The Origin of Mitochondria-Specific Outer Membrane β-Barrels from an Ancestral Bacterial Fragment

Joana Pereira1,* and Andrei N. Lupas1

1Department of Protein Evolution, Max-Planck-Institute for Developmental Biology, Tübingen, Germany

*Corresponding author: E-mail: joana.pereira@tuebingen.mpg.de.

Accepted: September 27, 2018

Abstract

Outer membrane β-barrels (OMBBs) are toroidal arrays of antiparallel β-strands that span the outer membrane of Gram-negative bacteria and eukaryotic organelles. Although homologous, most families of bacterial OMBBs evolved through the independent amplification of an ancestral ββ-hairpin. In mitochondria, one family (SAM50) has a clear bacterial ancestry; the origin of the other family, consisting of 19-stranded OMBBs found only in mitochondria (MOMBBs), is substantially unclear. In a large-scale comparison of mitochondrial and bacterial OMBBs, we find evidence that the common ancestor of all MOMBBs emerged by the amplification of a double ββ-hairpin of bacterial origin, probably at the time of the Last Eukaryotic Common Ancestor. Thus, MOMBBs are indeed descended from bacterial OMBBs, but their fold formed independently in the proto-mitochondria, possibly in response to the need for a general-purpose polypeptide importer. This occurred by a process of amplification, despite the final fold having a prime number of strands.

Key words: motif amplification, protein evolution, remote homology, mitochondria, outer membrane.

Introduction

Amplification of subdomain-sized fragments is a dominant phenomenon in the evolution of protein folds, resulting in repetitive proteins that adopt a pseudosymmetrical fold (Andrade et al. 2001; Söding and Lupas 2003; Alva and Lupas 2018). One of these is the outer membrane β-barrel (OMBB), a closed antiparallel β-sheet whose strands traverse the outer membrane in Gram-negative bacteria, but also mitochondria, mitochondria-related organelles and plastids (Duy et al. 2007; Remmert et al. 2010; Zeth and Thein 2010; Chaturvedi and Mahalakshmi 2017). OMBBs preform a wide array of functions, from solute transport to membrane protein assembly, and are composed of a variable number of β-strands.

Gram-negative OMBBs have an even number of β-strands between 8 and 26 (Koebnik et al. 2000; Chaturvedi and Mahalakshmi 2017). Internal sequence symmetry suggests that the major families of OMBBs arose independently through the amplification of a homologous pool of ancestral ββ-hairpins (Remmert et al. 2010). While bacterial outer membranes contain many families of OMBBs, mitochondrial ones contain only two (fig. 1). One, formed by the 16-stranded SAM50/TOB55, clearly belongs to the OMP85 family of bacterial OMBBs (Kozjak et al. 2003). The other, comprising the 19-stranded TOM40 and VDAC (Bay et al. 2012), is found only in mitochondria and its origins are as yet unclear (Cavalier-Smith 2006; Zeth and Thein 2010). In the following we will refer to this family as mitochondria-only OMBBs (MOMBBs). In addition to TOM40 and VDAC, which are present in almost all lineages of eukaryotes (supplementary fig. 1, Supplementary Material online), this family also contains three lineage-specific members: MDM10 from fungi and amoebozoa (Finner et al. 2013); and TAC40 and ATOM, both from trypanosomes (Pusnik et al. 2011; Zarsky et al. 2012; Schnarwiler et al. 2014). In order to shed light on the origins of this family, we carried out a broad survey of OMBBs in mitochondria and bacteria.

Results and Discussion

Using PSI-BLAST, we screened the nonredundant protein database at NCBI for homologs of known MOMBBs (see Materials and Methods). We retrieved a set of 1,394 sequences, which illustrate the distribution of the five MOMBB subfamilies in the major eukaryotic lineages (supplementary fig. 1, Supplementary Material online). No bacterial matches were found for VDAC, TOM40, TAC40, and MDM10, but searches with ATOM resulted in four incomplete matches to a family of
uncharacterized bacterial OMBBs. This family is putatively 12-stranded and distantly related to FapF, an OMBB involved in the secretion of amyloid subunits during biofilm formation (Rouse et al. 2017). These searches did not result in any, even marginally significant matches to SAM50.

When clustered based on their sequence similarity (fig. 2a), VDAC, TOM40, and MDM10 form a highly connected supercluster to which TAC40 links via VDAC. ATOM sequences connect only distantly to other MOMBBs and cluster closer to bacterial FapF-like OMBBs. In HMM-profile searches, all five MOMBBs make statistically significant, full-length matches to either VDAC or TOM40, but not to bacterial OMBBs (fig. 2b and supplementary fig. 3, Supplementary Material online). Only local matches (with a coverage of 20–40%) are found between mitochondrial and bacterial OMBBs, especially VDAC and TOM40. These results support the notion that all MOMBBs, including ATOM (Zarsky et al. 2012), and the five members of the 19-stranded OMBB family unique to mitochondria (MOMBBs) TOM40, VDAC, MDM10, ATOM, and TAC40 (Pusnik et al. 2011; Bay et al. 2012; Zarsky et al. 2012; Flinner et al. 2013; Schnarwiler et al. 2014). For those whose three-dimensional structure is known (VDAC and TOM40), the experimental structure is shown. For those whose structure was not yet experimentally determined, homology models are shown only for illustrative purposes. For that, the best templates for the reference sequences were identified with HHpred (Zimmermann et al. 2018) over the PDB70 (as of May 2018; supplementary table 1, Supplementary Material online) and the models built with SWISS-MODEL (Biasini et al. 2014) after target-template alignment with PROMALS3D (Pei and Grishin 2014). Long loops are shown in dashed cartoon lines for clarity and named Lx, where x refers to the ranking of the loop in the predicted structure.

have a prime number of strands, the sequence alignment of the repeats shows that the first one lacks the first strand, which may have been converted to a helix, in order to generate a plug (fig. 4a).

To test whether MOMBBs may have been amplified independently from the same structural unit, we compared each repeat with all the others in our set of MOMBBs. Almost invariably, where significant matches were obtained, repeat n of one MOMBB had its best match in repeat n’ of another MOMBB (figs. 3c and 5). From this we conclude that 19-stranded MOMBBs diverged from a fully amplified ancestor, rather than being amplified individually.

Next, we searched for clues to the origin of the fragment from which MOMBBs were amplified. Because VDAC and TOM40 are the only MOMBBs that still show recognizable internal sequence symmetry, we focused the search on their repeats. Searches with HHsearch over the PFAM (Finn et al. 2016), TiGRFAMs (Haft et al. 2003), COG (Galperin et al. 2015), and NCBI’s Conserved Domains (CD; Marchler-Bauer et al. 2015) databases identified numerous matches to OMBB families at a significance >50% (supplementary table 2, Supplementary Material online). These connect MOMBBs to OMBBs that are mostly involved in small molecule transport in a wide range of Gram-negative bacteria, especially proteobacteria. Where detectable, the repeats of these OMBBs correspond to single ββ-hairpins, not double ones as in VDAC and TOM40.

Although MOMBBs and OMBBs share a conserved C-terminal β-signal for membrane insertion (Kutik et al. 2008;
Walther et al. 2009), and indeed the majority of the matches covered the C-terminal hairpin of the OMBBs, they were all obtained with the fourth repeat from VDAC, not the fifth, C-terminal one (supplementary table 2, Supplementary Material online). The best match obtained was to the last three strands of the BcsC family of sugar transporters (Whitney and Howell 2013), which is also the only matched OMBB family found in α-proteobacteria. This seemed particularly attractive, since mitochondria are thought to have descended from α-proteobacteria (Andersson et al. 1998; Roger et al. 2017). BcsC proteins share local and global sequence similarity with several families of transporters, including PgaA and FapF-like proteins (fig. 2). While BcsC does not show detectable sequence repeats, PgaA has a clear ββ-hairpin repeat and is also the only one of known structure (fig. 3a and b). Using it as a structural prototype for the BcsC family, we find in comparisons to VDAC and TOM40 that all double ββ-hairpins have closely matching structures (fig. 3c and d). This suggests that the high level of similarity between the fourth repeat of VDAC and the C-terminal strands of BcsC is not the result of structural constraints.

Conclusions

In conclusion, our analysis confirms the monophyletic relationship of VDAC and TOM40 (Bay et al. 2012), and extends it to all MOMBBs including ATOM, for which our results confirm that it is a distant form of TOM40 and did not evolve independently from a bacterial OMBB (Zarsky et al. 2012). As MOMBBs and OMBBs match in sequence only locally, and VDAC and TOM40 were probably part of the Last Eukaryotic Common Ancestor (LECA) proteome, it seems likely that the ancestor of all MOMBBs emerged in the proto-mitochondrion and was not acquired from the proteobacterial endosymbiont. Instead, it evolved independently by the amplification of a double ββ-hairpin related to those of OMBBs. The evolution of a new outer-membrane pore may have been driven by the need for a general-purpose polypeptide importer, a function for which there are no prototypes in the bacterial outer membrane. This need would have arisen in the early stages of endosymbiosis, after an increasing number of genes were transferred from the symbiont to the host nucleus, requiring it to reimport the encoded proteins. If this scenario is correct, then the ancestral function of MOMBBs...
would have been polypeptide import, possibly facilitated by sensitivity to an electrochemical gradient. The electrochemically gated diffusion of small molecules mediated by VDAC would then have represented a subsequent evolutionary development. The de novo evolution of a new pore implies that it was initially independent of signal sequences, which would have gradually evolved with the acquisition of further TOM proteins to the import machinery (Garg et al. 2015).

As the best match between MOMBBs and OMBBs covers the C-terminal strands of BcsC and this family occurs in x-proteobacteria, it seems attractive to propose that the last four strands of a proteobacterial transporter related to BcsC were amplified during the transition from a free-living organism to an endosymbiotic organelle at the time of the LECA. The amplification of these strands would have been particularly advantageous as they already include the appropriate
The amplification of the 4-stranded fragment would have yielded a 20-stranded barrel, yet MOMBBs have 19 strands. Given the size of the N-terminal α-helix present in all MOMBBs (fig. 4), it is possible that this arose from the N-terminal strand, driven by the need to gate the newly evolved pore. This resulted in the present-day MOMBB architecture of a 19-stranded barrel surrounding an α-helical plug, which is an important determinant in the sensitivity of MOMBBs to electrochemical gradients (Tornroth-Horsefield and Neutze 2008). The 20-stranded barrel at the origin of MOMBBs would represent a fold not yet identified in any kingdom of life (Chaturvedi and Mahalakshmi 2017). While substantiating that MOMBBs descended from bacterial OMBBs, but their fold formed independently in the proto-mitochondria, our results also highlight the role of motif amplification in the de novo emergence of new forms for established protein architectures.

Materials and Methods

Assembly of the MOMBB and OMBB Sequence Set

We assembled our set of MOMBB and OMBB sequences by preforming four rounds of PSI-BLAST searches using the MPI Bioinformatics Toolkit (Zimmermann et al. 2018). Searches for MOMBB sequences were preformed over the nr database (as of May 2018) using the reference sequences of TOM40,
VDAC, MDM10, ATOM, and TAC40 (fig. 1) while searches for FapF-like and BcsC-like OMBBs were preformed over the bacterial part of \( nr \) (\( nr\_bac \)) (as of May 2018) using the sequences of the barrel regions identified in the sequences of *Rhodanobacter* sp. *Soil772* FapF-like (UniprotKB: A0A0Q9PBF2), *Pseudomonas* sp. *UK4* FapF (UniprotKB: C4IN73) and *Escherichia coli* BcsC (UniprotKB: P37650) and PgaA (UniprotKB: P69434). In order to identify these barrel regions, we searched for reference structures for these sequences on the PDB70 and SCOPe databases (as of May 2018) with HHpred, without scoring for secondary structure, and predicted their secondary structure content with Quick2D (Alva et al. 2016). In both cases, the parameters were set to default.

Identification and Comparison of Sequence Repeats

The repetitive nature of the HMM-consensus sequences was predicted with HHRepID (Biegert and Söding 2008; Zimmermann et al. 2018), using default parameters without the generation of a new multiple sequence alignment, and their secondary structure content predicted with Quick2D as described above. By extracting their corresponding regions in the alignments of the various barrels, we built HMM-profiles as described above for each of the repeats identified. The regions in MDM10, TAC40, and ATOM were assigned by mapping them to the VDAC and TOM40 consensus sequences, and those in BcsC by mapping to PgaA. To test the independent amplification of MOMBBs and OMBBs, the resulting HMM-profiles were aligned with hhalign, as described above, and the corresponding double \( \beta \beta \)-hairpins structurally compared by structural alignment with TMalign (Zhang and Skolnick 2005).

Identification of Bacterial OMBBs Matching MOMBB Repeats

To investigate the origins of the double \( \beta \beta \)-hairpins from the ancestor of all MOMBBs, the HMM-consensus sequence of the double \( \beta \beta \)-hairpins from VDAC and TOM40 were used for searches over the PFAM, TIGRFAM, CD, and COG databases (as of August 2018) with HHpred, without scoring for secondary structure. The secondary structure content and the repetitive nature of the protein families matched in the searches were predicted, respectively, with Quick2D and HHRepID as described above. The taxonomic distribution of these families was retrieved from PFAM and eggNOG (Huerta-Cepas et al. 2016) as of August 2018.

Supplementary Material

Supplementary data are available at Genome Biology and Evolution online.

Acknowledgments

We thank Vikram Alva and Jens Baßler for stimulating discussions. This work was supported by institutional funds from the Max Planck Society.

Literature Cited

Alva V, Lupas AN. 2018. From ancestral peptides to designed proteins. Curr Opin Struct Biol. 48:103–109.
Alva V, Nam S-Z, Söding J, Lupas AN. 2016. The MPI bioinformatics Toolkit as an integrative platform for advanced protein sequence and structure analysis. Nucleic Acids Res. 44(W1):W410–W415.
Andersson SGE, et al. 1998. The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. Nature 396(6707):133–140.
Andrade MA, Perez-Iratxeta C, Ponting CP. 2001. Protein repeats: structures, functions, and evolution. J Struct Biol. 134(2–3):117–131.
Bay DC, Hafez M, Young MJ, Court DA. 2012. Phylogenetic and coevolutionary analysis of the \( \beta \)-barrel protein family comprised of mitochondrial porin (VDAC) and Tom40. Biochim Biophys Acta – Biomembr. 1818(6):1502–1519.
Biasini M, et al. 2014. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. Nucleic Acids Res. 42(W1):W252–W258.
Biegert J, Söding J. 2008. De novo identification of highly diverged protein repeats by probabilistic consistency. Bioinformatics 24(6):807–814.
Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25(15):1972–1973.
Cavalier-Smith T. 2006. Origin of mitochondria by intracellular enslavement of a photosynthetic purple bacterium. Proc R Soc B Biol Sci. 273(1596):1943–1952.
