The Molecular Recognition Mechanism for Superoxide Dismutase Presequence Binding to the Mitochondrial Protein Import Receptor Tom20 from Oryza sativa Involves an LRTLA Motif

ARTICLE *in* THE JOURNAL OF PHYSICAL CHEMISTRY B · OCTOBER 2010

Impact Factor: 3.3 · DOI: 10.1021/jp103547s · Source: PubMed

The Molecular Recognition Mechanism for Superoxide Dismutase Presequence Binding to the Mitochondrial Protein Import Receptor Tom20 from *Oryza sativa* Involves an LRTLA Motif

Yubo Zhang,[†] Marc Baaden,[‡] Junjie Yan,[†] Jinzhen Shao,[†] Su Qiu,[§] Yingliang Wu,[§] and Yi Ding*[,†]

Key Laboratory of Ministry of Education for Plant Developmental Biology, College of Life Sciences, Wuhan University, Wuhan 430072, P. R. China, Institut de Biologie Physico-Chimique, Laboratoire de Biochimie Théorique, CNRS UPR 9080, 13, rue Pierre et Marie Curie, F-75005 Paris, France, and State Key Laboratory of Virology, College of Life Sciences, Wuhan University, Wuhan 430072, P. R. China

Received: April 20, 2010; Revised Manuscript Received: August 15, 2010

Most mitochondrial proteins are synthesized in the cytosol as precursor and imported into the mitochondria by Tom complexes (translocase of outer membrane complexes). Knowledge of the binding mechanism between precursor and Tom20 in plants is very limited. Here, computational methods are employed to improve our understanding of the interactions between both molecules. To this end, we model mitochondrial superoxide dismutase precursor (pSOD) in complex with Tom20 in *Oryza sativa* (OsTom20). In a first stage, five main binding modes were generated using clustering analysis, energy minimization, and expert knowledge. In a second stage, the quality and validity of the resulting complexes is assessed by molecular dynamics (MD) simulations with a generalized Born solvation model. The change in binding free energies is estimated using a computational alanine scanning technique. We identified a particularly favorable complex between pSOD and OsTom20, exhibiting the lowest binding free energy among all candidates and correlating well with experimental data. Furthermore, three independent explicit solvent MD simulations of this structure, each of 100 ns duration, reveal that hydrophobic interactions occur between pSOD and OsTom20, in particular between L158 of pSOD and W81 of OsTom20, as evidenced by analysis of intermolecular distances and corresponding relative free energy landscapes. L158 is part of an interacting LRTLA motif. These results provide new insight into the structural basis and dynamics of precursor recognition by Tom20 in plant, and their generality is supported by sequence alignments with seven other plants.

Introduction

Mitochondrial proteins are vital for maintaining mitochondrial biogenesis, which cannot take place in mitochondria themselves.[1] Most mitochondrial proteins are synthesized in the cytosol as precursor proteins and are subsequently imported into the mitochondria through translocases of outer mitochondrial membranes (Tom) such as Tom7, Tom20, Tom22, and Tom40.[2,3] Tom20 is the master receptor subunit of the Tom complex. It consists of an N-terminal membrane-anchor segment, a tetratricopeptide repeat (TPR) motif, and a C-terminal domain exposed to the cytosol.[1,4] Two novel isoforms of Tom20 were found in animals, and the type I form is similar to the fungal form of Tom20.[5] Tom20 directly recognizes the targeting sequence on mitochondrial precursor proteins and passes them on to the Tom40 channel.[6] Thus, obtaining reliable structures of Tom20 in complex with precursor proteins is essential in understanding the structure−function relationship of these assemblies. Some reports indicate that ionic interactions might mediate binding of positive charges on the presequence to acidic residues of Tom20.[7,8] However, the NMR structure of a general Tom20 import receptor from rat, in complex with a presequence peptide, revealed that hydrophobic rather than ionic interactions mediate precursor binding. This was confirmed by site-direct mutagenesis studies.[2]

The Tom20 receptor from plants was first identified in mitochondria from potato (*Solanum tuberosum*)[9,10] and was copurified on blue native/SDS-PAGE (sodium dodecyl sulfate polacrylamide gel electrophoresis) with the components of the plant Tom complex.[11] The three isoforms of plant Tom20 (*At*Tom20-2, *At*Tom20-3, and *At*Tom20-4) were identified with two-dimensional isoelectric focusing (IEF)/SDS-PAGE in *Arabidopsis thaliana*.[11] Recently, an NMR structure revealed that *At*Tom20 consists of seven helices in the plant mitochondria.[12] The mitochondrial precursor proteins may bind plant Tom20 through a groove formed by two TPR motifs in helices α2 and α4. This structural feature is also present in the structure of Tom20 in animals and fungi.[2,13,14] However, Tom20 in plants lacks a hydrophobic surface patch domain which functions as a binding site in the orthologous rat protein.[12]

To date, experimental attempts to determine the complex between plant Tom20 and mitochondrial precursor proteins have failed. Computational approaches can alternatively be used for predicting the interaction mechanisms between ligand and receptor.[15−17] Homology modeling combined with molecular dynamics (MD) simulations is a useful approach to design structural models based on available experimental data.[18] Mitochondrial superoxide dismutase forms such a precursor

* To whom correspondence should be addressed. Mailing address: Department of Genetics, College of Life Sciences, Wuhan University, Ba-Yi Street, Luojia Hill, Wuchang, Hubei Wuhan, 430072, P. R. China. E-mail: yiding@whu.edu.cn. Telephone: +86-27-68754319. Fax: +86-27-68754319.

† Key Laboratory of Ministry of Education for Plant Developmental Biology.

‡ Institut de Biologie Physico-Chimique, CNRS.

§ State Key Laboratory of Virology.

10.1021/jp103547s
Published on Web 10/11/2010

protein: pSOD. In this study, we use ZDOCK[19] to determine the spatial structures of pSOD in complex with OsTom20 from *Oryza sativa*. After clustering the candidate stuctures, five binding modes were selected. MD simulations and computational alanine-scanning based on the molecular mechanics generalized Born surface area (MM-GBSA) approach were used to assess the quality of these binding modes. Then, the most important residues involved in the formation of the pSOD-OsTom20 complex were characterized from the final equilibrated structure, evidencing a central role of an LRTLA sequence motif for binding.

Methods

Generation of Atomic Models of pSOD and OsTom20. In the first homology modeling step, template structures related to the OsTom20 and pSOD proteins were searched against the whole Protein Data Bank[20] (PDB) using the Blast algorithm. The amino acid sequence of the Tom20 protein from *O. sativa* was obtained from Swiss-Prot (ID: Q5JJI4). The atomic coordinates of Tom20 (PDB code: 1ZU2) in *A. thaliana* were downloaded from the PDB. *At*Tom20 consists of seven α helices, and the TPR in helices α2 and α4 is known to be important for recognition of the presequence binding site of mitochondrial proteins. We modeled Tom20 from *O. sativa* (OsTom20) by using the NMR structure of *At*Tom20 (PDB id: 1ZU2, sequence identity 43.67%) as a template through the SWISS-MODEL server.[21] This model was further refined and validated by subsequent MD simulations, as we have previously shown that such short MD simulations are useful to assess the quality of structural models.[18]

Previous reports indicated that presequences may form positively charged amphipathic α-helices, important for targeting the precursor protein to mitochondria.[22,23] Recently, circular dichroism (CD) spectroscopy of the F1b presequence from *Nicotiana plumbaginifolia* confirmed that this particular presequence could form an α-helical structure both in membrane mimetic environments (∼50% α-helix) and in acidic phospholipids (∼60% α-helix).[24] This N-terminal amphipathic α-helix could recognize the Tom20 receptor. Until now, structural data for the presequences from plant species were very limited, and we thus modeled the presequence of SOD (pSOD) from *O. sativa* by a similar motif (32.26% similarity) taken from the Y34F mutant of human mitochondrial manganese superoxide dismutase. This fragment is not in itself a presequence, yet the resulting model is structurally similar to the F1b presequence with both terminal regions being helical and connected by a short loop. The alignments between target and template protein structures are given in the Supporting Information.

In this work, residue numbering in the OsTom20-pSOD complex is consecutive, meaning that the first residue of the presequence in the complex is actually referred to as M156.

ZDOCK and Energy Minimization. The ZDOCK program, a fast Fourier transform (FFT)-based, initial-stage rigid-body molecular docking algorithm,[19] was employed for predicting pSOD-OsTom20 complexes. The top 1000 predicted ligand orientations from ZDOCK were classified into different groups by discriminating different recognition modes. Possible hits were obtained by clustering analysis within each group. To identify appropriate candidate complexes for further MD simulations, 3000 steps of energy minimization using the SANDER module in the AMBER 8 suite of programs[25] were employed for each selected candidate. The van der Waals (vdW) interaction energy between OsTom20 and pSOD was used to screen the candidates. The ligand−receptor vdW energy was calculated with the ANAL program of AMBER 8[25] for further assessment of the three-dimensional structure of the complex.

Calculation of Binding Free Energies by MM-GBSA. To further evaluate the quality of our models, we carried out 1000 ps of unrestrained MD using the Amber 8 program for each selected candidate. ZDOCK combined with GB MD simulations has been widely used to determine the structure of biological complexes.[16,26−28] This method has the advantage to estimate the binding free energy of the complex by the MM-GBSA approach.[29,30]

In this study, the implicit generalized Born solvation model was used (igb = 2). The temperature was set to 300 K. Nonbonded interactions were cut off at a distance of 12 Å. The ff99 force field (Parm99)[31] was applied throughout the energy minimization and MD simulations.

In the MM-GBSA implementation of Amber 8.0, the binding free energy of A + B → AB is calculated using the following thermodynamic cycle:

$$\begin{array}{ccccc} A_{aqu} & + & B_{aqu} & \xrightarrow{\Delta G_{binding}} & AB_{aqu} \\ \downarrow -\Delta G^{A}_{solv} & & \downarrow -\Delta G^{B}_{solv} & & \downarrow -\Delta G^{AB}_{solv} \\ A_{gas} & + & B_{gas} & \xrightarrow{\Delta G_{gas}} & AB_{gas} \end{array}$$

$$\begin{aligned} \Delta G_{binding} &= \Delta G_{gas} - \Delta G^{A}_{solv} - \Delta G^{B}_{solv} + \Delta G^{AB}_{solv} \\ &= \Delta H_{gas} - T\Delta S - \Delta G^{A}_{GBSA} - \Delta G^{B}_{GBSA} + \Delta G^{AB}_{GBSA} \\ &= \Delta H_{gas} - T\Delta S + \Delta\Delta G_{GB} + \Delta\Delta G_{SA} \end{aligned}$$

$$\Delta H_{gas} \approx \Delta E_{gas} = \Delta E_{intra} + \Delta E_{elec} + \Delta E_{vdw}$$

$$\Delta\Delta G_{GB} = \Delta G^{AB}_{GB} - \left(\Delta G^{A}_{GB} + \Delta G^{B}_{GB}\right)$$

$$\Delta\Delta G_{SA} = \Delta G^{AB}_{SA} - \left(\Delta G^{A}_{SA} + \Delta G^{B}_{SA}\right)$$

$$\begin{aligned} \Delta G_{\mathrm{binding}} &= \Delta G_{\mathrm{gas}} - \Delta G^{\mathrm{A}}_{\mathrm{solv}} - \Delta G^{\mathrm{B}}_{\mathrm{solv}} + \Delta G^{\mathrm{AB}}_{\mathrm{solv}} \\ &= \Delta H_{\mathrm{gas}} - T\Delta S - \Delta G^{\mathrm{A}}_{\mathrm{GBSA}} - \Delta G^{\mathrm{B}}_{\mathrm{GBSA}} + \Delta G^{\mathrm{AB}}_{\mathrm{GBSA}} \\ &= \Delta H_{\mathrm{gas}} - T\Delta S + \Delta\Delta G_{\mathrm{GB}} + \Delta\Delta G_{\mathrm{SA}} \end{aligned}$$

$$\Delta H_{\mathrm{gas}} \approx \Delta E_{\mathrm{gas}} = \Delta E_{\mathrm{intra}} + \Delta E_{\mathrm{elec}} + \Delta E_{\mathrm{vdw}}$$

$$\Delta\Delta G_{\mathrm{GB}} = \Delta G^{\mathrm{AB}}_{\mathrm{GB}} - (\Delta G^{\mathrm{A}}_{\mathrm{GB}} + \Delta G^{\mathrm{B}}_{\mathrm{GB}})$$

$$\Delta\Delta G_{\mathrm{SA}} = \Delta G^{\mathrm{AB}}_{\mathrm{SA}} - (\Delta G^{\mathrm{A}}_{\mathrm{SA}} + \Delta G^{\mathrm{B}}_{\mathrm{SA}})$$

where T is the temperature, S is the solute entropy, ΔG_{gas} is the interaction energy between A and B in the gas phase, and $\Delta G^{\mathrm{A}}_{\mathrm{solv}}$, $\Delta G^{\mathrm{B}}_{\mathrm{solv}}$, and $\Delta G^{\mathrm{AB}}_{\mathrm{solv}}$ are the solvation free energies of A, B, and AB, which are estimated using a GB surface area (GBSA) method.[29,30] That is, $\Delta G^{\mathrm{AB}}_{\mathrm{solv}} = \Delta G^{\mathrm{AB}}_{\mathrm{GBSA}} + G^{\mathrm{AB}}_{\mathrm{GB}} + \Delta G^{\mathrm{AB}}_{\mathrm{SA}}$, and so forth. ΔG_{GB} and ΔG_{SA} are the electrostatic and nonpolar terms, respectively. The bond, angle, and torsion energies constitute the intramolecular energy $\Delta E_{\mathrm{intra}}$ of the complex, while ΔE_{elec} and ΔE_{vdw} represent the receptor−ligand electrostatic and vdW interactions, respectively. Because of the constant contribution of $-T\Delta S$ for each docked complex, we refer to $\Delta G^{*}_{\mathrm{binding}}$ for $\Delta G_{\mathrm{binding}} + T\Delta S$ in the discussion. To verify the quality and validity of the resulting pSOD-OsTom20 complexes, the relative binding free energy $\Delta G^{*}_{\mathrm{binding}}$ was calculated by using MM-GBSA for post processing snapshots from the MD trajectories. The computational alanine scanning method in MM-GBSA was used to discriminate the candidates. The key residues were mutated to alanine, and subsequently the difference in the

binding free energies between mutated and wild-type complexes was calculated based on the MM-GBSA approach. The calculated results were compared with the experimental data.

MD Simulations with Explicit Solvent. The final selected complex was refined during three extended unrestrained simulation runs of 100 ns duration each, in explicit solvent, thereby introducing solvation effects that may affect the interaction between presequence and receptor. These MD simulations were performed with the GROMACS 3.3.0 software package[32] using the optimized potentials for liquid simulation (OPLS) force field[33] and the SPC216 water model. The protonation state of ionizable groups was chosen to correspond to pH 7.0. Counterions were added to compensate the net charge of the system. The initial structure of the complex was immersed in a periodic water box. The electrostatic interactions were calculated by using the particle-mesh Ewald (PME) algorithm,[34] and the vdW forces were treated with a cutoff distance of 10 Å. After 3000 steps of energy minimization using a steepest descent method, the system was subject to 100 ps of equilibration at 300 K and normal pressure, using harmonic position restraints with a force constant of 1000 kJ mol^{-1} nm^{-2}. The system was coupled to an external bath by the Berendsen pressure and temperature coupling method.[35] The three independent 100 ns production runs (Q1, Q2, Q3) were performed under the same conditions except that all position restraints were removed. The results were analyzed using the standard software tools provided by the GROMACS package.[32] Visualization and manipulation of the conformations was performed using the programs PyMOL 0.99 (http://www.pymol.org/) and Swiss-PdbViewer 3.7.[21] The sequences were aligned with ClustalW.[36] The distance analysis of the nonpolar residue pairs was performed using the *R* statistical software package.[37]

Results and Discussion

Screening of pSOD-OsTom20 Complexes by ZDOCK and Energy Minimization. During the ZDOCK calculation, the OsTom20 and pSOD binding partners were treated as rigid bodies, and 1000 conformations were generated by restricting the search to the recognition region. The α-helix of pSOD containing the key hydrophobic residues was required to associate with the TPR domain of Tom20, which is in agreement with experimental results. In this first step, 876 hits were abandoned because they did not satisfy this condition. After carefully screening the remaining candidates, five binding modes representing different orientations of the pSOD α-helix with respect to Tom20 were selected. As shown in Figure 1, the slope of the α-helix of pSOD presents left or right orientations around the central axis of the TPR groove of the OsTom20 partner in binding modes I (40 candidates) and II (43 candidates). Binding modes III (21 candidates) and IV (11 candidates) are different from the above modes, and the slope of the precursor α-helix is along the viewing direction, either coming out of or pointing inside the paper surface. In binding mode V (20 candidates), pSOD's α-helix is aligned with the OsTom20 helices and fits into the groove of the TPR region.

For the purpose of identifying near-native docking decoys, structural clustering and energy-based analysis was used to screen for reasonable tertiary protein structures. First, we used an all-atom root-mean-square deviation (rmsd) to quantify the pairwise similarity between any two ligand orientations of the candidates in each mode. The structures for which rmsd's from the central structure were smaller than a cutoff distance were selected in each group, leading to 5, 9, 9, 6, and 10 candidates for modes 1 to 5, respectively. The cutoff parameter was set to

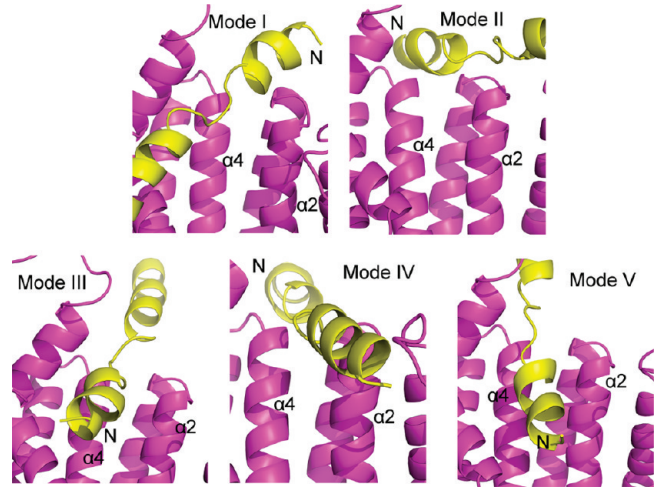

Figure 1. Representation of the five binding modes that were identified. These modes correspond to different spatial orientations of pSOD (yellow) in complex with OsTom20 (magenta). The N-terminus of pSOD is labeled, as well as helices 2 and 4 of OsTom20.

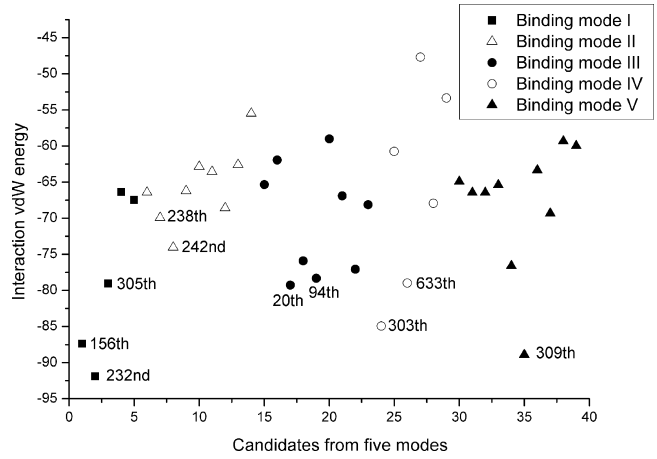

Figure 2. van der Waals interaction energy between pSOD and OsTom20 calculated for 39 candidate complex structures representing all five binding modes.

1.75 Å, 3 Å, 4 Å, or 5 Å, depending on the distribution of the cluster sizes.

Second, the vdW interaction energies were used to discriminate plausible pSOD-OsTom20 complexes since previous reports pointed out that hydrophobic interactions are essential for the recognition of mitochondrial proteins by Tom20. After using energy minimization to remove any steric clashes in the complexes, the conformations were ranked by vdW interaction energies ranging from −91.9 to −47.7 kcal mol^{-1} (Figure 2). Ten candidates with strong hydrophobic interactions from binding mode I (156th, 232nd, 305th), II (238th, 242nd), III (20th, 94th), IV (303rd, 633rd), and V (309th) were selected for further study. The selection was done in two steps: first, the top 6% of all candidates were selected (interaction energies stronger than −77 kcal mol^{-1}) since the vdW interaction energies between receptor and ligand in these ones were strongest. Then, as none of the candidates from mode II satisfied this condition, we selected the two lowest interaction energy structures from this mode to include them in the comparison.

Discrimination between Plausible Binding Modes by MD Simulations and Computational Alanine Scanning. To narrow the selection toward the best candidate complexes, 1000 ps of MD simulation of each complex were performed with the GB solvation model. This introduces flexibility and enables the

TABLE 1: Energy Components Contributing to Binding Free Energies[s]

binding mode	structure	ΔE_{elec}	ΔE_{vdw}	ΔE_{inter}	$\Delta\Delta G_{GB}$	$\Delta\Delta G_{SA}$	$\Delta G^{*}_{binding}$
I	232	−575.51	−76.52	−652.04	632.07	−12.52	−32.48
	156	−537.17	−75.11	−612.27	585.35	−13.28	−40.20
	305	−897.33	−59.26	−956.59	937.64	−12.68	−31.63
II	242	−544.28	−68.05	−612.33	592.97	−12.07	−31.43
	238	−633.83	−59.78	−693.62	677.00	−10.55	−27.16
III	20	−846.45	−67.58	−914.03	879.06	−12.30	−47.28
	94	−763.25	−61.74	−824.99	807.69	−11.92	−29.22
IV	303	−474.86	−66.74	−541.60	526.90	−12.33	−27.03
	633	−505.96	−70.01	−575.97	549.39	−11.90	−38.48
V	309	−735.76	−73.57	−809.33	783.70	−13.18	−38.80

[s] All energies are in kcal mol^{-1}.

residue side chains to adjust. Binding free energy calculations based on the MM-GBSA method were used to screen the resulting complexes. As shown in Table 1, vdW and electrostatic energies are major favorable contributions to binding, whereas nonpolar solvation energies contribute only slightly. The calculated binding free energy represents the binding affinity of the complex and successfully ranks the 20th conformation (mode III, −47.28 kcal mol^{-1}), 156th conformation (mode I, −40.2 kcal mol^{-1}), 309th conformation (mode V, −38.8 kcal mol^{-1}), and 633th conformation (mode IV, −38.48 kcal mol^{-1}) as most favorable structures among the 10 candidates. To further discriminate among these modes, we use the computational alanine scanning technique to calculate binding free energies $\Delta\Delta G^{*}_{binding}$ between mutated and wild-type complexes.

The computational data of the 309th structure (Figure 3A) were in good agreement with experimental data extracted from NMR chemical-shift experiments.[12] The helices 2 and 4 of Tom20 were located in the TPR-like motif[12] and the residues with the most significant chemical-shift perturbations located on the helices 2 and 4. This demonstrated that these residues were important for the precursor protein recognition. The residues from helix α5 were not essential for the recognition since none of them showed perturbation. Thus, we selected 10 residues from helices 2 and 4 for the computational alanine scanning, nine of which showed important perturbations in the experiment. We also selected three residues from helix α5 for the alanine mutation scanning.

In our study, $\Delta\Delta G^{*}_{binding}$ calculated for the 309th structure is affected when important residues in these regions are mutated to alanine. The residues L44, E45, and W48 of helix α2 show significant chemical-shift perturbations in the experiment,[12] which correspond to calculated values of 0.86, 2.35, and 2.07 kcal mol^{-1} in the $\Delta\Delta G^{*}_{binding}$ scan. The W81, S89, F92, and F93 residues show perturbations in the NMR chemical-shift measurements, which agree well with the large values of 2.8, 8.47, 2.75, and 7.18 kcal mol^{-1} in the $\Delta\Delta G^{*}_{binding}$ calculation.

A significantly lower effect on $\Delta\Delta G^{*}_{binding}$ values is observed for mutants of L43 in helix α2 and three hydrophilic residues K99, N101, and E102 in helix α5. The results on L43, K99, N101, and E102 are in line with experimental data as the mutation of these residues does not induce significant chemical-shift perturbations. Unexpectedly, the calculated data on W39 and T88 show a change of only 0.07 and −0.07 kcal mol^{-1} in $\Delta\Delta G^{*}_{binding}$, whereas experimental data indicate that these residues contribute to the binding affinity.[12]

In contrast to binding mode V (complex 309th), the binding affinities of modes I (complex 156th), III (complex 20th), and IV (complex 633th) do not correlate with the experimental data (Figure 3B). The low $\Delta\Delta G^{*}_{binding}$ values of four mutations (L44A, E45A, W81A, T88A) in the 20th, 156th, and 633th structures

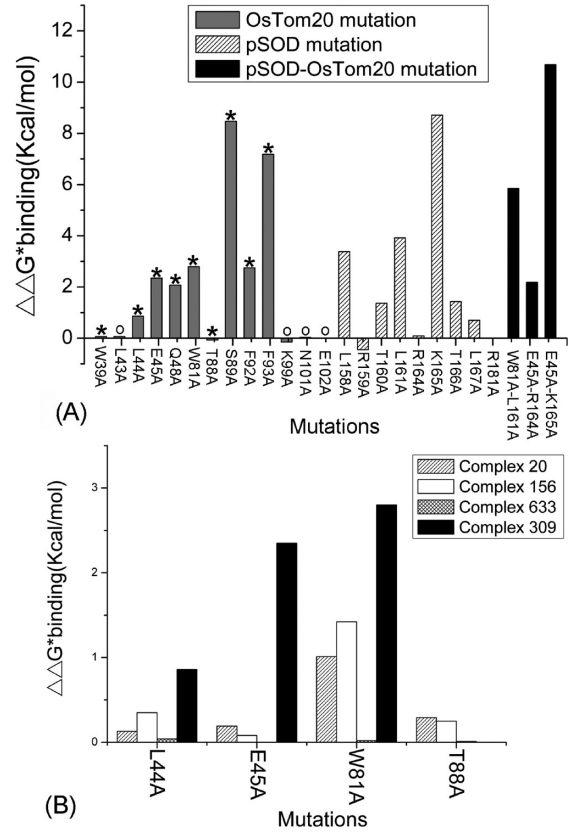

Figure 3. Computational alanine scanning mutagenesis results for pSOD-OsTom20 mutations on the binding affinity. ($\Delta\Delta G^{*}_{binding} = \Delta G^{*}_{binding(mutant)} - \Delta G^{*}_{binding(wild\text{-}type)}$). (A) $\Delta\Delta G^{*}_{binding}$ result for 22 single mutations for either OsTom20 or pSOD and three double mutations for pSOD and OsTom20 in the 309th structure. (B) Comparison of the $\Delta\Delta G^{*}_{binding}$ results for four mutations among modes I (156th), III (20th), IV (633th), and V (309th). The star (★) and the circle (○) indicate the residues that were shown experimentally to be important or to show no effect, respectively. These data were derived from either $^{13}C/^{1}H$ or $^{15}N/^{1}H$ chemical-shift perturbation.[12]

are in contradiction with a strong implication of these residues suggested by the experiment. Overall, the comparison of computational alanine scanning profiles with experimental data strongly suggests that binding mode V yields the most promising pSOD-OsTom20 complex model.

Stability and Interaction Energy Analysis of pSOD-OsTom20 during Explicit Solvent MD Simulations. The pSOD-OsTom20 complex from the 309th structure with binding mode V was further refined in three independent 100 ns explicit solvent MD simulations. To assess the stability of the complex, we determined its structural drift with respect to a reference obtained from the equilibration step just prior to the production

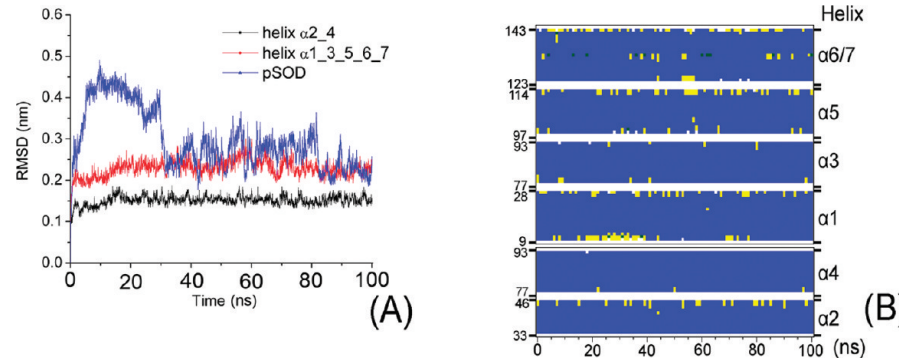

Figure 4. Characterization of the pSOD/OsTom20 complex in explicit solvent MD. The rmsd (A) and the secondary structure (B) (α-helix, blue; turn, yellow) of the OsTom20 helices in the pSOD-OsTom20 complex as a function of time for the 100 ns MD production run Q1.

MD run. The rmsd's for α-helices 2 and 4 and for the remaining ones of OsTom20, as well as for the recognition domain (from residue 158 to residue 169) of pSOD with respect to this reference structure, were calculated along the Q1 trajectory and are shown in Figure 4A. The rmsd analysis of the other two simulations Q2 and Q3 are given in the Supporting Information, Figure 2. Here, we mostly refer to Q1 as it remains closest to the original 309th starting structure. The rmsd for OsTom20 α-helices is stable during the 100 ns MD simulations. The α2 and α4 helices are essential for precursor protein recognition, and their rmsd leveled off at about 1.5 Å. A slightly higher rmsd value of about 2.2 Å was observed for the other OsTom20 helices. The recognition domain of pSOD initially showed a higher rmsd value, temporarily reaching 4.5 Å at 10 ns. Its rmsd then lowered to a plateau at about 2.5 Å after 30 ns of simulation time.

The secondary structure analysis confirms the very high structural integrity of the α-helices of the OsTom20 complex (Figure 4B). All helices maintain their α-helicity throughout the explicit solvent simulations.

The total solvent accessible surface area (SAS) can be used both as a direct measure of receptor−substrate interaction and to gauge convergence, which is achieved in the first quarter of the simulation (Supporting Information, Figure 3).

Using the final structures of the 100 ns explicit-solvent MD simulations, we re-evaluated the binding free energy of the pSOD-OsTom20 complex using a computational alanine scanning technique. The equilibrated complex was compared with the one obtained immediately after the ZDOCK step (Figure 5). The three sets of computational alanine scanning mutagenesis results obtained for each individual 100 ns MD simulation were averaged. We found a good agreement between this computational data of the complex after long equilibration and the experimental data.[12] It is interesting to note that the calculated effect on T88 showed a change of 0.72 kcal mol^{-1} in $\Delta\Delta G^{*}_{\text{binding}}$, whereas the previously calculated result of the complex after docking with ZDOCK only showed a change of −0.07 kcal mol^{-1} in $\Delta\Delta G^{*}_{\text{binding}}$. The computational data from the 100 ns MD equilibration confirms that residue T88 contributes to the binding affinity[12] and represents an improved model of the complex.

The high $\Delta\Delta G^{*}_{\text{binding}}$ values of 2.17, 4.04, and 4.18 kcal mol^{-1} for the hydrophobic residues W81, F92 and F93 suggest that these residues may contribute to the hydrophobic interaction during the recognition process. We furthermore observe a significantly higher effect for the polar residues E45, R164, and K165. The main reason is that E45 forms salt bridges and hydrogen bonds with the R164 and K165 residues during the MD simulation. In addition, the Q48 residue forms a stable

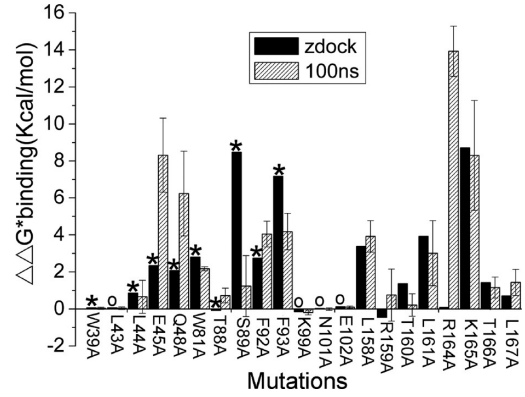

Figure 5. Computational alanine scanning result of the pSOD-OsTom20 complex after the ZDOCK stage and after 100 ns of explicit solvent MD simulation, averaged over simulations Q1, Q2, and Q3 with the standard deviation shown as error bars. As in Figure 3, the star (★) and the circle (○) symbols indicate residues that were shown experimentally to be important or to show no effect, respectively.

hydrogen bond with K165 during most of the MD simulation (See Supporting Information, Figure 4).

Hydrophobic Interactions Involved in the Recognition Mechanism of pSOD and OsTom20. In our final pSOD-OsTom20 structure (Figure 6A), OsTom20 recognizes pSOD mainly through the key residues E45 (α2), W81 (α4), S89 (α4), and F92 (α4). We aligned the α2, α3, and α4 helices from OsTom20 with seven other plants: *Sorghum vulgare* (C5XGB6), *Zea mays* (B6SZD1), *Vitis vinifera* (D1HJL4), *Physcomitrella patens* (A9TNW3), *S. tuberosum* (P92792), *Ricinus communis* (B9RBI0), and *Elaeis guineensis* (B3TM32). The sequence alignment (Figure 6B) indicates that the residues in helices α2 and α4 are more conserved than the ones in helix α3. The E45, W81, S89, and F92 residues are strictly conserved amino acids. It is plausible to assume that the plant Tom20 uses a similar recognition mechanism for all mitochondrial presequences.

We observe an LRTLA motif for pSOD, comprising three hydrophobic residues L158, L161, and A162 at its two ends. Furthermore, two polar residues, R159 and T160, are found in the middle of this motif, and the R159 residue has a long aliphatic side-chain. These characteristics match well with the general pattern of a previously postulated Tom20-recognition motif $\phi\chi\chi\phi\phi$.[38] This $\phi\chi\chi\phi\phi$ binding segment is very common in the mitochondrial presequences (Figure 6C), where ϕ is a hydrophobic residue and χ is any amino acid residue.[1,38] Our alanine scanning data indicates that the nonpolar residues L158 and L161 contribute 3.92 and 3.01 kcal mol^{-1} to the $\Delta\Delta G^{*}_{\text{binding}}$ affinity, respectively. These LRTLA motif residues thus form strong hydrophobic interactions which could be important for

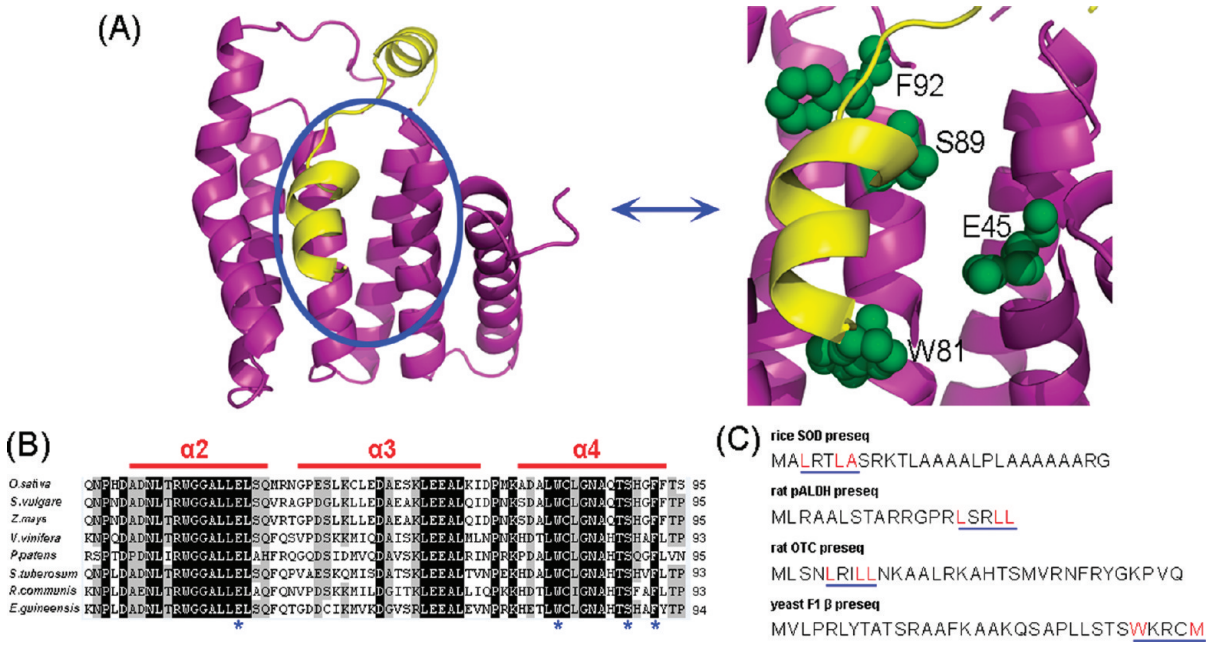

Figure 6. (A) A typical structure of the equilibrated pSOD-OsTom20 complex, left panel. The critical residues in the recognition mechanism are labeled on the right panel. Both molecules are represented as ribbon structures. (B) Multiple sequence alignment of OsTom20 with seven Tom20 proteins from other plants. The black boxes indicate strictly conserved amino acids, and gray boxes indicate closely related amino acids. The critical residues for recognition are annotated with an asterisk. (C) The $\phi\chi\chi\phi\phi$ motif of the rice SOD, rat ALDH, rat OTC, and yeast F1 β presequences is underlined. The hydrophobic residues ϕ from the $\phi\chi\chi\phi\phi$ motif are colored in red.

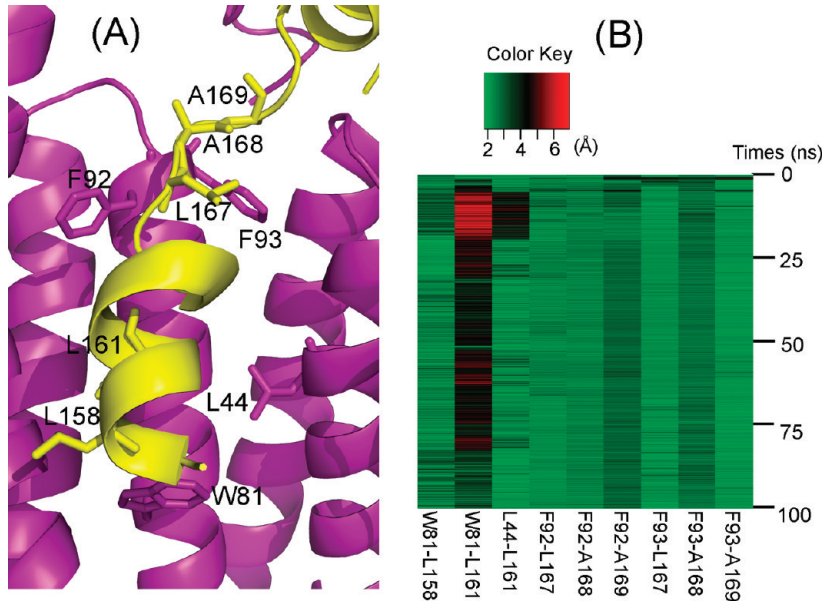

Figure 7. Interaction details of the hydrophobic residues in pSOD-OsTom20. (A) Hydrophobic residues of OsTom20 in contact with residues of pSOD. Panel B shows the distance between the hydrophobic residues as a function of time during the Q1MD simulation.

pSOD recognition by OsTom20. The central polar residues R159 and T160 only contribute 0.75 and 0.21 kcal mol^{-1} to the binding affinity compared to their Ala mutants. This may indicate that the χ residues of the $\phi\chi\chi\phi\phi$ motif do not directly influence pSOD-OsTom20 binding; they may however be essential to keep the motif in register.

In previous reports,[2,12] hydrophobic interactions were described as essential for the recognition mechanism of Tom20. We therefore analyzed the importance of the LRTLA motif for ligand−receptor interaction. The nonpolar W81 residue was suggested as a key residue for precursor binding, which is supported by significant NMR chemical-shift changes in a perturbation experiment.[12] This residue is located in helix α4 of OsTom20 (Figure 6A), and its importance for ligand−receptor interaction is confirmed by the corresponding 2.17 kcal mol^{-1} $\Delta\Delta G^{*}_{\text{binding}}$ value when mutated to alanine. We analyzed the closest interatomic distances between key nonpolar residues identified from the computational alanine scanning results.

Figure 7A and B show the distances between W81-L158 and W81-L161 in simulation Q1 as a function of time. Both residues come closer and were found within a contact distance of 4.0 Å at the end of the simulation. This may indicate that hydrophobic interactions may be important for W81 recognition within the LRTLA motif of pSOD. The large values of 4.04 and 4.18 kcal mol^{-1} for the nonpolar residues F92 and F93 in the $\Delta\Delta G^{*}_{\text{binding}}$ calculation suggest both of them contribute to the hydrophobic ligand−receptor interaction. As shown in Figure 7A, the nonpolar residues L167, A168, and A169 are located in the loop

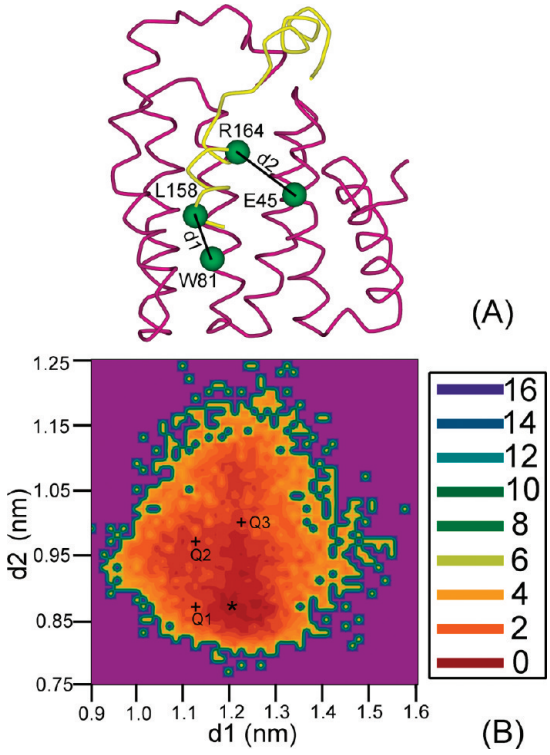

Figure 8. (A) The Cα atoms of the residues E45, W81, L158, and R164 used to define the d1 and d2 distances of panel B. (B) The two-dimensional free energy landscape as a function of the distances d1 and d2 (defined in the text and shown in panel A) for all three simulations. The final structures of simulations Q1, Q2, and Q3 are indicated as crosses. The star (★) indicates the global energy minimum. The color map indicates the energy in kT units, where k is the Boltzmann constant and T is the temperature, relative to the global minimum representing a reference energy of zero.

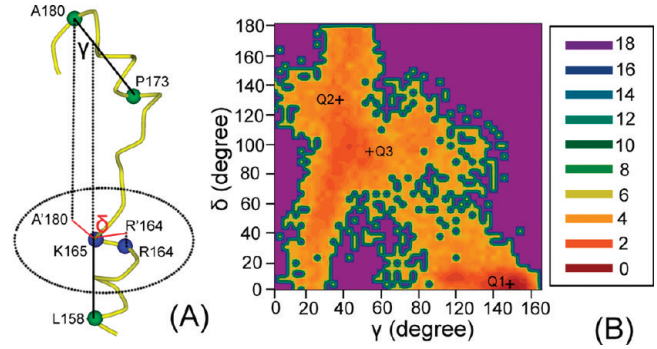

Figure 9. (A) The Cα atoms of the residues used for defining the angles γ and δ (described in the text). (B) The two-dimensional free energy landscape as a function of the kink angle γ and the swivel angle δ (defined in the text and shown in panel A) for all three simulations. The final structures of simulations Q1, Q2, and Q3 are indicated as crosses.

domain of the pSOD ligand. These three residues are in the vicinity of the F92 and F93 residues. We then analyzed the residue pairs F92-L167, F92-A168, F92-A169, F93-L167, F93-A168, and F93-A169. Figure 7B shows their distances to be stable around 3 Å during the whole Q1MD simulation. The Supporting Information Figure 5 provides the evolution of the interatomic distances during the Q2 and Q3MD simulations. The Q2MD simulation shows distance maps very similar to Q1, suggesting the repeatability of the experiment. In the Q3MD simulation, the F92 and F93 residues moved beyond a contact distance of 5 Å with respect to A168 and A169 of pSOD in the last half of the MD simulation, yet they still maintain contact with L167. We then superimposed the three independent 100 ns runs for a visual comparison provided as Supplementary Movie 1. The movie shows that the pSOD LRTLA motif is firmly anchored to the Tom20 receptor. However, the loop connecting the two α-helices of pSOD is very mobile, and some rearrangement of the peptide chain termini is observed during the MD simulation. These may contribute to the distance deviation of the F92 and F93 residues in the Q3MD simulation in particular.

Distance and Angular Analysis Using Free Energy Landscapes. To better understand the conformational dynamics that governs pSOD recognition, we analyzed the sampling of the LRTLA motif positioning close to its binding site as a function of two distances between key interacting residues (Figure 8A). The resulting potential of mean force is shown in Figure 8B. The two dimensions of this relative free energy landscape are the distance d1 between the Cα atoms of residues W81 and L158, both contributing significantly to the stabilizing interactions in the complex, and the distance d2 between the Cα atoms of residues E45 and R164. The latter residues show the highest contributions to binding affinity (Figure 5). The E45 residue actually forms hydrogen bonds with R164 in most of the MD simulations as indicated by previous analysis (Supporting Information Figure 4). The lowest free energy in the landscape (Figure 8B) centers around (1.20 nm, 0.87 nm), and the final structures from the Q1, Q2, and Q3 simulations are centered at (1.13 nm, 0.87 nm), (1.13 nm, 0.97 nm), and (1.23 nm, 1.00 nm), respectively. The distinct minimum centered around the highly populated 1.2 nm region for the d1 distance visible on the free energy contour map suggests a significant interaction between the involved residues W81 and L158. This observation is consistent with the suggestion (Figure 7B) that this hydrophobic interaction may be one of the driving forces for the recognition between pSOD and OsTom20. The shorter d2 minimum energy distance centered at about ~0.87 nm may reflect the contribution of the previously mentioned hydrogen bond to the interaction of pSOD-OsTom20.

Our MD simulation results further suggest a possible bending motion in the center of the pSOD peptide. To quantify this motion, we define two vectors α and β approximating the two helical axes of the pSOD peptide (Figure 9A). The first helix vector is taken between the Cα atoms of residues M158 and K165 and the second between the Cα atoms of residues P173 and A180. These two vectors form a kink angle γ as depicted in Figure 9A. We further define a plane perpendicular to the first helix and project the Cα atoms of A180 and R164 into this plane, as A′180 and R′164, respectively. The second swivel angle δ is measured with respect to the reference points A′180, K165, and R′164. The change of the δ angle represents the orientation of the projection of the second helix in the plane, relative to the K165−R′164 direction (Figure 9A). We can now describe the distortion of pSOD in terms of the rotation of its helical segments relative to one another expressed by the kink angle γ and the swivel angle δ. The corresponding free-energy landscape (Figure 9B) can be divided into two main regions: the top left region ($0° < \gamma < 50°$ and $40° < \delta < 140°$) with its minimum at (45°, 98°) and the bottom right region ($120° < \gamma < 160°$ and $0° < \delta < 15°$) centered around the global minimum at (150°, 9°). The graph illustrates the anisotropic distortion of both helices, echoing a previous more general study on helical kink and swivel distributions.[39] Figure 9B shows that the final structures of the Q1, Q2, and Q3 simulations center near the respective lowest energy regions. The significant difference of Q2 and Q3 is their wider angle distribution, with a less marked

minimum energy well. The visual analysis shown in Supplementary Movie 1 illustrates that the first pSOD helix remains parallel to the axis formed by α-helices 2 and 4 of OsTom20, firmly anchored to its recognition domain. The main reason for the observed angle distribution is thus the significant rearrangement of the second helix with respect to the complex.

Conclusion

In this work, a plausible structure of pSOD in complex with OsTom20 from *O. sativa* was obtained by combination of computational methods including ZDOCK, MD simulations, and MM-GBSA analysis. Five main binding modes were investigated, with different locations of the pSOD α-helix with respect to OsTom20. Combined with MD simulations and computational alanine scanning analysis, the 309th structure of mode V was selected as the best complex model.

This structural model was equilibrated by 100 ns of unrestrained MD simulation in explicit solvent. In three independent MD simulations, hydrophobic interactions exist between pSOD and OsTom20, which agrees with previous reports,[12] in particular providing evidence for a general $\phi\chi\chi\phi\phi$ precursor motif postulated earlier. We show that such an interaction motif exists for the pSOD:OsTom20 complex and involves L158, L161, and A162 of pSOD.

Additionally, important features of the pSOD:OsTom20 recognition mechanism were summarized, and their generality was assessed by a sequence alignment with Tom20 from seven other plants. The nonpolar residues W81, F92, and F93 of OsTom20 could form hydrophobic interactions with the nonpolar residues L161, A167, A168, and A169 of pSOD. The latter residues are located on an α helix on pSOD which locks into a stable position firmly held by the receptor.

To our knowledge, this is the first report relating structural modeling of mitochondrial precursor proteins recognizing the Tom20 complex in plant. Our study and the resulting structural model provide new information on the plant Tom20 recognition mechanism with the potential to help accelerate the research of structure−function relationships in Tom20. More generally, such structural models may present a great opportunity in biologically engineering more adaptable and cheaper rice grains.

Acknowledgment. This work was supported by 973 Program of China [2007CB108705] and by National Nature Science Foundation of China [30971740, 30821064].

Supporting Information Available: Supplementary Figures 1−5 with the sequence alignment of OsTom20 and the 1ZU2 template and pSOD and its template from PDB, the pSOD-OsTom20 rmsd, SAS of the pSOD-OsTom20 complex, E45 and Q48 residues, and the distance between the hydrophobic residues as a function of time, respectively. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Saitoh, T.; Igura, M.; Obita, T.; Ose, T.; Kojima, R.; Maenaka, K.; Endo, T.; Kohda, D. *EMBO J.* **2007**, *26*, 4777.
(2) Abe, Y.; Shodai, T.; Muto, T.; Mihara, K.; Torii, H.; Nishikawa, S.; Endo, T.; Kohda, D. *Cell (Cambridge, MA, U.S.)* **2000**, *100*, 551.
(3) Rapaport, D. *J. Cell Biol.* **2005**, *171*, 419.
(4) Iwahashi, J.; Yamazaki, S.; Komiya, T.; Nomura, N.; Nishikawa, S.; Endo, T.; Mihara, K. *J. Biol. Chem.* **1997**, *272*, 18467.
(5) Likic, V. A.; Perry, A.; Hulett, J.; Derby, M.; Traven, A.; Waller, R. F.; Keeling, P. J.; Koehler, C. M.; Curran, S. P.; Gooley, P. R.; Lithgow, T. *J. Mol. Biol.* **2005**, *347*, 81.
(6) Bohnert, M.; Pfanner, N.; van der Laan, M. *FEBS Lett.* **2007**, *581*, 2802.
(7) Komiya, T.; Rospert, S.; Koehler, C.; Looser, R.; Schatz, G.; Mihara, K. *EMBO J.* **1998**, *17*, 3886.
(8) Schleiff, E.; Shore, G. C.; Goping, I. S. *J. Biol. Chem.* **1997**, *272*, 17784.
(9) Heins, L.; Schmitz, U. K. *Plant J.* **1996**, *9*, 829.
(10) Perry, A. J.; Rimmer, K. A.; Mertens, H. D.; Waller, R. F.; Mulhern, T. D.; Lithgow, T.; Gooley, P. R. *Plant Physiol. Biochem.* **2008**, *46*, 265.
(11) Werhahn, W.; Niemeyer, A.; Jansch, L.; Kruft, V.; Schmitz, U. K.; Braun, H. *Plant Physiol.* **2001**, *125*, 943.
(12) Perry, A. J.; Hulett, J. M.; Likic, V. A.; Lithgow, T.; Gooley, P. R. *Curr. Biol.* **2006**, *16*, 221.
(13) Ramage, L.; Junne, T.; Hahne, K.; Lithgow, T.; Schatz, G. *EMBO J.* **1993**, *12*, 4115.
(14) Endo, T.; Kohda, D. *Biochim. Biophys. Acta* **2002**, *1592*, 3.
(15) Durrieu, M. P.; Lavery, R.; Baaden, M. *Biophys. J.* **2008**, *94*, 3436.
(16) Yi, H.; Qiu, S.; Cao, Z.; Wu, Y.; Li, W. *Proteins* **2008**, *70*, 844.
(17) Qiu, S.; Yi, H.; Liu, H.; Cao, Z.; Wu, Y.; Li, W. *J. Chem. Inf. Model.* **2009**, *49*, 1831.
(18) Law, R. J.; Capener, C.; Baaden, M.; Bond, P. J.; Campbell, J.; Patargias, G.; Arinaminpathy, Y.; Sansom, M. S. *J. Mol. Graphics Modell.* **2005**, *24*, 157.
(19) Chen, R.; Li, L.; Weng, Z. *Proteins* **2003**, *52*, 80.
(20) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *Nucleic Acids Res.* **2000**, *28*, 235.
(21) Guex, N.; Peitsch, M. C. *Electrophoresis* **1997**, *18*, 2714.
(22) von Heijne, G. *EMBO J.* **1986**, *5*, 1335.
(23) Roise, D.; Theiler, F.; Horvath, S. J.; Tomich, J. M.; Richards, J. H.; Allison, D. S.; Schatz, G. *EMBO J.* **1988**, *7*, 649.
(24) Moberg, P.; Nilsson, S.; Stahl, A.; Eriksson, A. C.; Glaser, E.; Maler, L. *J. Mol. Biol.* **2004**, *336*, 1129.
(25) 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.
(26) Han, S.; Yi, H.; Yin, S. J.; Chen, Z. Y.; Liu, H.; Cao, Z. J.; Wu, Y. L.; Li, W. X. *J. Biol. Chem.* **2008**, *283*, 19058.
(27) Wu, Y.; Cao, Z.; Yi, H.; Jiang, D.; Mao, X.; Liu, H.; Li, W. *Biophys. J.* **2004**, *87*, 105.
(28) Yi, H.; Cao, Z.; Yin, S.; Dai, C.; Wu, Y.; Li, W. *J. Proteome Res.* **2007**, *6*, 611.
(29) Qiu, D.; Shenkin, P.; Hollinger, F.; Still, C. *J. Phys. Chem. A* **1997**, *101*, 3005.
(30) Tsui, V.; Case, D. A. *J. Am. Chem. Soc.* **2000**, *122*, 2489.
(31) Wang, J.; Cieplak, P.; Kollman, P. *J. Comput. Chem.* **2000**, *21*, 1049.
(32) Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. *J. Comput. Chem.* **2005**, *26*, 1701.
(33) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225.
(34) Essmann, U.; Perera, L.; Berkowitz, M.; Darden, T.; Lee, H.; Pedersen, L. *J. Chem. Phys.* **1995**, *103*, 8577.
(35) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684.
(36) Thompson, J. D.; Higgins, D. G.; Gibson, T. J. *Nucleic Acids Res.* **1994**, *22*, 4673.
(37) R Development Core Team. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing: Vienna, Austria, 2009.
(38) Muto, T.; Obita, T.; Abe, Y.; Shodai, T.; Endo, T.; Kohda, D. *J. Mol. Biol.* **2001**, *306*, 137.
(39) Cordes, F. S.; Bright, J. N.; Sansom, M. S. *J. Mol. Biol.* **2002**, *323*, 951.

JP103547S