JCTC Journal of Chemical Theory and Computation

Minimalist Explicit Solvation Models for Surface Loops in Proteins

Ronald P. White and Hagai Meirovitch*

Department of Computational Biology, University of Pittsburgh School of Medicine,
3064 Biomedical Science Tower 3, Pittsburgh, Pennsylvania 15260

Received December 13, 2005

Abstract: We have performed molecular dynamics simulations of protein surface loops solvated by explicit water, where a prime focus of the study is the small numbers (e.g., ∼100) of explicit water molecules employed. The models include only part of the protein (typically 500−1000 atoms), and the water molecules are restricted to a region surrounding the loop. In this study, the number of water molecules (N_w) is systematically varied, and convergence with a large N_w is monitored to reveal N_w(min), the minimum number required for the loop to exhibit realistic (fully hydrated) behavior. We have also studied protein surface coverage, as well as diffusion and residence times for water molecules as a function of N_w. A number of other modeling parameters are also tested. These include the number of environmental protein atoms explicitly considered in the model as well as two ways to constrain the water molecules to the vicinity of the loop (where we find one of these methods to perform better when N_w is small). The results (for the root-mean-square deviation and its fluctuations for four loops) are further compared to much larger, fully solvated systems (using ∼10 000 water molecules under periodic boundary conditions and Ewald electrostatics) and to results for the generalized Born surface area (GBSA) implicit solvation model. We find that the loop backbone can stabilize with a surprisingly small number of water molecules (as low as five molecules per amino acid residue). The side chains of the loop require a somewhat larger N_w, where the atomic fluctuations become too small if N_w is further reduced. Thus, in general, we find adequate hydration to occur at roughly 12 water molecules per residue. This is an important result because, at this hydration level, computational times are comparable to those required for GBSA. Therefore, these “minimalist explicit models” can provide a viable and potentially more accurate alternative. The importance of protein loop modeling is discussed in the context of these, and other, loop models, along with other challenges including the relevance of an appropriate free-energy simulation methodology for the assessment of conformational stability.

I. Introduction

A great amount of work has been devoted in the past 20 years to understanding the function and determining the structure (or structures) of protein loops. The latter is particularly important in homology modeling where one generates initially a partial structure (a template) of unconnected chain segments of a target protein on the basis of the known X-ray structure of a homologous protein (or proteins); however, it still remains to determine the structure of the connecting (missing) loops. This endeavor, which is carried out by conformational search techniques or comparative modeling, is not a trivial task and is an unsolved problem for large loops;[1−3] the structure prediction of loops constitutes a challenge also in protein engineering.

Of special interest are surface loops that take part in protein−protein and protein−ligand interactions; such loops can form “lids” over active sites of proteins, and mutagenesis

* Corresponding author. Phone: 412-648-3338. Fax: 412-648-3163. E-mail: hagaim@pitt.edu.

10.1021/ct0503217 CCC: $33.50
Published on Web 05/02/2006

experiments show that residues within these loops are crucial for substrate binding or enzymatic catalysis.[4] Typically, these loops are flexible, and their flexibility is essential for protein function. Two general recognition mechanisms related to flexibility have been defined, *induced* and *selected fit.* Thus, the conformational change between a free and a bound antibody demonstrates the flexibility of the antibody combining site, which typically includes hypervariable loops; this provides an example of induced fit as a mechanism for antibody−antigen recognition (see, for example, refs 5 and 6). Alternatively, the *selected-fit* mechanism has been suggested, where a free loop interconverts among different microstates in thermodynamic equilibrium, and one of them is selected upon binding[7] (a microstate is a limited region in conformational space such as the helical region of a peptide). While loop flexibility can be detected by multidimensional nuclear magnetic resonance (NMR) and X-ray crystallography (in terms of elevated B factors in the latter method), using these methods to map the most stable microstates of an unbound loop (i.e., those with the lowest free energy) is problematic, and one has, therefore, to resort to molecular modeling techniques.

The interest in surface loops has yielded extensive theoretical work, where one avenue of research has been the classification of loop structures.[7,8−15] However, to understand various recognition mechanisms such as those mentioned above, it is mandatory to be able to predict the structure of a loop by theoretical/computational procedures. The commonly used methodologies in this category are comparative modeling based on known loop structures from the Protein Data Bank (PDB),[16,17] an energetic modeling (based on a force field), and methods that are hybrids of these two approaches. However, mapping the most stable microstates can only be achieved with the energetic approach that consists of calculating the loop−loop and loop−protein interaction energies. To be able to apply such calculations to a large number of loops, the entire protein structure has typically been kept fixed in its X-ray structure (and sometimes only part of it has been considered). Because of the exposure of surface loops to the solvent, the development of adequate modeling of solvation is mandatory. The most stable microstates can then be generated by a combination of conformational search techniques (simulated annealing, the bond relaxation algorithm, the local torsional deformation method, etc.); thermodynamic sampling methods, such as molecular dynamics (MD) simulation or Monte Carlo; and methods for calculating the free energy.[18−31]

Modeling of the solvent is of special importance. In some of the earlier studies, the solvation problem was not addressed at all, while others only use a distance-dependent dielectric function (i.e., $\epsilon = r_{ij}$ is substituted in the Coulomb potential, $E = q_iq_j/[r_{ij}\epsilon]$, making the interactions decay more rapidly as r_{ij}^{-2}). Better treatments of solvation were applied by Moult and James[23] and Mas et al.[32] A systematic comparison of solvation models was first carried out by Smith and Honig,[33] who tested the $\epsilon = r$ model against results obtained by the finite difference Poisson Boltzmann calculation including a hydrophobic term; the implicit solvation model of Wesson and Eisenberg[34] with $\epsilon = r$ was also studied by them. Later, the generalized Born surface area (GBSA) model[35] was applied to loops of ribonuclease A[36] and has been found by Blundell's group to discriminate better than other models between the native loop structures and close-to-native "decoy" structures.[37,38] Very recently, an extensive study of loops was carried out by Jacobson et al.,[39] who used the surface GB[40] and a nonpolar solvation model[41] (SGB−NP) with the OPLS force field.[42] Zhang et al.[43] have tested their knowledge-based statistical potential, DFIRE (distance-scaled, finite ideal gas reference state), by applying it to the loop sets studied in refs 37−39. Another interesting loop prediction algorithm has been suggested by Xiang et al.,[44] and finally, we mention our loop studies, using a simplified implicit model.[30,31]

The popularity of implicit solvent models for loops stems from their relative simplicity and the fact that the loops are applicable to a wide range of conformational search techniques, in particular, those that are based on energy minimization. At least in principle, an energy-minimized implicit model can be used as a gauge of loop stability (i.e., the free energy), because the solvent coordinates have been "averaged out". (Note, however, that this still does not account for the very important free-energy contribution associated with the movement of the loop atoms within a microstate.) On the other hand, explicit solvation—the more accurate modeling—is computationally expensive and allows application of limited types of search techniques. Therefore, systematic studies of loop structure prediction with explicit water have not been carried out; however, certain problems involving loops have been studied with explicit water.[45]

While the quality of these implicit models for loops has not been compared, most of them were found to be adequate for predicting the backbone structure of loops (in the known protein framework) of up to nine residues [i.e., a prediction within 1 Å root-mean-squared deviation (RMSD) from the X-ray crystal structure[23]]. However, the correlation between low free energy and low RMSD of structures generated by conformational search were found to be unsatisfactory (in particular, for highly charged loops), meaning that implicit modeling, in most cases, is not suitable for mapping the most stable microstates, and for that, one will have to resort to explicit solvation models. We have a special interest in such problems, as discussed in refs 30, 31, and 46 and in the Conclusions section.

Therefore, the objective of this article is to examine the validity (and efficiency) of explicit solvation models defined within the framework of the limited model mentioned above, where the loop moves in the presence of a fixed protein structure. Here, the loop is "capped" with a number of water molecules (N_w), and our aim is to determine the minimal N_w which still leads to reliable results. More specifically, we use the TIP3P model of water[47] and simulate the protein−loop−water system by MD,[48,49] where only the loop atoms and the surrounding waters are allowed to move while the rest of the protein atoms are kept in their X-ray coordinates; moreover, to further save computer time, we retain in the model only the part of the protein that is close to the loop. To gauge performance, the RMSDs of the heavy backbone and side-chain atoms from the X-ray structure are calculated together with the RMSD fluctuations and other quantities.

For the test cases studied here, the X-ray backbone loop structure is well-determined; that is, its atoms are defined with relatively low B factors; therefore, if the simulation starts from the X-ray structure, for a large enough N_w, one would expect the simulated backbone to demonstrate stability, that is, to remain close to this structure for long simulation times, while for a small N_w, the backbone might escape to another microstate. On the other hand, some of the coordinates of the side-chain atoms are typically poorly resolved (high B factors), and in general, the side-chain environment in the simulation could be expected to mimic the experimental solution environment better than that of the crystal; therefore, for the side chains, one would not expect the simulation to always reproduce the crystallographic data. However, as N_w is increased, the structure of the simulated side chains is expected to stabilize at some microstate. These are some of the criteria according to which the results are analyzed. (It should be emphasized, however, that during a long enough simulation the loop will change microstates, and therefore, such an analysis should be carried out with caution.) Finally, as further criteria to test the validity of the restricted (or “minimalist”) loop models studied here, we solvate the corresponding (entire) proteins with water under periodic boundary conditions, simulate them by MD, and compare the RMSD and fluctuations of the loops to those obtained from the restricted solvation models.

In this work, extensive MD simulation studies are carried out for four loops ranging in size from 8 to 10 residues; the loops are taken from the three proteins ribonuclease A (RNase A), ser-proteinase, and proteinase. (We also report results from less extensive tests conducted on several other loops.) As mentioned above, different N_w’s are tested for each loop (and other modeling parameters discussed below), and the minimal N_w [N_w(min)] which reproduces the large N_w behavior is determined. While N_w(min) depends on various properties of the loop and its associated nearby protein environment, for the four primary loops studied, we find N_w(min) ∼ 12 per residue, which (using the AMBER96 force field[50] programmed in the package TINKER[51]) requires comparable computer time to running MD based on the implicit solvent, GBSA.[35] It is also shown that for two loops the GBSA results deviate significantly from those obtained with the explicit solvent. While these results are expected to be typical, they should be validated for each loop studied.

It should be pointed out that approximate explicit solvation models, where only part of the protein (around the active site) is considered (and solvated), have been suggested before. One of the first was the stochastic boundary model of Karplus’ group, where the region of interest (including the protein and the solvent) is divided into subregions of decreasing importance;[52] we have used this model for calculating the backbone entropy of loops in the protein ras.[53] In many other studies of ligands in active sites, caps of water molecules were built around these sites, with the number of water molecules typically increasing as computers have become more powerful. For example, in 1986, Bash et al.[54] used only 168 waters to cover the active site of thermolysine in their calculations of the relative free energy of binding of two inhibitors, whereas in 1991, Merz used 300 waters for calculating the binding of CO_2 to human carbonic anyhdrase II.[55] In 1993, Miyamoto and Kollman used 205 waters to solvate the active site of streptavidin in their calculation of the absolute free energy of binding of biotin and other similar ligands to this protein.[56] In 1997, Jorgensen’s group capped 482 waters around the active site of trypsin and calculated the binding affinities of trypsin−benzamidine complexes;[57] however, in later publications of this group, caps including up to 1600 waters were used.[58] In most of these works, a systematic investigation of the effect of the number of water molecules has not been carried out. Our present study has been largely motivated by the work of Steinbach and Brooks,[59] who studied, by MD, the change in the RMSD of protein structures from their X-ray structures with an increasing number of water molecules; they found that a relatively small number of waters led to the behavior of the fully solvated system.

II. Methods

II.1. Models. Our investigations are focused on the solvation of protein surface loops with small numbers of explicit water molecules, N_w. The protein portion of these models is further limited to just the loop atoms, and only the protein atoms belonging to residues that are close to the loop. We will refer to this as the “partial-protein model”. To test the approximations inherent in this model (which are chiefly limited solvation, a reduced protein environment, and lack of flexibility in the template), we also model the entire protein, solvated under periodic boundary conditions with particle mesh Ewald electrostatics. This model is referred to as the “full-protein model”. Both models will be described in detail in the following sections.

All computational work associated with the partial-protein modeling (i.e., structure preparation and simulations) was performed using the TINKER software package (version 4.2),[51] which was modified to suit our specific needs. The computational work for the full-protein models (structure preparation, simulations, and analysis) was performed using a variety of programs in the AMBER software package (version 8). For both models, we used the AMBER96 force field,[50] where His is in the doubly protonated state (charge = +1) and four other residues are also modeled in their respective neutral pH charged states, Lys (+1), Arg (+1), Asp (−1), and Glu (−1). The water molecules are modeled with the three-site TIP3P potential.[47]

II.2. Construction of the Partial-Protein Model. The starting coordinates for the partial-protein model are taken from the PDB X-ray structure (where hydrogen atoms and disulfide bonds are added in the usual manner). As stated above, the loop atoms, and only the protein atoms that are close to the loop, are included in the model. The nonloop atoms which are retained in the model are collectively referred to as the “template”. To construct the template (see also Figure 1), the center of mass of the loop backbone atoms is calculated as a reference point. We denote the coordinates of this point as $\mathbf{x}_{\mathrm{cmb}}$. A distance ($R_{\mathrm{temp}}$) is chosen such that residues that are greater than R_{temp} from $\mathbf{x}_{\mathrm{cmb}}$ are not included

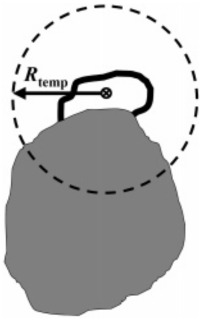

Figure 1. Diagram showing the region of the protein that is retained (the "template") in the partial-protein model. The loop is represented as the heavy black curve. The remainder of the protein is shown as a gray blob. The center of mass of the loop backbone, $\mathbf{x}_{cmb}$, is located at the position marked as ⊗. The protein template is "cut out" at the dashed circle (a sphere in three dimensions), which is defined by the distance R_{temp} measured from $\mathbf{x}_{cmb}$. All protein residues that are inside this region are considered in the model, thus defining the nearby protein environment for the loop.

Table 1. Diffusion Properties of Water Molecules Calculated for the Partial-Protein Model of RNase A [64−71][a]

N_w	R_{cap} (Å)	$\langle N_{surf} \rangle$	$\langle N_{surf}/N_w \rangle$	D_{all}	D_{surf}	τ_{all} (ps)	τ_{surf} (ps)
300	20	103.2	0.344	4.96	2.61	6.8	12.9
200	19	96.8	0.484	4.03	2.53	8.4	13.3
120	18	79.1	0.659	2.73	1.98	12.4	17.0
70	17	57.8	0.825	1.71	1.43	19.8	23.6
50	17	44.4	0.888	1.28	1.12	26.5	30.2

[a] N_w is the number of water molecules. R_{cap} is the radius of the spherical solvent restraining region (SPH restraint). The same protein template (R_{temp} = 15 Å) was used in all cases. $\langle N_{surf}/N_w \rangle$ is the (average) fraction of water molecules observed at the surface of the protein. D_{all} is the diffusion constant calculated for all N_w water molecules. D_{surf} is the diffusion constant calculated for just the water molecules at the protein surface. Units for D_{all} and D_{surf} are 10^{-5} cm^2/s. τ_{all} and τ_{surf} are estimated residence times defined by the time for a water molecule to diffuse a distance of 4.5 Å. τ_{all} is calculated for all N_w water molecules, and τ_{surf} is for the protein surface water only. Statistical uncertainties in $\langle N_{surf}/N_w \rangle$, D_{all} (D_{surf}), and τ_{all} (τ_{surf}) are typically less than 0.003, 0.05 × 10^{-5} cm^2/s, and 0.5 ps, respectively. Other details and definitions are given in the text.

in the template. More specifically, if any atom in a protein residue is less than the distance R_{temp}, from $\mathbf{x}_{cmb}$, the entire residue is included in the template. Otherwise, the residue is eliminated. Obviously, the choice of R_{temp} will determine the number of environmental protein atoms to be included in the model. Atom numbers for various R_{temp} values are given in Table 1 for each of the loops studied.

The starting (PDB) coordinates for the loop and template atoms are relaxed to a nearby geometry. This minimization is carried out using additional harmonic positional restraints ($k = 5$ kcal mol^{-1} Å^{-2}), which are applied to all heavy atoms. This eliminates bad atomic overlaps and strains in the original structure, while keeping the atoms still reasonably close to the PDB coordinates. These resulting relaxed coordinates are referred to as the "X-ray reference coordinates" and are denoted as $\mathbf{X}_{ref}$. [Note that $\mathbf{x}_{cmb}$ (above) is a single point in 3D space, whereas $\mathbf{X}_{ref}$ specifies the whole coordinate set for a group of atoms.] The loop coordinates from this configuration are used in the RMSD calculations, described below.

As outlined in the Introduction, MD simulations of the loop are carried out in the presence of the nearby template atoms, along with the N_w water molecules. Specifically, the coordinates of the loop atoms evolve in time under the influence of interactions with the template atoms, the water molecules, and each other. The water molecules are also mobile; they interact with each other, the protein atoms (in both the loop and template), and the boundary of a containment region (described below). The template atoms, however, are fixed in these simulations at their respective coordinates in $\mathbf{X}_{ref}$ (where the purpose of this approximation is to increase the computational efficiency, as it is then unnecessary to calculate template-atom−template-atom interactions).

II.3. Solvation of the Partial-Protein Model. To make best use of (the solvating effects of) the limited number of water molecules, they are restricted to a region that is close to the loop. This also prevents evaporation. The situation is similar to "capping" an active site, where one wishes to keep water molecules near the most critical region of the model investigation. Unlike many active sites, however, which tend to be concave, a solvation region around a surface loop tends to be more convex and, thus, can present more of a challenge. We have implemented two methods to restrain the water molecules to the vicinity of the loop. One involves a (semi-) spherical restraining region, which we call the SPH restraint. The other is a nearest-loop-atom-based restraint, which we call the NLA restraint. Both will be described in detail below.

II.3.1. Spherical Restraining Region. In the SPH restraint, water molecules are restrained with a flat-welled half-harmonic potential (force constant, $k = 5$ kcal mol^{-1} Å^{-2}), based on the distance from the "center" of the loop region. That is, the distance of each water molecule (in practice, the oxygen atom) is measured from a restraining center ($\mathbf{x}_{sph}$). If this distance is greater than a prescribed distance, R_{cap}, a harmonic restoring force is applied; otherwise, the restraining force is zero.

A reasonable restraining center could be, for example, the center of mass of the loop backbone atoms (i.e., $\mathbf{x}_{sph} = \mathbf{x}_{cmb}$). The choice of R_{cap}, on the other hand, should be roughly based on the number (N_w) of water molecules used. It is important to note, however, that a range of reasonable R_{cap} values can be found; but obviously, for large N_w's, small values of R_{cap} would be undesirable. (This scenario would be evidenced, for example, by a large average value for the restraining potential.) Some examples of R_{cap} at various values of N_w are available in Tables 2−7, where, in general, R_{cap} increases with N_w. In our modeling, the restraining volume is typically quite large for the given number of water molecules (i.e., much of the available volume is empty). A more detailed discussion describing the nature of the solvation within the partial-protein model will be given in section III.

In most cases, we have taken values of R_{cap} where $R_{cap} \geq R_{temp}$. (As described above, R_{temp} is the distance value used to determine the size of the template.) However, for large-enough values of R_{cap}, water molecules can migrate away

Table 2. Description of Loops and Modeling Parameters[a]

protein	number of atoms: protein	N_w^{pbc}	loop residues	sequence	R	number of atoms: loop	R_{temp}	number of atoms: loop + template
RNase A (1rat)	1860	6808	64−71 (8)	AC**K**NGQTN	3.2	107	14, 15, 16	526, 572, 590
ser-proteinase (2ptn)	3223	9320	143−151 (9)	NT**K**SSGTSY	4.9	117	13, 14, 15	498, 578, 738
proteinase (2apr)	4714	12393	128−137 loop1 (10)	**D**TITTV**R**GV**K**	4.3	158	11, 13, 15	497, 731, 1035
proteinase (2apr)	4714	12393	188−196 loop2 (9)	I**D**NS**R**GWWG	4.5	143	11, 13, 15	569, 775, 1034

[a] Atom numbers are provided for different portions of the system: the entire protein, the loop atoms only, and the loop together with the template. The number of atoms in the latter depends on the template radius parameter R_{temp} (in Å), where the values separated by commas give rise to the corresponding (comma separated) atom numbers. N_w^{pbc} is the number of water molecules used in the full-protein simulations. Loop sequences are given with the charged residues as bold-faced letters. R is the ratio between the length of the stretched loop and the distance between the C$^\alpha$ of the first and last residues of the loop.

Table 3. Partial-Protein Model Results for RNase A [64−71][a]

N_w	R_{temp}	water restraint	R_{cap} (or R_{nla})	RMSD(BB)	RMSD(SC)	σ(BB)	σ(SC)	σ^w(BB)	σ^w(SC)
300	15	SPH	20	0.57 (5)	1.31 (4)	0.19 (6)	0.48 (3)	0.14 (1)	0.28 (1)
200	15	SPH	19	0.54 (2)	1.13 (11)	0.17 (1)	0.38 (8)	0.15 (1)	0.24 (2)
200	15	SPH	16	0.55 (2)	1.23 (8)	0.17 (2)	0.44 (5)	0.15 (1)	0.26 (2)
200	16	SPH	19	0.56 (3)	1.22 (12)	0.17 (1)	0.37 (5)	0.15 (1)	0.25 (2)
120	14	SPH	18	0.64 (23)	1.21 (19)	0.18 (4)	0.31 (6)	0.15 (2)	0.21 (5)
120	15	SPH	18	0.54 (4)	1.02 (6)	0.17 (2)	0.29 (3)	0.15 (1)	0.20 (3)
120	15	NLA	8.5	0.67 (19)	1.09 (8)	0.24 (10)	0.36 (6)	0.15 (1)	0.23 (3)
120	16	SPH	18	0.61 (16)	1.35 (4)	0.21 (7)	0.34 (2)	0.14 (1)	0.23 (2)
100	15	SPH	16	0.52 (1)	1.00 (11)	0.15 (1)	0.27 (8)	0.14 (1)	0.20 (3)
70	14	SPH	14	0.50 (2)	1.27 (14)	0.15 (1)	0.30 (4)	0.13 (1)	0.21 (2)
70	15	SPH	17	0.50 (4)	1.02 (14)	0.17 (3)	0.26 (8)	0.14 (1)	0.17 (3)
70	15	NLA	7	0.58 (14)	1.00 (8)	0.19 (9)	0.27 (5)	0.14 (0)	0.19 (2)
70	16	SPH	16	0.52 (4)	1.40 (8)	0.18 (3)	0.33 (7)	0.15 (2)	0.21 (3)
50	15	SPH	17	0.50 (3)	0.99 (9)	0.18 (3)	0.22 (3)	0.15 (1)	0.15 (1)
50	15	NLA	7	0.49 (1)	1.10 (10)	0.14 (1)	0.28 (3)	0.12 (1)	0.18 (2)
50	15	SPH	16	0.57 (13)	1.09 (38)	0.21 (6)	0.29 (20)	0.15 (2)	0.16 (2)
40	15	SPH	16	0.55 (10)	1.10 (7)	0.20 (9)	0.27 (5)	0.14 (3)	0.16 (1)
30	15	SPH	16	0.55 (8)	1.07 (6)	0.21 (7)	0.21 (5)	0.15 (4)	0.15 (3)
20	15	SPH	16	0.51 (5)	0.99 (5)	0.19 (4)	0.17 (2)	0.14 (2)	0.13 (1)
10	15	SPH	16	0.67 (10)	1.21 (5)	0.22 (5)	0.18 (3)	0.18 (3)	0.15 (2)
5	15	SPH	16	1.03 (39)	1.59 (41)	0.27 (6)	0.24 (10)	0.21 (3)	0.16 (2)
0	15			2.30 (85)	2.60 (82)	0.48 (36)	0.44 (37)	0.18 (2)	0.13 (1)
GBSA	15			1.93 (48)	2.71 (62)	0.54 (11)	0.70 (16)	0.29 (5)	0.32 (4)

[a] N_w is the number of water molecules. R_{temp} (in Å) is a radius parameter defining the size of the template. The water restraint method is either "SPH" (spherical restraining region) or "NLA" (nearest-loop-atom-based restraint), which are described (respectively) by the parameters R_{cap} or R_{nla} (Å). The RMSD values (eq 2, averaged over all five trajectories) for the loop backbone (BB) and side-chain (SC) atoms are denoted by RMSD(BB) and RMSD(SC), respectively. The corresponding RMSD fluctuations (eqs 3 and 6, averaged over all five trajectories) are denoted σ(BB) and σ(SC), while the window-averaged RMSD fluctuations are denoted as σ^w(BB) and σ^w(SC). The numbers in parentheses are the standard deviations of the individual results from the five trajectories. For example, 1.31 (4) means that the standard deviation is 0.04, and 1.09 (38) implies a standard deviation of 0.38. All RMSD values and their fluctuations, σ, are reported in Å.

from the loop, around to the "back side" of the template, where their solvation effect is wasted. For this reason, we actually choose a restraining center such that

$$\mathbf{x}_{sph} = \mathbf{x}_{cmb} + (R_{cap} - R_{temp})(\mathbf{x}_{cmb} - \mathbf{x}_{cm})/|(\mathbf{x}_{cmb} - \mathbf{x}_{cm})| \quad (1)$$

where $\mathbf{x}_{cm}$ is the overall center of mass of the loop-template system. Here, the effect is to shift the center of the restraining sphere ($\mathbf{x}_{sph}$) toward the "loop side" of the loop-template system (see Figure 2). This serves to keep the water molecules away from the back of the template, because the van der Waals radii of the "back side" template atoms will now be closer to the wall of the restraining sphere. At the same time, there will be sufficient room for water on the "loop side" of the template.

II.3.2. Nearest-Loop-Atom-based Restraint. A slightly more elaborate restraint option for the water molecules is to employ a flat-welled half-harmonic potential ($k = 5$ kcal mol^{-1} Å^{-2}) that is based on the distance to the nearest loop atom (for an example, see ref 60). Specifically, for each water molecule, the distance to the nearest loop atom is calculated and then compared to a prescribed distance value, R_{nla}. If this distance is less than R_{nla}, then there are no restraining forces. If the distance is greater than R_{nla}, then a harmonic restoring force is applied to (the oxygen of) the water molecule and is directed along the vector between the water and the nearest loop atom (see Figure 3).

The NLA restraint is arguably advantageous compared to the SPH restraint, because the implementation can be

Table 4. Partial-Protein Model Results for Ser-Proteinase [143−151][a]

N_w	R_{temp}	water restraint	R_{cap} (or R_{nla})	RMSD(BB)	RMSD(SC)	σ(BB)	σ(SC)	σ^w(BB)	σ^w(SC)
300	13	SPH	20	0.69 (1)	1.51 (3)	0.14 (1)	0.26 (2)	0.13 (1)	0.20 (1)
200	13	SPH	19	0.69 (1)	1.53 (2)	0.14 (1)	0.25 (2)	0.13 (0)	0.20 (1)
200	15	SPH	19	0.69 (1)	1.44 (3)	0.12 (0)	0.26 (2)	0.12 (0)	0.20 (1)
120	13	SPH	18	0.68 (1)	1.51 (7)	0.14 (1)	0.29 (3)	0.12 (1)	0.20 (1)
120	13	NLA	8.5	0.67 (1)	1.50 (7)	0.13 (1)	0.26 (3)	0.12 (1)	0.19 (1)
120	14	SPH	18	0.67 (3)	1.54 (11)	0.14 (1)	0.29 (3)	0.12 (1)	0.20 (1)
120	15	SPH	18	0.64 (1)	1.39 (8)	0.12 (1)	0.27 (2)	0.11 (1)	0.19 (1)
70	13	SPH	17	0.80 (11)	1.49 (13)	0.22 (6)	0.32 (4)	0.15 (2)	0.18 (1)
70	13	NLA	7	0.69 (3)	1.46 (3)	0.17 (3)	0.28 (3)	0.13 (1)	0.20 (1)
70	14	SPH	17	0.83 (10)	1.57 (15)	0.25 (5)	0.37 (13)	0.16 (1)	0.19 (1)
70	14	NLA	7	0.65 (1)	1.40 (4)	0.13 (1)	0.27 (3)	0.12 (1)	0.19 (1)
50	13	SPH	17	1.28 (40)	1.88 (32)	0.30 (7)	0.35 (7)	0.17 (4)	0.17 (3)
50	13	NLA	7	0.75 (5)	1.55 (5)	0.21 (6)	0.32 (1)	0.15 (1)	0.19 (1)
GBSA	13			0.71 (3)	1.52 (9)	0.18 (2)	0.31 (2)	0.16 (1)	0.24 (2)

[a] The various parameters are defined in the captions of Table 3.

Table 5. Partial-Protein Model Results for Proteinase [128−137] (Loop 1)[a]

N_w	R_{temp}	water restraint	R_{cap} (or R_{nla})	RMSD(BB)	RMSD(SC)	σ(BB)	σ(SC)	σ^w(BB)	σ^w(SC)
300	13	SPH	20	0.74 (3)	2.19 (12)	0.12 (1)	0.37 (11)	0.09 (1)	0.17 (4)
200	13	SPH	19	0.73 (2	2.16 (19)	0.12 (0)	0.40 (7)	0.10 (1)	0.21 (5)
120	13	SPH	18	0.66 (2)	2.31 (14)	0.12 (2)	0.30 (7)	0.10 (1)	0.14 (3)
120	15	SPH	18	0.71 (5)	2.25 (13)	0.12 (1)	0.18 (3)	0.10 (1)	0.12 (2)
70	11	SPH	17	0.72 (2)	2.19 (2)	0.10 (0)	0.22 (3)	0.09 (0)	0.11 (2)
70	11	NLA	7	0.70 (3)	2.29 (13)	0.09 (1)	0.27 (5)	0.08 (0)	0.13 (2)
70	13	SPH	17	0.67 (3)	2.41 (8)	0.11 (1)	0.12 (3)	0.09 (1)	0.08 (1)
70	13	NLA	7	0.67 (3)	2.33 (6)	0.12 (1)	0.20 (3)	0.10 (1)	0.11 (2)
70	15	SPH	17	0.74 (3)	2.18 (12)	0.08 (1)	0.12 (4)	0.07 (1)	0.08 (1)
70	15	NLA	7	0.72 (4)	2.21 (11)	0.10 (1)	0.15 (5)	0.08 (1)	0.10 (3)
50	13	SPH	17	0.66 (2)	2.41 (3)	0.10 (1)	0.13 (1)	0.09 (1)	0.08 (1)
50	13	NLA	7	0.63 (1)	2.43 (9)	0.10 (1)	0.12 (6)	0.09 (1)	0.07 (1)
40	13	SPH	16	0.68 (4)	2.34 (16)	0.09 (1)	0.11 (3)	0.08 (1)	0.08 (1)
30	13	SPH	16	0.70 (4)	2.43 (4)	0.09 (2)	0.11 (2)	0.07 (1)	0.07 (0)
20	13	SPH	16	0.72 (5)	2.46 (6)	0.21 (6)	0.32 (1)	0.15 (1)	0.19 (1)
10	13	SPH	16	0.89 (11)	2.63 (9)	0.11 (4)	0.13 (4)	0.07 (1)	0.08 (1)
5	13	SPH	16	0.84 (17)	2.56 (12)	0.07 (1)	0.11 (6)	0.06 (1)	0.07 (1)
0	13			1.08 (13)	2.65 (24)	0.07 (4)	0.11 (2)	0.05 (1)	0.09 (2)
GBSA	13			0.79 (8)	2.88 (14)	0.18 (2)	0.31 (2)	0.16 (1)	0.24 (2)

[a] The various parameters are defined in the captions of Table 3.

somewhat less-dependent on the loop-template geometry, as it is able to effect a “glovelike” fit to the loop regardless of the conformation. Again, the choice of R_{nla} should be based roughly on the number of water molecules used in the model (and noting again, however, that acceptable performance can be obtained over a range of reasonable values). For very small N_w’s, we often choose R_{nla}’s to be roughly two water-molecule diameters, plus a little fluctuation room (e.g., 7 Å). For larger N_w’s, R_{nla} is increased somewhat. In general, the restraining volume is typically still large for the given number of water molecules.

II.4. Details of the Partial-Protein Simulations. Above, we described the initial preparation (from PDB coordinates) of the loop-template system, thus resulting in the coordinates $\mathbf{X}_{ref}$. A cluster of N_w water molecules is then added to this system. The center of mass of the water cluster is initially positioned, away from the protein atoms (such that there are no van der Waals overlaps), in the direction of $(\mathbf{x}_{cmb} - \mathbf{x}_{cm})/|(\mathbf{x}_{cmb} - \mathbf{x}_{cm})|$ (i.e., on the “loop side” of the loop-template system). The positions of the water molecules are then energy minimized, keeping all protein atoms fixed at $\mathbf{X}_{ref}$ (and subject to the water restraints described above). Following this minimization, 300 ps of MD simulation is performed to equilibrate the water molecules, keeping the protein atoms fixed at $\mathbf{X}_{ref}$. The first 50 ps is run at 600 K, followed by 50 ps at 450 K. These higher temperatures allow the water molecules to spread out and explore the entire protein surface (within the allowable restraining volume). The remaining 200 ps is run at 300 K.

As mentioned above, the main (production) MD simulations consist of the moveable loop atoms and water molecules (subject to the SPH or NLA restraints), in the presence of the fixed template. Therefore, following the above equilibration, the protein loop atoms are allowed to move (along with the water) and are equilibrated (at 300 K) for 30 ps. The production MD simulations are performed at 300 K and are

Table 6. Partial-Protein Model Results for Proteinase [188−196] (Loop 2)[a]

N_w	R_{temp}	water restraint	R_{cap} (or R_{nla})	RMSD(BB)	RMSD(SC)	σ(BB)	σ(SC)	σ^w(BB)	σ^w(SC)
300	13	SPH	20	0.63 (44)	1.50 (50)	0.12 (1)	0.37 (11)	0.09 (1)	0.17 (4)
200	13	SPH	19	0.49 (3)	1.45 (19)	0.18 (5)	0.43 (11)	0.12 (0)	0.23 (1)
120	13	SPH	18	0.46 (4)	1.42 (18)	0.15 (2)	0.32 (10)	0.11 (1)	0.17 (2)
120	15	SPH	18	0.54 (6)	1.69 (36)	0.16 (1)	0.43 (15)	0.11 (1)	0.17 (2)
120	15	NLA	8.5	0.52 (7)	1.67 (41)	0.16 (2)	0.36 (12)	0.12 (1)	0.19 (4)
70	11	SPH	17	0.45 (9)	1.63 (6)	0.18 (12)	0.24 (6)	0.11 (1)	0.16 (2)
70	11	NLA	7	0.44 (2)	1.57 (10)	0.15 (2)	0.29 (2)	0.11 (1)	0.17 (2)
70	13	SPH	17	0.65 (42)	1.95 (36)	0.19 (8)	0.24 (6)	0.12 (2)	0.15 (2)
70	13	NLA	7	0.52 (5)	1.54 (17)	0.13 (1)	0.25 (7)	0.12 (1)	0.17 (5)
70	15	SPH	17	0.46 (5)	2.25 (8)	0.12 (1)	0.23 (10)	0.10 (0)	0.15 (1)
70	15	NLA	7	0.53 (6)	1.73 (39)	0.18 (8)	0.39 (15)	0.12 (2)	0.18 (4)
50	13	SPH	17	0.45 (4)	1.84 (19)	0.14 (2)	0.25 (5)	0.11 (1)	0.13 (2)
50	13	NLA	7	0.57 (13)	1.46 (20)	0.17 (5)	0.22 (4)	0.11 (1)	0.13 (3)
GBSA	13			1.16 (50)	2.86 (70)	0.32 (7)	0.56 (26)	0.17 (2)	0.26 (3)

[a] The various parameters are defined in the captions of Table 3.

Table 7. Comparison of the Partial-Protein and Full-Protein Model Results[a]

N_w (or N_w^{pbc})	protein model	superpose	RMSD(BB)	RMSD(SC)	σ(BB)	σ(SC)	σ^w(BB)	σ^w(SC)
RNase A [64−71]								
6808	full-protein	yes	0.61 (13)	1.62 (39)	0.18 (8)	0.46 (22)	0.11 (2)	0.22 (2)
300	partial-protein R_{temp} = 15 Å	yes	0.42 (5)	1.03 (5)	0.15 (6)	0.37 (3)	0.11 (0)	0.19 (1)
300	partial-protein R_{temp} = 15 Å	no	0.57 (5)	1.31 (4)	0.19 (6)	0.48 (3)	0.14 (1)	0.28 (1)
Ser-Proteinase [143−151]								
9320	full-protein	yes	0.57 (13)	1.33 (22)	0.15 (7)	0.29 (11)	0.12 (2)	0.17 (2)
300	partial-protein R_{temp} = 13 Å	yes	0.44 (1)	1.12 (3)	0.10 (2)	0.22 (1)	0.09 (1)	0.15 (1)
300	partial-protein R_{temp} = 13 Å	no	0.69 (1)	1.51 (3)	0.14 (1)	0.26 (2)	0.13 (1)	0.20 (1)
Proteinase Loop 1 [128−137]								
12 393	full-protein	yes	1.04 (20)	2.47 (46)	0.24 (9)	0.54 (9)	0.13 (4)	0.22 (6)
300	partial-protein R_{temp} = 13 Å	yes	0.59 (2)	2.00 (12)	0.10 (2)	0.34 (10)	0.08 (1)	0.13 (3)
300	partial-protein R_{temp} = 13 Å	no	0.74 (3)	2.19 (12)	0.12 (1)	0.37 (11)	0.09 (1)	0.17 (4)
Proteinase Loop 2 [188−196]								
12 393	full-protein	yes	0.72 (27)	1.64 (36)	0.17 (6)	0.50 (15)	0.10 (1)	0.24 (6)
300	partial-protein R_{temp} = 13 Å	yes	0.48 (33)	1.29 (39)	0.17 (10)	0.43 (14)	0.09 (1)	0.18 (2)
300	partial-protein R_{temp} = 13 Å	no	0.63 (44)	1.50 (50)	0.12 (1)	0.37 (11)	0.09 (1)	0.17 (4)

[a] N_w and N_w^{pbc} denote the number of water molecules used in the partial- and full-protein models, respectively. Results for the partial-protein model were obtained using the spherical restraining method with a radius parameter of R_{cap} = 20 Å in all cases. The superpose column indicates whether RMSD values were minimized by superposing structures (see text). Other parameters are defined in the caption of Table 3.

run to a length of 5 ns. Five independent 5 ns production runs are carried out for each system investigated.

Other important simulation details are as follows. The velocity form of the Verlet algorithm[61] is used to integrate the equations of motion with a time step of 1 fs. The RATTLE[62] algorithm is used to fix all bonds involving hydrogen atoms in the loop and to maintain the rigid geometry of the TIP3P water molecules. The temperature is maintained using a Berendsen thermostat[63] (weak coupling method) with a time constant of 0.1 ps. No distance-based cutoffs are applied to the nonbonded [Lennard-Jones (LJ) and Coulombic] interactions.

As mentioned in the Introduction, the explicit water partial-protein results are compared with results obtained from MD calculations carried out with the GBSA implicit solvation model of Still and co-workers,[35] as implemented within TINKER (using the same simulation parameters described above).

II.5. The Full-Protein Model and Simulations. Starting with the PDB coordinates (with added hydrogens and disulfide bonds), the entire protein was solvated in a rectangular box, giving a 10 Å (11 Å for ser-proteinase) buffer distance to each wall of the box, as implemented in LEaP. All of the crystallographic waters for ser-proteinase, and some of the waters for proteinase (the interior waters), were kept from the PDB files. Counterions (Na^+ or Cl^-) were added to make the overall system charge neutral. The resulting numbers of water molecules are given for each protein in Table 1 (denoted N_w^{pbc}).

To eliminate any bad contacts/strains, the entire system is energy minimized with harmonic positional restraints (k = 100 kcal mol^{-1} Å^{-2}) applied to all protein atoms. This is followed by a second minimization under weaker positional restraints (k = 10 kcal mol^{-1} Å^{-2}). The coordinates resulting from these minimizations are used as a starting point for the MD simulations. They are also taken as the "X-ray reference

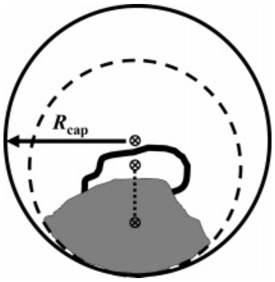

Figure 2. Two-dimensional diagram of the spherical water restraining region (the "SPH restraint"). The loop is represented as the heavy black curve, and the protein template is the region shown in gray. The dashed circle (radius = $R_{\rm temp}$), defining the edge of the template, is the same as that in Figure 1 and is shown here for convenience. Three positions are marked with the symbol ⊗ in the figure. These are, starting from the bottom, $\mathbf{x}_{\rm cm}$, $\mathbf{x}_{\rm cmb}$, and $\mathbf{x}_{\rm sph}$. $\mathbf{x}_{\rm cm}$ is the center of mass of all of the protein atoms considered explicitly in the model (the loop and template atoms), while $\mathbf{x}_{\rm cmb}$ (also shown in Figure 1) is the center of mass of the loop backbone. $\mathbf{x}_{\rm cm}$ and $\mathbf{x}_{\rm cmb}$ are connected by a dotted line, which defines the vector direction (pointing from $\mathbf{x}_{\rm cm}$ to $\mathbf{x}_{\rm cmb}$) that is used to determine the position of $\mathbf{x}_{\rm sph}$. (That is, $\mathbf{x}_{\rm sph}$ is shifted away from the template, see eq 1.) Water molecules are contained within a spherical region defined by the distance $R_{\rm cap}$ measured from $\mathbf{x}_{\rm sph}$. This containment region is represented by the large outer circle. Note that, generally, $R_{\rm cap} > R_{\rm temp}$, and therefore, the edge of this circle (sphere in 3D) is shifted to meet the (bottom) edge of the template so as to keep the majority of the water molecules on the "loop side" of the model system.

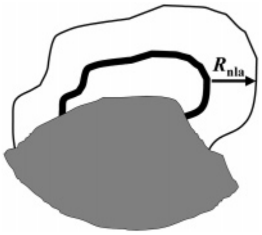

Figure 3. A two-dimensional diagram of the nearest-loop-atom-based restraining region (the "NLA restraint"). The loop is represented as the heavy black curve, and the protein template is the region shown in gray. Water molecules experience a restoring force only when the distance to the *nearest* loop atom becomes greater than a value, $R_{\rm nla}$. For this reason, the boundary of the surrounding containment region mimics the shape of the loop itself, as shown in the figure. Note that the loop side-chain atoms are also considered (as nearest atoms) in the implementation.

coordinates", used in the RMSD calculations. Several stages of MD equilibration are performed in addition to the production runs. All MD simulations are carried out under periodic boundary conditions, with a bath temperature of 300 K. Most of these simulations are also run under constant pressure (p) conditions, where p in all of these cases is set to 1 atm.

In the first stage of equilibration, the system is simulated for 10 ps at constant volume with the protein atoms under positional restraints ($k = 10$ kcal mol^{-1} Å^{-2}). The next stage consists of 40 ps of constant pressure simulation, again with the protein atoms under positional restraints ($k = 10$ kcal mol^{-1} Å^{-2}). This is followed by another 40 ps of constant pressure simulation under weaker positional restraints ($k = 2$ kcal mol^{-1} Å^{-2}). In the final equilibration stage, the positional restraints are removed and the system is again simulated at constant pressure for 40 ps. The production MD simulations (constant T and p) are run to a length of 2 ns. Five independent 2 ns production runs are carried out for each protein.

Other important simulation details are as follows. The leapfrog form of the Verlet algorithm is used to integrate the equations of motion with a time step of 2 fs. The SHAKE algorithm[64] is used to fix all bonds involving hydrogen atoms in the protein and to maintain the rigid geometry of the TIP3P water molecules. Berendsen coupling methods[63] are applied to maintain constant temperature and pressure, both with time constants of 1 ps. Coulombic interactions are modeled using particle mesh Ewald electrostatics[65] with a real space cutoff of 8 Å. (LJ interactions are also cutoff at 8 Å, with a long-range correction added to the energy and pressure.)

II.6. Calculation of RMSD Values. An important gauge of behavior in this investigation is the RMSD of the loop atoms, measured with respect to the X-ray reference coordinates ($\mathbf{X}_{\rm ref}$). We report two RMSD measures: the RMSD of the loop backbone atoms [which is denoted as RMSD(BB)] and the RMSD of the loop side-chain atoms [denoted as RMSD(SC)]. (Corresponding RMSD fluctuations, σ(BB) and σ(SC), will also be reported.) In all cases, only the heavy atoms are considered.

The methods used to calculate RMSD in the partial-protein and full-protein models are somewhat different. Because of the fixed template, RMSD values for the partial-protein model can be straightforwardly calculated in a fixed coordinate system. That is, given a coordinate set $\mathbf{X}_i$ for any structure i (sampled in the production runs), the (squared) distances of the loop atoms in $\mathbf{X}_i$ are simply measured from their positions in $\mathbf{X}_{\rm ref}$. (There is no superposing of structures.) In the case of the full-protein model, the value taken is the minimized RMSD resulting from superposing $\mathbf{X}_i$ and $\mathbf{X}_{\rm ref}$. Here, these superpositions are based on minimizing the RMSD of just the loop atoms and not the entire protein coordinate set. More specifically, for RMSD(BB), only the backbone atoms are superposed, and for RMSD(SC), only the side-chain atoms are superposed.

The quantities defined in this section can be applied for either the backbone or the side-chain atoms (or all of the loop atoms, etc.). Therefore, we will temporarily drop the "(BB)" and "(SC)" for compactness in the equations. In the discussion of the results, however, we will typically refer to the specific quantities (defined in eqs 2−6) by including this more detailed (BB) or (SC) notation.

We calculate RMSD$_i$ values as averages for the entire run (trajectory) i. Thus, for a single configuration $\mathbf{X}_t$, we have the "instantaneous" value, RMSD^t, which is superscripted with t for clarity, and the average value is therefore

$$\mathrm{RMSD}_i = \frac{1}{n}\sum_{t=1}^{n}\mathrm{RMSD}^t \qquad (2)$$

where n is the total number of configurations (snapshots) collected (evenly in time) over the course of the MD trajectory i. For each system, there are five independent runs, and in the tables, we provide values for the five-run average that, for simplicity, are denoted just RMSD. The standard deviation of the RMSD_i values (eq 2) for the five runs is also reported in the tables (in parentheses). These standard deviations can be helpful because, at times, there can be considerable variability in the results for individual runs (i). This, for example, can be due to changes in conformational microstates, which occur on time scales that are too long to be exhibited in all runs.

Even if the average RMSD_i is small for a given model, desirable behavior should also be manifested in the correct fluctuation properties. Therefore, the fluctuations in the instantaneous RMSD^t values (about the average RMSD_i) are also a useful property and are calculated (for run i) as follows:

$$\sigma_i = \left[\frac{1}{n}\sum_{t=1}^{n}(\text{RMSD}^t - \text{RMSD}_i)^2\right]^{1/2} \tag{3}$$

σ_i is (among other things) a reflection of the local motion of the system. If the system remains in a single conformational microstate, σ_i will converge to a well-defined value (as the loop atoms simply execute local motion confined within that microstrate). If, on the other hand, the system moves to another microstate (e.g., a major torsional change in the loop backbone), there will be a significant jump in the RMSD^t values as they will now tend to oscillate about a new average value. The fluctuations within the new microstate may not be that different from the previous one. However, σ_i calculated according to eq 3 will be shifted significantly, because of the large overall spread of RMSD^t values when both microstates are included.

Given the above points, it is helpful to also calculate fluctuations by window averaging. This is done by first defining

$$\sigma_{m(j)i} = \left[\frac{1}{m}\sum_{t=j}^{j+m-1}(\text{RMSD}^t - \text{RMSD}_{m(j)i})^2\right]^{1/2} \tag{4}$$

where

$$\text{RMSD}_{m(j)i} = \frac{1}{m}\sum_{t=j}^{j+m-1}\text{RMSD}^t \tag{5}$$

$\sigma_{m(j)i}$ is a value for the short-time-averaged RMSD fluctuations, where $m < n$. The average is taken over the jth window, consisting of m consecutive snapshots (configurations) recorded during the MD run. (There are $n - m + 1$ such windows.) All possible (contiguous) m-step windows are then averaged to give

$$\sigma_i^{\text{w}} = \frac{1}{(n - m + 1)}\sum_{j=1}^{n-m+1}\sigma_{m(j)i} \tag{6}$$

thus defining the "window-averaged RMSD fluctuations", σ_i^{w}. In this work, σ_i^{w} is calculated using time windows of 200 ps (i.e., the m steps cover a period of 200 ps).

We will report both fluctuation definitions. σ_i^{w} has the property of "smoothing over" the fluctuation effects of moving into (or perhaps flipping between) different microstates and, thus, more faithfully characterizes local motions. The gross changes resulting from different microstates (if they occur) will show up more strongly in σ_i^{w} and will also be reflected in the average RMSD. As for the reported RMSD values, the fluctuations will be averaged over five runs, and the standard deviations over the five runs appear in the tables in parentheses.

III. Results and Discussion

III.1. Solvation Properties of the Partial-Protein Model. Before discussing RMSD results for the individual loops, it is important to discuss some of the general aspects of solvation that we have observed within the partial-protein model. The number of water molecules can be very small, and it is thus helpful to note some of the differences in behavior compared to when larger numbers of water molecules are used. In the partial-protein model, water molecules experience two obviously different environmental influences compared to those in a bulk water environment. Most importantly is the contact/interaction with protein surface atoms. Furthermore, there is also the inevitable exposure to a vacuum due to the modest number of water molecules employed, coupled with the chosen boundary conditions (i.e., nonperiodic boundaries).

III.1.1. Analysis of Surface Coverage. One of the important general characteristics of the present partial-protein solvation model is that there is typically plenty of "extra room" for the water molecules within the allotted restraining region. We mentioned in section II.3 that parameters for both the SPH and NLA restraint methods have been chosen such that the restraining volume is somewhat large for the given N_{w}. This is further evidenced in our simulations by the fact that, at any instant, very few water molecules are experiencing a boundary restoring force (and by small values for the average restraining potential, in general). This is especially true for small N_{w} values, where the water molecules will typically migrate to charged and polar groups on the loop and nearby template, often leaving nonpolar regions bare (as described in ref 59). It is reasonable to assume that the screening/bridging of interactions with charged and polar groups is one of the most important solvating effects provided by the water molecules. It is thus expected that it is better to allow the water molecules to spread out (within reason) such that they can access the more strongly interacting protein atoms, rather than attempting to confine them to a much smaller volume in an effort, for example, to keep them at a density that is closer to the bulk density for water.

A good way to gain a sense for the behavior of the water molecules within these models is to identify those molecules that are considered to belong to the surface region of the protein, separately from those that reside farther away from the protein. Specifically, we choose to define a "protein surface water" as one whose center of mass is a distance of 3.3 Å or less from any protein atom. The total number of these surface waters found (at any given instant) is denoted as N_{surf}. One particularly insightful way to analyze the nature of these models is, thus, to monitor the number of surface water molecules (N_{surf}), or the fraction of protein surface water ($N_{\mathrm{surf}}/N_{\mathrm{w}}$), as the total number of water molecules (N_{w}) is varied. In Table 1, we provide some values for $\langle N_{\mathrm{surf}}\rangle$ and $\langle N_{\mathrm{surf}}/N_{\mathrm{w}}\rangle$ accumulated from simulations of the loop [64–71] of RNase A, modeled under the SPH solvent restraint. (Details of the behavior of the loop itself are deferred until section III.3.) It is seen that, when N_{w} is very small, nearly all of the molecules are directly on the surface of the protein. For example, $\langle N_{\mathrm{surf}}/N_{\mathrm{w}}\rangle$ is nearly 90% when $N_{\mathrm{w}} = 50$. It is not until $N_{\mathrm{w}} = 200$ that this ratio reaches 50%, thus corresponding (on average) to a situation where the protein surface waters are surrounded by a second outer layer of water. At $N_{\mathrm{w}} = 300$, roughly two-thirds of the water molecules are outside the inner hydrating layer. It is also important to note the trends in $\langle N_{\mathrm{surf}}\rangle$ itself, where it is seen that the protein surface appears to saturate with about 100 water molecules at $N_{\mathrm{w}} = 200$ (i.e., $\langle N_{\mathrm{surf}}\rangle$ remains at about 100 for $N_{\mathrm{w}} = 300$). This also implies, conversely, that even for $N_{\mathrm{w}} = 120$ (with $\langle N_{\mathrm{surf}}\rangle = 79$), significant bare regions remain on the protein surface.

III.1.2. Diffusion Properties and Residence Times. It is interesting to address some of the dynamical aspects of the solvation and to examine, in particular, how these properties are affected as N_{w} is increased within the partial-protein model. To do this, we have calculated diffusion constants for the water molecules using the Einstein relation $\langle r^2\rangle = 6Dt$, where D is the diffusion constant and $\langle r^2\rangle$ is the average squared distance that a particle will move in time t. Because the model is a finite system, the ratio $\langle r^2\rangle/t$ will go to zero at long times. Therefore, we estimate D from $\langle r^2\rangle$ values after a period of 10 ps (i.e., we take $D = \langle r^2\rangle/6t$ at $t = 10$ ps). This is a compromise between the effect of ballistic (nondiffusive) motion at very short times (less than 1 ps) and the onset of nonlinearity in $\langle r^2\rangle$ versus t, which is observed as $\langle r^2\rangle$ begins to approach the size of the system (at $t > \sim 30$ ps).

In Table 1, we show diffusion constants for the partial-protein model as the number of water molecules is increased. D_{all} is the value of D that is calculated using all N_{w} molecules. D_{surf}, on the other hand, is the diffusion constant calculated for just the protein surface waters. (Specifically, a molecule is included in the calculation of D_{surf} if it is a distance of 3.3 Å or less from any protein atom at the *beginning* of the 10 ps interval.) It is seen that the value of D_{all} systematically decreases as N_{w} decreases. Part of the reason for this is the high fraction of surface waters exhibited in the models with small N_{w} values (e.g., $N_{\mathrm{w}} = 50$ or 70). An important observation from other simulation studies of protein hydration[66−72] is that water molecules on the surface of the protein diffuse significantly more slowly than water molecules in the bulk. These studies have shown that D for protein surface water is lower (than the bulk value) by about a factor of 2 or more (see, for example, refs 66 and 67). In our calculations, the value of D_{all} at $N_{\mathrm{w}} = 300$ (4.96×10^{-5} cm^2/s) is approaching values that are typical of bulk TIP3P water. (Commonly calculated values at $T = 300$ K and $p = 1$ atm are about 5×10^{-5} cm^2/s but can vary depending on modeling details.[73]) In contrast, the value for D_{surf} (2.61×10^{-5} cm^2/s) is much lower (by about a factor of 2), and thus, it is in good agreement with the findings of the previous studies. The value of D_{surf} for $N_{\mathrm{w}} = 200$ (2.53×10^{-5} cm^2/s) is nearly the same as the value at 300, suggesting that the protein surface waters behave quite similarly in both models. This is despite the differences in D_{all}, which are thus mostly attributable to the difference in the relative amount of surface molecules ($\langle N_{\mathrm{surf}}/N_{\mathrm{w}}\rangle$).

Though there is good agreement for the cases of $N_{\mathrm{w}} = 200$ and 300, it is important to note, however, that D_{surf} becomes significantly lower as N_{w} is decreased further. Though there is still similarity in D_{surf} for the case of $N_{\mathrm{w}} = 120$, D_{surf} at 50 and 70 molecules, however, is roughly half the value of that at 200 or 300. The important distinction in these models ($N_{\mathrm{w}} = 50$ and 70) is that the water molecules generally lack neighboring water from a second layer (the $\langle N_{\mathrm{surf}}/N_{\mathrm{w}}\rangle$ values are 0.888 and 0.825). In view of the general observation of a lowered D for water in the first solvation shell of a protein, it is thus noted that the lack of a second solvation shell serves to lower D further. It is interesting to note, on the other hand, that despite the significantly lower D values for the case of $N_{\mathrm{w}} = 50$ and 70, the stability of the loops (discussed in the next sections) can often be surprisingly good at these very low hydration levels.

Inherent in the diffusion properties is information on the time scales of solvation. Specifically, these values can provide insight on residence times (τ) for water molecules near the surface of the protein. In earlier experimental (NMR) work,[74] an upper bound for residence times of protein surface water was placed at around 500 ps. In much better agreement with the simulation literature, more recent experimental work[75] has placed typical residence times roughly around 25 ps. Residence times have been investigated in simulation studies on a variety of solvated proteins such as BPTI,[68] myoglobin,[69] lysozyme,[70,71] and azurin.[72] Here, we will only briefly make some comparisons. In Brunne et al.,[68] detailed studies were carried out to determine the residence time of hydrating water molecules in specific regions on the protein surface (i.e., near specific types of atoms/groups). They found that the residence time of a surface water molecule is (on average) about 30 ps. (Specific results would vary depending on the nearby protein atoms—backbone atoms, side-chain atoms, charged, polar, nonpolar, etc.)

As a very rough comparison, we can estimate residence times (the time for a water molecule to leave the neighborhood of a solute protein atom) simply from the diffusion results. We take the residence time as the time for a water molecule to diffuse about one and a half molecular diameters, specifically, 4.5 Å (thus, $\tau = (4.5\ \text{Å})^2/6D$). These residence times are given in Table 1, where τ_{surf} is the residence time calculated for a protein surface water and τ_{all} is calculated for all N_{w} water molecules. For the case of $N_{\mathrm{w}} = 200$ or 300, the residence time for surface molecules is about 13 ps, which is in reasonable agreement with the 30 ps given by Brunne et al.,[68] especially when one accounts for the different modeling conditions. The modeling temperature in Brunne et al. was lower (277 K) to mimic NMR experimental conditions. Furthermore, these authors employed the SPC/E water model,[76] which is known to give a lower (more accurate) value for the bulk diffusion coefficient compared to TIP3P. Indeed, calculations in ref 71 using both the SPC/E and TIP3P models showed that the TIP3P residence times were a factor of 2 shorter (roughly 14 ps [TIP3P] as opposed to 27 ps [SPC/E]). (It should also be noted that the diffusion constants and residence times will also vary depending on the distances chosen to define a "protein surface water".)

In correspondence with their lowered D_{surf} values, the residence times for small N_{w} values (50 and 70) are longer. Though these values for τ_{surf} are closer to the values in some of the other studies, they should be interpreted as being "long" for the TIP3P water model (and therefore, they show a specific behavioral property of the partial-protein model at low hydration levels). It is thus expected that they would become much longer if a different water model was used, such as SPC/E or TIP4P,[47] both of which give more accurate diffusion properties.

One of the points discussed by Brunne et al.[68] was their, perhaps unintuitive, observation that the residence times near charged atoms were lower than those for polar, and even nonpolar, atoms. Though we did not carry out the detailed analysis as in that investigation, we did measure the diffusion time away from one specific charged group, the NH_3 group on the Lys side chain of the RNase loop, and for $N_{\mathrm{w}} = 200$ and 300, we also find a decreased residence time (about 10 ps). Interestingly, this effect reverses itself for the case of $N_{\mathrm{w}} = 50$. The value in this case is about 37 ps, which is longer than the average residence time for surface waters at this hydration level (N_{w}). The authors remarked that the shorter residence times for water molecules near charged groups must be related to the effects caused within the surrounding water. Obviously, the lack of outer layers in the case of small N_{w} values might suggest the possibility for different behavior. Here, at low hydration levels, arguments can more plainly be interpreted in terms of energetic benefits because more subtle entropic considerations (associated with surrounding water molecules) are less prevalent.

III.2. Some Properties of the Loops Studied. We now focus on the behavior of the protein loops. The four primary surface loops studied (ranging in size from 8 to 10 amino acid residues), and the related proteins, are presented in Table 2. The 3D structures of these proteins, taken from the PDB, have been determined with 1.5−1.8 Å resolution. The B factors of the loops of RNase A and the two loops 1 and 2 of proteinase are relatively small, where the maximal values obtained for the side chains are 35, 19, and 25, respectively; for ser-proteinase, the B factors of the backbone atoms of five residues range within 20−28, that is, still relatively low, while for some of the side-chain atoms, no significant electron density has been observed. It should be pointed out that side chains with a well-defined structure in the crystal environment (i.e., small B factors) might still be flexible in solution, the environment that is expected to better be described by our models.

While our tests require loops with well-defined structures, it is also imperative to verify that these loops are not stretched, as a stretched loop is insensitive to the model applied. Therefore, we present in Table 2 the ratio R = length of the stretched loop/distance between its ends, which is calculated between the C^{α} atoms of the first and last residues of the loop. The length (in Å) of the extended structure is obtained using the expressions $6.046(n_{\mathrm{res}}/2 - 1) + 3.46$ and $6.046(n_{\mathrm{res}} - 1)/2$ for an even and odd number of residues, n_{res}, respectively; the factors 6.046 and 3.46 Å are taken from Flory's book[77] (Chapter VII, p 251). To a large extent, R reflects the conformational freedom of the loop's backbone and, to a lesser extent, also that of the side chains; the larger R is, the greater the flexibility; indeed, the R values of the four loops are relatively large, ranging from 3.2 to 4.9. Notice, however, that the conformational freedom depends also on the structure of the surrounding protein template and the template−loop interactions. Typically, surface loops are hydrophilic and often charged; therefore, our chosen loops are predominantly polar, where those of RNase A and ser-proteinase each contain one charged residue (bold-faced in Table 2) and loops 1 and 2 of proteinase have three and two charged residues, respectively.

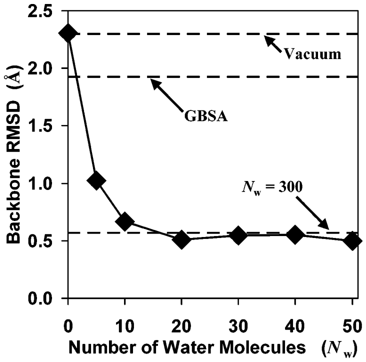

Figure 4. Plot of the average backbone RMSD [RMSD(BB)] as a function of the number of water molecules, N_{w}, for the loop [64−71] of RNase A. The dashed lines indicate the RMSD(BB) values obtained for 300 water molecules, the GBSA implicit solvation model, and simulation in vacuum.

III.3. The Loop of RNase A. We discuss, first, the partial-protein model results for the loop [64−71] of RNase A. Figure 4 is a "convergence plot" of (the backbone average) RMSD(BB) as the number of water molecules, N_{w}, is increased from 0 (vacuum) to 50. (All points are for the case of the SPH solvent restraint method and $R_{\mathrm{temp}} = 15$ Å.) Also marked in the figure is RMSD(BB) for $N_{\mathrm{w}} = 300$ (the largest N_{w} studied for the partial-protein model), as well as the result for GBSA. Though RMSD(BB) = 2.30 Å for the vacuum simulations is large (which is not unexpected), the figure suggests that only a handful of water molecules is necessary to stabilize it. RMSD(BB) is quite low for as little as $N_{\mathrm{w}} = 20$ (0.51 Å), and it is, furthermore, in excellent agreement (converged) with all larger values of N_{w}.

More extensive results, covering a wider range of modeling conditions, are presented in Table 3. Here, RMSD(BB) ranges between 0.51 and 0.67 Å for all N_{w} values between 20 and 300, further suggesting that the backbone behavior is reasonably reproducible and, thus, insensitive to increased levels of hydration. These (backbone) results appear, as well, to be relatively insensitive to the number of environmental protein atoms incorporated into the model (i.e., the template size R_{temp}) and the water containment method (SPH or NLA) and its associated restraining distances (R_{cap} or R_{nla}) (within the ranges tested). We note briefly that (for $N_w \geq 20$) the ranges of the average backbone fluctuations (over five runs), σ(BB), are small, 0.14−0.24 Å (0.19 Å for $N_w = 300$); the range of the corresponding σ^w(BB) is small as well, 0.13−0.15 Å (0.14 Å for $N_w = 300$). The above discussion suggests that, as far as the backbone is concerned, already, $N_w = 50$ (or less) is adequate.

It should be pointed out, however, that the standard deviations, for some of the runs in Table 3, are relatively high. This is due to individual runs that sample ("escape to") different conformational microstates. These transitions are manifested by a significant change in one or more of the (backbone) torsion angles (typically 90° or more). The large free-energy barriers associated with these transitions make the time scale (~1 ns or more) too long to straightforwardly sample/average over various possible conformations (microstates) within typical MD simulation runs. This is a common (and unavoidable) difficulty in testing and assessing potentially flexible regions in protein models. (For example, these "escapes" were also exhibited in the fully solvated, full-protein model.)

We take, as an example of this behavior, the set of trajectories for $N_w = 300$. Here, four runs fell within the range $0.53 \leq \text{RMSD}_i(\text{BB}) \leq 0.55$ Å. However, for one run, the (cumulative average) RMSD$_i$(BB) (eq 2) grows systematically as a function of time for $t > 3$ ns, becoming 0.66 Å for 5 ns, and would have increased further if the simulation had been continued (tending toward about 1 Å), meaning that the loop had transferred to a different microstate. Similar deviant runs were observed for $N_w = 120$, in sets 5, 7, and 8, and in sets 12 ($N_w = 70$) and 16 ($N_w = 30$) (numbering rows [sets] from the top of Table 3). It should be noted, however, that, for $N_w \geq 50$, 73 runs (out of 80 total, i.e., 91%) lead to very low RMSD$_i$(BB) values within 0.48−0.60 Å, with the seven most deviant runs still only averaging to about 0.85 Å. The number of "escaped" runs does not seem to depend on N_w, as one escaped run is found for $N_w = 50$, one for $N_w = 70$, four for $N_w = 120$, and one for $N_w = 300$.

Moving to the side-chain properties, we note, in general, that the side-chain RMSD(SC) values are relatively small. The values range from 1 to 1.4 Å (with 1.31 Å for $N_w = 300$), and thus, the difference between most runs is also relatively small (about 0.2 Å). These RMSD(SC) values are lower, for example, than the values obtained for other loops but larger, of course, compared to the backbone values. The side chains seem to show a dependence on the template size and slightly on N_w. For $R_{temp} = 15$ Å, RMSD(SC) decreases slightly from 1.31 Å (with a very small standard deviation) for $N_w = 300$ to 1.13 Å for $N_w = 200$, to 1.02 and 1.09 Å for $N_w = 120$, and to 1.02 and 1.00 Å for $N_w = 70$. On the other hand, for $R_{temp} = 14$ and 16 Å, the RMSD(SC) values are larger. In general, the corresponding average side-chain fluctuations, σ(SC) and σ^w(SC), tend to decrease as N_w decreases. (See also the discussion for proteinase, loop 1.) Also, on average, σ(SC) and σ^w(SC) for $N_w = 50$ and 70 appear to give somewhat better agreement with larger N_w values when the NLA solvent restraint is used.

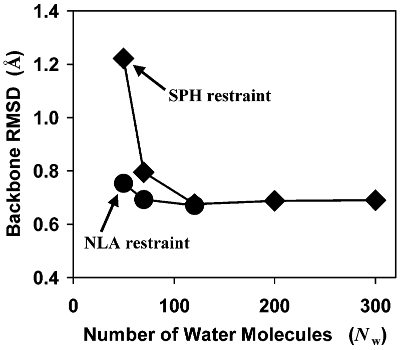

Figure 5. Plot of the average backbone RMSD [RMSD(BB)] as a function of the number of water molecules, N_w, for the loop [143−151] of ser-proteinase. The diamonds mark values obtained using the spherical water restraining method (marked "SPH restraint" in the figure). The circles are for values obtained using the nearest-loop-atom-based restraint (marked "NLA restraint").

The GBSA results, RMSD(BB) = 1.93 Å and RMSD(SC) = 2.71 Å, are significantly larger than those based on explicit water and are not much better than the vacuum results (see also Figure 4). It should be pointed out that, in two of the GBSA runs, RMSD$_i$(BB) and RMSD$_i$(SC) are still increasing significantly after 5 ns.

III.4. The Loop of Ser-proteinase. The results for the loop [143−151] of ser-proteinase are provided in Table 4. The RMSD(BB) values for $N_w = 300$, 200, and 120 are very similar, ranging from 0.64 to 0.69 Å with very small standard deviations (≤ 0.03 Å) for each set of five runs. The corresponding RMSD(SC) values are only slightly more dispersed and can still be considered as very close, ranging from 1.39 to 1.54 Å with a maximal standard deviation (over the five runs) of 0.11 Å. The average backbone fluctuations, σ(BB), are again very close, ranging from 0.12 to 0.14 Å, and the same applies to the average side-chain fluctuations, σ(SC), that vary between 0.26 and 0.29 Å; the corresponding ranges for σ^w(BB) and σ^w(SC) are again narrow, 0.11−0.13 and 0.19−0.20 Å. These results, which were calculated for different templates ($R_{temp} = 12$−15 Å), and with both solvent restraint methods (SPH and NLA), suggest that, already, $N_w = 120$ is sufficient to produce the results of full solvation.

Achieving adequate solvation for this loop becomes more problematic for $N_w < 120$, and it can, furthermore, depend on the modeling conditions. This is clearly shown in Figure 5, a convergence plot of RMSD(BB) as a function of N_w (all for the case of $R_{temp} = 13$ Å). While the points show convergence for $N_w = 120$−300 (as discussed above), the results for $N_w = 50$ and 70 using the SPH solvent cap clearly begin to diverge. Interestingly, the NLA restraint appears to maintain adequate solvation to lower N_w values. The contrast of the two solvent restraint methods at $N_w = 50$ is fairly significant. For the SPH restraint ($R_{cap} = 17$ Å), the results RMSD(BB) = 1.28 Å and RMSD(SC) = 1.88 Å (Table 4) are significantly larger than the 0.69 and 1.51 Å obtained, respectively, for $N_w = 300$. While, on the other hand, the NLA restraint at $N_w = 50$ ($R_{nla} = 7$ Å) is much closer, with RMSD(BB) = 0.75 and RMSD(SC) = 1.55 Å; only σ(BB) = 0.21 and σ(SC) = 0.32 Å (for NLA) are larger than the

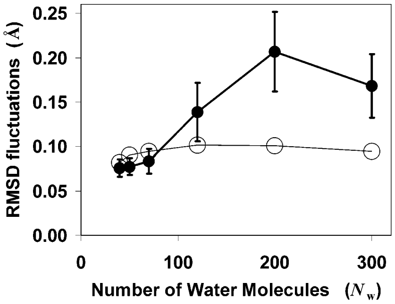

Figure 6. Plot of the window-averaged RMSD fluctuations for backbone and side-chain atoms [σ^w(BB) and σ^w(SC), respectively] as a function of the number of water molecules, N_w, for loop 1 of proteinase [128−137]. The values obtained for the side-chain RMSD fluctuations appear as solid circles and include error bars (the standard deviation of five trials). The backbone RMSD fluctuations appear as large open circles with a lighter trend line.

0.14 and 0.26 Å respectively obtained for $N_w = 300$. (The window-averaged fluctuations are fairly close, however, with σ^w(BB) = 0.15 and σ^w(SC) = 0.19 for (NLA) $N_w = 50$, and 0.13 and 0.20 Å, respectively, for $N_w = 300$.)

It should be pointed out that the GBSA results for RMSD(BB) and RMSD(SC) are actually equal to the corresponding $N_w = 300$ values, while the results σ(BB) = 0.18, σ(SC) = 0.31, σ^w(BB) = 0.16, and σ^w(SC) = 0.24 Å are slightly larger than their counterparts for $N_w = 300$.

III.5. Loop 1 of Proteinase. The results for loop 1 of proteinase [128−137] are summarized in Table 5. The table reveals that the backbone of this loop is very stable, where RMSD(BB) ~ 0.70 Å (with the standard deviation smaller than 0.05) already for $N_w \geq 20$. For $N_w = 10$, 5, and 0, RMSD(BB) increases to 0.89, 0.84, and 1.08 Å with relatively large standard deviations (of the five runs), 0.11, 0.17, and 0.13 Å, with maximal values of 1.01, 1.13, and 1.16 Å, respectively, where the first two maximal values have not been converged after 5 ns and are growing. The results for σ(BB) are small and similar for most N_w values: 0.12 Å for $N_w \geq 120$, 0.10 Å (on average) for $N_w = 70$ and 50, and 0.09−0.11 Å for $10 \leq N_w \leq 40$. Similar behavior is observed for σ^w(BB).

The RMSD(SC) values for this loop are significantly larger than those for the loop of ser-proteinase; that is, the side chains have moved significantly from their X-ray structure. For $N_w \geq 120$, RMSD(SC) ranges from 2.16 to 2.31 Å; for $N_w = 70$, the range is similar except in one case where RMSD(SC) = 2.41 Å is slightly larger. As N_w decreases further, RMSD(SC) increases moderately, becoming 2.65 Å for $N_w = 0$.

Though RMSD(SC) appears to be relatively converged at small N_w values, the side-chain fluctuations show a significant increase as N_w is increased. These trends are shown in Figure 6, which is a plot of the window-averaged side-chain fluctuations, σ^w(SC), as a function of N_w (all for the case of the SPH solvent restraint method and $R_{\rm temp}$ = 13 Å). σ^w(SC) is consistently small for all $N_w \leq 70$ compared to the higher solvation levels at $N_w = 200$ or 300. [Note, in contrast, that σ^w(BB), which is also given in the figure, appears to be converged for all N_w values shown.] To more clearly see how these results are manifested in the trajectories, we have plotted, as an example, the instantaneous (snapshot) values of RMSD^t(SC) over the course of a typical 5 ns run for $N_w = 70$ and compare that with a typical run for $N_w = 200$. These plots are shown in Figure 7. {Note that the y-axis scales [for RMSD^t(SC)] are the same in both plots.} Though the RMSD^t(SC) values for these two runs are similar, on average, the oscillations in these values (even over short times) show very different amplitudes. That is, the $N_w = 200$ run appears to visit a more diverse array of states, and even within those states, the atomic fluctuations are more broad, meaning higher entropy than in the $N_w = 70$ case.

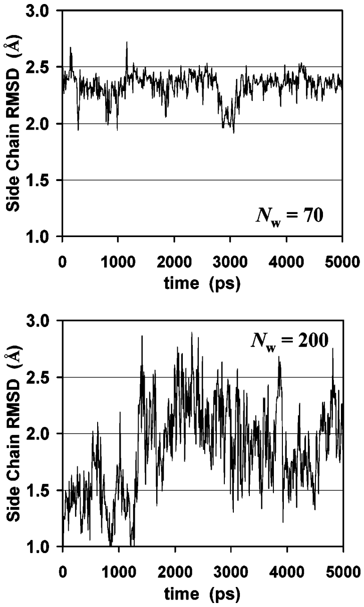

Figure 7. Instantaneous RMSD values of the side-chain atoms [RMSD^t(SC)] as a function of time for loop 1 of proteinase [128−137]. The upper plot is for a typical 5 ns trajectory for the case of $N_w = 70$. The lower plot is a typical trajectory with $N_w = 200$.

Some additional trends in the side-chain fluctuations are as follows. The table shows that the σ(SC) results for $N_w \geq 70$ decrease as the template radius $R_{\rm temp}$ is increased and N_w is decreased. Also, the NLA restraint leads to higher (i.e., better agreement with large N_w) σ(SC) and σ^w(SC) values than the spherical cap (SPH). Thus summarizing, for $R_{\rm temp}$ = 13 Å, we obtained almost the same σ(SC) values, 0.37, 0.40, and 0.30 Å, for $N_w = 300$, 200, and 120, respectively, and a slightly lower value, 0.20 Å, for $N_w = 70$ with a the NLA restraint. The corresponding σ^w(SC) values, 0.17, 0.21, 0.14, and 0.11 Å, are also close. The results for σ(SC) and σ^w(SC) for $N_w \leq 50$ are significantly smaller than the corresponding values for $N_w = 300$. Therefore, $N_w = 120$ (perhaps less with the NLA restraint) is necessary to solvate this loop.

It is noted that the GBSA values, RMSD$_i$(BB) = 0.90 Å and RMSD$_j$(SC) = 3.07 Å (from two different runs), are not converged after 5 ns, but they are in an increasing trend. Thus, the GBSA result, RMSD(BB) = 0.79 Å, is not converged, and the corresponding GBSA result, RMSD(SC) = 2.88 Å, that is already significantly larger (by 0.7 Å) than the 2.19 Å obtained for $N_w = 300$ is not converged either.

III.6. Loop 2 of Proteinase. The results for loop 2 of proteinase [188−196] are summarized in Table 6. The RMSD(BB) results in the table are similar for all of the N_w values. However, it should be pointed out that, for $N_w = 300$, one MD run escaped from the X-ray microstate, leading to RMSD_i(BB) = 1.42 Å, where from 3 to 5 ns RMSD_i(BB) is still increasing. Similar behavior is observed for single runs of the sets of $N_w = 70$ ($R_{\text{temp}} = 11$ Å and $R_{\text{cap}} = 17$ Å), where RMSD_i(BB) = 0.62 Å; $N_w = 70$ ($R_{\text{temp}} = 13$ Å and $R_{\text{cap}} = 17$ Å), where RMSD_i(BB) = 1.40 Å; and $N_w = 50$ ($R_{\text{temp}} = 13$ Å and $R_{\text{nla}} = 7$ Å), where RMSD_i(BB) = 0.73 Å. This suggests that the X-ray microstate of this loop may not be overwhelmingly stable (i.e., competing conformational microstates), as this instability is independent of the number of waters, N_w, occurring for $N_w = 50$ and 70 as well as for $N_w = 300$. Moreover, this behavior was exhibited in the full-protein model as well (see below). When the contribution of the "escaped" runs is omitted, all of the RMSD(BB) results are very close to 0.5 Å, and the fluctuations σ(BB) and σ^w(BB) are close to 0.12 and 0.15 Å, respectively.

The instability of the X-ray microstate is demonstrated even more strongly by the behavior of the side chains. While, for $N_w \geq 120$, the RMSD(SC) values are relatively close, ranging from 1.42 to 1.69 Å, the corresponding standard deviations are large, suggesting that the individual values, RMSD_i(SC), are very different. Indeed, for the two sets of $N_w = 120$ ($R_{\text{temp}} = 15$), the minimum and maximum RMSD_i(SC) values are 1.39 and 2.22 and 1.27 and 2.23 Å. Moreover, in some cases, the RMSD_i(SC) values have not been converged after 5 ns. For example, for $N_w = 300$, one MC run has led to a still unconverged value of RMSD_i(SC) = 2.37 Å, where for both $N_w = 200$ and 120 ($R_{\text{temp}} = 15$ Å and $R_{\text{cap}} = 18$ Å) two unconverged RMSD_i(SC) values occurred. A similar picture is observed for $N_w = 70$ and 50.

Even though this loop is not stable, it is evident that similar results are obtained for $N_w \geq 120$ and, for R_{nla}, also for $N_w = 70$. This is also demonstrated by the results for σ(BB) and σ^w(BB) that are close for these runs, that is, within the ranges 0.18−0.12 and 0.12−0.09 Å, respectively. The ranges of the side-chain fluctuations, σ(SC) and σ^w(SC), are also small, 0.43−0.25 and 0.23−0.17 Å, respectively.

It is of interest to point out that the GBSA results are significantly different from those obtained with explicit water. Thus, not only is RMSD(BB) = 1.16 Å considerably larger than the RMSD(BB) values obtained for explicit water but the standard deviation of the GBSA set is large because of elevated RMSD_i(BB) values within the range 0.67−1.71 Å; the same occurs also for the side chains, where RMSD(SC) = 2.86 Å is significantly larger than the corresponding values obtained for the explicit water, where the RMSD_i(SC) values for GBSA range within 2.31−3.63 Å.

III.7. Partial Study of Four More Loops. While the above study suggests that a relatively small number of waters is sufficient to solvate a loop, one would like to strengthen this conclusion by evidence from a larger number of loops. However, because of the extensive calculations required, we decided to carry out only partial studies of four extra loops, which indeed provide supportive evidence. We first treated the seven-residue loop [244−250] (ITTIYQA) of peptidase (5cpa) with a flexibility ratio, $R = 2.7$. Defining $R_{\text{temp}} = 13$ Å and using the spherical water restraint (SPH) with $N_w = 70$ waters, we obtained the relatively small RMSD(BB) = 0.69 Å, as the average of five MD runs. The second loop is of seven residues [57−63] (**EAKEH**C) of RNase H (2rn2), with a flexibility ratio $R = 1.6$, where again $R_{\text{temp}} = 13$ Å and $N_w = 70$. Here, the SPH restraint led to RMSD(BB) = 1.02 Å, as two deviating MD runs contributed RMSD_i(BB) values of 1.57 and 1.34 Å. However, the NLA restraint, which has been found to perform better for small N_w values, led to RMSD(BB) = 0.72 Å. Therefore, this loop is expected to stabilize with the SPH restraint at $N_w = 120$, similar to the case observed for ser-proteinase.

We also studied a seven-residue loop in porcine amylase (1pif) [304−310] (GHGAGGS) with a flexibility ratio $R = 3.2$ and the same loop in human amylase (1smd), where S is replaced by A and the flexibility ratio is $R = 2.3$. In the pig amylase, we used a template of $R_{\text{temp}} = 15$ Å with an SPH restraint. For $N_w = 70$, only two runs were generated, which led to RMSD(BB) = 0.47 Å, whereas for $N_w = 200$, the five MD runs led to RMSD(BB) = 0.45 Å. For the human amylase, we obtained RMSD(BB) = 0.73 Å using $R_{\text{temp}} = 15$ Å and the NLA restraint with $N_w = 70$ waters.

III.8. Results for the Full-Protein Model. The RMSD results for the full-protein model appear in Table 7 together with the corresponding $N_w = 300$ results obtained for the partial-protein model. However, because the RMSD was calculated differently for the two models, and to make the comparison between them on the same footing, we have recalculated the RMSD of the partial-protein model in the same way as that for the full-protein model (marked as "yes" in the "superpose" column of the table). The table reveals that, for all loops, the RMSD values (and fluctuations) of the full-protein model are always larger than the corresponding results of the partial-protein model. This effect is to be expected, on one hand, because of the nonfixed coordinates (of the nonloop atoms), thus promoting greater flexibility. On the other hand, however, there should be a mild but consistent effect to reduce the RMSDs because of the use of (minimized) superposition. This latter effect appears to reduce the backbone RMSD values by roughly 0.15 Å upon comparison of the superposed and nonsuperposed values for the partial-protein model in Table 7.

Not only are all the averages of the full-protein model larger than those of the partial model, but also the corresponding standard deviations (appearing in parentheses), which should be considered in the comparisons between the averages, are as well. Thus, for ser-proteinase, the values RMSD(BB) = 0.57(13) and 0.44(1) Å are equal within the standard deviations and all runs, on average, span the same microstate, where the most deviant single run for the full-protein model, RMSD_i(BB) = 0.80 Å, leads to the corresponding large standard deviation; this run also contributes to the large fluctuation [σ_i(BB) = 0.28 Å] and its large standard deviation (0.07 Å). A similar picture is seen for the side chains where RMSD(SC) = 1.33(22) and 1.12(3) Å are equal within the standard deviation, where one run contributes most significantly, RMSD_i(SC) = 1.70, σ_i(SC)

= 0.49 Å, and $\sigma_i^{\mathrm{w}}(\mathrm{SC}) = 0.20$ Å. Notice that the differences between the σ^{w}(BB) and σ^{w}(SC) values of the two models are small. In summary, for this loop, ignoring the effect of the run with largest results, both models lead to close results, RMSD(BB) = 0.52 and 0.44 Å, RMSD(SC) = 1.23 and 1.12, σ(BB) = 0.12 and 0.10, and σ(SC) = 0.25 and 0.22 Å.

Quite similar behavior is observed for RNase A, where only the results of RMSD(SC) of the two models differ significantly and are not covered by their standard deviations. Here again, the results for one trajectory *i* deviate significantly from the results of the other runs of the full-protein model, leading to $\mathrm{RMSD}_i(\mathrm{BB}) = 0.83$, $\mathrm{RMSD}_i(\mathrm{SC}) = 2.27$, $\sigma_i(\mathrm{BB}) = 0.32$, and $\sigma_i(\mathrm{SC}) = 0.82$ Å. Ignoring this run, the results for the two models are quite comparable, RMSD(BB) = 0.55 and 0.42 Å, RMSD(SC) = 1.47 and 1.03, σ(BB) = 0.15 and 0.15, σ(SC) = 0.38 and 0.37, σ^{w}(BB) = 0.11 and 0.11, and σ^{w}(SC) = 0.22 and 0.19 Å.

The results for loop 2 of proteinase for both models have relatively large standard deviations, reflecting differences among the results of the five runs. Thus, while the averages of the full-protein model are in most cases larger than those of the partial model, the differences are not large [e.g., 0.7 vs 0.5 Å for RMSD(BB) and 1.6 vs 1.3 Å for RMSD(SC)], where the average values are always covered by the error bars.

For loop 1 of proteinase, the results of the two models show the most disagreement among the four loops, where the error bars in most cases do not cover the average values. However, even in this case, the results are not very different, 1.0 vs 0.6 Å for RMSD(BB) and 2.5 vs 2 Å for RMSD(SC).

IV. Conclusions

We have shown that, for the present loops described in the framework of the partial-protein model, the results, in general, become less dependent on the parameters of the model as the number of waters is increased. Relatively small numbers of water molecules (120 and sometimes less) lead to results for RMSD and its fluctuations that are very similar to those obtained for 300 waters. It is expected that (similarly) ~12 waters per residue will be found adequate for other loops; however, this number should be checked for each individual loop. (We have already noted in the Introduction that Steinbach and Brooks have studied the effect of increasing the number of water molecules on the protein structure; examples of similar convergence studies performed on ions, water, and small molecules appear in refs 78 and 79). We have also found that, for a small number, N_{w}, of waters, the NLA restraint leads to slightly better results than the SPH restraint. The good performance obtained here with a relatively small number of waters is in accord with the free-energy calculations of Beglov and Roux,[60] who (originally) applied the NLA restraint to the alanine dipeptide and tripeptide molecules and have found good agreement with calculations based on bulk solvation. As expected, the RMSD (and fluctuation) values for the full-protein model are somewhat larger than their counterparts for the partial model. Indeed, the differences are not large, and it is not clear whether they stem from using more complete solvation (with particle mesh Ewald) or from modeling the entire protein with unfixed coordinates.

Still, the present partial-protein model can be made more realistic (1) by allowing residues neighboring the loop ends also to move, (2) by relaxing the fixed template atoms, by only restraining them harmonically to their X-ray positions, and (3) by increasing the template size; such changes would make the protein atom treatment in the present model more similar to the stochastic boundary MD approximation.[52] However, while, in principle, the partial-protein model with implicit solvation (such as GBSA) is inferior to that with explicit solvation, the long-range electrostatic interactions of the latter model are still not treated correctly. A more rigorous treatment is provided by sophisticated hybrid models where the region of interest is described by explicit solvent and the effect of the remote region by the reaction field of continuum solvation.[78−85] However, because of the complex and varying geometry of the *actual* outer surface of the protein−water system (e.g., this surface/boundary is not simply the boundary of the SPH or NLA restraining region), most of these techniques would be difficult to apply to the present partial-protein model, especially at small N_{w} values (see discussions in refs 84 and 85).

We intend to use the partial-protein model to study mobile loops that take part in binding processes. As mentioned in the Introduction, in the free protein, such a loop typically resides in an open (o) flexible microstate or it undergoes *intermediate flexibility*, that is, populates several microstates in thermodynamic equilibrium. Upon ligand binding, the loop moves to a structurally different (and less flexible) bound (b) microstate, sometimes creating a "lid" above the active site, thus protecting it from water. Several questions are of interest, for example: (i) Is the process of a selected-fit type? That is, is the microstate of the bound loop already included within those visited by the free protein (or otherwise the process is of an induced-fit type)? (ii) What is the loss in loop entropy in going from the open to the bound microstates, and what is the corresponding free-energy difference? (The backbone entropy can, in some cases, be compared with results obtained from NMR.) To study these problems, one would have to carry out MD simulations that cover both the bound and open microstates; such simulations are expected to become extremely long and, hence, prohibitive with the full-protein model.

However, with the partial-protein model (but not as easily with the full model), one can use replica-exchange or multicanonical techniques to carry out a conformational search more efficiently than with long MD simulations at constant temperature, and differences in free energies can be obtained from the relative duration of the trajectory in the microstates of interest. The feasibility of this approach (for the partial-protein model) is mainly due to the increased exchange acceptance that is concurrent with smaller system sizes. Still, the transition of a loop between microstates by simulations is typically difficult because of high energy barriers; therefore, procedures for calculating the *absolute* free energy are expected to be very effective, because they would lead to $\Delta F = F_{\mathrm{o}} - F_{\mathrm{b}}$ and $\Delta S = S_{\mathrm{o}} - S_{\mathrm{b}}$ by subtracting the values obtained from two separate simulations for the open and bound microstates without the need to "cover" the latter by a long trajectory. One such method, called HSMC or HSMD, was developed by us and has been applied thus far to argon, TIP3P water,[86] self-avoiding walks on a lattice,[87] and peptides,[88,89] and we intend to extend it to the present partial-protein model as well.

Acknowledgment. This work was supported by NIH grants R01 GM66090 and R01 GM61916.

References

(1) Kwasigroch, J. M.; Chomilier, J.; Mornon, J. P. *J. Mol. Biol.* **1996**, *259*, 855.

(2) Martin, A. C.; Toda, K.; Stirk, H. J.; Thornton, J. M. *Protein Eng.* **1995**, *8*, 1093.

(3) Oliva, B.; Bates, P. A.; Querol, E.; Aviles F. X.; Sternberg M. J. *J. Mol. Biol.* **1997**, *266*, 814.

(4) Fetrow, J. S. *FASEB J.* **1995**, *9*, 708.

(5) Getzoff, E. D.; Geysen, H. M.; Rodda, S. J.; Alexander, H.; Tainer, J. A.; Lerner, R. A. *Science* **1987**, *235*, 1191.

(6) Rini, J. M.; Schulze-Gahmen, U.; Wilson, I. A. *Science* **1992**, *255*, 959.

(7) Constantine, K. L.; Friedrichs, M. S.; Wittekind, M.; Jamil, H.; Chu, C. H.; Parker, R. A.; Goldfarb, V.; Mueller, L.; Farmer, B. T. *Biochemistry* **1998**, *37*, 7965.

(8) Bates, P. A.; Sternberg, M. J. *Proteins* **1999**, Suppl 3, 47.

(9) Mosimann, S.; Meleshko, R.; James, M. N. *Proteins* **1995**, *23*, 301.

(10) Sali, A. *Curr. Opin. Biotechnol.* **1995**, *6*, 437.

(11) Petrey, D.; Xiang, Z.; Tang, C. L.; Xie, L.; Gimpepev, M.; Mitros, T.; Soto, C. S.; Goldsmith-Fischman, S.; Kernytsky, A.; Schlessinger, A.; Koh, I. Y. Y.; Alexov, E.; Honig, B. *Proteins* **2003**, *53*, 430.

(12) Crasto, C. J.; Feng, J. *Proteins* **2001**, *42*, 399.

(13) Leszczynski, J. F.; Rose, G. D. *Science* **1986**, *234*, 849.

(14) Donate, L. E.; Rufino, S. D.; Canard, L. H.; Blundell, T. L. *Protein Sci.* **1996**, *5*, 2600.

(15) Fechteler, T.; Dengler, U.; Schomburg, D. *J. Mol. Biol.* **1995**, *253*, 114.

(16) Chothia, C.; Lesk, A. M. *J. Mol. Biol.* **1987**, *196*, 901.

(17) Chothia, C.; Lesk, A. M.; Tramontano, A.; Levitt, M.; Smith-Gill, S. J.; Air, G.; Sheriff, S.; Padlan, E. A.; Davies, D.; Tulip, W. R. *Nature* **1989**, *342*, 877.

(18) Gō, N.; Scheraga, H. A. *Macromolecules* **1970**, *3*, 178.

(19) Dudek, M. J.; Scheraga, H. A. *J. Comput. Chem.* **1990**, *11*, 121.

(20) Bruccoleri, R. E.; Karplus, M. *Biopolymers* **1987**, *26*, 137.

(21) Summers, N. L.; Karplus, M. *J. Mol. Biol.* **1990**, *216*, 991.

(22) Tappura, K. *Proteins* **2001**, *44*, 167.

(23) Moult, J.; James, M. N. *Proteins* **1986**, *1*, 146.

(24) Fine, R. M.; Wang, H.; Shenkin, P. S.; Yarmush, D. L.; Levinthal, C. *Proteins* **1986**, *1*, 342.

(25) Shenkin, P. S.; Yarmush, D. L.; Fine, R. M.; Wang, H. J.; Levinthal, C. *Biopolymers* **1987**, *26*, 2053.

(26) Higo, J.; Collura, V.; Garnier, J. *Biopolymers* **1992**, *32*, 33.

(27) Rosenfeld, R.; Zheng, Q.; Vajda, S.; DeLisi, C. *J. Mol. Biol.* **1993**, *234*, 515.

(28) Zheng, Q.; Rosenfeld, R.; Vajda, S.; DeLisi, C. *J. Comput. Chem.* **1993**, *14*, 556.

(29) Caralacci, L.; Englander, S. W. *J. Comput. Chem.* **1996**, *17*, 1002.

(30) Das, B.; Meirovitch, H. *Proteins* **2001**, *43*, 303.

(31) Das, B.; Meirovitch, H. *Proteins* **2003**, *43*, 470.

(32) Mas, M. T.; Smith, K. C.; Yarmush, D. L.; Aisaka, K.; Fine, R. M. *Proteins* **1992**, *14*, 483.

(33) Smith, K. C.; Honig, B. *Proteins* **1994**, *18*, 119.

(34) Wesson, L.; Eisenberg, D. *Protein Sci.* **1992**, *1*, 227.

(35) Qiu, D.; Shenkin, P. S.; Hollinger, F. P.; Still, W. C. *J. Phys. Chem A* **1997**, *101*, 3005.

(36) Rapp, C. S.; Friesner, R. A. *Proteins* **1999**, *35*, 173.

(37) de Bakker, P. I. W.; DePristo, M. A.; Burke, D. F.; Blundell, T. L. *Proteins* **2003**, *51*, 21.

(38) DePristo, M. A.; de Bakker, P. I. W.; Lovell, S. C.; Blundell, T. L. *Proteins* **2003**, *51*, 41.

(39) Jacobson, M. P.; Pincus, D. L.; Rapp, C. S.; Day, T. J. F.; Honig, B.; Shaw, D. E.; Friesner, R. A. *Proteins* **2004**, *55*, 351.

(40) Ghosh, A.; Rapp, C. S.; Friesner, R. *J. Phys. Chem.* **1998**, *102*, 10983.

(41) Gallicchio, E.; Zhang, L. Y.; Levy, R. M. *J. Comput. Chem.* **2002**, *23*, 517.

(42) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225.

(43) Zhang, C.; Liu, S.; Zhou, Y. *Protein Sci.* **2004**, *13*, 391.

(44) Xiang, X.; Soto, C. S.; Honig, B. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 7432.

(45) Tanner, J. J.; Nell, L. J.; McCammon, J. A. *Biopolymers* **1992**, *32*, 23.

(46) Szarecka, A.; Meirovitch, H. *J. Phys. Chem. B.* **2006**, *110*, 2869.

(47) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.

(48) Alder, B. J.; Wainwright, T. E. *J. Chem. Phys*. **1959**, *31*, 459.

(49) McCammon, J. A.; Gelin, B. R.; Karplus, M. *Nature* **1977**, *267*, 585.

(50) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M., Jr.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179.

(51) Ponder, J. W. *TINKER, Software Tools for Molecular Design*, version 4.2; Washington University: St. Louis, MO, 2004.

(52) Brooks, C. L., III; Brünger, A. T.; Karplus, M. *Biopolymers* **1985**, *24*, 843.

(53) Meirovitch, H.; Hendrickson, T. F. *Proteins* **1997**, *29*, 127.

(54) Bash, P. A.; Singh, U. C.; Brown, F. K.; Langridge, R.; Kollman, P. A. *Science* **1986**, *235*, 574.

(55) Merz, K. M., Jr. *J. Am. Chem. Soc*. **1991**, *113*, 406.

(56) Miyamoto, S.; Kollman, P. A. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 8402.

(57) Essex, J. W.; Severance, D. L.; Tirado-Rives, J.; Jorgensen, W. L. *J. Phys. Chem.* **1997**, *101*, 9663.

(58) Smith, R. H., Jr.; Jorgensen, W. L.; Tirado-Rives, J.; Lamb, M. L.; Janssen, P. A. J.; Michejda, C. J.; Kroeger Smith, M. B. *J. Med. Chem.* **1998**, *41*, 5272.

(59) Steinbach, P. J.; Brooks, B. R. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 9135.

(60) Beglov, D.; Roux, B. *Biopolymers* **1995**, *35*, 171.

(61) Swope, W. C.; Andersen, H. C.; Berens, P. H.; Wilson, K. R. *J. Chem. Phys.* **1982**, *76*, 637.

(62) Andersen, H. C. *J. Comput. Phys.* **1983**, *52*, 24.

(63) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684.

(64) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327.

(65) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089.

(66) Levitt, M.; Sharon, R. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 7557.

(67) Makarov, V. A.; Feig, M.; Andrews, B. K.; Pettitt, B. M. *Biophys. J.* **1998**, *75*, 150.

(68) Brunne, R. M.; Liepinsh, E.; Otting, G.; Wuthrich, K.; van Gunsteren, W. F. *J. Mol. Biol.* **1993**, *231*, 1040.

(69) Makarov, V. A.; Andrews, B. K.; Smith, P. E.; Pettitt, B. M. *Biophys. J.* **2000**, *79*, 2966.

(70) Sterpone, F.; Ceccarelli, M.; Marchi, M. *J. Mol. Biol.* **2001**, *311*, 409.

(71) Marchi, M.; Sterpone, F.; Ceccarelli, M. *J. Am. Chem. Soc.* **2001**, *124*, 6787.

(72) Luise, A.; Falconi, M.; Desideri, A. *Proteins* **2000**, *39*, 56.

(73) Feller, S. C.; Pastor, R. W.; Rojnukarin, A.; Bogusz, S.; Brooks, B. R. *J. Phys. Chem.* **1996**, *100*, 17011.

(74) Otting, G.; Liepinsh, E.; Wuthrich, K. *Science* **1991**, *254*, 974.

(75) Modig, K.; Liepinsh, E.; Otting, G.; Halle, B. *J. Am. Chem. Soc.* **2004**, *126*, 102.

(76) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. *J. Phys. Chem.* **1987**, *91*, 6269.

(77) Flory, P. J. *Statistical Mechanics of Chain Molecules*; Hasner: New York, 1988.

(78) Beglov, D.; Roux, B. *J. Chem. Phys.* **1994**, *100*, 9050.

(79) Sham, Y. Y.; Warshel, A. *J. Chem. Phys*. **1998**, *109*, 7940.

(80) King, G.; Warshel, A. *J. Chem. Phys.* **1989**, *91*, 3647.

(81) Lee, F. S.; Warshel, A. *J. Chem. Phys.* **1992**, *97*, 3100.

(82) Lounnas, V.; Ludemann, S. K.; Wade, R. C. *Biophys. Chem.* **1999**, *78*, 157.

(83) Alper, H.; Levy, R. M. *J. Chem. Phys*. **1993**, *99*, 9847.

(84) Lee, M. S.; Salsbury, F. R., Jr.; Olson, M. A. *J. Comput. Chem.* **2004**, *25*, 1967.

(85) Lee, M. S.; Olson, M. A. *J. Phys. Chem. B* **2005**, *109*, 5223.

(86) White, R. P.; Meirovitch, H. *J. Chem. Phys.* **2004**, *121*, 10889.

(87) White, R. P.; Meirovitch, H. *J. Chem. Phys.* **2005**, *123*, 214908.

(88) Cheluvaraja, S.; Meirovitch, H. *J. Chem. Phys.* **2005**, *122*, 054903.

(89) Cheluvaraja, S.; Meirovitch, H. *J. Phys. Chem. B* **2005**, *109*, 21963.

CT0503217