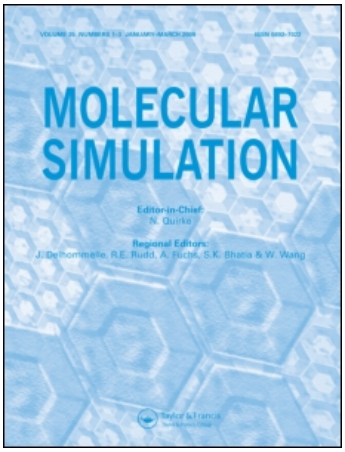

Molecular Simulation
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713644482

Protein Structure Prediction with a Combined Solvation Free Energy-Molecular Mechanics Force Field

Celia A. Schiffer[ab]; James W. Caldwell[c]; Peter A. Kollman[bc]; Robert M. Stroud[abc]
[a] Department of Biochemistry and Biophysics, University of California, San Francisco, California, USA [b] Graduate Group in Biophysics, University of California, San Francisco, California, USA [c] Department of Pharmaceutical Chemistry, University of California, San Francisco, California, USA

To cite this Article Schiffer, Celia A. , Caldwell, James W. , Kollman, Peter A. and Stroud, Robert M.(1993) 'Protein Structure Prediction with a Combined Solvation Free Energy-Molecular Mechanics Force Field', Molecular Simulation, 10: 2, 121 — 149
To link to this Article: DOI: 10.1080/08927029308022162
URL: http://dx.doi.org/10.1080/08927029308022162

Molecular Simulation, 1993, Vol. 10(2-6), pp. 121-149
Reprints available directly from the publisher
Photocopying permitted by license only

PROTEIN STRUCTURE PREDICTION WITH A COMBINED SOLVATION FREE ENERGY-MOLECULAR MECHANICS FORCE FIELD

CELIA A. SCHIFFER†‡, JAMES W. CALDWELL*,
PETER A. KOLLMAN†*, ROBERT M. STROUD†‡*

‡*Department of Biochemistry and Biophysics*, †*Graduate Group in Biophysics and* **Department of Pharmaceutical Chemistry University of California, San Francisco, San Francisco, California 94143, USA*

(Received January 1993, accepted January 1993)

Models of protein structure are frequently used to determine the physical characteristics of a protein when the crystal structure is not available. We developed a procedure to optimize such models, by use of a combined solvation free energy and molecular mechanics force field. Appropriately chosen atomic solvation parameters were defined using the criterion that the resulting protein model should deviate least from the crystal structure upon a forty picosecond molecular dynamics simulation carried out using the combined force field. Several tests were performed to refine the set of atomic solvation parameters which best complement the molecular mechanics forces. Four sets of parameters from the literature were tested and an empirically optimized set is proposed. The parameters are defined on a well characterized small molecule (alanyl dipeptide) and on the highly refined crystal structure of rat trypsin, and then tested on a second highly refined crystal structure of α-lytic protease. The new set of atomic solvation parameters provides a significant improvement over molecular mechanics alone in energy minimization of protein structures. This combined force field also has advantages over the use of explicit solvent as it is possible to take solvent effects into account during energetic conformational searching when modeling a homologous protein structure from a known crystal structure.

KEY WORDS: molecular dynamics, solvation free energy, protein structure, molecular mechanics, atomic solvation parameters

INTRODUCTION

From the folding process to the conformation of external side chains, protein structure is affected by the solvent. Therefore, the inclusion of the solvent when calculating physical properties derived from protein structure is essential. Explicitly modeling the solvent however, is computationally expensive, and can be impractical when deriving many physical properties. Thus, we sought an optimal method of representing the effects of water without including water molecules explicitly in the calculation. This can be very useful in fine tuning models of protein structure.

One method of approximating in an empirical fashion the effects of solvation assigns an atomic solvation parameter to each atom and multiplies it by the solvent exposed surface area to calculate a free energy of solvation ($\Delta G = \Delta\sigma^* A$) [1, 2]. This solvation free energy has been used to select between alternative models distinguishing between correctly and incorrectly folded static protein structures [3-5]. We previously developed this approach to distinguish between correctly and

incorrectly modeled side chain conformations [6]. However, there is a need to go beyond simply choosing between alternatives and to refine protein structure explicitly using these energy terms.

In carrying out energy minimization and molecular dynamics on a molecule, the derivative of the solvation free energy must be used as a component of the force to simulate a trajectory of the molecule which is relevant to its solvated state. Such a derivative of the solvation free energy, when added into a molecular mechanics force field, should be able to mimic more correctly the forces on a solvated molecule.

Wesson and Eisenberg [7] performed a series of MD simulations with a combined force field on the monomer and tetramer of melittin using water to vapor free energy of transfer [(w/v)W] atomic solvation parameters and the CHARMM [8] force field. Melittin is a 26 amino acid residue peptide that is helical when it is tetrameric and a coil when it is monomeric. Wesson and Eisenberg [7] observed that a helical monomer unfolds during a MD simulation with a combined force field, whereas the helical tetramer remains mainly helical. The tetramer, however, still deviates significantly from the crystal structure by 3.41 Å, indicating that these water to vapor atomic solvation parameters are not completely complementary with the CHARMM [8] molecular mechanics force field.

In this study we apply this combined force field to two protein structures, first to the crystal structure of rat trypsin [9] which was used to refine atomic solvation parameters, and then to the crystal structure of α-lytic protease [10]. We first evaluate four sets of atomic solvation parameters currently in the literature to determine which of these maintains the protein structure in the conformation most like the crystal structure, and then derive an optimal set that produces the least change from the crystal structure. Previously, we applied this combined force field to several conformations of alanyl dipeptide [11] whose structure can be defined by only two dihedral angles. We found that the addition of a surface area based solvation energy produces consistent results with both experimental [12, 13] and previous computational [14–18] studies. The work presented here includes further analysis of alanyl dipeptide, molecular dynamics simulations, and homology modeling using the lowest energy conformational search (LECS) algorithm we described previously [6] on the crystal structure of rat trypsin. These evaluations lead us to define a new set of atomic solvation parameters which are optimized to be complementary to the molecular mechanics force field. We then apply this new set of parameters to the crystal structure of α-lytic protease.

METHODS

All the molecular dynamics calculations were performed with AMBER force fields [19] on the crystal structure of rat trypsin [9]. The software for determining the derivatives of the solvation energy was provided by Wesson and Eisenberg [7] (a modification of the original derivation by Richmond [20]). The derivatives of the solvation energy with respect to atomic position were added directly to the rest of the molecular mechanics forces, where the force is the negative derivative of energy E, with respect to atomic position x:

$$F = -\frac{\delta E_{\text{total}}}{\delta x}$$

and

$$E_{\text{total}} = E_{\text{molecular mechanics}} + E_{\text{solvation}}$$

Where $E_{\text{molecular mechanic}}$ is the force field of Weiner *et al.*, [19] in AMBER and $E_{\text{solvation}}$ is the sum of the solvation energies from solvent exposed atoms.

$$E_{\text{molecular mechanics}} = \sum_{\text{bonds}} K_b(b - b_0)^2 + \sum_{\text{angles}} K_a(\theta - \theta_0)^2 +$$

$$\sum_{\text{dihedrals}} V_n/2[1 + \cos(n\phi - \gamma)] +$$

$$\sum_{i>j} [A_{ij}/R_{ij}^{12} + B_{ij}/R_{ij}^{6} + q_i q_j/\varepsilon R_{ij}] +$$

$$\sum_{H-\text{bonds}} [C_{ij}/R_{ij}^{12} + D_{ij}/R_{ij}^{10}]$$

and

$$E_{\text{solvation}} = \sum_{\text{all atoms}} \Delta\sigma A$$

All the dynamics calculations were performed at 300 K using a distance dependent dielectric to mimic solvent electrostatic shielding and an 8 Å non-bonded cutoff was used. The change in volume of the structure of rat trypsin was calculated from the radius of gyration $(4/3\pi(Rg^3_{\text{crystal}} - Rg^3_{\text{model}}))$. The volume overlap of the side chains for evaluation of the LECS calculation was determined by the method described in Schiffer *et al.* [6] where the side-chain volume was calculated from the van der Waals radii of their atoms, backbone atoms were not included in the calculations of the overlap. The water model used for the explicit solvent simulations was TIP3P. Calculations with the combined molecular mechanics solvation free energy were performed on an IBM RISC 6000/530, DEC 9000, and PSC Cray-YMP computers.

BACKGROUND: ORIGIN OF CURRENTLY USED ATOMIC SOLVATION PARAMETERS

Atomic solvation parameters have been determined from several experimental databases (Table 1a). The four sets that we assessed will be initially divided into two groups; two derived from water to octanol free energy of transfer data and two derived from water to vapor free energy of transfer data. In all four cases, partition coefficients were experimentally determined and transfer free energies were calculated from these values. The free energies are then partitioned into the relative contributions of five atom types (C, N/O, N+, O−, S). From these divided free energies the five atomic solvation parameters are determined. Reference solvent accessible surface areas were calculated for each amino acid residue X in a tripeptide

Gly-X-Gly. This technique was originally proposed by Shrake and Rupley [21] who took the conformation of the tripeptides from the backbone structures of insulin and lysozyme. This method was extended by Eisenberg *et al.* [2] when they found all the occurrences of the sequence Gly-X-Gly in proteins whose structure was deposited in the Brookhaven Protein Data Base and used the tripeptide conformations from those structures to calculate a reference solvent accessible surface area.

The set of parameters proposed by Eisenberg and McLachlan [1] (w/o)E came from the partition coefficient data of Fauchere and Pliska [22]. These partition functions were between water and octanol of "Na-acetyl-L-amino-acid amides" and from this data they calculated a hydrophobicity scale for all 20 side chains where p(side chain) = $\log P$ (acetyl-amino-acid amide) − $\log P$ (acytyl-glycine amide). P is the partition coefficient between water and octanol. The free energy of transfer was then calculated $\Delta G_{obs} = 2.3\,RT\pi$.

Kim [23, 24] took a different approach in obtaining the atomic solvation parameters (w/o)K. Rather than using amino acid analogues, tripeptides ala-X-ala (where X = gly, ala, phe, trp, pro, his, asp, and glu) which were N- and C-terminally blocked were used to obtain the free energy of transfer between water and octanol.

Two sets of atomic solvation parameters [7] have been calculated from a combination of two sets of experimental water to vapor free energy of transfer data [25–27]. This data is obtained by determining a compounds vapor pressure over dilute aqueous solutions [26]. To this experimental data two separate corrections were applied [28, 29] from which the two sets of atomic solvation parameters were derived (w/v)W and (w/v)S respectively.

Kyte and Doolittle [28] take small molecule free energy of transfer data from Wolfenden *et al.*, [25] and Hine and Mookerjee [27] and calculate a free energy of transfer between water and vapor. They correct the data so as "To eliminate any entropy of mixing from the values, the transfer must occur between standard states chosen in such a way that no change in volumes are involved." [28]. However, the appropriateness of this correction has been recently called into question. "... the widely used hydrophobicity scale of Kyte and Doolittle [28] involves an erroneous 'standard state' correction to the amino-acid side chain partition data ... This was apparently to account for differences in reference concentrations. However since the original data is already expressed in ratios of molarity in the vapor and water phases it is wrong to apply this correction." [29] Furthermore, Sharp *et al.* [29] suggest that the free energy of transfer values of Kyte and Doolittle [28] are greatly underestimated for another reason; that the individual solute and solvent molecules are of physically different sizes. This leads to another correction [29] applied to the free energy of transfer data of Wolfenden *et al.* [26] to properly take into account entropy of mixing. The experimental data for which both the Kyte and Doolittle [28] and the Sharp *et al.*, [29] adjustments were applied are a compilation of free energy of transfer data for small molecule amino acid analogues from several different laboratories. Kyte and Doolittle [28] directly combine four experimental partition coefficients from Wolfenden *et al.* [25] with thirteen experimental partition coefficients from Hine and Mookerjee [27]. Sharp *et al.* [29] use the water to vapor partition data from the Wolfenden *et al.* [26] study. The values of the partition coefficients in this study, however, are a compilation of seven coefficients they [26] measured themselves and twelve of the coefficients from six other laboratories [30–35].

Table 1a Atomic solvation parameters.

Reference	$\Delta\sigma(C)$	$\Delta\sigma(N/O)$	$\Delta\sigma(N^+)$	$\Delta\sigma(O^-)$	$\Delta\sigma(S)$
(w/o)E[1,2]	18+/−1	−9+/−3	−38+/−4	−37+/−7	5+/−6
(w/o)K[23]	13+/−2.2	−7+/−5.6	−87+/−10	−112+/−20	−3.6*
(w/v)W[7]	4+/−3	−113+/−14	−169+/−31	−166+/−38	−17+/−22
(w/v)S[29,7]	12+/−3	−116+/−13	−186+/−22	−175+/−3	−18+/−21

All atomic solvation parameters are in cal $Å^{-2}$ mol^{-1}, *the atomic solvation parameter value for S was approximated since none of the tripeptides in the study had sulphur in them.

Table 1b Atomic solvation parameters normalized to the $\Delta\sigma$(N/O) parameter.

Reference	$\Delta\sigma(C)$	$\Delta\sigma(N/O)$	$\Delta\sigma(N^+)$	$\Delta\sigma(O^-)$	$\Delta\sigma(S)$
(w/o)E[1,2]	2	−1	−4.2	−4.1	0.55
(w/o)K[23]	1.9	−1	−12.4	−16	−0.51
(w/v)W[7]	0.035	−1	−1.5	−1.5	−0.15
(w/v)S[29,7]	0.1	−1	−1.5	−1.5	−0.15

The major difference between the water to octanol free energy of transfer parameters ((w/o)E and (w/o)K) and the water to vapor free energy of transfer parameters ((w/v)W and (w/v)S) is the value of the $\Delta\sigma$(N/O). This parameter is twelve to sixteen times more negative in the water to vapor parameters than the water to octanol. (Table 1a). This large increase in negativity implies that neutral oxygen and nitrogen atoms will prefer to be solvated rather than buried if using either of the water to vapor atomic solvation parameters.

As well as the differences between the water to octanol and the water to vapor parameters, there are subtle differences between the two water to octanol parameter sets and the two water to vapor parameter sets. In Table 1b each set of atomic solvation parameters is divided by the magnitude of the respective $\Delta\sigma$(N/O) parameter. For the two water to octanol free energy of transfer sets the differences are clear. The values for $\Delta\sigma(N^+)$ and $\Delta\sigma(O^-)$ are four times more negative in the (w/o)K set of parameters then the (w/o)E set of parameters and the value of $\Delta\sigma$(S) has changed sign. The large increase in negativity of the (w/o)K parameters implies that the charged nitrogens and oxygens are preferentially solvated. For the two water to vapor sets of parameters (Table 1b) the magnitude of $\Delta\sigma$(C) is very small implying that carbon is not preferentially found either solvated or buried. Thus there is almost no penalty for exposing a carbon atom to solvent. Between the (w/v)W set of parameters and (w/v)S set of parameters the only notable difference is that in the (w/v)S parameters the $\Delta\sigma$(C) is three times larger than the (w/v)W set but is still very small in magnitude compared with the other solvation parameters.

We therefore determine empirically whether the parameters from water to octanol free energy of transfer or water to vapor free energy of transfer more accurately complement the molecular mechanics force field. Then we determine whether the differences between either (w/o)E and (w/o)K or (w/v)W and (w/v)S are significant, and establish that another set of parameters needs to be proposed. Finally we derived an optimal set of atomic solvation parameters for conjugation with the molecular mechanics force field.

RESULTS AND DISCUSSION

Part A: Three test cases

Complementation to the molecular mechanics force field was tested with four sets of atomic solvation parameters currently in the literature. These were then used in our combined force field on three test systems: on alanyl dipeptide, in an energetic conformational search homology modeling algorithm and on the high resolution crystal structure of rat trypsin. As a result of these tests we define a set of atomic solvation parameters which are complementary with the molecular mechanics force field.

I. *Test on alanyl dipeptide*

As an initial test of the combined force field, we chose alanyl dipeptide which has only two degrees of freedom and the structure of which has been extensively studied in various solvent environments both experimentally and computationally. We looked at four conformations of the alanyl dipeptide known to be low energy minima (C_{7eq}, C_{7ax}, C_5, α_r). We previously found that the addition of solvation free energy to the molecular mechanics force field destabilized the C_{7eq} conformation relative to the C_5 and C_{7ax}. [11] which is consistent with the experimental results of Madison and Kopple [13]. This study was done using (w/v)W atomic solvation parameters.

The energy of four conformations of alanyl dipeptide was minimized with the combined force field and the four separate sets of atomic solvation parameters (Table 1). The results of using all four sets of atomic solvation parameters are presented (Table 2) as well as using none (i.e., only molecular mechanics). The most striking difference between the parameter sets is that the water to vapor free energy of transfer atomic solvation parameters further stabilize the C_5 conformation of the alanyl dipeptide by 1.2 kcals/mole and the α_r conformation by 1.1 kcal/mole both relative to the C_{7eq} conformation. The extra stabilization conferred by the water to vapor parameters is due to the large magnitude of the atomic solvation parameter for the uncharged nitrogen and oxygen, relative to the water to octanol atomic solvation parameters.

For all four data sets, however, the ranking of the lowest energy conformations of the alanyl dipeptide remains the same C_5, C_{7eq}, C_{7ax}, α_r as expected in solution. This peptide model is not sufficient to completely evaluate the four parameter sets, since only two of the five atom types are represented (carbon and neutral nitrogen and oxygen) and the molecule is too small to have a hydrophobic core. Thus a protein model with a hydrophobic core is necessary to evaluate these four set of atomic solvation parameters.

II. *LECS search on 3 polar residues*

The advantage of being able to incorporate the effects of solvent without having solvent explicitly in the calculation, is that it is possible to perform an energetic conformational search on an external side chain or loop. This calculation, as implemented with the protocol of [6], is not practical with explicit water, although solvent is essential in determining the conformation of an external part of a protein. This is particularly relevant when building a homology model from a known crystal structure, when most of the differences in structure between homologous

Table 2 Minimized structures with a combined force field and a distance dependent dielectric constant.

Structure	*phi*	*psi*	*total E*	*solv E*	*ΔE*
a) (w/o)E - reference solvation energy* 2.89					
C_{7eq}	−76.9	71.0	9.43	1.06	0.00
C_{7ax}	69.6	−67.2	−8.48	0.90	0.95
C_5	−162.4	168.9	−9.89	0.85	−0.46
α_R	−64.7	−43.5	−5.23	0.83	4.20
b) (w/o)K - reference solvation energy* 2.13					
C_{7eq}	−77.1	70.6	−9.70	0.79	0.00
C_{7ax}	69.5	−67.5	−8.71	0.67	0.99
C_5	−162.8	170.2	−10.22	0.63	−0.41
α_R	−65.1	−42.7	−5.44	0.62	4.26
c) (w/v)W - reference solvation energy* −7.91					
C_{7eq}	−77.0	73.4	−9.03	1.37	0.00
C_{7ax}	69.8	−67.9	−8.34	1.00	0.69
C_5	−161.4	164.2	−10.63	0.02	−1.61
α_R	−68.3	−42.9	−5.78	0.23	3.25
d) (w/v)S - reference solvation energy* −6.49					
C_{7eq}	−76.9	73.8	−8.55	1.06	0.00
C_{7ax}	69.8	−68.3	−7.92	0.89	0.63
C_5	−161.3	163.8	−10.24	0.79	−1.69
α_R	−68.7	−42.8	−5.39	0.81	3.16
e) molecular mechanics alone					
C_{7eq}	−77.3	70.0	−10.49		0.00
C_{7ax}	69.8	−67.1	−9.38		1.11
C_5	−163.3	171.0	−10.74		−0.25
α_R	−65.4	−44.2	−6.06		4.43

*Reference solvation energy is the solvation energy of a completely extended conformation.

proteins occur on the surface of the protein. Thus, as a further test of the combined force field the four sets of atomic solvation parameters were applied in a homology modeling algorithm in which the lack of solvation has been shown previously to be the cause of inaccurate prediction of external side chain conformations [6]. This algorithm performs a lowest energy conformational search (LECS) on selected residues and chooses the lowest energy conformer to be the predicted structure. We applied this algorithm with all 5 solvent conditions (the 4 sets of atomic solvation parameters and molecular mechanics only) to three external side chains on the crystal structure of rat trypsin and then compared them to the crystal structure. The results listed in Table 3 compare both the volume occupied by each residue with respect to the crystal structure as well as the dihedral conformation of each residue. For ASN 130 all 5 simulations reproduce the crystal structure conformation and occupy 75–80% of the crystal volume (Figure 1a). The water and octanol atomic solvation parameters are able to correctly pick the crystal structure conformation of the dihedral angles for GLN 145 (Figure 1b), with the exception of χ_3 which is flipped by 180°. None of the solvent models, however, is able to correctly predict the conformation of ASP 147 (Figure 1c). Thus the water to octanol atomic solvation parameters are more successful in predicting the crystal structure conformation

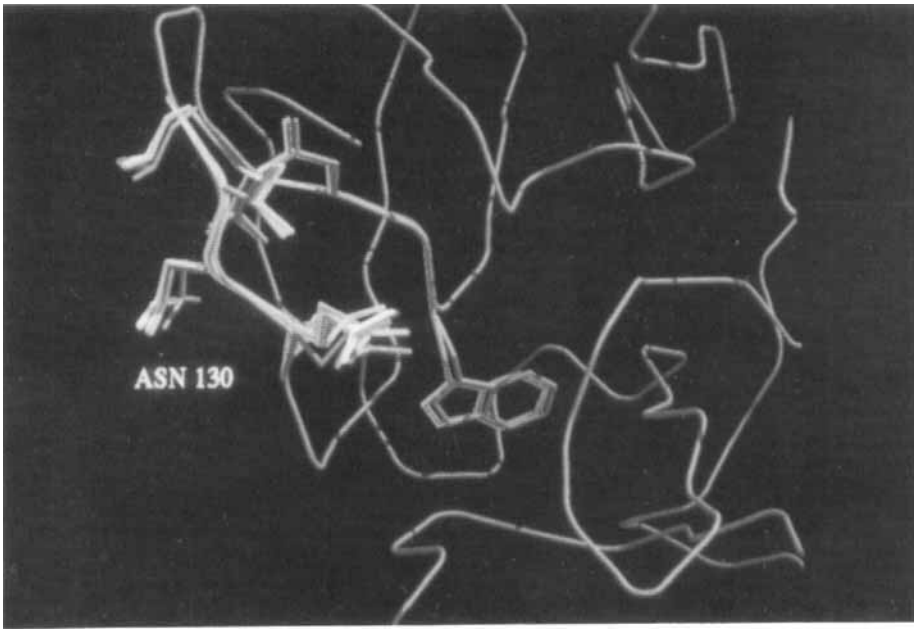

Figure 1a

Figure 1b

Figure 1 The superposition of 7 structures showing the sidechains as predicted by LECS as compared to the crystal structure. orange = crystal structure, blue = (w/v)S, cyan = (w/o)K, red = (w/o)E, white = (w/v)W, magenta = molecular mechanics alone, (yellow = (mm)S which we have derived) (see colour plates)
a) ASN 130. b) GLN 145. c) ASP 147.

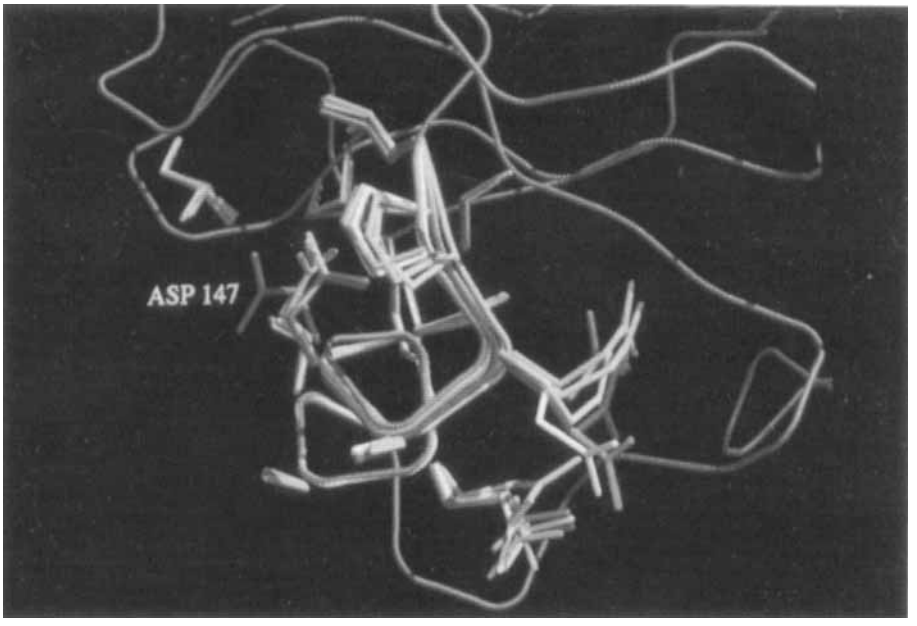

Figure 1c The superposition of 7 structures showing the sidechains as predicted by LECS as compared to the crystal structure. orange = crystal structure, blue = (w/v)S, cyan = (w/o)K, red = (w/o)E, white = (w/v)W, magenta = molecular mechanics alone, (yellow = (mm)S which we have derived) (see colour plates)
a) ASN 130. b) GLN 145. c) ASP 147.

Table 3 LECS dihederals and volume overlap with crystal structure conformation.

Residue	*Structure*	χ_1	χ_2	χ_3	*% overlap*
130 ASN	crystal	167.6	74.7		****
	(w/o)E	175.4	77.8		75.3
	(w/o)K	176.4	75.8		73.2
	(w/v)W	179.6	79.2		69.4
	(w/v)S	179.1	75.1		79.9
	mm	175.5	77.2		76.5
145 GLN	crystal	−154.9	74.2	167.5	****
	(w/o)E	−175.1	67.5	18.3	60.8
	(w/o)K	−176.7	67.7	17.5	68.7
	(w/v)W	67.7	174.7	103.9	41.1
	(w/v)S	−178.9	−177.2	−23.1	48.8
	mm	−73.7	−52.6	134.2	74.9
147 ASP	crystal	−67.8	−28.0		****
	(w/o)E	69.3	142.1		58.6
	(w/o)K	75.4	−125.0		61.8
	(w/v)W	71.9	−113.2		61.6
	(w/v)S	73.9	−118.8		62.5
	mm	70.0	144.7		58.0

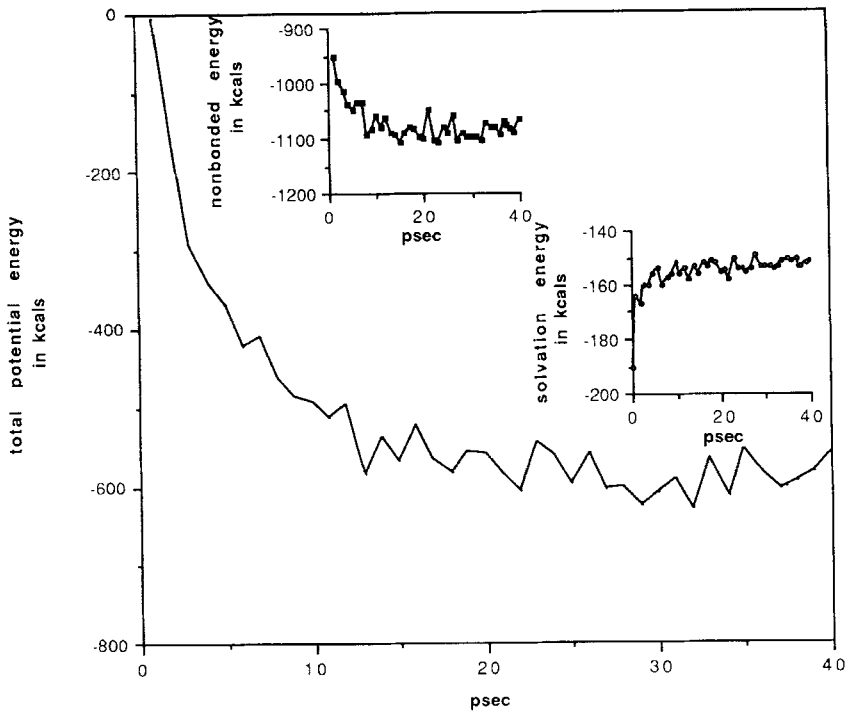

Figure 2 Components of the energy during the combined molecular mechanics solvation free energy 40 ps of MD calculation on the crystal structure of rat trypsin with (w/o)E atomic solvation parameters.

for external side chains than either of the water to vapor parameter sets or the molecular mechanics parameter set.

III. *Molecular dynamics simulations on rat trysin crystal structure*

Molecular Dynamic (MD) simulations were performed with the combined force field to further compare the four sets of atomic solvation parameters. The dynamic simulations were performed on the crystal structure of rat trypsin [9] at 300 K with a distance dependent dielectric and a time step of 2 femtoseconds. The first simulation was run for 40 ps with the (w/o)E atomic solvation parameters (Figure 2). During the course of the simulation the solvation energy increased quickly in the first 5 ps of the simulation and then stabilized for the remaining 35 ps. The total potential energy of the system decreased over the first 10 ps and before settling down for the remaining 30 ps. The initial sharp decrease in the total potential energy was due mainly to the change in the nonbonded energy of the system. The second simulation was also for 40 ps using the (w/o)K atomic solvation parameters (Figure 3). The course of this simulation was very similar to the previous one; once again the solvation energy increases quickly while the potential energy decreases as a result of the change in the nonbonded energy of the system. Thus the two sets of water to octanol free energy of transfer based atomic solvation

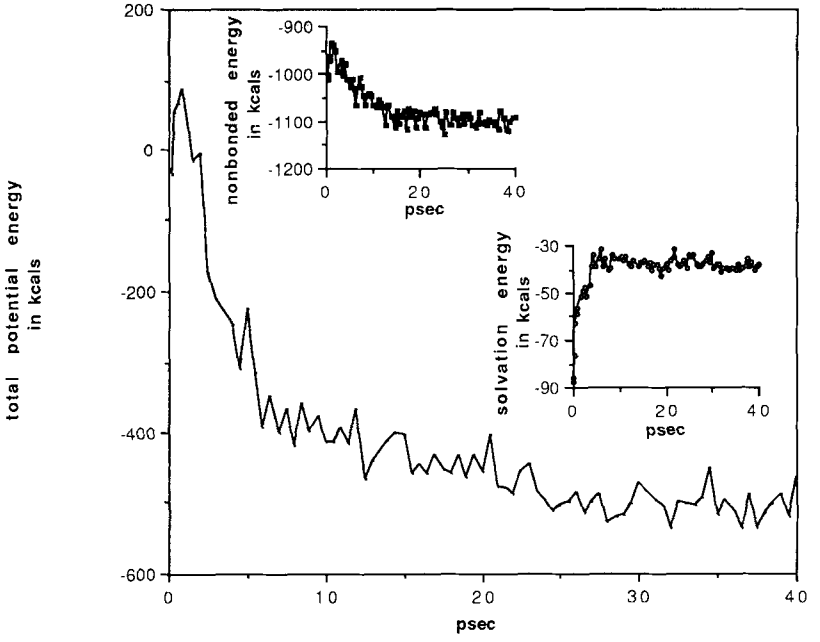

Figure 3 Components of the energy during the combined molecular mechanics solvation free energy 40 ps of MD calculation on the crystal structure of rat trypsin with (w/o)K atomic solvation parameters.

parameters have a similar effect on a 40 ps molecular dynamic trajectory of rat trypsin.

Two additional MD simulations were performed on the crystal structure of rat trypsin using the two sets of atomic solvation parameters derived from the water to vapor free energy of transfer data. A 10 ps MD simulation was performed using the set of (w/v)W atomic solvation parameters (Figure 4). In this case the solvation energy dropped quickly in the first 2 ps of the simulation and then stabilized for the remaining 8 ps. This drop in the solvation energy dominates the total potential energy. The nonbonded energy fluctuates much more during the first 10 ps of this simulation then was seen in the first two simulations. Another 18 ps MD simulation was performed using the (w/v)S atomic solvation parameters (Figure 5). Once again, the solvation energy decreased during the simulation, and as before, the nonbonded energy dominates the total potential energy. Thus the two sets of water to vapor free energy of transfer base atomic solvation parameters behaves similarly in that the solvation energy decreased during the course of the simulation.

Finally, a 40 ps MD simulation was performed on the crystal structure of rat trypsin using only the molecular mechanics force field (Figure 6). The solvation energy was calculated 15 times over the course of the simulation on intermediary structures with the (w/o)E atomic solvation parameters. During this simulation the total potential energy followed the nonbonded energy very closely.

To analyze the results of these five MD simulations the deviations of the structures over the course of the simulation were compared to the crystal structure

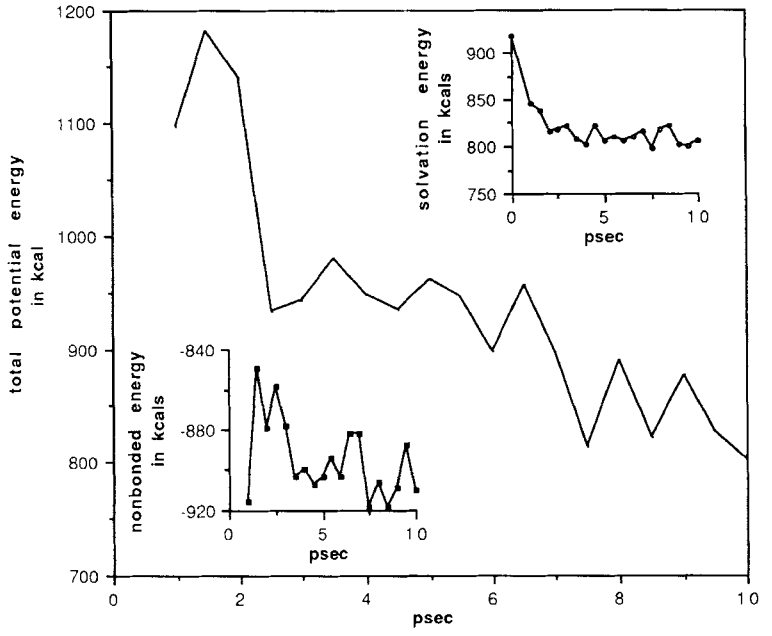

Figure 4 Components of the energy during the combined molecular mechanics solvation free energy 10 ps of MD calculation on the crystal structure of rat trypsin with (w/v) W atomic solvation parameters.

of rat trypsin. In Figures 7 and 8 the change in the compactness of the structure and the r.m.s. deviation with respect to the crystal structure is shown. Many of the changes in the structures that result from the simulations are in loop regions of the structure that do not have well defined pieces of secondary structure (Figure 9a, b, c). This is not surprising since the structures of these flexible regions are most susceptible to the solvent environment. In the first couple of picoseconds of the simulations, all of the structures expand. This expansion is most dramatic for the simulations in which the water to vapor transfer free energy atomic solvation parameters were used (Figure 9a). In both of these simulations the structure expands by more than 10% in the first 2 ps and remains at least as large for the following 8 ps of the simulation. The parameters (Table 1) for the neutral nitrogens and oxygens make it more favorable for the molecule to expose these atoms to solvent than to keep them buried in the interior. In addition, the magnitude of the atomic solvation parameters for carbon is relatively small and thus it is not particularly unfavorable to expose them to solvent. These parameters are not compatible with the molecular mechanics force field as the structure is essentially unfolding during the course of the simulation due to the denaturing solvent conditions.

In the other three simulations, with the two water to octanol atomic solvation parameter sets and molecular mechanics alone, the compactness of the structure, after the initial increase in volume, decreases to almost the crystal structure volume in 10 ps. The simulation with the (w/o)E atomic solvation parameters stays at a somewhat expanded volume of about 4% relative to the crystal structure. The structure from the simulation with only the molecular mechanics force field

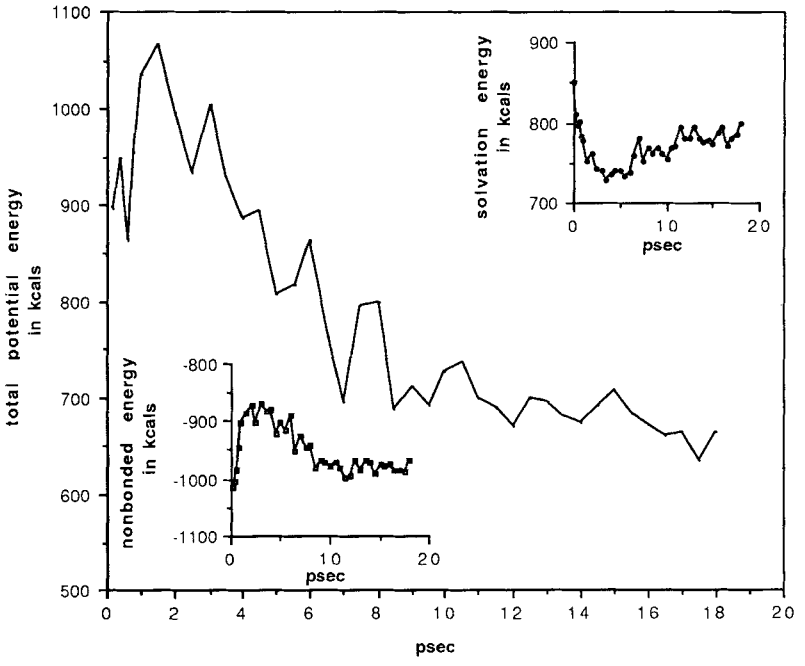

Figure 5 Components of the energy during the combined molecular mechanics solvation free energy 18 ps of MD calculation on the crystal structure of rat trypsin (w/v)S derived atomic solvation parameters.

contracted by about 2% from the crystal structure (Figure 9b). This is due to the lack of solvent in the calculation. As a result during the simulation, the polar and charged side chains at the surface of the protein find lower energy conformations by hydrogen bonding back to the protein. There is, however, a limit to how far the structure can compact because of the van der Waals interactions.

The simulation which maintains the structure closest to the crystal structure volume uses the (w/o)K atomic solvation parameters derived from the free energy of transfer of tripeptides from water to octanol [23, 24]. This parameter set balances the hydrophobicity of carbon with the hydrophilicity of the nitrogen and oxygen and best maintains the crystal structure volume after the first 10 ps of the simulation (Figure 9c).

Part B: In comparison to a shell of explicit water

For comparison with the combined force field calculations a simulation with a 4 Å shell of explicit water was performed on the structure of rat trypsin. The equilibration procedure of Guenot and Kollman [36] was followed to place the water shell, which includes using a distance dependent dielectric constant to damp the long range electrostatic effects. A total of 429 water molecules were included in the hydration shell. A 50 ps MD simulation at 300 K was performed and during the course of the simulation seven intermediary structures were compared to the

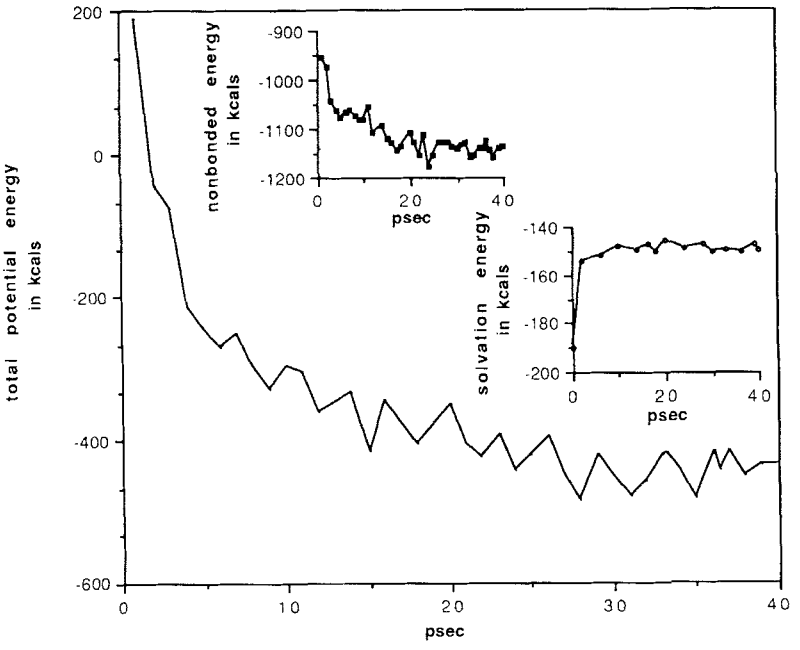

Figure 6 Components of the energy during the molecular mechanics alone 40 ps MD calculation on the crystal structure of rat trypsin. Note the solvation energy is calculated from static structures using the (w/o)E atomics solvation parameters.

crystal structure. The r.m.s. deviation to the crystal structure varied from 1.5–1.75 Å for all atoms, and the structures volumes ranged from an expansion of 1.48% to a contraction of −0.95%. Thus a thin shell of explicit waters maintains the crystal structure of the protein better than any of the combined force field solvation models.

Computationally the combined force field is slower than either MD without solvent or than MD with a 4 Å shell of explicit water. For 1 ps of MD on the structure of rat trypsin with an IBM RISC 6000/530 computer, molecular mechanics alone took 5,958 seconds, molecular mechanics with a 4 Å shell of water took 16,768 seconds, and the combined molecular mechanics solvation free energy force field took 47,479 seconds for a ratio of 1.0:2.8:8.0 respectively. The combined force field is still faster, however, than would be the case for completely solvating rat trypsin with explicit water and using periodic boundary conditions.

The advantages the combined force field has over the use of explicit solvent are that it can be applied to molecular modeling systems with more than a small number of sequence substitutions or insertions. Explicit water cannot be used in conjunction with molecular modeling that relies on conformational searches and energy minimization precisely because it requires an equilibration period to place the solvent molecules. This makes the comparison of the relative energies of the various conformations impossible. Furthermore, once the explicit waters are near low energy conformations, it is more difficult for parts of the proteins to change in conformation in a reasonable period of time. To accurately sample conforma-

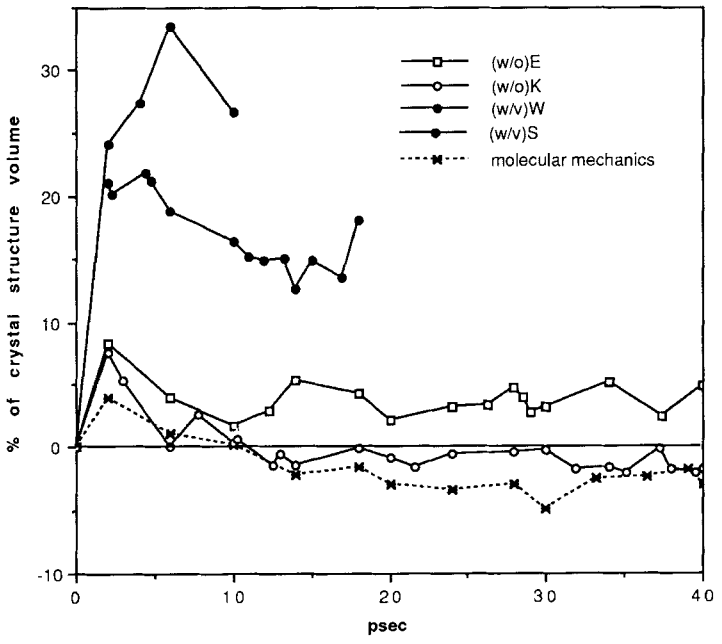

Figure 7 The change in volume relative to the crystal structure of rat trypsin during all 5 MD trajectories.

tional space it would be necessary to start the structure in a large variety of conformations and equilibrating the solvent every time, very quickly the computational advantage of using a small shell of explicit waters would be lost. Thus although explicit solvent is good for simulations where the structure of the protein is known it is not good for modeling the unknown structure of a protein. Therefore the combined force field makes energetic conformational searching with solvent effects feasible on side chains and loops, thus making it possible to model unknown structures of proteins with solvent affects.

Part C: Derivation of optimal atomic solvation parameters

I. *An* ab initio *method*

Since none of the experimental atomic solvation parameters are the ideal match with the molecular mechanics force field, we attempted to derive a set of atomic solvation parameters which exactly counteract the molecular mechanics forces on a well refined crystal structure. We selected a set of five globular proteins as test cases including: rat trypsin [9], bovine trypsin [37], α-lytic protease [10], streptomyces griseus protease [38], and Carboxypeptidase A [39]. The criteria for choosing these structures were as follows: (1) the proteins are at least 100 amino acid residues in length, (2) the structures have been solved to better than 2.0 Å resolution, (3) they are monomeric proteins, and (4) they are as spherical in overall shape as possible. The rationale for these criteria is to select proteins with well

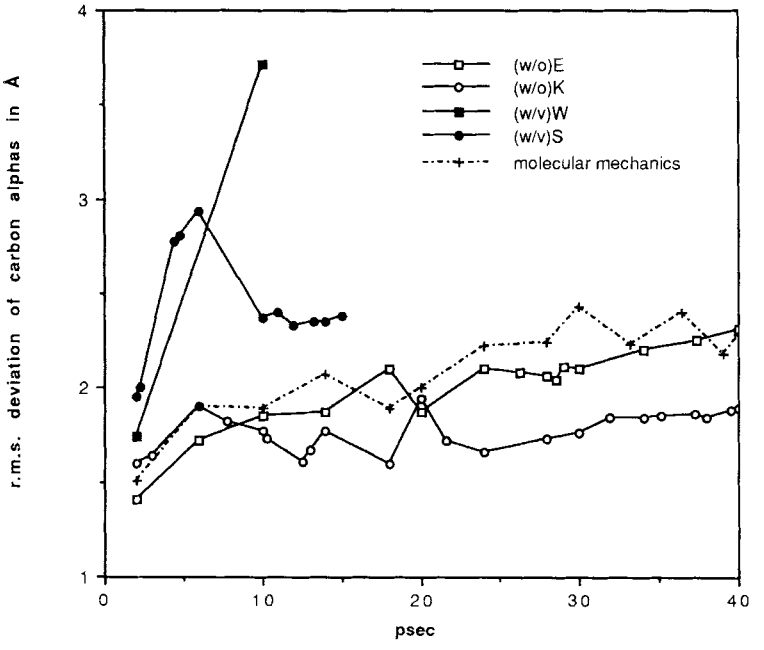

Figure 8a

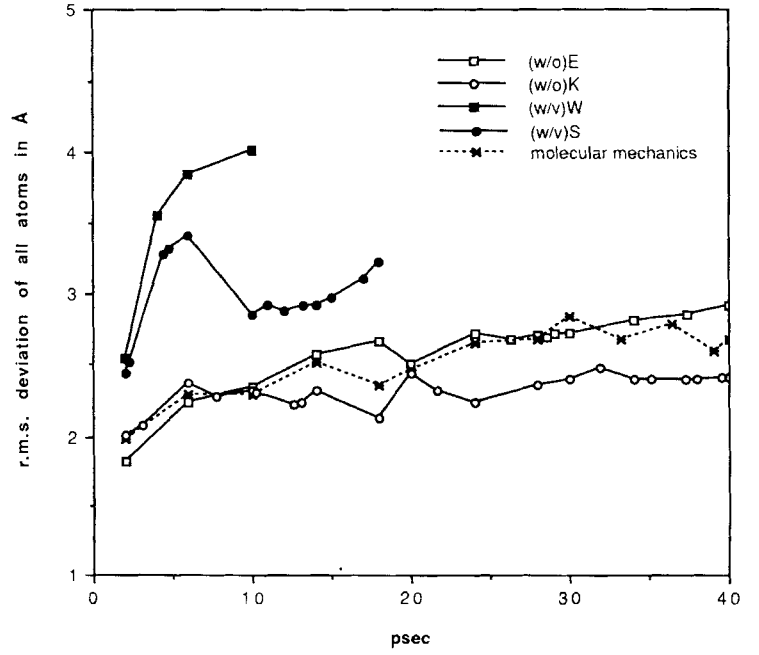

Figure 8 The rms deviation relative to the crystal structure of rat trypsin during all 5 MD trajectories. (a) of alpha carbons (b) of all non-hydrogen atoms.

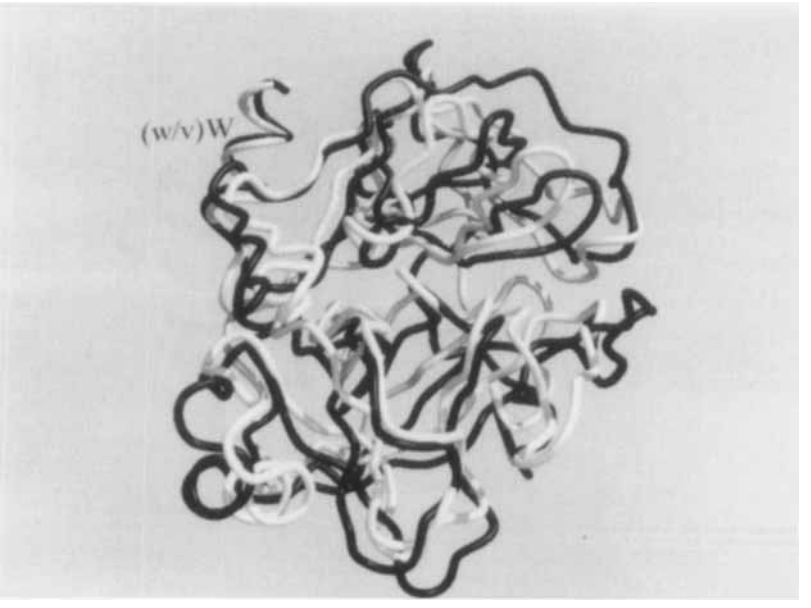

Figure 9a

Figure 9 In each comparison the white structure is the crystal structure of rat trypsin. (a) The black structure is the resulting expanded structure after 10 ps of MD with (w/v)W atomic solvation parameters. (b) The black structure is the resulting contracted structure after 40 ps of MD with the molecular mechanics force field. (c) The black structure is the resulting structure after 40 ps of MD with the (w/o)K atomic solvation parameters. (see colour plates)

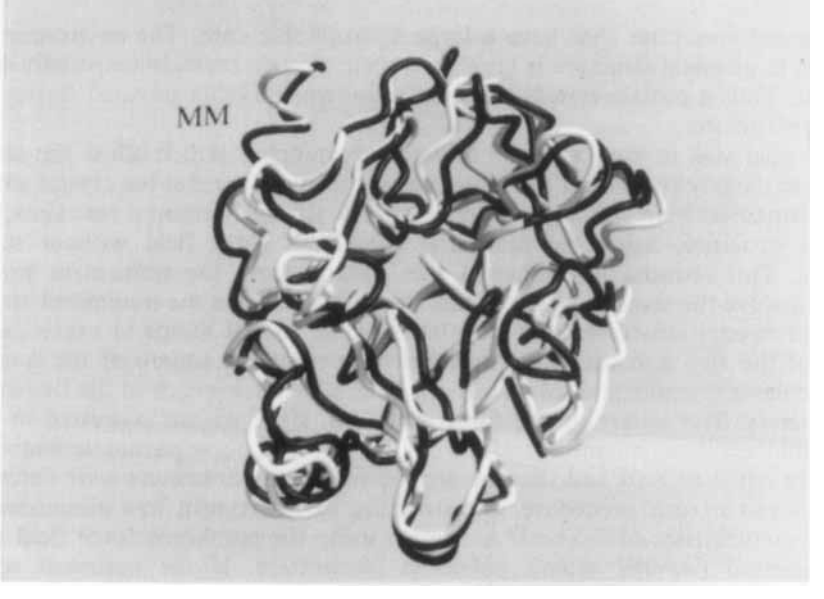

Figure 9b

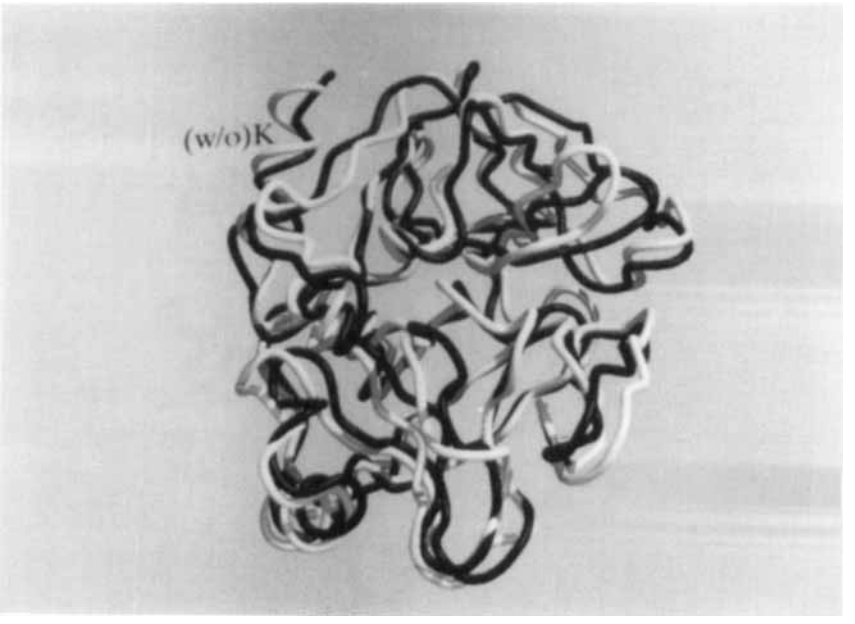

Figure 9c In each comparison the white structure is the crystal structure of rat trypsin. (a) The black structure is the resulting expanded structure after 10 ps of MD with (w/v)W atomic solvation parameters. (b) The black structure is the resulting contracted structure after 40 ps of MD with the molecular mechanics force field. (c) The black structure is the resulting structure after 40 ps of MD with the (w/o)K atomic solvation parameters. (see colour plates)

determined structures that have a large hydrophobic core. The environment of a protein in a crystal structure is highly solvated: protein crystals are usually 40–80% solvent. Thus a protein crystal structure represents a highly solvated conformation of the structure.

The goal was to find a set of solvation parameters which allow the solvation forces to exactly counteract the molecular mechanics forces at the crystal structure. The structures were initially minimized, with strong harmonic restraints, to the crystal structure, using the molecular mechanics force field without solvation energy. This reduced any inherent bias arising from the refinement procedure used to solve the structure. The r.m.s. deviation between the minimized structures and the crystal structures was less than 0.05 Å for all atoms in every case. For each of the five structures the partial derivative of the square of the total force (molecular mechanics and solvation) was determined using each of the five solvation parameters. The square of the force was used, since we are interested in finding the minimum in the magnitude of the overall force. The five partial derivatives were then set equal to zero and the five atomic solvation parameters were determined.

As a test of this procedure, the structure of rat trypsin was minimized to an r.m.s. convergence of 0.5 kcal/Å degree, using the combined force field and the experimental (w/v)W atomic solvation parameters. If the optimum solvation parameters from a given structure are correct, the minimized structure should give back the solvation parameters which were used in the minimization. This is

Table 4

	Solvation parameters used in the minimization*	*Solvation parameters* derived from the minimized structure*
$\Delta\sigma$(C)	4.0	9.0
$\Delta\sigma$(N/O)	−112	−109
$\Delta\sigma(O^-)$	−166	−150
$\Delta\sigma(N^+)$	−169	−164
$\Delta\sigma$(S)	−17	− 17.5

* Solvation parameters in cal $Å^{-2}$ mol^{-1}

Table 5 Crystal structure derived atomic solvation parameters to complement molecular mechanics.

Structure	$\Delta\sigma(C)$	$\Delta\sigma(N/O)$	$\Delta\sigma(N^+)$	$\Delta\sigma(O^-)$	$\Delta\sigma(S)$
4ptp	67	61	−77	−10	588
2trm	75	46	72	51	1051
2alp	43	−3.3	−5.9	121	498
2sga	60	8.8	71	140	2969
5cpa	83	38	36	45	761

All atomic solvation parameters are in cal $Å^{-2}$ mol^{-1}

indeed what was found (Table 4) where with the exception of the atomic solvation parameter for carbon all the parameters were within 10% of their starting values. Thus this procedure should give a set of atomic solvation parameters which minimizes the molecular mechanics force.

We applied this method to the five chosen protein structures. The results, however, did not give a consistent set of solvation parameters, (Table 5); on the contrary, they are physically unreasonable, with charged atoms sometimes having positive values. The inconsistencies in these five sets of parameters are probably due to the fact that the forces are only calculated for the surface atoms. The surface atoms do not reflect the fact that the interior of the protein has very few charged atoms or that there are many more buried carbon atoms. The weighting of the propensity of a certain atom type to be on the surface is needed to scale the solvation parameters in a physically reasonable manner. This method cannot therefore be used.

II. *An iterative method*

As a second attempt to determine an independent set of atomic solvation parameters which would better complement the molecular mechanics energy, we tried scaling one of the existing sets of atomic solvation parameters. In comparing the atomic solvation parameters, there are two types of scales. The first scale is the ratio of the relative magnitudes of the atomic solvation parameters of C, N/O, N^+, O^-, S to each other. The second scale is an overall scale of the solvation energy to the molecular mechanics energy. These two scales are essential to picking the most appropriate set of atomic solvation parameters. If the charged and polar parameters

Table 6 Testing Atomic Solvation Parameters with 2 ps MD simulations on rat trypsin.

	$\Delta\sigma(C)$	$\Delta\sigma(N/O)$	$\Delta\sigma(N^+)$	$\Delta\sigma(O^-)$	$\Delta\sigma(S)$	*rms Å*	*% vol*	*Δ Esolv*
(w/o)K	13	−7	−87	−112	−3.6	2.0	7.5	36.6
1)	0.1	−7	−87	−112	−3.6	2.0	8.5	9
2)	13	−113	−87	−112	−3.6	2.3	17	−79
3)	195	−105	−87	−112	−3.6	1.7	0.7	−171
4)	195	−105	−900	−999	−3.6	2.0	2.9	−751

are too large and negative relative to the magnitude of the carbon, the protein structure will unfold during the course of the MD trajectory as we see in the MD trajectories with the water to vapor free energy of transfer atomic solvation parameters (Figures 4, 5, 7). If the ratio of the various atomic solvation parameters is correct the tertiary structure will remain stable during the course of MD trajectory. Of the experimental atomic solvation parameters the (w/o)K parameters best succeeded in maintaining the crystal structure conformation.

To determine which of the individual atomic solvation parameters are the most critical for maintaining the stability of the protein structure, we varied the (w/o)K parameters and performed several 2 ps MD simulations on the structure of rat trypsin. In Table 6 we show various sets of atomic solvation parameters which we tested and their deviation in the resulting structure after 2 ps MD to the crystal structure. The r.m.s deviation of all nonhydrogen atoms, the change in volume and the change in solvation energy are shown. Ideally, the solvation energy should not change form that of the crystal structure during the simulation. In the first simulation we significantly reduced $\Delta\sigma$(C) so as to determine its effect on stabilization of the protein during an MD trajectory. In 2 ps the structure expanded by 8.6% which is more than with the original parameters. Thus the $\Delta\sigma$(C) does affect the balance of the other atomic solvation parameters and maintain the proteins structure. In the second simulation the $\Delta\sigma$(N/O) was significantly decreased. This time in the 2 ps simulation the structure quickly expanded by 17%. Thus the relative value of $\Delta\sigma$(N/O) is essential to the stability of the protein structure. In the third simulation the values of $\Delta\sigma$(C) and $\Delta\sigma$(N/O) were multiplied by fifteen and the other three atomic solvation parameters were left alone. In 2 ps the structure's volume expanded by only 0.68% with an r.m.s. deviation of 1.67 Å for all atoms. The solvation energy, however, decreased by 171 kcal/mole. This simulation shows that the ratio of these two parameters is critical in maintaining the crystal structure conformation; however their magnitudes affect the stability of the solvation energy. This is further shown in the fourth simulation where all the values, except for that of sulfur, were greatly increased in magnitude. In this simulation the structure also maintained a similar volume to the crystal structure with an expansion of only 2.9%. The solvation energy, however, once again decreased enormously by 751 kcal/mole. These simulations prove that the ratio of the magnitude of $\Delta\sigma$(C) and $\Delta\sigma$(N/O) are critical in maintaining the structure at the crystal structure during a MD simulation.

Although the (w/o)K set of atomic solvation parameters maintain the crystal structure better than the other four parameter sets the scale of the solvation energy calculated from these parameters relative to the molecular mechanics energy is not ideal. This is apparent since the solvation energy gets worse by 40 kcals in the

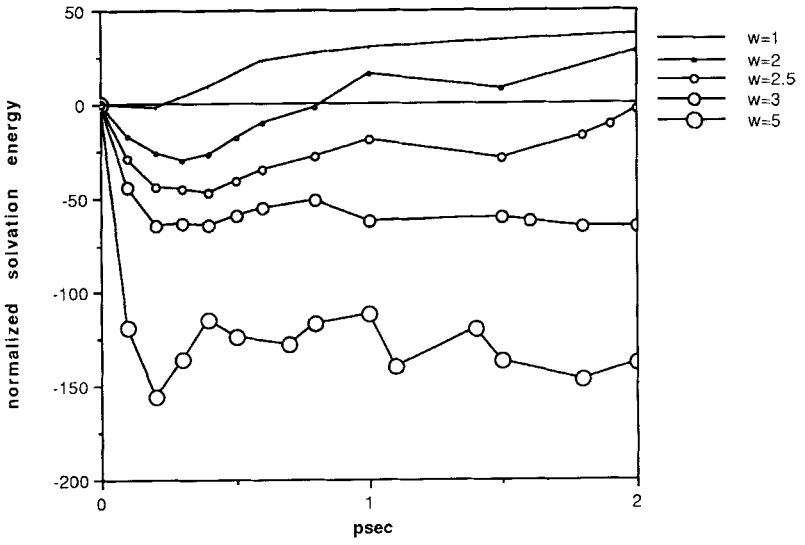

Figure 10 MD trajectories of 2 ps with (w/o)K scaled atomic solvation parameters normalized to the value of the solvation energy at the crystal structure.

first 2 ps of the MD trajectory (Figure 3). In all five MD trajectories (Figures 2-6) the solvation energy changed dramatically in the first 2 ps and then settled into an equilibrium state. Ideally the solvation energy would equilibrate at the value of the solvation energy of the crystal structure and remain relatively constant throughout the trajectory. This is expected since energetically it would be the most favorable to keep the relative exposure of hydrophilic and hydrophobic atoms constant. A series of 2 ps MD simulations were performed on the crystal structure of rat trypsin with various scales (ω).

$$E_{total} = E_{molecular\ mechanics} + \omega * E_{solvation\ energy}$$

These simulations were performed to determine which scale would minimize the gradient of the solvation energy. Figure 10 shows the solvation energy, normalized to the crystal structure solvation energy, with ω's ranging in value from $\omega = 1$ to $\omega = 5$ over a set of 2 ps trajectories. In the simulation with $\omega = 1$ the solvation energy increased steadily, whereas in the simulation with $\omega = 5$ the solvation energy decreased very quickly and stayed low. The simulation with $\omega = 2$ the energy first went down and then increased slowly over the 2 ps simulation. The best scale appears to be $\omega = 2.5$, (Table 6) where the solvation energy also initially went down and then slowly increased over the 2 ps. The structures volume expanded only by 4% in those 2 ps. This maintains the solvation energy at the crystal structure solvation energy while keeping the ratio of the various parameters at a ratio that keeps the structure intact. We will refer to these parameters which best complement the molecular mechanics force field as (mm)S atomic solvation parameters.

Table 7 Molecular mechanics complementary atomic solvation parameters (mm)S.

$\Delta\sigma(C)$	$\Delta\sigma(N/O)$	$\Delta\sigma(N^+)$	$\Delta\sigma(O^-)$	$\Delta\sigma(S)$	*rms Å*	*% vol*	*Δ Esolv*
32.5	−17.5	−217.5	−280.0	−9.0	1.8	4.1	−3

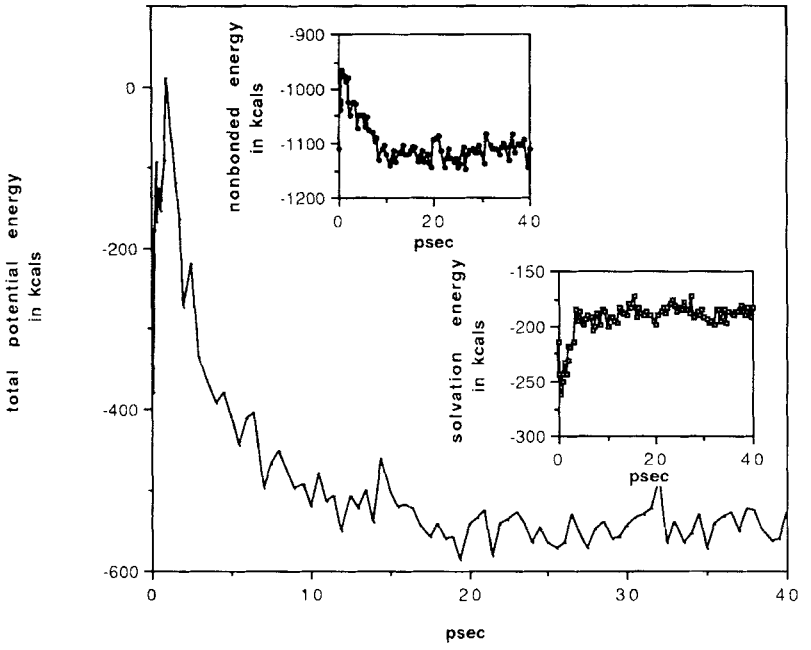

Figure 11 Components of the energy during the combined molecular mechanics solvation free energy 40 psec MD calculation on the crystal structure of rat trypsin with the (mm)S derived atomic solvation parameters.

Part D: Evaluation of the independently refined (mm)S atomic solvation parameters

I. *Extended molecular dynamics simulation on rat trysin*

The (mm)S set of atomic solvation parameters (Table 7) were further tested to determine their usefulness in the combined force field. A 40 ps MD trajectory was performed on the crystal structure of rat trypsin (Figure 11). Unlike any of the previous MD trajectories the solvation energy during the simulation initially dipped down and then increased equilibrating approximately 30 kcal/mole above the initial crystal structure solvation energy. This equilibration value of the solvation energy is 10 kcal/mole closer to the crystal structure value than the equilibration value of the solvation energy of any of the other MD trajectories. The deviations from the crystal structure were measured throughout the simulations (Figure 12a and b, Figure 13) and are compared with the deviations from the 40 ps simulations with molecular mechanics alone and with the (w/o)K parameters. The r.m.s. deviation to the crystal structure is consistently better than either (w/o)K simulation or the

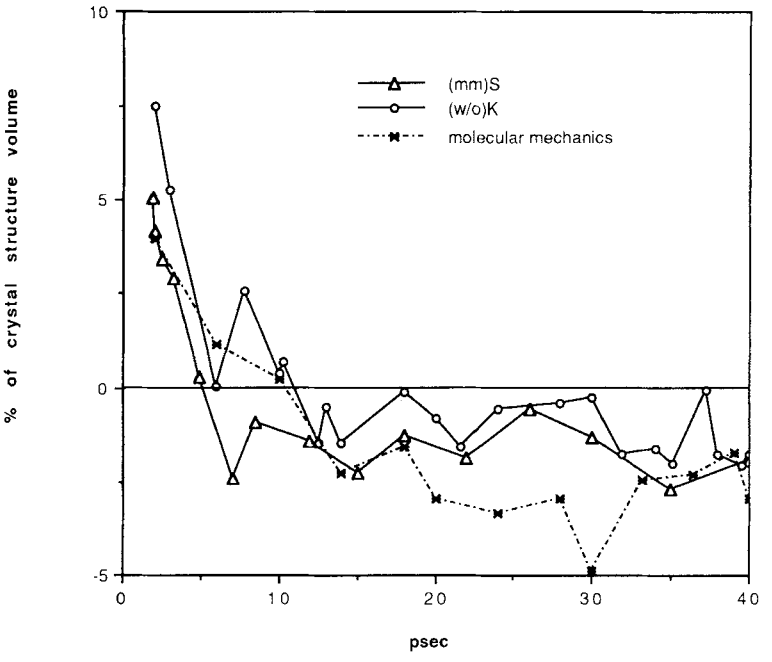

Figure 12(a) The change in volume relative to the crystal structure of rat trypsin during three MD trajectories: (mm)S, (w/o)K, and molecular mechanics.

molecular mechanics alone. The overall mean deviation between the crystal structure and the resulting structure after 40 ps MD with the (mm)S combined force field was 1.8 Å for all non-hydrogen atoms. Most of these differences are in the external loops. One of these loops which deviates is the calcium binding loop, in the crystal structure this loop is constrained by a calcium ion. This ion was not included in the calculation since when one is performing molecular modeling the position of an ion would probably not be known. The stability of this combined force field is shown in that for 80% of the resulting structure the mean deviation to the crystal structure is only 1.4 Å. Thus the (mm)S set of parameters clearly complements the molecular mechanics force field better than any of the experimental parameter sets.

II. *An MD simulation on another crystal structure: α-lytic protease*

As a further test of this (mm)S set of atomic solvation parameters we ran MD with the combined force field on the crystal structure of α-lytic protease. The rationale for this test was to determine whether or not we were indeed optimizing the atomic solvation parameters for the molecular mechanics force field rather than for the specific structure of rat trypsin. The structure of α-lytic protease was chosen since it is a globular protein whose structure has been well determined to high resolution. Two trajectories of 10 ps each were run; one with molecular mechanics alone and the other with the combined force field with the (mm)S parameters. At the end of 10 ps of MD with the combined force field and the (mm)S parameters

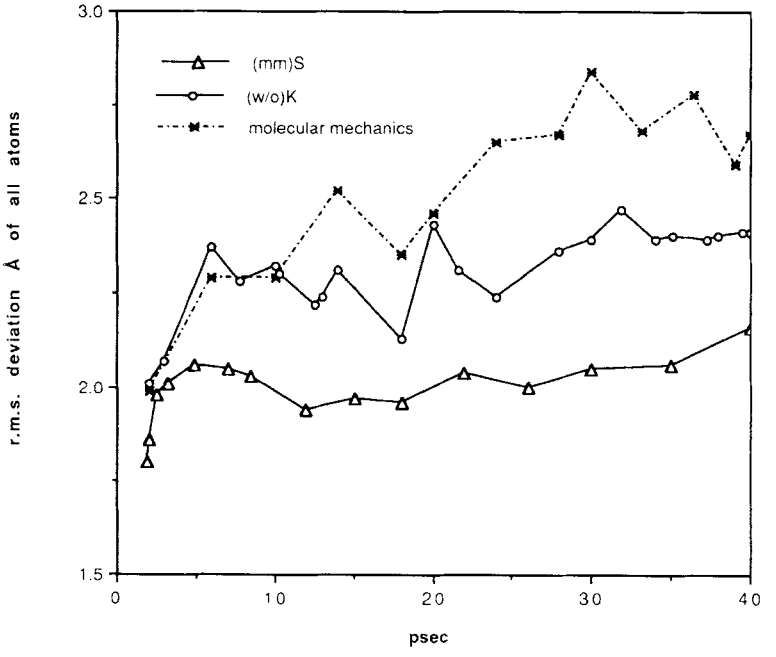

Figure 12(b) The rms deviation relative to the crystal structure of rat trypsin during the same three trajectories of all non-hydrogen atoms.

the structure was expanded by 2.7% and had an r.m.s. deviation of 1.10 Å for all atoms compared to the crystal structure. The solvation energy had increased by only 2 kcals compared to the value of 182 kcals of the crystal structure. In comparison, at the end of 10 ps of MD with only the molecular mechanics force field the structure also expanded by 2.7% and had an r.m.s. deviation of 1.42 Å of all atoms from the crystal structure. The largest difference between the two simulations, however, is that the solvation energy of the structure after 10 ps of MD with only molecular mechanics force field was 74 kcals worse in energy than it was to start with. Thus the (mm)S parameters are optimized to the molecular mechanics force field and their level of accuracy is not limited to just rat trypsin.

III. *A further LECS test*

To further test the (mm)S set of atomic solvation parameters a series of LECS conformational searches were performed on the structure of rat trypsin. We have previously [6] developed and shown LECS to be able to successfully predict the crystal structure conformation of internal side chains with the molecular mechanics force field alone. On external sidechains however LECS with molecular mechanics alone was not sufficient in predicting the conformation of these side chains, because solvent effects were not being taken into account. A total of 12 external residues (the 3 previously tested plus 9 others) were conformationally searched with the combined force field with the (mm)S set of atomic solvation parameters and compared to the crystal structure and their predicted conformation from LECS with

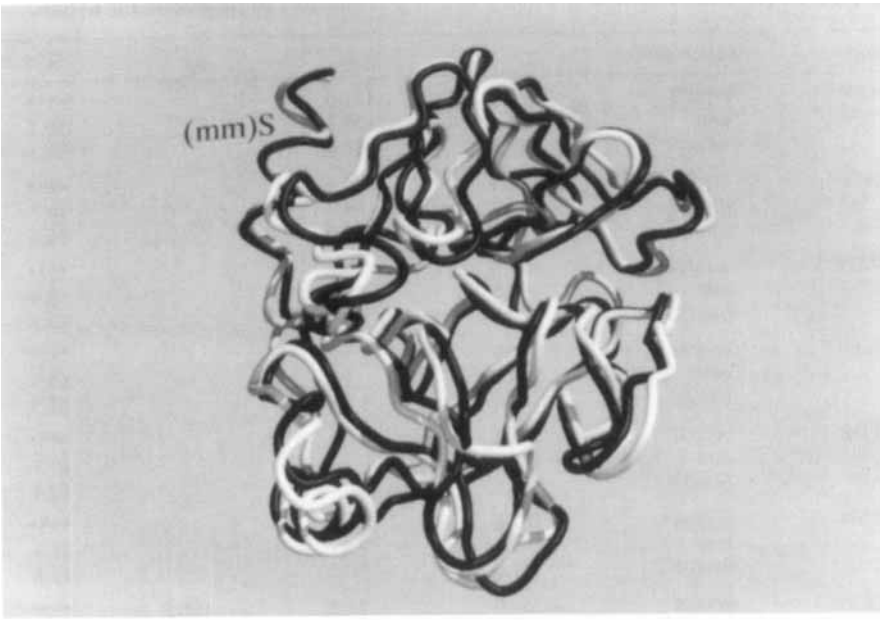

Figure 13 The superimposition of the rat trypsin crystal structure in white with the resulting structure after 40 ps with the resulting structure in black after 40 ps of MD with the (mm)S atomic solvation parameters. (see colour plates)

only the molecular mechanics force field (Table 8). The three previously tested side chains are shown in yellow in Figure 1. Of the twelve residues, two of them, HIS 53 and THR 102 were predicted by the LECS with the combined force field much more accurately then with LECS with molecular mechanics alone. For ASN 130 both conformational searches correctly predicted the crystal structure conformation. Neither of the conformational searches, however, were able to correctly predict the conformation of the remaining 9 residues. This indicates that although this method more accurately complements the molecular mechanics force field it is still not ideal for determining the conformation of all external residues.

CONCLUSIONS

In this study we evaluated a combined molecular mechanics solvation free energy force field where the analytical derivatives of solvation free energy are added to the molecular mechanics force field of AMBER. We seek to optimize the coupling between the molecular mechanics energy and the solvation free energy by determining which set of five atomic solvation parameters most accurately mimic the effects of solvent without having solvent explicitly in the calculation. A distance dependent dielectric constant was used to buffer the long range electrostatic interactions so as to mimic the long range buffering effect of explicit solvent. We evaluated 4 sets of experimental atomic solvation parameters and found that

Table 8 LECS dihederals and volume overlap with crystal structure conformation for (mm)S.

Residue	*Structure*	χ_1	χ_2	χ_3	*% overlap*
32 ASP	crystal	74.1	−177.0		****
	mm	63.4	126.8		58.3
	(mm)S	−47.9	−150.6		58.3
53 HIS	crystal	−173.0	61.8		****
	mm	−80.7	59.4		34.1
	(mm)S	180.0	57.0		90.1
66 ASN	crystal	−59.3	138.2		****
	mm	67.3	−79.9		50.4
	(mm)S	170.3	61.7		46.3
70 ILE	crystal	177.8			****
	mm	54.3			63.9
	(mm)S	53.0			63.9
102 THR	crystal	64.2			****
	mm	−177.3			67.7
	(mm)S	48.5			92.6
130 ASN	crystal	167.6	74.7		****
	mm	175.1	73.7		76.4
	(mm)S	177.6	80.2		69.6
145 GLN	crystal	−154.9	74.2	167.5	****
	mm	−65.5	−58.6	130.4	75.6
	(mm)S	−59.7	−62.6	128.6	73.2
147 ASP	crystal	−67.8	−28.0		****
	mm	69.9	140.8		57.8
	(mm)S	65.5	122.7		65.5
151 SER	crystal	−71.4			****
	mm	66.5			83.3
	(mm)S	60.0			85.2
158 ASP	crystal	−157.4	−135.2		****
	mm	116.9	81.9		54.6
	(mm)S	58.2	−171.6		58.5
184 ASN	crystal	−74.1	−72.8		****
	mm	65.9	−111.6		55.4
	(mm)S	65.0	−103.3		59.0
195 TYR	crystal	168.6	94.2	−179.3	****
	mm	62.1	−80.1	160.6	56.8
	(mm)S	60.9	−80.6	154.4	43.0

the ones derived from water to octanol free energy of transfer data were more compatible with the molecular mechanics force field than those derived from water to vapor free energy of transfer data. Of the four (w/o)K, which are derived from free energy of transfer data of tripeptides between water and octanol [23, 24], best maintained the crystal structure of rat trypsin during a 40 ps MD trajectory. The atomic solvation parameters derived from the water to vapor free energy of transfer overestimates the hydrophilicity of the neutral nitrogens and oxygens and underestimates the hydrophobicity of the carbon; this results in the structure of rat trypsin unfolding during a 40 ps MD trajectory.

To find the best fit of the atomic solvation parameters to the molecular mechanics force field, we have shown that there are two scales. The first is the relative ratio of the individual parameters to each other and the second is the overall scale of the solvation energy to the molecular mechanics energy. The relative ratio of the parameters maintains the structure of the molecule, whereas the overall scale maintains the equilibration solvation energy of the structure. We have shown that of the five parameters the values of carbon and neutral oxygen and nitrogen are the most important to the stability of the molecule. The set of parameters which we have found to maintain the protein's structure the best are those of (w/o)K times a scale of 2.5 which gives the (mm)S parameters. We have shown that these parameters maintain both the crystal structure and its solvation energy for the globular proteins rat trypsin and α-lytic protease. These results are very encouraging that it is possible to represent the effects of solvent without having solvent explicitly in the calculation.

Simulating the effects of solvent without having it explicitly in the calculation makes energetic conformational searching in protein structure feasible. The use of explicit water requires an equilibration period which makes comparison of the relative energy of various conformations impossible. However, the 40 ps MD simulation on the structure of rat trypsin with a 4 Å shell of water maintained the crystal structure better than the 40 ps MD simulation with the combined force field and the (mm)S atomic solvation parameters by an average reduction of 0.3 Å in the r.m.s. deviations and an average change of 1% in volume. Thus explicit water is a more accurate representation of solvent when performing simulations on a protein whose structure is known. However, often when modeling the unknown structure of a protein the use of explicit water is unfeasible. This combined force field then provides a more accurate representation of the effects of solvent when modeling the unknown structure of a protein.

Further optimization of the combined force field is, however, desirable and there exist several possible methods that may improve it. One possible method is to define more than five atom types to describe the types of atoms found in proteins. These new types could take into account what a particular atom is immediately bonded to, which directly affects its hydrophilicity. Another possible way to improve this combined force field is to use an additional Poisson-Boltzmann continuum solvent term to better handle the electrostatic effects of explicit solvent. The analytical derivatives of this energy would need to be added to the other forces so as to take advantage of this additional term. These additional optimizations may improve the success rate of the lowest energy conformational search (LECS) on predicting the conformation of external side-chains. Another place where this combined force field could be improved is in the speed of the calculation. Currently, the combined force field is eight times slower than molecular mechanics alone. One possible way of increasing the speed of the calculation is to replace the currently used analytical derivatives of the surface area with respect to position with a rapid method of calculating those same derivatives recently described by Perrot *et al.* [40] which has been suggested to be significantly faster. Thus with the improvement of both the accuracy and the speed of this combined force field the practicality of this technique will prove to be very useful in predicting protein structure.

Acknowledgements

This research was supported by NIH grants GM-24485 (R.M.S.) and GM-29072 (P.A.K.). Thanks to A. Kim and F.C. Szoka for their experimental results, M. Schiffer for organizational advice and J.A. Newdoll for help with graphics. The graphics pictures were made using the Midas Plus program written by Conrad Huang, Eric Pettersen, and Greg Couch at the UCSF Computer Graphics Laboratory and NEON by Thomas Hines of Genentech.

References

[1] D. Eisenberg, A.D. McLachlan, "Solvation energy in protein folding and binding" *Nature* **319**, 199-203 (1986).

[2] D. Eisenberg, M. Wesson, and M. Yamashita, "Interpretation of Protein Folding and Binding with Atomic Solvation Parameters" *Chemical Scripta* **29A**, 217-221 (1989).

[3] J. Novotny, A.A. Rashin, R.E. Bruccoleri, "Criteria that discriminate between native proteins and incorrectly folded models." *Proteins* **4**, 19-30 (1988).

[4] L. Chiche, C. Gaboriaud, A. Heitz, J.P. Mornon, B. Castro, P.A. Kollman, "Use of restrained molecular dynamics in water to determine three-dimensional protein structure: Prediction of the three-dimensional structure of Ecballium elaterium trypsin inhibitor II." *Proteins*, **6**, 405-417 (1989).

[5] L. Chiche, L.M. Gregoret, F.E. Cohen, P.A. Kollman, "Protein model structure evaluation using the solvation free energy of folding". *Proc. Natl. Acad. Sci.* **87**, 3240-3243 (1990).

[6] C.A. Schiffer, J.W. Caldwell, P.A. Kollman, R.M. Stroud, "Prediction of homologous protein structures based on conformational searches and energetics" *Proteins* **8**, 30-43 (1990).

[7] L. Wesson, and D. Eisenberg, "Atomic Solvation Parameters Applied to Molecular Dynamics of Proteins in Solution." *Protein Science* **1**, 227-235 (1992).

[8] B.R. Brooks, R.E. Bruccoleri, B.D. Olafson, D.J. States, S. Swaminatham, M. Karplus "CHARMM: A program for macromoleculer energy minimization, and dynamics calculations." *J. Comp. Chem.* **4**, 187-217 (1983).

[9] T.N. Earnest, E.B. Fauman, R.M. Stroud, "1.59 Å structure of trypsin at 120° K: Comparison of low temperature and room temperature structures." *Proteins* **10**, 171-187 (1991).

[10] M. Fujinaga, L.T.J. Delbaere, G.D. Brayer, M.N.G. James, "Refined structure of alpha-lytic protease at 1.7 angstrom resolution. Analysis of hydrogen bonding and solvent structure." *J. Mol. Biol.* **184**, 479 (1985).

[11] C.A. Schiffer, J.W. Caldwell, R.M. Stroud, P.A. Kollman, "Inclusion of solvation free energy with molecular mechanics energy: alanyl dipeptide as a test case." *Protein Science* **1**, 396-400 (1992).

[12] M. Avignon, C. Garrigou-Lagrange, P. Bothorel, "Conformational analysis of dipeptides in aqueous solution. II. molecular structure of glycine and alanine dipeptides by depolarized rayleigh scattering and laser raman spectroscopy." *Biopolymers*, **12**, 1651-1669 (1973).

[13] V. Madison, K.D. Kopple, "Solvent-dependent conformational distributions of some peptides." *J. Amer. Chem. Soc.*, **102**, 4855-4863 (1980).

[14] P.J. Rossky, M. Karplus, "Solvation: A molecular dynamics study of a dipeptide in water." *J. Amer. Chem. Soc.*, **101**, 1913-1937 (1979).

[15] A.T. Hagler, D.J. Osguthorpe, B. Robson, "Monte Carlo simulation of water behavior around the dipeptide N-acetylalanylN-methylamide." *Science* **208**, 599-601 (1980).

[16] B.M. Pettitt, M. Karplus, "The potential of mean force surface for the alanine dipeptide in aqueous solution: A theoretical approach." *Chem. Phys. Let.*, **121**, 194-201 (1985).

[17] W.F. Lau, B.M. Pettitt, "Conformations of the Glycine Dipeptide." *Biopolymers*, **26**, 1817-1831 (1987).

[18] A.G. Anderson, J. Hermans, "Microfolding: conformational probability map for the alanine dipeptide in water from molecular dynamic simulations." *Protein*, **3**, 262-265 (1988).

[19] S.J. Weiner, P.A. Kollman, D.T. Nguyen, D.A. Case, "An all atom force field for simulations of proteins and nucleic acids." *J. Comp. Chem.* **7(2)**, 230-252 (1986).

[20] T.J. Richmond, "Solvent accessible surface area and excluded volume in proteins: Analytical equations for overlapping spheres and implications for the hydrophobic effect." *J. Mol. Biol.* **178**, 63-89 (1984).

[21] A. Shrake, J.A. Rupley, "Environment and exposure of solvent of protein atoms, lysozyme and insulin." *J. Mol. Biol.* **79**, 351-371 (1973).
[22] J.L. Fauchere, V. Pliska, "Hydrophobic parameters p of amino-acid side chains from the partitioning of N-acetyl-amino-acid amides." *Eur. J. Med. Chem.- Chim. Ther.* **18**, 369-375 (1983).
[23] A. Kim, "Amino Acid Side Chain Contribution to Free Energy of Transfer of Tripeptides from Water to Octanol." PhD Dissertation, UCSF Department of Pharmaceutical Chemistry, Advisor: Szoka, F. (1990).
[24] A. Kim, F.C. Szoka, "Amino acid side chain contributions to free energy of transfer of tripeptides from water to octanol." *Pharmaceutical Research* **9**, 504-514 (1992).
[25] R.V. Wolfenden, P.M. Cullis, C.C.F. Southgate, "Water, Protein Folding, and the Genetic Code." *Science* **206**, 575-577 (1979).
[26] R. Wolfenden, L. Andersson, P.M. Cullis, C.C.B. Southgate, "Affinities of amino acid side chains for solvent water." *Biochemistry* **20**, 849-855 (1981).
[27] J. Hine, P.K. Mookerjee, "The intrinsic hydrophilic character of organic compounds. Correlations in terms of structural contributions." *J. Org. Chem.* **40**, 292-298 (1975).
[28] J. Kyte, R.F. Doolittle, "A simple method for displaying the hydropathic character of a protein" *J. Mol. Biol.* **157**, 105-132 (1982).
[29] K. Sharp, A. Nicholls, R. Friedman, and B. Honig, "Extracting Hydrophobic Free Energies From Experimental Data: Relationship to Protein Folding and Theoretical Models." *Biochemistry* **30**, 9686-9697 (1991).
[30] J.A.V. Butler, C.N. Ramchandani, "The solubility of non-electrolytes. Part II. The influence of the polar group on the free energy of hydration of aliphatic compounds." *J. Chem. Soc.* 952-955 (1935).
[31] N.C. Deno, H.E. Berkeimer, "Activity coefficients as a function of structure and media." *J. Chem. Eng. Data* **1**, 1-5 (1960).
[32] J. Hine, R.D. Jr. Weimar, "Carbon Basicity." *J. Amer. Chem. Soc.* **87**, 3387-3396 (1965).
[33] C. McAuliffe, "Solubility in water of paraffin, cycloparaffin, olefin, acetylene, cycloolefin, and aromatic hydrocarbons." *J. of Phys. Chem.* **70**, 1267-1275 (1966).
[34] R.C. Weast, Ed. CRC Handbook of Chemistry and Physics, Chemical Rubber Co., Cleveland OH (1972).
[35] G.H. Parsons, C.H. Rochester, A. Rostron, P.C. Sykes, "The thermodynamics of hydration of phenols" *J. Chem. Soc. Perkins* **II** 136-138 (1972).
[36] J. Guenot, and P.A. Kollman, "Molecular dynamics studies of a DNA binding protein: 2. An evaluation of implicit and explicit solvent models for the molecular dynamics simulation of the Escherichia coli trp Repressor." *Protein Science* **1**, 1185-1205 (1992).
[37] J. Finer-Moore, A.A. Kossiakoff, J.H. Hurley, R.M. Stroud, "Water structure in crystals of trypsin determined by xray and neutron diffraction." *Proteins* **12**, 203-222 (1992).
[38] J. Moult, F. Sussman, M.N.G. James, "Electron density calculations as an extension of protein structure refinement. Streptomyces Griseus Protease at 1.5 Angstroms Resolution." *J. Mol. Biol.* **182**, 555-566 (1985).
[39] D.C. Rees, M. Lewis, W.N. Lipscomb, "Refined crystal structure of carboxypeptidase A at 1.54 angstroms resolution." *J. Mol. Biol.* **168**, 367-387 (1983).
[40] G. Perrot, B. Cheng, K.D. Gibson, J. Vila, K.A. Palmer, A. Nayeem, B. Maigret, H.A. Sheraga, "MSEED: A program for the rapid analytical determination of accessible surface areas and their derivatives." *J. Comp. Chem.* **13**, 1-11 (1992).