Study of the electrostatics treatment in molecular dynamics simulations

ARTICLE *in* PROTEINS STRUCTURE FUNCTION AND BIOINFORMATICS · NOVEMBER 1999

Impact Factor: 2.63 · DOI: 10.1002/(SICI)1097-0134(19991115)37:3<417::AID-PROT9>3.0.CO;2-U

PROTEINS: Structure, Function, and Genetics 37:417–428 (1999)

Study of the Electrostatics Treatment in Molecular Dynamics Simulations

Robert Garemyr and Arne Elofsson*
Department of Biochemistry, Stockholm University, Stockholm, Sweden

ABSTRACT **This article considers the treatment of long-range interactions in molecular dynamics simulations. We investigate the effects of using different cutoff distances, constant versus distance-dependent dielectric, and different smoothing methods. In contrast to findings of earlier studies, we find that increasing the cutoff over 8 Å does not significantly improve the accuracy (Arnold and Ornstein, Proteins 1994;18:19–33), and using a distance-dependent dielectric instead of a constant dielectric also does not improve accuracy (Guenot and Kollman, Protein Sci 1992;1:1185–1205). This might depend on differences in simulation protocols or force fields, or both, because we use the CHARMM22 force field with stochastic boundary conditions, whereas earlier studies used other protocols and energy functions. We also note that the stability of the simulations is highly dependent on the starting structure, showing that accurate molecular simulations not only depend on a realistic simulation protocol but also on correct initial conditions. Proteins 1999;37: 417–428.** © 1999 Wiley-Liss, Inc.

Key words: molecular dynamics simulations; cutoff; dielectric; electrostatic; thioredoxin

INTRODUCTION

Computer simulation techniques, such as molecular dynamics (MD), are nowadays used as a standard tool for studying the structure and dynamics of biological systems. Earlier studies,[1,2] have shown that MD simulations using explicit water molecules describe protein dynamics more accurately than do in vacuo simulations. One reason for this is that without explicit water the hydrophilic side chains have no solvent molecules to interact with and are therefore often found compacted onto the protein surface rather than being extended out into the solvent. Although computer power is rapidly increasing, inclusion of explicit solvent molecules put severe restrictions on the length of feasible sampling periods. Aqueous simulations are still limited to sampling periods of nanoseconds or less for most systems of biological interest. This stems from the fact that most of the computational burden results from the evaluation of nonbonded interactions, which scales as N^2. A number of different methods have been developed aiming to reduce the computational cost for evaluation of long-range interactions. Examples are truncation of the nonbonded interactions beyond a certain cutoff radius; reaction field methods, which replace long-range electrostatic interactions by those with a dielectric continuum;[3] and Ewald summation techniques,[4] which assume periodicity and long-range order. The most common method is the truncation method; however, there are several problems with this method, including the introduction of nonphysical forces, resulting from the discontinuity in the gradient of the energy function, and the problem of switching of one atom-atom interaction before the other in dipolar systems such as water. To overcome the first problem, a smoothening function is often used to get a smooth transition of the energy function to zero. The second problem can be handled by not allowing dipolar groups to be split by the cutoff criterion, so called group-based truncation.[5]

Several studies have aimed at finding simulation methods that strikes a good balance between computational efficiency and high accuracy. These studies include different methods to incorporate water and treating the long-range interactions. Examples are studies on pure water systems,[6] relatively small polypeptide-water systems,[3,7] and larger systems using either a water shell around the protein[2,8–10] or periodic boundary conditions.[1,10] This study is similar to the latter studies, but it also includes the dependency of the starting structure. We have used two different starting structures from the two different molecules in the asymmetric unit of 2trx and in the oxidized/reduced state of the active disulfide bridge to study the accuracy of different simulation protocols. We have chosen to perform our simulations with the stochastic boundary method[11] and to use the CHARMM param22. This method places a water sphere around the protein and treats the outer layer of the water with Langevin dynamics, i.e., trying to reproduce interactions with water outside the sphere. To the best of our knowledge this is the first extensive study of cutoff interaction using the CHARMM param22 force field in conjunction with the stochastic boundary method.

Two earlier studies, those by Arnold and Ornstein[9] and Guenot and Kollman,[2] render special interest for comparison of results. Arnold and Ornstein studied several solvent models, including a series of nonbonded cutoff distances ranging from 9 to 20 Å, using constant or distance-

Abbreviations: MD, molecular dynamics simulations; th1 and th2, first and second molecule from the 2trx PDB set.

*Correspondence to: Arne Elofsson, Department of Biochemistry, Stockholm University, 106 91 Stockholm, Sweden. E-mail: arne@biokemi.su.se

Received 16 March 1999; Accepted 21 June 1999

dependent dielectric with or without a switching function. Solvent was incorporated into the models either by x-ray crystal waters alone or in combination with different water shells with thicknesses in the range of 3 to 10 Å. The protein used in this study was phage T4 lysozyme, and 100 ps molecular dynamics trajectories were calculated using the Discover program and the consistent valence force field.[12] Among the results of this study was that (1) for models using a constant dielectric, a minimum cutoff radius of 15 Å and no switching function was required to get reasonable agreement with the crystal structure; (2) models using a constant dielectric and a cutoff radius of less than 15 Å gave large temperature differences between protein and water; and (3) using a distance-dependent dielectric resulted in significant reduction of atomic fluctuations compared with results obtained using a constant dielectric. Guenot and Kollman[2] used two different proteins, the trp repressor and BPTI, to confirm that the results were not protein dependent. Twelve different models were used for the trp repressor. Two models involved explicit water, modeled by a 4-Å thick water shell, using constant and distance-dependent dielectric, respectively. Forty picosecond trajectories were calculated using the AMBER program,[13] and the all-atom force field by Weiner et al.[14] Among the main results of the trp simulations were that using a distance-dependent dielectric resulted in smaller root mean square (rms) deviations, smaller rms fluctuations, and a significantly smaller temperature difference between protein and water compared with results using a constant dielectric. In the BPTI simulations, four different models were used, all using a 4-Å water shell but differing in that they used constant or distance-dependent dielectric and either an 8-Å cutoff or an infinite cutoff. Fifty picosecond trajectories were calculated for these models, and for the distance-dependent dielectric model with an 8-Å cutoff, the simulation was continued to 121.5 ps. The BPTI results confirmed the results from the trp study. Further, it was found that, in contrast to the constant dielectric model, the distance-dependent dielectric model showed essentially no conformational dependence on the nonbonded cutoff.

Evaluation of the ability to reproduce experimentally obtained structural and dynamic properties is complicated by several issues, e.g., (1) the significance of using crystal structures for evaluation of simulations performed in water and (2) the lack of experimental data on structural or dynamic parameters with atomic resolution on MD time scales. Concerning the first point, the following motivates our choice of simulating in solution: firstly, we wanted to examine conditions that are commonly used in molecular dynamics studies. Secondly, it is not obvious that the same set of parameters that is best for a simulation in an aqueous environment is best for a simulation in a crystal environment. Lastly, high coincidence has been shown between structural data for globular proteins obtained by x-ray crystallography and that for proteins obtained by solution NMR, despite the different environments.[15] Concerning the second point mentioned earlier, one can argue that the production of stable trajectories, not deviating far from the crystal structure, is certainly a necessary condition for a good simulation. This motivates, for instance, the use of the rms deviation between generated structures and the crystal structure as one evaluation parameter. Another frequently used measure is the change in accessible surface area. Further, it seems plausible to demand agreement between simulation rms fluctuations about the average structure and experimental temperature factors, at least qualitatively, even though the former lack contributions from external motions. Most other experimental data lack the atomic resolution to be useful to distinguish the small differences in accuracy among different simulation models.

In this article we address the behavior of MD simulations using a series of different cutoff distances, constant versus distance-dependent dielectric, and different methods for smoothing of the long-range potentials at the cutoff distance. We have used the following measures for evaluation of different methods: (1) the rms deviation of the calculated structures from the experimental x-ray structure, (2) correlations between calculated rms fluctuations about the average coordinates and the experimentally obtained temperature factors, (3) accessible surface area, (4) temperature gradient between protein and water, (5) "rms-drift," and (6) computational cost.

MATERIALS AND METHODS

We chose to perform all our simulations using *Escherichia coli* thioredoxin (Trx) because it is a fairly small and well characterized redox protein. Trx contains 108 amino acid residues, with a single disulfide bridge (C32–C35). The molecule is globular and has five β-pleated sheet strands and four α-helices. It has been shown that the structural changes occurring with the reduction of the disulfide bond do not affect the general structure of Trx but that the structure around the active site is slightly changed.[16]

The simulation method has been thoroughly described in previous studies,[11,17] and is therefore described only briefly in this article. Starting from the crystal coordinates of oxidized Trx (Trx-S_2, Protein Data Bank [PDB] code 2trx), refined at 1.7 Å,[18] the following process was applied to each of the two asymmetric crystal units, denoted by th1 and th2. A 25-Å sphere centered at a point between the aromats (Y49, Y70, W28, W31) and the active site was filled by overlaying a 25-Å radius sphere of TIP3 waters. The system was then partitioned into two parts, a full MD region of both protein and water atoms included in a 23-Å radius sphere and a Langevin region between 23 and 25 Å. After the overlaying and the partitioning of the system, the positions of the water molecules were energy minimized for 50 steps of steepest descent minimization while the protein conformation was constrained. Reorientation of the water ball was then performed followed by repartitioning of the system.[11] The entire system was then energy minimized for 200 steepest descent steps with harmonic constraints on the protein atoms that were gradually decreased every 40 steps. The whole process was repeated for two cases in which the disulfide bridge had been

reduced. The resulting four systems, which we denote by th1ox, th2ox, th1red and th2red, were the starting points for the MD runs.

All simulations were performed using CHARMM[19] parameters (param22) including all hydrogens. Initiation of the MD simulations was performed by instantaneously assigning all atoms a random velocity yielding an overall kinetic energy corresponding to 300 K (Maxwell Boltzmann distribution).

Stochastic boundary conditions were used by propagating all water atoms in the Langevin region with a friction constant of 50 ps^{-1},[17] whereas atoms within the reaction zone were propagated by the Verlet algorithm.[20] The frequency for checking whether an atom is in the Langevin region was set to 5 time steps, i.e., every 10 fs. For all atoms in the Langevin region the temperature was checked every 20 fs and allowed to vary ±10 K. All covalent bonds between hydrogens and heavy atoms were constrained using the SHAKE method.[21]

Average coordinates were calculated for each trajectory by sampling coordinates every ps, rotating each frame for best fit relative to the starting structure, followed by 200 steps steepest descent minimization for the resulting structures.

One of the most frequently used measures to assess the stability of an MD simulation over the course of time is the rms deviation (rmsd) between experimental coordinates and the generated structures, as follows:

$$rmsd = \sqrt{\frac{1}{N}\Sigma\,(r_i^{exp} - r_i^{gen})^2}$$

where $r_i^{exp.}$ and $r_i^{gen.}$ denote the cartesian coordinates for atom i in the experimental and generated structures, respectively. After rotation for best fit, averages were calculated over all nonhydrogen atoms and all backbone atoms, respectively.

To assess the agreement between atomic fluctuations and experimental B-factors we calculated the rms fluctuations (rmsf) about the average structure, as follows:

$$rmsf = \sqrt{\frac{1}{N}\Sigma < (r_i^t - r_i^{aver})^2 >_t}$$

where $<\ldots>_t$ denotes time average. The fluctuations were averaged over nonhydrogen atoms for each residue. If the fluctuations are isotropic there is an exact connection between the rms values for an individual atom i and the corresponding experimental temperature factor B_i.

$$rmsf_i = \sqrt{3B_i/8\pi^2}.$$

The B-factors for all nonhydrogen atoms, all backbone atoms, and all side chain atoms, respectively, were averaged per residue and converted to rms fluctuations using the formula above.

As a measure of how fast each model spans conformational space, we have chosen the following measure: for each trajectory the rmsd between the structure at time t and the structure at time t + Δt was calculated.[22] Averages were calculated for different values of Δt. Solvent accessible surface areas were calculated using a 1.4-Å probe.[23]

TABLE I. Simulation Models Tried in the Initial Phase of This Study

Model name	Dielectric	Cutoff (Å)	Smoothing	CPU time relative to time for 8r	rmsd th1ox
NoCut	1.0	100.0	fshift/atom	23.0	1.26
14c	1.0	14.0	fshift/atom	3.4	1.16
8c	1.0	8.0	fshift/atom	1.1	1.34
7c	1.0	7.0	fshift/atom	0.9	1.25
6c	1.0	6.0	fshift/atom	0.8	1.46
14r	$1.0r_{ij}$	14.0	shift/atom	2.9	1.35
8r	$1.0r_{ij}$	8.0	shift/atom	1.0	1.46
14c1fs[a]	1.0	14.0	fshift/atom	6.9	1.22
8r1fs[a]	$1.0r_{ij}$	8.0	shift/atom	2.3	1.10
14p19[b]	1.0	14.0	fshift/atom	2.8	1.52
14p20[c]	1.0	14.0	fshift/atom	3.6	1.53
14g	1.0	14.0	fswitch/group	3.6	1.40
8g	1.0	8.0	fswitch/group	1.2	1.45
8ge[d]	1.0	8.0	fswitch/group/ extended	2.4	2.55
14w8[e]	1.0	14.0	fshift/atom	2.3	2.90
14rsw	$1.0r_{ij}$	14.0	switch/atom	3.2	1.11
8–14rsw	$1.0r_{ij}$	14.0	switch/atom	3.3	1.27
8–14csw	1.0	14.0	fswitch/atom	8.3	1.21
10–14rswg	$1.0r_{ij}$	14.0	switch/group	2.9	1.25
8–14rswg	1.0_{ij}	14.0	switch/group	3.0	1.04
9–12csw	1.0	12.0	fswitch/atom	5.5	1.11

[a] 14c and 8r, respectively, but with 1fs timestep.
[b] Param19 force field.
[c] Param20 force field.
[d] Uses multipole approximation if dist. ($r_i \rightarrow r_j$) greater than cutoff.
[e] 8-Å cutoff used for water-water interactions, 14-Å cutoff for all other interactions.

In the initial phase of this work we tried 21 different models, simulating for 50 ps and collecting data from the last 25 ps (see Table I). The simulation models tested the following parameters: (1) constant versus linear distance-dependent dielectricity, (2) a series of nonbonded cutoff distances ranging from 6.0 Å to infinity, (3) smoothing the nonbonded truncation by using a shifting or a switching function and doing this on a group-by-group basis or on an atom-by-atom basis, (4) using a shorter cutoff distance for water-water interactions than for protein-water interactions and protein-protein interactions, and (5) different force fields. Table I summarizes the different parameters used. For models using a switching function, the default cuton-cutoff distance is 2.0 Å. However, the last five models use a larger switching region, e.g., the 8-14rswg model with cuton = 8 Å and cutoff = 14 Å. The 8ge-model invokes the extended electrostatics method built into CHARMM. This method uses a time-saving multipole approximation for electrostatic interactions between pairs of particles separated by distances larger than the cutoff radius.

We decided to proceed further with seven of the models from trial 1 plus three new models, this time simulating

TABLE II. Simulation Models Tried in the Second Phase of This Study

Model name	Dielectric	Electrostatic cutoff (Å)	Smoothing	CPU time relative to time for 8r
NoCut	1.0	100.0	fshift/atom	23.0
14c	1.0	14.0	fshift/atom	3.4
8c	1.0	8.0	fshift/atom	1.1
6c	1.0	6.0	fshift/atom	0.8
4c	1.0	4.0	fshift/atom	0.5
14r	$1.0r_{ij}$	14.0	shift/atom	2.9
8r	$1.0r_{ij}$	8.0	shift/atom	1.0
6r	$1.0r_{ij}$	6.0	shift/atom	0.7
4r	$1.0r_{ij}$	4.0	shift/atom	0.5
8–14rswg	$1.0r_{ij}$	14.0	switch/group	3.0

for 250 ps and collecting data from the last 150 ps. Table II shows the parameters for the models used in trial 2. MD trajectories were calculated for each model using each of the four starting systems, saving coordinates at 1-ps intervals. All average values were taken over the last 150 ps. In the last stage of this study we performed four 1ns simulations. This was applied only to the 8c and 14c models, using th1ox and th1red.

RESULTS

Simulating for 50 ps is clearly too short to obtain sufficient sampling of conformational space. Further, the results from trial 1 showed that for most models the rmsd was still increasing at the end of the trajectory (data not shown). However, based on average rmsd, see Table I, and rmsf, we used the results from trial 1 as a discriminator, sorting out clearly unstable models and models with questionable gain in performance, taking computational cost into account. The following models were disregarded in further studies because they were less stable than other models: 14p19, 14p20, 8ge, and 14w8. The 14c1fs and 8r1fs models were abandoned because of large computational cost and limited gain in accuracy. The 14g and 8g models were abandoned because they did not improve the results compared with the 14c and 8c models. Among the remaining models using a switching function, we decided to proceed further with the 8-14rswg model because the results indicated both computational efficiency and good stability (lowest rmsd of all models considered). Models 8r and 6c were kept for further studies solely because they are interesting as reference models. To estimate the error in our data we performed ten simulations of the 8c model (th2red) using different seed numbers for the assignment of initial velocities.

Root Mean Square Deviations

Time series for the rmsd, averaged over either all heavy atoms or backbone atoms, were calculated relative to the starting structure. Figure 1 shows the rmsd for all heavy atoms, plotted as a function of time and smoothed by averaging over 5-ps windows, for all simulations starting from th1red. The 4c model exhibits a steady increase for the first 170 ps and then reaches a plateau of about 3.9 Å. The 4r model never reaches such a plateau; the rmsd increases over the whole 250-ps interval of the simulation, ending at 5.9 Å. With the 6c and 6r models, the rmsd gets considerably lower, although the 6r model exhibits the same behavior as the 4r model, i.e., a steady increase over the 250 ps of the simulation. The NoCut, 14c, 14r, 8c, 8r, and 8-14rswg models all show lower rmsds than does the 6c model over the last 150 ps, even though the stability of these simulations is still questionable, with a possible exception for the 8-14rswg simulation. We observed the same trends for all simulations starting from the th1ox structure. For simulations using th2ox as starting structure, the main difference is that the 4c model exhibits a steady increase and ends up with a higher rms value than does the 4r model (data not shown). For th2red we observed very high rmsds for the 6c model over the last 150 ps, reaching a plateau of about 4.5 Å (data not shown).

Figure 2 summarizes the average rmsd for the last 150 ps of each trajectory relative to the crystal structure, for backbone atoms (bottom panels) and heavy atoms (top panels). It can be seen that the rmsd are larger for th2 than for th1, for all simulations using a cutoff of at least 8 Å. For the models using very short cutoff radiuses, i.e., 6c, 4c, 6r, and 4r, we cannot see this pattern.

To see if it was possible to discriminate the behavior of models using cutoffs of 8 Å and 14 Å, we decided to use the 8c and 14c models and simulate for 1 ns. Figure 3 shows the backbone rmsd as a function of time for th1ox and th1red, smoothed by averaging over windows of 5 ps. For th1ox it is hard to see any significant difference in behavior resulting from different cutoff radiuses, even though there is a weak indication of larger instability for the 8c simulation over the last 200 ps. For th1red we observe a large bulge starting at approximately 840 ps, reaching a first peak at approximately 900 ps and its maximum at approximately 940 ps for the 8c simulation, whereas the 14c model remains quite stable over the last 250 ps. To investigate the reasons for the rmsd bulge in the th1red 8c simulation, we looked at the backbone rmsds averaged by residue over the 841–901 ps period. These results showed that the largest contributions to the rmsd stems from one region, residue 30-41, with the largest values for residues 37-38. Further, concentrating on this region and the 841-ps and 901-ps structure, a dihedral angle analysis revealed large changes for Gly33 and Met37, $\Delta\phi \geq 70°$, and for Cys32, Gly33, Cys35, and Pro40, $\Delta\psi \geq 70°$. Figure 4 shows the differences between the 841-ps and 901-ps structures for residues 30-42.

Root Mean Square Fluctuations

Reproduction of the experimental structure alone is not enough to distinguish a good MD model from a bad model. Looking only at rmsds from the crystal structure would judge a model that stays close to the crystal structure as good, even if this positive result is the result of damping of important intramolecular motions or freezing out molecular degrees of freedom. For this reason, an important measure of the correct dynamic behavior is the correlation

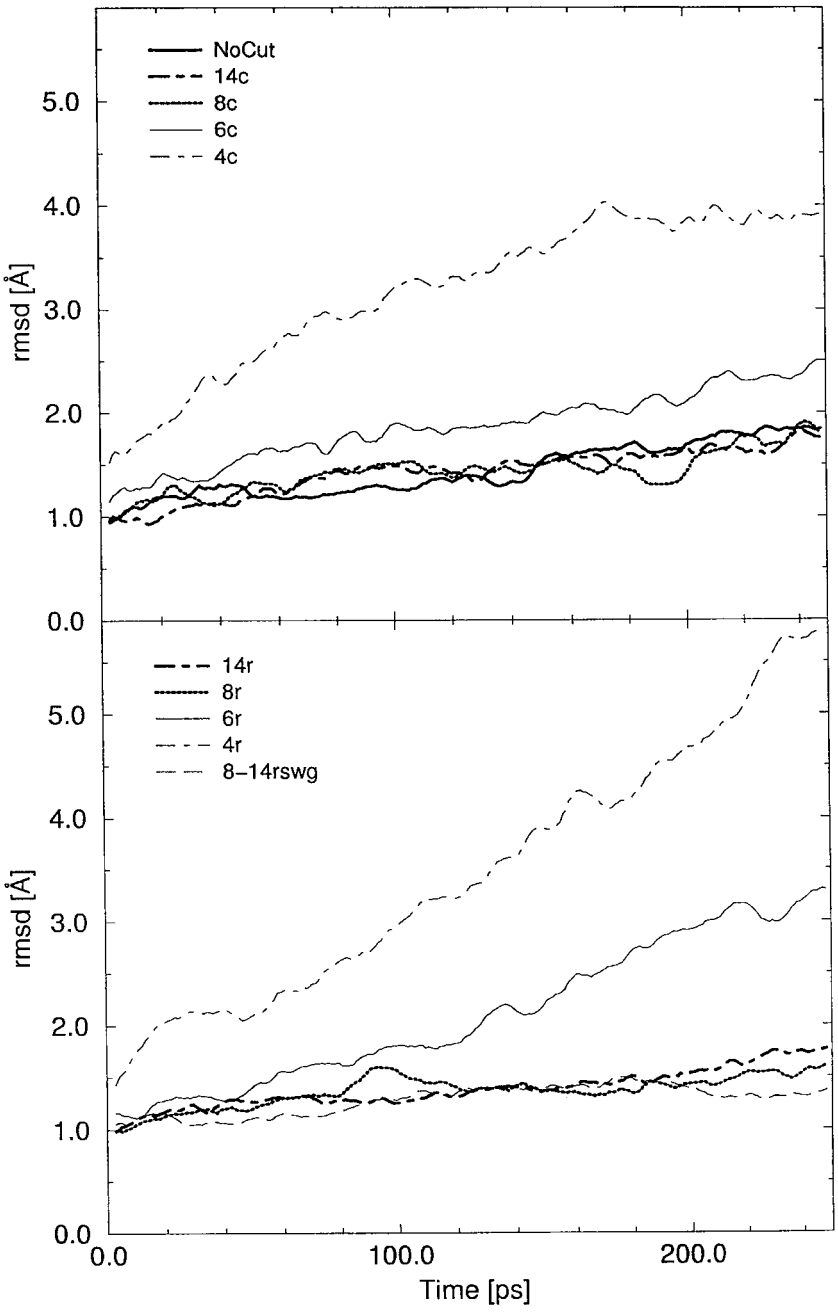

Fig. 1. Root mean square deviation development for all heavy atoms of different simulation models for th1red.

between the rmsfs about the average coordinates and experimental B-factors. Figure 5 shows the correlation coefficients between converted B-factors and the rmsfs around the average structure for the last 150 ps of each simulation.

In Table III we report the average rmsfs over the last 150 ps for each model and the average B-factors converted to rmsfs from the crystal structures. The estimated uncertainty for the average rmsfs is ± 0.03. All models using an inner cutoff radius of at least 8 Å show equal or lower rmsfs compared with the experimental data, whereas the opposite is true for the 6c, 6r, 4c, and 4r models. The 4c and 4r models shows very large average rmsfs, indicating substantial loss of stability because of the truncation of long-range electrostatic interactions.

Accessible Surface Areas

Figure 6 summarizes the average accessible surface area over the last 150 ps for each model. The estimated error is 84.3 Å^2, and the solvent accessible surface areas for the starting structures are 5,687 Å^2 (th1ox), 5,685 Å^2 (th1red), 5,762 Å^2 (th2ox), and 5,760 Å^2 (th2red). Averages for the th1ox/th1red and th2ox/th2red starting structures are shown as horizontal lines. It can be seen that the accessible surface area increases for all models and, with

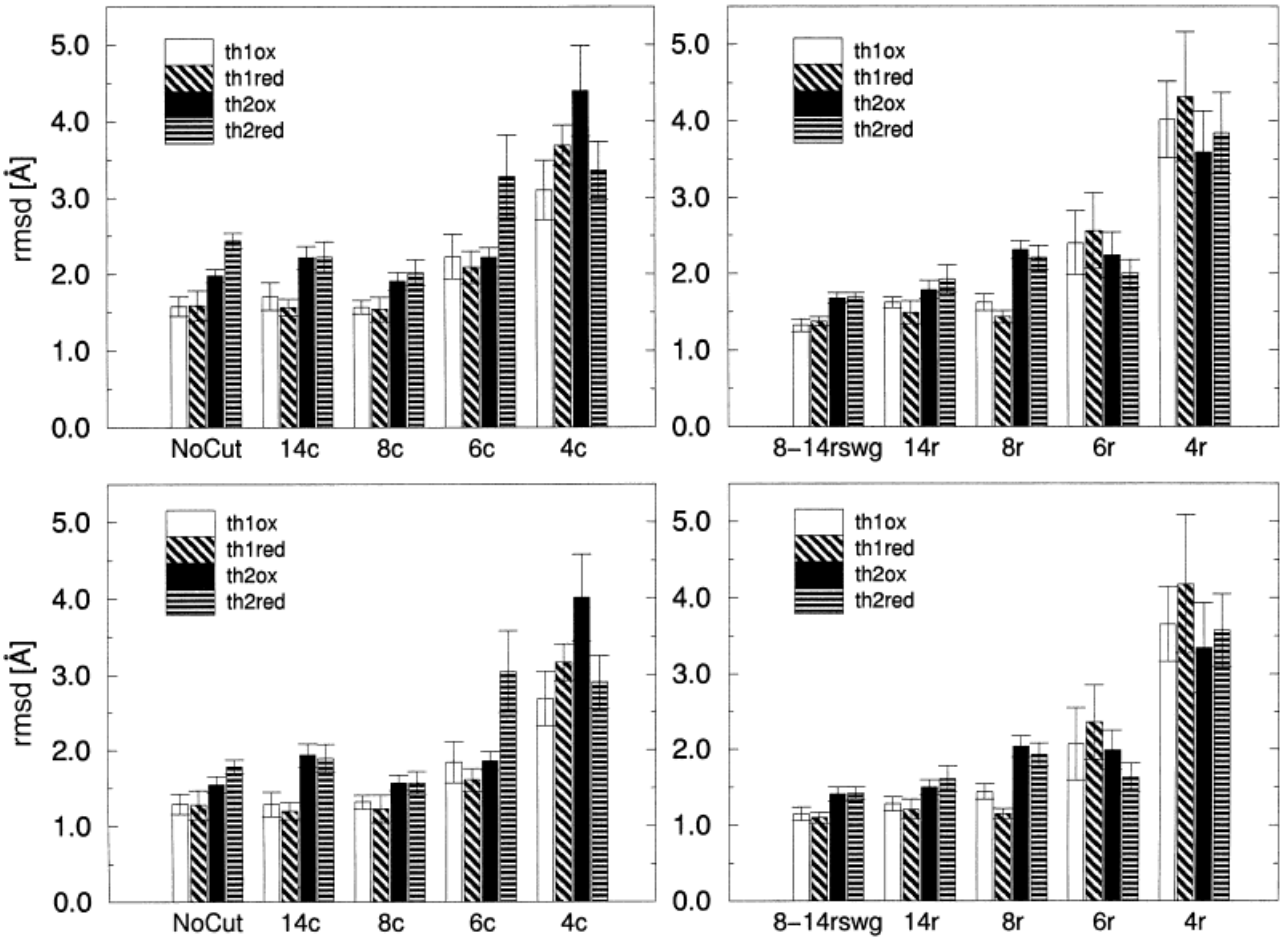

Fig. 2. Root mean square deviation between the average structure and the starting structure for all models. All averages were taken over the last 150 ps of each trajectory. **Top panels,** all heavy atoms. **Bottom panels,** only backbone atoms.

an exception for the models using a 4-Å cutoff, that the expansion is larger with a constant dielectric than with a distance-dependent dielectric. Using a cutoff radius of 4 Å leads to very large expansions even with a distance-dependent dielectric.

The behavior as a function of time (data not shown) was similar for all models using a cutoff radius of at least 8 Å. For these models the accessible surface area was very stable during the whole simulation time. The 6c and 6r models were quite stable, whereas the 4r and especially the 4c model showed dramatic expansion of the accessible surface area.

Temperature Difference Protein-Water

Earlier studies,[2,9] have shown that, especially for constant dielectric models, the truncation of the nonbonded interactions may lead to severe problems in partitioning the kinetic energy between water and protein. In both these studies the whole system was coupled to a single temperature bath. This "hot solvent–cold solute problem" can be avoided by separately coupling the solvent and solute to two temperature baths, as in later studies by Fox and Kollman.[10] Because we did not use separate coupling for protein and water in this study, we were worried that our models might exhibit large solvent-protein temperature differences. The result showed that the temperature difference was about the same, approximately 10 K, for the 14c, 6c, and 14r models. We were a bit surprised that the 6c model showed this small temperature difference because we intuitively expected the short cutoff radius to cause bad coupling between water and protein.

Root Mean Square Drift

Time averages over the last 150 ps were calculated for six different values of Δt. The results are summarized in Figure 7 for the th1ox simulations. The qualitative behavior for the other three cases was similar (data not shown). The behavior, both qualitatively and quantitatively, is very much the same for all models using a cutoff radius of at least 8 Å. The 6c, 6r, 4c, and 4r models show very large rmsds for all six values of Δt relative to the other models, with a possible exception for the 6c model, $\Delta t = 100$ ps.

Structural Differences Between the Oxidized and Reduced Forms

The solution structures of oxidized and reduced trx have been shown to be similar.[24] Slight differences were located

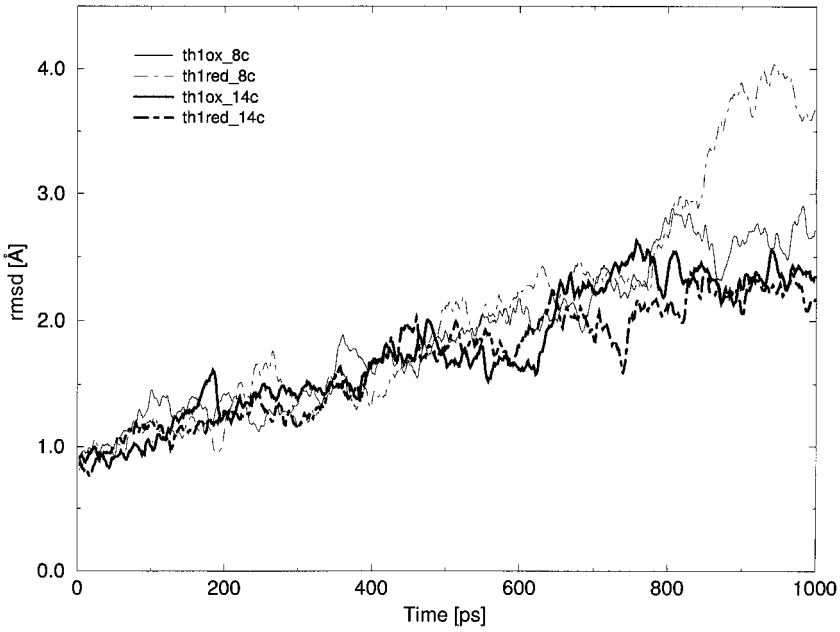

Fig. 3. Root mean square deviation from the starting structure for the 1-ns simulations of th1ox and th1red.

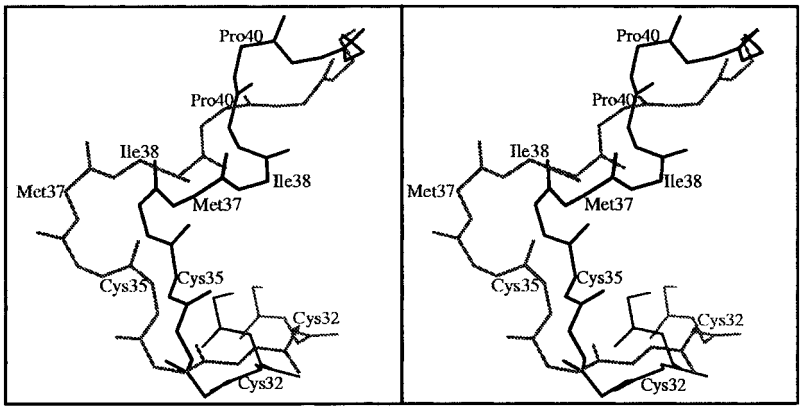

Fig. 4. The 841-ps and 901-ps structures from the th1red 8c simulation, residues 30 to 42.

in the region of the active site, which includes Cys32 and Cys35, with an increase of the S-S distance upon reduction of the disulfide. The χ^1 angles of the two cysteines were found to remain the same in the two forms. Further, it was found that the overall backbone structure of oxidized *E. coli* thioredoxin in solution is very similar to that of molecule A in the asymmetric crystal unit, i.e., th1, whereas the difference between it and molecule B, i.e., th2, is considerably larger. This motivates a comparison between the solution structures and the structures of the present study, considering only th1ox and th1red.

We have calculated the S-S distance and χ^1 angles for Cys32 and Cys35 for (1) the average solution structures of Jeng et al.,[24] PDB codes 1xoa and 1xob, (2) the th1ox and th1red starting structures, and (3) the average structures from the simulations using a cutoff radius of at least 8 Å. The results are summarized in Table IV.

DISCUSSION

Considering only the 250-ps simulations, almost all of the measures used for evaluation in this study have showed the same picture: the results for th2 are not as good as those for th1. To investigate the origin of this difference we calculated the rmsd between the th1 and th2 crystal structures and found the largest values for residues 6, 10–13, 18–20, 52, 73, 82–85, 101, 104, and 107–108 (see Fig. 8, top panel). We also note that, for all models using a cutoff radius of at least 8 Å, there is a clear tendency for higher correlations between generated rmsfs and experimental B-factors for th1ox and th1red than for th2ox and th2red (see Fig. 5). A closer look on a residue basis reveals a large bulge in the experimental B-factors for the thio2 crystal structure, located between res 10 and res 20 (see Fig. 8, bottom panel). This bulge is not reproduced by our models. Further, this is the same region Katti et al.[18] refer to as the "disordered region" of th2. It is possible that there are strains inherent in the th2 structure, possibly because of crystal contacts, especially in the region between residues 10 and 20. Therefore, simulating in a different medium than the actual crystal environment gives results deviating significantly from experimental values, e.g., large rmsds between the crystal structure and generated structures. Further, as mentioned in Results, experimental data have shown that the overall backbone

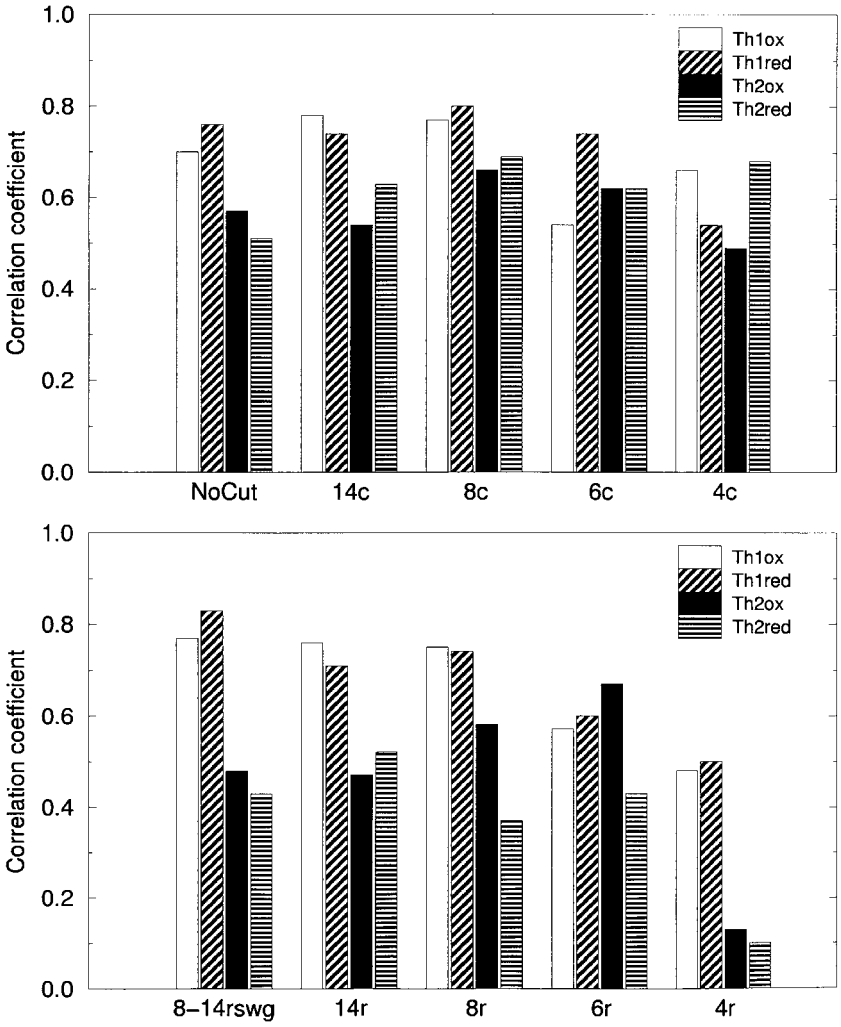

Fig. 5. Correlation between root mean square fluctuations and temperature factors. The estimated uncertainty for these correlations is ± 0.13.

structure of oxidized *E. coli* thioredoxin in solution is very similar to the th1 crystal structure whereas the difference to th2 is considerably larger.[24] This is a motivation for using only the results for th1 in the evaluation of the different models.

Considering only th1, it seems as if an 8-Å cutoff is sufficient to obtain rmsds comparable to results achieved with larger cutoffs in earlier studies. The averages for th1ox/th1red are 1.28 Å and 1.56 Å for backbone and heavy atoms, respectively. As a comparison, Arnold and Ornstein[9] report backbone rmsds of approximately 1.5 Å from simulations performed using cutoff radiuses in the range of 11.5 to 20 Å. Still, one can argue that the rmsds in the present study are significantly larger than the ones reported by Kitson et al.[25] or Guenot and Kollman,[2] for example. However, the results by Kitson et al. were obtained simulating in the actual crystal environment and without any smoothing at the cutoff. In fact, on comparison of two simulations, both using a cutoff radius of 15 Å, but one with and one without a switching function, Kitson et al.[25] found significantly lower rmsds for the model using no switching function. We also tried to use abrupt truncation, with cutoff radiuses set to 8 Å and 14 Å, but the simulations turned out to be too unstable, forcing us to use smoothing functions. Further, the protein used in the second part of the study by Guenot and Kollman,[2] (BPTI, PDB code 5pti), is smaller than trx and contains three cysteine bridges, which may explain the smaller rmsds compared with the ones in the present study. For the trp

TABLE III. Average Root Mean Square Fluctuations for All Heavy Atoms

Model	th1ox	th1red	th2ox	th2red
NoCut	0.74	0.75	0.76	0.75
14c	0.68	0.64	0.72	0.88
8c	0.70	0.79	0.74	0.81
6c	0.97	0.99	0.96	1.29
4c	1.45	1.39	1.61	1.42
14r	0.67	0.67	0.69	0.70
8r	0.67	0.66	0.76	0.86
6r	1.04	1.11	1.00	1.06
4r	1.61	1.77	1.65	1.54
8–14rswg	0.60	0.56	0.62	0.60
B-factor → rmsf	—	0.83	—	0.88

rmsf, root mean square fluctuation.

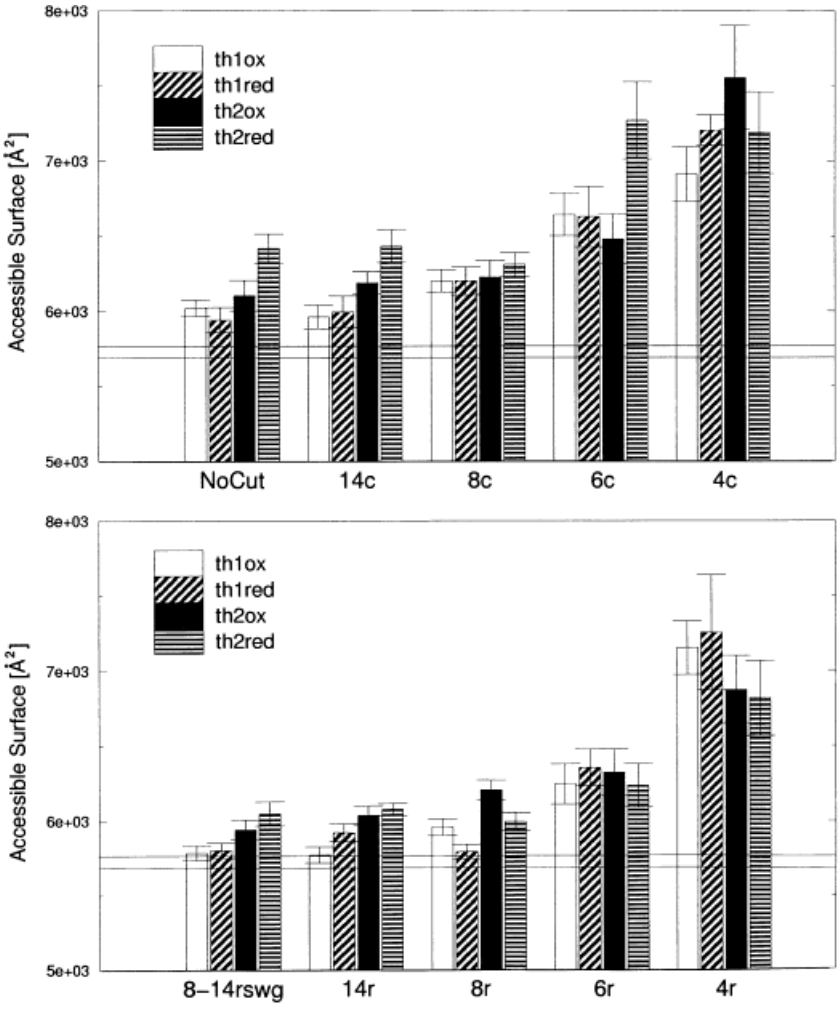

Fig. 6. Solvent-accessible surface area of the average structures.

repressor, using distance-dependent dielectric, Guenot and Kollman report rmsds to the crystal structure that are similar to our results. However, the corresponding results using a constant dielectric are considerably larger (2.23 Å).

In contrast to the results from earlier studies,[2,9] we did not observe smaller rmsfs when using a distance-dependent dielectric compared with using a constant dielectric, although the model that exhibits the lowest fluctuations is the 8-14rswg model. This particular model seems to be quite stiff, with average rmsfs of 67 to 72% of the experimental data for th1, indicating that there may be inappropriate damping of intramolecular motions. However, even though there is some damping, Figure 5 shows that the 8-14rswg model has good correlation with the experimental B-factors for th1ox and th1red. Further, it is interesting that models using a distance-dependent dielectric exhibit smaller rmsds for the rms drift, for $\Delta t \leq 10$ ps and for each value of the cutoff radius, than the corresponding model using a constant dielectric.

Earlier results[9] showed that rmsds from the crystal structure, fluctuations about the mean coordinates, and fluctuations in solvent-accessible surface area were highly correlated. The results in the present study seems to support this, as Figure 2, Table III, and the error bars in Figure 6 indicate. A thorough calculation of the correlation coefficients between the three measures gave results in the range 0.80 to 0.99. The lowest correlation is found for th1red and has its origin in the surprisingly low rmsfs in the accessible surface area for the 4c-th1red simulation.

The structural comparisons of the two cysteines in oxidized and reduced form (see Table IV) show that data for the models of the present study using a cutoff of at least 8 Å are in good agreement with the experimental solution data. Considering the th1red simulations, it is noteworthy that the S-S bond distance for the average structures is longer but quite close to the experimental value, despite the considerably shorter S-S bond distance for the starting structure. Further, for the th1ox simulations, we note a tendency in which the C32 χ^1 angles approach the experimental value, with an exception for the 14r simulation. It has to be stressed that these results are to be taken as indications of tendencies and not in too much detail because the sampling is limited.

Taking computational efficiency into account, the above results seem to suggest that the optimal cutoff radius is 8 Å. It is hard to see any improvement in results if the cutoff is extended to 14 Å, or even further as in the NoCut simulation. This is a bit surprising taking the results of

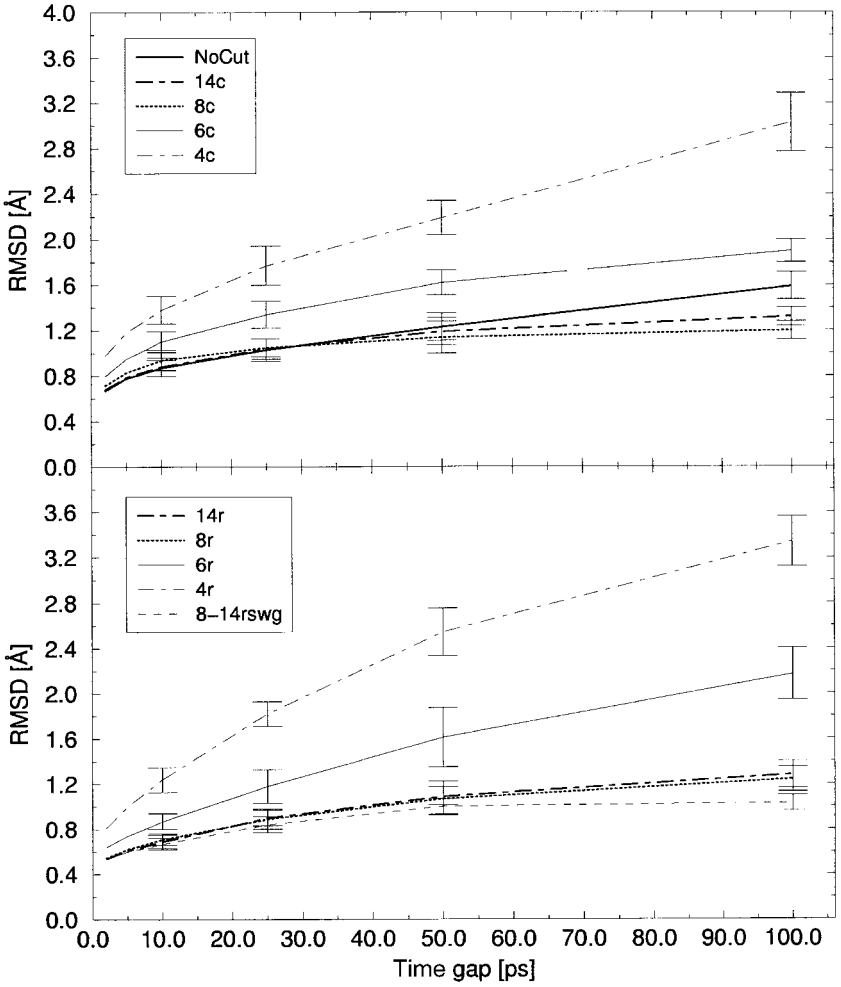

Fig. 7. Root mean square deviation drift for the different simulation models.

TABLE IV. Structure Data for the Two Cysteines, Cys32 and Cys35†

	th1ox			th1red		
	S-S distance (Å)	Cys32 χ^1 (degrees)	Cys35 χ^1 (degrees)	S-S distance (Å)	Cys32 χ^1 (degrees)	Cys35 χ^1 (degrees)
Experimental	2.07	−175.5	−59.7	3.67	−171.1	−62.5
Start	2.03	−189.1	−60.2	3.19	−171.6	−70.1
NoCut	2.03	−171.1	−54.3	4.07	−169.2	−63.5
14c	2.03	−183.1	−56.6	3.97	−165.4	−60.6
14r	2.03	−189.9	−54.5	3.82	−173.3	−61.8
8c	2.02	−177.1	−53.7	3.74	−168.1	−60.0
8r	2.03	−178.8	−48.6	3.92	−168.4	−57.0
8–14rswg	2.03	−185.5	−56.1	3.85	−169.4	−64.3

†The experimental structures are average structures of 20 experimental solution structures in oxidized and reduced form, respectively.

Arnold and Ornstein[9] into account. Using a constant dielectric they found reasonably good results only for models using a cutoff radius of 15 Å or more. This difference might depend on the difference in simulation protocol and/or potential function.

Concerning the 1-ns simulations, the aim was to further support the conclusion that no significant improvement of the results is obtained by increasing the cutoff radius from 8 Å to 14 Å. However, the results of these studies were not conclusive. We observe a tendency for less instability for simulations using a 14-Å cutoff than for the ones using an 8-Å cutoff during the last 200 ps (Fig. 3). Further studies are required to determine whether this difference is significant or not, e.g., running a series of simulations with different assignments of the initial velocities. The bulge in the rmsd for the th1red-8c simulation may be a result of

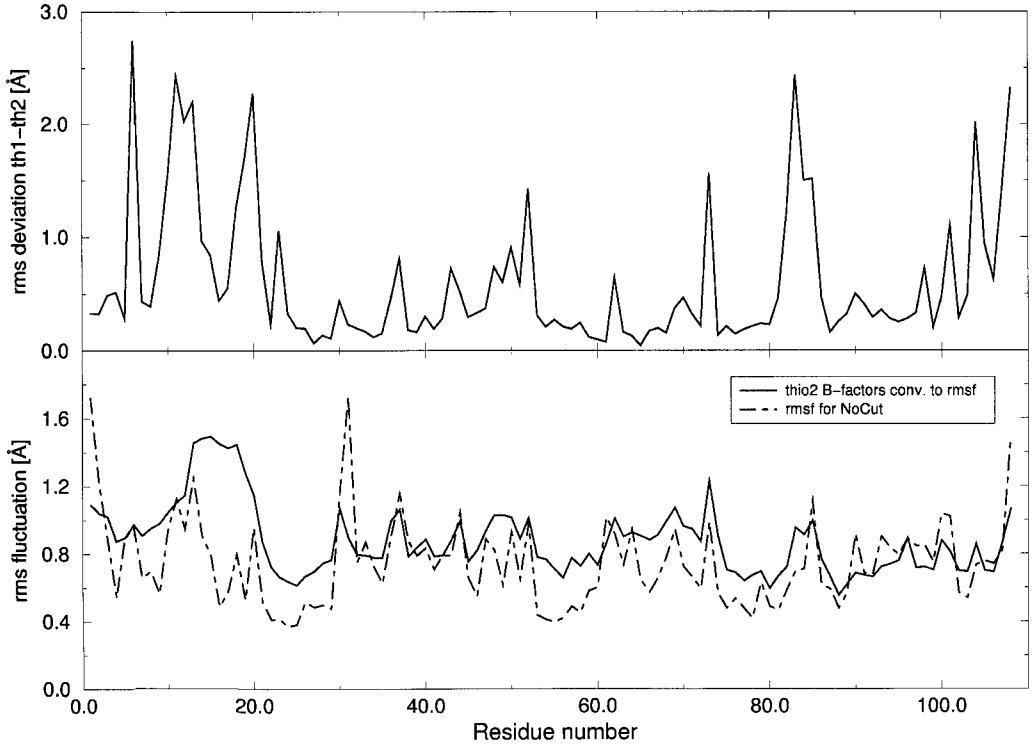

Fig. 8. **Top panel,** root mean square deviation between the th1 and th2 crystal structures versus residue number. **Bottom panel,** converted B-factors for th2 and root mean square fluctuation (rmsf) for the corresponding NoCut simulation versus residue number.

the initial assignment of velocities, or it may, in fact, be a direct consequence of neglecting important stabilizing long-range interactions. Further, one might ask if the increased stability for the th1ox-8c simulation is linked to the stabilizing effect of the disulfide bond.

CONCLUSIONS

The primary objective of this study was to find a model that strikes a good balance between computational efficiency and good agreement with experimental data. The overall picture from the measures used in this study is that the models using a cutoff distance of at least 8 Å fall into a different class than the rest of the models. Further, using a cutoff of 14 Å does not seem to improve the results significantly over those obtained using an 8-Å cutoff. The 8-Å cutoff models are three times faster than the corresponding 14-Å cutoff model (Table II). The models using a 6-Å cutoff radius show reasonably good results for average rmsds to the crystal structure, correlations with experimental B-factors, and protein-water temperature differences. However, the time development of the trajectories from these simulations indicates great instability, which is also revealed by the accessible surface areas. Models using a cutoff radius of 4 Å exhibit really poor performances for all measures used in this study, with a possible exception for the protein-water temperature difference, which was not calculated for these models. At least within the framework of stochastic boundary simulations we do not recommend using a cutoff radius of 6 Å or shorter. Finally, almost all measures used for evaluation in this study showed that the stability of the simulations was highly dependent on the starting structure. The results for th2 were not as good as those for th1.

REFERENCES

1. Levitt M, Sharon R. Accurate simulation of protein dynamics in solution. Proc Natl Acad Sci USA 1988;85(20):7557–7561.
2. Guenot J, Kollman PA. Molecular dynamics studies of a dna-binding protein: 2. An evaluation of implicit and explicit solvent models for the molecular dynamics simulation of the *Escherchia coli* trp repressor. Protein Sci 1992;1(9):1185–1205.
3. Schreiber H, Steinhauser O. Taming cut-off induced artifacts in molecular dynamics studies of solvated polypeptides. The reaction field method. J Mol Biol 1992;228(3):909–923.
4. York DM, Darden TA, Pedersen LG. The effect of long-range electrostatic interactions in simulations of macromolecular crystals: a comparison of the Ewald and truncated list methods. J Chem Phys 1993;99:8345–8348.
5. Loncharich RJ, Brooks BR. The effects of truncating long-range forces on protein dynamics. Proteins 1989;6(1):32–45.
6. Tasaki K, Mcdonald S, Brady JW. Concerning the treatment of long-range interactions in molecular dynamics simulations. J Comp Chem 1993;14(3):278–284.
7. Schreiber H, Steinhauser O. Molecular dynamics studies of solvated polypeptides—why the cut-off scheme does not work. Chem Phys 1992;168(1):75–89.
8. Guenot J, Kollman PA. Effects of truncating nonbonded interactions in an aqueous protein dynamics simulation. J Comp Chem 1993;14(3):295–311.
9. Arnold GE, Ornstein RL. An evaluation of implicit and explicit solvent model systems for the molecular dynamics simulation of bacteriophage t4 lysozyme. Proteins 1994;18:19–33.
10. Fox T, Kollman PA. The application of different solvation and electrostatic models in molecular dynamics simulations of ubiqui-

tin: how well is the x-ray structure "maintained"? Proteins 1996;25:315–334.
11. MacKerell AD, Nilsson L, Rigler R, Saenger W. Molecular dynamics simulations of ribonuclease t1: analysis of the effect of solvent on the structure, fluctuations and active site of the free enzyme. Biochemistry 1988;27:4547–4556.
12. Dauber-Osguthorpe P, Roberts VA, Osguthorpe DL, Wolff J, Genest M, Hagler AT. Structure and energetics of ligand binding to proteins: *escherichia coli* dihodryfolate reductase-trimethoprim, a drug-receptor system. Proteins 1998;4:31–47.
13. Singh UC, Weiner PK, Caldwell J, Kollman PA. AMBER 3A. San Francisco: University of California; 1989.
14. Weiner SJ, Kollman PA, Nguyen DT, Case DA. An all-atom force field for simulations of proteins and nucleic acids. J Comp Chem 1986;7:230–252.
15. Billeter M. Comparison of protein structures determined by NMR in solution and by x-ray diffraction in single crystals. Q Rev Biophys 1992;25:325–377.
16. Dyson HJ, Gippert GP, Case DA, Holmgren A, Wright PE. Three-dimensional solution structure of the reduced form of *Escherichia coli* thioredoxin determined by nuclear magnetic resonance spectroscopy. Biochemistry 1990;29:4129–4136.
17. Brooks CL, Brunger A, Karplus M. Active site dynamics in protein molecules, a stochastic boundary molecular dynamics approach. Biopolymers 1985;24:843–865.
18. Katti SK, Lemaster DM, Eklund H. Crystal structure of thioredoxin from *Escherichia coli* at 1.68 Å resolution. Biochemistry 1990;212:167–184.
19. Brooks BR, Bruccoleri RE, Olafsen BD, States DJ, Swaminathan S, Karplus M. CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. J Comp Chem 1983;4:187–217.
20. Verlet L. Computer "experiments" on classical fluids. i. Thermodynamical properties of Lennard-Jones molecules. Physiol Rev 1977;159:98–105.
21. Ryckaert JP, Ciccotti G, Berendsen HJC. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. J Comp Phys 1977;23:327–337.
22. Elofsson A, Nilsson L. How consistent are molecular dynamics simulations? Comparing structure and dynamics in reduced and oxidized *Escherichia coli* thioredoxin. J Mol Biol 1993;233:766–788.
23. Lee B, Richards FM. The interpretation of protein structures: estimation of static accessibility. J Mol Biol 1971;55(3):379–400.
24. Jeng MF, Campbell AP, Begley T, et al. High-resolution solution structures of oxidized and reduced *Escherichia coli* thioredoxin. Structure 1994;2:853–868.
25. Kitson DH, Avbelj F, Moult J, et al. On achieving better than 1-Å accuracy in a simulation of a large protein: *Streptomyces griseus* protease a. Proc Natl Acad Sci USA 1993;90:8920–8924.