JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Article

pubs.acs.org/JACS

Disorder-to-Order Transition of an Intrinsically Disordered Region of Sortase Revealed by Multiscale Enhanced Sampling

Kei Moritsugu,*[,†] Tohru Terada,[†,‡] and Akinori Kidera[†,§]

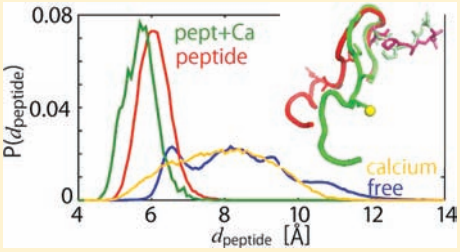
[†]Research Program for Computational Science, RIKEN, 2-1 Hirosawa, Wako, Saitama 351-0198, Japan

[‡]Graduate School of Agricultural and Life Sciences, The University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku, Tokyo 113-8657, Japan

[§]Graduate School of Nanobioscience, Yokohama City University, 1-7-29 Suehiro-cho, Tsurumi-ku, Yokohama 230-0045, Japan

Supporting Information

ABSTRACT: Molecular functions of intrinsically disordered proteins (IDPs) or intrinsically disordered regions (IDRs), such as molecular recognition and cellular signaling, are ascribed to dynamic changes in the conformational space in response to binding of target molecules. Sortase, a transpeptitase in Gram-positive bacteria, has an IDR in a loop which undergoes a disordered-to-ordered transition (called "disordered loop"), accompanying a tilt of another loop ("dynamic loop"), upon binding of a signal peptide and a calcium ion. In this study, all-atom conformational ensembles of sortase were calculated for the four different binding states (with/without the peptide and with/without a calcium ion) by the multiscale enhanced sampling (MSES) simulation to examine how the binding of the peptide and/or calcium influences the conformational ensemble. The MSES is a multiscale and multicopy simulation method that allows an enhanced sampling of the all-atom model of large proteins including explicit solvent. A 100 ns MSES simulation of the ligand-free sortase using 20 replicas (in total 2 μs) demonstrated large flexibility in both the disordered and dynamic loops; however, their distributions were not random but had a clear preference which populates the N-terminal part of the disordered loop near the bound form. The MSES simulations of the three binding states clarified the allosteric mechanism of sortase: the N- and C-terminal parts of the disordered loop undergo a disorder-to-order transition independently of each other upon binding of the peptide and a calcium ion, respectively; however, upon binding of both ligands, the two parts work cooperatively to stabilize the bound peptide.



INTRODUCTION

Intrinsically disordered proteins (IDPs) are a class of proteins which are biologically functional but lack in well-defined structures in contrast to the conventional "structure−function paradigm".[1−5] Low hydrophobicity of the amino acids composition results in the "intrinsic disorder" of IDPs or intrinsically disordered regions (IDRs) in folded proteins.[3,6] Such flexibility is considered to facilitate conformational adaptability depending on various binding partners (such as ligands and receptor proteins) via so-called "coupled folding and binding" or "fly casting",[7,8] yielding the associated functions of IDPs/IDRs, such as molecular recognition and cellular signaling. Since molecular recognition may be determined by subtle atomic interactions, it is necessary to understand the molecular mechanism of the allosteric regulation of IDPs/IDRs on the basis of the conformational ensembles of a complete set of binding states at all-atom resolution under physiological condition, which allows the dynamic changes of the free-energy landscape via bindings of the target molecules to be fully described.

The conformational ensemble of IDPs/IDRs has been studied by various experimental methods,[3,9] such as spectroscopic techniques (nuclear magnetic resonance (NMR),[10] electron paramagnetic resonance,[11] and circular dichroism),[12] atomic force microscopy,[13] Förster resonance energy transfer,[14,15] and small-angle X-ray scattering.[16] Combined with these experimental methods, two classes of computer simulations have also been successfully utilized to construct high-resolution structural models. The first class aims at modeling a few representative structures so as to reproduce the associated experimental results.[17−20] Although such structural modeling has clarified how the disordered regions are flexible, the structural models were limited to a small number and thus not sufficient to describe the coupling between binding of a target molecule and local folding of the disordered region. The other class of the simulations for IDPs/IDRs uses enhanced sampling simulations, such as generalized ensemble methods,[21−23] which generate broad ranges of structures that are not trapped by local potential minima. However, due to the poor scalability of conventional enhanced sampling methods, the simulation system sizes were limited to those of small peptides, such as amyloid-β,[24] histone tails,[25] a neural restrictive silencer factor,[26] and the N-terminal domain of tumor repressor p53.[27] To alleviate the scalability problem, various techniques, such as using the implicit solvent model,[8] restricting the conformational space by restraining the target protein,[26] or fixing a part of the disordered region[28] and using a

Received: January 26, 2012
Published: April 2, 2012

   dx.doi.org/10.1021/ja3008402 | J. Am. Chem. Soc. 2012, 134, 7094−7101

coarse-grained model[27] or simplified force field,[29] have been adopted.

In the present study, we propose a full conformational sampling of a protein with an IDR at atomic resolution with explicit solvent using the multiscale enhanced sampling (MSES) simulation.[30] The MSES consists of a combination of two systems: an all-atom system containing a protein molecule and surrounding solvents (MM) and a coarse-grained system (CG), described by Hamiltonian H, given as

$$H = H_{MM} + H_{CG} + k_{MMCG}[\theta(\mathbf{r}_{MM}) - \mathbf{r}_{CG}]^2 \quad (1)$$

where H_{MM} and H_{CG} are the Hamiltonians for MM and CG, respectively, and the number of degrees of freedom in CG, M, is considered to be much smaller than that in MM (N; where $N \gg M$). The CG model can be arbitrarily chosen according to prior knowledge or experimental information. In this study, the CG model may represent the disordered region undergoing the disorder-to-order transition using only a small number of degrees of freedom, i.e., C_α atoms in this study. The last term in eq 1 defines the coupling between the MM and CG coordinates ($\mathbf{r}_{MM}$ and $\mathbf{r}_{CG}$, respectively) with force constant k_{MMCG} for driving the MM system by the accelerated dynamics of the CG system, where the M-dimensional vector $\boldsymbol{\theta}(\mathbf{r}_{MM})$ is a projection of $\mathbf{r}_{MM}$ onto the CG space. When the CG model is coupled with only a limited number of degrees of freedom in the MM model, e.g., the disordered region in the present study, the enhancement of sampling in the MM model mostly occurs in the corresponding part of the MM model, or the disordered region, while the other part moves in the same manner as the standard MD simulation since the coupling with the other part does not exist.

The ultimate goal of the simulation is to derive the free energy surface (FES) solely from H_{MM} without any bias due to the coupling with CG. It is therefore necessary to eliminate the influence from the coupling or to extrapolate the system to the coupled Hamiltonian with $k_{MMCG} = 0$. For this purpose, the Hamiltonian replica exchange method was adopted in which many replicated systems are assigned by different k_{MMCG} values ranging from a large value to zero.[31] The exchange probability between replicas m and n satisfying the detailed balance condition is given by

$$p_{mn} = \min(1, \exp(\Delta_{mn})) \quad (2)$$

with

$$\Delta_{mn} = \beta(k^m_{MMCG} - k^n_{MMCG})([\boldsymbol{\theta}(\mathbf{r}^m_{MM}) - \mathbf{r}^m_{CG}]^2 - [\boldsymbol{\theta}(\mathbf{r}^n_{MM}) - \mathbf{r}^n_{CG}]^2) \quad (3)$$

where β is inverse temperature. Equation 3 indicates that the exchange probability is determined by the difference between $\boldsymbol{\theta}(\mathbf{r}_{MM})$ and $\mathbf{r}_{CG}$, both of which are defined in a small dimension relating to the CG model (i.e., M). Because $M \ll N$, Δ_{mn} can be kept small enough to give a high exchange probability p_{mn} irrespective of the size of the system, N. This condition guarantees excellent scalability superior to that of the conventional temperature replica exchange method, where the difference in the potential energy of MM of the neighboring replicas (scaling up with N^2) determines the exchange probability, Δ_{mn}. In our previous study, MSES simulation was applied to the folding dynamics of a miniprotein, chignolin.[30] Simulation of the fully solvated system using the same number of replicas as the implicit solvent simulation demonstrated significantly improved scalability.

In the present study, MSES was applied to conformational sampling of a folded protein with an IDR, sortase, which is a member of the IDPs/IDRs database, DISPROT.[32] Sortase is a transpeptidase in Gram-positive bacteria and cleaves a C-terminal sorting signal of surface proteins at a conserved LPXTG motif with the help of a calcium ion, leading to anchoring of the bacteria to the cell wall peptideglycan.[33] NMR solution structures with and without a signal peptide (LPAT)[34] revealed the existence of a disordered loop in the unbound state that undergoes a disorder-to-order transition upon a peptide and a calcium ion bindings together with a large tilt of another loop (named dynamic loop) (Figure 1). It was reported that

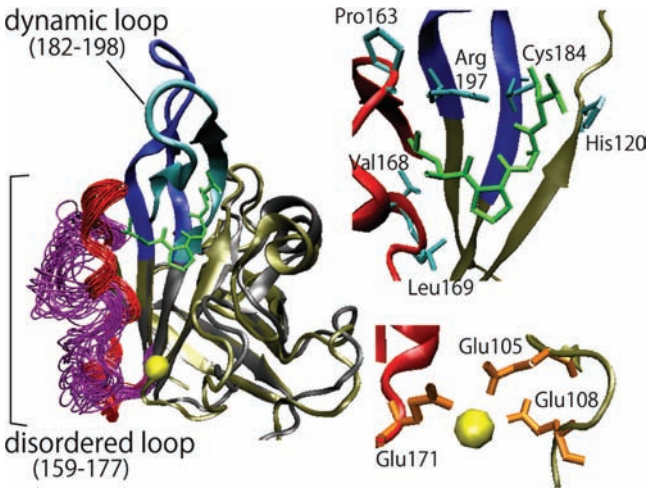


Figure 1. (left) NMR solution structures of the unbound (gray: 1IJA) and bound (dark yellow: 2KID) sortase. The disordered and dynamic loops are colored by magenta/red and cyan/blue for the unbound/bound forms, respectively. The bound peptide and calcium ion are also shown by green and yellow. (right) Close-up views of the residues forming the peptide binding pocket (Pro163, Val168, Leu169/Arg197, and His120/Cyc184) and those binding to a calcium ion (Glu105/Glu108 and Glu171).

binding of a calcium ion functions as an allosteric regulator, increasing the enzyme activity by eight times.[35] In the present study, a large-scale all-atom conformational sampling of the disordered loop was conducted in the ligand-free form as well as the three binding states (peptide, calcium ion, and both). Through the disorder-to-order transitions between the ligand-free form and each of the three binding states observed in the four simulations, we tried to elucidate the influence of binding of the peptide and/or a calcium ion to the free-energy landscape of sortase, particularly focusing on the allosteric effects of a calcium ion on the peptide binding.

RESULTS

Disordered and Dynamic Loops Undergo Structural Change via Substrate Binding. Comparison of the NMR models in the ligand-unbound and the ligand-bound forms (Figure 1) shows the existence of a remarkably flexible loop and a substrate-induced conformational change in the two loop regions, while the β-sheet-rich rigid domain is relatively stable.[34] To clarify these features, the 24 residues comprising the four β-sheets (residues 74−78, 83−87, 141−146, and 149−156) were designated as "rigid domain", since the average root-

dx.doi.org/10.1021/ja3008402 | J. Am. Chem. Soc. 2012, 134, 7094−7101

mean-square deviation (RMSD) for C_α atoms of the rigid domain between the 20 bound and 25 unbound NMR models in the PDB entry 1IJA[36] is only 0.66 Å. Figure S1 in the Supporting Information shows the RMSD values obtained with the bound and unbound NMR models after fitting them to the rigid domain, indicating large conformational changes in residues 159−177 ("disordered loop", corresponding to the disordered region defined in DISPROT) and residues 182−198 ("dynamic loop"). The structural changes in the loops were thus evaluated as the relative position to the rigid domain or by superimposing the rigid domain, i.e., the RMSD values include the contributions from the external translation and rotation of the loops.

MSES Simulation. The MSES simulation of the ligand-free sortase in explicit solvents (the MM system contains approximately 35 000 atoms) was performed for 100 ns using 20 replicas (total simulation time: 2 μs) for enhancing the sampling of the disordered and dynamic loops. This aims at identifying the distribution of the flexible disordered loop and the dynamical correlation that realizes the allosteric influence on the disordered and dynamic loops.

The energy distribution of the coupling term (the third term on the right-hand side of eq 1 determining exchange probability) showed significant overlaps of the probability distributions of neighboring replicas (Figure S2A in the Supporting Information). This guarantees high exchange probability (see eq 2) or a successful Hamiltonian exchange simulation; the average acceptance ratio of the exchange resulted in 0.17. The unbiased MM ensemble (k_{MMCG} = 0) showed that the rigid domain was stable (C_α RMSD for the rigid domain was 1.4 Å) (Figure S2B in the Supporting Information) and that the disordered and dynamic loops fluctuated largely (Figures S2C and D in the Supporting Information), indicating that the sampling achieved large enhancement; C_α RMSD from the NMR-bound form extended over the ranges of 2−10 Å (the disordered loop) and 1−15 Å (the dynamic loop). The convergence of the distribution was confirmed by comparison of the distributions calculated by the first half and the second half of the trajectory (Figure S3 in the Supporting Information).

The FES along the reaction coordinate linearly connecting the bound and unbound structures (d_{BU} = 0 and 1 for the bound and unbound states, respectively; Figure 2) illustrates the sampling process used in the MSES simulations. The potential energy of the CG model used here was a mixture of two harmonic potentials around the unbound and bound structures, respectively, with a high mixing temperature, β^{-1} in eq 4 (see Methods Section for details). Using a high mixing temperature erased the trace of the potentials of the bound and unbound states and made frequent back and forth motions between the two structures, as seen in the FES of the simulation using only the CG potential, or the single peaked distribution between the unbound and bound structures. This, together with relatively small force constants for the two harmonic potentials for the bound and unbound structures, allows the CG model to cover a wide range including both structures, with having almost no influence from the choices of the associated reference structures. The strong coupling with the CG model (a large value of k_{MMCG}) drives the MM model to sample the conformational space, which is closely similar to that of the CG model. Decreasing the coupling constant k_{MMCG} or changing the potential to the intrinsic MM force field derives flat distribution of the structural ensemble covering both the unbound and bound structures and extending further beyond the unbound structure, which may represent the real feature of disorder. Hamiltonian replica exchange with gradually decreasing k_{MMCG} will relax the V_{CG} bias, eliminating the dependences of the function of V_{CG} and the choices of the reference structures for the CG force field.

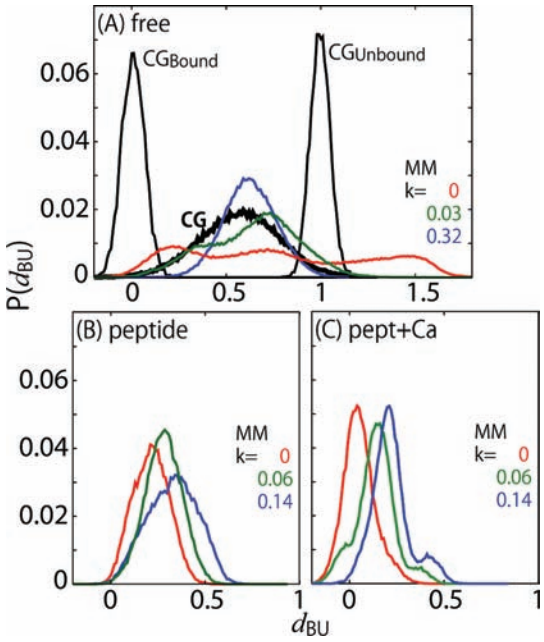


Figure 2. Probability distributions along the reaction coordinate, d_{BU}, obtained in the MSES simulations of (A) free, (B) peptide, and (C) pept + Ca states; shown are the distributions for three values of k_{MMCG} (in units of kcal/mol/Å^2) and those of the CG model used together with those of the CG models of the bound and unbound structures.

In addition to the free state, the MSES simulations were also applied to the three bound states, namely, with the signal peptide (designated as "peptide"), a calcium ion ("calcium"), and both the signal peptide and a calcium ion ("pept + Ca"), respectively, for 50 ns each. Since the dynamic loop has to be in the tilted bound form to accommodate the peptide, the MSES simulations of the peptide bound states used the restraints between MM and CG for the disordered loop only, namely, not for the dynamic loop, and the initial MM structure of the dynamic loop were set in the bound form. This condition made it possible to decrease the number of replicas to 12. Total simulation time was thus 0.6 × 3 μs. High exchange rates of replicas were also confirmed through the overlaps of the probability distributions of the coupling energy between neighboring replicas; the average acceptance ratios of the exchange were 0.16, 0.14, and 0.15 for the peptide, calcium, and pept + Ca states, respectively.

Ensembles of peptide and pept + Ca are shown in Figure 2B,C. These simulations used the same CG potential as that used in the simulation in the free state; however, they differ in terms of the existence of the bound peptide and calcium ion in the MM systems. This difference derived much narrower distributions than the free state, but the simulation of the peptide state could not achieve the distribution corresponding to the NMR bound structures (Figure 2B). Instead, an addition of the calcium ion bound was found to induce the perfect bound state (Figure 2C). This role of the allosteric effector of the calcium ion will be discussed below in more detail.

Disordered Loop Is Flexible in the Free State. In Figure 3A, the energy landscape observed in the one-dimensional reaction coordinate in Figure 2 is extended to the two-dimensional space spanned by the two RMSD values for the disordered and dynamic loops, respectively, from the bound structure. It was confirmed that the choice of the reaction coordinate did not affect the overall view of the landscape. The conformational ensemble obtained by the MSES simulation contains the ensemble of the NMR unbound structures and extends over a much larger conformational space, but it does not attain the NMR bound structures. The root-mean-square fluctuation (RMSF) of the C_α atoms in the disordered loop is 3.9 Å, much larger than the RMSF value, 2.1 Å, observed in the NMR models (1IJA).[34] The snapshot structures in the simulation are given in Figure S4 in the Supporting Information. The conventional equilibrium MD simulations of the free state on the same time scale (100 ns), starting from either the unbound structure or the bound structure, did not converge to a single ensemble but stayed at separate local areas (Figure 3B,C, respectively). In the simulation started from the bound structure (Figure 3B), the disordered loop largely deviated from the initial bound structure, though the dynamic loop almost maintained its bound form. This result may be due to the flexibility of the disordered loop as well as the charge repulsion of three glutamic acids at the calcium binding site because the ligands were removed in the simulation (see below for more detail). The simulation started from the unbound structure (Figure 3C) could not reach the bound structure, favoring the entropic contributions from the flexible loops. These equilibrium MD results demonstrate the necessity of the MSES to achieve a broad conformational sampling in sortase.

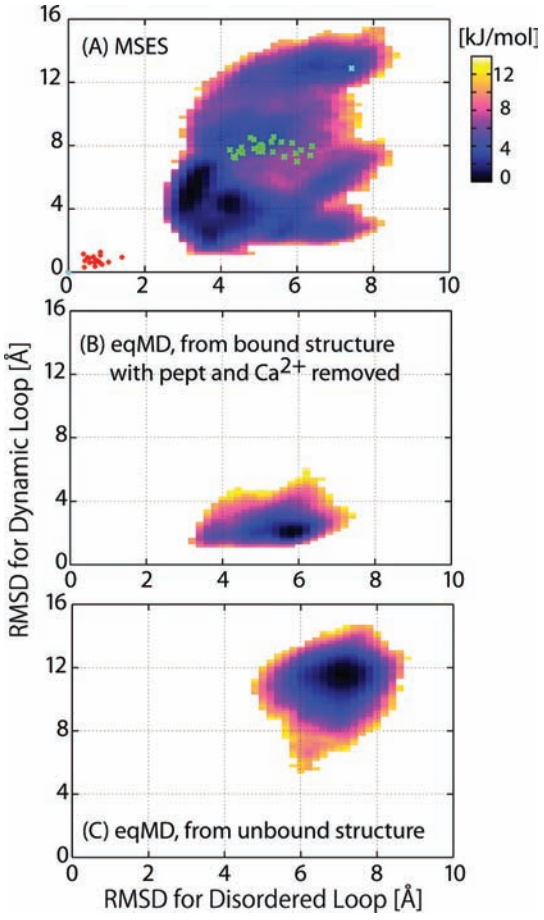


Figure 3. (A) Two-dimensional FESs along the C_α RMSDs of both the disordered and dynamic loops from the bound structure, calculated from the MSES simulation of the free state. The NMR bound and unbound structures and the two reference structures for the CG force field are depicted on the surface by red, green, and cyan dots, respectively. Those from the equilibrium MD simulations starting at (B) the bound NMR structure (but with the bound peptide and Ca^{2+} removed) and (C) the unbound NMR structure are also shown.

Hereafter, the dynamic correlation between the disordered and the dynamic loops in the free state is investigated. The correlation coefficient between the two RMSD values in Figure 3A is 0.40, although no correlation was imposed on the CG model. This significant correlation was originated from a large population close to the bound structure, i.e., the area of 2−5 and 1−6 Å RMSD for disordered and dynamic loops, respectively. Figure S5 in the Supporting Information shows more clearly that the disordered loop is likely to be close to the bound form when the dynamic loop is also close to the bound form, while it shows little correlation when the dynamic loop is far from the bound form. This large population, or high correlation between the disordered and dynamic loops around the bound form, was originated from the interactions between Pro163 in the disordered loop and Arg197 in the dynamic loop (see Figure 1), which occurred when the dynamic loop approached the bound form and eventually produced the peptide binding pocket in the bound structure. This feature was also seen in the highly positive motional correlation between the N-terminal residues 159−165 in the disordered and the dynamic loops. It suggests that the two parts moved together (see the residue-level correlation map in Figure 4), although the other part in the disordered loop (C-terminal residues 166−177) had negative correlation, indicating the two loops' association and dissociation motions. This result indicates that, in the case of sortase, the disordered state is not fully disordered; instead, it has a significant propensity staying near the bound form at the N-terminal part as if it is ready to bind ligand molecules.

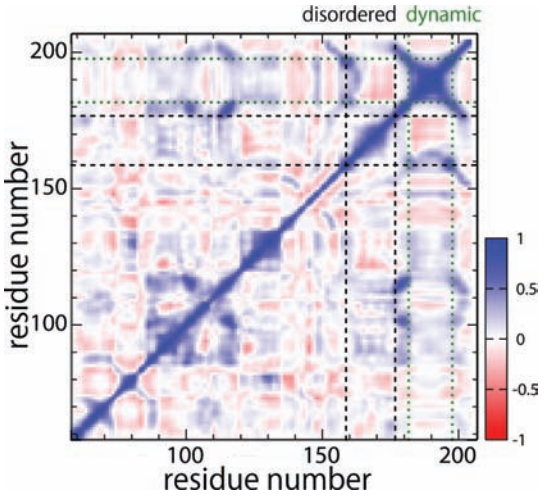


Figure 4. Dynamic correlation map of sortase C_α atoms, calculated from the MSES simulation of the free state.

Dynamic Loop Moves as a Rigid Body. During the MSES simulation, the dynamic loop was found to move almost as a rigid body (Figure S6 in the Supporting Information); C_α RMSD of the loop from the bound structure is <2.5 Å (Figure

dx.doi.org/10.1021/ja3008402 | J. Am. Chem. Soc. 2012, 134, 7094−7101

S2E in the Supporting Information). This rigidity comes from the hydrogen bonds forming a β-sheet (drawn in Figure S6 in the Supporting Information). Figure S6A in the Supporting Information shows that the loop region is also very rigid, with the local C_α RMSD being <0.8 Å. The dynamic loop therefore moves as a rigid body upon the peptide binding. This behavior is completely different from that of the disordered loop.

Disorder-to-Order Transition upon Ligand Binding. The disorder-to-order transition of the disordered loop upon binding of the signal peptide and/or the calcium ion is then discussed. The transition can be observed in the C_α RMSF values of the disordered loop; the fluctuations were significantly suppressed upon binding from 3.9 to 2.0, 1.8, and 1.6 Å for calcium, peptide, and pept + Ca states, respectively. The snapshots from these simulations demonstrate these ordering transitions (Figure S4 in the Supporting Information). This ordering transition from the free state to the peptide or the pept + Ca state has already been observed, see Figure 2, which shows that binding of the peptide was not sufficient for the disordered loop to form the stable bound structure and that the additional binding of a calcium ion was needed.

To characterize the role of calcium binding in the ordering transition of the disordered loop at the local atomic interaction level, the two-dimensional FES describing the formation of the peptide and the calcium binding pockets was calculated as a function of two distances: first, the C_α distance between Pro163 and Arg197, indicating the formation of the peptide binding site (designated as $d_{peptide}$), and, second, the distance between the side-chain C_δ atoms of Glu171 and Glu105 or Glu171 and Glu108 as an indicator of the calcium-binding site (the smaller value of the two distances was designated as d_{ca}).[35] Figure 5A shows a broad FES for the free state, which also covers most of the NMR models in the unbound form, while this is apart from the distribution of the NMR bound structures, as is also in Figure 3A. The dominant distribution of $d_{peptide}$ being approximately 6 Å corresponds to the distributions around a small RMSD for the disordered loop (see Figure S5 in the Supporting Information), indicating the formation of the interactions between Pro163 and Arg197. Once the signal peptide binds to sortase (the peptide state; Figure 5B), the FES converges to a single basin with small values of $d_{peptide}$ corresponding to the bound form. However, this basin is located in the area of large d_{ca} values, even larger than the lower limit in the free state (Figure 5A). Lower values of d_{ca} were obtained only with finite values of k_{MMCG} or in the CG-biased replicas. This result clearly shows that the peptide binding does not induce the formation of the pocket for a calcium ion. Incidentally, the crystal structures of the free state (PDB: 1T2P (chains A, B, and C) and 1T2W (chains B and C)) and the peptide state (with no calcium ion; 1T2W (chain A)) show similar trends: a decrease in $d_{peptide}$ from 7.0 to 7.2 Å in the free state to 6.6 Å in the peptide state, while d_{ca} did not show a definite trend from 14.1−17.7 to 17.3 Å.[30] When looking at the calcium state in Figure 5C, we found that the two distributions of $d_{peptide}$ in the free and the calcium sates cover almost the same region, though the calcium state extended the lower limit of $d_{peptide}$ to approximately 5 Å. This means that the induction of the formation of the peptide binding pocket by calcium binding is limited to a small range. In summary, binding of a single ligand, either the peptide or a calcium ion, appears not to confine the conformational region to favor binding of the other ligand. However, when comparing peptide with pept + Ca, we observed a shift of the distribution to the lower value of $d_{peptide}$

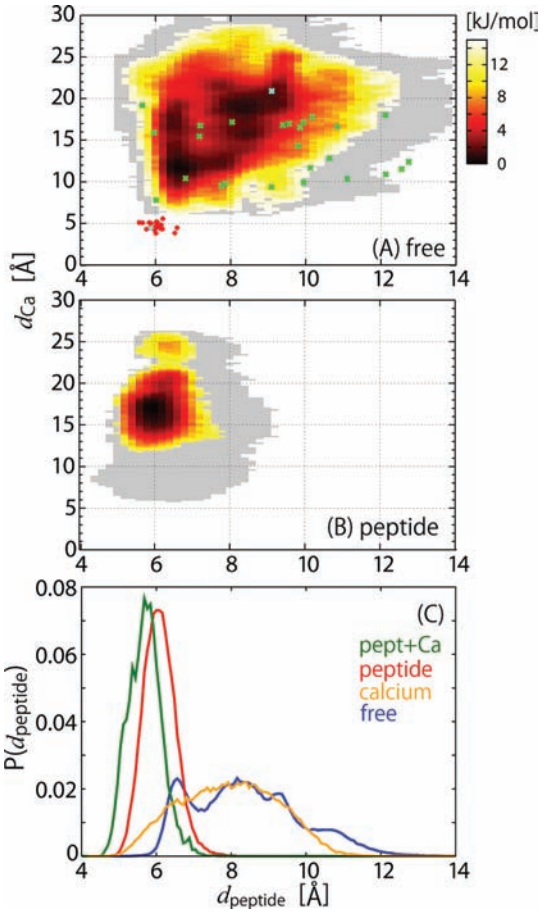


Figure 5. (A) Free energy profiles along $d_{peptide}$ (C_α distance between Pro163 and Arg197; see Figure 1) and d_{ca} (minimum value of two distances of side-chain C_δ atoms, one between Glu171 and Glu105 and the other between Glu171 and Glu108) are shown from the MSES simulations of sortase for the free and peptide states. The NMR bound and unbound structures and the two reference structures for the CG force field are depicted on the surface by red, green, and cyan dots, respectively. Those of the replicas with the finite k_{MMCG} values are shown in gray. (B) Probability distributions along $d_{peptide}$ for free, peptide, calcium, and pept + Ca states.

when a calcium ion binds. This clearly indicates an allosteric effect of the calcium binding to the peptide binding, which is in good agreement with experimental data that shows a calcium ion increases peptide binding affinity by eight times.[35]

The atomic details of the allosteric effect of calcium binding can be seen in the comparison of the disordered loop between peptide and pept + Ca. Figure 6 shows that the binding of Glu171 to the calcium ion caused a partial folding of a 3_{10}-helix in the disordered loop (residues 166−168 according to DSSP (Define Secondary Structure of Proteins)),[38] making the methyl groups of Leu in the peptide form a tighter hydrophobic binding to the disordered loop. In addition, the calcium binding caused the shift of the C-terminal of the loop in the direction of the peptide, and this motion in turn made Val168 participate in the peptide binding; the distance between the C_β atom of Val168 in the disordered loop and the C_α atom of Leu in the peptide decreased from 7.5 ± 0.8 to 5.5 ± 0.7 Å. In a mutation study,[39] these two residues were shown to be important factors in the enzymatic activity.

As shown in the dynamic correlation (Figure 4), the disordered loop can be divided into N-terminal (159−165)

dx.doi.org/10.1021/ja3008402 | J. Am. Chem. Soc. 2012, 134, 7094−7101

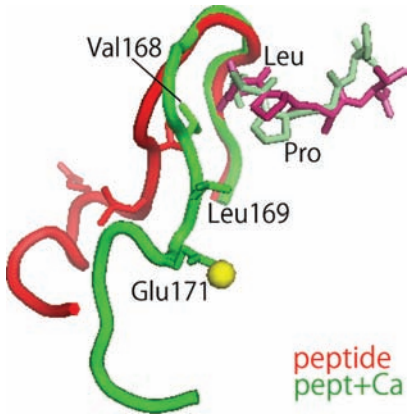


Figure 6. Representative structures of the disordered loop and the bound peptide for the peptide (red) and the pept + Ca (green) states.

and C-terminal (166−177) parts. The N-terminal part moves together with the dynamic loop to form the peptide binding site even in the free state (also see Figure 3), and it makes hydrophobic interactions with the peptide upon binding. It is also noticed that the distribution of the free state (Figure 5A) covers the bound form of the NMR structure with regard to the peptide binding site, or $d_{peptide}$. Therefore, the distribution and the correlation in motion together indicate that the peptide binding process can be explained by the population-shift model.[40] On the other hand, the C-terminal part maintains its flexibility even after the peptide binds and requires the calcium binding to form the ligand bound structure by neutralizing the charge repulsions between the negative charges of the side chains of Glu171 and Glu105/108. This may be classified as the induced-fit model.[41] These findings are consistent with the previous coarse-grained simulation result, which shows that stronger and longer-ranged interaction favors the induced-fit model, while weaker and short-ranged interactions lead to the population-shift model.[42] These observations indicate that the disordered loop consists of two parts, the N- and C-terminal, which undergo the disorder-to-order transition independently of each other upon binding of one of the ligands and work cooperatively to stabilize the bound peptide when the two ligands bind.

Finally, the side-chain packing between the disordered loop and the peptide, which was not included in the C_α-level CG model, is examined on the basis of the nuclear Overhauser effect (NOE) data of the NMR experiment.[35] Table 1 compares the simulated distances averaged by the weight of r^{-6} for pept + Ca and peptide states with the experimental data.[35] In the peptide state, the number of distances satisfying the NMR experimental data for the bound state is only seven. This number increases to 12 in the pept + Ca state. The number of distances which show decreases upon binding of calcium amounts to 14. The agreement with the experimental data indicates that the MSES simulation predicted sufficiently reliable side-chain conformations at the peptide binding site.

■ CONCLUDING REMARKS

Conformational sampling of sortase, a folded protein having an intrinsically disordered region, in the four binding states with the signal peptide and/or a calcium ion bound/unbound clarified the mechanism of the disorder-to-order transition at atomistic resolution. Ligand-free sortase was shown to have large flexibility in both disordered and dynamic loops, but these

Table 1. Experimental and Simulation Hydrogen−Hydrogen Distances for Peptide Binding[a]

hydrogen in disordered loop	hydrogen in peptide	$\langle r\rangle_{exp}$ upper bound	$\langle r\rangle_{MSES}$ pept + Ca	$\langle r\rangle_{MSES}$ peptide	$\langle r\rangle_{MSES}$ pept + Ca− peptide
Pro163HA	Leu207HD	3.2	**2.6**	3.30	**-0.68**
Pro163HA	Leu207HB	5.0	**3.2**	**4.79**	**-1.62**
Pro163HA	Leu207HG	5.0	**2.7**	**3.53**	**-0.86**
Pro163HB	Leu207HD	5.0	**4.2**	**4.03**	0.21
Asp165HN	Leu207HD	3.4	3.9(3.5)	4.86	**-0.99**
Asp165HN	Leu207HG	3.3	**2.3**	4.65	**-2.40**
Val166HA	Leu207HD	3.2	4.6 (4.4)	3.27	1.34
Val166HA	Leu207HB	3.3	3.9(**3.0**)	3.60	0.26
Val166HA	Leu207HG	2.7	**2.0**	4.20	**-2.21**
Val166HN	Leu207HD	5.0	**4.5**	**4.07**	0.41
Val166HN	Leu207HG	3.5	3.9(**3.5**)	4.84	**-0.98**
Gly167HN	Leu207HD	5.0	5.1 (**4.8**)	**4.61**	0.52
Val168HB	Leu207HD	3.8	**2.9**	4.49	1.63
Val168HG	Leu207HG	3.2	**2.9**	4.54	**-1.66**
Val168HG	Leu207HB	5.5	**2.6**	**3.70**	**-1.08**
Val168HG	Leu207HD	3.3	**3.2**	3.44	**-0.21**
Val168HG	Pro208HD	5.5	**2.4**	**5.23**	**-2.79**
Leu168HD	Leu207HA	3.3	4.3(4.0)	15.1	**-10.76**
Leu168HD	Pro208HD	3.3	4.3(3.9)	11.2	**-6.86**
Leu168HD	Pro208HG	3.2	3.6(3.3)	12.2	**-8.61**

[a]Hydrogen−hydrogen distances (in units of Å) between the disordered loop and the peptide derived from NOE data of the NMR experiment for the bound structure ($\langle r\rangle_{exp}$, ref 34) and from the MSES simulations for the pept + Ca and the peptide states ($\langle r\rangle_{MSES}$); $\langle r\rangle_{exp}$ is the upper bound distance derived from the NOE intensity, and $\langle r\rangle_{MSES}$ was calculated as the average of $1/r^6$ over the MSES trajectory. Bold characters indicate $\langle r\rangle_{MSES}$ satisfying the experimental range. Bold characters in the right-most column indicate that the binding of calcium improved the distance evaluations. The averages of $1/r^6$ over the smaller half of the distances are shown in parentheses.

loops are not fully random; the N-terminal part (residues 159−165, including the peptide binding site) moves in correlation with the dynamic loop due to the interactions between Pro163 and Arg197, while the C-terminal part of the disordered loop (residues 166−177, where the calcium ion binds) has negative correlation to the dynamic loop. The binding of the peptide and/or a calcium ion drastically narrows the distribution of the disordered loop relative to that in free state, inducing the disorder-to-order transition as well as the side-chain packing between the peptide and the disordered loop. The N-terminal part moves together with the dynamic loop to form the peptide binding site even in free state, suggesting the population shift-like behavior,[40] while the C-terminal part requires calcium binding to neutralize the electrostatic interactions and to form the ligand bound structure, reminding us of the induced-fit model.[41] Each of the peptide and the calcium ion was found not to induce the formation of the binding site for the other ligand; the distribution of the N-terminal of the disordered loop was only slightly altered by calcium binding and that of the C-terminal was almost unaffected by binding of the peptide. However, as shown in the simulation of pept + Ca, the ordered structure of the N-terminal induced by the peptide binding and that of the C-terminal induced by a calcium ion appear to work cooperatively to stabilize the bound peptide. This apparent cooperativity is originated from the fact that Val168 contributing to the peptide binding is in the C-terminal of the disordered loop, and its peptide binding structure is formed

dx.doi.org/10.1021/ja3008402 | J. Am. Chem. Soc. 2012, 134, 7094−7101

by calcium binding, not by peptide binding. This picture of the allosteric regulation is different from the conventional definition. Usually two distant parts of a protein performing allostery have dynamic correlation with each other, and this correlation is the fundamental basis to explain the mechanism of the allostery. On the contrary, in the case of sortase, two building blocks are formed almost independently by the two ligands, and dynamic correlation is small; however, the effector binding results in the formation of an additional binding site for the ligand. This will result in an experimentally observed allosteric mechanism of sortase that the bound calcium increases the affinity of the peptide binding.

These biological implications of the ligand-binding mechanism in sortase were successfully derived by MSES simulation. This multiscale simulation is based on the coupling of MM with CG as the driving force. The CG force field as a prior knowledge, which can be based on experimental data, is a key of the definition of the low-dimensional "essential" subspace, which is tuned subsequently via the fine MM force field. Most importantly, the smallness of the essential subspace guarantees the scalability in the Hamiltonian exchange. The flexibility of the CG model or the essential subspace should provide an optimal reaction coordinate (RC), which may be much superior to predefined RCs, such as a linear interpolation of the reactant and the product. In future work we will extend the application of the MSES simulation to full IDPs using a CG model which allows a much larger sampling ability.

METHODS

Simulation Models. The MSES simulations were performed for the conformational samplings of sortase in explicit solvent in the free, pept + Ca, calcium, and peptide states. The starting structure of the free state was taken from Model 1 of the NMR structure in PDB entry 1IJA (NMR unbound structure)[34] and from Model 1 of the NMR structure in 2KID (NMR bound structure)[35] for the three bound states. In the two starting structures, the N-terminal 58 residues were not included in the model, which was reported to be unstructured and embedded in the membrane and thus irrelevant to the binding affinity of the peptide.[43] The starting structures of the peptide and calcium states were set by removing the calcium ion and the peptide, respectively, from the bound structure. The peptide analogue, LPAT, in the bound structure was replaced by the corresponding amino acids with capping the termini by amino and carboxylate groups. Rectangular simulation boxes were constructed with a margin of 12 Å from the boundary of the simulation box. The solution systems of the 4 states contained about 10 000 TIP3P water molecules[44] together with 4 (for the pept + Ca and calcium states) or 2 (for the peptide and free states) chloride ions to neutralize the simulation system.

Potential Energy Functions and Kinetic Parameters for MSES. The Hamiltonian functions, H_{MM} (= V_{MM} + K_{MM}), H_{CG} (= V_{CG} + K_{CG}), and the coupling term, in eq 1 were defined as follows: For the all-atom potential energy V_{MM}, the AMBER ff03 force field was used.[45] The CG potential V_{CG} was prepared by the following three steps: (1) A high-temperature equilibrium MD at 600 K was performed starting from the NMR unbound structure; (2) a representative open form of the disordered loop was chosen from the trajectory (called "unbound structure"; this choice was shown not to affect the results); and (3) V_{CG} was constructed on the basis of the mixed-elastic network model[46] connecting the NMR bound structure and the derived unbound structure by two harmonic potentials, i.e.,

$$V_{CG} = -\beta^{-1}\ln[\exp(-\beta V_{bound}) + \exp(-\beta V_{unbound})] \quad (4)$$

where V_{bound} and $V_{unbound}$ are the potential functions of the C_α elastic network model around the bound and unbound structures, respectively, and β^{-1} was set to be sufficiently high (0.014 kcal/mol). The force constant and the cutoff length in the elastic network model were set to 1.8 kcal/mol/Å^2 and 11.5 Å, respectively. The coupling term (eq 1 and eq 3) contains 456 pairwise distances between 24 C_α atoms in the rigid body and 19 C_αs in the disordered loop and 240 distances between 24 C_αs in the rigid body and 10 C_αs in a part of the dynamic loop (residues 187−196), which constrain MM to be the same as the CG model in terms of the relative configurations of both the disordered and the dynamic loops from the rigid domain. The kinetic term of the CG model, K_{CG}, was determined after several test calculations as T_{CG} = 600 K, k_{CG} = 2, and m_{CG} = 1000. These values were chosen so that a large inertial force of the CG model drives the MM model.

Computations. The MSES simulations were performed using the class library for multicopy and multiscale MD simulations.[47] The MM simulations were performed under constant temperature and pressure (NPT) condition (T = 300 K and P = 1 atm), using Berendsen's thermostat and barostat[48] with a relaxation time of 1 ps and with the particle mesh Ewald method[49] for electrostatic interactions. The simulation time step was 2 fs, the extension of which was allowed by constraining bonds involving hydrogen atoms via the SHAKE algorithm.[50] The positions of the peptide and/or calcium ion were maintained at the proper binding site of the protein by distance restraints[51] with the potential V_{const}

$$\begin{aligned} V_{const} &= 0 && d < d_{up} \\ &= k(d - d_{up})^2 && d \geq d_{up} \end{aligned} \quad (5)$$

where atomic distance d is bounded by upper bound d_{up}. The peptide was restrained by all atom pairs between the peptide and the rigid domain within 4 Å from any atom of the peptide in the bound structure. Thus, the motion of the disordered loop did not receive a direct influence from the restraints, but it was affected by the nonbonded interactions with the peptide. On the contrary, the calcium ion was required to be restrained by all atom pairs including the disordered loop. This is because the simulation time, 50 ns, was not sufficient to form the correct coordination structure of the calcium ion including many surrounding side-chains. Thus, the restraints were imposed for all pairs of atoms within 5 Å from the calcium ion. In this sense, the calcium and pept + Ca simulations were considered as the samplings of the disordered loop including the constraint to the calcium binding site. The parameters used were: k = 0.5 kcal/mol/Å^2 and d_{up} = 5 Å for the peptide and d_{up} = 6 Å for the calcium ion. The CG simulation was performed by Langevin dynamics with a friction constant of 5 ps^{-1}. For the MSES simulation of the free state, 20 replicas were used with k_{MMCG} = 0, 0.0004, 0.002, 0.004, 0.0064, 0.001, 0.014, 0.02, 0.028, 0.038, 0.05, 0.064, 0.08, 0.1, 0.12, 0.15, 0.18, 0.22, 0.27, and 0.32 kcal/mol/Å^2. For the other three bound states, 12 replicas were used with k_{MMCG} = 0, 0.0016, 0.004, 0.007, 0.012, 0.02, 0.03, 0.046, 0.06, 0.08, 0.0106, and 0.14 kcal/mol/Å^2. The replica was exchanged every 50 ps. Total simulation time was 100 ns for free state and 50 ns for the other three states.

■ ASSOCIATED CONTENT

Supporting Information

Six supporting figures on the RMSD between the unbound and bound NMR sortase, the MSES simulation, and the MSES simulation structures and FESs. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

moritsuguk@riken.jp

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We acknowledge support by the MEXT grand challenge program using next-generation supercomputing. A.K. was supported by MEXT, Grant-in-Aid for Scientific Research, 23247027. The computations were performed on the RIKEN Integrated Cluster of Clusters (RICC) and on TSUBAME2.0 at the Global Scientific Information and Computing Center of Tokyo Institute of Technology.

■ REFERENCES

(1) Dunker, A. K.; Lawson, J. D.; Brown, C. J.; Williams, R. M.; Romero, P.; Oh, J. S.; Oldfield, C. J.; Campen, A. M.; Ratliff, C. M.; Hipps, K. W.; Ausio, J; Nissen, M. S.; Reeves, R.; Kang, C.; Kissinger, C. R.; Bailey, R. W.; Griswold, M. D.; Chiu, W.; Garner, E. C.; Obradovic, Z. *J. Mol. Graphics Modell.* **2001**, *19*, 26−59.

(2) Dyson, H. J.; Wright, P. E. *Nat. Rev. Mol. Cell Biol.* **2005**, *6*, 197−208.

(3) Tompa, P. *Structure and Function of Intrinsically Disordered Proteins*; CRC Press: New York, 2010.

(4) Dunker, A. K.; Gough, J. *Curr. Opin. Struct. Biol.* **2011**, *21*, 1−3.

(5) Uversky, V. N.; Int., J. *Biochem. Cell Biol.* **2011**, *43*, 1090−1103.

(6) Vucetic, S.; Brown, C. J.; Dunker, A. K.; Obradovic, Z. *Protein* **2003**, *52*, 573−584.

(7) Shoemaker, B. A.; Portman, J. J.; Wolynes, P. G. *Proc. Natl. Acad. Sci.U.S.A.* **2000**, *97*, 8868−8873.

(8) Chen, J. *J. Am. Chem. Soc.* **2009**, *131*, 2088−2089.

(9) Uversky, V; Longhi, S. *Instrumental Analysis of Intrinsically Disordered Proteins: Assessing Structure and Conformation*: Wiley-VCH: Weinheim, Germany, 2010.

(10) Dyson, H. J.; Wright, P. E. *Chem. Rev.* **2004**, *104*, 3607−3622.

(11) Jeganathan, S.; von Bergan, M.; Brutlach, H.; Steinhoff, H. J.; Mandelkow, E. *Biochemistry* **2006**, *45*, 2283−2293.

(12) Reeve, R. *Gene* **2001**, *277*, 63−81.

(13) Miyagi, A; Tsunaka, Y.; Uchihashi, T.; Mayanagi, K.; Hirose, S.; Morikawa, K.; Ando, T. *Chem. Phys. Chem.* **2008**, *9*, 1859−1866.

(14) Beechem, J. M.; Haas, E. *Biophys. J.* **1989**, *55*, 1225−1236.

(15) Huang, F.; Rajagopalan, S.; Settanni, G.; Marsh, R. J.; Armoogum, D. A.; Nicolaou, N.; Bain, A. J.; Lerner, E.; Haas, E.; Ying, L.; Fersht, A. R. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *106*, 20758−20763.

(16) Svergun, D. I.; Koch, M. H. J. *Curr. Opin. Struct. Biol.* **2002**, *12*, 654−660.

(17) Mittag, T.; Forman-Kay, J. D. *Curr Opin Struct Biol.* **2007**, *17*, 3−14.

(18) Bernadó, P.; Mylonas, E.; Petoukhov, M. V.; Blackledge, M.; Svergun, D. I. *J. Am. Chem. Soc.* **2007**, *129*, 5656−5664.

(19) Rauscher, S; Pomès, R. *Biochem. Cell Biol.* **2010**, *88*, 269−90.

(20) Fisher, C. K.; Stultz, C. M. *Curr. Opin. Struct. Biol.* **2011**, *21*, 426−431.

(21) Hansmann, U. H. E. *Chem. Phys. Lett.* **1997**, *281*, 140−150.

(22) Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **1999**, *314*, 141−151.

(23) Wang, F.; Landau, D. P. *Phys. Rev. Lett.* **2001**, *86*, 2050−2053.

(24) Sgourakis, N. G.; Merced-Serrano, M.; Boutsidis, C.; Drineas, P.; Du, Z.; Wang, C.; Garcia, A. E. *J. Mol. Biol.* **2011**, *405*, 570−583.

(25) Potoyan, D. A.; Papoian, G. A. *J. Am. Chem. Soc.* **2011**, *133*, 7405−7415.

(26) Higo, J.; Nishimura, Y.; Nakamura, H. *J. Am. Chem. Soc.* **2011**, *133*, 10448−10458.

(27) Terakawa, T.; Takada, S. *Biophys. J.* **2011**, *101*, 1450−1458.

(28) Wells, M.; Tidow, H.; Rutherford, T. J.; Markwick, P.; Jensen, M. R.; Mylonas, E.; Svergun, D. I.; Blackledge, M.; Fersht, A. R. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 5762−5767.

(29) Wu, K. P.; Weinstock, D. S.; Narayanan, C.; Levy, R. M.; Baum, J. *J. Mol. Biol.* **2009**, *391*, 784−796.

(30) Moritsugu, K.; Terada, T; Kidera, A. *J. Chem. Phys.* **2010**, *133*, 224105.

(31) Fukunishi, H.; Watanabe, O.; Takada, S. *J. Chem. Phys.* **2002**, *116*, 9058−9067.

(32) Sickmeier, M.; Hamilton, J. A.; LeGall, T.; Vacic, V.; Cortese, M. S.; Tantos, A.; Szabo, B.; Tompa, P.; Chen, J.; Uversky, V. N.; Obradovic, Z.; Dunker, A. K. *Nucleic Acids Res.* **2007**, *35*, D786−793.

(33) Marraffini, L. A.; Dedent, A. C.; Schneewind, O. *Microbiol. Mol. Biol. Rev.* **2006**, *70*, 192−221.

(34) Suree, N.; Liew, C. K.; Villareal, V. A.; Thieu, W.; Fadeev, E. A.; Clemens, J. J.; Jung, M. E.; Clubb, R. T. *J. Biol. Chem.* **2009**, *284*, 24465−24477.

(35) Naik, M. T.; Suree, N.; Ilangovan, U.; Liew, C. K.; Thieu, W.; Campbell, D. O.; Clemens, J. J.; Jung, M. E.; Clubb, R. T. *J. Biol. Chem.* **2006**, *281*, 1817−1826.

(36) Ilangovan, U.; Ton-That, H.; Iwahara, J.; Schneewind, O.; Clubb, R. T. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 6056−6061.

(37) Zong, Y.; Bice, T. W.; Ton-That, H.; Schneewind, O.; Narayana, S. V. *J. Biol. Chem.* **2004**, *279*, 31383−31389.

(38) Kabsch, W.; Sander, C. *Biopolymers* **1983**, *22*, 2577−2637.

(39) Bentley, M. L.; Lamb, E. C.; McCafferty, D. G. *J. Biol. Chem.* **2008**, *283*, 14762−14771.

(40) Tsai., C. J.; Kumar, S.; Ma, B.; Nussinov, R. *Protein Sci.* **1999**, *8*, 1181−1190.

(41) Koshland, D. *Proc. Natl. Acad. Sci. U.S.A.* **1958**, *44*, 98−104.

(42) Okazaki, K; Takada, S. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 11182−11187.

(43) Ton-That, H.; Liu, G.; Mazmanian, S. K.; Faull, K. F.; Schneewind, O. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 12424−12429.

(44) Jorgensen, W. D.; Chandrasekhar, J.; Madura, J. D. *J. Chem. Phys.* **1983**, *79*, 926−935.

(45) Ponder, J. W.; Case, D. A. *Adv. Protein Chem.* **2003**, *66*, 27−85.

(46) Zheng, W.; Brooks, B. R.; Hummer, G. *Proteins* **2007**, *69*, 43−57.

(47) Terada, T.; Matsunaga, Y.; Moritsugu, K.; Kidera, A. in preparation.

(48) Juffer, A. H.; Botta, E. F. F.; Vankeulen, B. A. M.; Vanderploeg, A.; Berendsen, H. J. C. *J. Comput. Phys.* **1991**, *97*, 144−171.

(49) Darden, T.; York, D.; Pedersen, L. G. *J. Chem. Phys.* **1993**, *98*, 10089−10092.

(50) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327−341.

(51) 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. Comput. Chem.* **2005**, *26*, 1668−1688.