Protein Science (1992), *1*, 1185–1205. Cambridge University Press. Printed in the USA.

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

JEANMARIE GUENOT AND PETER A. KOLLMAN
Department of Pharmaceutical Chemistry, University of California at San Francisco, San Francisco, California 94143
(RECEIVED October 23, 1991; REVISED MANUSCRIPT RECEIVED March 27, 1992)

Abstract

Although aqueous simulations with periodic boundary conditions more accurately describe protein dynamics than in vacuo simulations, these are computationally intensive for most proteins. *Trp* repressor dynamic simulations with a small water shell surrounding the starting model yield protein trajectories that are markedly improved over gas phase, yet computationally efficient. Explicit water in molecular dynamics simulations maintains surface exposure of protein hydrophilic atoms and burial of hydrophobic atoms by opposing the otherwise asymmetric protein–protein forces. This properly orients protein surface side chains, reduces protein fluctuations, and lowers the overall root mean square deviation from the crystal structure. For simulations with crystallographic waters only, a linear or sigmoidal distance-dependent dielectric yields a much better trajectory than does a constant dielectric model. As more water is added to the starting model, the differences between using distance-dependent and constant dielectric models becomes smaller, although the linear distance-dependent dielectric yields an average structure closer to the crystal structure than does a constant dielectric model. Multiplicative constants greater than one, for the linear distance-dependent dielectric simulations, produced trajectories that are progressively worse in describing *trp* repressor dynamics. Simulations of bovine pancreatic trypsin were used to ensure that the *trp* repressor results were not protein dependent and to explore the effect of the nonbonded cutoff on the distance-dependent and constant dielectric simulation models. The nonbonded cutoff markedly affected the constant but not distance-dependent dielectric bovine pancreatic trypsin inhibitor simulations. As with *trp* repressor, the distance-dependent dielectric model with a shell of water surrounding the protein produced a trajectory in better agreement with the crystal structure than a constant dielectric model, and the physical properties of the trajectory average structure, both with and without a nonbonded cutoff, were comparable.

Keywords: aqueous simulation; bovine pancreatic trypsin inhibitor; counterions; hydrogen bond; nonbonded cutoff; protein electrostatics; sigmoidal dielectric; solvent free energy

This is the second paper in a series of computational studies on the *Escherichia coli trp* repressor, a DNA-binding protein. The first study (Howard & Kollman, 1991) examined the usefulness of molecular dynamics (MD) for describing the structural and dynamic differences between the *trp* repressor and *trp* aporepressor proteins. MD was shown to provide a qualitative description of the flexibilities of both the tryptophan-bound and unbound states that was in good agreement with the crystallographic *B*-factors for *trp* repressor (Schevitz et al., 1985) and the NMR nuclear Overhauser effect structural uncertainties (Jardetzky, pers. comm.) for both the tryptophan-bound and unbound protein models, despite the short time scales used and the lack of periodic boundary conditions. Our long-term interest in studying site-specific mutants, conformational changes, and the relative binding free energies of tryptophan and its analogues to the protein effector site prompted us to use the *trp* repressor as our model system to evaluate simple solvent models for protein dynamics simulations.

Reprint requests to: Peter A. Kollman, Department of Pharmaceutical Chemistry, University of California at San Francisco, San Francisco, California 94143.

How accurately the dynamics describes the physical properties of a macromolecule in the condensed phase depends upon the level of atomic detail included in the model system. Simulations without explicit solvent err in their aqueous representation of macromolecules by not adequately exposing protein polar surface groups to the solvent environment and by allowing hydrophobic groups to become overexposed. Solvent-exposed side chains, which interact with water in the protein crystal, are unable to find their enthalpic counterparts in vacuo, and will readily establish interactions within the protein that may globally distort the structure. These problems can be avoided by simulating the molecule in water, with periodic boundary conditions (Levitt & Sharon, 1988), which eliminates the artifacts arising from the system–vacuum interface. However, the size of the protein *and* the quantity of water needed to solvate the molecule contribute to the computation time, which increases approximately as n^2, where n is the number of atoms in the system. As a consequence, protein simulations with the solvent environment represented by explicit waters may become unmanageably large. The computation time can further be reduced to be proportional to n by using a cutoff for the nonbonded interactions.

Molecular modeling of the solvent environment must conform to experimental reality as well as possible to insure that the dynamics trajectories can provide useful physical information. However, due to assumptions inherent in the method and limitations imposed to reduce the computational burden, the description of intramolecular interactions obtained by MD simulations is approximate, and increasing the detail in the molecular model may not necessarily provide better results. Our objective for this study was to find a computationally efficient solvent model that accurately describes the experimental properties of the *trp* repressor. We will use this model in later studies to study the structural and energetic effects of a variety of active-site ligands, site-specific mutants, and starting conformations for comparative *trp* repressor simulations. Decreasing the computational expense per dynamics calculation will potentially increase the length of the dynamic trajectories and the number of calculations per study. This is highly desirable for a protein like *trp* repressor, which is so well studied experimentally and has many exciting structural avenues to explore.

The *trp* repressor is an all-helical dimer with no cysteine and therefore, no disulfide bridges. Each 107-amino acid monomer contains six helices that intertwine to form a dimer with an unusual monomer–monomer interface (Schevitz et al., 1985). The *trp* repressor regulates the transcription of the trpR, aroH, and trpABCDE operons by preventing the binding of RNA polymerase to the promoter sites (Brown, 1968; Rose et al., 1973; Squires et al., 1975; Gunsalus & Yanofsky, 1980). These three DNA regulatory sequences encode the aporepressor protein (Morse & Yanofsky, 1969), one enzyme involved in aromatic amino acid synthesis (Pittard et al., 1969), and several enzymes directly involved in tryptophan biosynthesis (Yanofsky, 1971), respectively. The protein interacts nonspecifically with DNA, in the absence of bound tryptophan, and with sequence specificity upon binding two molecules of L-tryptophan, or a tryptophan analogue, noncooperatively (McGeough et al., 1973; Rose et al., 1973; Shimizu et al., 1973; Rose & Yanofsky, 1974; Marmorstein et al., 1987; Marmorstein & Sigler, 1989). The allosteric change that occurs imparts sequence specificity by repositioning the conserved helix-turn-helix structural motifs or DNA reading heads for the protein–DNA interaction (Schevitz et al., 1985; Zhang et al., 1987; Lawson et al., 1988; Otwinoski et al., 1988). Because the conformational change occurs in a flexible region of the protein, which is also where tryptophan binds, and because this region lies on the DNA-binding surface and interacts with the DNA, it is crucial that solvent effects on the protein dynamics are modeled realistically to ensure the usefulness of any model for future calculations. In addition, with such an unusual shape, and no disulfide bonds, we expected that maintaining the structural integrity of the *trp* repressor with limited solvent would be quite a challenge, and that these results can be applied generally to other protein dynamics simulations.

We sought to distinguish between those protocols that explicitly incorporate water into the starting model and hence exert their stabilizing effect on the protein through both electrostatic and van der Waals interactions, primarily with surface residues, and those models that involve an implicit description of solvent, which mimic charge screening by altering the Coulomb potential only. The important difference between the implicit and explicit solvent models is that water can specifically orient the polar surface side chains through protein–solvent interactions and keep them from forming strong protein–protein interactions. The electrostatic interactions of the explicit water may contribute to protein stability, indirectly. In a minimization study of trimethoprim binding to dihydrofolate reductase, Dauber-Osguthorpe and co-workers (1988) found a significant interaction energy between trimethoprim and the crystallographic waters included in their simulation model, independent of whether the waters were in contact with the ligand. Explicit ions, in addition to locally stabilizing the protein through electrostatic and van der Waals interactions, are important in reducing numerical errors that arise from solving the equations of motion for a non-neutral system. A detailed treatise on modeling electrostatic interactions in proteins is beyond the scope of this work. Harvey (1989) and Sharp and Honig (1990) offer two excellent reviews on this subject.

The conformationally dependent properties used to evaluate the agreement of the trajectory average struc-

tures with the crystal structure were solvent exposure of hydrophobic and hydrophilic atoms, root mean square (RMS) deviation from the crystal structure, RMS fluctuations from the average structure, volume and surface area changes, and radius of gyration. MD temperatures of both the protein and solvent were also calculated. Including waters and ions in the starting model dramatically improved the linear distance-dependent dielectric and the constant dielectric simulations. Of the models examined, the constant and linear distance-dependent dielectric simulation models with the solvent shell best retained the physical properties of the crystal structure. Surprisingly, the linear distance-dependent dielectric model kept the trajectory average structure much closer to the crystal structure than did the constant dielectric model. Explicit water in the simulation model decreased the RMS fluctuations from the average structure, and the trajectory remained closer to the crystal structure. This underlines the importance of including *some* explicit water in protein dynamics simulations.

Simulation models with crystallographic waters and ions only, and different constant and distance-dependent dielectric constants were examined to explore the effect of changing the relative contribution of the Coulomb potential to the total potential function with the Weiner et al. (1986) all-atom force field. For the purposes of this work, constant dielectric refers to $\varepsilon = 1.0$, linear distance-dependent dielectric describes $\varepsilon = 1.0$ Rij (where 1.0 is the dielectric multiplicative constant), and for all other dielectrics, ε is stated explicitly. The contribution of the electrostatic term to the total potential energy is larger for the constant dielectric than for the distance-dependent dielectrics and decreases proportionately with increasing multiplicative constant for the linear distance-dependent dielectrics. Our results show that using either a linear or sigmoidal distance-dependent dielectric keeps the trajectory much closer to the crystal structure than does using a constant dielectric. This difference between the dielectric functions is more dramatic when just explicit crystallographic water is included in the simulation model than when a 4-Å water shell is also included. The larger the multiplicative constant for the linear distance-dependent dielectric, the larger the fluctuations from the average structure and the greater the RMS deviation of that model from the crystal structure. Our results with the distance-dependent dielectrics suggest that the stronger weighting of the short-range interactions is primarily responsible for keeping the trajectory close to the crystal structure.

Bovine pancreatic trypsin inhibitor (BPTI) was used to further explore the differences between the constant and linear distance-dependent dielectrics with an explicit water model. BPTI was chosen as a second model, because due to its small size, it was feasible to calculate 50-ps trajectories without a nonbonded cutoff. The effects of a nonbonded cutoff on evaluation of the linear distance-dependent and constant dielectric simulations was evaluated. Four BPTI trajectories were calculated using either a constant or linear distance-dependent dielectric with either an 8-Å or infinite nonbonded cutoff. As with *trp* repressor, for the simulations with nonbonded cutoffs, the distance-dependent model remained closer to the crystal structure during the trajectory, than did the constant dielectric model. However, the average properties of the distance-dependent dielectric models, and not the constant dielectric model with an 8-Å cutoff, were more similar to those of the constant dielectric simulation with an infinite nonbonded cutoff. The constant dielectric simulation was much more sensitive to the nonbonded cutoff than was the distance-dependent dielectric simulation. We supplemented this analysis with an examination of the hydrogen bonds formed in the trajectory average structures and in each coordinate dump during the postequilibration phase of the trajectory.

In order to obtain good structures from MD simulations, it is important to strike a balance among the protein-protein, protein-solvent, and solvent-solvent interactions. Because the water-vacuum interface is present when periodic boundary conditions are not applied, the water in the solvent shell may not accurately describe the same water surrounding the protein in a periodic box simulation. Locally, the vacuum would directly affect the water-water and the water-protein interactions and will likely lead to differences between the protein simulations. To examine the global effect of the solvent-solvent interactions with a linear distance-dependent dielectric, we compared the dynamics trajectories for a periodic box of TIP3P water (Jorgensen, 1981), with a linear distance-dependent and a constant dielectric. We were particularly interested in characterizing the effect of the dielectrics on the energy and density of bulk water. As expected, the distance-dependent dielectric increased the density of bulk water over that of the constant dielectric model by allowing the intermolecular water oxygens to come closer together, with lower energy.

Results

The trp repressor

The *trp* repressor models with a water shell required less than twice as much computer time to complete (Table 1) than the in vacuo simulations, although all 12 calculations were not run on the same machine, and the conversion factors used to transform the times into Cray YMP equivalents are approximate. A factor of between 2 and 4 would be a more conservative estimate of the difference in computational times between the in vacuo and solvated simulations. The physical properties of the 12 30-ps average structures have been summarized in Table 3.

Table 1. *The trp repressor electrostatic models*[a]

Models	ε	X-ray waters	Counterions	Water shell	Neutralized R, E, D, K	Computer	CPU time (h/ps)
PN1	1				N	Convex C1	0.20
PXC1	1	X	C			Stellar	0.42
PXCS1	1	X	C	S		Cray-YMP	0.36
PXC50	50	X	C			Stellar	0.81[b]
P1r	1.0 Rij					Covex C1	0.14
PX1r	1.0 Rij	X				Stellar	0.51[b]
PXC1r	1.0 Rij	X	C			Convex C1	0.20
PC1r	1.0 Rij		C			Convex C1	0.13
PXCS1r	1.0 Rij	X	C	S		Cray-YMP	0.34
PXC2r	2.0 Rij	X	C			Stellar	0.21
PXC4.5r	4.5 Rij	X	C			Stellar	0.33
PXCh	SIGM	X	C			FPS 264	0.46

[a] Definitions of the *trp* repressor electrostatic models used in the text. R, E, D, and K in the last column are the one-letter codes for the charged amino acids arginine, glutamate, aspartate, and lysine. The letters X, C, and S in our model-naming convention indicate whether the model contains X-ray waters, 14 counterions, and/or a 4-Å water shell. N indicates that the charged amino acids were neutralized (see text). Ten cations and four anions were used in the counterion simulations (C). One hundred and eighteen X-ray waters were included in the simulations with X-ray waters (X). The water shell (S), which surrounded the protein, X-ray waters, and ions, was 4 Å thick. The dielectric constant, ε, has been appended to each name for clarity in the text, tables, and graphs, as indicated in this table. Central processing unit (CPU) time is recorded as Cray YMP equivalents. The CPU time to complete 40 ps of molecular dynamics for each simulation model was divided by 40 and multiplied by a conversion factor. The conversion factors used were: Stellar = 1/10.5 Cray YMP, Convex C1 = 1/15 Cray YMP, and FPS = 1/9 Cray YMP.

[b] Inconsistencies in CPU times are likely due to variations in machine loads.

The all heavy atom CORE and CRYSTAL deviations follow the same trends as the NCαCO backbone. The RMS fluctuations from the 30-ps average structures are in good qualitative agreement with the crystallographic *B*-factors. When the CORE RMS deviations and fluctuations are large, the CRYSTAL RMS deviations are also large. The all heavy atom RMS deviation follows the same trends as the NCαCO and Cα backbone RMS deviations. With regard to selecting the simulation model that affords the best description of the protein dynamics, the large CORE and CRYSTAL RMS values in Table 3 for models PN1, P1r, PC1r, PX1r, PXC50, PXC2r, PXC4.5r, and PXC1 re-

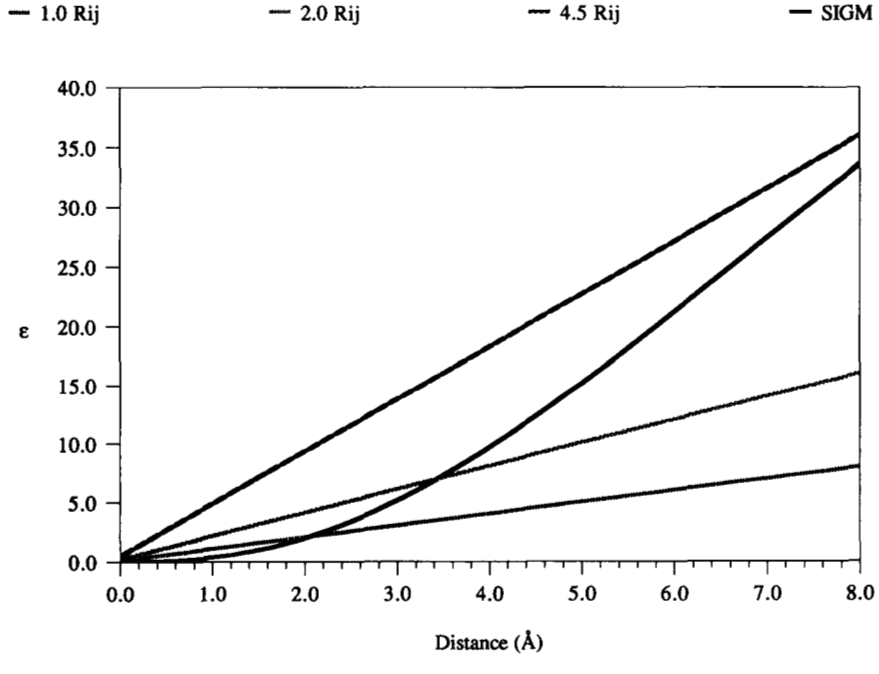

Fig. 1. Distance-dependent dielectrics as a function of distance. Since the nonbonded cutoff is residue-based, the electrostatic interactions are likely to extend beyond 8 Å. SIGM is defined in the text. Refer to Table 1 for the model definitions.

moved them from further consideration (compare to Table 2). Elimination of these models based on the 30-ps average structure RMS deviations is further substantiated by Figure 2, which shows that the CORE RMS deviations of the average structures calculated over 5-ps intervals are unreasonably large for these models. Although only the CORE RMS deviations from the crystal structure have been presented in Figure 2, the same trends were also present for the CRYSTAL all heavy atom and Cα RMS deviations. The N-termini and A helices remained extended in the four best *trp* repressor models. The other models showed dramatic structural changes of the A, D, and E helices toward the protein core. Of the models that provide a poor description of *trp* repressor dynamics, PXC50, PXC2r, PXC4.5r, and PXC1 differed in dielectric only, from two relatively good models, PXCh and PXC1r. The fact that dielectric alone is not enough to provide good conformational results in a protein simulation is exemplified by the series PC1r, PXC1r, and PXCS1r, where increasing the amount of water in the simulation shows a marked improvement on the conformational results obtained (Table 3). Comparing PX1r, PC1r, and P1r shows that including crystallographic waters in the simulation model leads to a greater improvement of the gas phase model than is found by simply including counterions, although including counterions when only crystallographic water is present, is important for obtaining even better results (compare PX1r, PC1r, and PXC1r). Neutralizing the charged amino acids and using a constant dielectric model without any water produces low CORE RMS deviations from the crystal structure but substantially higher CRYSTAL RMS deviations, which, in addition to the higher solvation free energies compared to the PXCS1 model (Table 3), emphasizes the importance of the orientational effect of water on keeping hydrophilic functional groups exposed to the solvent, thereby maintaining better protein surface properties.

It is quite clear that, of the models remaining, those containing the 4-Å water shell have the smaller RMS deviations from the crystal structure, even though one model has a constant, and the other, a distance-dependent dielectric. The NCαCO backbones of the 30-ps average structures have been overlayed with the crystal structure for these models (Fig. 3). Although the CORE RMS deviations are larger than the core deviation between 1wrp and 2wrp, the NCαCO CRYSTAL RMS deviations are smaller (Table 2). Of those models containing explicit crystallographic waters, only those with distance-dependent dielectrics fare well with respect to the CORE and CRYSTAL RMS deviation. The latter, PXCh, and PXC1r, with dielectrics that employ a sigmoidal and linear dependence on distance, respectively, exhibit very similar average RMS values, except that the CORE deviation for PXC1r is *less than*, and the CRYSTAL deviation is *greater than*, that for PXCh. This, in conjunction with the larger fluctuations in the 30-ps average structure for both CORE and CRYSTAL regions of PXC1r (Fig. 4), suggests that the PXC1r DNA reading heads sample a larger region of conformational space than do the PXCh reading heads, while retaining a better CORE structure. For comparison, the difference between the CORE and CRYSTAL RMS deviations for the models that are closer to the crystal structure indicate that *both* the PXCS1r CORE and the CRYSTAL fluctuations are consistently lower than those for PXCS1. The RMS fluctuation trend is PXC1r > PXCS1 > PXCh > PXCS1r. The RMS fluctuations of PXC1r and PXCS1 are both larger than the fluctuations for PXCh and PXCS1r, despite the fact that both the

Table 2. *Summary of trp repressor RMS deviations for the* CORE *and* CRYSTAL *residues*[a]

Model	*trp* repressor trigonal crystal 1wrp	*trp* repressor orthorhombic crystal 2wrp	Aporepressor orthorhombic 3wrp	Indole-3-propionate pseudorepressor orthorhombic crystal
trp repressor trigonal crystal 1wrp	–	2.67	3.19	2.80
trp repressor orthorhombic crystal 2wrp	0.51	–	1.80	0.81
Aporepressor orthorhombic crystal 3wrp	0.74	0.83	–	1.82
Indole-3-propionate pseudorepressor orthorhombic crystal	0.53	0.26	0.81	–

[a] The CORE RMS deviations are below the diagonal and the CRYSTAL RMS deviations are above the diagonal. All values are in Å.

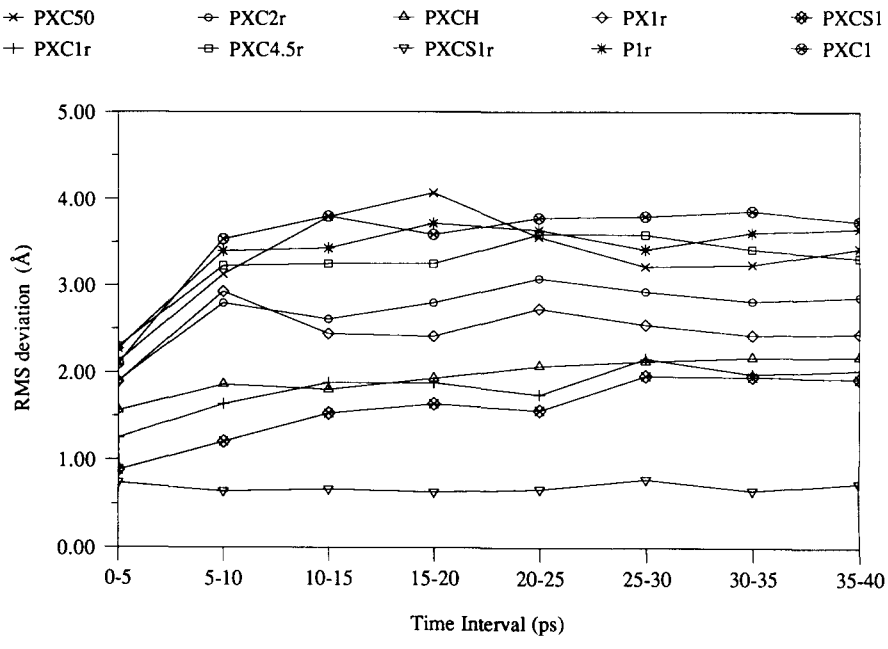

Fig. 2. The *trp* repressor CORE RMS deviations from the 1wrp crystal structure. These values were calculated for the NCαCO backbone of structures averaged over 5-ps time intervals. The CORE residues are defined in the text. The two L-tryptophan ligands are included in the RMS deviations. Refer to Table 1 for the model definitions.

PXCS1 and PXCS1r trajectories remain closer to the crystal structure. In addition, the RMS deviations of the average structures calculated from the coordinates collected during 5-ps intervals show that our choice of using a 10-ps equilibration period was, in fact, reasonable, since the RMS deviation shows very little fluctuation for all of the models after 10 ps.

The influence of those distance-dependent dielectrics plotted in Figure 1 on the *trp* repressor RMS fluctuations is (data not shown) PXC4.5 > PXC2r > PXC1r > PXCh. In this series, decreasing the short-range interactions increases the RMS fluctuations. This suggests that the stronger weighting of the short-range, rather than the long-range, interactions (see Fig. 1) is responsible for

Table 3. *Global structural and energetic properties of the average structures calculated from 30-ps trajectories*[a]

Model	% Volume compaction	% Surface area compaction	RMS deviation (Å)		ENERGY (kcal/mol)	NET ENERGY (kcal/mol)	ΔG_{STAB} (kcal/mol)
			CORE	CRYSTAL			
1wrp	–	–	–	–	14.8	−220.0	−400.9
2wrp	–	−1	0.44	3.19	18.6	−216.2	−397.1
PN1	8.58	4	2.75	4.76	83.8	−151.0	−332.0
PXC1	1.03	6	3.71	4.74	84.2	−150.6	−331.5
PXCS1	3.06	−4	1.63	2.23	44.1	−190.4	−368.1
PXC50	2.89	14	3.39	5.59	64.1	−170.7	−351.6
P1r	−8.20	8	3.48	4.91	115.5	−119.3	−300.3
PX1r	−2.24	4	2.45	3.66	92.8	−142.0	−322.9
PXC1r	−6.44	4	1.88	3.08	78.2	−156.6	−337.6
PC1r	−9.58	3	3.68	4.03	107.8	−127.0	−307.9
PXCS1r	3.76	2	0.63	1.27	26.6	−207.9	−385.6
PXC2r	−9.24	7	2.70	4.97	88.6	−146.2	−327.1
PXC4.5r	−4.29	7	3.25	4.72	89.1	−145.7	−326.7
PXCh	−1.05	6	1.99	2.52	75.7	−159.3	−340.2

[a] The tabulated properties are for the 30-ps average structures calculated from the molecular dynamics trajectory. In the Eisenberg-McLachlan method, the solvation free energy involves calculating the solvent-accessible surface area of a molecule in the folded state and in an extended conformation (see text for further details). The total reference energy and the total reference area for the *trp* repressor are 234.8 kcal/mol and 39,526 Å^2, respectively. These values reflect an identical *extended* conformational state for the 1wrp and 2wrp structures. The reference volume and surface area for 1wrp are 26,985 Å^3 and 12,759 Å^2, respectively. Percent volume and surface area compactions are relative to 1wrp, the starting model. The 1wrp reference model contains the N- and C-termini model-built in an extended conformation.

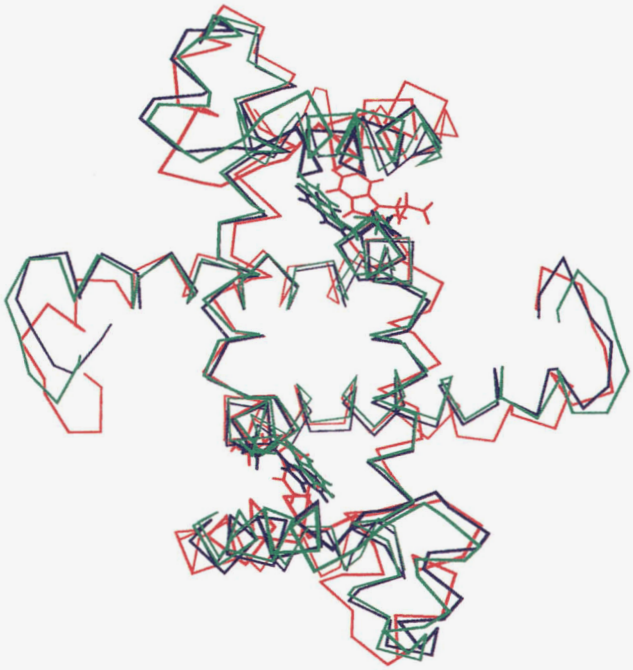

Fig. 3. The *trp* repressor 30-ps average structure Cα backbones overlaid with the 1wrp crystal structure (green) for the two best dynamics models: PXCS1 (red) and PXCS1r (blue). See Table 1 for model definitions. The L-tryptophan ligands are also displayed.

keeping the fluctuations from the trajectory average small and the trajectory average similar to the starting structure. The RMS deviations for these models reported in Table 3 and Figure 2 support this conclusion, as do the data for the constant dielectric model when only the crystallographic water is included in the simulation. Reduction of the long-range interactions led to better agreement between the model and the crystallographic structure and lower RMS fluctuations from the average structure. Adding water lowers the RMS fluctuations from the trajectory average and RMS deviations from the crystal structure when a distance-dependent dielectric is used. Comparison of models PXCS1r, PXC1r, and PC1r shows that increasing the amount of explicit water also increases protein volume compaction and lowers the solvent free energy.

A net volume compaction for the constant dielectric models and a volume expansion for the distance-dependent dielectric models, except for PXCS1r, occurs during the simulation (Table 3). Contraction of the water shell around the protein may have countered what otherwise may have been a volume expansion, leading to an overall volume contraction, since this was the only distance-dependent dielectric model that contracted. The solvation free energy is lower for the protein models that contain water, which suggest that when water is present in the simulation model, the hydrophilic atoms are more exposed, and the hydrophobic atoms are less exposed than in the absence of water. Also, the solvent free energy is lower for the models with a 4-Å shell than for those with just crystallographic waters. Comparison of these models shows that although the solvation free energy for PXCS1 is larger than for PXCS1r, the surface area for this model expands, and incidentally, is the only surface area expansion. This could rationalize the difference in solvation energies between these models. The PXC50

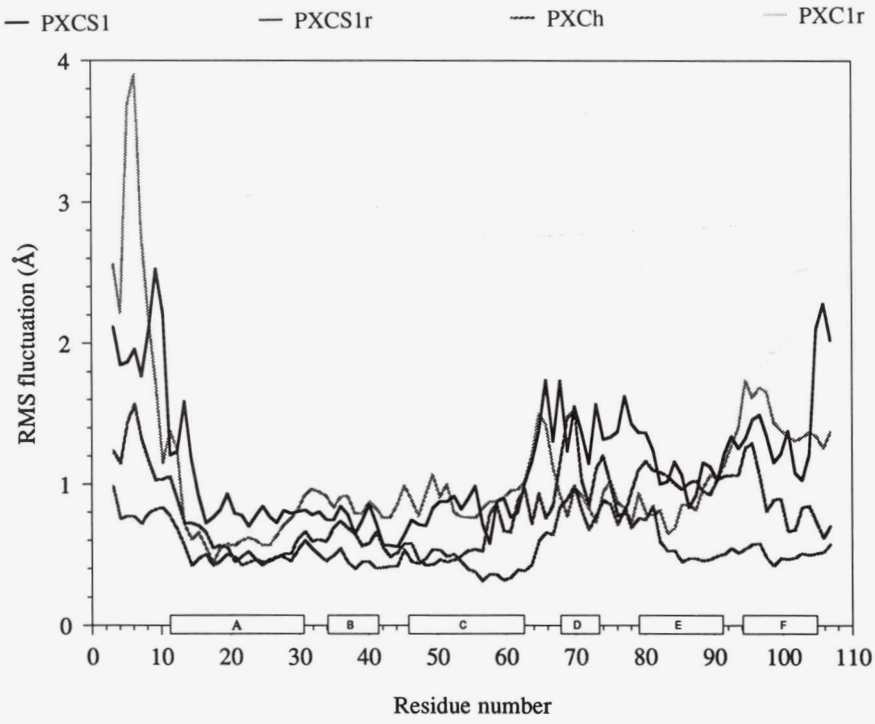

Fig. 4. The *trp* repressor RMS fluctuations from the average structures for the four best electrostatic models. The averages were calculated from the final 30 ps of 40-ps MD trajectories. The fluctuations for the Cα of each crystallographically resolved residue in each monomer were averaged and plotted as a function of residue number. The secondary structure assignments for the protein are shown with A, B, C, D, E, and F representing the six helices, according to Schevitz et al. (1985). The two helices in the helix–turn–helix DNA-binding motif are D and E. Refer to Table 1 for the model definitions.

model has a lower solvation free energy than either PXC1r or PXCh but larger than either model containing a solvent shell. This model also underwent a 14% decrease in the surface area of the average structure, reducing the exposure of hydrophilic atoms to the solvent (NET ENERGY) and increasing the the exposure of hydrophobic groups (ENERGY). Comparison with PXCh and PXC1r shows that the reduction of surface area for PXC50 led to a reduction of hydrophobic surface area, relative to the other two. Also, that PC1r underwent less surface area compaction than did PXC1r, yet has a much larger solvation free energy, underlines the importance of including some water in MD simulations. Excluding PXC50 from the consideration, the order of solvation free energies parallels the RMS results.

The MD temperatures were calculated separately for the solvent and solute and for the total solute-solvent system at the end of 5-ps intervals during the trajectory. Because the system was coupled to a temperature bath, the total solute-solvent temperature remained close to 300 K for every simulation. However, when we examined the temperatures of the individual components, we found unequal partitioning of the velocities, and hence energies, between solute and solvent. Specifically, for the constant dielectric *trp* repressor models, the water was nearly 100 K hotter than the protein throughout the trajectory (Fig. 5). Although the temperature difference of the models with crystallographic waters shows larger fluctuations than those models with a 4-Å water shell, there is a clear division between the constant and distance-dependent dielectric models. The temperature difference is most dramatic for those models with a 4-Å water shell.

The *trp* repressor simulation with more explicit water—a 6-Å water shell—and a constant dielectric (Howard & Kollman, 1991) remained closer to the crystal structure than did PXCS1 but not as close as PXCS1r. The RMS fluctuations for this model were also less than those for the PXCS1 model but greater than the fluctuations for PXCS1r. This supports the contention that more water in the simulation model leads to lower RMS fluctuations and RMS deviations although it is not clear whether this is due to an artifact from the water-vacuum interface or the stabilization of the protein via more water-protein and water-water interactions. The average MD temperature difference between solute and solvent for this simulation was 110 K, which is a little larger than that for PXCS1. A detailed quantitative comparison of these models with the *trp* repressor model in the Howard study is not entirely appropriate, because the starting model in the latter simulation differed in the number, type, and parameters used for the ions and did not include the crystallographic waters prior to adding the water shell.

BPTI

The four BPTI trajectory average structures are in good agreement with the crystal structure (Table 4). This is also apparent from overlaying the NCαCO backbones of the 45-ps average structures (Fig. 6). The solvation energies, surface areas, and volumes (Table 5) are essentially the same for all four BPTI models. The volume of all models contracted. The constant dielectric model with an 8-Å nonbonded cutoff differs most from the crystal structure and from the other three models investigated. Figure 7, however, shows substantial variation in the RMS devia-

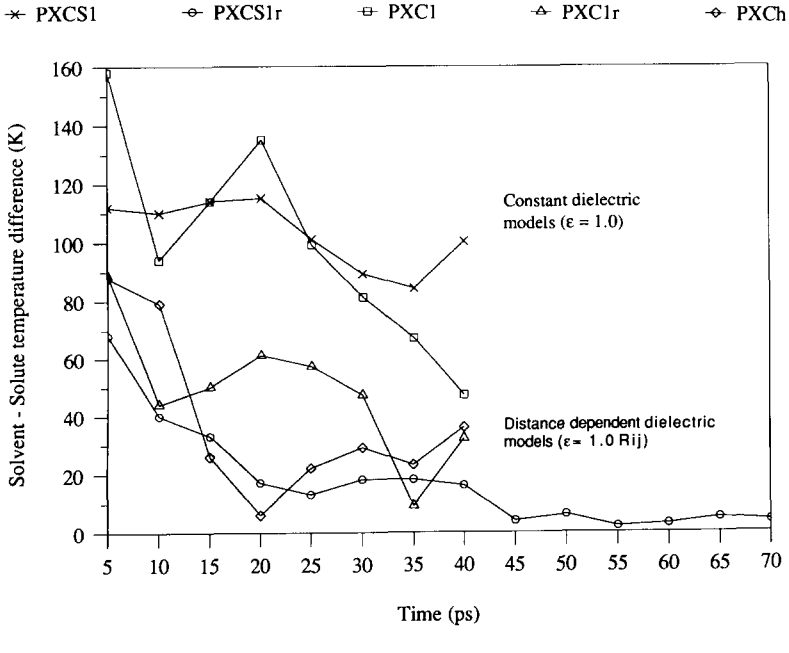

Fig. 5. The *trp* repressor MD temperature differences for distance-dependent and constant dielectric models with explicit water, both with and without a 4-Å water shell. The solute and solvent temperatures were calculated from the atomic velocities at the end of 5-ps time intervals. Refer to Table 1 for the model definitions.

Table 4. *Physical properties calculated from 45-ps average structures, for constant and linear distance-dependent dielectrics with both an 8-Å and infinite nonbonded cutoff*[a]

Model	Model description: Dielectric	Model description: Nonbonded cutoff (Å)	Radius of gyration (Å)	Cα RMS (Å)	NCαCO RMS (Å)	All heavy atom RMS (Å)
B8S1r	1.0 Rij	8	11.15	0.79	0.86	1.16
BinfS1r	1.0 Rij	Infinite	11.13	0.95	0.95	1.09
B8S1	1.0	8	11.08	1.10	1.23	1.42
BinfS1	1.0	Infinite	11.22	0.77	0.85	1.12

[a] The BPTI naming convention is analogous to that used for the *trp* repressor. The initial volume and radius of gyration for the 5pti crystal structure are 6,063 Å^3 and 11.31 Å, respectively.

tions of the 4.5-ps average structures from the crystal structure for the constant dielectric model with an 8-Å nonbonded cutoff, despite the close agreement of the trajectory average with the crystal structure.

The distance-dependent dielectric BPTI simulation with an 8-Å cutoff was continued to 121.5 ps to determine whether the RMS deviation to the crystal structure remained small after longer time periods. The all heavy atom and the NCαCO backbone RMS deviations between the 4.5-ps average structures and the crystal structure do not exceed 1.5 Å and 1.0 Å, respectively, during the course of the trajectory. Extending Figure 7 to 121.5 ps shows that the RMS deviation of the average structures from the crystal structure does not vary more than 0.2 Å during the 45–121.5-ps time interval (data not shown). The RMS deviations were also calculated for the entire simulation, as well as for both halves of the simulation. The all heavy atom RMS deviations were 1.18, 1.40, and 1.27 Å for the 4.5–63-ps, 63–121.5-ps, and 4.5–121.5-ps time intervals, respectively. The NCαCO backbone RMS deviations for the same time intervals, were 0.88, 1.02, and 0.93 Å. The RMS deviations obtained from this model were comparable to those of the detailed periodic box simulation of Levitt and Sharon (1988), despite the differences in force field and the use of a smoothing function in the latter. Structural equilibration was reached much more quickly in our study than in theirs. Levitt and Sharon noted an approximately 0.7-Å difference in RMS deviation during the initial 30 ps of their periodic box simulation of BPTI. The change in RMS deviation in our study was approximately half as large as theirs, with the largest change for the distance-dependent dielectric simulations of BPTI and *trp* repressor, and the constant dielectric simulation of BPTI with no cutoff occurring in the first 5–10 ps. However, the RMS deviation at the start of our simulation is larger than that observed in the Levitt and Sharon study, and both studies yield structures that are similar in quality after 100 ps. This is likely due to the differences in equilibration procedure.

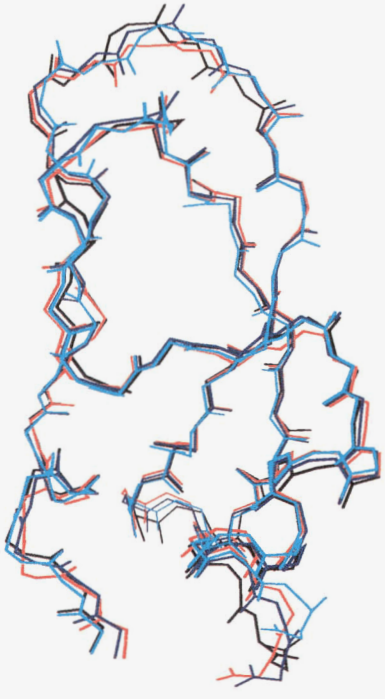

Fig. 6. The BPTI 45-ps average structure NCαCO backbones overlaid with the 5pti crystal structure (green) for the four BPTI models: B8S1 (red), B8S1r (blue), BinfS1 (black), and BinfS1r (cyan). See Table 4 for model definitions.

The hydrogen bonds in the BPTI trajectory averages were also examined (Table 6). Of the hydrogen bonds that appeared in each of the average structures, 67% of the hydrogen bonds found in the crystal structure were common to all four of the BPTI trajectories when a 120° hydrogen bond angle criterion is used. Thus, the dielectric model and nonbonded cutoff had no effect on the formation of most of the hydrogen bonds. The hydrogen bonds present in all of the trajectories and the crystal structure had consistently high occupancies, and were

Table 5. *Solvation free energies calculated from 45-ps average structures, with constant and distance-dependent dielectrics and 8-Å and infinite nonbonded cutoffs*[a]

Model	% Volume compaction	Surface area (Å^2)	ENERGY (kcal/mol)	NET ENERGY (kcal/mol)	ΔG_{STAB} (kcal/mol)
5pti	–	4,063	0.2	−55.7	−103.9
B8S1r	4.33	4,018	3.2	−52.8	−101.0
BinfS1r	4.81	3,897	1.2	−54.8	−103.0
B8S1	6.01	3,977	1.2	−54.7	−103.0
BinfS1	2.31	4,001	3.4	−52.5	−100.7

[a] The tabulated properties are for the 45-ps average structures calculated from the molecular dynamics trajectory. The total reference energy and the total reference area for the BPTI starting structures are 55.9 kcal/mol and 9,770 Å^2, respectively.

predominately found in the protein secondary structure. One exception is the constant dielectric BPTI simulation with an 8-Å cutoff, which had consistently lower hydrogen bond occupancies in the first helix. The hydrogen bond occupancies were high in the second helix, so the low occupancy can be attributed to the fact that helix 1 is a small helix, rather than this model being unable to form good helical structure. The hydrogen bonds in the distance-dependent dielectric models more closely match those found in the crystal structure, which may reflect the

Table 6. *Hydrogen bond occupancies for all hydrogen bonds that are present in the crystal structure, 45-ps BPTI trajectory average structures, and/or have a trajectory occupancy greater than 25%*[a]

Hydrogen bond donor-hydrogen-acceptor		Crystal structure	8-Å cutoff ε = 1.0	No cutoff ε = 1.0	8-Å cutoff ε = 1.0 R	No cutoff ε = 1.0 R
1 ARG (NH1)-HN11-56 GLY (O)				**49**	28	
1 ARG (NH1)-HN12-55 CYX (O)		**X**	32	**100**	**100**	**99**
5 CYX (N)-HN-2 PRO (O)	H	**X**	54	**90**	**55**	**94**
6 LEU (N)-HN-3 ASP (O)	H	**X**	**53**	**92**	**94**	**94**
7 GLU (N)-HN-4 PHE (O)	H	**X**	**75**	**96**	**93**	**83**
11 THR (N)-HN-9 PRO (O)				28		
11 THR (OG1)-HOG-34 VAL (O)		**X**			**54**	**98**
12 GLY (N)-HN-11 THR (OG1)				26		
14 CYX (N)-HN-12 GLY (O)						**84**
16 ALA (N)-HN-14 CYX (O)				**61**	37	
16 ALA (N)-HN-36 GLY (O)		**X**	**71**	**35**	**72**	
18 ILE (N)-HN-16 ALA (O)			**50**			
18 ILE (N)-HN-35 TYR (O)	S	**X**	**84**	**95**	**98**	**99**
20 ARG (N)-HN-33 PHE (O)	S	**X**	**98**	**99**	**100**	**99**
20 ARG (NH1)-HN12-44 ASN (OD1)		**X**	**87**	**99**	**77**	**100**
21 TYR (N)-HN-45 PHE (O)	S	**X**	**100**	**100**	**97**	**100**
22 PHE (N)-HN-31 GLN (O)	S	**X**	**100**	**100**	**100**	**100**
23 TYR (N)-HN-43 ASN (OD1)		**X**	**99**	**100**	**100**	**100**
24 ASN (N)-HN- 29 LEU (O)	S	**X**	**98**	**100**	**98**	**100**
24 ASN (ND2)-HND2-31 GLN (OE1)		**X**	**76**	64	**94**	**95**
26 LYS (N)-HN-24 ASN (OD1)		**X**	25		40	54
27 ALA (N)-HN-24 ASN (O)		**X**	–	–	–	–
27 ALA (N)-HN-24 ASN (OD1)		**X**	**78**	**82**	**60**	**51**
28 GLY (N)-HN-24 ASN (O)		**X**	**97**	**97**	**98**	**95**
29 LEU (N)-HN-24 ASN (O)	S		**93**	**80**	**98**	**91**
31 GLN (N)-HN-22 PHE (O)	S	**X**	**98**	**90**	**100**	**96**
33 PHE (N)-HN-20 ARG (O)	S	**X**	**99**	**100**	**100**	**100**
35 TYR (N)-HN-18 ILE (O)	S	**X**	**59**	**56**	**93**	**100**
35 TYR (N)-HN-33 PHE (O)			**43**	**51**		
36 GLY (N)-HN-11 THR (O)		**X**	**80**	**80**	**98**	
37 GLY (N)-HN-35 TYR (O)				28		
38 CYX (N)-HN-36 GLY (O)				**51**		
40 ALA (N)-HN-35 TYR (OH)		**X**	30		37	
43 ASN (ND2)-HND1-7 GLU (O)		**X**	**78**	**88**	**84**	**94**

(*continued*)

Table 6. *Continued*

Hydrogen bond donor-hydrogen-acceptor		Crystal structure	8-Å cutoff ε = 1.0	No cutoff ε = 1.0	8-Å cutoff ε = 1.0 R	No cutoff ε = 1.0 R
43 ASN (ND2)-HND2-23 TYR (O)		X	**100**	**89**	**100**	**92**
44 ASN (N)-HN-42 ARG (O)		X	**60**	**94**	**93**	**93**
45 PHE (N)-HN-21 TYR (O)		X	**100**	**100**	**100**	**100**
46 LYS (N)-HN-50 ASP (OD2)			**100**			
47 SER (N)-HN-50 ASP (OD1)						39
47 SER (OG)-HOG-46 LYS (O)					33	
50 ASP (N)-HN-47 SER (O)			39	33	**41**	35
50 ASP (N)-HN-47 SER (OG)		X	–	–	–	–
51 CYX (N)-HN-47 SER (O)	H	X	**99**	**82**	**88**	**98**
52 MET (N)-HN-48 ALA (O)	H	X	**99**	**100**	**100**	**100**
53 ARG (N)-HN-49 GLU (O)	H	X	**97**	**96**	**100**	**100**
53 ARG (NH2)-HN21-50 ASP (OD1)					41	
53 ARG (NH2)-HN21-50 ASP (OD2)						25
54 THR (N)-HN-50 ASP (O)	H	X	**99**	**100**	**94**	**91**
54 THR (N)-HN-51 CYX (O)	H	X				33
54 THR (OG1)-HOG-50 ASP (O)		X	**98**	**96**	**97**	**99**
55 CYX (N)-HN-51 CYX (O)	H	X	**99**	**100**	**100**	**100**
56 GLY (N)-HN-51 CYX (O)			**99**	**54**	**95**	**92**
56 GLY (N)-HN-52 MET (O)	H	X		**60**		
57 GLY (N)-HN-52 MET (O)				67	**67**	
58 ALA (N)-HN-56 GLY (O)				**73**		
Number of hydrogen bonds in the average structure		36	27	28	31	29
% Crystal structure hydrogen bonds formed in the average structure		100	84	76	89	91

[a] Percent occupancy is defined as the number of structures that contain a particular hydrogen bond of the 300 structures collected during the trajectory, multiplied by 100. Values in bold type indicate those hydrogen bonds also present in the trajectory average for that model. The hydrogen bonds are defined by a donor-acceptor distance less than or equal to 3.5 Å and a donor-hydrogen-acceptor angle greater than or equal to 120°. – refers to those hydrogen bonds that are present in the crystal structure but not observed during any of the dynamics trajectories. H indicates those residues involved in alpha-helix, and S refers to residues involved in beta-sheet secondary structures. The simulation model included a 4-Å shell of explicit water (see text for more details).

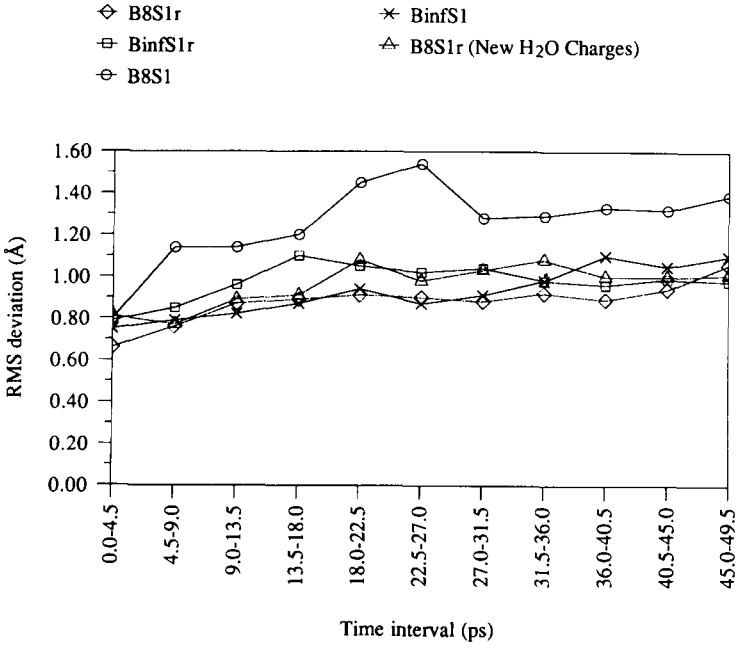

Fig. 7. The BPTI NCαCO RMS deviations from the crystal structure. The protein simulation models had a 4-Å water shell and employed either a constant or distance-dependent dielectric and an 8-Å or infinite nonbonded cutoff. Average structures from 4.5-ps time intervals were compared to the initial protein model. Refer to Table 4 for the model-naming convention. The B8S1r (new H_2O charges) model is more defined in the text in the Results.

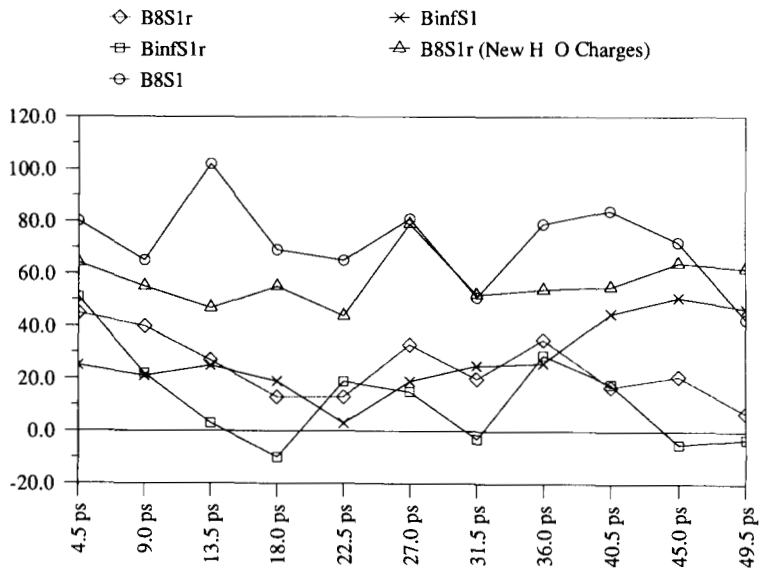

Fig. 8. The BPTI MD temperature differences for distance-dependent and constant dielectric models with explicit water, with and without a nonbonded cutoff. The solute and solvent temperatures were calculated from the atomic velocities at the end of 5-ps time intervals. Refer to Table 4 for the model definitions. The B8S1r (new H_2O charges) model is more defined in the text in the Results.

lower RMS fluctuations from the average structure for the distance-dependent dielectric models. The constant dielectric models had a tendency to form i to $i+2$ backbone hydrogen bonds, which was pronounced when a either a 90° or 100° hydrogen bond angle criterion was used (data not shown), although only five remained with lower trajectory occupancies when the more stringent 120° angle criterion was used (Table 6).

The constant and distance-dependent dielectric BPTI simulations exhibited analogous behavior to the *trp* repressor simulations with respect to solute-solvent temperature differences (Fig. 8). The temperature difference for the constant dielectric simulation with an 8-Å nonbonded cutoff was near 80 K throughout the trajectory. The temperature difference for both distance-dependent dielectric simulations and the constant dielectric simulation without a nonbonded cutoff fluctuated much closer to 0 K with an average temperature difference for each of these simulations at about 15 K.

Near the end of this study, an error was found in the AMBER code, where the use of a temperature coupling constant between the recommended values, 0.1 ps^{-1} and 0.4 ps^{-1}, coupled the system too tightly to the temperature bath. The temperature coupling constant was 20 times smaller than what it should have been. To ensure that this would have little effect on the conclusions made in this study, we compared the dynamics of both *trp* repressor and BPTI with coupling constants of 0.2 and 4.0 ps^{-1} for the PXCS1r and B8S1r simulations. The average properties examined in this study for the TAUTP = 4.0 ps^{-1} coupling constant simulations are within 10% of the TAUTP = 0.2 ps^{-1} simulations (data not shown). The RMS fluctuations from the average *trp* repressor structure were essentially the same with both temperature coupling constants, although the TAUTP = 0.2 ps^{-1} trajectory consistently showed a 0.05–0.10-Å larger RMS fluctuation from the BPTI average structure than did the TAUTP = 4.0 ps^{-1} simulation (data not shown), although both sets of fluctuations show the same qualitative behavior. This trend was also reflected in the trajectory average properties. The RMS fluctuations for the *trp* repressor model were essentially the same for both temperature coupling constants, with the fluctuations from the TAUTP = 4.0 ps^{-1} simulation slightly larger in some regions of the sequence, than for the TAUTP = 0.2 ps^{-1} simulation. This trend is also borne out in the RMS deviations, which are slightly larger for the TAUTP = 4.0 ps^{-1} simulation. Because the two simulations showed opposite trends in the average properties, the effects of this error are likely protein dependent. Although we tested the linear distance-dependent *trp* repressor and BPTI models with a 4-Å shell only, we have assumed that the temperature coupling parameter will affect each simulation in the same manner, and that the trends found in this study will be similar regardless of whether TAUTP is set to 0.2 or 4.0 ps^{-1}.

Water

The TIP3P water-water interaction energy is substantially larger when a distance-dependent dielectric is used instead of a constant dielectric. The final potential energies for the water dimers, minimized with $\varepsilon = 1.0$, 1.0 Rij, and SIGM, were −6.49, −9.43, and −9.65 kcal/mol, respectively. The experimental association energy of a water dimer, relative to the water monomers, is -5.44 ± 0.7 kcal/mol (Curtiss et al., 1979). The TIP3P water model with a constant dielectric overestimates this value, whereas both distance-dependent dielectric models greatly overestimate it. The nonbonded energy was larger and the electrostatic energy was smaller for the linear and sigmoidal distance-dependent than for the constant dielectric

models. Because the intermolecular oxygen distance of the dimer was shorter when minimized with the distance-dependent dielectrics, 2.59 Å compared to 2.75 Å, than with $\varepsilon = 1.0$, we suspected the water would undergo a global contraction relative to the $\varepsilon = 1.0$ model. The effect of the linear distance-dependent dielectric on pure water with periodic boundary conditions is to increase the strength of the water-water interactions, which leads to an 8% volume contraction, relative to the constant dielectric simulation. The density for the constant and the distance-dependent dielectric models was 1.0 g/cm^3 and 1.1 g/cm^3, respectively. Although the constant dielectric more closely reproduces the water dimer energies and properties of bulk solvent for a periodic boundary simulation, errors introduced by the inadequate description of the vacuum-solvent interface in the protein models may alter this comparison between constant and distance-dependent dielectrics for water.

BPTI with new water charges

As noted above, the TIP3P water model with a constant dielectric overestimates the experimental association energy of a water dimer by approximately 1 kcal/mol, whereas the distance-dependent dielectrics overestimate the stabilization interaction by about 4 kcal/mol. The stronger water-water interactions may be improperly balanced against the protein-protein and protein-solvent interactions, such that the stronger water-water interactions from using a distance-dependent dielectric model would reduce the protein fluctuations by creating an ice-like water structure around the protein. By changing the water charges such that the minimized energy of the water dimer with a distance-dependent dielectric is close to that for the constant dielectric model and rerunning the simulation, we could determine whether the results obtained for the protein simulations were due to the stronger than normal water-water interactions.

The water hydrogen and oxygen charges were changed from the TIP3P value (+0.417 for H and −0.834 for O) to three-point model values that would reproduce the potential energy of a constant dielectric water dimer with TIP3P charges. This corresponds to a +0.365 charge for hydrogen and −0.730 charge for oxygen (data not shown). The BPTI model with the water shell and a distance-dependent dielectric was rerun using exactly the same protocol as used for B8S1r, but with the altered charges for water. The conformational results of this control simulation were very similar to those obtained for the B8S1r simulation and the TIP3P water charges (Fig. 7), rather than the B8S1 model, indicating that the protein conformational properties arise from the dielectric effect on the protein-protein interactions, rather than the potentially stronger water-water interactions.

However, the solute-solvent temperature difference for the BPTI model with the altered water charges was significantly larger than the solute-solvent temperature difference for the equivalent model with TIP3P water charges (Fig. 8). This has several implications. First, the solute-solvent temperature difference for the constant dielectric model may not simply be an effect of the discontinuities in the forces at the nonbonded cutoff but may also be related to the charge on water and hence the balance of interactions between the solvent with other solvent molecules and with the protein. Second, no evaporation occurred in the simulation with the altered charges, and so water most likely remains in the simulation model because of a dielectric effect, rather than the increased strength of the water-water interactions. And third, the higher energy of the water, as expressed by the water dimer results, which are closer to the experimental values, may be responsible, in part, for the solute-solvent temperature difference observed for the constant dielectric model and the BPTI model with the altered charges. It is not clear that the water in a protein simulation with a simple explicit solvent shell and a vacuum-solvent interface should behave in the same manner as bulk water. These data suggest that the protein conformational disparity between the dielectric models arises from the effect of charge screening on the protein rather than from artifacts arising from the differences between the description of water-water and water-protein interactions.

Discussion

Although long simulation times are required in order to insure adequate conformational sampling, for the purposes of this study, it would have been an inefficient use of computational resources to simulate the poorer models for longer time periods. We have simulated the *trp* repressor models for 40 ps and the BPTI models for 50 ps. Although this may not have been long enough to distinguish which was best among the physically reasonable models, it was long enough to reject the bad *trp* repressor models and to determine whether the nonbonded cutoff had affected the dynamics of the constant and distance-dependent dielectric trajectories. Large-scale conformational changes were not expected on this small time scale.

Although the RMS deviation from the crystal structure was the single best evaluation criterion, the RMS fluctuations from the average structure were able to distinguish among models with similar RMS deviations. With constant and distance-dependent dielectrics and explicit water, the trajectories remain close to the crystal structure, although the constant dielectric exhibits larger fluctuations from the average structure. This implies larger conformational sampling about the crystal structure. We have assumed that the differences between crystal structures (Table 2) provide a reasonable guide to the RMS deviations that can be expected during a dynamics simulation. This may or may not be valid, depending upon the role of crystal packing forces in insuring that the protein core remain relatively invariant.

Including explicit water in the simulation model maintains the proper orientation of the protein hydrophilic and hydrophobic atoms with respect to the aqueous environment. In MD, this occurs through *both* van der Waals and electrostatic interactions. Although it is unclear which type of nonbonded interactions plays a larger role, the electrostatic effects are exerted over longer distances and comprise about 70% of the total potential energy with the Weiner et al. all-atom force field and a constant or distance-dependent dielectric. The solvation free energy data indicate that models which lack explicit water do not effectively discriminate between polar and nonpolar atoms on the protein surface. The solvation free energy is a reliable indicator of a good model provided that the RMS deviation from the crystal structure is low and surface area compaction is not extensive. A change in surface area from the crystal structure will result in a different proportion of hydrophobic and hydrophilic groups exposed to the solvent environment. When the change in surface area is taken into account, the solvent free energies are low, as determined by Chiche et al. (1990), and follow the trends of the backbone and all heavy atom RMS deviations from the crystal structure. Maintaining solvent exposure of polar atoms may, in part, keep the protein close to the crystal structure. Alternately, keeping the protein close to the crystal structure may maintain proper solvent exposure of the protein surface groups.

There are two problems associated with calculations that do not include periodic boundary conditions. The first is the presence of a system–vacuum interface, which leads to artifacts at the system surface, because an atom at the interface establishes interactions with the protein system as opposed to extending into the vacuum where it has fewer interactions. Including some explicit water in the model, to some extent, corrects for this. The protein surrounded by a water shell is simulated in a droplet, which approximates the aqueous simulation. However, the second problem is particular to all simulations without periodic boundary conditions. In a gas phase protein simulation, the number of nonbonded interactions for each atom at the interior of the protein is *always* greater than the number of nonbonded interactions at the protein surface. Neglect of an explicit solvent environment reduces the number of nonbonded interactions because these are pairwise additive, and this number depends upon the number of atoms surrounding it. Without explicit water, the energetic contribution of the interior atoms is greater than that of the exterior atoms. In addition, proteins are more densely packed in the interior than at the exterior. If water is not included in the simulation, the protein–solvent interactions can be further underestimated. This is because the forces are asymmetrical for the surface residues but not for the residues at the interior. When explicit water is included in the calculation, the number of nonbonded interactions at the protein surface is increased, and depending upon the amount of explicit water included in the simulation, and the nonbonded cutoff used, the number of protein interactions at the surface can be increased until it is equal to the number in the protein interior. At this point, the energetic contribution of the surface atoms would be approximately that of the interior atoms, depending upon the surface composition relative to the interior. If an 8-Å nonbonded cutoff is used, then an 8-Å water shell would be required in order for all of the protein atoms to have essentially the same number of interactions contributing to the total potential energy. The water, however, would not have the same number of interactions as the protein, and its energetic contribution would decrease with increasing distance to the protein. Simulating the protein with periodic boundary conditions eliminates this problem because the atoms near the boundary would interact with an image atom, and hence, atoms at the protein surface and in the aqueous environment, and atoms in the protein interior would have essentially equivalent nonbonded contributions to the energy of the system. Simply balancing the forces on protein surface residues may very well lead to better protein trajectories.

Periodic boundary conditions were not used in the BPTI or *trp* repressor simulations. Including more explicit water in the *trp* repressor model kept the trajectory close to the crystal structure. This is consistent with more explicit water increasing the energetic contributions of the protein surface atoms relative to that of the simulations with less water. The RMS deviations of the distance-dependent dielectric *trp* repressor models were smaller than those for the constant dielectric models, for models with the same amount of water, because the distance-dependent dielectric weights the short-range interactions more heavily than the long-range interactions, both in the interior and at the surface of the protein. With a 4-Å water shell, an 8-Å nonbonded cutoff, and a distance-dependent dielectric, the neglect of a number of protein–water interactions would lead to a smaller error than with a constant dielectric, which has a larger energy contribution for these interactions. For a distance-dependent dielectric simulation, the electrostatic contribution due to protein–water interactions for a protein atom at the protein–solvent interface is small beyond 4 Å and still includes a large number of protein–protein interactions to 8 Å. The electrostatic contributions from shorter distances may play an important role in maintaining the structure, and truncating the long-range interactions may lead to smaller errors. This is substantiated by the distance-dependent dielectric BPTI trajectories, for which the structural properties examined in this study are nearly identical, despite the fact that one had a nonbonded cutoff and the other did not. Because the forces between atoms are weaker, errors accumulated in the calculated forces, due to the nonbonded cutoff, may make the distance-dependent dielectric model less sensitive to the cutoff than the constant dielectric model.

The solvent–solute temperature difference also arises

from truncation of forces at the nonbonded cutoff. The forces are smaller at the nonbonded cutoff for the distance-dependent dielectric, and consequently, the errors in the forces are also smaller. The temperature, therefore, remains more equally distributed between components, in contrast to the constant dielectric simulations where we found unequal partitioning of the kinetic energy. For the constant dielectric model, switching or smoothing functions (Levitt & Sharon, 1988) can be used to gradually truncate the forces at the cutoff in order to circumvent these errors. The difference between using a distance-dependent dielectric and a constant dielectric with a switching function is that the former progressively reduces the strength of the electrostatic interactions from the value for which the charges were originally derived. Although the optimal solution would be to not use a nonbonded cutoff at all, this is not a computationally efficient alternative, at present.

The lack of periodic boundary conditions likely contributed to solvent evaporation, which occurred for all *trp* repressor models except PXCS1r. This includes PXC1r—the same model with less explicit water. This suggests that the water–water interactions with a distance-dependent dielectric are stronger when more water is included in the simulation model, most likely due to the increased number of solvent–solvent interactions. For the distance-dependent dielectric simulations with crystallographic waters only, the solvent may have evaporated because the protein surface interaction with the water was not strong enough to keep the water associated with the system. Water evaporation occurred with all models in this study except for the two BPTI simulations with distance-dependent dielectrics and the *trp* repressor simulation with a 4-Å shell of water and a distance-dependent dielectric.

Other studies that have examined the effect of dielectric on in vacuo dynamics simulations have arrived at results that conflict with the simulations containing crystallographic water reported here. Wendoloski and Matthew (1989) compared four in vacuo AMBER-generated simulations for tuna cytochrome *c* with $\varepsilon = 1.0$ Rij, 4.0 Rij, 50, and 80 and one simulation, which included 800 explicit waters and had an $\varepsilon = 1.0$ dielectric, and found that a dielectric of 50 gave the best agreement with the crystal structure and the explicit water simulation. In the present study, this was the worst model we found for modeling *trp* repressor dynamics. Water evaporation was most dramatic for this model, which lost 90% of the explicit crystallographic water and 13 of the 14 initially placed ions, so it is likely that the crystallographic waters and ions had little effect on the difference in protein behavior with the $\varepsilon = 50$ model. The Weiner et al. force field was parameterized for use with $\varepsilon = 1.0$ and 1.0 Rij dielectrics and yields much better dynamics trajectories with these models. Reducing the electrostatic energy by a factor of 50 will greatly alter the contribution of the electrostatic energy to the potential energy. As reported, the study by Wendoloski and Matthew (1989) lacks sufficient information concerning their equilibration procedure, force field, and nonbonded cutoff, to determine why their results were different from ours. A second study (Ornstein, 1990) found that for gas phase simulations of crambin, $\varepsilon = 1.0$, 1.0 Rij, 2.0 and 2.0 Rij provided good agreement between the crystal structure and structures generated from minimization of dynamics structures along the trajectory. Consistent with our results, a 4.0 multiplicative constant did not provide good results compared with the other models examined. The best model was $\varepsilon = 2.0$, a model that we did not test. However, we would not have expected $\varepsilon = 1.0$ to have yielded a good model. The Ornstein study used a different force field with an infinite nonbonded cutoff on a protein $\frac{1}{4}$ the size of *trp* repressor. In addition, neutralization of the charges on all acidic and basic residues is likely to have made a large contribution to the difference between the results of this study (Ornstein, 1990) and ours. The equilibration procedure and structural analysis was also performed differently. A nonbonded cutoff is commonly used to reduce the computational burden of larger proteins. The constant dielectric models (Ornstein, 1990) may not provide better results than the distance-dependent dielectric models when a nonbonded cutoff is used. This result is likely to hold true for other force fields, although it should be tested.

The effect of the long-range interactions is to increase the RMS fluctuations in the series: $\varepsilon = 1.0 > \varepsilon = 1.0$ Rij $> \varepsilon =$ SIGM. By decreasing the long-range interactions, the sigmoidal dielectric function yields a trajectory that remains close to the crystal structure, with low RMS fluctuations from the average structure. Decreasing the effect of the long-range interactions, at its limit, may be the same as using a shorter nonbonded cutoff. Although this may keep the protein close to the crystal structure, conformational sampling about that structure may be too restrictive. The fluctuations are necessary in order to derive experimentally useful and testable information from the simulation. If in spite of large fluctuations, the system retains the average properties close to the crystal structure, then it is an appropriate simulation model for extensive conformational sampling about the crystal structure.

The use of the linear distance-dependent dielectric for MD simulations is likely dependent upon amount of water surrounding the protein. There was a larger difference between the quality of the model when only crystallographic waters were included than when crystallographic waters and a 4-Å shell were used. However, use of a linear distance-dependent dielectric with a small water shell may lead to good results due to compensating errors. The distance-dependent dielectric models all expanded except for PXCS1r, whereas the constant dielectric models compacted. For the distance-dependent dielectrics, reducing the contribution of the long-range interactions relative to the constant dielectric model may not hold the protein together as tightly, allowing it to expand during the simulation. The two pure water simulations showed a global

contraction of TIP3P water, resulting from stronger water–water interactions when a distance-dependent dielectric is applied. Including water in the protein simulations may, therefore, reduce the tendency for protein expansion with a distance-dependent dielectric, through strong water–protein and water–water interactions. Hence the primary effect of using a distance-dependent dielectric with a small water shell may be to strike a balance among the water–water, water–solute, and solute–solute interactions. Alternatively, not calculating all of the surface interactions may provide an improper balance between the water–water, water–protein, and protein–protein interactions. It is necessarily true that without the periodicity of the solvent and with a solvent shell smaller than the nonbonded cutoff, the potential accounts for more protein–protein than protein–water interactions, and more protein–water interactions than water–water interactions. The fewer interactions at the protein surface may be compensated by stronger water–water interactions using the distance-dependent dielectric, in addition to the decrease in strength of interaction with distance.

Conclusions

Our primary objective was to evaluate MD simulation models with computationally efficient solvent models in order to find a model that retains the known physical properties of the protein. For *trp* repressor, this means a low CORE RMS deviation to the crystal structure and sufficient fluctuation in the helix–turn–helix units, in qualitative agreement with the crystallographic *B*-factors. A constant and distance-dependent dielectric, each with a 4-Å water shell, yielded structures closest to the crystal structure and were *computationally efficient*, requiring only two to four times more time than the model without the water shell. However, the linear and sigmoidal distance-dependent dielectric models with crystallographic waters only provided results in good agreement with the experimentally known features of the protein and required less computer time to complete than the models with a 4-Å shell of water.

Reasonable agreement with the crystal structure can also be obtained by including only the crystallographic waters. The best linear distance-dependent simulation had a multiplicative constant = 1.0 and therefore assumed no scaling of the electrostatic term. When crystallographic waters alone are included in the simulation model, both the linear distance-dependent and the sigmoidal distance-dependent dielectric models remained equally close to the crystal structure, although the RMS fluctuations from the average for the sigmoidal distance-dependent dielectric model were much smaller than those for the linear distance-dependent dielectric. These models fared far better in every respect than the constant dielectric model, which indicates that either a linear or sigmoidal distance-dependent dielectric should be used for simulations containing crystallographic waters and not a constant dielectric. It is interesting to note that the linear distance-dependent dielectric fares significantly better than a constant dielectric when only crystallographic waters are included in the simulation, but only marginally better when a 4-Å water shell is included.

A second objective of this study was to determine whether the *trp* repressor results could be extended to other proteins and to explore the effect of the nonbonded cutoff on the constant and distance-dependent dielectric trajectories. The dielectric dependence on the conformational results obtained for BPTI with a nonbonded cutoff is similar to that observed with *trp* repressor. However, this disparity arises from the nonbonded cutoff dependence of the constant dielectric model. The distance-dependent dielectric model shows essentially no conformational dependence on the nonbonded cutoff and is in closer agreement with the results obtained from a constant dielectric simulation with no nonbonded cutoff than is the constant dielectric model with a cutoff. The 8-Å nonbonded cutoff may have had a similar effect on the *trp* repressor simulations, and this should be taken into account in evaluating these results.

Our results were consistent with those of Levitt and Sharon (1988): water is critical for increasing the agreement between the trajectory average structures and the crystal structure and for lowering the overall fluctuations of the protein. However, periodic boundary conditions do not appear to be necessary in order to obtain good agreement with the crystal structure, although the physical properties of the hydration water may not be accurately represented without periodic boundary conditions. We suggest simulating protein dynamics with as much water as is computationally feasible with either a constant ($\varepsilon = 1.0$) or linear distance-dependent dielectric ($\varepsilon = 1.0$ Rij), rather than altering the dielectric to compensate for lack of solvent.

Materials and methods

Protein models

The trp repressor

The *trp* repressor coordinates (entry 1wrp, version July 1987 [Schevitz et al., 1985]) were obtained from the Protein Data Bank (PDB) at Brookhaven National Laboratory (Bernstein et al., 1977; Abola et al., 1987). The 3 C-terminal residues, that were not crystallographically resolved were model-built in an extended conformation. N-terminal residues 2–4 were not resolved in the electron-density map either, and so we modeled these in an extended conformation. The terminal methionine, which is present on the protein gene product but is cleaved after translation in vivo (Gunsalus & Yanofsky, 1980), was also modeled onto the N-terminus. This was an oversight in

the initial model-building. Because the amino-terminus is extended away from the body of the protein in the crystal structure, and because residues 2–4 were also model-built, we doubt this oversight will have a substantial impact on our results. Using the rotation-translation matrix in the PDB, the second monomer was generated from the first, after the missing amino acids and atoms had been added. The L-tryptophan ligands were modeled as zwitterions. In the counterion simulations, 14 ions were added in the regions of highest positive and negative charge. The electrostatic potential for the protein was calculated using a pairwise additive Coulomb potential with AMBER charges and $\varepsilon = 1.0$. Counterions were placed where the potential was less than -0.5 e or greater than $+0.5$ e. The resulting ion placement led to 10 cations and 4 anions at the protein surface. At pH 7.0, the protein has an overall -6 charge, and so the protein models that include counterions have a zero net charge. Because 59 crystallographic waters were associated with each protein monomer in the PDB, a total of 118 explicit crystallographic waters were included in those models with crystallographic waters. The 4-Å water shell, which contained 695 waters, was added around the protein after the ions, and crystallographic waters were placed.

BPTI

The coordinates for BPTI were obtained from the PDB (entry 5pti, version July 1987 [Wlodawer et al., 1987]). We included the 63 crystallographic D_2O found in the coordinate file and replaced all deuterium with hydrogen. The protein and crystallographic waters were then solvated with a 4-Å water shell. Although the protein has an overall +6 charge at pH 7.0, no ions were added to the model.

Methods

In order to determine how weighting the long-range electrostatic interactions against the short-range interactions affected the different *trp* repressor trajectories, we constructed the distance-dependent MD models with the following dielectric constants: 1.0 Rij, 2.0 Rij, 4.5 Rij, and an exponential dielectric, modified by Ramstein and Lavery (1988), from Hingerty (1985):

$$D - \frac{D}{2}[(RS)^2 + 2RS + 2]e^{-(RS)}, \qquad (1)$$

where the solvent dielectric, D, is 78, R is the distance between charges, and S is the sigmoidal parameter. For the purposes of this paper, this sigmoidal distance-dependent dielectric function will be referred to as SIGM, and the model that employs it as PXCh. Details concerning the use of $S = 0.30$ have been provided previously (Daggett et al., 1991). The rationales for using $\varepsilon = 4.5$ Rij and $\varepsilon = 50.0$ have been provided by the Pickersgill (1988) and Wendoloski and Matthew (1989) studies, respectively. The 2.0 Rij dielectric was chosen to fall between the 1.0 Rij and 4.5 Rij dielectrics, where it would weight the long-range interactions to the same degree as SIGM, but would put much less emphasis on the short-range interactions (Fig. 1). The presence of crystallographic waters provides the dielectric descriptions with a best-case scenario, since the waters are expected to aid in the orientation of polar groups at the protein surface. For comparison, we also evaluated a model in which the charged side chains were neutralized (Aqvist et al., 1985). The *trp* repressor models used in this study have been tabulated in Table 1.

The crystallographic waters, ions, and AMBER-added waters were minimized to a 0.5-kcal/mol·Å RMS gradient with the protein held rigid. This was followed by 10 ps of MD on the ions and waters, again with the protein held rigid. Improperly placed waters and ions are likely to destabilize the protein and contribute to errors at the start of the simulation. This equilibration protocol (A.E. Howard, unpubl.) allows for reorientation of the water electric vectors to avoid spurious protein motions during minimization. The structure thus obtained served as the starting model for minimization of the protein, ions, and waters. To keep the initial MD structure close to the crystal structure, we minimized each model with positional constraints on the NCαCO backbone. A 100-kcal/mol·Å^2 force constant was used to minimize the system to a 0.5-kcal/mol·Å RMS gradient. The constraints were gradually removed, by subsequent minimizations to a 0.1-kcal/mol·Å RMS gradient using force constants of 50, 15, 2, and 0 kcal/mol·Å^2. This minimization protocol was determined to be superior to minimization without backbone constraints for keeping the MD starting structure close to the crystal structure (A.E. Howard, unpubl.). Forty-picosecond trajectories were calculated for each *trp* repressor model using a 1.5-fs time step with all bonds constrained using the SHAKE algorithm (van Gunsteren & Berendsen, 1977). An 8-Å residue-based cutoff was used for all nonbonded interactions. The nonbonded list was updated every 10 steps. The system was coupled to a constant temperature bath (Berendsen et al., 1984) with the temperature coupling constant set to 0.2 ps^{-1}. The coordinates were collected every 0.3 ps during the trajectory, and an average structure was calculated after energy equilibration had been established. Our choice of a 10-ps equilibration period was sufficient, in part because most of the models, certainly the better models, had reached energy equilibration by the end of 5 ps. The RMS deviations had also stabilized by this point (Fig. 2). Four 50-ps MD trajectories were calculated for BPTI. These models employed either a linear distance-dependent or constant dielectric with an 8-Å or infinite cutoff for the nonbonded interactions. The coordinates were collected every 0.15 ps. We used a 4.5-ps equilibration period for the BPTI simulations. The trajectories were otherwise calcu-

lated in exactly the same manner as for the *trp* repressor simulations.

To examine the effect of dielectric on the solvent–solvent interactions, two MD trajectories were calculated for pure water, each consisting of 216 TIP3P waters (Jorgensen, 1981) in a 6,600-Å^3 periodic box. The periodic boundary conditions were applied in order to eliminate artifacts from the solvent–vacuum interface. An 8-Å residue-based nonbonded cutoff was used for both simulations. Both $\varepsilon = 1.0$ and $\varepsilon = 1.0$ Rij dielectric constants were used. Ten picoseconds of constant volume dynamics was run to equilibrate the model prior to applying constant pressure. Equilibrating with constant pressure dynamics requires more time due to the structural changes generated by rescaling the volume at each step. The constant volume dynamics was followed by 20 ps of constant pressure dynamics. Average water structures were calculated from the final 10 ps of MD. The density of these average structures and the final potential energies are reported for both simulations. To localize the energetic effects of the dielectric to specific terms in the potential function, we minimized a TIP3P water dimer to a 0.001-kcal/mol·Å RMS gradient, using $\varepsilon = 1.0$, 1.0 Rij and SIGM, compared this to the experimental value, and examined the interatomic distances for all three dielectric models.

The all-atom force field (Weiner et al., 1986) was used as implemented in the AMBER 3.0 Revision A molecular mechanics and dynamics software (Singh et al., 1989) where the potential function is defined as:

$$V_{total} = \sum K_r(r - r_o)^2 + \sum K_q(q - q_o)^2 + \sum \frac{V_n}{2}[1 + \cos(\eta\phi - \gamma)] + \sum \varepsilon^*_{ij}\left(\frac{R^*_{ij}}{R^{12}_{ij}} - \frac{R^*_{ij}}{R^6_{ij}}\right) + \sum \frac{q_i q_j}{\varepsilon R_{ij}} + \sum \left(\frac{C_{ij}}{R^{12}_{ij}} - \frac{D_{ij}}{R^{10}_{ij}}\right), \quad (2)$$

where $\varepsilon^*_{ij} = \sqrt{\varepsilon^*_i \varepsilon^*_j}$ and $R^*_{ij} = R^*_i + R^*_j$. The harmonic force constants for bond stretching and angle bending are K_r and K_q, respectively; r_o and q_o are the equilibrium bond and angle values. The torsional term is defined by the V_n, η, and γ parameters, which correspond to the force constant, periodicity, and phase of the torsional potential. For the nonbonded terms, the distance between atoms i and j is R_{ij}, and C_{ij} and D_{ij} are the hydrogen-bonding parameters for a particular interaction between atom types. For the Lennard-Jones potential, ε^*_i is the well depth and R^*_i is an effective van der Waals radius for atom i. The electrostatic interaction is described by a pairwise additive Coulomb potential, where q_i and q_j refer to the charges on atom i and atom j, and ε is the dielectric. The charges were previously derived using a least-squares fit to the electrostatic potential of individual amino acids (Singh & Kollman, 1984). The 1–4 nonbonded terms, which include the Lennard-Jones potential and the Coulomb potential, were scaled by a factor of two. We used sulfur parameters modified by Seibel (1990). The ions were parameterized with $\varepsilon^* = 0.1$ kcal/mol and $R^* = 4.0$ Å, to simulate hydrated ions (Singh et al., 1985). Computational times and resources for the *trp* repressor models are listed in Table 1.

Analysis

The coordinate averages were calculated for each *trp* repressor trajectory from the final 30 ps of dynamics and from the average structures from 5-ps intervals. The latter was used to follow the time course of the trajectory. The following structural properties were calculated for the 30-ps average structures: RMS deviation to the crystal structure, RMS fluctuations from the trajectory average structure, radius of gyration, % volume, and % surface area compaction relative to the crystallographic structure, solvent-accessible surface area, stabilization free energy, and MD temperatures. All structures and dynamics trajectories were also visualized using the updated version of the MIDAS graphical display software (Huang et al., 1985).

RMS deviation between two structures was evaluated as:

$$\text{RMS deviation} = \sqrt{\frac{\sum (x_1 - x_2)^2}{n}}, \quad (3)$$

where x_1 is the atomic coordinate set for molecule 1, x_2 is the atomic coordinate set for molecule 2, and n is the number of atoms in each structure. The AMBER analysis (ANAL) module aligns the two structures using a least-squares minimization algorithm prior to calculating the RMS deviation between the two. This yields the smallest possible RMS difference between the structures. Because regional flexibility is likely to bias the molecular orientation, RMS deviations calculated from a least-squares structural alignment may not be the best way to assess changes between two structures. Hence, to locate the regions of largest change, we calculated the following RMS deviations: CRYSTAL, which includes all but the model-built 3 C-terminus and 4 N-terminus residues; CORE, referring to the protein core, which is defined as residues 12–67 and 92–105 of each monomer (Schevitz et al., 1985); and ENTIRE, which includes all residues. The two L-tryptophan ligands are included in the RMS deviations. All heavy atom, Cα, and NCαCO backbone RMS differences were calculated for each model, as well as for the individual monomers, although only the CORE and CRYSTAL NCαCO backbone RMS deviations have been reported. The coordinate average for the BPTI sim-

ulations was calculated from the final 45 ps, because the energy and RMS deviation had plateaued for all four trajectories by 5 ps. Averages were also calculated for 4.5-ps intervals. All residues were included in the BPTI RMS calculations.

We used RMS deviation from the crystal structure as our primary elimination criterion. In particular, low values for the CORE RMS deviation were critical for establishing whether a trajectory provided an accurate description of the protein dynamics. Comparison of the protein in the available crystal structures (Schevitz et al., 1985; Lawson & Sigler, 1988; Lawson et al., 1988) shows that the core remains relatively invariant, despite whether the protein contains bound tryptophan or indole-3-propionate (Table 2). The ligands were not included in the crystal structure comparisons. The differences among the *trp* repressor crystal structures have been localized primarily to the helix–turn–helix DNA reading heads and the N- and C-termini (Lawson et al., 1988). Of the structures compared in Table 2, two contain bound tryptophan, one does not, and the fourth has a tryptophan analogue, indole-3-propionate, bound with a different binding mode. Yet the protein core NCαCO RMS deviation between any two crystal structures does not exceed 1 Å. We therefore expect models that provide an accurate description of the dynamics to have deviations not much greater than 1 Å. Further support may be gained from the relative magnitude of the *B*-factors, which are low for the 1wrp and 2wrp core residue Cα. The protein core remains invariant despite the crystal form. This is indicative of the inherent flexibility of the DNA reading heads. If these structures do define distinct low energy conformations of the protein (Lawson et al., 1988), we do not expect them to interconvert on the 40-ps MD time scale, since the energy barriers are likely to be greater than kT, the interconversion itself may be an improbable event, and because the crystal packing forces may influence the conformation of the reading heads. To substantiate this, the 30-ps trajectory average structures were also compared to the *trp* repressor structure from the orthorhombic crystal (PDB entry 2wrp, 1988 version [Lawson et al., 1988]). In all cases, the RMS deviations of the simulation averages were smaller when compared with 1wrp, the starting structure (data not shown). We expect the differences between the model CORE and the crystal structure core RMS deviations to reflect how closely the trajectories model the crystal structure. The larger fluctuations in the DNA reading heads and N- and C-termini should show a high degree of conformational sampling to be consistent with the regions that show the most experimental disorder.

The RMS fluctuations from the trajectory average for each residue Cα were examined in order to assess backbone conformational sampling and its agreement with the crystallographic *B*-factors. The RMS fluctuation from the 30-ps coordinate average was evaluated as:

$$\text{RMS fluctuation} = \sqrt{\frac{\sum (x_1 - \boldsymbol{x})^2}{n}}, \qquad (4)$$

where x_1 is the Cα coordinate position and $\boldsymbol{x}$ is the coordinate average of that position from the coordinate sets generated during the simulation. Because the protein is a dimer, composed of identical subunits, the RMS fluctuations for both monomers were averaged, so that the fluctuations per residue number include the fluctuations of both monomers.

To measure the relative extent of exposure of hydrophilic and hydrophobic atoms at the protein surface, the Eisenberg and McLachlan (1986) solvent free energies were calculated for the *trp* repressor 30-ps average structures and the BPTI 45-ps average structures. This relies on a semiempirical formula, which relates the solvent-exposed surface area in the extended protein chain to that of the folded structure. The solvent free energy, ΔG_{STAB}, has two components and is relative to the extended chain reference state: the ENERGY reflects the energy cost of exposing hydrophobic groups to the solvent and the NET ENERGY is the favorable contribution to the solvent free energy that is obtained from exposing hydrophilic atoms to the solvent. Chiche and co-workers (1989) showed that solvation free energy and polar versus apolar surface ratios could be used to distinguish between correct and incorrect protein structural models. In addition, Chiche et al. (1990) demonstrated a linear relationship between these solvation free energies and the protein size, with the caveat that incorrect structures may show solvent free energies that are similar to correct structures. We used this property to examine the trajectory averages because it is a convenient way to express the solvent exposure of protein hydrophobic and hydrophilic surface groups, and because it is sensitive to changes in the protein surface composition. We have assumed that the results from the Chiche studies are also applicable to MD average structures. However, it is important to note that the solvent free energies of the protein models in this study were evaluated for the MD average structures, rather than a single structure. The MD average structures represent a coordinate average structure rather than one structure, and the uncertainty involved in the actual structure alignment will lend itself to errors in certain flexible protein regions, and particularly for amino acids on the protein surface. In addition, because the solvent free energy is a function of surface area, a change in surface area is likely to alter the relative exposure of hydrophilic and hydrophobic protein functional groups, leading to a change in the solvent free energy, relative to the starting structure. In light of this, it is not too surprising that the stabilization energies for the MD average structures is higher than that for the crystal structures. However, taken in the context that the models with the lowest RMS deviation from the crystal structure are the ones with the lowest stabil-

ization free energy (see Results) suggests that this quantity does provide a good measure for evaluating good versus poor MD models.

We used 3.5 Å for the maximum donor–acceptor distance and 90°, 100°, and 120° for the minimum donor–hydrogen–acceptor angle as the cutoff criteria for defining a hydrogen bond. Hydrogen bonds were examined in the BPTI crystal structure, trajectory average structures, and the instantaneous structures collected during the trajectory.

Acknowledgments

This work was supported by grant GM-29072 to P.A.K. We thank the UCSF Computer Graphics Laboratory supported by NIH-RR-1081 to Robert Langridge, Principal Investigator, for the graphical displays in this text. We thank the San Diego Supercomputer Center for time on their Cray YMP computer. We also thank Allison Howard for providing the energy analysis program, the ion placement program, and the program to calculate MD temperatures; David Eisenberg for providing us with the solvent free energy program; and Paul Sigler for the crystal structure of the indole-3-propionate pseudorepressor. We are also grateful to Allison Howard for her assistance in running the PXCh calculation. This study benefitted from stimulating discussions with Allison Howard, George Seibel, and David Spellmeyer.

References

Abola, E.E., Bernstein, F.C., Bryant, S.H., Koetzle, T.F., & Weng, S. (1987). In *Journal of Crystallographic Databases—Information Content, Software Systems, Scientific Applications* (Allen, F.H., Bergerhoff, G., & Seivers, R., Eds.), pp. 107–132. Data Commission of the International Union of Crystallography, Bonn, Cambridge, Chester.

Aqvist, J., van Gunsteren, W.F., Leijonmarck, M., & Tapia, O. (1985). A molecular dynamics study of the C-terminal fragment of the L7/L12 ribosomal protein: Secondary structure motion in a 150 picosecond trajectory. *J. Mol. Biol. 183*, 461–477.

Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F., DiNola, A., & Haak, J.R. (1984). Molecular dynamics with coupling to an external bath. *J. Chem. Phys. 81*, 3684–3690.

Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.F., Jr., & Tasumi, M. (1977). The Protein Data Bank: A computer-based archival file for macromolecular structures. *J. Mol. Biol. 12*, 535–542.

Brown, K.D. (1968). Regulation of aromatic amino acid biosynthesis in *Escherichia coli* K12. *Genetics 60*, 31–48.

Chiche, L., Gaboriaud, C., Heitz, A., Mornon, J.P., Castro, B., & Kollman, P.A. (1989). 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.

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

Curtiss, L.A., Frurip, D.J., & Blander, M. (1979). Studies of molecular association in H_2O and D_2O vapors by measurement of thermal conductivity. *J. Chem. Phys 71*, 2703–2711.

Daggett, V., Kuntz, I.D., & Kollman, P.A. (1991). Molecular dynamics simulations of small peptides: Dependence on dielectric model and pH. *Biopolymers 31*, 285–304.

Dauber-Osguthorpe, P., Roberts, V.A., Osguthorpe, D.J., Wolff, J., Genest, M., & Hagler, A.T. (1988). Structure and energetics of ligand binding to proteins: *Escherichia coli* dihydrofolate reductase–trimethoprim, a drug–receptor system. *Proteins 4*, 31–47.

Eisenberg, D. & McLachlan, A.D. (1986). Solvation energy in protein folding and binding. *Nature 319*, 199–203.

Gunsalus, R.P. & Yanofsky, C. (1980). Nucleotide sequence and expression of *Escherichia coli* trpR, the structural gene for the trp aporepressor. *Proc. Natl. Acad. Sci. USA 77*, 7117–7121.

Harvey, S.C. (1989). Treatment of electrostatic effects in macromolecular modeling. *Proteins 5*, 78–92.

Hingerty, B.E., Ritchie, R.H., Ferrell, T.L., & Turner, J.E. (1985). Dielectric effects in biopolymers: The theory of ionic saturation revisited. *Biopolymers 24*, 427–439.

Howard, A.E. & Kollman, P.A. (1991). Molecular dynamics studies of a DNA-binding protein: 1. A comparison of the *trp* repressor and *trp* aporepressor aqueous simulations. *Protein Sci. 1*, 1173–1184.

Huang, C., Pettersen, E., Ferrin, T., & Langridge, R. (1989 version revised from the 1985 version). *MidasPlus—Molecular Interactive Display and Simulation*. University of California, San Francisco.

Jorgensen, W.L. (1981). Transferable intermolecular potential functions for water, alcohols, and ethers. Application to liquid water. *J. Am. Chem. Soc. 103*, 335–340.

Lawson, C.L. & Sigler, P.B. (1988). The structure of trp pseudorepressor at 1.65 Å shows why indole propionate acts as a trp 'inducer.' *Nature 333*, 869–871.

Lawson, C.L., Zhang, R.G., Schevitz, R.W., Otwinoski, Z., Joachimiak, A., & Sigler, P.B. (1988). Flexibility of the DNA-binding domains of trp repressor. *Proteins 3*, 18–31.

Levitt, M. & Sharon, R. (1988). Accurate simulation of protein dynamics in solution. *Proc. Natl. Acad. Sci. USA 85*, 7557–7561.

Marmorstein, R.Q., Joachimiak, A., Sprinzl, M., & Sigler, P.B. (1987). The structural basis for the interaction between L-tryptophan and the *Escherichia coli* trp aporepressor. *J. Biol. Chem. 262*, 4922–4927.

Marmorstein, R.Q. & Sigler, P.B. (1989). Stereochemical effects of L-tryptophan and its analogues on trp repressor's affinity for operator-DNA. *J. Biol. Chem. 264*, 9149–9154.

McGeough, D., McGeough, J., & Morse, D.E. (1973). Synthesis of tryptophan operon RNA in a cell-free system. *Nature New Biol. 245*, 137–140.

Morse, D.E. & Yanofsky, C. (1969). Amber mutants of the trpR regulatory gene. *J. Mol. Biol. 44*, 185–193.

Ornstein, R.L. (1990). Using molecular dynamics simulations on crambin to evaluate the suitability of different continuum dielectric and hydrogen atom models for protein simulations. *J. Biomol. Struct. Dyn. 7*, 1019–1041.

Otwinoski, Z., Schevitz, R.W., Zhang, R.G., Lawson, C.L., Joachimiak, A., Marmorstein, R.Q., Luisi, B.F., & Sigler, P.B. (1988). Crystal structure of trp repressor/operator complex at atomic resolution. *Nature 335*, 321–329.

Pickersgill, R.W. (1988). A rapid method for calculating charge–charge interaction energies in proteins. *Protein Eng. 2*, 247–248.

Pittard, J., Camakaris, J., & Wallace, B.J. (1969). Inhibition of 3-deoxy-D-arabinoheptulosonic acid-7-phosphate synthetase (trp) in *E. coli*. *J. Bacteriol. 97*, 1242–1247.

Ramstein, J. & Lavery, R. (1988). Energetic coupling between DNA bending and base pair opening. *Proc. Natl. Acad. Sci. USA 85*, 7231–7235.

Rose, J.K., Squires, C.L., Yanofsky, C., Yang, H.L., & Zubay, G. (1973). Regulation of in vitro transcription of the tryptophan operon by purified RNA polymerase in the presence of partially purified repressor and tryptophan. *Nature New Biol. 245*, 133–137.

Rose, J.K. & Yanofsky, C. (1974). Interaction of the operator of the tryptophan operon with repressor. *Proc. Natl. Acad. Sci. USA 71*, 3134–3138.

Schevitz, R.W., Otwinoski, Z., Joachimiak, A., Lawson, C.L., & Sigler, P.B. (1985). The three-dimensional structure of trp repressor. *Nature 317*, 782–786.

Seibel, G. (1990). Computer simulation of the effects of point mutations on enzyme binding and catalysis. Ph.D. Thesis, University of California, San Francisco.

Sharp, K.A. & Honig, B. (1990). Electrostatic interactions in macromolecules: Theory and applications. *Annu. Rev. Biophys. Biophys. Chem. 19*, 301–332.

Shimizu, Y., Shimizu, N., & Hayashi, M. (1973). In vitro repression of

transcription of the tryptophan operon by trp repressor. *Proc. Natl. Acad. Sci. USA 70*, 1990–1994.

Singh, U.C. & Kollman, P.A. (1984). An approach to calculating electrostatic charges for molecules. *J. Comput. Chem. 5*, 129–145.

Singh, U.C., Weiner, P.K., Caldwell, J., & Kollman, P.A. (1989). *AMBER 3A* (G. Seibel, revision author of AMBER 3.0). University of California, San Francisco.

Singh, U.C., Weiner, S.J., & Kollman, P. (1985). Molecular dynamics simulations of d(C-G-C-G-A)·d(T-C-G-C-G) with and without "hydrated" counterions. *Proc. Natl. Acad. Sci. USA 82*, 755–759.

Squires, C.L., Lee, F.D., & Yanofsky, C. (1975). Interaction of the trp repressor and RNA polymerase with the trp operon. *J. Mol. Biol. 92*, 93–111.

van Gunsteren, W.F. & Berendsen, H.J.C. (1977). Algorithms for macromolecular dynamics and constraint dynamics. *Mol. Phys. 34*, 1311–1327.

Weiner, S.J., Kollman, P.A., Nguyen, D.T., & Case, D.A. (1986). An all atom force field for simulations of proteins and nucleic acids. *J. Comput. Chem. 7*, 230–252.

Wendoloski, J.J. & Matthew, J.B. (1989). Molecular dynamics effects on protein electrostatics. *Proteins 5*, 313–321.

Wlodawer, A., Deisenhofer, J., & Huber, R. (1987). Comparison of two highly refined structures of bovine pancreatic trypsin inhibitor. *J. Mol. Biol. 193*, 145–156.

Yanofsky, C. (1971). Tryptophan biosynthesis in *Escherichia coli*: Genetic determination of the proteins involved. *J. Am. Med. Assoc. 218*, 1026–1035.

Zhang, R.G., Joachimiak, A., Lawson, C.L., Schevitz, R.W., Otwinoski, Z., & Sigler, P.B. (1987). The crystal structure of trp aporepressor at 1.8 Å shows how binding tryptophan enhances DNA affinity. *Nature 327*, 591–597.