Published in final edited form as:

J Phys Chem B. 2018 December 13; 122(49): 11355–11362. doi:10.1021/acs.jpcb.8b07264.

Peptide and Protein Structure Prediction with a Simplified Continuum Solvent Model

Peter J. Steinbach

Center for Molecular Modeling, Center for Information Technology, National Institutes of Health, Bethesda, MD 20892, USA

Abstract

A continuum solvent model based on screened Coulomb potentials has been simplified and parameterized to sample native-like structures in replica-exchange simulations of each of six different peptides and mini-proteins. Low-energy, native and non-native structures were used to iteratively refine eleven parameter values. The centroid of the largest cluster of structures sampled in simulations initiated from an extended conformation represents the predicted structure. The main-chain rms deviation of this prediction from the experimental structure was 0.47 Å for the 12-residue Trp-*zip2*, 0.86 Å for the 14-residue MBH12, 2.53 Å for the 17-residue U(1–17)T9D, 2.03 Å for the 20-residue BS1, 1.08 Å for the 20-residue Trp-*cage*, and 3.64 Å for the 35-residue villin headpiece subdomain HP35. The centroid of the sixth largest cluster sampled for HP35 deviated by 0.91 Å. The CHARMM22/CMAP force field was used, with an additional ψ torsion term for residues other than glycine and proline. Six parameters govern the dielectric response of the continuum solvent, and four values of surface tension approximate nonpolar effects. An atom's self-energy and interaction energies are screened independently, each depending on whether the atom is part of a charged group, a neutral hydrogen-bonding main-chain group, or any other neutral group. The parameters inferred result in strong main-chain hydrogen bonds, consistent with the view that protein folding is dominated by the formation of these bonds^{1–2}. Conformations of MBH12 and BS1 were excluded from the energy-function refinement, suggesting the parameters, referred to as SCP18, are transferable. An efficient estimate of solvent-accessible surface area is also described.

Graphical Abstract

Telephone: (301) 496-1100, steinbac@mail.nih.gov.

Supporting Information. Representative interactions assessed during parameter optimization. Plots of rms deviation as a function of simulation step. Plots of clusters of sampled configurations and their deviations from structure determined experimentally.

Introduction

Significant progress has been made in recent years in the use of Newtonian simulations to predict biomolecular structure and dynamics. The native conformations of small proteins have been repeatedly sampled in folding simulations using both explicit³ and implicit⁴ solvent. The goal of this study is to assess how simple a continuum solvent model can be made that results in the sampling of native conformations in folding simulations of a diverse set of peptides and mini-proteins. Several solvent models have been developed to avoid the computational expense incurred when calculating water-water and water-solute forces explicitly or when applying Poisson-Boltzmann theory^{5–6}. These solvent descriptions include those based on solvent accessible surface area (SASA)^{7–8}, SASA and buried solute volume⁹, solvent exclusion^{10–11}, generalized Born approaches^{12–14}, and screened Coulomb potentials^{15–17}.

The implicit solvent model based on screened Coulomb potentials (SCPISM) is the focus of this study because of its physically reasonable foundation¹⁶ and its computational efficiency¹⁸. It is reparametrized here with considerably fewer parameters by screening atomic interactions and self-energies based on the net charge of atomic groups. This simplified description was used to repeatedly sample native-like conformations for each of six peptides and mini-proteins in simulations initiated from an extended conformation. For each of the six systems, the most-sampled cluster is native-like.

First, the estimation of SASA was refined to improve estimates of effective radii and the nonpolar free energy of solvation. An iterative procedure was then used in which dynamics was simulated for peptides of known structure, low-energy conformations collected, and parameter values biased to promote sampling of native-like conformations for multiple peptides. Hundreds of simulations were performed, initiated from both extended and folded conformations. Ultimately, a simple description was obtained that utilizes only six electrostatic screening parameters and four surface tensions, plus an extra main-chain torsion term added for use with the CHARMM22 parameter set¹⁹ with the CMAP adjustment²⁰. The CHARMM22/CMAP force field has been used successfully to predict the structure of the trp-cage mini-protein in implicit solvents^{17, 21} and in explicit-water folding simulations of the villin headpiece²².

The six peptides and mini-proteins studied were chosen because they adopt a variety of stable structures in solution, some involving beta strands (two or three), others with helical segments. In order of increasing sequence length, they are trp-zip²³: SWTWENGKWTWK, MBH12²⁴: RGKWTYNGITYEGR, U(1–17)T9D²⁵: MQIFVKTLDGKTITLEV, BS1 (also referred to as beta3s)^{26–27}: TWIQNGSTKQYQNGSTKIYT, trp-cage²⁸: NLYIQWLKDGGPSSGRPPPS, and HP35²⁹: LSDEDFKAVFGMTRSAFANLPLWKQQLKKEKGLF. Computational techniques have been used to study trp-zip^{30–31}, MBH12^{32–33}, U(1–17)T9D^{34–35}, BS1^{36,37}, the trp-cage^{17, 21, 31, 38–44}, and variants of the villin headpiece^{22, 45–50}.

Ranging from 12 to 35 amino acids in length, these model systems provide a stringent test of both the calculated energy landscapes and the conformational explorations thereof. At the

protein level, the villin headpiece subdomain is an ultrafast-folding protein that has emerged as a model system for gaining insights into the folding of larger, more complex proteins. This protein is large enough to have a hydrophobic core that is significantly buried by the association of its three helices, requiring a balance between modeled hydrophobic and polar effects. Its prominence in dynamics simulations and theoretical folding studies prompted Davies and coworkers to revisit its structure, solving it at high resolution²⁹. Explicit-solvent simulations have indicated that this crystal structure of the 35-residue subdomain, HP35, provides a better reference structure for the subdomain than a structure determined previously by NMR⁴⁸. In addition to this high-resolution structure, experimental and theoretical studies by Eaton and coworkers have contributed significantly to our understanding of villin folding^{51–52}, as well as the folding of other peptides⁵³ and proteins^{54–56}. Considerably larger than the other systems simulated here, convergence of HP35 to its native structure in a simulation initiated from an extended conformation poses the greatest challenge.

The continuum solvent model is summarized next, followed by a description of the revised approximation of atomic surface areas. Results obtained with the simplified parameterization of dielectric effects introduced here are then presented, and possibilities for future improvements are discussed.

The Screened Coulomb Potential Implicit Solvent Model

The SCPISM is a continuum model based on the Lorentz-Debye-Sack-Onsager theory of polar liquids⁵⁷. It is formulated in terms of distance-dependent screening functions that approximate the dielectric response to an arbitrary charge distribution immersed in the continuum solvent. Whereas the dielectric permittivity ϵ accounts for the damping of the vacuum electric field E , the screening function D attenuates the electrostatic potential ϕ . For an isolated point charge q in a dielectric of permittivity ϵ , $q/(\epsilon r^2) = |E| = -d\phi/dr = -d/dr[q/(D r)]$, which gives the relation between $D(r)$ and $\epsilon(r)$. This screening is a sigmoidal function of the distance from the point source, approaching the bulk-solvent value at large distances⁵⁷.

The electrostatic energy of a system of N atoms immersed in a continuum dielectric is given in the SCPISM¹⁵ by:

$$E_e = \frac{1}{2} \sum_{i \neq j}^N \frac{q_i q_j}{D_{ij}(r_{ij}; \alpha_{ij}) r_{ij}} + \frac{1}{2} \sum_{i=1}^N \frac{q_i^2}{\mathcal{R}_i} \left\{ \frac{1}{D_i(\mathcal{R}_i; \alpha_i)} - 1 \right\} \quad (1)$$

where the first term is the sum of pair-wise electrostatic interactions, the second is the sum of atomic self-energies, r_{ij} is the distance between atoms i and j , and q_i and $\mathcal{R}_i$ are the charge and effective radius of atom i , respectively. A sigmoidal form is used for D that requires one screening parameter, α , to control the rate at which D increases from unity at zero separation to the bulk dielectric constant, ϵ_0 , at large distances:

$$D(r; \alpha) = \frac{1 + \epsilon_0}{1 + ke^{-\alpha r}} - 1 \quad (2)$$

where $k \equiv (\epsilon_0 - 1)/2$. Screening increases with the value of α ; the sigmoidal D approaches bulk behavior at shorter distances as α increases.

The screening parameters that govern interactions are taken to be $\alpha_{ij} = (\alpha_i \alpha_j)^{1/2}$. In principle, the interaction between two atoms in a biomolecule is screened depending on the extent to which solvent is excluded by other solute atoms. For simplicity, this conformational dependence of the α parameters arising from crowding and solvent exclusion¹⁷ is ignored here.

In contrast to typical Generalized Born models of solvation^{13–14, 58}, the effective radius of an atom governs only its self-energy in the SCPISM and does not affect its interactions with other atoms. For an atom in a solvated protein, this radius is approximated as a linear combination of two “bulk” radii¹⁶, one characterizing the size of the atom when completely immersed in aqueous solution, R_w and another associated with the atom immersed in a homogeneous, “bulk” protein interior, R_p :

$$R_i = \xi_i R_{wi} + (1 - \xi_i) R_{pi}, \quad (3)$$

where ξ_i is the fractional SASA of the atom: $\xi_i = \text{SASA}_i / (4\pi R_i^2)$, $R_i = R_{vdW,i} + R_{probe}$, $R_{vdW,i}$ is the van der Waals radius of atom i , and R_{probe} was taken to be 1.4 Å.

The nonpolar effects of solvation to be captured by an implicit solvent model are the work done to create a solute-shaped cavity in the solvent and the van der Waal’s interactions between the solute and the solvent molecules that are not explicitly represented. For solutes of modest size, these effects are approximately proportional to the total SASA of the solute⁵⁹. If atom-dependent surface tensions σ are considered, this observation is expressed as:

$$G_{np} = a + \sum_i \sigma_i \text{SASA}_i \quad (4)$$

Thus, the SCPISM requires an estimate of atomic surface areas to compute the effective radii that govern self-energies and to compute the nonpolar free energy of solvation, G_{np} .

Methods

Estimating Solvent Accessible Surface Area

The surface area of a spherical atom i buried by a spherical atom j is given by^{60–61}:

$$a_{ij} = 2\pi R_i [R_i - r_{ij}/2 - (R_i^2 - R_j^2)/(2r_{ij})] \quad r_{ij} < R_i + R_j, \\ = 0 \quad r_{ij} \geq R_i + R_j \quad (5)$$

Note that the derivative of a_{ij} is discontinuous at $r_{ij} = R_i + R_j$. To ensure that all surface areas and their derivatives are continuous everywhere, a smooth switching function S is introduced here to approximate a_{ij} as

$$A_{ij}(r_{ij}) = a_{ij}(r_{on}) S(r_{ij}); \quad (6)$$

$$\begin{aligned} S(r_{ij}) &= 1 & r_{ij} &\leq r_{on}, \\ S(r_{ij}) &= (r_{off} - r_{ij})^2 (r_{off} + 2r_{ij} - 3r_{on}) / (r_{off} - r_{on})^3 & r_{on} < r_{ij} < r_{off}, \\ S(r_{ij}) &= 0 & r_{ij} &\geq r_{off}. \end{aligned} \quad (7)$$

In the switching region from r_{on} to r_{off} , S is the cubic function of r_{ij} that satisfies $S(r_{on}) = 1$, $S(r_{off}) = 0$, $dS/dr(r_{on}) = 0$, and $dS/dr(r_{off}) = 0$. To calculate atomic SASA, S is switched to zero at the separation distance

$$r_{off} \equiv R_i + R_j = R_{vdW,i} + R_{vdW,j} + 2R_{probe},$$

and the separation at which S is turned on was chosen to be

$$r_{on} = 0.85 (R_{vdW,i} + R_{vdW,j}).$$

Thus, r_{on} and r_{off} depend on the (i,j) atom pair being considered.

The SASA of atom i is then approximated as:

$$SASA_i = SASA0_i - \sum_{k=1,2,3} c_{k,i} \sum_j A_{ij}(r_{ij}), \quad (8)$$

where $SASA0_i = 4\pi R_i^2 c_{0,i}$ represents the surface area of atom i that is not buried by atoms separated from atom i by 1 or 2 bonds (“1–2” and “1–3” interactions). This term is approximated as being independent of conformation; the small, high-frequency changes in the burial of atoms due only to local bond stretching and angle bending are neglected. The k^{th} sum is over all atoms j that are separated from atom i by at least 3 bonds and are bonded to k heavy (non-hydrogen) atoms.

Because the portions of the surface of atom i buried by different neighboring atoms may not be distinct, three different sums are used, each with a different scale factor. Any given atom j contributes only to the sum associated with the number of heavy atoms to which atom j is bonded. The constants $c_{1,i}$, $c_{2,i}$, and $c_{3,i}$ are introduced to reduce errors associated with the double-counting of surface areas of atom i that are buried by more than one atom. The three sums involve only pairs of atoms that are considered in typical evaluations of nonbonded interactions, providing an efficient SASA estimate for use in molecular dynamics simulations.

This revised model of atomic SASA requires four parameters (c_k , for $k = 0,1,2,3$) for every atom of a given van der Waals type bonded to a given number of heavy atoms. It depends linearly on these parameters, which were determined by fitting the exact SASA calculated numerically⁶² for 341 high-resolution x-ray structures. Single-chain, ligand-free protein structures determined at a resolution of 1.8 Å or better, sharing sequence identity of 50 % or less, and found to have reasonable bond strain according to the CHARMM22 force field, were used in the parameterization. No energy minimization of the structures was performed as a diversity of geometries, with different levels of local strain, was desired for parameter determination.

Parameter refinement and replica-exchange simulations

Different forms of the energy function were considered, and hundreds of parameter optimizations were performed. In the end, eleven parameters were estimated, including six screening parameters, four atomic surface tensions, and the amplitude of an additional ψ torsion with a minimum at 150° for residues other than glycine and proline. The surface tensions were assigned by atom type as a straightforward extension to a surface-area model of hydrophobicity using a single surface tension. Aliphatic carbon and hydrogen atoms, excluding α -carbons, plus proline carbons were assigned a surface tension σ_1 . Aromatic carbon atoms and their bound hydrogens in phenylalanine residues were assigned σ_F ; other aromatic carbon and hydrogen atoms and carbon atoms in tryptophan side chains were assigned σ_W . All other atoms were assigned a surface tension σ_0 .

Parameters in the energy function were optimized using the subroutine NEWUOA⁶³ to simultaneously bias the energy gap between native and non-native conformations toward a user specified value for three peptides simultaneously: trp-zip2, U(1–17)T9D, and HP35. In the later iterations, the α and σ values were varied while fixing the torsional amplitude. This biasing of the three energy landscapes was performed subject to restraints that five characteristic hydrogen bonding interactions (Figure S1) have minimum energies in the range –4.0 to –0.1 kcal/mol and while limiting the range of each parameter. Energy gaps were estimated as the difference in the averages of the lowest 1000 energies for previously sampled native and non-native structures. Native structures were designated as those within 1.25 Å rmsd of the experimental structure for trp-zip2 and those within 3.0 Å for U(1–17)T9D and for HP35. Structures with greater rmsd values were considered non-native.

Simulations were initiated from the experimentally determined structures of three model systems and from extended conformations. Low-energy conformations sampled in the simulations were repeatedly added to the accumulated ensemble of decoy structures that

were used in subsequent parameter estimations. It was observed that trp-zip2 and U(1–17)T9D folded without the estimated energy gaps actually favoring native conformations, indicating entropic stabilization of these two-stranded beta sheets. Therefore, the optimization ultimately imposed a native-conformation bias for HP35 while limiting the destabilization of the other two systems.

Replica-exchange⁶⁴ Langevin dynamics simulations were performed using an interface to CHARMM¹⁸ developed previously⁶⁵, with 32 replicas for HP35 and 24 replicas for the other five systems. Temperatures were spaced approximately exponentially, ranging from the temperature at which the structure was determined experimentally to a maximum temperature of about 1480 K. Replicas simulated at adjacent temperatures were swapped with a probability of about 0.25 or more. Swaps were considered after 100 steps of Langevin dynamics using a collision frequency of 2 ps^{-1} for non-hydrogen atoms. Electrostatic interactions were shifted⁶⁶ to zero at $10 \text{ \AA}$, and Lennard-Jones interactions were switched off from 8 to $10 \text{ \AA}$. The SHAKE algorithm was applied to bonds involving hydrogen atoms, and the integration time step used to propagate dynamics was 1.5 fs for temperatures below 700 K and 1.0 fs for higher temperatures. The average temperature simulated for the first replica was observed to be 1 to 2 degrees below the nominal (experimental) temperature.

All systems were simulated with charged termini, except for trp-zip2, a C-terminal amide. The histidine of HP35 was protonated at the δ -nitrogen. Conformations sampled in dynamics simulations were characterized by the root mean-square deviation (rmsd) of main-chain (N, Ca, and C) atoms from an experimentally determined structure and by an effective energy, W . The first and last residues were ignored in calculating rmsd. The potential of mean force, W , is neither a potential energy nor a free energy, as W implicitly accounts for solvent entropy but not for the configurational entropy of the solute¹⁰.

Results

The self-energies calculated with the contact model (equation 8) are compared to those calculated with an analytical expression for atomic SASA in figure 1. This contact model more closely approximates the true SASA than does the previous model expressed in terms of exponential functions¹⁶.

The surface tensions estimated by simultaneously biasing different energy landscapes are $\sigma_0 = 5.26$, $\sigma_1 = 14.01$, $\sigma_w = 6.91$, $\sigma_F = 39.45 \text{ cal/\AA}^2$, and the screening parameters inferred are listed in Table 1. The amplitude of the additional ψ torsion is 0.14 kcal/mol (peak-to-trough difference = 0.28 kcal/mol).

The parameters in Table 1 call for self- and interaction-energies of atoms in charged groups to be screened significantly differently than the energies of other atoms. Electrostatic attraction and repulsion are weakened considerably by the large screening ($\alpha = 2.6 \text{ \AA}^{-1}$) of charge-charge interactions. By contrast, the self-energy of atoms in uncharged groups is highly screened relative to the self-energy of charged groups, diminishing the forces that move neutral groups to and from the protein surface. Thus, charges are especially favored on the protein surface while salt-bridges are only modestly stabilizing (Figure 2).

Structures sampled in folding simulations begun from an extended conformation were clustered by backbone configuration. For HP35, the first 48000 replica-exchange steps were ignored, and the structure saved every 500th step thereafter was analyzed, yielding 12000 structures to be clustered. For the other systems, the first 92000 were ignored, and every 200th step was analyzed, yielding 9500 structures.

Cluster hubs were added, one at a time, while calculating the distance R , equal to one half of the average hub-hub separation. An arbitrarily chosen structure was designated as the first hub, and the structure most distant from it was chosen as the second. Each structure was assigned to the cluster of its nearest hub (smallest backbone rmsd). If any structure was more than R from its hub, then the structure farthest from its hub was designated the third hub. Hubs were added, one at a time, until every structure was within R of its hub. The process was repeated with the first hub redefined as the structure having the most neighbors within the distance R . The value of R was found to be insensitive to the choice of the first hub. This simple approach⁶⁷ provides a convenient measure of the size of clusters sampled. It was extended here as follows. For each cluster, the hub was reassigned to be the centroid, the structure possessing the minimum rmsd to all members. The following three steps were then repeated until no clusters were merged: reassign each structure to its nearest hub, merge clusters with hubs closer than R , and center the hubs again. Finally, all clusters smaller than 10 % of the largest were discarded, and the entire process was repeated to focus with finer resolution on the space of significantly sampled structures. The resulting clusters are depicted in figures S4 to S9. The centroid of the largest cluster (Table 2, Figure 3) represents an objective prediction of the protein's structure in aqueous solution.

For each of these systems, the centroid of the largest cluster was a native-like structure. For trp-zip 2, MBH12, and trp-cage, this objective prediction was within 1.1 Å of the known structure. For U(1–17)T9D and BS1, the prediction was within 2.6 Å, and for the 35-residue HP35 protein, the prediction was within 3.7 Å. The sixth largest cluster sampled for HP35 represents 5.7% of the structures sampled, and its centroid is within 0.91 Å of the crystal structure (Figure S9). Low-energy structures from three of the six systems (trp-zip 2, U(1–17)T9D, and HP35) were utilized in the latter iterations of parameter refinement, and BS1 and MBH12 structures were never used to refine the energy function. These parameters are therefore transferrable, at least to some degree, and are likely to produce realistic simulations for other small proteins. By simplifying the SCP parameterization in a physically appealing way (*i.e.*, by using electrostatic parameters based on group charge), a much smaller set of parameters has been found that captures physical realism across a diverse group of structures.

The centroids of the six largest clusters are shown for each of the systems simulated in figures S4 to S9. For the shortest of the β peptides, trp-zip2, these structures vary considerably (figure S4), ranging from the prediction (centroid 1) that agrees very well with the NMR structure to a fully helical structure (centroid 5). By contrast, each of the six centroids shown in figure S7 for the BS1 peptide resembles the native structure. While the tryptophan side chain is correctly packed in the “cage” of trp-cage in the predicted structure (centroid 1), fairly modest changes in the fold are accompanied by incorrect packing of this side chain in clusters 2 and 5 (figure S8).

A second set of simulations using this same, optimized energy function with 15 replicas and a maximum temperature of 550 K did not sample HP35 structures with $\text{rmsd} < 2 \text{ \AA}$ (data not shown).

Discussion

The dielectric response of a polar liquid to either a point charge or a point dipole is a sigmoidal function of the distance from the source⁵⁷. Consequently, a simple continuum model formulated in terms of sigmoidal screening functions, the SCPISM, has been shown to be a physically reasonable and computationally efficient alternative to explicit water in the simulation of proteins and peptides^{16–17}. Here, the implementation of the model is simplified by requiring that the distance dependence of the screening of the electrostatic potential surrounding an atomic partial charge, *e.g.* the value of α in Eq. 1, is identical for all atoms within a given group and for all other atoms within electrostatically similar groups. Only three types of group are distinguished here: groups carrying a net charge, neutral groups in the protein backbone that participate in hydrogen bonding, and all other neutral groups. Another parameter is used for each type of group to govern self-energies. This use of separate α values for the self-energy allows for some imprecision in the estimation of the effective atomic radii. Thus, six screening parameters are used, replacing more than 30 values employed in previous implementations based on the atom type designated by the force field. This simplification is a significant step toward the ideal of a parameter-free continuum description⁵⁷.

A few additional physical approximations are explored here to simplify the model. The effective radius of a polar hydrogen atom was not modified to modulate hydrogen bonding, as in previous work¹⁶. Here, main-chain hydrogen bonding was modulated by using values of α for main-chain groups different from those used for other neutral groups. Moreover, the conformational dependence of the α parameters due to solvent exclusion¹⁷ was neglected here. The six screening parameters and four surface tensions reported were chosen using an iterative procedure to promote sampling of native-like conformations of peptides and mini-proteins.

The representative interactions depicted in Figure 2 indicate that charge-charge interactions are well screened. Salt bridges are only modestly stabilizing according to these parameters. This trend is consistent with the neutralization of charged groups used as part of other implicit solvent models (EEF1¹⁰, SASA⁸), but here, no charges were altered in the CHARMM force field. The weakening of salt bridges is achieved through the atomic screening parameters.

Despite the simplicity of the newly parameterized solvent model, the current results compare favorably to earlier results obtained with other implicit-solvent models used with CHARMM force fields²¹ and with more complicated parameterizations of the SCPISM¹⁷. For example, the current model predicted a trp-cage structure within 1.1 $\text{\AA}$ of the NMR structure and resulted in greater sampling of such structures (Figure S2e) than observed in our previous simulations that included solvent-exclusion effects¹⁷.

The inferred screening parameters and surface tensions approximate the forces exerted by an aqueous solution on a polymer in an arbitrary conformation. Estimating the strength of a given hydrogen bond, for example, within an arbitrary conformation is complicated by the presence of other residues proximal to the interacting donor and acceptor groups. The inference of interaction strengths from multiple simulations of peptides and mini-proteins in native and non-native conformations is one way to quantify many-body effects in an average way. The structure predicted by replica-exchange simulations was within 1.1 Å for three of the systems and within 2.6 Å for two more. These five peptides and mini-proteins include the two that were excluded from the parameter refinement. The results obtained for HP35 are encouraging but leave room for improvement.

The solvent model that has emerged here, referred to as SCP18, calls for strong main-chain hydrogen bonds and a pronounced bias for the burial of phenylalanine side chains. The ranges in hydrogen-bond strengths and in atomic surface tensions might result from oversimplifications in the current protein and solvent models, but they are not without justification. The drive to form backbone hydrogen bonds echoes the view of Rose and coworkers that protein folding is dominated by main-chain hydrogen bonding¹ and that the energetics of main-chain hydrogen bonding controls the shifts in protein stability induced by osmolytes².

The relatively large surface tension used for the atoms of phenylalanine side chains may reflect the limitations in the surface-area approximation to nonpolar energetics and a compensation for other imperfections in the energy landscape. Still, the value of $\sigma_F = 39.45$ cal/mol/Å² is less than the estimate of 47 cal/mol/Å² obtained in an analysis of alkane solubilities⁶⁸. Preliminary simulations of HP35 using the newer CHARMM36 force field⁶⁹ also call for deep minima for main-chain hydrogen bonds and phenylalanine burial (data not shown).

As protein and solvent models are refined, the amplitude of any additional torsional term used with implicit solvents might be diminished, but in practice, such modifications have proven useful. The torsion added here to favor β conformations is a minimum at $\psi = 150^\circ$, with an amplitude that is modest relative to similar adjustments^{11, 31, 70}.

This modification to the ψ torsion can be improved, presumably, along with the other ten parameters estimated here. Work continues in pursuit of a simple, balanced, physically realistic continuum model that better samples native conformations in folding simulations, particularly for HP35. The current results might be improved by reintroducing the conformational dependence of electrostatic screening¹⁷. Differentiating the dielectric screening around groups with a net electrostatic dipole moment from that surrounding apolar groups might warrant the addition of two parameters. Accounting for solute-solvent dispersion forces⁷¹ is another extension to be explored. The enhanced sampling achievable with a GPU implementation of this continuum solvent model would expedite further refinement and application to larger systems.

Conclusions

A simplified parameterization of a continuum solvent model based on screened Coulomb potentials was found that resulted in the sampling of native-like conformations for each of six peptides/mini-proteins in folding simulations initiated from an extended conformation. An efficient estimation of atomic solvent accessible surface area was also implemented. The parameters obtained reflect the need for strong backbone-backbone hydrogen bonds and for strong screening of charged interactions. A substantial energetic reward for the burial of phenylalanine side chains facilitated sampling of conformations of the villin headpiece subdomain within 1 Å of the x-ray structure. This simple parameterization of dielectric screening based on group charge should facilitate application to non-peptide systems, including simulations of nucleic acids and ligand binding.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

I thank Sergio A. Hassan for many stimulating discussions and for comments on this manuscript. This work utilized the computational resources of the NIH HPC Biowulf cluster (<http://hpc.nih.gov>) and was supported by the Intramural Research Program of the National Institutes of Health (NIH, CIT).

References

1. Rose GD; Fleming PJ; Banavar JR; Maritan A A Backbone-Based Theory of Protein Folding. *Proc. Natl. Acad. Sci. USA* 2006, 103 (45), 16623–16633. [PubMed: 17075053]
2. Bolen DW; Rose GD Structure and Energetics of the Hydrogen-Bonded Backbone in Protein Folding. *Annu. Rev. Biochem* 2008, 77 (1), 339–362.
3. Lindorff-Larsen K; Piana S; Dror RO; Shaw DE How Fast-Folding Proteins Fold. *Science* 2011, 334 (6055), 517–520. [PubMed: 22034434]
4. Nguyen H; Maier J; Huang H; Perrone V; Simmerling C Folding Simulations for Proteins with Diverse Topologies Are Accessible in Days with a Physics-Based Force Field and Implicit Solvent. *J. Am. Chem. Soc* 2014, 136 (40), 13959–13962. [PubMed: 25255057]
5. Gilson MK; Davis ME; Luty BA; McCammon JA Computation of Electrostatic Forces on Solvated Molecules using the Poisson-Boltzmann Equation. *J. Phys. Chem* 1993, 97 (14), 3591–3600.
6. Boda D; Valiskó M; Henderson D; Gillespie D; Eisenberg B; Gilson MK Ions and Inhibitors in the Binding Site of HIV Protease: Comparison of Monte Carlo Simulations and the Linearized Poisson-Boltzmann Theory. *Biophys. J* 2009, 96 (4), 1293–1306. [PubMed: 19217848]
7. Wesson L; Eisenberg D Atomic Solvation Parameters Applied to Molecular-Dynamics of Proteins in Solution. *Protein Sci* 1992, 1 (2), 227–235. [PubMed: 1304905]
8. Ferrara P; Apostolakis J; Caflisch A Evaluation of a Fast Implicit Solvent Model for Molecular Dynamics Simulations. *Proteins: Struct., Funct., Genet* 2002, 46 (1), 24–33. [PubMed: 11746700]
9. Allison JR; Boguslawski K; Fraternali F; van Gunsteren WF A Refined, Efficient Mean Solvation Force Model that Includes the Interior Volume Contribution. *J. Phys. Chem. B* 2011, 115 (15), 4547–4557. [PubMed: 21434626]
10. Lazaridis T; Karplus M Effective Energy Function for Proteins in Solution. *Proteins: Struct., Funct., Genet* 1999, 35 (2), 133–152. [PubMed: 10223287]
11. Bottaro S; Lindorff-Larsen K; Best RB Variational Optimization of an All-Atom Implicit Solvent Force Field To Match Explicit Solvent Simulation Data. *J. Chem. Theory Comput* 2013, 9 (12), 5641–5652. [PubMed: 24748852]

12. Haberthür U; Caflisch A FACTS: Fast Analytical Continuum Treatment of Solvation. *J. Comput. Chem* 2008, 29 (5), 701–715. [PubMed: 17918282]
13. Nguyen H; Roe DR; Simmerling C Improved Generalized Born Solvent Model Parameters for Protein Simulations. *J. Chem. Theory Comput* 2013, 9 (4), 2020–2034. [PubMed: 25788871]
14. Arthur EJ; Brooks CL Parallelization and Improvements of the Generalized Born Model with a Simple sWitching Function for Modern Graphics Processors. *J. Comput. Chem* 2016, 37 (10), 927–939. [PubMed: 26786647]
15. Hassan SA; Guarnieri F; Mehler EL A General Treatment of Solvent Effects Based on Screened Coulomb Potentials. *J. Phys. Chem. B* 2000, 104 (27), 6478–6489.
16. Hassan SA; Mehler EL; Zhang DQ Weinstein H, Molecular Dynamics Simulations of Peptides and Proteins with a Continuum Electrostatic Model Based on Screened Coulomb Potentials. *Proteins: Struct., Funct., Genet* 2003, 51 (1), 109–125. [PubMed: 12596268]
17. Hassan SA; Steinbach PJ Water-Exclusion and Liquid-Structure Forces in Implicit Solvation. *J. Phys. Chem. B* 2011, 115 (49), 14668–14682. [PubMed: 22007697]
18. Brooks BR; Brooks CL; Mackerell AD; Nilsson L; Petrella RJ; Roux B; Won Y; Archontis G; Bartels C; Boresch S et al. CHARMM: The Biomolecular Simulation Program. *J. Comput. Chem* 2009, 30 (10), 1545–1614. [PubMed: 19444816]
19. MacKerell AD; Bashford D; Bellott M; Dunbrack RL; Evanseck JD; Field MJ; Fischer S; Gao J; Guo H; Ha S et al. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* 1998, 102 (18), 3586–3616. [PubMed: 24889800]
20. MacKerell AD; Feig M; Brooks CL Extending the Treatment of Backbone Energetics in Protein Force Fields: Limitations of Gas-Phase Quantum Mechanics in Reproducing Protein Conformational Distributions in Molecular Dynamics Simulations. *J. Comput. Chem* 2004, 25 (11), 1400–1415. [PubMed: 15185334]
21. Steinbach PJ Exploring Peptide Energy Landscapes: A Test of Force Fields and Implicit Solvent Models. *Proteins: Struct. Funct. Bioinf* 2004, 57 (4), 665–677.
22. Freddolino PL; Schulten K Common Structural Transitions in Explicit-Solvent Simulations of Villin Headpiece Folding. *Biophys. J* 2009, 97 (8), 2338–2347. [PubMed: 19843466]
23. Cochran AG; Skelton NJ; Starovasnik MA Tryptophan Zippers: Stable, Monomeric β -hairpins. *Proc. Natl. Acad. Sci. USA* 2001, 98 (10), 5578–5583. [PubMed: 11331745]
24. Pastor MT; López de la Paz M; Lacroix E; Serrano L; Pérez-Payá E Combinatorial Approaches: A New Tool to Search for Highly Structured β -Hairpin Peptides. *Proc. Natl. Acad. Sci. USA* 2002, 99 (2), 614–619. [PubMed: 11782528]
25. Zerella R; Chen PY; Evans PA; Raine A; Williams DH Structural Characterization of a Mutant Peptide Derived from Ubiquitin: Implications for Protein Folding. *Protein Sci* 2000, 9 (11), 2142–2150. [PubMed: 11152124]
26. De Alba E; Santoro J; Rico M; Jimenez MA De Novo Design of a Monomeric Three-Stranded Antiparallel Beta-Sheet. *Protein Sci* 1999, 8 (4), 854–865. [PubMed: 10211831]
27. Santiveri CM; Santoro J; Rico M; Jimenez MA Factors Involved in the Stability of Isolated Beta-Sheets: Turn Sequence, Beta-Sheet Twisting, and Hydrophobic Surface Burial. *Protein Sci* 2004, 13 (4), 1134–1147. [PubMed: 15044739]
28. Neidigh JW; Fesinmeyer RM; Andersen NH Designing a 20-Residue Protein. *Nature Structural Biology* 2002, 9 (6), 425–430. [PubMed: 11979279]
29. Chiu TK; Kubelka J; Herbst-Irmer R; Eaton WA; Hofrichter J; Davies DR High-Resolution X-ray Crystal Structures of the Villin Headpiece Subdomain, an Ultrafast Folding Protein. *Proc. Natl. Acad. Sci. USA* 2005, 102 (21), 7517–7522. [PubMed: 15894611]
30. Asim O; Bentley S; Viktor H; Carlos S Using PC Clusters to Evaluate the Transferability of Molecular Mechanics Force Fields for Proteins. *J. Comput. Chem* 2003, 24 (1), 21–31. [PubMed: 12483672]
31. Chen JH; Im WP; Brooks CL Balancing Solvation and Intramolecular Interactions: Toward a Consistent Generalized Born Force Field. *J. Am. Chem. Soc* 2006, 128 (11), 3728–3736. [PubMed: 16536547]

32. Matthes D; de Groot BL Secondary Structure Propensities in Peptide Folding Simulations: A Systematic Comparison of Molecular Mechanics Interaction Schemes. *Biophys. J* 2009, 97 (2), 599–608. [PubMed: 19619475]
33. Kim E; Yang C; Jang S; Pak Y Free Energy Landscapes of a Highly Structured β -Hairpin Peptide and its Single Mutant. *J. Chem. Phys* 2008, 129 (16), 165104. [PubMed: 19045319]
34. Jang S; Shin S; Pak Y Molecular Dynamics Study of Peptides in Implicit Water: Ab Initio Folding of Beta-Hairpin, Beta-Sheet, and Beta-Beta-Alpha Motif. *J. Am. Chem. Soc* 2002, 124 (18), 4976–4977. [PubMed: 11982359]
35. Ulmschneider JP; Jorgensen WL Polypeptide Folding using Monte Carlo Sampling, Concerted Rotation, and Continuum Solvation. *J. Am. Chem. Soc* 2004, 126 (6), 1849–1857. [PubMed: 14871118]
36. Kalgin IV; Caflisch A; Chekmarev SF; Karplus M New Insights into the Folding of a β -Sheet Miniprotein in a Reduced Space of Collective Hydrogen Bond Variables: Application to a Hydrodynamic Analysis of the Folding Flow. *J. Phys. Chem. B* 2013, 117 (20), 6092–6105. [PubMed: 23621790]
37. Ferrara P; Caflisch A Folding Simulations of a Three-Stranded Antiparallel Beta-Sheet Peptide. *Proc. Natl. Acad. Sci. USA* 2000, 97 (20), 10780–10785. [PubMed: 10984515]
38. Simmerling C; Strockbine B; Roitberg AE All-Atom Structure Prediction and Folding Simulations of a Stable Protein. *J. Am. Chem. Soc* 2002, 124 (38), 11258–11259. [PubMed: 12236726]
39. Snow CD; Zagrovic B; Pande VS The Trp cage: Folding Kinetics and Unfolded State Topology via Molecular Dynamics Simulations. *J. Am. Chem. Soc* 2002, 124 (49), 14548–14549. [PubMed: 12465960]
40. Carnevali P; Toth G; Toubassi G; Meshkat SN Fast Protein Structure Prediction using Monte Carlo Simulations with Modal Moves. *J. Am. Chem. Soc* 2003, 125, 14244–14245. [PubMed: 14624550]
41. Zhou RH Trp-cage: Folding Free Energy Landscape in Explicit Water. *Proc. Natl. Acad. Sci. USA* 2003, 100 (23), 13280–13285. [PubMed: 14581616]
42. Nymeyer H; Garcia AE Diversity of Structure in the TRPcage Protein Studied by Replica Exchange Molecular Dynamics. *Biophys. J* 2004, 86 (1), 624A–624A.
43. Paschek D; Nymeyer H; García AE Replica Exchange Simulation of Reversible Folding/Unfolding of the Trp-cage Miniprotein in Explicit Solvent: On the Structure and Possible Role of Internal Water. *J. Struct. Biol* 2007, 157 (3), 524–533. [PubMed: 17293125]
44. Okur A; Miller BT; Joo K; Lee J; Brooks BR Generating Reservoir Conformations for Replica Exchange through the Use of the Conformational Space Annealing Method. *J. Chem. Theory Comput* 2013, 9 (2), 1115–1124. [PubMed: 23585739]
45. Duan Y; Kollman PA Pathways to a Protein Folding Intermediate Observed in a 1-Microsecond Simulation in Aqueous Solution. *Science* 1998, 282 (5389), 740–744. [PubMed: 9784131]
46. Fernández A; Shen M.-y.; Colubri A; Sosnick TR; Berry RS; Freed KF Large-Scale Context in Protein Folding: Villin Headpiece. *Biochemistry* 2003, 42 (3), 664–671. [PubMed: 12534278]
47. Jayachandran G; Vishal V; Pande VS Using Massively Parallel Simulation and Markovian Models to Study Protein Folding: Examining the Dynamics of the Villin Headpiece. *J. Chem. Phys* 2006, 124 (16), 164902. [PubMed: 16674165]
48. Wickstrom L; Bi Y; Hornak V; Raleigh DP; Simmerling C Reconciling the Solution and X-ray Structures of the Villin Headpiece Helical Subdomain: Molecular Dynamics Simulations and Double Mutant Cycles Reveal a Stabilizing Cation- π Interaction. *Biochemistry* 2007, 46 (12), 3624–3634. [PubMed: 17338549]
49. Piana S; Lindorff-Larsen K; Shaw DE Protein Folding Kinetics and Thermodynamics from Atomistic Simulation. *Proc. Natl. Acad. Sci. USA* 2012, 109 (44), 17845–17850. [PubMed: 22822217]
50. Best RB; Hummer G Microscopic Interpretation of Folding ϕ -Values using the Transition Path Ensemble. *Proc. Natl. Acad. Sci. USA* 2016, 113 (12), 3263–3268. [PubMed: 26957599]
51. Kubelka J; Eaton WA; Hofrichter J Experimental Tests of Villin Subdomain Folding Simulations. *J. Mol. Biol* 2003, 329 (4), 625–630. [PubMed: 12787664]

52. Henry ER; Best RB; Eaton WA Comparing a simple theoretical model for protein folding with all-atom molecular dynamics simulations. *Proc. Natl. Acad. Sci. USA* 2013, 110 (44), 17880–17885. [PubMed: 24128764]
53. Muñoz V; Thompson PA; Hofrichter J; Eaton WA Folding dynamics and mechanism of β -hairpin formation. *Nature* 1997, 390, 196. [PubMed: 9367160]
54. Schuler B; Lipman EA; Eaton WA Probing the free-energy surface for protein folding with single-molecule fluorescence spectroscopy. *Nature* 2002, 419, 743. [PubMed: 12384704]
55. Henry ER; Eaton WA Combinatorial modeling of protein folding kinetics: free energy profiles and rates. *Chemical Physics* 2004, 307 (2), 163–185.
56. Chung HS; Eaton WA Protein folding transition path times from single molecule FRET. *Current Opinion in Structural Biology* 2018, 48, 30–39. [PubMed: 29080467]
57. Hassan SA; Mehler EL From Quantum Chemistry and the Classical Theory of Polar Liquids to Continuum Approximations in Molecular Mechanics Calculations. *Int. J. Quantum Chem* 2005, 102 (5), 986–1001.
58. Still WC; Tempczyk A; Hawley RC; Hendrickson T Semianalytical Treatment of Solvation for Molecular Mechanics and Dynamics. *J. Am. Chem. Soc* 1990, 112 (16), 6127–6129.
59. Reynolds JA; Gilbert DB; Tanford C Empirical Correlation Between Hydrophobic Free Energy and Aqueous Cavity Surface Area. *Proc. Natl. Acad. Sci. USA* 1974, 71 (8), 2925–2927. [PubMed: 16578715]
60. Wodak SJ; Janin J Analytical Approximation to the Accessible Surface Area of Proteins. *Proc. Natl. Acad. Sci. USA* 1980, 77 (4), 1736–1740. [PubMed: 16592793]
61. Weiser J; Shenkin PS; Still WC Approximate Atomic Surfaces from Linear Combinations of Pairwise Overlaps (LCPO). *J. Comput. Chem* 1999, 20 (2), 217–230.
62. Richmond TJ Solvent Accessible Surface Area and Excluded Volume in Proteins: Analytical Equations for Overlapping Spheres and Implications for the Hydrophobic Effect. *J. Mol. Biol* 1984, 178 (1), 63–89. [PubMed: 6548264]
63. Powell MJD A View of Algorithms for Optimization Without Derivatives. In Cambridge University Technical Report DAMPT 2007/NA03, 2007.
64. Sugita Y; Okamoto Y Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett* 1999, 314 (1), 141–151.
65. Maragakis P; Spichty M; Karplus M Optimal Estimates of Free Energies from Multistate Nonequilibrium Work Data. *Phys. Rev. Lett* 2006, 96 (10), 100602. [PubMed: 16605720]
66. Steinbach PJ; Brooks BR New Spherical-Cutoff Methods for Long-Range Forces in Macromolecular Simulation. *J. Comput. Chem* 1994, 15 (7), 667–683.
67. Clason R Finding Clusters: An Application of the Distance Concept. *The Mathematics Teacher* 1990, 83 (4), 301–307.
68. Sharp K; Nicholls A; Fine R; Honig B Reconciling the Magnitude of the Microscopic and Macroscopic Hydrophobic Effects. *Science* 1991, 252 (5002), 106–109. [PubMed: 2011744]
69. Best RB; Zhu X; Shim J; Lopes PEM; Mittal J; Feig M; MacKerell AD Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone ϕ , ψ and Side-Chain χ_1 and χ_2 Dihedral Angles. *J. Chem. Theory Comput* 2012, 8 (9), 3257–3273. [PubMed: 23341755]
70. Best RB; Hummer G Optimized Molecular Dynamics Force Fields Applied to the Helix–Coil Transition of Polypeptides. *J. Phys. Chem. B* 2009, 113 (26), 9004–9015. [PubMed: 19514729]
71. Hassan SA Implicit Treatment of Solvent Dispersion Forces in Protein Simulations. *J. Comput. Chem* 2014, 35 (22), 1621–1629. [PubMed: 24919463]
72. Kraulis PJ MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. *Journal of Applied Crystallography* 1991, 24, 946–950.
73. Merritt EA; Bacon DJ Raster3D: Photorealistic molecular graphics. In *Macromolecular Crystallography*, Pt B, 1997; Vol. 277, pp 505–524.

Figure 1.
Atomic self-energies calculated for 810,675 non-polar-hydrogen atoms in high-resolution structures.

Figure 2.

Representative interactions illustrated in figure S1, as given by the parameters inferred from iterative refinement. The Glu-Arg interaction energy is plotted in red, Glu-Lys in blue, Glu-His in light blue, Gln-His in green, NMA-NMA in orange, and Phe-Phe in gray. For the five polar interactions, x represents the heavy-atom separation along a linear hydrogen bond. NMA = *N*-methyl acetamide.

Figure 3: Energy landscapes and predicted native conformations for diverse set of model systems.

Left) Effective energy *versus* backbone deviation from experimental structure, for conformations sampled in simulations started from an extended conformation, for a) trp-zip 2, b) MBH12, c) U(1-17)T9D, d) BS1, e) trp-cage, and f) villin subdomain HP35. The extent of sampling is indicated by color, from red (most) to purple (least). Right) Centroid of the largest cluster of sampled structures (gold), superimposed on the experimental structure (green), shown as α -carbon traces. Structures were rendered using the programs MolScript⁷² and Raster3D⁷³.

Table 1.

Electrostatic screening parameters, α , in $\text{\AA}^{-1}$.

	Charged group	H-bonding, main-chain group	Other Neutral group
self-energy α	0.5443	2.1767	6.7873
interaction-energy α	2.6281	0.4797	0.3969

Table 2

Deviation of centroid from experiment and cluster size for two largest clusters of structures.

Peptide/Mini-protein	# residues	Cluster 1		Cluster 2 %	
		Centroid		Centroid	
		rmsd (Å)	% sampled	rmsd (Å)	% sampled
trp-zip 2	12	0.47	17.7	4.30	6.9
MBH12	14	0.86	24.1	2.11	6.7
U(1–17)T9D	17	2.53	28.0	2.30	22.0
BS1	20	2.03	12.9	3.01	10.8
trp-cage	20	1.08	38.6	1.51	28.9
villin HP35	35	3.64	13.6	3.99	9.7