Supplemental Information

Structural basis of outer membrane protein biogenesis in bacteria

Reinhard Albrecht and Kornelius Zeth

Figure Legends of Supplemental Figures

Fig. S1. *Sequence conservation of BamA*

The sequence of BamA from *E. coli* was the basis to generate a multiple sequence alignment against the UNIPROT database. Using FRpred conserved residues are marked by blue colours with dark blue representing strong similarity and light blue representing weaker similarity (1).

Fig. S2. *Structure of the nine-bladed β-propeller*

A Architecture of the nine bladed β-propeller through assembly of four BamB protein fragments comprising WD40 repeats B1 - B3 (in red), B2 - B3 (in orange) and two B5 - B6 fragments (shown in blue). The structure in cartoon representation is displayed as top and after tilting the model around the x-axis by 90 degrees as side view. *B* Surface representation of the nine-bladed propeller structure which clearly indicates the pore dimensions based on this new blade architecture.

Fig. S3. *Subtilisin treatment of full length BamC*

Preparative subtilisin treatment of BamC has been performed as described (2). Full-length BamC (lane 1) was treated with subtilisin and the reaction mixture (lane 2) was subjected to NiNTA-chromatography. The N- and C-terminal domains were separated by NiNTA affinity chromatography and the resulting proteins are obtained in the flow through (lane 3 - $BamC_{ND}$) and elution fraction (lane 4 - $BamC_{CD}$). Both domains were further used for crystallization.

Fig. S4. *Structural details of $BamC_{ND}$ and $BamC_{CD}$*

(A) The structure of the artificial $BamC_{ND}$ dimer is shown in two different views. Termini (NT and CT) and secondary structure elements are marked ($\alpha1 - \alpha2, \beta1 - \beta5$) as well as residues involved in dimerization. The two domains adopt an anti-parallel orientation. (B) Topology diagram of $BamC_{ND}$ with strands ($\beta1 - \beta5$) are marked in blue and helices ($\alpha1 - \alpha2$) in red. (C) Topology diagram of the C-terminal domain in the same representation as defined for (B) with strands ($\beta1 - \beta6$) are marked in blue and helices ($\alpha1 - \alpha3$) in red. The

topology diagram indicates two additional secondary structure elements (β1 and α2). (D) Representation of $BamC_{ND}$ as surface plot illustrated from two different orientations. Conserved residues are marked in orange and numbered according to their occurrence in the protein sequence. (E) Surface representation of the $BamC_{CD}$ domain shown from two sides. Residues conserved in BamC proteins are marked in red and numbers assign the individual positions in the sequence.

Fig. S5. - *Protein complexes with domains structurally related to $BamC_{CD}$*
Structure of $BamC_{CD}$ is kept in the same orientation for (A) - (C).
(A) Superposition of the Roc-COR complex (PDB-entry: 3DPU) onto $BamC_{CD}$ (3). (B) Superposition of $BamC_{CD}$ onto AMP-activated protein kinase (PDB-entry: 2QRD) (4). (C) Related AMP-activated protein kinase superimposed onto $BamC_{CD}$ (PDB-entry: 2V8Q) (4).

Fig. S6. *Topology of BamE and conservation pattern of the protein*
(A) Topology model of the dimeric BamE structure. The two subunits are marked in blue and orange. Secondary structure elements are marked (β1 – β3) as well as protein termini (NT and CT). The intertwining and pseudoknot-formation of the protein subunits is visualized. Conserved residues of one protein chain are marked with blue dots according to the conservation analysis given in (B). (B) Conservation pattern of the BamE class of lipoproteins using multiple alignments and the probability weighting scheme for conservation as defined in the program FRpred (1).

Fig. S7**.** *Analysis of BamE in the context of known structure models*
(A) Superposition of the OmlA lipoprotein (in red; PDB-entry 2PXG; (5) and the monomeric structure of BamE determined by NMR methods (in light blue; PDB-entry 2KXX; (6) onto the dimeric crystal structure of BamE (in orange and dark blue). The OmlA lipoprotein was superimposed with an r.m.s.d of 3.2 Å for 33 Cα atoms onto the BamE crystal structure. The NMR structure showed a deviation of 2.1 Å for 40 Cα atoms which were aligned due to distance restraints (< 3.5 Å). (B) Superposition of the beta-lactamase inhibitor structures (BLI; PDB-entries 3GMV in forest green and 3GMX in red) onto a monomer BamE of *E. coli* (7). (C) Topology diagram of the beta-lactamase inhibitor (PDB-entry: 2G2U) with □-helices marked in red and β-sheets (β1 – β8) assigned in blue (8). (D) For comparison with (C) the topology diagram of BamE is provided comprising the same nomenclature as in Figure S3A. (E) Superposition of the 2G2U beta-lactamase inhibitor onto the dimeric BamE structure. It is

apparent that by a small movement of the two BamB monomers relative to each other the superposition of β-strands between BamB and the BLI would fit and the similarity between the two molecules would become more evident. (F) Superposition of BamE onto the 2G2U structure in complex with the corresponding beta-lactamase (BL). The active site of the BL is encircled and nearby the BLI and a monomer of the BamE structure. Selected conserved residues of BamE at the interface are highlighted.

Fig. S8. *Analysis of conservation patterns in the BamD protein group*

(A) Analysis of conserved residues on BamD using the program FRpred (1). These residues are marked on the surface representation of BamD shown in Figure 6D. (B) Table showing distances between BamD and the His6-tag of the neighbouring monomer (see Figure 6B for comparison) of the crystal packing. The table was generated using the program PISA (http://www.ebi.ac.uk/msd-srv/prot_int/pistart.html). Hydrogen bonds and salt bridges were marked with colours (red and magenta) if the interaction was based on conserved residues.

Fig. S9. *Analysis of conservation patterns on BamD*

(A) Assignment of the five TPR domains to the sequence of BamD. In the lower panel the structure alignment of the five TPR domains is shown. (B) Superposition of two TPR structures - BamD (in orange) and the structurally closest homolog co-chaperone Hop (PDB-entry: 1ELR; (9)). The superposition reveals the proximity of the two peptides (C-terminus of Hsp90 and the His6-terminus of BamD) in the binding scaffold of TPR_{1-3}. (C) Structure of BamD with the C-terminal portion marked in grey. This part of the protein was demonstrated to be essential for the interaction with BamA and termed *yfio4* or *yfio5*, respectively.

Supplemental Reference

1. Fischer, J. D., Mayer, C. E., and Soding, J. (2008) *Bioinformatics* **24**, 613-620
2. Albrecht, R., and Zeth, K. (2010) *Acta Crystallogr Sect. F Struct. Biol. Cryst. Commun.* **66**, 1586-1590
3. Gotthardt, K., Weyand, M., Kortholt, A., Van Haastert, P. J., and Wittinghofer, A. (2008) *EMBO J.* **27**, 2239-2249
4. Xiao, B., Heath, R., Saiu, P., Leiper, F. C., Leone, P., Jing, C., Walker, P. A., Haire, L., Eccleston, J. F., Davis, C. T., Martin, S. R., Carling, D., and Gamblin, S. J. (2007) *Nature* **449**, 496-500
5. Vanini, M. M., Spisni, A., Sforca, M. L., Pertinhez, T. A., and Benedetti, C. E. (2008) *Proteins* **71**, 2051-2064
6. Kim, K. H., Kang, H. S., Okon, M., Escobar-Cabrera, E., McIntosh, L. P., and Paetzel, M. (2011) *Biochemistry* **50**, 1081-1090

7. Gretes, M., Lim, D. C., de Castro, L., Jensen, S. E., Kang, S. G., Lee, K. J., and Strynadka, N. C. (2009) *J. Mol. Biol.* **389**, 289-305
8. Reynolds, K. A., Thomson, J. M., Corbett, K. D., Bethel, C. R., Berger, J. M., Kirsch, J. F., Bonomo, R. A., and Handel, T. M. (2006) *J. Biol. Chem.* **281**, 26745-26753
9. Scheufler, C., Brinker, A., Bourenkov, G., Pegoraro, S., Moroder, L., Bartunik, H., Hartl, F. U., and Moarefi, I. (2000) *Cell* **101**, 199-210