THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

pubs.acs.org/JPCB

Energetics of the Presequence-Binding Poses in Mitochondrial Protein Import Through Tom20

Yasuaki Komuro,[§,‡] Naoyuki Miyashita,[†] Takaharu Mori,[†] Eiro Muneyuki,[§] Takashi Saitoh,[∥] Daisuke Kohda,[∥] and Yuji Sugita*[,‡,†,⊥]

[§]Graduate School of Science and Engineering, Chuo University, 1-13-27, Kasuga, Bunkyo-ku, Tokyo 112-8551, Japan

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

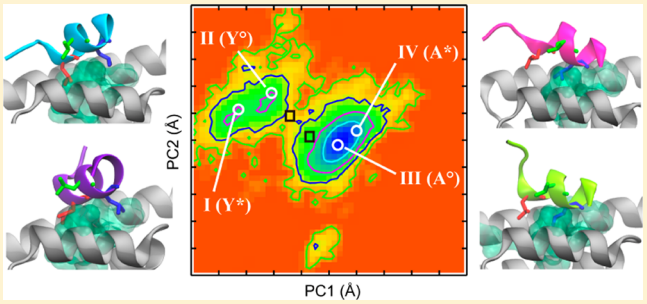
[†]RIKEN Quantitative Biology Center, 7-1-26 minatojima-minamimachi, Chuo-ku, Kobe, Hyogo 650-0047, Japan

[∥]Division of Structural Biology, Medical Institute of Bioregulation, Kyushu University, 3-1-1, Maidashi, Higashi-ku, Fukuoka 812-8582, Japan

[⊥]RIKEN Advanced Institute for Computational Science, 7-1-26 minatojima-minamimachi, Chuo-ku, Kobe, Hyogo 650-0047, Japan

Supporting Information

ABSTRACT: Tom20 is located at the outer membrane of mitochondria and functions as a receptor for the N-terminal presequence of mitochondrial-precursor proteins. Recently, three atomic structures of the Tom20-presequence complex were determined using X-ray crystallography and classified into A-, M-, and Y-poses in terms of their presequence-binding modes. Combined with biochemical and NMR data, a dynamic-equilibrium model between the three poses has been proposed. To investigate this mechanism in further detail, we performed all-atom molecular dynamics (MD) simulations and replica-exchange MD (REMD) simulations of the Tom20-presequence complex in explicit water. In the REMD simulations, one major distribution and another minor one were observed in the converged free-energy landscape at 300 K. In the major distribution, structures similar to A- and M-poses exist, whereas those similar to Y-pose are located in the minor one, suggesting that A-pose in solution is more stable than Y-pose. A *k*-means clustering algorithm revealed a new pose not yet obtained by X-ray crystallography. This structure has double salt bridges between Arg14′ in the presequence and Glu78 or Glu79 in Tom20 and can explain the binding affinity of the complex in previous pull-down assay experiments. Structural clustering and analyses of contacts between Tom20 and the presequence suggest smooth conformational changes from Y- to A-poses through low activation barriers. M-pose lies between Y- and A-poses as a metastable state. The REMD simulations thus provide insights into the energetics of the multiple-binding forms and help to detail the progressive conformational states in the dynamic-equilibrium model based on the experimental data.



INTRODUCTION

Most mitochondrial proteins are synthesized by ribosomes in the cytosol and imported into mitochondria by the TOM (translocase of outer mitochondrial membrane) complex and the TIM (translocase of inner mitochondrial membrane) complex. The TOM complex is a large protein complex, which contains three subunits, named Tom20, Tom22, and Tom40, as main components. Tom40, a 40-kDa subunit of TOM, forms a protein-conducting channel embedded in the outer membrane.[1] Tom20 and Tom22 are 20 kDa and 22 kDa subunits, respectively, of the TOM complex, whose cytosolic domains recognize the N-terminal presequence of mitochondrial preproteins. The preproteins are then transferred to the Tom40 channel, and finally delivered to the mitochondrial inner membrane or matrix through the TIM complex.[2] The N-terminal presequence has a tendency to form an amphiphilic helical conformation with an abundance of positively charged residues on the one side and hydrophobic residues on the other.[3]

A solution NMR structure of the cytosolic domain (residues 51−145) of rat Tom20 in complex with a presequence peptide derived from rat mitochondrial aldehyde dehydrogenase (ALDH) was reported.[4] Due to the low affinity, simple mixing of Tom20 and presequences has not produced cocrystals. Crystal structures of the same Tom20-presequence complex were later determined by introducing an additional intermolecular disulfide bond between Tom20 and the presequence [A-pose (PDB entry: 2V1T) and Y-pose (PDB entry: 2V1S)],[5] or an intramolecular disulfide bond in the presequence [M-pose (PDB entry: 3AX3)].[6] In Figure 1, these three X-ray crystal structures are shown.

Received: January 4, 2013
Revised: February 19, 2013
Published: February 22, 2013

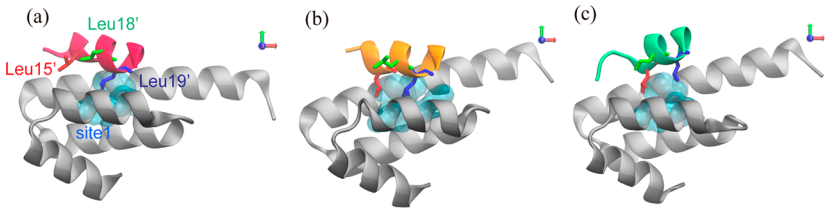


Figure 1. X-ray crystal structures of the Tom20-presequence complex. (a) A-pose, (b) M-pose, and (c) Y-pose. Surface representation (cyan) on Tom20 shows hydrophobic site1 (Ile74, Leu106, Leu110). Leu15′, Leu18′, and Leu19′ of the presequence are shown in red, green, and blue, respectively. Hereafter, residue numbers in the presequence are marked prime.

The structure analysis and pull-down assays on the Tom20-presequence complex showed that hydrophobic interactions were more important than ionic forces for the specific interaction.[7] NMR titration and peptide-library experiments revealed a five-residue sequence motif for the recognition of Tom20 in the N-terminal presequences. The consensus pattern is defined as "$\phi\chi\chi\phi\phi$", where ϕ is a hydrophobic amino acid and χ is any amino acid.[7,8] The NMR ^{15}N relaxation analysis showed the existence of residual motions on a submillisecond time scale at the molecular interface between Tom20 and presequence in the complex.[5] Based on these experimental results, there appears to be a dynamic equilibrium among multiple binding poses in the Tom20-presequence complex. Since this "dynamic-equilibrium model" is distinct from other protein−ligand binding mechanisms,[9,10] it may be meaningful to investigate the molecular mechanisms involving multiple binding poses. For this, detailed structural and energetic information on the different binding poses is fundamental.

We therefore performed all-atom molecular dynamics (MD) simulations and replica-exchange MD (REMD) simulations of the Tom20-presequence complex in explicit water. In REMD simulations,[11] a number of noninteracting copies of the original system (or replicas) at different temperatures are simulated independently and simultaneously by conventional MD simulations. Every few steps, pairs of replicas are exchanged with a transition probability according to the Metropolis criteria.[12] This exchange process generates random walks of replicas in temperature space, which, in turn, induce random walks in potential energy space. Because the random walk helps the simulation to escape from local energy minimum states, REMD enhances conformational sampling in the MD simulations to ensure statistically reliable free-energy landscapes at different temperatures by using single-histogram[13] or multiple histogram techniques.[14,15] We obtained a converged free-energy landscape starting from two different X-ray crystal structures and compared the stability of different binding poses in explicit water. Utilizing *k*-means clustering[16] and contact analysis, we investigated the transition pathway between different X-ray structures, which is useful for understanding the molecular mechanisms underlying the dynamics of binding in the Tom20-presequence complex in solution.

MATERIALS AND METHODS

Setting up Initial Structures for MD and REMD Simulations. To prepare starting structures for the current simulations, we used two crystal structures (PDB entries: 2V1T and 2V1S)[5] consisting of the cytosolic domain of Tom20 (residues 58−126) and a presequence (residues 12′−23′: GPRLSRLLSXAC). Hereafter, the residue number of the presequence is marked with a prime. The presequence peptide was tethered onto Tom20 using an intermolecular disulfide bond between Cys100 of Tom20 and Cys23′ of the presequence in order to facilitate the formation of the cocrystals. The amino acid sequence of the presequence in the two crystals differed at the position 21′: X = Ala in 2V1T and X = Tyr in 2V1S, which resulted in the different structures of the complex (A-pose for 2V1T and Y-pose for 2V1S). The crystals of 2V1T and 2V1S contain two and seven Tom20-presequence complexes in the asymmetric unit, respectively. The chain A (Tom20) and chain C (presequence) were selected from 2V1T, and the chain C (Tom20) and chain L (presequence) were from 2V1S. Note that intertwined dimers were formed in the 2V1S crystal, and thus the Tom20 and the presequence peptide were not directly connected via the intermolecular disulfide bond. The side chain of the Tyr21′ residue appeared to be responsible for the formation of the intertwined dimers. To prepare the starting structures, the Cys23′ residues were deleted, and the Tyr21′ in 2V1S was replaced by Ala to make the two starting structures have identical amino acid sequences. The N- and new C-termini of the presequence were blocked with an acetyl group and a methyl amide group, respectively. The cavities inside the two structures were filled with water molecules using *DOWSER* software.[17] Missing hydrogen atoms of the X-ray crystal structures were added using MMTSB Tool Set.[18] The two simulation systems (Systems A and Y in Table S1) were solvated with the same number of water molecules and 150 mM NaCl added after neutralization. To shorten computational time of the REMD simulations, we reduced the number of water molecules in Systems A and Y and made smaller systems (Systems AS and YS in Table S1). All the molecular graphics in this paper were drawn using *VMD 1.9.1*.[19]

MD Simulations and REMD Simulations of the Tom20-Presequence Complex. All-atom MD simulations were carried out using NAMD2.[20] CHARMM27 force field parameters with ϕ, ψ cross-term map correction (CMAP)[21,22] were used for proteins. TIP3P model[23] was used for water molecules. All bonds involving hydrogen atoms in proteins were constrained by SHAKE,[24] and water molecules were treated as rigid bodies by using SETTLE.[25] The equation of motion was integrated with a time step of 2 fs using velocity verlet. Electrostatic interactions were calculated with particle-mesh Ewald summation[26] and Lennard-Jones interactions were truncated at a cutoff distance of 12 Å (with a switching function being effective at 10 Å). Constant temperature (300 K) and constant pressure (1 atm) were controlled by using Langevin dynamics with a damping coefficient of 5 ps^{-1} and the Langevin piston Nosé-Hoover method,[27,28] respectively. Prior to productive simulations, the system was gradually equilibrated in several steps. Energy minimization was performed using the conjugate gradient method with an harmonic potential restraint of 10.0 kcal/molÅ^2 on the heavy atoms of proteins and 1.0

dx.doi.org/10.1021/jp400113e | J. Phys. Chem. B 2013, 117, 2864−2871

kcal/molÅ^2 on ions and water molecules inserted by *DOWSER*. Under the same restraint, the system was equilibrated in a NVT ensemble for 200 ps, and then in a NPT ensemble for 800 ps while gradually decreasing the restraint to zero. No restraint was imposed in subsequent productive runs. We performed two 500 ns MD simulations (MD_A and MD_Y) using Systems A and Y (see Table S1). The coordinates were saved every 10 000 steps in all MD simulations (Table 1).

Table 1. Summary of MD Simulations in This Study[a]

simulation	system (pose)	method	simulation length	temperature
MD_A	A (A-pose)	MD (NPT)	500 ns	300 K
MD_Y	Y (Y-pose)	MD (NPT)	500 ns	300 K
REMD_A	AS (A-pose)	REMD (NVT)	56 × 50 ns = 2.8 μs	300−400 K
REMD_Y	YS (Y-pose)	REMD (NVT)	56 × 50 ns = 2.8 μs	300−400 K

[a]The number of atoms in each system is listed in Table S1.

For REMD simulations, we used in-house software called REIN (Replica-Exchange Interface)[29] for exchanging replicas and NAMD2 for MD simulation of each replica. We simulated 56 replicas in a NVT ensemble with different temperatures between 300 and 400 K, which were generated by the REMD temperature simulator.[30] Replica exchanges were attempted every 1000 steps. For equilibration, each replica was simulated for 200 ps and then a 50 ns productive run performed. We performed two REMD simulations (REMD_A and REMD_Y) using Systems AS and YS. The total simulation time was 2.8 μs (= 50 ns × 56 replicas) for both systems. The coordinates were saved every 1000 steps for the REMD simulation and only the coordinates at 300 K were used for the free-energy and contact analyses described below. In Table 1, we summarize the results of the REMD simulations.

Structural Analysis of the Presequence in MD and REMD Trajectories. In this study, we focused on both the intra- and intermolecular motions of the presequence in the Tom20-presequence complex. Since the first two residues (residues 12′ and 13′) are structurally disordered, we did not analyze this region. To examine the intramolecular motions of the rest of the presequence, we superimposed the Cα atoms in residues 14′−22′ in the MD trajectories on those in the X-ray structures and computed the structural deviations. Tom20 has four α-helices: α1 (residues 62−83), α2 (residues 87−98), α3 (residues 105−112), and α4 (residues 117−123) helices in the cytosolic domain. We calculated the root-mean-square deviation (RMSD) of the presequence (residues 14′−22′) after superimposing the Cα atoms in the four helical regions of Tom20 in the MD trajectories on those in the X-ray structures. This value is hereafter referred to as "RMSD_pose" and is used to investigate the intermolecular motions of the presequence. To get detailed information of structural distributions, we performed the *k*-means clustering in MMTSB Tool Set.[18] In the clustering, we focused only on the heavy atoms of the consensus motif in the presequence (residues 15′−19′). The concept and detailed procedures of the analysis is described in Supporting Information.

RESULTS

Conformational Stability of the Tom20-Presequence Complex in MD Simulations. In Table 2, the RMSDs and RMSD_poses of the presequence in the X-ray crystal structure and in the MD simulations (MD_A and MD_Y) are shown. Although the RMSD between different binding poses is less than 1.0 Å, RMSD_pose is larger than 2.5 Å in the crystal structures. This means that only intermolecular motions are essential between different binding poses. In the MD simulations, the same trends were observed: the orientation relative to Tom20 was changed, while the presequence kept its α-helical conformation. In MD_A, the structure of the complex remained around the starting structure (A-pose) and was distinct from Y-pose. In contrast, we observed a conformational transition of the complex at 4.9 ns in MD_Y. We separately averaged out the RMSD values for the first 4.9 ns (0−4.9 ns) and for the rest of the simulation (4.9−500 ns) for the MD_Y trajectory. After the transition at 4.9 ns in MD_Y, the complex became closer to A-pose.

Changes in binding pose were investigated by examining the center of mass (COM) distance between Leu15′ or Leu19′ and the hydrophobic site 1 consisting of Ile74, Leu106, and Leu110 (see Figure 2).[4] We denote the distances as Dis(L15′, site1) and Dis(L19′, site1), respectively, which are used as reaction coordinates to draw the potential of mean force (PMF) (Figure 2a−c). The pair of the two distances, Dis(L15′, site1) and Dis(L19′, site1), are (9.65 Å, 3.11 Å), (5.84 Å, 2.14 Å), and (5.01 Å, 5.19 Å) for A-, M-, and Y-poses, respectively. In MD_A, we observed only one major distribution, which contains the A- and M-poses, in the PMF. A similar distribution in PMF was observed in MD_Y as a major stable distribution after the transition at 4.9 ns. We can see a minor distribution including Y-pose in the first 4.9 ns trajectory before the transition. No transition from A- to Y-pose was observed in MD_A and MD_Y. The free-energy landscapes obtained in the two MD simulations starting from different initial structures do not agree with each other. This result suggests that the length of the conventional MD simulations is not sufficient to examine the free-energy landscape of the Tom20-presequence complex.

REMD Simulations and Convergence of Free-Energy Landscapes. To explore all the binding poses in solution and

Table 2. RMSD and RMSD_pose of the Presequence in the X-Ray Crystal Structures and MD Simulations[a]

		X-ray crystal structures			MD simulations		
RMSD (fitted region)	reference (X-ray) structure	A-pose	Y-pose	M-pose	MD_A (500 ns)	MD_Y (0−4.9 ns)	MD_Y (4.9−500 ns)
RMSD (presequence)	A-pose	-	0.99	0.92	0.93 (0.89)	0.97 (0.98)	0.94 (1.10)
	Y-pose	0.99	-	0.31	0.39 (0.31)	0.42 (0.52)	0.41 (0.47)
	M-pose	0.92	0.31	-	0.45 (0.33)	0.47 (0.60)	0.48 (0.53)
RMSD_pose (Tom20)	A-pose	-	3.45	2.50	1.53 (1.65)	3.95 (2.75)	1.66 (1.29)
	Y-pose	3.45	-	2.95	3.32 (2.81)	2.47 (2.43)	3.37 (3.11)
	M-pose	2.50	2.95	-	2.32 (2.26)	3.90 (2.78)	2.30 (2.36)

[a]In the MD simulations, the averaged values as well as values for the final structure are listed in parentheses.

dx.doi.org/10.1021/jp400113e | J. Phys. Chem. B 2013, 117, 2864−2871

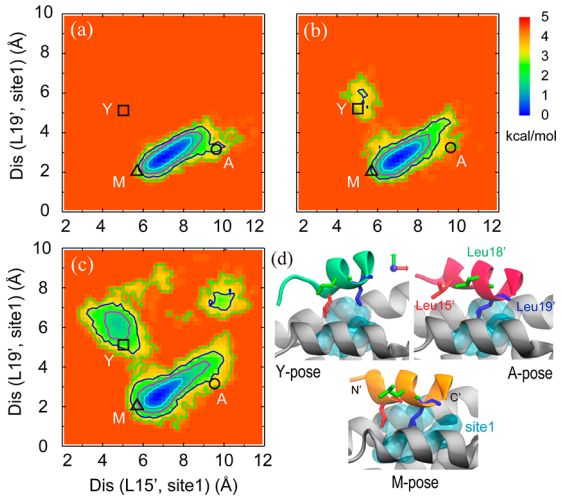


Figure 2. PMF along the distance between Tom20 and the presequence in (a) MD_A, (b) MD_Y, and (c) combined trajectories of REMD_A and REMD_Y. Circle, triangle, and square on the PMF are the position of A-pose, M-pose, and Y-pose, respectively. All the contour lines are separated by 1.0 kcal/mol. (d) Close-up view of the binding poses in X-ray crystal structures. Surface representation (cyan) on the Tom20 shows hydrophobic site1. Leu15′, Leu18′, and Leu19′ are shown in red, green, and blue, respectively.

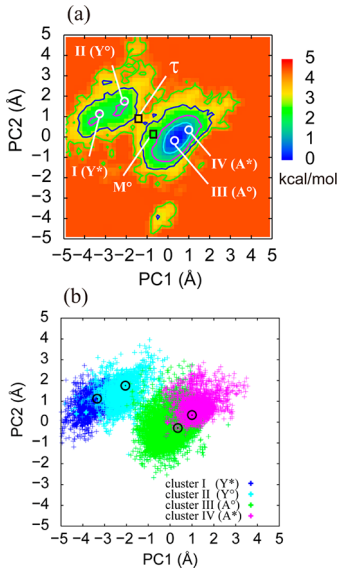


Figure 3. PMF of the first two principal components (PC1 and PC2) of the presequence (a) and mapping of the four representative structural clusters (b). The merged trajectories of REMD_A and REMD_Y at 300 K were used for the analysis. Four white circles indicate centroid (representative structure) of four clusters I, II, III, and IV. Black squares indicate transition state τ, and metastable state M° between representative structures II and III on the free energy landscape. Distribution of four clusters I (blue), II (cyan), III (green), and IV (magenta) obtained from the *k*-means cluster analysis. Ratio of each cluster I, II, III, and IV is 4.76%, 5.13%, 80.0%, and 4.23%, respectively. All the contour lines are separated by 1.0 kcal/mol.

to examine their conformational stability, we performed REMD simulations of the Tom20-presequence complex, starting from A-pose (REMD_A) and Y-pose (REMD_Y) in explicit solvent. Here, we discuss the numerical performance of the REMD simulations. If REMD algorithms work properly, random walks in potential energy space should be observed for all replicas. This, indeed, was the case. In Figure S2, random walks in temperature space and potential-energy space for replicas 1, 30, and 56 in REMD_A and REMD_Y at 300 K are shown. The average acceptance ratios were 50.7% and 50.3% in REMD_A and REMD_Y, respectively. Thus, the algorithm performed satisfactorily. We also checked the convergences of PMF along Dis(L15′, site1) and Dis(L19′, site1) in Figure S3, taking an average every 12.5 ns for each REMD simulation. There are similar trends after 25 ns. Since the PMF converged for both REMD simulations, we merged two trajectories and used all the snapshots obtained at 300 K for further analysis (Figure 2c).

Binding Motions of the Presequence Relative to Tom20. Here, we carried out principal component analysis[31,32] (PCA) on the atomic structures of the presequence viewed from Tom20. The variance-covariance matrix was constructed using the Cα coordinates of the presequence (residues 14′−22′) by superimposing the four helices of Tom20 (similar to how RMSD_pose was computed). In this way, we emphasize the binding motions of the presequence as relative to Tom20 in the bound states. The first and second principal components (PC1 and PC2) were used as reaction coordinates to draw a free-energy landscape at 300 K (Figure 3a). Contributions to the mean square fluctuation from the PC1 and PC2 modes were 60.7% and 18.2%, respectively, implying that these two coordinates are sufficient to describe the structural characteristics. We can see the two distributions on the PC plane, like the PMF along the distances between Leu15′ or Leu19′ and site 1 of Tom20 as shown in Figure 2c. The two distributions connect with each other at a point that appears to be a transition state between two different states (Figure 3a).

To get further information about the complex structures in solution, we carried out conformational cluster analysis using MMTSB Tool Set.[18] At first, 29 clusters were obtained by the toolset and we then merged some clusters that contained almost the same structures (Table S2). In addition, we discarded clusters that contained unfolded or dissociated presequence or Tom20 structures. The percentage of the discarded structures is 3.49%, thus having little effect on the final conclusion. Finally, four clusters, designated I, II, III, and IV, were obtained and mapped on the free-energy landscape along PC1 and PC2 (Figure 3b). The centroid structures are named structures I, II, III, and IV, respectively (Figure 4). In addition to these structures, we detected the transition state (structure τ) along the minimum energy path between structures II and III (Figure 3a). Along the path, the free energy differences between structures I and II and between structures III and VI were less than 0.5 kcal/mol, since these structures are located in the same basins of the free-energy landscape. The estimated free-energy changes between the two distributions and the activation free energy are estimated to be 1.8 and 1.6 kcal/mol, respectively (Figure 4).

In Figure 5, we show the number of contact atom pairs between hydrophobic residues of Tom20 and three hydrophobic residues (Leu15′, Leu18′, and Leu19′) of the presequence. In A-pose, Leu19′ has the largest number of contacts, whereas Leu15′ does not contact any hydrophobic residues in Tom20. In contrast, M- and Y-poses have a large number of contacts between Leu15′ and hydrophobic residues of Tom20. The number of contacts between Leu19′ and Tom20 is the smallest in Y-pose. In the simulations, the number of contacts between Leu15′ in the presequence and hydrophobic residues in Tom20 gradually decreases from

dx.doi.org/10.1021/jp400113e | J. Phys. Chem. B 2013, 117, 2864−2871

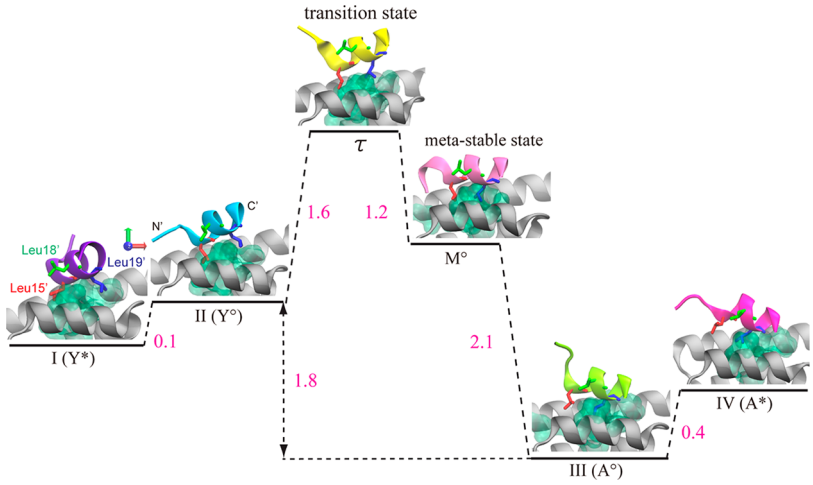


Figure 4. Energetics of the representative structures of the complex observed in REMD simulations. Free-energy differences obtained from the structural transition pathway are in magenta (in kcal/mol). Each color of the presequence of the four structures corresponds to the distribution in Figure 3b. Surface representation in green on the Tom20 shows hydrophobic residue contact with Leu15′, Leu18′, or Leu19′. Leu15′, Leu18′, and Leu19′ are red, green, and blue residue, respectively.

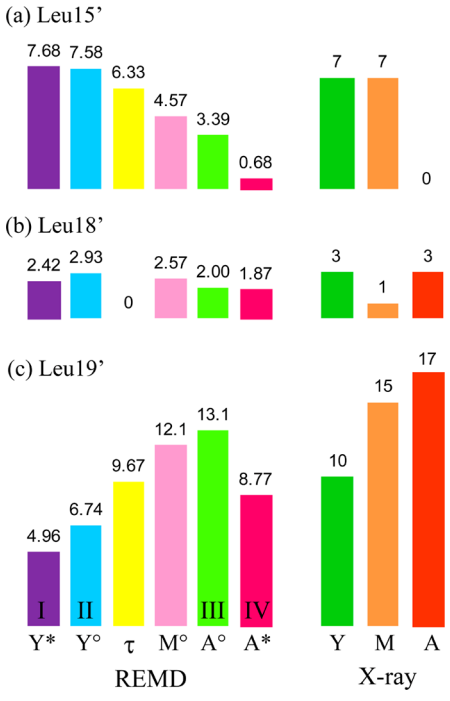


Figure 5. Number of contact atom pairs between hydrophobic residues of Tom20 and Leu15′, Leu18′, and Leu19′ of the presequence in REMD simulation (left) and X-ray crystal structures (right). Contact atom pairs were counted when the distance between hydrophobic residues of Tom20 and Leu15′, Leu18′, and Leu19′ of the presequence is less than 4.5 Å. The height of the bars and the number above the bar show the number of contact atom pairs.

structures I to IV, whereas the number with Leu19′ increases. Together with the information of the contacts and the RMSD_pose of the centroid structures I–IV from the X-ray structures (Table 3), we concluded that structure II corresponds to Y-pose and III represents A-pose. Although structure I has similar contact patterns to Y-pose, the structure itself deviates slightly from Y-pose. Hereafter, we refer to structures I, II, and III as Y*, Y°, and A°. Structure IV has a different contact pattern from the existing X-ray structures, implying a new complex structure. Structure τ shows the least RMSD_pose from Y-pose, but contact structures intermediate between Y° and A°. We also found a structure that is similar to M-pose between structures τ and III (referred to as M°).

DISCUSSION

In the free-energy landscape at 300 K, we found three major structural clusters that contain similar structures to the X-ray crystal structures and another minor cluster (cluster IV) whose structure is different from any X-ray structure. Careful examination of the structural characteristics of cluster IV revealed that most of the structures form double salt bridges between Arg14′ in the presequence and Glu78 and Glu79 in Tom20 (Figure 6). The probability of formation of the salt bridges at 300 K is shown in Table 4. We refer to this double salt bridged structure as A*. Interestingly, Glu78 and Glu79 of Tom20 are conserved in various organisms.[4] A pull-down assay was conducted to examine the effect of the substitution of Glu78 and/or Glu79 in Tom20 on the binding of the COX IV presequence (pCOX IV). The results showed that three mutants of Tom20, Glu78→Ala, Glu79→Ala, Glu78, 79→Ala

Table 3. RMSD_pose of the Representative Structures in REMD Simulations[a]

	RMSD_pose					
X-ray structure	I (Y*)	II (Y°)	τ	M°	III (A°)	IV (A*)
A-pose	4.63 (4.55)	3.66 (3.54)	3.02 (3.36)	2.01 (2.03)	1.78 (1.19)	2.04 (1.61)
M-pose	3.28 (3.36)	3.04 (3.15)	2.57 (3.19)	1.63 (1.57)	2.32 (1.89)	3.61 (3.05)
Y-pose	2.67 (2.52)	1.71 (1.40)	1.30 (1.16)	2.05 (2.48)	3.30 (3.33)	3.99 (3.78)

[a]The averaged values and the value for the centroid structure (in parentheses) in each structural cluster (I–IV) and the transition state (τ) and metastable state (M°) are shown.

dx.doi.org/10.1021/jp400113e | J. Phys. Chem. B 2013, 117, 2864–2871

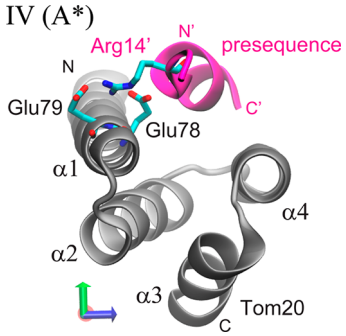


Figure 6. Double salt bridges between Tom20 and the presequence observed in cluster IV in REMD simulation. The amino acid residues, Glu78, Glu79, and Arg14′, that are involved in the two salt bridges between Tom20 and the presequence in cluster IV, are shown in the stick model. The structure is viewed from the N-terminus of the presequence (magenta). Probabilities of formation of salt bridges observed in clusters I–VI are summarized in Table 4.

Table 4. Probability of Formation of Salt Bridges between Glu78 and Arg14′, and between Glu79 and Arg14′, Observed at 300 K in the REMD Simulations[a]

	probability of formation of salt bridge (%)	
cluster	Tom20 (Glu78) - Presequence (Arg14′)	Tom20 (Glu79) - Presequence (Arg14′)
I (Y*)	90.5	0.08
II (Y°)	94.3	25.4
τ	100	33.3
M°	85.7	0
III (A°)	65.7	12.4
IV (A*)	99.1	95.3

[a]The same clustering is used as that in Table 3.

bound the pCOX IV with 80%, 60%, and 65% affinity of the wild-type, respectively.[4] This reduction in binding affinities suggests that formation of salt bridges between Arg14′ and Glu78 and/or Glu79 is important for the stabilization of the complex. We conclude that A* is an important binding pose of the Tom20-presequence complex in solution, although there is currently no supporting X-ray crystal structure.

Finally, we discuss the interrelationship of the five representative structures observed in REMD simulations at 300 K (A°, Y°, M°, A*, Y*). In the X-ray structures, the number of contact atom pairs between Tom20 hydrophobic residues and Leu15′ gradually decreases from Y- or M-pose to A-pose, whereas those between the same residues in Tom20 and Leu19′ increase from Y-pose to A-pose. In Figure 5, the same trends can be seen from structure I to IV, suggesting that the transition likely happens in this order and vice versa. We can draw a minimum energy pathway in the PMF along the first and second PCs (Figure 3). In the pathway, these structures align in the same order. Since A* can be found to be close to A° in the PMF, we propose that the conformational transitions between multiple binding poses of the Tom20-presequence occur in the following manner:

$$Y^* \Leftrightarrow Y^\circ \Leftrightarrow M^\circ \Leftrightarrow A^\circ \Leftrightarrow A^*$$

Except for A*, these binding poses are stabilized mainly via hydrophobic interactions, which likely avoid from water penetration at the interface between Tom20 and presequence. The activation free-energy barrier from Y° to A° is less than 1.6 kcal/mol, whereas the barrier in the reverse transition is a bit greater. However, taking M° as a metastable intermediate state, Y° and A° states are exchangeable in solution, supporting the dynamic equilibrium model proposed from the X-ray crystallography (Figure 4). Since the first two principal components can describe more than 80% of atomic motions of Tom20 and presequence in solution, we consider this analysis to be valid and it provides a reliable transition pathway between intermediate states. However, to get the free-energy barriers in a more quantitative manner, combinations of other conformational sampling techniques like the string method[33−37] or the Markov state sampling method[38−42] are very useful.

Each pose recognizes different combinations of the two hydrophobic side chains of the three hydrophobic residues in the Tom20 binding consensus, $\phi_1\chi\chi\phi_4\phi_5$: Y-pose (Y* and Y°), M-pose (M°), and A-pose (A° and A*) recognizes ϕ_1 and ϕ_4, ϕ_1 and ϕ_5, and ϕ_4 and ϕ_5, respectively. The rapid equilibrium between these poses enables Tom20 to recognize the three hydrophobic residues on ensemble average. This novel dynamic mechanism allows each pose to adapt to the size and chemical differences of two side chains individually. The adaptation is considered to be more straightforward compared to the static recognition, where a single binding state must cope with the differences of the three side chains simultaneously. Thus, we consider that the dynamic recognition mechanism is the molecular basis of the promiscuous binding of Tom20 toward diverse mitochondrial presequences with similar affinities, and facilitates the efficient translocation of a huge variety of preproteins across the mitochondrial membranes. Finally, we suggest that the dynamic recognition is a general mechanism for proteins with promiscuous binding specificity.

CONCLUSIONS

We investigated conformational dynamics of the Tom20-presequence complex in solution by performing atomistic MD and REMD simulations. In contrast to conventional MD simulations, the REMD simulations starting from two different initial structures provided converged PMF along the first two principal components. The initial structures (A-pose and Y-pose) were derived from two of the three crystal structures (in addition, M-pose) determined previously.[5,6] The PMF shows the major and minor structural clusters at 300 K (Figure 3). The major cluster contains A-pose-like structures (A° and A*), whereas the minor cluster contains Y-pose-like structures (Y* and Y°). An M-pose-like structure (M°) was also found as a metastable structure between the two clusters. Thus, in the free-energy landscape at 300 K, the three crystal poses are not the most stable structures in the MD simulations, but they are located close to the stable regions. The PMF reveals the binding energetics of Tom20-presequence interactions. A° is the most stable binding pose of the complex in solution, and Y° appears to be less stable than A°. Low activation barriers of 1−2 kcal/mol suggest smooth conformational changes between A- and Y-poses. Thus, our simulation analyses elucidate details of the exchanges of conformational states in the dynamic-equilibrium recognition based on the experimental data.

ASSOCIATED CONTENT

Supporting Information

Table of summary of simulation systems in this study, time series of RMSD and RMSD_pose in MD_A and MD_Y, numerical performance of the REMD simulations, convergence of PMF obtained with REMD_A and REMD_Y, description of

dx.doi.org/10.1021/jp400113e | J. Phys. Chem. B 2013, 117, 2864−2871

k-means clustering analysis, list of all clusters and their properties. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*Tel: +81-48-462-1407. Fax: +81-48-467-4532. E-mail: sugita@riken.jp.

Present Address

Takashi Saitoh, Center for Research and Education on Drug Discovery, Faculty of Pharmaceutical Sciences, Hokkaido University, Sapporo 060−0812, Japan.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This research was supported in part by a Grant for Scientific Research on a Priority Area 'Transient macromolecular complex' (to Y.S.), the Development and Use of the Next-Generation Supercomputer Project of the Ministry of Education, Culture, Sports, Science and Technology (MEXT) (to Y.S.), and the Fund from the High Performance Computing Infrastructure (HPCI) Strategic Program of MEXT (to Y.S.). We also thank the RIKEN Integrated Cluster of Clusters (RICC) for providing computational resources.

REFERENCES

(1) Hill, K.; Model, K.; Ryan, M. T.; Dietmeier, K.; Martin, F.; Wagner, R.; Pfanner, N. Tom40 forms the hydrophilic channel of the mitochondrial import pore for preproteins. *Nature* **1998**, *395*, 516−521.

(2) Chacinska, A.; Koehler, C. M.; Milenkovic, D.; Lithgow, T.; Pfanner, N. Importing mitochondrial proteins: machineries and mechanisms. *Cell* **2009**, *138*, 628−644.

(3) Heijne, G. The distribution of positively charged residues in bacterial inner membrane proteins correlates with the trans-membrane topology. *EMBO J.* **1986**, *5*, 3021−3027.

(4) Abe, Y.; Shodai, T.; Muto, T.; Mihara, K.; Torii, H.; Nishikawa, S.; Endo, T.; Kohda, D. Structural basis of presequence recognition by the mitochondrial protein import receptor Tom20. *Cell* **2000**, *100*, 551−560.

(5) Saitoh, T.; Igura, M.; Obita, T.; Ose, T.; Kojima, R.; Maenaka, K.; Endo, T.; Kohda, D. Tom20 recognizes mitochondrial presequences through dynamic equilibrium among multiple bound states. *EMBO J.* **2007**, *26*, 4777−4787.

(6) Saitoh, T.; Igura, M.; Miyazaki, Y.; Ose, T.; Maita, N.; Kohda, D. Crystallographic Snapshots of Tom20-Mitochondrial Presequence Interactions with Disulfide-Stabilized Peptides. *Biochemistry* **2011**, *50*, 5487−5496.

(7) Muto, T.; Obita, T.; Abe, Y.; Shodai, T.; Endo, T.; Kohda, D. NMR identification of the Tom20 binding segment in mitochondrial presequences. *J. Mol. Biol.* **2001**, *306*, 137−143.

(8) Obita, T.; Muto, T.; Endo, T.; Kohda, D. Peptide Library Approach with a Disulfide Tether to Refine the Tom20 Recognition Motif in Mitochondrial Presequences. *J. Mol. Biol.* **2003**, *328*, 495−504.

(9) Koshland, D. E.; Nemethy, G.; Filmer, D. Comparison of Experimental Binding Data and Theoretical Models in Proteins Containing Subunits. *Biochemistry* **1966**, *5*, 365−385.

(10) Monod, J.; Wyman, J.; Changeux, J.-P. On the nature of allosteric transitions: A plausible model. *J. Mol. Biol.* **1965**, *12*, 88−118.

(11) Sugita, Y.; Okamoto, Y. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* **1999**, *314*, 141−151.

(12) Hukushima, K.; Nemoto, K. Exchange Monte Carlo Method and Application to Spin Glass Simulations. *J. Phys. Soc. Jpn.* **1996**, *65*, 1604−1608.

(13) Ferrenberg, A.; Swendsen, R. Optimized Monte Carlo data analysis. *Phys. Rev. Lett.* **1989**, *63*, 1195−1198.

(14) Ferrenberg, A.; Swendsen, R. New Monte Carlo technique for studying phase transitions. *Phys. Rev. Lett.* **1988**, *61*, 2635−2638.

(15) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem.* **1992**, *13*, 1011−1021.

(16) MacQueen, J. B. Some Methods for classification and Analysis of Multivariate Observations; In *Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability*; University of California Press: Berkeley, 1967; Vol *1*, pp 281−297.

(17) Gumbart, J.; Trabuco, L. G.; Schreiner, E.; Villa, E.; Schulten, K. Regulation of the protein-conducting channel by a bound ribosome. *Structure* **2009**, *17*, 1453−1464.

(18) Feig, M.; Karanicolas, J.; Brooks, C. L., III MMTSB Tool Set: enhanced sampling and multiscale modeling methods for applications in structural biology. *J. Mol. Graph. Model.* **2004**, *22*, 377−395.

(19) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graph.* **1996**, *14*, 33−38.

(20) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **2005**, *26*, 1781−1802.

(21) MacKerell, A. D.; Bashford, D.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; et al. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **1998**, *102*, 3586−3616.

(22) Mackerell, A. D.; Feig, M.; Brooks, C. L., III Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.* **2004**, *25*, 1400−1415.

(23) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926−935.

(24) Ryckaert, J.; Ciccotti, G.; Berendsen, H. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **1977**, *23*, 327−341.

(25) Miyamoto, S.; Kollman, P. A. Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *J. Comput. Chem.* **1992**, *13*, 952−962.

(26) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995**, *103*, 8577−8593.

(27) Feller, S. E.; Zhang, Y.; Pastor, R. W.; Brooks, B. R. Constant pressure molecular dynamics simulation: The Langevin piston method. *J. Chem. Phys.* **1995**, *103*, 4613−4621.

(28) Hoover, W. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A* **1985**, *31*, 1695−1697.

(29) Miyashita, N.; Sugita, Y. ISLiM Molecular Scale code ID: M-2, Replica-exchange Molecular Dynamics program (REIN-K), http://www.islim.org/islim-dl_e.html.

(30) Patriksson, A.; van der Spoel, D. A temperature predictor for parallel tempering simulations. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2073−2077.

(31) Kitao, A.; Hirata, F.; Gō, N. The effects of solvent on the conformation and the collective motions of protein: Normal mode analysis and molecular dynamics simulations of melittin in water and in vacuum. *Chem. Phys.* **1991**, *158*, 447−472.

(32) Amadei, A.; Linssen, A. B. M.; Berendsen, H. J. C. Essential dynamics of proteins. *Proteins: Struct., Funct., Bioinformatics* **1993**, *17*, 412−425.

(33) E, W.; Ren, W.; Vanden-Eijnden, E. String method for the study of rare events. *Phys. Rev. B* **2002**, *66*.

(34) E, W.; Ren, W.; Vanden-Eijnden, E. Finite Temperature String Method for the Study of Rare Events. *J. Phys. Chem. B* **2005**, *109*, 6688−6693.

(35) Vanden-Eijnden, E.; Venturoli, M. Revisiting the finite temperature string method for the calculation of reaction tubes and free energies. *J. Chem. Phys.* **2009**, *130*, 194103.
(36) Maragliano, L.; Fischer, A.; Vanden-Eijnden, E.; Ciccotti, G. String method in collective variables: minimum free energy paths and isocommittor surfaces. *J. Chem. Phys.* **2006**, *125*, 24106.
(37) Maragliano, L.; Vanden-Eijnden, E. On-the-fly string method for minimum free energy paths calculation. *Chem. Phys. Lett.* **2007**, *446*, 182−190.
(38) Chodera, J.; Swope, W.; Pitera, J.; Dill, K. Long-Time Protein Folding Dynamics from Short-Time Molecular Dynamics Simulations. *Multiscale Model. Simul.* **2006**, *5*, 1214−1226.
(39) Chodera, J. D.; Singhal, N.; Pande, V. S.; Dill, K. A.; Swope, W. C. Automatic discovery of metastable states for the construction of Markov models of macromolecular conformational dynamics. *J. Chem. Phys.* **2007**, *126*, 155101.
(40) Voelz, V. A.; Bowman, G. R.; Beauchamp, K.; Pande, V. S. Molecular Simulation of ab Initio Protein Folding for a Millisecond Folder NTL9(1−39). *J. Am. Chem. Soc.* **2010**, *132*, 1526−152.
(41) Noé, F.; Fischer, S. Transition networks for modeling the kinetics of conformational change in macromolecules. *Curr. Opin. Struct. Biol.* **2008**, *18*, 154−162.
(42) Singhal, N.; Snow, C. D.; Pande, V. S. Using path sampling to build better Markovian state models: predicting the folding rate and mechanism of a tryptophan zipper beta hairpin. *J. Chem. Phys.* **2004**, *121*, 415−425.