Cryo-electron microscopy-based Integrative atomic-resolution modeling of the TOM GIP complex

A.V. DUDKO

Belarusian State University, Minsk, Belarus

e-mail: hannah.dud94@gmail.com

V.G. VERESOV

Institute of Biophysics and Cell Engineering, Minsk, Belarus.

e-mail: veresov@ibp.org.by

More than 98% of mitochondrial proteins are nuclear encoded, synthesized on cytosolic ribosomes, and imported into the organelle posttranslationally. The translocase of the outer mitochondrial membrane (TOM) complex is the main entry gate for proteins imported into mitochondria [1]. The TOM complex is an assembly of seven subunits acting together at different stages of protein import: the surface receptors TOM20 and TOM70, the general import pore (GIP) consisting of TOM40, TOM22, and the small TOM proteins TOM5, TOM6, and TOM7 [1]. The preproteins are initially recognized by the receptor TOM proteins, TOM20 or TOM70, and are then transferred to the GIP complex. Information on the three-dimensional (3D) structure of the TOM complex is essential to understand the molecular mechanisms underlying its multiple functions. Whereas the structures of the cytosolic domains of rat Tom20 in complex with a presequence peptide [2, 3], Arabidopsis TOM20 without ligand [4], S. cerevisiae TOM70 [5], together with the yeast cryo-electron microscopy structure [6] and yeast cross-linking data [7] have been reported, but, so far, a high-resolution structure of the entire assembly is absent. While conventional high-resolution techniques in structural biology are challenged by the size and flexibility of many biological assemblies, the combination of low-resolution techniques, such as cryo-electron microscopy, SAXS or cross-linking protein interaction analysis with computational structure-based modeling have opened up new avenues to define the structures of such assemblies at atomic resolution. In this study, we used computational structural biology tools and the most recent structural information that was obtained through electron-microscopy and cross-linking experiments to obtain the atomic-resolution structural model of the yeast GIP, a core part of the TOM-complex. The 3D-structures of yeast TOM40, TOM22, TOM5, TOM6 were obtained using the combination of the I-TASSER protocol [8], ROSETTA 3.6 Loop Closure approach [9] and the atomic-resolution refinement by ModRefiner algorithm [10]. The prediction of the 3D-structures of protein-protein complexes was performed in a stepwise fashion with an initial rigid-body global search using the program Piper [11] and subsequent steps using the FlexPepDock [12], HADDOCK [13], and ROSETTADOCK [14] programs to refine the rigid-body stage predictions. At first, the 3D-structures of the TOM40 and

TOM22 transmembrane and cytosolic portions were individually modeled using the I-TASSER protocol. Next the full-length structure of TOM22 was generated, by integrating the transmembrane and cytosolic portions into a single structure with the use of the ROSETTA 3.6 Loop Closure protocol, followed by the refinement with the ModRefiner algorithm. To obtain the 3D-structure of the TOM40/OM22 complex, only the TOM22 transmembrane domain (TOM22TM) was used at first, because of availability of cross-linking data only for the TOM40/TOM22(TM) complex. The modeling of this complex was performed using the program PIPER with the enhanced electrostatic coefficient option to avoid energetically unfavorable exposure of TOM22(TM) hydrophilic surfaces to a hydrophobic membrane environment. The TOM40/TOM22(TM) complex with the best fit to cross-linking data was selected and next subjected to a refinement procedure with the use of the FlexPepDock program. Next, three copies of the top-ranked FlexPepDock TOM40/TOM22(TM) dimer were subjected to PIPER symmetric docking with rotational symmetry of order 3, in accordance with electron-microscopy data [6] showing a three-fold rotational symmetry of the TOM complex. The structure with the best fit to cross-linking data among the set of the symmetric protein complexes provided by PIPER was selected and subjected to a FlexPepDock refinement protocol using TOM22TM as the peptide-ligand and the two adjacent TOM40 monomers as the receptor. The best-scored TOM40/TOM40/TOM22TM structure was integrated into the $(TOM40/TOM22TM)_3$ symmetric hexameric complex, resulting in the basic $(TOM40/TOM22TM)_3$ structure, which was used in further modeling. Thereafter, the entire structure of TOM22 that was obtained at the previous stage was superimposed onto the TOM22TM structures of the $(TOM40/TOM22TM)_3$ hexamer to model $(TOM40/TOM22)_3$ structure. Thereafter the structure that was obtained was subjected to a refinement by ROSETTADOCK using the TOM40 dimer and TOM22 as the receptor and ligand, respectively. Finally, the stepwise docking of small components TOM5, TOM6 to the (TOM40/TOM22)3 was carried out. The method of individual modeling of the cytosolic amd transmembrane portions with I-TASSER protocol, followed by their integration into single structure by ROSETTA-Loop-Closure approach and a refinement with ModRefiner was applied to modeling the 3D-structures of these small TOM components, too. These subunits were next docked to the $(TOM40/TOM22)_3$ hexamer using PIPER – HADDOCK – ROSETTADOCK docking protocol.

This strategy has resulted in the structure of the yeast GIP complex with high shape- and electrostatic complementarities between the components of the complex. 11 polar contacts took place between TOM40 and TOM22 monomers in the final model. These were: His346TOM40A - Tyr114TOM40C, Asp345TOM40A – Trp100TOM22B, Tyr309TOM40A – Leu115TOM22B, Asn266TOM40A –Ala119TOM22B, Arg310TOM40C –Leu118TOM22A, Arg310TOM40C – Ala119TOM22A, Tyr309TOM40C – Ser114TOM22C, Asn266TOM40B – Ala119TOM22C,

Tyr309TOM40B – Leu118TOM22C, His346TOM40B– Tyr114TOM40C, Asp345TOM40B – Trp100TOM22C. As expected, the predicted structure is consistent with recent cryo-electron microscopy data [6] and cross-linking data [7]. The most important results of docking for this complex are summarized in Table 1. As to the small TOM components, such as TOM5 and TOM6, one salt bridge (Arg261TOM40 - Asp55TOM6) and five hydrogen bonds (Thr220TOM40 – Leu25TOM5, Thr220TOM40– Gln27TOM5, Gln294TOM40 – Asn38TOM6, Met279TOM40 – Leu32TOM6, Thr273TOM40 – Gln50TOM6) were observed between each TOM40 monomer and small TOM components, TOM5 and TOM6. The results that were obtained for interactions of TOM5 and TOM6 are also in good agreement with crosslinking data for these components [7]. The structure of the full GIP complex is shown in Fig.1.

Table 1. The ROSETTADOCK interface energy scores (I_sc), Buried Surface Area (BSA), ROSETTADOCK van der Waals interaction energy score ($\Delta VDWS_{RD}$), ROSETTADOCK desolvation energy score ($\Delta solv_{RD}$), number of intermolecular salt bridges (N_{sb}) and intermolecular hydrogen bonds (N_{HB}) for the highest- rank complexes between different TOM components

Protein pairs	Isc	BSA Å^2	Evdw kcal/ mol	Edesolv kcal/ mol	N(sb)	N(hb)
TOM40-TOM22	-8.0	2031.0	-7.3	2.1	0	5
$(TOM40)_2$-TOM22	-8,3	1529.4	-8.4	1.8	0	11
TOM40-TOM5	-7.1	1539.1	-8.7	1.4	0	2
TOM40-TOM6	-7.2	1779.5	-9.4	3.4	1	3

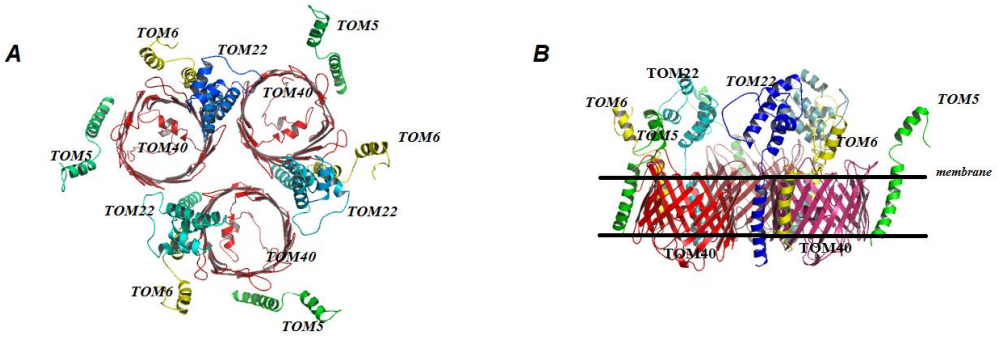

Fig.1. Structural model of the GIP-complex. A: a top view of the GIP complex; B: a side view of the GIP complex

1. M.Bohnert, N.Pfanner, M. van der Laan (2015) Mitochondrial machineries for insertion of membrane, *Curr Opin Struct Biol*., **33:**92-102.
2. T.Saitoh et al. (2007) Tom20 recognizes mitochon-drial presequences through dynamic equilibrium among multiple bound states, *EMBO J*., **26:**4777–4787.
3. Y.Abe et al. (2000) Structural basis of presequence recognition by the mitochondrial protein import receptor Tom20, *Cell,* **100:**551–560.
4. A.J.Perry et al. (2006) Convergent evolution of receptors for protein import into mitochondria, *Curr. Biol.,* **16:**221–229.
5. Y.Wu, B. Sha (2006) Crystal structure of yeast mitochondrial outer membrane translocon member Tom70p, *Nat. Struct. Mol. Biol*., **13:**589-593.
6. K.Model, C.Meisinger, W.Kühlbrandt (2008) Cryo-electron microscopy structure of a yeast mitochondrial preprotein translocase, *J. Mol. Biol.,* **383:**1049–1057.
7. T.Shiota et al. (2015) Molecular architecture of the active mitochondrial protein gate, *Science,* **349:**1544-1548.
8. A.Roy, A.Kucukural, Y.Zhang (2010) I-TASSER: a unified platform for automated protein structure and function prediction, *Nat. Protocols,* **5:**725-738.
9. D.J.Mandell, E.A.Coutsias, T.Kortemme (2009) Sub-angstrom accuracy in protein loop reconstruction by robotics-inspired conformational sampling, *Nat Methods,* **6:**551-552.
10. D.Xu, Y.Zhang (2011) Improving the Physical Realism and Structural Accuracy of Protein Models by a Two-step Atomic-level Energy Minimization, *Biophys. J.,***101:**2525-2534.
11. D.Kozakov (2006) PIPER: an FFT-based protein docking program with pairwise potentials, *Proteins,* **65:**392-406.
12. B.Raveh, N.London, O.Schueler-Furman (2010) Sub-angstrom modeling of complexes between flexible peptides and globular proteins, *Proteins*, **78:**2029-2040.
13. S.J. de Vries, M. van Dijk, A.M.J.J. Bonvin (2010) The HADDOCK web server for data-driven biomolecular docking, *Nature Protocols,* **5**:883-897.
14. S.Lyskov, J.J. Gray (2008) The RosettaDock server for local protein-protein docking, *Nucleic Acids Research*, **36:**233-238.