THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

pubs.acs.org/JPCB

Molecular Dynamics Simulation of Autotaxin: Roles of the Nuclease-like Domain and the Glycan Modification

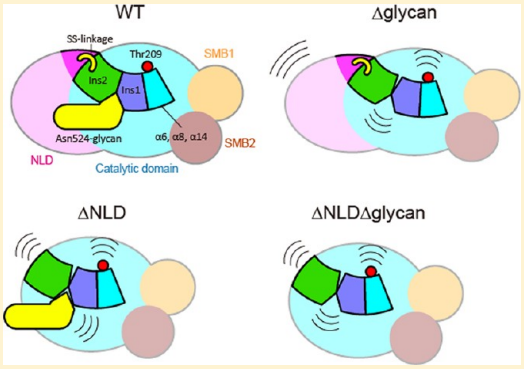
Michio Koyama,[†] Hiroshi Nishimasu,[†] Ryuichiro Ishitani,*[,†,‡] and Osamu Nureki*[,†]

[†]Department of Biophysics and Biochemistry, Graduate School of Science, The University of Tokyo, 2-11-16 Yayoi, Bunkyo-ku, Tokyo 113-0032, Japan

[‡]RIKEN Advanced Science Institute, 2-1 Hirosawa, Wako-shi, Saitama 351-0198, Japan

Supporting Information

ABSTRACT: Autotaxin (ATX) is a secreted lysophospholipase D that produces lysophosphatidic acid, a lipid mediator that activates G protein-coupled receptors to evoke various cellular responses. The nuclease-like domain of ATX and the Asn524-linked glycan are reportedly critical for the catalytic activity. Recently, the crystal structures of ATX were determined, but the means by which the nuclease-like domain and the *N*-glycosylation participate in the catalytic activity still remain undetermined. To address this question, we conducted molecular dynamics simulations of ATX. The simulation trajectories starting from the full-length structure and from structures lacking the nuclease-like domain and/or the glycan were compared. The results suggested that an allosteric interaction pathway, formed by the catalytic domain, including the two insertion regions, the essential glycan modification, and the nuclease-like domain, may stabilize the proper location of the catalytic threonine residue. The results complement those from previous biochemical experiments.



INTRODUCTION

The ectonucleotide pyrophophatase/phosphodiesterase (Enpp) family includes membrane-associated or secreted ectoenzymes that hydrolyze pyrophosphate or phosphodiester bonds in various extracellular compounds, such as nucleotides and lysophospholipids. The mammalian Enpp family comprises seven members (Enpp1−Enpp7), which are divided into two subgroups based on their domain organization: Enpp1−Enpp3 and Enpp4−Enpp7. ATX, which is also called Enpp2, is a secreted lysoPLD that hydrolyzes lysophosphatidylcholine (LPC) to generate lysophosphatidic acid (LPA) and choline.[1,2] LPA activates LPA-specific GPCRs (LPA_{1-6}) to evoke a wide range of cellular responses, including cellular survival, proliferation, differentiation, and migration.[3,4] Thus, ATX is involved in a number of physiological processes, such as brain development,[5] blood vessel formation,[6,7] and lymphocyte trafficking,[8] as well as in pathophysiological conditions.[9] While ATX knockout mice are embryonic lethal with multiple defects, including abnormal vascular development,[7] aberrant ATX expression has been found in malignant tumor tissues, such as breast cancer,[10] Hodgkin's lymphoma,[11] hepatocellular carcinoma,[12] follicular lymphoma, and glioblastoma.[13] In addition, ATX and LPA stimulate fibroblast migration[14] and are implicated in lung and kidney fibrosis. Therefore, ATX is an attractive target for novel anticancer drugs and antifibrotic treatments.[15]

Recently, the mouse and rat ATX crystal structures[16,17] were solved. The structures revealed substantially the same ATX architectures, in which the catalytic domain (CAT; residues 161−536) interacts with the SMB-like domains 1 and 2 (residues 51−94 and 95−139) on one side, and the nuclease-like domain (NLD; residues 586−855) on the other side (Figure 1a). The CAT domain has the hydrophobic lipid-binding pocket that accommodates LPA molecules with various chain lengths and saturations in different conformations. A catalytic threonine residue, Thr209, is located in the vicinity of two essential zinc ions in the hydrophobic pocket (Figure 1b). The crystal structure revealed that the CAT domain contains the other lipid-binding region, the hydrophobic channel, suggesting that the LPAs are delivered from the active site to the cognate GPCRs through this channel (Figure 1b).

Among the Enpp family members, only Enpp1−Enpp3 have the NLD domain. Previous results showed that the NLD domain is important for the catalytic activities of ATX[18] and Enpp1.[19] The domain interface between the NLD domain and the CAT domain is covalently linked by the Cys413−Cys801 disulfide bond, which was shown to be critical for the catalytic activity.[20] However, despite the availability of the crystal structure of ATX, it was difficult to understand how the NLD domain and the S−S linkage facilitate its catalytic activity. The NLD domain is located on the opposite side of the entrance to the catalytic site, and thus it is unlikely to be directly involved in

Received: April 4, 2012
Revised: August 31, 2012
Published: September 11, 2012

dx.doi.org/10.1021/jp303198u | *J. Phys. Chem. B* 2012, 116, 11798−11808

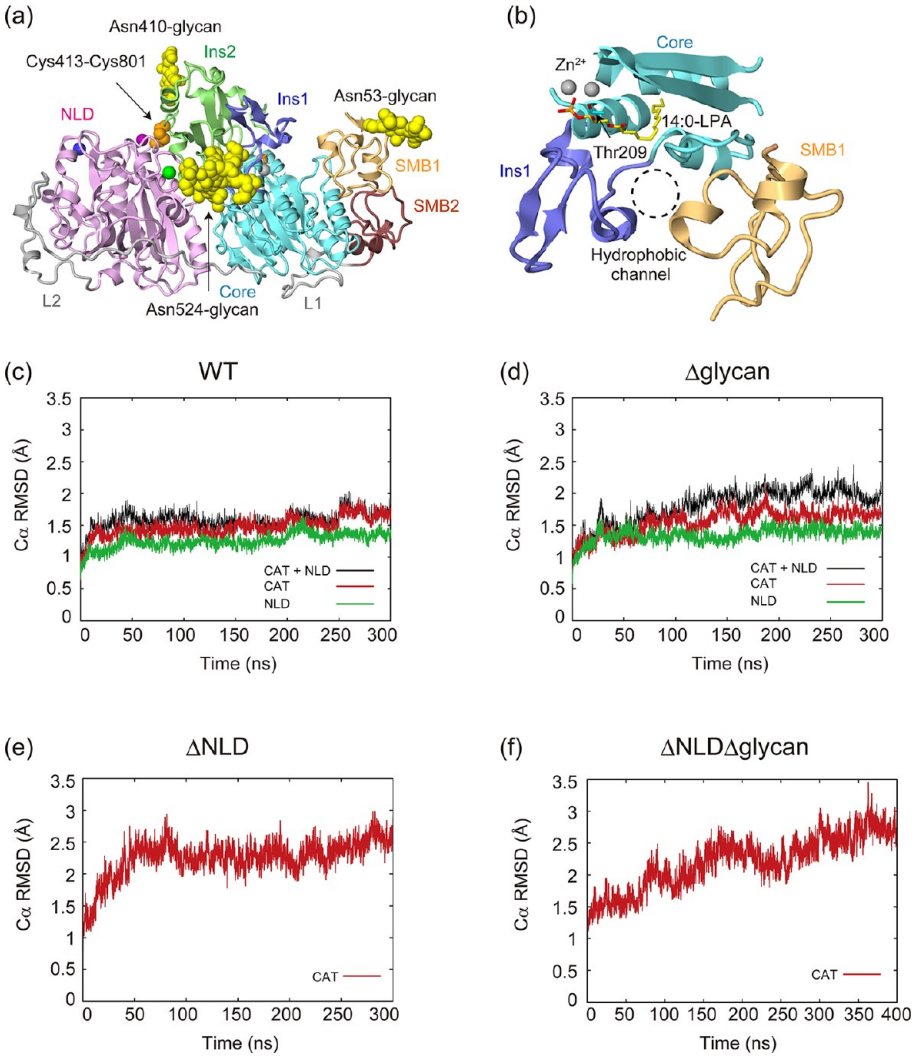


Figure 1. (a) Crystal structure of ATX. Protein backbones are depicted by ribbon models. The somatomedin B-like domain 1 (SMB1), SMB2, the core region, the Ins1 region, the Ins2 region, the NLD domain, and the two linker regions (L1 and L2) are colored beige, brown, cyan, blue, green, magenta, and gray, respectively. The *N*-glycans and the sulfate ion are shown as yellow CPK models and an orange stick model, respectively. Zn^{2+}, Ca^{2+}, Na^{+}, and K^{+} ions are shown as gray, green, magenta, and blue spheres, respectively. (b) Active site of ATX. Protein backbones and Zn^{2+} ions are depicted with the same coloring scheme as that in (a). The hydrophobic channel, formed by the SMB-like domain 1, the core region, and the Ins1 region, is shown as dashed lines. (c–f) Residue-averaged RMSD of the Cα atoms in each simulation. The black line shows the RMSD calculated by superposing the CAT and NLD domains together, and the red and green lines show the RMSD calculated by superposing each domain individually.

either substrate recognition or catalysis. The NLD domain forms several hydrophilic and hydrophobic interactions, as well as the S–S linkage, with the CAT domain behind the entrance of the catalytic site. Therefore, the NLD domain and its S–S linkage were speculated to prevent the collapse of the hydrophobic lipid-binding pocket, by indirectly stabilizing the CAT domain. However, this hypothesis cannot explain the inactivation mechanism of Enpp1, which does not hydrolyze lipid substrates and thus may lack a hydrophobic pocket.

The crystal structure revealed that ATX is *N*-glycosylated at Asn53, Asn410, and Asn524, among which only the Asn524-linked glycan is essential for the catalytic activity.[21] Similar to the NLD domain, the Asn524-linked glycan is also located behind the entrance of the catalytic site and thus is not involved in the formation of the lipid-binding pocket or the hydrophobic channel. Therefore, the static crystal structure still cannot explain the mechanism by which the Asn524-linked glycan affects the catalytic activity of ATX.

To address these questions, we performed molecular dynamics (MD) simulations of ATX, starting from the initial structures of the full-length molecule and those lacking the NLD domain and/or the glycan. The results suggested the existence of an allosteric interaction pathway spanning from the NLD domain to the catalytic Thr209 residue, including the two insertion regions in the CAT domain, as well as the Asn524-linked glycan. The disruption of the pathway, by the removal of either the NLD domain or the Asn524-linked glycan, destabilized the two insertion regions and Thr209, which may consequently impair the catalytic activity of ATX. In addition, the hydrophobic channel of the wild-type ATX collapsed without the bound lipid molecules in the simulation, suggesting that the channel is structurally flexible and open only when lipid molecules pass through it.

dx.doi.org/10.1021/jp303198u | *J. Phys. Chem. B* 2012, 116, 11798–11808

Table 1. Summary of the Simulations Performed in This Study

simulation run	residues	glycan modification	length (ns)	system size (Å)
WT	51−855	Asn53, Asn410, Asn524	300	128 × 128 × 128
Δglycan	51−855	none	300	128 × 128 × 128
ΔNLD	51−534	Asn53, Asn410, Asn524	300	96 × 96 × 96
ΔNLDΔglycan	51−534	none	400	96 × 96 × 96

MATERIALS AND METHODS

Initial Structures and System Preparation. The atomic coordinates of the crystal structure of ATX in the free form (Protein Data Bank (PDB) entry 3NKM[16]) were retrieved from the PDB. Although the structure of ATX in the free form contains $GlcNAc_2Man_5$ at Asn524, we added two extra mannoses to the Asn524-linked glycan, based on the electron densities. The WT simulations were performed using this modified structure as the initial structure. The Δglycan simulation was performed using the crystal structure without the *N*-glycans on Asn53, Asn410, and Asn524. The ΔNLD simulation was performed using the crystal structure lacking the NLD domain (535−855). The ΔNLDΔglycan simulation was performed after removing both the *N*-glycans and the NLD domain. All of the structures included two Zn^{2+} ions and a sulfate ion at the active site and Ca^{2+}, K^+, and Na^+ ions at the NLD domain, which are observed in the crystal structure. All of the water molecules observed in the crystal structure were kept. The missing side chains and hydrogen atoms were built with the program LEAP.[22] The protonation states of the histidine residues were determined according to their chemical environment in the crystal structure: His315, His348, His359, His474, and His793 are protonated at the δ position, and the others are protonated at the ε position. An explicit solvent periodic boundary system was prepared. The resulting sizes of the simulation boxes are listed in Table 1. The net charge of the solute was neutralized through the addition of chloride ions. The molecular topologies and parameters from the AMBER03[23] and GLYCAM06[24] force fields were used.

Molecular Dynamics Simulation Protocol. Molecular dynamics simulations were performed with the program NAMD 2.7.[25] The systems were first energy-minimized for 1000 steps with fixed positions of the non-hydrogen atoms, and then for another 1000 steps with 10 kcal/mol restraints for the non-hydrogen atoms. Next, equilibrations were performed for 0.1 ns under NVT conditions, with 10 kcal/mol restraints for the non-hydrogen atoms. Finally, equilibrations were performed for 10, 0.5, and 0.5 ns, under NPT conditions with the 10, 1.0, and 0.1 kcal/mol restraints, respectively. In the minimization and equilibration processes, the restraint constants of the atoms coordinated to the Zn^{2+} ions were set to 10 times the value of each restraint. Production runs were performed for 300 ns in the WT, Δglycan, and ΔNLD simulations, and for 400 ns in the ΔNLDΔglycan simulation. To emulate the coordination bonds between the Zn^{2+} ions and their coordinating atoms, their distances were harmonically constrained with force constants of 100 kcal/mol/Å^2. In the equilibration and production processes, the pressures and temperatures were set to 1.0 atm and 300 K, respectively. Constant temperature was maintained by using Langevin dynamics. Constant pressure was maintained by using the Langevin piston Nose−Hoover method.[26] Long-range electrostatic interactions were calculated using the particle mesh Ewald method.[27] The one to four electrostatic and van der Waals interactions were scaled by the standard AMBER values (SCEE = 1.2, SCNB = 2.0) in the all simulations according to ref 28.

Trajectory Analysis. *RMSD and RMSF.* The residue-averaged root-mean-square deviation (RMSD) and the time-averaged root-mean-square fluctuation (RMSF) were calculated to evaluate the trajectory stability and the flexible regions, respectively. The residue-averaged root-mean-square deviation (RMSD) for the *t*th time step and the time-averaged root-mean-square fluctuation (RMSF) for the *i*th atom are defined as follows

$$\mathrm{RMSD}=\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left\{(x_i^t-X_i)^2+(y_i^t-Y_i)^2+(z_i^t-Z_i)^2\right\}}$$

and

$$\mathrm{RMSF}=\sqrt{\frac{1}{T}\sum_{t=1}^{T}\left\{(x_i^t-X_i)^2+(y_i^t-Y_i)^2+(z_i^t-Z_i)^2\right\}}$$

where N denotes the number of atoms, T denotes the total time steps, x_i^t, y_i^t, and z_i^t denote the coordinates of the ith atom at the tth time step, and X_i, Y_i, and Z_i denote the reference coordinates of the ith atom. The calculations were performed after superposing each trajectory frame on the reference coordinates. For the RMSD calculation, the crystal structure was used as the reference coordinates. For the RMSF calculation, the averaged structure was used as the reference coordinates, unless otherwise noted.

Domain Rotation. The rotational angle between the two domains (e.g., domains A and B) in the simulation trajectories was calculated as follows. Each structure from the simulation trajectory was rotated and translated to fit over the Cα atoms in domain A of the crystal structure by the least-squares method. The Euler angles (ϕ, θ, ψ) for rotation and the translation vector to fit the Cα atoms in domain B of the crystal structure were then calculated for each structure from the simulation trajectory. The Euler angles were then converted to the rotational angle χ (see ref 29 for details).

Principal Component Analysis. The PCA is based on the 3*N*-dimensional covariance matrix with elements $\mathbf{C}_{ij}$ for coordinates i and j

$$\mathbf{C}_{ij}=\langle(\mathbf{r}_i-\langle\mathbf{r}_i\rangle)(\mathbf{r}_j-\langle\mathbf{r}_j\rangle)\rangle$$

where $\mathbf{r}_i$, $\mathbf{r}_j$ are the x, y, or z coordinates of the Cα atoms in the ith and jth residues during the MD simulation, and $\langle\mathbf{r}_i\rangle$, $\langle\mathbf{r}_j\rangle$ are the average coordinates of the ith and jth residues in the simulation.[30,31] The covariance matrix $\mathbf{C}$ can be decomposed as

$$\mathbf{C}=\mathbf{P}\Delta\mathbf{P}^T$$

where the eigenvectors, $\mathbf{P}$, represent the principal components (PCs), and the eigenvalues are the elements of the diagonal matrix, Δ. The eigenvalues are sorted in descending order. The first few PCs typically describe the collective, global motions of the system. We performed PCA for each simulation. We used 50−300 ns of the trajectories for the calculation of the principal

dx.doi.org/10.1021/jp303198u | J. Phys. Chem. B 2012, 116, 11798−11808

components. Each trajectory was superposed on the CAT domain of the crystal structure, and the principal components were calculated only for the CAT domain.

Correlation Analysis. The correlation matrix was calculated for all Cα atoms. The elements of this matrix, $\mathbf{C}_{ij}$, are defined as follows: $\mathbf{C}_{ij} = \langle(\mathbf{r}_i - \overline{\mathbf{r}}_i)\cdot(\mathbf{r}_j - \overline{\mathbf{r}}_j)\rangle/\{\langle(\mathbf{r}_i - \overline{\mathbf{r}}_i)\cdot(\mathbf{r}_i - \overline{\mathbf{r}}_i)\rangle\langle(\mathbf{r}_j - \overline{\mathbf{r}}_j)\cdot(\mathbf{r}_j - \overline{\mathbf{r}}_j)\rangle\}^{1/2}$, where $\overline{\mathbf{r}}_i$ is the mean position of atom i. The value of element $\mathbf{C}_{ij}$ indicates the degree to which the fluctuations of atom i are correlated with those of atom j over the course of the MD trajectory. We used 50−300 ns of the MD trajectory for the calculation of the correlation matrices.

RESULTS

MD Simulations. To clarify the roles of the NLD domain and the Asn524-linked glycan in the activity of ATX, we performed a series of molecular dynamics (MD) simulations. In the WT simulation, the crystal structure (PDB ID: 3NKM, Figure 1a) was used as the initial structure, after some modifications (see the Materials and Methods section). In the initial structure of the Δglycan simulation, all *N*-glycans were removed. In the ΔNLD simulation, the NLD domain was removed. In the ΔNLDΔglycan simulation, both the *N*-glycans and the NLD domain were removed. A summary of the simulations is provided in Table 1.

Dynamics of the Overall Structures. To examine the structural changes, we calculated the residue-averaged root-mean-square deviation (RMSD) for the CAT (residues 161−536), NLD (residues 586−855), and CAT + NLD (residues 161−536 and 586−855) domains (Figure 1c−f). In the WT simulation, all of the RMSD values remained in the range of 1−2 Å (Figure 1c), indicating that no drastic structural change occurred during the 300 ns simulation. In the Δglycan simulation, the RMSD values of the CAT and NLD domains were also in a similar range to those of the WT simulation, whereas that of the CAT + NLD domains exhibited significantly higher values than those of the CAT and NLD domains (Figure 1d). Thus, the structure of each domain was stable throughout the 300 ns simulation, but the relative orientations of the CAT and NLD domains slightly changed. To further analyze the changes in the orientation between the CAT and NLD domains, we calculated the rotational angle between them (Figure S1a in the Supporting Information). The results showed that the domain orientation in the Δglycan simulation deviated more from the initial crystal structure than that in the WT simulation. Among the glycans in ATX, the Asn524-linked glycan exists between the CAT and NLD domains, whereas the Asn53-linked and Asn410-linked glycans are exposed to the solvent. Thus, it is likely that the Asn524-linked glycan is a major factor for the stabilization of the orientation between the CAT and NLD domains.

We also calculated the RMSD values of the CAT domain in the ΔNLD and ΔNLDΔglycan simulations (Figure 1e,f). The results showed that the RMSD values in both simulations are much higher than those in the WT and Δglycan simulations, indicating that the NLD domain has a significant impact on the stabilization of the CAT domain. Particularly, in the ΔNLDΔglycan simulation, the RMSD value continued to increase during the 400 ns simulation, and finally reached more than 3.0 Å, suggesting that the glycan modifications also affect the stability of the CAT domain, besides the NLD domain. Altogether, the RMSD analysis suggested that the NLD domain and the glycan modification function to stabilize the overall structure of ATX.

Stabilization of the Catalytic Domain by the Nuclease-like Domain and the Glycan Modification. The topology of the CAT domain comprises a repeat of the α/β segments, except for the two insertion regions, Gly219-Gln257 and Thr365-Phe470, which contain auxiliary β strands (Figure 1a). Thus, we divided the CAT domain into three regions: Ins1 (Gly219-Gln257), Ins2 (Thr365-Phe470), and core (Val161-Thr218, Pro258-Val364, and Gln471-Thr536) regions (Figure 1a). The Ins1 region is also conserved in the bacterial Xac phosphatases,[32] while the Ins2 region is a specific insertion for the NPP superfamily.

To analyze the mechanism of the CAT domain stabilization by the glycan and the NLD domain, we calculated the RMSF values of the CAT domain. The region Leu458-Cys468 exhibited especially high RMSF values over 2 Å in all simulations (Figure S2, Supporting Information), and it is disordered or has high *B* factors in the crystal structures.[16,17] Except for this region, the RMSF value was 1−2 Å in the WT simulation, indicating that no large deformation of the CAT domain occurred (Figure S2a, Supporting Information). To clearly understand the differences in the fluctuations by the simulations, we calculated the average RMSF values for the Ins1, Ins2, and core regions (Figure 2). The results showed

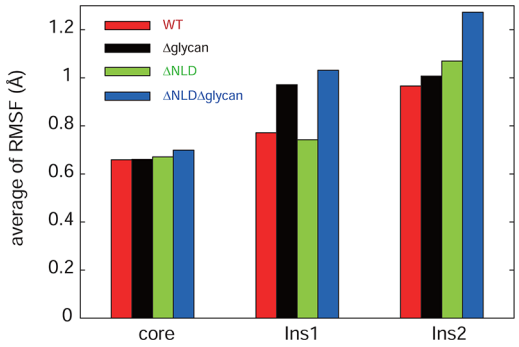


Figure 2. Average RMSF values of the core, Ins1, and Ins2 regions.

that, in all four simulations, the RMSF values of the core region are almost the same, whereas those of the Ins1 and Ins2 regions are different. The Ins1 region exhibited a similar level of destabilization (i.e., increase of RMSF) in both the Δglycan and the ΔNLDΔglycan simulations. Among the glycan modifications, only the Asn524-linked glycan extensively interacted with both the Ins1 and the Ins2 regions (Figure 3a and Figure S3 (Supporting Information)), while the other glycans are exposed to the solvent. Thus, the absence of the Asn524-linked glycan in the Δglycan and ΔNLDΔglycan simulations is probably a major factor for the destabilization of the Ins1 region. On the other hand, the Ins2 domain exhibited small and large destabilizations in the ΔNLD and ΔNLDΔglycan simulations, respectively. This suggested that the NLD domain is important for the stabilization of the Ins2 region. In addition, slight destabilization of the Ins2 region was observed in the Δglycan simulation, despite the presence of the NLD domain. Thus, it is likely that the Asn524-glycan also contributes to the stabilization of the Ins2 region. Altogether, the NLD domain mainly affects the flexibility of the Ins2 region, while the Asn524-linked glycan affects the flexibility of both the Ins1 and the Ins2 regions.

dx.doi.org/10.1021/jp303198u | J. Phys. Chem. B 2012, 116, 11798−11808

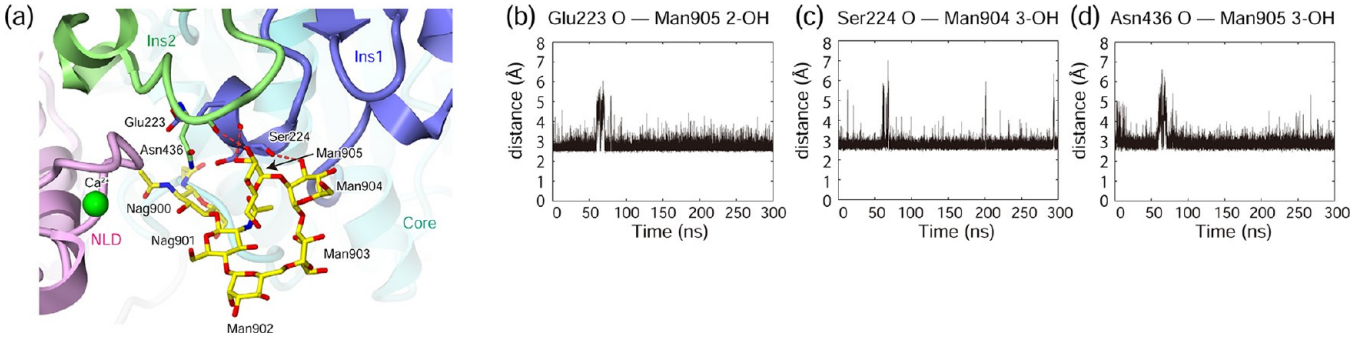


Figure 3. (a) The hydrogen bonds between the Asn524-linked glycan and the Ins1 or Ins 2 regions retained during the WT simulation. The hydrogen bonds are shown by dashed red lines. The protein backbones of the crystal structure are depicted by ribbon models, with the same coloring scheme as that in Figure 1a. The protein side chains and the glycan residues discussed in the text are shown as stick models. (b–d) The distance between atom pairs in the ATX protein and the Asn524-linked glycan during the WT simulation. Each figure shows the distance between (b) the main chain carbonyl group of Glu223 and the 2-OH group of Man905, (c) the main chain carbonyl group of Ser224 and the 3-OH group of Man904, and (d) the main chain carbonyl group of Asn436 and the 3-OH group of Man905.

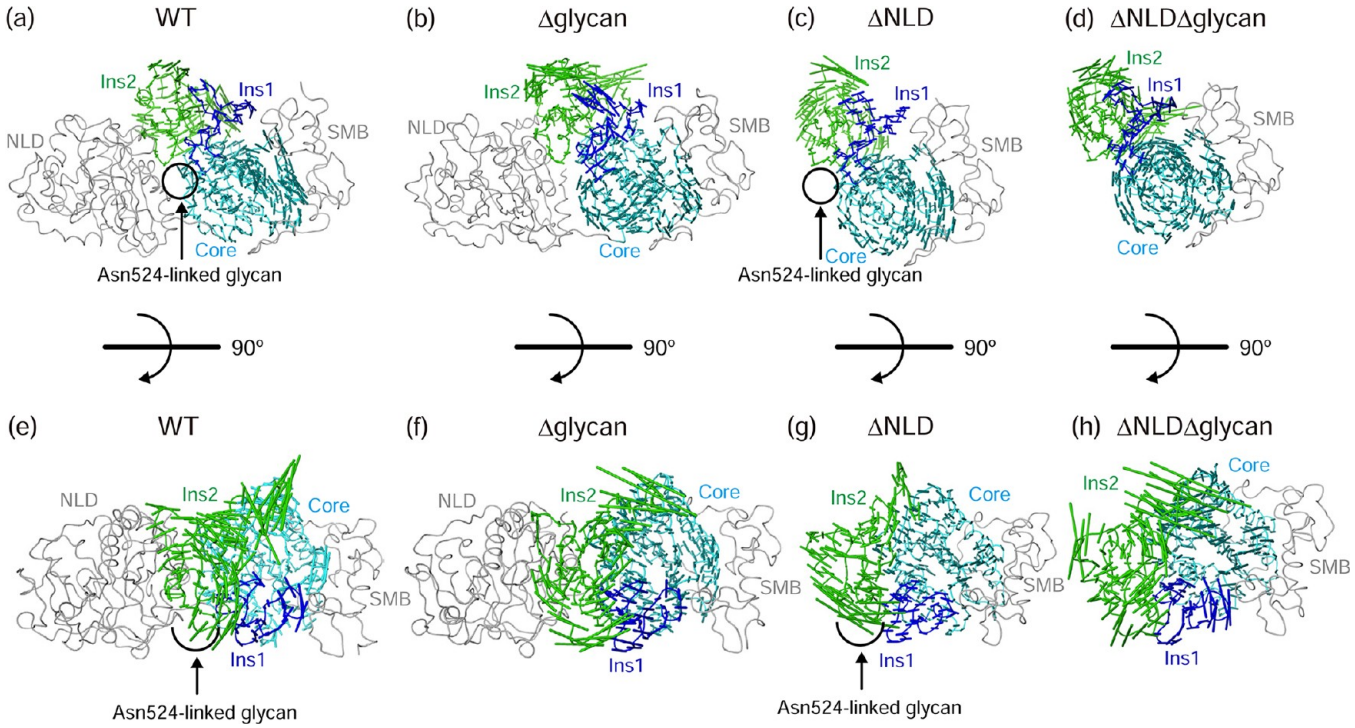


Figure 4. Visualizations of the PC1 motions calculated from the WT, Δglycan, ΔNLD, and ΔNLDΔglycan simulations, viewed from two different directions. The Ins1, Ins2, and core regions are depicted with the same coloring scheme as that in Figure 1a, and the other domains are colored gray. The motions of PC1 are shown by sticks. The stick lengths represent the relative amplitudes of the fluctuations of the corresponding residue. For the WT and ΔNLD simulations, the locations of the Asn524-glycans are indicated by circles.

To further analyze the mechanism of the CAT domain stabilization by the glycan and the NLD domain, we performed a principal component analysis (PCA; Figure 4 and Figure S4 (Supporting Information)). The principal component with the largest eigenvalue (PC1) is generally used to characterize the dominant motion in protein dynamics.[30,31] The magnitudes of the PC1 motions of the Ins2 region in the ΔNLD and ΔNLDΔglycan simulations were larger than those in the WT and Δglycan simulations (Figure 4). This is consistent with the above discussion that the NLD domain plays a role in stabilizing the Ins2 region. In addition, the core domain in the ΔNLD and ΔNLDΔglycan simulations exhibited a similar concerted motion (Figure 4c,d), which differed from those in the WT and Δglycan simulations (Figure 4a,b). This indicated that the deletion of the NLD domain affected the dynamics of the core domain as well. The glycan modifications seemed to affect the overall dynamics of the CAT domain. The overall PC1 motion of the CAT domain in the Δglycan simulation was quite different from that in the WT simulation (Figure 4a,b,e,f). The PC1 motion of the Ins1 and Ins2 regions in the Δglycan simulation tended to move toward the NLD domain (Figure 4b,f), and its amplitude and direction were larger than and different from those in the WT simulation (Figure 4a,e). This is consistent with the above notions that the Asn524-linked glycan stabilizes both the Ins1 and the Ins2 regions, as well as the orientation of the CAT and NLD domains.

Throughout the WT simulation, the three hydrogen bonds between the Asn524-linked glycan and the ATX protein

dx.doi.org/10.1021/jp303198u | J. Phys. Chem. B 2012, 116, 11798–11808

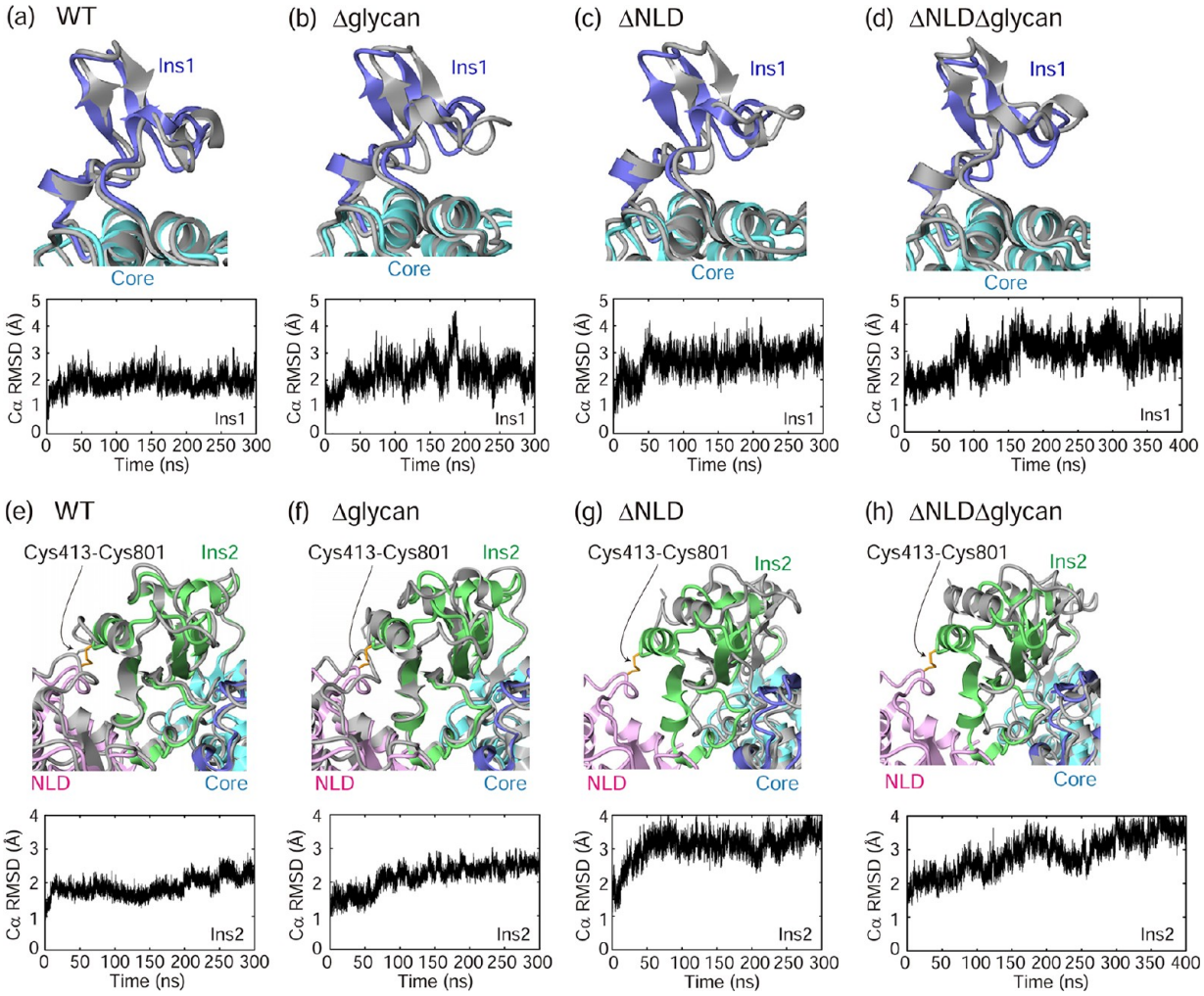


Figure 5. Snapshots of the Ins1 (a–d) and Ins2 (e–h) regions at 300 ns of each MD simulation, as compared with the crystal structure (top), and RMSDs of the Ins1 (a–d) and Ins2 (e–h) regions of the Cα atoms relative to the crystal structure of each simulation (bottom). Snapshots of each MD simulation at 300 ns are colored gray. The crystal structure of ATX, used as the initial structure for the MD simulations, is depicted with the same coloring scheme as that in Figure 1a.

observed in the crystal structure (Figure 3a) were maintained (Figure 3b–d). The 2- and 3-OH groups of the Man905 residue hydrogen bond with the main chain carbonyl groups of Glu223 and Asn436, which belong to the Ins1 and Ins2 regions, respectively. The 3-OH group of the Man904 residue hydrogen bonds with the main chain carbonyl group of Ser224 in the Ins1 region. Taken together, these interactions mediated by the Asn524-linked glycan may stabilize the interaction between the Ins1 and the Ins2 regions.

Flexibility of the Ins1 and Ins2 Regions. To explore the structural flexibility of the Ins1 and Ins2 regions, we calculated the residue-averaged root-mean-square deviation (RMSD), by superposing the core region in the CAT domain of each frame onto that of the crystal structure (Figure 5).

In the WT simulation, the RMSD value of the Ins1 region gradually increased during the initial 40 ns, and then reached a plateau (Figure 5a). The structures deviated from the initial crystal structure (Figure 5a), but no large structural change was observed after the relaxation. By contrast, in the Δglycan, ΔNLD, and ΔNLDΔglycan simulations, the RMSD significantly fluctuated and increased over the simulation (Figure 5b–d). Notably, similar to the Δglycan and ΔNLDΔglycan simulations, the RMSD of the Ins1 region in the ΔNLD simulation significantly drifted, despite the presence of the Asn524-linked glycan. These results suggested that the Ins1 region is stabilized not only by the Asn524-linked glycan but also by the NLD domain.

The RMSD values of the Ins2 region exhibited similar drifts to those of the Ins1 region (Figure 5e–h). In the WT simulation, the RMSD of the Ins2 region became stable after 15 ns (Figure 5e), and no large structural change was observed at the end of the simulation (Figure 5e). By contrast, in the Δglycan, ΔNLD, and ΔNLDΔglycan simulations, the RMSDs significantly fluctuated (Figure 5f–h). Especially, the largest structural deviation was observed in the ΔNLDΔglycan simulation, in which the Ins2 region slightly moved toward the catalytic site (Figure 5h). These results suggested that the Ins2 region has a tendency to approach the catalytic site, in the absence of the NLD domain. Furthermore, in the Δglycan simulation, the RMSD of the Ins2 region significantly drifted, despite the presence of the NLD domain (Figure 5f), suggesting that the interaction between the Asn524-linked glycan and the Ins2 region also plays an important role in stabilizing the Ins2 region. Thus, these results suggested that the Ins2 region is stabilized by interactions with both the NLD domain and the Asn524-linked glycan.

dx.doi.org/10.1021/jp303198u | J. Phys. Chem. B 2012, 116, 11798–11808

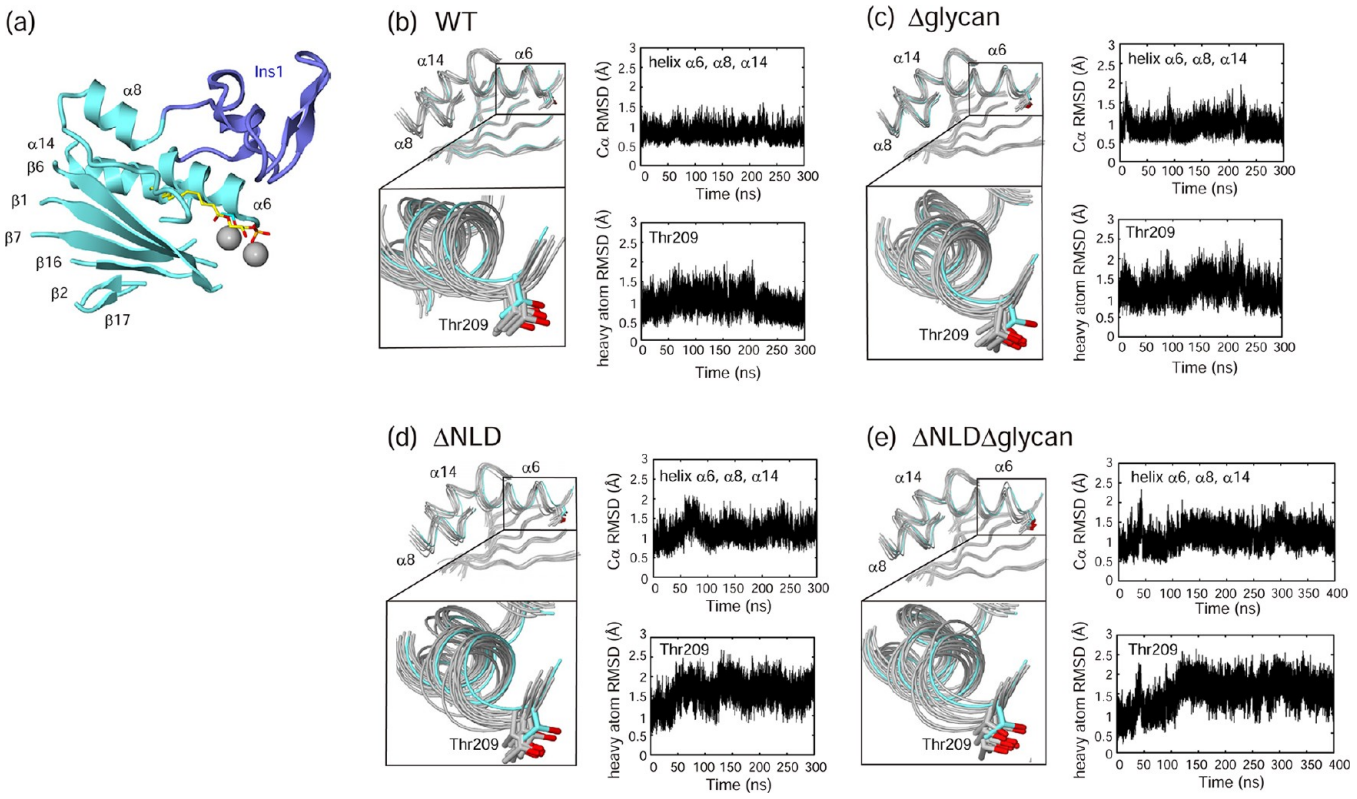


Figure 6. Structural flexibility of the catalytic site and the lipid-binding pocket. (a) The hydrophobic lipid-binding pocket. The protein backbones and the side chain of Thr209 are depicted with the same coloring scheme as that in Figure 1a. The two Zn^{2+} ions are shown as gray spheres, and 14:0-LPA is depicted by a stick model. (b–e) Snapshots of each MD simulation at every 30 ns from 30 to 300 ns, as compared with the crystal structure (left), and RMSDs of the Cα atoms of helices α6, α8, and α14 relative to the crystal structure and of Thr209 and all heavy atoms relative to the crystal structure (right). Snapshots of each MD simulation are colored gray. The crystal structure of ATX, used as the initial structure for MD simulations, is depicted with the same coloring scheme as that in Figure 1a. Protein backbones are depicted by tube models.

Flexibility of the Catalytic Site and the Lipid-Binding Pocket. To investigate the effect of the Ins1 and Ins2 regions on the catalytic site, we calculated the residue-averaged RMSD values of helices α6, α8, and α14 (Figure 6) and of the catalytic Thr209, by superimposing the core β sheets (β1, β2, β6, β7, β16, and β17) in the catalytic core. These α helices constitute the lipid-binding pocket with the core β sheets, and helix α6 harbors the catalytic Thr209 residue. In the WT simulation, the RMSD values of helices α6, α8, and α14 and of Thr209 remained almost constant over the simulation, indicating the lack of structural changes around the catalytic core (Figure 6b). By contrast, in the ΔNLDΔglycan simulation, changes in the RMSD values were observed for the three α helices and Thr209 (Figure 6e). A small shift of helix α6 toward its C-terminal side was detected at the end of the simulation (Figure 6e), and the position and orientation of the Thr209 side chain also deviated from those of the initial crystal structure (Figure 6e). The Ins1 region is inserted between helices α6 and α8, and thus it is likely that the motion of the Ins1 region directly influenced the stability of helices α6 and α8, further affecting the catalytic Thr209 residue. Similar tendencies were also observed for the ΔNLD simulation (Figure 6d). In the Δglycan simulation, comparable fluctuations of the three α helices were observed, but the amplitudes and frequencies were lower than those of the ΔNLD and ΔNLDΔglycan simulations (Figure 6c). In addition, the position and orientation of the Thr209 side chain deviated from those of the initial crystal structure, but the amplitude was smaller (Figure 6c). However, the RMSD values of the three α helices and Thr209 increased at around 150–200 ns of the simulation (Figure 6c), which coincides with the increased RMSD of the Ins1 region observed around 150–200 ns (Figure 5b). Altogether, these results suggested that the lack of the NLD domain and the Asn524-linked glycan destabilizes the three α helices in the core region, resulting in the pronounced fluctuation of the catalytic Thr209 residue.

The Importance of the Interaction Pathway in the Stabilization of the Catalytic Site. Correlated motions between the residues in proteins are essential for their functions, such as ligand interactions, catalysis, and allostery.[33−37] To visualize the correlated motions between the catalytic site and the other domains, we calculated the correlation coefficient between each Cα atom in each simulation (Figure 7). In the WT simulation, the correlation coefficients for the Ins1–Ins2 and Ins1–helix α6 interfaces became slightly lower than those in the other simulations (Figure 7a). On the basis of the RMSF analysis results discussed above, the structure of the CAT domain in the WT simulation was more rigid than those in the other simulations (Figure 2), which may have resulted in the small random fluctuation, and thus the low correlation coefficients, in the WT simulation.

In the Δglycan, ΔNLD, and ΔNLDΔglycan simulations, several pairs of residues between helix α6 and the Ins1 region with correlation coefficients greater than 0.6 were observed (Figure 7b–d). Especially, a high correlation coefficient between Thr215 and Leu220 was observed in all four simulations (Figure 7a–d,f), which can be ascribed to the hydrogen bond between the side chain OH group of Thr215

dx.doi.org/10.1021/jp303198u | J. Phys. Chem. B 2012, 116, 11798–11808

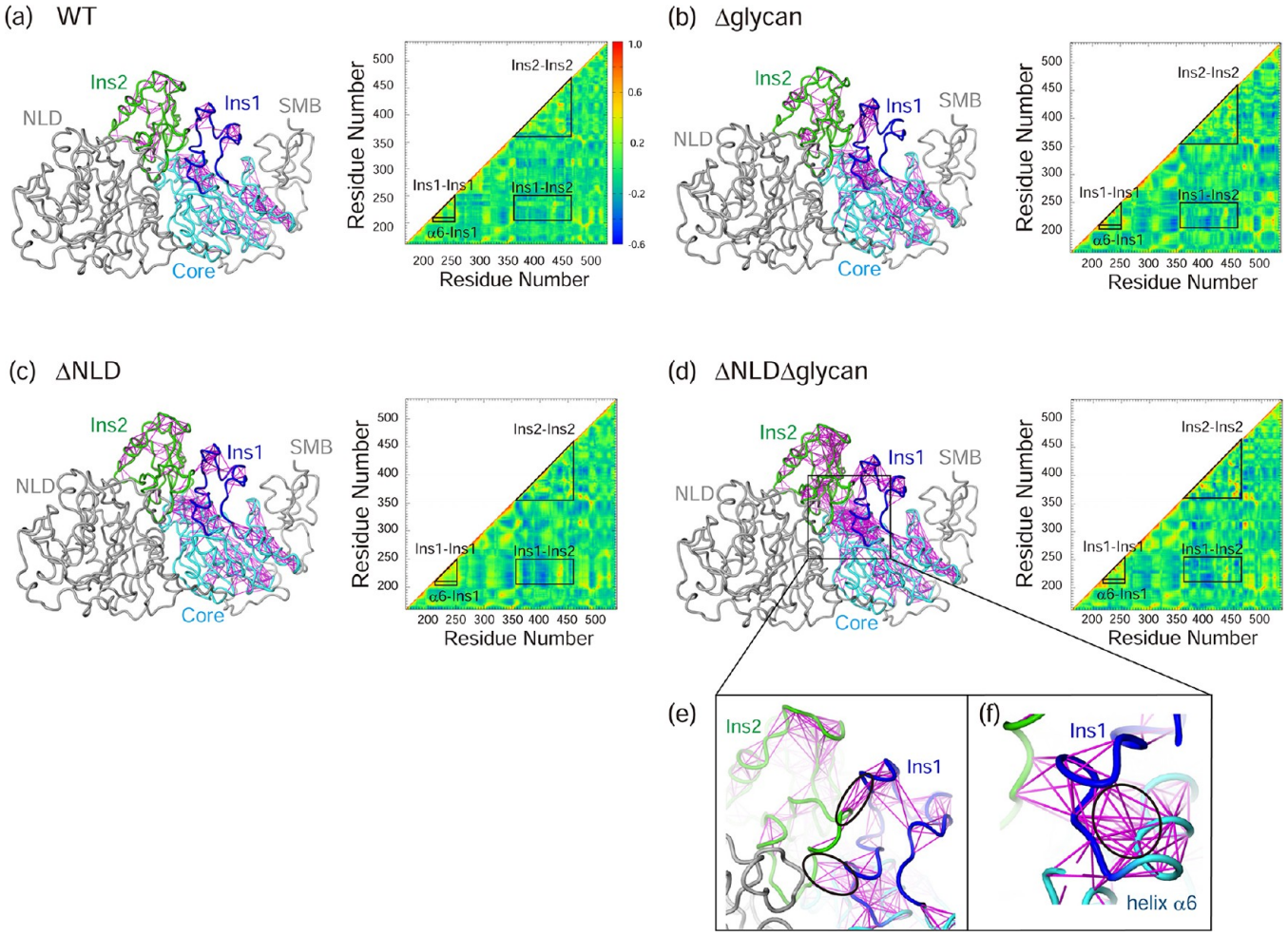


Figure 7. Network of residues with correlated motion. (a–d) Residue pairs with correlation coefficients over 0.6, connected by magenta lines (left), and Cα correlation matrices of the CAT domain of each simulation (right). The Ins1 and Ins2 regions and the core region are depicted with the same coloring scheme as that in Figure 1a, and the other domains are colored gray. (e, f) Correlation network of (e) the Ins1 and Ins2 regions and (f) helix α6 and the Ins1 region of the ΔNLDΔglycan simulation. Residue pairs between the In1 and the Ins2 regions and between the Ins1 region and the helix α6 are shown by circles.

and the main chain carbonyl group of Leu220. This Thr215 residue is well conserved among the Enpp family members, and thus the mutation of this Thr215 to a nonpolar Ala or Val residue may affect the catalytic activity of ATX. Furthermore, the interface between the Ins1 and Ins2 regions also exhibited a high correlation coefficient (over 0.6), consisting of two parts: the interactions between Asp234-Val236 in the Ins1 region and Arg438 and Arg439 in the Ins2 region, and the interactions between Tyr221-Glu223 in the Ins1 region and His433 and Tyr434 in the Ins2 region (Figure 7e). Taken together, in the ATX variants lacking the NLD domain and/or the Asn524-linked glycan, highly correlated motions were observed in the interaction pathway spanning from the Ins2 region to helix α6 and the catalytic center.

Structural Change of the Hydrophobic Channel. In the previous structural analyses of ATX,[16,17] the hydrophobic channel formed by the SMB-like domain 1 and the Ins1 region was shown to be important for the cell migration activity in vivo. Electron density peaks with elongated shapes were observed in this hydrophobic channel, in both the lipid-bound and the free-form crystal structures. Therefore, some lipid molecules derived from the culture medium may be bound to the hydrophobic channel even in the free-form crystal structure, as confirmed by mass spectrometry.[16,17]

In the present simulation, we analyzed the dynamics of the hydrophobic channel (Figure 8) in the absence of bound lipid molecules. The C-terminal part (241–257) of the Ins1 region forms a wall of the hydrophobic channel (Figure 8a). In all of the simulations, this C-terminal part of the Ins1 region exhibited higher RMSF values than those of the N-terminal part (Figure 8). In addition, the rotational angle between the CAT and the SMB-like domains also significantly deviated from that of the crystal structure (Figure S1b, Supporting Information). In this relaxed conformation, as compared to that in the crystal structure, this C-terminal part of the Ins1 region and the SMB-like domain 1 gradually approached each other, and two new van der Waals contacts, between Tyr82 and Arg246 and between Leu78 and Phe249, were formed (Figure 8). Therefore, these results suggested that the hydrophobic channel is flexible and possibly closed in the absence of bound lipid molecules. The channel might only open when lipid molecules are bound within it. In contrast, the structure of the lipid-binding pocket did not collapse in the present simulation without binding lipid molecules. Furthermore, the closure of the hydrophobic channel occurred in all three simulations,

dx.doi.org/10.1021/jp303198u | J. Phys. Chem. B 2012, 116, 11798–11808

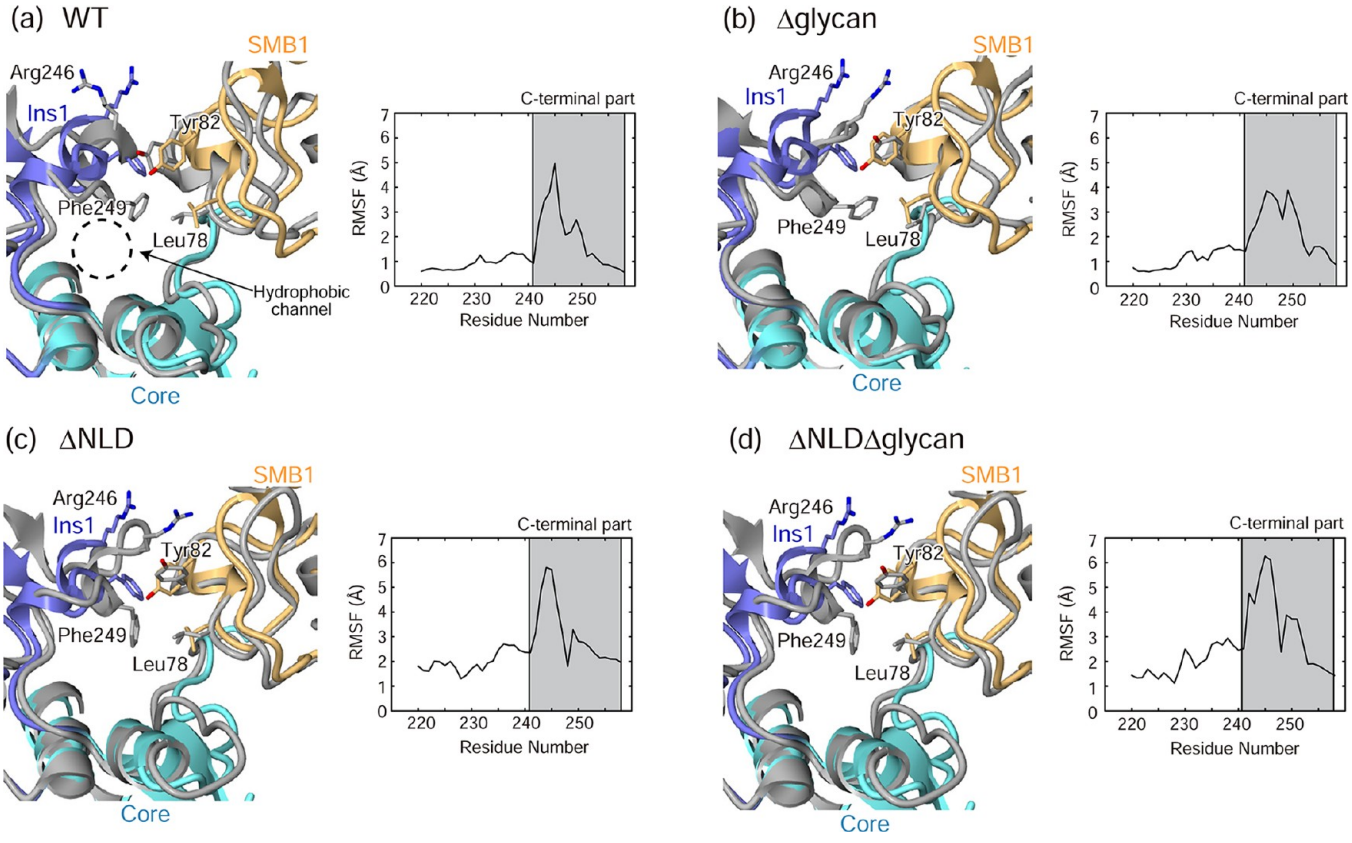


Figure 8. Structural change of the hydrophobic channel. (a–d) Snapshots of the hydrophobic channel, including the C-terminal part of the Ins1 region, of each simulation at 300 ns, as compared with the crystal structure (left), and the time-averaged RMSF of the C-terminal part of Ins1 relative to the crystal structure of each simulation (right). The side chains forming new van der Waals contacts during the simulations are shown as stick models. Snapshots of each MD simulation are colored gray. The Ins1 and Ins2 regions and the core region in the crystal structure are depicted with the same coloring scheme as that in Figure 1a.

suggesting that the stabilization of the Ins1 region by the NLD domain and the glycan modification has no effect on this channel closure.

DISCUSSION

On the basis of the MD simulation results, we propose a mechanism for the stabilization of the ATX structure, mediated by the NLD domain and the glycan modification. The present results suggested that the interaction pathway, consisting of the NLD domain, the Ins1 and Ins2 regions, the glycan modification, and the catalytic core (helices α6, α8, and α14), is critical for the structural rigidity of the intact ATX (Figure 9a). The lack of the glycan modification and/or the NLD domain may disrupt this pathway, which could result in the large correlated motions of the Ins1 and Ins2 regions and the catalytic core, as seen in the Δglycan, ΔNLD, and ΔNLDΔglycan simulations (Figure 9b–d). Particularly, the increased flexibility of the Ins1 region may lead to the destabilization of helix α6 and the catalytic Thr209 residue. Among the conserved residues in the catalytic site of ATX, the Thr209 residue is unique, in that it forms a covalent bond with the substrate phosphate group. The large fluctuation of the Thr209 residue may hinder the correct positioning of its OH group for the nucleophilic attack on the substrate phosphate group, leading to the catalytic inactivation of ATX. Therefore, the NLD domain and the glycan modification may function in the stabilization of the Thr209 residue through the interaction

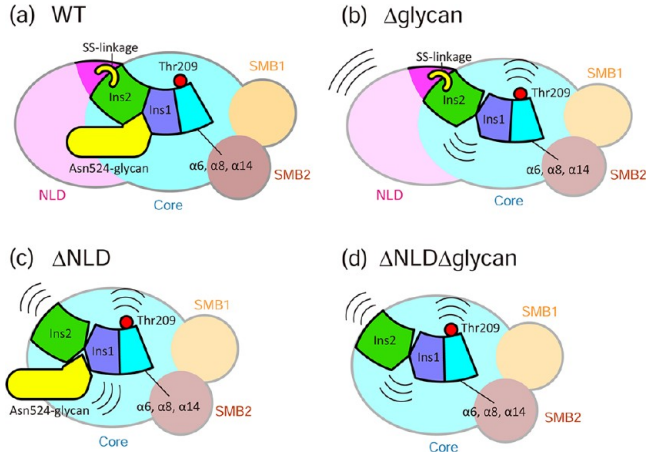


Figure 9. Proposed mechanism of the structural stabilization of the lipid-binding pocket by the NLD domain, Ins1, Ins2, and the Asn524-linked glycans, based on the present simulations. The schematic representations are drawn based on the results of (a) the WT simulation, (b) the Δglycan simulation, (c) the ΔNLD simulation, and (d) the ΔNLDΔglycan simulation. The triple black arcs in (b)–(d) indicate that each (sub)domain becomes flexible by the removal of Asn524-linked glycans and/or the NLD domain.

pathway, rather than in the maintenance of the hydrophobic pocket, as speculated from the static crystal structures.

Recently, del Sol et al. proposed a new view of allostery: all proteins have multiple pre-existing pathways through which

dx.doi.org/10.1021/jp303198u | J. Phys. Chem. B 2012, 116, 11798–11808

allosteric signals are transmitted, and perturbation events at any site in a protein will not create new pathways, but only shift the pre-existing ensemble of pathways.[38] The shift of the ensemble of pre-existing pathways will not necessarily elicit conformational changes, but will influence the dynamics of the protein. From this point of view, the interaction pathway observed in this study can be interpreted as a type of allosteric pathway. The disruptions of this pathway shifted the ensemble of protein conformations, and then caused the destabilization of the catalytic Thr209 residue.

The above hypothesis, based on the present simulation study, seems to be further applicable to Enpp1 and Enpp3. Enpp1 and Enpp3 lack the hydrophobic lipid-binding pocket and hydrolyze various nucleotide triphosphates and their derivatives, but not LPCs. These enzymes have the conserved catalytic Thr residue and thus are considered to catalyze the hydrolysis reaction in a similar manner to ATX. The components involved in the allosteric stabilization pathway (i.e., the NLD domain, the Ins1 and Ins2 regions, and the glycan modification) are also conserved in Enpp1 and Enpp3. Therefore, the catalytic Thr residue may be allosterically stabilized in a similar manner to that in ATX, and the lack of the NLD domain or the glycan modification may compromise the correct positioning of the OH group of the catalytic Thr residue, leading to catalytic inactivation.

It should be noted here that our discussion on the role of the NLD domain does not preclude the possibility that the NLD domains of ATX, Enpp1, and Enpp3 have an "other function", which is not directly related to their catalytic activity. It seems unlikely that the positive selection pressure to preserve the NLD domains of these enzymes is only to stabilize their CAT domains, since Enpp4−Enpp7 lack the NLD domain.[39] For example, the NLD domain of ATX, together with the CAT domain, forms a flat molecular surface, which may be involved in the interaction with the plasma membrane.[16] Further structural and functional studies of the Enpp family members are required to understand the "other function" of the NLD domain of the Enpp family.

CONCLUSIONS

To elucidate the roles of the NLD domain and the glycan modifications in the catalytic activity of ATX, we performed molecular dynamics (MD) simulations, starting from the initial structures of the full-length molecule and those lacking the NLD domain and/or the glycans. The results revealed that the Asn524-linked glycan stabilizes the orientation between the CAT and the NLD domains, while the NLD domain stabilizes the structure of the CAT domain. The detailed analyses further suggested the existence of an allosteric interaction pathway spanning from the NLD domain to the catalytic Thr209 residue, including the Ins1 and Ins2 regions in the CAT domain. The disruption of the pathway, by the removal of the NLD domain or the Asn524-linked glycan, destabilized the Ins1 and Ins2 regions and Thr209, and thus may consequently impair the catalytic activity of ATX.

ASSOCIATED CONTENT

Supporting Information

Graphs of the domain rotations, schematic diagram of the hydrogen bonds around the Asn524-linked glycan, and table of the interactions between the CAT and NLD domains conserved during the WT simulation. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*Phone: +81-3-5841-4392. Fax: +81-3-5841-8057. E-mail: nureki@biochem.s.u-tokyo.ac.jp (O.N.), ishitani@biochem.s.u-tokyo.ac.jp (R.I.).

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by the Japan Society for the Promotion of Science (JSPS) through its "Funding Program for World-Leading Innovative R&D on Science and Technology (FIRST program)" to O.N., by a grant for the National Project on Protein Structural and Functional Analyses from the Ministry of Education, Culture, Sports, Science and Technology (MEXT) to O.N., by a Grant-in-Aid for Scientific Research on Innovative Areas from MEXT to R.I. and O.N., and by a grant for HPCI STRATEGIC PROGRAM Computational Life Science and Application in Drug Discovery and Medical Development by MEXT. We are grateful to Dr. Y. Sugita (Advanced Science Institute, RIKEN, Japan) and Dr. J. Aoki (Tohoku University, Japan) for helpful discussions and comments on the manuscript. We thank the RIKEN Integrated Cluster of Clusters (RICC) for providing computational resources.

REFERENCES

(1) Umezu-Goto, M.; Kishi, Y.; Taira, A.; Hama, K.; Dohmae, N.; Takio, K.; Yamori, T.; Mills, G. B.; Inoue, K.; Aoki, J.; et al. *J. Cell Biol.* **2002**, *158*, 227−233.

(2) Tokumura, A.; Majima, E.; Kariya, Y.; Tominaga, K.; Kogure, K.; Yasuda, K.; Fukuzawa, K. *J. Biol. Chem.* **2002**, *277*, 39436−39442.

(3) Noguchi, K.; Herr, D.; Mutoh, T.; Chun, J. *Curr. Opin. Pharmacol.* **2009**, *9*, 15−23.

(4) Moolenaar, W. H.; van Meeteren, L. A.; Giepmans, B. N. G. *BioEssays* **2004**, *26*, 870−881.

(5) Contos, J. J. A.; Fukushima, N.; Weiner, J. A.; Kaushal, D.; Chun, J. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 13384−13389.

(6) Tanaka, M.; Okudaira, S.; Kishi, Y.; Ohkawa, R.; Iseki, S.; Ota, M.; Noji, S.; Yatomi, Y.; Aoki, J.; Arai, H. *J. Biol. Chem.* **2006**, *281*, 25822−25830.

(7) van Meeteren, L. A.; Ruurs, P.; Stortelers, C.; Bouwman, P.; van Rooijen, M. A.; Pradère, J. P.; Pettit, T. R.; Wakelam, M. J. O.; Saulnier-Blache, J. S.; Mummery, C. L.; et al. *J. Mol. Cell Biol. (Oxford, U.K.)* **2006**, *26*, 5015−5022.

(8) Kanda, H.; Newton, R.; Klein, R.; Morita, Y.; Gunn, M. D.; Rosen, S. D. *Nat. Immunol.* **2008**, *9*, 415−423.

(9) Inoue, M.; Rashid, M. H.; Fujita, R.; Contos, J. J. A.; Chun, J.; Ueda, H. *Nat. Med. (Tokyo, Jpn.)* **2004**, *10*, 712−718.

(10) Liu, S.; Umezu-Goto, M.; Murph, M.; Lu, Y.; Liu, W.; Zhang, F.; Yu, S.; Stephens, L. C.; Cui, X.; Murrow, G.; et al. *Cancer Cell* **2009**, *15*, 539−550.

(11) Baumforth, K. R. N.; Flavell, J. R.; Reynolds, G. M.; Davies, G.; Pettit, T. R.; Wei, W.; Morgan, S.; Stankovic, T.; Kishi, Y.; Arai, H.; et al. *Blood* **2005**, *106*, 2138−2146.

(12) Zhang, G.; Zhao, Z.; Xu, S.; Ni, L.; Wang, X. *Chin. Med. J.* **1999**, *112*, 330−332.

(13) Masuda, A.; Nakamura, K.; Izutsu, K.; Igarashi, K.; Ohkawa, R.; Jona, M.; Higashi, K.; Yokota, H.; Okudaira, S.; Kishimoto, T.; et al. *Br. J. Haematol.* **2008**, *143*, 60−70.

(14) Hama, K.; Aoki, J.; Fukaya, M.; Kishi, Y.; Sakai, T.; Suzuki, R.; Ohta, H.; Yamori, T.; Watanabe, M.; Chun, J.; et al. *J. Biol. Chem.* **2004**, *279*, 17634−17639.

(15) Parrill, A. L.; Baker, D. L. *Anti-Cancer Agents Med. Chem.* **2008**, *8*, 917−923.

(16) Nishimasu, H.; Okudaira, S.; Hama, K.; Mihara, E.; Dohmae, N.; Inoue, A.; Ishitani, R.; Takagi, J.; Aoki, J.; Nureki, O. *Nat. Struct. Mol. Biol.* **2011**, *18*, 205−212.
(17) Hausmann, J.; Kamtekar, S.; Christodoulou, E.; Day, J. E.; Wu, T.; Fulkerson, Z.; Albers, H. M. H. G.; van Meeteren, L. A.; Houben, A. J. S.; van Zeijl, L.; et al. *Nat. Struct. Mol. Biol.* **2011**, *18*, 198−204.
(18) Cimpean, A.; Stefan, C.; Gijsbers, R.; Stalmans, W.; Bollen, M. *Biochem. J.* **2004**, *381*, 71−77.
(19) Gijsbers, R.; Ceulemans, H.; Bollen, M. *Biochem. J.* **2003**, *371*, 321−330.
(20) Jansen, S.; Andries, M.; Derua, R.; Waelkens, E.; Bollen, M. *J. Biol. Chem.* **2009**, *284*, 14296−14302.
(21) Jansen, S.; Callewaert, N.; Dewerte, I.; Andries, M.; Ceulemans, H.; Bollen, M. *J. Biol. Chem.* **2007**, *282*, 11084−11091.
(22) Case, D. A.; Cheatham, T. E., III; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M., Jr.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. *J. Comput. Chem.* **2005**, *26*, 1668−1688.
(23) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; et al. *J. Comput. Chem.* **2003**, *24*, 1999−2012.
(24) Kirschner, K. N.; Yongye, A. B.; Tschampel, S. M.; González-Outeiriño, J.; Daniels, C. R.; Foley, B. L.; Woods, R. J. *J. Comput. Chem.* **2008**, *29*, 622−655.
(25) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. *J. Comput. Chem.* **2005**, *26*, 1781−1802.
(26) Feller, S. E.; Zhang, Y.; Pastor, R. W.; Brooks, B. R. *J. Chem. Phys.* **1995**, *103*, 4613−4621.
(27) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089−10092.
(28) Xu, D.; Newhouse, I. E.; Amaro, E. R.; Pao, C. H.; Cheng, S. L.; Markwick, R. L. P.; McCammon, A. J.; Li, W. W.; Arzberger, W. P. *J. Mol. Biol.* **2009**, *387*, 465−491.
(29) Ishitani, R.; Sugita, Y.; Dohmae, N.; Furuya, N.; Hattori, M.; Nureki, O. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 15393−15398.
(30) Amadei, A.; Linssen, A.; Berendsen, H. *Proteins* **1993**, *17*, 412−425.
(31) Yang, L.; Song, G.; Carriquiry, A.; Jernigan, R. *Structure* **2008**, *16*, 321−330.
(32) Zalatan, J. G.; Fenn, T. D.; Brunger, A. T.; Herschlag, D. *Biochemistry* **2006**, *45*, 9788−9803.
(33) Brandman, R.; Lampe, J.; Brandman, Y.; Montellano, P. *Arch. Biochem. Biophys.* **2011**, *509*, 127−132.
(34) Eisenmesser, E.; Millet, O.; Eisenmesser, E.; Millet, O.; Labeikovsky, W; Korzhnev, D.; Wolf-Watz, M.; Bosco, D.; Skalicky, J.; Kay, L.; Kern, D. *Nature* **2005**, *438*, 117−121.
(35) Lange, O.; Grubmuller, H. *Proteins* **2008**, *70*, 1294−1312.
(36) Rod, T.; Radkiewicz, J.; Brooks, C. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 6980−6985.
(37) Agarwal, P.; Geist, A.; Gorin, A. *Biochemistry* **2004**, *43*, 10605−10608.
(38) del Sol, A.; Tsai, C.-J.; Ma, B.; Nussinov, R. *Structure* **2009**, *17*, 1042−1050.
(39) Stefan, C.; Jansen, S.; Bollen, M. *Trends Biochem. Sci.* **2005**, *30*, 542−550.