PROTEINS: Structure, Function, and Genetics 14:110–119 (1992)

Empirical Solvation Models in the Context of Conformational Energy Searches: Application to Bovine Pancreatic Trypsin Inhibitor

Roger L. Williams, Jorge Vila, Georges Perrot, and Harold A. Scheraga
Baker Laboratory of Chemistry, Cornell University, Ithaca, New York 14853-1301

ABSTRACT **Continuum solvation models that estimate free energies of solvation as a function of solvent accessible surface area are computationally simple enough to be useful for predicting protein conformation. The behavior of three such solvation models has been examined by applying them to the minimization of the conformational energy of bovine pancreatic trypsin inhibitor.**

The models differ only with regard to how the constants of proportionality between free energy and surface area were derived. Each model was derived by fitting to experimentally measured equilibrium solution properties. For two models, the solution property was free energy of hydration. For the third, the property was NMR coupling constants. The purpose of this study is to determine the effect of applying these solvation models to the nonequilibrium conformations of a protein arising in the course of global searches for conformational energy minima. Two approaches were used: (1) local energy minimization of an ensemble of conformations similar to the equilibrium conformation and (2) global search trajectories using Monte Carlo plus minimization starting from a single conformation similar to the equilibrium conformation.

For the two models derived from free energy measurements, it was found that both the global searches and local minimizations yielded conformations more similar to the X-ray crystallographic structures than did searches or local minimizations carried out in the absence of a solvation component of the conformational energy. The model derived from NMR coupling constants behaved similarly to the other models in the context of a global search trajectory. For one of the models derived from measured free energies of hydration, it was found that minimization of an ensemble of near-equilibrium conformations yielded a new ensemble in which the conformation most similar to the X-ray determined structure PTI4 had the lowest total free energy.

Despite the simplicity of the continuum solvation models, the final conformation generated in the trajectories for each of the models exhibited some of the characteristics that have been reported for conformations obtained from molecular dynamics simulations in the presence of a bath of explicit water molecules. They have smaller root mean square (rms) deviations from the experimentally determined conformation, fewer incorrect hydrogen bonds, and slightly larger radii of gyration than do conformations derived from search trajectories carried out in the absence of solvent.

Key words: solvation, Monte Carlo, minimization, protein folding

INTRODUCTION

Solute–solvent interactions, like solute–solute interactions, can be estimated with the same force fields used to estimate conformational energies of proteins in the absence of solvent. However, the interactions between a protein molecule in solution and given water molecules are in general short-lived. Hence, it is only by averaging over a large number of configurations of a protein–solvent system, using Monte Carlo or molecular dynamics techniques, that we can obtain information about probable solution conformations of proteins. Simulations of proteins in baths of explicit water molecules can yield information regarding the dynamics of solutes and solvent. However, the computational demands imposed by such simulations preclude their applica-

Received August 22, 1991; revision accepted November 22, 1991.

Address reprint requests to Harold A. Scheraga, Department of Chemistry, Baker Laboratory, Cornell University, Ithaca, NY 14853-1301.

Present address of Roger L. Williams: Center for Advanced Biotechnology and Medicine, Rutgers University, Piscataway, NJ 08854.

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

Georges Perrot is on leave from I.B.M. France, Service 3393, 40 rue Dussoubs, 75002 Paris, France (1990).

tion to studies involving *extensive* searches of the conformational space of a protein.

We have developed several empirical solvation models[1] that are intended to mimic the effect of solvation. The assumption inherent in these models is that the principal contribution to hydration free energy derives from an interaction between surface-accessible regions of a protein and the surrounding solvent. We assume that this interaction is specific for each of the various types of atoms of the protein and that the total interaction of a protein with solvent can be expressed as a sum of the free energies of solvation for each of the component groups of the protein molecule.

Previously,[1] we examined the correlation between several empirical solvation estimates and deviations from an X-ray determined conformation for an ensemble[2] of near-native conformations. These studies suggested several models that appeared to correlate low solvation free energies with small deviations from an experimentally determined conformation. With the development of considerably faster methods to calculate analytical surface areas and their derivatives with respect to atomic coordinates,[3,4] it is now possible to carry out extensive searches of the conformational space of proteins by incorporating these continuum solvation models into empirical force fields. We describe here searches with several such models that indicate that the empirical solvation models in conjunction with estimates of the solute conformational potential energy are capable of arriving at most probable solution conformations that are in substantial agreement with experimentally determined protein structures.

METHODS

Envelope Hydration Models

We have recently described several continuum hydration models for conformational energy calculations.[1] Previously, these models were evaluated without reference to conformational energy functions for proteins in the absence of solvent. We have now evaluated the models by using the *total* conformational energy for the small protein bovine pancreatic trypsin inhibitor, BPTI. This total energy for a given conformation k, E^k_{tot}, of a protein is estimated as

$$E^k_{tot} = E^k_{hyd} + E^k_{int} \quad (1)$$

$$E^k_{hyd} = \sum_{i=1}^{N} A^k_i \sigma_i \quad (2)$$

where E^k_{hyd} is the free energy of hydration, E^k_{int} is the conformational energy of the protein in the absence of solvent, A^k_i is the solvent accessible surface area of the ith group (computed with the algorithm of ref. 4), σ_i is an empirically derived parameter for the ith group, and N is the number of groups in the protein. E^k_{int} was estimated by using the Empirical Conformational Energy Program for Peptides (ECEPP/2)[5-7]; this program includes a loop-closing potential to form the correct disulfide-bond pairings. The empirical hydration parameters, σ_i, were derived from two sources: experimentally measured free energies of hydration for small nonpeptide molecules or from NMR coupling constants measured for peptides in water. Derivation of both types of parameters has been described in detail elsewhere.[1]

Because the solvation free energy described by Eq. (2) is a function of surface areas, these models have been designated as "envelope" models. On the basis of the relationships that we have observed, during the course of the current work, between rms deviation from the X-ray-determined structure and the solvation free energies calculated with several hydration models, we selected three of these models for use in estimating the total free energies of Eq. (1). These are the SRFOPT, JRF, and OONS parameter sets.[1,8] The SRFOPT and OONS parameter sets were derived from experimentally measured free energies of hydration, while the JRF set was based on NMR coupling constants. The solvation parameters were described in detail by Vila et al.[1] The radii used in the derivation of the JRF and OONS parameter sets were reported by Ooi et al.[8] while the radii used for derivation of the SRFOPT parameter set were described by Vila et al.[1]

Minimization of Total Free Energies

For parameter sets that were derived independently of the ECEPP/2 potential function, SRFOPT and OONS, we have minimized E_{tot} as expressed by Eq. (1).

For the JRF parameter set, initial experience with minimizing the total free energy for large polypeptides indicated that the parameter set produced unrealistically large perturbations of the conformations (results not shown). However, if the JRF solvation free energy is simply added to the minimized ECEPP/2 energy, these drastic perturbations encountered in Monte Carlo minimization trajectories (see below) are avoided. Application of the JRF solvation potential in this manner is consistent with the procedure used to derive the parameters for that set.

The parameters of the JRF set were derived so as to minimize the sum of the deviations between an experimentally observed and calculated average solution property for a series of tetrapeptides. This property, an NMR coupling constant, was calculated as a Boltzmann average over an ensemble of probable solution conformations. The ensemble of solution conformations was initially obtained by minimizing the ECEPP/2 energies of each member of a set of combinatorially generated conformations in the absence of water. From this large ensemble of confor-

mations, only unique conformations with relatively low ECEPP/2 energies (i.e., local minima, with *assumed* identical widths of local potential wells) were retained in a smaller ensemble for optimization of the calculated coupling constants. The JRF parameter set was derived by using this smaller set of conformations. A nonlinear least-squares optimization was carried out to derive solvation free energies that, when combined with the ECEPP/2 energies, would give optimal Boltzmann weighting factors to minimize the difference between the observed and calculated coupling constants. To make the optimization computationally tractable, the smaller ensemble was not changed in the course of the parameter optimization, i.e., the solvation simply changed the relative probabilities of the members of the ensemble without changing their conformations.

Exploration of the Conformational Energy Hyperspace in the Vicinity of an Experimentally Determined Conformation

To investigate the behavior of the total energy estimates, using Eq. (1) and the SRFOPT, JRF, and OONS envelope hydration models, we have employed two approaches. The first approach was to generate an ensemble of conformations similar to the X-ray crystallographically determined conformations and to minimize E_{tot} for each of these conformations. If the total energy estimate is a reasonable approximation of the conformational free energy, then the lowest energy conformation should approach more closely the experimentally determined conformation. The ensemble of conformations that we used for this purpose consisted of the 39 conformations of the 58 amino acid protein bovine pancreatic trypsin inhibitor (BPTI) described by Vila et al.[1] and by Ripoll et al.[2] These conformations were generated by the electrostatically driven Monte Carlo (EDMC) procedure described by Ripoll et al.[2] Each conformation is a local minimum of the ECEPP/2 function in the absence of solvent.

The second approach was to carry out limited searches of the conformational space in the vicinity of the standard-geometry version[2] of an X-ray-derived conformation of BPTI, PTI5,[9] to determine if this structure represents a highly probable conformation as estimated by E_{tot}. These searches were carried out with the Monte Carlo plus minimization (MCM) procedure of Li and Scheraga,[10,11] using a modification by Ripoll and Scheraga.[12] This search procedure employs a Metropolis-like algorithm combined with conformational energy minimization to proceed through the conformational space along a trajectory of local minima.

In the original implementation of this algorithm,[10,11] the ith nonminimum energy conformation in the trajectory is generated from the $(i-1)$th local minimum-energy conformation by a random change, ε, of a randomly chosen dihedral angle. The ith local minimum-energy conformation is then generated from the ith nonminimum energy conformation by secant unconstrainted minimization.[13] The Metropolis criterion[14] is applied to the difference in energies between the ith and $(i-1)$th minimum-energy conformations. If the criterion is satisfied, then the ith conformation is retained, otherwise the ith conformation is replaced by the $(i-1)$th conformation. For application to BPTI, Ripoll et al.[2,12] modified the MCM procedure to change dihedral angles ϕ for residue j and ψ for residue $j-1$ simultaneously so as to produce a crank-like motion of a given peptide unit. The change in ϕ and ψ, $\Delta\phi$ and $\Delta\psi$, is given by $\Delta\phi = \varepsilon = -\Delta\psi$. The purpose of these crank-like motions was to prevent disruption of previously formed favorable interactions between the parts of the protein proximal and distal to the site of perturbation. An additional modification[2,12] of the MCM procedure that produced movements to align peptide-bond dipoles (i.e., the EDMC procedure) was not employed in the searches reported here because we wanted the sampling to be less biased toward native-like structures. The sampling that we employed resulted in more structural diversity in the trajectories and as such constitutes a better test of the effects of the solvation models.

RESULTS

Minimization of the Total Energy for the Ensemble of 39 Conformations

Table I describes the relationship between total free energy and the rms deviations with respect to an X-ray-determined structure of BPTI, PTI4,[15] for the ensemble of 39 conformations. These concordance coefficients express the monotonicity of the relationship and were calculated as described by Kendall.[16] The coefficients can take on values from 0 to 1, with 1 indicating a perfect positive, monotonic relationship. The coefficients are uniformly higher for conformations whose energies were minimized in the presence of solvent than in the absence of solvent. The highest of these concordance coefficients is 0.77 (for all side-chain atoms) and 0.74 (for all atoms), for the SRFOPT parameter set [concordance between the RMS deviation for all nonhydrogen atoms and total free energy as given by Eq. (1)]. The values of these concordance coefficients indicate considerable deviation from monotonicity both before and after minimization. However, minimization of the total energy does increase the concordance over that obtained using the total energies of the 39 conformations before minimization for the parameter sets for which the total energy was minimized (SRFOPT and OONS).

Figure 1 illustrates the relationship between total free energy and rms deviation for the SRFOPT parameter set both before and after minimization. The best relationship between rms deviation and energy for the set of 39 conformations is that for this pa-

TABLE I. Concordance Coefficients for the Relationship Between Total Free Energy and rms Deviation From the PTI4[15] Structure Determined by X-Ray Crystallography for Nonhydrogen Atoms in the Ensemble of 39 Conformations

Energy function	All atoms	C^{α}	All backbone atoms	All side-chain atoms	Lowest rms has lowest energy*	Rank of lowest energy†
ECEPP/2‡	NA§	NA§	NA§	NA§	NA§	NA§
ECEPP/2**	0.56	0.55	0.49	0.63	No	21
OONS‡	0.58	0.57	0.51	0.63	No	21
OONS**	0.65	0.61	0.59	0.69	No	3
SRFOPT‡	0.57	0.56	0.49	0.61	No	19
SRFOPT**	0.74	0.66	0.67	0.77	Yes	0
JRF‡	0.48	0.52	0.55	0.43	No	28
JRF**	NA††	NA††	NA††	NA††	NA††	NA††

*This is the answer to the question: Is the lowest-energy conformation the conformation with the lowest rms deviation with respect to PTI4.
†This is the number of conformations having lower rms *deviations* than the lowest *energy* conformation.
‡The energy function as expressed in Eq. (1) before minimization.
§In the EDMC procedure, from which the 39 conformations of BPTI were obtained, each *accepted* conformation corresponds to a local minimum of the ECEPP/2 potential function. Hence, any energy minimization of these conformations using the ECEPP/2 potential function (without considering hydration) would not alter the conformation. Consequently, the first two lines of the table would be identical for ECEPP/2 only.
**The energy function as expressed in Eq. (1) after minimization.
††The JRF solvation free energies were added to the ECEPP/2-minimized energies, i.e., the total energies were not minimized, to be consistent with the manner in which the JRF parameters were derived (see Methods).

rameter set, in which the conformation most similar to PTI4 is the conformation with lowest total free energy (Fig. 1b). As indicated in Table I, a similar tendency was found for OONS. For the OONS set, before minimizing E_{tot}, 21 conformations had lower rms deviations than the lowest total energy conformation of the 39. However, after minimization of the total free energy, only three conformations had lower rms deviations than the conformation of lowest total free energy. As noted in Methods, the JRF solvation was simply added to the previously minimized ECEPP/2 energy without minimization of E_{tot}. Using this total energy, it can be seen from Table I that the lowest total-energy conformation has among the *highest* rms deviations with respect to PTI4.

Search of Conformational Space in the Vicinity of the X-Ray Conformation

The MCM search started from the conformation indicated by the arrow in Figure 1a (this is the standard-geometry version of PTI5). Figure 2 illustrates the results of such a 200-conformation search carried out for each of the three parameter sets (ECEPP/2 plus hydration) and for the ECEPP/2 function alone in the *absence* of solvent effects. There is a tendency for each of the searches carried out in the *presence* of a solvation potential to yield conformations closer to the PTI4[15] X-ray conformation when the rms deviation is calculated over *all* atoms (residues 1–58). Figure 2 shows that while the searches tend to yield conformations more similar to PTI4[15] (Fig. 2a), they tend to diverge from an alternative X-ray-determined conformation, PTI5[9] (Fig. 2b). However, this divergence from PTI5 is less for the potential functions containing solvation than for the ECEPP/2 potential function alone. PTI4 and PTI5 are both X-ray-derived conformations of BPTI. The PTI4 structure is based on the protein crystallized from 2.25 M K_2HPO_4 while the PTI5 structure is based on the protein crystallized from 0.5 M K_2HPO_4. The differences in the structures represent the effects of different crystal packing and, to some extent, different ionic strength.

Table II lists the rms deviations for the final accepted conformation for each of the solvation parameter sets and for ECEPP/2 in the absence of solvent. As shown in Table II, when the two N- and C-terminal residues are removed from consideration, all of the rms deviations decrease with respect to both PTI4 and PTI5, implying that much of the difference among the conformations listed arises from these terminal residues. Although the larger rms deviation for trajectories carried out in the absence of solvent (last line of Table II) is less pronounced when only residues 3–56 are considered, the trajectories carried out in the presence of a solvation function still yield conformations that deviate less with respect to either of the X-ray structures than those obtained in the absence of solvent. The solvation parameter set for which the rms deviation of the final accepted conformation is closest to PTI4, when ter-

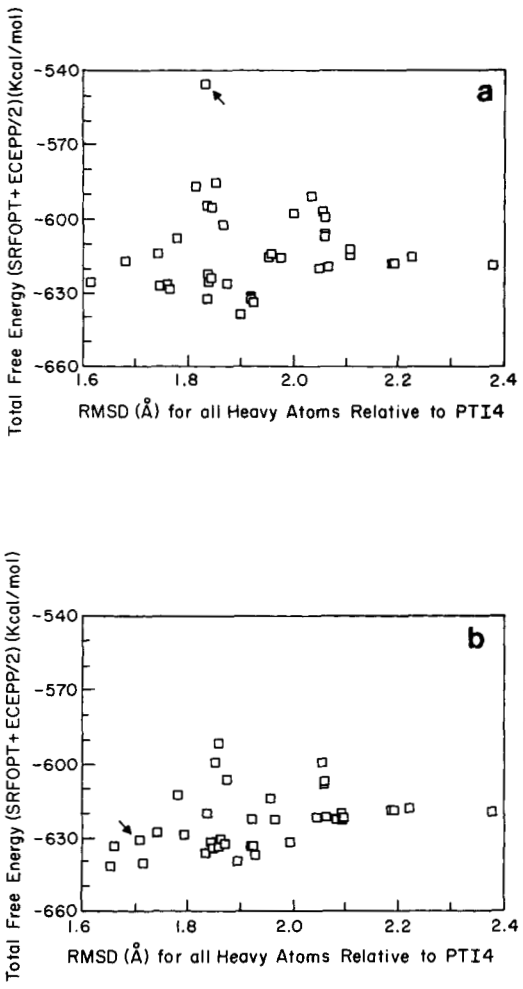

Fig. 1. **(a)** A scatterplot showing total free energy *before minimization* as calculated with Eq. (1) with the SRFOPT parameter set versus rms deviation from the PTI4 crystallographically determined structure[15] for the ensemble of 39 conformations generated by Ripoll et al.[2] The arrow indicates the position of the initial conformation used for the MCM conformational search illustrated in Figure 2 (this conformation is a standard-geometry version of PTI5). **(b)** The same relationship is illustrated *after* the total energy of each conformation has been minimized. The arrow indicates the position of the conformation (pointed to by the arrow in **a**) after minimization.

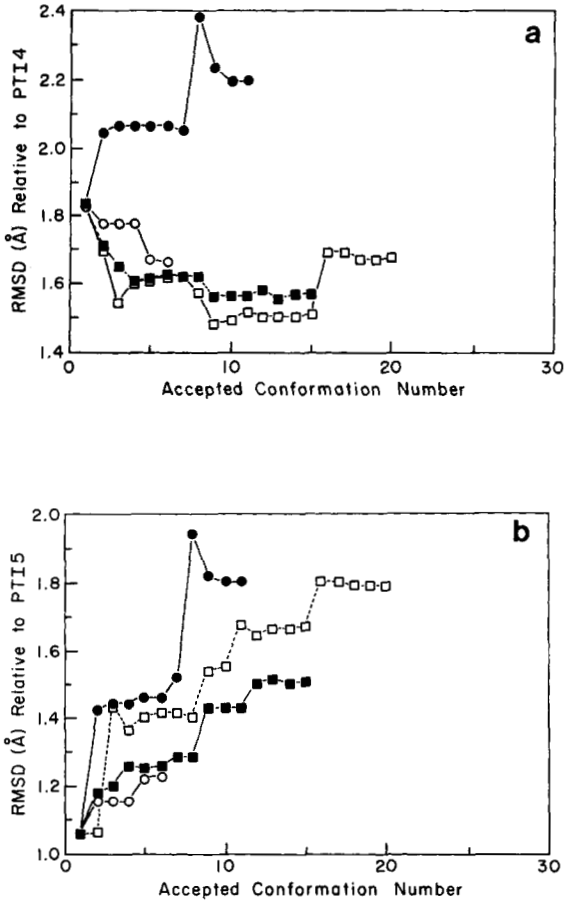

Fig. 2. Trajectories of the 200 step Monte Carlo minimization using E_{tot} as calculated in Eq. (1) for each of the parameter sets SRFOPT (■) and OONS (□). For JRF (○), the ECEPP/2 energy was minimized in each step of the trajectory, and the JRF free energy was then added to the resulting ECEPP/2 energy at each local minimum. Also illustrated is the EDMC trajectory (●) (carried out with the ECEPP/2 potential function in the *absence* of solvent), as described by Ripoll et al.[2] The accepted conformation number is indicated on the abscissa. **(a)** The ordinate is the rms deviation with respect to PTI4.[15] **(b)** The ordinate is the rms deviation with respect to PTI5,[9] with the same symbols as in **a**.

minal residues are included, is SRFOPT. When the terminal residues are excluded from the analysis, the trajectory yielding a conformation most similar to PTI4 is that carried out with the OONS parameter set. The trajectory that yielded a final accepted conformation most similar to PTI5 was JRF—either excluding or including the terminal residues from the analysis. Despite these differences, the three solvation models seem to behave similarly in the conformational energy search with respect to the overall rms deviation from the crystallographically determined conformations.

The three models also appear similar in terms of the component energies for the conformations generated in the trajectories. Each of the trajectories yielded a monotonic decrease in the total energy from start to finish. The differences in the total energies between the initial and final conformations were fairly similar for the three types of trajectories: 111, 116, and 126 kcal/mol for the JRF, SRFOPT, and OONS parameter sets, respectively. The dominant contribution to the change in energy for each trajectory was the change in the ECEPP/2 energy. The differences between the ECEPP/2 energies between the initial and final conformations were 76, 79, and 96 kcal/mol for JRF, SRFOPT, and OONS, respectively. The differences for the hydration component were 34, 37, and 29 kcal/mol for the same parameter sets. For OONS and SRFOPT, the ECEPP/2 energies decreased monotonically throughout the trajectories. However, the solvation

TABLE II. rms Deviation With Respect to PTI4 and PTI5 for All Nonhydrogen Atoms for the Last-Accepted Conformations of Each of the 200-Step Trajectories and for the Initial Conformation

	Rmsd with respect to PTI4 (Å)		Rmsd with respect to PTI5 (Å)	
	Residues 1–58	Residues 3–56	Residues 1–58	Residues 3–56
Initial	1.83	1.42	1.06	0.80
OONS*	1.68	1.32	1.79	1.14
SRFOPT*	1.57	1.38	1.48	1.11
JRF*	1.67	1.46	1.27	0.99
ECEPP/2†	2.19	1.50	1.80	1.18

*MCM trajectory.
†EDMC trajectory reported by Ripoll et al.[2]

component for OONS and SRFOPT decreased monotonically throughout only a portion of the trajectory: the first 7 steps for OONS and the first 11 steps for SRFOPT. The trajectory carried out with the JRF parameter set did not give monotonic decreases for either ECEPP/2 or hydration, although the total energy did decrease monotonically.

Ripoll et al.[2] reported that EDMC trajectories carried out in the absence of solvent were accompanied by a *decrease* in the radius of gyration for the C^{α}s with respect to both the PTI4 and PTI5 conformations. However, the radii of gyration tend to *increase* in the course of a 200-step MCM trajectory when *any* of the three parameter sets are used to estimate solvent effects. This effect was observed by examining the radius of gyration of the *last* accepted conformation in the 200 steps of each trajectory. The increase in radius of gyration with respect to PTI4 ranges from 0.5% for JRF to 2.0% for SRFOPT.

The hydrogen-bonding information for the last accepted conformation of the trajectories is summarized in Table III. The total numbers of errors in hydrogen bonding are similar for all of the solvation models. Furthermore, for each model, the total number of errors is only slightly less than the number of errors resulting from a trajectory carried out in the absence of solvent. For the trajectory carried out in the absence of solvent, there are 8 extra and 6 missing hydrogen bonds. OONS and SRFOPT also miss 6 hydrogen bonds while JRF misses 7. However, none of the solvent models results in as many extra hydrogen bonds as the trajectory carried out in the absence of solvent. Among the solvent models, the number of hydrogen bond errors involving only side chains is constant and involves the same errors present in the initial conformation.

The most notable differences among the solvent models is seen in the main chain–main chain hydrogen bonds and in the main chain–side chain (mixed) hydrogen bonds. All of the solvent models result in fewer main chain errors than in the initial conformation. The most dramatic difference is seen with the OONS model. On the other hand, this same model results in a larger number of mixed hydrogen bond errors. The OONS solvation model yielded a final accepted conformation (relative to PTI4) with 5 extra hydrogen bonds between side chains and backbone (5 extra, plus 2 missing hydrogen bonds, yields the 7 "mixed errors" indicated in Table III). The initial conformation had no extra hydrogen bonds of this type, and neither of the other solvent parameters nor ECEPP/2 produced such an increase. The main chain hydrogen-bonding errors are roughly equally divided between errors of absent and extra bonds for SRFOPT and JRF whereas the main chain errors for the initial model are principally extra in nature. The OONS model has yielded a reduction of the extra main chain hydrogen bonds by forming some mixed hydrogen bonds. The source of the additional mixed bonds with the OONS model will be discussed below. These results suggest that the solvation parameters for all of the models have sufficient energetic contribution to break a potential protein–protein hydrogen bond.

Figure 3 illustrates the rms deviation with respect to PTI4 for the starting conformation of the MCM trajectories and for the last accepted conformation for SRFOPT, the parameter set that yielded the lowest rms deviation with respect to the PTI4 conformation. Figure 3 shows that the decrease in rms deviation was largely confined to the region from residues 20 to 25 and 30 to 52 when the two N-terminal and two C-terminal residues are removed from consideration. Residues 45 to 52 comprise a region that shows both relatively large differences between PTI4 and PTI5 as well as somewhat larger temperature factors in PTI4 as compared to PTI5. This suggests that this region of the molecule is somewhat flexible.

DISCUSSION

Figure 2 indicates that there is a tendency for the trajectories carried out in the presence of the continuum solvation models to diverge from PTI5 and

TABLE III. Differences in Intramolecular Hydrogen Bonding With Respect to PTI4 in the Last Accepted Conformations of the MCM Trajectories*

Function	Total "errors"†	Total matches‡	Total hydrogen bonds	Side chain errors§	Main chain errors**	Mixed errors††
Initial‡‡	12	19	23	3	8	1
OONS§§	13	21	28	3	3	7
SRFOPT§§	12	21	27	3	6	3
JRF§§	13	20	26	3	6	4
ECEPP/2§§	14	21	29	4	5	5
PTI4 (exptl)	—	—	27	—	—	—

*These trajectories were carried out with the total free energies as defined for each of the three solvent parameter sets and for the EDMC trajectory carried out in the absence of solvent (ECEPP/2 potential). Also shown are the differences in hydrogen bonding between the initial conformation for the trajectories and the PTI4 structure. The last line gives the total number of hydrogen bonds for PTI4. A hydrogen bond was defined as a donor acceptor pair with H . . . acceptor atom distance less than 2.6 Å and donor-H . . . acceptor angle greater than 145°.
†An "error" can be either a hydrogen bond that is present in PTI4 but absent in the calculated conformation or a hydrogen bond present in the calculated conformation but absent in the crystallographic structure.
‡A hydrogen bond present in both the observed and calculated structure.
§Errors for hydrogen bonds involving only side chain atoms. An error is as defined in footnote †.
**Errors for hydrogen bonds involving only main chain atoms. An error is as defined in footnote †.
††Errors for hydrogen bonds between side chain and main chain atoms. An error is as defined in footnote †.
‡‡The initial conformation for all of the trajectories was the rigid geometry analog of PTI5 as described by Ripoll et al.,[2] indicated by the arrow in Figure 1a.
§§The OONS, SRFOPT and JRF potentials were used in MCM runs, and only the ECEPP/2 potential was used in an EDMC run.

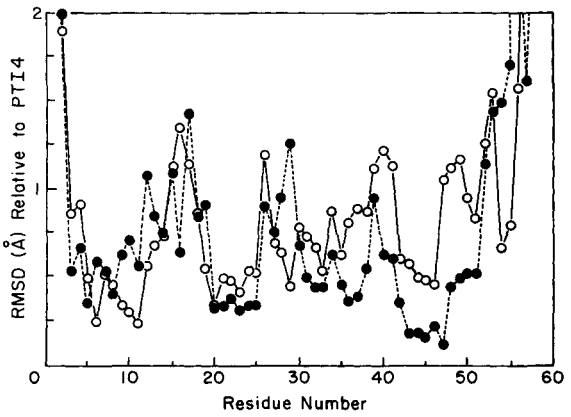

Fig. 3. Rms deviations for nonhydrogen backbone atoms for each residue versus residue number in the protein sequence. Residues 1–2 and 57–58 are omitted. The open circles represent the initial conformation of the MCM trajectory. The solid circles represent the last accepted conformation for the MCM trajectory using the SRFOPT parameter set.

to become progressively more similar to PTI4 while monotonically decreasing the conformational energy. However, the magnitude of the rms deviation is strongly influenced by the fact that the conformations of the two C-terminal residues in the final conformations resulting from the trajectories carried out in the presence of the solvent models are more similar to PTI4 than PTI5. Considering the remainder of the polypeptide exclusive of the C-terminus, the conformations resulting from the searches conducted in the presence of the continuum solvation models are overall somewhat more similar to PTI5 than to PTI4 (see Table II). Solution structures of BPTI were reported by Wagner et al.[17] These conformations, generated by using distance geometry and the observed and inferred distance constraints, show much greater deviations among themselves than do the deviations between the X-ray conformations PTI4 and PTI5. However, there is indication from the crystallographic comparison[9] that some of the variation between PTI4 and PTI5 derives from crystal packing.

Figure 3 illustrates that the initial conformation of the trajectory, a conformation very similar to PTI5 (rmsd of 0.77 Å for all backbone nonhydrogen atoms), differs considerably from the last accepted conformation (using the SRFOPT solvation parameters) in a contiguous stretch from residues 43 to 52. There is a relatively large difference between the two X-ray-determined structures in this region, although temperature factors for these residues are not especially large. This region is one of the better defined regions among the NMR conformations. Variation in conformation for this region within the collection of conformations reported by Wagner et al.[17] is somewhat less than the average variation for all residues. Furthermore, deviation of NMR-determined conformations in this region with respect to PTI4 is less than the average deviation found in the protein as a whole. It is possible that the NMR conformations are more similar to PTI4 in this region than to PTI5. This would suggest that the increased similarity to PTI4 in this region in the final accepted conformations of searches carried out in the pres-

ence of an empirical solvation model is a reflection of the ability of the solvation model to mimic the effect of water reasonably well.

Both the MCM trajectories and the minimizations of the ensemble of 39 conformations suggest that the three solvation models described here are useful components of an empirical conformational energy. Trajectories carried out with these potentials result in conformations more similar to the experimentally observed conformation than trajectories carried out in their absence. The effects of the solvation free energies on the most probable conformation encountered in the trajectories are similar in some respects to the effects that Levitt and Sharon[18] reported for simulation of BPTI in a bath of 2,607 water molecules. Levitt and Sharon carried out long molecular dynamics simulations of the protein both in vacuo and in the bath of explicit water molecules. For these simulations, the protein had an initial conformation identical to the PTI4 conformation.

Levitt and Sharon reported that simulations in the presence of solvent resulted in conformations that were slightly more expanded than simulations without solvent. The in vacuo simulations yielded an average radius of gyration about 5% smaller than that observed in the PTI4 structure or in the simulation with explicit water. This is the same trend seen with the MCM searches in the absence and presence of empirical solvation models. However, the shrinkage in the absence of solvent is not as large as seen in the in vacuo BPTI simulations of Levitt and Sharon.

The molecular dynamics simulations in a bath of explicit water molecules, by Levitt and Sharon, yielded conformations with fewer backbone hydrogen bonds than the in vacuo simulations and a greater number of *correct* backbone hydrogen bonds (with respect to PTI4). As indicated in Table III, fewer hydrogen-bonding errors were also observed in the MCM trajectories for all types of hydrogen bonds using the continuum solvation models described here. The tendency of the OONS solvation parameters to lead to a somewhat larger increase in the number of side chain–backbone hydrogen bonds is perhaps a consequence of the very slight negative free energy of hydration associated with the carbonyl oxygen as defined by these parameters. Examination of these extra hydrogen bonds indicated that all but one of these were between carbonyl oxygens of side chains (Asn or Glu) and backbone N–H. These extra hydrogen bonds are destabilized by the *increase* in solvation free energy when the carbonyl *oxygens* are buried in the formation of hydrogen bonds.

This destabilizing contribution is opposed by stabilization from two sources: decrease in potential energy arising from hydrogen-bond enthalpy and decrease in solvation free energy obtained by burying the nearby carbonyl *carbons*. Apparently, the destabilizing influence is not enough to offset the stabilizing ones. Such anomalies are likely to arise in any empirical solvation model derived from sets of small-molecule free energies of hydration due to the limited conformational repertoire represented by the compounds (in terms of solvent accessible area). The numbers of errors in backbone hydrogen bonds relative to PTI4 are larger for any of the last-accepted conformations in the MCM trajectories reported here than indicated by Levitt and Sharon for the simulation of BPTI with explicit water molecules. Trajectories using SRFOPT, OONS, and JRF, yielded 6, 3, and 6 total errors (sum of omitted and surplus hydrogen bonds) for the backbone–backbone hydrogen bonds as compared with 1 for the simulation of Levitt and Sharon. These differences may be due to the inherent oversimplification of the envelope solvation model or they may be attributable to a starting model that was far from the PTI4 conformation. The initial conformation for each of our trajectories was a conformation with 8 errors in backbone hydrogen bonding (arising from the imposition of standard geometry on the X-ray structure of *PTI5*) whereas the initial conformation for the simulations of Levitt and Sharon with explicit water molecules was the PTI4 X-ray conformation with *no* errors in hydrogen bonding.

Both PTI4 and PTI5 crystals contain ordered solvent. Wlodawer et al.[9] reported 60 and 63 solvent peaks for the PTI4 and PTI5, respectively. About half of these water molecules were in identical locations in the two crystal forms indicating very stable association with the protein. Four internal water molecules were observed in each of the two conformations. Implicit in the continuum solvation models employed here is the assumption that the interaction of the protein groups with solvent can be represented as an average over all configurations of the solvent. The observation that some water molecules form very long-lived interactions with the protein that are independent of the crystal environment indicates that this assumption is not entirely valid. However, the majority of the solvent in the crystal is not ordered on the time scale of the X-ray experiment.

Among the three solvation models, the JRF parameter set appears to perform the most poorly in terms of the rms deviations with respect to PTI4. However, all three models appear to improve the MCM trajectories relative to the trajectories carried out in the absence of solvent, and the differences among the models are slight.

The very different behavior of the JRF solvation parameters when applied throughout the course of a minimization as opposed to being added to the ECEPP/2 energy minimized in the absence of solvent suggests that derivation of parameters from solution properties of peptides should probably allow for conformational change of the ensemble of confor-

mations in the course of the parameter optimization. The MCM trajectories reported here indicate that the JRF model is comparable to the other solvation models only *when the trajectory is carried out in a manner consistent with the manner in which the parameter set was derived.* The manner in which the MCM trajectory was carried out for the JRF parameters was computationally considerably less demanding than the trajectories carried out for either the OONS or SRFOPT parameter sets.

The OONS and SRFOPT parameters sets were derived independently of ECEPP/2; however, the results of the MCM trajectories and the minimization of the ensemble suggest that the total energy, as given by using ECEPP/2 to respresent E_{int} in Eq. (1), leads to conformations that agree well with experiment. However, only the nonelectrostatic components of the ECEPP/2 function are truly gas phase potentials. The electrostatic component of the ECEPP/2 energy function is the potential in a medium of low dielectric constant. This contrasts with other potentials that have been developed to represent conformational energies in the gas phase, e.g., AMBER[19–21] or CHARMM.[22] The solvation potentials described here, with the exception of JRF, have been derived to represent the free energy of transfer to water with respect to a gas reference state. This suggests that these parameters may not be directly additive to the ECEPP/2 energy because the reference state of the ECEPP/2 electrostatic energy is a medium of low dielectric constant. The assumption made by using ECEPP/2 to estimate the electrostatic component of E_{int} in Eq. (1) is that the electrostatic free energy of transfer from gas to a medium of low dielectric constant is either negligible or conformationally invariant. A second assumption made in application of these solvation parameters to BPTI is that the conformation is independent of the state of ionization of the amino acids. This was inherent in our use of neutral residues to represent the protein. Because the experimentally observed structures occur in high-salt aqueous environments, it is likely that this assumption is not too unrealistic. We are now incorporating the solvation models reported here into calculations using gas phase potentials[23] and reaction field estimates of electrostatic energy of solvation for charged residues.

We are currently carrying out trajectories on conformations distant from the X-ray conformation[24] to see if the parameter sets described here are reliable in their intended role as a component of a conformational energy function useful for predicting protein folding on the basis of the thermodynamic hypothesis.

ACKNOWLEDGMENTS

We thank George Némethy for helpful comments on this 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 Grants DMB84-01811 and DMB90-15815 from the National Science Foundation. Support was also received from the National Foundation for Cancer Research. R.L.W. was supported by an Individual National Research Service Award from the Public Health Service (Grant F32 GM-12111) (1987–1990). J.V. was supported by a fellowship from Consejo Nacional de Investigaciones Cientificas y Tecnicas (CONICET) of Argentina (1988–1990). 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 its Corporate Research Institute.

REFERENCES

1. Vila, J., Williams, R. L., Vásquez, M., Scheraga, H. A. Empirical solvation models can be used to differentiate native from near-native conformations of bovine pancreatic trypsin inhibitor. Proteins: Struct. Funct. Genet. 10: 199–218, 1991.
2. 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: Struct. Funct. Genet. 10:188–198, 1991.
3. Perrot, G., Maigret, B. New determinations and simplified representations of macromolecular surfaces. J. Mol. Graphics 8:141–144, 1990.
4. Perrot, G., Cheng, B., Gibson, K. D., 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. J. Comput. Chem. 13:1–11, 1992.
5. 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.
6. Némethy, G., Pottle, M. S., Scheraga, H. A. Energy parameters in polypeptides. 9. Updating of geometrical parameters, nonbonded interactions, and hydrogen bond interactions for the naturally occurring amino-acids. J. Phys. Chem. 87:1883–1887, 1983.
7. Sippl, M. J., Némethy, G., Scheraga, H. A. Intermolecular potentials from crystal data. 6. Determination of empirical potential for O-H . . . O=C hydrogen bonds from packing configurations. J. Phys. Chem. 88:6231–6233, 1984.
8. 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.
9. Wlodawer, A., Deisenhofer, J., Huber, R. Comparison of two highly refined structures of bovine pancreatic trypsin inhibitor. J. Mol. Biol. 193:145–156, 1987.
10. Li, Z., Scheraga, H. A. Monte Carlo-minimization approach to the multiple-minima problem in protein folding. Proc. Natl. Acad. Sci. U.S.A. 84:6611–6616, 1987.
11. Li, Z., Scheraga, H. A. Structure and free energy of complex thermodynamic systems, J. Mol. Struct. (Theochem). 179:333–352, 1988.
12. 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). Biopolymers 27:1283–1303, 1988.
13. Gay, D. M. Algorithm 611. Subroutines for unconstrained minimization using a model trust-region approach. ACM Trans. Math. Software 9:503–524, 1983.

14. Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., Teller, E. Equation of state calculations by fast computing machines. J. Chem. Phys. 21:1087–1092, 1953.
15. Deisenhofer, J., Steigemann, W. Crystallographic refinement of the structure of bovine pancreatic trypsin inhibitor at 1.5 Å resolution. Acta Crystallogr. Sect. B 31:238–250, 1975.
16. Kendall, M. G. "Rank Correlation Methods." Charles Griffin, 1962:92–100.
17. Wagner, G., Braun, W., Havel, T. F., Schaumann, T., Gō, N., Wüthrich, K. Protein structure in solution by nuclear magnetic resonance and distance geometry. J. Mol. Biol. 196:611–639, 1987.
18. Levitt, M., Sharon, R. Accurate simulation of protein dynamics in solution. Proc. Natl. Acad. Sci. U.S.A. 85:7557–7561, 1988.
19. Weiner, P. K., Kollman, P. A. AMBER: Assisted model building with energy refinement. A general program for modeling molecules and their interactions. J. Comp. Chem. 2:287–303, 1981.
20. Weiner, S. J., Kollman, P. A., Case, D. A., Singh, U. C., Ghio, C., Alagona, G., Profeta, S. J., Weiner, P. A. A new force field for molecular mechanical simulation of nucleic acids and proteins. J. Am. Chem. Soc. 106:765–784, 1984.
21. Weiner, S. J., Kollman, P. A., Nguyen, D. T., Case, D. A. An all atom force field for simulations of proteins and nucleic acids. J. Comp. Chem. 7:230–252, 1986.
22. Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., Karplus, M. CHARMM: A program for macromolecular energy, minimization, and dynamic calculations. J. Comp. Chem. 4:187–217, 1983.
23. No, K. T., Kang, Y. K., Scheraga, H. A. in preparation.
24. Simon, I., Glasser, L., Scheraga, H. A. Calculation of protein conformation as an assembly of stable overlapping segments: Application to bovine pancreatic trypsin inhibitor. Proc. Natl. Acad. Sci. U.S.A. 88:3661–3665, 1991.