PROTEINS: Structure, Function, and Genetics 10:199–218 (1991)

Empirical Solvation Models Can Be Used to Differentiate Native From Near-Native Conformations of Bovine Pancreatic Trypsin Inhibitor

J. Vila, R. L. Williams, M. Vásquez, and H. A. Scheraga
Baker Laboratory of Chemistry, Cornell University, Ithaca, New York 14853-1301

ABSTRACT **Several hydration models for peptides and proteins based on solvent accessible surface area have been proposed previously. We have evaluated some of these models as well as four new ones in the context of near-native conformations of a protein. In addition, we propose an empirical site–site distance-dependent correction that can be used in conjunction with any of these models.**

The set of near-native structures consisted of 39 conformations of bovine pancreatic trypsin inhibitor (BPTI) each of which was a local minimum of an empirical energy function (ECEPP) in the absence of solvent. Root-mean-square (rms) deviations from the crystallographically determined structure were in the following ranges: 1.06–1.94 Å for all heavy atoms, 0.77–1.36 Å for all backbone heavy atoms, 0.68–1.33 Å for all α-carbon atoms, and 1.41–2.72 Å for all side-chain heavy atoms.

We have found that there is considerable variation among the solvent models when evaluated in terms of concordance between the solvation free energy and the rms deviations from the crystallographically determined conformation. The solvation model for which the best concordance (0.939) with the rms deviations of the C^{α} atoms was found was derived from NMR coupling constants of peptides in water combined with an exponential site–site distance dependence of the potential of mean force.

Our results indicate that solvation free energy parameters derived from nonpeptide free energies of hydration may not be transferrable to peptides. Parameters derived from peptide and protein data may be more applicable to conformational analysis of proteins. A general approach to derive parameters for free energy of hydration from ensemble-averaged properties of peptides in solution is described.

Key words: empirical potentials, energy calculations, surface area, protein stability, protein folding, protein structure prediction

INTRODUCTION

One can expect to be able to compute solution conformations of polypeptides and proteins only by including the effects of the solvent. For calculations of structures of proteins, estimation of free energy of hydration using explicit water molecules is currently computationally too complex to be of use in the protein folding problem where free energies must be estimated for many conformations. Several continuum models of solvation have been proposed that are computationally more tractable.[1–6]

Continuum models based on solvent accessible surface area make the implicit assumption that, once an atom is buried, it makes no additional contribution to the free energy of hydration regardless of its environment in the interior. In addition, such models assume that the potential of mean force between a pair of atoms is zero once the two atoms are farther apart than the sum of their van der Waals radii plus the thickness of one hydration shell. However, there are theoretical reasons and experimental observations to support the existence of long-range hydrophobic and hydrophilic interactions.[7–13] Because the available experimental evidence is obtained from interactions between uniform and macroscopic surfaces, it may be questioned whether these observations are entirely relevant to intramolecular interactions (involving hydration) within a macromolecular solute consisting of both hydrophobic and hydrophilic groups. Nevertheless, two generalizations can be deduced from these experiments. The strength and range of the forces of interaction depend on the nature of the surfaces, and the decay of the force of interaction with distance can be modeled with an exponential form.[9]

Received June 22, 1990; revision accepted December 3, 1990.
Address reprint requests to Harold A. Scheraga, Baker Laboratory of Chemistry, Cornell University, Ithaca, New York 14853-1301.

J. Vila is on leave from the National University of San Luis, Faculty of Science and Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Instituto de Matemática Aplicada—San Luis. Ejército de los Andes 950, 5700 San Luis, Argentina, 1988–1991.

M. Vásquez's present address is: Protein Design Lab, 2375 Garcia Avenue, Mountain View, CA 94043.

Conformational energy calculations carried out in the absence of solvent effectively subject the conformations to artificially large surface forces. When a set of near-native local minimum-energy conformations is generated, it can be impossible to discern the native conformation on the basis of such conformational energies alone. The principal purpose of the work described here is to evaluate the utility of several models of solvation free energy for discriminating the native conformation of a protein among a collection of near-native conformations whose energy was minimized in the absence of solvent.

We have investigated the utility of several solvation models that depend only on solvent accessible surface area, and also two modifications of these models in which an explicit dependence on the distance between a pair of groups (either atoms or unified side chains) is included. We will refer to a model that depends only on solvent accessible surface area as an envelope model. We report the results for six envelope models that differ only in the radii used to describe atoms and the atomic solvation parameters. These are the models of Eisenberg and McLachlan[1] (E&M), Ooi et al.[2] (OONS), and four models that we have developed recently: three derived by fitting to small-molecule free energies of hydration [one with radii as adjustable parameters (SRA) and two with radii fixed: one with a zero radius solvent sphere (ZRSS), and one with a 1.4 Å radius solvent sphere with atomic solvation parameters optimized using nonpeptide thermodynamic data (SRFOPT)], and one fixed-radii model (JRF) in which solvation parameters were derived from vicinal NH-C$^{\alpha}$H coupling constants observed for peptides in water.

The two pairwise-distance-dependent modifications that we have investigated are of the following forms: the AIPD (atomic interaction using pairwise distances) modification in which interactions between all atoms are taken into account, and the UISC (unified interacting side chains) modification in which the side chains are represented as unified interacting groups. As will be described below, each AIPD modification is based on an envelope model. Consequently, there is a corresponding AIPD modification for any envelope model. We have tested the AIPD modification corresponding to each of the six envelope models mentioned above. We have tested only one UISC modification—one that uses the parameters of Fauchere and Pliska[14] (F&P).

Even though the concept of an exponential decay in the potential of mean force was deduced from interacting macroscopic surfaces, we have incorporated this concept, in a microscopic theory, for a pair of interacting groups in all of the AIPD and UISC modifications, viz.

$$U_{ij} = -S_{ij}\,(e^{-\gamma^{-1}d_{ij}}) \qquad (1)$$

where U_{ij} is the potential of interaction of group i with group j in water, d_{ij} is the distance between i and j, and S_{ij} and γ are empirically determined parameters. The strength of the interaction at any given distance depends upon a prefactor S_{ij} determined by the identity of the interacting groups. For the AIPD modifications, the prefactor S_{ij} was taken to be the free energy of hydration or transfer of atoms i and j for a given conformation, as determined by an empirical solvent accessible hydration model. The details for calculating this prefactor will be described below. For the UISC modification, S_{ij} was taken to be the sum of the *experimentally* measured[14] free energies of transfer of group i and j, i.e., the free energy averaged over *all* conformations rather than that for a *particular* conformation; one could perhaps *calculate* average solvation free energies of i and j from an envelope model, but we have not done so.

Both the envelope models and the AIPD modifications thereof rely on an accurate determination of the free energies of hydration for unit surface areas for the various types of atoms found in a protein. Similar considerations would apply to the unified residues of the UISC modification if the average free energies were derived from calculations with an envelope model. The atomic solvation parameters have typically been derived from experimental studies of hydration of small molecules. One of the difficulties in parameterizing a solvent model based on free energies of hydration or free energies of transfer of small molecules is that extrapolation to proteins is not assured. A characteristic of the small-molecule data used to parameterize such models as those of Ooi et al.[2] and Kang et al.[3-6] is the predominance of atoms with large solvent exposures. However, the deduced parameters may have limited applicability to conformations with reduced exposure to solvent. The average solvent exposures of the atoms are only slightly less when *extended* peptide conformations are used, as in the work of Eisenberg and McLachan.[1] In folded proteins, on the other hand, many of the atoms are completely or nearly completely buried. Our initial investigations of low-energy conformations of peptides in the absence of solvent suggested that similar small solvent exposures, for at least some types of atoms, could also be observed in some conformations of even relatively small peptides. These peptides also exhibit ranges of solvent accessibility uncommon in other small nonpeptide molecules. Consequently, we have attempted to derive atomic solvation parameters from free energies of hydration of peptides, rather than other small molecules. Because free energies of hydration for peptides cannot be measured directly, we have obtained this information from conformationally dependent solution properties, viz. NMR coupling constants measured in water. Application of the NMR-derived parameters to near-native confor-

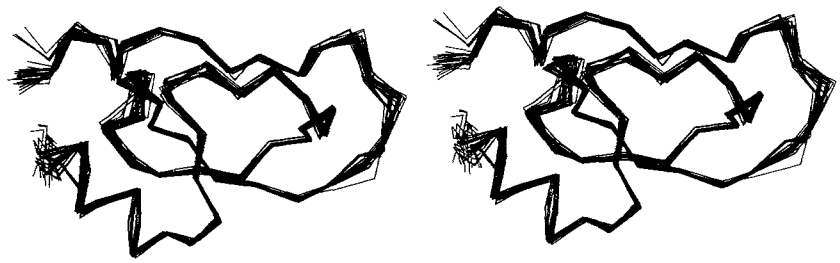

Fig. 1. Stereo view of the superposition of backbone atoms (N, C^{α}, C', and O) of the 39 structures of BPTI used in the tests reported here.

mations of bovine pancreatic trypsin inhibitor (BPTI) is described here.

In order to evaluate the effectiveness of the envelope models and their AIPD modifications as well as the UISC model, in the context of a protein, we have chosen a set of 39 near-native conformations of BPTI. These conformations were generated as local minima of the empirical conformational energy program for peptides (ECEPP/2)[15–17] by the electrostatically driven Monte Carlo (EDMC) algorithm of Ripoll and Scheraga.[18] The EDMC procedure involves an iterative deformation of the conformation, and then a minimization of the energy of the deformed structure. The local energy minimizations of the deformed conformations were carried out with the SUMSL algorithm (secant unconstrained minimization solver) of Gay.[19] The α-carbon traces for all of the 39 structures superimposed on each other are shown in Figure 1. The rms deviations as a function of residue number are shown in Figure 2.

Each of the members of the set of 39 BPTI conformations selected for this study has a lower potential energy in the absence of solvent than the rigid-geometry analog of the native structure. The rigid-geometry analog of the native structure[20,21] was generated by minimizing the objective function:

$$O = w\sum_{i=1}^{N_d} [d(i)^2 - d(i)^2_{xray}]^2 + U_{ECEPP} \qquad (2)$$

where U_{ECEPP} is the ECEPP/2 potential energy of the conformation, $d(i)$ is the distance between two atoms in the molecule, $d(i)_{xray}$ is the distance between the same two atoms in the X-ray crystallographic structure,[22] and w is a weighting factor. N_d is the number of pairwise distances restrained for the fit; a distance restraint was added for each 1–4 and 1–5 nonhydrogen atom pair. The fitting was carried out in stages with w very large (100 kcal/mol/Å^4) for the initial stage and zero for the final stage. The dihedral angles ω were constrained to 180°. If the thermodynamic hypothesis is correct, then these 39 conformations are of lower energy than the native analog only because of inadequacies of the function used to estimate the relative energies. One obvious inadequacy is that no solvation

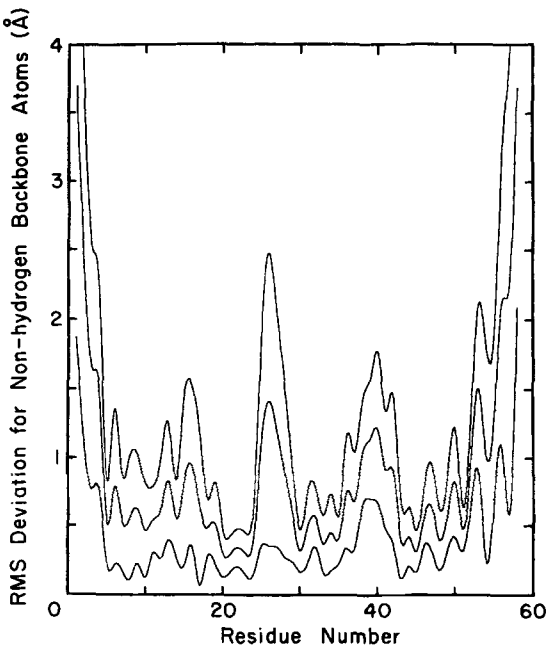

Fig. 2. The upper, middle, and lower curves show the maximum, average, and minimum rms deviations, respectively, for each C^{α} atom in the set. The curves have been smoothed by spline interpolation.

component was included in the ECEPP/2 potential used in the EDMC algorithm.

Figure 3 illustrates the ECEPP/2 energies and rms deviations from the crystallographically determined structure.[22] The conformation with the smallest rms deviation has the *largest* energy. It is clear that the ECEPP/2 energies of the 39 structures do not allow us to distinguish the native analog from the remaining 38 conformations nor is there any general *positive* correlation between ECEPP/2 energy and deviation from the native analog. On the contrary, there is a fairly strong *negative* correlation between ECEPP/2 energy and rms deviation from the native analog. This set of conformations is probably not particularly unusual, and such collections would likely be the general consequence of carrying out a search for the global minimum of the ECEPP/2

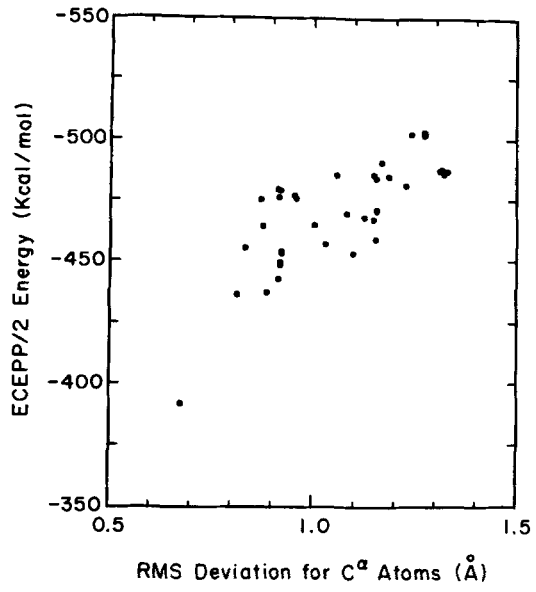

Fig. 3. The ECEPP/2 energies for the 39 conformations of BPTI used in this study vs. the C^{α} rms deviation from the crystallographically determined structure.[22]

energy while attempting to predict the conformation of a protein.

We sought a means to be able to distinguish the native conformation from among such a set of conformations. Because the solvation free energy was not included in the energy function that was minimized for these conformations, it would seem that a solvation potential might provide a reasonable discriminator for these conformations. We have evaluated the effectiveness of several solvation functions as discriminators by examining the correlation of the solvation free energy of each model with the rms deviation between the calculated and crystallographically determined structures.

All of the 39 conformations in the set are fairly close to the native conformation. Even though the rms deviations are not large, it is important to be able to establish criteria that can discriminate between such incorrect structures and the native one. The range of rms deviations of the 39 conformations used in this study lies outside the typical range of variation observed for the same protein in different crystalline environments and is well beyond the radius of convergence of conventional least-squares crystallographic refinement algorithms. Some of the conformations are probably beyond the radius of convergence of crystallographic refinement procedures that incorporate molecular dynamics and simulated annealing.[23] Several molecular dynamics simulations of small proteins in water have been carried out recently.[24-26] The average deviations from the high-resolution X-ray structures that were obtained in these simulations range from about 0.8 to 1.5 Å. On the other hand, structures obtained experimentally for the same proteins found in different crystalline environments indicate that the variation for the average structure is between 0.2 and 0.4 Å.[27] The much lower deviations obtained experimentally cannot be explained by assuming some bias in interpreting the experimental data. Brunger et al.[23] carried out refinements of X-ray structures starting from several different initial models. The refinements yielded structures with rms deviations between 0.09 and 0.58 Å. The larger deviations seen in the simulations could be the consequence of differences between the crystal (any crystal form) and solution, but the crystals typically contain at least 50% water. A more plausible explanation for the larger *average* rms deviations seen in the simulations is inadequacy of the simulation arising from sources such as inaccuracies of the potential energy function, insufficient length of simulation, periodic boundary conditions, or cutoff distances. The crystallographic root-mean-square displacement, μ, for an atom can be estimated with the Debye relationship[28] as $\mu^2 = 3B/8\pi^2$, where B is the isotropic temperature factor. The crystallographically determined temperature factors suggest that root-mean-square displacements of the main chain are significantly less than the average rms deviations from the native conformation for these 39 conformations. This can be seen by examining the differences between the C^{α} mean crystallographic displacements and corresponding average rms deviations from the native conformation. The hypothesis that the differences have a random distribution of signs must be rejected based on a nonparametric sign test[29] (the probability of incorrectly rejecting the null hypothesis is less than 0.001). Consequently, the 39 conformations of BPTI described here do not represent likely variants of the native structure in solution. They arose in the EDMC algorithm as low-energy conformations presumably because of the inadequacy of the ECEPP/2 potential when used in the absence of a solvation potential. However, there is a significant *concordance* between rms deviations from the native conformation and the crystallographic rms displacements (concordance coefficient 0.86). This implies that, although the magnitudes of the deviations are large, the deviations of the 39 structures from the native conformation occur principally in regions in which the crystallographic data indicate there is flexibility. These conformations represent a useful system with which empirical models can be evaluated for their ability to distinguish native from near-native conformations after minimization in the absence of solvent. A solvent model that successfully discerns the native among a set of possible conformations should prove useful in protein folding calculations, in general, and in the calculation of the conformations of homologous proteins in particular. This is most clearly evident

when one considers that low-energy conformations of surface loops, as determined in predictions of homologous structures, often fall within the range of deviations seen for the BPTI set used here.[30]

We have compared several empirical solvation models for their abilities to discern the native in a set of 39 near-native conformations, and to exhibit a correlation between the calculated solvation free energy and the rms deviation from the crystallographic conformation. With the exception of the UISC model, all of the models that we examine here rely on the same computational algorithm, i.e., solvent accessible surface area (the UISC model uses prefactors that are *experimentally* measured free energies of transfer from *n*-octanol to water rather than values calculated from a model that depends on solvent-accessible surface area).

METHODS

The Near-Native Structures

The set of near-native structures obtained as the local minima of four trajectories of the EDMC procedure is composed of 39 conformations of BPTI.[31] The rms deviations of this set from the native (crystallographic) structure[22] of BPTI lie in the following ranges: 1.06–1.94 Å, for all heavy atoms, 0.68–1.33 Å, for all C^{α} atoms, 0.77–1.36 Å, for all backbone heavy atoms, and 1.41–2.72 Å, for all side-chain heavy atoms.

All of these structures (Figs. 1 and 2) are characterized by a low total ECEPP/2 energy—even lower than the rigid-geometry version of the native structure.[20,21] The ECEPP/2 energy represents the energy of the protein in a medium of low dielectric constant. No correlation was found between this total energy and rms deviations for this set (see Fig. 3).

If the native conformation were a local minimum of a harmonic potential function and the near-native collection of structures were sampled within the same local minimum, then there would be a positive monotonic relationship between deviation from the native conformation and conformational energy. However, each of these 39 conformations represents a unique local minimum-energy conformation. As such, there is no reason *a priori* to expect that the energies of these conformations should be correlated with their deviations from the X-ray structure. However, we have found that several functions that express free energy of solvation are positively correlated with these deviations. Such positive correlations can have a practical role in structure prediction. If such a positive correlation exists, then it is possible to discern the correct structure with a greater probability of being correct than if no correlation existed. This does not necessarily imply that the free energies of solvation that yield the best correlations are those that are the most accurate estimators of free energy of solvation. This observation implies only that such solvation functions have the most utility for discerning the native conformation.

To be most useful for discerning the native conformation, there should be a monotonic relationship between the empirical solvation free energies and the rms deviations from the native conformation. We have employed a nonparametric coefficient of concordance[32] to express the monotonicity of the relationship between solvation free energies and rms deviations. This nonparametric statistic is the Kendall coefficient of concordance. It must be emphasized that, in searching for a correlation between solvation free energies and rms deviations, we do not alter the 39 conformations of BPTI in the original set. We are simply reordering them according to their solvation free energies and then examining whether the solvation free energy correlates with the rms deviations from the native structure and whether the structure with lowest solvation free energy corresponds to the native one.

UISC Approximation

We have investigated several modifications for expressing solvent effects in terms of a pairwise potential of mean force. For these modifications, we have assumed that hydration interactions decay exponentially with distance. This assumption is consistent with experimental observations.[7–13] The form that we have used is

$$U_{ij} = -(S_i + S_j)e^{-\gamma^{-1}d_{ij}} \tag{3}$$

where γ is the characteristic decay length, and d_{ij} is the distance between groups i and j. In the UISC modification, the groups are unified side chains and d_{ij} refers to the distance between the centroids of side chains i and j. S_i and S_j describe the propensities of i and j to be found in the interior or exterior of a protein. They could be free energies of hydration or free energies of transfer from nonpolar to polar environments. With this interpretation of the prefactor, the interaction potential given by Eq. (3) represents a distance-dependent correction to the free energy of solvation or free energy of transfer of the individual groups. In this simplified modification, the total hydration free energy required to hydrate a pair of groups, i and j, and bring them to a distance d_{ij} from each other would be given by

$$G_{ij} = U_{ij} + S_i + S_j \tag{4a}$$

or

$$G_{ij} = (S_i + S_j)\,(1-e^{-\gamma^{-1}d_{ij}}). \tag{4b}$$

At large separations, Eq. (4b) simplifies to $G_{ij} = S_i + S_j$. The quantities S_i and S_j may be regarded as the contributions to G_{ij} to bring groups i and j from the gas phase into the solvent, and the U_{ij} is then the contribution to bring them from infinite separation to a separation of d_{ij} in the solvent. In the UISC

approximation, the Ss are assumed to be constant for each amino acid independent of the conformation. For our UISC calculations, we took the prefactors to be the side-chain hydrophobicities reported by Fauchere and Pliska.[14] A similar approach was used by Fauchere et al.[33] However, rather than treating a side chain as a unified group, as we have for our UISC approximation, they used fragments of amino acids as the interacting groups and derived the prefactors from the hydrophobicity parameters of Rekker and de Kort.[34] Fauchere et al.[33] used a characteristic decay length, γ, of 1 Å. Previous experimental evidence suggests values between 2.6 Å[9] and 10 Å.[8] The significance of the wide range of decay lengths reported for various systems is beyond the scope of this work. For this work, we have used γ = 2.0 Å unless otherwise noted. This choice was based on a preliminary survey of concordance coefficients obtained using values ranging from 1 to 10. We have found that the best concordance was obtained with values from 1.5 to 2.5. To be consistent with previous studies,[12] we have assumed that the characteristic decay length is constant for all interactions. However, it should be noted that there is considerable experimental evidence to suggest that the decay length depends on the nature of the interacting surfaces, e.g., it has been shown[13] that the effective decay length would be different for surfaces formed of saturated and unsaturated hydrocarbons.

The Envelope Models and Their AIPD Modifications

In the framework of envelope models, the free energy of hydration of a conformation can be expressed as

$$G_i = \sum_{k=1}^{N} \sigma_k A_k \tag{5}$$

where A_k is the solvent accessible surface area of the kth atom obtained with our implementation of the analytical surface area algorithm of Connolly,[35] σ_k is the free energy of hydration per unit surface area for atom k, and N is the number of atoms of the amino acid i. In the formulation of Ooi et al.,[2] the free energy, G_i, is the free energy of hydration. For the model of Eisenberg and McLachlan,[1] G_i represents the free energy of transfer from n-octanol to water. If S of Eq. (4b) is taken to be G of Eq. (5) with the sum over all atoms of an amino acid, then

$$S_i = G_i = \sum_{k\epsilon i} \sigma_k A_k \tag{6}$$

where the sum is taken over all atoms k in amino acid i. Using Eq. (6), we can rewrite Eq. (4b) as

$$G_{ij} = [\sum_{k\epsilon i} \sigma_k A_k + \sum_{l\epsilon i} \sigma_l A_l](1 - e^{-\gamma^{-1} d_{ij}}). \tag{7}$$

Because $\Sigma_{k\epsilon i}^{N_i} \Sigma_{l\epsilon j}^{N_j} \sigma_k A_k = N_j \Sigma_{k\epsilon i} \sigma_k A_k$, Eq. (7) can be written as a sum over all pairwise interactions as

$$G_{ij} = \sum_{k\epsilon i} \sum_{l\epsilon j} [\sigma_k A_k / N_j + \sigma_l A_l / N_i](1 - e^{-\gamma^{-1} d_{ij}}) \tag{8}$$

where N_i and N_j are the numbers of atoms in residues i and j. The d_{ij} in Eq. (8) can be viewed as an effective distance between amino acids i and j that can be replaced by individual pairwise distances *between atoms* as follows:

$$G_{ij} = \sum_{k\epsilon i} \sum_{l\epsilon j} (\sigma_k A_k / N_j + \sigma_l A_l / N_i)(1 - e^{-\gamma^{-1} d_{kl}}) \tag{9a}$$

At large distances relative to γ, Eq. (8) and (9a) will be approximately equivalent. However, in a protein, intergroup distances may not in general be large relative to γ and Eq. (9a) may result in more realistic potentials. Equation (9a) means that there is a conformational dependence in both the exponential factor and in the prefactors (because the prefactors vary with the exposed surface areas of the groups). It is clear that Eq. (9a) will tend to Eq. (5) when two residues are far from each other (i.e., $\lim_{d_{ij} \to \infty} G_{ij} = G_i + G_j$).

The formulation of Eq. (9a) would lead to nonzero free energies for the interaction between a completely buried residue and a solvent-accessible one [as seen from Eq. (9a) by setting $\Sigma_{k\epsilon i} A_k / N_j = 0$, with $\Sigma_{l\epsilon j} A_l / N_i \neq 0$]. Because this would seem physically unreasonable, we have also employed an alternative form of Eq. (9a) that places a constraint on interactions of residues that have buried surface areas, viz.

$$G_{ij} = \begin{cases} \Sigma_{k\epsilon i} \Sigma_{l\epsilon j} (\sigma_k A_k / N_j + \sigma_l A_l / N_i)(1 - e^{-\gamma^{-1} d_{kl}}), \\ \quad \text{both } A_i^{\text{tot}} \text{ and } A_j^{\text{tot}} > 0; \\ 0 \quad \text{otherwise} \end{cases} \tag{9b}$$

where A_i^{tot} is the total accessible surface area of residue i.

The envelope models are characterized by free energies of hydration calculated using Eq. (5), and the corresponding models with the AIPD modification are characterized by Eqs. (9a) or (9b). They differ only by the inclusion of the pairwise exponential decay term in the AIPD modification.

We have tested several parameter sets for the envelope models and their AIPD modifications including those of Ooi et al.,[2] Eisenberg and McLachlan,[1] and four new sets that we have derived (SRA, ZRSS, SRFOPT, and JRF). The new sets were derived from two different sources: small molecule free energies of hydration (SRA, ZRSS, and SRFOPT) and NMR J-coupling constants for peptides (JRF). SRA, ZRSS, SRFOPT, and JRF differ only in the radii and atomic solvation parameters employed.

Atomic Solvation Parameters Derived From Small-Molecule Free Energies of Hydration

Scheraga and co-workers have previously reported two continuum models for solvation.[2-6] In

the model of Ooi et al.,[2] the free energy of solvation of a peptide is based on the solvent-accessible surface areas of the groups of the peptide in a given conformation as expressed in Eq. (5). The σ_k of Eq. (5) is an empirically determined weighting factor which applies to the kth type group in whatever molecular context it may appear. The second model which Scheraga et al. have applied to peptides is based on an analogous formalism that replaces A_k with V_k, the volume of the hydration shell exposed to solvent.[3-6] Given a set of compounds, C, the σ_k for each unique atom type of the set can be derived by minimizing the objective function Δ,

$$\Delta = \sum_{j \epsilon C} [G(j)_{\text{exp}} - G(j)_{\text{calc}}]^2 \tag{10}$$

where $G(j)_{\text{exp}}$ and $G(j)_{\text{calc}}$ are the experimental and calculated free energies of hydration for the jth compound. We have derived several parameter sets that minimize this function. These parameter sets differ in terms of the number and types of variables used in the minimization. Two approaches were adopted. In the first approach, hydrated radii were chosen as the sum of reasonable van der Waals radii plus the radius of a solvent sphere. For the work described here, the van der Waals radii were taken to be half of the minimum-energy distance of the OPLS Lennard–Jones type potential of Jorgensen and Gao.[36] We have investigated several values for the solvent-sphere radius, and a few of the results based on these radii are described here. The ZRSS parameter set was derived by using a solvent-sphere radius of 0.0. The SRFOPT parameter set was derived with a solvent-sphere radius of 1.4 Å.

In a first approach, only the σs were treated as variables in the minimization. In a second approach, the hydrated radii were also varied. The solvent-accessible surface areas, i.e., the values of A_k, were calculated using our implementation of the analytical surface area algorithm of Connolly.[35] We describe here one good (in terms of minimizing the objective function) parameter set derived with variable radii SRA and two parameter sets derived with fixed radii: ZRSS and SRFOPT. The parameter sets SRA, ZRSS, and SRFOPT were derived by minimizing the objective function expressed in Eq. (10) using the set of 158 neutral compounds used by Kang et al.[4,6] for deriving optimal volume parameters.

When the radii are fixed (as for the ZRSS and SRFOPT parameter sets), the objective function expressed in Eq. (10) is minimized by solving the following nonlinear least-squares system for Δ to determine the optimal set of σs:

$$\Delta = \min \|\mathbf{M}\boldsymbol{\sigma} - \mathbf{R}\|^2 \tag{11}$$

where $\mathbf{M}$ is a $(158 \times N_e)$ matrix such that $M(i,j)$ is the total solvent-accessible surface area of the jth type of group contained in the ith compound; $\boldsymbol{\sigma}$ is a vector of weights such that $\sigma(j)$ is the free energy of solvation per unit surface area for the jth group; $R(i)$ is the residual for the ith compound, i.e., the difference between the experimental and calculated [using Eq. (5)] free energy for the ith compound; and N_e is the number of types of groups used to describe the compounds (21 for SRA, ZRSS, and SRFOPT). For a few compounds in which more than one conformation is stable, e.g., the rotamers of n-butane, the matrix element $M(i,j)$ is

$$M(i,j) = \sum_{k=1}^{N_r(i)} \frac{e^{-E_t(k,i)/RT}}{Q(i)} S(j,k,i) \tag{12}$$

$$Q(i) = \sum_{k=1}^{N_r(i)} e^{-E(k,i)/RT} \tag{13}$$

$$E_t(k,i) = E_g(k,i) + G(k,i) \tag{14}$$

where $N_r(i)$ is the number of conformations of compound i, $E_g(k,i)$ is an estimate of the energy of the kth conformation in the absence of solvent, $G(k,i)$ is the free energy of solvation estimated with Eq. (5). $S(j,k,i)$ is the total area for the jth type of atom in the kth conformation of the ith compound. In general, however, the compounds were modeled by a single most stable conformation.

When the radii, as well as the σs, are allowed to vary (as for the SRA parameter set), the objective function Δ can no longer be minimized by solving a single least-squares system. To carry out the SRA optimization, the simplex minimizer AMOEBA[37] was used to optimize the radii. Each function evaluation carried out in the simplex minimization involved solving the above non-linear system for optimal $\sigma(j)$ at the current radii.

The radii and solvation parameters for the three sets are given in Table I. In all three cases, as with the OONS[2] parameter set where these parameters were applied to peptides, all residues were modeled in their uncharged form, i.e., there is only one nitrogen parameter regardless of whether it occurs in the backbone or in the side chains of the basic residues, and, correspondingly, the same parameter is used for both carbonyl and carboxylate oxygen. This constraint was necessary to keep the number of parameters to a minimum. The atomic radii derived as optimal in the SRA set are clearly nonphysical, with hydrated radii being smaller than the van der Waals radii for several of the atoms. We have investigated the performance of the SRA model in the BPTI test case only to see if such a test system can resolve differences between this model and ones with more physical meaning. The average absolute errors between the experimental and the computed free energies of hydration for the SRA, ZRSS, and SRFOPT parameter sets are 0.39, 0.54, and 0.61 kcal/mol, respectively. The average absolute error reported by Kang et al.[4] was 0.46 kcal/mol.

TABLE Ia. Radii and Solvation Parameters for the SRA Hydration Models

	Atom type	Envelope model SRA Radius (Å)*	σ†
1	Hydroxyl, carboxyl H	2.85	−0.0487
2	Amine, or amide H	2.85	−0.0487
4	Thiol H	2.85	0.0690
5	Aliphatic CH_3	0.946	0.0676
6	Aliphatic CH_2	0.946	0.0193
7	Aliphatic CH	0.946	0.00438
8	Aliphatic or alicyclic C	0.946	0.01691
9	Alicyclic CH_2 or CH	0.946	0.0190
10	Aromatic CH	0.946	−0.0110
11	Aromatic C	0.946	−0.137
12	Aromatic C of fused ring	0.946	−0.103
13	Aromatic CH of fused ring	0.946	0.2342
14	Carbonyl or carboxylic C	0.946	−0.104
15	N of primary amine	4.10	−0.0225
17	N of secondary amine	4.10	−0.0213
19	N of cyclic amine	4.10	−0.0254
20	N of aromatic system	4.10	−0.0234
22	N of amide	4.10	−0.0337
23	Ether or hydroxyl O	2.83	−0.0282
24	Carboxylic O	2.83	0.0394
25	Carbonyl O of ester	2.83	−0.0489
26	Amide carbonyl O	2.83	−0.0920
28	Thiol or sulfide S	7.37	−0.0020

*Hydrated radius, i.e., in the solvent-accessible model, the distance between the atom center and the center of the solvent sphere. This is half the minimum energy separation distance for the OPLS potential[36] plus 1.4 Å for the SRFOPT parameter set and plus 0.0 for the ZRSS set. The radii that are equal for the SRA parameter set were constrained in the optimization to be equal.

†Atomic solvation parameter (kcal/mol/Å²).

NMR-Derived Solvation Parameters

The NMR-derived parameters were the result of minimizing the following nonlinear objective function:

$$\sum_{i=1}^{13} [J^{obs}_{NHC^{\alpha}H}(X_i) - J^{calc}_{NHC^{\alpha}H}(X_i)]^2. \quad (15)$$

$J^{obs}_{NHC^{\alpha}H}(X_i)$ is the observed value of the vicinal coupling constant for the X residue of the Gly-Gly-X-Ala tetrapeptide in water as reported by Bundi and Wüthrich.[38] The sum was taken over the 13 neutral amino acids reported by these authors. The values for $J^{calc}_{NHC^{\alpha}H}(X_i)$ were obtained by calculating a Boltzmann averaged coupling constant using all low-energy conformations for a given tetrapeptide:

$$J^{calc}_{NHC^{\alpha}H}(X_i) = \sum_{k=1}^{N_c(i)} J^{calc}_{NHC^{\alpha}H}(k,X_i) e^{-E_i(k)/kT}/Q_i \quad (16)$$

$$Q_i = \sum_{k=1}^{N_c(i)} e^{-E_i(k)/kT} \quad (17)$$

$$E_i(k) = E_i(k)^{hyd} + E_i(k)^{ECEPP/2} \quad (18)$$

$$J^{calc}_{NHC^{\alpha}H}(k,X_i) = 7.9 \cos^2 \theta - 1.5 \cos \theta + 1.3 \sin^2 \theta \quad (19)$$

$$\theta = |\phi - 60.0| \quad (20)$$

TABLE Ib. Radii and Solvation Parameters for the SRFOPT and ZRSS Hydration Models*

	Atom type	Envelope model SRFOPT Radius (Å)	σ†	ZRSS Radius (Å)	σ†
1	Hydroxyl, carboxyl H	2.40	0.0498	1.00	0.312
2	Amine, or amide H	2.40	−0.00830	1.00	−0.00829
4	Thiol H	1.40	0.000453	0.00	0.000453
5	Aliphatic CH_3	3.59	0.00119	2.200	0.00984
6	Aliphatic CH_2	3.59	−0.00154	2.200	0.00505
7	Aliphatic CH	3.56	0.0321	2.161	0.0375
8	Aliphatic or alicyclic C	3.53	0.000448	2.13	0.00630
9	Alicyclic CH_2 or CH	3.59	−0.0301	2.19	−0.0440
10	Aromatic CH	3.50	−0.00510	2.10	−0.00837
11	Aromatic C	0.946	−0.137	2.10	−0.0996
12	Aromatic C of fused ring	3.50	−0.0917	2.10	−0.0942
13	Aromatic CH of fused ring	3.56	0.0219	2.16	0.0472
14	Carbonyl or carboxylic C	3.50	0.162	2.10	0.573
15	N of primary amine	3.22	−0.105	1.82	−0.289
17	N of secondary amine	3.22	−0.364	1.82	−0.637
19	N of cyclic amine	3.22	−0.228	1.82	−0.493
20	N of aromatic system	3.22	−0.171	1.82	−0.365
22	N of amide	3.22	−0.142	1.82	−0.439
23	Ether or hydroxyl O	3.12	−0.125	1.72	−0.329
24	Carboxylic O	3.06	−0.116	1.66	−0.409
25	Carbonyl O of ester	3.06	−0.151	1.66	−0.637
26	Amide carbonyl O	3.06	−0.138	1.66	−0.560
28	Thiol or sulfide S	3.39	−0.0216	1.99	−0.0716

*The numbering of the atom types is that of Kang et al.[6] Types 3, 18, 21, and 27 are used for charged species. No charged species were used in the present work.

†Atomic solvation parameter (kcal/mol/Å²).

$$E_i(k)^{hyd} = \sum_{l=1}^{l=N_g(i)} A_{ik}(l)\sigma_l \quad (21)$$

where $N_c(i)$ is the number of low-energy conformations of tetrapeptide i. The relationship between ϕ and coupling constant [Eq. (19)] is an empirical one derived by Ramachandran et al.[39] Although several such relationships have been proposed, they differ little from each other—especially in the low energy regions of the dihedral angle ϕ. $N_g(i)$ is the number of groups present in tetrapeptide i, and $A_{ik}(l)$ is the solvent-accessible surface area of the lth group of the ith tetrapeptide in the kth low-energy conformation as calculated by our implementation of Connolly's[35] analytical surface area algorithm.

We obtained the ensemble of tetrapeptide conformations by using the low-energy conformations for terminally-blocked single amino acid residues reported by Vásquez et al.[40] Low-energy terminally-blocked tetrapeptide conformations were generated by exhaustively combining the low-energy conformations for single amino acids, and then minimizing each resulting conformation with the ECEPP/2 potential function. For example, for the tetrapeptide GGGA, Vásquez et al. reported 7 conformations for blocked glycine and 7 for alanine, yielding $7\times7\times7\times7$ or 2401 initial conformations for this tetrapeptide, the energy of each of which was minimized. This was done for each of the 20 naturally occurring amino acids in the X position. The energies of a total of 315,516 conformations of the various peptides were minimized. Table II lists the number of conformations whose energies were minimized for each tetrapeptide. Most of the resulting conformations are of high energy. Of the low-energy conformations, many are nearly identical. Consequently, the resulting low-energy conformations, i.e., those for which the energy was less than or equal to 4 kcal/mol above the minimum-energy conformation for a given tetrapeptide [$e^{(-4000/RT)} = 0.00121$ at 300°], were then culled to remove redundancies using the clustering algorithm of Koontz et al.[41] as implemented by Dudek and Scheraga.[30] The clustering yielded a total of 2,202 unique conformations for the 20 types of tetrapeptides.

In order to illustrate the differences between the peptide and nonpeptide data sets in terms of solvent-accessible areas for the various types of atoms, Tables III and IV describe the two data sets. Table III lists the accessible areas observed for the various types of atoms in the 158 neutral compounds used in deriving the SRFOPT parameter set. Table IV illustrates the accessible surface areas for the various types of atoms comprising the 20 types of tetrapeptides. In order to facilitate comparison of the areas in the peptide and nonpeptide data sets, the areas for both were calculated with a common set of radii, although different solvated radii were used for the peptide and nonpeptide data sets in the actual parameterization. Tables III and IV illustrate the fundamental differences in the areas observed in the two types of data sets. In general, the minimum, maximum, and average areas are smaller for the peptide conformations than for the nonpeptide molecule data set. The differences in the two data sets is particularly striking for the backbone atoms N, C^{α}, C′, and O (types 22, 7, 14, and 26, respectively, in Table I). Relative to their respective means, the standard deviation of the atomic areas in these peptides are larger than in the nonpeptide compounds.

We fitted the experimentally observed coupling constants for only the uncharged amino acids that were reported by Bundi and Wüthrich.[38] Consequently, the nonlinear least-squares system that we solved contained only 13 equations. The assumption in this process is that the ensemble of low-energy conformations will consist of the same conformations both in the presence or absence of solvent but that the presence of solvent will modify the relative weights of the conformations in any ensemble-averaged property. The nonlinear least-squares system was optimized for the best set of atomic solvation parameters (i.e., σs) using a modified Levenberg–Marquardt algorithm[42,43] and a finite difference Jacobian as implemented in the IMSL[44] algorithm UNLSF. Seven different types of atoms were defined for the optimization. The initial σs were all set to 1 kcal/mol/Å^2. The derived σs and the identity of the atom types are given in Table V. The rms deviation between the calculated and observed coupling constants for the 13 tetrapeptides was 0.91 Hz. For comparison, the set of σs derived by Ooi et al.[2] gave an rms deviation of 1.75 Hz. The volume parameters of Kang et al.[4,6] yielded an rms deviation of 1.09 Hz. The ECEPP/2 energy without hydration results in an rms deviation of 1.89 Hz. The experimental error in the measurements was estimated to be 0.2 Hz.[38]

Minimization of Total Energy (ECEPP/2 + Solvation Free Energy)

We have implemented the analytical surface area algorithm described by Connolly[35] along with analytical derivatives of the surface area as a function of the internal coordinates of a molecule. We have incorporated this into ECEPP/2 so as to be able to minimize the conformational energy in the presence of water, i.e., the sum of the ECEPP/2 energy and the solvation free energy. The calculation of the analytical derivatives adds little time to the calculation of the surface area itself, yet dramatically speeds up a minimization process as compared to a numerical estimation of derivatives. This algorithm makes it feasible to apply such techniques as EDMC[18] to proteins in the presence of solvent (R.L.

TABLE II. Numbers of Conformations Used to Represent the Ensemble of Conformations for Each of the Tetrapeptides of Sequence Gly-Gly-X-ALA

X	Number minimized	Number final*
Gly	2401	95
Arg	72030	35
Glu	25382	112
Lys	61054	279
Gln	26068	173
Pro	1715	184
Ser	14406	46
Met	18179	74
His	12691	45
Asn	13377	105
Ala	2401	80
Phe	4802	102
Tyr	8918	56
Trp	10290	60
Asp	12005	164
Cys	12691	132
Val	3087	107
Leu	5831	91
Ile	4802	60
Thr	3430	202
Total	315560	2202

*The final number was obtained by removing all conformations with energies greater than 4 kcal/mol above the minimum-energy conformation for a given peptide and by removing redundant conformations. All minimizations were carried out with the neutral forms of the residues.

Williams, J. Vila, G. Perrot, and H.A. Scheraga, in preparation).

RESULTS

The various solvation models considered were evaluated with respect to two criteria: (1) Is there a monotonic relationship between the solvation free energy and the magnitude of the deviation from the native structure? (2) Does a given model discriminate the native analog from the near-native conformations?

The answer to the second question and a quantitative measure of the first criterion are presented in Table VI. The quantitative measure of monotonicity was the coefficient of concordance calculated according to the method of Kendall.[32] These coefficients can range between 0 and 1, with 1 signifying a perfect positive monotonic relationship.

Comparison of Main-Chain and Side-Chain Deviations for the 39 Conformations Irrespective of Solvation Energies

The results presented in Table VI are divided into three sets in order to examine the discriminatory power of the potentials for side-chain atoms separately from main-chain atoms. Because there is a possibility that there is a covariance between the side chains and main chains, it is necessary to evaluate the relationship of the main-chain to the side-chain deviations irrespective of any solvation models. The last three lines of Table VI indicate that there is little relationship between the main-chain and side-chain deviations of the 39 BPTI conforma-

TABLE III. The Atomic Areas Observed in the Data Set of Small Molecules From Which the SRFOPT Parameter Set Was Derived*,†

Type	No. obs.	Min. (Å^2)	Max. (Å^2)	Avg. (Å^2)	SD (Å^2)
1	55	6.61	10.99	9.08	1.32
4	3	0.00	0.00	0.00	0.00
5	276	43.59	113.57	79.40	9.39
6	258	22.23	69.16	44.42	9.39
7	31	4.21	25.91	15.54	5.16
8	91	34.11	77.46	49.90	9.16
9	7	9.63	21.16	17.91	4.22
10	200	2.51	78.12	38.30	7.52
11	41	1.25	8.15	5.26	1.42
12	32	3.76	9.72	5.37	0.99
13	5	13.59	58.58	23.95	19.38
14	31	2.19	25.57	13.45	4.60
15	10	43.53	66.76	52.76	8.40
17	5	10.22	26.41	14.15	6.88
19	8	16.99	44.73	24.14	9.76
20	13	6.23	24.68	18.11	5.84
22	5	24.42	59.06	44.65	15.14
23	120	2.61	56.91	28.21	15.76
24	17	5.50	46.54	14.88	15.04
25	11	25.96	43.67	33.44	5.39
26	19	21.05	46.30	32.43	7.04
28	4	31.59	87.12	61.01	25.80

*The definitions of the atom types are those given in Table I.
†The areas are solvent accessible areas (Å^2) calculated using the radii listed for the SRA model in Table I.

TABLE IV. The Atomic Areas Observed in the Peptide Conformations Used to Derive the JRF Parameter Set*,†

Type	No. obs.	Min. (Å^2)	Max. (Å^2)	Avg. (Å^2)	SD (Å^2)
1	2955	0.00	0.00	0.00	0.00
2	12124	0.00	11.82	3.25	3.47
5	3023	31.31	94.22	67.94	9.98
6	7136	2.98	61.22	42.32	12.14
7	4538	0.00	20.38	9.12	2.92
8	168	17.86	48.20	38.44	10.02
9	56	9.18	13.10	11.95	0.93
10	2063	7.13	63.35	28.96	9.69
11	480	2.17	3.50	2.33	0.37
12	120	2.68	8.05	3.80	1.33
13	202	5.30	10.53	7.65	2.07
14	9443	0.00	20.36	7.11	4.40
15	2246	3.16	44.21	28.86	8.31
17	164	0.32	8.70	6.54	2.03
20	90	1.09	25.24	14.55	8.06
22	7153	0.00	53.05	4.53	9.77
23	2642	1.61	49.80	38.94	12.79
24	313	8.49	45.49	34.53	10.77
26	9279	0.41	45.46	25.64	10.69
28	192	26.40	74.33	53.37	15.54

*The definitions of atom types are those given in Table I.
†The areas are solvent accessible areas (Å^2) calculated using the radii listed for the SRA model in Table I.

TABLE V. Atomic Solvation Parameters Derived From Vicinal NH-C$^{\alpha}$H Coupling Constants (JRF)

Group	σ (kcal/mol/Å^2)
C$^{\alpha}$*	0.216
Carbonyl or carboxyl C	−0.732
Aromatic C	−0.678
Nitrogen†	−0.312
Carbonyl or carboxyl O	−0.262
Noncarbonyl, noncarboxyl O‡	−0.910
S	−0.281

*Also CH_3, CH_2, and CH.
†Also NH and NH_2.
‡OH.

tions. This is seen directly by examining the coefficients of concordance among the three sets of deviations themselves. If the deviations of the main chain were related to those of the side chain, then there would be a high concordance coefficient between the two of them in contrast to the inconsequential concordance (0.593) shown in Table VI. This suggests that side chains have rotated independently of the main chain in the search procedure that generated these conformations.

The Relationship of Solvation Free Energy to Main-Chain and Side-Chain Deviations

Table VI also shows the results obtained for the coefficients of concordance between the empirical solvation free energy given by Eqs. (5), (9a), or (9b) and the rms deviation for the six different sets of parameters (used in both envelope and AIPD mod-

TABLE VI. Coefficients of Concordance for Various Solvation Models

	Concordance coefficient				Discrimination*	
Model	C$^{\alpha}$s	Main chain	Side chain	γ	Native	Analog
OONS†	0.803	0.832	0.391	2.0	No	No
OONS‡	0.810	0.833	0.397	2.0	No	No
OONS§	0.805	0.835	0.414	NA†††	No	No
E&M†	0.841	0.825	0.358	2.0	Yes	No
E&M‡	0.823	0.822	0.420	2.0	Yes	Yes
E&M§	0.804	0.814	0.542	NA†††	Yes	Yes
SRA†	0.777	0.831	0.428	2.0	No	No
SRA‡	0.760	0.806	0.457	2.0	No	No
SRA§	0.726	0.785	0.448	NA†††	Yes	Yes
JRF§	0.855	0.902	0.494	NA†††	Yes	Yes
JRF†	0.939	0.936	0.431	2.0	Yes	Yes
JRF‡	0.906	0.928	0.438	2.0	Yes	Yes
JRF§,**	0.870	0.908	0.776	NA†††	Yes	Yes
JRF†,**	0.878	0.917	0.788	2.0	Yes	Yes
F&P††	0.318	0.262	0.698	10	No	No
F&P††	0.515	0.487	0.880	2.5	Yes	No
ECEPP/2‡‡‡	0.115	0.091	0.512	NA	No	No
ZRSS†	0.886	0.926	0.612	2.0	Yes	Yes
ZRSS§	0.688	0.695	0.901	NA†††	Yes	No
SRFOPT†	0.820	0.843	0.419	2.0	Yes	Yes
SRFOPT§	0.823	0.849	0.420	NA†††	Yes	Yes
C$^{\alpha}$ RMS‡‡	1.00	0.975	0.549	NA	NA	NA
Bk. Ats. RMS§§	0.975	1.00	0.593	NA	NA	NA
S.C. Ats. RMS***	0.549	0.593	1.00	NA	NA	NA

*The "native" column contains the answer to the question: Does the lowest solvation-free-energy conformation have the lowest rms deviation from the crystallographically determined structure? The "analog" column is the answer to the question: Is the rigid-geometry native analog[20] the lowest solvation-free-energy conformation among the 39 conformations?
†An AIPD form. Solvation free energy calculated using Eq. (9a), i.e., distance dependence without constraint on buried surface areas.
‡An AIPD form. Solvation free energy calculated using Eq. (9b), i.e., distance dependence with constraint on buried surface areas.
§Solvation free energy calculated using Eq. (5), i.e., no distance dependence.
**All atoms were present for the surface area calculations, but only residues 3–56 were used in the calculation of free energy and concordance.
††Solvation free energy calculated using Eq. (3), i.e., UISC approximation.
‡‡Concordance of the C$^{\alpha}$ rms deviation of the 39 BPTI conformations with the C$^{\alpha}$ rms deviations, the main-chain rms deviations, and the side-chain rms deviations. These are concordances between pairs of deviations, i.e., completely independent of any solvation free energy.
§§Concordance of the rms deviation for all backbone atoms with the C$^{\alpha}$ rms deviations, the main-chain rms deviations, and the side-chain rms deviations. These are concordances between pairs of deviations, i.e., completely independent of any solvation free energy.
***Concordance of the rms deviation for all side-chain atoms with the C$^{\alpha}$ rms deviations, the main-chain rms deviations, and the side-chain rms deviations. These are concordances between pairs of deviations, i.e., completely independent of any solvation free energy.
†††Not applicable. The exponential decay constant γ is not applicable to envelope models, i.e., energy calculated using Eq. (5).
‡‡‡The concordance between rms deviation and the ECEPP/2 conformational energy in the absence of solvent.

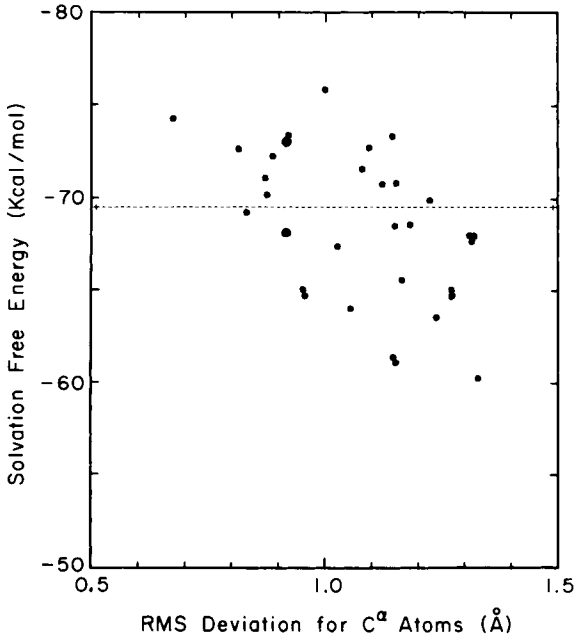

Fig. 4. Solvation free energy vs. rms deviation from the crystallographically determined structure for a model that depends on solvent accessibility. The dashed line represents the solvation free energy of the crystallographically determined conformation. Solvation free energy using the SRA parameter set in the distance-modified form of the solvation free energy described by Eq. (9a) [i.e., AIPD modificaton without the buried residue constraint specified by Eq. (9b)].

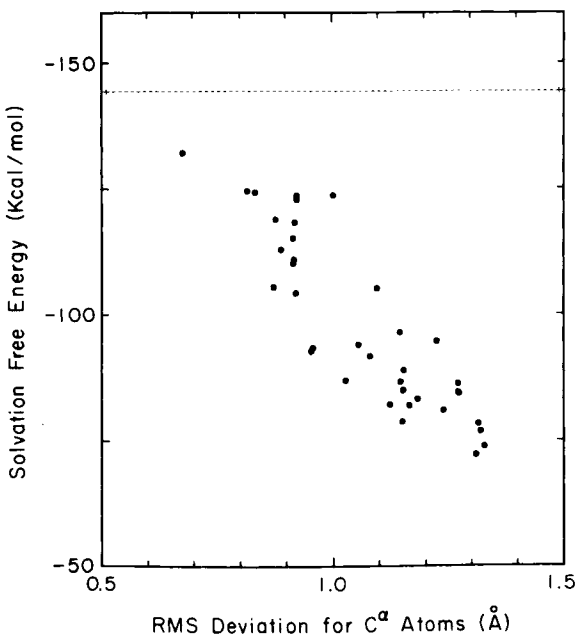

Fig. 5. Same as Figure 4, but for the NMR-derived (JRF) parameter set.

els), and the UISC model. Two of these correlations are illustrated graphically in Figures 4 and 5. The results for the UISC model [Eq. (3), summed over all pairs of residues] are given in Table VI for two different decay lengths, $\gamma = 2.5$ Å, and $\gamma = 10$ Å, using as prefactors the side-chain hydrophobicity scale of Fauchere and Pliska.[14] Although one might employ any other scale of hydrophobicity, we were particularly interested in comparing this residue–residue interaction modification with the atomic hydration models because Eisenberg and McLachlan[1] derived their atomic solvation parameters from the scale of Fauchere and Pliska.[14]

The solvation models evaluated here appear to sort into two categories: those that are good predictors of main-chain deviations, and those that are good predictors of side-chain deviations.

Category I

The first category includes the OONS, E&M, SRA, JRF, and SRFOPT models. These all yield free energies that have a high concordance with the deviations of the main-chain atoms and of the α-carbons, but they have a far worse concordance with the deviations for the side-chain atoms. For all of these AIPD and envelope models, the side-chain concordance coefficients are very low, and the hypothesis that there is no concordance could not be rejected (based upon a χ-squared statistic and a level of confidence of 0.05[32]). The ZRSS AIPD model is also of category I, but the envelope model corresponding to this parameter set is in category II. The better concordance with main-chain deviations observed for this parameter set, when applied in AIPD form, is probably the consequence of effectively extending the extremely short hydrated radii by use of the exponential distance dependence.

The apparently disparate behavior between the main chain and side chains for these models is probably a consequence of relatively freely rotatable groups on the side chains. If the end of a side chain is not within the hydration shell of any other group, then a solvent-accessible-surface-area-based model would impart no energetic penalty to rotation around the bond connecting this group to the rest of the protein. If this explanation were the cause of the observed behavior, we might expect similarly poor discrimination of deviations at the termini of the molecule. When the two residues at each of the N- and C-termini are removed from the calculation of the concordance coefficients, the coefficients for the side-chain atoms are much larger (as shown in Table VI for the JRF parameterization). However, the largest rms deviations for the 39 conformations of BPTI are seen at the termini, as illustrated in Figure 2 and there are considerable differences between two crystallographically determined conformations at the termini.[22] Consequently, poor concordance at the termini could also be indicative of a natural flexibility of these regions.

Category II

The ZRSS envelope free energies (Table VI) show a moderate concordance with the deviations of the main-chain atoms but a very high concordance with the side-chain deviations. The UISC model, F&P, is also a good predictor of side-chain deviations but a very bad predictor of the main-chain deviations. For the UISC model, this is perhaps not surprising given that the primary variables in the UISC model are the separations between side-chain centroids and that the results shown in the last three lines of Table VI indicate that the side-chain and main-chain deviations are largely independent of one another.

The very high concordance between the energies, as estimated with the ZRSS model, and the side-chain deviations is likely a specific consequence of having previously minimized the energies of the 39 conformations in the absence of solvent. This procedure results in the formation of hydrogen bonds involving the side-chain atoms with each other and with the main chain. There will be a loss of van der Waals surface area only when two atoms are closer than their minimum-energy position. This occurs principally when the atoms are hydrogen bonded. Consequently, the ZRSS parameter set is very sensitive to the formation of hydrogen bonds. Because the variation among the conformations in the main chain is considerably less than in the side chains, and because the main chain–main chain hydrogen bonding pattern is very similar among these conformations, the ZRSS parameter set is a poor predictor of the main-chain deviations.

Discrimination of the Native and Native Analog Conformations

The last two columns of Table VI indicate whether a given model is able to discern the native or native analog as the conformation of lowest solvation free energy. Figures 4 and 5 illustrate the energies and deviations for each of the 39 conformations. The dashed line on each of Figures 4 and 5 corresponds to the solvation free energy of the crystallographically-determined conformation. In both Figures 4 and 5, the point with lowest rms deviation corresponds to the rigid-geometry native analog of Vásquez and Scheraga.[20]

Envelope Models

The crystallographically determined conformation has the lowest solvation energy for all the envelope models except OONS. The native analog is the lowest solvation energy rigid-geometry conformation for all of the envelope models except ZRSS and OONS.

AIPD Models

The distance-dependent formulation did not affect the discrimination of the crystallographic structure for any of the models. For the E&M model, the native analog was not the rigid-geometry conformation with the lowest solvation free energy. For the remainder of the AIPD models, the native analog was the lowest-energy rigid-geometry conformation.

Distinctions Between AIPD and Envelope Models

It is clear from Table VI that the distance-modified solvent accessibility models (the AIPD formulations of E&M, OONS, SRA, ZRSS, and JRF denoted by footnotes † and ‡ in the table) described by Eq. (9a) or (9b) generally result in better main-chain concordances. The only envelope model showing no increase in main-chain concordance in the AIPD form is the SRFOPT model.

The difference in the concordance coefficients between the envelope model and the corresponding AIPD modification for the parameter set with the largest concordance coefficients, the JRF set, becomes smaller for the main chain when the termini are removed. This is seen in Table VI by comparing entries labeled with footnote ** with those not so labeled. Removing these termini increases the concordance obtained with the envelope model while decreasing the concordance of the AIPD modification. This suggests that the AIPD modifications estimate conformational energy at the termini better than elsewhere in the sequence, while the unmodified envelope models perform worse in these regions. This behavior at the termini is to be expected since the primary effect of an AIPD modification is to extend the ranges of interactions beyond the first hydration shell of the corresponding envelope model because the interaction energies as calculated by Eqs. (9a) and (9b) do not become zero beyond the distance at which the hydration shells no longer intersect. As the interaction range is extended, there are greater numbers of interactions that will be affected by rotation around any given bond. Consequently, free rotations around terminal bonds cannot occur without affecting the total energy of the protein.

Comparison of the New Solvation Models in the Role of Discerning Near-Native Conformations From Native

Regardless of whether one uses the distance-modified solvent accessibility model or the conventional solvent accessibility model, the new parameter sets presented in Tables I and V can be ranked with respect to their concordance coefficients for the main-chain atoms. This ranking is JRF > ZRSS > SRFOPT > SRA. The ZRSS model has very disparate behavior in the AIPD and envelope forms. In the AIPD form, it is similar to SRFOPT. As an envelope model it seems to be unusually sensitive to side-chain deviations.

Additional Comments on the Models

The following are some additional detailed evaluations of the various models, based on the data of Table VI.

UISC model

With $\gamma = 2.5$ in the UISC model, the concordance for the side-chain atoms is reasonably good, but poor for the backbone atoms. Furthermore, the structure of the lowest solvation free energy is *not* the one with the lowest rms deviation. Increase of γ to 10.0 for this model degrades the quality of the results, which scatter above and below the native structure.

OONS model

In the model of Ooi et al., either without or with introduction of the distance dependence, the concordance for the backbone atoms is reasonably good, but the results scatter above and below the native structure.

E&M model

The E&M model has concordances slightly higher than the OONS model both in envelope and AIPD forms. Furthermore, the X-ray structure is distinctively the lowest energy structure in both envelope and AIPD forms. Only the AIPD form does not discriminate the native *analog* as the lowest energy rigid-geometry conformation.

JRF model

The JRF model exhibits good concordance for the C^{α} atoms, with the structure of lowest solvation free energy being closest to the native structure, both with (Fig. 5) and without inclusion of distance dependence. This model appears to offer the greatest sensitivity to deviation from the native in the main chain. Without the termini, this model also has significant concordance with the side-chain deviations. This is particularly significant because the termini and especially the side chains of the terminal residues are probably not strongly constrained in the solution structure.

SRA model

The SRA envelope model (Fig. 4) exhibits a poorer concordance for the C^{α} atoms than do the SRFOPT, E&M, and JRF models, but the structure with the lowest rms deviation is among those of lowest solvation free energy, although not the lowest when distance dependence is included.

ZRSS and SRFOPT models

The ZRSS and SRFOPT models do not have especially large concordances between C^{α} deviations and free energies. Using the ZRSS model, the native analog has the second lowest solvation free energy while the native analog has the lowest solvation free energy for the SRFOPT parameter set. Both of these parameter sets distinguish the crystallographic conformation as having considerably lower solvation free energy than any other conformation in the set.

Concordance Between Total Energy and Deviation From the Native Conformation

The SRA, SRFOPT, JRF, and ZRSS parameter sets were derived to express the free energy of transfer from gas to dilute aqueous solution. In practice, they were intended to be added to the ECEPP/2 energy to yield a total conformational energy. The concordances obtained using the total of ECEPP/2 plus the envelope and AIPD models corresponding to these parameter sets are similar to the concordance observed for ECEPP/2 itself (data not shown). These total energies do not have a significant concordance with the rms deviations of either the side-chain or the main-chain atoms. As such, the *total* energy would be a poor discriminator for near-native conformations that have previously been minimized in the absence of solvent. Apparently, the addition of none of these solvation estimates to the ECEPP/2 energy is sufficient to overcome the *negative* correlation between the ECEPP/2 energy and rms deviation. However, several of the solvation functions by themselves have significant concordances with the rms deviations. This would suggest that the high concordance illustrated by a solvation function may not necessarily imply that the solvation function is a good estimator of the free energy of hydration of a conformation of BPTI. This implies only that such a function is useful for distinguishing the native from near-native conformations that have previously been minimized in the absence of solvent.

DISCUSSION AND CONCLUSIONS

There are three principal conclusions that can be drawn from the results presented here. First, the test system described here seems to differentiate reasonably among solvent models. The crystallographic conformation is the one of lowest solvation free energy for most models, and for several models, the rigid-geometry native analog is the rigid-geometry conformation of lowest solvation free energy. For the parameterization with the best concordance between rms deviation and solvation free energy, JRF, the rigid-geometry native analog is the conformation of lowest solvation free energy when applied in both envelope and AIPD forms. Second, the parameter set derived from an average solution property of peptides, the JRF parameter set, seems to yield the best concordances and to distinguish both the native and rigid-geometry native analogs as the conformations of lowest solvation free energy. Finally, for all of the parameter sets, the AIPD-modified models seem to yield better concordance with the rms deviations of the main chain than the corresponding envelope models.

The Relationship Between Solvation Models Applied in the BPTI Test System to the Performance of the Models in Estimating Free Energies of Hydration of Small Molecules

There is considerable variation in sensitivity to deviations from the native conformation among the various parameter sets. This variation would not necessarily be predicted from an analysis of the parameters in smaller nonpeptide systems. For example, as noted previously, the SRA set showed the best agreement with the experimental free energies of hydration for the 158 compounds used by Kang et al.,[4,6] although this set does not yield as high a concordance with the BPTI deviations as SRFOPT. The parameter sets SRFOPT and SRA were derived from the same 158 compounds. Examination of Table 1 illustrates the differences between these two parameter sets. The optimal hydrated radii of SRA do not seem physically reasonable, e.g., the hydrated radius of the carbon atom is smaller than a reasonable van der Waals radius. It is likely that the set of radii derived for SRA is optimal only in the small, rigid-molecule environment. That this model does not perform very well in the protein test case indicates that the test is able to discriminate a nonphysical model from more physically sensible models. It is perhaps understandable how such anomalous solvation parameters as the radii of the SRA set can arise when deriving parameters from the free energies of hydration of small molecules. When considering small, rigid molecules for which there is an experimentally measurable free energy of hydration, one encounters the difficulty that there is not as extensive a variation in solvent accessibilities for atoms among these compounds as there would be among the various conformations of a peptide or protein, as illustrated in Table III in comparison to Table IV. The solvent accessibilities for the atoms of the 158 neutral compounds used are relatively large, and the inaccessible surfaces arise principally because of neighboring covalently bonded atoms. In a protein, however, many atoms are entirely inaccessible to solvent and the differences in exposures of atoms are determined largely by noncovalent interactions. Because of these differences in the two types of data, it would appear that some caution should be exercised when transferring atomic solvation parameters from small nonpeptide molecules to peptide systems and that a good fit to experimental free energies of hydration for the nonpeptide data set does not guarantee that the parameter set will be useful for discriminating native from near-native conformations (the SRA parameter set was the best parameter set found for the small molecule data set).

Similar caution is suggested by the results of the ZRSS parameter set. This parameter set fits the nonpeptide thermodynamic data better than the SRFOPT set, but it shows very poor discrimination for main-chain deviations. In the envelope form of this model, very good side-chain discrimination was observed, but this is largely due to sensitivity to the nonnative hydrogen bonds in the collection of near-native BPTI structures.

Comparison of AIPD and Envelope Models

Comparison of the results for the envelope models and the AIPD modifications thereof, shown in Table VI, indicates that models that use atomic solvation parameters and atomic solvent accessibility are slightly more effective when applied in the distance-modified (AIPD) form than in the conventional (envelope) form. It also indicates that calculations of interactions with Eq. (9b) do not lead to a better concordance between solvation free energy and rms deviations, i.e., in this test system, setting all solvation interactions with buried residues to zero does not improve the AIPD formulations. The differences between the AIPD and envelope performances in this system become smaller when termini are removed. This may suggest that the AIPD formulation would not necessarily serve as a better predictor than an envelope model of deviation from native conformation in all systems.

Derivation of Parameters From Average Properties of Peptides in Water

The performance of the JRF parameter set suggests that parameter sets derived from a variety of conformations of peptides might be more appropriate for peptide hydration. The difficulty with this approach is that free energy-of-hydration information is not readily available for such systems.

We have outlined here a procedure for obtaining such a parameter set from an ensemble-averaged property, the NMR coupling constants. Any such conformationally dependent solution property that could be measured for small peptides could be used in a similar approach. An alternative would be to use free energies of transfer from a less polar solvent to water. This approach would, however, render the resulting solvation model nonadditive with conformational energy calculations carried out in the asence of solvent. The total free energy of a peptide in water can be written as

$$G_{\text{total}} = G_{\text{gas}} + \Delta G_{\text{gas}\rightarrow\text{octanol}} + \Delta G_{\text{octanol}\rightarrow\text{water}}. \quad (22)$$

Therefore, the assumption that one could add the conformational energy calculated in the absence of solvent to the free energy of transfer from octanol to water and obtain the hydrated free energy for a conformation is tantamount to assuming that the free energy of transfer from gas to octanol is zero. A similar assumption is implicitly made when free energy of solvation is added to the ECEPP/2 energy.[3-6] Be-

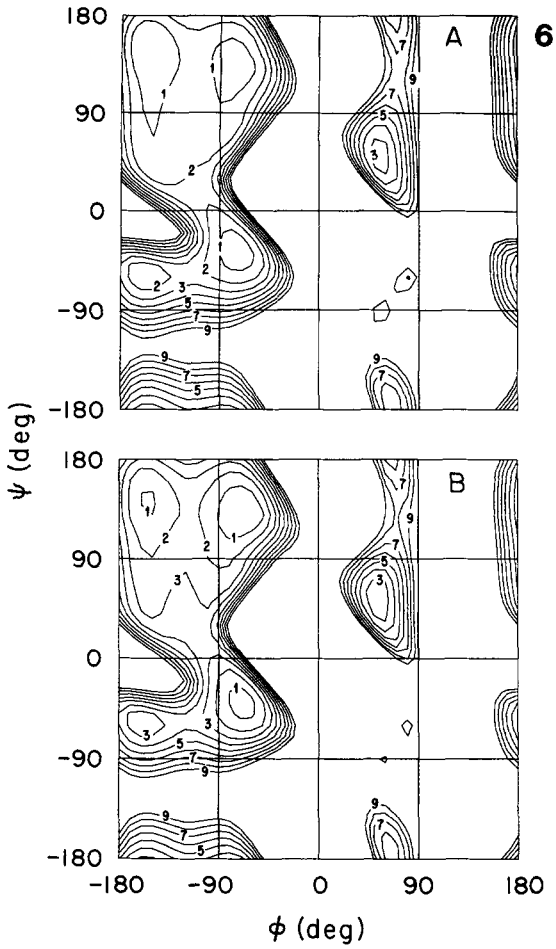

Fig. 6. (**A**) Adiabatic (ϕ,ψ) map for *N*-acetyl-*N'*-methyl-alaninamide computed from the sum of the ECEPP/2 energy plus the ZRSS envelope model solvation free energy [Eq. (5)]. At 10° increments in ϕ and ψ, all dihedral angles other than ϕ and ψ were allowed to vary in the minimization. Minimization of the total of ECEPP/2 plus hydration energy was carried out using analytical first derivatives of the energies with respect to the cartesian coordinates and the secant unconstrained minimization solver, SUMSL.[19] Contours are drawn at one kcal/mol intervals with the highest contour being at 9 kcal/mol above the minimum energy conformation in the map. (**B**) A partially relaxed ϕ–ψ map for terminally-blocked alanine. The map was calculated as for **A** except that only the ECEPP/2 energy was minimized. The solvent free energy was then added to the ECEPP/2 energy after minimizing in the absence of solvent.

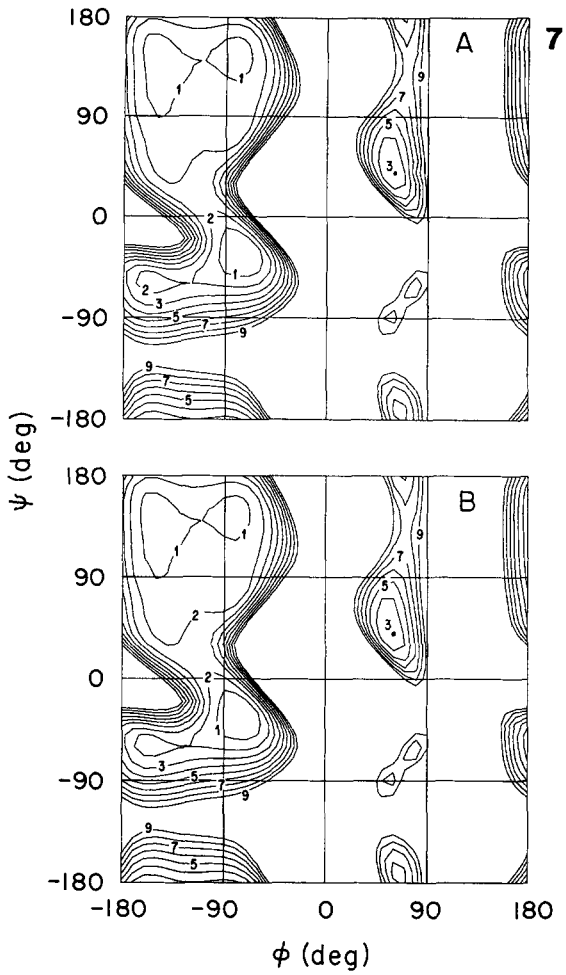

Fig. 7. (**A**) Same as Figure 6A, but for the SRFOPT parameter set. (**B**) Same as Figure 6B, but for the SRFOPT parameter set.

cause ECEPP/2 is not a vacuum potential, the assumption is made that the energy of transfer from vacuum to the low dielectric medium characteristic of ECEPP/2 is either zero or conformationally invariant. The ability of the NMR-derived parameter set to distinguish the native conformation and the high coefficients of concordance—particularly for the main-chain atoms—suggests that a parameter set derived from peptide data, albeit indirectly, might also perform better in the context of conformational energy calculations on proteins than small-molecule-based atomic solvation parameters.

Minimization of Conformational Energy in the Presence of a Continuum Representation of Water

Although the primary focus of this work is to demonstrate solvation free energy models that have utility in discerning the native among near-native conformations that have been energy-minimized in the absence of a solvent function, it might be questioned whether the general prescription of minimizing without solvent and then adding the solvation free energy for the minimized conformation is a reasonable approach. Earlier computational strategies that we employed[2,3–6] simply computed the solvation free energy for a peptide as a term to be added only after minimizing the energy in the absence of solvent. This strategy was adopted because of the tremendous computational effort involved in calculating the analytical volumes[45] or in calculating surface areas numerically.[2] It is conceivable that conformations minimized in the absence of solvent

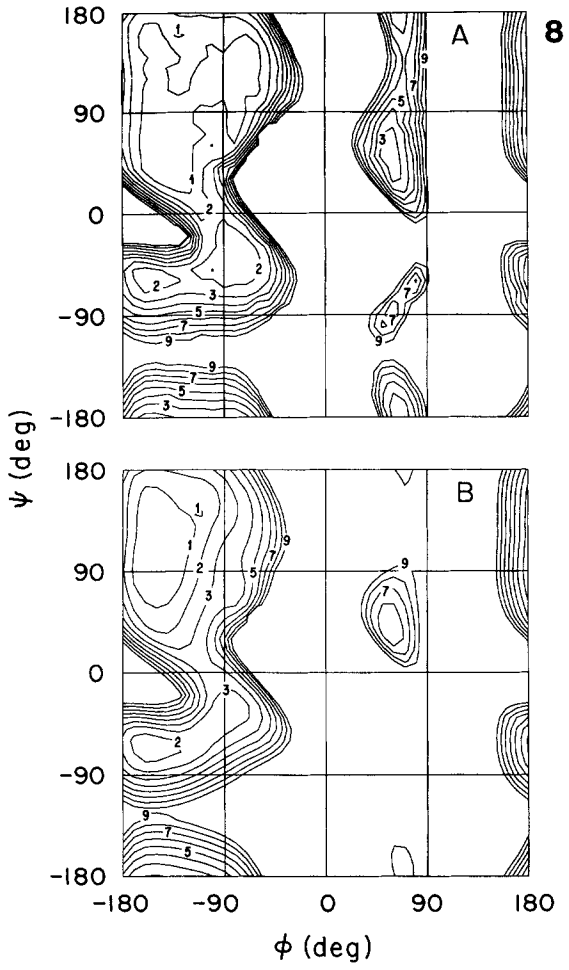

Fig. 8. (**A**) Same as Figure 6A, but for the JRF parameter set. (**B**) Same as Figure 6B, but for the JRF parameter set.

are quite different than those that would be obtained by minimizing in the presence of solvent. We have implemented an efficient parallel algorithm for analytically calculating solvent accessible surface areas and their derivatives with respect to dihedral angles.[46] This allows us to examine the validity of this approach by carrying out minimizations in the presence and in the absence of the solvation function. We are currently evaluating this procedure for BPTI, but we will describe here some results for a much simpler system.

Terminally-Blocked Alanine

The example that we will treat is *N*-acetyl-*N'*-methyl-alanineamide, a terminally-blocked alanine. We calculated adiabatically relaxed ϕ–ψ maps for the ZRSS, SRFOPT, and JRF parameter sets. These maps are shown in Figure 6–8. For each parameter set, two maps are shown. The first map for each parameter set corresponds to an adiabatically relaxed map. This was calculated by fixing the dihedral angles ϕ and ψ and using the secant unconstrained minimization solver[19] to minimize the total energy (ECEPP/2 plus hydration free energy) with respect to the remaining variable dihedral angles. The second map shown for each parameter set was calculated by minimizing the ECEPP/2 energy in the absence of a solvent model for each point in the map, and then adding the solvation free energy for the minimized conformation. We will refer to this second set of maps as partially relaxed maps. The purpose of these maps was to help evaluate the usefulness of the partial relaxation procedure in a very small system. In addition, examination of these maps can help to suggest whether we can expect these solvation models, that we have heretofore used only to discriminate among the 39 near-native BPTI conformations, to be useful to calculate solution properties for other systems. This second aim can be achieved by asking whether the total energy for a given model is consistent with experimental observations of terminally-blocked alanine in water. The experimental data available for this molecule would suggest that more than one conformation is present in solution. NMR coupling constants[47] suggest that the population of conformations in solution is most likely dominated by conformations with $-70° > \phi > -80°$. Although there is little experimental evidence about conformational preferences for this molecule in solution, it seems reasonable that the conformations accessible to this molecule would be similar to those of alanine residues with no hydrogen bonds in proteins.[48] For each of the parameter sets, there is a strong similarity between the partially and fully relaxed ϕ–ψ maps. This would suggest that the partial relaxation procedure is a fairly safe expedience in this system.

Figures 6 and 7 illustrate the effects of the ZRSS and SRFOPT parameter sets. The two parameter sets produce maps that are very similar to each other. They both produce a shift away from the intramolecularly hydrogen-bonded minimum C_7^{eq}. These models also result in a broadening and flattening of the C_5 minimum. Both of these effects are consistent with the observed distribution of alanine residues in proteins. However, the average coupling constants calculated with the ZRSS and SRFOPT parameters are 7.0 and 7.1 Hz, respectively, as compared to an experimental value of 6.1 Hz.[47] Figure 8A and B illustrates the fully and partially relaxed maps for the JRF parameter set. A shift from the C_7^{eq} minimum is very pronounced. These maps seem to underestimate the population of α_R relative to the minimum energy C_5 region considerably. This underestimate is quantitatively evident by examining the average vicinal NMR coupling constant. The Boltzmann-averaged NH-C$^{\alpha}$H vicinal coupling constant for this map is 7.45 Hz.

It is clear that the JRF, ZRSS, and SRFOPT parameter sets produce maps more strongly dominated

by the ECEPP/2 potential. Because of this, the α' conformation at $\phi = -150°$, $\psi = -40°$ that appears in the ECEPP/2 maps but is never observed in proteins is present in the ZRSS, SRFOPT, and JRF maps.

The Shape of the Conformational Energy Hypersurface in the Vicinity of a Local Minimum

Most of the solvation free energy functions that we have examined do have a positive monotonic relationship with rms deviation from the native conformation. This implies that one might generate conformations by minimizing energy in the absence of solvent and choose from among a group of similar low-energy conformations the most likely native conformation on the basis of such a solvent free energy function. In addition to this empirical utility, the observed positive monotonic relationships could have implications about the nature of the conformational energy hypersurface of a protein. Levitt et al.[49] and Nishikawa and Gō[50] have carried out normal mode analyses of BPTI. They both found that normal modes (especially low-frequency modes) were largely independent of conformation for local minimum-energy conformations near the native conformation. This suggests that although the conformational energy hypersurface is anharmonic as indicated by the presence of many local minima near the native conformation, these minima lie within a broad depression of the hypersurface. Examination of the current set of 39 conformations of BPTI indicates that the potential energy hypersurface may have a similar shape in the vicinity of the minimum. This can be seen by calculating the concordance between ECEPP/2 potential energy and rms deviation from the *lowest energy* conformation rather than from the native conformation. In contrast to the *negative* correlation illustrated in Figure 3, the concordance coefficient for this relationship is 0.82. As indicated in Table VI, there is a significant positive concordance between the JRF solvent free energies and the deviations from the native conformation. There is a *negative* correlation (concordance ceofficient 0.17) between rms deviations from the ECEPP/2 minimum-energy conformation and the JRF free energies. These observations suggest that minimizing in the absence of solvent results in local minima clustered around a regional minimum that is displaced from the X-ray conformation. The JRF solvent free energies have an apparent local minimum in the vicinity of the X-ray conformation. It is perhaps not surprising that the JRF solvent free energies have this property. They were derived as adjustments to the ECEPP/2 conformational energies minimized for a group of peptides in the absence of solvent. This feature of the JRF solvation free energies does not imply that they would be the most useful solvation parameters if one were to minimize the total conformational energies (solvent plus ECEPP/2 energies).

Use of Solvation Models in Searches for the Global Minimum-Energy Conformation for Polypeptides

We have found two important properties of the SRFOPT and JRF models and especially their corresponding AIPD forms. First, because they appear to have the ability to identify misfolded structures they may be very useful to evaluate computer-generated structures and can be incorporated into an existing empirical energy function, such as ECEPP/2. Second, from a computational point of view, all of these solvation models have approximately the same complexity: considerably simpler than simulation with discrete water molecules but more complex than an empirical pairwise energy calculation (a typical minimization of the total of ECEPP/2 plus an envelope solvation model requires about 10 times as long as minimization of the ECEPP/2 energy alone).

The utility of such models in protein folding will probably be realized when combined with other computationally simpler methods. A reasonable prescription for searching the conformational energy surface for a protein in the presence of solvent would probably be to conduct a rough search using a computationally simple solvation function such as that in the UISC form. This could then be followed by calculations of intermediate complexity involving envelope or AIPD modifications thereof for low-energy regions. The last stages in such a hierarchical approach to structure calculation would probably involve simulation in a bath of explicit water molecules. Work is currently in progress to build such a hierarchical scheme into the global search procedures that Scheraga and co-workers have developed. In addition, we are in the process of testing models combining envelope and AIPD formulations with reaction field methods to describe more realistically the influence of charged amino acids on protein structure.

ACKNOWLEDGMENTS

We thank D.R. Ripoll for providing us with the coordinates of the 39 EDMC-generated conformations of BPTI, M.J. Dudek for providing us with his implementation of the Koontz clustering algorithm, and G. Némethy and Y.K. Kang for their comments on the manuscript. This work was supported by Grant GM-14312 from the National Institute of General Medical Sciences of the National Institutes of Health, and by Grant DMB84-01811 from the National Science foundation. J.V. was supported by a fellowship from Consejo Nacional de Investigaciones Cientificas y Tecnicas (CONICET) of Argentina (1988–1990). R.L.W. was supported by an Individual National Research Service Award from the Public Health Service. (Grant F32 GM-12111) (1987–1990).

The computations were carried out at the Cornell National Supercomputer Facility, a resource of the Cornell Center for Theory and Simulation in Science and Engineering, which receives major funding from the National Science Foundation and IBM corporation, with additional support from New York State and members of the Corporate Research Institute.

NOTE ADDED IN PROOF

Improvements of the algorithm for calculating the analytical surface areas and their derivatives with respect to atomic coordinates[51] have resulted in dramatic increases in computational speed. With these improvements, it has been possible to carry out searches of the conformational space in the vicinity of the conformation of the rigid-geometry native analog. For the JRF, SRFOPT, and OONS parameter sets, these searches yielded conformations with lower *total* energy (ECEPP + Solvent) and lower rms deviations from the BPTI4 x-ray-determined conformations[46] than did a search that started from the rigid-geometry native analog without inclusion of solvation.

REFERENCES

1. Eisenberg, D., McLachlan, A.D. Solvation energy in protein folding and binding. Nature (London) 319;199–203, 1986.
2. Ooi, T., Oobatake, M., Némethy, G., Scheraga, H.A. Accessible surface areas as a measure of the thermodynamic parameters of hydration of peptides. Proc. Natl. Acad. Sci. U.S.A. 84:3086–3090, 1987.
3. Kang, Y.K., Némethy, G., Scheraga, H.A. Free energies of hydration of solute molecules. 1. Improvement of the hydration shell model by exact computations of overlapping volumes. J. Phys. Chem. 91:4105–4109, 1987.
4. Kang, Y.K., Némethy, G., Scheraga, H.A. Free energies of hydration of solute molecules. 2. Application of the hydration shell model to nonionic organic molecules. J. Phys. Chem. 91:4109–4117, 1987.
5. Kang, Y.K., Némethy, G., Scheraga, H.A. Free energies of hydration of solute molecules. 3. Application of the hydration shell model to charged organic molecules. J. Phys. Chem. 91:4118–4120, 1987.
6. Kang, Y.K., Gibson, K.D., Némethy, G., Scheraga, H.A. Free energies of hydration of solutes molecules. 4. Revised treatment of the hydration shell model. J. Phys. Chem. 92:4739–4742, 1988.
7. Parsegian, V.A., Fuller, N., Rand, R.P. Measured work of deformation and repulsion of lecithin bilayers. Proc. Natl. Acad. Sci. U.S.A. 76:2750–2754, 1979.
8. Israelachvili, J., Pashley, R. The hydrophobic interaction is long range, decaying exponentially with distance. Nature (London) 300:341–342, 1982.
9. Pashley, R.M., McGuiggan, P.M., Ninham, B.W., Attractive forces between uncharged hydrophobic surfaces: direct measurements in aqueous solution. Science 229:1088–1089, 1985.
10. Israelachvilli, J. Solvation forces and liquid structure, as probed by direct force measurements. Acc. Chem. Res. 20: 415–421, 1987.
11. Cristenson, H.K., Gruen, D.W.R., Horn R.G., Israelachvili, J.N. Structuring in liquid alkanes between solid surfaces: Force measurements and mean-field theory. J. Chem. Phys. 87:1834–1841, 1987.
12. Israelachvili, J.N., Kott, S.J., Gee, M.L., Witten, T.A. Entropic orientational forces between surfaces in anisotropic liquids. Langmuir 5:1111–1113, 1989.
13. Marsh, D. Water adsorption isotherms and hydration forces for lysolipids and diacyl phospholipids. Biophys. J. 55:1093–1100, 1989.
14. Fauchere, J.L., Pliska, V. Hydrophobic parameters π of amino-acid side chains from the partitioning of N-acetyl-amino-acid amides. Eur. J. Med. Chem. Chim. Ther. 18: 369–375, 1983.
15. Momany, F.A., McGuire, R.F., Burgess, A.W., Scheraga, H.A. Energy parameters in polypeptides. VII. Geometric parameters, partial atomic charges, nonbonded interactions, hydrogen bond interactions, and intrinsic torsional potentials for the naturally occurring amino acids. J. Phys. Chem. 79:2361–2381, 1975.
16. Némethy, G., Pottle, M.S., Scheraga, H.A. Energy parameters in polypeptides. 9. Updating of geometrical parameters, nonbonded interactions, and hydrogen bond interations for the naturally occurring amino acids. J. Phys. Chem. 87:1883–1887, 1983.
17. Sippl, M.J., Némethy, G., Scheraga, H.A. Intermolecular potentials from crystal data. 6. Determination of empirical potentials for O-H . . . O=C hydrogen bonds from packing configurations. J. Phys. Chem. 88:6231–6233, 1984.
18. Ripoll, D.R., Scheraga, H.A. On the multiple-minima problem in the conformational analysis of polypeptides. II. An electrostatically driven Monte Carlo method-tests on poly(L-alanine). (1988) Biopolymers 27:1283–1303, 1988.
19. Gay, D.M. Algorithm 611. Subroutines for unconstrained minimization using a model/trust-region approach. ACM Trans. Math. Software 9:503–524, 1983.
20. Vásquez, M., Scheraga, H.A. Calculation of protein conformation by the buildup procedure. Application to bovine pancreatic trypsin inhibitor using limited simulated nuclear magnetic resonance data. J. Biomolec. Struc. Dyn. 5:705–755, 1988.
21. Vásquez, M., Scheraga, H.A. Variable-target-function and build-up procedures for the calculation of protein conformation. Application to bovine pancreatic trypsin inhibitor using limited simulated nuclear magnetic resonance data. J. Biomolec. Struc. Dyn. 5:757–784, 1988.
22. Wlodawer, A., Deisenhofer, J., Huber, R. Comparison of two highly refined structures of bovine pancreatic trypsin inhibitor. J. Mol. Biol. 193:145–156, 1987.
23. Brunger, A.T., Karplus, M., Petsko, G.A. Crystallographic refinement by simulated annealing: application to crambin. Acta Crystallogr. A45:50–61, 1989.
24. Van Gunsteren, W.F., Berendsen, H.J.C. Computer simulation as a tool for tracing the conformational differences between proteins in solution and in the crystalline state. J. Mol. Biol. 176:559–564, 1984.
25. Levitt, M., Sharon, R. Accurate simulation of protein dynamics in solution. Proc. Natl. Acad. Sci. U.S.A. 85:7557–7561, 1988.
26. Tirado-Rives, J., Jorgensen, W.L. Molecular dynamics of proteins with the OPLS potential functions. Simulation of the third domain of silver pheasant ovomucoid in water. J. Am. Chem. Soc. 112:2773–2781, 1990.
27. Chothia, C., Lesk, A.M. The relation between the divergence of sequence and structure in proteins. EMBO J. 5: 823–826, 1986.
28. Debye, P.J. Interferenz von Röntgenstrahlen und Wärmebewegung Annl. Phys. series 4. 43:49–95, 1914.
29. Kraft, C.H., van Eeden, C. "A Nonparametric Introduction to Statistics," chapter 9. New York: Macmillan, 1968.
30. Dudek M.J., Scheraga, H.A. Protein structure prediction using a combination of sequence homology and global energy minimization. I. Global energy minimization of surface loops. J. Comput. Chem. 11:121–151, 1990.
31. Ripoll, D.R., Piela, L., Vásquez, M., Scheraga, H.A. On the multiple-minima problem in the conformational analysis of polypeptides. V. Application of the self consistent electrostatic field and the electrostatically driven Monte Carlo methods to bovine pancreatic trypsin inhibitor. Proteins, 10:188–198, 1991.
32. Kendall, M.G. "Rank Correlation Methods." Charles Griffin & Co., 1962:94–100.
33. Fauchere, J.L., Quarendon, P., Kaetterer, L. Estimating and representing hydrophobic potential. J. Mol. Graph. 6: 203–206, 1988.
34. Rekker, R.F., de Kort, H.M. The hydrophobic fragmental constant; an extension to a 1000 data point set. Eur. J. Med. Chem. Chim. Ther. 14:479–488, 1979.

35. Connolly, M.L. Analytical molecular surface calculation. J. Appl. Crystallogr. 16:548–558, 1983.
36. Jorgensen, W.L., Gao, J. Cis-trans energy difference for the peptide bond in the gas phase and in aqueous solution. J. Am. Chem. Soc. 110:4212–4216, 1988.
37. Press, W.H., Flannery, B.P., Teukolsky, S.A., Vetterling, W.T. "Numerical Recipes. The Art of Scientific Computing." Cambridge: Cambridge Univ. Press, 289–293, 1986.
38. Bundi, A., Wüthrich, K. H-NMR Parameters of the common amino acid residues measured in aqueous solutions of the linear tetrapeptides H-Gly-Gly-X-L-Ala-OH. Biopolymers 18:285–297, 1979.
39. Ramachandran, G.N., Chandrasekaran, R., Kopple, K.D. Variation of the NH-C$^{\alpha}$H coupling constant with dihedral angle in NMR spectra of peptides. Biopolymers 10:2113–2131, 1971.
40. Vásquez, M., Némethy, G., Scheraga, H.A. Computed conformational states of the 20 naturally occurring amino acid residues and of the prototype residue α-aminobutyric acid. Macromolecules 16:1043–1049, 1983.
41. Koontz, W.L.G., Narendra, P.M., Fukunaga, K. A graph-theoretic approach to nonparametric cluster analysis. IEEE Trans. Comput. C25:936–944, 1976.
42. Levenberg, K. A method for the solution of certain non-linear problems in least squares. Quart. Appl. Math. 2: 164–168, 1944.
43. Marquardt, D.W. An algorithm for least-squares estimation of non-linear parameters. SIAM Journal on Applied Mathematics 11:431–441, 1963.
44. IMSL International Mathematics Subroutine Library, Inc., Houston, TX.
45. Gibson, K.D., Scheraga, H.A. Exact calculation of the volume and surface area of fused hard-sphere molecules with unequal atomic radii. Mol. Phys. 62:1247–1265, 1987.
46. Williams, R.L., Vila, J., Perrot, G., Scheraga, H.A. In preparation.
47. Madison, V., Kopple, K.D. Solvent-dependent conformational distribution of some dipeptides. J. Am. Chem. Soc. 102:4855–4863, 1980.
48. Roterman, I.K., Lambert, M.H., Gibson, K.D., Scheraga, H.A. A Comparison of the CHARMM, AMBER and ECEPP potentials for peptides. II. φ-ψ maps for N-acetyl alanine N′-methyl amide: Comparisons, contrasts and simple experimental tests. J. Biomolec. Struc. Dyn. 7:421–453, 1989.
49. Levitt, M., Sander, C., Stern, P.S. Protein normal-mode dynamics: Trypsin inhibitor, crambin, ribonuclease, and lysozyme. J. Mol. Biol. 181:423–447, 1985.
50. Nishikawa, T., Gō, N. Normal modes of vibration in bovine pancreatic trypsin inhibitor and its mechanical property. Proteins 2:308–329, 1987.
51. Perrot, G., Gibson, K.D., Cheng, B., Vila, J., Palmer, K.A., Nayeem, A., Maigret, B., Scheraga, H.A. MSEED: A program for the rapid analytical determination of accessible surface areas and their derivatives with respect to atomic positions. J. Comput. Chem., to be submitted.