Molecular Dynamics Simulations of the Rehydration of Folded and Unfolded Cytochrome *c* Ions in the Vapor Phase

Yi Mao, Mark A. Ratner,* and Martin F. Jarrold*

Contribution from the Department of Chemistry, Northwestern University, 2145 Sheridan Road, Evanston, Illinois 60208

Received August 17, 2000

Abstract: Molecular dynamics (MD) simulations have been performed to study the rehydration of compact and unfolded cytochrome *c* ions in the vapor phase. Experimental studies have shown that the compact conformations adsorb many more water molecules than unfolded ones when exposed to water vapor. MD simulations performed with up to 150 water molecules reproduce the key experimental observations, including a partial refolding caused by hydration. According to the calculations it is more energetically favorable to hydrate the compact conformation in the initial stages of hydration, because it is easier for a water molecule to interact simultaneously with several polar groups (due to their proximity). The protonated side chains are not favored hydration sites in the simulations because they have "self-solvation" shells which must be disrupted for the water to penetrate. For both conformations, the adsorbed water molecules are mainly located in surface crevices.

Protein−water interactions play a central role in determining the stability, structure, and function of proteins.[1−3] Water not only interacts with the protein surface but also occupies internal cavities and deep clefts, and so one may speak of the external and internal hydration.[4] The effects of hydration depend on the number of water molecules involved and simulations have been performed to study protein behavior over a wide range of hydration levels.[5] Binding a few water molecules at specific sites can play a critical role in determining the specificity and function of proteins as catalysts and regulators.[6] The differential solvation of the hydrophobic core and hydrophilic surface of a protein is believed to play an important role in defining a native structure in solution. How protein−water interactions are affected by the conformation of a protein is important in understanding protein folding.[7]

Recently it has become possible to probe the hydration of peptides and proteins in the vapor phase,[8−16] and this approach can provide detailed information about the hydration process as a function of the number of adsorbed water molecules. The rehydration of compact and unfolded cytochrome *c* ions was examined in several recent studies.[10,13] In these experiments, the ions were produced by electrospray which generates a distribution of $(M + nH)^{n+}$ charge states. Cross section measurements have shown that the higher charge states of proteins unfold (due to Coulomb repulsion).[17,18] For cytochrome *c* the unfolding transition occurs around $n = 6$: the $(M + 5H)^{5+}$ ion remains folded in the gas phase while the $(M + 7H)^{7+}$ ion unfolds when heated. Free energy changes have been measured for the first few steps in the rehydration of the compact $(M + 5H)^{5+}$ ion and the unfolded $(M + 7H)^{7+}$.[10] The results indicate that adsorption of the first few water molecules is more favorable for the compact +5 conformation than for the unfolded +7. In addition, the average number of water molecules adsorbed under near saturation conditions has been measured as a function of charge state. There is a sharp *drop* in the average number of adsorbed water molecules (from ∼50 to ∼28) at the unfolding transition.[13] Thus both sets of results indicate that rehydration of the folded conformation is significantly more favorable than rehydration of the unfolded ones. In an effort to understand the origin of this behavior, we have performed molecular dynamics (MD) simulations of the rehydration of the compact $(M + 5H)^{5+}$ and unfolded $(M + 7H)^{7+}$ ions.

MD simulations were performed using the MACSIMUS molecular modeling software (Kolafa, J. http://www.icpf.cas.cz/jiri/macsimus/default.htm) employing CHARMM potentials (21.3 parameter set) (Brooks et al., 1983).[19] The water molecules are represented by the TIP3 model.[20] To model gas-phase

(1) Nemethy, G.; Scheraga, H. A. *J. Phys. Chem.* **1962**, *66*, 1773.
(2) Kuntz, P. T., Jr.; Kauzmann, W. *Adv. Protein Chem.* **1974**, *28*, 239.
(3) Finney, J. L. *Faraday Discuss.* **1996**, *103*, 1.
(4) Denisov, V. P.; Halle, B. *Faraday Discuss.* **1996**, *103*, 227.
(5) Steinbach, P. J.; Brooks, B. R. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 9135.
(6) Singer, P. T.; Smalas, A.; Carty, R. P.; Mangle, W. F.; Sweet, R. M. *Science* **1993**, *259*, 669.
(7) Dill, K. A. *Biochemistry* **1990**, *29*, 7133.
(8) Chowdhury, S. K.; Katta, V.; Chait, B. T. *Rapid Commun. Mass Spectrom.* **1990**, *4*, 81.
(9) Klassen, J. S.; Blades, A. T.; Kebarle, P. *J. Phys. Chem.* **1995**, *99*, 15509.
(10) Woenckhaus, J.; Mao, Y.; Jarrold, M. F. *J. Phys. Chem. B* **1997**, *101*, 851.
(11) Woenckhaus, J.; Hudgins, R. R.; Jarrold, M. F. *J. Am. Chem. Soc.* **1997**, *119*, 9586.
(12) Rodriguez-Cruz, S. E.; Klassen, J. S.; Williams, E. R. *J. Am. Soc. Mass Spectrom.* **1997**, *8*, 566.
(13) Fye J. L.; Woenckhaus, J.; Jarrold, M. F. *J. Am. Chem. Soc.* **1998**, *120*, 1327.
(14) Lee, S. W.; Freivogel, P.; Schindler, T.; Beauchamp, J. L. *J. Am. Chem. Soc.* **1998**, *120*, 11758.
(15) Rodriguez-Cruz, S. E.; Klassen, J. S.; Williams, E. R. *J. Am. Soc. Mass Spectrom.* **1999**, *10*, 958.
(16) Mao Y.; Ratner M. A.; Jarrold, M. F. *J. Am. Chem. Soc.* **2000**, *122*, 2950.
(17) Covey T. R.; Douglas D. J. *J. Am. Soc. Mass Spectrom.* **1993**, *4*, 616.
(18) Shelimov, K. B.; Clemmer, D. E.; Hudgins, R. R.; Jarrold, M. F. *J. Am. Chem. Soc.* **1997**, *119*, 2240.
(19) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus M. *J. Comput. Chem.* **1983**, *4*, 187.

10.1021/ja0030717 CCC: $20.00
Published on Web 06/15/2001

conditions the dielectric constant is set to unity, no cutoff is used in computing the electrostatic interactions, and we assume that basic and acidic sites are neutral (except for the basic sites that are protonated to account for the overall charge). There are 23 basic sites in cytochrome *c* and numerous ways of distributing the five or seven protons (required for the $(M + 5H)^{5+}$ and $(M + 7H)^{7+}$ ions) between them. The issue of how the charge is distributed has been examined by several groups.[21,22] We have used low-energy charge permutations and seed conformations derived in the same way as in our previous work on the thermal unfolding of unsolvated cytochrome *c*.[23] Related studies have been performed by Tapia and collaborators.[24] In the charge permutations that we focus on here, the following residues were protonated: K13, K27, H33, K39, and K55 in the +5 charge state and K13, K25, R38, K55, K73, R91 and K100 in the unfolded +7 charge state. The seed conformation for the +5 charge state resulted from a 480 ps MD simulation at 300 K starting from the crystal structure, while the +7 charge state was unfolded by a 480 ps simulation at 600 K (see ref 23 for details). Average collision cross sections calculated for these conformations (1360 Å^2 for the +5 and 2010 Å^2 for the +7) are close to values determined by ion mobility measurements (1200 Å^2 for the +5 and 2050 Å^2 for the +7). The +5 charge state is less compact than observed in the experiments, but the other +5 charge permutations examined in ref 23 had larger cross sections. For the +7, the other charge permutations had smaller cross sections. In addition to the specific charge permutations mentioned above, some simulations were performed using other charge permutations/seed conformations (K13, K27, H33, R38, K55 and K27, R38, K55, R91, K100 for +5 and K13, K27, H33, R38, K39, K55, R91 and K13, H26, H33, R38, K55, K73, R91 for +7). For the +5 charge state, where all the conformations have similar cross sections, the hydration behavior was not significantly affected by using different charge permutations/seed conformations. For the +7 charge state, where the other conformations are less compact, there were some significant differences. Generally the more-folded +7 conformations had hydration properties between those of the compact +5 and the unfolded +7. Other small differences will be mentioned below.

In the saturation experiment the dehydrated protein ions are exposed to water vapor at close to its saturation vapor pressure. It is not possible to simulate directly the equilibrium between adsorbed and vapor phase water under the conditions employed in the experiments (0.73 Torr at 253 K) because of the low water vapor pressures employed. Instead, we use an isolated protein ion surrounded by a shell of water molecules as the starting point for the simulations. For the compact +5 conformation we typically started with a shell of 70 water molecules, while for the unfolded +7 we typically started with 150 waters. The ratio of the number of water molecules to the surface area is about the same for both conformations. During the simulations some water molecules fly away from the protein and they are then removed from the system. The waters that leave are ones that start off in locations where they are not strongly bound to the protein. The simulations were run for at least 0.5 ns, after which there were about 60 (out of the initial 70) and 130 (out of the initial 150) water molecules left on the +5 and +7 charge states, respectively. In the experiments, an average of 50 water molecules adsorbed on the compact +5 charge state and an average of 28 adsorbed on the unfolded +7 charge state. The excess water molecules used in the simulations help to make up for their limited time scale. Figure 1 shows examples of the compact +5 and unfolded +7 conformations from the simulations. For clarity we show examples (see below) with fewer waters than described above: the +5 conformation is associated with 52 water molecules and the +7 conformation has 45 waters.

During the simulations we monitor the interaction between the water molecules and the protein by the water coordination number, N_c, which we define as the number of protein atoms within 3 Å of any atom in the water molecule. $N_c \geq 10$ indicates that the water is completely buried inside the protein, while $N_c < 1$ indicates that the water is not directly connected to the protein. If N_c is between 1 and 10 the water is either on the surface of the protein or in surface crevices. In the initial state (the protein ion surrounded by a shell of water molecules) N_c is ≤ 3 for all waters, and the average water coordination numbers are 1.5 and 1.2 for the +5 and +7 charge states, respectively. The water coordination numbers for the initial states and after the first 120 ps of the simulations are shown in Figure 2. For both charge states, the water coordination numbers increase as some of the water molecules gradually penetrate into the protein. The average coordination numbers after the first 120 ps are 2.8 and 2.4 for the +5 and +7 charge states, respectively.

After 120 ps, about 80% of the water molecules are involved in water−water interactions and form water clusters and chains throughout and around the protein (see Figure 1). The water molecules in the clusters and chains have at least one hydrogen bond with each other. The average N_c for the waters in the clusters and chains (2.4 for +5 and 2.2 for +7) is less than that in isolated water molecules. Obviously, there is a competition between water−water interactions and water−protein interactions. The water chains curve around the protein, following its surface. In some cases, additional water molecules, which are not in direct contact with the protein, cluster above the water chains. A few water chains are partially buried inside the protein. The buried waters usually have two or three hydrogen bonds to the protein. Several clusters of up to three water molecules are often completely buried inside the compact +5 conformations. Overall, the compact $(M + 5H)^{5+}$ and unfolded $(M + 7H)^{7+}$ do not differ much in terms of the amount of water-penetration into the protein or in the number of water−water interactions.

The average hydration enthalpy for water molecules in the chains and clusters was estimated from the difference between the energies of the hydrated and dehydrated conformations. The energies were obtained by minimizing (by steepest descent and conjugate gradient methods) the conformations taken from the MD simulations. The average hydration enthalpy (per water molecule) is around −46 kJ mol^{-1} for the water chains on the surface of the protein, around −59 kJ mol^{-1} for the partially buried water chains, and around −67 kJ mol^{-1} for the fully buried water clusters. The larger values are close to those recently measured for the initial steps in the rehydration of $(BPTI + 6H)^{6+}$ (bovine pancreatic trypsin inhibitor).[11]

When protein ions become dehydrated there are expected to be structural changes. MD simulations suggest that the main changes are due to self-solvation of the charged groups by oxygen atoms from backbone carbonyl groups and side chains.[23] When the water molecules are added back, typically one or two

(20) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.*, **1983**, *79*, 926.
(21) Schnier, P. D.; Gross, D. S.; Williams, E. R. *J. Am. Chem. Soc.* **1995**, *117*, 6747.
(22) Miteva, M.; Demirev, P. A.; Karshikoff, A. D. *J. Phys. Chem. B* **1997**, *101*, 9645.
(23) Mao, Y.; Woenckhaus, J.; Kolafa, J.; Ratner, M. A.; Jarrold, M. F. *J. Am. Chem. Soc.* **1999**, *121*, 2712.
(24) See, for example: Reimann, C. T.; Velazquez, I.; Tapia, O. *J. Phys. Chem. B* **1998**, *102*, 9344. Reimann, C. T.; Velazquez, I.; Bittner, M.; Tapia, O. *Phys. Rev. E* **1999**, *60*, 7277.

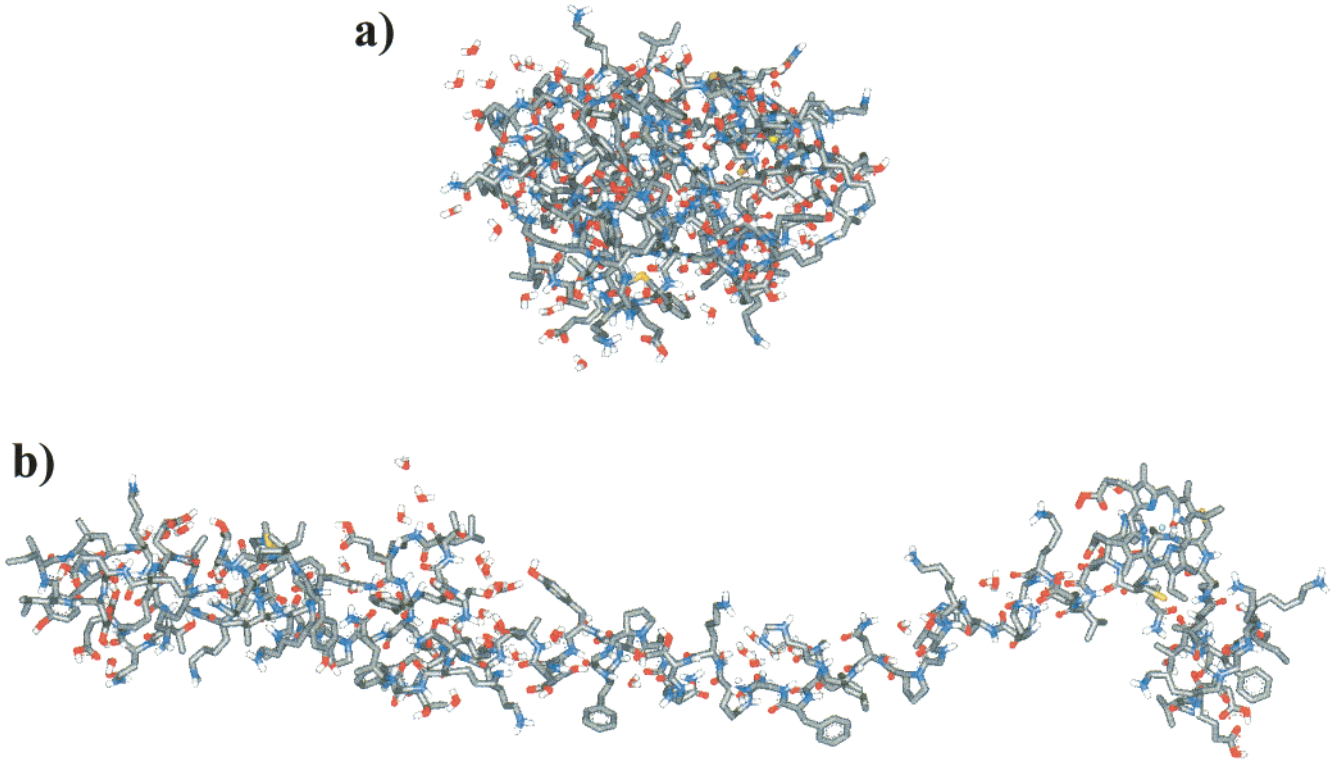

Figure 1. Examples of (a) the compact +5 and (b) unfolded +7 conformations from the MD simulations. In the examples shown here, the +5 conformation has 52 water molecules associated with it and the +7 conformation has 45.

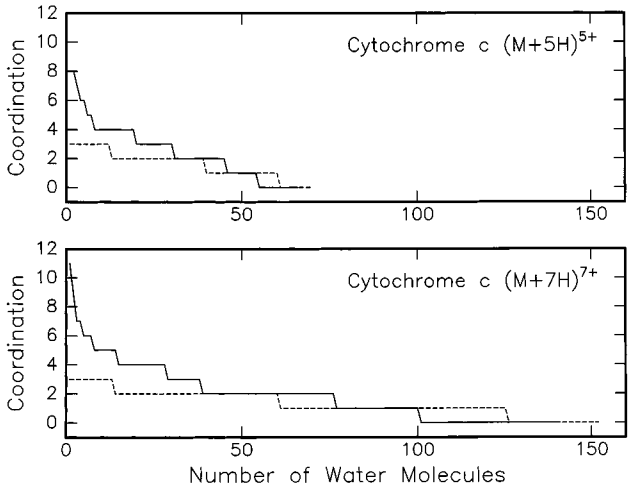

Figure 2. The coordination number of the water molecules at the beginning of the simulations (dashed lines) and at the end of a 120 ps run (solid line) for the compact +5 (top) and unfolded +7 (bottom) conformations. Note that only the distribution of coordination number is shown here and there is no correspondence between solid lines and dotted lines for individual water molecules.

of them resolvate each charged group. Some of the oxygen atoms in the original self-solvation shell remain while others are pushed away by the intrusion of the water molecules. The oxygen atoms that are pushed away often form hydrogen bonds to the intruding water molecules, so that the water ends up acting as a bridge. This behavior is observed for both the compact +5 and the unfolded +7 conformations.

Rehydration of the compact +5 conformation does not affect the secondary structure and the measured and calculated cross sections remain unchanged. Rehydration of the unfolded +7 conformation causes the cross section to decrease from 2010 Å^2 to 1900 Å^2 by the end of a 480 ps simulation. A corresponding decrease in the cross section is seen in the experiments, where addition of only 29 water molecules decreases the average cross section by around 220 Å^2. In the simulations, the decrease in the cross section results from the formation of a turn between LYS60 and LEU68. The turn is induced by a single water molecule. Each water hydrogen interacts with a pair of backbone carbonyl oxygen atoms (MET65O and GLU62O, and LEU64O and GLU61O). To facilitate the interactions with the water, the carbonyl oxygen atoms point in the same direction and the backbone twists to maximize hydrogen bond formation. This type of behavior, where a polypeptide chain forms a turn by bending back on itself to hydrogen bond with a water molecule, has been described previously.[25] It seems that in this initial stage of hydration, the water acts as a nonspecific glue which aids collapse, but otherwise plays little role in dictating the specific architecture of the protein. In simulations performed for other +7 charge permutations/seed conformations a decrease in the cross section was not observed. However, the other +7 charge permutations/seed conformations were initially more compact than the one that showed the collapse, and so the absence of further collapse is perhaps not surprising.

For the addition of the first few water molecules to a protein, water−water interactions can be ignored because the rehydration is driven mainly by water−protein interactions. Our calculations also show that for the adsorption of the first few water molecules, the average hydration enthalpy per water is more negative for individual water molecules than for water molecules involved in clusters. A procedure for determining energetically favorable hydration sites on a macromolecule has been described.[26,27] The basic idea is to calculate the hydration energy

(25) Robert, C. H.; Ho, P. S. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 7600.
(26) Goodford, P. J. *J. Med. Chem.* **1985**, *28*, 849.
(27) Boobbyer, D. N.; Goodford, P. J.; McWhinnie, P. M.; Wade, R. C. *J. Med. Chem.* **1989**, *32*, 1083.

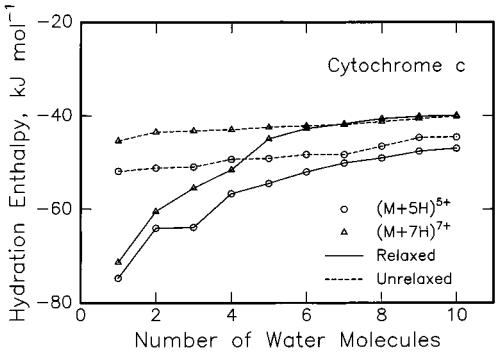

Figure 3. The hydration enthalpies of the 10 most favorable hydration sites for the compact +5 (circles) and the unfolded +7 (triangles) conformations. Results are shown for unrelaxed conformations (dashed lines) and after relaxation induced by hydration (solid lines).

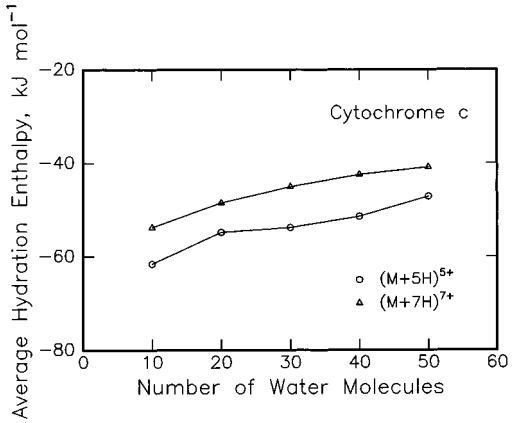

Figure 4. Estimated average hydration enthalpies for the compact +5 and the unfolded +7 conformations as a function of the number of adsorbed water molecules. The values were determined from MD simulations (see text).

as a function of the position of the oxygen atom as the water is moved throughout and around the protein. This type of approach has recently been used to identify the first hydration site on the +6 charge state of gas-phase BPTI.[28] Here, energy contours were generated for the +5 and +7 charge states of cytochrome *c* by computing the interaction between the protein ion and an individual water molecule on a 1 Å grid within 5 Å from the protein. To investigate the effects of structural relaxation upon hydration, conformations from MD simulations performed before and after the addition of water molecules were employed (with all waters removed). Only the positions of the water hydrogen atoms were optimized during these calculations.

Hydration enthalpies were estimated from the difference between the potential energies, after minimization, of the protein with and without water. The most energetically favorable site found for hydration of the unfolded +7 charge state is the heme iron, which is only five-coordinate in the unfolded conformation. The calculated hydration enthalpy of this site is −117 kJ mol^{-1}. The measured hydration free energies of the +7 charge state are not consistent with such a large hydration enthalpy,[10] which is probably the result of an artifact in the potential used for the calculations. For this reason we ignore hydration of the heme iron. The calculated hydration enthalpies for the 10 sites with the most negative hydration enthalpies (excluding the heme iron) are shown in Figure 3 for compact +5 and unfolded +7 conformations. Results are shown for both the unrelaxed conformations and conformations that were relaxed after hydration. The hydration enthalpies are larger for the relaxed conformations, and they are significantly more negative for the compact +5 conformation than for the unfolded +7 conformation. In the most favorable hydration sites found for both charge states, the water forms two or three hydrogen bonds with oxygen atoms and hydrogen atoms in the backbone or on polar side chains. The sites with large hydration enthalpies are usually in surface crevices where it is easier to form multiple hydrogen bonds and where there are not significant space constraints. The buried charged groups are not energetically favorable sites for hydration because they are already "solvated" by oxygen atoms from the backbone and side chains. The existing "self-solvation" shells must be disrupted for the water to penetrate. In this sense, the surface of a gas-phase protein ion is still more hydrophilic than the interior. The surface of the compact +5 conformation is more hydrophilic than that of the unfolded +7 conformation because multiple interactions with the water molecule are stronger in the compact conformation due to the proximity of the polar groups on the surface of the protein. As a result, the compact +5 conformation has lower hydration enthalpies than the unfolded +7. Thus the simulations reproduce the behavior observed in the experiments.[13]

The model employed above, an individual water molecule interacting with the protein, is valid when the number of water molecules adsorbed is small (so that the waters do not interact) and when the conformational changes associated with hydration are small and localized. If more than a few water molecules interact with the protein, significant conformational changes occur. These changes involve the reorganization of "self-solvation" shells around the charged groups and the gathering of hydrogen bond partners in the protein to maximize the interactions with the water molecule. This explains why the conformations that were relaxed in the presence of water have significantly more negative hydration enthalpies (see Figure 3).

For adsorption of more than a few water molecules, a more realistic estimate of the hydration energetics is obtained from MD simulations where the multiple protein−water interactions as well as water−water interactions are taken into account. The average hydration enthalpy per water is then estimated from the energy difference between the hydrated and dehydrated protein divided by the number of adsorbed water molecules. The initial conformations for these simulations were obtained as follows. Starting with the final conformations from the simulations described above with a large number of water molecules, we removed the waters that had the least contact with the protein until the required number of water molecules remained. Figure 4 shows the average hydration enthalpy per water molecule, calculated from these MD simulations, plotted against the number of water molecules. For at least the first 50 water molecules adsorbed, the average hydration enthalpy per water molecule obtained from MD simulations is more negative for the +5 charge state than for the +7 charge state. This result is in agreement with the experimental observation that more water molecules adsorb on the compact +5 conformation than on the unfolded +7 conformation. The physical origin of this behavior is that the multiple hydrogen bond potential of water molecules can be better satisfied in the compact conformation than in the unfolded one.

So far we have only considered the hydration enthalpy, not the hydration free energy. NMR studies have indicated that the entropy of buried water molecules engaged in three hydrogen bonds is comparable to that of the bulk water,[29] so the hydration entropies for different sites are probably quite similar, in general.

(28) Mao, Y.; Ratner, M. A.; Jarrold, M. F. *J. Am. Chem. Soc.* **2000**, *122*, 2950.

This is consistent with studies of the initial steps in the hydration of $(BPTI + 6H)^{6+}$. The measured entropy changes for adding individual water molecules to $(BPTI + 6H)^{6+}$ in the gas phase are close to the $\Delta S°$ for adding a water molecule to bulk water, except for the first water molecule adsorbed where $\Delta S°$ and $\Delta H°$ are both anomalously large. The special behavior of the first water molecule has been attributed to it being tetrahedrally coordinated to the protein at an internal site.[16] This type of tetrahedrally coordinated water was not found in our studies of cytochrome *c*.

Hydrating a gas-phase protein at low water vapor pressure is not equivalent to solvating a protein in the bulk solution. Previous simulations of myoglobin suggest that it is effectively fully hydrated by 350 water molecules (myoglobin with 153 residues is around 50% larger than cytochrome *c* which has 104). The hydration shell formed by these 350 waters is not a uniform monolayer but a patchwork of water clusters that hydrate the charged and polar groups; around 37% surface is left uncovered.[5] In our simulations hydration of the protonated side chains is not favored energetically because the "self-solvation" shell must be disrupted to accommodate the water. Even at high levels of hydration, only one or two waters interact with each charged group. The preferred hydration sites appear to be in water chains or clusters in surface crevices and to some extent in the interior of the protein. The penetration of water into the protein interior is much easier in the gas phase than in solution due to lack of strong water−water interactions in the gas phase.[30] An issue that we have not considered is to what extent addition of water to the protein promotes the ionization of acidic and basic groups. It has been suggested that this occurs at very low water coverage in dehydrated crystals, where it is also assumed that the ionized groups are the most favorable sites for initial hydration.[31] In our simulations, the charged groups are not favored sites for initial hydration because they are self-solvated and to incorporate a water molecules into the solvation shell, other groups must be displaced. If self-solvation is taken into account, it is not clear that the acidic and basic groups should be expected to ionize at low coverage.

In the studies described here we focused mainly on one charge permutation/seed conformation for each charge state. We also examined (in less detail) two additional charge permutations/seed conformations for each charge state, and the results were not significantly different (except for differences that could be attributed to varying degrees of unfolding). This suggests that the main features of the results do not depend on the precise conformations and charge permutations. There are other examples of this behavior. In our previous studies of the charge-induced unfolding of gas-phase cytochrome *c*, it was found that the total number of protons, rather than their precise location, was the main factor determining the conformation (unfolded or compact). The reason the collective behavior is not strongly influenced by the microscopic details is that proteins are large enough that statistical averaging plays a role. In the present case, the observation that hydration of the compact conformations is more favorable than hydration of the unfolded ones is attributed to the polar sites being further apart in the unfolded conformations, making it more difficult to form multiple hydrogen bonds to water. It is possible that in a few places the polar sites move closer together when the protein unfolds, leading to regions of enhanced water binding. But on average the polar sites move further apart and it is this averaging that makes our observations robust and largely independent of the details of the conformations. This behavior is quite different from that found with small molecules where the precise location of the atoms is often an issue of paramount importance.

In conclusion, MD simulations of the rehydration of compact and unfolded conformations of cytochrome *c* show that the initial hydration properties of gas-phase proteins are strongly influenced by the multiple hydrogen-bonding potential of water. With the compact conformation, the close proximity of polar groups on the surface facilitates the formation of multiple hydrogen bonds, thus it is more energetically favorable to hydrate the compact conformation than the unfolded one. Overall there is good qualitative agreement between the predictions of the simulations and previous experimental results, including the presence of a water-induced conformational change, which in the simulations results from a single bridging water molecule inducing a loop. Hydration of the protonated side chains disrupts their self-solvation shells, so these are not energetically favored hydration sites. The most favorable hydration sites appear to be in surface crevices which accommodate chains or clusters of water molecules.

Acknowledgment. We are grateful to the National Institutes of Health for partial support of this work. We are also grateful to Dr. Jiri Kolafa for the use of his MACSIMUS molecular modeling program.

JA0030717

(29) Denisov, V. P.; Venu, K.; Peters, J.; Horlein, H. D.; Halle, B. *J. Phys. Chem.* **1997**, *101*, 9380.
(30) Wolfenden, R.; Radzicka, A. *Science* **1994**, *265*, 936.
(31) Rupley, J. A.; Careri, G. *Adv. Protein Chem.* **1991**, *41*, 37.