Variational Optimization of an All-Atom Implicit Solvent Force Field To Match Explicit Solvent Simulation Data

Sandro Bottaro,[†,§] Kresten Lindorff-Larsen,*[,†] and Robert B. Best*[,‡,¶]

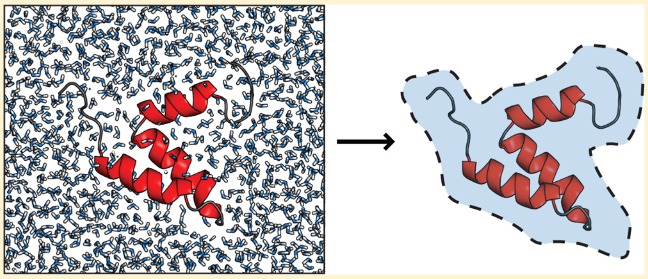
[†]Department of Biology, University of Copenhagen, Copenhagen, Denmark

[‡]Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, United Kingdom

[¶]Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, 9000 Rockville Pike, Bethesda, Maryland, United States

[§]SISSA-Scuola Internazionale Superiore di Studi Avanzati, Trieste, Italy

S *Supporting Information*



ABSTRACT: The development of accurate implicit solvation models with low computational cost is essential for addressing many large-scale biophysical problems. Here, we present an efficient solvation term based on a Gaussian solvent-exclusion model (EEF1) for simulations of proteins in aqueous environment, with the primary aim of having a good overlap with explicit solvent simulations, particularly for unfolded and disordered states − as would be needed for multiscale applications. In order to achieve this, we have used a recently proposed coarse-graining procedure based on minimization of an entropy-related objective function to train the model to reproduce the equilibrium distribution obtained from explicit water simulations. Via this methodology, we have optimized both a charge screening parameter and a backbone torsion term against explicit solvent simulations of an α-helical and a β-stranded peptide. The performance of the resulting effective energy function, termed EEF1-SB, is tested with respect to the properties of folded proteins, the folding of small peptides or fast-folding proteins, and NMR data for intrinsically disordered proteins. The results show that EEF1-SB provides a reasonable description of a wide range of systems, but its key advantage over other methods tested is that it captures very well the structure and dimensions of disordered or weakly structured peptides. EEF1-SB is thus a computationally inexpensive (~10 times faster than Generalized-Born methods) and transferable approximation for treating solvent effects.

■ INTRODUCTION

The aqueous environment plays an important role in determining the structure, function, and dynamics of biomolecules.[1,2] For this reason, the solute-water interaction has been the subject of many recent theoretical, computational, and experimental studies.[3−5] The most realistic way to treat solvation effects in computer simulations is through the inclusion of explicit water molecules. The high level of detail provided by this approach, however, has substantial computational costs, due to the larger system size (usually 90% water molecules) and need to compute long-range forces, which often become prohibitive for molecular systems undergoing significant structural rearrangements. Examples include conformational changes of allosteric proteins, molecular motors, protein folding, and folding/binding of intrinsically disordered proteins.

Implicit solvent representations provide a simplified, although in practice less accurate, alternative to explicit water models for treating solvent effects. In implicit solvent models the average influence of water molecules is described by an effective free energy that depends only on the atomic coordinates of the solute. The formulation of an accurate and computationally efficient description of solvent effects is a nontrivial theoretical problem, and the development of implicit water models for biomolecular simulations has progressed along many different lines of research. There are two main challenges in the development of implicit solvent models, common to force field development in general. The first is to develop a physically motivated approximation to the solvation free energy as a function of atomic coordinates, and the second is to determine, in some sense, the best possible parameters for that function. We discuss these issues in turn below.

One of the simplest approaches to approximating solvation free energy is given by solvent-accessible surface area (SASA) models,[6] in which the solvent effect is taken to be proportional to the area of the solute atom that is accessible to solvent molecules. Typically, the proportionality constants are determined by matching the experimental free energy of hydration of small molecules[6,7] or by reproducing the relative hydrophobic and hydrophilic solvent accessible surface areas in simulations of a selected set of folded proteins, compared with their known crystal structures.[8,9] Even for hydrophobic solutes,

Received: June 5, 2013
Published: November 5, 2013

however, the surface area approximation is questionable, as solvation free energy is only proportional to surface area for solutes larger than ∼1 nm in radius. For small hydrophobic solutes, more typical of protein-like functional groups, solvation free energy is better approximated using the volume of the solute than the surface area;[10,11] this problem has been recently addressed in a refined version of the SASA model.[12]

Another group of methods are the contact models, where the solvation free energy depends on the number of contacts that each atom makes with other solute atoms,[13] and the contacts are weighted according to some function of their distance.[14] This is related to the approach which we consider in more detail here, namely the Gaussian solvent-exclusion models, in which the solvent effects are assumed to be proportional to the volume of the first hydration shell that is accessible to the solvent;[15] interactions with nearby groups exclude solvent and thus reduce the solvation free energy. This approximation has been motivated by theoretical work on solvation thermodynamics[16,17] and is described in more detail below. The solvation free energy of the solute in the absence of any solvent exclusion is derived from empirical solvation free energies for small model compounds. Because of its good accuracy and computational efficiency (only about 50% more computer time with respect to a *vacuum* simulation), the Gaussian solvent-exclusion model EEF1 (hereafter called EEF1-C19)[18] has been applied to a wide range of biological problems.

Several studies have shown EEF1-C19 to provide a reasonably accurate description of solvent effects,[18−22] and it has been shown in certain cases to yield comparable results with respect to explicit water simulations[23] and recently was used in successful applications in protein structure prediction[24,25] and folding studies.[26] Furthermore, it is often used as a component of the highly successful ROSETTA energy function for structure prediction and design.[27] A conceptually similar model based on empirical solvation free energies has been employed in the ABSINTH force field;[28] the principal difference is that in this model the electrostatic interactions are screened by a function which also is dependent on the degree of burial of an atom. Although the short-range contribution to solvation free energy in EEF1-C19 is quite well developed, the treatment of electrostatic interactions through a simple distance-dependent dielectric is much cruder. Another deficiency which has been identified is the treatment of the protein backbone in the underlying force field (CHARMM19), and it has been suggested that EEF1 might be profitably combined with a CHARMM CMAP-style[29] backbone energy function.[22] Given these limitations of EEF1-C19, other implicit solvent approaches have sometimes been found to be superior (e.g., in stabilizing the native structure of a folded protein[30] and in reproducing the unfolding behavior of an amyloid beta fragment[31] − note, however, that stabilization of the structure of a folded protein is not by itself a sufficient test for the quality of an implicit solvent model, as we will show later).

In all of the aforementioned effective potentials, the electrostatic effects are usually crudely approximated (e.g., using a distance-dependent dielectric constant[9,18]) or completely ignored. The effect of the solvent on electrostatic interactions may be more accurately described through continuum electrostatic models, where the solute is assumed to be a low-dielectric cavity immersed in a high-dielectric and featureless environment. The electrostatic (polar) free energy of solvating a molecule is then calculated by solving the Poisson−Boltzmann (PB) equation[32−34] or estimated by using the popular Generalized-Born (GB) equation.[35,36] While a more accurate description of electrostatic solvation free energy, continuum models are nonetheless still an idealization and cannot distinguish features dependent on the molecular details of the solvent (e.g., the difference in solvation free energy for otherwise identical positively and negatively charged ions[37]), without adopting artificial parameter values. It is also worth noting that the computational cost of most PB and GB methods scales extremely poorly with the system size and is comparable to explicit water simulations for large, globular molecules.[38] For this reason, several approximations and variations to the original GB approach have been introduced in order to improve the computational efficiency and the accuracy of the method. Many popular implicit solvent models such as the analytical continuum electrostatic method (ACE)[39] and the fast analytical continuum treatment of solvation (FACTS)[40] as well as the accurate GBSW model[41] all belong to this category. In the screened Coulomb potential implicit solvent model (SCPISM),[42] instead, the electrostatic contribution to solvation free energy is efficiently estimated by employing a distance dependent, sigmoidal dielectric function. While such continuum electrostatic models all provide a theoretical formulation for polar interactions, the nonpolar effects (e.g., hydrophobicity) are modeled by empirical potentials proportional to the solvent-accessible surface area.

In the present work, our goal was to derive a computationally inexpensive implicit solvent model for use in conjunction with the CHARMM36[43] all-atom model. Our principal aim in doing this was to improve the description of unfolded and disordered states, which in current implicit solvent models are much too structured and compact (as our results below will show), making them unsuitable for applications to structure, binding and folding of intrinsically disordered proteins.[44] We also wanted a model which could be used in multiscale applications, for example to perform an initial exploration of the conformation space of interest before explicit simulations are performed. For this purpose, we have chosen to use the EEF1 implicit solvent functional form from the original EEF1-C19 force field,[18] which was based on the united-atom CHARMM19 force field.[45] The significant differences in the molecular representation as well as in the parametrization of the two force fields do not allow a direct transfer of the EEF1-C19 model parameters. Therefore, we devised a modified version of EEF1-C19, which we term EEF1-SB, where the model parameters are adjusted so as to mimic the equilibrium ensemble obtained from explicit water simulations. Following the ideas of Shell and co-workers,[46−49] we minimize an objective function known as the relative entropy, which is a measure of the "overlap" between the ensembles sampled with the implicit and explicit water models. We have recently used this method to obtain additive force fields by an effective coarse-graining of polarizable models.[50] This approach is conceptually similar (although in practice very different) to the force-matching[51] procedure employed by Fraternali and co-workers[52] to optimize the parameters in a SASA implicit solvent model; force matching has also recently been applied to deriving implicit solvent models for ionic solutions from atomistic simulations.[53]

This paper is structured as follows: first, we describe the EEF1-SB effective energy function and briefly outline the relative entropy minimization approach used for parametrization. The procedure is used to determine optimal values for two parameters in EEF1-SB. We initially considered optimization of many more parameters, as discussed, but found that a charge screening parameter and backbone torsion term yielded the greatest improvement in our target function. As "training" data, we used explicit water simulations of the α-helical peptide Ac-(AAQAA)$_3$-NH$_2$ (which we will refer to as (AAQAA)$_3$)[54] and the GB1 β-hairpin.[55,56] The optimization is able to match the distribution for the helical peptide particularly well, although the results for the hairpin are less ideal. For both peptides, the optimal parameters result in an increased sampling of expanded over collapsed structures, in agreement with the explicit solvent data. In this respect, the new force field represents a significant improvement over previous implicit solvent models. Finally, we test the accuracy of EEF1-SB on systems not used in the parametrization, by performing molecular dynamics simulations on the native state of globular proteins and by conducting folding studies on short peptides. EEF1-SB produces stable trajectories when simulating the native state of two globular proteins and good agreement with NMR scalar couplings for the folded state. In the folding studies, EEF1-SB results in near-native conformations of the Trp-cage and WW-domain but fails to fold the helical Villin headpiece protein to the correct structure. Finally, EEF1-SB accurately captures the secondary structure propensity of an unstructured fragment of hen lysozyme, both with respect to experimental data and to the sampling in explicit solvent simulations, yielding much improved accuracy relative to existing implicit solvent models. Overall, our results demonstrate that the presented model is an extremely fast and reasonable approximation for treating solvent effects and holds particular advantages relative to other models when considering disordered or unstructured proteins.

■ THEORY AND METHODS

Description of the Solvent Model. The functional form of the solvent exclusion model in EEF1-SB is identical to the EEF1-C19 approach, which we briefly summarize here.[18] The total solvation free energy of a protein is expressed as a sum of atomic contributions:

$$\Delta G^{\text{solv}} = \sum_i \Delta G_i^{\text{solv}} \tag{1}$$

Each individual term ΔG_i^{solv} is equal to a reference solvation free energy ΔG_i^{ref}, obtained by dissecting into group contributions the experimental free energy for a set of small compounds,[57] minus a reduction due to the presence of surrounding atoms.

$$\Delta G_i^{\text{solv}} = \Delta G_i^{\text{ref}} - \sum_{j \neq i} f_i(r_{ij}) V_j \tag{2}$$

Here, the sum runs over the neighboring atoms j with volume V_j. The solvation free energy density $f_i(r)$ is chosen such that volume integral over the first solvation shell accounts for ≈85% of the solvation energy, in accordance with theoretical and computational studies on fluids.[16,17,58] In their original study,[18] Lazaridis and Karplus found the Gaussian function

$$f_i(r) 4\pi r^2 = \frac{2}{\sqrt{\pi}} \frac{\Delta G_i^{\text{free}}}{\lambda_i} \exp\left\{ -\frac{(r - R_i)^2}{\lambda_i^{\,2}} \right\} \tag{3}$$

to have the desired behavior when setting λ equal to the width of the first solvation shell (λ =3.5 Å). Here, ΔG^{free} is the solvation free energy of the isolated group, and R is the van der Waals radius.

Model Parameters. The atom types used in the original EEF1-C19 implementation are those from the united-atom CHARMM19 force field,[45] in which only hydrogen atoms belonging to polar groups are explicitly included. In the present work we instead use the CHARMM36 all-atom representation.

The reference solvation free energies ΔG^{ref}, taken from the EEF1-C19 model,[18] were originally obtained by dissecting the experimental solvation free energy (at T = 298.15 K) for a set of model compounds into group contributions[57] and subsequently corrected to account for long-range van der Waals effects.[18] A similar procedure was used to determine the solvation enthalpy ΔH[59] and heat capacity ΔC_p.[60] These quantities make it possible to obtain an approximate expression for ΔG^{ref} as a function of the temperature. It is assumed that the solvation free energies and heat capacities can be accurately approximated as a sum over contributions from different chemical groups in a molecule.

The volume of each atom type is calculated as the van der Waals volume, reduced by the overlap volume between covalently bonded atoms[61] (triple or higher overlap volumes are neglected). We have used the CHARMM36 van der Waals radii and bond lengths for the volume calculation, and all hydrogens are assumed not to contribute to the solvation energy: their volumes are therefore set to zero.

Finally, as in the original EEF1-C19 parametrization, the thickness of the hydration shell, λ, is set to 3.5 Å except for the atoms in charged groups, for which a value of 6 Å is used. The chosen values of the solvation parameters are listed in Table S1 of the Supporting Information.

Electrostatic Interactions. The electrostatic screening effect of water is not considered directly by the solvent-exclusion model and is here approximated using a linear, distance-dependent dielectric constant (i.e., $\varepsilon = k_\varepsilon r$). As pointed out by Lazaridis and Karplus,[18] the distance-dependent dielectric constant does not screen the electrostatic interactions for charged groups to a sufficient degree. In order to account for this effect, ionic side-chains are neutralized by adjusting the partial atomic charges, as detailed in Table S2 of the Supporting Information. To account for the strong interactions between atoms in ionic groups and solvent molecules, the correlation length λ for these atom types is set to 6 Å, and the reference solvation free energies ΔG^{ref} are arbitrarily set to large values, in order to increase their hydrophilic propensity. Electrostatic and van der Waals interactions are smoothly switched-off between 7 and 9 Å, and interactions between atoms separated by three covalent bonds (1−4 pairs) are not rescaled.

Parameter Optimization. As we show below, the direct transfer of the original EEF1-C19 model parameters to the all-atom representation used in the CHARMM36 force field does not accurately model the solvent effects, and a suitable reparameterization is necessary. In the present work, the optimal values of the EEF1-SB parameters were determined using the relative entropy approach introduced by Shell and co-workers.[46] This coarse graining technique has previously been successfully applied in different contexts[49] and makes it possible to optimize the parameters in a coarse-grained potential so as to reproduce the equilibrium Boltzmann distribution obtained in reference, all-atom simulations. In this work we used the all-atom CHARMM36 force field in combination with the TIP3P[62] explicit water model for deriving the optimal parameters in the EEF1-SB effective force field.

The approach is based upon variational minimization of the relative entropy, S_{rel} (also known as the Kulback-Leibler divergence), between the configurational ensembles produced in a reference explicit water (E) and implicit water (I) simulation

$$S_{\mathrm{rel}} = \sum_i p_E(i) \log \frac{p_E(i)}{p_I(i)} \tag{4}$$

$p_E(i)$ and $p_I(i)$ are respectively the probability of protein configuration i in the explicit and implicit solvent ensembles, and the index i runs over the all-atom configurations. In the canonical ensemble, the relative entropy is given by

$$S_{\mathrm{rel}} = \beta\langle U_I - U_E\rangle_E - \beta(A_I - A_E) + \langle S_{\mathrm{map}}\rangle_E \tag{5}$$

Here, U_E and U_I are the all-atom and coarse-grained potentials, respectively, $A = -k_B T \log Z$, where Z is the partition function and $\beta = 1/k_B T$ is the inverse temperature T multiplied by the Boltzmann's constant k_B. The mapping entropy $\langle S_{\mathrm{map}}\rangle_E$ is the average entropy that results from degeneracies in the target-model mapping. According to eq 5, calculating S_{rel} requires an impractical estimation of free energies. However, assuming the coarse-grained potential to be a function of some parameters, k, the derivatives of the relative entropy with respect to k can be expressed as simple averages over the two ensembles

$$\begin{aligned} \frac{\partial S_{rel}}{\partial k} &= \beta\left\langle \frac{\partial U_I}{\partial k} \right\rangle_E - \beta\left\langle \frac{\partial U_I}{\partial k} \right\rangle_I \\ \frac{\partial^2 S_{rel}}{\partial k^2} &= \beta\left\langle \frac{\partial^2 U_I}{\partial k^2} \right\rangle_E - \beta\left\langle \frac{\partial^2 U_I}{\partial k^2} \right\rangle_I + \beta^2\left\langle \frac{\partial U_I}{\partial k}^2 \right\rangle_I \\ &\quad - \beta^2\left\langle \frac{\partial U_I}{\partial k} \right\rangle_I^2 \end{aligned} \tag{6}$$

Hence standard numerical techniques can be employed to minimize the relative entropy with respect to the model parameters, for example by iterative application of the Newton–Raphson update rule

$$k_{j+1} = k_j - \gamma\left[\frac{\partial^2 S_{rel}}{\partial k^2}\right]^{-1}\left[\frac{\partial S_{rel}}{\partial k}\right] \tag{7}$$

where k and $k + 1$ are successive values of the parameter and γ is a parameter controlling the step size.

Performing the minimization can be challenging in practical applications. First, because it may be difficult to monitor the progress of the optimization, as the absolute value of the relative entropy in eq 5 cannot be easily calculated. As proposed by Shell and co-workers,[48] an approximate expression for the relative entropy is obtained from eq 5 via standard free energy perturbation[63]

$$S_{\mathrm{rel}} \approx \log(\langle \exp(\Delta - \langle\Delta\rangle_E)\rangle_E) \tag{8}$$

where $\Delta = \beta(U_I - U_E)$. The approximation holds only as long as a substantial overlap exists between the coarse-grained and all-atom ensembles. Moreover, the average of the exponential in eq 8 is dominated by individual contributions with large Δ and can therefore be affected by statistical errors. It is worth highlighting that for parameters linear in the potential U_I, the second derivative of the relative entropy in eq 6 reduces to $(\partial^2 S_{rel})/(\partial k^2) = \beta^2(\langle(\partial U_I^{\,2})/(\partial k)\rangle_I - \langle(\partial U_I)/(\partial k)\rangle_I^2)$. This quantity is positive definite, therefore requiring the gradient $|(\partial S_{rel})/(\partial k)|$ to be zero is a sufficient condition for optimality. As a consequence, for parameters linear in the potential, it is not strictly necessary to monitor the absolute value of the relative entropy during the minimization. Second, the procedure becomes inaccurate when the individual implicit solvent simulations are not sufficiently equilibrated, causing large fluctuations during successive iteration of parameter optimization. Therefore, in the present work we make extensive use of replica exchange molecular dynamics (REMD) simulations.[64]

As formulated here, the relative entropy approach can be easily extended to perform a simultaneous optimization of multiple parameters. The optimization in a high-dimensional space can however be numerically unstable and may lead to overfitting when multiple atom-types are considered. For this reason, in the present work we used a minimal number of adjustable parameters.

Simulation Methods. The explicit water simulations used in this work are taken from the study of Best et al.[43] For ease of reference, we briefly report the simulation conditions. All simulations were performed with the CHARMM36 force field and TIP3P water model using GROMACS 4.5.3.[65] Long range electrostatics were treated using Particle-Mesh Ewald[66] summation with a real-space cutoff of 12 Å and a 1 Å grid spacing, while the Lennard-Jones interactions were treated with a switching function from 10 to 12 Å. The equations of motion were integrated with a 2 fs time step. All bond lengths were constrained using the LINCS algorithm.[67] The Ac-(AAQAA)$_3$-NH$_2$ peptide was solvated with 1833 water molecules in a truncated octahedron cell with a distance between nearest faces of 42 Å. The peptide was first unfolded using a 5 ns constant volume simulation at 800 K. Subsequently, a constant volume replica exchange MD simulation was run, with 32 replicas spanning a temperature range from 278 to 416 K and exchange attempts every 10 ps, for a total of 150 ns per replica, of which only the last 50 ns were used as target in the relative entropy procedure. A Langevin thermostat with a friction coefficient of 1 ps^{-1} was used. The GB1 hairpin (residues 41–56 of protein G)[55] in a completely unfolded structure obtained from high temperature vacuum simulations was solvated in a 49 Å truncated octahedron box with 2225 water molecules. A 500 ns REMD simulation was conducted utilizing 40 replicas from 278 to 595 K, attempting exchanges every 1 ps.

Simulations in implicit water were performed with the CHARMM36 force field using the CHARMM software package.[68] Langevin dynamics with a friction coefficient of 1 ps^{-1} was used, and all bond lengths were constrained using the SHAKE algorithm. REMD simulations were performed utilizing 16 replicas spanning a temperature range from 285 to 570 K and with exchange attempts every 0.4 ps. Simulations were analyzed using the software package wordom.[69] ^3J backbone scalar couplings and backbone order parameters were calculated as described in previous studies.[70,71]

■ RESULTS

Relative Entropy Parametrization of EEF1-SB. In order to obtain the optimized parameters in EEF1-SB, we use the structural ensemble obtained from explicit solvent molecular dynamics simulations of the α-helical (AAQAA)$_3$ peptide and of the GB1 hairpin. (AAQAA)$_3$ is a weakly structured peptide that significantly populates helical states under physiological conditions,[54] while the GB1 hairpin is known to populate 40%–60% conformations similar to the β-hairpin of the full structure.[55,56] The characteristic secondary structure propensity of the two systems, together with their small size, has made them systems of choice for force field optimization,[72,73] validation and comparison.[74−76] Furthermore, it was recently shown that force fields that were corrected to improve the balance between helical and coil structures in $(AAQAA)_3$ were able to fold both α-helical and β-sheet proteins,[71,77] suggesting this peptide to be useful for force field parametrization.

Since the CHARMM36 force field reproduces the behavior of $(AAQAA)_3$ and GB1 hairpin in solution as inferred from chemical shift data, in the present work we employ the equilibrium ensemble obtained from explicit water simulations as target distributions in the relative entropy minimization.

There are a large number of parameters which can be optimized in the force field, even if attention is restricted only to the aspects related to the implicit solvent model. However, considering too many parameters can lead to numerical instability when handling noisy data. We therefore considered a number of small sets of parameters with the potential to improve the agreement between the implicit and explicit solvent simulations. For example, during the developmental stage, the following parameters were included in the minimization procedure: (i) the solvation free energies ΔG^{free}, (ii) the thickness of the hydration shell λ in eq 3, and (iii) the 1−4 scaling of electrostatic interactions. However, their values were subject to small variations during optimization - and were therefore kept fixed.

However, we found the linear factor k_ε of the distance-dependent dielectric constant ($\varepsilon = k_\varepsilon r$) to be critically suboptimal when combining the original EEF1-C19 parameters with the CHARMM36 force field. Furthermore, we adjusted the force constant k_ψ in a dihedral correction term of the form

$$E_\psi = k_\psi(1 + \cos(\psi - \delta)) \tag{9}$$

The phase shift $\delta = 227°$ was chosen to allow us to tune the balance between extended and helical structures. Previous works suggest this correction to be of particular importance, as it is crucial to reproduce the correct balance between secondary structures in proteins.[71−73,75] We note that this torsional term would have been zero in the "bare" CHARMM36 force field.

The relative entropy-based optimization procedure can be summarized as follows:

1. Generation of Target Trajectories. With the purpose of generating an equilibrated ensemble, we used a previously conducted 150 ns REMD simulation (32 replicas spanning a temperature range from 278 to 416 K) on $(AAQAA)_3$ with the CHARMM36 force field in combination with the TIP3P water model, while for the GB1 hairpin we used a 500 ns REMD simulation spanning a temperature range 278−595 K and utilizing 40 replicas. The last 50 ns (for $(AAQAA)_3$) and 250 ns (for GB1 hairpin) of the trajectories at 300 K were used as reference simulations in the relative entropy procedure.

2. Generation of Model Trajectories. A 100 ns REMD simulation in implicit solvent on the two systems is performed, starting from an extended conformation and with initial parameters $k_{\varepsilon,0} = 1.0$, $k_{\psi,0} = 0.0$.

3. Parameter Update. The parameters, k_ε and k_ψ, are updated using the Newton−Raphson rule (eq 7). The gradient and Hessian are calculated after discarding the first 50 ns. The numerical stability of the minimization is further improved by dynamically adjusting the step size in parameter space (i.e., the value of γ in eq 7). More precisely, the stepsize is reduced until two heuristic criteria are met: i) the absolute change in the parameter at each iteration step is smaller than 50% of its initial value and ii) the change $\Delta S_{\text{rel}} = S_{\text{rel}}(k_{j+1}) - S_{\text{rel}}(k_j)$ is smaller than the 20% of $S_{\text{rel}}(k_j)$, where the variation ΔS_{rel} is extimated via Zwanzig perturbation as

$$\Delta S_{\text{rel}} = \log(\langle \exp(-\beta \Delta U)\rangle) + \beta\langle \Delta U\rangle \tag{10}$$

Here, $\Delta U = U(k_{j+1}) - U(k_j)$, and the averages are computed over the k_j ensemble.

4. Reweighting. Instead of performing a new simulation every iteration step, we used an efficient scheme,[48,63,78] that makes it possible to compute the parameter update at step $j + l$ by reweighting the simulation performed at iteration j. The reweighting strategy is accurate as long as a substantial overlap between the simulations j and $j + l$ exists. To avoid poor overlap, we reweighted the trajectory j until the effective fraction of frames contributing to reweighting, f_{RW}, dropped below 0.6. Here, $f_{RW} = \exp(\sum p(x_i) \log p(x_i))/n$, n is the total number of frames and $p(x_i) = w_i/\sum_q w_q$, with $w_i = \exp(\beta(U(k_j, x_i) - U(k_{j+l}, x_i)))$. Note that in principle, several rounds of sampling can be included in the reweighting via a multiple histogram procedure.[79]

5. Convergence. The resampling-reweighting procedure described above is iterated until convergence of the parameters as well as of the relative entropy gradient.

Before optimizing EEF1-SB, we tested the validity of the this relative entropy approach and the correct implementation of the minimization procedure using synthetic data (Figure S1) generated using known parameters.

Finally, the above procedure was applied to a global optimization against the real explicit solvent training data. The optimization converged after 7 resampling and a total of 493 reweighting iterations, producing the optimal values $k_\varepsilon = 1.5$ and $k_\psi = 0.5$ kcal/mol (Figure 1).

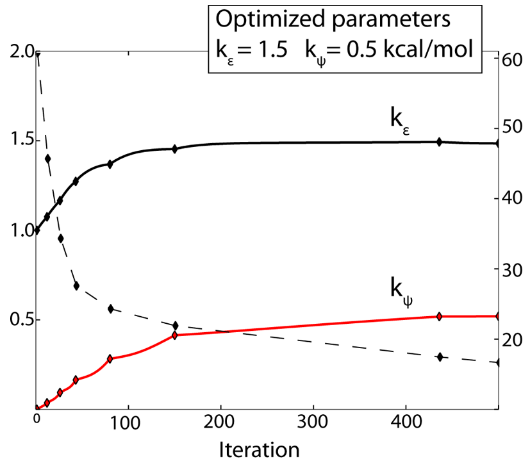


Figure 1. Relative entropy minimization. The relative entropy minimization converged after 500 iterations (7 resampling steps, marked by points), yelding the optimized parameters $k_\varepsilon = 1.5$ and $k_\psi = 0.5$ kcal/mol. The convergence of the procedure is ensured by monitoring the behavior of the gradient (dashed line).

We assess the effect of the relative entropy minimization by comparing the structural ensembles obtained using both the optimized and initial parameter set with the target, explicit solvent simulations. As shown in Figure 2 the changes in the parameters − although relatively small in magnitude − greatly affect the folding free energy surface of $(AAQAA)_3$ at $T = 298$ K. The unoptimized model favors compact, hairpin-like

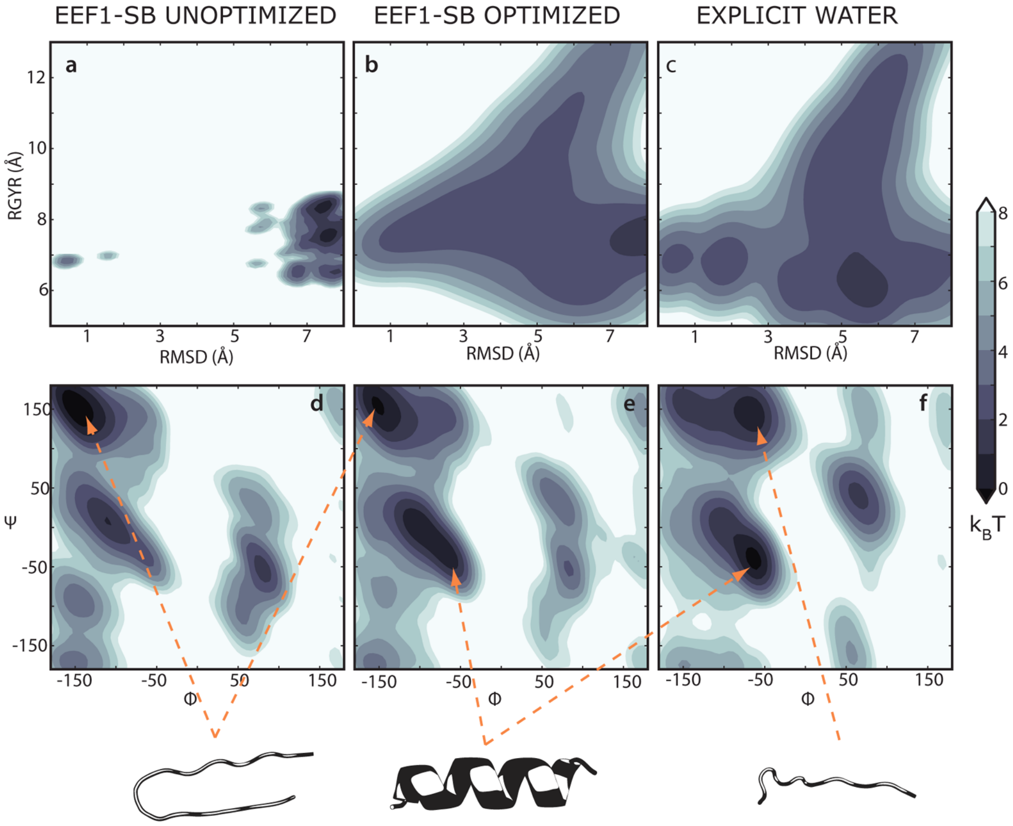


Figure 2. Relative entropy minimization on $(AAQAA)_3$. (a-c) Folding free energy surfaces of $(AAQAA)_3$ projected onto the radius of gyration and RMSD from α-helical state for the unoptimized/optimized EEF1-SB implicit solvent and for the target explicit water simulation at 298 K. (d-f) Ramachandran map for the 9 central residues and representative structures associated with β-sheet, helix, and polyproline (PPII) regions.

structures, while the broad distribution observed in the explicit water simulation is reproduced by the optimized implicit solvent model (Figure 2a-2c). The structural diversity of the equilibrium ensemble observed in explicit water simulations, in accordance with experimental data,[54] is composed of a mixture of elongated polyproline II (PPII) and α-helical configurations. A significant helical content is observed for the optimized implicit solvent, while helix is completely absent in the unoptimized model (Figure 2d). When helices are not formed in the optimized implicit solvent, we instead observe turn or hairpin structures, suggesting that the implicit solvent model does not stabilize sufficiently the unstructured, solvent-exposed states associated with the PPII region of the Ramachandran map (Figure 2e-2f). As described in a number of experimental studies,[80,81] these PPII conformations are stabilized by direct interactions between the main chain and water molecules, that may be difficult to capture with a simple solvent-exclusion model. It is also important to note that only helical or very compact conformations are observed in vacuum (i.e., using the "bare" CHARMM36 force field), thus proving the solvent model to play a major role in determining the behavior of the $(AAQAA)_3$ peptide (Figure 3a). Finally, we highlight that the helix−coil balance of $(AAQAA)_3$ is modeled poorly by many force fields in explicit solvent,[71] unless specific corrections are introduced. We show in Figure 3b−d that other implicit solvent models suffer from similar weaknesses, as they favor the formation of beta strands (as in the original EEF1-C19 model) or overstabilize helical, compact structures (as in the case of FACTS[40] and SCPISM[42]). Conversely, EEF1-SB is able to capture the correct fraction helix near ∼300 K (Figure 4), where it was parametrized; however, the temperature-dependence of helix formation is too weak, given that the explicit solvent model does capture the temperature-dependence of helix formation.[82] This may be because the reference free

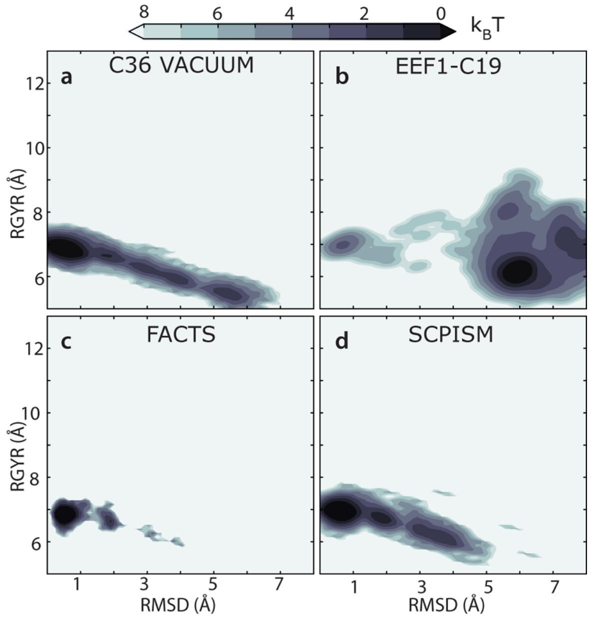


Figure 3. Folding free energy surfaces of $(AAQAA)_3$ at 298 K projected onto the radius of gyration and RMSD from α-helical state for vacuum simulations with (a) CHARMM 36 and for different implicit solvent models: (b) EEF1-C19, (c) FACTS, and (d) SCPISM.

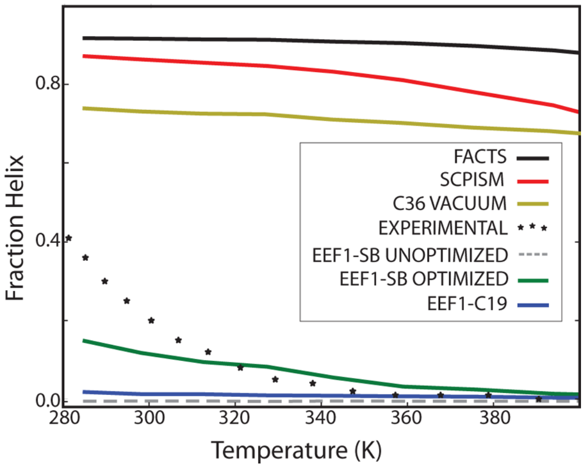


Figure 4. Comparison between calculated and experimental[54] helical fraction of the $(AAQAA)_3$ peptide in simulations and experiment as a function of temperature. The optimized and unoptimized EEF1-SB models are compared with FACTS,[40] SCPISM[42] with the EEF1-C19[18]force field and with CHARMM36 in vacuum.

energies used in all replicas were those at 300 K (although these may in principle be varied), or alternatively because many-body effects arising from the elimination of solvent degrees of freedom are not well described by the model.

The free energy surface of GB1 hairpin observed in explicit water simulations reveals a broad distribution, with a clear "folded' basin (Figure 5c).

Although the relative entropy procedure greatly improves the agreement with the explicit solvent result, the correctly folded β-hairpin is only partially stabilized in the optimized EEF1-SB model (Figure 5a-5b). It is likely that the neutralization of charged groups in EEF1-SB diminishes the formation of the salt bridge between the charged termini, that is known to play a role in the stabilization of hairpin structures.[75] By contrast, the implicit solvents EEF1-C19, SCPISM, and FACTS are able to describe well the dominant folded state (Figure 5e-5g). Nevertheless, the extended unfolded configurations observed in EEF1-SB as well as in explicit water simulations are completely absent in EEF1-C19, SCPISM, and FACTS, suggesting that such models overstabilize compact conformations with well-defined secondary structure relative to more disordered states.

Model Validation. *Simulations of Native Proteins.* We first show EEF1-SB to yield stable trajectories by performing molecular dynamics simulations at room temperature for two folded proteins, ubiquitin (76 residues, pdb code 1UBQ[83]) and GB3 (56 residues, pdb code 1P7E[84]). As a number of experimental and computational studies suggest, both systems are very stable but at the same time undergo small conformational changes occurring on the microsecond timescale.[85,86] In the present test, we assess the ability of the solvent model to maintain a native-like structure during the simulation. Starting from the experimentally solved structures, we perform 100 ns molecular dynamics simulations at a temperature of 300 K using the EEF1-SB model. Both systems remain in the vicinity of the native structure throughout the simulations, with an average C_α root mean squared deviation (RMSD) of 2.7 Å (residues 1−71) and 4.0 Å for ubiquitin and GB3, respectively. We, however, observe a partial loss of the native secondary structure content, as helices are not stabilized to a sufficient degree (Figure S2). Since we have established that EEF1-SB provides a good balance between helix and coil for $(AAQAA)_3$,

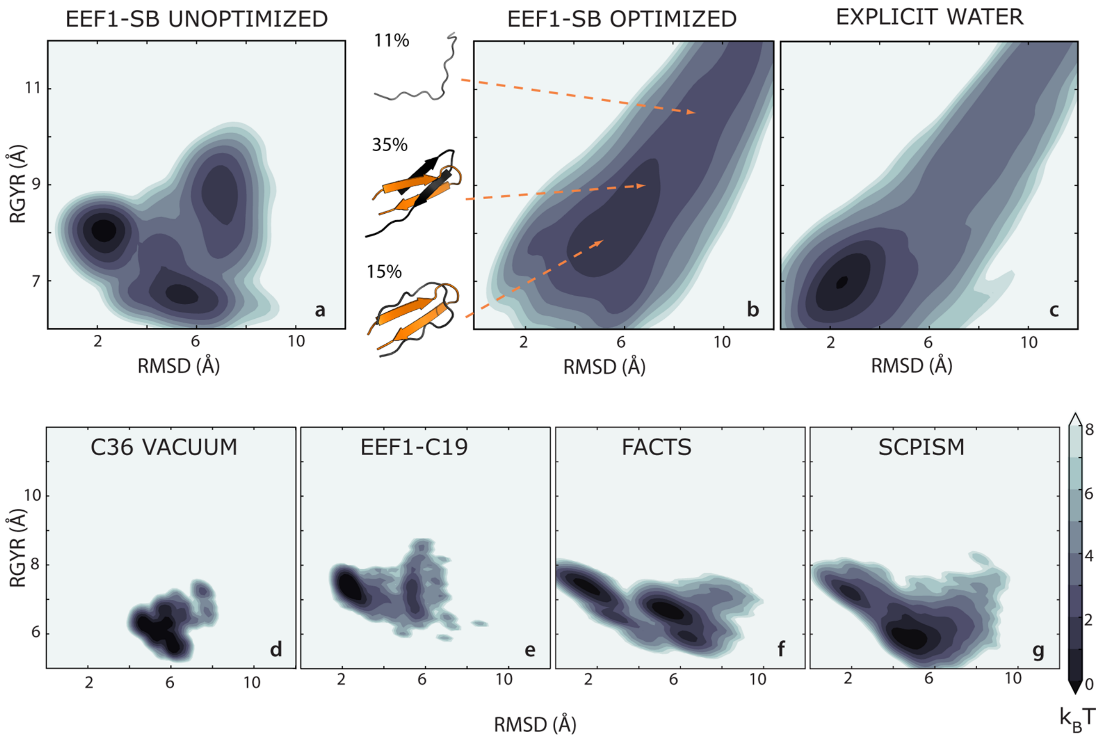


Figure 5. (a-c) Folding free energy surfaces of GB1 projected onto the radius of gyration and RMSD from the folded state for the unoptimized/optimized EEF1-SB implicit solvent and for the target explicit water simulation at 298 K. For the optimized EEF1-SB model, the centers of the first three clusters (in gray) are superimposed on the native structure (orange). Free energy surfaces obtained in vacuum simulations with (d) CHARMM 36 and using the implicit solvent models (e) EEF1-C19, (f) FACTS, and (g) SCPISM.

this most likely arises because the helix is insufficiently stabilized by tertiary packing in our model.

To compare EEF1-SB with existing models, we performed molecular dynamics simulations on the two folded proteins using a number of implicit solvents currently implemented within the CHARMM software package: the EEF1-C19 effective energy function[18] with CHARMM19, the analytic continuum electrostatics (ACE) model,[39] the fast analytical continuum treatment of solvation (FACTS),[40] the Generalized-Born model GBSW,[41] and the screened Coulomb potential SCPISM,[42] as well as simple simulations in vacuum. All simulations were conducted using the simulation protocol described in the original papers.

The results, summarized in Table 1, show the RMSD for EEF1-SB to be comparable to that observed for the CHARMM19 EEF1-C19 model and with ACE. Other implicit solvents such as FACTS, SCPISM, and GBSW appear to perform significantly better, partly because they were parametrized with an emphasis on folded proteins and more importantly because such models treat electrostatic interactions more rigorously.

Table 1. Simulations on Native Proteins

force-field[a]	solvent model	Rgyr[b] (Å)	RMSD[b] (Å)	ΔRMSD[c] (Å)	speed[d] (ns/day)
		ubiquitin			
C36	EEF1-SB	11.4	2.7	2.1	12
C36	VACUUM	10.7	2.2	0.2	10
C22	FACTS	10.7	0.7	0.03	3.4
C22	EEF1[e]	29.2	25.8	22.3	13
C22	SCPISM	10.8	1.0	0.1	6.5
C22	GBSW	10.8	1.0	0.06	1.3
C19	EEF1	11.2	3.8	2.0	23
C19	ACE	10.5	2.2	1.6	5.5
		GB3			
force-field	solvent model	Rgyr (Å)	RMSD (Å)	ΔRMSD (Å)	speed (ns/day)
C36	EEF1-SB	11.1	4.0	1.6	20
C36	VACUUM	10.5	3.8	1.0	17
C22	FACTS	10.3	1.6	0.2	5.6
C22	EEF1[e]	33.4	30.7	6.4	20
C22	SCPISM	10.3	1.5	0.03	11
C22	GBSW	10.3	1.6	0.2	2.1
C19	EEF1	10.9	4.0	1.5	35
C19	ACE	10.1	3.7	0.5	9.2

[a]The implicit solvent models were used in combination with the associated force field described in the original paper: the united-atom CHARMM19 force field was used for ACE and EEF1 and the all-atom force field CHARMM22 for FACTS, SCPISM, and GBSW. [b]Average radius of gyration and C_α RMSD from native. For ubiquitin, residues 71−76 are not included in the RMSD and radius of gyration calculations. [c]RMSD difference between the average calculated on the last and the first 10 ns of the 100 ns simulations. [d]Approximate computational time on a single 2.83 GHz Intel Xeon processor expressed in ns/day. [e]C22 EEF1 is a preliminary version of EEF1 in combination with the CHARMM22 force field, provided with the CHARMM simulation package.[68]

The ultimate test of accuracy is comparison with primary experimental data. Thus, to further assess the ability of the different force fields to model accurately the folded state, we report the agreement between the simulations and measurements of ^3J backbone scalar couplings and S^2 amide order parameters (Figure 6). J-couplings were calculated using a Karplus relation as described previously.[71,72] The overall agreement between experiment and simulation was assessed by computing the statistic

$$\chi^2 = \frac{1}{N}\sum_i \frac{(J_{\mathrm{obs},i} - J_{\mathrm{calc},i})^2}{\sigma_i^2} \tag{11}$$

where the sum runs over N experimentally measured scalar couplings $J_{\mathrm{obs},i}$ which are compared with the corresponding $J_{\mathrm{calc},i}$ calculated from the simulations. The quantity σ_i is an estimate of the uncertainty in computing the scalar coupling from the simulation data, which is the main source of error. All force-fields except CHARMM22 with EEF1-C19 and the vacuum simulation provide a reasonably accurate description of the native state of ubiquitin and GB3, although EEF1-SB is less accurate compared to GBSW, FACTS, and SCPISM. In particular, large deviations between calculated and experimental values of S^2 are observed in the helical regions of the two proteins.

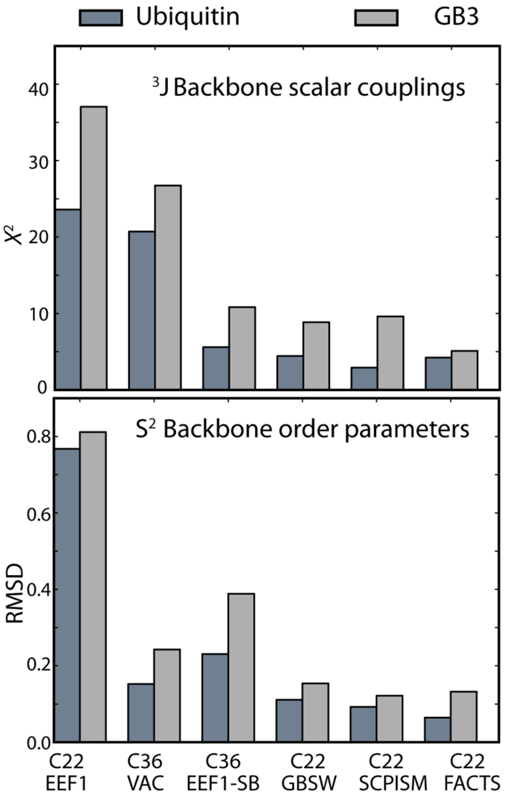


Figure 6. Consistency of folded protein simulations with NMR data. Upper panel: χ^2 between experimental and back-calculated ^3J backbone scalar couplings. Lower panel: root mean squared deviation (RMSD) between experimental and back-calculated S^2 backbone amide order parameters.

Test on a Disordered Peptide. While good performance of the solvent model for folded proteins is important, the strength of EEF1-SB is expected to be for weakly structured (or "intrinsically disordered") peptides. This is because data for such peptides was used in the parametrization; as with any coarse graining, it is inevitable that not all features of the original model can be captured faithfully, but by using data for small peptides to derive the parameters, we expect the properties of similar peptides to be best reproduced.

Table 2. *J*-Couplings for 19-Residue Peptide Derived from Hen Lysozyme (HEWL19)[a]

residue/J	J_{expt}	J (C36 explicit water)	J (EEF1-SB unopt.)	J (EEF1-SB)	J (FACTS)	J (SCPISM)
A10 $^1J_{\text{NC}\alpha}$	10.58	10.35	9.60	10.53	9.52	9.81
A10 $^2J_{\text{NC}\alpha}$	7.24	7.02	6.51	7.25	6.29	6.30
A10 $^3J_{\text{H}\alpha\text{C}}$	1.72	2.10	6.33	2.50	0.53	1.27
A10 $^3J_{\text{HNC}}$	1.33	0.88	1.10	0.88	1.86	0.55
A10 $^3J_{\text{HNC}\beta}$	2.19	2.86	2.09	1.80	3.81	3.69
A10 $^3J_{\text{HNH}\alpha}$	5.10	6.19	6.15	8.13	2.55	5.34
A10 $^3J_{\text{HNH}\alpha}$	0.46	0.37	0.37	0.53	0.10	0.16
$\chi^2(J)$ [Hz]		0.67	12.0	2.4	6.7	4.69
% α_+		57.3	49.4 (2.8)	55.3 (2.0)	100.0 (0.0)	99.2 (0.3)
% β		14.6	17.9 (3.1)	36.3 (1.3)	0.0 (0.0)	0.2 (0.2)
% ppII		18.2	0.7 (0.2)	2.8 (0.4)	0.0 (0.0)	0.0 (0.0)
% helix		39.4	0.0 (0.0)	10.4 (1.0)	100.0 (0.0)	93.1 (0.8)

[a] Top part of the table gives the scalar couplings measured experimentally and calculated from the 298 K replica of REMD simulations using the "DFT2" Karplus relation from ref 72. Results are only shown for the central Ala-10 for brevity; full results are listed in the Supporting Information. Data for C36 with explicit solvent are taken from ref 43. The lower part of the table shows the overall χ^2 between experiment and simulation, the populations of different regions of the Ramachandran map, and the fraction helix averaged over Ala-9, Ala-10, and Ala-11, as defined in ref 72.

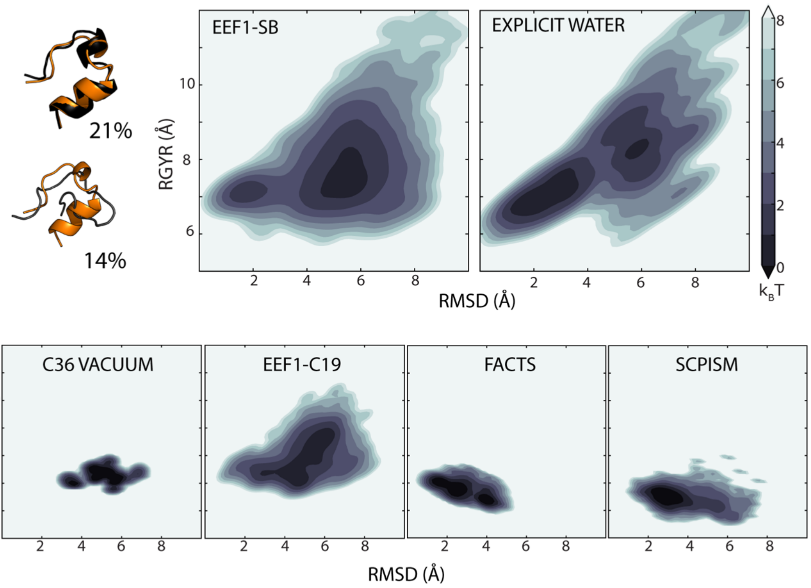


Figure 7. Explicit water and implicit solvent folding free energy surfaces of Trp-cage at T = 298 K. Upper row shows explicit water and EEF1-SB, and the lower row a selection of other implicit solvents, and results for CHARMM36 in vacuum, as labeled. The two-dimensional surface is calculated by projecting onto the radius of gyration and C_α RMSD from native structure. For the EEF1-SB simulation, the first two clusters are shown in gray and superimposed to the native structure (shown in orange).

We have chosen a 19-residue fragment of hen lysozyme (HEWL19), studied by NMR by Schwalbe and co-workers,[87] as a test case for a disordered peptide. This peptide is helical in the full-length lysozyme native structure but only partially helical by itself. NMR scalar couplings have been recorded which reflect the (ϕ, ψ) torsion angles of the central three Ala residues, allowing these to be directly compared with sampling from simulations. The peptide was sampled by running REMD simulations for at least 100 ns for each force field, of which the first 50 ns was discarded as equilibration. The calculated scalar couplings (Table 2) show that the optimization was successful in bringing the sampled distribution of (ϕ, ψ) angles closer to the sampling obtained with CHARMM36, despite this peptide not being used as part of the optimization. The calculated scalar couplings are also much closer to experiment as evident from the reduction of χ^2 from 12.0 in the unoptimized EEF1-SB to 2.4. The sampling of the α-helical region of the Ramachandran map (α_+ in Table 2) is much closer to that in CHARMM36, and the fraction of helix is also much improved (a helical residue is defined as part of sequence of three consecutive residues in the helical region of the Ramachandran map). However, the ppII region is still undersampled, as also evident for the (AAQAA)$_3$ peptide in Figure 2. For reference, we have also run REMD simulations of the HEWL19 peptide with the FACTS and SCPISM models. In both cases, essentially only helix is populated at 298 K (despite the REMD simulations being initiated from disordered structures), and so the deviation from experiment is consequently substantial.

Folding Simulations. As a final test, we assess the ability of EEF1-SB to fold different peptides. The aim of this experiment is not to perform a full thermodynamic characterization of the system but rather to assess the correct qualitative behavior of the force field at room temperature. Here, three fast-folding peptides with different secondary structure propensities are considered: the α-helical mini protein Trp-cage,[88] the 3-β stranded GTT variant of the FiP35 WW domain,[89] and the 35-

residue Nle/Nle double mutant of the villin headpiece,[90] whose native fold consists of 3 helices. For each system, we report the results from 500 ns REMD simulations (16 replicas spanning the temperature range 278−525 K) using EEF1-SB. All simulations were initialized from an extended conformation, and the first 150 ns were discarded in the analysis. Figures 7 and 8 show the folding free energy surfaces of the three proteins. EEF1-SB is able to fold to near-native conformations both Trp-Cage and WW domain but fails in reproducing the folded state of the double mutant of Villin headpiece. Experimental studies on the latter protein suggest the Nle/Nle mutant to be stabilized by salt bridges,[89] that are poorly modeled in EEF1-SB.

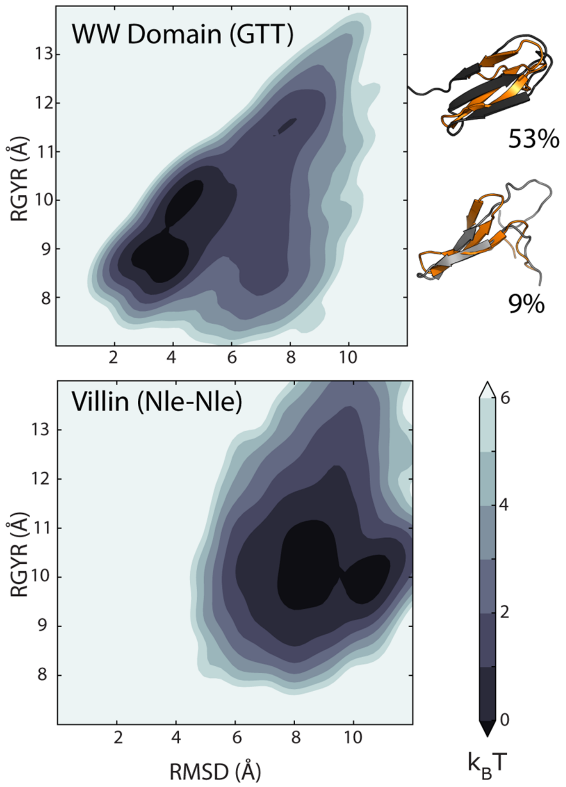


Figure 8. EEF1-SB folding free energy surface of WW domain and Villin at T = 298 K. The two-dimensional surface is calculated by projecting onto the radius of gyration and C_α RMSD from native structure. For WW domain, the first two clusters are shown in gray and superimposed to the native structure (shown in orange).

Because it is small, and readily sampled, we have used Trp Cage to compare EEF1-SB against other implicit solvent models and against simple vacuum simulations in CHARMM36 (Figure 7). These free energy surfaces show that the vacuum simulations will clearly not fold to the native state, while for EEF1-C19, a native-like state is only marginally stable. While FACTS and SCPISM do obtain the correct folded structure, the unfolded structure is much too compact, relative to the explicit solvent simulations (in fact, more compact than the folded state). By contrast, EEF1-SB captures much better the overall dimensions of the unfolded protein seen in explicit solvent simulations, as we have also seen above for the $(AAQAA)_3$ and GB1 peptides.

DISCUSSION

Describing the effect of the aqueous environment is of fundamental importance in molecular simulations of biomolecules. While explicit water simulations provide a high level of detail, implicit solvent models represent a fast and approximate way to describe the behavior of proteins in solution, that can be useful for systems undergoing large conformational transitions, (e.g unfolding experiments, simulations of unstructured peptides) or aggregation studies, that typically require a large simulation box when explicit solvent is included.

In the present work we have employed an approximate but extremely fast solvent-exclusion model and combined it with the all-atom CHARMM36 force field. CHARMM36 was extensively optimized against experimental data, by performing long molecular dynamics simulations in explicit water. In order to retain the accuracy of the original force field, we optimized the parameters in the solvent-exclusion model so as to mimic the behavior of explicit water simulations, using a coarse-graining technique based on the minimization of the relative entropy. In contrast to standard parametrization approaches, this methodology is not aimed at stabilizing only native conformations but at reproducing the full equilibrium distribution as observed in explicit water simulations. This is a desirable target if non-native or disordered states are of interest and also the ideal reference if the implicit solvent simulations are intended for use in conjunction with the explicit solvent in a multiscale approach.

Various tests conducted with the optimized model were shown to produce favorable results. EEF1-SB was shown to result in stable trajectories of native proteins in room temperature molecular dynamics simulations, although for that purpose the alternative FACTS and SCPISM models are more accurate. Where EEF1-SB appears to have a distinct advantage over other implicit models is in its description of the dimensions and secondary structure formation in disordered peptides and non-native states. Finally, EEF1-SB is able to fold Trp-Cage and WW domain to near-native structures but not the Villin headpiece subdomain. When compared to the other implicit solvent models considered here, EEF1-SB is much better able to capture the overall dimensions of the chain, and the features of the folding free energy landscapes, relative to the explicit solvent simulations.

The design of an implicit solvent model often entails three distinct but connected elements. First, a force field describing the solute−solute interactions must be chosen. Second, a specific physical model describing the effects arising from the presence of the solvent must be assumed. Lastly, the parameters of the implicit solvent model and possibly also of the solute−solute model must be optimized according to some procedure. In principle, the solute model should not need to be adjusted when moving to implicit solvent, but in practice (as we have shown here), such adjustments can yield signficant improvements. These adjustments should be thought of as part of the implicit solvent model, as they effectively project some of the solvation effects onto the internal degrees of freedom of the solute.

For the first element, we have chosen one of the latest generations of protein force fields, representing many years of careful parametrization, and for the second element we have chosen a computationally efficient, yet accurate implicit solvent model. For the third element, we have demonstrated that the relative entropy formalism is able to systematically improve the quality of the EEF1-C19 force field, resulting in a final model which reproduces well the free energy landscapes of small peptides. The remaining discrepancies are most likely due to the functional form of the implicit solvent model we have chosen to optimize, particularly the description of electrostatic energy. This would be a likely direction for future improvement of the model.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

Table S1, solvation parameters used in EEF1-SB. **Figure S1**, validation of the relative entropy approach on synthetic data. **Table S2**, partial atomic charges for ionic groups in EEF1-SB. **Figure S2**, RMSD and average secondary structure content of GB3 and ubiquitin using EEF1-SB. **Table S3**, *J*-couplings for 19-residue peptide derived from hen lysozyme (HEWL19). This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: lindorff@bio.ku.dk (K.L.-L.); robertbe@helix.nih.gov (R.B.B.).

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

R.B.B. was supported by a Royal Society University Research Fellowship while at Cambridge and by the Intramural Research Program of the National Institute for Diabetes and Digestive and Kidney Diseases of the National Institutes of Health. S.B. and K.L.-L. were supported by a Hallas-Moller stipend from the Novo-Nordisk Foundation. This study utilized the high-performance computational capabilities of the Biowulf Linux cluster at the National Institutes of Health, Bethesda, MD (http://biowulf.nih.gov).

■ REFERENCES

(1) Dill, K. *Biochemistry* **1990**, *29*, 7133−7155.
(2) Prabhu, N.; Sharp, K. *Chem. Rev.* **2006**, *106*, 1616−1623.
(3) Bryant, R. *Annu. Rev. Biophys. Biomol. Struct.* **1996**, *25*, 29−53.
(4) Tarek, M.; Tobias, D. *Biophys. J.* **2000**, *79*, 3244−3257.
(5) Zhang, L.; Wang, L.; Kao, Y.; Qiu, W.; Yang, Y.; Okobiah, O.; Zhong, D. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 18461−18466.
(6) Eisenberg, D.; McLachlan, A. *Nature* **1986**, *319*, 199−203.
(7) Ooi, T.; Oobatake, M.; Nemethy, G.; Scheraga, H. *Proc. Natl. Acad. Sci. U. S. A.* **1987**, *84*, 3086−3090.
(8) Fraternali, F.; Van Gunsteren, W. *J. Mol. Biol.* **1996**, *256*, 939−948.
(9) Ferrara, P.; Apostolakis, J.; Caflisch, A. *Proteins: Struct., Funct., Bioinf.* **2002**, *46*, 24−33.
(10) Hummer, G. *J. Am. Chem. Soc.* **1999**, *121*, 6299−6305.
(11) Chandler, D. *Nature* **2005**, *437*, 640−647.
(12) Allison, J. R.; Boguslawski, K.; Fraternali, F.; van Gunsteren, W. F. *J. Phys. Chem. B* **2011**, *115*, 4547−4557.
(13) Colonna-Cesari, F.; Sander, C. *Biophys. J.* **1990**, *57*, 1103−1107.
(14) Irbäck, A.; Mohanty, S. *Biophys. J.* **2005**, *88*, 1560−1569.
(15) Stouten, P.; Frömmel, C.; Nakamura, H.; Sander, C. *Mol. Simul.* **1993**, *10*, 97−120.
(16) Lazaridis, T. *J. Phys. Chem. B* **1998**, *102*, 3531−3541.
(17) Lazaridis, T. *J. Phys. Chem. B* **1998**, *102*, 3542−3550.
(18) Lazaridis, T.; Karplus, M. *Proteins: Struct., Funct., Bioinf.* **1999**, *35*, 133−152.
(19) Lazaridis, T.; Karplus, M. *Science* **1997**, *278*, 1928−1931.
(20) Inuzuka, Y.; Lazaridis, T. *Proteins: Struct., Funct., Bioinf.* **2000**, *41*, 21−32.
(21) Hassan, S.; Mehler, E. *Proteins: Struct., Funct., Bioinf.* **2002**, *47*, 45−61.
(22) Steinbach, P. J. *Proteins* **2004**, *57*, 665−677.
(23) Huang, A.; Stultz, C. *Biophys. J.* **2007**, *92*, 34−45.
(24) Cavalli, A.; Salvatella, X.; Dobson, C.; Vendruscolo, M. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 9615−9621.
(25) Kaufmann, K.; Lemmon, G.; DeLuca, S.; Sheehan, J.; Meiler, J. *Biochemistry* **2010**, *49*, 2987−2998.
(26) Ding, F.; Tsao, D.; Nie, H.; Dokholyan, N. *Structure* **2008**, *16*, 1010−1018.
(27) Davis, I. W.; Baker, D. *J. Mol. Biol.* **2009**, *385*, 381−392.
(28) Vitalis, A.; Pappu, R. V. *J. Comput. Chem.* **2008**, *30*, 673−699.
(29) Mackerell, A. D.; Feig, M.; Brooks, C. L. *J. Am. Chem. Soc.* **2004**, *126*, 698−699.
(30) Pu, M.; Garrahan, J.; Hirst, J. *Chem. Phys. Lett.* **2011**, *515*, 283−289.
(31) Juneja, A.; Ito, M.; Nilsson, L. *J. Chem. Theory Comput.* **2012**, *9*, 834−846.
(32) Warwicker, J.; Watson, H. *J. Mol. Biol.* **1982**, *157*, 671−679.
(33) Nicholls, A.; Honig, B. *J. Comput. Chem.* **1991**, *12*, 435−445.
(34) Luo, R.; David, L.; Gilson, M. *J. Comput. Chem.* **2002**, *23*, 1244−1253.
(35) Constanciel, R. *Theor. Chem. Acc.* **1986**, *69*, 505−523.
(36) Still, W.; Tempczyk, A.; Hawley, R.; Hendrickson, T. *J. Am. Chem. Soc.* **1990**, *112*, 6127−6129.
(37) Hummer, G.; Pratt, L. R.; García, A. E. *J. Phys. Chem. A* **1998**, *102*, 7885−7895.
(38) Feig, M.; Reiher, M.; Wolf, A. *Modeling solvent environments*; Wiley Online Library: 2010.
(39) Schaefer, M.; Karplus, M. *J. Phys. Chem.* **1996**, *100*, 1578−1599.
(40) Haberthür, U.; Caflisch, A. *J. Comput. Chem.* **2008**, *29*, 701−715.
(41) Im, W.; Lee, M.; Brooks, C., III *J. Comput. Chem.* **2003**, *24*, 1691−1702.
(42) Hassan, S.; Guarnieri, F.; Mehler, E. *J. Phys. Chem. B* **2000**, *104*, 6478−6489.
(43) Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E.; Mittal, J.; Feig, M.; MacKerell, A. D., Jr. *J. Chem. Theory Comput.* **2012**, *8*, 3257−3273.
(44) Baker, C. M.; Best, R. B. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2013**, DOI: dx.doi.org/10.1002/wcms.1167.
(45) Neria, E.; Fischer, S.; Karplus, M. *J. Chem. Phys.* **1996**, *105*, 1902−1922.
(46) Shell, M. *J. Chem. Phys.* **2008**, *129*, 144108−144115.
(47) Chaimovich, A.; Shell, M. *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.* **2010**, *81*, 060104−060108.
(48) Chaimovich, A.; Shell, M. *J. Chem. Phys.* **2011**, *134*, 094112−0941127.
(49) Carmichael, S.; Shell, M. *J. Phys. Chem. B* **2012**, *116*, 8383−8393.
(50) Baker, C. M.; Best, R. B. *J. Chem. Theory Comput.* **2013**, *9*, 2826−2837.
(51) Izvekov, S.; Parrinello, M.; Burnham, C. J.; Voth, G. A. *J. Chem. Phys.* **2004**, *120*, 10896.
(52) Kleinjung, J.; Scott, W. R.; Allison, J. R.; van Gunsteren, W. F.; Fraternali, F. *J. Chem. Theory Comput.* **2012**, *8*, 2391−2403.
(53) Cao, Z.; Dama, J. F.; Lu, L.; Voth, G. A. *J. Chem. Theory Comput.* **2013**, *9*, 172−178.
(54) Shalongo, W.; Dugad, L.; Stellwagen, E. *J. Am. Chem. Soc.* **1994**, *116*, 8288−8293.
(55) Blanco, F.; Rivas, G.; Serrano, L. *Nat. Struct. Mol. Biol.* **1994**, *1*, 584−590.
(56) Muñoz, V.; Thompson, P. A.; Hofrichter, J.; Eaton, W. A. *Nature* **1997**, *390*, 196−199.
(57) Privalov, P.; Makhatadze, G. *J. Mol. Biol.* **1993**, *232*, 660−679.
(58) Lazaridis, T.; Karplus, M. *J. Chem. Phys.* **1996**, *105*, 4294.
(59) Makhatadze, G.; Privalov, P. *J. Mol. Biol.* **1993**, *232*, 639−659.
(60) Privalov, P.; Makhatadze, G. *J. Mol. Biol.* **1992**, *224*, 715−723.

(61) Di Qiu, M.; Shenkin, P.; Hollinger, F.; Still, W. *J. Phys. Chem. A* **1997**, *101*, 3005−3014.
(62) Jorgensen, W. *J. Am. Chem. Soc.* **1981**, *103*, 335−340.
(63) Zwanzig, R. *J. Chem. Phys.* **1954**, *22*, 1420−1427.
(64) Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **1999**, *314*, 141−151.
(65) Hess, B.; Kutzner, C.; Van Der Spoel, D.; Lindahl, E. *J. Chem. Theory Comput.* **2008**, *4*, 435−447.
(66) York, D.; Darden, T.; Pedersen, L. *J. Chem. Phys.* **1993**, *99*, 8345−8349.
(67) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. *J. Comput. Chem.* **1997**, *18*, 1463−1472.
(68) Brooks, B. R.; et al. *J. Comput. Chem.* **2009**, *30*, 1545−1614.
(69) Seeber, M.; Felline, A.; Raimondi, F.; Muff, S.; Friedman, R.; Rao, F.; Caflisch, A.; Fanelli, F. *J. Comput. Chem.* **2011**, *32*, 1183−1194.
(70) Best, R. B.; Vendruscolo, M. *J. Am. Chem. Soc.* **2004**, *126*, 8090−8091.
(71) Lindorff-Larsen, K.; Maragakis, P.; Piana, S.; Eastwood, M.; Dror, R.; Shaw, D. *PLoS One* **2012**, *7*, e32131.
(72) Best, R.; Hummer, G. *J. Phys. Chem. B* **2009**, *113*, 9004−9015.
(73) Piana, S.; Lindorff-Larsen, K.; Shaw, D. *Biophys. J.* **2011**, *100*, L47.
(74) Irbäck, A.; Mitternacht, S.; Mohanty, S. *BMC Biophys.* **2009**, *2*, 2.
(75) Best, R.; Mittal, J. *J. Phys. Chem. B* **2010**, *114*, 8790−8798.
(76) Best, R.; Mittal, J. *Proteins: Struct., Funct., Bioinf.* **2011**, *79*, 1318−1328.
(77) Mittal, J.; Best, R. B. *Biophys. J.* **2010**, *99*, L26−L28.
(78) Norgaard, A.; Ferkinghoff-Borg, J.; Lindorff-Larsen, K. *Biophys. J.* **2008**, *94*, 182−192.
(79) Wang, L.-P.; Chen, J.; van Voorhuis, T. *J. Chem. Theory Comput.* **2013**, *9*, 452−460.
(80) Rucker, A.; Creamer, T. *Protein Sci.* **2002**, *11*, 980−985.
(81) Liu, Z.; Chen, K.; Ng, A.; Shi, Z.; Robert, W.; Kallenbach, N. *J. Am. Chem. Soc.* **2004**, *126*, 15141−15150.
(82) Best, R. B.; Mittal, J.; Feig, M.; Mackerell, A. D. *Biophys. J.* **2012**, *103*, 1045−1051.
(83) Cornilescu, G.; Marquardt, J. L.; Ottiger, M.; Bax, A. *J. Am. Chem. Soc.* **1998**, *120*, 6836−6837.
(84) Ulmer, T. S.; Ramirez, B. E.; Delaglio, F.; Bax, A. *J. Am. Chem. Soc.* **2003**, *125*, 9179−9191.
(85) Markwick, P.; Bouvignies, G.; Blackledge, M. *J. Am. Chem. Soc.* **2007**, *129*, 4724−4730.
(86) Lange, O.; Lakomek, N.; Farès, C.; Schröder, G.; Walter, K.; Becker, S.; Meiler, J.; Grubmüller, H.; Griesinger, C.; de Groot, B. *Science* **2008**, *320*, 1471−1475.
(87) Graf, J.; Nguyen, P. H.; Stock, G.; Schwalbe, H. *J. Am. Chem. Soc.* **2007**, *129*, 1179−1189.
(88) Qiu, L.; Pabit, S.; Roitberg, A.; Hagen, S. *J. Am. Chem. Soc.* **2002**, *124*, 12952−12953.
(89) Kubelka, J.; Chiu, T. K.; Davies, D. R.; Eaton, W. A.; Hofrichter, J. *J. Mol. Biol.* **2006**, *359*, 546−553.
(90) Piana, S.; Sarkar, K.; Lindorff-Larsen, K.; Guo, M.; Gruebele, M.; Shaw, D. E. *J. Mol. Biol.* **2011**, *405*, 43−48.