A new transfer free energy based implicit solvation model for the description of disordered and folded proteins

THE JOURNAL OF PHYSICAL CHEMISTRY B
A JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

pubs.acs.org/JPCB

Article

A New Transfer Free Energy Based Implicit Solvation Model for the Description of Disordered and Folded Proteins

Published as part of The Journal of Physical Chemistry virtual special issue "Pablo G. Debenedetti Festschrift".

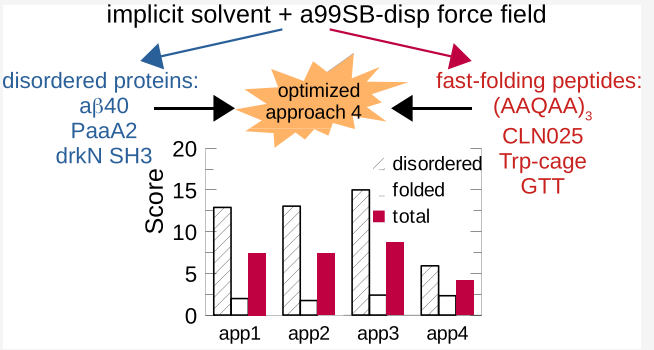
Andrea Arsiccio, Roberto Pisano,* and Joan-Emma Shea*

Cite This: *J. Phys. Chem. B* 2022, 126, 6180–6190

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Most biological events occur on time scales that are difficult to access using conventional all-atom molecular dynamics simulations in explicit solvent. Implicit solvent techniques offer a promising solution to this problem, alleviating the computational cost associated with the simulation of large systems and accelerating the sampling compared to explicit solvent models. The substitution of water molecules by a mean field, however, introduces simplifications that may penalize accuracy and impede the prediction of certain physical properties. We demonstrate that existing implicit solvent models developed using a transfer free energy approach, while satisfactory at reproducing the folding behavior of globular proteins, fare less well in characterizing the conformational properties of intrinsically disordered proteins. We develop a new implicit solvent model that maximizes the degree of accuracy for both disordered and folded proteins. We show, by comparing the simulation outputs to experimental data, that in combination with the a99SB-disp force field, the implicit solvent model can describe both disordered (aβ40, PaaA2, and drkN SH3) and folded ($(AAQAA)_3$, CLN025, Trp-cage, and GTT) peptides. Our implicit solvent model permits a computationally efficient investigation of proteins containing both ordered and disordered regions, as well as the study of the transition between ordered and disordered protein states.



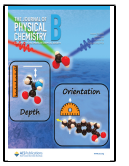
INTRODUCTION

Molecular dynamics (MD) is a powerful technique for the description of molecular-scale phenomena, and its application to the study of proteins has yielded important insights regarding folding and aggregation mechanisms.[1,2] While significant information can be extracted from MD data, and one can, in principle, obtain molecular details that cannot be captured by current experimental techniques,[3] it is nonetheless important to keep in mind two key points that, in practice, limit the ability of MD simulations to successfully capture these details. First, the conformational transitions involved in the phenomena being studied often occur on time scales (microseconds or more) that are challenging to access using conventional MD simulations. Recent advances in high performance and quantum computing[4] have paved the way for the computational investigation of previously unaccessible biological events, but in most cases, the description of complex phenomena still requires techniques that increase computational efficiency by reducing the degrees of freedom of the system,[5,6] or by enhancing the sampling.[7,8] The second factor that needs to be considered is that the accuracy of the results obtained at the end of an MD simulation depends on the accuracy of the physical description (i.e., the force field) used to model the system.[9,10]

One means of achieving computational efficiency in protein simulations is to simplify the representation of the solvent through the use of an implicit solvent model.[11] This approach reduces the degrees of freedom of the system and further enhances conformational transitions by removing viscous effects due to friction with the solvent.[12] However, a challenge with existing implicit solvent models is that they often lack accuracy compared to their explicit solvent counterpart, and in some instances cannot predict certain physical properties due to the absence of explicit solvent molecules.

With the aim of developing and characterizing a simulation approach for proteins that enhances both computational efficiency and accuracy, we recently proposed an implicit solvent model that had the unique feature among implicit solvent models of being able to describe heat-, cold-, and pressure-denaturation of proteins,[13,14] as well as the effect of

Received: June 8, 2022
Revised: July 7, 2022
Published: August 15, 2022



https://doi.org/10.1021/acs.jpcb.2c03980
J. Phys. Chem. B 2022, 126, 6180–6190

osmolytes on protein stability.[15] Our implicit solvent approach combines a Generalized-Born (GB) method to model electrostatic interactions[16] with the commonly used SASA-based description of the nonpolar solvation term, based on arguments from scaled particle theory.[17,18] The SASA-based description defines the nonpolar hydration contribution as a surface tension value γ that multiplies the solvent accessible surface area (SASA) of the protein. However, while most implicit solvent models use a single value of γ (independent of both temperature and atom type), we proposed a more complex treatment of the nonpolar solvation term that involves temperature- and residue-dependent values of surface tension.[13] The relation between γ and temperature was obtained by mining a large data set of protein structures resolved by nuclear magnetic resonance (NMR) in the range 265−335 K. Specifically, we proposed two different mathematical derivations to take into account the effect of temperature. In the first one, the nonpolar solvation term is kept unaltered at 298 K, and the surface tension values are used to describe differences in behavior compared to this reference temperature. In the second case, on the other hand, a completely new set of surface tension values has been derived at any value of temperature.[13] We showed that a model with a constant surface tension term could fold globular proteins at room temperature but was unable to capture cold unfolding. Our models with temperature-dependent surface tension terms, on the other hand, captured both cold and hot denaturation, as well as protein stability at room temperature.[13]

In this paper, we further refine the implicit model to accurately describe not only the free energy landscape of folding of globular proteins, but also that of intrinsically disordered proteins. Achieving a correct description of natively ordered and disordered proteins is a challenge, even in simulations in which explicit solvent is used. Previous studies showed that combinations of popular force fields and water models predicted disordered states that were too compact.[19] More realistic dimensions of unfolded states of proteins could be obtained by rescaling protein−water interactions in the a03w force field.[20] Piana et al.[19] found that a more balanced description of dispersion and electrostatic interactions in the water models used for MD simulations improved the prediction of disordered states, but sometimes worsened the description of folded proteins. The same research group[9] further addressed this issue by optimizing torsion parameters and introducing changes in the protein/water van der Waals interaction terms, eventually developing a force field (a99SB-disp) that displays an optimal balance between disordered and folded protein state prediction.

To assess our implicit solvent model, we consider a selection of different protein models, including both disordered (aβ40, PaaA2, and drkN SH3) and folded ($(AAQAA)_3$, CLN025, Trp-cage, and GTT) peptides. The implicit solvent technique is combined with the a99SB-disp force field[9] for the proteins, as this was shown to outperform previous force fields for the description of both disordered and folded protein structures.

More details on the mathematical derivation and potential applications of our approach can be found in the original publication[13] and in the Theoretical Background section below. Simulations are performed using the parallel tempering replica exchange molecular dynamics (PT-REMD) approach,[21] that enhances sampling by periodically exchanging configurations between a ladder of replicas at different temperatures. The conformational ensembles of the disordered peptides and the temperature stability of the proteins selected for this work are extracted from the simulations, and the simulation outputs are compared with explicit solvent simulations and with experimental data, to quantitatively assess the degree of accuracy of the proposed description. We first considered the original implicit models (that differ in the way that the surface tension term is parametrized) that we introduced in our earlier paper in the context of folding globular proteins and assessed their ability to capture the conformational states of disordered proteins. This allows us to highlight both the advantages and the disadvantages of the existing implicit solvent approaches and enabled us to develop an optimized implicit solvent model that maximizes the accuracy for both disordered and folded protein states.

MATERIALS AND METHODS

Theoretical Background. The free energy of a peptide dissolved in water can be described as the sum of various terms, involving both polar and nonpolar contributions,

$$G^{\mathrm{tot}} = E^{\mathrm{vac}} + G^{\mathrm{el}} + G_0^{\mathrm{np}} + G^{\mathrm{tr}}(T) \tag{1}$$

Here, E^{vac} is the energy of the peptide in vacuum, i.e., the sum of both internal bonded contributions (bonds, angles, dihedrals) and nonbonded van der Waals interactions. G^{el} and G_0^{np} represent instead the polar and nonpolar contributions to the free energy of hydration, respectively. The nonpolar contribution G_0^{np} was developed to describe protein behavior at ambient temperature (T_0 = 298 K) and is generally expressed as

$$G_0^{\mathrm{np}} = \gamma_0 \mathrm{SASA} \tag{2}$$

where γ_0 is a surface tension, independent of temperature and residue type, while SASA is the total solvent accessibility of the protein. A value γ_0 = 5 cal mol^{-1} Å^{-2}[22] was used in the present work.

We recently proposed the addition of a further energy term $G^{\mathrm{tr}}(T)$[13] that depends on temperature T and improves the description of the temperature dependence of protein stability, allowing, for instance, the prediction of cold denaturation in implicit solvent trajectories.

Briefly, the term $G^{\mathrm{tr}}(T)$ was obtained by mining a large set of pdb (Protein Data Bank) files obtained at different temperatures by NMR. The probability of each side chain and of the backbone atoms to be surface exposed at different temperatures was extracted from the pdb files and converted to a temperature-dependent free energy term. In our previous study,[13] we proposed three different approaches for the description of protein folding at different temperatures.

Approach 1 is the standard implementation, where a single, temperature-independent value of surface tension is employed. Approach 2 preserves the standard implementation at T_0 = 298 K but makes use of different values of surface tension γ at $T \neq T_0$. A completely new set of γ values is instead employed in approach 3.

In summary, G_0^{np} and $G^{\mathrm{tr}}(T)$ are described according to the following expressions depending on the approach considered.

Approach 1:

$$G_0^{\mathrm{np}} = \gamma_0 \mathrm{SASA} \tag{3}$$

$$G^{\mathrm{tr}}(T) = 0 \quad \forall\ T \tag{4}$$

https://doi.org/10.1021/acs.jpcb.2c03980
J. Phys. Chem. B 2022, 126, 6180−6190

Approach 2:

$$G_0^{np} = \gamma_0 SASA \tag{5}$$

$$G^{tr}(T) = \begin{cases} 0, & T = T_0 \\ \sum_{k=1}^{n} (\Delta g_{tr,2}^{sc,k}(T)\alpha^{sc,k}) + \Delta g_{tr,2}^{bb}(T) \sum_{k=1}^{n} \alpha^{bb,k}, & T \neq T_0 \end{cases} \tag{6}$$

where the summation runs over the n residues of the protein, and $\Delta g_{tr,2}^{sc,k\ or\ bb}(T)$ is a free energy term that represents the transfer of a side chain (superscript sc) or backbone (superscript bb) group from T_0 = 298 K to a different temperature T. $\alpha^{sc,k\ or\ bb,k}$ is instead the fractional solvent accessibility, defined as the ratio between the SASA of each side chain or backbone, and the corresponding SASA in the tripeptide GLY-k-GLY

$$\alpha^{sc,k\ or\ bb,k} = \frac{SASA^{sc,k\ or\ bb,k}}{SASA_{GLY\text{-}k\text{-}GLY}^{sc,k\ or\ bb,k}} \tag{7}$$

Approach 3:

$$G_0^{np} = 0 \tag{8}$$

$$G^{tr}(T) = \sum_{k=1}^{n} \Delta g_{tr,3}^{sc,k}(T)\alpha^{sc,k} + \Delta g_{tr,3}^{bb}(T) \sum_{k=1}^{n} \alpha^{bb,k} \ \forall\ T \tag{9}$$

Here the expression is functionally identical to eq 6, but the free energy of transfer terms $\Delta g_{tr,3}^{sc,k\ or\ bb}(T)$ are different, because they were obtained in a different way. Briefly, the $\Delta g_{tr,2}^{sc,k\ or\ bb}(T)$ energies were obtained by keeping unaltered the description of the nonpolar contribution G_0^{np} already implemented in MD simulation suites, while a completely new set of transfer free energies was derived in the case of the $\Delta g_{tr,3}^{sc,k\ or\ bb}(T)$ values. The interested reader can refer to our previous publication[13] for a comprehensive description of the derivation and meaning of the free energy terms $\Delta g_{tr,2}^{sc,k\ or\ bb}(T)$ and $\Delta g_{tr,3}^{sc,k\ or\ bb}(T)$. Moreover, a script for the computation of the free energy of transfer terms $\Delta g_{tr,2}^{sc,k\ or\ bb}(T)$ and $\Delta g_{tr,3}^{sc,k\ or\ bb}(T)$ is freely available at https://github.com/andrea-arsiccio/DeltaG-calculation.

Simulation Details. *Model Systems.* The objective of the present work was to compare the different implicit solvent approaches described in the Theoretical Background section. For this purpose, different protein models were selected, as detailed in Table 1.

- aβ40: aβ40 is a disordered protein fragment associated with Alzheimer's disease.[23] It was simulated at pH 8, in 20 mM salt at 298 K, using either the Amber 99SB-disp force field[9] (sim 1) or the Amber 99SB-ILDN description[24] (sim 8). The trajectories were started from an extended conformation (Figure 1A).
- PaaA2: PaaA2 is an intrinsically disordered antitoxin from the human pathogen *E. coli* O157.[25] It was simulated at pH 6.6, in 500 mM salt at 298 K, using either the Amber 99SB-disp force field[9] (sim 2) or the Amber 99SB-ILDN description[24] (sim 9). The trajectories were started from pdb 3ZBE[25] (Figure 1B).
- drkN SH3: The N-terminal SH3 domain of the *Drosophilia* signal transduction protein drk (drkN SH3) is another disordered protein.[26] It was simulated at pH 6, in 50 mM salt at 278 K, using either the Amber 99SB-disp force field[9] (sim 3) or the Amber 99SB-ILDN description[24] (sim 10). The trajectories were started from an extended configuration (Figure 1C).
- $(AAQAA)_3$: $(AAQAA)_3$ is a short peptide with a known α-helix structure.[27] It was simulated at neutral pH, in the temperature range 270.0−350.1 K using the PT-REMD technique.[21] It was capped by an acetyl group and an amide moiety at the N- and C-termini, respectively, and was described using the Amber 99SB-disp force field[9] (sim 4). The initial configuration for the trajectories was fully extended (Figure 1D).
- CLN025: CLN025 is a 10-residue peptide with known β-sheet structure.[28] It was simulated at neutral pH, in the temperature range 278.0−363.9 K using the PT-REMD technique and the Amber 99SB-disp force field[9] (sim 5). The initial configuration was taken from pdb 5AWL[28] (Figure 1E).
- Trp-cage: Trp-cage is a well-known model protein, with only 20 amino acids, displaying a α-helical[29] secondary structure. It was simulated at neutral pH, in the temperature range 285.0−347.9 K using the PT-REMD technique and the Amber 99SB-disp force field[9] (sim 6). The initial configuration was taken from pdb 1L2Y[29] (Figure 1F).
- GTT: GTT is a 35-residue fast-folding variant of the WW domain FiP35,[30] with a β-sheet secondary structure. It was simulated at neutral pH, in the temperature range 290.0−369.7 K using the PT-REMD technique and the Amber 99SB-disp force field[9] (sim 7). The initial configuration was taken from a mutated variant (N26G, A27T, and S28T) of pdb 2F21[31] (Figure 1G).

Table 1. List of the Simulations Performed in This Work

sim no.	protein	force field	duration, ns	temperature, K
1	aβ40	a99SB-disp	250	298
2	PaaA2	a99SB-disp	250	298
3	drkN SH3	a99SB-disp	250	278
4	$(AAQAA)_3$	a99SB-disp	300	270.0, 294.8, 321.5, 350.1
5	CLN025	a99SB-disp	300	278.0, 304.5, 333.1, 363.9
6	Trp-cage	a99SB-disp	300	285.0, 304.8, 325.7, 347.9
7	GTT	a99SB-disp	300	290.0, 304.7, 319.9, 335.9, 352.4, 369.7
8	aβ40	a99SB-ILDN	250	298
9	PaaA2	a99SB-ILDN	250	298
10	drkN SH3	a99SB-ILDN	250	278

Simulation Approach. In all cases, the protonation state of the proteins was adjusted using the H++ server, version 3.2 (http://newbiophysics.cs.vt.edu/H++/[32]), and the simulations were performed using all the three different implicit solvent approaches described in the Theoretical Background section. The implicit solvent results were then compared to the explicit solvent outputs obtained with the a99SB-disp force field

https://doi.org/10.1021/acs.jpcb.2c03980
J. Phys. Chem. B 2022, 126, 6180−6190

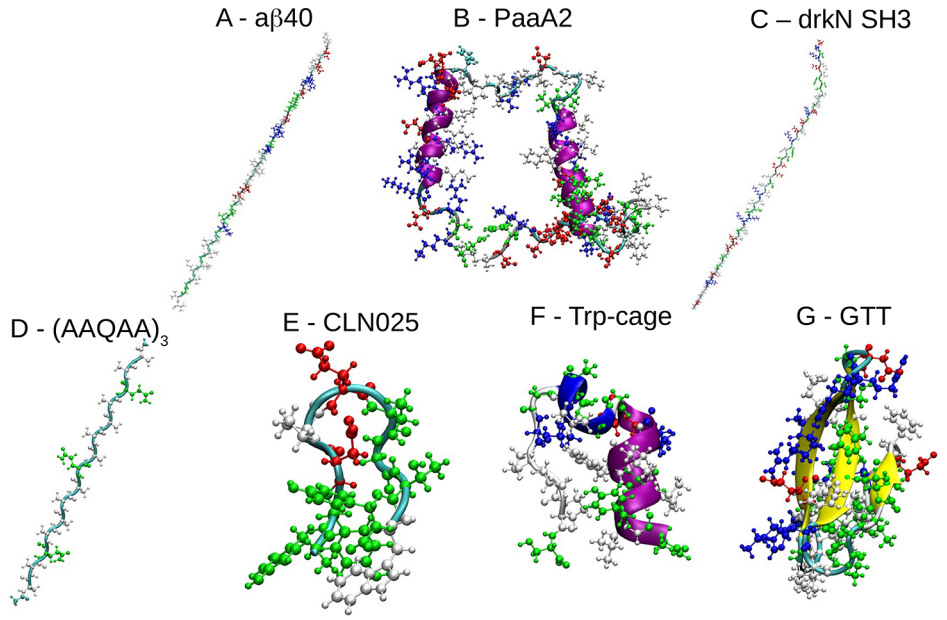


Figure 1. Starting configurations of the different proteins simulated in this work. The cartoon representation highlights the presence of different secondary structure patterns (purple, α-helix; yellow, β-sheet; blue, 3_{10}-helix), while the stick and ball representation distinguishes between polar (green), nonpolar (white), and positively (blue) and negatively (red) charged amino acids.

described in ref 9. The trajectories were run using the AMBER 20 simulation suite,[33] combined with Plumed 2.4.7.[34] The generalized Born/surface area model of AMBER 20 was used to model the first three terms of eq 1, with the OBC(II) model[16] being employed to estimate the Born radii (IGB = 5). The free energy of transfer term $G^{tr}(T)$ was instead added as an external bias using the module SASA of Plumed, which is available starting from version 2.8.

The SASA module implements two different algorithms for the computation of the fractional solvent accessibility $\alpha^{sc,k \text{ or } bb,k}$ in the $G^{tr}(T)$ term, i.e., either the fast algorithm by Hasel et al.[35] or the linear combination of pairwise overlaps (LCPO) algorithm.[36] The Supporting Information file presents a comparison between the two different algorithms, performed on a statistically relevant database of protein configuration files (including 720 structures from the pdb database). The comparison shows that the LCPO algorithm is more accurate but significantly slower than the simplified formula by Hasel et al. Since the fractional solvent accessibility has to be computed at each time step during the implicit solvent trajectories, the faster algorithm by Hasel et al. was employed in this work for the sake of computational efficiency. The effect of temperature on the dielectric constant was also taken into account, using the equations proposed in ref 37 for this purpose.

In all cases, temperature was controlled using Langevin dynamics, with a collision frequency of 1.0 ps^{-1}. All bonds linking to hydrogen atoms were constrained using the SHAKE algorithm.[38] The centers of mass translation and rotation were removed every 500 steps (1 ps), and no cutoff was used for the Coulombic and Lennard-Jones interactions. A time step of 2.0 fs was used, and configurations were saved every 2 ps.

All of the systems were first energy minimized for 3000 steps using the steepest descent algorithm and subsequently simulated for the duration listed in Table 1. In all cases, the first 100 ns were deemed as an equilibration and not considered for the subsequent analyses.

The PT-REMD approach[21] was used for simulations 4−7, and the temperature values used for each replica are listed in the last column of Table 1. Each replica was first equilibrated for 5 ns at the desired temperature, without exchanging configurations. After that, 4.5 ns long REMD runs, where replica swaps were attempted every 3 ps, were performed, in order to check the average exchange probabilities. The average acceptance ratio was adjusted to a value between 0.2 and 0.3. Finally, REMD simulations were carried out for 300 ns, again at a 3 ps exchange frequency.

Analysis of the Trajectories. *Cluster Analysis.* The peptide conformations during the equilibrated trajectories were grouped together by performing a cluster analysis based on the Daura algorithm.[39] The conformations were grouped together if the root-mean-square deviations (RMSD) of the N−C_α−C atoms were less than a given cutoff to each other (0.20 nm for aβ40, 0.15 nm for PaaA2 and drkN SH3, 0.35 nm for $(AAQAA)_3$, 0.25 nm for CLN025, 0.35 nm for Trp-cage, and 0.45 nm for GTT). The most probable conformations were subsequently visualized using VMD (Visual Molecular Dynamics).[40]

α-Helix and Parallel/Antiparallel β-Sheet Content. The α-helix or parallel/antiparallel β-sheet (β) content used in the remainder of this work is defined as the number of 6 residue sections of the peptide having an α-helical or parallel/antiparallel β-sheet configuration,[41]

$$\alpha/\beta = \sum_{\mu} g[r_{\text{dist}}(\{R_i\}_{i\in\Omega_\mu}, \{R^0\})] \tag{10}$$

The summation runs over all possible segments involved in the α-helix or parallel/antiparallel β-sheet, while $\{R_i\}_{i\in\Omega_\mu}$ are the atomic coordinates of a set Ω_μ of 6 residues of the protein, and $g(r_{\text{dist}})$ is the following switching function

https://doi.org/10.1021/acs.jpcb.2c03980
J. Phys. Chem. B 2022, 126, 6180−6190

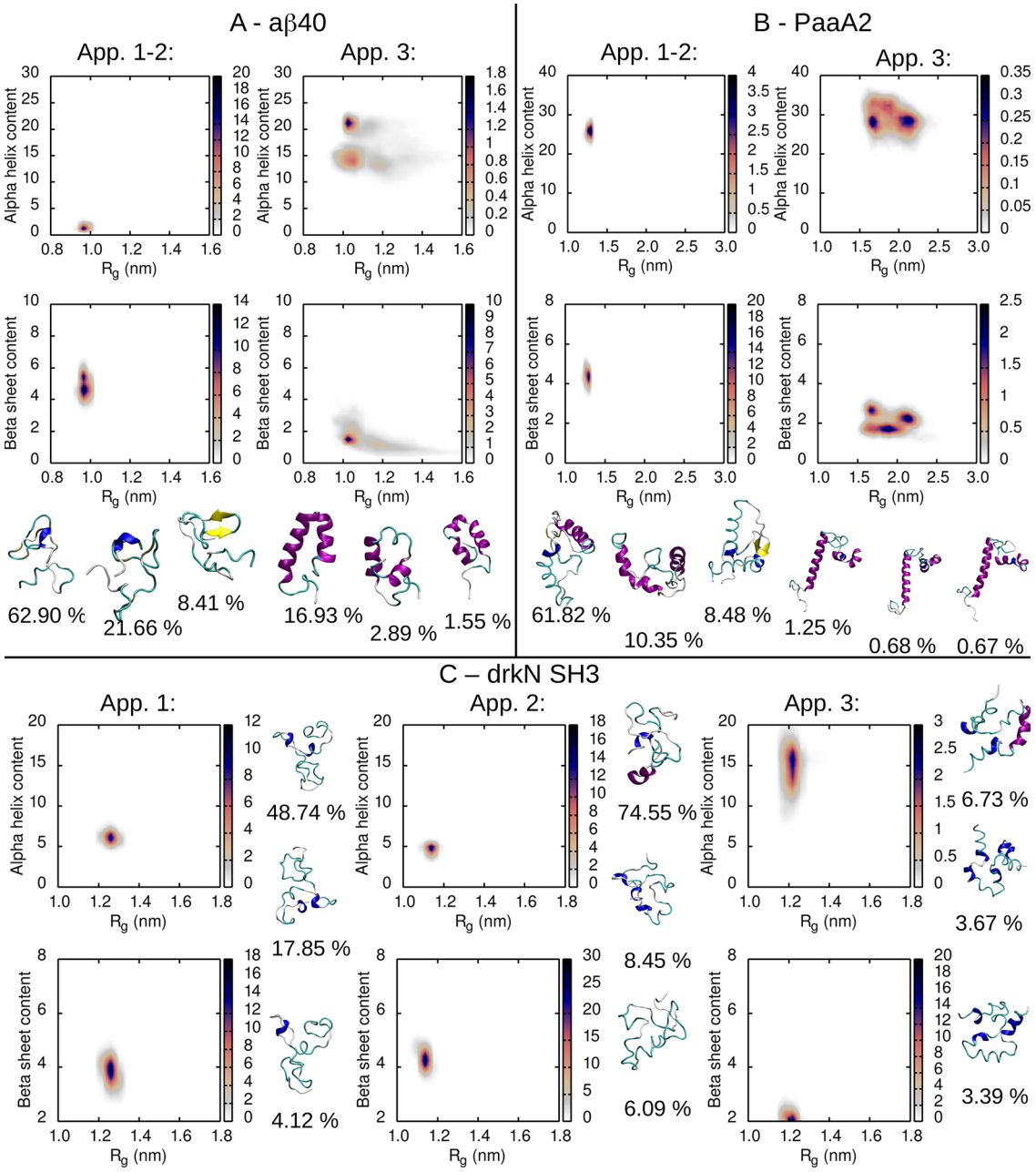


Figure 2. Distribution of α-helix and β-sheet content vs radius of gyration (R_g), and the most likely protein configurations, with corresponding probabilities (as determined using the Daura algorithm[39]) for (A) aβ40, (B) PaaA2, and (C) drkN SH3. The implicit solvent approaches 1, 2 or 3, in combination with the a99SB-disp force field for the proteins, have been considered. In the cartoon representation of the proteins, α-helices are in purple, β-sheets in yellow, and 3_{10}-helices in blue.

$$g(r_{dist}) = \frac{1 - \left(\frac{r_{dist}}{r_0}\right)^8}{1 - \left(\frac{r_{dist}}{r_0}\right)^{12}} \quad (11)$$

A cutoff distance of r_0 = 0.08 nm was used, and r_{dist} is the distance RMSD with respect to a reference α-helix or parallel/antiparallel β-sheet structure $\{R^0\}$.

Folded Fraction. The folded fractions of $(AAQAA)_3$, CLN025, Trp-cage, and GTT during the equilibrated trajectories (last 200 ns) were extracted from the backbone root-mean-square deviation (RMSD) compared to a reference structure (the reference structures were the pdb files 5AWL[28] for CLN025, 1L2Y[29] for Trp-cage, the mutated file 2F21[31] for GTT, and the most sampled structure at 270 K when using approach 1 for $(AAQAA)_3$). The peptides were deemed to be folded when the backbone RMSD was less than a given cutoff (0.35 nm for $(AAQAA)_3$, 0.25 nm for CLN025, 0.35 nm for Trp-cage, and 0.45 nm for GTT).

RESULTS AND DISCUSSION

Evaluation of the Original Implicit Solvent Models: Approaches 1 and 2 Predict Too Collapsed Ensembles for Disordered Peptides, while Approach 3 Overestimates their Helical Propensity. As a first objective, we compared the three implicit solvent approaches described in the Theoretical Background section, using a set of

https://doi.org/10.1021/acs.jpcb.2c03980
J. Phys. Chem. B 2022, 126, 6180−6190

Table 2. Summary of the Results Obtained in This Work for the Implicit Solvent Approaches 1−4, in Combination with the a99SB-disp Force Field[a]

protein	parameter	app. 1	app. 2	app. 3	app. 4	a99SB-disp in explicit solvent	exptl
aβ40[b]	R_g, nm	0.96 ± 0.003	0.96 ± 0.003	1.15 ± 0.047	1.03 ± 0.005	1.39	1.20 ± 0.13
	$RMSD_\beta$	0.198	0.198	0.081	0.122	0.093	
	$RMSD_{helix}$	0.127	0.127	0.713	0.404	0.016	
PaaA2[c]	R_g, nm	1.27 ± 0.005	1.27 ± 0.005	1.90 ± 0.068	1.47 ± 0.017	2.14	2.24 ± 0.4
	$RMSD_\beta$	0.182	0.182	0.116	0.003	0.015	
	$RMSD_{helix}$	0.309	0.309	0.424	0.294	0.312	
drkN SH3[d]	R_g, nm	1.25 ± 0.004	1.13 ± 0.002	1.39 ± 0.163	1.31 ± 0.003	1.95	1.41 ± 0.05
	$RMSD_\beta$	0.009	0.010	0.005	0.005	0.047	
	$RMSD_{helix}$	0.230	0.278	0.290	0.448	0.087	
$(AAQAA)_3$[e]	$RMSD_{folded\text{-}fraction}$	0.555	0.468	0.442	0.453	0.165	
CLN025[f]	$RMSD_{folded\text{-}fraction}$	0.631	0.632	0.627	0.635	0.581	
Trp-cage[g]	$RMSD_{folded\text{-}fraction}$	0.173	0.135	0.368	0.335	0.501	
GTT[h]	$RMSD_{folded\text{-}fraction}$	0.411	0.381	0.581	0.553	0.190	

[a]The results obtained with the a99SB-disp force field in explicit solvent (ref 9) are also shown for comparison. [b]Experimental data from refs 23 and 42. [c]Experimental data from refs 25. [d]Experimental data from ref 26 and 43. [e]Experimental data from ref 27. [f]Experimental data from ref 28. [g]Experimental data from ref 29. [h]Experimental data from ref 30.

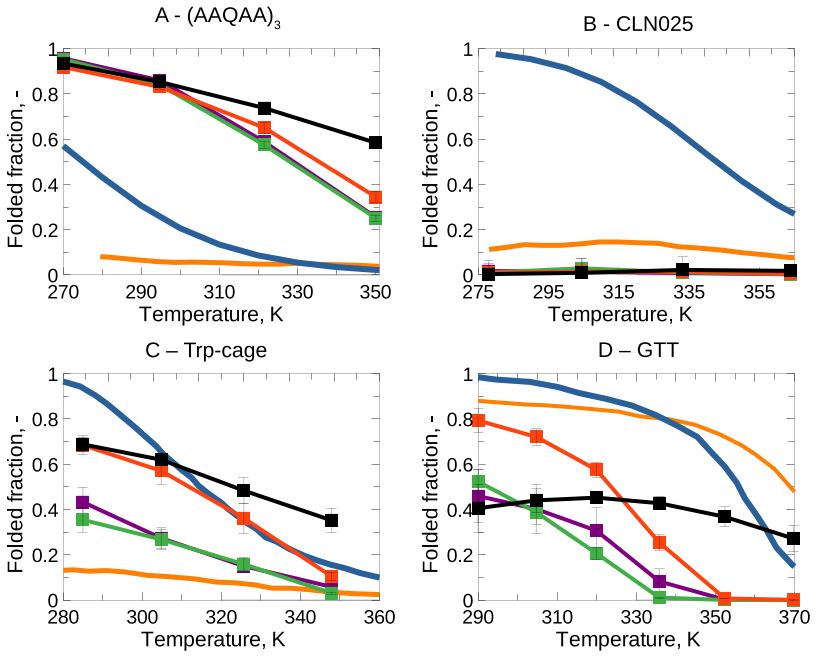


Figure 3. Folded fraction of (A) $(AAQAA)_3$, (B) CLN025, (C) Trp-cage, and (D) GTT as a function of temperature. Black line: approach 1. Red line: approach 2. Green line: approach 3. Purple line: approach 4. Orange line: original a99SB-disp force field in excplicit solvent. Blue line: experimental values. Errors were estimated by block averaging. Briefly, the equilibrated trajectories were divided into 4 blocks, and the standard deviation was computed over the average values of the folded fraction in each of the blocks.

disordered (aβ40, PaaA2, and drkN SH3) and fast-folding ($(AAQAA)_3$, CLN025, Trp-cage, and GTT) peptides as model systems.

Figure 2 and Figure S2 show how the three approaches behave in describing the free energy landscape of disordered proteins (aβ40, PaaA2, and drkN SH3), when combined with the a99SB-disp force field. Approaches 1 and 2 are identical at T_0 = 298 K, and for this reason their outputs are displayed together for aβ40 and PaaA2.

The radius of gyration (R_g) and the secondary structure content (in terms of β-sheets and helices) were evaluated for each protein and each implicit solvent approach (Figure 2), and the most likely conformations sampled in each condition were extracted from the trajectories. The β-sheet and helical propensities as a function of residue number (Figure S2) were also compared to the experimental curves, and the corresponding values of root-mean-square deviation ($RMSD_\beta$ and $RMSD_{helix}$, respectively) were extracted from this comparison (Table 2).

We observed that approach 3 predicts conformational ensembles of the proteins that are, on average, more expanded (larger R_g) compared to approaches 1 and 2. From the point of view of the radius of gyration, approach 3 is more in line with the experimental data, as shown in Table 2, while approaches 1 and 2 tend to predict too collapsed conformations of the peptides. However, when approach 3 was combined with the a99SB-disp force field for the protein, a significant overstimation of the helical content of aβ40 and drkN SH3 was observed (accompanied by a corresponding decrease in β-sheets, as shown in Figure 2A,C).

https://doi.org/10.1021/acs.jpcb.2c03980
J. Phys. Chem. B 2022, 126, 6180−6190

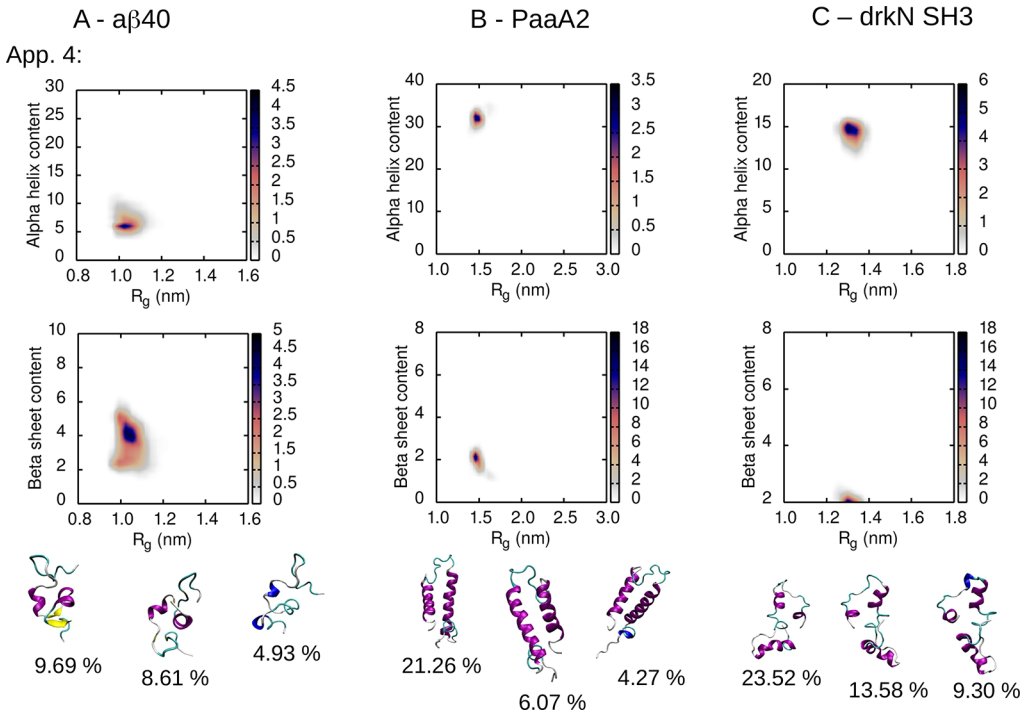


Figure 4. Distribution of α-helix and β-sheet content vs radius of gyration (R_g), and the most likely protein configurations, with corresponding probabilities (as determined using the Daura algorithm[39]) for (A) aβ40, (B) PaaA2, and (C) drkN SH3. The implicit solvent approach 4, in combination with the a99SB-disp force field for the proteins, has been used for these simulations. In the cartoon representation of the proteins, α-helices are in purple, β-sheets in yellow, and 3_{10}-helices in blue.

When comparing the helical propensities predicted by approach 3 to the experimental curves (Figure S2), the values of $RMSD_{helix}$ obtained were, unfortunately, noticeably higher than for approaches 1 and 2 (Table 2). We further observed that this overestimation of the helical propensity was due to the specific combination of approach 3 with the a99SB-disp force field. When approach 3 was instead combined with the a99SB-ILDN description for the protein (simulations 8−10), no overestimation of the helicity of aβ40 and drkN SH3 could be detected, as illustrated in Figures S3 and S4. In line with this, the combination a99SB-ILDN/approach 3 resulted in small values of $RMSD_{helix}$ for aβ40 and drkN SH3 (Table S2).

Approach 2 Shows the Best Agreement with the Temperature-Dependent Stability of Fast-Folding Peptides. For what concerns the fast-folding peptides $(AAQAA)_3$, CLN025, Trp-cage, and GTT, Figure 3 shows their folded fraction, as a function of temperature, as predicted by the three different implicit solvent approaches, as well as by the original a99SB-disp description in explicit solvent.[9] The folded fraction values over temperature were also compared to the experimental curves,[27−30] and corresponding root-mean-square deviations ($RMSD_{folded\text{-}fraction}$) were computed to quantitatively compare the different approaches (Table 2). Figures S5−S8 further display the evolution of the secondary structure content and radius of gyration for the different peptides studied as a function of temperature, together with the most likely conformations sampled in each condition.

The implicit solvent approaches overestimated the folded fraction of $(AAQAA)_3$ (Figure 3A) and GTT (Figure 3D), while they underestimated the stability of CLN025 (Figure 3B) and Trp-cage (Figure 3C). The implicit solvent description improved the prediction of Trp-cage stability compared to the explicit solvent counterpart (Figure 3C), but it was not as effective as the explicit solvent model in matching the experimental values for GTT (Figure 3D).

Overall, we found that approach 2 was generally the best at predicting the temperature stability of fast-folding peptides (as evidenced by the lowest $RMSD_{folded\text{-}fraction}$ values in Table 2), in line with what was observed already in our previous study.[13] Approach 3 generally was the least accurate, as it predicted slightly decreased temperature stability of proteins compared to approaches 1 and 2.

It is interesting to observe that both explicit and implicit solvent models still fail in providing a perfect agreement with the temperature dependence of protein stability. This is due to the intrinsic difficulty of simulating such systems where the number of degrees of freedom, including protein−protein and protein−water interactions as well as their temperature dependence, is huge. Identifying which of these several degrees of freedom is key for reproducing the experimental behavior is a challenging task that still lacks a definitive answer.

Developing a New Implicit Solvent Approach: Approach 4. The results shown so far indicate that approaches 1 and 2 predict too collapsed conformations for the disordered peptides, while approach 3 slightly underestimates the stability of folded proteins and, when combined with the a99SB-disp force field, is biased toward the helical secondary structure. We therefore set out to develop a new implicit solvent approach, in the following referred to as approach 4, that could allow a good description of both disordered and folded protein states.

For this purpose, we started from approach 3 with the objective to understand what contribution to the solvation term $G^{tr}(T)$ in eq 9 mostly led to the overestimation of helicity when combined with the a99SB-disp force field. Looking at eq 9, the two main contributions come from the side chains (first

https://doi.org/10.1021/acs.jpcb.2c03980
J. Phys. Chem. B 2022, 126, 6180−6190

summation) and from the backbone (second summation), respectively. We therefore ran short-time simulations (20 ns) for aβ40 at 298 K and using the a99SB-disp description for the protein, during which the nonpolar solvation term was described mixing approach 3 with approach 1. Specifically, we considered three possible combinations: (i) approach 3 for the backbone, and approach 1 for the apolar side chains (Ala, Ile, Leu, Met, Phe, Pro, Trp, Tyr, and Val); (ii) approach 3 for the backbone, and approach 1 for the polar side chains (Arg, Asn, Asp, Cys, Gln, Glu, Gly, His, Lys, Ser, and Thr); (iii) approach 3 for the side chains, and approach 1 for the backbone (Figure S9A).

The helical content of aβ40 rapidly increased when approach 3 was used to describe solvation of the side chains, while approach 1 was employed for the backbone (blue curve in Figure S9A). In contrast, when approach 1 was used for either the polar or apolar side chains, while approach 3 described the backbone solvation, no noticeable increase in helicity was observed (red and green curves in Figure S9A). This suggests that the side chain solvation is mostly involved in the incorrect prediction of helicity when approach 3 is combined with the a99SB-disp force field. This likely occurs because approach 3 predicts a stronger water−side chain interaction compared to approaches 1 and 2. This leads to less collapsed conformations, in line with experiments, but it also fosters the formation of helices.

We further asked ourselves if some side chains contributed more than the others to the overestimation of helicity. For this purpose, we conducted additional short-time simulations (again for aβ40 at 298 K and using the a99SB-disp force field for the protein) where approach 1 was employed for a single side chain (the selected side chain was different for the different simulations performed), while approach 3 was used for the backbone and all the remaining side chains. The results of this analysis are displayed in Figure S9B,C and suggest that Asn, Asp, Ile, Leu, and Lys influence the α-helix content of aβ40 the most.

On the basis of this preliminary analysis, we decided to define the new implicit solvent approach 4 as a combination of: (i) approach 1 for Asn, Asp, Ile, Leu, and Lys; and (ii) approach 3 for the backbone and the remaining side chains' solvation. Figure 4 and Figure S2 show the results for the combination a99SB-disp/approach 4 applied to the disordered peptides aβ40, PaaA2, and drkN SH3.

Approach 4 still predicts expanded conformational ensembles for the disordered peptides, but partially corrects for the overestimation of helicity observed when using approach 3. For what concerns the folded peptides $(AAQAA)_3$, CLN025, Trp-cage, and GTT, approach 4 does not behave dramatically differently compared to approach 3 (Figure 3 and Figures S5−S8), although, especially for GTT (Figure 3D), it predicts a slightly higher temperature stability. It is important to note that approach 4 was developed using aβ40 as a model system (Figure S9) and was then tested on a separate and statistically relevant set of model proteins (PaaA2, drkN SH3, $(AAQAA)_3$, CLN025, Trp-cage, and GTT). This should ensure adequate transferability of approach 4 to different proteins. In the following section, a quantitative comparison will be performed between approach 4 and the pre-existing approaches 1−3.

A Quantitative Comparison between the Implicit Solvent Approaches. We defined a score to better compare the different implicit solvent approaches, similar to what was done in ref 9 to compare different force fields. The score provides a quantitative estimate of the ability of the selected approach to describe the experimental data. Specifically, a value of the score close to 1 is indicative of a good prediction of experimental values. The experimental data considered for the disordered peptides aβ40, PaaA2, and drkN SH3 were the radius of gyration R_g and the RMSD between the simulated and experimental secondary structure, for both the helical ($RMSD_{helix}$) and β-sheet ($RMSD_\beta$) fraction. All of these values are listed in Table 2. The score for the disordered peptides was then computed as follows:

$$\text{disordered score} = \frac{1}{3}\sum_{a\beta40,\text{PaaA2},\text{drkNSH3}}\left[\frac{\left(\frac{\text{RMSD}_{\text{helix}}}{\text{min}_{\text{app.}}\text{RMSD}_{\text{helix}}}+\frac{\text{RMSD}_{\beta}}{\text{min}_{\text{app.}}\text{RMSD}_{\beta}}\right)}{2}+R_{\text{g,penalty}}\right] \quad (12)$$

where $\text{min}_{\text{app.}}\text{RMSD}_{\text{helix}}$ and $\text{min}_{\text{app.}}\text{RMSD}_{\beta}$ indicate the smallest observed RMSD among all of the different simulation approaches considered in this study, while $R_{\text{g,penalty}}$ is defined as

$$R_{\text{g,penalty}} = \max\left(\frac{|R_{\text{g,exp}} - R_{\text{g,sim}}| - R_{\text{g,exp error}}}{R_{\text{g,exp}}}; 0\right) \quad (13)$$

Here, $R_{\text{g,exp}}$ and $R_{\text{g,sim}}$ are the experimental and simulated radius of gyration, respectively, while $R_{\text{g,exp error}}$ is the error associated with the experimental measurement.

For the folded proteins $(AAQAA)_3$, CLN025, Trp-cage, and GTT, the experimental values considered were the respective folded fractions over temperature. The RMSD between the experimental and simulated values ($RMSD_{\text{folded-fraction}}$, listed in Table 2) was computed for each protein, and the folded score was then defined as

$$\text{folded score} = \frac{1}{4}\sum_{(AAQAA)_3,\text{CLN025},\text{Trp-cage},\text{GTT}}\left[\frac{\text{RMSD}_{\text{folded-fraction}}}{\text{min}_{\text{app.}}\text{RMSD}_{\text{folded-fraction}}}\right] \quad (14)$$

where, again, $\text{min}_{\text{app.}}\text{RMSD}_{\text{folded-fraction}}$ indicates the smallest observed RMSD among all of the different simulation approaches considered in this study.

Finally, a total score was defined as

$$\text{total score} = \frac{\text{disordered score} + \text{folded score}}{2} \quad (15)$$

The results of this analysis are displayed in Figure 5. Approach 3 is worse than approaches 1 and 2 for what

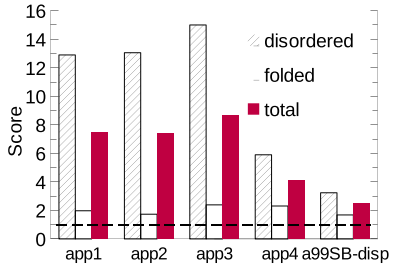


Figure 5. Scores for the four different implicit solvent approaches considered in this study (in combination with the a99SB-disp force field), compared to the score of the original a99SB-disp description in explicit solvent. Values close to 1 (dashed line) are indicative of a good agreement with experimental data.

https://doi.org/10.1021/acs.jpcb.2c03980
J. Phys. Chem. B 2022, 126, 6180−6190

concerns both the disorded and folded peptides, due to the overestimation of helicity and the underestimation of temperature stability.

In contrast, approach 4, while not performing quite as well as approaches 1 and 2 for folded proteins, shows the best (i.e., closest to 1) disordered score among the proposed implicit solvent approaches because it predicts expanded conformational ensembles without dramatically overestimating the helical content. The total score of the combination a99SB-disp/approach 4 (4.10) is not dramatically different from the score of the original a99SB-disp description in explicit solvent (2.46), making approach 4 a good candidate for the description of both disordered and folded peptides.

CONCLUSIONS

In this paper, we have compared the ability of three different implicit solvent approaches (approaches 1−3)[13] to describe the free energy landscape of both disordered (aβ40, PaaA2 and drkN SH3) and folded ($(AAQAA)_3$, CLN025, Trp-cage and GTT) peptides. The implicit solvent description has been combined with the a99SB-disp[9] force field for the proteins, because this force field proved to be the most accurate for both disordered and folded peptides.

We have found that the implicit solvent approach 3, that uses amino-acid-dependent values of surface tensions at any temperature, correctly predicts expanded ensembles for the disordered peptides, but overestimates their helicity and slightly underestimates the stability of folded peptides. Approaches 1 and 2 lead to conformational ensembles of disordered peptides that are too collapsed but show a reasonable agreement with the temperature stability of fast-folding peptides (this is particularly true for approach 2 that, as shown in our previous work,[13] can also predict cold-induced unfolding).

We have further proposed a new implicit solvent approach, namely, approach 4, that results from the combination of approaches 1 and 3, and shows the best agreement with experimental data among the tested implicit solvent approaches. We observed that the role of some side chains, such as Asn, Asp, Ile, Leu, and Lys, is crucial for the correct prediction of helical content when implicit solvent approaches are combined with the a99SB-disp description. Approach 4 predicts expanded conformational ensembles for the disordered peptides, without overestimating their helical content. We therefore recommend its use, in combination with the a99SB-disp force field, as this results in a balanced description of both disordered and folded protein states. Such balanced description could allow a computationally efficient investigation of proteins containing both ordered and disordered regions, or proteins that transition between ordered and disordered states.

The results discussed in this work demonstrate that the implicit solvent approach offers an extremely valid alternative to its explicit solvent counterpart for the simulation of protein folding. While some caveats still remain, the advantages in terms of speed and computational efficiency are unquestionable and extremely appealing. On one hand, the elimination of the solvent drastically decreases the degrees of freedom to be simulated, reducing the number of CPUs necessary for each simulation and increasing the throughput. Second, the elimination of the viscous effects due to friction with the solvent accelerates sampling, making the exploration of the conformational space significantly more efficient. Furthermore, the elimination of a consistent number of degrees of freedom (those due to the solvent) allows the use of a consistently smaller number of replicas in REMD simulations, dramatically reducing the computational cost associated with their execution. In this framework, this work provides some useful suggestions for the implementation of implicit solvent simulations, suggesting a force field/implicit description combination that maximizes accuracy for both folded and disordered peptides.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jpcb.2c03980.

Comparison between different analytical algorithms for the computation of solvent accessibility; β and helical propensities observed with the a99SB-disp force field; results for the disordered peptides aβ40, PaaA2, and drkN SH3 simulated with the a99SB-ILDN description; evolution of secondary structure content, radius of gyration and most likely conformations for $(AAQAA)_3$, CLN025, Trp-cage, and GTT as a function of temperature (PDF)

Text file of the database (TEX)

AUTHOR INFORMATION

Corresponding Authors

Roberto Pisano − *Molecular Engineering Laboratory, Department of Applied Science and Technology, Politecnico di Torino, Torino 10129, Italy;* orcid.org/0000-0001-6990-3126; Email: roberto.pisano@polito.it

Joan-Emma Shea − *Department of Chemistry and Biochemistry, University of California, Santa Barbara, California 93106, United States; Department of Physics, University of California, Santa Barbara, California 93106, United States;* orcid.org/0000-0002-9801-9273; Email: shea@ucsb.edu

Author

Andrea Arsiccio − *Department of Chemistry and Biochemistry, University of California, Santa Barbara, California 93106, United States;* orcid.org/0000-0003-3809-4957

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jpcb.2c03980

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors acknowledge support from the Center for Scientific Computing at the California Nanosystems Institute (CNSI, NSF grant CNS-1725797) and from the hpc@polito team (http://www.hpc.polito.it) for the availability of high performance computing resources and support. This work used the Extreme Science and Engineering Discovery Environment, which is supported by the National Science Foundation grant ACI-1548562 (MCA05S027). The authors aknowledge support from the NSF (MCB-1716956 and CHE-1800352) and the NIH (R01-GM118560-01A).

REFERENCES

(1) Scheraga, H. A.; Khalili, M.; Liwo, A. Protein-Folding Dynamics: Overview of Molecular Simulation Techniques. *Annu. Rev. Phys. Chem.* **2007**, *58*, 57−83.

(2) Morriss-Andrews, A.; Shea, J.-E. Computational Studies of Protein Aggregation: Methods and Applications. *Annu. Rev. Phys. Chem.* **2015**, *66*, 643−666.

(3) Durrant, J. D.; McCammon, J. A. Molecular dynamics simulations and drug discovery. *BMC Biology* **2011**, *9*, 71.

(4) Cheng, H.-P.; Deumens, E.; Freericks, J. K.; Li, C.; Sanders, B. A. Application of Quantum Computing to Biochemical Systems: A Look to the Future. *Front. Chem.* **2020**, *8*, 1066.

(5) Kmiecik, S.; Gront, D.; Kolinski, M.; Wieteska, L.; Dawid, A. E.; Kolinski, A. Coarse-Grained Protein Models and Their Applications. *Chem. Rev.* **2016**, *116*, 7898−7936.

(6) Onufriev, A. In Chapter 7—Implicit Solvent Models in Molecular Dynamics Simulations: A Brief Overview. *Annual Reports in Computational Chemistry*; Wheeler, R. A., Spellmeyer, D. C., Eds.; Elsevier: Amsterdam, Netherlands, 2008; Vol. *4*; pp 125−137.

(7) Bernardi, R. C.; Melo, M. C.; Schulten, K. Enhanced sampling techniques in molecular dynamics simulations of biological systems. *Biochim. Biophys. Acta* **2015**, *1850*, 872−877.

(8) Yang, Y. I.; Shao, Q.; Zhang, J.; Yang, L.; Gao, Y. Q. Enhanced sampling in molecular dynamics. *J. Chem. Phys.* **2019**, *151*, 070902.

(9) Robustelli, P.; Piana, S.; Shaw, D. E. Developing a molecular dynamics force field for both folded and disordered protein states. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115*, E4758−E4766.

(10) Samantray, S.; Yin, F.; Kav, B.; Strodel, B. Different Force Fields Give Rise to Different Amyloid Aggregation Pathways in Molecular Dynamics Simulations. *J. Chem. Inf. Model.* **2020**, *60*, 6462−6475.

(11) Roux, B.; Simonson, T. Implicit solvent models. *Biophys. Chem.* **1999**, *78*, 1−20.

(12) Anandakrishnan, R.; Drozdetski, A.; Walker, R. C.; Onufriev, A. V. Speed of Conformational Change: Comparing Explicit and Implicit Solvent Molecular Dynamics Simulations. *Biophys. J.* **2015**, *108*, 1153−1164.

(13) Arsiccio, A.; Shea, J.-E. Protein Cold Denaturation in Implicit Solvent Simulations: A Transfer Free Energy Approach. *J. Phys. Chem. B* **2021**, *125*, 5222−5232.

(14) Arsiccio, A.; Shea, J.-E. Pressure Unfolding of Proteins: New Insights into the Role of Bound Water. *J. Phys. Chem. B* **2021**, *125*, 8431−8442.

(15) Arsiccio, A.; Ganguly, P.; Shea, J.-E. A Transfer Free Energy Based Implicit Solvent Model for Protein Simulations in Solvent Mixtures: Urea-Induced Denaturation as a Case Study. *J. Phys. Chem. B* **2022**, *126*, 4472−4482.

(16) Onufriev, A.; Bashford, D.; Case, D. A. Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins* **2004**, *55*, 383−394.

(17) Lebowitz, J. L.; Helfand, E.; Praestgaard, E. Scaled Particle Theory of Fluid Mixtures. *J. Chem. Phys.* **1965**, *43*, 774−779.

(18) Reiss, H.; Frisch, H. L.; Lebowitz, J. L. Statistical Mechanics of Rigid Spheres. *J. Chem. Phys.* **1959**, *31*, 369−380.

(19) Piana, S.; Donchev, A. G.; Robustelli, P.; Shaw, D. E. Water Dispersion Interactions Strongly Influence Simulated Structural Properties of Disordered Protein States. *J. Phys. Chem. B* **2015**, *119*, 5113−5123.

(20) Best, R. B.; Zheng, W.; Mittal, J. Balanced Protein-Water Interactions Improve Properties of Disordered Proteins and Non-Specific Protein Association. *J. Chem. Theory Comput.* **2014**, *10*, 5113−5124.

(21) Sugita, Y.; Okamoto, Y. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* **1999**, *314*, 141−151.

(22) Sitkoff, D.; Sharp, K. A.; Honig, B. Accurate Calculation of Hydration Free Energies Using Macroscopic Solvent Models. *J. Phys. Chem.* **1994**, *98*, 1978−1988.

(23) Granata, D.; Baftizadeh, F.; Habchi, J.; Galvagnion, C.; De Simone, A.; Camilloni, C.; Laio, A.; Vendruscolo, M. The inverted free energy landscape of an intrinsically disordered peptide by simulations and experiments. *Sci. Rep.* **2015**, *5*, 15449.

(24) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **2010**, *78*, 1950−1958.

(25) Sterckx, Y. G.; Volkov, A. N.; Vranken, W. F.; Kragelj, J.; Jensen, M. R.; Buts, L.; Garcia-Pino, A.; Jové, T.; Van Melderen, L.; Blackledge, M.; et al. Small-angle X-ray scattering- and nuclear magnetic resonance-derived conformational ensemble of the highly flexible antitoxin PaaA2. *Structure* **2014**, *22*, 854−865.

(26) Choy, W.-Y.; Mulder, F. A.; Crowhurst, K. A.; Muhandiram, D.; Millett, I. S.; Doniach, S.; Forman-Kay, J. D.; Kay, L. E. Distribution of molecular size within an unfolded state ensemble using small-angle X-ray scattering and pulse field gradient NMR techniques11Edited by P. E. Wright. *J. Mol. Biol.* **2002**, *316*, 101−112.

(27) Shalongo, W.; Dugad, L.; Stellwagen, E. Distribution of Helicity within the Model Peptide Acetyl(AAQAA)3amide. *J. Am. Chem. Soc.* **1994**, *116*, 8288−8293.

(28) Honda, S.; Akiba, T.; Kato, Y. S.; Sawada, Y.; Sekijima, M.; Ishimura, M.; Ooishi, A.; Watanabe, H.; Odahara, T.; Harata, K. Crystal Structure of a Ten-Amino Acid Protein. *J. Am. Chem. Soc.* **2008**, *130*, 15327−15331.

(29) Neidigh, J. W.; Fesinmeyer, R. M.; Andersen, N. H. Designing a 20-residue protein. *Nat. Struct. Biol.* **2002**, *9*, 425−430.

(30) Piana, S.; Sarkar, K.; Lindorff-Larsen, K.; Guo, M.; Gruebele, M.; Shaw, D. E. Computational Design and Experimental Testing of the Fastest-Folding β-Sheet Protein. *J. Mol. Biol.* **2011**, *405*, 43−48.

(31) Jäger, M.; Zhang, Y.; Bieschke, J.; Nguyen, H.; Dendle, M.; Bowman, M. E.; Noel, J. P.; Gruebele, M.; Kelly, J. W. Structure−function−folding relationship in a WW domain. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 10648−10653.

(32) Anandakrishnan, R.; Aguilar, B.; Onufriev, A. V. H++ 3.0: Automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. *Nucleic Acids Res.* **2012**, *40*, W537−W541.

(33) Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E.; DeBolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Comput. Phys. Commun.* **1995**, *91*, 1−41.

(34) Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. PLUMED 2: New feathers for an old bird. *Comput. Phys. Commun.* **2014**, *185*, 604−613.

(35) Hasel, W.; Hendrickson, T. F.; Still, W. C. A rapid approximation to the solvent accessible surface areas of atoms. *Tetrahedron Comput. Methodol.* **1988**, *1*, 103−116.

(36) Weiser, J.; Shenkin, P. S.; Still, W. C. Approximate atomic surfaces from linear combinations of pairwise overlaps (LCPO). *J. Comput. Chem.* **1999**, *20*, 217−230.

(37) Bradley, D. J.; Pitzer, K. S. Thermodynamics of electrolytes. 12. Dielectric properties of water and Debye-Hueckel parameters to 350.degree.C and 1 kbar. *J. Phys. Chem.* **1979**, *83*, 1599−1603.

(38) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **1977**, *23*, 327−341.

(39) Daura, X.; Gademann, K.; Jaun, B.; Seebach, D.; van Gunsteren, W. F.; Mark, A. E. Peptide Folding: When Simulation Meets Experiment. *Angew. Chem., Int. Ed.* **1999**, *38*, 236−240.

(40) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graph.* **1996**, *14*, 33−38.

(41) Pietrucci, F.; Laio, A. A Collective Variable for the Efficient Exploration of Protein Beta-Sheet Structures: Application to SH3 and GB1. *J. Chem. Theory Comput.* **2009**, *5*, 2197−2201.

(42) Camilloni, C.; De Simone, A.; Vranken, W. F.; Vendruscolo, M. Determination of Secondary Structure Populations in Disordered

https://doi.org/10.1021/acs.jpcb.2c03980
J. Phys. Chem. B 2022, 126, 6180−6190

States of Proteins Using Nuclear Magnetic Resonance Chemical Shifts. *Biochemistry* **2012**, *51*, 2224−2231.
(43) Marsh, J. A.; Forman-Kay, J. D. Structure and Disorder in an Unfolded State under Nondenaturing Conditions from Ensemble Models Consistent with a Large Number of Experimental Restraints. *J. Mol. Biol.* **2009**, *391*, 359−374.

https://doi.org/10.1021/acs.jpcb.2c03980
J. Phys. Chem. B 2022, 126, 6180−6190