Molecular Basis for the pH Dependent Structural Transition of Nitrophorin 4

Marcelo A. Martí,*,†,‡ Dario A. Estrin,† and Adrián E. Roitberg§

Departamento de Química Inorgánica, Analítica, y Química Física, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, INQUIMAE-CONICET, Buenos Aires, Argentina, Departamento de Química Biológica, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Ciudad Universitaria, Pab. 2, C1428EHA, Buenos Aires, Argentina, and Quantum Theory and Project and Department of Chemistry, University of Florida, Gainesville, Florida 32611-8435

Received: September 10, 2008; Revised Manuscript Received: December 10, 2008

Allostery can be defined in a broad sense as a structural change in a protein. The theoretical framework for allostery includes several formulations. In the stereochemical view, the activation event causes a local conformational change that is propagated through residue-to-residue contacts to the rest of the protein through well-defined structural pathways. The thermodynamic, or population shift model, instead implies that the "activated conformation" is already present with non-negligible population in the nonactivated conformational ensemble, and therefore the activation merely shifts the equilibrium. Nitrophorins (NPs) are heme proteins that store and transport NO in a pH dependent manner, due to a conformational change. Using MD simulations, we show that the NP structural transition occurs in two different conformational free energy landscapes, each one corresponding to a pH condition and characterized by specific residue−residue interactions that characterize them. We also show that when the protonation state of the equilibrium state is modified the conformation becomes unstable and proceeds very fast to an intermediate stable state that is different for each pH condition. Finally, we will discuss that allosteric transition in NP4 does not occur due to a change in the relative population of both end states, but due to a drastic change in the free energy landscape of its conformational ensemble.

Introduction

Allostery can be defined in a broad sense as a structural change in a protein. In many cases, the structural changes are located in regions far apart from the activation sites.[1] Understanding how such local modifications are propagated to the rest of the protein leading to significant transitions in its structure and activity is of great value from the fundamental and biomedical viewpoints.[2] The theoretical framework for thinking about allosteric transitions includes thermodynamic[3,4] and stereochemical formulations.[5] Mechanistically, in the stereochemical view, the activation event causes a local conformational change that is propagated through residue-to-residue contacts to the rest of the protein. The propagation occurs through well-defined structural pathways and involves specific aminoacids.[5,6] The other scheme is also known as the population shift model[7,8] and implies that the "activated conformation" (the dominant conformation after the allosteric transition) is already present with non-negligible population in the nonactivated conformational ensemble, and therefore the activation merely shifts the equilibrium.[1] This view has gained considerable support given the observed sensitivity of the ensemble to changes in the environmental conditions (i.e., ligand binding, pH, temperature).[7] It should be noted, however, that both formulations are not necessarily excluding each other, and some specific characteristics of each model may operate at the same time in some cases. The most relevant implication of the population shift model is that in order to populate the activated state, protein equilibrium dynamics must include motions related to the allosteric transition. Therefore, the nature of the transition could in principle be predicted from the equilibrium dynamics of the protein in the deactivated state. [1,2,9,10]

The relationship between protein dynamics and function has been extensively studied in the past years with the main focus on understanding how protein motion determines and assists enzymatic function.[11−13] Understanding protein dynamics is a difficult task due to the enormous number of degrees of freedom involved. However, there are several computational methods that allow rationalizing protein dynamics in terms of a few collective motions such as normal mode analysis (NMA)[12,14−17] and essential dynamics (ED) or quasi-harmonic analysis.[18] Recently it was shown that the deformational space with either approach is quite similar if a sufficient number of modes is taken into account. In this context, it has recently been reported that the space described by ED is wider but simpler, so fewer modes are needed to describe the conformational space.[19] As mentioned above, in the thermodynamic model of allostery the essential motions (EM) of the protein are expected to play a predominant role by allowing exploration of a wide conformational space. However, to our knowledge only few studies have focused on the relation between protein essential motions and allostery, and all of them were based on a NMA.[1,2,9,10] In this scenario, a comparative analysis of the conformational space explored during an allosteric transition, with the one explored during equilibrium dynamics, is a key missing element in the above-mentioned theoretical framework. To bridge this gap we decided to study the pH dependent structural transition of Nitrophorin 4 (NP4).

Nitrophorins (NPs), salivary proteins of blood-sucking insects,[20] are heme proteins that store and transport NO, which when released in the victim's tissue produces vasodilatation and

* To whom correspondence should be addressed. E-mail: marcelo@qi.fcen.uba.ar.
† Departamento de Química Inorgánica, Analítica, y Química Física, Universidad de Buenos Aires.
‡ Departamento de Química Biológica, Universidad de Buenos Aires.
§ University of Florida.

10.1021/jp808055e CCC: $40.75
Published on Web 01/26/2009

inhibition of blood coagulation.[21] Some of the most studied NPs (NP1−4) are those obtained from the salivary glands of the kissing bug *Rhodnius prolixus*. They have been cloned,[22] characterized spectroscopically and kinetically,[21,23] and three of them also structurally[24−27] with the most intensively studied being NP4.

These studies showed that NPs function by binding NO in a pH sensitive manner. NO is synthesized in the salivary glands by a typical NOS[28] where it binds tightly to NP at a low pH of around 5. Once NPs are injected in the victims blood stream with a pH of approximately 7.5, a conformational change occurs that allows NO to be released.[20,23] The X-ray structural data showed that all 4 NPs have a conserved structure consisting of an eight-stranded antiparallel β-barrel typical of the lipocalin family.[29] The heme group is tightly buried in the barrel, contacting the invariant heme proximal ligand His59.[24,25,27,30] The heme side end of the barrel is capped by the four loops AB, CD, EF, and GH. High resolution structures of NP4 in different pH conditions reveal that the large pH-induced conformational change involves mainly loops AB and GH, which opens at high pH to release NO. Other interesting data on NP4 is that it shows multiphase ligand kinetics, which is an indication of the existence of multiple protein conformations.[23,30,31]

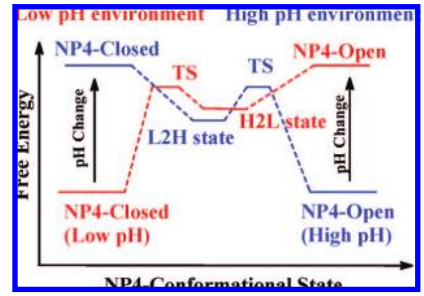
The relationship between pH, NP4 structure, and NO release was also studied using computer simulation techniques by several groups[32] including our own.[33] The results showed that differential protonation of Asp30 and possibly Asp35 are key elements of the pH dependent structural transition.[32] Keeping this difference is enough for maintaining the open and closed conformations respectively during molecular dynamics (MD) simulations.[33] Furthermore, our previous results showed that NO is only able to escape in the high pH open conformation and that several differential residue−residue interactions are characteristic of each conformational state. Among them the experimentally observed Asp30-Leu130 hydrogen bond (HB) that contributes to maintaining the AB and GH loops closed at low pH condition and that is broken upon Asp30 deprotonation, triggering the allosteric transition.[33]

In this work, we will use MD simulations to show that the NP structural transition occurs in two different conformational free energy landscapes, with each one corresponding to each pH condition. We will demonstrate this by showing that the conformational space explored by NP4 in the stable states, "open high pH" and "closed low pH" conformations during the MD simulations are significantly different and have specific residue−residue interactions that characterize them. We will also show that when in any of the stable states, the protonation state is modified to that of the corresponding target pH and the starting structure is unstable and proceeds very fast to an intermediate stable state. These intermediate states are different from each other and correspond to intermediate conformations along the transition. They are characterized by having lost the specific residue−residue interactions that characterize the end state but not having formed the new target structure specific interactions (as shown schematically in the resulting Scheme 1). Lastly, we discuss how the studied NP4 allosteric transition is related to the "population shift" and stereochemical model of allostery.

Computational Methods

System Setup and Equilibration. The initial structure for the MD simulations was built starting from the crystal structures of NO bound NP4 (pH = 5.6) pdb code 1x8o and NH_3 bound

SCHEME 1



NP4 (pH = 7.4) pdb code 1x8p.[25,30,31,34] These structures present the closed low pH and open high pH conformation of AB and GH loops, respectively. Special attention was paid to the protonation state of the titratable residues since they are a key issue in modulating AB and GH loop conformation. The protonation state was assigned as suggested from previous experimental[30] and theoretical works.[35] The two residues with differential protonation states are Asp30 and Asp35, which are both protonated at low pH and deprotonated (charged) at high pH. As shown previously, this difference is enough to maintain the open and closed conformations along the simulation.[33]

To study further the transition between the high and low pH structures, we performed two additional simulations. The first was called low-to-high (L2H) and corresponds to a MD simulation starting from the low pH structure, but in which Asp30 and Asp35 have been deprotonated. The second simulation is called high-to-low (H2L) simulation, and the starting structure corresponds to the high pH structure, but Asp30 and Asp35 are protonated.

To set up the systems for the MD simulations, the above-mentioned structures were immersed in a pre-equilibrated octahedral box of TIP3P water molecules. All simulations were performed at 1 atm and 300 K, maintained with the Berendsen barostat and thermostat,[36] using periodic boundary conditions and Ewald sums (grid spacing of 1 Å) for treating long-range electrostatic interactions with a 10 Å cutoff. The SHAKE algorithm was used to keep bonds involving H atoms at their equilibrium length.[37] A 2 fs time step was used for the integration of Newton's equations. The Amber ff99SB[38] force field parameters were used for all residues, except the heme. The heme parameters used in this work were developed and thoroughly validated by our group in previous works.[39,40] All simulations were performed with the PMEMD module of the AMBER9 program.[38,41] Equilibration protocols consisted in performing an initial optimization of the initial structures, followed by a slow heating up to the desired temperature. The heating was performed in 200 ps of constant volume MD, followed by 200 ps of MD at constant pressure. Once the system was equilibrated, different production 50 ns long MD simulations were performed. Frames were collected at 1 ps intervals, which were subsequently used to analyze the trajectories.

Essential Dynamics Analysis. In order to get insight into the dynamical properties of each structure and their influence on the coordination transitions, several essential dynamics (ED) analysis were performed for all production MD runs using the ptraj module of the AMBER program.[18] The ED for each run (or combination of runs) is determined by diagonalizing the coordinate covariance matrices (Cov_{ij}) of the backbone atomic positions along the desired trajectory, obtaining the corresponding eigenvalues and eigenvectors (eq 1)

$$\mathrm{Cov}_{ij} = (1/M)\sum_{k}^{M}[X_i(k) - \langle X_i\rangle][\{\{X_j(k) - \langle X_j\rangle\}\}] \quad (1)$$

where the sum goes over the "M" configurations or snapshots from the dynamics, $X_i(k)$ corresponds to i^{th} Cartesian coordinate of the system in snapshot number k, $\langle X_i\rangle$ represents the mean value of X_i along the MD simulation.

After diagonalization, each obtained eigenvector ($\boldsymbol{\nu}_i$) corresponds to an EM of the protein. Together all the EM describe the motions of the protein along the MD run used to generate the computed matrix. The eigenvalues (λ_i) obtained represent the relative contribution of each EM to the overall motions with the first EM being the one with major contribution or larger eigenvalue. The EM for each individual MD run were computed and analyzed. Then the EM of combined trajectories are computed to gain deeper insight into the structural transition between both simulations. In this combined MD analysis, the first EM (V_1) is able to extract the concerted movement corresponding to the transition between both states, used to compute the covariance matrix and will be called a transition EM.

Projections "$P_N(t)$" along the MD runs onto selected EM were also performed to analyze the configurational space explored along the MD run (eq 2)

$$P_N(t) = \nu_N r(t) \quad (2)$$

where $\boldsymbol{\nu}_N$ is the N^{th} EM, and $r(t)$ is the protein conformation at time t. Projections are measured in angstroms, and the value corresponds to the overall deviation from the mean structure along the projected mode. After projecting the selected MD simulation along the selected mode, the normalized histograms were computed.

Comparison between different pairs of EM were performed using the similarity index (SI), which is the scalar product between both the eigenvectors corresponding to two selected EM (eq 3)

$$\mathrm{SI}_{\mathrm{A}_i\mathrm{B}_j} = \nu_{\mathrm{A}_i}\nu_{\mathrm{B}_j} \quad (3)$$

where A_i corresponds to EM number "i" of MD simulation "A". A value of 1 means both modes are identical, whereas a value of zero indicates orthogonal modes. To have a global comparison of the dynamics for different protein states (A and B), a global SI ($\mathrm{SI}_{\mathrm{AB}}$) was computed as the sum of a desired number of SIs weighted by their eigenvalue and normalized by the total contribution of the compared modes to the overall dynamics (eq 4)

$$\mathrm{SI}_{\mathrm{AB}} = (1/\Sigma\lambda_i)\Sigma\,\mathrm{SI}_{\mathrm{A}_i\mathrm{B}_j}(\boldsymbol{\lambda}_i) \quad (4)$$

where $\mathrm{SI}_{\mathrm{AB}}$ in this case represents the similarity between MD A and B, and $\boldsymbol{\lambda}_i$ is the mean of the eigenvectors of A and B simulations for the i^{th} EM.

On the basis of the comparison of the equilibrium and transition EM, a slightly modified version of the involvement coefficients (IC) of each EM to the transition are computed (eq 5)[10]

$$\mathrm{IC} = \upsilon_{\mathrm{Aj}}\bullet\mathrm{V}_{\mathrm{AB}} \quad (5)$$

In equation 5, $\boldsymbol{\nu}_{\mathrm{A}_j}$ is the j^{th} EM of equilibrium simulation of the A state, whereas V_{AB} corresponds to the transition EM computed by combining the A and B trajectories. In this case, IC expresses how much of the structural transition between states A and B is contained in the EM number "j" of MD corresponding to the A state. This method therefore allows the analysis of how much of the natural motions of the protein in a given state contribute to the selected conformational transition.

This type of analysis based on the essential dynamics has already proven useful in the study of the structural and dynamical relationships in several proteins including the study of structural transitions.[7,9,11,16−19]

Results

We first analyze the dynamics of both stable end states at low and high pH. Second, we analyzed the dynamics of the intermediate states, L2H and H2L, and compared them to the stable states. Finally, the relation of each conformational ensemble is analyzed in relation to the pH dependent structural transition, and the results are discussed in relation with the population shift and stereochemical models of allostery.

MD Simulations of the Low and High pH Protein States with NO Bound. We start by analyzing the structural and dynamical characteristics of NP4 in both end states, corresponding to the low and high pH conditions by analyzing a 50 ns MD simulation for each case. As already mentioned in the Introduction, these simulations are an extension of the simulations presented in our previous work, where it was shown that only the high pH "open" conformation was able to release NO into the solvent. Figure 1A shows superposition between the average structures of both high and low pH states. Figure 1B shows the RMSD in the CA positions between the high and low pH average structures along the sequence.

As can be seen from Figure 1, the differences between the low and high pH structures are highly localized in very small segments, which correspond mainly to loop regions. These differences are consistent with those observed in the corresponding X-ray structures.[30] The most relevant differences are observed for residues located in the AB, CD, GH, and α2α3 loops. These regions include the previously identified key residues that open the NO exit path Leu130 and Leu133 in the GH loop and Val36 and Pro37 in the AB loop. The low CA-RMSD against each corresponding average structure along the 50 ns (see Figure SM1 in Supporting Information) shows that each conformation is stable during the whole simulation. Notably, the structural differences between both states' average structures are significantly larger than the fluctuations along the MD simulation of each state, as shown in Figure 1B.

As mentioned in the introduction, there are several key residue−residue interactions that define and characterize each conformational state (Table 1). One of these critical interactions is that between Asp30 (located in the AB Loop) and Leu130 (in the GH Loop). At low pH, Asp30 is protonated and forms a tight HB with the carbonyl of Leu130. When the pH is increased and Asp30 becomes charged, the HB is broken, and it moves into the solvent. This also promotes a rotation of Leu130, leaving space for NO to escape into the distal cavity. In this conformation Leu130 carbonyl forms a strong HB with the NH_3^+ moiety of Ala1.[33] This differential interaction is strictly maintained along the whole simulation length (Figure SM2A,B in Supporting Information). The other key residue in the AB loop is Asp35. At low pH, the neutral Asp35 can be seen to be temporarily buried and interacting with Asp129 bringing together the AB and GH loops and closing the NO escape path.

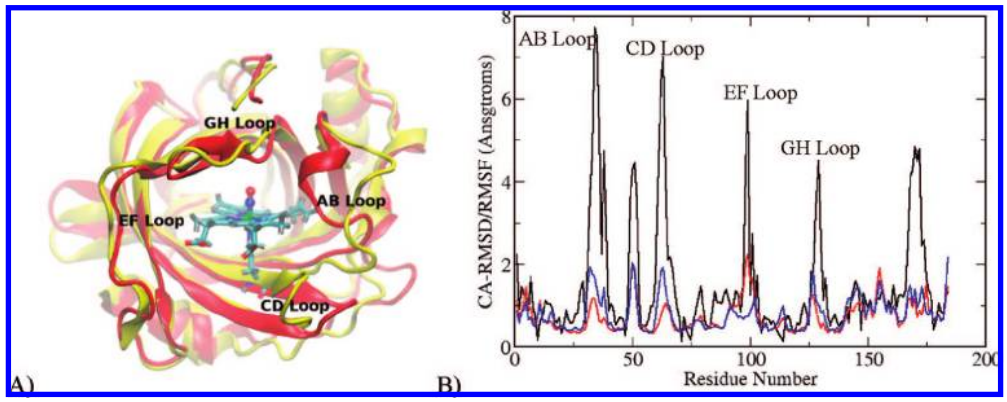


Figure 1. (A) Structural comparison between the average structures corresponding to low pH (red), and high pH (yellow) MD simulations, (B) RMSD vs residue plot between low and high pH average structures obtained from the 50 ns MD simulation (black line) and of the RMSF vs residue plot data for the low pH (blue) and high ph (red) simulation.

TABLE 1: Average Distance (Angstrom) during the MD Simulation for the Characteristic Interactions in NP4

interaction	low pH	high pH
Leu130-Asp30	1.8	9.6
Leu130-Ala1	6.6	2.9
Asp129- Asp35	7.4	13.2
Val36-HemeCD	2.0	9.0
Leu133-Val36	5.5	9.0

When the pH is raised, Asp35 becomes charged and fully solvated, maintaining the open conformation of the AB loop.

After NO is released from the heme by thermal breaking of the Fe−N bond, it remains located in the distal cavity. To exit the protein, it must pass next to Leu133 (GH loop) and then through a tunnel lined on one side by the heme CD side and by Val36 and Pro37 (AB loop) on the other. At low pH, there are close interactions between Leu113 and Val36 and between the heme CD side and Val36 Pro37 residues (Table 1) closing the exit path. After the low to high pH transition, these interactions are lost, leaving enough space for NO to pass between these residues and allowing it to escape (Figure SM2B−D in Supporting Information).

To further analyze both end states from a dynamical viewpoint, we analyzed the essential dynamics of both simulations. To calculate the EM we used the last 40 ns, in which the protein remains dynamically stable. The ED analysis shows that more than half of the overall protein motions can be described with the first six EM (Ssee Table in Supporting Information). For all these EM, motions are mostly localized in the loop flexible regions, as expected. Histograms of the projections of the low and high pH simulations along the six EM show that for the low pH simulation the first two EM present two conformations, whereas the others have a Gaussian-like distribution. The first and second EM also show a wider distribution than the others. In the high pH simulation, only the first EM presents bimodal wider distribution. Analysis of the projection along the MD simulation shows that in these modes in which two conformations are involved more than one transition is observed, suggesting that extensive exploration of the modes have been accomplished along the MD simulation. The multiple conformations revealed by the EM analysis are consistent with the kinetic data mentioned in the introduction, which suggested the existence of multiple conformations for NP4.

MD Simulations of the Transition (Intermediate) States. To study the transition between the high and low pH structures, we performed two simulations. The first was called low-to-high (L2H), which corresponds to a MD simulation starting from the low pH structure but Asp30 and Asp35 have been deprotonated. On the other hand, in the so-called high-to-low (H2L) simulation, the starting structure corresponds to the high pH structure but Asp30 and Asp35 are protonated.

The first characteristic of both the L2H and H2L simulations is that they equilibrate very fast and remain stable along the whole simulation (as evidenced from the RMSD versus time plot shown in Figure SM3 in Supporting Information) in intermediate structures between those that characterize the closed low pH or open high pH characteristic structure, as shown in Figure 2A. This is confirmed by comparing the RMSD between the average structures of the four MD simulations as shown in Table 2. Both end states present the largest structural difference, as judged from the RMSD of the whole protein, or only the mobile loops. As judging from the overall RMSD, it is not clear whether the transition structures are closer to any of the end structures. However, when looking only at the loop RMSD they seem to be slightly closer to the target state. These data strongly suggest that along the simulation the L2H and H2L states are trapped in between the conformational transition, and therefore they can be thought as intermediate states.

As mentioned above, there are several contacts between residues located in the AB and CG loops that characterize and maintain the conformation of each state. Figure 2B shows the corresponding interactions along the L2H, and H2L simulations. For comparison, the equilibrium distances values for the interactions in each end state are also shown as red (low pH) and black (high pH) arrows.

In the L2H transition (left panel red lines of Figure 2B), it can be seen that immediately after equilibration Asp30 moves away from Leu130 to an intermediate position (about 6 Å distance) where it remains for about 12 ns, then it moves completely away into the solvent (like in the open conformation) during the rest of the simulation. This exit movement of Asp30 allows Leu130 carbonyl to form the proper contact with the with the N-terminal end of Ala1 at about the same time. On the contrary in the H2L simulation (Figure 2B right panel black line) no change toward the low pH structure is observed for the Leu130 interactions for about 40−45 ns, during that time it remains bound to Ala1 and away from Asp30. Only in the last 10 ns Asp30 comes closer and gets buried and in the last 5 ns the Leu130-Ala1 HB starts weakening. Probably a much longer simulation time is needed for a full change in the interaction pattern.

The interaction between Asp129 and Asp35 in the L2H transition is broken frequently during the simulation, showing that once Asp35 becomes charged it moves away into the solvent

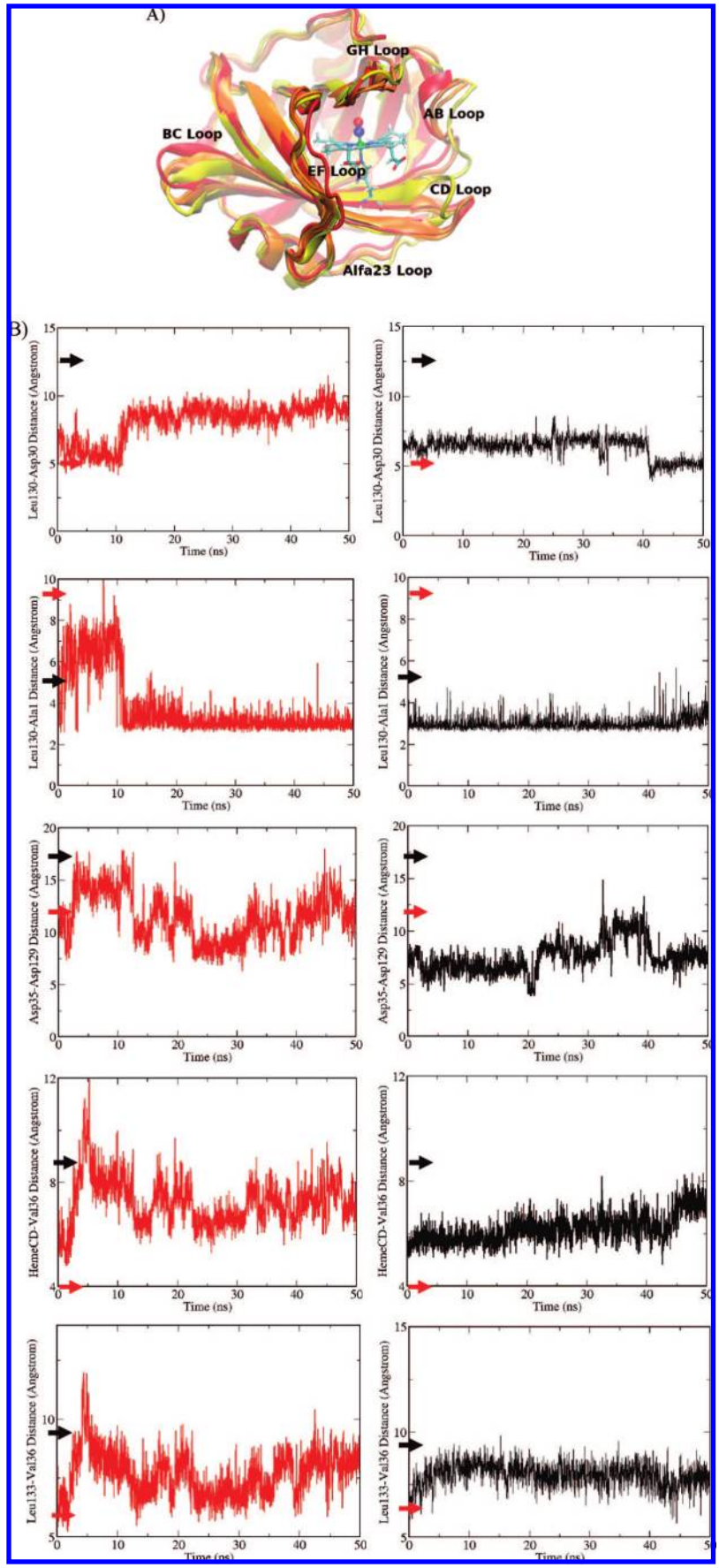


Figure 2. (A) Structural comparisons of low pH, high pH, L2H, and H2L average structures. (B) Distance vs time plot along the L2H pH simulation (left panel red line). Distance vs time plot along the H2L pH simulations (right panel black line). For both cases, the red arrow indicates low pH equilibrium value, and the black arrow indicates high pH equilibrium value. Distances correspond to Leu130CO-Asp30COO, Leu130COO-Ala1NH_3^+, Asp129COO^-Asp35COO Val36CA-HemeCD, and Val36-Leu133 sidechains.

TABLE 2: RMSD between the Corresponding Average Structures, Backbone RMSD Upper Right, AB, CD, EF and GH Loops Power Left Triangle (All Values Are in Angstrom)

	low pH	high pH	L2H	H2L
low pH		1.99	1.38	1.34
high pH	3.53		1.44	1.38
L2H	2.24	2.48		1.02
H2L	1.94	2.68	1.43	

TABLE 3: Global and First EM SI for Comparison of the Different Protein States[a]

	Low pH	High pH	L2H	H2L
Low pH		0.32	0.22	0.32
High pH	0.40		1.16	0.22
L2H	0.27	0.03		0.06
H2L	0.50	0.39.	0.02	

[a] Upper right global SI, Lower left SI fort the 1st EM.

immediately. However, it does not remain stably exposed and comes closer to Asp129 several times along the simulation. In the H2L, although the now protonated Asp35 comes closer to Asp129, a proper HB is not formed during the time scale of the simulation. Interestingly, both simulations show a high mobility between the two extreme conditions. Last but not least, Val36 in L2H simulation moves away from the Heme and Leu133 and exhibits significant fluctuations, again showing the mobility of this conformation. In the H2L Val36 comes closer to the Heme and to Leu133 but is still further that in the low pH conformation, remaining stable in this intermediate distance along the simulation.

The results show that in both cases the interactions that should be lost upon pH change are lost very fast, except for Leu130-Ala1, but that the target structure interactions need more time to be firmly established. In most of the cases, they appear temporarily or intermediate states are observed. This data together with the RMSD values presented above further confirms that in both cases the transition simulation result in intermediate structures along the structural transition.

We also performed an ED analysis of the L2H and H2L MD simulations. As in both end state simulations around 65% of the overall protein motions are captured in the first six EM. (Table S1 in Supporting Information). The four sets of EM allow a final comparison of the different protein states from a dynamical perspective. To do this comparison, the SI were computed for all of the EM in combination of protein states. The global index for all pair of structures, together with the SI for the comparison of the first EM are shown in Table 3

The data in Table 3 shows that from a dynamical viewpoint low pH and high pH states share about 30% of the dynamical behavior, and their SI for first and second EM is about 0.4. Interestingly, the least similar states are the L2H and H2L states with less than 10% global SI.

Connection of the L2H and H2L Intermediate States with the pH Dependent Structural Transition. To complete the picture of the pH dependent structural transition and analyze the contribution of each state conformational ensemble to it, we performed an ED analysis of selected combined trajectories in order to obtain the transition modes. Then we projected the trajectories for each state on to these transition modes and analyzed the histogram of the resulting time series.

We started by computing the EM for the combined low pH and high pH end state trajectories. As expected, in this case the first EM corresponds to the low to high pH transition, it accounts

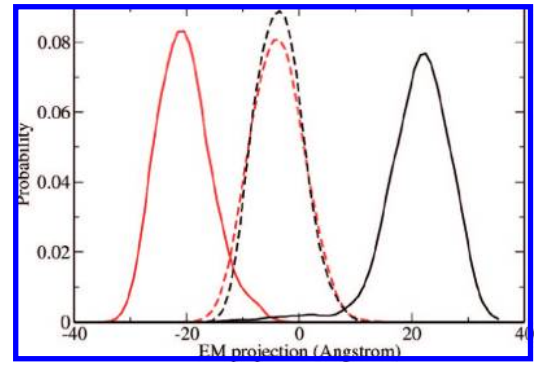


Figure 3. Histogram of the first EM for the low to high pH structural transition. Solid red line, low pH simulation; black solid line, high pH simulation; red dotted line L2H simulation; black dotted line, H2L simulation.

for 63% of the motion of the concerted run, and the RMSD versus residue plot along the mode shows the same pattern as the RMSD versus residue plot between average structures (Figure SM4 in Supporting Information). Projection of both trajectories along the EM shows that the mode is able to distinguish both structures clearly. Using this EM we also projected both transition runs; the histogram for the EM projections along the four MD simulations is shown in Figure 3. This data shows that both low and high pH simulations are far apart with no conformational overlap along the simulations.

The transition simulations show however some very interesting results. First and consistently with the average structure RMSD comparison, both L2H and H2L simulations are in the middle of the structural transition; moreover in both cases they explore very similar conformational space along the transition described by both end states. The two intermediate structures, as shown by the interactions analysis above, represent an ensemble which is a kind of mixture of both end states which is shown by Figure 3. The other interesting point is that by combining the low and high pH simulations with either of both transition trajectories some overlap is observed. Also noteworthy is the fact that the high pH structure is the one that is furthest away from the others and with a wider histogram, a fact that is consistent with a looser conformation.

Given that both transition dynamics (L2H and H2L) display a low RMSD between their average structures but different interactions and different dynamical behavior as suggested by the EM comparison, and given the fact that both display a very similar conformational space distribution along the complete structural transition, the question arises of whether they represent the same or different conformational ensemble and how this ensemble is related to the transition. To shed light on this issue, we computed the ED for the combined L2H and H2L trajectories.

As expected, the first EM captures the transition, and represents 42% of the combined structural variation. Interestingly when projecting this EM on the four trajectories and analyzing their population, now both L2H and H2L states are well separated with both low and high pH states populations located at intermediate values, as shown in Figure 4.

This clearly shows that the L2H and H2L states display different conformational distributions, but their difference is not along the low to high pH transition. To further confirm that the difference between the L2H and H2L states is not in the direction of the low to high pH transition, we computed the SI between the EM for the low to high transition. The L2H to H2L transition yielded a value of 0.006, clearly demonstrating that both transitions are not related.

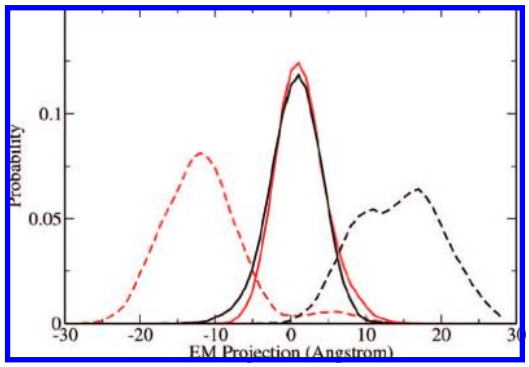


Figure 4. Histogram of the first EM for the L2H and H2L combined trajectories. Solid red line, low pH simulation; black solid line, high pH simulation; red dotted line, L2H simulation; black dotted line, H2L simulation.

TABLE 4: Involvement Coefficients (IC) for the Six EM, Corresponding to the Protein State Dynamics for the Low to High pH Conformational Transition

System	Low pH	High pH	L2H	H2L
1st EM	18.5	25.3	13.5	12.4
2nd EM	10	6.6	12.8	9.9
3rd EM	18.6	37.7	16.7	25.7
4th EM	25	28	46.3	17.3
5th EM	13	4.1	5	30.4
6th EM	19.9	6	16.3	6.4
$\sum IC_i\lambda_i$[a]	17.2	20.3	17.6	14.8

[a] Overall contribution to the transition computed as the sum of each IC multiplied by its weight to the overall dynamics (λ).

Finally, in order to analyze the contribution of the equilibrium protein motions to the pH dependent structural transition we computed the involvement coefficients for the six EM of each state to the transition. For this, we computed the SI between the EM for a given protein state to the transition EM obtained from the combined low and high pH trajectories; the results are shown in Table 4.

The data from Table 4 shows that all analyzed EM contribute to the structural change. No single EM contributes more than 40% (except mode four in the L2H state). Moreover, the overall contribution of the fluctuations to the structural transition is also small. The fact that the structural change is not described by one or two EM but is rather uniformly distributed in all EM suggests that the transition is not just a concerted motion of some secondary structure elements but a complex conformational change involving motions of several protein regions.

Discussion

Building a Model for the NP pH Dependent Structural Transition. As mentioned in the Introduction, the theoretical framework for thinking allosteric transitions includes thermodynamic[3,4] and stereochemical formulations,[5] also discussed as "new" population shift or "old" induced fit mechanism.[1] Briefly, the key differences between them concerns whether there exists a significant population of the target structure prior to the activation event, or whether there is a defined structural pathway or an ordered occurrence of events that lead from the activation event to the transition. How does the study of NP4 structural change fit into this theoretical framework?.

The population of the different conformations along the structural transition (shown in Figure 4) seems to contradict the most relevant implication of the population shift model, since there is no overlap between the histograms of both low and high pH states. Moreover, when the high pH structure is assigned a low pH protonation state or vice versa, as in both L2H and H2L intermediate state simulations, the results show that the protein conformation changes very fast and remains trapped in intermediate structures along the low to high pH structural transition, which is significantly different from the initial state (See Figures 1 and 2 and Tables 2 and 3). This fact suggests that in a low pH condition the high pH conformation represents a high energy unstable structure, and that under high pH condition the low energy structure is a high energy unstable conformation. The results for the L2H and H2L simulation also show that both these states represent conformational ensembles which are in between both end state conformations (Figure 4), but that that they are different from each other. Moreover, it can be seen that their difference does not lie along the transition path and is not related to it. Taking this data altogether the following model (shown in Scheme 1) is proposed for the NP4 structural transition.

Under low pH condition (red free energy curve), the conformational space is dominated by the low pH closed conformation, (left in Scheme 1) when the protein is released in the victims blood and encounters a high pH environment, (left black vertical arrow in Scheme 1) several residues (mainly Asp30 and Asp35) change their protonation state, and a fast structural transition occurs due to the lost of some interactions (Asp30-Leu130, Asp35-Asp129), leading to the L2H state (blue curve in Scheme 1). This state must then slowly proceed to the fully open high pH conformation, by overcoming a transition state (TS). In the reverse transition, when the open conformation stable under a high pH environment (right side of blue curve in Scheme 1) is transferred to a low pH media (right black vertical arrow), both Asp30 and Asp35 are protonated and neutralized. As a consequence, both residues try to escape from the solvent and start to close the AB Loop, yielding the intermediate H2L state (red free energy curve); this step also occurs very fast. Again in a second slower step, passing another putative TS must occur to close the protein and reach the low pH stable conformation with Leu130-Asp30, Asp35-Asp129, Val36-HemeCD, and Val36-Leu133 interactions fully established.

The model reflects the fact that both intermediate states lie between the stable end states, along the transition, but are different; the first step is fast since it is barrierless, and the second is slow due to the presence of a free energy barrier. The fact that in the second step new specific residue interactions must be established possibly explains why it is slower than the first step in which interactions are broken due to a change in the charge of the involved residues. The conformational space explored by the intermediate states in our model can be considered as the group of structures lying in a high energy basin and separated by a barrier from the most stable end sate. So the L2H will correspond to high energy population of the high pH ensemble, and the H2L will correspond to a high energy population of low pH ensemble. The existence of multiple basins in energy landscape of NP4 at both pH conditions is further supported by the bimodal distribution of the first mode in the high pH condition and the first and second EM in the low pH condition.

The proposed mechanism for the NP4 pH dependent structural transition is in clear opposition to what is expected from the population shift paradigm. The fact that the transition occurs in two significantly different free energy landscapes (low and high pH) avoids significant population of the "target state" to be observed prior to the activation event, that is, during the equilibrium dynamics of the "starting state". The allosteric

transition in NP4 does not occur due a change in the relative population of both end states but due to a drastic change in the free energy landscape of NP4 conformational ensemble. Furthermore, as already mentioned, in the thermodynamic model of allostery the EM of the protein are expected to play a predominant role, particularly along the conformational transition. The values of the involvement coefficients observed for NP4 are low with overall value of less than 25% of the transition represented in the equilibrium dynamics. As a comparison, we considered that previous studies found values between up to 0.86, for the highest coefficient of only one EM in a study of 20 proteins in both the open and closed conformations.[10]

The NP4 pH dependent transition fits slightly better in the stereochemical model of allostery, since for both transitions some ordering of the molecular events is observed. In both cases, the activation event triggers a fast loosing of specific key contacts in the first place, which should be followed by a slow gaining of new specific contacts. However, the fact that two different intermediate states are observed that are not contiguous along the transition clearly demonstrates that the structural change path is not symmetric. So the molecular events are differently ordered when the transition occurs in the forward or backward direction.

Can the conclusions drawn here for NP4 allosteric transition be extended to other systems? What is clear from the data presented and resumed in Scheme 1 is that whenever a significant change in the proteins conformational free energy landscape is expected as a consequence of the activation event, the population shift model will probably be unfit due to the high energy of the target structure in the starting condition. On the other hand, when the activation event is a more subtle change, like a phosphorilation, the population shift model will probably represent the process better.

Conclusion

Using MD simulations we have studied the pH dependent structural transition of NP4. On the basis of the results, we proposed a dual free energy landscape description of NP4 allosteric mechanism. In this scenario, the "target" structure corresponds to an unstable high energy conformation prior to the activation state and cannot be observed. The "population shift" framework is therefore unable to describe the transition. The data presented suggest that when the activation event conveys a significant change in the free energy landscape of the protein conformation the population shift model in not a proper description for the allosteric transition. In summary, our results show that the allosteric transition in NP4 does not occur due a change in the relative population of both end states, but due to a drastic change in the free energy landscape of the NP4 conformational ensemble.

Acknowledgment. This work was supported by grants from the ANPCYT (National Science Agency of Argentina) and University of Buenos Aires to D.A.E and M.A.M. D.A.E. and M.A.M. are members of CONICET (National Research Council Argentina). Computer resources were provided by the Large Allocations Resource Committee through Grant TGMCA05S010 to A.E.R.

Supporting Information Available: RMSD versus time plot for the both low, high L2H, and H2L pH simulations using the average structure as a reference. Distance versus time plot for Leu130CO-Asp30COO, Leu130COO-Ala1NH_3^+, Leu129COO^- Asp35COO, Val36CA-HemeCD, and Val36-Leu133 side chain interactions, in both low and high pH simulations. Contributions of the first EM to the overall protein dynamics. RMSD versus residue plot between the average and the 1st EM projected structure. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Formaneck, M. S.; Ma, L.; Cui, Q. *Proteins* **2006**, *63*, 846.
(2) Ma, L.; Cui, Q. *J. Am. Chem. Soc.* **2007**, *129*, 10261.
(3) Szabo, A.; Karplus, M. *J. Mol. Biol.* **1972**, *72*, 163.
(4) Monod, J.; Wyman, J.; Changeux, J. P. *J. Mol. Biol.* **1965**, *12*, 88.
(5) Perutz, M. F. *Nature* **1970**, *228*, 726.
(6) Yu, E. W.; Koshland, D. E. *Proc. Nat. Acad. Sci. U.S.A.* **2001**, *98*, 9517.
(7) Kern, D.; Zuiderweg, E. R. *Curr. Opin. Struct. Biol.* **2003**, *13*, 748.
(8) Swain, J. F.; Gierasch, L. M. *Curr. Opin. Struct. Biol.* **2006**, *16*, 102.
(9) Xu, C.; Tobi, D.; Bahar, I. *J. Mol. Biol.* **2003**, *333*, 153.
(10) Tama, F.; Sanejouand, Y. H. *Protein Eng.* **2001**, *14*, 1.
(11) Tousignant, A.; Pelletier, J. N. *Chem. Biol.* **2004**, *11*, 1037.
(12) Yang, L.-W.; Bahar, I. *Structure (London)* **2005**, *13*, 893.
(13) Frauenfelder, H.; Fenimore, P. W.; Young, R. D. *IUBMB Life* **2007**, *59*, 506.
(14) Bahar, I.; Atilgan, A. R.; Erman, B. *Folding Des.* **1997**, *2*, 173.
(15) Yang, L.-W.; Eyal, E.; Chennubhotla, C.; Jee, J.; Gronenborn, A. M.; Bahar, I. *Structure* **2007**, *15*, 741.
(16) Go, N.; Noguti, T.; Nishikawa, T. *Proc. Natl. Acad. Sci. U.S.A.* **1983**, *80*, 3696.
(17) Levitt, M.; Sander, C.; Stern, P. S. *J. Mol. Biol.* **1985**, *181*, 423.
(18) Andrea Amadei, A. B. M. L. H. J. C. B. *Proteins: Struct., Funct., Genet.* **1993**, *17*, 412.
(19) Rueda, M.; Ferrer-Costa, C.; Meyer, T.; Pérez, A.; Camps, J.; Hospital, A.; Gelpí, J. L.; Orozco, M. *Proc. Nat. Acad. Sci. U.S.A.* **2007**, *104*, 796.
(20) Montfort, W. R.; Weichsel, A.; Andersen, J. F. *Biochim. Biophys. Acta* **2000**, *1482*, 110.
(21) Ribeiro, J. M. C.; Hazzard, J. M. H.; Nussenzveig, R. H.; Champagne, D. E.; Walker, F. A. *Science* **1993**, *260*, 539.
(22) Champagne, D. E.; Nussenzveig, R. H.; Ribeiro, J. M. C. *J. Biol. Chem.* **1995**, *270*, 8691.
(23) Andersen, J. F.; Ding, X. D.; Balfour, C.; Shokhireva, T. K.; Champagne, D. E.; Walker, F. A.; Montfort, W. R. *Biochemistry* **2000**, *39*, 10118.
(24) Andersen, J. F.; Montfort, W. R. *J. Biol. Chem.* **2000**, *275*, 30496.
(25) Andersen, J. F.; Weichsel, A.; Balfour, C. A.; Champagne, D. E.; Montfort, W. R. *Structure* **1998**, *6*, 1315.
(26) Weichsel, A.; Andersen, J. F.; Champagne, D. E.; Walker, F. A.; Montfort, W. R. *Nat. Struct. Biol.* **1998**, *5*, 304.
(27) Andersen, J. F.; Champagne, D. E.; Weichsel, A.; Ribeiro, J. M. C.; Balfour, C. A.; Dress, V.; Montfort, W. R. *Biochemistry* **1997**, *36*, 4423.
(28) Ribeiro, J. M. C.; Nussenzveig, R. H. *FEBS Lett.* **1993**, *330*, 165.
(29) Flower, D. R.; North, A. C. T.; Sansom, C. E. *Biochim. Biophys. Acta* **2000**, *1482*, 9.
(30) Kondrashov, D. A.; Roberts, S. A.; Weichsel, A.; Montfort, W. R. *Biochemistry* **2004**, *43*, 13637.
(31) Maes, E. M.; Weichsel, A.; Andersen, J. F.; Shepley, D.; Montfort, W. R. *Biochemistry* **2004**, *43*, 6679.
(32) Kondrashov, D. A.; Montfort, W. R. *J. Phys. Chem. B* **2007**, *111*, 9244.
(33) Marti, M. A.; Lebrero, M. C. G.; Roitberg, A. E.; Estrin, D. A. *J. Am. Chem. Soc.* **2008**, *130*, 1611.
(34) Maes, E. M.; Roberts, S. A.; Weichsel, A.; Montfort, W. R. *Biochemistry* **2005**, *44*, 12690.
(35) Menyhárd, D. K.; Keserü, G. M. *FEBS Lett.* **2005**, *579*, 5392.
(36) Berendsen, H. J. C.; Postma, J. P. M.; Van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684.
(37) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327.
(38) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins: Struct., Funct., Genet.* **2006**, *65*, 712.
(39) Marti, M. A.; Crespo, A.; Capece, L.; Boechi, L.; Bikiel, D. E.; Scherlis, D. A.; Estrin, D. A. *J. Inorg. Biochem.* **2006**, *100*, 761.
(40) Bikiel, D. E.; Boechi, L.; Capece, L.; Crespo, A.; De Biase, P. M.; Di Lella, S.; Gonzalez Lebrero, M. C.; Marti, M. A.; Nadra, A. D.; Perissinotti, L. L.; Scherlis, D. A.; Estrin, D. A. *Phys. Chem. Chem. Phys.* **2006**, *8*, 5611.
(41) Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham Iii, T. E.; DeBolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. *Comput. Phys. Commun.* **1995**, *91*, 1.

JP808055E