PROTEINS: Structure, Function, and Genetics 35:133–152 (1999)

RESEARCH ARTICLES

Effective Energy Function for Proteins in Solution

Themis Lazaridis[1] **and Martin Karplus**[1,2]*
[1]*Department of Chemistry and Chemical Biology, Harvard University, Cambridge, Massachusetts*
[2]*Laboratoire de Chimie Biophysique, ISIS, Université Louis Pasteur, Strasbourg, France*

ABSTRACT **A Gaussian solvent-exclusion model for the solvation free energy is developed. It is based on theoretical considerations and parametrized with experimental data. When combined with the CHARMM 19 polar hydrogen energy function, it provides an effective energy function (EEF1) for proteins in solution. The solvation model assumes that the solvation free energy of a protein molecule is a sum of group contributions, which are determined from values for small model compounds. For charged groups, the self-energy contribution is accounted for primarily by the exclusion model. Ionic side-chains are neutralized, and a distance-dependent dielectric constant is used to approximate the charge–charge interactions in solution. The resulting EEF1 is subjected to a number of tests. Molecular dynamics simulations at room temperature of several proteins in their native conformation are performed, and stable trajectories are obtained. The deviations from the experimental structures are similar to those observed in explicit water simulations. The calculated enthalpy of unfolding of a polyalanine helix is found to be in good agreement with experimental data. Results reported elsewhere show that EEF1 clearly distinguishes correctly from incorrectly folded proteins, both in static energy evaluations and in molecular dynamics simulations and that unfolding pathways obtained by high-temperature molecular dynamics simulations agree with those obtained by explicit water simulations. Thus, this energy function appears to provide a realistic first approximation to the effective energy hypersurface of proteins. Proteins 1999;35:133–152.**

Key words: implicit solvation; hydration thermodynamics; binding energy; molecular dynamics; computer simulations

INTRODUCTION

Knowledge of the effective energy hypersurface of biological macromolecules in solution is of fundamental importance for understanding their properties. The effective energy (potential of mean force) for a given conformation of the macromolecule is the free energy of the system consisting of the macromolecule and the solvent with an average over all solvent degrees of freedom at a given temperature.[1] It consists of the intramolecular energy (the energy of the macromolecule in vacuum) and the solvation free energy (the free energy of transfer of the macromolecule from the gas phase to solution). The total free energy of the macromolecule-solvent system in a given region of the hypersurface is the sum of the average effective energy and the configurational entropy. Under physiological conditions, it is generally assumed that proteins are stable in the neighborhood of the global minimum (the native conformation) despite the large configurational entropy gain in the denatured state. This is referred to as the "thermodynamic hypothesis" of protein stability, and available evidence suggests that it is valid for the majority of small, single-domain proteins.[2] Even where ergodicity seems to break down,[3,4] the native state must correspond to a deep minimum on the effective energy hypersurface, which is separated from the global minimum by barriers too high to be traversed on experimental time scales. Theoretical predictions of the structure of a protein from the amino acid sequence and analysis of the mechanism of folding requires knowledge of the effective energy hypersurface and a method for efficient search of the conformation space to determine the location of the minima.

The all-atom force fields used in molecular mechanics and dynamics simulations of macromolecules[5–8] give the energy of a protein in vacuum. Only by introducing explicit solvent and doing simulations including both the macromolecule and solvent can solvation effects be accounted for. It is clear from simulation results and theoretical considerations that the effective energy hypersurface including solvent is significantly different from the intramolecular energy hypersurface.[9] Although the intramolecular energy hypersurface often has a deep minimum near the native conformation, it may have other equally deep or deeper minima in distant regions of the conformation space. This can arise from a number of effects. For example, polar–

Themis Lazaridis' present address is Department of Chemistry, City College of New York, Convent Avenue & 138th Street, New York, NY 10031.
*Correspondence to: Martin Karplus, Department of Chemistry and Chemical Biology, Harvard University, 12 Oxford Street, Cambridge, MA 02138. E-mail: marci@tammy.harvard.edu
Received 7 July 1998; Accepted 17 November 1998

nonpolar group interactions in the gas phase are as favorable energetically as nonpolar–nonpolar interactions. In water, however, polar–nonpolar interactions are effectively repulsive because of the high penalty of desolvating the polar group. Moreover, hydrophobic hydration makes the effective interactions between nonpolar groups in water stronger than in the gas phase.[9] Thus, water aids in stabilizing the native state with buried nonpolar groups and helps to guarantee its uniqueness. Opposite charge–-charge interactions are much less stabilizing in water than in vacuum.[10] Thus, problems related to protein folding and stability cannot be addressed without taking account of solvation.

Although significant progress has been made in the statistical theory of liquids during the past 40 years,[11,12] simple and accurate models for aqueous solvation are still lacking. Currently, the most reliable method of taking solvation into account is to simulate a protein in the presence of explicit water molecules. This approach appears to give reasonable dynamic and thermodynamic properties with the presently available force fields, but it incurs a large computational expense. In fact, most of the computer time in such simulations is devoted to calculating solvent–solvent interactions; e.g., in simulations of barnase denaturation, the system consisted of 1,100 protein atoms and about 9,000 solvent atoms.[13] Having to include such a large number of solvent atoms places severe limits on the type of problems that can be studied. For example, although the dynamics of native states can be simulated effectively with explicit solvent models,[14] simulations of unfolding (and of unfolded states) over the time scales required for sampling a large enough number of different initial states to obtain meaningful converged results is not yet possible, except perhaps for small peptides.

Another limitation of explicit solvation simulations is that free energy differences are not obtained in a straightforward way. For example, explicit water simulations of the native and a misfolded structure do not show which one has the lowest free energy. Only by special simulations involving the reversible transition from one conformation to the other or umbrella sampling[15] can the required free energy differences be calculated. Such simulations have been carried out so far mainly for small molecules, such as butane,[16,17] dipeptides,[18–20] and other small peptides.[21–23] Recently, the potential of mean force for a small protein with respect to the radius of gyration as an order parameter has been calculated with explicit solvent;[24] the simulation required 450 hr of CPU time on a 64-node T3E supercomputer (the equivalent of 4 months on a standard workstation) (C.L. Brooks, personal communication). Thus, the use of this approach to study the details of the folded and unfolded states of a series of proteins, as well as many other problems of current interest, is still not possible.

To overcome the limitations of vacuum calculations on the one hand and explicit solvent simulations on the other, a variety of implicit solvation models for proteins have been developed that combine an empirical force field for the intramolecular interactions in vacuum with a solvation correction. The latter is obtained by considering the transfer of the entire protein or appropriate constituent groups from the gas phase to water. One very simple model[25] used the CHARMM potential energy function [5,26] and an atom-based surface area solvation correction; five different atom types were included. The same solvation model was combined with the AMBER force field by Schiffer et al.[27] Fraternali and van Gunsteren[28] proposed an even simpler model for use in molecular dynamics (MD) simulations; this model is also based on accessible surface areas and uses only two parameters—one for nonpolar and one for polar groups.

Although the above models evaluated the solvation effect in terms of the exposed surface area, there is no fundamental theoretical justification for this choice. A model that does not use the surface area is the hydration shell model.[29] It assumes that the hydration free energy of a group arises from the first hydration shell and that it is proportional to the volume of the hydration shell that is accessible to the solvent (i.e., that not occupied by other solute atoms). Another type of solvation model is based on the contacts each group makes with other solute atoms.[30] The more contacts there are, the smaller is the magnitude of the solvation free energy of the group, and the contacts are weighted according to some function of their distance from the group. This model is similar to approaches used earlier by Gibson and Scheraga[31] and by Levitt.[32] Physically, the Colonna-Cesari-Sander model[30] is similar in spirit to the surface area models, but it is much faster to use because counting the number of contacts takes considerably less time than even the most efficient analytical methods for calculating the surface area.[33,34] Furthermore, analytical derivatives of the solvation energy can be obtained easily to evaluate the forces necessary for energy minimization and molecular dynamics simulations. A version of this model was parametrized based on solvation free energies of small molecules.[35] The model assumes a linear relationship between the solvation free energy and a weighted sum of the contacts that the group makes with other solute atoms. A solvation parameter is assigned to each group. The model was combined with the GROMOS force field and used in a stochastic dynamics simulation of bovine pancreatic trypsin inhibitor (BPTI). It was found that the resulting structures were reasonable, although the deviation from the crystal structure was somewhat larger than in simulations with explicit water.

Another set of solvation models treats the entire protein at once and is based on continuum electrostatics and the linearized Poisson-Boltzmann equation.[36] Because numerical solutions of the Poisson-Boltzmann equation are expensive, semianalytical or analytical approximations have been proposed. Still et al.[37] introduced a simple generalization of the Born formula to polyatomic molecules. More recently, the generalized Born equation was combined with an integrated field method for self-energies to give a completely analytical treatment of electrostatic energies and forces.[38] Applications have been made to a number of simple systems, and it is likely that this approach will be used more extensively in the future.

In this article, we develop a new solvent-exclusion model for the solvation free energy and combine it with the CHARMM polar hydrogen energy function.[5,26] The purpose of the present implementation is to provide a realistic and reasonably accurate approximation to the effective energy hypersurface of proteins that is fast to use for calculation of energies and energy derivatives. Unlike many of the statistically based effective energy functions,[39–41] which appear to be limited to the evaluation of protein conformations obtained by threading and related procedures,[42] the resulting energy function is applicable to protein-folding and protein-unfolding studies. The formulation proposed in this article is only 50% slower than a vacuum simulation and thus makes possible many studies for which simulations in explicit water are prohibitively expensive. Given its physical basis and decomposability, the effective energy function can be used for approximate thermodynamic analysis of contributions to protein stability. It is expected to have broad applicability and complement the analytical Poisson-Boltzmann-based methods.

The model is presented in the next section, and applications of the model that test its accuracy are described in the section entitled Tests and Applications of EEF1. It is shown that the model is a significant improvement for molecular dynamics simulations of proteins, relative to the use of the vacuum potential (section entitled Simulations of Native Proteins). A more stringent test is provided by comparing the effective energy of the native structure and structural models for the unfolded state (in the section on Enthalpy and Free Energy of Unfolding). In the section on Helix Unfolding Enthalpy, the model is applied to determine the enthalpy of unfolding of a polyalanine helix. The Appendix provides a concise definition of the effective energy of a macromolecule in solution and shows that it determines its conformational distribution. In addition to the tests and applications described here, the method has been used to determine multiple pathways for CI2 unfolding at high temperatures.[43] A separate description of its use to distinguish native from misfolded protein structures is given elsewhere.[44] Overall, the applications indicate that the effective energy function (EEF1) developed here is a realistic approximation for proteins and polypeptides in aqueous solution. Possible improvements are outlined briefly in the Discussion.

DESCRIPTION OF THE MODEL

The effective energy $W(\mathbf{R}^M)$ of the macromolecule with coordinates $\mathbf{R}^M$ in solution can be written (see Appendix)

$$W(\mathbf{R}^M) = H_{intra}(\mathbf{R}^M) + \Delta G^{slv}(\mathbf{R}^M) \quad (1)$$

where H_{intra} is the intramacromolecule energy, and ΔG^{slv} is the solvation free energy. To obtain Equation 1, the only assumption is that the Hamiltonian is separable; that is, it is the sum of solute-solute, solute-solvent, and solvent-solvent terms. This is the case in most empirical energy functions that do not include polarization. Recent theoretical work on solvation thermodynamics[45,46] has shown that the solvation free energy ΔG^{slv} of a given conformation $\mathbf{R}^M$ can be written as an integral over the space around it; that is,

$$\Delta G^{slv} = \int f(\mathbf{r})\, d\mathbf{r} \quad (2)$$

where $f(\mathbf{r})$ is the solvation free energy density at point $\mathbf{r}$. It contains contributions from solute-solvent energy, solvent reorganization energy, solute-solvent entropy, and solvent reorganization entropy.[45,46] The solvation free energy density is expected to be a strong function of distance. Its magnitude is largest close to the solute and decays to zero far from the solute.[45,46]

When two solute molecules approach each other or the conformation of a polyatomic solute changes, the solvation of each group changes because of two effects: (a) exclusion of solvent from the space occupied by other solute groups, and (b) modification of the density and orientational distribution of the solvent in the remaining, nonsolute occupied space. For electrostatic interactions, the self-energy of the charges belongs to the first category and dielectric screening belongs to the second. In the present model, we neglect the second effect for the nonpolar groups because it is expected to be small and partly take it into account for polar groups through the use of a distance-dependent dielectric constant.

We assume that for a polyatomic solute, we can write the solvation free energy as a sum over group contributions, that is,

$$\Delta G^{slv} = \sum_i \Delta G_i^{slv} \quad (3)$$

where ΔG_i^{slv} is the solvation free energy of group i. Equation 3 can be formally derived by considering the solute-solvent interaction energy as a sum of group-solvent interactions and the solute-solvent correlation function as a product of group-solvent correlation functions (T.L., M.K, unpublished). Taking into account only the solvent exclusion effect, we can write

$$\Delta G_i^{slv} = \Delta G_i^{ref} - \sum_j \int_{Vj} f_i(\mathbf{r})\, d\mathbf{r} \quad (4)$$

where ΔG_i^{ref} (reference solvation free energy) is the solvation free energy of i in a suitably chosen small molecule in which group i is essentially fully solvent-exposed (but see below). The integral in Equation 4 is over the volume Vj of group j and the summation is over all groups j around i. To simplify the calculation, the integral over $f_i(\mathbf{r})$ is approximated by the product $f_i(r_{ij})V_j$, so that

$$\Delta G_i^{slv} = \Delta G_i^{ref} - \sum_{j \neq i} f_i(r_{ij})V_j \quad (5)$$

where r_{ij} is the distance between i and j. Equation 5 says that the solvation free energy of group i is that in the model system (ΔG_i^{ref}) minus the reduction in solvation because of the presence of surrounding groups. The solvation free energy density is assumed to be given by the Gaussian

function (see below).

$$f_i(r)4\pi r^2 = \alpha_i \exp(-x_i^2), \qquad x_i = \frac{r - R_i}{\lambda_i} \tag{6}$$

where R_i is the van der Waals radius of i (1/2 of the distance to the energy minimum in the Lennard-Jones potential), λ_i is a correlation length, and α_i is a proportionality coefficient given by

$$\alpha_i = 2\Delta G_i^{free}/\sqrt{\pi}\lambda_i \tag{7}$$

where ΔG_i^{free} is the solvation free energy of the free (isolated) group i; ΔG_i^{free} is close to ΔG_i^{ref} but not identical to it and is determined empirically by requiring that the solvation free energy of deeply buried groups be zero (see below). The parameters used in the model are given in Table I.

Given Equations 3–6, the solvation free energy evaluation requires a choice of atom types i, their volumes V_i, their correlation length λ_i , ΔG_i^{ref}, and ΔG_i^{free}. The specific atom types and solvation parameters are chosen to be used in conjunction with the CHARMM 19 polar hydrogen energy function. The atom types (Table I) are those used in CHARMM 19. To simplify the treatment of electrostatic interactions (see below), the ionic side-chains are neutralized and a linear distance-dependent dielectric constant is used ($\epsilon = r$). Also, certain choices were made for calculating the nonbonded interactions. Both van der Waals and electrostatic interactions are cut off at 9 Å with a switching function used between 7 and 9Å. Electrostatic interactions are calculated on a group-by-group basis. The value of λ_i was taken to be the thickness of one hydration shell (3.5 Å) except for the neutralized ionic groups, for which a value of 6 Å was used.

TABLE I. Solvation Parameters†

Atom types[a]	Volume	ΔG_i^{ref} [b]	ΔG_i^{free} [c]	ΔH_i^{ref} [b]	ΔCp_i^{ref} [d]
C	14.7	0.000	0.00	0.000	0.00
CR	8.3	−0.890	−1.40	2.220	6.90
CH1E	23.7	−0.187	−0.25	0.876	0.00
CH2E	22.4	0.372	0.52	−0.610	18.60
CH3E	30.0	1.089	1.50	−1.779	35.60
CR1E	18.4	0.057	0.08	−0.973	6.90
NH1	4.4	−5.950	−8.90	−9.059	−8.80
NR	4.4	−3.820	−4.00	−4.654	−8.80
NH2	11.2	−5.450	−7.80	−9.028	−7.00
NH3	11.2	−20.000	−20.00	−25.000	−18.00
NC2	11.2	−10.000	−10.00	−12.000	−7.00
N	0.0	−1.000	−1.55	−1.250	8.80
OH1	10.8	−5.920	−6.70	−9.264	−11.20
O	10.8	−5.330	−5.85	−5.787	−8.80
OC	10.8	−10.000	−10.00	−12.000	−9.40
S	14.7	−3.240	−4.10	−4.475	−39.90
SH1E	21.4	−2.050	−2.70	−4.475	−39.90

†Volumes in Å,[3] free energies and enthalpies in kcal/mol, heat capacities in cal/mol · K. λ_i in Equation 6 is 3.5 Å for all except for NH3, NC2, OC, for which it is 6.0Å. For polar hydrogens, the parameters in this table are zero. All free energy, enthalpy, and heat capacity values refer to 298.15 K.

[a]Atom types: C: carbonyl carbon, CR: carbon with no hydrogens, CR1E: extended aromatic carbon with 1 H, CH1E: extended aliphatic carbon with 1 H, CH2E: extended aliphatic carbon with 2 H, CH3E: extended aliphatic carbon with 3 H, NH1: amide nitrogen, NR: aromatic nitrogen with no hydrogens, NH2: nitrogen bound to two hydrogens, NH3: nitrogen bound to three hydrogens, NC2: guanidinium nitrogen, N: proline nitrogen, OH1: hydroxyl oxygen, O: carbonyl oxygen, OC: carboxyl oxygen, S: sulphur, SH1E: extended sulphur with one hydrogen; see Neria et al.[26]

[b]ΔG_i^{ref} from Table 2 of Privalov and Makhatadze[47] and ΔH_i^{ref} from Table 2 of Makhatadze and Privalov[48] plus 0.2 kcal/mol except for (a) NH3, NC2, and OC, which are such that the solvation free energy of the pseudoionic side-chains is −20 kcal/mol (see text). The enthalpies for these groups are given a value larger in magnitude than the free energy to account for a negative solvation entropy; (b) proline N, for which no parameters were available; it is given values smaller in magnitude than for the other, more polar nitrogens. For the carbonyl group, all solvation free energy is assigned to the O.

[c]ΔG_i^{free} is determined by increasing the magnitude of ΔG_i^{ref} until the solvation free energy of deeply buried groups becomes zero (see text).

[d]ΔCp_i^{ref} obtained from Table 1 of Privalov and Makhatadze[49] and the surface areas of the amino acid side-chains in Table 5 of Makhatadze and Privalov.[50]

Other Thermodynamic Quantities

The same model can be used to calculate other solvation properties, such as the solvation enthalpy, entropy, and heat capacity with the appropriate reference solvation values for the enthalpy and heat capacity. These values make it possible to calculate the solvation free energies at any temperature T. Assuming that the heat capacity is independent of temperature, we have

$$\frac{\partial \Delta G_i^{ref}}{\partial T} = -\Delta S_i^{ref}(T) \tag{8}$$

$$\Delta S_i^{ref}(T) = \Delta S_i^{ref}(To) + \int_{To}^{T} \Delta Cp_i/T dT = \Delta S_i^{ref}(To) + \Delta Cp_i \ln \frac{T}{To} \tag{9}$$

$$\Delta G_i^{ref}(T) = \Delta G_i^{ref}(To) - \int_{To}^{T} \Delta S_i^{ref}(T)\, dT = \Delta G_i^{ref}(To) - \Delta S_i^{ref}(To)[T - To] - \Delta Cp_i T \ln \frac{T}{To} + \Delta Cp_i[T - To] \tag{10}$$

where To is taken to be 298.15 K (25°C) and

$$\Delta S_i^{ref}(To) = \frac{\Delta H_i^{ref}(To) - \Delta G_i^{ref}(To)}{To}. \tag{11}$$

Details of the Model and Parameters

Thermodynamic parameters

The reference solvation free energies were obtained from Table 2 of Privalov and Makhatadze,[47] after subtraction of the long-range van der Waals contribution, −0.2 kcal/mol (see below). They are shown in Table I. The ΔG_i^{ref}, ΔG_i^{free} values in Table I refer to 298.15 K. Values at other temperatures are determined using the ΔH_i^{ref} and ΔCp_i^{ref} values. $\Delta H_i^{ref}(T)$ are obtained from Table 2 of Makhatadze

and Privalov[48] after subtracting the long-range van der Waals contribution. The values of ΔH_i^{ref} for the pseudoionic groups in Table I were chosen in analogy to those of the polar groups; i.e., the enthalpy values are larger in magnitude than the free energy values, the difference being accounted for by negative solvation entropies. This is qualitatively consistent with the Born model, although the Born model predicts smaller solvation entropies. In any case, the choice does not have a strong effect on the difference between folded and unfolded conformations, because ionic side-chains are seldom buried in the interior of proteins. For ΔCp_i^{ref} (T), we used the surface area proportionality coefficients of Privalov and Makhatadze (Table 1 of ref. 49) and the surface areas of the amino acid side-chains given in Table 5 of ref. 50 to derive values for each group type. The solvation enthalpy and heat capacity are determined based on the calculated solvation free energy:

$$\Delta H_i^{slv} = \Delta H_i^{ref} * \Delta G_i^{slv}/\Delta G_i^{ref} \quad (12)$$

$$\Delta Cp_i^{slv} = \Delta Cp_i^{ref} * \Delta G_i^{slv}/\Delta G_i^{ref} \quad (13)$$

Functional form of $f_i(r_{ij})$

The functional form of f_i could be determined, in principle, by statistical mechanical and computer simulation studies for various types of groups[45] or other theoretical approaches. To date, only a limited number of such studies have been made. For methane in water, about 90% of the solute-solvent orientational entropy[51] and 70% of the solvent reorganization energy[52] were found to arise from the first solvation shell . In pure water, 85% of the energy and 94% of the entropy arise from the first neighbor shell.[53] In Lennard-Jones fluids, about 80% of the solvation free energy comes from the first solvation shell.[46] Although the various terms contributing to the solvation free energy (i.e., the solute-solvent interaction, the solute-solvent entropy, and the solvent reorganization energy and entropy) may have different distance dependencies, in the present model we use a single functional form for f_i. The form is chosen such that approximately 80–90% of the solvation free energy comes from the first solvation shell; as shown below, 84% of the solvation free energy comes from the first solvation shell for all groups except the pseudo-ionic groups.

If λ is set equal to the width of the first solvation shell (3.5Å), simple exponential forms [$f \sim \exp(-r/\lambda)$ or $f4\pi r^2 \sim \exp(-r/\lambda)$] were found to decay too slowly (26 and 63% of the solvation free energy comes from the first solvation shell, respectively). The Gaussian form (Eq. 6) gave a contribution from the first solvation shell closer to the desired 80–90%. The Gaussian function has been used previously (without justification) to characterize the solvent exposure of protein atoms by Nauchitel and Somorjai[54] and Stouten et al.[35]

Inserting Equation 6 to Equation 2, we obtain

$$\Delta G_i^{slv} = \int \alpha_i \exp(-x_i^2)\, dr = \alpha_i\lambda_i \int \exp(-x_i^2)\, dx. \quad (14)$$

This integral is related to the error function

$$\mathrm{erf}(y) = \frac{2}{\sqrt{\pi}} \int_0^y \exp(-x^2)\, dx \quad (15)$$

which for $y = \infty$ is equal to 1 and for $y = 1$ is equal to 0.8427. Therefore, this function gives about 84% of the solvation free energy from the region $0 < x < 1$ (the region within $R_i + \lambda_i$ from the solute, which, for $\lambda_i = 3.5$Å, is the first solvation shell). Thus, the solvation forces in the Gaussian model are relatively short ranged, although less so than in surface area models. The solvation effects associated with the pseudo-ionic groups (see below), for which λ_i was chosen to be 6 Å, are somewhat longer ranged.

Determination of the α coefficients

If f_i is the solvation free energy density of group i, integration of f_i over all space gives the solvation free energy of free group i; that is,

$$\Delta G_i^{free} = \int_0^\infty f_i 4\pi r^2\, dr. \quad (16)$$

From the form of this integral, it is evident that the solvation free energy density is assumed to be isotropically distributed around each group. This is probably a good approximation for nonpolar groups, but it may not be entirely appropriate for polar groups (e.g., those involved in hydrogen bonding), which have more directional interactions with the solvent. Any inaccuracies are restricted to partly exposed groups, because in the limit of a fully buried group the results are correct (see below).

From Equations 6 and 16 and

$$\int_0^\infty \exp(-x_i^2)\, dx = \sqrt{\pi}/2 \quad (17)$$

we obtain

$$\alpha_i = 2\Delta G_i^{free}/\sqrt{\pi}\lambda_i. \quad (18)$$

It is important to note that ΔG_i^{free} is not exactly equal to ΔG_i^{ref}, the solvation free energy of i in the model compound. This is true because other groups are bonded to group i in the model compound and exclude solvent from its vicinity. For this reason, we expect ΔG_i^{free}, which corresponds to the isolated atom in solution, to be larger in magnitude than ΔG_i^{ref}. In the present model, we determined ΔG_i^{free} by requiring that the solvation free energy of groups deeply buried inside a protein be approximately zero. For each group, the magnitude of ΔG_i^{free} was incremented (starting from ΔG_i^{ref}) until the solvation free energy of some groups of that type in the protein CI2 became approximately zero (it is not possible to make all fully buried groups have exactly zero solvation free energy because even fully buried groups differ somewhat in the number of atoms surrounding them). This procedure led to the ΔG_i^{free} values listed in Table I. They are all larger in magnitude than

ΔG_i^{ref}. This procedure was not followed for the pseudo-ionic groups because such groups are rarely buried deeply in the protein interior. For these groups, ΔG_i^{free} was set equal to ΔG_i^{ref}.

Calculation of atomic volumes

To determine the atomic volumes V_i for Equation 5, we use the atomic radii of the CHARMM 19 parameter set. These radii, R_i, correspond to half of the distance where the Lennard-Jones interaction for two i particles has a minimum. At first sight, it might seem that the appropriate volume to be used in Equation 5 is the volume around an atom from which solvent is excluded; i.e., the volume determined with a radius equal to the sum of the atomic radius plus the solvent radius, analogous to the solvent-accessible surface. However, this neglects packing effects, because of which the region just outside this excluded volume (corresponding to the first peak of the radial distribution function) has a solvent density that is higher than in the bulk. This buildup of density compensates the solvent exclusion effect to some extent. As a result, the partial molar volume of an atom is much smaller than its excluded volume. For example, the partial molar volume of methane in water is about 60 Å^3 but its excluded volume is about 160 Å^3. Use of Ri gives smaller effective excluded volumes. For methane, this choice gives an effective excluded volume of about 38 Å^3. We could have used somewhat larger radii to make the effective excluded volume equal to the partial molar volume, but we decided in EEF1 to limit the number of model parameters and use the standard CHARMM 19 radii.

Because the volumes occupied by covalently bonded atoms overlap significantly, we take these overlaps into account in calculating the solvent exclusion effect. The overlap between the van der Waals volumes of two atoms covalently bonded to each other was calculated by Equation14 of Kang et al.[29] Each of the two partners was assigned half of the overlap volume. Thus, the volume of the atom of type i was calculated as

$$V_i = \frac{4}{3}\pi R_i^3 - \sum_j \frac{1}{2} V_{ij}^{(2)} \qquad (19)$$

where j goes over all atoms covalently bonded to i and $V_{ij}^{(2)}$ is the overlap volume between atoms i and j. When the same atom type is found in different covalent arrangements (e.g., atom type C, which is used for both the carbonyl and carboxyl carbon), the most common of these was used for the volume determination. Standard bond lengths and angles were used and their variation during a dynamics simulation was neglected. Triple and higher overlap volumes (i.e., overlaps between three or more atoms) were neglected because the correction caused by them is small with respect to the general level of approximations in the model. The values of V_i are given in Table I. Hydrogen atom volumes were neglected because they are small and almost entirely overlap with the volumes of the heavy atoms to which they are bonded.

Treatment of long-range van der Waals interactions

The computational cost of calculating all nonbonded interactions in a system scales as the square of the number of atoms and becomes expensive for large molecules. It is common practice to use a cutoff beyond which these interactions are neglected. In the present model, we use a cutoff of 9 Å with a switching function between 7 and 9 Å.[5] The experimental solvation free energy and enthalpy data include contributions from long-range interactions, which correspond to an integral over the entire volume of the system. To maintain the balance between intramolecular and solute-solvent interactions, we adjusted the experimental data to remove approximately the long-range contributions from the latter. Assuming that the dominant contribution to the solvation free energy at long range is the solute-solvent interaction, we estimate the long-range van der Waals contribution by the integral of the Lennard-Jones potential from 9Å to infinity:[55]

$$E_{sw}^{lr} = \rho \int_9^\infty g 4\epsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right] 4\pi r^2 \, dr. \qquad (20)$$

Assuming g = 1 and using the density of water (0.033 Å^{-3}) and typical values of Lennard-Jones parameters (σ= 3.3 Å, ϵ= 0.2 kcal/mol), the Lennard-Jones interaction beyond 9Å is calculated to be about −0.2 kcal/mol. We subtract this value from all experimental ΔG^{ref} and ΔH^{ref} values.[47,48]

Treatment of electrostatic interactions and ionic side-chains

The solvent has two types of effects on the charged and polar groups. The first is the self-energy contribution for each charge and the second is the screening of the interactions between charges. The self-energy contribution arises from the transfer of a charge from a high to a low dielectric environment. The screening of interactions arises from the perturbation of water structure by charged groups. Solvent screening reduces dramatically the interactions between charged groups in water. Site-directed mutagenesis studies [56,57] as well as theoretical studies[10] have shown that surface salt bridges usually make only small contributions to stability (on the order of 1 kcal/mol or less), in contrast to the very large interaction obtained from the unscreened coulomb term. Also, theoretical potentials of mean force of oppositely charged groups in water show little attraction (1–2 kcal/mol)[58] and in MD simulations of solvated proteins, surface water salt bridges often break (e.g., refs. 13 and 59).

The self-energy contribution corresponds to the loss of charge-solvent interactions and is well described by the solvent exclusion model (Eq. 5). Solvent screening is not given directly by the solvent exclusion model, but it can be approximated by a combined use of a variable dielectric parameter and reduced charges for the ionic groups. In the present model, we use a linear distance-dependent dielectric constant (RDIE, ϵ = r).[60,61] This choice has the important effect that it leaves the short-range hydrogen bonding interactions almost unaltered while screening

TABLE II. Partial Charges for the Pseudo-ionic Sidechains[†]

ARG		LYS		ASP (GLU)	
CD	0.10	CE	0.0	CB (CG)	−0.15
NE	−0.40	NZ	−1.35	CG (CD)	1.35
HE	0.15	HZ1, 2, 3	0.45	OD1, 2 (OE1, 2)	−0.60
CZ	0.25				
NH1, 2	−0.85				
HH11, 12, 21, 22	0.40				
NTER		CTER			
N	−1.35	C	1.20		
HT1, 2, 3	0.45	OT1, 2	−0.60		

[†]The neutral form of His is used in all calculations reported herein.

groups at larger distances. For example, the backbone CO-NHCα electrostatic interactions in RNAse A were calculated to be about 20% stronger on average with RDIE than with $\epsilon = 1$, although one might have expected diminution of the interactions by about a factor of 3, the approximate value of r ($\epsilon = r$) for these interactions. The reason that such diminution does not occur is that when dipoles are aligned head to head or head to tail, as in hydrogen bonding configurations, the unlike and like partial charges are at different distances and are screened to a different extent. Thus, for short-range attractive interactions, RDIE is approximately equivalent to using $\epsilon = 1$. In contrast, medium- and long-range electrostatic interactions are reduced significantly by the use of RDIE.

Even with RDIE, the electrostatic interaction of charged groups with each other and with polar groups is considerably stronger than the shielded interactions observed in solution, as described above. To obtain an approximately correct free energy of interaction, neutralized models for the ionic side-chains were used. The partial charges on ionic side-chains were chosen to make them neutral, but, in contrast to previous studies,[9] the polarity was increased significantly to obtain the correct stabilizing interaction for the salt bridges. In the actual model, the solvation free energy of the pseudo-ionic groups was arbitrarily set to −20 kcal/mol, a value large enough to prevent the burial of these groups in the protein interior. The charge distributions were adjusted to make salt bridges on the protein surface marginally stable. Test simulations were performed on an analog of the S-peptide of RNAse A,[62] which provides a good sample of ionic and ion-polar interactions. The objective was to observe reasonable stability (breakage and reformation) of the salt bridges of the S peptide in the molecular dynamics simulations. The final partial charges used for the ionic side-chains are given in Table II. In addition, the correlation length λ (see Equations 6 and 7) was set to 6 Å (rather than 3.5 Å), to account for the larger contribution from long-range solvent interactions for charged groups.

The treatment of ionic side-chains, although empirical, was found to give reasonable results. The model used leads to a small attraction between oppositely charged side-chains, whereas like charged side-chains repel each other via dipole–dipole interactions. The effective interaction energies are modest in magnitude, as one would expect for potentials of mean force in water. For example, the effective energy of forming the Lys8-Glu12 salt bridge in the S peptide analog is about −4 kcal/mol, whereas the same energy calculated with the vacuum energy function with charged side-chains and $\epsilon = 1$ is about −94 kcal/mol.

Implementation in the CHARMM Program

The solvation model has been developed for use with a slightly modified version of the CHARMM 19 polar hydrogen potential energy function.[26] It could, in principle, be used with other empirical potential functions (AMBER, GROMOS, etc.), but some of the parameters might have to be altered. The modifications introduced into CHARMM 19 are: (a) the atom type CR is added to represent a carbon atom bonded to no hydrogens; it is used for CG of Tyr, Phe and His, CG, CD2 and CE2 for Trp, and the CZ of arginine. CR has the same properties as atom type C (carbonyl carbon), except its solvation free energy (Table I); and (b) the charge distributions of Arg, Lys, Asp, Glu, and the N and C termini were modified to render them electrically neutral but polar in accord with the discussion above (Table II).

The solvation model has been introduced into the CHARMM program as a set of separate routines (about 400 lines of code). Atoms one or two bonds apart are excluded in the calculation of each other's solvation free energy; i.e., they are not included in Equation 21 below, because such covalently bonded atoms are present in the model compounds used to derive the reference free energies. This is consistent with the nonbonded electrostatic and van der Waals energy exclusions in the CHARMM energy function[5] and allows the use of the nonbonded and exclusion lists of CHARMM for the solvation free energy calculation.

The solvation free energy can be decomposed into a sum of pairwise interactions. It has the form (see Eqs. 3 and 5)

$$\begin{aligned}\Delta G^{slv} &= \sum_i \Delta G_i^{slv} = \sum_i \Delta G_i^{ref} - \sum_i \sum_{j\neq i} f_i(r_{ij})V_j \\ &= \sum_i \Delta G_i^{ref} - \sum_{j>i} (f_i(r_{ij})V_j + f_j(r_{ij})V_i) \\ &= \sum_i \Delta G_i^{ref} - \sum_{j>i} \left\{ \frac{2\Delta G_i^{free}}{4\pi\sqrt{\pi}\lambda_i r_{ij}^2} \exp(-x_{ij}^2)V_j \right. \\ &\quad \left. + \frac{2\Delta G_j^{free}}{4\pi\sqrt{\pi}\lambda_j r_{ij}^2} \exp(-x_{ji}^2)V_i \right\}. \end{aligned} \quad (21)$$

Analytical derivatives of the solvation free energy are readily obtained. For example,

$$\begin{aligned}\frac{\partial \Delta G^{slv}}{\partial z_i} = \sum_{j\neq i} \left(\frac{\Delta G_i^{free}}{\pi\sqrt{\pi}\lambda_i r_{ij}^3} \exp(-x_{ij}^2) \left[\frac{x_{ij}}{\lambda_i} + \frac{1}{r_{ij}} \right] V_j \right. \\ \left. + \frac{\Delta G_j^{free}}{\pi\sqrt{\pi}\lambda_j r_{ij}^3} \exp(-x_{ji}^2) \left[\frac{x_{ji}}{\lambda_j} + \frac{1}{r_{ij}} \right] V_i \right) * (z_i - z_j). \end{aligned} \quad (22)$$

They are included in the CHARMM program so that minimization and dynamics can be done as with the vacuum energy function. Even though the solvation term is recalculated every step, the CPU time required for simulations with EEF1 is only about 50% larger than that for a vacuum simulation with the same cutoff (9Å). Thus, it is much more efficient than surface-area-based solvation models. The implementation of the Wesson and Eisenberg[25] solvation model in CHARMM takes about 35 times longer than the corresponding vacuum simulation. For the same model implemented into AMBER, a factor of 8 has been reported.[27]

On the Development of the Model

The model was developed by use of an iterative procedure. We describe briefly what was done to aid others in developing corresponding models. A given version of the model was tested, problems were encountered, and the physical basis of the problems was identified. This understanding was used to devise a physically justifiable solution, which led to a revised version. Below we discuss what variations were attempted and which features of the earlier versions of the model proved unsatisfactory.

The solvation parameters of Privalov and Makhatadze[47] were used as published, except for the subtraction of the long-range van der Waals component and the introduction of specific solvation parameters for the pseudo-ionic groups. The Gaussian form of f_i was chosen from the beginning (Equation 6) and was not changed. For λ_i, only the values 3 and 3.5 Å were tried with similar results. The physically reasonable larger value 6Å for the ionic groups was found to give smaller deviations from the crystal structure in some native protein simulations.

For the ionic side-chains, two other treatments were tried: ionic side-chains with full charges and the solvation free energies pertaining to such groups (of the order of −80 kcal/mol) and the neutral side-chains of Lazaridis et al.[9] with the solvation free energies given by Privalov and Makhatadze. Use of fully charged side-chains did not work because it led to the existence of nonnative, semicompact states with energies lower than that of the native state. The reason is that electrostatic interactions between charged groups and their solvation free energies are so large that even small errors in the balance of intramolecular and solvation terms can create deep artificial minima on the effective energy surface. With the neutral side-chains of low polarity,[9] the main problem was that the side-chains would readily become buried in the protein interior. Also, salt bridges were not as stable as they should be. For the pseudo-ionic groups, solvation free energy values other than −20 kcal/mol were not tested; different values in the same range might work as well.

Initially, a constant dielectric of 1 was tried. This resulted in semicompact conformations with lower energies than the native conformation. In such conformations, the polar groups would have almost their full solvation free energy (i.e., they were essentially exposed), while at the same time they had large stabilizing medium-range electrostatic interactions with other polar groups of the protein.

Comments on the Model

The obvious advantage of the model over explicit solvent calculations is an approximately two orders of magnitude increase in the speed of the calculations. Its obvious disadvantage is that it is a cruder approximation than that obtained with explicit solvent models. Compared with the usual surface area models, the excluded volume model has many advantages. Not only is it simpler and computationally less expensive, but in many cases it is physically more correct. For example, for slightly exposed polar groups, such as the peptide groups in α-helices, the surface area model gives a solvation free energy that is almost zero. This is incorrect, because polar groups near the surface can interact significantly with the solvent even if their calculated accessible surface area is zero.[9] The solvent exclusion model gives a substantial residual solvation free energy for the peptide groups in helices. Surface area models or the hydration shell model of Kang et al.[29] assume that all solvation free energy comes from the first hydration shell. In the present model, such an assumption is not necessary. Depending on the type of group, the solvation free energy density f_i can be such that small but significant contributions may come from the second or higher hydration shells;[52] in the present implementation, 16% of the solvation free energy comes from beyond the first solvation shell. In this respect, the solvent exclusion model, despite its simplicity, is physically superior to surface area models.

Several limitations of excluded volume-type models have been pointed out by Stouten et al.[35] One basic assumption is that "nonprotein space" is "solvent space".[30] Thus, the model does not take into account that water has a finite size and cannot fit into any cavity, or that during dynamics, even if a cavity is large enough, water may not have had enough time to diffuse into it. However, this is not likely to be a serious problem. First, cavities too small for solvent have been eliminated, in part, in an average sense by the use of slightly larger atomic radii. Second, larger cavities are not likely to be empty in proteins, even if they appear to be hydrophobic.[63–65] Also, in unfolding simulations it was observed that whenever a cavity opened up, water rapidly moved in to occupy the empty space.[13,66] Quantitatively, the error is not large if cavities are surrounded by nonpolar groups (which is most likely) because the solvation free energy of such groups is small in magnitude.

A related problem is the assumption that solvation by "confined" water is thermodynamically equivalent to solvation by bulk water. This could be of importance in semicompact protein conformations. There one might have instances where a few water molecules are surrounded by protein groups, polar or nonpolar. It is likely that such water molecules do not solvate the protein groups in the same way as water on the protein surface, for example. This is essentially a water structure perturbation effect; i.e., the absence of bulk water may affect the way the water

molecules orient relative to the protein groups. Waters in the interior of a protein (e.g., the four fully buried waters in BPTI) and tightly bound surface waters could be introduced explicitly in native state simulations with some adjustments of the implicit solvation model.

A fundamental assumption of the solvation model is that the solvation free energy can be decomposed into a sum of group contributions. This assumption has been shown to work reasonably well for small molecules, in which all of the atoms are exposed to solvent.[47,48] Here we extend the additivity assumption to large peptides and proteins by introducing group contributions that depend on solvent exposure. This is analogous to the widely used surface-area-based models for solvation. Although the decomposability of the free energy has been questioned recently,[67,68] it has been shown that meaningful decompositions are possible and, if made carefully, can be very useful.[69–72] The inhomogeneous approach to solvation thermodynamics[45] is the basis of the present model. It can be used to define group contributions (see above). In such a theory, the contributions are context dependent and thus not completely transferable from one molecule to the other; that is, decomposability does not necessarily imply transferability. For example, a CO group next to an NH group may have a different group solvation free energy than a CO group next to a CH3 group. Such context dependencies, in the form of interference between neighboring polar groups, have been pointed out previously.[9,73–75] In the present model, we introduce the group additivity assumption as a first approximation. Improvements will come from a more detailed treatment of how the solvation of one group is affected by its neighbors. Introduction of such effects would complicate the model, but is likely to make it more accurate.

Difficulties with the surface area model, when applied to polar groups, have been pointed out previously.[9] Some of these criticisms apply to the present model. For example, polar group–water interactions tend to be highly directional, so that the isotropic distribution of the free energy density assumed in Equation 16 may not be realistic. The linear relationship between solvation and exposure implied by the present model holds only in an average sense (see Figs. 1 and 2 of ref. 9); i.e., groups with similar solvent accessibility can have widely different interactions with the solvent but increased solvent accessibility is, on average, associated with lower solvation free energy. Solvent packing effects (the oscillations in the pairwise potentials of mean force[76]) are not accounted for by the model. Consequently, the model does not reproduce accurately, for example, potentials of mean force between small molecules in water, or conformational equilibria of small molecules, although the solvation free energies are correct, essentially by construction. The primary purpose of the model is to capture the global balance between intramolecular and protein-solvent interactions that underlies protein stability.

Some of these problems are illustrated by applying the present model to the ϕ,ψ conformational map (Ramachandran plot) of the alanine dipeptide. The excluded volume model presented here, as well as any accessible surface-area-based model (e.g., ref. 27) causes only minimal changes in the vacuum energy map. In the present model, the solvation free energy varies by no more than 1 kcal/mol throughout the alanine dipeptide ϕ,ψ map. This is because the solvent exposure of the polar groups is nearly invariant for all conformations of the dipeptide. Explicit water simulations,[19,20,77] integral equation theory,[78] and continuum electrostatics calculations[37,79,80] predict much larger variation in the solvation free energy with ϕ and ψ. For example, continuum electrostatics calculations give (electrostatic) solvation free energy -13.85 kcal/mol for $C7_{eq}$, and -21.09 kcal for α_R.[37,79] The value of the solvation free energy in the present model is about -16.5 kcal/mol throughout the ϕ-ψ map. This shows that the model is correct on average, but does not capture the local variations caused by polar group interference.

These deficiencies may have quantitative consequences in certain applications. First, because the ϕ-ψ map of the dipeptides is more structured than in reality, the entropy of the unfolded state could be underestimated with respect to local variations of ϕ and ψ; this could also affect the local dynamics of the unfolded state. For the folded protein, where solvent effects are much smaller, this does not cause any problems. Also, tertiary interactions should not be affected because they involve primarily solvent exclusion. Global dynamics should not be substantially affected, to the extent that they are decoupled from local dynamics. Second, the energetic cost of dipole alignment, when these dipoles are exposed, will be overestimated. For example, in the case of helix formation, the energy barrier for helix initiation will be overestimated, because it involves alignment of dipoles. Therefore, the model is not expected to be very accurate for modeling small peptide conformations or loops in proteins. Tertiary structure formation, on the other hand, which involves large changes in the solvent exposure of many protein groups, should be captured correctly. To avoid such problems, a term could be added to describe the variation of the solvation free energy of the backbone with local conformation. One possibility, which also leads to additive two-body terms, is a modified version of the superposition approximation for site–site correlation functions[78] (work in progress).

TESTS AND APPLICATIONS OF EEF1

We first test the ability of EEF1 to yield stable structures for native proteins by performing molecular dynamics simulations at room temperature on several proteins and examining the deviations from the crystal structure. We compare the performance of EEF1 with the results from the same CHARMM energy function without a solvation correction. We also constructed model unfolded conformations and confirmed that their effective energies and enthalpies are consistent with the observed folding thermodynamics. In separate articles, it is shown that EEF1 discriminates native from misfolded conformations[44] and gives unfolding pathways at high temperatures, in agreement with explicit water simulations.[43]

TABLE III. Deviations From Crystal Structure, Radius of Gyration After 50-ps MD Simulation[abc]

		RMSD, EFF1		RMSD, VAC		ϕ-ψ RMSD		R_g		
	Nres	Back	All atoms	Back	All atoms	EFF1	VAC	Exp	MD	MD, VAC
Melittin	26	1.55	2.15	4.44	5.71	16	52	11.35	11.28	10.89
Melittin	2 × 26	1.91	2.20	5.51	6.59	17	56	13.24	13.80	12.21
Melittin	4 × 26	3.04	3.53	3.63	3.93	26	52	13.98	15.05	13.38
Crambin	46	1.13	1.45	2.04	2.76	29	40	9.63	10.07	9.14
Arc repressor	2 × 53	1.32	2.01	2.39	2.89	40	56	14.32	14.44	14.03
Rubredoxin	54	2.08	2.52	3.54	4.31	23	63	10.12	10.39	9.82
G protein B1	56	1.86	2.44	1.85	2.40	30	38	10.59	10.70	10.45
BPTI	58	2.20	2.85	3.41	4.20	39	55	11.04	11.20	10.17
CI2	64	1.05	2.01	1.71	2.47	29	38	11.27	11.32	10.88
CTF	68	1.32	1.82	3.00	3.66	29	49	11.21	11.37	10.55
Eglin c	70	1.83	2.59	3.10	3.98	46	59	11.18	11.10	10.74
Ubiquitin	76	1.35	1.95	2.60	3.18	32	51	11.76	11.75	11.32
Cytochrome c	103	1.95	2.29	3.53	4.18	32	49	12.54	12.75	12.41
Barnase	110	2.46	2.83	3.12	3.81	101	98	13.60	14.03	13.13
RNAse A	124	2.40	2.91	3.72	4.02	31	49	14.35	14.72	13.07
Lysozyme	129	1.91	2.44	2.44	3.00	39	51	13.99	14.03	13.37
Staph nuclease	149	1.58	2.39	3.06	3.58	43	49	14.50	14.48	14.14
Myoglobin	153	2.08	2.55	2.83	3.46	89	91	15.18	15.73	15.11
Interleukin 1b	153	1.96	2.55	2.62	3.40	41	49	15.07	14.98	14.16
Chymotrypsin	245	2.96	3.40	3.91	4.63	40	49	16.65	16.96	15.93
Chymotrypsin	2 × 245	3.59	3.60	3.80	4.11	39	48	22.41	23.59	21.78
Average		1.98	2.50	3.15	3.82	39	54			

[a]PDB codes: melittin: 2mlt, crambin: 1crn, eglin c: 1egl (NMR, first model used), myoglobin: 1mbc, RNAse A: 8rat, ubiquitin: 1ubq, chymotrypsin inhibitor 2: 2ci2, G protein B1 domain: 2gb1 (NMR), lysozyme: 2lyz, chymotrypsin: 5cha (first monomer used), cytochrome c: 5cyt, rubredoxin: 5rxn, interleukin: 6i1b (NMR), bpti: 5pti, barnase: 1rnb, staphylococcal nuclease: 1stn, arc repressor: 1arr.

[b]The RMSD and R_g values for eglin c do not include the disordered N terminal heptapeptide. The RMSD for chymotrypsin does not include the disordered segment 9–13. For barnase, the N-terminal tetrapeptide is not included. For staph nuclease, residues 1–5 and 142–149 (unresolved in crystal) were built in an extended conformation and are not included in Rg and RMSD calculations. For the arc repressor, residues 1–7 and 47–53 are not included in RMSD calculation.

[c]For the simulations all water (including internal waters), counterions, or substrates were deleted (except for the heme in cyt c and myoglobin and the CO in myoglobin). HBUILD is used to construct the missing hydrogens; 300 steps of ABNR minimization are done. The cutoff is 9Å, based on groups, switched off from 7Å. Timestep 2fs.

Simulations of Native Proteins

An essential requirement for a realistic effective energy function is that native protein structures be stable during molecular dynamics simulations at room temperature and do not exhibit large deviations from the experimental (x-ray or NMR) structure. The proteins studied are listed in Table III. They were chosen to cover a broad range of sizes and architectures; some were chosen because of the availability of simulations with an explicit water model for comparison. All structures were first energy minimized (300 adapted basis Newton-Raphson, ABNR, steps) with EEF1 and then subjected to a 50-ps molecular dynamics simulation at 300 K. This simulation length is deemed sufficient for the purposes of this study; longer simulations show that most of the root mean square deviation (RMSD) occurs within the first 50 ps (see below). For comparison, equal-length simulations were performed with the vacuum CHARMM 19 polar hydrogen energy function ($\epsilon = 1$). Comparison is made to the vacuum force field with $\epsilon = 1$ because, first, it is a widely accepted standard in the field, and second, it corresponds to an extreme. The use of $\epsilon = r$ or the neutralization of side-chains alone would improve the results compared with vacuum simulations with $\epsilon = 1$, but the best results are obtained by including the additional solvation term.

Table III shows the RMSD from the crystal structure and the radius of gyration (Rg) with EEF1 and with CHARMM19 alone calculated from the last coordinate set. A well-known artifact of vacuum simulations is an overall contraction of the molecules.[81,82] This is clearly evident in the vacuum results in Table III. For all proteins, the radius of gyration after the vacuum simulations is smaller than in the experimental structure. Such contraction does not occur with the EEF1 force field. On the contrary, in most cases Rg increases slightly during the MD simulation. This effect may be realistic and result from the absence of crystal contacts. Small increases in Rg have been observed in explicit water simulations.[66,83,84] Reproduction of the correct radius of gyration in MD simulations was used as one quality criterion in a recent implicit solvation model described in the Introduction.[28]

The backbone RMSDs from the EEF1 simulations are in general between 1 and 2 Å; for the larger proteins, the RMSDs tend to be greater. The observed range of RMSD is similar to the range obtained in explicit water simulations. In comparing the EEF1 simulations to explicit water simulations, it is important to note that the effective time span of implicit solvent simulations is longer than the actual duration of the simulation because of the absence of solvent caging and friction forces in EEF1.[43] This means

that for equal simulation time, explicit water simulations are expected to stay closer to the experimental structure. The solvent frictional forces could be included by performing Langevin dynamics,[14] but in many applications such as conformational searching, the speedup of the dynamics is actually desirable.

Among the proteins in Table III, the most extensively simulated is BPTI. van Gunsteren and Berendsen[85] reported a 20-ps simulation with a final C_α RMSD of 1.5 Å; Levitt and Sharon[82] and Levitt[86] reported 210-ps and 180-ps simulations, respectively, with final C_α RMSD of about 1.4 Å; Daggett and Levitt[87] reported a final C_α RMSD of 1.95 Å after 550 ps. Stouten et al.[35] obtained 2.13 Å with the GROMOS force field after 200 ps. Finally, Brunne et al.[88] reported 1.7 Å RMSD after 1.4 ns. The backbone RMSD obtained with the present energy function is slightly larger, 2.2Å.

In many of the proteins in Table III, the RMSD obtained is similar to or smaller than that obtained in explicit water simulations. For example, a simulation of lysozyme for 550 ps[84] gave 2.2 Å, larger than the one in Table III. Another simulation of the same protein (Buck and Karplus, to be published) gave backbone RMSD less than 1Å after 1.6 ns. For ubiquitin, Alonso and Daggett[89] report 1.5Å after 500 ps, similar to the value 1.35 Å in Table III. For CTF, Daggett and Levitt[90] obtained 1.24Å C_α RMSD after 200 ps, similar to the value 1.32 Å obtained here. In other cases, the deviations observed in the present simulations were somewhat larger. For example, a simulation of barnase[66] gave 1.5 Å backbone RMSD after 250 ps. The value obtained in Table III (among the highest) is significantly higher, 2.46 Å for the backbone. For interleukin 1β, Chandrasekhar et al.[91] obtained an all-atom RMSD of 1.7Å after 450 ps, also smaller than the 2.55 Å value obtained here.

To our knowledge, no simulation of the entire chymotrypsin molecule in explicit water has been reported. Simulation of trypsin in explicit solvent for 19 ps with periodic boundary conditions yielded an all-atom RMSD of 2.18 Å for the average structure.[92] Stochastic boundary simulations of the active site of chymotrypsin based on crystal structures different from the one used here have been performed.[93,94] One of them[94] reported considerable shifts (1.35 Å all-atom RMSD) in the active site structure, even in the presence of explicit solvation. The RMSD values with EEF1 and the vacuum potential are significantly larger than for the other systems, probably because of the size of the molecule. Because chymotrypsin forms a dimer in the crystal, we examined the possibility that these deviations are caused by the lack of contacts with the other subunit by performing molecular dynamics (MD) on the full dimer (482 residues in all). The RMS deviations observed were similar to those obtained for the monomer. A simulation of the acetylcholinesterase dimer (10,438 protein atoms) in explicit solvent showed that the all-atom RMSD for one subunit reached 3 Å and did not level off after 500 ps of molecular dynamics.[95]

To investigate the effect of longer simulations, CI2 (among the lowest RMSD) and barnase (among the highest RMSD) were studied for 200 ps. Figure 1 shows the backbone RMSD as a function of time for these two cases. The RMSD clearly stabilized during the 200-ps run. The final values are 1.46 Å (2.33 Å all atom) for CI2 and 2.91 Å (3.54 Å all atom) for barnase. The corresponding values in vacuum are 1.88 Å (2.69 Å all atom) for CI2 and 3.94 Å (4.54 Å all atom) for barnase. Longer simulation times are sometimes required for stabilization of the RMSD in explicit solvent simulations;[96] the barnase simulation seems to have stabilized by 250 ps.[13]

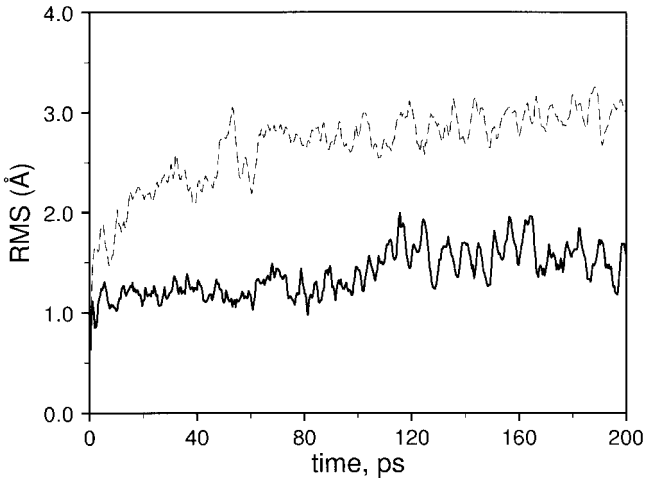

Fig. 1. Backbone RMSDs from the crystal structure for CI2 (solid line) and barnase (dashed line); see text.

From the results in Table III, we can conclude that native proteins are stable with the EEF1 force field and exhibit deviations from the experimental structure comparable to simulations in explicit water. The simulations in vacuum exhibit larger deviations from the experimental structures in all cases except GB1, where similar deviations are observed. The same conclusion follows from the RMSD of values for the backbone ϕ and ψ angles that are also listed in Table III.

Table IV lists the number of hydrogen bonds in the experimental structure and at the end of the vacuum and EEF1 simulations. The number of hydrogen bonds increases substantially (from 40% to 250%) in the vacuum simulations. A more modest increase is observed for the simulations with EEF1. In all cases, the number of hydrogen bonds with EEF1 is between that in the experimental structures and in the vacuum simulations. Most of the hydrogen bonds formed in the simulations with EEF1 involve the external polar side-chains. It is possible that the strength of interaction between these side-chains is overestimated in the present model. Alternatively, this may be a solvent effect that involves specific water molecule interactions and so is not reproduced by the model. The analysis of hydrogen bonding shows that secondary structural elements are stable during the simulations.

Table IV also shows the quenched energies (energies after 300 steps of ABNR minimization) before (i.e., minimization of the crystal structure) and after 50 ps of molecular

TABLE IV. Number of Hydrogen Bonds and Energies After 50-ps MD Simulation

	Hbonds			Quenched energy (kcal/mol)[a]	
	Exp	EFF1	VAC	Exp	After MD
Melittin 1	25	26	40	−721 (148, 406, 229)	−728 (135, 413, 238)
Melittin 2	50	52	71	−1,464 (317, 824, 453)	−1,491 (305, 865, 453)
Melittin 4	98	114	144	−2,983 (726, 1,637, 891)	−3,075 (698, 1,786, 869)
Crambin	36	40	45	−1,187 (254, 708, 339)	−1,212 (245, 737, 346)
Arc repressor	97	122	166	−3,371 (805, 1,732, 1,142)	−3,591 (777, 1,953, 1,144)
Rubredoxin	32	38	55	−1,601 (366, 759, 620)	−1,674 (349, 864, 601)
G protein B1	38	52	72	−1,747 (399, 860, 599)	−1,820 (400, 941, 589)
BPTI	35	44	78	−1,659 (374, 812, 608)	−1,765 (360, 929, 605)
CI2	49	51	87	−1,905 (470, 986, 617)	−1,963 (459, 1,064, 607)
CTF	58	63	79	−2,023 (445, 1,004, 687)	−2,070 (420, 1,077, 689)
Eglin c	26[b]	67	87	−2,007 (476, 1,046, 674)	−2,148 (540, 1,186, 611)
Ubiquitin	63	82	100	−2,400 (547, 1,241, 778)	−2,480 (533, 1,361, 761)
Cytochrome c	92	106	125	−3,126 (839, 1,630, 1,015)	−3,260 (816, 1,807, 1,003)
Barnase	94	110	154	−3,380 (856, 1,756, 1,062)	−3,517 (811, 1,954, 1,044)
RNAse A	109	140	174	−3,824 (852, 2,045, 1,199)	−3,974 (847, 2,257, 1,167)
Lysozyme	111	139	175	−4,008 (932, 2,196, 1,217)	−4,128 (882, 2,343, 1,233)
Staph nuclease	130	159	217	−4,640 (1,076, 2,332, 1,588)	−4,847 (1,105, 2,614, 1,490)
Myoglobin	146	173	213	−4,680 (1,286, 2,394, 1,419)	−4,837 (1,210, 2,585, 1,438)
Interleukin 1b	84[b]	143	196	−4,534 (1,095, 2,246, 1,540)	−4,809 (1,111, 2,633, 1,445)
Chymotrypsin	172	216	276	−6,628 (1,673, 3,645, 1,943)	−6,919 (1,571, 3,997, 1,967)
Chymotrypsin 2	360	428	548	−13,315 (3,455, 7,318, 3,811)	−13,906 (3,265, 8,031, 3,849)

[a]For the energies the following notation is used throughout: the first number is the total effective energy in kcal/mol. In parentheses are given the van der Waals component, the electrostatic component, and the solvation free energy (the negative sign is omitted for brevity). The bonded energy terms can be obtained by subtracting the above three components from the total energy. Quenching is done by 300 ABNR steps. A hydrogen bond is considered present when the hydrogen-acceptor distance is less or equal to 2.5 Å and the X-H . . . Y and H . . . Y-Z angles are greater than 90°.
[b]These are NMR structures where hydrogen bonding is not as good as in crystal structures.

dynamics. This comparison shows the terms that contribute to making the final structure have a lower energy than the minimized crystal structure, given the energy function. Protein configurational entropy may also cause deviations from the crystal structure, but its role is difficult to quantify. The quenched energy after dynamics is always lower than that of the crystal. The major contribution to the energy difference comes from the electrostatic component, which invariably improves upon MD. The van der Waals term deteriorates in most cases, whereas the solvation term sometimes improves and sometimes deteriorates. In a room temperature simulation of barnase, the intraprotein energy increased during the simulation, whereas the protein–solvent interaction energy (one component of the solvation free energy) decreased.[13]

As an interesting additional example, simulations were performed on melittin, a small 26-residue peptide present in honeybee venom. In its monomeric form it is a random coil, but at high ionic strengths it forms tetramers, in which each monomer adopts a helical structure. The crystal structure of its tetrameric form has been solved[97] and the thermodynamics of tetramerization under a variety of solution conditions has been analyzed.[98] The peptide is highly amphiphilic and upon tetramerization the hydrophobic faces of the helices are buried. It is also amphiphilic along its length, having four consecutive positively charged residues near its C terminus. Molecular dynamics simulations of the melittin monomer and tetramer have been performed by Wesson and Eisenberg[25] as a test of their surface area based solvation model. The results of molecular dynamics simulations of the melittin monomer, dimer, and tetramer with EEF1 and the vacuum energy function are shown in Table III and Table IV. Because the monomeric and dimeric forms are not stable, larger RMS deviations might have been expected for these simulations than in the simulation of the tetramer. However, this was not found to be the case. The RMSD was smallest for the monomer, larger for the dimer, and largest for the tetramer. This is likely to be partly a size effect, corresponding to that observed in the chymotrypsin simulations. If individual helices are superposed, the backbone RMSD values are 1.55 Å for the monomer, 1.79 Å and 1.57Å for the dimer, and 2.05, 1.74, 0.98, and 2.1Å for the tetramer. The all-atom RMSD values obtained by Wesson and Eisenberg after 110 ps of molecular dynamics simulation are 6.93 Å for the monomer and 3.41 Å for the tetramer. This is to be compared with the present all-atom values of 2.15 Å for the monomer and 3.53 Å for the tetramer.

The quenched energies before and after dynamics are shown in Table IV. The quenched crystal energies show that the energy of the tetramer (−2,983 kcal/mol) is somewhat more negative than that of two isolated dimers (−1,464 × 2 = −2,928 kcal/mol). This yields an effective energy change for the association of 2 dimers into a tetramer:

$$\Delta W = -55(-90, +11, +15) \text{ kcal/mol}$$

where the numbers in parentheses are the van der Waals, electrostatic, and solvation contributions, in that order; the total includes the bonded terms. The van der Waals (essentially hydrophobic) interactions are highly favorable to tetramerization, whereas solvation and electrostatic contributions are unfavorable. This is consistent with the experimental fact that melittin is tetrameric only at high ionic strengths, whereas at low ionic strength it is monomeric.[98] The experimental free energy for the association of four monomers to a tetramer at 25°C is −15.5 kcal/mol at pH 7.4 in Tris/NaCl buffer.[98] The change in effective energy for the association of four monomers to a tetramer from Table IV is $\Delta W = -99$ kcal/mol. The quantity ΔW includes only the change in effective energy and not the change in translational and rotational entropy on binding. The latter are expected to be in the range +15–20 kcal/mol for dimer formation;[99,100] there may also be a change in side-chain conformational entropy, which can contribute up to +15 kcal/mol.[99] The total translational/rotational/conformational entropy change for the association of four monomers to a tetramer is estimated to be about 90 kcal/mol. Thus, the experimental value is of the order expected from the calculations, including the crude estimate of the entropic term. The unfavorable electrostatic interaction is caused by the repulsive interaction between Arg and Lys residues that come close in the tetramer. Although in the present model the ionic side-chains do not carry a net charge, the unfavorable interaction between these side-chains is reproduced in the minimized structure. These repulsions would be alleviated by the presence of counter ions; in fact, in the crystal structure, a sulfate ion is located near Arg 24, which was not included in the simulations.

Examining the molecular dynamics results for the melittin tetramer revealed a problem with the neutral-polar model for the Arg residue. In the simulation, Lys and Arg side-chains found configurations where they could interact favorably with each other. For a hydrogen bond donating group approaching perpendicular to the guanidinium plane, the guanidinium nitrogens of Arg can act as a hydrogen bond acceptor. This interaction is probably not realistic; a rigorous calculation of the potential of mean force in explicit solvent would be of interest here. Such like–like charge interactions are not observed for Glu and Asp side-chains, probably because the oxygens are bulky and for Lys because the three hydrogens have a tetrahedral arrangement and the N is not accessible. In Arg, the hydrogens form a plane so that a hydrogen bond donor can approach the nitrogens from below that plane. This problem could be alleviated by increasing the van der Waals radius of the nitrogens.

Dynamics

To determine the effect of the solvation potential on the dynamics, we consider the results for CI2 and barnase because they are the only simulations that are long enough (200 ps) to calculate meaningful fluctuations. The RMS fluctuations during the last 40 ps of the 200-ps simulation of CI2 in the vacuum and solvated simulations are given in

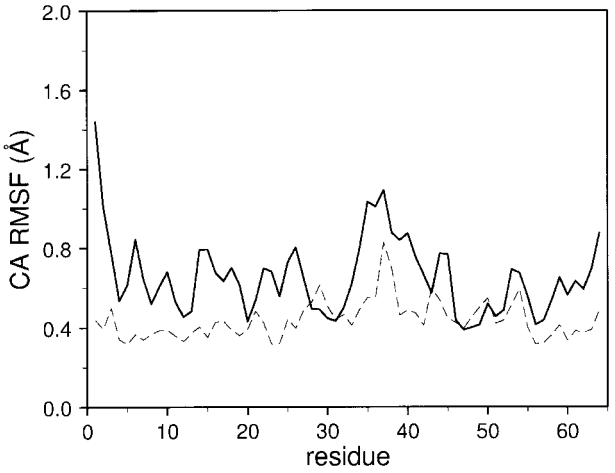

Fig. 2. RMS fluctuations for the Cα atoms of CI2. Vacuum simulation (dashed line) and implicit solvent simulation (solid line) are shown; see text.

Figure 2. The average RMSF of Cα atoms are 0.44 ± 0.1 in vacuum and 0.65 ± 0.2 for the solvated simulation. The corresponding numbers for all atoms are 0.6 ± 0.3 and 0.9 ± 0.5 Å, respectively. Thus, not only are the fluctuations on average larger in implicit solvent, but the range of the values increases somewhat. Such an increase in fluctuations on adding a solvation term was also found by Stouten et al.[35] In explicit water simulations, some increase in fluctuations was observed in a stochastic boundary simulation of the active site of lysozyme.[101] Steinbach and Brooks[102,103] found that at room temperature, hydration increases the anharmonicity and the mobility of surface groups in myoglobin. Levitt reported that fluctuations of BPTI decreased in explicit water simulation relative to vacuum.[82,86] Experimental results show that hydration increases protein mobility.[104] For example, increase in the hydration level of protein powder films increases the rate of hydrogen exchange and the mobility of bound ligands,[105] Mössbauer spectroscopy shows a characteristic enhancement of mobility above 210 K for wet but not dry myoglobin samples,[106] and dielectric relaxation measurements show enhancement of protein flexibility with hydration.[107] However, a direct comparison with the simulation results is not possible because many of the acid and basic groups are neutralized at very low hydration.

Enthalpy and Free Energy of Unfolding

To obtain the solvation enthalpy of a given configuration, the same methods are used with solvation enthalpy group contributions substituted for the solvation free energy contributions (see section on Description of the Model). The total enthalpy is then $H = H_{intra} + \Delta H^{slv}$, where H_{intra} is the intramolecular energy and ΔH^{slv} the solvation enthalpy. The enthalpy difference between the native and unfolded conformations can be compared directly with experimental data. In this section, we consider the protein CI2 for which the experimental enthalpy of unfolding at

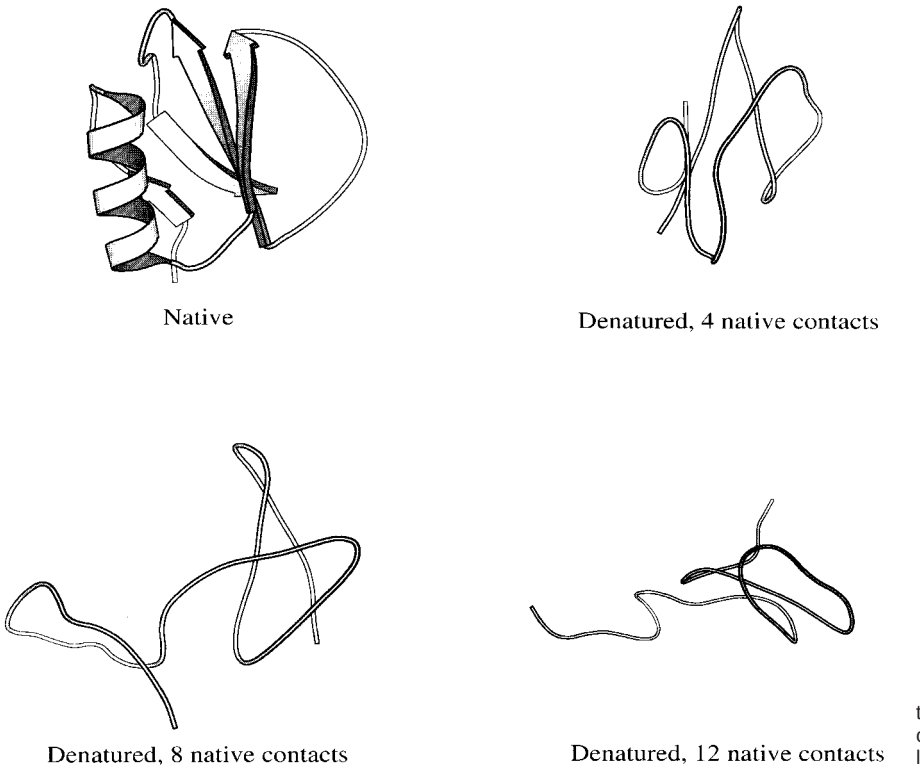

Fig. 3. The native and three denatured structures of CI2. The denatured structures were obtained from high-temperature unfolding simulations.[43]

room temperature is 30.5 kcal/mol and the free energy of unfolding is 7 kcal/mol.[108]

To obtain a model for the unfolded state, we use an ensemble of 400 unfolded conformations extracted from high-temperature unfolding trajectories,[43] simulated at 300 K for 10 ps, and then energy minimized for 300 steps with ABNR (see Fig. 3).[5] These conformations have less than 40% of native contacts; their Rg values range from 11.3 to 18 Å, in agreement with other studies.[109,110] The effective energy of these conformations is 85 ± 15 kcal/mol higher than that of the native state and the enthalpy is 45 ± 15 kcal/mol higher than that of the native state. Thus, the enthalpy value is of the same order but somewhat larger than that found experimentally.

To estimate the free energy of unfolding, we combine the effective energy change, $\Delta_N^U W$, which includes the solvation free energy, with the free energy contribution of the configurational entropy, $T\Delta_N^U S$ (see Appendix):

$$\Delta_N^U G = \Delta_N^U W - T\Delta_N^U S \tag{23}$$

Experimentally, $\Delta_N^U G$ is on the order of 5–15 kcal/mol for many proteins;[111] the value for CI2 is 7 kcal/mol.[108] Thus, the change in effective energy on unfolding must be of the same order of magnitude as the change in configurational entropy. As stated in the previous paragraph, $\Delta_N^U W$ is estimated to be 85 ± 15 kcal/mol for the 400 configurations used for the unfolding enthalpy evaluation. Thus, the configurational entropy of unfolding should be about 260 ± 50 cal/mol K, or 4 ± 0.8 cal/mol K per residue (CI2 has 64 residues). Empirical and theoretical estimates for the configurational entropy of unfolding range from 1 to 12 cal/mol K per residue.[109,112–114] The large estimate was obtained assuming that the denatured state is highly expanded and fully solvated and the low estimate assuming a compact denatured state. Thus, even though quantitative comparisons are not possible, the calculated difference in effective energy between folded and unfolded conformations is within the expected range.

It is of interest to compare the enthalpy value obtained from the ensemble of structures with those found when various simpler models of the unfolded state are used; the results are summarized in Table V. An ideal β-strand was constructed ($\phi = -140$, $\psi = 135$) and the energy was minimized for 300 steps. Clearly, the ideal β conformation has a much higher effective energy. This is because of the absence of intramolecular stabilization. The solvation term is more favorable for the unfolded conformation because the favorable solvation free energy of the exposed polar groups dominates the unfavorable solvation free energy of the exposed nonpolar groups.

The second conformation is that obtained from the ideal β chain after a few ps of MD simulation followed by energy minimization; this conformation is characterized by the values $\phi \approx -72$ and $\psi \approx +79$ and has a clear right-handed twist; the driving force for this right-hand twist is electrostatic. When we build a CI2 chain with these values of ϕ and ψ and minimize, we obtain an extended chain with lower energy than the ideal β chain, mainly because of electrostatic interactions between backbone CO-CO and

TABLE V. Energetics of CI2 Conformations

Structure[a]	W (vdW, elec, solv)	ΔH^{slv}	H
	Native		
N1	−1905 (−470, −986, −617)	−860	−2148
N2	−1963 (−459, −1064, −607)	−851	−2207
	Unfolded		
U1	−1617 (−176, −765, −848)	−1207	−1976
U2	−1731 (−174, −921, −788)	−1126	−2069
	Differences		
U2-N2	232 (285, 143, −181)	−275	138

[a]N1: Crystal minimized, N2: Native quenched after 50 ps of MD. U1: Minimized ideal β-strand; U2: Minimized twisted β-strand. All values in kcal/mol.

CO-NH groups. This is consistent with the explanation proposed recently[115] for the origin of the right-handed twist in protein β sheets. The enthalpy difference between the minimized native structure (after dynamics) and the minimized twisted β conformation is 138 kcal/mol, significantly higher than the experimental value. This suggests that completely extended chains are not good models for the denatured state at room temperature.

The heat capacity of unfolding is also experimentally measurable. An analysis of the change in heat capacity of proteins has been made based on the model developed here.[116] The results show that more realistic (collapsed) models for the denatured state are consistent with experimental heat capacity data.

Helix Unfolding Enthalpy

For polyalanine we calculated the enthalpy of unfolding from the average energies of the helical and the unfolded conformations. The helix was built in the ideal conformation ($\phi = -57$, $\psi = -47$) and minimized for 300 steps. For the unfolded chain, the twisted β chain ($\phi = -72$, $\psi = 79$) was constructed. Then 20 ps simulations at 300 K for both the helical and the twisted β conformation were performed, and the average properties over the last 2 ps were calculated. The enthalpy of unfolding at 300 K is found to be 30 kcal/mol, or 1.5 kcal/mol per residue. This value is close to the experimental value for a 50-residue alanine-rich peptide, which is 1.3 kcal/mol per residue.[117] Extensive sampling of unfolded conformations is needed for quantitative results (see Enthalpy and Free Energy of Unfolding, above).

DISCUSSION

An atom-based function (EEF1) is proposed for describing the effective energy hypersurface of proteins. It uses the CHARMM19 polar hydrogen potential energy function complemented by a simple Gaussian model for the solvation free energy. The solvation model is based on theoretical considerations that are used to determine how neighboring atoms affect the solvation free energy of a given atom by excluding solvent from the surrounding space. The functional form of the solvation free energy density was chosen based on statistical mechanical calculations. Ionic side-chains are treated as highly polar but neutral, so as to avoid the problem arising from the insufficiently accurate compensation between the very large coulombic interactions and solvation free energies of ionic groups. The model uses the linear distance-dependent dielectric constant in CHARMM. This factor was found to play the useful role of leaving short-range (hydrogen bonding) interactions almost unaffected while reducing significantly medium- and long-range interactions. Given the form of the solvation model, it was parametrized with the empirical group contributions to the solvation free energy given by Privalov and Makhatadze.[47]

Although many aspects of the model, particularly those regarding electrostatic interactions, are heuristic and were chosen to make the model work, it has an underlying physical basis. The present model differs substantially from "knowledge-based" potentials[39–41] and can be used for molecular dynamics simulations as well as for comparing "reasonable" folded structures.

A wide range of problems has been studied to ensure that the model gives realistic results. The model has been shown to give stable native proteins in room temperature molecular dynamics simulations, reasonable energies for unfolded conformations, high-temperature unfolding pathways[43] in agreement with those obtained in explicit water simulations, and discriminates native from nonnative conformations.[44]

Possible applications of the solvation model involve areas where explicit water modeling is expensive, awkward, or simply unthinkable. They include molecular dynamics simulations of very large proteins and protein complexes, e.g., viruses, conformational searches for structural prediction and inverse folding studies, NMR structure refinement, dynamics of protein folding and unfolding, and the study of the denatured state at an atomic level of detail.

For many of the above applications, EEF1 appears to be sufficiently accurate. Improvements will have to be introduced if higher accuracy and more specific structural details are required. The largest errors are likely to arise from the polar groups because they have the largest solvation free energy. An improvement would be to incorporate the effect of the perturbation of solvent structure into the model. Dielectric screening is just one form of such a perturbation. Packing effects in liquids (the familiar oscillations in the radial distribution functions) are also caused by solvent perturbation. Inclusion of such effects might necessitate the use of nonpairwise additive effective interactions. The finite size of water and specific structural water molecules could be taken into account in more sophisticated models.

A continuum electrostatics treatment for polar group solvation has many advantages over solvent exclusion models because it can account for solvent orientational polarization effects. Poisson-Boltzmann calculations, for example, give reasonable results for the alanine dipeptide conformational map,[37,79,80] which is not accurately represented by the present model. However, certain problems exist in that Poisson-Boltzmann calculations appear to

overestimate the solvation free energy of ionic side-chains separately solvated relative to their ion pairing on the protein surface. Salt bridge formation is too unfavorable with the CHARMM19 potential energy function combined with a Poisson-Boltzmann model for solvation.[10,13] As a specific example, we found from calculations on a Lys-Ala-Glu tripeptide that the Lys-Glu ion-paired configuration had a total energy about 10 kcal/mol higher than a non-ion-paired configuration; the total energy of the ion-paired configuration was −284 kcal/mol (internal −249, solvation −34 kcal/mol) and that of the ion-separated conformation was −295 kcal/mol (internal −177, solvation −117 kcal/mol). The result is unrealistic, given the great number of ion pairs observed at the surface of proteins in crystal structures.[118] This difficulty with the Poisson-Boltzmann approach could perhaps be rectified by using a larger value for the protein dielectric constant for regions of the protein that are exposed to solvent.[119] Recent calculations of peptide-folding equilibria with an analytical approximation to the Poisson-Boltzmann equation with a hydrophobic correction term have yielded results in agreement with experiments.[120]

Possible extensions of the model are of interest. To introduce the effect of pH, protons could be added to Asp, Glu, His, or the C-terminus and subtracted from Lys, Arg, and the N-terminus, and a partial charge distribution and a solvation free energy pertaining to the neutral residues would be assigned. Ionic strength effects could be incorporated by using a Debye-Huckel type shielding of the Coulomb interactions. The solvent-exclusion model could also be used for solvents other than pure water; e.g., at finite concentration of denaturants. Data on the solvation free energies of small model compounds in solutions of denaturant (e.g., ref. 121) could be used to derive the dependence of group solvation free energies on denaturant concentration. This would allow simulations of chemical denaturation in addition to thermal denaturation and could lead to an improved understanding of the effects of denaturants on protein stability.

ACKNOWLEDGMENTS

This work was supported by a grant from the National Science Foundation and a Burroughs Wellcome PMMB postdoctoral fellowship to T.L.

APPENDIX

Effective Energy and Conformational Distribution

For a system described by classical mechanics, the partition function in the canonical (NVT) ensemble for a macromolecule of M atoms dissolved in N solvent molecules is[122]

$$Q = \frac{Z}{N!\Lambda^{3M}\Lambda^{3N}} \tag{A1}$$

where λ is the thermal de Broglie wavelength. The quantity Z is the classical canonical configurational integral

$$Z = \int \exp(-\beta H)\, d\mathbf{r}^N d\mathbf{R}^M \tag{A2}$$

where $\mathbf{r}^N$ are the solvent coordinates, $\mathbf{R}^M$ the macromolecular coordinates, H is the Hamiltonian, and $\beta = 1/kT$. From a knowledge of Q, the thermodynamic properties of the system can be calculated by standard methods.[122] Of primary interest here are the free energy, the energy, and the entropy. The Helmholtz free energy, energy, and entropy are

$$A = -kT \ln Q = -kT \ln Z + kT \ln (N!\Lambda^{3M}\Lambda^{3N}) \tag{A3}$$

$$E = kT^2\left(\frac{\partial \ln Q}{\partial T}\right)_{NV} \tag{A4}$$

$$S = k \ln Q + kT\left(\frac{\partial \ln Q}{\partial T}\right)_{NV}. \tag{A5}$$

In most empirical force fields without polarization, the Hamiltonian is additive, so that we can write

$$H = H_{intra} + H_{mw} + H_{ww}. \tag{A6}$$

The three terms in Equation A6 represent the intramacromolecule, H_{intra}, the macromolecule-solvent, H_{mw}, and the solvent-solvent interactions, H_{ww}. One can formally perform the integration over the solvent coordinates in Equation A1 by defining the potential of mean force W:

$$\exp(-\beta W) = Z_{ww}^{-1} \int \exp(-\beta H)\, d\mathbf{r}^N \tag{A7}$$

where

$$Z_{ww} = \int \exp(-\beta H_{ww})\, d\mathbf{r}^N, \tag{A8}$$

is the pure solvent configurational integral. The configurational integral becomes:

$$Z = Z_{ww} \int \exp(-\beta W)\, d\mathbf{R}^M. \tag{A9}$$

Using Equation A6 we obtain

$$\exp(-\beta W) = \exp(-\beta H_{intra}) Z_{ww}^{-1} \int \times \exp(-\beta H_{mw} - \beta H_{ww})\, d\mathbf{r}^N. \tag{A10}$$

The last integral in Equation A10 can be written

$$Z_{ww}^{-1} \int \exp(-\beta H_{mw} - \beta H_{ww})\, d\mathbf{r}^N = \langle \exp(-\beta H_{mw}) \rangle_0 \tag{A11}$$

where the ensemble average is taken over the pure solvent. This ensemble average is related to the excess chemical potential, or standard solvation free energy.[123]

$$-kT \ln \langle \exp(-\beta H_{mw}) \rangle_o = \Delta G^{slv} \tag{A12}$$

Equation 10 becomes

$$\exp(-\beta W) = \exp(-\beta H_{intra}) \exp(-\beta \Delta G^{slv}) \tag{A13}$$

or

$$W = H_{intra} + \Delta G^{slv}. \quad (A14)$$

Thus, the potential of mean force consists of two terms: the first is the intramacromolecular energy and the second is the solvation free energy. In the following, instead of the term potential of mean force, we use the more physically appealing term *effective energy*. Previous definitions of the potential of mean force[1,124–126] do not include the Z_{ww} factor. The potential of mean force defined here differs from the ones given in the above references by a constant equal to the free energy of pure solvent; this has no effect on the results.

For the macromolecule, it is more convenient to use the internal coordinates $\mathbf{q}$ instead of the Cartesian coordinates of its atoms, $\mathbf{R}^M$. The Jacobian for the transformation depends only on bond lengths and bond angles[124,127] and is, therefore, approximately constant for all conformations and can be taken out of the integral in Equation A9. For notational simplicity, we can include this factor implicitly in the differential $d\mathbf{q}$. Because the system is homogeneous, the integration over the six external coordinates can be performed to give $V8\pi^2$; i.e., we have

$$Z = Z_{ww} V8\pi^2 \int \exp(-\beta W)\, d\mathbf{q}. \quad (A15)$$

The probability of finding the system at the configuration $(\mathbf{q},\mathbf{r}^N)$ is

$$p(\mathbf{q}, \mathbf{r}^N) = \frac{\exp(-\beta H)}{\int \exp(-\beta H)\, d\mathbf{q} d\mathbf{r}^N}. \quad (A16)$$

This probability distribution is normalized to unity. The marginal probability distribution of finding the macromolecule in state $\mathbf{q}$ (i.e., the probability distribution of the macromolecule) is

$$p(\mathbf{q}) = \int p(\mathbf{q}, \mathbf{r}^N)\, d\mathbf{r}^N. \quad (A17)$$

Using Equation A7, we obtain

$$p(\mathbf{q}) = \frac{\exp(-\beta W(\mathbf{q}))}{\int \exp(-\beta W(\mathbf{q}))\, d\mathbf{q}}. \quad (A18)$$

Equation 18 shows that the conformational distribution of the macromolecule in solution is governed by the effective energy W. From Equations A18 and A15 we obtain

$$\int p(\mathbf{q}) \ln p(\mathbf{q})\, d\mathbf{q} = \int p(\mathbf{q}) [-\beta W(\mathbf{q}) - \ln Z + \ln Z_{ww} + \ln V8\pi^2]\, d\mathbf{q} = -\ln Z + \ln Z_{ww} + \ln V8\pi^2 - \beta \int p(\mathbf{q}) W(\mathbf{q})\, d\mathbf{q}. \quad (A19)$$

Combining Equation A3 with Equation A19, we have the Helmholtz free energy in the form

$$A = A^o + kT \ln\left(\frac{\Lambda^{3M}}{V8\pi^2}\right) + \int p(\mathbf{q}) W(\mathbf{q})\, d\mathbf{q} + kT \int p(\mathbf{q}) \ln p(\mathbf{q})\, d\mathbf{q} \quad (A20)$$

where A° is the free energy of the pure solvent ($-kT \ln Z_{ww}/N!\Lambda^{3N}$) and the second term is the ideal contribution from macromolecular translation and rotation; these two terms are independent of conformation and so play no role in determining relative stabilities. The third term in Equation A20 is the *average effective energy*, which is equal to the average intramolecular energy plus the average solvation free energy. The last term (divided by $-T$) is the *configurational entropy* of the macromolecule, including the conformational and vibrational entropy contributions. We suggest, as have others,[128] that the term effective energy be used for quantities that do not include the configurational entropy and that the term free energy be reserved for quantities that include such contributions. It should be noted that the conformational distribution that should be used in the calculation of the configurational entropy is the conformational distribution *in solution*, and not that in vacuum. The change in free energy resulting from a conformational change, such as folding, can be similarly decomposed into the change in configurational entropy plus the change in the effective energy averaged over the conformations identified with a given state.

REFERENCES

1. Karplus M, Shakhnovich E. Protein folding: theoretical studies of thermodynamics and dynamics. In: Creighton TE, editor. Protein folding. New York: Freeman; 1992. p 127–195.
2. Kim PS, Baldwin RL. Intermediates in the folding reactions of small proteins. Ann Rev Biochem 1990;59:631–660.
3. Baker D, Agard DA. Kinetics versus thermodynamics in protein folding. Biochemistry 1994;33:7505–7509.
4. Dinner AR, Karplus M. A metastable state in folding simulations of a protein model. Nature Struct Biol 1998;5:236–240.
5. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. CHARMM: a program for macromolecular energy minimization and dynamics calculations. J Comput Chem 1983; 4:187–217.
6. Weiner SJ, Kollman PA, Nguyen DT, Case DA. An all atom force field for simulations of proteins and nucleic acids. J Comput Chem 1986;7:230–252.
7. Jorgensen WL, Tirado-Rives J. The OPLS potential functions for proteins. Energy minimizations for crystals of cyclic peptides and crambin. J Am Chem Soc 1988;110:1657–1666.
8. Hagler AT, Maple JR, Thacher TS, Fitzgerald GB, Dinur U. Potential energy functions for organic and biomolecular systems. In: van Gunsteren WF, Weiner PK, editors. Computer simulation of biomolecular systems. Leiden: ESCOM; 1989. p 149–167.
9. Lazaridis T, Archontis G, Karplus M. Enthalpic contributions to protein stability: insights from atom-based calculations and statistical mechanics. Adv Protein Chem 1995;47:231–306.
10. Hendsch ZS, Tidor B. Do salt bridges stabilize proteins? A continuum electrostatic analysis. Protein Sci 1994;3:211–226.
11. Montroll EW, Lebowitz JL, editors. The liquid state of matter: fluids, simple and complex. Amsterdam: North-Holland; 1982.
12. Hansen JP, McDonald IR. Theory of simple liquids. London: Academic Press; 1986.
13. Caflisch A, Karplus M. Acid and thermal denaturation of barnase

investigated by molecular dynamics simulations. J Mol Biol 1995;252:672–708.
14. Brooks CL III, Karplus M, Pettitt BM. Proteins: a theoretical perspective of dynamics, structure, and thermodynamics. Adv Chem Phys 1988;71.
15. Torrie GM, Valleau JP. Nonphysical sampling distribution in Monte Carlo free energy estimation: umbrella sampling. J Comput Phys 1977;23:187–199.
16. Jorgensen WL, Buckner JK. Use of statistical perturbation theory for computing solvation effects on molecular conformation. Butane in water. J Phys Chem 1987;91:6083–6085.
17. Tobias DJ, Brooks CL III. The thermodynamics of solvophobic effects: a molecular dynamics study of n-butane in carbon tetrachloride and water. J Chem Phys 1990;92:2582–2592.
18. Elber R. Calculation of the potential of mean force using molecular dynamics with linear constraints. An application to a conformational transition in a solvated dipeptide. J Chem Phys 1990;93:4312–4321.
19. Anderson AG, Hermans J. Microfolding: conformational probability map for the Ala dipeptide in water from molecular dynamics simulations. Proteins 1988;3:262–265.
20. Mezei M, Mehrotra PK, Beveridge DL. Monte Carlo simulation of the free energy and internal energy of hydration for the Ala dipeptide at 25°C. J Am Chem Soc 1985;107:2239–2245.
21. Tobias DJ, Brooks CL III. Thermodynamics and mechanism of α helix initiation in Ala and Val peptides. Biochemistry 1991;30:6059–6070.
22. Lazaridis T, Tobias DJ, Brooks CLI, Paulaitis ME. Reaction paths and free energy profiles for conformational transitions: an internal coordinate approach. J Chem Phys 1991;95:7612–7625.
23. Fraternali F, van Gunsteren WF. Conformational transitions of a dipeptide in water. Biopolymers 1994;34:347–355.
24. Boczko EM, Brooks CLI. First-principles calculation of the folding free energy of a three-helix bundle protein. Science 1995;269:393–396.
25. Wesson L, Eisenberg D. Atomic solvation parameters applied to molecular dynamics of proteins in solution. Protein Sci 1992;1:227–235.
26. Neria E, Fischer S, Karplus M. Simulation of activation free energies in molecular systems. J Chem Phys 1996;105:1902–1921.
27. Schiffer CA, Caldwell JW, Stroud RM, Kollman PA. Inclusion of solvation free energy with molecular mechanics energy: alanyl dipeptide as a test case. Protein Sci 1992;1:396–400.
28. Fraternali F, van Gunsteren WF. An efficient mean solvation force model for use in molecular dynamics simulations of proteins in aqueous solution. J Mol Biol 1996;256:939–948.
29. Kang YK, Nemethy G, Scheraga HA. Free energies of hydration of solute molecules 1. Improvement of the hydration shell model by exact computations of overlap volumes. J Chem Phys 1987;91:4105–4109.
30. Colonna-Cesari F, Sander C. Excluded volume approximation to protein-solvent interaction. The solvent contact model. Biophys J 1990;57:1103–1107.
31. Gibson KD, Scheraga HA. Minimization of polypeptide energy I. Preliminary structures of bovine pancreatic RNAse S-peptide. Proc Natl Acad Sci USA 1967;58:420–427.
32. Levitt M. A simplified representation of protein conformations for rapid simulation of protein folding. J Mol Biol 1976;104:59–107.
33. Wodak SJ, Janin J. Analytical approximation to the accessible surface area of proteins. Proc Natl Acad Sci USA 1980;77:1736–1740.
34. Hasel W, Hendrickson TF, Still WC. A rapid approximation to the solvent accessible surface area of atoms. Tetrah Comput Methodol 1988;1:103–116.
35. Stouten PFW, Frommel C, Nakamura H, Sander C. An effective solvation term based on atomic occupancies for use in protein simulations. Mol Simul 1993;10:97–120.
36. Sharp K, Honig B. Electrostatic interactions in macromolecules: theory and applications. Ann Rev Biophys Biophys Chem 1990;19:301–332.
37. Still WC, Tempczyk A, Hawley RC, Hendrickson T. Semianalytical treatment of solvation for molecular mechanics and dynamics. J Am Chem Soc 1990;112:6127–6129.
38. Schaefer M, Karplus M. A comprehensive analytical treatment of continuum electrostatics. J Phys Chem 1996;100:1578–1599.
39. Miyazawa S, Jernigan R. Estimation of effective interresidue contact energies from protein crystal structures: quasi-chemical approximation. Macromolecules 1985;18:534–552.
40. Sippl MJ. Calculation of conformational ensembles from potentials of mean force. J Mol Biol 1990;213:859–883.
41. Kolinski A, Godzik A, Skolnick J. A general method for the prediction of the three dimensional structure and folding pathway of globular proteins: application to designed helical proteins. J Chem Phys 1993;98:7420–7433.
42. Marchler-Bauer A, Bryant SH. A measure of success in fold recognition. TIBS 1997;22:236–40.
43. Lazaridis T, Karplus M. "New view" of protein folding reconciled with the old through multiple unfolding simulations. Science 1997;278:1928–1931.
44. Lazaridis T, Karplus M. Discrimination of the native from misfolded protein models with an energy function including implicit solvation. J Mol Biol 1999; in press.
45. Lazaridis T. Inhomogeneous fluid approach to solvation thermodynamics 1. Theory. J Phys Chem 1998;102:3531–3541.
46. Lazaridis T. Inhomogeneous fluid approach to solvation thermodynamics 2. Applications to simple fluids. J Phys Chem 1998;102:3542–3550.
47. Privalov PL, Makhatadze GI. Contribution of hydration to protein folding thermodynamics II. The entropy and Gibbs energy of hydration. J Mol Biol 1993;232:660–679.
48. Makhatadze GI, Privalov PL. Contribution of hydration to protein folding thermodynamics. I. The enthalpy of hydration. J Mol Biol 1993;232:639–659.
49. Privalov PL, Makhatadze GI. Contribution of hydration and non-covalent interactions to the heat capacity effect on protein unfolding. J Mol Biol 1992;224:715–723.
50. Makhatadze GI, Privalov PL. Heat capacity of proteins I. Partial molar heat capacity of individual amino acid residues in aqueous solution: hydration effect. J Mol Biol 1990;213:375–384.
51. Lazaridis T, Paulaitis ME. Simulation studies of the hydration entropy of simple hydrophobic solutes. J Phys Chem 1994;98:635–642.
52. Matubayasi N, Reed LH, Levy RM. Thermodynamics of the hydration shell 1. Excess energy of a hydrophobic solute. J Phys Chem 1994;98:10640–10649.
53. Lazaridis T, Karplus M. Orientational correlations and entropy in liquid water. J Chem Phys 1996;105:4294–4316.
54. Nauchitel VV, Somorjai RL. Gaussian neighborhood: a new measure of accessibility for residues of protein molecules. Proteins 1993;15:50–61.
55. Allen MP, Tildesley DJ. Computer simulation of liquids. Oxford: Clarendon Press; 1989.
56. Serrano L, Kellis JT Jr, Cann P, Matouschek A, Fersht AR. The folding of an enzyme II. Substructure of barnase and the contribution of different interactions to protein stability. J Mol Biol 1992;224:783–804.
57. Matthews BW. Structural and genetic analysis of protein stability. Ann Rev Biochem 1993;62:139–60.
58. Huston SE, Rossky PJ. Free energies of association for the sodium-dimethyl phosphate ion pair in aqueous solution. J Phys Chem 1989;93:7888–7895.
59. Tirado-Rives J, Jorgensen WL. Molecular dynamics simulations of an α-helical analogue of RNAse A S-peptide in water. Biochemistry 1991;30:3864–3871.
60. Gelin BR, Karplus M. Side-chain torsional potentials: effect of dipeptide, protein and solvent. Biochemistry 1979;18:1256–1268.
61. McCammon JA, Wolynes PG, Karplus M. Picosecond dynamics of tyrosine sidechains in proteins. Biochemistry 1979;18:928–942.
62. Mitchinson C, Baldwin RL. The design and production of semisynthetic RNAses with increased thermostability by incorporation of S-peptide analogues with enhanced helical stability. Proteins 1986;1:23–33.
63. Rashin AA, Iofin M, Honig B. Internal cavities and buried waters in globular proteins. Biochemistry 1986;25:3619–3625.
64. Ernst JA, Clubb RT, Zhou H-Z, Gronenborn AM, Clore GM. Demonstration of positionally disordered water within a protein hydrophobic cavity by NMR. Science 1995;267:1813–1816.
65. Buckle AM, Cramer P, Fersht AR. Structural and energetic

responses to cavity-creating mutants in hydrophobic cores: observation of a buried water molecule and the hydrophilic nature of such hydrophobic surfaces. Biochemistry 1996;35:4298– 4305.
66. Caflisch A, Karplus M. Molecular dynamics simulation of protein denaturation: solvation of the hydrophobic cores and secondary structure of barnase. Proc Natl Acad Sci USA 1994;91:1746–1750.
67. Mark AE, van Gunsteren WF. Decomposition of the free energy of a system in terms of specific interactions. J Mol Biol 1994;240:167–176.
68. Dill KA. Additivity principles in biochemistry. J Biol Chem 1997;272:701–704.
69. Fersht AR, Matouschek A, Serrano L. The folding of an enzyme I. Theory of protein engineering analysis of stability and pathway of protein folding. J Mol Biol 1992;224:771–782.
70. Gao J, Kuczera K, Tidor B, Karplus M. Hidden thermodynamics of mutant proteins: a molecular dynamics analysis. Science 1989;244:1069–1072.
71. Archontis G, Karplus M. Cumulant expansion of the free energy: application to free energy derivatives and component analysis. J Chem Phys 1996;105:11246–11260.
72. Brady GP, Szabo A, Sharp KA. On the decomposition of free energies. J Mol Biol 1996;263:123–125.
73. Cabani S, Gianni P, Mollica V, Lepori L. Group contributions to the thermodynamic properties of non-ionic organic solutes in dilute aqueous solution. J Soln Chem 1981;10:563–595.
74. Ben-Naim A. Solvation effects on protein association and protein folding. Biopolymers 1990;29:567–596.
75. Yu H-A, Pettitt BM, Karplus M. Aqueous solvation of N-methyl acetamide conformers: comparison of simulations and integral equation theories. J Am Chem Soc 1991;113:2425–2434.
76. Pratt LR, Chandler D. Theory of the hydrophobic effect. J Chem Phys 1977;67:3683–3704.
77. Hagler AT, Osguthorpe DJ, Robson B. Monte Carlo simulation of water behavior around the dipeptide N-acetylalanyl-N-methylamide. Science 1980;208:599–601.
78. Pettitt BM, Karplus M. Conformational free energy of hydration for the Ala dipeptide: Thermodynamic analysis. J Phys Chem 1988;92:3994–3997.
79. Jean-Charles A, Nicholls A, Sharp K, et al. Electrostatic contributions to solvation energies. Comparison of free energy perturbation and continuum calculations. J Am Chem Soc 1991;113:1454–1455.
80. Osapay K, Young WS, Bashford D, Brooks CL III, Case DA. Dielectric continuum models for hydration effects on peptide conformational transitions. J Phys Chem 1996;100:2698–2705.
81. van Gunsteren WF, Karplus M. Protein dynamics in solution and in a crystalline environment: a molecular dynamics study. Biochemistry 1982;21:2259–2274.
82. Levitt M, Sharon R. Accurate simulation of protein dynamics in solution. Proc Natl Acad Sci USA 1988;85:7557–7561.
83. Vijayakumar S, Vishveshwara S, Ravishanker G, Beveridge DL. Differential stability of β-sheets and α-helices in β-lactamase: a high temperature molecular dynamics study of unfolding intermediates. Biophys J 1993;65:2304–2312.
84. Mark AE, van Gunsteren WF. Simulation of the thermal denaturation of hen egg white lysozyme: trapping the molten globule state. Biochemistry 1992;31:7745–7748.
85. van Gunsteren WF, Berendsen HJC. Computer simulation as a tool for tracing the conformational differences between proteins in solution and in the crystalline state. J Mol Biol 1984;176:559–564.
86. Levitt M. Molecular dynamics of macromolecules in water. Chem Scripta 1989;29A:197–203.
87. Daggett V, Levitt M. A model of the molten globule state from molecular dynamics simulations. Proc Natl Acad Sci USA 1992;89:5142–5146.
88. Brunne RM, Liepinsh E, Otting G, Wuthrich K, van Gunsteren WF. Hydration of proteins. A comparison of experimental residence times of water molecules solvating BPTI with theoretical model calculations. J Mol Biol 1993;231:1040–1048.
89. Alonso DOW, Daggett V. Molecular dynamics simulations of protein unfolding and limited refolding: characterization of partially unfolded states of ubiquitin in 60% methanol and in water. J Mol Biol 1995;247:501–520.
90. Daggett V, Levitt M. A molecular dynamics simulation of the C-terminal fragment of the L7/L12 ribosomal protein in solution. Chem Phys 1991;158:501–512.
91. Chandrasekhar I, Clore GM, Szabo A, Gronenborn AM, Brooks BR. A 500 ps molecular dynamics simulation study of interleukin-1β in water. J Mol Biol 1992;226:239–250.
92. Wong CF, McCammon JA. Computer simulation and the design of new biological molecules. Isr J Chem 1986;27:211–215.
93. Nakagawa S, Yu H-A, Karplus M, Umeyama H. Active site dynamics of acyl-chymotrypsin. Proteins 1993;16:172–194.
94. Lazaridis T, Paulaitis ME. Computational study of conformational transitions in the active site of tosyl-α-chymotrypsin. J Am Chem Soc 1994;116:1546–1556.
95. Wlodek ST, Clark TW, Scott LR, McCammon JA. Molecular dynamics of acetylcholinesterase dimer complexed with tacrine. J Am Chem Soc 1997;119:9513–9522.
96. Li A, Daggett V. Investigation of the solution structure of chymotrypsin inhibitor 2 using molecular dynamics: comparison to x-ray crystallographic and NMR data. Protein Eng 1995;8:1117–1128.
97. Terwilliger TC, Eisenberg D. The structure of melittin II. Interpretation of the structure. J Biol Chem 1982;257:6016–6022.
98. Wilcox W, Eisenberg D. Thermodynamics of melittin tetramerization determined by circular dichroism and implications for protein folding. Protein Sci 1992;1:641–653.
99. Finkelstein AV, Janin J. The price of lost freedom: entropy of bimolecular complex formation. Protein Eng 1989;3:1–3.
100. Tidor B, Karplus M. The contribution of vibrational entropy to molecular association. J Mol Biol 1994;238:405–414.
101. Brooks CLI, Karplus M. Solvent effects on protein motions and protein effects on solvent motions. Dynamics of the active site region of lysozyme. J Mol Biol 1989;208:159–181.
102. Steinbach PJ, Brooks BR. Protein hydration elucidated by molecular dynamics simulations. Proc Natl Acad Sci USA 1993;90:9135–9139.
103. Steinbach PJ, Brooks BR. Hydrated myoglobin's anharmonic fluctuations are not primarily due to dihedral transitions. Proc Natl Acad Sci USA 1996;93:55–59.
104. Rupley JA, Careri G. Protein hydration and function. Adv Prot Chem 1991;41:37–172.
105. Rupley JA, Gratton E, Careri G. Water and globular proteins. Trends Biochem Sci 1983;8:18–22.
106. Parak F, Frolov EN, Mossbauer RL, Goldanskii VI. Dynamics of metmyoglobin crystals investigated by nuclear gamma resonance absorption. J Mol Biol 1981;145:825–833.
107. Bone S, Pethig R. Dielectric studies of protein hydration and hydration-induced flexibility. J Mol Biol 1985;181:323–326.
108. Jackson SE, Fersht AR. Folding of CI2 1. Evidence for a two-state transition. Biochemistry 1991;30:10428–10435.
109. Dill KA. Theory for the folding and stability of globular proteins. Biochemistry 1985;24:1501–1509.
110. Calmettes P, Roux B, Durand D, Desmadril M, Smith JC. Configurational distribution of denatured phosphoglycerate kinase. J Mol Biol 1993;231:840–848.
111. Makhatadze GI, Privalov PL. Energetics of protein structure. Adv Protein Chem 1995;47:307–425.
112. Schellman JA. The stability of hydrogen-bonded peptide structures in aqueous solution. C R Trav Lab Carlsberg Ser Chim 1955;29:230–259.
113. Lee KH, Xie D, Freire E, Amzel LM. Estimation of changes in side chain configurational entropy in binding and folding: general methods and application to helix formation. Proteins 1994;20:68–84.
114. Makhatadze GI, Privalov PL. On the entropy of protein folding. Protein Sci 1996;5:507–510.
115. Maccallum PH, Poet R, Milner-White EJ. Coulombic attractions between partially charged main-chain atoms stabilize the right-handed twist found in most β-strands. J Mol Biol 1995;248:374–384.
116. Lazaridis T, Karplus M. Heat capacity and compactness of denatured proteins. Biophys Chem 1999; in press.
117. Scholtz JM, Marqusee S, Baldwin RL, et al. Calorimetric determination of the enthalpy change for the α-helix to coil transition of an alanine peptide in water. Proc Natl Acad Sci USA 1991;88:2854–2858.
118. Barlow DJ, Thornton JM. Ion-pairs in proteins. J Mol Biol 1983;168:867–885.

119. Delepierre M, Dobson CM, Karplus M, Poulsen FM, States DJ, Wedin RE. Electrostatic effects and hydrogen exchange behaviour in proteins. The pH dependence of exchange rates in lysozyme. J Mol Biol 1987;197:111–122.
120. Schaefer M, Bartels C, Karplus M. Solution conformations and thermodynamics of structured peptides: molecular dynamics simulation with an implicit solvation model. J Mol Biol 1998;284: 835–847.
121. Zou Q, Habermann-Rottinghaus SM, Murphy KP. Urea effects on protein stability: hydrogen bonding and the hydrophobic effect. Proteins 1998;31:107–115.
122. McQuarrie DA. Statistical mechanics. New York: Harper & Row; 1976.
123. Ben-Naim A. Standard thermodynamics of transfer. uses and misuses. J Phys Chem 1978;82:792–803.
124. DeVoe H. Theory of the conformations of biological macromolecules in solution. In: Timascheff SN, Fasman GD, editors. Structure and stability of biological macromolecules. New York: Marcel Dekker; 1969.
125. Gibson KD, Scheraga HA. Minimization of polypeptide energy V. Theoretical aspects. Physiol Chem Phys 1969;1:109–126.
126. Wang J, Purisima EO. Analysis of thermodynamic determinants in helix propensities of nonpolar amino acids through a novel free energy calculation. J Am Chem Soc 1996;118:995–1001.
127. Go N, Scheraga HA. On the use of classical statistical mechanics in the treatment of polymer chain conformations. Macromolecules 1976;9:535–542.
128. Dobson C.M, Sali A, Karplus M. Protein folding: a perspective from theory and experiment. Angew Chem Int Ed 37:868–93, 1998.