Protein Structures under Electrospray Conditions

ARTICLE *in* BIOCHEMISTRY · FEBRUARY 2007

Impact Factor: 3.02 · DOI: 10.1021/bi061182y · Source: PubMed

3 AUTHORS, INCLUDING:

Erik Marklund
University of Oxford
17 PUBLICATIONS **387** CITATIONS
SEE PROFILE

David Spoel
Uppsala University
154 PUBLICATIONS **20,255** CITATIONS
SEE PROFILE

BIOCHEMISTRY

 Volume 46, Number 4 January 30, 2007

Articles

Protein Structures under Electrospray Conditions

Alexandra Patriksson, Erik Marklund, and David van der Spoel*

Department of Cell and Molecular Biology, Uppsala University, Husargatan 3, Box 596, SE-751 24 Uppsala, Sweden

Received June 14, 2006; Revised Manuscript Received November 20, 2006

ABSTRACT: During electrospray ionization (ESI), proteins are transferred from solution into vacuum, a process that influences the conformation of the protein. Exactly how much the conformation changes due to the dehydration process, and in what way, is difficult to determine experimentally. The aim of this study is therefore to monitor what happens to protein structures as the surrounding waters gradually evaporate, using computer simulations of the transition of proteins from water to vacuum. Five different proteins have been simulated with water shells of varying thickness, enabling us to mimic the entire dehydration process. We find that all protein structures are affected, at least to some extent, by the transfer but that the major features are preserved. A water shell with a thickness of roughly two molecules is enough to emulate bulk water and to largely maintain the solution phase structure. The conformations obtained in vacuum are quite similar and make up an ensemble which differs from the structure obtained by experimental means, and from the solution phase structure as found in simulations. Dehydration forces the protein to make more intramolecular hydrogen bonds, at the expense of exposing more hydrophobic area (to vacuum). Native hydrogen bonds usually persist in vacuum, yielding an easy route to refolding upon rehydration. The findings presented here are promising for future bio-imaging experiments with X-ray free electron lasers, and they strongly support the validity of mass spectrometry experiments for studies of intra- and intermolecular interactions.

Over the past two decades, mass spectrometry (MS) has proven to be a very valuable tool for studying biomolecular interactions in the gas phase (*1*−*3*). In combination with different fragmentation techniques (*4*, *5*), MS has turned out to be a sensitive analytical method in the field of proteomics as well. A crucial part of all mass spectrometric methods is the injection process, and electrospray ionization (ESI) (*6*−*8*), which was developed in the late 1960s (*1*) and in 2002 rewarded the Nobel Prize in Chemistry (*6*−*8*), is the method used routinely for this purpose. Upon leaving the electrospray, the (bio)molecules are immersed in solvent, but since they are exposed to air and/or vacuum conditions, the solvent will gradually evaporate (*9*) and hence the protein becomes dehydrated. Even though the ionization−dehydration process of the ESI is believed to preserve specific noncovalent interactions that exist in solution, it is not certain how much such changes in environmental conditions affect the conformation of a protein. It is known that unfolding of proteins in vacuum is possible, in particular when the proteins are highly charged (*10*−*12*). On the other hand, it has also been shown that a virus particle can be injected by electrospray into a mass spectrometer and collected at the end and still be infective (*13*), which proves that the particle as a whole is undamaged or at least that any changes inflicted by dehydration are reversible. In addition, it has been shown that very large molecular complexes can be detected after they travel through air and or vacuum (*14*−*16*). Indeed,

* To whom correspondence should be addressed. Telephone: 46-18-4714205. Fax: 46-18-511755. E-mail: spoel@xray.bmc.uu.se.

10.1021/bi061182y CCC: $37.00
Published on Web 01/05/2007

complex interactions can be studied in detail by mass spectrometry (*17*−*19*).

Because of its documented effectiveness within MS, ESI has been proposed as a method for particle injection in future X-ray free electron lasers (XFEL) as well (*20*, *21*). The goal of these experiments is to determine structures of biomolecules or small aggregates, e.g., nanocrystals in the gas phase or within a small droplet of water. Today, the predominant technique for determining atomic-resolution structures of macromolecules, such as proteins, is X-ray crystallography (*22*). However, the need for diffracting crystals also forms the major bottleneck of the method. Many biologically important molecules such as membrane proteins, which are the most common targets for pharmaceuticals, are very difficult to crystallize, causing systematic vacancies in the understanding of the function of important molecular networks. Henceforth, there is a growing demand for other techniques with the ability of determining even the most difficult structures, and the XFEL is an instrument with the potential to revolutionize the field of structural biology. On the basis of the use of extremely intense femtosecond pulses of X-rays, it may be possible to record diffraction patterns of small biomolecular aggregates or even (large) single particles (*20*, *21*). In such experiments, the sample may survive the radiation damage due to the X-rays if the experimental parameters are carefully chosen (*21*, *23*−*25*). Data from many scattering experiments will have to be collected and combined to reconstruct the original structure (*26*, *27*). This in turn relies on the reproducibility of the samples, and although one can be fairly confident that proteins in solvent have the same conformation, this is not given in vacuum. A requirement for XFEL single-particle imaging is an injection process that can provide a purified stream of molecules into a vacuum chamber where the interaction with the highly focused X-ray pulse takes place. Because of its ability to form intact protein ions under gentle conditions, and its proven adequacy within mass spectrometry, electrospray ionization has been selected as the injection method for the XFEL.

On the basis of the requirements of both standard mass spectrometry and the XFEL imaging technique, we ask ourselves the following question in this study: What happens to a protein or protein complex on a molecular and structural level, when transferred from a fully solvated environment to vacuum, through a successive dehydration process using an electrospray?. A straightforward way to study this is to use molecular dynamics simulations. The time the protein spends in the electrospray before being detected by a mass spectrometric instrument is between microseconds and milliseconds, depending on the electrospray used, a time scale that is beyond the scope of simulations of full-size proteins. Therefore, we emulate the various steps of the electrospray process by simulating proteins covered with water shells of varying thickness in vacuum. The proteins that we use (Table 1) have been studied by various experimental methods (*5*, *28*−*37*) and are thus well-characterized, which makes them potential targets for initial XFEL experiments. From our vacuum simulations, we can study the dynamics of the structure over time and investigate how the transition from a solvated environment to vacuum affects the protein conformation and other properties.

Table 1: Overview of the Simulated Proteins

PDB entry	protein	no. of amino acids	resolution (Å)	source
1L2Y (*42*)	Trp-cage	20	NMR	synthetic construct
4INS (*39*)	insulin	51	1.50	pig
1CTF (*38*)	ribosomal L7/L12	68	1.70	*E. coli*
1UBQ (*40*)	ubiquitin	79	1.80	*Homo sapiens*
1AKI (*41*)	lysozyme	129	1.50	hen egg white

MATERIALS AND METHODS

Proteins. Five globular proteins (Table 1) of varying sizes were selected for this study. The structures that were used have been determined either by X-ray crystallography (*38*−*41*) or by NMR spectroscopy (*42*) and were retrieved from the Protein Data Bank. Since insulin is known to be in a monomeric state in its active form (*43*), a monomeric structure was used in our study. However, as the monomers tend to aggregate in solution, no crystal structure of an insulin monomer has hitherto been reported. Therefore, a dimer structure was selected, from which only the first of the two monomers was used for simulation and analysis. The L7/L12 ribosomal protein exists as a dimer in their biologically functional form, the monomers of which are identical except for the acetylated N-terminal serine of L7. The C-terminal fragment used in this study (termed Ctf) is, however, known to be remarkably stable in the isolated form as well (*37*) and thus suitable as a model protein.

Water Simulation. Initially, all five proteins were subject to relatively short (10−20 ns) simulations in solvent using periodic boundary conditions. The aim was to let the PDB structure equilibrate in solvent and to produce slightly different starting structures for the vacuum runs, thereby taking the natural dynamics of the proteins into account explicitly. The simulations were run for either 10 or 20 ns, long enough for the structures to stabilize (see the Results and Figure 1). Charges corresponding to pH 7 were used for the solution simulation. Sodium or chloride ions, using the force field parameters developed by Åqvist (*44*), were added as necessary to neutralize the system. All five systems were simulated using the OPLS-AA force field (*45*−*47*) and the TIP4P (*48*) water model. Prior to each production simulation, an energy minimization was performed, followed by a 10 ps simulation with position restraints on the protein. To maintain the temperature at 300 K and the pressure at 1 bar, Berendsen weak coupling (*49*) was used, with coupling constants of 0.1 ps for the temperature and 20 ps for the pressure. A twin-range cutoff of 0.9/1.4 nm for van der Waals interactions was applied, where the long-range interactions were updated every fifth time step, and analytic corrections to the energy and pressure beyond the cutoff were used (*50*). The smooth particle mesh Ewald algorithm (*51*, *52*) was used for Coulomb interactions, with a switching distance of 0.9 nm. Furthermore, all hydrogen atoms were replaced with virtual (dummy) particles to allow for longer time steps of 4 fs (*53*−*55*). Neighbor lists were utilized and updated every fifth integration step, and constraints were applied for bond lengths using the LINCS algorithm for the protein (*56*) and SETTLE (*57*) for the water. All simulations were parallelized to run on four processors, and everything was performed with GROMACS, version 3.3 (*58*−*60*). An overview of the water simulations is provided in Table 2.

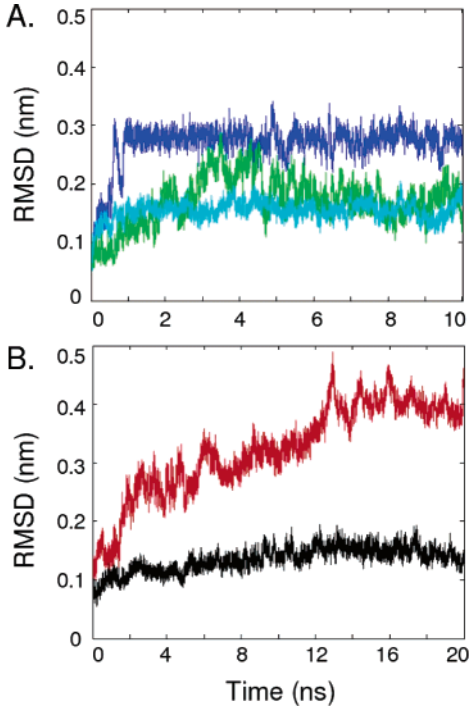

FIGURE 1: Structural drift of the five proteins, when run in bulkwater, presented as the Cα rmsd from the PDB structure. Presented in panel A are Trp-cage (blue), Ctf (cyan), and ubiquitin (green) and in panel B insulin (red) and lysozyme (black).

Table 2: Overview of the Simulations[a]

	no. of water molecules				charge			
protein	bulk	0.6 nm	0.3 nm	0 nm	bulk	0.6 nm	0.3 nm	0 nm
Trp-cage	1849	342	95	0	+1	+1	+1	+2
insulin	3521	586	161	0	−2	−2	−2	+5
Ctf	5126	681	217	0	−2	−2	−2	+5
ubiquitin	5743	770	254	0	0	0	0	+7
lysozyme	8499	1042	317	0	+8	+8	+8	+8

[a] The three vacuum systems (0.6, 0.3, and 0 nm) were simulated three times with different starting conformations.

Vacuum Simulation. From the last part of the water simulations, three structures with a reasonable rmsd with respect to the native PDB structure were extracted together with surrounding waters and used as starting structures for the vacuum runs. By deleting the outermost waters, we created three different systems with different sizes of water shells (0, 0.3, and 0.6 nm) from each starting structure, giving nine systems in total for each protein (Table 2). On the basis of the findings that protonation of macromolecules in ESI takes place during the very last step of the evaporation, as stated by the charge residue model (CRM) (*1*, *61*−*63*), the solution phase charges used in the bulk water simulations were also preserved in the 0.6 and 0.3 nm systems. However, for the 0 nm systems, the charges were changed (Table 2 and 1) in accordance with published experimental data (*5*, *32*, *64*−*66*) on the proteins in vacuum. For those proteins where no such data have been published, the locations of the vacuum charges for a particular charge state were estimated by electrostatic calculations using

$$E_{\text{coul}} = \sum_{i=1}^{N} \sum_{j=i+1}^{N} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \quad (1)$$

Table 3: Charged Residues of the 0 nm Simulations[a]

protein	positive	negative
Trp-cage	Q5, R16	
insulin	G1, Q15, H26, H31, R43, K50	E4
Ctf	K7, K13, K18, K32, K56	E1, E36
ubiquitin	K6, K11, K27, K29, K33, R42, K48, R54, K63, R72, R74	E16, E34, D39, E64
lysozyme	R21, R45, R68, R73, K96, R114, K116, R128	

[a] The numbering starts with 1 from the N-terminus. The two chains of insulin are counted as one single amino acid chain, numbered from 1 to 51, with chain B starting at residue 22. All remaining residues are left neutral.

The net charge corresponding to the most abundant ion found in ESI experiments (*5*, *32*, *64*−*68*) was selected, and a set of potential protonation sites were permutated and the Coulomb energies calculated. The charge configuration with the lowest energy was then selected for subsequent vacuum simulations. Note that even though the charge configuration with a minimal E_{coul} is not necessarily the one with highest probability, it is likely that it is at least a configuration of high probability and consequently a reasonable descriptor of the protein in vacuum. The charges used in the vacuum simulations are given in Table 3. Similar but more extensive methods for determining protonation states in vacuo have been reported (*66*).

Prior to the final production simulations of 10 ns, the nine systems for each protein were energy minimized and then run for 50 ps in vacuum with the temperature kept constant at 300 K. During the production simulations, the temperature coupling was turned off, and all bonds containing hydrogens were constrained using the SHAKE algorithm (*69*) with a relative tolerance of 10^{-8}, and the integration time step was reduced to 1 fs. For the remaining parameters, the same settings were used as in the water simulations, except that the periodic boundary conditions were turned off and no cutoffs whatsoever were used. The larger systems were all run in parallel on two processors, whereas the smaller systems were run on single processors. All simulations were run in double precision.

Analysis. The total energy of the simulated systems was monitored to ensure energy conservation, which was found to be very good ($\Delta E_{\text{tot}}/E_{\text{tot}} < 5 \times 10^{-4}$ during 10 ns of simulation). Structure analyses were done using available GROMACS programs except as otherwise stated; the conserved hydrogen bond analysis was implemented specifically for this study, as described below.

We first define the following reference sets of hydrogen bonds: A, the hydrogen bonds between protein and specific solvent molecules (crystal waters) that are present in the experimental structure; B, the intramolecular hydrogen bonds present for at least 50% of the last nanosecond in the bulk simulations; C, the intramolecular hydrogen bonds present in the experimental structure; and D, same as C, but restricted to those hydrogen bonds involved in β-sheets or α-helices. With this definition, we calculate the fraction of time χ that these hydrogen bonds exist. χ is averaged over the last nanosecond of a simulation, and over the hydrogen bonds. For the vacuum simulations, the average and standard deviation are determined from averaging over the three replicas with different starting conformations.

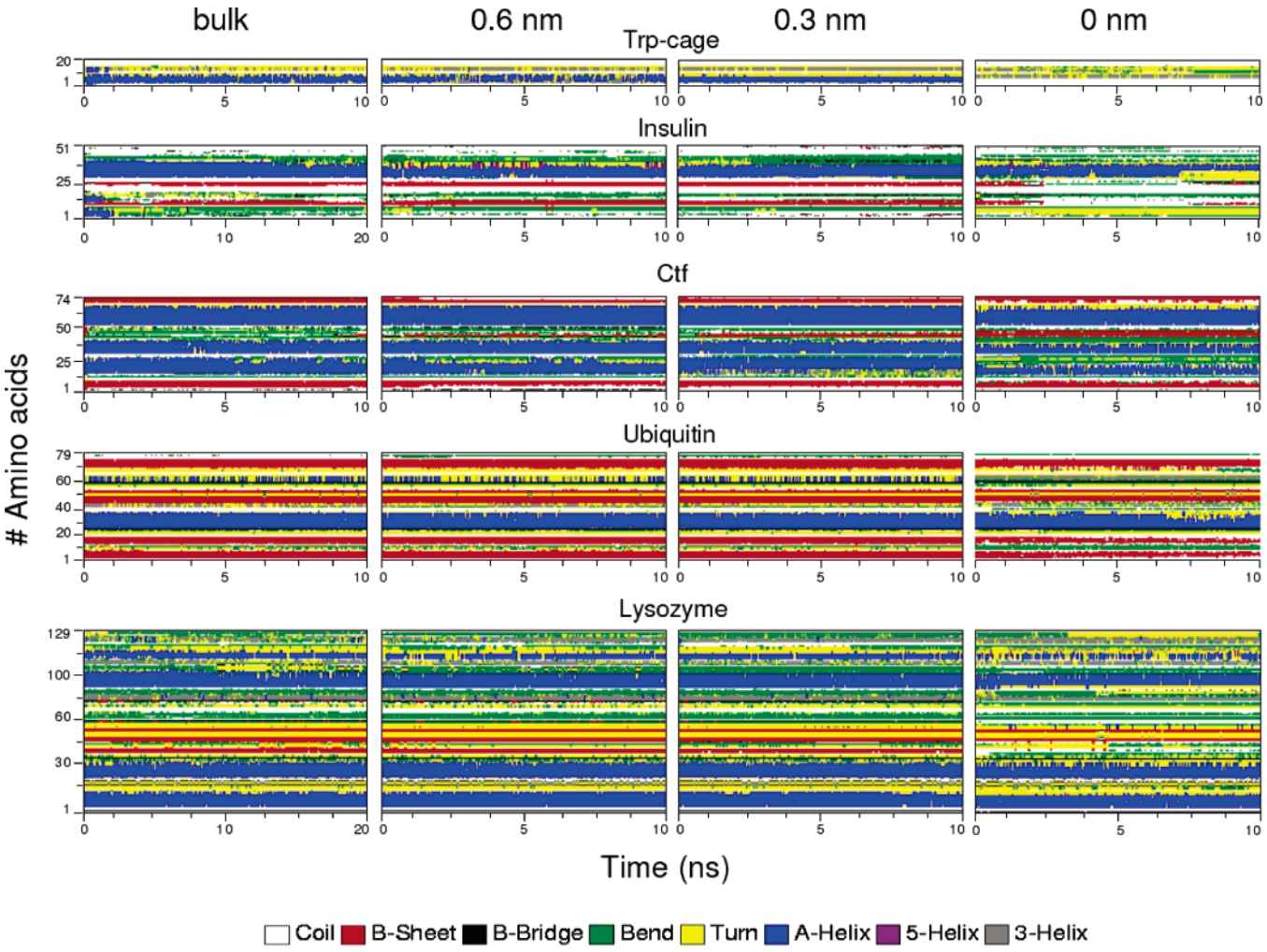

FIGURE 2: Secondary structure content of the five proteins as a function of time, when run in bulk, with a 0.6 nm water shell, a 0.3 nm shell, and in vacuum (0 nm).

Hydrogen bonds were defined using a geometrical criterion based on a donor−acceptor distance cutoff of 0.35 nm and a maximum hydrogen−donor−acceptor angle of 30°. As donor groups, OH and NH groups were used. No distinction was made between the oxygens of the carboxylate ions of aspartate, glutamate, and the C-terminus or between the amine groups of arginine.

RESULTS

Water Simulation

The preparatory water simulations were run as long as required for the dynamics of the proteins to stabilize, and convergence of the Cα rmsd from the PDB structure was used as a stabilization criterion. For Trp-cage, Ctf, and ubiquitin, 10 ns was enough, whereas insulin and lysozyme needed 20 ns to converge (Figure 1A). The average rmsd's of the last 3 ns of the simulations are 0.28 nm for Trp-cage, 0.40 nm for insulin, 0.16 nm for Ctf, 0.17 nm for ubiquitin, and 0.14 nm for lysozyme. The extreme increase in rmsd for insulin can be explained by the fact that as a monomer, this protein is known to have highly flexible N-terminal and C-terminal ends (*70*, *71*). A recent study by Zoete et al. (*70*) shows a rmsd of 0.3 nm for a monomeric insulin in bulk water, which is exactly what we obtain after the same time of simulation (Figure 1). The rmsd calculated for Ctf is in good agreement with an early simulation study in water by Daggett and Levitt (*72*).

To realize a qualitative understanding of the structural changes causing the increase in rmsd, the secondary structure content was analyzed at every 50th ps, using the DSSP program (*73*) (Figure 2). The major structural features are well-preserved, with minor local distortions. The only protein subject to serious structural changes is insulin. Furthermore, β-sheets are found to be more stable than α-helices (*74*), although the second β-sheet of Ctf is disrupted almost immediately and recovers only at the very end of the simulation. The most serious loss of α-helical structure is observed for insulin, which after simulation for only 1 ns has lost two of its three helices. The helix that is preserved is part of chain B, in agreement with previous studies, stating that this part of the protein is found to be very stable (*70*). It is believed that the structure of chain B is independent of the structure of chain A such that it can remain intact even if chain A has unfolded (*75*). The transition of the C-terminal part of the helix into a β-turn, found at the end of the simulation, has also been observed in a recent simulation study of the isolated B chain (*76*). The instability of chain A observed in our simulation is, however, in disagreement with the results presented by Zoete et al. (*70*), who found it to be the most stable part of the entire monomer, together with the helix of chain B.

Table 4: Average Temperature of the Final Protein−Water Cluster during the Last Nanosecond of the Simulation (starting temperature of 300 K)[a]

	temperature (K)					
	0.6 nm			0.3 nm		
Trp-cage	255 (3.6)	260 (4.3)	258 (5.3)	238 (7.5)	239 (6.6)	252 (9.9)
insulin	267 (1.5)	263 (3.9)	266 (5.4)	261 (5.0)	260 (8.5)	262 (1.6)
Ctf	263 (5.0)	263 (4.8)	266 (2.7)	262 (5.3)	262 (6.4)	261 (4.7)
ubiquitin	266 (5.4)	266 (3.5)	268 (3.7)	266 (4.2)	261 (5.3)	265 (3.8)
lysozyme	265 (3.0)	265 (3.2)	267 (2.8)	267 (4.7)	263 (4.7)	263 (1.4)

[a] The standard deviation is given in parentheses.

Table 5: Absolute Number and Fraction of the Total Number of Water Molecules that Have Evaporated after Simulation for 10 ns

	evaporation (no./%)					
	0.6 nm			0.3 nm		
Trp-cage	43/13	37/11	37/11	22/23	20/21	21/22
insulin	60/10	63/11	63/11	31/19	31/19	32/20
Ctf	68/10	69/10	70/10	29/13	32/15	33/15
ubiquitin	69/9	68/9	68/9	33/13	41/16	38/16
lysozyme	112/11	116/11	114/11	55/17	57/18	52/16

For the other four proteins, Trp-cage, Ctf, ubiquitin, and lysozyme, the secondary structures are well-preserved. Only minor disruptions, like a few transitions from short 3_{10}-helices into normal α-helical structure and vice versa, and interconversions between turns and bends are observed. Some temporary loss of structure at the ends of the α-helices can also be found. Trp-cage is slightly more affected than the other four proteins, probably because of its very small size and because it consists almost exclusively of helices and turns.

Evaporation

Immediately as the protein-containing droplets are extracted from the equilibrated bulk water simulations and transferred into vacuum, the surrounding waters start to evaporate. Evaporation implies a transfer of kinetic energy from the big droplet to the leaving waters, which causes a natural decrease in temperature (Table 4). An interesting observation is that the decrease is very similar for the four "normal sized'' proteins (insulin, Ctf, ubiquitin, and lysozyme), although the total sizes of the systems are quite different. For Trp-cage, a much larger decrease in temperature is found, especially for the 0.3 nm clusters. The absolute number of waters that has evaporated at the end of the simulation is system specific and increases with the size of the protein. There is good agreement between the three replicas of each system. However, the evaporation in terms of the fraction of the initial total number of waters is similar for all systems. The net evaporation from the 0.3 nm simulations is between 15 and 20% versus 10% from the 0.6 nm simulations (Table 5).

Structural Drift

To investigate how the evaporation and phase transition from solution to vacuum affect the stability of the native structure, the following structural parameters were calculated from all vacuum simulations: the Cα root-mean-square deviation (rmsd), the number of hydrogen bonds within the protein (HB_{pp}) and between the protein and the surrounding water molecules (HB_{ps}), the radius of gyration (R_g), and the total (A_{tot}) and hydrophobic (A_{phob}) surface area of the proteins. To obtain an appropriate estimate, only the last 3 ns of each simulation was considered, and the results, given as time averages, are presented in Tables 6 (rmsd) and 7 (HB_{pp}, HB_{ps}, R_g, A_{tot}, and A_{phob}).

Table 6: Structural Drift of the Different Vacuum Simulations Indicated by Cα Root-Mean-Square Deviation (rmsd) from Two Different Reference Structures, the PDB Structure (PDB) and a Structure Representative of the Protein when Equilibrated in Bulk Solvent (sol)[a]

	bulk	0.6 nm			0.3 nm			0 nm		
Trp-cage										
PDB	0.28	0.24	0.25	0.24	0.24	0.23	0.28	0.37	0.39	0.36
sol	−	0.13	0.13	0.12	0.16	0.13	0.05	0.22	0.19	0.27
insulin										
PDB	0.40	0.46	0.41	0.43	0.55	0.44	0.41	0.47	0.56	0.50
sol	−	0.26	0.25	0.25	0.41	0.26	0.20	0.31	0.48	0.44
Ctf										
PDB	0.16	0.14	0.15	0.15	0.16	0.17	0.22	0.39	0.37	0.35
sol	−	0.12	0.11	0.13	0.16	0.19	0.22	0.44	0.37	0.36
ubiquitin										
PDB	0.17	0.18	0.19	0.28	0.19	0.29	0.19	0.43	0.47	0.49
sol	−	0.13	0.13	0.19	0.13	0.19	0.13	0.44	0.45	0.47
lysozyme										
PDB	0.13	0.14	0.13	0.14	0.13	0.15	0.16	0.26	0.33	0.34
sol	−	0.10	0.09	0.08	0.12	0.11	0.11	0.27	0.35	0.34

[a] The rmsd that is given is the time average of the last 3 ns of the simulations. Values are given in nanometers.

Root-Mean-Square Deviation (rmsd). The rmsd was calculated with respect to two different reference structures: the PDB structure (termed the native structure) and a structure representing the solvated conformation of the protein. This solvated conformation was selected from a cluster analysis of the last 3 ns of the water simulation using the algorithm of Daura et al. (*77*) with a 0.1 nm Cα rmsd as a cutoff criterion.

From the results (Table 6), we can see that the protein structures are indeed changing as the waters start to evaporate. The PDB structure is well preserved when the protein is kept in bulk solvent (column bulk); rmsd values of ∼0.15 nm can be considered as normal fluctuations. The only protein showing a significant rmsd is insulin, as explained in this section. However, as the number of surrounding waters decreases, the rmsd starts to increase. For all five proteins, the most significant structural drift, both from the PDB structure and from the solution conformation, is observed when all water is gone, as studied by the 0 nm simulations. It seems that the gas phase structure is different from both the solution phase structure and the crystal structure, or in other words, the more water that is kept around the protein, the more structural similarity to either of the references is preserved.

Table 7: Number of Hydrogen Bonds within the Protein (HB_{pp}) and between the Protein and the Water Molecules (HB_{ps}), Radius of Gyration (R_g, nanometers), and the Total (A_{tot}) and Hydrophobic (A_{phob}) Surface Area (square nanometers) of the Proteins[a]

	bulk		0.6 nm		0.3 nm			0 nm			PDB
					Trp-cage						
HB_{pp}	10	10	10	10	10	12	13	13	15	14	9
HB_{ps}	47	44	44	44	38	37	37	–	–	–	–
R_g	0.70	0.72	0.72	0.71	0.71	0.71	0.70	0.70	0.71	0.72	0.73
A_{tot}	17	18	18	18	18	18	17	17	17	17	13
A_{phob}	10	11	11	11	11	10	10	12	11	12	8
					Insulin						
HB_{pp}	12	13	12	14	16	18	18	38	37	38	34
HB_{ps}	154	150	152	154	127	126	138	–	–	–	4
R_g	1.03	0.98	1.00	1.01	0.96	0.98	0.98	0.98	0.96	1.05	1.03
A_{tot}	39	35	38	39	35	36	35	36	35	37	37
A_{phob}	22	19	22	22	21	22	21	23	23	23	20
					Ctf						
HB_{pp}	39	41	40	39	46	45	46	60	54	56	53
HB_{ps}	182	176	183	187	147	148	153	–	–	–	23
R_g	1.12	1.10	1.10	1.11	1.11	1.11	1.10	1.14	1.11	1.11	1.12
A_{tot}	46	44	44	46	43	42	43	41	43	42	42
A_{phob}	22	21	21	23	23	22	23	27	27	27	21
					Ubiquitin						
HB_{pp}	56	57	58	63	61	62	60	73	69	73	51
HB_{ps}	180	179	182	175	156	154	158	–	–	–	28
R_g	1.19	1.16	1.17	1.16	1.16	1.16	1.16	1.21	1.17	1.16	1.18
A_{tot}	48	48	48	47	47	47	47	47	46	44	48
A_{phob}	22	22	23	23	22	23	23	27	27	27	21
					Lysozyme						
HB_{pp}	94	92	95	97	105	102	97	123	121	122	103
HB_{ps}	276	285	284	279	227	235	242	–	–	–	58
R_g	1.44	1.43	1.44	1.43	1.41	1.40	1.41	1.41	1.45	1.35	1.40
A_{tot}	73	73	73	72	71	71	72	67	69	64	64
A_{phob}	32	32	33	32	33	33	34	38	38	36	28

[a] Time averages of the last 3 ns of each simulation are given.

Deviating rmsd values that are observed in two of the ubiquitin simulations (one 0.3 nm and one 0.6 nm replica) can be attributed to flexibility in the last five residues of the C-terminal tail. These residues are not well defined in the crystal structure used as a starting conformation for our simulations and have no significant intramolecular hydrogen bonding or close packing contacts with the rest of the molecule (*40*). In new rmsd calculations using the first 71 Cα atoms only, the rmsd was lowered to ∼0.25 nm for the 0 nm systems and to 0.075−0.1 nm for the remaining systems, and no outliers were observed.

Hydrogen bonds (HB_{pp} and HB_{ps}). In Table 7, we present the number of hydrogen bonds within the proteins and between the proteins and the surrounding waters. There is a profound increase in the number of intramolecular hydrogen bonds as the number of waters decreases, a pattern that can be observed for all proteins, and in particular for the very flexible insulin. The increase in HB_{pp} is of course reflected by a decrease in HB_{ps}, as the protein−water complex strives to maximize the total number of hydrogen bonds. The calculated number for each specific shell is almost constant for the separate proteins, with only a few bonds being different among the three replicas. The 0.6 nm simulations have approximately the same values of HB_{pp} and HB_{ps} as the solution phase structures, indicating that a layer with a thickness of four to six water molecules is enough to mimic bulk solvent with respect to the hydrogen bonding pattern. Also, three of the five PDB structures have an HB_{pp} that is similar to those of the solvated structures: 9 versus 10 for Trp-cage, 51 versus 56 for ubiquitin, and 103 versus 94 for lysozyme. Insulin and ctf have a much larger number of intramolecular bonds in the crystal structure than in solution.

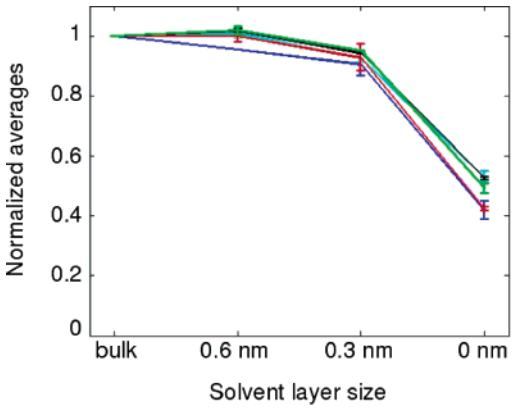

FIGURE 3: Total number of protein groups involved in hydrogen bonds in the 0, 0.3, and 0.6 nm and bulk water simulations of Trp-cage (blue), insulin (red), Ctf (cyan), ubiquitin (green), and lysozyme (black). Presented are the averages of the three replicas from each simulation, normalized to the bulk value. Standard deviations are given as error bars.

The total number of protein groups involved in hydrogen bonding in all solvated simulations (bulk, 0.6 and 0.3 nm) was calculated using the formula

$$HB = 2HB_{pp} + HB_{ps}$$

and the results are presented in Figure 3 (the values are normalized with respect to the bulk water simulation). All five proteins give very similar curves; they all have a plateau

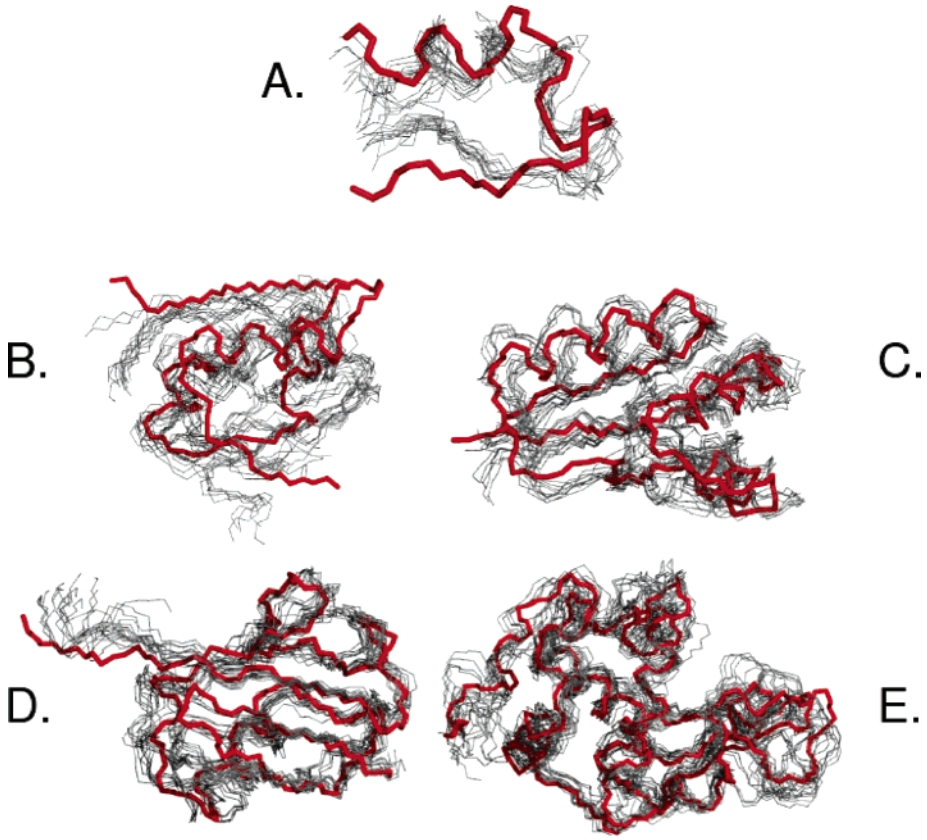

FIGURE 4: Snapshots taken from every second nanosecond of the three 0 nm simulations (gray), superimposed on the PDB structure (red): (A) Trp-cage, (B) insulin, (C) Ctf, (D) ubiquitin, and (E) lysozyme.

between the bulk and 0.3 nm level and then decrease sharply between the 0.3 and 0 nm simulations. Insulin, Ctf, ubiquitin, and lysozyme have their maximum HB at 0.6 nm, whereas Trp-cage has its maximum level of occupied hydrogen bonding sites in the bulk water simulation.

Radius of Gyration (R_g). The overall size of the proteins, estimated by the radius of gyration (R_g, Table 7), is affected very little by the presence or absence of a surrounding water shell. However, for the larger proteins, a slight decrease in R_g can be observed as the water shell becomes thinner, a behavior that is in good agreement with the pronounced increase in the number of intramolecular hydrogen bonds (HB_{pp}) observed when the protein goes from being fully solvated to dry. The size of the protein in the PDB structure is approximately the same as that in bulk water.

Surface Area (A_{tot} and A_{phob}). The total solvent accessible surface area (A_{tot}, Table 7) of the proteins, as determined using the double-cube lattice method (*78*), shows a decrease that is correlated with the evaporation of the waters; as the number of waters decreases, so does the total surface area. The areas calculated from the 0.6 nm simulations are comparable to the areas obtained from the bulk water, indicating once again that a layer of as few as one or two water molecules is sufficient to mimic the solvent environment. Only the PDB structures deviate from the pattern as they also have surface areas that are comparable to the solution structure (only lysozyme is an exception). For the hydrophobic part of the surface area (A_{phob}, Table 7), the observed trend is the opposite: the more exposed the protein is to the gas phase, the larger the hydrophobic surface area becomes. The similarity between the PDB structure and the solution phase structures can be observed here as well.

Secondary Structure Analysis

A secondary structure analysis, using DSSP (*73*) was applied to the vacuum simulations. For brevity, only one replica from each protein and each water shell, including the 0 nm systems, are presented (Figure 2).

First, some general conclusions can be drawn. For instance, when the proteins are transferred from solution to vacuum directly, the effect on the structures is more severe. Second, the β-sheets are somewhat better preserved, compared to the α-helices (*74*), even for the fully naked proteins. Only insulin experiences significant loss of β-sheet structure. The minor structural distortions observed for the 0.3 and 0.6 nm simulations are very similar; hence, the degree of protection is much the same for both water layers.

Looking at the proteins separately, one can state that a majority of the specific features are indeed preserved, even in the 0 nm case. The biggest challenge seems to be the preservation of α-helices, which mostly lose some of their structure at their ends. There also seems to be a frequent transition from normal α-helix to 3_{10}-helix (see Trp-cage and lysozyme) or, much less often, 5-helix, as the protein gradually becomes desolvated.

To visualize the structural changes caused by the vacuum environment, single snapshots from every second nanosecond of the three 0 nm simulation replicas have been extracted and superimposed on the PDB structure (Figure 4). In general, the vacuum conformations are more similar to each other than to the native structure, especially within the separate simulations, and the secondary structure elements are preserved but dislocated in space. The average rmsd (Figure 5) obtained from pairwise calculations using 30

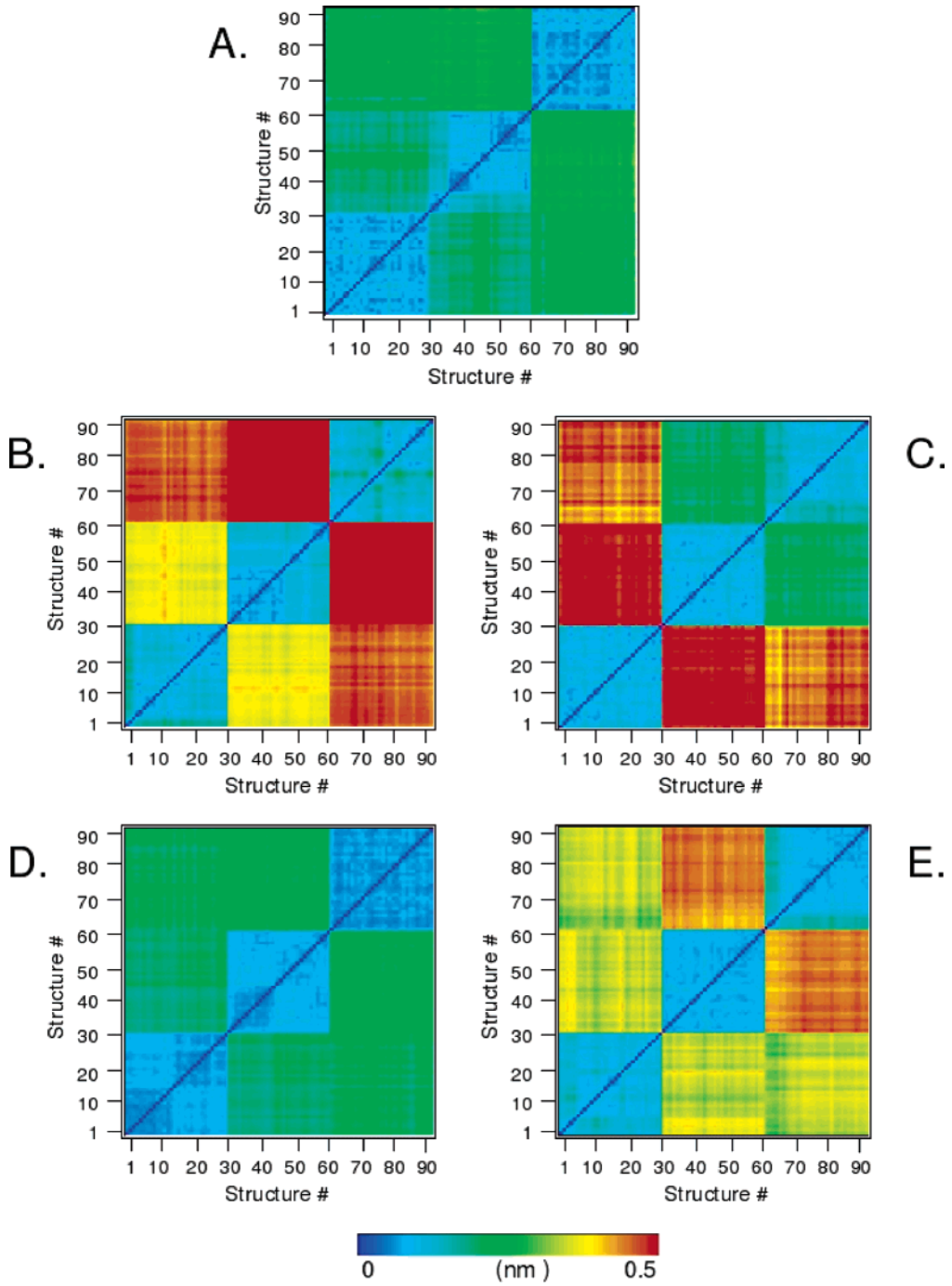

FIGURE 5: rmsd matrices of structures taken every 100 ps of the last 3 ns of the three 0 nm simulations: (A) Trp-cage, (B) insulin, (C) Ctf, (D) ubiquitin, and (E) lysozyme.

structures taken every 100 ps of the last 3 ns of each separate simulation is much lower than the average rmsd (Table 6) from both the solution reference and the native PDB structure: 0.08, 0.10, and 0.08 nm for Trp-cage, 0.11, 0.10, and 0.12 nm for insulin, 0.09, 0.09, and 0.12 nm for Ctf, 0.07, 0.08, and 0.06 nm for ubiquitin, and 0.11, 0.08, and 0.10 nm for lysozyme, respectively. Also, the average rmsd obtained from pairwise calculations using all 90 structures from all three simulations is quite low, 0.18 nm for Trp-cage, 0.36 nm for insulin, 0.29 nm for Ctf, 0.16 nm for ubiquitin, and 0.29 for lysozyme. Notice that the values given for ubiquitin are based on the first 71 amino acids; only the flexible C-terminus has not been taken into account.

Conserved Hydrogen Bonds

Protein−Water. Crystal structures of proteins usually contain water molecules, often termed structural waters. The presence of these waters may play a role in the activity and structure preservation of the proteins and has therefore attracted much interest. For instance, the influence on the protein structure (*79*, *80*), kinetics (*81*), and thermodynamics (*82*, *83*) of the hydration has been investigated. However, the results are not unified, and some claim that buried waters, found in inner cavities of proteins, could even destabilize the structure due to entropic and energetic costs of hydration (*83*, *84*). Here, we have analyzed whether the water−protein

Table 8: Number of Structural Waters Found in the PDB File (not Counting Water Molecules with an Occupancy of <1)

protein	PDB	protein	PDB
Trp-cage	0	ubiquitin	58
insulin	48	lysozyme	78
Ctf	62		

interactions found in the PDB files (Table 8) selected as starting conformations still could be found after simulation for several nanoseconds, first in bulk solvent and later in vacuum. Our hypothesis was that if there are some waters that are more important for the native structure than the bulk waters, then the hydrogen bonds connecting them to the protein should be slightly stronger than an average hydrogen bond, and these waters should therefore be preserved better and thus among the last to evaporate.

The results (Figure 6A) indicate that a minor fraction, 25−50%, of the water molecules bound to the protein in the PDB structure can be found at the end of the simulations, and this fraction is almost constant over the different water layers of each protein. Thus, the binding of these waters seems to be very little affected by the evaporation, at least as long as there are other waters left in the surrounding water layer. Although it cannot be expected that the same water molecule will be bound in the same position forever, the fact that they exist for the whole simulation indicates that these hydrogen bonds are stronger than others. A more involved analysis of hydrogen bond thermodynamics (*85*) that could in principle be done falls beyond the scope of this work however.

Protein−Protein. Since all analyses suggest that a thin layer of water molecules would be enough to give the same "protection'' to the protein structures as bulk water, we used the conserved hydrogen bond analysis to investigate how much the binding patterns of intramolecular hydrogen bonds of the vacuum simulations differ from that of the fully solvated protein. The set of intramolecular hydrogen bonds found to exist during more than 50% of the last nanosecond of each bulk water simulation was selected as a reference set, and the analysis of the vacuum simulations was restricted to these protein atoms only. In other words, we monitor the fraction of time (χ) for which this specific set of hydrogen bonds exists.

In Figure 6B, we have plotted the average fraction of time (χ) that these hydrogen bonds are present. For ubiquitin, χ is almost identical in the bulk, 0.6 and 0.3 nm simulations, 84% or more during the last nanosecond. Lysozyme follows the same pattern as ubiquitin but with a slightly lower overall χ. Ctf has a large χ in bulk water and with a 0.6 nm shell, but it drops quickly as the number of surrounding waters decreases further. Trp-cage and insulin exhibit a different trend with a larger χ at 0.3 nm than at 0.6 nm. The uncertainties in the data for these two proteins are also quite large.

Next, the existence and preservation of the native intramolecular hydrogen bonds, defined as the bonds present in the PDB structure, were investigated. We wanted to find out whether these native bonds were conserved both in solution and after evaporation. The intramolecular hydrogen bonds of the PDB structure were selected as a reference, and the analysis of the simulations was restricted to these atoms only (Figure 6C). Both ubiquitin and lysozyme have their native hydrogen bonds intact during 60−70% of the last nanosecond of the bulk water, as well 0.6 and 0.3 nm simulations, on average. In complete vacuum, the existence χ decreases to ∼45%. For insulin, χ is highest in bulk water and with a 0.3 nm water shell, whereas Trp-cage has its maximum at 0 nm.

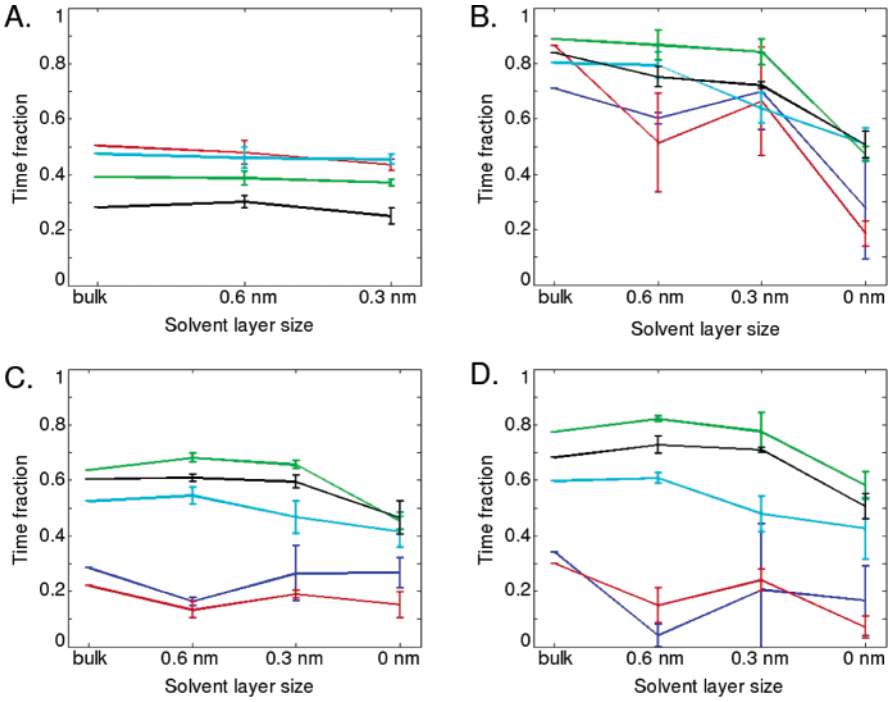

FIGURE 6: Conserved hydrogen bond analysis. Time fraction χ that hydrogen bonds exist in the last nanosecond of the simulations, where the reference set of hydrogen bonds is determined from (A) the crystal waters found in the PDB structure, (B) the intramolecular hydrogen bonds of the bulk water structures, (C) the native intramolecular hydrogen bonds of the PDB structure, and (D) the same as panel C, but restricted to those hydrogen bonds found in α-helices and/or β-sheets. Colors correspond to Trp-cage (blue), insulin (red), Ctf (cyan), ubiquitin (green), and lysozyme (black).

Ctf follows the same curve as ubiquitin and lysozyme in bulk water and with a 0.6 nm water layer but drops once again quickly as the number of waters decreases to 0.3 nm. In complete vacuum, Ctf has a χ value similar to those of Trp-cage and insulin.

When the analysis is further restricted, by considering only the native hydrogen bonds present in α-helices and/or β-sheets (Figure 6D), χ increases to >70% for both ubiquitin and lysozyme. However, the variation also increases somewhat. The abundance observed for Ctf has increased to ∼60%. Trp-cage and insulin still follow their own pattern, but with increased uncertainties in the numbers. This result indicates that the hydrogen bonds that stabilize the secondary structure are more stable upon dehydration than the other native hydrogen bonds, at least for the larger proteins.

DISCUSSION

The aim of this study was to simulate the transfer from water to vacuum that takes place during ESI and to monitor what happens to a protein structure as the surrounding waters start to evaporate until it becomes fully dehydrated. Our results indicate that the protein structures are indeed affected by the transition. The evaporation of the waters forces the protein to adjust its conformation such that the number of hydrogen bonds, and thus the internal energy, is optimized. This implies that the radius of gyration and total surface area decrease at the expense of some extra hydrophobic surface area. However, these conformational changes are not very severe since the secondary structure elements are largely intact and resemble what they look like in the native state. α-Helices are affected more than β-sheets, in agreement with earlier work on lactate dehydrogenase (*74*), with some loss of structure at their termini and/or transitions into 3_{10}-helices or, less frequently, 5-helices. In accordance with a study by Nagendra et al. (*86*) on the amount of water necessary in a lysozyme crystal to maintain the structure, we find that the amount of water required to protect the structure is small: a 0.3 nm shell seems to have much the same effect as a 0.6 nm shell. Furthermore, we see that the structural features at the end of these simulations are very similar to those of the fully solvated proteins. From this, we draw the conclusion that a shell with a thickness of not more than a few water molecules is enough to mimic bulk water. Our results (Figures 4 and 5) also show that in the gas phase, ensembles of structures can be found, which are very similar internally but which differ, both from the "native'' structure as obtained by experimental means and from the solution phase structure, as seen in our MD simulations. These gas phase ensembles are fairly close to each other in rmsd. It therefore seems as if the final evaporation causes a collapse of the solution structure into a vacuum structure which thereafter experiences very limited dynamics. Experimental data from ion mobility measurements (*36*, *87*, *88*) show, however, that both +7 ubiquitin and +8 lysozyme exist in several, more or less nativelike, conformational ensembles in vacuum. In our simulations of these proteins, we find two and three ensembles, respectively, predominantly compact and nativelike ones. Extended sampling through longer simulations or, for instance, replica exchange molecular dynamics simulations (*89*) could result in other, less compact structures. It seems plausible that there is a connection between the location of the charges and the conformational ensemble as well.

Evaporation seems to be faster for larger proteins, as observed by the increasing absolute number of evaporated waters with the increasing size of the protein. However, if we consider the evaporation as a percentage of the initial amount of water, the trend is that systems with a thin water layer evaporate faster than systems containing more waters, independent of the size of the water−protein cluster: the 0.6 nm simulations lose ∼10% of the water molecules in 10 ns, whereas the 0.3 nm simulations lose ∼15−20%. The drop in temperature is very similar for the four normal sized' proteins, as they converge to much the same temperature, independent of size, a finding that is in good agreement with observations by Caleman and Van der Spoel for evaporation of water clusters (*9*). Only the small Trp-cage experiences a temperature loss that is significantly larger. For all proteins, the evaporation from the three replicas of each system is reproducible in magnitude.

From the conserved hydrogen bond analysis (Figure 6A), we can see that for Ctf, ubiquitin, and insulin, 40−50% of the structural waters found in the PDB files remains bound to the protein after simulation 10 ns, and for lysozyme, 30%. Trp-cage has no structural waters and thus cannot be included in the analysis. The fraction that is found to be bound is constant over the different simulations, suggesting that there are certain water molecules which bind more tightly to the protein than the bulk water does and that these waters, hence, are affected only little by the evaporation taking place during our simulations. Moreover, the intramolecular hydrogen bonds that are identified in the fully solvated protein are to a large extent also present in vacuum (Figure 6B). For ubiquitin, which shows the most nativelike structure in vacuum, these intramolecular hydrogen bonds, identified in the solution structure, are found during more than 84% of the last nanosecond of the bulk water and the 0.6 and 0.3 nm simulations. At 0 nm of water, i.e., after complete evaporation, χ drops significantly, although the total number of hydrogen bonds within the protein increases. The intramolecular hydrogen bonds found in the PDB structures, defined as native hydrogen bonds, are well preserved in the simulations (Figure 6C): χ is between 60 and 70% for ubiquitin and lysozyme in the bulk, 0.6 and 0.3 nm simulations, and 45% in vacuo. For Trp-cage and insulin, the numbers are below 30%. Ctf has a large existence in bulk and with 0.6 nm, ∼55%, but it drops quickly as the number of waters decreases further. The native hydrogen bonds from α-helices and/or β-sheets are much more stable (Figure 6D); almost 60% is preserved even in vacuo for ubiquitin. This finding suggests that if a protein is deformed due to transfer to vacuum, but nevertheless has most of its hydrogen bonds intact, in particular in the secondary structure elements, then the native structure can be recovered with limited refolding only upon resolvation. This result is very promising, both for the interpretation of mass spectrometry experiments and for future bio-imaging experiments at the XFEL. The internal consistency between dehydrated structures (Figure 4), corresponding to a systematic deviation from the native structure, strongly suggests that if structures of dry proteins can be determined from XFEL experiments, a computational resolvation procedure (*90*) could be used to obtain the true structure (*91*), for instance, by employing replica exchange molecular dynamics simulations (*89*).

A comparison of panels B and C of Figure 6 also shows that the fraction of hydrogen bonds that is found in the vacuum simulations with respect to those found in bulk simulations is consistently higher than the fraction of hydrogen bonds in vacuum with respect to those in the experimental structure. From this, we can conclude that there is a slightly different hydrogen bonding pattern in solvent simulations compared to experiment which may be due to the force field or to the difference in environment between crystal and solution or, most likely, a combination of the two. If there were a systematic bias due to the force field, this would not affect the trend that we observe here. Therefore, we conclude that when the sample moves from bulk solution to vacuum, a significant fraction of the protein−protein hydrogen bonds is conserved (Figure 6), in particular in secondary structure elements.

ACKNOWLEDGMENT

The Lawrence Livermore National Laboratory (LLNL) is acknowledged for allocation of computer time on Thunder; Richard London, Abraham Szöke, Daniel Barsky, and Janos Hajdu are thanked for fruitful discussions.

REFERENCES

1. Dole, M., Mach, L. L., Hines, R. L., Ferguson, L. P., and Alice, M. B. (1968) Molecular beams of macroions, *J. Chem. Phys. 49*, 2240−2249.
2. McLafferty, F. W., Fridriksson, E. K., Horn, D. M., Lewis, M. A., and Zubarev, R. A. (1999) Biomolecule mass spectrometry, *Science 284*, 1289−1290.
3. Jarrold, M. F. (2000) Peptides and proteins in the vapor phase, *Annu. Rev. Phys. Chem. 51*, 179−207.
4. Zubarev, R. A., Kelleher, N. L., and McLafferty, F. W. (1998) Electron capture dissociation of multiply charged protein cations. A non-ergodic process, *J. Am. Chem. Soc. 120*, 3265−3266.
5. Breuker, K., Oh, H., Horn, D. M., Cerda, B. A., and McLafferty, F. W. (2002) Detailed unfolding and folding of gaseous ubiquitin ions characterized by electron capture dissociation, *J. Am. Chem. Soc. 124*, 6407−6420.
6. Yamashita, M., and Fenn, J. B. (1984) Electrospray ion source. Another variation on the free-jet theme, *J. Phys. Chem. 88*, 4451−4459.
7. Yamashita, M., and Fenn, J. B. (1984) Negative ion production with the electrospray ion source, *J. Phys. Chem. 88*, 4671−4675.
8. Fenn, J. B., Mann, M., Meng, C. K., Wong, S. F., and Whitehouse, C. M. (1989) Electrospray ionization for mass spectrometry of large biomolecules, *Science 246*, 64−71.
9. Calemand, C., and van der Spoel, D. (2006) Temperature and structural changes of water clusters in vacuum due to evaporation, *J. Chem. Phys. 125*, 154508.
10. Jarrold, M. F. (1999) Unfolding, refolding and hydration of protein ions in the gas phase, *Acc. Chem. Res. 32*, 360−367.
11. Velazquez, I., Reimann, C. T., and Tapia, O. (1999) Proteins in vacuo: Relaxation of unfolded lysozyme leads to folding into native and non-native structures. A molecular dynamics study, *J. Am. Chem. Soc. 121*, 11468−11477.
12. Iavarone, A., and Parks, J. (2005) Conformational change in unsolvated Trp-cage protein probed by fluorescence, *J. Am. Chem. Soc. 127*, 8606−8607.
13. Tito, M. A., Tars, K., Valegård, K., Hajdu, J., and Robinson, C. V. (2000) Electrospray time of flight mass spectrometry of the intact MS2 virus capsid, *J. Am. Chem. Soc. 122*, 3550−3551.
14. Benjamin, D. R., Robinson, C. V., Hendrick, J. P., Hartl, F. U., and Dobson, C. M. (1998) Mass spectrometry of ribosomes and ribosomal subunits, *Proc. Natl. Acad. Sci. U.S.A. 95*, 7391−7395.
15. Loo, J. A. (1997) Studying noncovalent protein complexes by electrospray ionization mass spectrometry, *Mass Spectrom. Rev. 16*, 1−23.
16. Rostom, A. A., Fucini, P., Benjamin, D. R., Juenemann, R., Nierhaus, K. H., Hartl, F. U., Dobson, C. M., and Robinson, C. V. (2000) Detection and characterization of intact ribosomes in the gas phase of a mass spectrometer, *Proc. Natl. Acad. Sci. U.S.A. 97*, 5185−5190.
17. Robinson, C. V., Gross, M., Eyles, S. J., Ewbank, J. J., Mayhew, M., Hartl, F. U., Dobson, C. M., and Radford, S. E. (1994) Conformation of GroEL-bound α-lactalbumin probed by mass spectrometry, *Nature 372*, 646−651.
18. Ruotolo, B. T., Giles, K., Campuzano, I., Sandercock, A. M., Bateman, R. H., and Robinson, C. V. (2005) Evidence for macromolecular protein rings in the absence of bulk water, *Science 310*, 1658−1661.
19. Heck, A. J. R., and van den Heuvel, R. H. H. (2004) Investigation of intact protein complexes by mass spectrometry, *Mass Spectrom. Rev. 23*, 368−389.
20. Neutze, R., Wouts, R., van der Spoel, D., Weckert, E., and Hajdu, J. (2000) Potential for biomolecular imaging with femtosecond X-ray pulses, *Nature 406*, 752−757.
21. Neutze, R., Huldt, G., Hajdu, J., and van der Spoel, D. (2004) Potential impact of an X-ray free electron laser on structural biology, *Radiat. Phys. Chem. 71*, 905−916.
22. Drenth, J. (1994) *Principles of Protein X-ray Crystallography*, Springer, New York.
23. Hau-Riege, S. P., London, R. A., and Szöke, A. (2004) Dynamics of biological molecules irradiated by short X-ray pulses, *Phys. Rev. E 69*, 051906.
24. Jurek, G. O., and Faigel, G. (2004) Imaging atom clusters by hard X-ray free-electron lasers, *Europhys. Lett. 65*, 491−497.
25. Bergh, M., Tîmneanu, N., and van der Spoel, D. (2004) A model for the dynamics of a water cluster in a X-ray FEL beam, *Phys. Rev. E 70*, 051904.
26. Miao, J. W., Charalambous, P., Kirz, J., and Sayre, D. (1999) Extending the methodology of X-ray crystallography to allow imaging of micrometre-sized non-crystalline specimens, *Nature 400*, 342−344.
27. Huldt, G., Szöke, A., and Hajdu, J. (2003) Diffraction imaging of single particles and biomolecules, *J. Struct. Biol. 144*, 219−227.
28. Katta, V., and Chalt, B. T. (1993) Hydrogen/deuterium exchange electrospray ionization mass spectrometry: A method for probing protein conformational changes in solution, *J. Am. Chem. Soc. 115*, 6317−6321.
29. Loo, J. A., Edmonds, C. G., Udseth, H. R., and Smith, R. D. (1990) Effect of reducing disulfide-containing proteins on electrospray ionization mass spectra, *Anal. Chem. 62*, 693−698.
30. Reimann, C. T., Velázquez, I., Bittner, M., and Tapia, O. (1999) Proteins in vacuo: A molecular dynamics study of the unfolding behavior of highly charged disulfide-bond-intact lysozyme subjected to a temperature pulse, *Am. Phys. Soc. 60*, 7277−7284.
31. Patriksson, A., Adams, C., Kjeldsen, F., Raber, J., van der Spoel, D., and Zubarev, R. A. (2006) Prediction of N-C_O bond cleavage frequencies in electron capture dissociation of Trp-cage dications by force-field molecular dynamics simulations, *Int. J. Mass Spectrom. 248*, 124−135.
32. Adams, C. M., Kjeldsen, F., and Zubarev, R. A. (2004) Electron capture dissociation distinguishes a single D-amino acid in a protein and probes the tertiary structure, *Am. Soc. Mass Spectrom. 15*, 1087−1098.
33. Adams, C. M., Kjeldsen, F., Patriksson, A., van der Spoel, D., Gräslund, A., Papadopoulos, E., and Zubarev, R. A. (2006) Probing solution- and gas-phase structures of Trp-cage cations by chiral substitution and spectroscopic techniques, *Int. J. Mass Spectrom. 253*, 263−273.
34. Kalthasov, I. A., and Mohimen, A. (2005) Estimates of protein surface areas in solution by electrospray ionization mass spectrometry, *Anal. Chem. 77*, 5370−5379.
35. Danell, A. S., and Parks, J. H. (2003) FRET measurements of trapped oligonucleotide duplexes, *Int. J. Mass Spectrom. 229*, 35−45.
36. Koeniger, S. L., Merenbloom, S. I., and Clemmer, D. E. (2006) Evidence for many resolvable structures within conformation types of electrosprayed ubiquitin ions, *J. Phys. Chem. B 110*, 7017−7021.
37. Luer, C. A., and Wong, K. (1980) Conformational stability of ribosomal protein L7/L12: Effects of pH, temperature and guanidinium chloride, *Biochemistry 19*, 176−183.
38. Leijonmarck, M., and Liljas, A. (1987) Structure of the C-terminal domain of the ribosomal protein L7/L12 from *Escherichia coli* at 1.7 ångströms, *J. Mol. Biol. 195*, 555.
39. Baker, E. N., Blundell, T. L., Cutfield, J. F., Cutfield, S. M., Dodson, E. J., Dodson, G. G., Hodgkin, D. M., Hubbard, R. E., Isaacs, N. W., et al. (1988) The structure of 2 Zn pig insulin crystals at 1.5 Å resolution, *Philos. Trans. R. Soc. London, Ser. B 319*, 369−456.

40. Vijay-Kumar, S., Bugg, C. E., and Cook, W. J. (1987) Structure of ubiquitin refined at 1.8 Å resolution, *J. Mol. Biol. 194*, 531−544.
41. Artymiuk, P. J., Blake, C. C. F., Rice, D. W., and Wilson, K. S. (1982) The structures of the monoclinic and orthorombic forms of hen egg-white lysozyme at 6 Ångstroms resolution, *Acta Crystallogr. B38*, 778.
42. Neidigh, J. W., Fesinmeyer, R. M., and Andersen, N. H. (2002) Designing a 20-residue protein, *Nat. Struct. Biol. 9*, 425−430.
43. Dodson, G., and Steiner, D. (1998) The role of assembly in insulin's biosynthesis, *Curr. Opin. Struct. Biol. 8*, 189−194.
44. Åqvist, J. (1990) Ion-water interaction potentials derived from free energy perturbation simulations, *J. Phys. Chem. 94*, 8021−8024.
45. Jorgensen, W. L., and Tirado-Rives, J. (1988) The OPLS potential functions for proteins. Energy minimizations for crystals of cyclic peptides and crambin, *J. Am. Chem. Soc. 110*, 1657−1666.
46. Kaminski, G. A., Friesner, R. A., Tirado-Rives, J., and Jorgensen, W. L. (2001) Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides, *J. Phys. Chem. B 105*, 6474−6487.
47. Jorgensen, W. L. (1998) in *Encyclopedia of Computational Chemistry*, Vol. 3, pp 1986−1989, Wiley, New York.
48. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983) Comparison of simple potential functions for simulating liquid water, *J. Chem. Phys. 79*, 926−935.
49. Berendsen, H. J. C., Postma, J. P. M., DiNola, A., and Haak, J. R. (1984) Molecular dynamics with coupling to an external bath, *J. Chem. Phys. 81*, 3684−3690.
50. Allen, M. P., and Tildesley, D. J. (1987) *Computer Simulations of Liquids*, Oxford Science Publications, Oxford, U.K.
51. Darden, T., York, D., and Pedersen, L. (1993) Particle mesh Ewald: An N-log(N) method for Ewald sums in large systems, *J. Chem. Phys. 98*, 10089−10092.
52. Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H., and Pedersen, L. G. (1995) A smooth particle mesh ewald method, *J. Chem. Phys. 103*, 8577−8592.
53. Feenstra, K. A., Hess, B., and Berendsen, H. J. C. (1999) Improving efficiency of large time-scale molecular dynamics simulations of hydrogen-rich systems, *J. Comput. Chem. 20*, 786−798.
54. Feenstra, K. A., Peter, C., Scheek, R. M., van Gunsteren, W. F., and Mark, A. (2002) A comparison of methods for calculating NMR cross-relaxation rates (NOESY and ROESY intensities) in small peptides, *J. Biomol. NMR 23*, 181−194.
55. van der Spoel, D., and Lindahl, E. (2003) Brute force molecular dynamics simulations of villin headpiece: Comparison with NMR parameters, *J. Phys. Chem. B 107*, 11178−11187.
56. Hess, B., Bekker, H., Berendsen, H. J. C., and Fraaije, J. G. E. M. (1997) LINCS: A linear constraint solver for molecular simulations, *J. Comput. Chem. 18*, 1463−1472.
57. Miyamoto, S., and Kollman, P. A. (1992) SETTLE: An analytical version of the SHAKE and RATTLE algorithms for rigid water models, *J. Comput. Chem. 13*, 952−962.
58. Berendsen, H. J. C., van der Spoel, D., and van Drunen, R. (1995) GROMACS: A message-passing parallel molecular dynamics implementation, *Comput. Phys. Commun. 91*, 43−56.
59. Lindahl, E., Hess, B. A., and van der Spoel, D. (2001) GROMACS 3.0: A package for molecular simulations and trajectory analysis, *J. Mol. Model. 7*, 306−317.
60. van der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E., and Berendsen, H. J. C. (2005) GROMACS: Fast, Flexible and Free, *J. Comput. Chem. 26*, 1701−1718.
61. Felitsyn, N., Peschke, M., and Kebarle, P. (2002) Origin and number of charges observed on multiply-protonated native proteins produced by ESI, *Int. J. Mass Spectrom. 219*, 39−62.
62. Verkerk, U. H., Peachke, M., and Kebarle, P. (2002) Effect of buffer cations and of H_3O^+ on the charge states of native proteins. Significance to determinations of stability constants of protein complexes, *J. Mass Spectrom. 38*, 618−631.
63. de la Mora, J. F. (2000) Electrospray ionization of large multiply charged species proceeds via Dole's charged residue mechanism, *Anal. Chim. Acta 406*, 93−104.
64. Nettleton, E. J., Tito, P., Sunde, M., Bouchard, M., Dobson, C. M., and Robinson, C. V. (2000) Characterization of the oligomeric states of insulin in self-assembly and amyloid fibril formation by mass spectrometry, *Biophys. J. 79*, 1053−1065.
65. Oh, H. B., Breuker, K., Sze, S. K., Ge, Y., Carpenter, B. K., and McLafferty, F. W. (2002) Secondary and tertiary structures of gaseous protein ions characterized by electron capture dissociation mass spectrometry and photofragment spectroscopy, *Proc. Natl. Acad. Sci. U.S.A. 99*, 15863−15868.
66. Miteva, M., Demirev, P. A., and Karshikoff, A. D. (1997) Multiply-protonated protein ions in the gas phase: Calulation of the electrostatic interactions between charged sites, *J. Phys. Chem. B 101*, 9645−9650.
67. Carbeck, J. D., Severs, J. C., Gao, J., Wu, Q., Smith, R. D., and Whitesides, G. M. (1998) Correlation between the charge of proteins in solution and in the gas phase investigated by protein charge ladders, capillary electrophoresis and electrospray ionization mass spectrometry, *J. Phys. Chem. B 102*, 10596−10601.
68. Konermann, L., and Douglas, D. J. (1998) Equilibrium unfolding of proteins monitored by electrospray ionization mass spectrometry: Distinguishing two-state from multi-state transitions, *Rapid Commun. Mass Spectrom. 12*, 435−442.
69. Ryckaert, J. P., Ciccotti, G., and Berendsen, H. J. C. (1977) Numerical integration of the Cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes, *J. Comput. Phys. 23*, 327−341.
70. Zoete, V., Meuwly, M., and Karplus, M. (2004) A comparison of the dynamic behavior of monomeric and dimeric insulin shows structural rearrangements in the active monomer, *J. Mol. Biol. 342*, 913−929.
71. Zhang, Y., Whittingham, J. L., Turkenborg, J. P., Dodson, E. J., Brange, J., and Dodson, G. G. (2002) Crystallization and preliminary crystallographic investigation of a low-pH native insuline monomer with flexible behavior, *Act. Crystallogr. D58*, 186−187.
72. Daggett, V., and Levitt, M. (1991) A molecular dynamics simulation of the C-terminal fragment of the L7/L12 ribosomal protein in solution, *Chem. Phys. 158*, 501−512.
73. Kabsch, W., and Sander, C. (1983) Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features, *Biopolymers 22*, 2577−2637.
74. van der Spoel, D., Vogel, H. J., and Berendsen, H. J. C. (1996) Molecular dynamics simulations of N-terminal peptides from a nucleotide binding protein, *Proteins: Struct., Funct., Genet. 24*, 450−466.
75. Qiao, Z., Min, C., Hua, Q., Weiss, M., and Feng, Y. (2003) In vitro refolding of human proinsulin, *J. Biol. Chem. 278*, 17800−17809.
76. Legge, F. S., Budi, A., Treutlein, H., and Yarovsky, I. (2006) Protein flexibility: Multiple molecular dynamics simulations of insuline chain b, *Biophys. Chem. 119*, 146−157.
77. Daura, X., Gademann, K., Jaun, B., Seebach, D., van Gunsteren, W. F., and Mark, A. E. (1999) Peptide folding: When simulation meets experiment, *Angew. Chem., Int. Ed. 38*, 236−240.
78. Eisenhaber, F., Lijnzaad, P., Argos, P., Sander, C., and Scharf, M. (1995) The double cube lattice method: Efficient approaches to numerical integration of surface area and volume and to dot surface contouring of molecular assemblies, *J. Comput. Chem. 16*, 273−284.
79. Williams, M. A., Goodfellow, J. M., and Thornton, J. M. (1994) Buried waters and internal cavities in monomeric proteins, *Protein Sci. 3*, 1224−1235.
80. Park, S., and Saven, J. G. (2005) Statistical and molecular dynamics studies of buried waters in globular proteins, *Proteins: Struct., Funct., Bioinf. 60*, 450−463.
81. Garcia, A. E., and Hummer, G. (2000) Water penetration and escape in proteins, *Proteins: Struct., Funct., Genet. 38*, 261−272.
82. Fischer, S., and Verma, C. S. (1999) Binding of buried structural waters increases the flexibility of proteins, *Proc. Natl. Acad. Sci. U.S.A. 96*, 9613−9615.
83. Zhang, L., and Hermans, J. (1996) Hydrophilicity of cavities in proteins, *Proteins: Struct., Funct., Genet. 24*, 433−438.
84. Dunitz, J. D. (1994) The entropic cost of bound water in crystals and biomolecules, *Science 264*, 670.
85. van der Spoel, D., van Maaren, P. J., Larsson, P., and Timneanu, N. (2006) Thermodynamics of hydrogen bonding in hydrophilic and hydrophobic media, *J. Phys. Chem. B 110*, 4393−4398.
86. Nagendra, H. G., Sukumar, N., and Vijayan, M. (1998) Role of water in plasticity, stability, and action of proteins: The crystal structures of lysozyme at very low levels of hydration, *Proteins: Struct., Funct., Genet. 32*, 229−240.

87. Valentine, S. J., Anderson, J. G., Ellington, A. D., and Clemmer, D. E. (1997) Disulfide-intact and reduced lysozyme in the gas phase: Conformations and pathways of foldng and unfolding, *J. Phys. Chem. B 101*, 3891−3900.
88. Myung, S., Badman, E. R., Lee, Y. J., and Clemmer, D. E. (2002) Structure transitions of electrosprayed ubiquitin ions stored in an ion trap over 10 ms to 30 s, *J. Phys. Chem. A 106*, 9976−9982.
89. Hukushima, K., and Nemoto, K. (1996) Exchange monte carlo method and application to spin glass simulations, *J. Phys. Soc. Jpn. 65*, 1604−1608.
90. Mao, Y., Ratner, M. A., and Jarrold, M. F. (2001) Molecular dynamics simulations of the rehydration of folded and unfolded cytochrome C ions in the vapor phase, *J. Am. Chem. Soc. 123*, 6503−6507.
91. Seibert, M., Patriksson, A., Hess, B., and van der Spoel, D. (2005) Reproducible polypeptide folding and structure prediction using molecular dynamics simulations, *J. Mol. Biol. 354*, 173−183.

BI061182Y