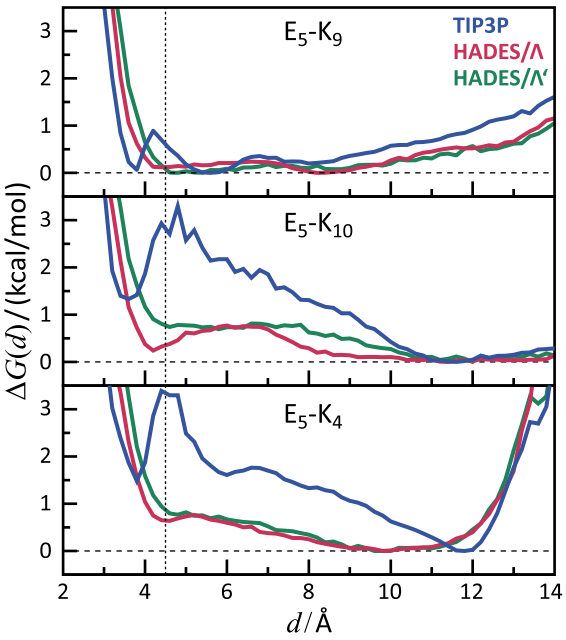

Figure S8: $\Delta G(d)$ for the three possible salt bridges E_5-K_9, E_5-K_{10}, and E_5-K_4 in $\mathcal{P}$; the colors code different simulation setups; blue: explicit solvent REST-MD, red: HADES-RE-MD with parameter set Λ, green: HADES-RE-MD with Λ'. The 4.5 Å cutoff employed to compute the salt bridge frequencies listed in the paper is indicated by the dotted line.

ion pairs is that the minimum at d_S is global for E_5-K_9 and only local for the other two.

HADES-RE-MD(Λ) predicts qualitatively different profiles $\Delta G(d)$ (red curves), which, however share one important quantitative aspect with the blue explicit water curves. Here, the $\Delta G(d_S)$ values are equal for E_5-K_4 and E_5-K_{10} and only a little larger for E_5-K_9. For larger distances, however, $\Delta G(d_S)$ keeps decreasing until about 4.2 Å in all three cases indicating that the continuous solvent can intrude into the salt bridge without any excluded volume, which is typical for a molecular solvent. Within the region up to d_S the red profile is seen to be uniformly shifted toward larger distances d as compared to the blue profile indicating that, at close contact, the salt bridges are stronger in TIP3P than in HADES. On the other hand, at distances between d_S and 4.5 Å the lacking granularity of HADES strongly favors smaller ion pair distances.

With the alternative parameter set Λ', which was introduced to generate an enhanced solubility of the ions, one clearly recognizes that all profiles $\Delta G(d)$ (green curves) are shifted (as compared to the red curves) toward larger distances indicating that the intended weakening of ion pair interactions actually took place.

S5. Calculation of RF energies

Approximate RF energies $\tilde{W}^{\mathrm{RF}}(\mathbf{R})$ of $\mathcal{P}$ were calculated by the well-known program DelPhi [28, 29], which numerically solves the dielectric Poisson equation (PE) on a hierarchy of cubic grids, and by HADES for a set of 112 000 configurations $\mathbf{R}$, which had been randomly chosen from the ensemble of the HADES-RE-MD simulation $\mathcal{P}.\mathrm{H}.\alpha/\mathrm{u}$.

The atomic partial charges needed for the DelPhi [28, 29] and HADES calculations of $\tilde{W}_{\mathrm{D}}^{\mathrm{RF}}(\mathbf{R})$ and $\tilde{W}_{\mathrm{H}}^{\mathrm{RF}}(\mathbf{R})$, respectively, were taken from the CHARMM22 force field. As suggested in Ref. [27] the atomic radii, which define the atomic volumes in DelPhi, were chosen as the van der Waals equilibrium distances $2^{1/6}\sigma_X^{\mathrm{vdW}}$ of the van der Waals parameters σ_X^{vdW} given by CHARMM22 for the various atoms types X occurring in $\mathcal{P}$. For an accurate computation of the $W^{\mathrm{RF}}(\mathbf{R})$ with DelPhi a three step focusing scheme was employed, which used 65^3 grid points and a 20%, 70%, and 90% filling (perfil parameter of DelPhi) filling of the grid. The dielectric constants were set to $\varepsilon_s = 1.0$ and $\varepsilon_c = 80.0$, respectively.

S5.1. Linear regression

To investigate the correlation of the $\tilde{W}^{\mathrm{RF}}(\mathbf{R})$ obtained by DelPhi and HADES we calculated the linear correlation coefficient

$$r = \frac{\sum_i (x_i - \langle x\rangle)(y_i - \langle y\rangle)}{\sqrt{\sum_i (x_i - \langle x\rangle)^2 \sum_i (y_i - \langle y\rangle)^2}} \qquad \text{(S3)}$$

which is one for perfect correlation. Here, the x_i represent the DelPhi RF energies of the structures $\mathbf{R}_i$ and the y_i the corresponding HADES results. $\langle x\rangle$ and $\langle y\rangle$ are the respective averages. The slope of the linear regression was then calculated from

$$b = \frac{\sum_i (x_i - \langle x\rangle)(y_i - \langle y\rangle)}{\sum_i (x_i - \langle x\rangle)^2}. \qquad \text{(S4)}$$

Correlations were considered, for the original speculative HADES parameter set Λ, its variant $\Lambda\prime$ supposedly yielding a higher solubility, and other variants as described in the main text.

Acknowledgement

This work has been supported by the Deutsche Forschungsgemeinschaft (SFB 749/C4).

[1] R. Reichold, Rechnergestützte Beschreibung der Struktur und Dynamik von Peptiden und ihren Bausteinen, Dissertation, Fakultät für Physik, Ludwig-Maximilians-Universität (2009).

[2] G. Schaftenaar, J. H. Noordik, Molden: A pre- and post-processing program for molecular and electronic structures, J. Comput. Aid. Mol. Des. 14 (2000) 123–134.

[3] S. R. Durell, B. R. Brooks, A. Ben-Naim, Solvent-induced forces between two hydrophilic groups, J. Phys. Chem. 98 (1994) 2198–2202.

[4] A. D. MacKerell, 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, M. Karplus, All-atom empirical potential for molecular

modeling and dynamics studies of proteins, J. Phys. Chem. B 102 (1998) 3586–3616.
[5] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, Comparison of simple potential functions for simulating liquid water, J. Chem. Phys. 79 (1983) 926–935.
[6] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, J. R. Haak, Molecular dynamics with coupling to an external bath, J. Chem. Phys. 81 (1984) 3684–3690.
[7] M. Eichinger, H. Grubmüller, H. Heller, P. Tavan, FAMUSAMM: A new algorithm for rapid evaluation of electrostatic interactions in molecular dynamics simulations, J. Comp. Chem. 18 (1997) 1729–1749.
[8] G. Mathias, B. Egwolf, M. Nonella, 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 (2003) 10847–10860.
[9] K. Lorenzen, M. Schwörer, P. Tröster, S. Mates, P. Tavan, Optimizing the accuracy and efficiency of fast hierarchical multipole expansions for MD simulations, J. Chem. Theory Comput. 8 (2012) 3628–3636.
[10] M. Schwörer, B. Breitenfeld, P. Tröster, K. Lorenzen, P. Tavan, G. Mathias, Coupling DFT to polarizable force fields for efficient and accurate Hamiltonian molecular dynamics simulation, J. Chem. Phys. 138 (2013) 244103.
[11] S. Bauer, P. Tavan, G. Mathias, Electrostatics of proteins in dielectric solvent continua. II. Hamiltonian reaction field dynamics, J. Chem. Phys. 140 (2014) 104103.
[12] K. Lorenzen, C. Wichmann, P. Tavan, Including the dispersion attraction into structure-adapted fast multipole expansions for MD simulations, J. Chem. Theory Comput. 10 (2014) 3244.
[13] A. D. MacKerell, Jr., M. Feig, C. L. Brooks, III, Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulation, J. Comp. Chem. 25 (2004) 1400–1415.
[14] C. Niedermeier, P. Tavan, A structure adapted multipole method for electrostatic interactions in protein dynamics, J. Chem. Phys. 101 (1994) 734–748.
[15] C. Niedermeier, P. Tavan, Fast version of the structure adapted multipole method — efficient calculation of electrostatic forces in protein dynamics, Mol. Simul. 17 (1996) 57–66.
[16] M. P. Allen, D. J. Tildesley, Computer simulations of liquids, Clarendon, Oxford, 1987.
[17] V. Kräutler, W. F. van Gunsteren, P. H. Hünenberger, A fast SHAKE algorithm to solve distance constraint equations for small molecules in molecular dynamics simulations, J. Comp. Chem. 22 (2001) 501–508.
[18] M. Lingenheil, R. Denschlag, R. Reichold, P. Tavan, The "hot-solvent/cold-solute" problem revisited, J. Chem. Theory Comput. 4 (2008) 1293–1306.
[19] K. Hukushima, K. Nemoto, Exchange Monte Carlo method and application to spin glass simulations, J. Phys. Soc. Jpn. 65 (1996) 1604–1608.
[20] U. H. E. Hansmann, Free energy landscape and folding mechanism of a β-hairpin in explicit water: A replica exchange molecular dynamics study, Chem. Phys. Lett. 281 (1997) 140.
[21] Y. Sugita, Y. Okamoto, Replica-exchange molecular dynamics method for protein folding, Chem. Phys. Lett. 314 (1999) 141–151.
[22] P. Liu, B. Kim, R. A. Friesner, B. J. Berne, Replica exchange with solute tempering: A method for sampling biological systems in explicit water, Proc. Natl. Acad. Sci. U.S.A. 102 (2005) 13749–13754.
[23] R. Denschlag, M. Lingenheil, P. Tavan, G. Mathias, Simulated solute tempering, J. Chem. Theory Comput. 5 (2009) 2847–2857.
[24] R. Denschlag, M. Lingenheil, P. Tavan, Optimal temperature ladders in replica exchange simulations, Chem. Phys. Lett. 473 (2009) 193–195.
[25] A. E. Garcia, K. Y. Sanbonmatsu, α-Helical Stabilization by Side Chain Shielding of Backbone Hydrogen Bonds, Proc. Natl. Acad. Sci. U.S.A. 99 (2002) 2782–2787.
[26] W. Kabsch, C. Sander, Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features, Biopolymers 22 (1983) 2577–2637.
[27] S. Bauer, G. Mathias, P. Tavan, Electrostatics of proteins in dielectric solvent continua. I. An accurate and efficient reaction field description, J. Chem. Phys. 140 (2014) 104102.
[28] W. Rocchia, E. Alexov, B. Honig, Extending the applicability of the nonlinear Poisson-Boltzmann equation: A multiple dielectric constants and multivalent ions, J. Phys. Chem. B 105 (2001) 6507–6514.
[29] A. Nicholls, B. Honig, A rapid finite difference algorithm, utilizing successive over-relaxation to solve the Poisson–Boltzmann equation, J. Comp. Chem. 12 (1991) 435–445.

4 Résumé und Ausblick

"Kids, you tried your best and you failed miserably. The lesson is; never try."

Homer Simpson,
The Simpsons – Burns' Heir

Ziel dieser Arbeit war es, anknüpfend an die Arbeiten von ET [103] und ST [104, 105] ein RF-Verfahren zur Beschreibung der Kontinuumselektrostatik bei MD-Simulationen zu entwickeln. Der ursprüngliche Plan, RF-Kräfte aus dem in [103] gegebenen Ausdruck für die RF-Energie zu berechnen, stellte sich aber leider nicht als zielführend heraus. Zwar liefert diese Darstellung für das wichtige Beispiel der Kirkwood-Kugel sehr gute Ergebnisse, für andere sehr einfache Beispiele versagt die gegebene Approximation jedoch (vgl. Abschnitt II.B.2 in [97]).

Eine genaue Untersuchung der Arbeit von ET anhand der einfachsten möglichen und analytisch lösbaren Beispiele zur Kontinuumselektrostatik, nämlich des Born Ions [61] und der sphärischen Kavität im dielektrischen Kontinuum, zeigten, dass dieses Versagen auf die Vernachlässigung der radialen Anteile der atomaren Antipolarisationen zurückzuführen ist. Eine entsprechende Neuformulierung der exakten Theorie von ET lieferte den gesuchten exakten und atomaren Ausdruck für das RF-Potential [97].

Darauf aufbauend konnten die von ET vorgeschlagenen Approximationen so erweitert werden, dass sich ein einfacher analytischer Ausdruck für die RF-Energie ergab. Dieser Ausdruck bildete dann die Grundlage für die in dieser Arbeit vorgestellte Simulationsmethode des Hamilton'schen dielektrischen Lösungsmittels (HADES, von engl. *Hamiltonian dielectric Solvent*) [109]. Anhand einiger relevanter Beispiele wird in [97] gezeigt, mit welcher Genauigkeit die erweiterte Approximation die PG löst und in welchen Fällen sie der ursprünglichen Formulierung durch ET überlegen ist.

Ausgehend von der in [97] beschriebenen Approximation der RF-Energie war es nun, in der direkt daran anknüpfendenden Arbeit [109], möglich analytische Ausdrücke für die atomaren RF-Kräfte abzuleiten, die Hamilton'sche und damit stabile MD-Simulationen erlauben [109]. Inspiriert von der Ähnlichkeit der RF-Darstellung der Kontinuumselektrostatik für gelöste Proteine zu gewöhnlichen polarisierbaren Kraftfeldern [110] wurde dazu die Methode der Lagrange-Multiplikatoren verwendet. Diese Methode erlaubt es, die Zwangskräfte, die sich durch die verschiedene in [97] gegebenen Selbstkonsistenzbedingung ergeben, zu berücksichtigen. Die abgeleiteten Lagrange-Multiplikatoren sind ihrerseits durch Selbstkonsistenzbedingung gegeben und müssen ebenfalls durch Iteration bestimmt werden.

Um HADES für MD-Simulationen anwendbar zu machen, wurde eine Reihe von Methoden implementiert, welche die Konvergenz der Selbstkonsistenziterationen beschleunigen. Die *direct inversion in the iterative subspace* (DIIS) [111] verbessert dabei direkt die Konvergenz

der Iterationen, während die Extrapolation durch Lagrange-Polynome (LPE) bessere Startwerte für die Iterationen durch Extrapolation aus vorangegangenen Zeitschritten erzeugt. Diese beiden Methoden wurden in dem Programmpaket IPHIGENIE, das in unerer Arbeitsgruppe entwickelt wird, neben HADES auch für polarisierbare Kraftfelder vom Typ I [110] und Typ II [112] implementiert, so dass auch hier die Konvergenz der nötigen Iterationen beschleunigt werden konnte.

Eine erste Anwendung von HADES auf das kleine Modellpeptid Ace-Ala-NHMe in [109] zeigte, dass HADES tatsächlich Hamilton'sche MD-Simulationen erlaubt. Ferner wurden in dieser Arbeit Konvergenzparameter von HADES bestimmt, welche zu einem optimalen Kompromiss zwischen Genauigkeit und Effizienz führen. Dazu wurden maximale Parameterwerte so bestimmt, dass die Verletzung der Energieerhaltung vernachlässigbar, d.h. vergleichbar mit einer MD-Simulation im Vakuum, ist. Ein Vergleich der freien Energielandschaft dieses Peptids mit einem Scan der RF-Energie über seine möglichen Konformationen zeigt weiterhin die Kompatibilität der HADES Kräfte mit der RF-Energielandschaft. Diese wichtige Übereinstimmung war in der Vorgängermethode von ST nicht gegeben [105], da die RF-Kräfte hier nicht direkt aus der RF-Energie abgeleitet wurde.

Die Anwendung von HADES auf das α-helikale Dekapeptid Ace-AAAKEAAAKK-HN_2 in [113] zeigt, dass die für Ace-Ala-NHMe beobachtete Effizienz auch für größere Peptide erhalten bleibt und vergleichbar mit der von etablierten GB Methoden ist. Da HADES das linear skalierende $SAMM_{4,1}$ [43, 48] Verfahren zur Beschreibung der elektrostatischen Wechselwirkungen verwendet, bleibt seine Effizienz auch mit steigender Größe des untersuchten Peptids oder Proteins bestehen. Im Gegensatz dazu sinkt die Effizienz von GB Methoden für große Systeme unter Umständen unter die von expliziten Lösungsmittelsimulationen [114].

Wegen der angesprochenen Effizienz von HADES war es in [113] möglich, für das Dekapeptid mit Replika-Austausch-Methoden das temperaturabhängige Profil des α-Helixanteils [115] zu berechnen, den Einfluss der Parametrisierung von HADES auf die resultierende Schmelzkurve zu untersuchen und diese mit den Ergebnissen von Reichold [108] aus expliziten Lösungsmittelsimulationen zu vergleichen. Dieser Vergleich hat gezeigt, dass HADES bereits mit der äußerst einfachen *ad-hoc* Parametrisierung aus [97] eine Schmelzkurve liefert, die qualitativ das Verhalten des Modellpeptids bei einer expliziten Lösungsmittelbeschreibung wiederspiegelt. Quantitative Unterschiede, die durch eine leicht modifizierte Wahl der Parameter von HADES verringert werden konnten, machten jedoch deutlich, dass eine genauere und differenziertere Opimierung der HADES Parameter möglich und nötig ist.

4.1 Parametrisierung von HADES

Um die Parameter von HADES, also die atomaren Größen σ_i der Atome zu optimieren, kann ein Vorgehen, wie es bei ST beschrieben ist [105], herangezogen werden. Dazu sollten für unterschiedliche Proteine und Peptide die mittleren Lösungsmittelkräfte für ein möglichst realistisches Wassermodell, wie etwa das polarisierbare Sechspunktmodell TL6P von Tröster et al. [116], bestimmt werden, die dann mit den RF-Kräften aus HADES verglichen werden

können. Diese Optimierung ist Gegenstand der laufenden Masterarbeit von Benedikt Hoock in unserer Arbeitsgruppe.

Die nötigen mittleren elektrostatischen Lösungsmittelkräfte $\langle \mathbf{F}_i^e \rangle$ auf die N Proteinatome i werden dabei für eine „eingefrorene“ Struktur des zellulären Prion-Proteins PrP^C berechnet, während die Dynamik der explizit modellierten Wasserumgebung integriert wird. Die $\langle \mathbf{F}_i^e \rangle$ ergeben sich dann aus der Mittelung

$$\langle \mathbf{F}_i^e \rangle \equiv \frac{1}{N_k} \sum_{j=1}^{N_k} \mathbf{F}_i^e(\mathbf{R}_j) \tag{4.1}$$

über ein Ensemble von N_k Wasserkonfigurationen $\mathbf{R}_j$. Um die Qualität eines Parametersatzes Λ abschätzen zu können, wird der mittlere quadratische Abstand

$$\Delta\mathbf{F}(\Lambda, \mathcal{G}) \equiv \frac{1}{N} \sum_{i=1}^{N} \left[\langle \mathbf{F}_i^e \rangle - \mathbf{F}_i^{\mathrm{RF}}(\Lambda, \mathcal{G}) \right]^2 \tag{4.2}$$

der HADES Kräfte $\mathbf{F}_i^{\mathrm{RF}}(\Lambda, \mathcal{G})$ von den $\langle \mathbf{F}_i^e \rangle$ verwendet. Der Parameter $\mathcal{G}$ gibt dabei eine Gruppierung der Proteinatome in Klassen ähnlicher Atome an, die sich an den in MM Kraftfeldern verwendeten Gruppierungen orientieren sollte. Eine optimale Parameterisierung Λ^{opt} kann nun durch Minimierung von $\Delta\mathbf{F}(\Lambda, \mathcal{G})$, etwa durch Gradientenabstieg, gefunden werden.

4.2 Erweiterung von HADES auf ionenhaltige Lösungsmittel

Für die Simulation von polaren oder schwach geladenen Peptiden und Proteinen ist es meist ausreichend, das Lösungsmittel als dielektrisches Kontinuum zu beschreiben. Realistischer, und für die Simulation stark geladener Biomoleküle (etwa von DNA) auch nötig ist jedoch die Verwendung eines ionenhaltigen Lösungsmittelkontinuums. Die gelösten Ionen werden in der Kontinuumselektrostatik durch eine „Ionenwolke“ dargestellt, die in einer durch die Boltzmannverteilung beschriebenen Ionenladungsdichte resultiert [117, 118]. Das elektrostatische Potential, das von einer Ladungsverteilung $\rho(\mathbf{r})$ erzeugt wird, ergibt sich dann aus der Lösung der Poisson-Boltzmann-Gleichung (PBG)

$$\nabla\varepsilon(\mathbf{r})\nabla\Phi(\mathbf{r}) = -4\pi \left[\rho(\mathbf{r}) + \sum_i c_i Q_i \lambda(\mathbf{r}) \exp\left(-\frac{Q_i \Phi(\mathbf{r})}{k_{\mathrm{B}} T} \right) \right] \tag{4.3}$$

mit der Konzentration c_i der Ionensorte i, welche die Ladung Q_i besitzt. Die charakteristische Funktion der Ionen $\lambda(\mathbf{r})$ beschreibt das Volumen, das für die gelösten Ionen zugänglich ist.

Die nichtlineare PBG kann nur in wenigen Fällen analytisch gelöst werden [119, 120]. Für geringe Ionenkonzentrationen ist es jedoch möglich, sie zu linearisieren und man erhält mit der Definition des Debye-Hückel-Parameters [119, 121]

$$\kappa^2 = \frac{4\pi}{\varepsilon_c k_{\mathrm{B}} T} \sum_i c_i Q_i \tag{4.4}$$

die linearisierte PBG (LPBG)

$$\nabla\varepsilon(\mathbf{r})\nabla\Phi(\mathbf{r}) = -4\pi\rho(\mathbf{r}) + \varepsilon_c\kappa^2\lambda(\mathbf{r})\Phi(\mathbf{r}). \tag{4.5}$$

Dabei wird für das Lösungsmittelkontinuum Ladungsneutralität, d.h.

$$\sum_i c_i Q_i = 0 \tag{4.6}$$

vorausgesetzt, so dass die niedrigste Ordnung der Entwicklung der Expontentialfunktion aus Gleichung (4.3) verschwindet.

Um auch für die LPBG eine geeignete Approxmiation abzuleiten, ist es sinnvoll, diese zunächst in eine RF-Darstellung zu überführen [97]. Analog zur Darstellung in [103] ist es auch für die LPBG (4.5) möglich, den Einfluss des Dielektrikums durch eine Antipolarisation

$$\mathbf{P}(\mathbf{r}) = \chi_\varepsilon\Theta(\mathbf{r})\nabla\Phi(\mathbf{r}) \tag{4.7}$$

mit der durch Gleichung (1.15) gegebenen Suszeptibilität χ_ε darzustellen und man erhält

$$\Delta\Phi(\mathbf{r}) = -\frac{4\pi}{\varepsilon_c}[\rho(\mathbf{r}) - \nabla\cdot\mathbf{P}(\mathbf{r})] + \kappa^2\lambda(\mathbf{r})\Phi(\mathbf{r}). \tag{4.8}$$

Für eine Ladungsverteilung (1.8) aus atomaren Punktladungen liefert eine unscharfe Zerlegung der charakteristischen Funktion $\Theta(\mathbf{r})$, wie sie schon in [103] oder [97] beschrieben ist, mit

$$\Delta\Phi(\mathbf{r}) = -\frac{4\pi}{\varepsilon_c}\sum_i [q_i\delta(\mathbf{r}-\mathbf{r}_i) - \nabla\cdot\mathbf{P}_i(\mathbf{r})] + \kappa^2\lambda(\mathbf{r})\Phi(\mathbf{r}) \tag{4.9}$$

eine atomisierte Darstellung der Form (4.8) der ursprünglichen LPBG (4.5).

Analog zur Herleitung der RF-Darstellung mit den Gleichungen (24) bis (28) in [97], liefert die Separation des radialen Anteils der atomaren Antipolarisationsdichten $\mathbf{P}_i(\mathbf{r})$ und die Einführung der atomaren Abschirmladungsdichten $\hat{\rho}_i(\mathbf{r})$ und der skalierten nichtradialen Antipolarisationdichten $\hat{\mathbf{P}}_i(\mathbf{r})$ die RF-Darstellung

$$\Delta\Phi(\mathbf{r}) = -\frac{4\pi}{\varepsilon_s}\sum_i \left[q_i\delta(\mathbf{r}-\mathbf{r}_i) + \hat{\rho}_i(\mathbf{r}) - \nabla\cdot\hat{\mathbf{P}}_i(\mathbf{r})\right] + \kappa^2\lambda(\mathbf{r})\Phi(\mathbf{r}) \tag{4.10}$$

der LPBG.

Da mit der Definition (1.10) der charakteristischen Funktion bereits eine Beschreibung des Proteinvolumen $\mathfrak{V}_s$ zur Verfügung steht und die Ionen des Lösungsmittels nicht in $\mathfrak{V}_s$ eindringen können, ist es naheliegend, die charakterische Funktion der Ionen durch

$$\lambda(\mathbf{r}) = 1 - \Theta(\mathbf{r}), \tag{4.11}$$

auszudrücken und die Verteilung der Ionen so auf das dielektrisches Kontinuum zu beschränken [107]. Die Form (4.11) der charakteristischen Funktion ergibt nun, zusammen mit der unscharfen Zerlegung von $\Theta(\mathbf{r})$ und mit den durch

$$\hat{\rho}_i^\kappa(\mathbf{r}) \equiv \frac{\kappa^2\varepsilon_s}{4\pi}\vartheta_i(\mathbf{r})\Phi(\mathbf{r}) \tag{4.12}$$

definierten atomaren Ladungsverteilungen, die gesuchte exakte RF-Darstellung

$$(\Delta - \kappa^2)\Phi(\mathbf{r}) = -\frac{4\pi}{\varepsilon_s}\sum_i \Big[q_i\delta(\mathbf{r}-\mathbf{r}_i) + \hat{\rho}_i(\mathbf{r}) + \hat{\rho}_i^\kappa(\mathbf{r}) - \nabla\cdot\hat{\mathbf{P}}_i(\mathbf{r})\Big] \tag{4.13}$$

der LPBG. Der Einfluss des Dielektrikums wird hier, wie schon in [97], vollständig durch die atomaren Abschirmladungsdichten $\hat{\rho}_i(\mathbf{r})$ und Antipolarisationsdichten $\mathbf{P}_i(\mathbf{r})$ beschrieben. Letztere müssen dazu die in [97] angegebenen Selbstkonsistenzbedingungen erfüllen. Dagegen kann der Einfluss der Ionendichte durch die atomaren, auf $\mathfrak{V}_s$ beschränkten Ladungsverteilungen $\hat{\rho}_i^\kappa(\mathbf{r})$ dargestellt werden. Die Ladungsverteilung $\hat{\rho}_i^\kappa(\mathbf{r})$ ist dabei durch die neue Selbstkonsistenzbedingung (4.12) bestimmt.

Um zu einer approximativen Lösung $\tilde{\Phi}(\mathbf{r})$ für das Potential $\Phi(\mathbf{r})$ zu gelangen, müssen geeignete Näherungen für die Quellen $\hat{\rho}_i(\mathbf{r})$, $\hat{\rho}_i^\kappa(\mathbf{r})$ und $\hat{\mathbf{P}}_i(\mathbf{r})$ gefunden werden. In [97] konnten wir bereits zeigen, dass die Gauß'schen Ladungsverteilungen

$$\tilde{\rho}_i(\mathbf{r}) = \hat{q}_i G(\mathbf{r}\,|\,\mathbf{r}_i,\hat{\sigma}_i) \tag{4.14}$$

und Dipolverteilungen

$$\tilde{\mathbf{P}}_i(\mathbf{r}) = \tilde{\mathbf{p}}_i G(\mathbf{r}\,|\,\mathbf{r}_i,\hat{\sigma}_i) \tag{4.15}$$

als Näherungen für die Dichten $\hat{\rho}_i(\mathbf{r})$ und $\hat{\mathbf{P}}_i(\mathbf{r})$ ausgezeichnete Ergebnisse liefern. Die Stärken $\hat{q}_i$ und $\tilde{\mathbf{p}}_i$ dieser atomaren Dichten müssen dabei die durch die Gleichungen (30), (31) und (77) aus [97] gegebenen Bedingungen erfüllen.

Zur Approximation der neuen Ladungsdichten $\hat{\rho}_i^\kappa(\mathbf{r})$ ist es, den Argumenten von ET folgend [107], sinnvoll, die atomaren charakterischen Funktionen $\vartheta_i(\mathbf{r})$ durch ihre Gauß'schen Näherungen

$$\tilde{\vartheta}_i(\mathbf{r}) = v_i G(\mathbf{r}\,|\,\mathbf{r}_i,\sigma_i) \tag{4.16}$$

und das Potential in der Nähe der Atome i durch die lokale lineare Näherung

$$\tilde{\phi}(\mathbf{r}) \equiv \langle\phi\rangle_i + (\mathbf{r}-\mathbf{r}_i)\langle\nabla\phi\rangle_i \tag{4.17}$$

darzustellen. Dabei sind $\langle\phi\rangle_i$ und $\langle\nabla\phi\rangle_i$ lokale Mittelwerte des Potentials $\tilde{\Phi}(\mathbf{r})$ beziehungsweise seines Gradienten $\nabla\tilde{\Phi}(\mathbf{r})$ über das durch $\tilde{\vartheta}_i(\mathbf{r})$ definierte Gauß'sche Volumen.

Die Näherung (4.17) des lokalen Potentials kann nun mit

$$(\mathbf{r}-\mathbf{r}_i)G(\mathbf{r}\,|\,\mathbf{r}_i,\sigma_i) = -\sigma_i\nabla G(\mathbf{r}\,|\,\mathbf{r}_i,\sigma_i) \tag{4.18}$$

verwendet werden, um die Ladungsverteilung $\hat{\rho}_i^\kappa(\mathbf{r})$ durch

$$\tilde{\rho}_i^\kappa(\mathbf{r}) = \tilde{q}_i^\kappa G(\mathbf{r}\,|\,\mathbf{r}_i,\sigma_i) - \nabla\cdot\tilde{\mathbf{p}}_i^\kappa G(\mathbf{r}\,|\,\mathbf{r}_i,\sigma_i) \tag{4.19}$$

zu approximieren. Das Potential $\tilde{\Phi}(\mathbf{r})$ wird also neben den Gauß'schen Abschirmladungsverteilungen (4.14) und den anti-polarisierbaren Dipoldichten (4.15) durch die zusätzlichen Ladungsverteilungen

$$\tilde{\rho}_i^\kappa(\mathbf{r}) \equiv \tilde{q}_i^\kappa G(\mathbf{r}\,|\,\mathbf{r}_i,\sigma_i) \tag{4.20}$$

und die ebenfalls anti-polarisierbaren Dipoldichten

$$\tilde{\mathbf{P}}_i^\kappa(\mathbf{r}) \equiv \tilde{\mathbf{p}}_i^\kappa G(\mathbf{r} \,|\, \mathbf{r}_i, \sigma_i) \tag{4.21}$$

erzeugt. Die Stärken $\tilde{q}_i^\kappa$ und $\tilde{\mathbf{p}}_i^\kappa$ sind durch die neuen Selbstkonsistenzbedingungen

$$\tilde{q}_i^\kappa = \frac{\kappa^2 \varepsilon_s}{4\pi} v_i \langle \phi \rangle_i \tag{4.22}$$

und

$$\tilde{\mathbf{p}}_i^\kappa = \frac{\kappa^2 \varepsilon_s}{4\pi} v_i \sigma_i \langle \nabla \phi \rangle_i \tag{4.23}$$

gegeben. Es ist also auch im Fall der LPBG möglich, sowohl den Einfluss des Dielektrikums, als auch den der gelösten Ionendichte durch atomare Ladungs- und Dipolverteilungen zu beschreiben, die ausschließlich auf das Volumen des Proteins beschränkt sind.

Für eine effiziente Implementierung der hier kurz dargestellten Näherung im Rahmen der $\text{SAMM}_{4,1}$ Methode ist jedoch noch einiges an Arbeit zu leisten. Zum einen muss für die Potentiale und Felder, die sich für die auftretenden Ladungs- und Dipolverteilungen aus der Green'schen Funktion

$$G^\kappa(\mathbf{r}, \mathbf{r}') \equiv \frac{1}{|\mathbf{r} - \mathbf{r}'|} \exp\left(-\kappa |\mathbf{r} - \mathbf{r}'|\right) \tag{4.24}$$

zum Differentialoperator $(\Delta - \kappa^2)$ ergeben, eine Multipol-Darstellung abgeleitet werden, damit diese in die $\text{SAMM}_{4,1}$ Methode integriert werden können. Zum anderen führen die neuen Selbstkonsistenzbedingungen (4.22) und (4.23) zu Zwangskräften, die ebenfalls noch berücksichtigt werden müssen. Diese können jedoch wie schon die Selbstkonsistenzbedingung (77) aus [97] durch die Methode der Lagrange-Multiplikatoren erfasst werden, für die entsprechende Selbstkonsistenzbedingungen abgeleitet werden müssen.

Literaturverzeichnis

[1] J. M. Berg, J. L. Tymoczko und L. Stryer. *Biochemie*. Spektrum Akademischer Verlag GmbH, Heidelberg – Berlin, 5. Auflage (2003).

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

[3] G. E. Schulz und R. H. Schirmer. *Principles of Protein Structure*. Springer-Verlag, New York (1979).

[4] C. Branden und J. Tooze. *Introduction to Protein Structure*. Garland Publishing, New York (1991).

[5] S. Y. Tan und M. B. Pepys. Amyloidosis. *Histopathology*, **25**, 403–414 (1994).

[6] J. Hardy und D. J. Selkoe. The amyloid hypothesis of alzheimer's disease: Progress and problems on the road to therapeutics. *Science*, **297**, 353–356 (2002).

[7] S. B. Prusiner. Molecular biology of proin diseases. *Science*, **252**, 349–358 (1991).

[8] M. Goedert. Alpha-synuclein and neurodegenerative diseases. *Nat. Rev. Neurosci.*, **2**, 492–501 (2001).

[9] T. S. Ulmer, A. Bax, N. B. Cole und R. L. Nussbaum. Structure and dynamics of micelle-bound human alpha-synuclein. *J. Biol. Chem.*, **280**, 9595–9603 (2005).

[10] C. A. Opitz, M. Kulke, M. C. Leake, C. Neagoe, H. Hinssen und R. J. Hajjar. Damped elastic recoil of the titin spring in myofibrils of human myocardium. *Proc. Natl. Acad. Sci. U.S.A.*, **100**, 12688–12693 (2003).

[11] C. Branden und J. Tooze. *Introduction to Protein Structure*. Garland New York, 2. Auflage (1991).

[12] W. Humphrey, A. Dalke und K. Schulten. VMD - Visual molecular dynamics. *J. Mol. Graphics*, **14**, 33–38 (1996).

[13] M. Karplus. The levinthal paradox: Yesterday and today. *Folding and Design*, **2**, S69–S75 (1997).

[14] B. Honig. Protein folding: From the levinthal paradox to structure prediction. *J. Mol. Biol.*, **293**, 283–293 (1999).

[15] C. Levinthal. How to fold graciously. In P. Debrunner, J. C. M. Tsibris und E. Münck (Herausgeber), *Mössbauer Spectroscopy in Biological Systems*. University of Illinois (1969).

[16] B. Berger und T. Leighton. Protein folding in the hydrophobic-hydrophilic (HP) model in NP-complete. *J. Comput. Biol.*, **5**, 27–40 (1998).

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

[18] So much more to know. *Science*, **309**, 78–102 (2005).

[19] P. D. Boyer. The ATP synthase – A splendid molecular machine. *Ann. Rev. Biochem.*, **66**, 717–749 (1997).

[20] T. L. Blundell und L. N. Johnson. *Protein Crystallography*. Academic Press, London (1976).

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

[22] D. Neuhaus und M. Williamson. *The Nuclear Overhauser Effect in Structural and Conformational Analysis*. VCH, Weinheim (1989).

[23] W. A. Eaton, 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. Struct.*, **29**, 327–359 (2000).

[24] E. Garman und T. Schneider. Macromolecular cryocrystallography. *J. Appl. Cryst.*, **30**, 211–237 (1997).

[25] R. Winter und F. Noll. *Methoden der Biophysikalischen Chemie*. Teubner Studienbücher, Stuttgart (1998).

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

[27] B. Alberts, D. Bray, J. Lewis, M. Raff, K. Roberts und J. D. Watson. *Molekularbiologie der Zelle*. VCH, Weinheim (1990).

[28] W. Zinth und D. Oesterhelt. Primary photochemical processes in bacteriorhodopsin. In E. Riklis (Herausgeber), *Photobiology, the Science and its Applications*, S. 531–535. Plenum Press, New York (1990).

[29] F. Siebert, W. Mäntele und W. Kreutz. Evidence for the protonation of two internal carboxylic groups during the photocycle of bacteriorhodopsin. *FEBS Letters*, **141**, 82–87 (1982).

[30] R. Kaptein, R. Boelens, R. M. Scheek und W. F. van Gunsteren. Protein structure from NMR. *Biochemistry*, **27**, 5389–5395 (1988).

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

[32] F. López García, R. Zahn, R. Riek und K. Wüthrich. NMR structure of the bovine prion protein. *Proc. Natl. Acad. Sci. U.S.A.*, **97**, 8334–8339 (2000).

[33] S. Spörlein, H. Carstens, H. Satzger, C. Renner, R. Behrendt, L. Moroder, P. Tavan, W. Zinth und J. Wachtveitl. Ultrafast spectroscopy reveals sub-nanosecond peptide conformational dynamics and validates molecular dynamics simulation. *Proc. Natl. Acad. Sci. U.S.A.*, **99**, 7998–8002 (2002).

[34] L. Nilsson, M. Karplus, A. Brünger, M. Clore und A. Gronenbornt. Determination of 3d molecular structures from NMR/NOE distance data: Application to nucleic acids. *J. Mol. Graphics*, **4**, 182 (1986).

[35] S. Kirkpatrick, C. D. Gelatt und M. P. Vecchi. Optimization by simulated annealing. *Science*, **220**, 671–680 (1983).

[36] W. F. van Gunsteren 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).

[37] D. Rapaport. *The Art of Molecular Dynamics Simulation*. Cambridge University Press, 2. Auflage (2004).

[38] A. D. MacKerell, 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 und M. Karplus. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B*, **102**, 3586–3616 (1998).

[39] D. A. Pearlman, D. A. Case, J. W. Caldwell, W. R. Ross, T. E. Cheatham III, S. DeBolt, D. Ferguson, G. Seibel und P. Kollman. AMBER, a computer program for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to elucidate the structures and energies of molecules. *Comput. Phys. Commun.*, **91**, 1–41 (1995).

[40] M. Christen, P. H. Hünenberger, D. Bakowies, R. Baron, R. Bürgi, D. P. Geerke, T. N. Heinz, M. A. Kastenholz, V. Kräutler, C. Oostenbrink, C. Peter, D. Trzesniak und W. F. van Gunsteren. The GROMOS software for biomolecular simulation: GROMOS05. *J. Comp. Chem.*, **26**, 1719–1751 (2005).

[41] J. W. Ponder und D. A. Case. Force fields for protein simulations. *Advances in Protein Chemistry*, **66**, 27–85 (2003).

[42] L. Verlet. Computer "experiments" on classical fluids. I. Thermodynamical properties of Lennard–Jones molecules. *Phys. Rev.*, **159**, 98–103 (1967).

[43] K. Lorenzen, C. Wichmann und P. Tavan. Including the dispersion attraction into structure-adapted fast multipole expansions for MD simulations. *J. Chem. Theory Comput.*, **10**, 3244–3259 (2014).

[44] T. A. Darden, 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).

[45] C. Niedermeier und P. Tavan. A structure adapted multipole method for electrostatic interactions in protein dynamics. *J. Chem. Phys.*, **101**, 734–748 (1994).

[46] C. Niedermeier 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).

[47] M. Eichinger, H. Grubmüller, H. Heller und P. Tavan. FAMUSAMM: A new algorithm for rapid evaluation of electrostatic interactions in molecular dynamics simulations. *J. Comp. Chem.*, **18**, 1729–1749 (1997).

[48] K. Lorenzen, M. Schwörer, P. Tröster, S. Mates und P. Tavan. Optimizing the accuracy and efficiency of fast hierarchical multipole expansions for MD simulations. *J. Chem. Theory Comput.*, **8**, 3628–3636 (2012).

[49] J. A. McCammon, B. J. Gelin und M. Karplus. Dynamics of folded proteins. *Nature*, **267**, 585–590 (1977).

[50] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey und M. L. Klein. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, **79**, 926–935 (1983).

[51] H. J. C. Berendsen, J. R. Grigera und T. P. Straatsma. The missing term in effective pair potentials. *J. Phys. Chem.*, **91**, 6269–6271 (1987).

[52] W. Weber, P. H. Hünenberger und J. A. McCammon. Molecular dynamics simulations of a polyalanine octapeptide under Ewald boundary conditions: Influence of artificial periodicity on peptide conformation. *J. Phys. Chem. B*, **104**, 3668–3675 (2000).

[53] G. Mathias, 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).

[54] P. Tavan, 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*, S. 1170–1195. Wiley-VCH, Weinheim (2005).

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

[56] S. Geibel, J. H. Kaplan, E. Bamberg und T. Friedrich. Conformational dynamics of the Na^+/K^+–ATPase probed by voltage clamp fluorometry. *Proc. Natl. Acad. Sci. U.S.A.*, **100**, 964–969 (2003).

[57] M. E. Davis und J. A. McCammon. Electrostatics in biomolecular structure and dynamics. *Chem. Rev.*, **90**, 509–521 (1990).

[58] B. Roux und T. Simonson. Implicit solvent models. *Biophys. Chem.*, **78**, 1 – 20 (1999).

[59] J. Tomasi und M. Persico. Molecular interactions in solution: An overview of methods based on continuous distributions of the solvent. *Chem. Rev.*, **94**, 2027–2094 (1994).

[60] R. Jimenez, G. R. Fleming, P. V. Kumar und M. Maroncelli. Femtosecond solvation dynamics of water. *Nature*, **369** (1994).

[61] M. Born. Volumen und Hydratationswärme der Ionen. *Zeitschrift für Physik*, **1**, 45–48 (1920).

[62] J. G. Kirkwood. Theory of solutions of molecules containing widely separated charges with special application to zwitterions. *J. Chem. Phys.*, **2**, 351–361 (1934).

[63] M. E. Davis, 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).

[64] A. Nicholls und B. Honig. A rapid finite difference algorithm, utilizing successive over-relaxation to solve the Poisson–Boltzmann equation. *J. Comp. Chem.*, **12**, 435–445 (1991).

[65] M. K. Gilson, 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).

[66] Z. Zhou, P. Payne, M. Vasquez, N. Kuhn und M. Levitt. Finite-difference solution of the Poisson–Boltzmann equation: Complete elimination of self-energy. *J. Comp. Chem.*, **17**, 1344–1351 (1996).

[67] W. Rocchia, E. Alexov und B. Honig. Extending the applicability of the nonlinear Poisson–Boltzmann equation: A multiple dielectric constants and multivalent ions. *J. Phys. Chem. B*, **105**, 6507–6514 (2001).

[68] R. Luo, L. David und M. K. Gilson. Accelerated Poisson-Boltzmann calculations for static and dynamic systems. *J. Comp. Chem.*, **23**, 1244–1253 (2002).

[69] W. Geng, S. Yu und G. Wei. Treatment of charge singularities in implicit solvent models. *J. Chem. Phys.*, **127**, 114106 (2007).

[70] A. H. Juffer, E. F. F. Botta, B. A. M. van Keulen, A. van der Ploeg und H. J. C. Berendsen. The electric potential of a macromolecule in a solvent: A fundamental approach. *J. Chem. Phys.*, **97**, 144–171 (1991).

[71] D. Horvath, D. van Belle, G. Lippens und S. J. Wodak. Development and parametrization of continuum solvent models. I. Models based on the boundary element method. *J. Chem. Phys.*, **104**, 6679–6695 (1996).

[72] J. Liang und S. Subramaniam. Computation of molecular electrostatics with boundary element methods. *Biophys. Journal*, **73**, 1830–1841 (1997).

[73] A. H. Boschitsch, M. O. Fenley und H. X. Zhou. Fast boundary element method for the linear Poisson–Boltzmann equation. *J. Phys. Chem. B*, **106**, 2741–2754 (2002).

[74] A. J. Bordner 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).

[75] M. D. Altman, J. P. Bardhan, J. K. White und B. Tidor. Accurate solution of multi-region continuum biomolecule electrostatic problems using the linearized Poisson–Boltzmann equation with curved boundary elements. *J. Comp. Chem.*, **30**, 132–153 (2009).

[76] R. Yokota, J. P. Bardhan, M. G. Knepley, L. A. Barba und T. Hamada. Biomolecular electrostatics using a fast multipole BEM on up to 512 GPUs and a billion unknowns. *Comput. Phys. Commun.*, **182**, 1272–1283 (2011).

[77] W. Im, D. Beglov und B. Roux. Continuum solvation model: Computation of electrostatic forces from numerical solutions to the Poisson-Boltzmann equation. *Comput. Phys. Commun.*, **111**, 59–75 (1997).

[78] Q. Lu und R. Luo. A Poisson-Boltzmann dynamics method with nonperiodic boundary condition. *J. Chem. Phys.*, **119**, 11035 (2003).

[79] W. Geng und G. W. Wei. Multiscale molecular dynamics using the matched interface and boundary method. *J. Comput. Phys.*, **230**, 435–457 (2011).

[80] W. C. Still, 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).

[81] M. Schaefer und M. Karplus. A comprehensive analytical treatment of continuum electrostatics. *J. Phys. Chem.*, **100**, 1578–1599 (1996).

[82] D. Bashford und A. Case. Generalized Born models for macromolecular solvation effects. *Ann. Rev. Phys. Chem.*, **51**, 129–152 (2000).

[83] M. S. Lee, F. R. Salsbury Jr. und C. L. Brooks III. Novel generalized Born methods. *J. Chem. Phys.*, **116**, 10606–10614 (2002).

[84] W. Im, M. S. Lee und C. L. Brooks III. Generalized Born model with a simple smoothing function. *J. Comput. Chem.*, **24**, 1691–1702 (2003).

[85] T. Grycuk. Deficiency of the Coulomb-field approximation in the generalized Born model: An improved formula for Born radii evaluation. *J. Chem. Phys.*, **119**, 4817–4826 (2003).

[86] M. Wojciechowski und B. Lesyng. Generalized Born model: Analysis, refinement, and applications to proteins. *J. Phys. Chem. B*, **108**, 18368–18376 (2004).

[87] E. Gallicchio und R. M. Levy. AGBNP: an analytic implicit solvent models suitable for molecular dynamics simulations and high-resolution modeling. *J. Comput. Chem.*, **25**, 479–499 (2004).

[88] G. Sigalov, P. Scheffel und A. Onufriev. Incorporating variable dielectric environments into the generalized Born model. *J. Chem. Phys.*, **122**, 094511 (2005).

[89] G. Sigalov, A. Fenley und A. Onufriev. Analytical electrostatics for biomolecules: Beyond the generalized Born approximation. *J. Chem. Phys.*, **124**, 124902 (2006).

[90] J. Mongan, A. Svrcek-Seiler und A. Onufriev. Analysis of integral expressions for effective Born radii. *J. Chem. Phys.*, **127**, 185101 (2007).

[91] A. Fenley, J. C. Gordon und A. Onufriev. An analytical approach to computing biomolecular electrostatic potential. I. Derivation and analysis. *J. Chem. Phys.*, **129**, 075101 (2008).

[92] C. D. Snow, N. Nguyen, V. S. Pande und M. Gruebele. Absolute comparison of simulated and experimental protein-folding dynamics. *Nature*, **420**, 102–106 (2002).

[93] R. Geney, M. Layten, R. Gomperts, V. Hornak und C. Simmerling. Investigation of salt bridge stability in a generalized Born solvent model. *J. Chem. Theory Comput.*, **2**, 115–127 (2006).

[94] D. R. Roe, A. Okur, L. Wickstrom, V. Hornak und C. Simmerling. Secondary structure bias in generalized Born solvent models: Comparison of conformational ensembles and free energy of solvent polarization from explicit and implicit solvation. *J. Phys. Chem.*, **111**, 1846–1857 (2007).

[95] H. Nguyen, D. R. Roe und C. Simmerling. Improved generalized Born solvent model parameters for protein simulations. *J. Chem. Theory Comput.*, **9**, 2020 – 2034 (2013).

[96] A. Onufriev, D. A. Case und D. Bashford. Effective Born radii in the generalized Born approximation: The importance of being perfect. *J. Comp. Chem.*, **23**, 1297–1304 (2002).

[97] S. Bauer, G. Mathias und P. Tavan. Electrostatics of proteins in dielectric solvent continua. I. An accurate and efficient reaction field description. *J. Chem. Phys.*, **140**, 104102 (2014).

[98] R. Zhou 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).

[99] H. Nymeyer und A. E. García. Simulation of the folding equilibrium of α-helical peptides: A comparison of the generalized Born approximation with explicit solvent. *Proc. Natl. Acad. Sci. U.S.A.*, **100**, 13934–13939 (2003).

[100] A. Okur, L. Wickstrom und C. Simmerling. Evaluation of salt bridge structure and energetics in peptides using explicit, implicit, and hybrid solvation models. *J. Chem. Theory Comput.*, **4**, 488–498 (2008).

[101] S. Chaudhury, M. A. Olson, G. Tawa, A. Wallqvist und M. S. Lee. Efficient conformational sampling in explicit solvent using a hybrid replica exchange molecular dynamics method. *J. Chem. Theory Comput.*, **8**, 677–687 (2012).

[102] H. Sklenar, F. Eisenhaber, M. Poncin und R. Lavery. Including solvent and counterion effects in the force fields of macromolecular mechanics: The field integrated electrostatic approach (FIESTA). In D. L. Beveridge und R. Lavery (Herausgeber), *Theoretical Biochemistry & Molecular Biophysics, 2. Proteins*, S. 317–335. Adenine Press, New York (1991).

[103] B. Egwolf und P. Tavan. Continuum description of solvent dielectrics in molecular-dynamics simulations of proteins. *J. Chem. Phys.*, **118**, 2039–2056 (2003).

[104] M. Stork und P. Tavan. Electrostatics of proteins in dielectric solvent continua. I. Newton's third law marries $q\mathbf{E}$ forces. *J. Chem. Phys.*, **126**, 165105 (2007).

[105] M. Stork und P. Tavan. Electrostatics of proteins in dielectric solvent continua. II. First applications in molecular dynamics dimulations. *J. Chem. Phys.*, **126**, 165106 (2007).

[106] M. Schwörer, B. Breitenfeld, P. Tröster, K. Lorenzen, P. Tavan und G. Mathias. Coupling DFT to polarizable force fields for efficient and accurate Hamiltonian molecular dynamics simulation. *J. Chem. Phys.*, **138**, 244103 (2013).

[107] B. Egwolf und P. Tavan. Continuum description of ionic and dielectric shielding for molecular-dynamics simulations of proteins in solution. *J. Chem. Phys.*, **120**, 2056–2068 (2004).

[108] R. Reichold. *Rechnergestützte Beschreibung der Struktur und Dynamik von Peptiden und ihren Bausteinen*. Dissertation, Fakultät für Physik, Ludwig-Maximilians-Universität (2009).

[109] S. Bauer, P. Tavan und G. Mathias. Electrostatics of proteins in dielectric solvent continua. II. Hamiltonian reaction field dynamics. *J. Chem. Phys.*, **140**, 104103 (2014).

[110] P. Ahlström, A. Wallquist, S. Engström und B. Jönsson. A molecular dynamics study of polarizable water. *Mol. Phys.*, **68**, 563–581 (1998).

[111] C. Császár und P. Pulay. Geometry optimization by direct inversion in the iterative subspace. *J. Mol. Struct.*, **114**, 31 (1984).

[112] V. Schultheis, R. Reichold, B. Schropp und P. Tavan. A polarizable force field for computing the infrared spectra of the polypeptide backbone. *J. Phys. Chem. B*, **112**, 12217–12230 (2008).

[113] S. Bauer, P. Tavan und G. Mathias. Exploring Hamiltonian dielectric solvent molecular dynamics. *Chem. Phys. Lett.*, **612**, 20–24 (2014).

[114] D. E. Tanner, K. Y. Chan, J. C. Phillips und K. Schulten. Parallel generalized Born implicit solvent calculations with NAMD. *J. Chem. Theory Comput.*, **7**, 3635–3642 (2011).

[115] W. Kabsch und C. Sander. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, **22**, 2577–2637 (1983).

[116] P. Tröster, K. Lorenzen und P. Tavan. Polarizable six-point water models from computational and empirical optimization. *J. Phys. Chem. B*, **118**, 1589–1602 (2014).

[117] M. Gouy. Sur la constitution de la charge électrique à la surface d'un électrolyte. *J. Phys. Theor. Appl.*, **9**, 457–468 (1910).

[118] D. L. Chapman. A contribution to the theory of electrocapillarity. *Phil. Mag.*, **25**, 475–481 (1913).

[119] F. Fogolari, P. Zuccato, G. Esposito und P. Viglino. Biomolecular electrostatics with the linearized Poisson–Boltzmann equation. *Biophys. Journal*, **76**, 1–16 (1999).

[120] I. A. Shkel, O. V. Tsodikov und M. T. Record. Complete asymptotic solution of cylindrical and spherical Poisson–Boltzmann equations at experimental salt concentrations. *J. Phys. Chem. B*, **104**, 5161–5170 (2000).

[121] P. Debye und E. Hückel. Zur Theorie der Elektrolyte. *Phys. Zeitschrift*, **24**, 185–206 (1923).

Danksagung

Humor ist, wenn man trotzdem lacht.

Otto Julius Bierbaum,
Yankeedoodle-Fahrt und andere Reisegeschichten

An dieser Stelle möchte ich herzlich all denen danken, die direkt oder indirekt zum Gelingen dieser Arbeit beigetragen haben.[1]

Mein besonderer Dank gilt hier meinem Doktorvater Prof. Dr. Paul Tavan, der mir dieses spannende, aber auch anstrengende und stellenweise frustrierende Thema anvertraut hat. „Anvertraut“ trifft hier auch deshalb den Nagel auf den Kopf, weil er, trotz diverser Fehlschläge und längerer Durststrecken, nie sein Vertrauen in meine Fähigkeiten und seinen Glauben an den Erfolg dieser Arbeit verloren hat. Ich danke ihm für seine ausdauernde Unterstützung, die er mir, nicht nur beim wissenschaftlichen Schreiben, gab und nicht zuletzt für seinen Einsatz bei der Finanzierung meines Projekts. Danke Paul!

Bedanken möchte ich mich auch bei Gerald Mathias, der mir stets mit wertvollen Tips und Ratschlägen zur Seite stand. Seine „nervigen“ Fragen halfen mir, mein Projekt immer wieder aus einem anderen Blickwinkel zu betrachten und er hat so nicht unwesentlich zum Erfolg dieser Arbeit beigetragen. An dieser Stelle sollen auch diverse Besprechungen am chinesischen Turm nicht unerwähnt bleiben, die immer wieder Platz für neue Ideen in unseren Köpfen geschaffen haben. Ohne seine unerschrockene Hilfe bei der Programmierung wäre IPHIGENIE sicher immer noch der Doktoranden-Alptraum, der es unter dem Namen EGO viele Jahre war.

Benjamin Rieff, meinem langjährigen Leidensgenossen, danke ich für viele Jahre gemeinsam durchlebter Höhen und Tiefen im Studium und schließlich im C–Flügel des BMO, der jetzt schon eine Weile D-Flügel heißt. Lieber Benny, auch ich habe nun fertig. Es war ein langer, langer Weg und niemand sagte es wird leicht, aber in diesem Fall bist du vorne weg gegangen und das Licht am Ende des Tunnels war – Gott sei’s gedankt – kein Zug.

Allen Mitarbeitern und Studenten des BMO, die mich über die Jahre begleitet haben gilt mein Dank für das gute Arbeitsklima und die unkomplizierte Hilfe bei allen möglichen und unmöglichen Problemen. Danke Galina Babizki, Thorben Cordes, Robert Denschlag, Cord Elsner, Franziska Graupner, Thomas Hirschberger, Nils Krebs, Florian Lederer, Martin Lingenheil, Konstantin Lorenzen, Christine Lutz, Stephan Maciej, Karl-Heinz Mantel, Alexandra Michaelis, Rudolf Reichold, Gerald Ryseck, Bernhard Schropp, Verena Schultheiss, Rudolf Schwarz, Magnus Schwörer, Martina Stork, Philipp Tröster, Christoph Wichmann und Marianne Widmann-Diermeier.

[1] Die folgende Liste erhebt keinerlei Anspruch auf Vollständigkeit.

Zuletzt danke ich meiner Familie für ihre Unterstützung. Ohne sie hätte ich das hier sicher nie fertig bekommen. Speziell danke ich meiner Frau Rita für ihre fast grenzenlose Geduld, die sie während der letzten Jahre für mich aufgebracht hat, den ein oder anderen spurerhaltenden Tritt und ihre Unterstützung bei allem was sich außerhalb der Mauern des BMO abgespielt hat. All das kann aber nicht im geringsten gegen den kleinen Lorenz anstinken, den sie mir geschenkt hat und dessen noch zahnloses Lächeln mich täglich daran erinnert was wirklich wichtig ist. Danke euch!