Nuclear pore-like structures in a compartmentalized bacterium

Evgeny Sagulenko[1], Amanda Nouwens[1]*, Richard I. Webb[2]*, Kathryn Green[2], Benjamin Yee[1], Garry Morgan[2], Andrew Leis[3], Kuo-Chang Lee[1], Margaret K. Butler[1], Nicholas Chia[4], Uyen Thi Phuong Pham [1], Stinus Lindgreen[5], Ryan Catchpole[5,6], Anthony M. Poole[5,6,7] and John A Fuerst[1]

[1]School of Chemistry and Molecular Biosciences, The University of Queensland, Brisbane, Queensland 4072, Australia

[2] Centre for Microscopy and Microanalysis, The University of Queensland, Brisbane, Queensland 4072, Australia

[3]CSIRO - Livestock Industries, Australian Animal Health Laboratory, Biosecurity Microscopy Facility (ABMF)
Geelong, Victoria 3220 Australia

[4]Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

[5]School of Biological Sciences, [6]Biomolecular Interaction Centre, [7]Allan Wilson Centre, University of Canterbury, Private Bag 4800, Christchurch 8140, New Zealand

*These authors contributed equally to this work

Corresponding author: Emeritus Professor John A. Fuerst, School of Molecular and Microbial Sciences, The University of Queensland, Brisbane, Queensland 4072, Australia

Phone: +61 402 481 714

Email: j.fuerst@uq.edu.au

ABSTRACT

Planctomycetes are distinguished from other Bacteria by compartmentalization of cells via internal membranes, interpretation of which has been subject to recent debate regarding potential relations to Gram-negative cell structure. In our interpretation of the available data, the planctomycete *Gemmata obscuriglobus* contains a nuclear body compartment, and thus possesses a type of cell organization with parallels to the eukaryote nucleus**.** Here we show that pore-like structures occur in internal membranes of *G.obscuriglobus* and that they have elements structurally similar to eukaryote nuclear pores, including a basket, ring-spoke structure, and eight-fold rotational symmetry. Bioinformatic analysis of proteomic data reveals that some of the *G. obscuriglobus* proteins associated with pore-containing membranes possess structural domains found in eukaryote nuclear pore complexes. Moreover, immuno-gold labelling demonstrates localization of one such protein, containing a β-propeller domain, specifically to the *G. obscuriglobus* pore-like structures. Finding bacterial pores within internal cell membranes and with structural similarities to eukaryote nuclear pore complexes raises the dual possibilities of either hitherto undetected homology or stunning evolutionary convergence.

INTRODUCTION

A nucleus surrounded by a double membrane envelope is a universal characteristic of eukaryote cells [1] and is thought to be universally absent from the prokaryote domains Bacteria and Archaea. The nucleus is accompanied by a complex apparatus for transport of macromolecules, including a multi-protein nuclear pore complex embedded in the nuclear envelope, and a soluble transport system [2]. The nuclear pore complex and many of its component proteins appear universal among eukaryotes, spanning yeast, trypanosomes and vertebrates [3], and the Last Eukaryotic Common Ancestor already possessed a complex version of the nuclear pore complex, nuclear envelope and connected endomembrane system [4,5,6]. *Gemmata obscuriglobus,* a member of the bacterial phylum *Planctomycetes*, possesses compartments including the nuclear body containing DNA and ribosomes, riboplasm containing ribosomes but no DNA, and the paryphoplasm, a ribosome-free compartment [7,8]. By whole-cell tomography and conventional transmission electron microscopy (TEM), we have previously established that the riboplasm and nucleoid compartments of *G. obscuriglobus* cells are bounded by membranes [7,8,9]. Confocal fluorescence micrographs of cells where the nuclear region has been stained with DiOC6 membrane stain and DAPI DNA stain are also consistent with the membrane-bounded nature of the DNA in this organism [10]. An earlier study of *G. obscuriglobus* internal membranes differs in its conclusions from those of ours, proposing that only one invaginated membrane exists in such cells and that there is no membrane enclosure of the *Gemmata* chromosome [11], and instead a tubulovescular model for internal membranes has been proposed [12]. Our tomography analysis of *G.obscuriglobus* cells demonstrated that the internal membranes do not display continuity with the cytoplasmic membrane apposed to the cell wall [7]. Such a cell plan implies specialized internal membrane(s)

distinct from the cytoplasmic membrane and would also require some form of transport system (e.g. pore structures) for macromolecules passing between the internal compartments and the rest of the cytoplasm. This hypothesis is consistent with the recent finding of confinement of translation to non-nucleoid regions of *G. obscuriglobus* cells [13]. A corollary of our study of *G. obscuriglobus* internal compartments was that several different types of membranes might be isolatable from lysed cells, and we have confirmed this concept here. There has been extensive debate regarding the evolutionary significance of compartmentation in *G. obscuriglobus* [14,15,16]. However, such discussions have been limited by lack of knowledge about *Gemmata* membrane composition and the structure of internal membranes in particular. Recently, components of cell walls characteristic for Gram-negative bacteria such as peptidoglycan and lipopolysaccharide have been found in *Gemmata obscuriglobus* [17,18] correlating with other data on occurrence of peptidoglycan in *Planctomyces limnophilus* [17]and an anammox planctomycete species[19]. The exact location of these components within planctomycete is yet unknown, but the results suggest a potential for planctomycete cell plan to relate more closely to Gram-negative cell wall and structure than previously thought as outlined in published hypotheses[20,21]. The implications of these results for interpretation of planctomycete internal membranes and their evolutionary significance are not yet clear. Here we present evidence that some of the internal membranes of *G. obscuriglobus* possess pores with complex structure. Moreover we identify proteins specific to these membranes, some of which possess structural domains also found in eukaryote nucleoporins. The evolutionary implications of these results are considered, both from the perspective of common ancestry with the eukaryote nuclear pore complex, and from the viewpoint of convergent evolution.

RESULTS AND DISCUSSION

Planctomycetes possess pores in internal membranes

Pore-like structures (termed 'pores' throughout the remainder of the text) in the internal membranes of the planctomycete bacterium *Gemmata obscuriglobus* can be observed in transmission electron microscopy images of thin or thick sections of whole cells (Fig 1A, 1B, S1 Fig). When thin sections of cells prepared either via high-pressure freezing or cryosubstitution without high-pressure are examined, favourable planes of section reveal pore-like structures along sectioned nuclear envelopes. In Fig 1B a circular complex (arrowhead) appears in a gap between folded regions of the nuclear envelope membranes (arrows). Appearance of the pore is comparable to that of sectioned eukaryote yeast pores prepared by high-pressure freezing [22]. The diameter *ca.* 35 nm of the circular complex in Fig 1B is consistent with the diameter of pore structures in negatively- stained preparations (see below). One ring of the pore complex analogous to that of the nuclear and cytoplasmic rings of the eukaryote nuclear pore may be visible *en face* by a favourable tangential plane of section (Fig 1B). The position of such a pore within the membrane in a whole cell can be seen in S1A Fig, and the details of ring and central plug and spoke-like structures within such pores seen in face are also evident in S1B Fig.

When cells of *G. obscuriglobus* are gently lysed, internal membrane can be released through the broken cell walls (Fig 1C), the overall shape of which is consistent with a sphere or compressed sphere (Fig 1C and 1D**)**. The released membranes are covered with circular ring-like structures, possessing a dense circular centre (Fig 1E).

We have also demonstrated the presence of pores on the internal membranes of intact *G. obscuriglobus* cells using the freeze-fracture technique (Fig 2A and 2B). In freeze-fracture replicas, it is possible to unambiguously identify the membrane surfaces displaying pores on a single major internal membrane-

bounded organelle, since the position of this membrane-bounded structure is clearly visible within the cross-fractured cells. The pores, circular structures, appear on one outer surface or fracture surface of the organelle envelope, which clearly overlies another fractured membrane of this envelope, indicated by the splits visible in the envelope. The complexity of the organelle membranes has been observed in *Gemmata* in tomographic and thin-sectioning studies [7,8,10]. In enlarged view (Fig 2A insets, Fig 2C) each pore represents a circular complex in which an outer ring surrounds a centre which usually appears darker than the outer ring to a degree depending on the metal shadow angle relative to the pore aspect. The size of pores in the freeze-fracture replica in Fig 2C are *ca.* 32 - 45 nm, which is consistent with dimensions obtained from negative staining but such dimensions would be expected to be more variable due to metal shadow used to prepare the replicas**.** It is of interest that we obtained similar freeze-fracture replica images of pore-like structures with central core and ring structure on internal membranes in another *Gemmata* strain, CJuql4 (ACM5157) (S2 Fig), closely related to the general *Gemmata* cluster phylogenetically, and a member of the same roughly genus-level phylogenetic group with high bootstrap support in a phylogenetic tree of planctomycetes [23]. These pore-like structures have a central core and ring structure and are 23-38 nm in diameter. Cells of this additional *Gemmata* group strain have been shown to possess an internal membrane-bounded nucleoid-containing compartment in thin sections, similar to that of *Gemmata obscuriglobus* [23]. These results therefore reinforce the conclusions from *Gemmata obscuriglobus* regarding the significance of pores as structures associated with the *Gemmata* group-characteristic internal membrane envelopes often surrounding the nucleoid region. A membrane-bounded region surrounding the nucleoid represents a type of structure known only from the *Gemmata* group strains so far among either planctomycetes or among any other domain Bacteria.

In negatively-stained membrane fragments released from cells lysed by sonication, the pores consist of a thin inner ring immediately surrounding the electron-dense pore centre, and an outer thicker ring distinguishable from the inner ring (Fig 3A). Assuming uniform distribution, there are ca. 87 such pores per μm^2 of such a membrane sheet. At higher magnification the pores of *G. obscuriglobus* appear to be composed of subunits with internal "plug" (Fig 3B and 3C). In membrane fragments released by mechanical lysis and negatively stained with uranyl acetate, the outer diameter of the pore based on the outer ring diameter was calculated as 33.5 ± 2 nm, the diameter of the inner ring as 17.5 ± 2 nm, and that of the electron-dense pore centre (equivalent to an 'inner pore' or 'central plug') as 9.5 ± 0.8 nm (S3 Fig).

After treating of the membranes released from lysed cells and isolated (as 'fraction 3', see below) via density gradient fractionation with detergent, aggregates of individual pores could be detected (S4 Fig) with only degraded membrane between them. Individual pores in such preparations show a central dense core surrounded by a light ring and in some cases material projecting from the outer rim of the ring possibly representing spokes normally connecting inner to outer ring in intact pore complexes.

In addition to these pores, two smaller classes of pores are also found. They may be seen most clearly in membrane fragments released via sonication (Fig 3A). The larger of these two classes consists of rings that are 14.5 ± 2 nm in diameter with an inner dense centre of 5 ± 1 nm wide, while the smaller class consists of pores 6± 0.9 nm wide and which also possess a dense centre and can appear in clusters. Neither of these smaller pore types seem comparable in structure to eukaryote nuclear pore structures in the sense of possession of both inner and outer rings as well as an inner dense centre. It is also possible that at least the 14.5 nm diameter class is the result of a reverse view ('basket' side view) of only part of

the larger pore structures seen because of a folded membrane, since they appear most clearly in examinations of negatively stained membrane sheets rather than via other EM preparation methods. If true additional pore types, they suggest either specialization in pore function in these membranes or some stages in assembly for the largest pores. It is the largest pore type we have seen which is in any case most relevant to the argument of this paper since it is the class displaying most complex structure and is the class comparable with that of eukaryotic nuclear pores.

Thus, we have determined here from studies of whole cells and membranes released from lysed whole cells that at least one type of internal membrane contains pores. This has been demonstrated by three distinct electron microscope preparative techniques – thin-sectioning and freeze-fracture replica technique of the whole cryosubstituted cells, and negative staining of membranes released from lysed cells. If pores are genuine structures, one would predict that several different methods should be consistent e.g although negatively stained internal membranes released from lysis display the clearest examples of pores, methods such as freeze-fracture and thin-sectioning should also reveal entities with comparable structure on internal membranes in whole cells, and this is indeed the case. If such structures are significant functionally, one would expect they would be associated with a specific type of internal membrane and be able to find differences in protein composition of that distinct type of internal membrane.

To confirm the association of pores with a specific type of internal membrane and to investigate the composition of such membranes further we fractionated the internal membranes and applied proteomics and EM methods including immunoelectron microscopy to localize a specific protein of potential relevance to pore structure.

There are three types of membranes in *G.obscuriglobus* cells, one of which contains pores

To purify membranes for structural and proteome studies we applied a two-step density gradient fractionation technique (S5 Fig). The discontinuous and subsequent continuous gradient fractionations aimed to separate the membranes at the highest possible level. This procedure resulted in appearance of three distinct membrane types (S6 - S8 Fig). Only one of the fractions (fraction 3, consisting of characteristic 'canoes'), displayed pores on the membrane surfaces when examined via TEM after negative staining (S7 Fig). The large pores are present at high density (ca. 200 / μm^2 of membrane sheet) and similar in size and structure to the complex pores seen in unfractionated membrane released from lysed cells (e.g. Fig. 3A). We were able to apply clear markers for two of these fractions via an antibody against a beta-propeller-containing protein (a protein exclusive to fraction 3, see below and S9A Fig), and via a specific antibody against *G. obscuriglobus* clathrin-like membrane coat (MC) protein gp4978 (shown in a previous study to react specifically with membrane vesicles associated with endocytosis-like protein uptake in *G. obscuriglobus* [24]. The anti-beta-propeller-containing protein reacted only with fraction 3 but not with fractions 2 or 6, and the anti-MC protein antibody demonstrated reactivity only against fraction 2 (S9B Fig)**.**

The gradient fractionation technique demonstrates that membranes related to distinct internal structures of *G.obscuriglobus* can be separated, and the probing with antibodies clearly shows that the fractions do not contain significant amount of cross-contamination.

Structural analyses of the pores reveal their similarity to the nuclear pores of eukaryotes

Electron tomography of the fraction 3 membranes shows that, topologically, these pores are membrane insertions (Fig 4A and 4B, S1 Video and S2 Video). By analysing digital slices through the tomogram, we see continuous membrane envelope followed by a pore structure and then followed again by continuous membrane (Fig 4B). Such pores display a projection on one side and in some slices the plug could be seen (Fig 4A and 4B)**.** Cryo-EM analyses (S10 Fig) and a modified Markham rotation analysis of the electron micrographs from frozen-hydrated pore-containing membranes (Fig 4C and 4D) suggests an eight-fold symmetry for the organization of these pores, indicating that these complexes are likely to be modular constructions. It should be noted that we have used the same technique here as applied in the first demonstration of eight-fold symmetry in nuclear pores from invertebrate and vertebrate animal species[25]. Eight-fold rotation gave strongest reinforcement e.g. relative to 7-fold rotation. Markham rotation cannot by itself prove 8-fold symmetry of pore structure, but taking both the Markham analysis and the clear radial and octagonal symmetry visible in the original micrographs into account, on current evidence, 8-fold symmetry is the most likely possibility.

3-D reconstructions of the pores inserted into the membranes confirms a basket-like structure, with struts connecting the pore to a distal ring, and projecting from the part of the pore inserted in membrane (Fig 5 and S3 Video).

In summary, the pores embedded in the internal membranes of *G.obscuriglobus* display startling similarities to the nuclear pore complex of eukaryotes [26,27,28,29]. For structural comparison of the *Gemmata* pore with a eukaryotic pore model [2] see S11 Fig. The pores are comparable to the appearance and distribution of nuclear pores on negatively stained isolated nuclear membranes of yeast, plant and mammalian liver cells [30,31,32]. Those similarities include apparent eight-fold symmetry of

its components, the presence of a basket structure projecting asymmetrically from the membrane plane, and the presence of at least two rings within that plane. These structurally complex pores of *G.obscuriglobus* are considerably smaller than nuclear pores on isolated nuclear membranes of eukaryotic cells. At *ca.* 35 nm in frozen-hydrated membranes, they are only approximately a third the diameter of characterized nuclear pores [28]. The frozen-hydrated NPCs of yeast are 96 nm wide [33], the *Xenopus* or yeast nuclear pore complexes in thin-sectioned cells are *ca.* 120 nm and 103 nm wide respectively [22,34], negatively stained NPCs of *Xenopus* are 107 nm wide [35], and detergent-released negatively stained are 133 nm wide [36], and finally intact *Dictyostelium* slime mould NPCs studied by cryo-electron tomography are 125 nm wide [26]. However, despite the difference in size of the *Gemmata* pores compared to those of eukaryote nuclear envelopes, there are clear analogies in structure, since the eukaryote pores also have a central plug and a central ring-like assembly composed of spokes sandwiched between a cytoplasmic ring and nuclear ring[22], and in negatively stained detergent-treated nuclear envelopes of *Xenopus* oocytes pores are visible as rings containing a central plug connected to the ring by spokes [27].

Our structural studies of the *Gemmata* pores are summarised in a deduced model (Fig 6). From Cryo-EM (Fig 4C and 4D) the pores from the "basket side" have two rings of ca 20 nm (inner ring) and 35 nm (outer ring) in diameter. The pores from this side display an octagonal symmetry, consistent with rotational folding analysis of cryo-electron micrographs from frozen-hydrated TEM of pore-containing membranes (Fig 4C and 4D). Within the central core region is a central 'plug' of *ca.* 10 nm in diameter (Fig 4C and 4D). At the higher magnification the central plug is seen as a structure connected to the inner side of the pore (Fig 3C). From the opposite side the pores also have two rings. As from transmission electron micrographs of the whole compartments lysed by grinding in liquid N_2 (Fig 1C

and 1E), Pt/C-shadowed replicas of whole cells (Fig 2C), and 3D reconstructions of the pores (Fig 5A and 5E) the outer ring is *ca.* 25 nm and the inner ring is *ca.* 20 nm in diameter. As calculated from the isolated membranes used for tomography studies (Fig 4B) the maximum distance from the top side of the pore to the end of the basket is 25 nm.

Whole membrane proteome analyses reveal distinct protein content for three types of internal membranes and the cell wall

Proteomics was applied to cross-compare proteins from the fractions obtained via gradient centrifugations (see above) in order to distinguish proteins unique to the pore-containing membranes. Protein composition appeared different for pore-containing membranes (fraction 3) compared to those with no visible pores (fractions 2 and 6) (Fig 7A). 342 membrane and membrane-associated proteins were identified from all fractions examined (Fig 7B, S1 Table), 46% of which are annotated as hypothetical proteins. In non-pore-containing membranes (fractions 2 and 6), many constituents of the respiratory chain, ABC transporters and secretion system components were identified. The recently described MC-like vesicle-associated protein [37] was found exclusively as a constituent of fraction 2 (S1 Table). Fraction 2 appears enriched with vesicles, which are present at high number within the paryphoplasm. Some vesicles may be formed during endocytosis and are derived from the cytoplasmic membrane [24,38]. Proteins such as glycosyl transferases, a number of dehydrogenases, including the NADH-dependent dehydrogenase, and the periplasmic solute-binding protein, were restricted to fraction 6, which suggests that this fraction is enriched for the cytoplasmic membranes, and the other two fractions do not contain detectable amount of cytoplasmic membrane debris. ATP synthase was found *only* in this cytoplasmic membrane fraction (fraction 6) and in the vesicle-enriched membrane fraction (fraction 2), and *not* in the pore-containing membrane fraction (fraction 3). This indicates that the

function of the pore-containing membranes is probably not that of electron transport or energy generation. In the pore-containing membranes (fraction 3), we identified 128 proteins, 39 (30.5%) of which were unique to this fraction (S2 Table). Common proteins found by proteomics analysis in all three fractions (S1 Table) are predicted pilins and predicted ribosomal proteins, which we believe are the result of contamination since actual pili structures were detected in small number by electron microscopy in all three fractions, and heavy ribosomal subunits which might be distributed along the entire gradient. In our proteomic analysis regarding significant proteins correlated with presence of pore structures, we only consider proteins exclusive to fraction 3.

It should be noted that the pores described here structurally resemble crateriform structures, a characteristic signature structure of planctomycete surfaces [39] and isolated cell walls[40] including those of *G. obscuriglobus*[41]. These crateriform structures could be seen on the surface of whole cells as circular regions and in the cell slices used for TEM as pits protruding the cell wall and cytoplasmic membrane, confirming published data on *G. obscuriglobus* [42]. It might be proposed that the pores described here represent in fact the crateriform structures as seen in purified cell walls (Fig S12). However, these crateriform structures when negatively stained display a dense centre but only one surrounding ring, so apparently differ from the more complex internal membrane pores. The membrane fractions we used as starting material for the gradient fractionations should not have contained significant amounts of wall, as they are effectively eliminated during preparations of the membranes for initiating the gradient fractionations; the walls are denser than any of the membranes and are pelleted and thus separated at relatively low centrifugation speed. Any walls remaining after this step are pelleted during fractionation in any case. The walls also cannot be lysed in the buffers usually used for dissolving the membranes for Western blot analysis, so their proteins would not be detected in isolated gradient-

fractionated membranes. Compared to the membranes, the walls are highly resistant to boiling in 10% SDS, and such boiling is used for purification of the walls from the membranes during wall isolation. In a separate experiment we have isolated the walls via boiling the *G.obscuriglobus* cells with 10% SDS and analysed their protein content. The cell walls revealed some proteins homologous to proteins which have in previous studies been shown to be characteristic of the wall of planctomycetes [43,44]. Via proteomics we have identified the major constituents of the walls, the so-called YTV-proteins (Gobs_U38067, GobsU_28375, and GobsU_21360). The bacterial cell wall marker peptidoglycan has now been reported to comprise at least part of the wall composition in *G. obscuriglobus* and some other planctomycetes [17,19], and lipopolysaccharide has been reported in *Gemmata obscuriglobus* [18] but proteins appear to comprise significant proportions of wall in *G. obscuriglobus* as well as other planctomycetes [40,41,45]. However, no planctomycete cell wall/surface protein homologs were found via proteomics in any of the three membrane fractions isolated from fractionation of lysed cells, reinforcing evidence from analysis of enzyme markers that all of these three fractions are from membranes other than cell wall/cell surface structures [42]. In addition, sectioned whole cells immuno-gold-labelled with a fraction 3 specific antibody (see below) were shown to label only internal membranes, and there was no labelling of walls or other surface structures. Thus we conclude that the pores in fraction 3 membranes are not wall/surface crateriform structures either on the criterion of location within the cell or on a criterion of protein composition. The pores represent structures embedded into internal membranes and do not represent crateriform structures or any other wall components. This implies potential performance by such internal membranes of functions such as transport of material between internal membrane-bounded cell compartments.

Thus, we have demonstrated here that the three membrane types separated by density gradient centrifugation are distinct from each other in their composition of specific proteins. Continuity of cytoplasmic membrane with internal *Gemmata* membranes has been proposed as part of the concept of the planctomycete cell plan as essentially one of a classical Gram-negative cell [11,46]. However, our data concerning protein composition of isolated membrane fractions does not support this concept, but rather is consistent with a concept that genuine internal compartmentalisation exists within the *Gemmata* planctomycete cell, where the compartments are separated from each other by different types of membranes. Our data is also consistent with distinction of cytoplaamic membrane from two different types of internal membranes.

Bioinformatic analyses of the membrane proteome

We characterised the identified set of proteins using a range of bioinformatics-based analyses (S Text and S3 - S6 Tables). Initial blast analyses indicated that many of the proteins showed little or no similarity to proteins outside *G. obscuriglobus*. We therefore performed Blast cluster analysis (VisBLAST) to establish whether any proteins in the pore-containing membranes exhibit sequence similarities to one another. We also performed profile-based (phmmer) screens to search for more distant similarities, and ran structural predictions using Phyre2 [47] to assess similarity of proteins to known folds. Sequence and structural analyses of our membrane proteomics data revealed the presence of a number of bacterial transmembrane proteins, including outer membrane efflux proteins, translocons and porins (S Text), underscoring the bacterial nature of these membranes. However, none of these were unique to pore-containing fraction 3, so their origin as contaminants from cell structures or components other than membranes of fraction 3 is possible, and they cannot be implied as characterizing any specific fraction 3 membranes. In addition, published structural data are not consistent with any of these

transmembrane or 'outer membrane' protein homologs forming pores with dimensions and attributes similar to the pores we observe in *G.obscuriglobus*.

Cluster analyses performed on all proteins identified through proteomics revealed two groups of proteins from fraction 3 with substantial sequence similarity (S13 and S14 Fig). Phyre2 models generated for one of these clusters identified a conserved C-terminal beta-propeller fold for 8 of the 11 proteins making up this cluster (Fig 7C; S15 - S18 Fig, S4 Table). Beta-propeller folds are found in protein constituents of eukaryotic nuclear pore complexes and coated vesicles, and it has been proposed that the eukaryotic nuclear pore and coatomer complexes evolved from a suite of membrane-curving proteins with common structural elements [48]. We therefore searched for evidence of other folds associated with eukaryotic nuclear pore complexes. Most notably, we identified two fraction 3 proteins (ZP_02735673 and ZP_02736511) that model well (>95% confidence, Phyre2) against clathrin adaptor core proteins, exhibiting an alpha-solenoid architecture (S Text, S18 Fig, S3 Table).

The presence of protein folds characteristic of the eukaryote nuclear pore complex [49] is intriguing in light of our deduced pore model, since in the eukaryote nuclear pore complex, beta-propeller- and alpha solenoid (stacked alpha-helices)-containing proteins act as scaffold proteins [50,51]. Beta-propeller proteins form vertices in a lattice-like model of the NPC and have special sequence-independent protein—protein interaction functions [52] while stacked alpha-helices of other scaffold nucleoporins are central to the lattice model interactions, forming edges of the NPC scaffold lattice [50] and are also structurally related to soluble proteins significant to nucleocytoplasmic transport through nuclear pores [53]. While it is remarkable that structural prediction yields folds known from the eukaryote nuclear pore complex, neither β-propeller folds nor alpha-solenoids are unique to the eukaryote nuclear pore

complex or endomembrane system, and examples of both folds are known from both Bacteria and Archaea [54]. For all *G. obscuriglobus* proteins carrying folds also found in eukaryote nuclear pore proteins, we find no evidence of substantive sequence similarity with eukaryote counterparts. Evidence of such similarity might be expected if recent horizontal gene transfer from eukaryotes was their origin. We therefore conclude these genes are not the result of recent transfer from eukaryotes. Our data instead indicate these are genes of bacterial origin.

Immuno-gold labelling confirms association of the pores with internal membranes

One of the proteins identified in fraction 3 exhibiting a β-propeller fold (ZP_02736670), was selected for antibody generation with the aim of using the antibody to immunolocalize the protein. This antibody (ab6670) showed high specificity as established by Western Blot, and reacted specifically with fraction 3, which comprises pore-containing membranes (S9B Fig). It was used for immunolocalization experiments to assess localization of the pore-containing membranes within the cells. On whole sectioned cells, gold particles were observed exclusively at membranes within the cell cytoplasm and internal to both cytoplasmic membrane and paryphoplasm (ribosome-less cytoplasm). The antibody labelled membranes comprising the nuclear body envelope and membranes associated with riboplasm (ribosome-containing cytoplasm) (Fig 8 and S19 Fig). Consistent with this result, the antibody also recognised pores in the purified membranes from fraction 3 (Fig 9 and S19 Fig) Membranes from other fractions, without such pores, were not labelled when tested. Such labelling is consistent with the presence of the ZP_02736670 β-propeller fold protein within pore complexes found via proteomics exclusively in fraction 3 (the origin of the protein data from which the peptide antibody was prepared).

Evolutionary implications of the findings

Our results imply two alternative possibilities, both equally profound. One is that the architectural similarity of pores in *Gemmata* bacteria and eukaryotes may be the result of a shared deep evolutionary origin, so that they are homologous and related by vertical inheritance. This would be consistent with past analyses that have placed *Planctomycetes* among the deepest branching phyla within domain Bacteria [55], and more recent analyses support the view that other members of the PVC (Planctomycetes, Verrucomicrobia and Chlamydiae) superphylum also carry compartmentalized cell plans [56,57]. In our view, the lack of detectable sequence similarity precludes homology resultant from recent horizontal gene transfer from a eukaryotic source. More generally, the lack of an unambiguous signal for common descent, together with the notable differences in size, weakens the case for common descent of the *G. obscuriglobus* pores and eukaryotic nuclear pores. An alternative interpretation to deep shared evolutionary origin of pores and pore complex proteins is that the similarities we observe between the *G. obscuriglobus* pore complex and the eukaryote nuclear pore complex are the result of convergence spanning all the way from gross architecture down to the recruitment of individual protein components with shared domain architecture**.** Under this interpretation the *G. obscuriglobus* pore complex is an analog of the eukaryote nuclear pore complex rather than a homologous structure derived by descent from a common ancestor. In other words, similar solutions to macromolecular transport across internal membranes and between cell compartments may have evolved independently, once in eukaryotes and once in planctomycete bacteria, via the parallel co-option of protein folds present in both bacteria and eukaryotes. The structures we report support the view that these solutions are architecturally similar, and in this regard, further characterization of the *G. obscuriglobus* pore will be critical for understanding the evolution of internal cell structure in this bacterium. The pore structures we have found are not straightforward to explain within the context of recent exciting results regarding Gram-negative bacteria cell wall components in planctomycetes, but there is no evidence from our data

consistent with a concept that pore-containing membranes are outer membrane or other cell wall components, or that the pores contain any wall-specific proteins. Such pore structures are however consistent with the known unique internal cell structure features of planctomycetes, and consistent with distinctive planctomycete cell organization and functional properties. Significantly, when considered alongside the recent discoveries of endocytosis [24] and compartmentalized transcription and translation [13], the parallels between the cell biology of *Gemmata obscuriglobus* and that of eukaryotes are nothing short of remarkable, whether due to hidden homology or an analogous reuse of a similar set of protein folds. Detailed analysis of the composition of the internal bacterial pores will no doubt enable the evolutionary origin of these structures to be definitively established.

ACKNOWLEDGEMENTS

The authors acknowledge the facilities, and the scientific and technical assistance, of the Australian Microscopy & Microanalysis Research Facility (AMMRF) at the Centre for Microscopy and Microanalysis, The University of Queensland. We thank Dr. Vitaliya Sagulenko for help with Western-blot analyses. We thank Tony Romeo of Electron Microscope Unit (EMU), University of Sydney, for help with preparing some of the freeze-fracture replicas. JAF acknowledges assistance in developing some preliminary density gradient protocols and donation of gp4978 antibody from Jody Franke now of Creighton University, Omaha, Nebraska USA.

MATERIALS AND METHODS

Cell lysis and electron microscopy of nuclear body envelopes

In the case of mechanical lysis experiments, *Gemmata obscuriglobus* was grown on M1 agar for 6 days at 28 ˚C, and growth harvested by washing into filtered Milli-Q water. For mechanical lysis via grinding with alumina, suspensions were mixed with alumina powder (Sigma Alumina Type A-5) in an eppendorf tube and a plastic pestle manually rotated within the tube for ca. 60 sec to lyse cells. The supernatant from this homogenate after allowing particles to settle was either harvested directly for negative staining or this supernatant purified by centrifugation using a microfuge at 20160 g. Negative staining was performed by mixing a homogenate drop on a pioliform–coated specimen support grid with 1% uranyl acetate containing 0.4% sucrose. In the case of preparations that were sonicated, *G. obscuriglobus* culture was grown on M1 medium for 8 days at 28˚C. Cells were suspended in 1 ml of sterile Milli-Q water, and sonicated in a Branson Sonifier 250 at amplitude output level 1 for ten thirty second intervals, separated by 30 sec rests. The resultant suspension was pelleted in a benchtop centrifuge for five mins at 20160 g, and the pellet was resuspended in 50μL sterile Milli-Q H_2O. Cells from these suspensions were negatively stained using 2% ammonium molybdate pH6.5 (mixture of 5μl each of suspension and stain was prepared on a carbon- and pioloform-coated specimen support grid followed by removal of excess fluid with filter paper and air-drying).

Cryo-electron microscopy

4 μL aliquots of density-gradient-purified membrane fractions from sonication-lysed *G. obscuriglobus* cells were deposited onto holey carbon films on hexagonal 200 mesh copper grids (‘C-flat’, Protochips, NC) in the humidified chamber of a commercial thin-film vitrification apparatus (CP3, Gatan,

Pleasanton, CA). An additional 4 µL of colloidal gold (nominal size 10-12 nm) was added to the membrane fraction and the grid was blotted from both sides for 3.5 seconds before automatic propulsion into liquefied ethane. Grids were transferred under liquid nitrogen to a cryo-sample holder (Model 914, Gatan) for transfer to the microscope and observation at a stable temperature of approximately -175 Celsius. A JEM-1400 transmission electron microscope (JEOL, Japan) fitted with a high-contrast polepiece and LaB6 cathode, and operating at an accelerating voltage of 120 kV was used for data acquisition. Micrographs were recorded at a nominal defocus value of -10 µm and an electron dose of approximately 4 000 electrons/nm2/micrograph. Detector noise was reduced by means of a median filter with a radius of 1 pixel.

Electron tomography from thick sections

For electron tomography of isolated membrane samples, 300nm thick sections were cut using Leica EM UC6 ultramicrotome (Leica Microsystems, Austria). The membranes were cryofixed as described for thin-sectioning. Dual-axis tilt-series data was collected at 39000x magnification on an FEI Tecnai F30 (FEG)TEM (FEI Company, the Netherlands) operating at 300kV, over a tilt range of +/-66º at 1º increments for the a-axis and 2º increments for the b-axis, using SerialEM software (The Boulder Lab for 3D Electron Microscopy, USA).

Membrane fractionations

Cells were collected from two to three M1-agar plates, washed once, and resuspended in 500 µL of bt-DMSO buffer [10mMbis-Tris (pH 7.5), 0.1mMMgCl2, and 20% DMSO] supplemented with 10 µL of

protease inhibitor mix (Protease Inhibitor Mixture Set 3; Merck), 10 μg of DNase, and 10 μg of RNase. Cells were then sonicated using a Branson Sonifier 250, and unbroken cells were spun down in a microfuge at 5,000 × g for 10 min. Supernatant was centrifuged at 100,000 × g in a Beckman Coulter tabletop ultracentrifuge (Optima TLX). The pellet was resuspended in 500 μL of bt-DMSO buffer, loaded onto a five-step sucrose gradient, and centrifuged in an SW60 rotor on a Beckman Coulter L8-60M ultracentrifuge at 215,000 × g for 4 h. Visible pink or white membrane fractions were collected via puncturing the side of the tube and to remove sucrose, centrifuged at 100,000 × g in a Beckman Coulter tabletop ultracentrifuge (Optima TLX), and the pellets were resuspended in ≈500 μL of bt-DMSO buffer. The material was loaded onto a 20–60% or 30-70% sucrose/bt-DMSO continuous gradient. After centrifugation using SW60 rotor, for 16 h at 215,000 × g, ≈400 μL fractions were collected from a puncture at the bottom of the tube , To concentrate and remove sucrose, fractions (in all cases containing a visible band) were diluted with bt-DMSO buffer and centrifuged at 100,000 × g in a Beckman Coulter tabletop ultracentrifuge (Optima TLX). The pellets were resuspended in 5mM Tris, pH7.5 and immediately used for electron microscopy experiments or proteomics. Each fractionation was performed twice using separate culture batches of cells, and 2-3 technical replicates from each fractionation were used in proteomics experiments.

Analysis of structural symmetry

To assess symmetry of structure within pore-like structures within a pore-containing membrane fraction purified from membranes released from lysed *G. obscuriglobus* cells, Markham rotation [58](Markham et al., 1963) was performed according to the modifications suggested by Friedman [59](Friedman, 1970)

by assuming any possible symmetry up to and including 8-fold symmetry. Pores in isolated pore-containing membranes from fraction 3 from the membrane fractionation described above were imaged *en face* by cryo- electron microscopy and were extracted from images and rotated in silico by set increments corresponding to the assumed symmetry. For example, reinforcement of possible 8-fold symmetry was assessed by superimposition of the corresponding 45.0 degree rotations, 7-fold by 51.4 degrees and so on. The procedure was conducted on raw data only rather than the bandpass-filtered structures. ImageJ was used for all feature extraction, rotation and summation.

Preparation of material used for Markham rotation: Isolated membranes from fraction 3 of the density gradient fractionation of membranes from lysed *Gemmata* cells were processed for cryo-electron microscopy and symmetry analysis via Markham rotation. Briefly, isolated membrane sheets were vitrified by rapid immersion in liquid ethane prior to mounting the samples in a cryo- sample holder (Model 914, Gatan, Pleasanton, CA) and imaging the frozen-hydrated specimens in a JEM-1400 transmission electron microscope (JEOL Ltd, Japan) equipped with a charge-coupled device detector (Gatan) and low-dose exposure conditions (< 4,000 electrons/nm2) to avoid radiation damage. The accelerating voltage was 120 kV and the nominal magnification of 15,000 corresponded to a pixel size of 0.69 nm at the detector. A minority of pores appeared to be markedly ellipsoid in projection. This was found to be the result of a tilted or folded membrane, which was taken into account by manual tilting of the specimen. Note that this differs from slight deviations in circularity that probably represent the respective functional states of transport-competent pores[60].This tilt was always less than 10 degrees, indicating that the untilted membrane sheets were mounted approximately orthogonal to the beam.

Freeze-fracture electron microscopy of whole cells

Cells from a 14-day culture of *G. obscuriglobus* ACM 2246 were grown on soil extract agar medium. Cells were harvested directly without chemical fixation into 20% (v/v) aqueous glycerol as cryoprotectant prior to freezing in liquid Freon 22. Fracturing was performed using a Balzers BAF 301 apparatus fitted with a complementary fracturing device, at -115 ˚C and 10^{-7} torr (1 torr = 133 Pa). Replicas were produced using platinum/carbon and stabilized with a layer of carbon.

Cryosubstitution and thin-sectioning

For cells cryofixed by plunging into liquid propane, cells of *G. obscuriglobus* were cryofixed using a Reichert-Jung KF80 cryofixation system. Cryosubstitution was performed with 2% osmium tetroxide in molecular-sieve-dried acetone at -79 ˚C (dry ice/acetone bath) for 50 hr. The temperature was increased to -20 ˚C over 14 hr. Specimens were brought to room temperature and then washed in acetone. Specimens were embedded in Epon resin, then sectioned and stained on pioloform-covered specimen support grids with aqueous uranyl acetate and lead citrate. For cells cryofixed by high-pressure freezing, cells of *G. obscuriglobus* were cryofixed by first mixing with 2% agarose and placing the suspension between hexadecene-soaked planchettes, then frozen using a BAL-TEC HPM 010 High-Pressure Freezer. Frozen cells at -160 ˚C were warmed to -85 in 1.9 hrs at 4 C/hr , stored at -85 ˚C for 52 hrs and then raised to -20 ˚C over 11 hrs in a Leica EM AFS cryosubstitution apparatus. Cells were then embedded in Epon resin and sectioned and stained as above.

Electron microscopy

For experiments other than those involving tomography, specimens were examined using a JEOL 1010 transmission electron microscope operated at 80 kV.

Proteomics

Before proteomics, pellets of membrane fractions in Tris buffer were dissolved using Laemmli buffer, protein concentration was measured using BCA Protein assay (ThermoFisher Scientific) and 20 µg of each suspension was loaded onto PAA gels. **T**he proteins, separated by 1-D SDS-PAGE electrophoresis, were cut out from the 8-12% PAA gels for mass spectroscopy. Gel slices were destained with 50% ACN in 50 mM ammonium bicarbonate (ABC) followed by dehydration in 100% acetonitrile (ACN). Trypsin (80 ng) in 50 mM ABC was added and gel slices rehydrated at 4 °C for 10 min, followed by incubation at 37 °C overnight. Peptides were extracted three times with 50 ul of 50% ACN / 0.1% formic acid, followed by clean up with a ZipTip (Millipore). Peptides were separated using reversed-phase chromatography on a Shimadzu Prominence nanoLC system. Using a flow rate of 30 µl/min, samples were desalted on an Agilent C18 trap (0.3 x 5 mm, 5 µm) for 3 min, followed by separation on a Vydac Everest C18 (300 A, 5 µm, 150 mm x 150 µm) column at a flow rate of 1 µl/min. A gradient of 10-60% buffer B over 30 min where buffer A = 1 % ACN / 0.1% FA and buffer B = 80% ACN / 0.1% FA was used to separate peptides. Eluted peptides were directly analysed on a TripleTof 5600 instrument (ABSciex) using a Nanospray III interface. Gas and voltage settings were adjusted as required. MS TOF scan across m/z 350-1800 was performed for 0.5 sec followed by information dependent acquisition of the top 20 peptides across m/z 40-1800 (0.05 sec per spectra). Data were converted to mgf format and searched in MASCOT accessed via the Australian Proteomics Computational Facility and searched against the LudwigNR database, limited to 'other bacteria', using trypsin as enzyme, 2 mis-cleavages, MS tolerance of 0.5 Da and MS/MS tolerance of 0.2 Da. Oxidation (met, variable) and carbamidomethylation (cys, fixed) modifications were also included.

The MS analyses performed were strictly qualitative, not quantitative, and therefore no normalisation of protein amount prior to trypsin digested was performed. A nominal amount (eg 80 ng) trypsin is added

per gel slice, as it is not feasible to determine the amount of protein in each gel band processed for MS. This is typical practice in the proteomics field. Samples were ziptipped after digestion, prior to MS, in part to desalt/concentrate the samples, but also to ensure the LCMS system was not overloaded (ziptips have a limited loading capacity (5ug)).

Antibodies

A polyclonal antibody (designated as Anti-Protein 6670) was raised by GenScript Inc. (Piscataway, NJ, USA). The antigen used for immunization was VPVTDDTRKEPTETC, derived from the translated protein ZP_02736670. Immunogen was a Peptide-KLH conjugate, and host strain was New Zealand rabbit. The antibody was affinity purified and stored in Phosphate Buffered Saline (PBS, pH 7.4) with 0.02% Sodium Azide at -20^{0}C. Membrane preparations of *G. obscuriglobus* were resolved by SDS/PAGE (10%) and the specificity of the antibody was tested via western blotting at 1:2000 dilution (S10 Fig). HRP-conjugated goat anti-rabbit antibody (Cell Signalling Technology, Australia, 1:5000 dilution) was used as secondary antibody. Detection was done by using ECL Western Blotting System (GE Healthcare Life Science). The 'anti-MC protein' antibody against the Gemmata obscuriglobus protein gp4978, a clathrin heavy chain-like membrane coat (MC) protein already shown to be present on internal membrane vesicles of *Gemmata obscuriglobus* associated with endocytosis-like protein uptake in this species[24]. This antibody to gp4978 was a rabbit polyclonal antibody raised against recombinantly expressed and purified gp4978 protein identified as a eukaryotic MC coatomer protein (National Center for Biotechnology Information reference sequence: ZP_02732338.1; see [37]).

Immuno-gold labelling

Ultrathin sections of high-pressure frozen and cryosubstituted intact *G. obscuriglobus* cells or density gradient-purified fraction 3 membranes on formvar-carbon-coated copper grids were floated onto drops of Block solution containing 0.2% (wt/vol) fish skin gelatin, 0.2% (wt/vol) BSA, 200 mM glycine, and 1× PBS on a sheet of Parafilm, and treated for 1 min at 150 W in a Biowave microwave oven. The grids were then transferred onto 8 μL of primary antibody, diluted 1:25 in blocking buffer, and treated in the microwave at 150 W, for 2 min with microwave on, 2 min off, and 2 min on. The grids were then washed on drops of Block solution three times and treated in the microwave at 150 W each time for 1 min before being placed on 8 μL of goat anti-mouse IgG Fc (γ)-specific antibody conjugated with 10 nm gold (British Biocell International, catalog no. EM GAM10) diluted 1:50 in Block solution and treated in the microwave at 150 W, for 2 min with microwave on, 2 min off, and 2 min on. Then grids were washed three times in 1× PBS, each time being treated for 1 min each in the microwave at 150 W, and four times in water for 1 min each in the microwave at 150 W, and examined via transmission electron microscopy.

REFERENCES

1. Wilson KL, Dawson SC Evolution: functional evolution of nuclear structure. J Cell Biol 2011;195: 171-181.
2. Hoelz A, Debler EW, Blobel G () The structure of the nuclear pore complex. Annu Rev Biochem 2011;80: 613-643.
3. DeGrasse JA, DuBois KN, Devos D, Siegel TN, Sali A, et al. Evidence for a shared nuclear pore complex architecture that is conserved from the last common eukaryotic ancestor. Mol Cell Proteomics 2009;8: 2119-2130.
4. Field MC, Koreny L, Rout MP Enriching the pore: splendid complexity from humble origins. Traffic 2014; 15: 141-156.
5. Neumann N, Lundin D, Poole AM Comparative genomic evidence for a complete nuclear pore complex in the last eukaryotic common ancestor. PLoS One 2010;5: e13241.

6. Hirst J, Schlacht A, Norcott JP, Traynor D, Bloomfield G, et al. Characterization of TSET, an ancient and widespread membrane trafficking complex. Elife 2014:3: e02866.
7. Sagulenko E, Morgan GP, Webb RI, Yee B, Lee KC, et al. Structural studies of planctomycete *Gemmata obscuriglobus* support cell compartmentalisation in a bacterium. PLoS One 2014; 9: e91344.
8. Lindsay MR, Webb RI, Strous M, Jetten MS, Butler MK, et al. Cell compartmentalisation in planctomycetes: novel types of structural organisation for the bacterial cell. Arch Microbiol 2001;175: 413-429.
9. Fuerst JA, Webb RI Membrane-bounded nucleoid in the eubacterium *Gemmata obscuriglobus*. Proc Natl Acad Sci U S A 1991; 88: 8184-8188.
10. Lee KC, Webb RI, Fuerst JA The cell cycle of the planctomycete *Gemmata obscuriglobus* with respect to cell compartmentalization. BMC Cell Biol 2009;10: 4.
11. Santarella-Mellwig R, Pruggnaller S, Roos N, Mattaj IW, Devos DP Three-dimensional reconstruction of bacteria with a complex endomembrane system. PLoS Biol 2013; 11: e1001565.
12. Acehan D, Santarella-Mellwig R, Devos DP A bacterial tubulovesicular network. J Cell Sci 2014;127: 277-280.
13. Gottshall EY, Seebart C, Gatlin JC, Ward NL Spatially segregated transcription and translation in cells of the endomembrane-containing bacterium *Gemmata obscuriglobus*. Proc Natl Acad Sci U S A 2014;111: 11067-11072.
14. Forterre P, Gribaldo S Bacteria with a eukaryotic touch: a glimpse of ancient evolution? Proc Natl Acad Sci U S A 2010;107: 12739-12740.
15. McInerney JO, Martin WF, Koonin EV, Allen JF, Galperin MY, et al. Planctomycetes and eukaryotes: a case of analogy not homology. Bioessays 2011; 33: 810-817.
16. Fuerst JA, Sagulenko E Keys to eukaryality: planctomycetes and ancestral evolution of cellular complexity. Front Microbiol 2012; 3: 167.
17. Jeske O, Schuler M, Schumann P, Schneider A, Boedeker C, et al. Planctomycetes do possess a peptidoglycan cell wall. Nat Commun 2015;6: 7116.
18. Mahat R, Seebart C, Basile F, Ward NL Global and targeted lipid analysis of *Gemmata obscuriglobus* reveals the presence of lipopolysaccharide, a signature of the classical Gram-negative outer membrane. J Bacteriol. 2015; 198(2):221-36. doi: 10.1128/JB.00517-15.
19. van Teeseling MC, Mesman RJ, Kuru E, Espaillat A, Cava F, et al. Anammox Planctomycetes have a peptidoglycan cell wall. Nat Commun 2015;6: 6878.
20. Devos DP Re-interpretation of the evidence for the PVC cell plan supports a Gram-negative origin. Antonie Van Leeuwenhoek 2014;105: 271-274.
21. Devos DP PVC bacteria: variation of, but not exception to, the Gram-negative cell plan. Trends Microbiol 2014; 22: 14-20.
22. Stoffler D, Fahrenkrog B, Aebi U The nuclear pore complex: from molecular architecture to functional dynamics. Curr Opin Cell Biol 1999;11: 391-401.
23. Wang J, Jenkins C, Webb RI, Fuerst JA Isolation of *Gemmata*-like and *Isosphaera*-like planctomycete bacteria from soil and freshwater. Appl Environ Microbiol 2002; 68: 417-422.
24. Lonhienne TG, Sagulenko E, Webb RI, Lee KC, Franke J, et al. Endocytosis-like protein uptake in the bacterium *Gemmata obscuriglobus*. Proc Natl Acad Sci U S A 2010;107: 12883-12888.
25. Gall JG Octagonal nuclear pores. J Cell Biol 1967;32: 391-399.
26. Beck M, Forster F, Ecke M, Plitzko JM, Melchior F, et al. Nuclear pore complex structure and dynamics revealed by cryoelectron tomography. Science 2004;306: 1387-1390.

27. Unwin PN, Milligan RA A large particle associated with the perimeter of the nuclear pore complex. J Cell Biol 1982;93: 63-75.
28. Rout MP, Blobel G Isolation of the yeast nuclear pore complex. J Cell Biol 1993;123: 771-783.
29. Frenkiel-Krispin D, Maco B, Aebi U, Medalia O Structural analysis of a metazoan nuclear pore complex reveals a fused concentric ring architecture. J Mol Biol 2010;395: 578-586.
30. Strambio-de-Castillia C, Blobel G, Rout MP Isolation and characterization of nuclear envelopes from the yeast Saccharomyces. J Cell Biol 1995;131: 19-31.
31. Franke WW, Deumling B, Baerbelermen, Jarasch ED, Kleinig H Nuclear membranes from mammalian liver. I. Isolation procedure and general characterization. J Cell Biol 1970;46: 379-395.
32. Franke WW Isolated nuclear membranes. J Cell Biol 1966;31: 619-623.
33. Yang Q, Rout MP, Akey CW Three-dimensional architecture of the isolated yeast nuclear pore complex: functional and evolutionary implications. Mol Cell 1998;1: 223-234.
34. Fahrenkrog B, Hurt EC, Aebi U, Pante N Molecular architecture of the yeast nuclear pore complex: localization of Nsp1p subcomplexes. J Cell Biol 1998;143: 577-588.
35. Hinshaw JE, Milligan RA Nuclear pore complexes exceeding eightfold rotational symmetry. J Struct Biol 2003;141: 259-268.
36. Hinshaw JE, Carragher BO, Milligan RA Architecture and design of the nuclear pore complex. Cell 1992;69: 1133-1141.
37. Santarella-Mellwig R, Franke J, Jaedicke A, Gorjanacz M, Bauer U, et al. The compartmentalized bacteria of the planctomycetes-verrucomicrobia-chlamydiae superphylum have membrane coat-like proteins. PLoS Biol 2010;8: e1000281.
38. Fuerst JA, Sagulenko E Protein uptake by bacteria: An endocytosis-like process in the planctomycete Gemmata obscuriglobus. Commun Integr Biol 2010;3: 572-575.
39. Schmidt JM Isolation and ultrastructure of freshwater strains of Planctomyces. Current Microbiology 1978;1: 65-70.
40. Liesack W, Konig H, Schlesner H, Hirsch P Chemical composition of the peptidoglycan-free cell envelopes of budding bacteria of the *Pirella /Planctomyces* group. Archives of Microbiology 1986;145: 361-366.
41. Stackebrandt E, Wehmeyer U, Liesack W 16S ribosomal RNA- and cell wall analysis of *Gemmata obscuriglobus*, a new member of the Order Planctomycetales. Fems Microbiology Letters 1986;37: 289-292.
42. Ward N, Staley JT, Fuerst JA, Giovannoni S, Schlesner H, et al. The Order Planctomycetales, Including the Genera Planctomyces, Pirellula, Gemmata and Isosphaera and the Candidatus Genera Brocadia, Kuenenia and Scalindua In: Dworkin Mea, editor. The Prokaryotes: An Evolving Electronic Resource for the Microbiological Community. 3rd, release 3.18 , December 21,2004, ed. New York: Springer-Verlag, .
43. Voigt B, Hieu CX, Hempel K, Becher D, Schluter R, et al. Cell surface proteome of the marine planctomycete Rhodopirellula baltica. Proteomics 2012;12: 1781-1791.
44. Hieu CX, Voigt B, Albrecht D, Becher D, Lombardot T, et al. Detailed proteome analysis of growing cells of the planctomycete Rhodopirellula baltica SH1T. Proteomics 2008;8: 1608-1623.
45. Konig E, Schlesner H, Hirsch P Cell wall studies on budding bacteria of the *Planctomyces*/*Pasteuria* group and on a *Prosthecomicrobium* sp. Archives of Microbiology 1984;138: 200-205.
46. Speth DR, van Teeseling MC, Jetten MS Genomic analysis indicates the presence
of an asymmetric bilayer outer membrane in planctomycetes and verrucomicrobia.
. Frontiers in Microbiology 2012;3: 304.

47. Kelley LA, Sternberg MJ Protein structure prediction on the Web: a case study using the Phyre server. Nat Protoc 2009;4: 363-371.
48. Devos D, Dokudovskaya S, Alber F, Williams R, Chait BT, et al. Components of coated vesicles and nuclear pore complexes share a common molecular architecture. PLoS Biol 2004;2: e380.
49. Devos D, Dokudovskaya S, Williams R, Alber F, Eswar N, et al. Simple fold composition and modular architecture of the nuclear pore complex. Proc Natl Acad Sci U S A 2006;103: 2172-2177.
50. Brohawn SG, Schwartz TU Molecular architecture of the Nup84-Nup145C-Sec13 edge element in the nuclear pore complex lattice. Nat Struct Mol Biol 2009;16: 1173-1177.
51. Brohawn SG, Schwartz TU A lattice model of the nuclear pore complex. Commun Integr Biol 2009;2: 205-207.
52. Leksa NC, Schwartz TU Membrane-coating lattice scaffolds in the nuclear pore and vesicle coats: commonalities, differences, challenges. Nucleus 2010;1: 314-318.
53. Andersen KR, Onischenko E, Tang JH, Kumar P, Chen JZ, et al. Scaffold nucleoporins Nup188 and Nup192 share structural and functional properties with nuclear transport receptors. Elife 2013;2: e00745.
54. Andrade MA, Petosa C, O'Donoghue SI, Muller CW, Bork P Comparison of ARM and HEAT protein repeats. J Mol Biol 2001;309: 1-18.
55. Brochier C, Philippe H Phylogeny: a non-hyperthermophilic ancestor for bacteria. Nature 2002;417: 244.
56. Fuerst JA The PVC superphylum: exceptions to the bacterial definition? Antonie Van Leeuwenhoek 2013;104: 451-466.
57. Lee KC, Webb RI, Janssen PH, Sangwan P, Romeo T, et al. Phylum Verrucomicrobia representatives share a compartmentalized cell plan with members of bacterial phylum Planctomycetes. BMC Microbiol 2009;9: 5.
58. Markham R, Frey S, Hills GJ Methods for enhancement of image detail and accentuation of structure in electron microscopy. Virology 1963;22: 88-102.
59. Friedman MH A reevaluation of the Markham rotation technique using model systems. J Ultrastruct Res 1970;32: 226-236.
60. Beck M, Lucic V, Forster F, Baumeister W, Medalia O Snapshots of nuclear pore complexes in action captured by cryo-electron tomography. Nature 2007;449: 611-615.

FIGURE LEGENDS

Fig 1. Pores are inserted into the internal membranes of *Gemmata obscuriglobus* cells

(***A***) Transmission electron micrograph of a thin-section of a cryosubstituted cell of *G. obscuriglobus*, showing a portion of the nuclear body envelope, apparently consisting of two closely apposed membranes enclosing the fibrillar nucleoid DNA (N) (for evidence of DNA fibrillar nature in *G. obscuriglobus* see [9]. The membranes (arrows) are interrupted by a disc-like structure (indicated by arrowhead within the boxed region) consistent with a pore complex inserted between the membranes on either side. Bar, 50 nm. **(*B*)** Enlargement of the sectioned cell of *G. obscuriglobus* seen in Fig 1A, showing a disc structure (arrowhead) seen *en face*, situated between the folded double membranes of the nuclear body envelope on either side (arrows). Bar, 50 nm. **(*C*)** Transmission electron micrograph of cell lysed by grinding in liquid N_2, followed by negative staining of thawed cells with uranyl acetate. An internal membrane fragment (IM) possibly representing the nuclear body envelope or other internal compartment membranes appears to have been released from a lysed cell, and the mostly intact cell wall (CW) can also be seen. The membrane displays numerous evenly distributed pore structures on its surface, enlarged views of which can be seen in the inset. Bar, 500 nm. Inset shows enlargement of pore structures, which display a dense core surrounded by a light ring further surrounded by a dense ring. Bar, 50 nm. (***D***) Transmission electron micrograph of negatively stained preparation of a completely released internal compartment from cells lysed as in C. Pore structures are widely distributed over the membrane surface including within the boxed region. The 'canoe' shape is typical for pore-containing membranes. Bar, 500 nm. **(*E*)** An enlarged view of the boxed region in Fig 1D showing the large pore structures (arrows), each displaying dark pore centre regions, and lighter inner and outer ring structures, distributed densely on the membrane surface. Bar, 100 nm.

Fig 2. *Gemmata obscuriglobus* internal membrane pores as seen in freeze-fractured cells

(*A*) Transmission electron micrograph of a platinum/carbon (Pt/C)-shadowed replica of a whole cell of *G. obscuriglobus* which has been prepared via the freeze-fracture technique. Bar, 100 nm. Inside the cell, a large spherical internal organelle consistent with the nuclear body organelle surrounding the nucleoid has been fractured (split) along and through the surface membranes of its envelope. Pores with a central core and at least one surrounding ring are visible on one region of one of the membranes of this organelle surface. Insets represent successive enlarged views of the boxed region in the main image displaying the pores at higher magnification. Bars, 100nm. At the highest enlargement the substructure of each of several pores can be resolved including central core and surrounding inner dark and outer light rings (right inset). (***B***) This micrograph of the whole cell reveals an apparently cross-fractured major internal organelle compartment and a membrane surface (boxed) representing a fracture through the membrane surrounding the organelle. Bar, 200 nm. **(*C*)** An enlarged view of the boxed region of the freeze-fractured cell seen in Fig 2B showing a region of a membrane surface where roughly circular pore structures (arrowheads) are visible, in some cases with two light rings surrounding a dark centre,. Bar, 50 nm.

Fig 3. Pores in the membranes of *Gemmata obscuriglobus* released via sonication

(*A*) Transmission electron micrograph of a membrane fragment released from a lysed cell via sonication and negatively stained with ammonium molybdate. Large pores (arrows) with relatively electron-dense pore centers surrounded by a thin lighter inner ring and a thicker outer ring are seen. Smaller pore structures (arrowheads) are also visible and may represent either another class of pores or a result of a reverse view of the same large pores resulting from overlapping folds in the membrane (evidence for

such structures is not derived from other microscopy methods). Bar, 100 nm. **Inset:** enlargement of boxed large pore in main Fig where a pore centre (PC), an inner ring (IR) and an outer ring (OR) can be distinguished. Bar, 50 nm. (***B***) TEM of a pore seen in negatively stained membrane fraction isolated from sonication-lysed cells, showing pore complex structure including outer ring (OR), inner ring (IR), spokes connecting inner and outer rings (S) and central plug (CP). Bar, 30 nm. **(*C*)** Enlarged view of the inner ring (IR) and central plug (CP) of the boxed pores in Fig 3A, the octagonal shape of the rings (especially visible if the outer edge of the outer ring is traced) is consistent with an eight-fold symmetry. Bar, 15 nm.

Fig 4. Architecture of the *Gemmata obscuriglobus* pore

(***A***) Structure of pores embedded into membranes from fraction 3 purified via density gradient centrifugations and visualized via TEM of thin-sections. The spiral seen consists of a membrane (arrows) in which pores are embedded interrupting dense-light-dense layers of the trilaminar membrane. Basket structures (arrowheads) of each pore complex project only from one side of the membrane. The inner and outer dense leaflets of the membrane are seen to be connected forming a continuous folded membrane (on each side of the pore) of extreme membrane curvature (arrows). Bar, 100 nm.(See also S1 Video and S2 Video for 3D reconstructions of the membranes and S3 Video for 3D reconstruction of the pore). **(*B*)** Transmission electron micrographs from a tilt-series of one pore. In panel 1 intact membrane without pore is seen, while in panels 2-6 passing through progressive slices generated via the tilt-series, the pore appears, interrupting the trilaminar membrane on either side, and most clearly indicated by a basket structure projecting below the plane of the membrane (arrowhead). In panels 3 and 4, the central plug region of the pore can be seen (arrows). In panel 6 the trilaminar membrane is again continuous, but some parts of the basket structure are still visible (arrow). This series is consistent with the interruption of membrane by embedded pore structures, the basket component of which projects

beyond the membrane plane. Bar, 20 nm. (***C***) Micrograph from cryo-EMof a frozen-hydrated preparation of the isolated and sucrose-purified fraction 3 membranes. Two randomly selected pores (see Fig S10 supplement 1for cryo-EM of fraction 3 membrane sheets from which these pores were selected) clearly display inner ring (IR), outer ring (OR) and central plug (CP).This image has been processed via uniform application of a conservative bandpass filter (respective low- and high-frequency cut-offs of 40 and 3 pixels). Bar, 30 nm. (***D***) Modified Markham rotation analysis of one of the pores from Figure 4C showing reinforcement of 8-fold symmetry of pore structure. Bar, 10 nm.

Fig 5. 3-D reconstructions of the pore complex

(*A* and *B*) Views of the 3-D reconstructions based on one spiral membrane from fraction 3 membranes (see Fig 4A). Pore complexes (arrows) are visible as embedded structures in the surface of the envelope, shown as viewed from the inner side of the spiral in Fig **5A** and from the outer side in Fig **5B**. Fig **5C** shows the basket structure of one of these pores projecting from the inner side of the membrane spiral. Bars, 20 nm. **(*D* and *E*)** Reconstruction of architecture of a single pore seen from two different angles. In panel ***D***, a side view of the pore displays the basket structure with its distal ring (arrowhead) and a series of struts (arrow) connecting with the main pore rings. In panel ***E***, a top view shows the ring-like element (arrowhead) of the main part of the pore and a central plug structure is visible within the pore connected to the ring's inner rim via spokes.

Fig 6. Model of the pore complex of *Gemmata obscuriglobus*

The pore complex is composed of at least two concentric upper rings (blue), and a lower ring (light blue) connected by struts to a distal ring (green) to form a basket structure. The central plug (purple) rests within the inner ring and spans the length of the pore. The whole pore complex rests within membrane

(orange). The structure and dimensions are based on available data from all EM methods applied, from both whole cells and fraction 3 isolated membranes, and with minimal extrapolation, so that although the pore is probably not a hollow structure the space within the pore has not been filled in.

Fig 7. Protein composition of *Gemmata obscuriglobus* pore-containing membrane

(*A*) SDS-PAGE gel showing that *G.obscuriglobus* cells have three different types of membranes. Exclusively pore-containing membranes (fraction 3) display a characteristic protein profile distinct from that of membrane fractions which do not possess pore structures. **(*B*)** Venn diagram showing the number and distribution of proteins among the fractions and (in brackets) the number of proteins with the beta-propeller folds. The members of the beta-propeller cluster belong either exclusively to fraction 3 (4 proteins), or to fractions 3 and 2 (2 proteins), and to fractions 2, 3 and 6 (2 proteins). No beta-propeller containing proteins were found exclusively in fractions 2 or 6. **(*C*)** A beta-propeller family found in fraction 3 (pore-containing membranes), including some exclusive to fraction 3. Cluster analyses revealed a set of proteins with conserved C-terminal regions (Figs S13 - S16) that model beta-propeller folds with high (>95%) confidence. Models 3 (for protein ZP_02737072), 4 (ZP_02736670), 5 (ZP_02734776) and 6 (ZP_ZP_02734577) were deduced from proteins found exclusively in fraction 3 (pore-containing fraction); models 2 (for ZP_02737073) and 7 (for ZP_02733245) were deduced from proteins found in fractions 3 and 2 only; models 1 (for ZP_02737797) and 8 (for ZP_02731113) – for proteins found in fractions 3, 2, and 6 (Table S4).

Fig 8. The antibody 6670 recognises internal membranes in *G.obscuriglobus* cells

(*A*) TEM of a thin-sectioned cryosubstituted cell labelled with the antibody 6670. The majority of the gold particles (arrows) are seen to be bound to intracytoplasmic membrane (ICM, arrowheads). This

membrane separates the electron-dense ribosome-free paryphoplasm (P) from relatively electron-transparent riboplasm (R), as well as riboplasm vesicles from each other. A few particles label the border envelope between NB and riboplasm including double membrane regions. Bar, 500nm. **(*B*)** An enlarged view of the boxed region B in Fig 8A showing the nuclear body (NB) with nucleoid DNA. A few gold particles (arrows) are visible on the envelope membranes (arrowheads), separating NB and riboplasm. Bar, 200 nm. **(*C*)** An enlarged view of the boxed region C in Fig 8A showing an electron-transparent region continuous with riboplasm, surrounded by paryphoplasm which is separated from the riboplasm-continuous region by ICM. Gold particles (arrows) unambiguously label the ICM (arrowheads). Bar, 200 nm.

Fig 9. The antibody 6670 recognises pores in the isolated membranes

(***A***) Immuno-gold labelling of membrane sheets from membrane fraction 3 with the antibody 6670. In the majority of cases the gold particles indicating antibody can be seen as associated with the outer ring of the pores. Panels (***B, C, D***, and ***E*)** show enlarged areas of **(*A*)**, which are marked as boxes in (***A***). In all the cases the gold particles can be seen at the edge of the pores. In some cases more than one gold particle is associated with the pores (for example see box **(*D*)**, the bottom pore which is surrounded by three particles). For statistical analyses approximately the same areas were used for counting the particles: 397 particles were observed as associated with pores (distance from a pore does not exceed 20nm) and 45 particles were considered as not associated. Arrows indicate gold particles, black or white arrowheads (depending on background) – pores. Bars, **A** – 1 μm, **B** -200nm, **C**, **D**, and **E** – 100nm.

LEGENDS FOR SUPPLEMENTARY VIDEOS AND FIGURES

S1 Video: Electron tomography of the fraction 3 membranes containing pores

Curving sheets comprising isolated membranes from fraction 3 (isolated via density gradient fractionation from lysed *G. obscuriglobus* cells) can be seen in section. As tomogram slices of the spirally coiled sheets are passed through during the movie's course (in effect passing through successive and different planes of a thick section), they can be seen to be interrupted periodically by non-membranous pore structures some regions of which project from one side of each of the membranes which comprise a coil. If one membrane sheet is examined at different points of the movie, several pores can be identified as slices of the membrane sheet are passed through successively. Bar, 50 nm.

S2 Video: 3D reconstruction of an internal pore-containing membrane of *Gemmata obscuriglobus*

The electron tomography membrane reconstruction shown here is derived from a representative of the fraction 3 membranes in the thick section tilt-series.

S3 Video: 3D reconstruction of a pore embedded in the internal pore-containing membrane of *Gemmata obscuriglobus*. The reconstruction is derived from a pore in the membrane shown in S1 Video and S2 Video.

S1 Fig. Pores embedded in internal membranes of *Gemmata obscuriglobus*

(***A***)Transmission electron micrograph of a tomographic slice of high-pressure-frozen cryosubstituted thick-sectioned cell showing a pore (boxed region) embedded in internal membranes situated within the cytoplasm and bounding the nuclear body region containing the cell's nucleoid. Bar, 1 μm. **Inset:** (***B***)

Enlarged view of the pore outlined by the box in (A) – the circular pore (arrowheads) is seen tilted *en face* and displays a complex ring structure and a central plug. Bar, 100 nm.

S2 Fig. *Gemmata* CJuql4 internal membrane pores as seen in the freeze-fractured cells

(*A*) Transmission electron micrograph of a platinum/carbon (Pt/C)-shadowed replica of a whole cell of *Gemmata* CJuql4 (ACM5157) which has been prepared via the freeze-fracture technique. The fractured whole cell contains a large spherical organelle taking up most of the cell volume, the surface of which has been fractured along a membrane. Pores are visible on the fractured membrane surface of this major cell compartment, interpreted as the nuclear body (e.g. in the boxed region) Bar, 200 nm. (***B***) An enlarged view of the boxed region of the freeze-fractured cell seen in (*A*) showing a region of a membrane surface where pore structures are visible. Several pores display substructure consistent with complex structure including a dark central core and a lighter ring surrounding the core (arrows). Other circular structures in the same size range are also visible but do not present this complex core-ring structure as clearly, presumably reflecting angle at which Pt/C metal shadow has been deposited during formation of the replica after fracture of the frozen cell. Bar, 100 nm.

S3 Fig. Dimensions of the *Gemmata obscuriglobus* internal pores

The dimensions were calculated from transmission electron micrographs of the membrane fragments released from lysed cells via sonication and negatively stained with ammonium molybdate. The pores usually appear as circular structures with dense pore centers surrounded by a thin lighter inner ring and a thicker outer ring (see Fig 3 for example). The bars are generated automatically and calculated by microscope software.

S4 Fig. Pore-containing membrane of *Gemmata obscuriglobus* disintegrates and pores aggregate after detergent treatment

(*A*) Transmission electron micrograph of negatively stained gradient-fractionated pore-containing membranes purified from sonicated *G. obscuriglobus* acting as control for detergent treatments shown in (*B*) and (*C*). A "canoe" structure with pores (arrowheads) is visible. Bar, 500 nm. **(*B*)** Transmission electron micrograph of negatively stained gradient-fractionated pore-containing membranes purified from sonicated *G. obscuriglobus* after treatment with 1% Triton X-100 and 1% sodium deoxycholate detergent for 5 min. Pores (arrowheads) are visible within a partially degraded membrane background. Bar, 200 nm. **(*C*)** Transmission electron micrograph of negatively stained aggregated pore complexes seen after treatment of gradient-fractionated pore-containing membranes with 1% Triton X-100 and 1% sodium deoxycholate detergent for 30 min. Individual pores show a central dense core surrounded by a light ring and in some cases material projecting from the outer rim of the ring possibly representing spokes normally connecting inner to outer ring in intact pore complexes (arrowheads). Bar, 50 nm.

S5 Fig. Sucrose gradient fractionation of *Gemmata obscuriglobus* membranes

(*A*) Schematic diagram showing bands resulting from density gradient fractionation of membrane fractions released from cells of *G. obscuriglobus* lysed via sonication. On the left is the initial distribution of sucrose concentrations in the gradient before ultracentrifugation and the initial position of the total membrane mixture. On the right are the resulting bands that were visible after ultracentrifugation – fractions collected from the whole length of the gradient are indicated by numbers 1-8 and the resulting protein bands are indicated as *a-d*. **(*B*)** SDS-PAGE gel of continuous-gradient purified fractions corresponding to fractions 2-6 of membrane bands described in (**A**). Bands resulting from electrophoresis of the different membrane fractions show that purified fractions 3 and 4 (band *b*)

contain a distinctive pattern of a limited number of proteins relative to fractions 2 (band *a*), 5 (band *c*) and 6 (band *d*). Fractions 1, 7, and 8 did not contain any material and were excluded from further work. Purified fractions 3 and 4 were shown to contain only 'canoe' membranes with pore structures via TEM of negatively stained membranes (Fig S8). Protein fraction 5 after continuous gradient fractionation formed a "smear" band which was collected and examined under electron microscope. The collected fraction was found to contain a mixture of membranes morphologically similar to those from fractions 2, 3 (and 4), and 6, and was therefore excluded from proteome analysis. Fraction 4 after preliminary Mass-spec analysis revealed the same protein content as fraction 3, thus for the analysis of the whole protein content of the band *b* we used fraction 3 only.

S6 Fig. Transmission electron microscopy of the membranes enriched in fraction 2 and SDS PAGE of the proteins obtained from this fraction

(*A*) (1, **2** and **3) -** transmission electron micrographs of negatively stained membranes of fraction 2 (see Fig S6) containing membranes which do not display pore complexes. Bar A1, 1 μm, Bar A2, 500 nm, Bar A3, 200 nm. **(B)** SDS-PAGE of membrane fraction 2 proteins. All these individual bands were cut out for proteomic analysis (for results see S1 Table).

S7 Fig. TEM of the membranes enriched in fraction 3 and SDS PAGE of the proteins obtained from this fraction

(*A*) TEM of negatively stained membranes of fraction 3 (see Fig S6) containing membranes which display pore complexes. **1** and **2** show appearance of aggregates of membranes at relatively low magnification while **3** shows the characteristic 'canoe' shape of pore-containing membranes comprising this fraction. The enlarged view in A3 shows the typical appearance of the large pore ring structures

(arrows) on these 'canoe'-shaped membranes. Bar A1, 5 µm, Bar A2, 1 µm, Bar A3, 200 nm. **(*B*)** SDS-PAGE of membrane fraction 3 proteins. All the individual bands were cut out for proteomic analysis (for results see S1 Table). Fraction 4 contained the same 'canoe' shaped membranes and proteomics analysis revealed no difference between those fractions at protein level.

S8 Fig. Transmission electron microscopy of the membranes enriched in fraction 6 and SDS PAGE of the proteins obtained from this fraction

(*A*) Transmission electron microscopy of negatively stained membranes of fraction 6 (see Fig S6) containing membranes which do not display pore complexes. Bar A1, 10 µm, Bar A2, 500 nm, Bar A3, 200 nm. **(*B*)** SDS-PAGE of membrane fraction 6 proteins. All the individual bands were cut out for preparation for proteomic analysis (for results see S1 Table).

S9 Fig. The antibodies 6670 binds specifically to the beta-propeller-containing protein from fraction 3

(*A*) Amino acid sequence of the protein annotated as *Na-Ca exchanger/integrin-beta4* (NCBI Reference Sequence: ZP_02736670, later renumbered as the synonymous WP_010049031.1) was used for generation of an antibody (antibody 6670). The protein was identified by mass-spectrometry analyses as a unique protein for Fraction 3 and the full sequence was retrieved from the NCBI Database. **(*B*)** *G.obscuriglobus* fractions were used for testing the antibody specificity. The antibody does not react with proteins from fractions 2 and 6, and only one band was detected in fraction 3 at ca. 40- 45 kDa (arrowhead), which is consistent with the predicted MW for the *Na-Ca exchanger/integrin-beta4*. As a control for purity of fractionations the antibody against MC-like protein was tested. The antibody

recognizes specifically a protein from fraction 2 with the approximate molecular mass of 120 kDa, which is in accordance with the calculated mass for this protein.

S10 Fig. Cryo-EM of fraction 3 isolated membrane sheets

Cryo-EM image of frozen-hydrated sucrose-purified membranes from fraction 3 isolated from lysed cell preparation by density gradient centrifugation. Membrane sheets are indicated by arrows and large pore structures are marked by arrowheads. The two pores in the boxed region are seen in Figure 4C.

S11 Fig. Comparison of *Gemmata* pores with the nuclear pores of eukaryotes

(*A*) and **(*B*)** Reconstruction of architecture of a single pore from two different angles. In panel **(*A*)**, a side view of the pore displays the basket structure with a series of struts (arrows) connecting with the main pore rings. In panel **(*B*)**, a top view shows the ring-like element (arrow) of the main part of the pore and a central plug structure is visible within the pore connected to the ring's inner rim via spokes. The same major pore structural elements (plug, ring and spokes) are indicated in the eukaryote nuclear pore shown in Fig S11C and the pores are shown at similar angles (***C***) The image in (***C***) represents a cryo-electron tomographic reconstruction of the *Dictyostelium discoideum* nuclear pore complex published previously by Beck et al. [26].

S12 Fig. Crateriform structures on the surfaces of *G.obscuriglobus* cell walls

(*A*) TEM of cell walls of *G.obscuriglobus* isolated via boiling of bacteria in 10% SDS for 1hr. Bar, 2 µm. (***B***) One of the cell walls of *G.obscuriglobus* with clearly recognizable crateriform structures (arrowheads). The electron dense core regions are variable in shape. Bar, 200 nm. Inset: an enlarged

view of the boxed area in A showing a single crateriform structure with central electron-dense core surrounded by a single electron-transparent ring. There is no indication of division of the ring region into an inner and outer ring. Bar, 50 nm.

S13 Fig. Clustering of membrane-related proteins

All 128 proteins associated with the membrane fractions were clustered using VisBLAST (E = 0.001, i = 2). Proteins are identified by Genbank accession numbers. Lines indicate detectable sequence similarity between proteins. Colour key indicates membrane fractions in which proteins were detected. The 91 singleton proteins that did not show any significant sequence similarity to the other proteins are listed in S3 Table. The cluster in the top left corner (cluster 1) corresponds to the cluster containing fraction 3 proteins from Fig S14. Structural modelling using Phyre2 indicates that constituents of this cluster carry beta-propeller folds (Fig 7C and S4 Table). The cluster on the bottom left is dominated by pili proteins (cluster 2) (see also Fig S18 and S5 Table).

S14 Fig. Clustering of all proteins identified through proteomics

All 512 identified in our proteomics analysis were clustered using VisBLAST (E = 0.001, i = 2). This revealed several large clusters, though only one of these contained proteins specific to fraction 3 (cluster 1). Proteins are identified by Genbank accession numbers. Lines indicate detectable sequence similarity between two proteins. Colour key indicates in which membrane fractions proteins were detected.

S15 Fig. Multiple alignment of the proteins with significant beta-propeller structures

The multiple alignment of the 8 proteins from cluster 1 (top left cluster in Fig S13 and S4 Table) that gave significant structure predictions using Phyre2 is shown. The alignment was made using MAFFT,

option L-ins-i, and was evaluated using the T-Coffee CORE server (see Supplementary Text for details). The modelled structures are indicated by the highlighted part of the alignment at the C-terminal end. Structural models (Fig 3) correspond to the conserved part of the sequences.

S16 Fig. Multiple alignment of the proteins with significant structure predictions in cluster 2

The 10 pili proteins in cluster 2 (bottom left in Fig S13 and S5 Table) that gave significant structure predictions using Phyre2 were aligned using MAFFT, option L-ins-i, and the alignment was evaluated using the T-Coffee CORE server (see text for details). The predicted structures are all clearly located in the conserved subsequence in the N-terminal end of the alignment (see the highlighted part).

S17 Fig. Structures of the pili-proteins from membrane fraction 3

Using PyMOL, the predicted structures for the 11 proteins from membrane fraction 3 that were clustered together and shared a pili-like structure were visualized. None of these are unique to fraction 3 (S5 Table). The structures are for (from left to right and top to bottom): ZP_02731198, ZP_02731806, ZP_02732451, ZP_02732467, ZP_02733038, ZP_02733041, ZP_02735033, ZP_02735132, ZP_02735532, ZP_02735914 and ZP_02737610.

S18 Fig. Structures of the two α-solenoids

Two proteins showed a potential α-solenoid structure with stacked α-helices. Left: ZP_02735673 (constituent of fraction 2 and fraction 3) and right: ZP_02736511, unique to pore-containing membrane fraction 3.

S19 Fig. Immuno-gold particles distribution

(***A***) Distribution of gold particles within *Gemmata obscuriglobus* cells. The bars represent a number of particles associated with the intracytoplasmic membranes (blue bars) vs with no visible association with the membranes (red bars). A total of 50 cells were used for the counting; 549 particles were found as associated with the membranes and 60 as not associated. The particles were counted according to the description in (*B*). (***B***) An example of gold particle distribution within the cells of *Gemmata obscuriglobus*, labelled with 6670 antibody and then with 10 nm gold protein A. The majority of the particles were found associated with the intracytoplasmic membrane (black arrows). The particles were considered as membrane-associated if the distance between the membrane and the particle did not exceed 20 nm. Blue arrowheads indicate particles which were counted as a background or cell wall-associated, black arrowheads show non-membrane associated particles. Bar, 1 μm.

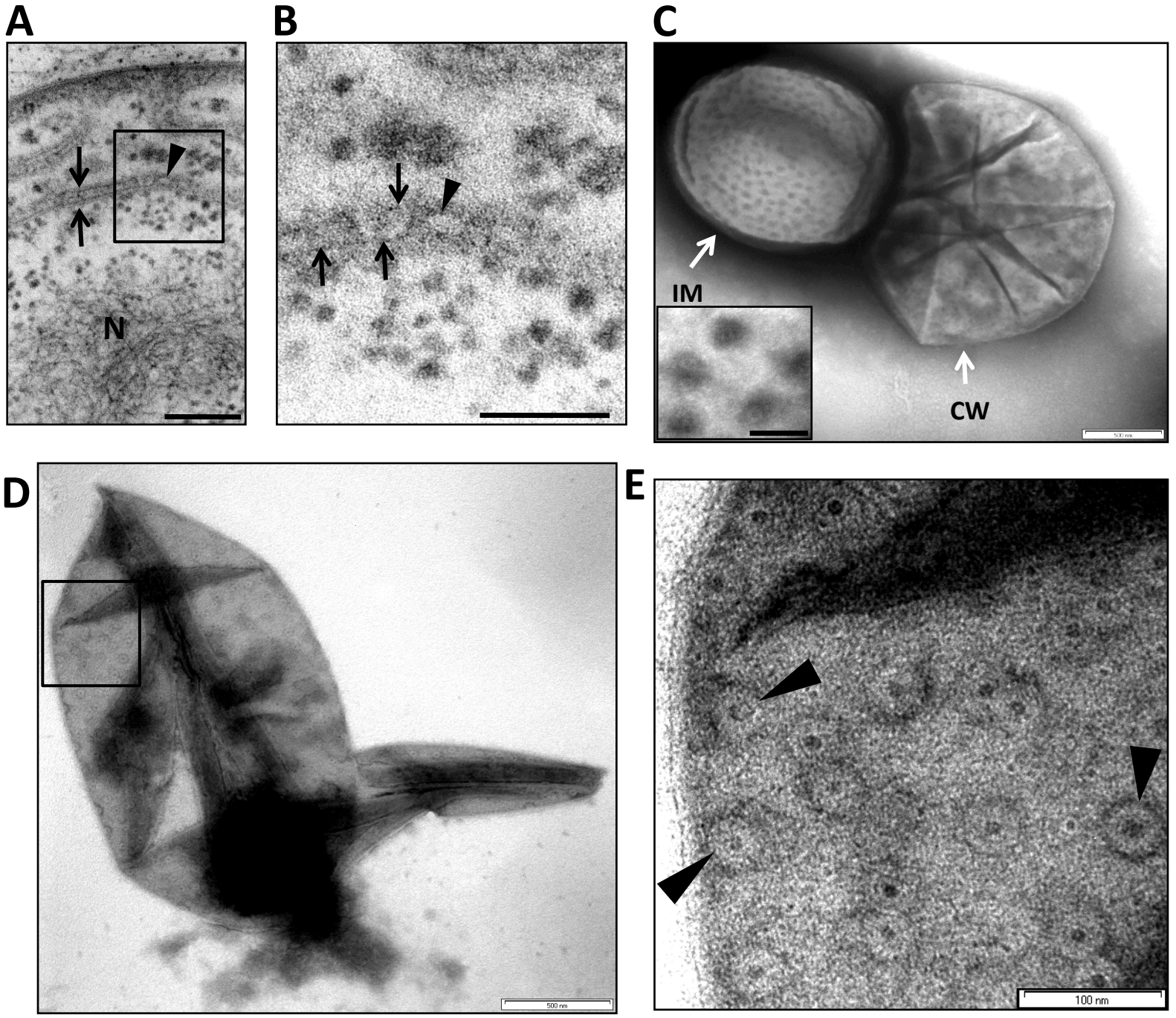


Fig. 1

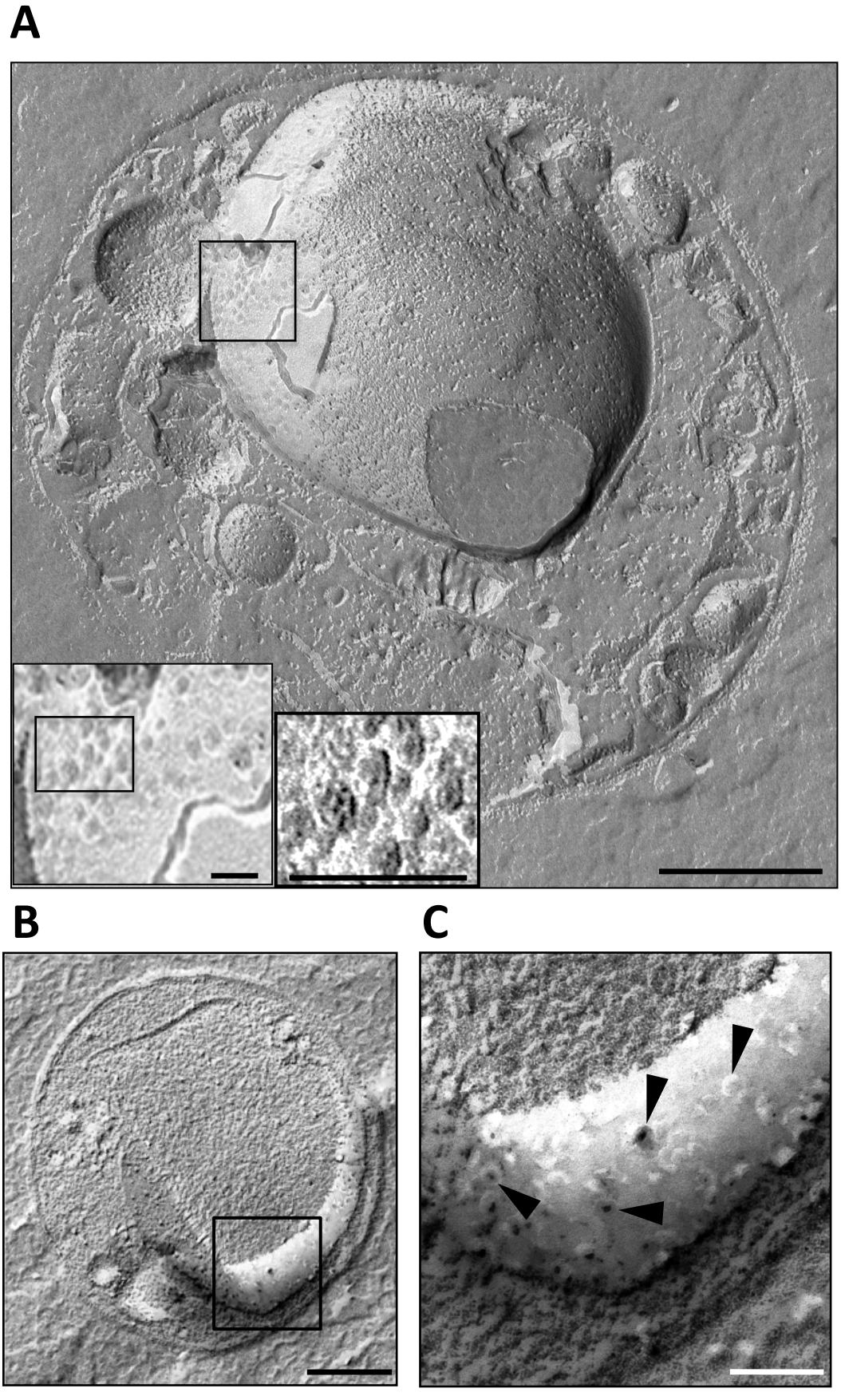


Fig. 2

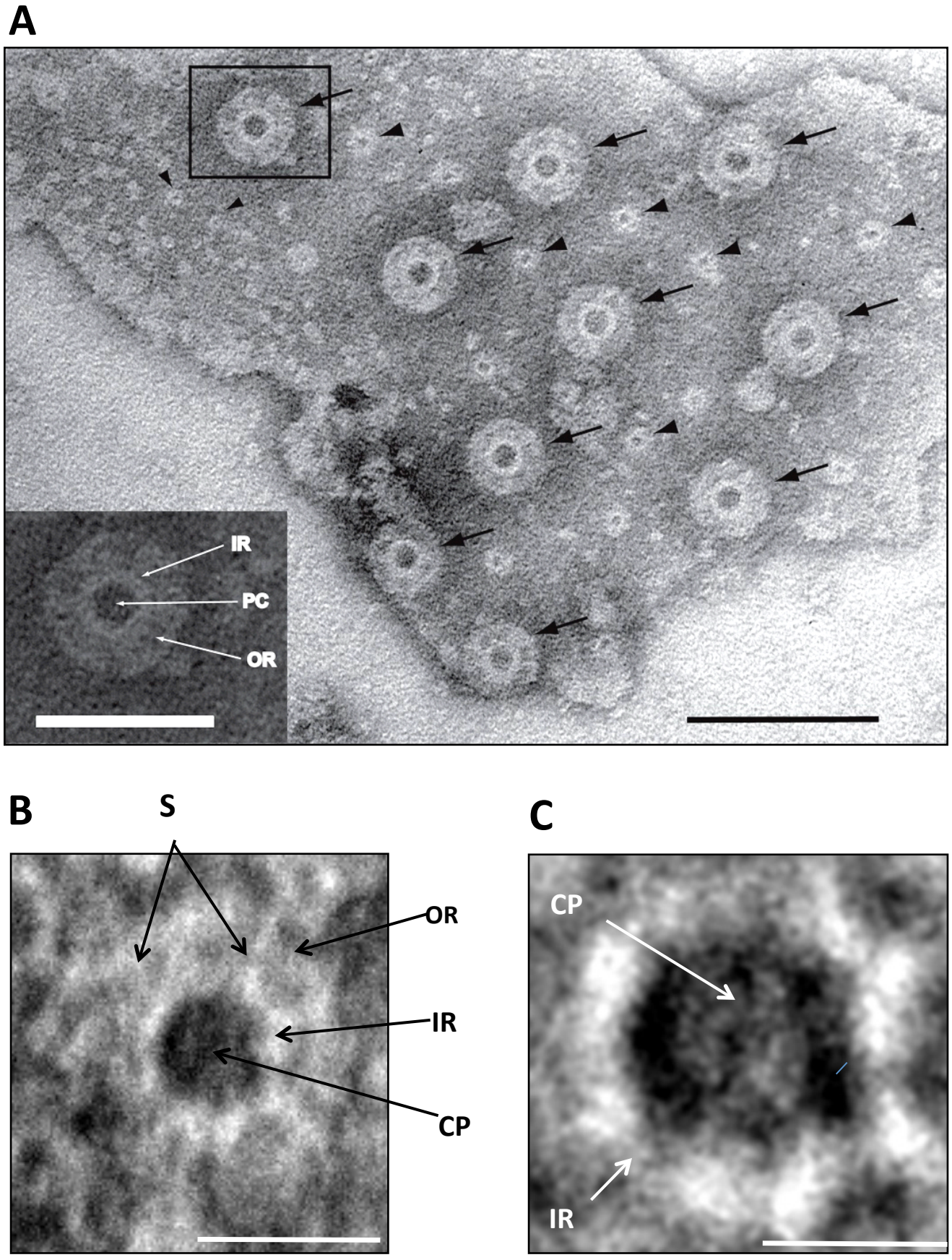


Fig. 3

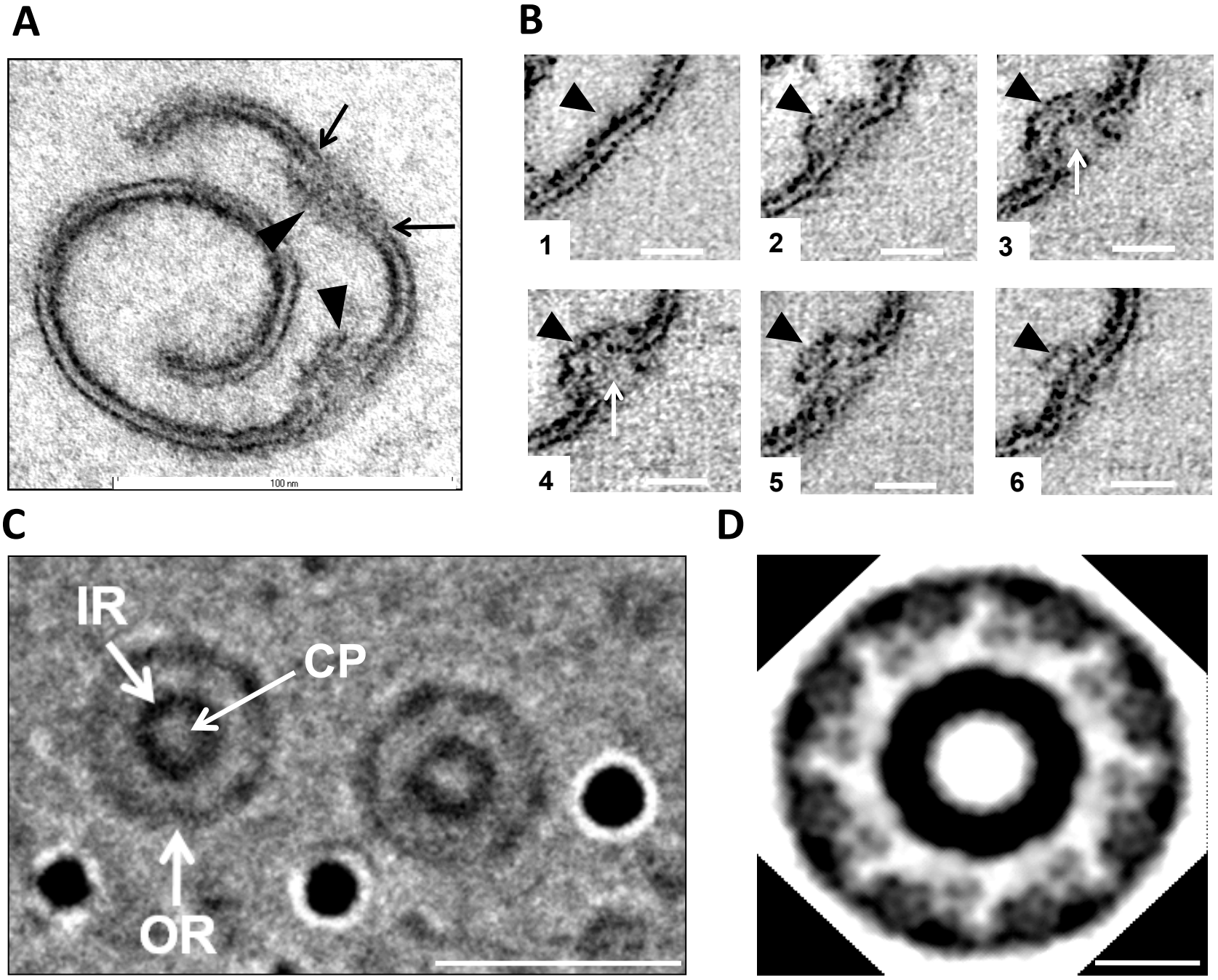


Fig. 4

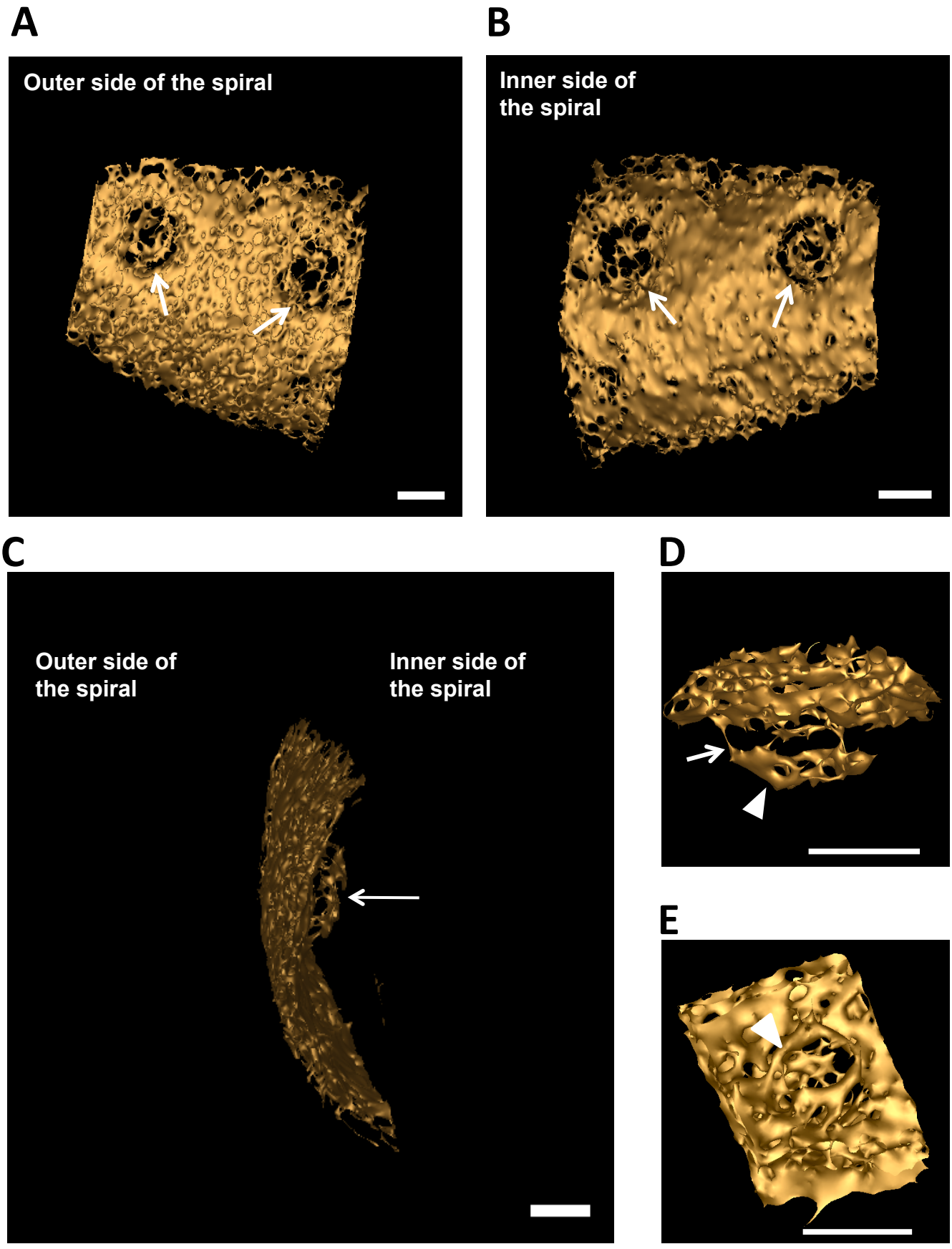

A
Outer side of the spiral
B
Inner side of
the spiral
C
Outer side of
the spiral
Inner side of
the spiral
D
E

Fig. 5

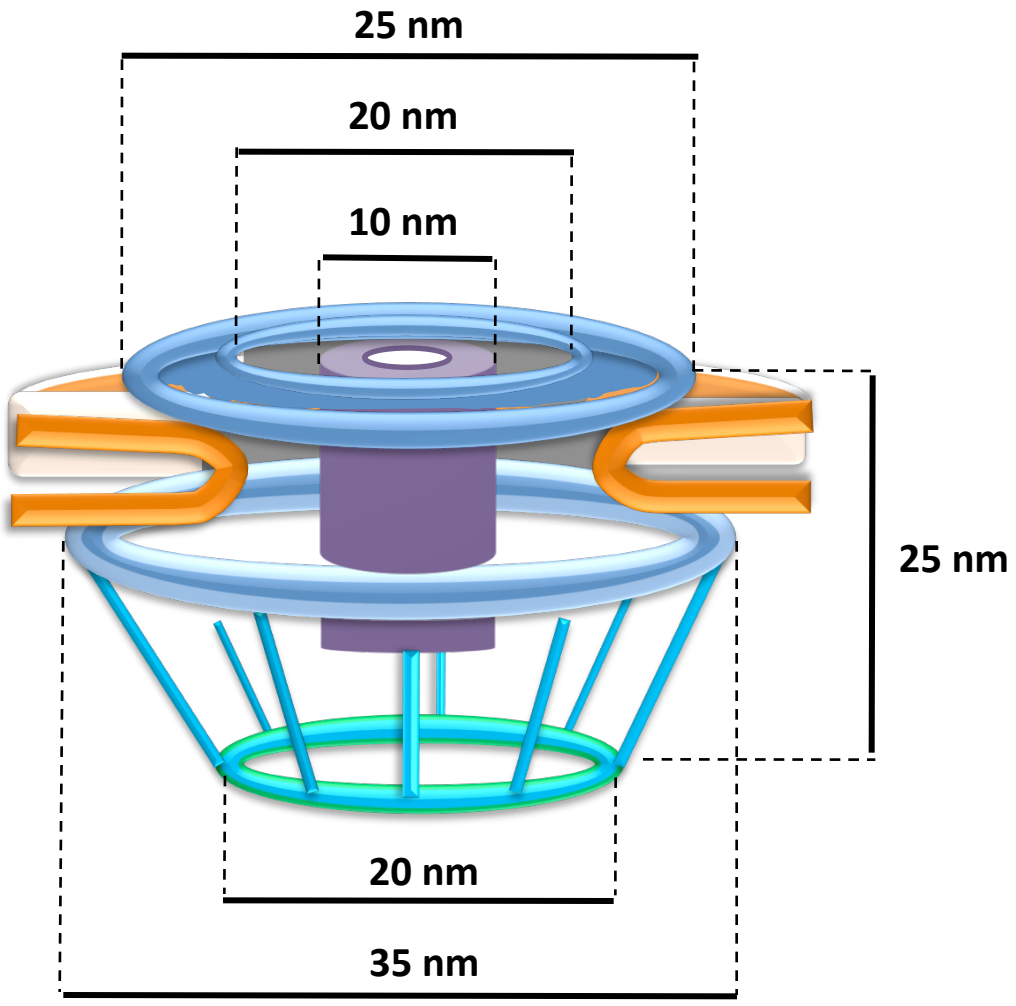


Fig. 6

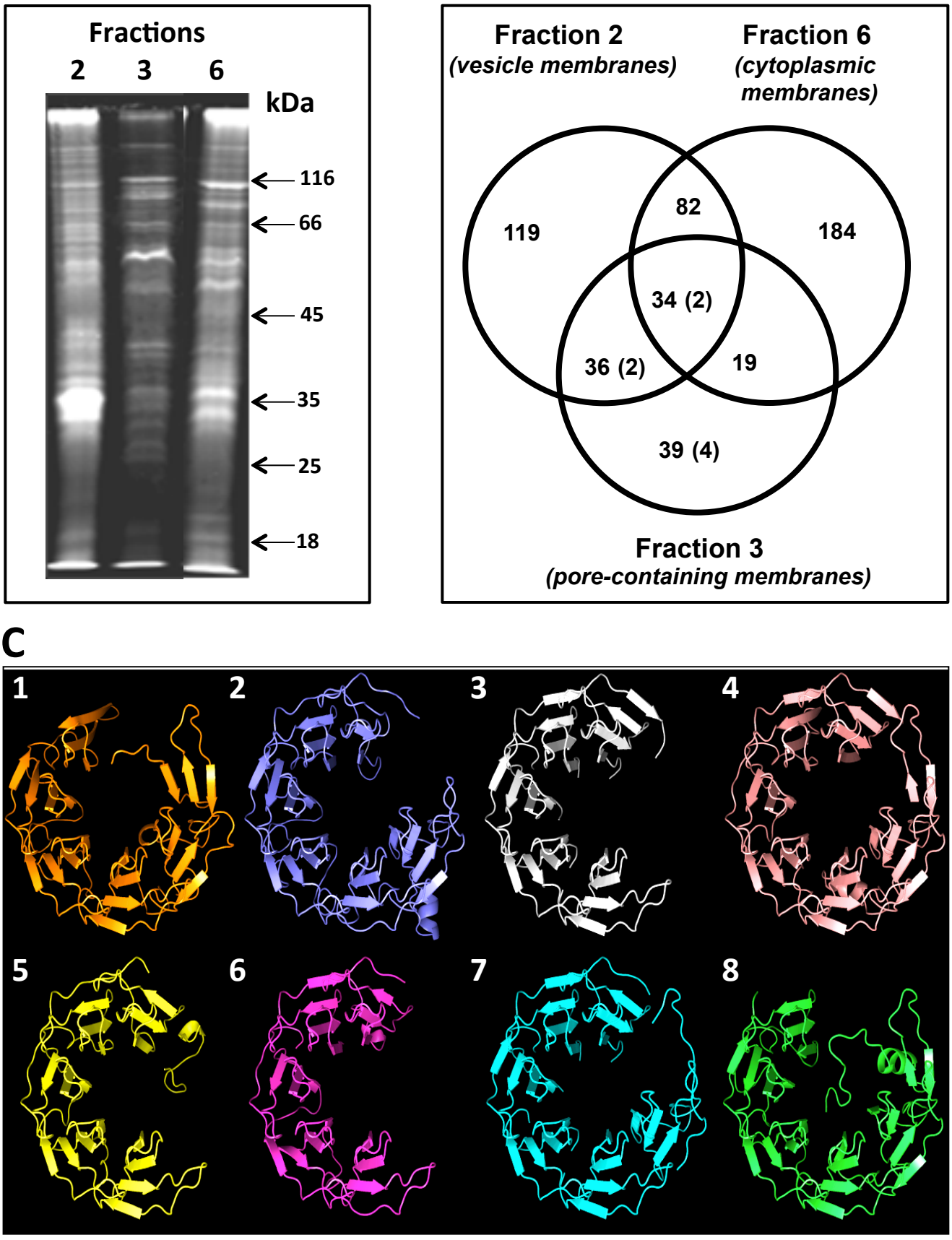

A
Fractions
2
3
6
kDa
116
66
45
35
25
18
B
Fraction 2
(vesicle membranes)
Fraction 6
(cytoplasmic membranes)
119
82
184
34 (2)
36 (2)
19
39 (4)
Fraction 3
(pore-containing membranes)
C
1
2
3
4
5
6
7
8

Fig. 7

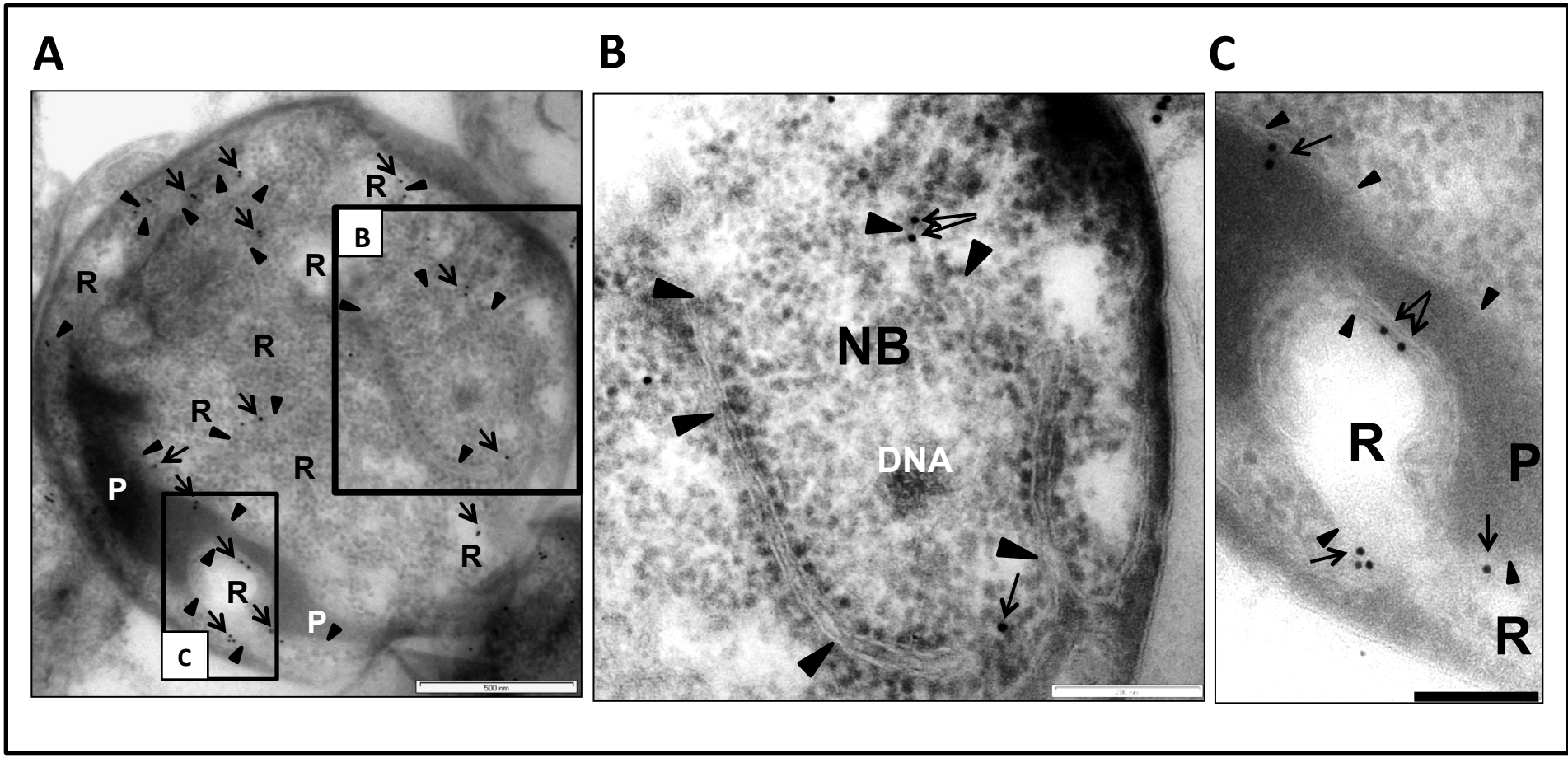


Fig. 8

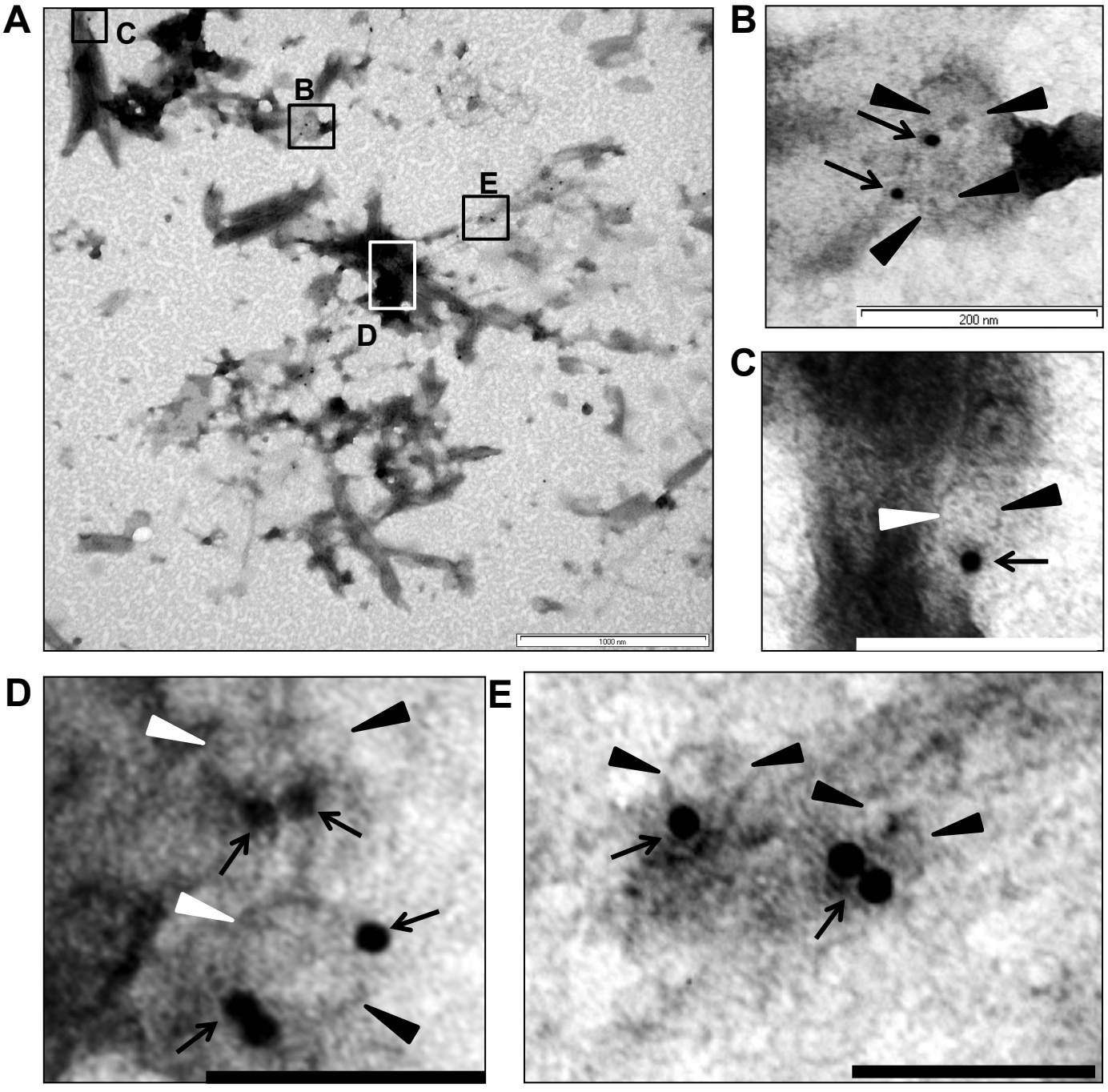
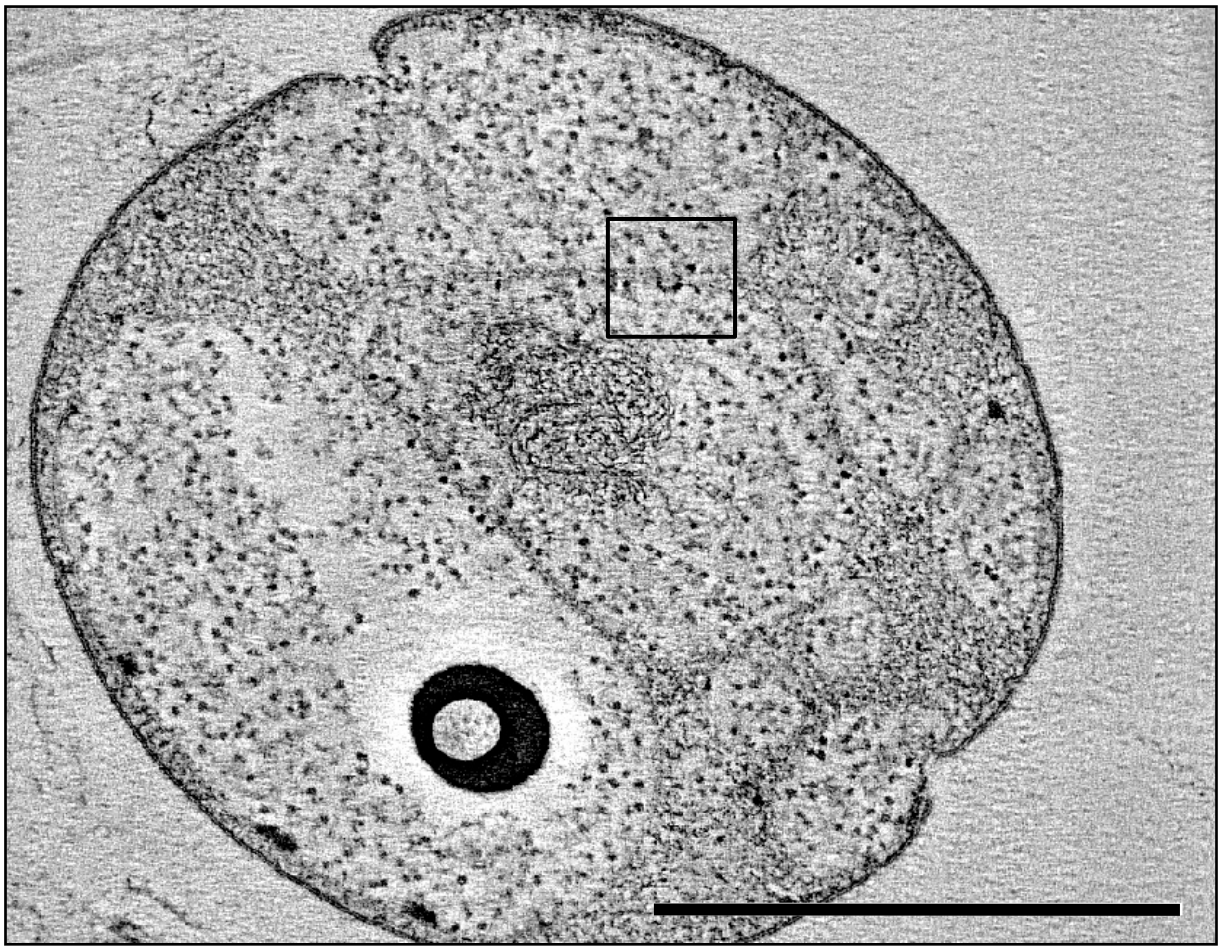
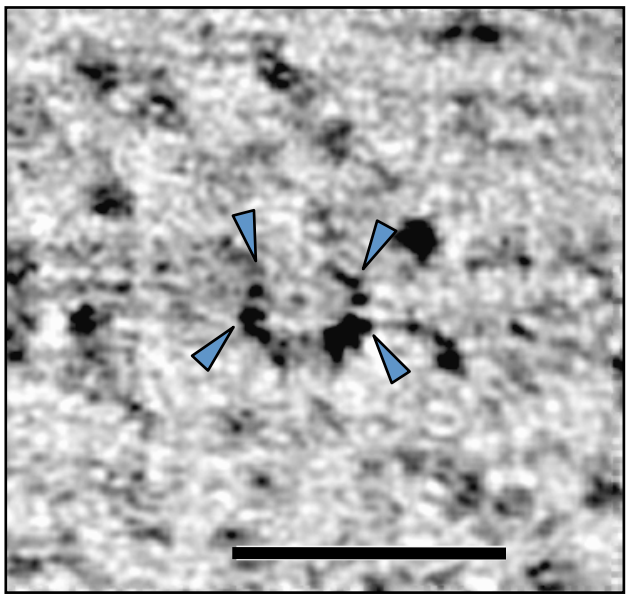


Fig. 9

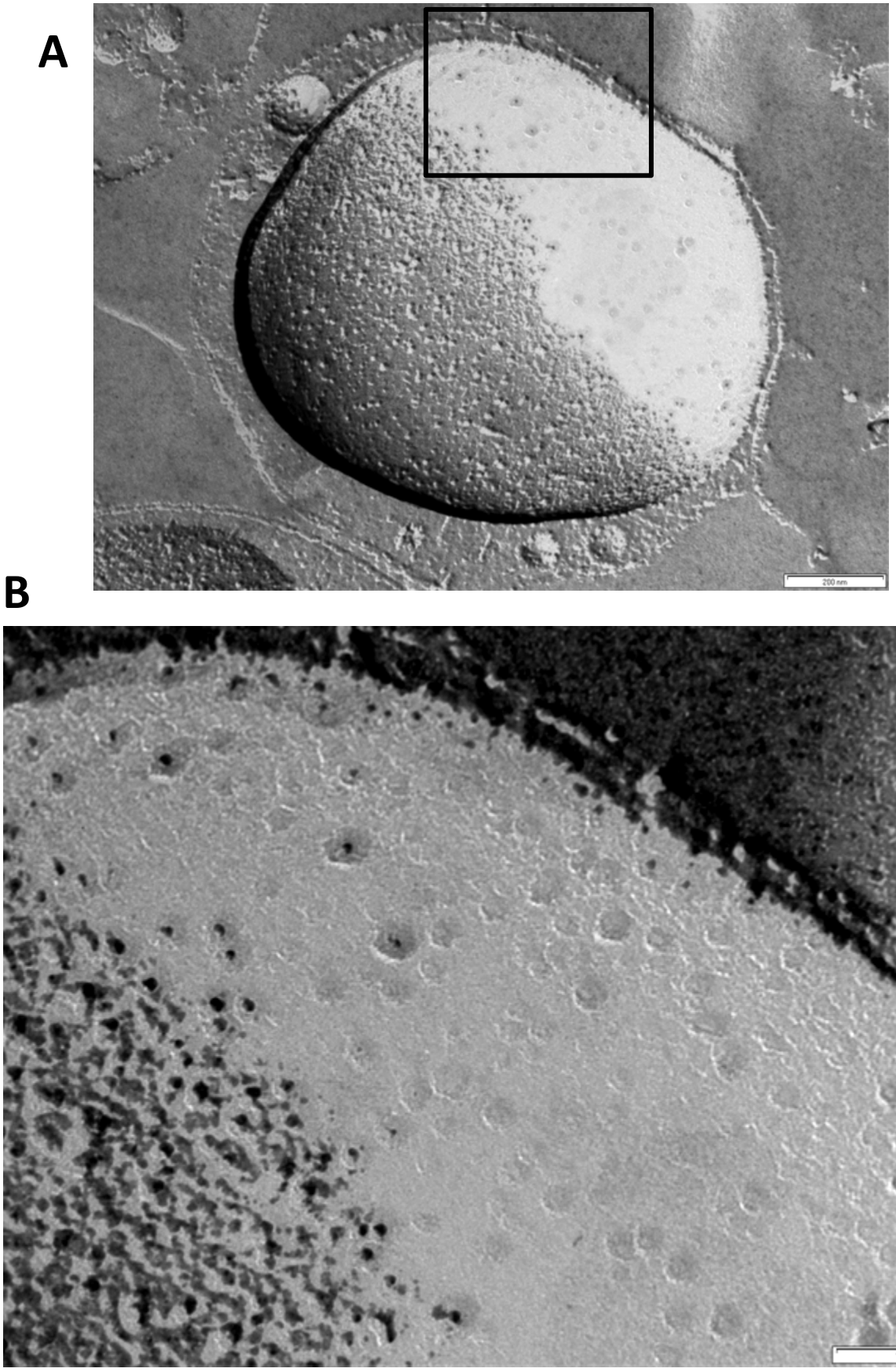
A



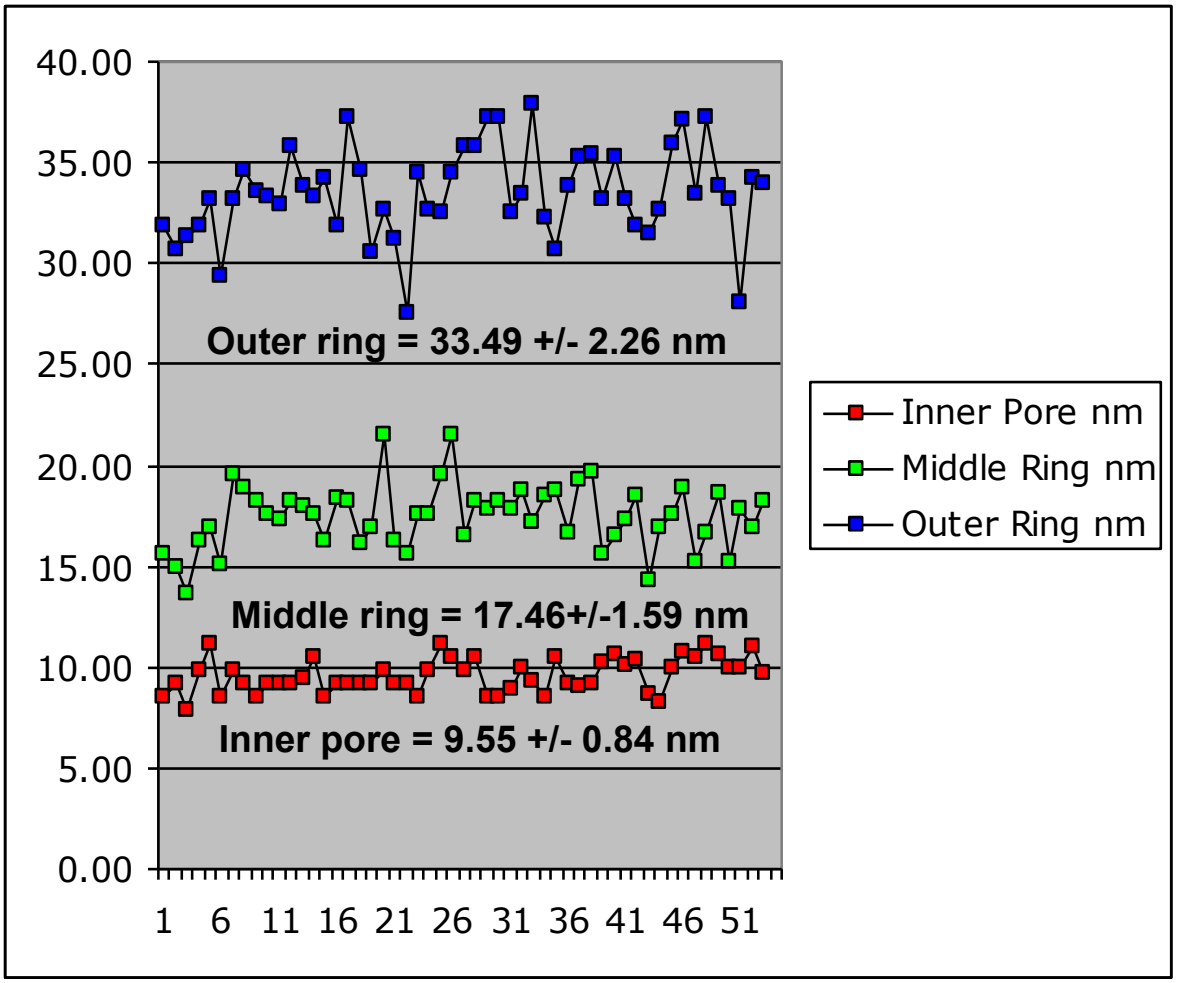
B



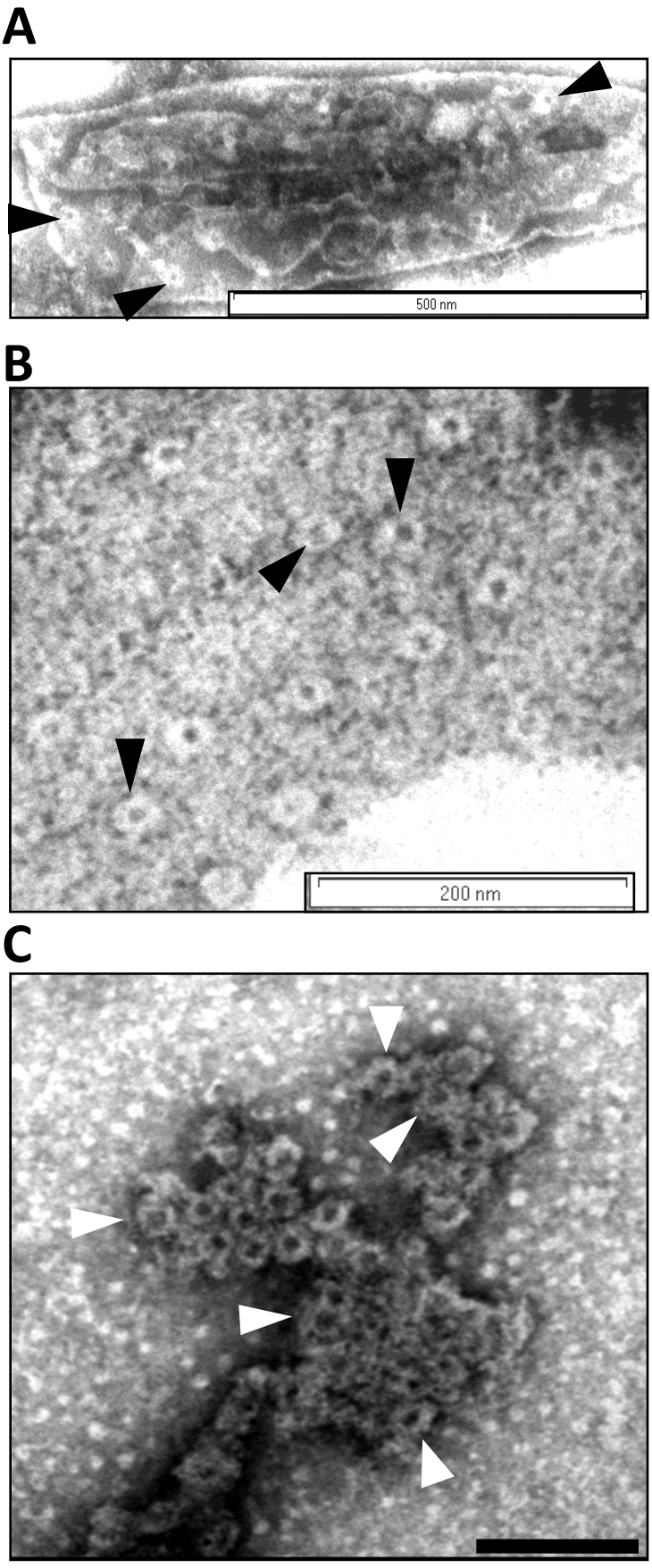
S1 Fig



S2 Fig

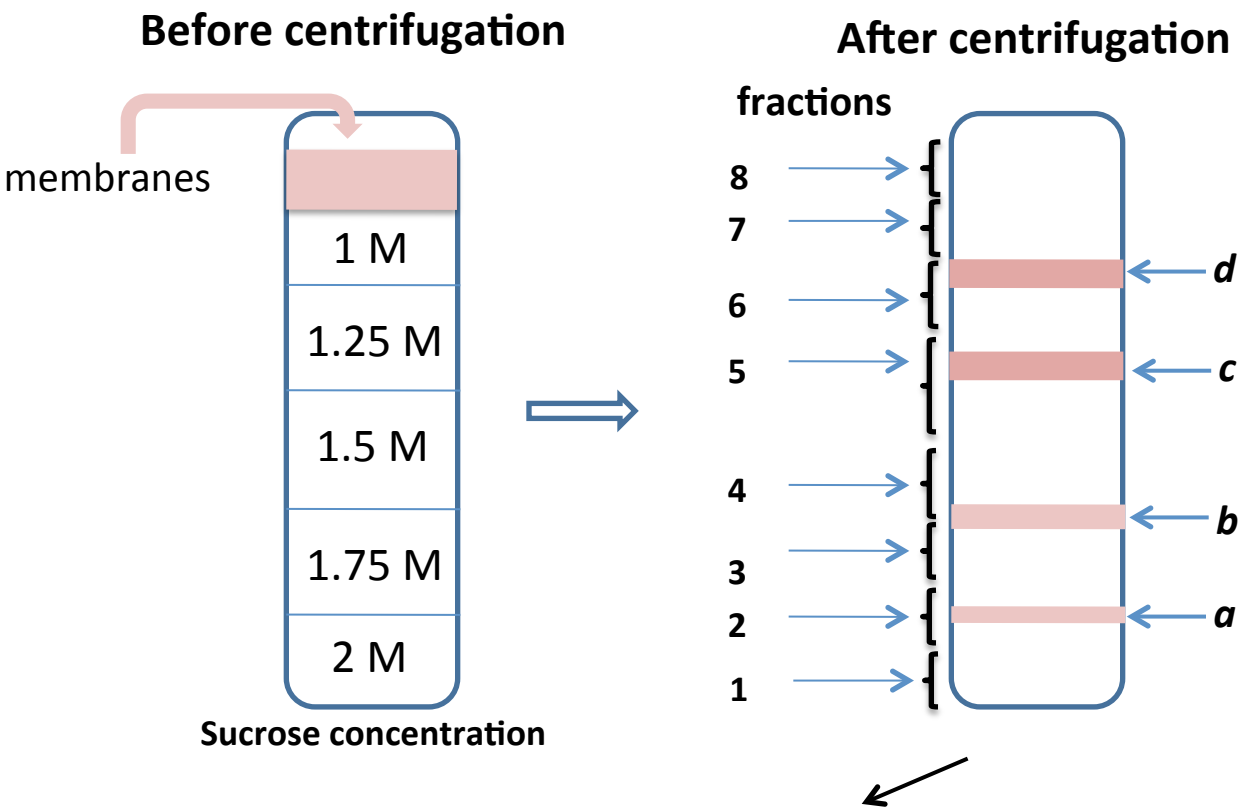

40.00
35.00
30.00
25.00
20.00
15.00
10.00
5.00
0.00
1
6
11
16
21
26
31
36
41
46
51
Outer ring = 33.49 +/- 2.26 nm
Middle ring = 17.46+/-1.59 nm
Inner pore = 9.55 +/- 0.84 nm
Inner Pore nm
Middle Ring nm
Outer Ring nm

S3 Fig



S4 Fig

A



Further purification via continues gradient centrifugations of fractions 2 (band *a*), 3 and 4 (band *b*), 5 (band *c*), and 6 (band *d*)

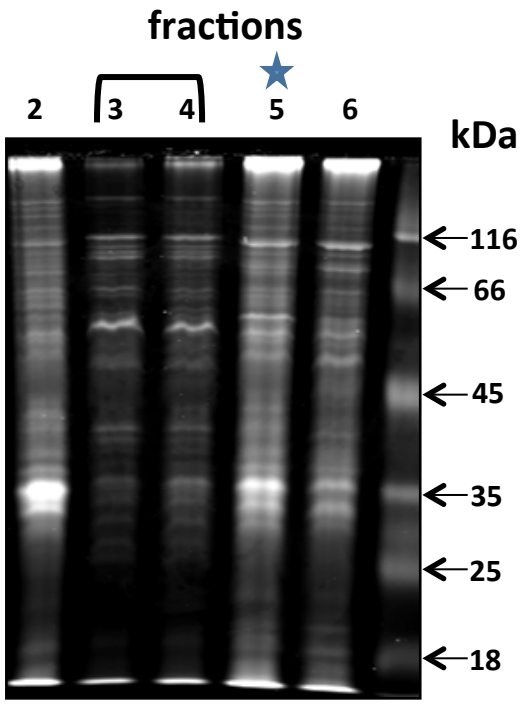
B

Fraction 2 corresponds to band a

Fractions 3 and 4 – to band b

★ *Fraction 5* – corresponds to band c (this fraction represents not separated Membrane aggregates

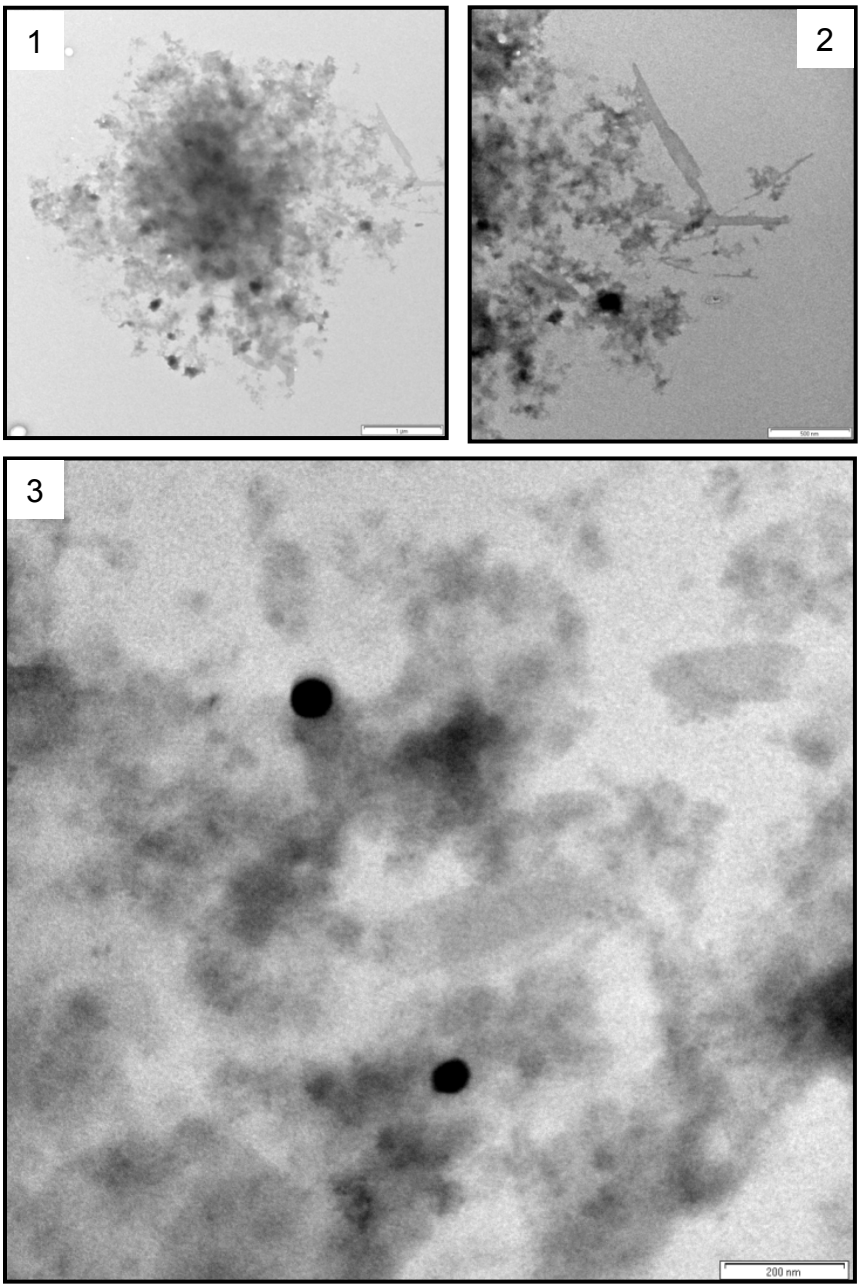
Fraction 6 – band d



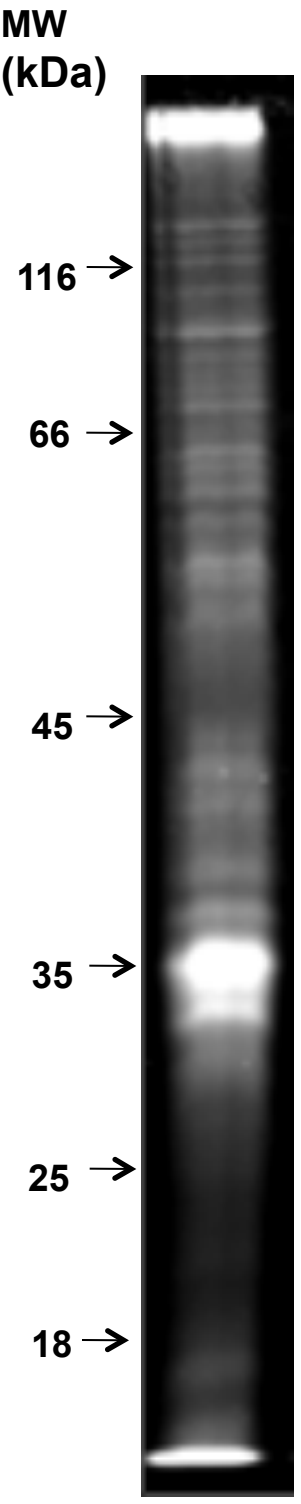
S5 Fig

Sucrose gradient *Fraction 2*

A

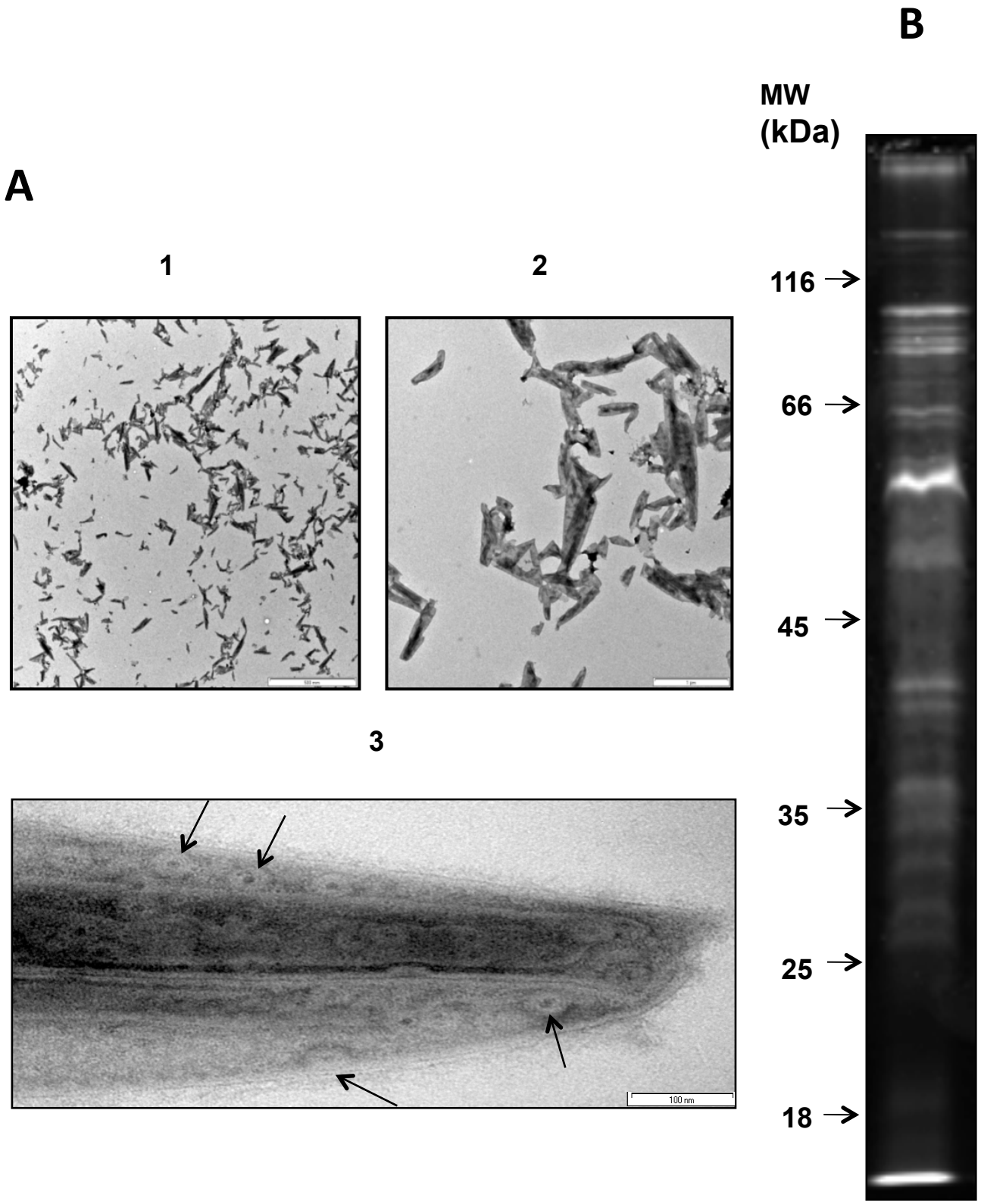


B



S6 Fig

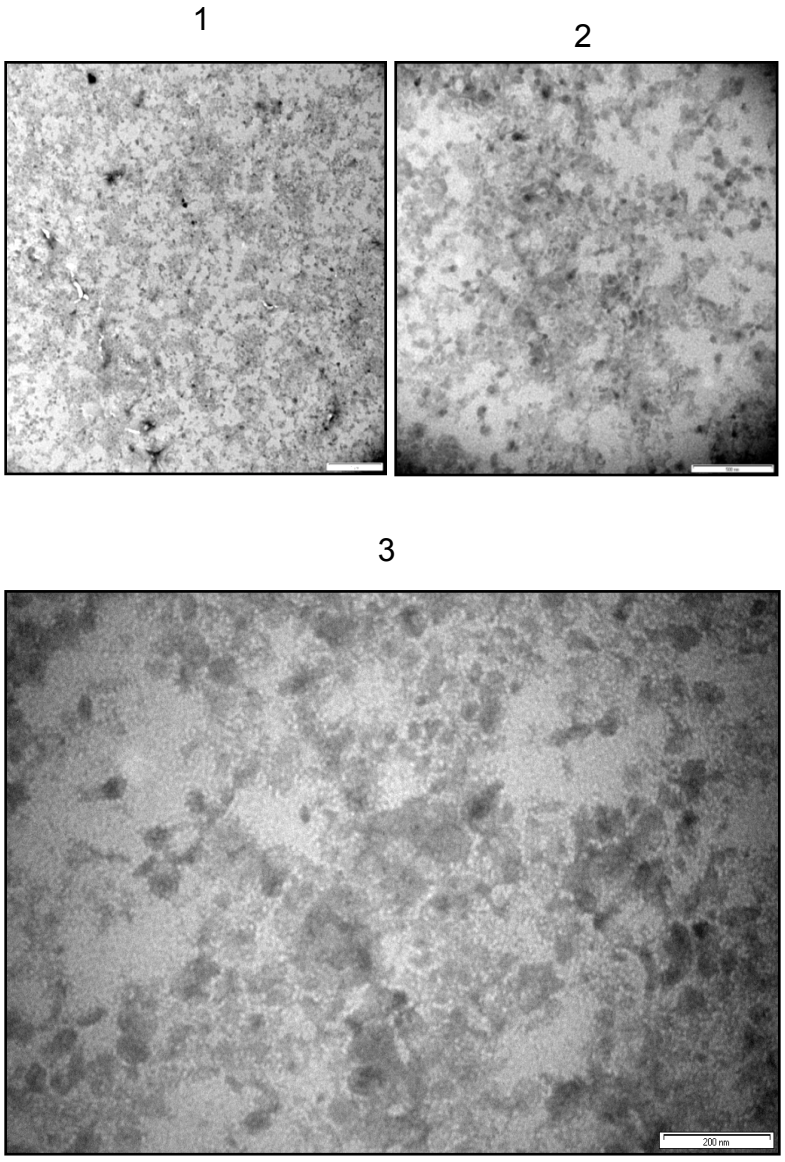
Sucrose gradient *Fraction 3*



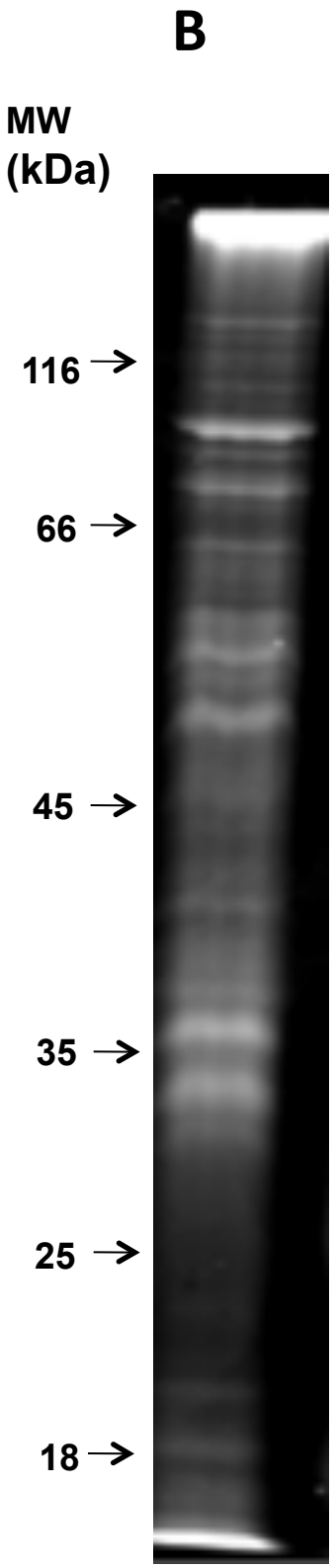
S7 Fig

Sucrose gradient *Fraction 6*

A



B



S8 Fig

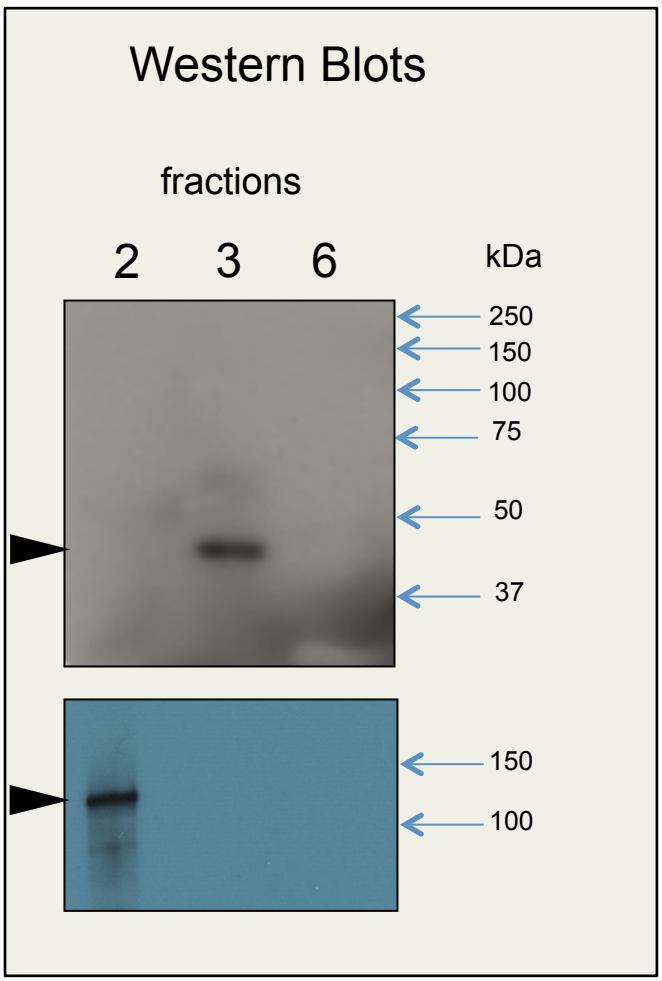
Annotation: Na-Ca exchanger/integrin-beta4 [Gemmata obscuriglobus]

NCBI Reference Sequence: **ZP_02736670** (WP_010049031.1)

>gi|497734847|ref|WP_010049031.1| Na-Ca exchanger/integrin-beta4 [Gemmata obscuriglobus]
MSFLPSPLRARLNCEPFEDRTVPAAGVMEISFAGNAFETPGAAVNVTVIRTDGSDGIVSA
TLTATGGTATSGTDFAAGPFTVTFAPGEVTKVVTVPVTDDTRKEPTETATFTLSNPTGGAT
LGTSVSATATIFDDDGTLVASAQPPSQFASPQQTTVNISAANASEILPTGFRSSFVPFAGFN
GTLSVALGDINRDGIRDVIVGASVNGHVKVLDGATGAELRSFLAYPGYPGAVRVAAGDV
NGDGVADIITGADINAHVKVFDGATGAELRSFLAYPGFPGAVDVAAADVTGDGVADIITG
AGINGHVKVFDGKTNAEVRSFLAYSGYTGPALVAAGDVNGDGFADIITGAPAAGHIKVFD
GKTNAESLSFLDTRTTPTVAFVALAAVDADGDGLADIIAGESSIYGSYNGRTGVKLPISGSID
RLNTNFVSVG

product="Na-Ca exchanger/integrin-beta4"

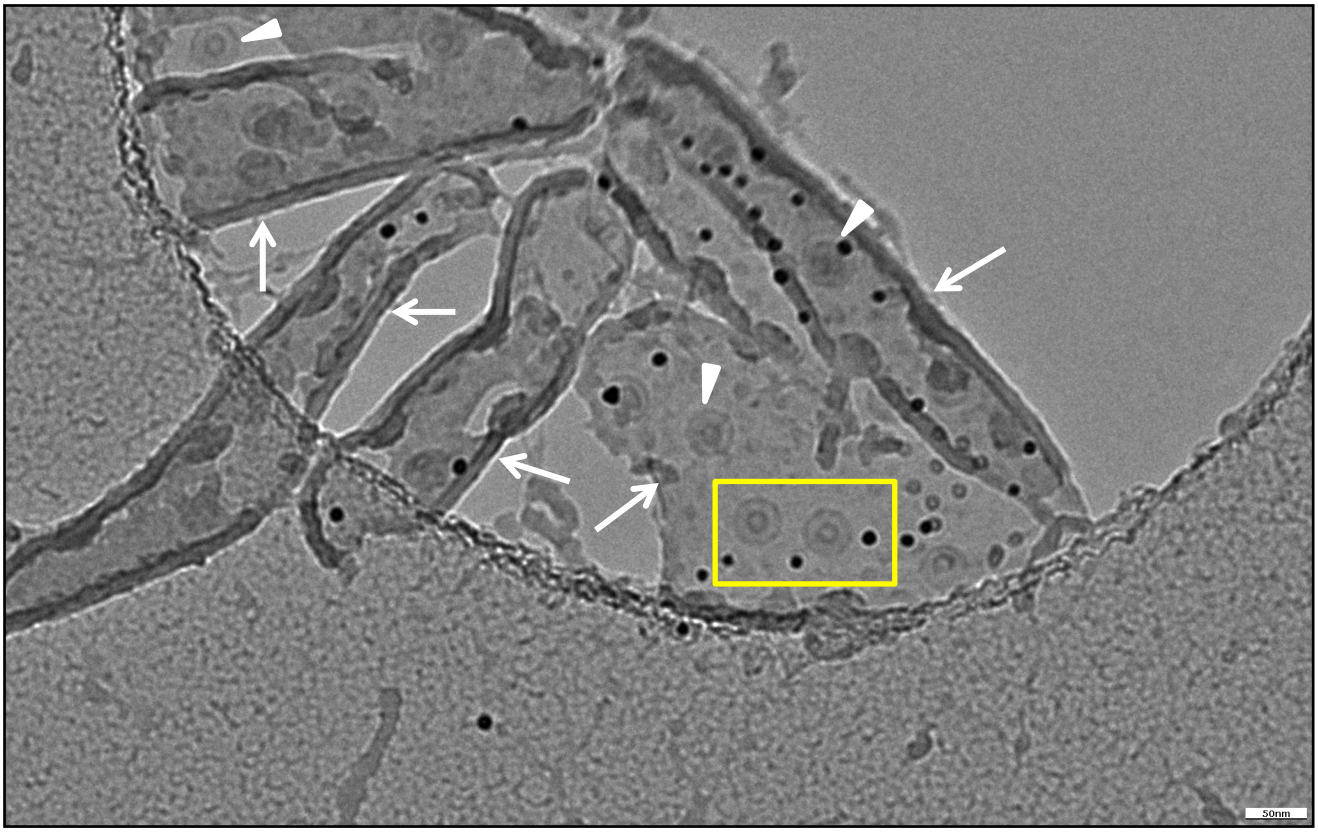
/calculated_mol_wt=43599



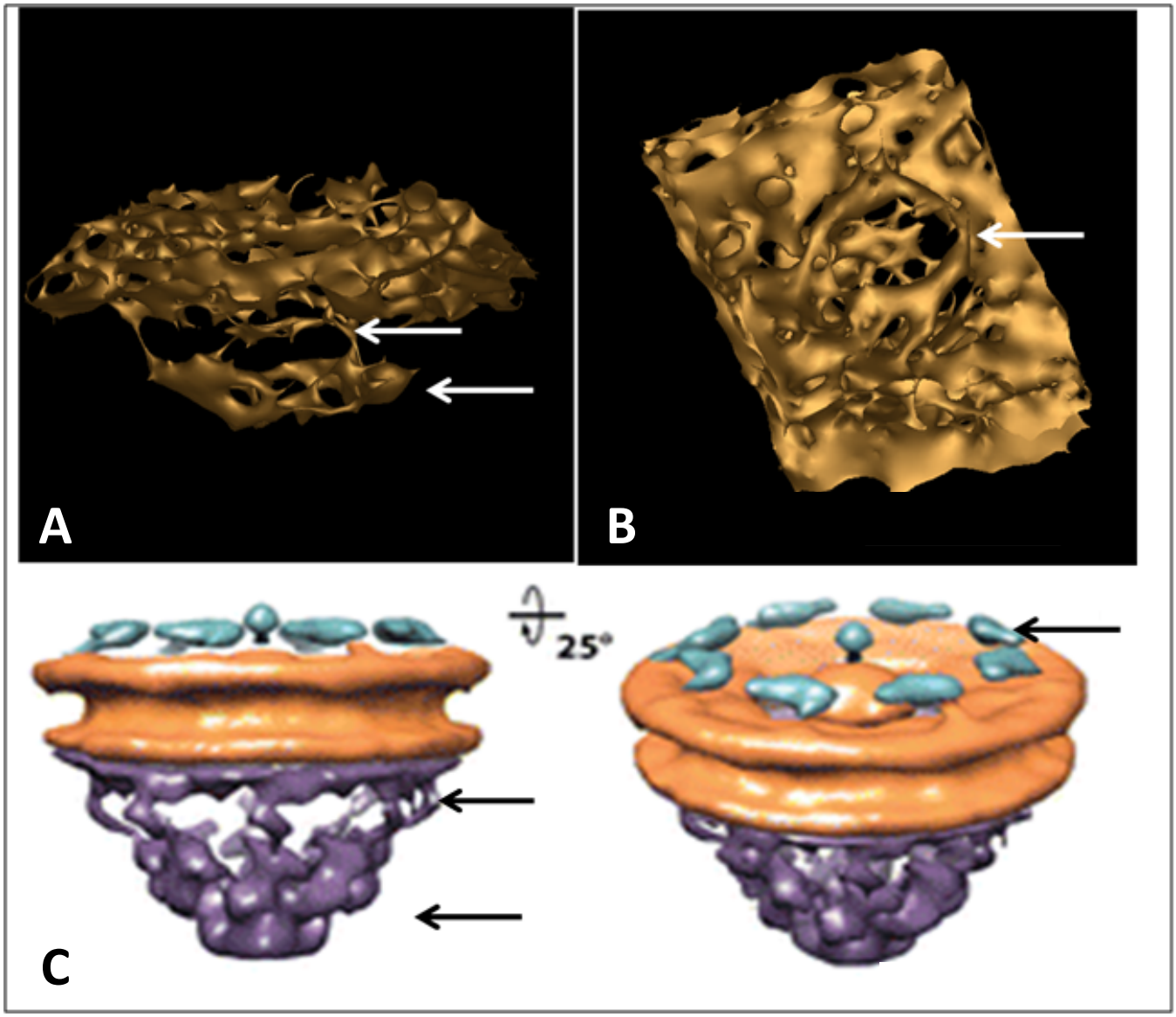
Anti-Protein 6670 antibody
Peptide sequence –
VPVTDDTRKEPTETC

Anti-MC-like protein antibody

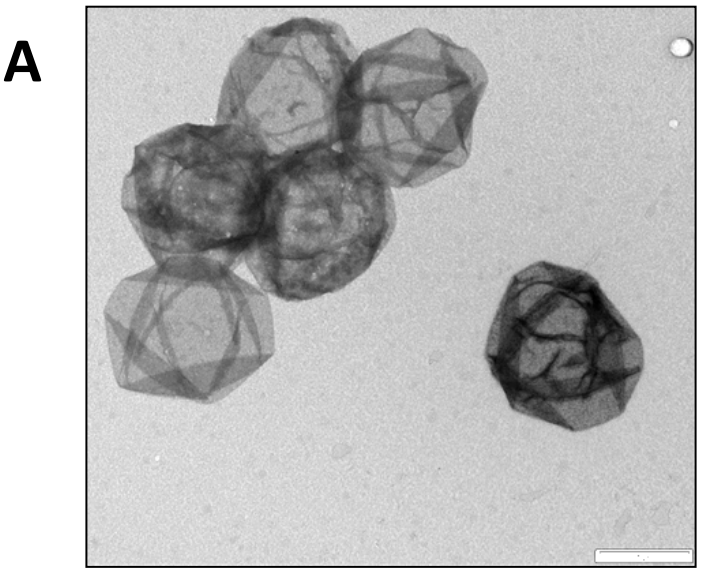
S9 Fig

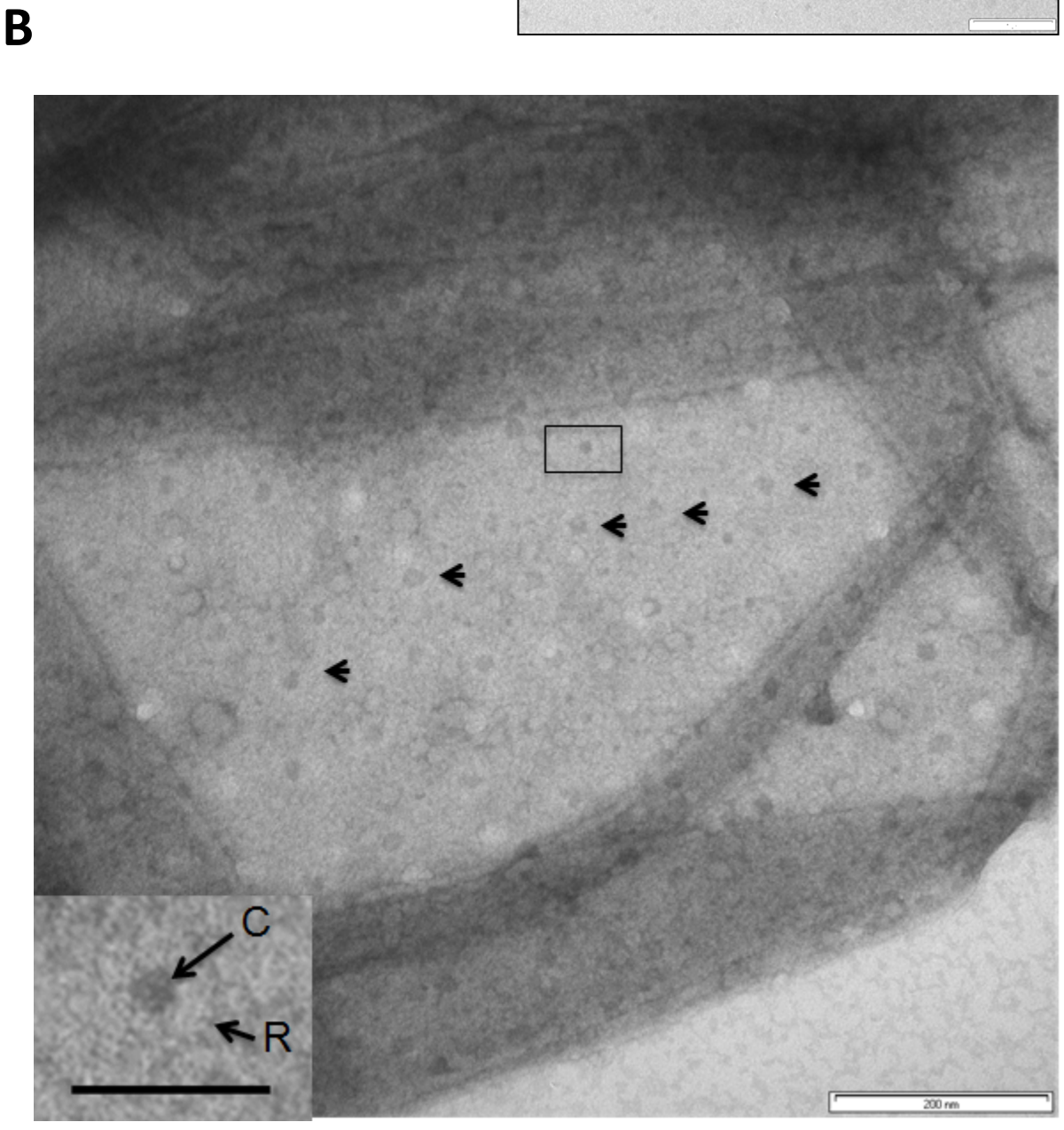


S10 Fig

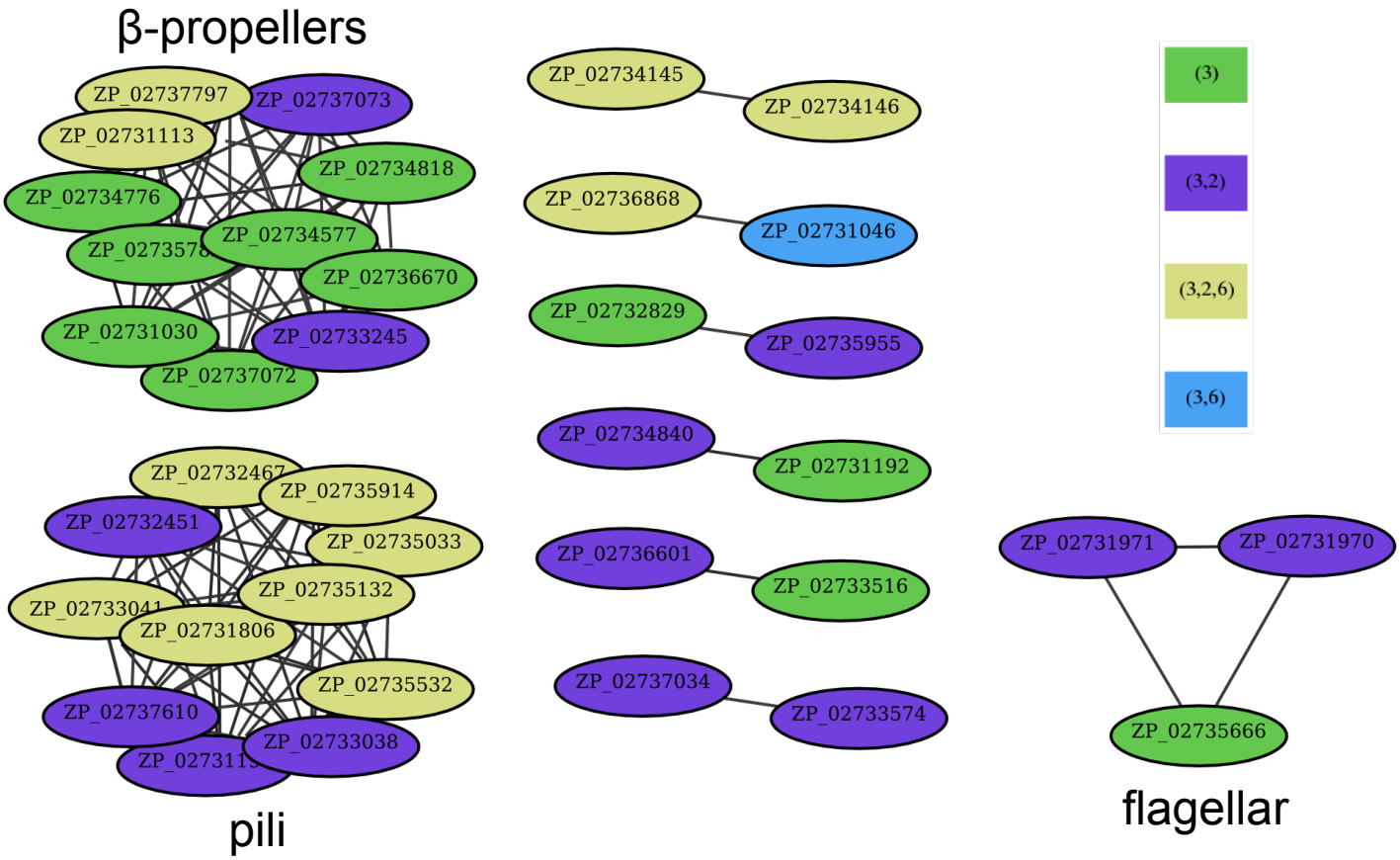


S11 Fig

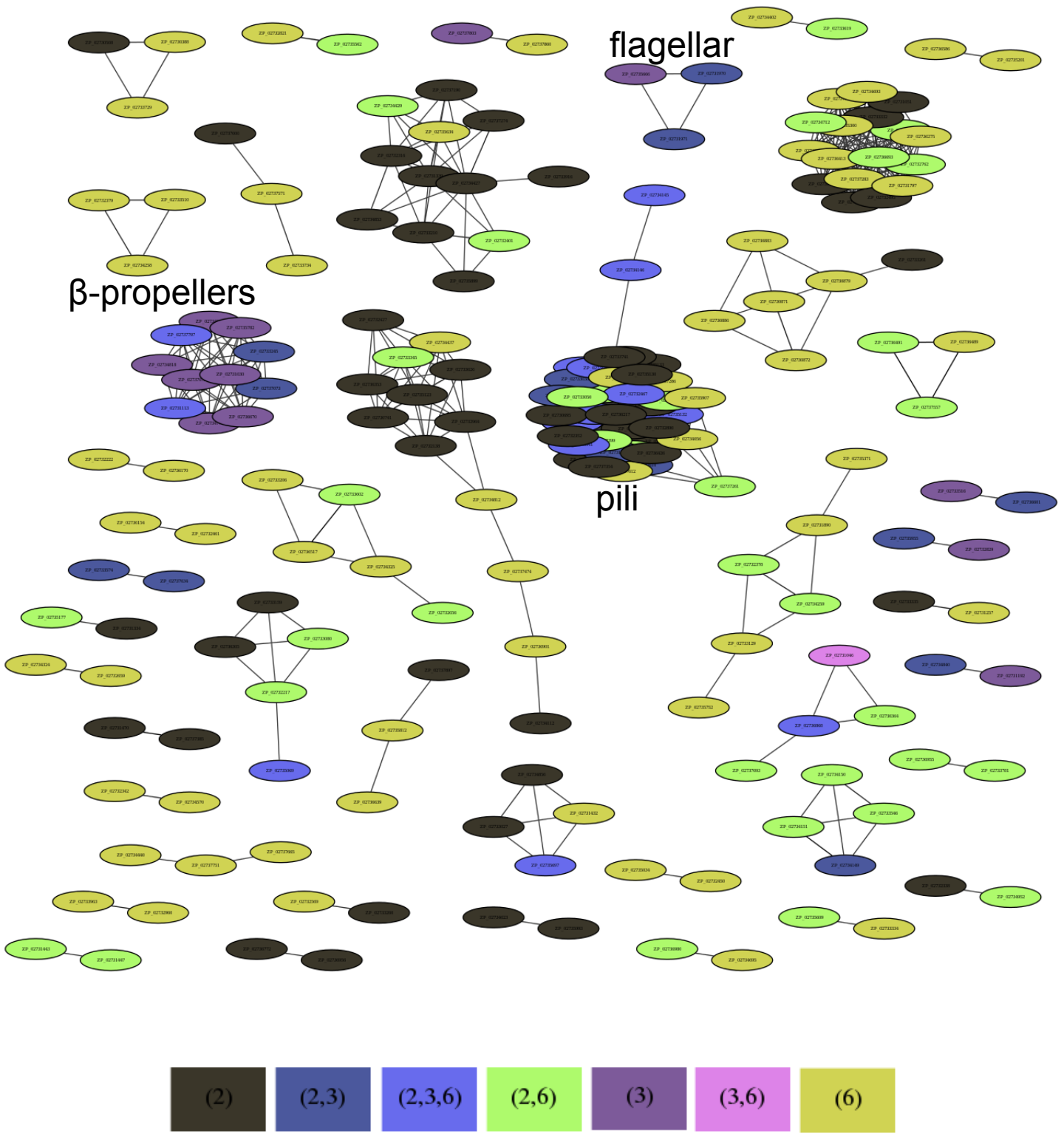




S12 Fig


β-propellers
ZP_02737797
ZP_02737073
ZP_02731113
ZP_02734818
ZP_02734776
ZP_02734577
ZP_02736670
ZP_02731030
ZP_02733245
ZP_02737072
ZP_02734145
ZP_02734146
ZP_02736868
ZP_02731046
ZP_02732829
ZP_02735955
ZP_02734840
ZP_02731192
ZP_02736601
ZP_02733516
ZP_02737034
ZP_02733574
(3)
(3,2)
(3,2,6)
(3,6)
ZP_02732467
ZP_02735914
ZP_02732451
ZP_02735033
ZP_02735132
ZP_02731806
ZP_02735532
ZP_02737610
ZP_02733038
pili
ZP_02731971
ZP_02731970
ZP_02735666
flagellar

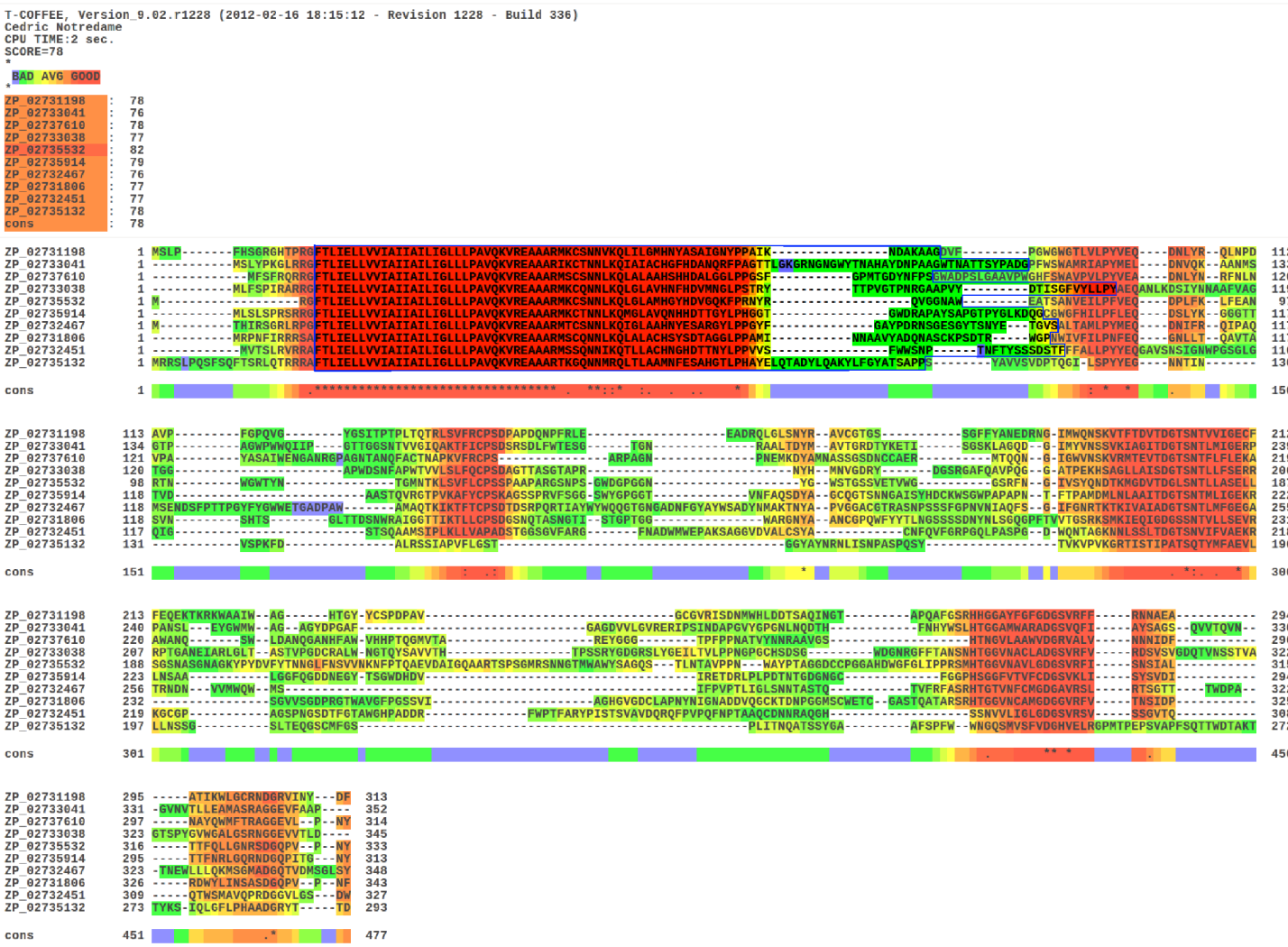
S13 Fig



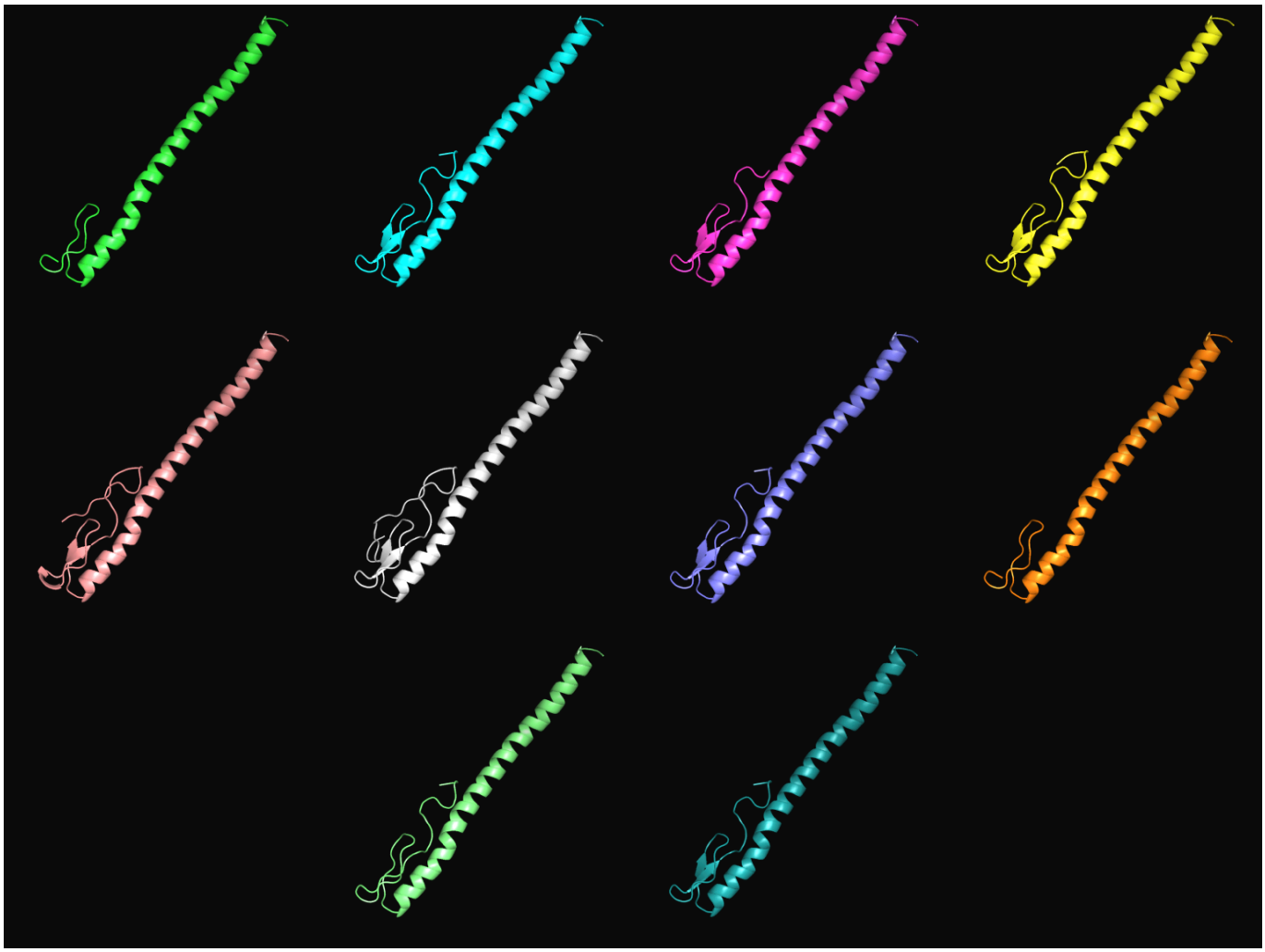
S14 Fig


T-COFFEE, Version_9.02.r1228 (2012-02-16 18:15:12 - Revision 1228 - Build 336)
Cedric Notredame
CPU TIME:5 sec.
SCORE=69
BAD AVG GOOD
ZP_02731113 : 70
ZP_02733245 : 67
ZP_02734776 : 72
ZP_02736670 : 73
ZP_02737797 : 72
ZP_02734577 : 70
ZP_02737072 : 68
ZP_02737073 : 57
cons : 69

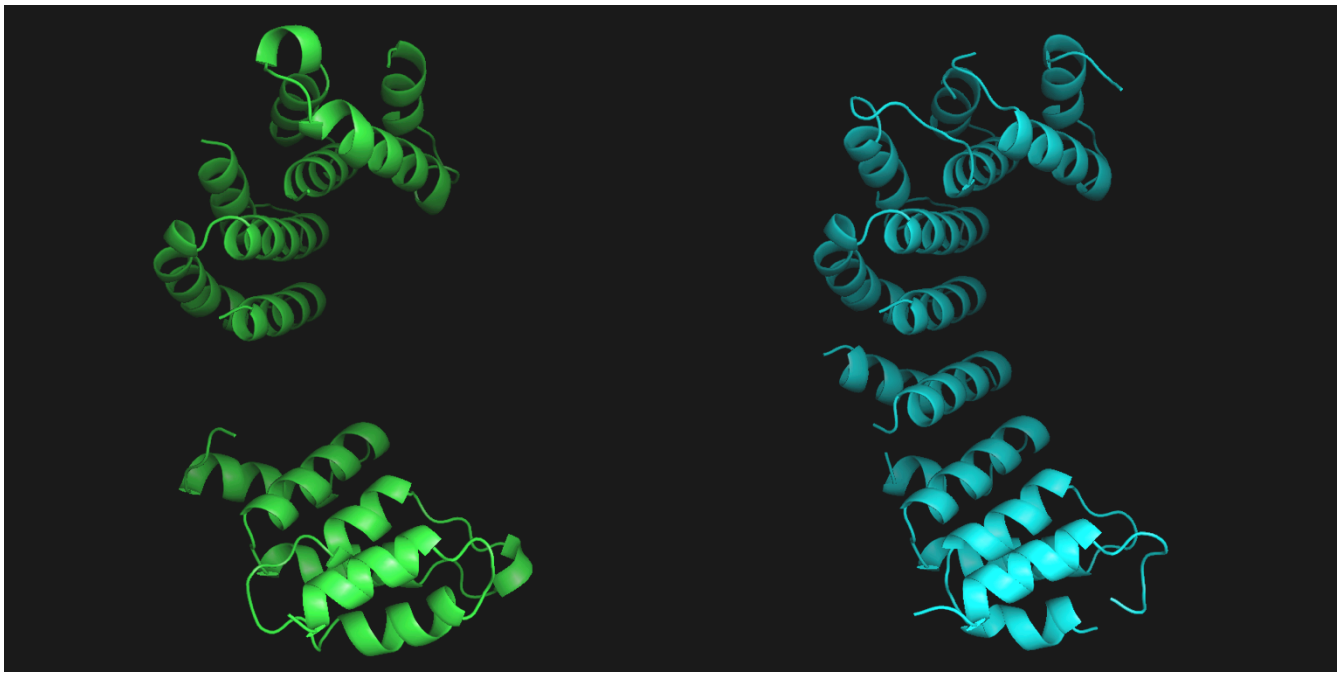
S15 Fig


T-COFFEE, Version_9.02.r1228 (2012-02-16 18:15:12 - Revision 1228 - Build 336)
Cedric Notredame
CPU TIME:2 sec.
SCORE=78
BAD AVG GOOD
ZP_02731198 : 78
ZP_02733041 : 76
ZP_02737610 : 78
ZP_02733038 : 77
ZP_02735532 : 82
ZP_02735914 : 79
ZP_02732467 : 76
ZP_02731806 : 77
ZP_02732451 : 77
ZP_02735132 : 78
cons : 78

S16 Fig

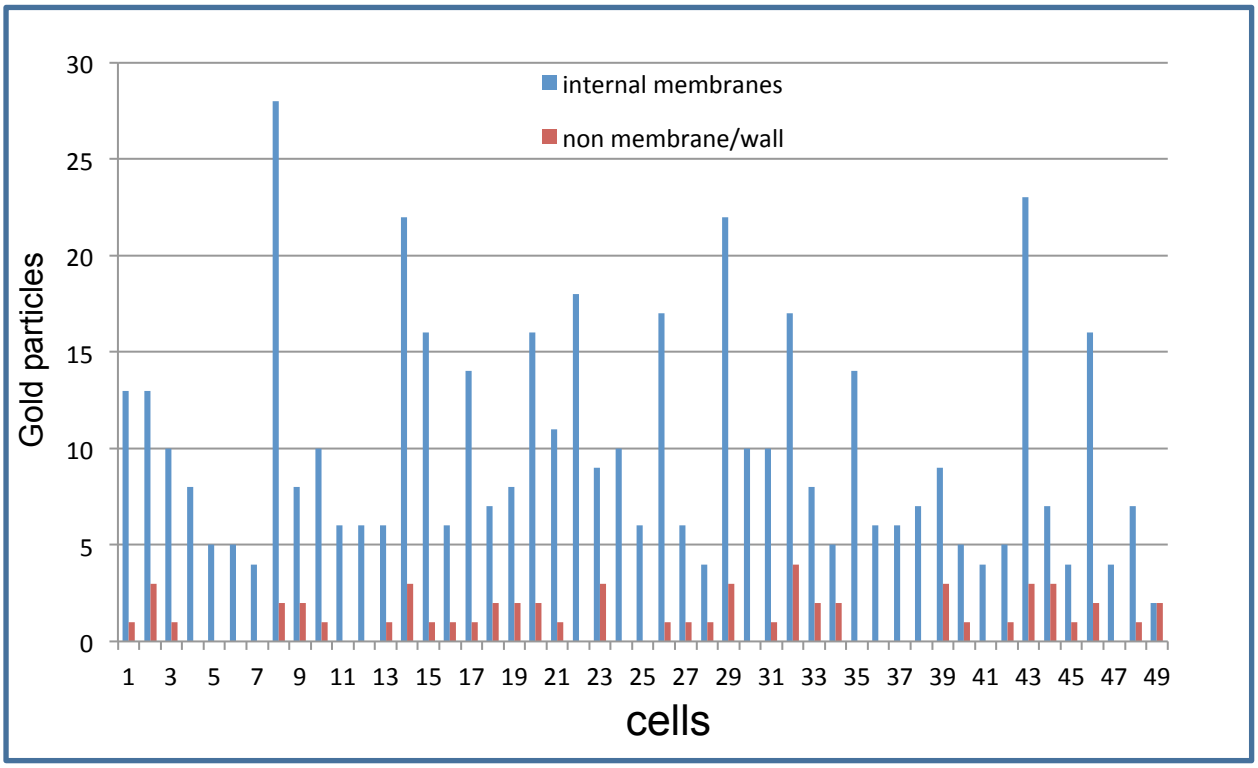


S17 Fig



S18 Fig

A


internal membranes
non membrane/wall
Gold particles
cells
0
5
10
15
20
25
30
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39 41 43 45 47 49

B

S19 Fig

S1 Table. Proteins identified in membrane fractions by MALDI-TOF.

Fraction 2 (if protein found in other fractions it is indicated in brackets)	**Protein name**	**NCBI accession numbers**	**Calculated molecular weight (Da)***	**MOWSE score**	**Peptides matched**	**Sequence coverage (%)**
2	hypothetical protein GobsU_11075 (MC-like protein)	ZP_02732338	124479	969	18	19.3
2 (6)	4Fe-4S ferredoxin, iron-sulfur binding domain protein	ZP_02733602	122387	1335	18	26.2
2 (3,6)	hypothetical protein GobsU_29906	ZP_02736062	76943	1341	71	40.8
2 (3)	type IV fimbrial assembly protein PilB	ZP_02734149	70569	1068	39	30.3
2 (6)	probable DNA-directed RNA polymerase alpha chain	ZP_02736955	51701	762	15	39.1
2 (6)	F0F1 ATP synthase subunit beta	ZP_02731447	52132	939	14	40.3
2 (3,6)	hypothetical protein GobsU_20253	ZP_02734146	26677	693	74	53.9
2 (6)	oxidoreductase, short chain dehydrogenase/reductase family protein	ZP_02733345	24580	509	23	42.8
2	short chain dehydrogenase	ZP_02732964	28767	258	13	18.4
2 (3,6)	hypothetical protein GobsU_27236	ZP_02735532	35749	883	94	50.8
2 (6)	hypothetical protein GobsU_17456	ZP_02733593	12531	232	3	31.6
2 (6)	hypothetical protein GobsU_11415	ZP_02732406	14995	394	5	36.6
2 (6)	DNA-directed RNA polymerase beta chain	ZP_02733087	141275	760	13	15.4
2 (6)	DNA-directed RNA polymerase beta chain	ZP_02733088	164212	514	14	9.2
2 (6)	hypothetical protein GobsU_34912	ZP_02737058	127649	1141	18	17.9
2	putative exported protease	ZP_02737000	118801	486	10	13.9
2	signal transduction histidine kinase with CheB and CheR activity	ZP_02734469	104719	480	8	12.9
2 (6)	hypothetical protein GobsU_35960	ZP_02737261	103499	446	9	11.4
2	dihydrolipoamide dehydrogenase	ZP_02733335	50314	706	10	31.4
2	hypothetical protein GobsU_14594	ZP_02733027	56438	326	5	12.5
2	hypothetical protein GobsU_25171	ZP_02735122	52187	280	5	14.3
2	hypothetical protein GobsU_14434	ZP_02732995	52892	274	8	22
2 (6)	Cobalamin synthesis protein/P47K	ZP_02730707	41940	201	7	13
2	glycogen synthase	ZP_02733261	56208	220	3	11.2
2 (6)	multidrug efflux system, HlyD family subunit	ZP_02734259	47627	199	4	11.2
2 (6)	6-phosphogluconate dehydrogenase	ZP_02734911	52647	155	3	11.2
2	hypothetical protein GobsU_04689	ZP_02731070	52224	165	7	9.8
2	hypothetical protein GobsU_20468	ZP_02734189	25555	206	2	15.4
2 (3)	Uridylate kinase	ZP_02736756	26457	425	20	34.1
2 (6)	Sulphate transport system permease protein 1	ZP_02732762	39652	326	16	19.4
2 (3,6)	hypothetical protein GobsU_06435	ZP_02731416	103608	654	15	17
2 (3,6)	aconitate hydratase 1	ZP_02730459	98013	514	10	15.5
2 (6)	protein-export membrane protein secD	ZP_02733003	123971	699	10	11.8
2 (3,6)	putative peptidase	ZP_02736310	42703	766	62	39
2 (3,6)	hypothetical protein GobsU_20233	ZP_02734142	48877	686	56	34
2	methyltransferase	ZP_02734785	43932	352	18	11
2	chorismate mutase	ZP_02736115	41559	346	13	11
2 (6)	methylcitrate synthase	ZP_02732426	42530	250	14	11
2 (3)	flagellar basal body rod protein FlgG	ZP_02731970	27437	418	12	29.7
2 (6)	putative ABC transporter ATP-binding protein	ZP_02736693	34576	1208	93	67.9
2 (6)	hypothetical protein GobsU_14709	ZP_02733050	37118	312	25	20.9
2 (3,6)	50S ribosomal protein L1	ZP_02733084	30488	343	35	21.5
2	succinic semialdehyde dehydrogenase	ZP_02737897	17115	159	8	30.2
2 (6)	cobalt-zinc-cadmium resistance protein	ZP_02733511	54820	194	5	11.7
2	bifunctional GMP synthase/glutamine amidotransferase protein	ZP_02737792	56752	216	9	11.3
2	branched-chain amino acid transport ATP-binding protein	ZP_02731798	25852	267	11	23.4
2	Streptomyces cyclase/dehydrase	ZP_02734441	20855	174	10	23
2 (6)	ketol-acid reductoisomerase	ZP_02737871	38760	457	14	28.7
2 (3,6)	hypothetical protein GobsU_20643	ZP_02734224	92712	526	15	12.7
2 (6)	hypothetical protein GobsU_04644	ZP_02731061	47023	306	17	8
2 (3,6)	hypothetical protein GobsU_31654	ZP_02736410	87810	452	10	11.5
2 (3,6)	oxidoreductase domain protein	ZP_02735697	47995	1426	157	113
2 (6)	serine proteinase, HtrA/DegQ/DegS family	ZP_02735252	42708	236	11	11
2	xylose isomerise domain protein TIM barrel	ZP_02734509	30645	110	4	11.5

2 (3,6)	translation-associated GTPase	ZP_02730857	39663	200	7	6
2 (6)	elongation factor Tu	ZP_02733080	47629	359	9	16.4
2 (6)	pyruvate dehydrogenase complex, dihydrolipoamide acetyltransferase E2 component	ZP_02735609	56948	483	18	17.4
2 (6)	ATP synthase F1, alpha subunit	ZP_02731443	62190	688	24	23.5
2 (6)	hypothetical protein GobsU_23532	ZP_02734797	61448	711	25	22.2
2 (3,6)	Flagellar FliF M-ring protein	ZP_02730987	54974	136	11	7.5
2 (6)	Ferredoxin	ZP_02732656	63129	523	30	20.2
2 (3,6)	hypothetical protein GobsU_04904	ZP_02731113	61669	881	47	29.3
2 (6)	sigma-54 dependent transcriptional regulator/response regulator	ZP_02732401	53500	754	31	30.5
2 (3)	hypothetical protein GobsU_28805	ZP_02735845	67828	775	35	25.4
2 (6)	RNA binding S1 domain protein	ZP_02737093	118263	588	24	12.7
2 (6)	sialic acid-specific 9-O-acetylesterase	ZP_02737096	57483	960	49	32
2 (6)	hypothetical protein GobsU_34522	ZP_02736980	59238	472	12	18.4
2 (6)	hypothetical protein GobsU_14549	ZP_02733018	54836	467	18	18.1
2 (3,6)	hypothetical protein GobsU_38147	ZP_02737694	46334	350	36	17.9
2 (6)	WD-40 repeat	ZP_02737050	72113	472	15	12
2 (6)	hypothetical protein GobsU_16609	ZP_02733426	56115	201	11	8.1
2 (6)	Signal Transduction Histidine Kinases (STHK)	ZP_02734429	58312	663	12	28.2
2 (6)	fumarate hydratase	ZP_02735455	54808	194	7	13.5
2 (6)	succinate dehydrogenase flavoprotein subunit	ZP_02737002	70751	370	20	14
2 (6)	heat shock protein GroEL	ZP_02737557	58429	167	6	9.9
2 (6)	probable chemotaxis transducer	ZP_02735177	59822	935	36	35.6
2 (6)	hypothetical protein GobsU_24941	ZP_02735076	56127	258	7	15.5
2 (6)	hypothetical protein GobsU_27341	ZP_02735553	61963	172	4	9.4
2 (3,6)	hypothetical protein GobsU_38668	ZP_02737797	58509	755	60	26.5
2 (6)	30S ribosomal protein S1	ZP_02730364	57995	729	27	28.6
2 (3,6)	30S ribosomal protein S1	ZP_02736868	69112	198	3	8.4
2 (6)	heat shock protein 70 family protein	ZP_02735562	66391	593	22	23.2
2 (6)	2-isopropylmalate synthase	ZP_02732952	57498	324	17	14.8
2 (6)	PrkA AAA domain protein	ZP_02733099	79001	1430	82	33.5
2 (6)	hypothetical protein GobsU_19384	ZP_02733973	47872	555	19	24.6
2 (6)	hypothetical protein GobsU_20458	ZP_02734187	63184	934	39	35
2 (6)	flagellin FliC	ZP_02737314	61923	1470	90	45
2 (6)	type IV fimbrial assembly protein PilB	ZP_02734151	63301	663	33	29.2
2 (6)	trigger factor	ZP_02733942	55785	1081	57	27
2 (6)	hypothetical protein GobsU_21120	ZP_02734319	66262	130	3	5.3
2 (6)	alkaline phosphatase	ZP_02732407	66257	277	8	11.2
2 (6)	60 kDa chaperonin	ZP_02736491	61375	679	33	26.6
2 (6)	hypothetical protein GobsU_30770	ZP_02736234	68086	306	10	10
2 (6)	glycyl-tRNA synthetase	ZP_02736496	63211	198	7	11.2
2	hypothetical protein GobsU_15760	ZP_02733259	66647	426	15	15.4
2	hypothetical protein GobsU_34902	ZP_02737056	63883	582	16	18.3
2 (3,6)	glutamine synthetase, catalytic region	ZP_02737578	79979	1309	78	37.1
2 (6)	threonyl-tRNA synthetase	ZP_02734724	70497	387	12	13.1
2 (3,6)	negative regulator of genetic competence ClpC/MecB	ZP_02733990	94501	870	76	20.9
2 (6)	hypothetical protein GobsU_30765	ZP_02736233	70604	199	9	7.5
2 (3)	hypothetical protein GobsU_38353	ZP_02737734	62324	223	10	10.5
2	hypothetical protein GobsU_07582	ZP_02731641	70264	237	12	11.2
2	hypothetical protein GobsU_35775	ZP_02737224	79841	274	11	8.9
2 (3,6)	elongation factor G	ZP_02735069	63641	276	9	10.3
2	hypothetical protein GobsU_17665	ZP_02733634	75700	295	11	8.3
2	probable chemotaxis sensory transducer	ZP_02731334	64063	296	7	12
2 (3)	Na-Ca exchanger/integrin-beta4	ZP_02733245	73354	579	20	24.3
2 (6)	GTP-binding elongation factor	ZP_02732217	67411	600	34	19.7
2	hypothetical protein GobsU_25226	ZP_02735133	29992	600	21	30.7
2	putative multi-domain protein	ZP_02734623	28378	544	65	39.4
2 (3)	50S ribosomal protein L25/general stress protein Ctc	ZP_02736396	23172	504	38	42.6
2	hypothetical protein GobsU_17625	ZP_02733626	26218	493	28	48
2 (6)	hypothetical protein GobsU_26266	ZP_02735339	27891	453	23	23
2 (3)	hypothetical protein GobsU_27941	ZP_02735673	28277	413	15	26.9
2	hypothetical protein GobsU_36335	ZP_02737336	32111	368	9	23.8
2	D-mannonate oxidoreductase	ZP_02732427	27806	354	21	23
2	hypothetical cytosolic protein	ZP_02733142	27103	353	15	25.2
2 (6)	hypothetical protein GobsU_30964	ZP_02736272	23254	311	11	29.5
2 (3,6)	30S ribosomal protein S4	ZP_02736977	22489	514	77	38.8

2	hypothetical protein GobsU_37972	ZP_02737659	31736	281	16	18.7
2	ABC transporter ATP-binding protein	ZP_02732495	26694	271	7	22.2
2	ABC transporter ATP-binding protein	ZP_02731083	71649	241	9	7.8
2	short-chain dehydrogenase/reductase SDR	ZP_02732138	23811	143	7	11.9
2	short-chain dehydrogenase/reductase SDR	ZP_02736353	27181	266	7	18.1
2	short-chain dehydrogenase/reductase SDR	ZP_02730741	30327	217	4	19.5
2 (3,6)	hypothetical protein GobsU_25221	ZP_02735132	31969	337	34	24.6
2	hypothetical protein GobsU_35950	ZP_02737259	31056	252	8	20.6
2	SSU ribosomal protein S9P	ZP_02737482	22532	244	24	29
2	Sulfate/thiosulfate import ATP-binding protein cysA	ZP_02731051	27037	236	10	26.7
2	dihydrodipicolinate reductase	ZP_02734700	28100	233	8	13.7
2	tryptophan synthase alpha chain	ZP_02731237	28505	265	10	23
2	NIPSNAP family containing protein	ZP_02732973	28688	286	14	21.7
2	ribosomal protein L9	ZP_02736392	20902	342	16	28.1
2 (6)	hypothetical protein GobsU_11080	ZP_02732339	35572	384	14	25.5
2	hypothetical protein GobsU_09481	ZP_02732020	34215	215	7	17.9
2 (3,6)	hypothetical protein GobsU_24726	ZP_02735033	29233	568	27	33.7
2	hypothetical protein GobsU_28900	ZP_02735864	29227	125	5	16.7
2	hypothetical protein GobsU_13877	ZP_02732890	35419	313	34	23.3
2	probable phosphoesterase	ZP_02734400	28951	118	4	18.5
2	hypothetical protein GobsU_10698	ZP_02732263	27768	115	8	14.7
2	phosphoesterase, PA-phosphatase related protein	ZP_02730579	58312	682	28	26.7
2 (3,6)	hypothetical protein GobsU_29154	ZP_02735914	33558	380	21	25.2
2	hypothetical protein GobsU_02788	ZP_02730695	33058	305	15	26.1
2 (3,6)	hypothetical protein GobsU_08407	ZP_02731806	36568	145	10	10.2
2	hypothetical protein GobsU_27841	ZP_02735653	63110	305	8	12.2
2	hypothetical protein GobsU_29553	ZP_02735993	26987	119	4	15.6
2	5-formyltetrahydrofolate cyclo-ligase	ZP_02732856	23328	115	3	14.5
2	3-ketoacyl-(acyl-carrier-protein) reductase	ZP_02735123	25082	156	4	11.3
2	translation initiation factor IF-3	ZP_02734237	22323	153	11	16.7
2	hypothetical protein GobsU_03729	ZP_02730882	20378	206	5	24.9
2	hypothetical protein GobsU_01552	ZP_02730453	31476	198	7	22.1
2	Protein kinase:GAF	ZP_02730508	73025	316	6	8.1
2	peptidase S9, prolyl oligopeptidase active site domain protein	ZP_02736956	110508	902	38	34
2	hypothetical protein GobsU_17520	ZP_02733605	32828	298	9	19.2
2	acetylglutamate kinase	ZP_02735209	31245	206	6	16.6
2	hypothetical protein GobsU_22402	ZP_02734571	25373	312	13	24.7
2	4Fe-4S ferredoxin iron-sulfur binding domain protein	ZP_02735878	27903	3217	479	48.3
2 (6)	chaperone protein HtpG	ZP_02733601	27919	176	17	13.1
2 (6)	ABC transporter (glutamine transport ATP-binding protein)	ZP_02735279	25792	160	5	21.7
2 (3)	ribosomal protein L17	ZP_02733782	20725	290	19	21.8
2 (3,6)	30S ribosomal protein S3	ZP_02734657	28381	107	3	16
2	acetolactate synthase III	ZP_02737021	64739	428	16	20.5
2 (3,6)	probable serine/threonine protein kinase related protein	ZP_02734134	84113	872	27	24.1
2 (3)	flagellar basal body rod protein	ZP_02731971	25399	107	4	18.5
2	ABC transporter, ATPase subunit	ZP_02733332	27639	436	10	24.3
2 (3)	hypothetical protein GobsU_27311	ZP_02735547	39432	209	6	12.2
2	glucose-1-phosphate thymidylyltransferase (strD)	ZP_02737087	24780	177	7	14.9
2 (6)	lipoprotein releasing system ATP-binding protein lolD	ZP_02734712	24120	168	5	21.9
2 (3)	hypothetical protein GobsU_26961	ZP_02735478	20743	303	10	26.4
2	thioredoxin peroxidase	ZP_02735335	22012	310	27	26.3
2 (3)	hypothetical protein GobsU_17515	ZP_02733604	21816	255	13	33.7
2 (3)	hypothetical protein GobsU_03025	ZP_02730742	21233	235	27	23.6
2 (3)	nitroreductase	ZP_02737361	23271	636	29	42.3
2	LexA repressor	ZP_02737915	26890	209	7	18.9
2	acetyl-CoA carboxylase (biotin carboxyl carrier subunit) accB	ZP_02731261	18550	170	5	25.9
2	ATP--cobalamin adenosyltransferase	ZP_02736319	21124	136	4	14.6
2	Dipeptidyl aminopeptidase	ZP_02736772	83036	522	20	14.5
2	ribonuclease E	ZP_02731873	108200	590	35	14.4
2	hypothetical protein GobsU_29289	ZP_02735941	68122	708	22	24.7
2	bifunctional sulfate adenylyltransferase subunit 1/adenylylsulfate kinase protein	ZP_02733159	70580	172	6	10.7
2	putative signal transduction protein	ZP_02734853	22925	196	12	37.6

2	multi-sensor hybrid histidine kinase	ZP_02734427	108654	244	6	3.9
2	multi-sensor hybrid histidine kinase	ZP_02732316	68813	299	8	14.5
2	multi-sensor hybrid histidine kinase	ZP_02731339	72445	250	8	8.4
2	multi-sensor hybrid histidine kinase	ZP_02737274	39793	204	6	14.8
2	probable sensor kinase	ZP_02735899	61961	374	14	18.9
2	phosphomannomutase	ZP_02732506	71596	305	13	16.9
2 (3)	hypothetical protein GobsU_29361	ZP_02735955	91983	738	43	19.5
2	2-oxoglutarate ferredoxin oxidoreductase alpha subunit	ZP_02733920	69353	680	37	25.7
2	hypothetical protein GobsU_12802	ZP_02732679	78483	626	35	19.9
2 (6)	probable secreted glycosyl hydrolase	ZP_02734952	155793	844	26	11.6
2	transcription initiation factor sigma 70	ZP_02737037	64210	386	11	17.1
2	hypothetical protein GobsU_11425	ZP_02732408	74604	533	31	16.5
2	probable adenylate cyclase	ZP_02735779	70674	498	17	17.8
2	hypothetical protein GobsU_06118	ZP_02731353	75215	476	13	18.1
2	heat shock protein 90	ZP_02737391	69763	456	18	16.2
2	signal transduction histidine kinase with CheB and CheR activity	ZP_02733916	66728	333	8	13.3
2	hypothetical protein GobsU_38503	ZP_02737764	68151	290	8	10.6
2	DNA gyrase subunit B	ZP_02732070	71463	275	15	14
2	translation initiation factor IF-2	ZP_02736305	124152	270	9	5.9
2	D-galactarate dehydratase/altronate hydrolase-like protein	ZP_02735035	55918	399	12	17.6
2	hypothetical protein GobsU_36582	ZP_02737385	59802	372	8	11.9
2	hypothetical protein GobsU_30335	ZP_02736147	58651	308	9	17.5
2	thiamine-phosphate pyrophosphorylase	ZP_02732066	52913	373	13	23.5
2	hypothetical protein GobsU_06705	ZP_02731470	57820	199	7	11
2	phosphoglycerate dehydrogenase	ZP_02733260	56312	371	11	13.2
2	two-component system sensory histidine kinase	ZP_02737190	58308	202	4	9.5
2	cetoacetate metabolism regulatory protein atoC	ZP_02733210	52917	328	10	15.1
2	cyanophycinase	ZP_02733322	57350	208	5	8.7
2 (3)	hypothetical protein GobsU_09903	ZP_02732104	49991	307	7	13.8
2	histidinol dehydrogenase	ZP_02736566	51372	292	8	19
2 (6)	probable auxin-responsive-like protein	ZP_02733619	63167	216	6	9.6
2	Amidase	ZP_02736331	59678	591	9	13
2 (3,6)	hypothetical protein GobsU_14664	ZP_02733041	37661	426	32	19.9
2 (3)	hypothetical protein GobsU_23747	ZP_02734840	47608	761	25	23
2 (3)	hypothetical protein GobsU_34987	ZP_02737073	41782	652	20	18
2 (3)	Collagen triple helix repeat	ZP_02737845	43510	329	9	6
2 (3)	hypothetical protein GobsU_33449	ZP_02736767	25070	98	3	13.6
2 (3)	hypothetical protein GobsU_39193	ZP_02737902	37851	226	7	15.4
2 (3,6)	hypothetical protein GobsU_20648	ZP_02734225	74282	309	11	10.3
2 (3)	hypothetical protein GobsU_28690	ZP_02735822	38321	316	8	14.8
2 (6)	hypothetical protein GobsU_29901	ZP_02736061	51041	287	6	6
2 (3)	probable tolQ protein	ZP_02733303	29569	171	12	12.3
2	protein translation elongation factor P (EF-P)	ZP_02732409	21479	188	17	23.6
2	sigma-24, ECF subfamily protein	ZP_02735550	19550	109	4	16.8
2 (3)	hypothetical protein GobsU_11650	ZP_02732451	35030	655	40	31.8
2 (3)	hypothetical protein GobsU_25326	ZP_02735153	23124	200	5	10.9
2 (6)	efflux transporter, RND family, MFP subunit	ZP_02732378	48939	226	8	10.7
2 (3)	hypothetical protein GobsU_34792	ZP_02737034	161460	595	20	16
2 (6)	DNA-directed RNA polymerase subunit alpha	ZP_02733781	37755	914	45	32.7
2 (6)	twitching mobility protein PilT	ZP_02734150	43895	513	13	23.5
2 (6)	twitching motility protein PilT	ZP_02733546	43722	353	8	13.9
2 (6)	hypothetical protein GobsU_26261	ZP_02735338	45236	207	6	9.2
2	probable NADH-dependent dehydrogenase	ZP_02734856	51921	194	6	7.9
2 (3)	hypothetical protein GobsU_16177	ZP_02733342	43879	548	15	31.5
2 (6)	Aldose 1-epimerase	ZP_02734020	41900	303	8	13.2
2 (6)	hypothetical protein GobsU_20243	ZP_02734144	41028	386	15	18.3
2 (3)	hypothetical protein GobsU_17361	ZP_02733574	145135	553	23	6.6
2	hypothetical protein GobsU_27771	ZP_02735639	36936	435	16	22.2
2 (3)	polysaccharide export protein	ZP_02736601	40175	346	15	17.5
2 (3)	hypothetical protein GobsU_32114	ZP_02736502	41942	238	9	10
2	hypothetical protein GobsU_17136	ZP_02733529	31873	222	7	13
2	GDP-mannose 4,6-dehydratase	ZP_02734112	37597	221	14	13.9
2 (6)	cytochrome c oxidase, subunit II	ZP_02733609	41126	816	36	36
2 (3,6)	hypothetical protein GobsU_12210	ZP_02732563	41758	104	3	6.2
2 (6)	hypothetical protein GobsU_14991	ZP_02733106	39437	639	18	33

2 (6)	hypothetical protein GobsU_20238	ZP_02734143	37132	413	21	23.1
2	hypothetical protein GobsU_30685	ZP_02736217	36258	288	7	21.1
2 (3)	hypothetical protein GobsU_25386	ZP_02735165	44361	272	8	10.2
2 (6)	PfkB domain protein	ZP_02737066	39152	260	7	12.7
2 (3,6)	30S ribosomal protein S2	ZP_02737376	26492	310	10	30.4
2	hypothetical protein GobsU_30580	ZP_02736196	36997	145	4	6.9
2 (6)	hypothetical protein GobsU_17530	ZP_02733607	26219	339	20	25.9
2	hypothetical protein GobsU_11145	ZP_02732352	39951	324	18	8.7
2	hypothetical protein GobsU_31734	ZP_02736426	37527	237	23	14.5
2 (6)	hypothetical protein GobsU_32284	ZP_02736536	34391	216	8	15.4
2 (6)	hypothetical protein GobsU_36654	ZP_02737399	34611	614	31	43
2 (3)	hypothetical protein GobsU_14649	ZP_02733038	36583	475	19	29.3
2 (6)	hypothetical protein GobsU_30670	ZP_02736214	35019	463	17	33.1
2 (3,6)	hypothetical protein GobsU_11730	ZP_02732467	38243	428	27	30.2
2	hypothetical protein GobsU_16057	ZP_02733318	33108	314	13	21.1
2 (3)	hypothetical protein GobsU_05341	ZP_02731198	34681	278	8	23
2 (3,6)	hypothetical protein GobsU_26271	ZP_02735340	33045	237	9	14.5
2	hypothetical protein GobsU_25211	ZP_02735130	31614	123	3	11.6
2 (3,6)	hypothetical protein GobsU_20248	ZP_02734145	35110	143	5	9.7
2	hypothetical protein GobsU_18200	ZP_02733741	34127	336	11	15.5
2	hypothetical protein GobsU_36425	ZP_02737354	31866	241	6	18
2	hypothetical protein GobsU_38508	ZP_02737765	34806	121	3	7.5
2 (3)	hypothetical protein GobsU_37727	ZP_02737610	34268	195	8	11.5
2 (3)	hypothetical protein GobsU_21390	ZP_02734373	28539	167	3	15.7
2	hypothetical protein GobsU_31319	ZP_02736343	30456	112	3	11.9
Summary for Fraction 2	**119 proteins are unique for this fraction (in green) – 44 % of total**	**34 proteins overlap with proteins from fractions 3 and 6**	**36 proteins overlap with proteins from fraction 3**	**82 proteins overlap with proteins from fraction 6**		**271 proteins in total**
Fraction 3 (Pore-containing membrane fraction; if protein found in other fractions it is indicated in brackets)	**Protein name**	**NCBI accession numbers**	**Calculated molecular weight (Da)***	**MOWSE score**	**Peptides matched**	**Sequence coverage (%)**
3 (2,6)	aconitate hydratase 1	ZP_02730459	98013	767	26	23.1
3 (2)	hypothetical protein GobsU_16177	ZP_02733342	43879	651	40	35.5
3 (2,6)	30S ribosomal protein S3	ZP_02734657	28381	250	3	19.8
3 (6)	50S ribosomal protein L5	ZP_02734651	20975	158	11	18
3 (2,6)	hypothetical protein GobsU_27236	ZP_02735532	35749	662	31	36.3
3 (2,6)	hypothetical protein GobsU_20248	ZP_02734145	35110	359	5	25.9
3 (2,6)	hypothetical protein GobsU_14664	ZP_02733041	37661	313	6	16.5
3 (2,6)	hypothetical protein GobsU_24726	ZP_02735033	29233	310	5	19.3
3 (2,6)	hypothetical protein GobsU_29154	ZP_02735914	33558	297	3	17.9
3 (2)	hypothetical protein GobsU_11650	ZP_02732451	35030	414	19	20.8
3 (6)	hypothetical protein GobsU_10668	ZP_02732257	29414	569	45	32.4
3 (2,6)	hypothetical protein GobsU_11730	ZP_02732467	38243	243	4	23
3 (2,6)	hypothetical protein GobsU_08407	ZP_02731806	36568	225	3	12.8
3 (6)	50S ribosomal protein L4	ZP_02734662	25590	210	3	17
3 (2)	hypothetical protein GobsU_37727	ZP_02737610	34268	178	3	11.1
3 (2)	hypothetical protein GobsU_14649	ZP_02733038	36583	166	3	11
3 (2)	hypothetical protein GobsU_05341	ZP_02731198	34681	160	3	16.6
3 (2,6)	50S ribosomal protein L1	ZP_02733084	30488	169	12	10
3 (2)	hypothetical protein GobsU_38353	ZP_02737734	62324	382	8	19.4
3 (2,6)	hypothetical protein GobsU_38668	ZP_02737797	58509	860	74	26.5
3 (2)	hypothetical protein GobsU_25326	ZP_02735153	23124	332	18	23.5
3 (2,6)	30S ribosomal protein S4	ZP_02736977	22489	472	25	28.6
3	hypothetical protein GobsU_26936	ZP_02735473	32285	524	7	36.1
3	probable polysaccharide export protein	ZP_02733516	30207	264	5	32.9
3 (2)	polysaccharide export protein	ZP_02736601	40175	549	24	30.1
3 (2)	hypothetical protein GobsU_25386	ZP_02735165	44361	697	10	33.6
3 (2,6)	30S ribosomal protein S2	ZP_02737376	26492	152	4	23.8
3 (6)	putative small-conductance mechanosensitive ion channel	ZP_02737296	96665	372	10	11.3
3 (2)	hypothetical protein GobsU_23747	ZP_02734840	47608	828	47	30
3	hypothetical protein GobsU_28460	ZP_02735776	43365	492	24	24.7
3	hypothetical protein GobsU_27906	ZP_02735666	43560	334	4	17.4
3 (2)	hypothetical protein GobsU_34987	ZP_02737073	41782	995	145	49.6
3	Na-Ca exchanger/integrin-beta4	ZP_02736670	43599	374	6	15.4

3	hypothetical protein GobsU_12560	ZP_02732631	49640	650	37	26.1
3	hypothetical protein GobsU_03779	ZP_02730892	39742	495	22	24.9
3	hypothetical protein GobsU_05311	ZP_02731192	42925	559	9	27
3 (2)	hypothetical protein GobsU_29361	ZP_02735955	91983	921	47	26.3
3 (2,6)	negative regulator of genetic competence ClpC/MecB	ZP_02733990	94501	893	36	25.8
3	hypothetical protein GobsU_06270	ZP_02731383	94980	1312	55	35.1
3	hypothetical protein GobsU_23427	ZP_02734776	96132	605	19	14.1
3	hypothetical protein GobsU_27926	ZP_02735670	21683	296	10	25.5
3 (6)	Scramblase family protein	ZP_02730790	21656	250	5	32.5
3 (2)	hypothetical protein GobsU_17515	ZP_02733604	21816	216	10	29.6
3 (2)	hypothetical protein GobsU_03025	ZP_02730742	21233	138	6	18.6
3	hypothetical protein GobsU_32159	ZP_02736511	106755	2423	182	42.7
3	hypothetical protein GobsU_17066	ZP_02733515	138451	896	25	16.4
3	hypothetical protein GobsU_09718	ZP_02732067	50348	660	17	25.8
3	probable divalent cation resistant determinant protein C	ZP_02731891	46316	471	11	19.5
3 (2,6)	oxidoreductase domain protein	ZP_02735697	47995	583	14	23.9
3 (2,6)	translation-associated GTPase	ZP_02730857	39663	417	19	19.6
3 (2)	Collagen triple helix repeat	ZP_02737845	43510	349	10	17.1
3 (2,6)	putative peptidase	ZP_02736310	42703	313	11	18.5
3	hypothetical protein GobsU_29613	ZP_02736005	41551	238	5	11.2
3 (2)	hypothetical protein GobsU_27941	ZP_02735673	28277	611	24	41.3
3 (2)	flagellar basal body rod protein FlgG	ZP_02731970	27437	495	29	29.7
3 (2)	flagellar basal body rod protein	ZP_02731971	25399	182	4	18.5
3 (2)	50S ribosomal protein L25/general stress protein Ctc	ZP_02736396	23172	262	8	24.5
3 (2,6)	hypothetical protein GobsU_20253	ZP_02734146	26677	233	9	16.2
3 (2,6)	hypothetical protein GobsU_25221	ZP_02735132	31969	157	8	12.6
3 (2)	hypothetical protein GobsU_32114	ZP_02736502	41942	638	33	32.5
3	hypothetical protein GobsU_33214	ZP_02736720	43019	186	6	11.8
3 (6)	hypothetical protein GobsU_01252	ZP_02730395	38353	198	6	14.7
3 (6)	hypothetical protein GobsU_20093	ZP_02734114	28967	213	6	17.3
3 (2)	hypothetical protein GobsU_26961	ZP_02735478	20743	356	12	34.2
3 (2)	nitroreductase	ZP_02737361	23271	307	8	31.2
3 (2)	hypothetical protein GobsU_33449	ZP_02736767	25070	115	4	13.6
3	hypothetical protein GobsU_36674	ZP_02737403	13718	108	4	10.4
3	hypothetical protein GobsU_22432	ZP_02734577	92982	788	82	16.3
3	hypothetical protein GobsU_18530	ZP_02733807	114144	566	36	13.8
3	autotransporter-associated beta strand repeat protein	ZP_02735782	111523	983	150	15.2
3 (2,6)	hypothetical protein GobsU_06435	ZP_02731416	103608	428	15	13.1
3 (6)	hypothetical protein GobsU_09596	ZP_02732043	81831	378	19	13.7
3 (2)	hypothetical protein GobsU_27311	ZP_02735547	39432	286	8	12.7
3 (6)	hypothetical protein GobsU_14884	ZP_02733085	19370	237	7	19.7
3 (2)	hypothetical protein GobsU_39193	ZP_02737902	37851	196	19	10.4
3 (6)	hypothetical protein GobsU_20228	ZP_02734141	177838	1152	38	18.9
3	hypothetical protein GobsU_01822	ZP_02730507	122792	1726	87	29.9
3	hypothetical protein GobsU_02172	ZP_02730573	81027	1136	55	32.4
3	hypothetical protein GobsU_28980	ZP_02735880	102196	1106	42	21.8
3 (2,6)	hypothetical protein GobsU_20643	ZP_02734224	92712	1032	48	26.2
3 (2,6)	hypothetical protein GobsU_31654	ZP_02736410	87810	939	33	23
3 (2)	Na-Ca exchanger/integrin-beta4	ZP_02733245	73354	813	23	29.5
3	peroxidase/catalase	ZP_02730801	86701	786	25	22.7
3 (2,6)	probable serine/threonine protein kinase related protein	ZP_02734134	84113	694	42	22.3
3 (6)	probable fimbrial assembly protein PilM	ZP_02734226	88149	684	23	19
3 (2,6)	hypothetical protein GobsU_29906	ZP_02736062	76943	609	17	24.7
3 (2,6)	hypothetical protein GobsU_20648	ZP_02734225	74282	591	20	19.2
3	hypothetical protein GobsU_22867	ZP_02734664	90756	422	12	17.3
3 (6)	polynucleotide phosphorylase/polyadenylase	ZP_02731046	80887	402	16	15.8
3 (2)	hypothetical protein GobsU_17361	ZP_02733574	145135	862	58	10.1
3 (6)	hypothetical protein GobsU_34877	ZP_02737051	82715	326	9	10.2
3	hypothetical protein GobsU_21510	ZP_02734397	37960	583	17	29.5
3	hypothetical protein GobsU_04229	ZP_02730982	28031	404	14	32.6
3 (2)	hypothetical protein GobsU_21390	ZP_02734373	28539	435	10	33.6
3 (2)	hypothetical protein GobsU_28690	ZP_02735822	38321	350	22	18.4
3 (6)	hypothetical protein GobsU_19324	ZP_02733961	107113	527	16	13.7
3 (2)	probable tolQ protein	ZP_02733303	29569	175	5	12.3
3	hypothetical protein GobsU_23637	ZP_02734818	254017	1257	74	7.9
3	FG-GAP repeat protein	ZP_02731030	163956	668	23	8.4
3	peptidase S8 and S53, subtilisin, kexin,	ZP_02737072	64444	432	7	13.6

	sedolisin					
3 (2,6)	hypothetical protein GobsU_12210	ZP_02732563	41758	229	7	11.7
3 (6)	flagellar motor switch protein G	ZP_02730986	32972	459	16	31.8
3 (2)	type IV fimbrial assembly protein PilB	ZP_02734149	70569	573	17	15.8
3 (2,6)	30S ribosomal protein S1	ZP_02736868	69112	191	7	8.3
3 (6)	50S ribosomal protein L3	ZP_02737246	32224	113	2	8.6
3 (2,6)	glutamine synthetase, catalytic region	ZP_02737578	79979	581	16	17.6
3 (2,6)	hypothetical protein GobsU_20233	ZP_02734142	48877	296	10	17.5
3 (2,6)	elongation factor G	ZP_02735069	63641	337	10	13
3 (2)	Uridylate kinase	ZP_02736756	26457	149	4	12
3	hypothetical protein GobsU_02177	ZP_02730574	73578	561	25	23.3
3 (2)	ribosomal protein L17	ZP_02733782	20725	217	3	11.7
3 (2)	hypothetical protein GobsU_34792	ZP_02737034	161460	1209	155	11.1
3 (2,6)	hypothetical protein GobsU_04904	ZP_02731113	61669	798	34	28.5
3	probable outer membrane lipoprotein IbeB	ZP_02736193	58744	1021	86	27.7
3 (2,6)	hypothetical protein GobsU_38147	ZP_02737694	46334	280	26	14
3	hypothetical protein GobsU_31139	ZP_02736307	52674	195	5	16.9
3 (2,6)	Flagellar FliF M-ring protein	ZP_02730987	54974	329	9	16.4
3 (2)	hypothetical protein GobsU_28805	ZP_02735845	67828	1291	122	38.2
3 (2)	hypothetical protein GobsU_09903	ZP_02732104	49991	817	29	32.4
3	hypothetical protein GobsU_13562	ZP_02732829	57915	595	45	20.9
3 (6)	30S ribosomal protein S7	ZP_02733090	17959	462	15	44.3
3 (6)	30S ribosomal protein S8	ZP_02734649	17113	334	9	52.6
3	UspA domain protein	ZP_02737803	13406	183	3	19.8
3	hypothetical protein GobsU_05481	ZP_02731226	17416	174	4	19.1
3 (6)	ribosomal protein L21	ZP_02733065	13902	134	4	24.6
3	hypothetical protein GobsU_31204	ZP_02736320	21793	96	4	15
3 (2,6)	hypothetical protein GobsU_26271	ZP_02735340	33045	361	13	24.8
3	hypothetical protein GobsU_09169	ZP_02731958	30276	294	6	13.4
Summary for Fraction 3	**39 proteins are unique for this fraction (in green)-30.5% of total**	**34 proteins overlap with proteins from fractions 2 and 6**	**36 proteins overlap with proteins from fraction 2**	**19 proteins overlap with proteins from fraction 6**		**128 proteins in total**
Fraction 6 (if protein found in other fractions it is indicated in brackets)	**Protein name**	**NCBI accession numbers**	**Calculated molecular weight (Da)***	**MOWSE score**	**Peptides matched**	**Sequence coverage (%)**
6 (2)	ATP synthase F1, alpha subunit	ZP_02731443	62190	2083	265	54.9
6 (2)	putative ABC transporter ATP-binding protein	ZP_02736693	34576	1208	93	67.9
6	Extracellular ligand-binding receptor	ZP_02731794	45810	1061	62	49.7
6	NuoF2 NADH I CHAIN F	ZP_02732659	50133	763	25	35.8
6 (2,3)	hypothetical protein GobsU_27236	ZP_02735532	35749	836	108	46.2
6 (2)	DNA-directed RNA polymerase beta chain	ZP_02733087	141275	2279	76	29.7
6	preprotein translocase subunit SecA	ZP_02731074	145943	1616	49	23.9
6 (2)	protein-export membrane protein secD	ZP_02733003	123971	2335	88	37.4
6	transporter, hydrophobe/amphiphile efflux-1 (HAE1) family protein	ZP_02734258	134298	1010	42	20.1
6 (2)	hypothetical protein GobsU_35960	ZP_02737261	103499	1007	41	21.8
6	hypothetical protein GobsU_31659	ZP_02736411	80096	754	26	23.4
6 (2)	cytochrome c oxidase, subunit II	ZP_02733609	41126	584	6	33.2
6	hypothetical protein GobsU_09963	ZP_02732116	33154	270	8	14.7
6 (2)	DNA-directed RNA polymerase beta chain	ZP_02733088	164212	2279	76	29.7
6 (3)	hypothetical protein GobsU_20228	ZP_02734141	177838	1868	95	26.9
6	hypothetical protein GobsU_32939	ZP_02736665	22754	121	4	13.7
6	ATP-binding protein	ZP_02731038	141780	916	27	20
6 (2,3)	hypothetical protein GobsU_06435	ZP_02731416	103608	1683	102	32.6
6 (3)	hypothetical protein GobsU_19324	ZP_02733961	107113	1500	53	33.5
6 (2)	probable secreted glycosyl hydrolase	ZP_02734952	155793	1256	36	18.7
6 (3)	putative small-conductance mechanosensitive ion channel	ZP_02737296	96665	1289	39	28.8
6 (2)	4Fe-4S ferredoxin, iron-sulfur binding domain protein	ZP_02733602	122387	3144	378	48.1
6	cyclic nucleotide-binding domain (cNMP-BD) protein	ZP_02734325	103286	1287	46	25.5
6 (2,3)	aconitate hydratase 1	ZP_02730459	98013	1003	35	29.3
6	pyruvate phosphate dikinase	ZP_02732609	96664	866	30	19.1
6	probable chaperone protein DnaK	ZP_02732821	102245	832	30	22.5
6	hypothetical protein GobsU_04604	ZP_02731053	94790	808	26	19.4
6 (3)	hypothetical protein GobsU_09596	ZP_02732043	81831	472	13	13.7

6	hypothetical protein GobsU_32934	ZP_02736664	70018	261	9	11.7
6 (2,3)	hypothetical protein GobsU_31654	ZP_02736410	87810	2886	229	52.7
6	Protease	ZP_02733734	90452	1453	61	36.2
6	hypothetical protein GobsU_36485	ZP_02737366	87089	1147	36	32.1
6 (2,3)	negative regulator of genetic competence ClpC/MecB	ZP_02733990	94501	1019	37	22.6
6 (3)	probable fimbrial assembly protein PilM	ZP_02734226	88149	601	18	13.7
6	acylglycerophosphoethanolamine acyltransferase	ZP_02736225	95762	471	13	13.8
6	tetratricopeptide TPR_2	ZP_02736723	78679	1022	35	31.3
6 (3)	hypothetical protein GobsU_34877	ZP_02737051	82715	918	28	26.2
6 (2,3)	elongation factor G	ZP_02735069	63641	767	23	27.4
6 (2,3)	probable serine/threonine protein kinase related protein	ZP_02734134	84113	465	13	15.1
6 (3)	polynucleotide phosphorylase/polyadenylase	ZP_02731046	80887	687	20	18.4
6	hypothetical protein GobsU_06560	ZP_02731441	20832	538	80	37.2
6	Thioredoxin peroxidase	ZP_02734436	18170	227	4	21.6
6	hypothetical protein GobsU_12710	ZP_02732661	14951	188	8	26.4
6	30S ribosomal protein S5	ZP_02734646	18356	118	7	21.1
6	hypothetical protein GobsU_05059	ZP_02731142	15020	625	61	49
6 (3)	ribosomal protein L21	ZP_02733065	13902	127	4	24.6
6	Redoxin domain protein	ZP_02732745	19451	130	8	11.6
6	heat shock protein, HSP20 family	ZP_02734270	16043	100	7	16.7
6 (3)	30S ribosomal protein S8	ZP_02734649	17113	388	35	42.8
6	hypothetical protein GobsU_27776	ZP_02735640	14871	300	33	35
6 (3)	30S ribosomal protein S7	ZP_02733090	17959	298	9	26.6
6	hypothetical protein GobsU_10558	ZP_02732235	17920	278	10	21
6	hypothetical protein GobsU_29743	ZP_02736031	19895	250	16	28.8
6	hypothetical protein GobsU_29149	ZP_02735913	16715	242	7	30.2
6	hypothetical protein GobsU_35850	ZP_02737239	15435	215	12	26.2
6	UspA domain protein	ZP_02737860	15996	186	11	32
6	50S ribosomal protein L16	ZP_02734656	17323	158	4	18.4
6	NADH (or F420H2) dehydrogenase, subunit C	ZP_02732663	19152	158	14	21.6
6	probable general stress protein 26	ZP_02736999	17931	154	8	18.3
6	hypothetical protein GobsU_09469	ZP_02732018	18718	129	6	15
6	hypothetical protein GobsU_15977	ZP_02733302	17748	88	2	10.7
6	hypothetical protein GobsU_24731	ZP_02735034	13764	290	62	41
6 (3)	hypothetical protein GobsU_10668	ZP_02732257	29414	321	25	26.3
6 (2,3)	hypothetical protein GobsU_14664	ZP_02733041	37661	299	8	14.8
6 (2,3)	hypothetical protein GobsU_20248	ZP_02734145	35110	215	5	13.1
6 (2)	oxidoreductase, short-chain dehydrogenase/reductase family protein	ZP_02733345	24580	405	13	26.6
6 (2,3)	50S ribosomal protein L1	ZP_02733084	30488	215	9	14.2
6 (2)	hypothetical protein GobsU_30670	ZP_02736214	35019	460	38	26.7
6 (2,3)	hypothetical protein GobsU_08407	ZP_02731806	36568	428	15	13.7
6	probable ABC-type transport system ATP-binding protein	ZP_02735300	31963	410	21	29.5
6	hypothetical protein GobsU_12240	ZP_02732569	33038	399	12	32
6 (2)	hypothetical protein GobsU_11080	ZP_02732339	35572	357	28	18.9
6 (2)	hypothetical protein GobsU_36654	ZP_02737399	34611	335	15	28.2
6 (2)	hypothetical protein GobsU_14991	ZP_02733106	39437	691	30	30.5
6 (2,3)	hypothetical protein GobsU_26271	ZP_02735340	33045	327	17	17.8
6	oxidoreductase, short-chain dehydrogenase/reductase family protein	ZP_02734437	42703	368	20	19.2
6	peptidylprolyl isomerase FKBP-type	ZP_02737067	31340	273	22	15
6 (3)	50S ribosomal protein L4	ZP_02734662	25590	263	7	20
6	probable transport ATP-binding protein	ZP_02737283	33853	250	6	14.6
6 (2,3)	hypothetical protein GobsU_11730	ZP_02732467	38243	243	22	16.7
6	hypothetical protein GobsU_26666	ZP_02735419	36271	404	16	21.2
6	hypothetical protein GobsU_36085	ZP_02737286	36120	186	3	12.9
6	Translation elongation factor Ts (EF-Ts)	ZP_02737377	30314	180	8	14.8
6	hypothetical protein GobsU_18952	ZP_02733887	35360	238	25	16.6
6	hypothetical protein GobsU_14996	ZP_02733107	26205	160	4	17.1
6 (2,3)	hypothetical protein GobsU_24726	ZP_02735033	29233	156	6	16.7
6 (2,3)	30S ribosomal protein S3	ZP_02734657	28381	143	8	12.5
6 (3)	hypothetical protein GobsU_20093	ZP_02734114	28967	116	2	11.4
6	ABC transporter, ATP-binding protein	ZP_02732799	35037	822	38	40.7
6	ABC transporter, ATP-binding protein	ZP_02737368	35182	617	33	35.5
6	ATP synthase gamma subunit	ZP_02731445	33052	1064	90	47.2
6 (2)	hypothetical protein GobsU_17530	ZP_02733607	26219	634	95	51
6	ABC transporter, ATPase subunit	ZP_02736275	34092	433	15	25.8

6 (3)	50S ribosomal protein L3	ZP_02737246	32224	615	30	32.5
6 (2)	hypothetical protein GobsU_14709	ZP_02733050	37118	339	35	25.1
6	hypothetical protein GobsU_35638	ZP_02737197	34223	272	8	13.5
6	probable protein kinase yloP	ZP_02730388	33059	474	15	34.6
6 (2)	hypothetical protein GobsU_32284	ZP_02736536	34391	263	10	21.5
6	putative serine/threonine-protein kinase	ZP_02735907	35707	249	10	13.7
6	hypothetical protein GobsU_37737	ZP_02737612	38778	223	6	10
6	hypothetical protein GobsU_13797	ZP_02732874	29652	214	10	15
6	Squalene/phytoene synthase	ZP_02731127	34549	208	7	12.6
6	Ribose transporter, periplasmic binding protein	ZP_02736325	35409	1030	73	51.2
6 (2)	F0F1 ATP synthase subunit beta	ZP_02731447	52132	1133	40	51.1
6	hypothetical protein GobsU_05738	ZP_02731277	31560	189	8	16.6
6	hypothetical protein GobsU_19803	ZP_02734056	30577	584	36	38.5
6	periplasmic solute binding protein	ZP_02732461	34514	627	24	33.8
6	ferrochelatase	ZP_02735861	36562	183	8	13.7
6	succinic semialdehyde dehydrogenase	ZP_02736639	27647	235	6	20.3
6 (2)	cobalt-zinc-cadmium resistance protein	ZP_02733511	54820	512	14	24.3
6	formate dehydrogenase beta subunit	ZP_02733206	35765	127	4	11
6 (2,3)	30S ribosomal protein S2	ZP_02737376	26492	319	14	32.9
6	hypothetical protein GobsU_18957	ZP_02733888	29436	106	4	10
6	Alcohol dehydrogenase, zinc-binding domain protein	ZP_02737665	35836	749	23	42.6
6	putative zinc ABC transporter, zinc-binding protein	ZP_02736154	36873	398	20	18.7
6 (2)	PfkB domain protein	ZP_02737066	39152	614	7	34.9
6 (2)	hypothetical protein GobsU_20238	ZP_02734143	37132	461	4	29.2
6 (2,3)	hypothetical protein GobsU_12210	ZP_02732563	41758	454	5	25.1
6	Peptidase S1 and S6, chymotrypsin/Hap	ZP_02736370	44435	259	7	16.6
6	MoxR-related ATPase, AAA superfamily protein	ZP_02732960	36904	250	7	16.7
6 (3)	hypothetical protein GobsU_01252	ZP_02730395	38353	387	5	15.8
6	phosphate ABC transporter, substrate-binding protein PstS	ZP_02733764	37171	347	2	20.1
6 (3)	flagellar motor switch protein G	ZP_02730986	32972	237	3	20.4
6	UDP-glucose 4-epimerase	ZP_02736901	36518	209	6	12.6
6	hypothetical protein GobsU_23012	ZP_02734693	34897	208	6	11
6	hypothetical protein GobsU_29321	ZP_02735947	36283	193	8	10.7
6	sulfate-binding protein precursor	ZP_02733577	37509	192	7	10
6	ROK family protein	ZP_02735314	34351	181	2	9.9
6	hypothetical protein GobsU_19329	ZP_02733962	36927	180	2	11.1
6	Alcohol dehydrogenase	ZP_02737751	35170	177	2	16.5
6	hypothetical protein GobsU_39208	ZP_02737905	37781	640	26	26.4
6	type IV fimbrial assembly protein PilC	ZP_02734147	45854	591	28	24.4
6	2-hydroxyglutarate dehydrogenase	ZP_02730858	43266	157	7	12.1
6	branched-chain amino acid transport ATP-binding protein	ZP_02731797	41111	172	6	8.8
6 (2,3)	hypothetical protein GobsU_20233	ZP_02734142	48877	605	28	25.7
6	hypothetical protein GobsU_34582	ZP_02736992	44739	180	5	9.4
6	ABC transporter related protein	ZP_02736413	36171	569	43	37.2
6	glycosyl transferase, group 1 family protein	ZP_02730883	36479	331	6	16.8
6	glycosyl transferase, group 1 family protein	ZP_02730879	40001	224	8	10.9
6	glycosyl transferase, group 1 family protein	ZP_02730886	42261	167	3	10.3
6	glycosyl transferase, group 1	ZP_02730871	41119	626	22	38.8
6	glycosyl transferase, group 1	ZP_02730872	39488	280	10	21.5
6	glycosyl transferase, group 1	ZP_02731829	44120	311	7	18.4
6	glycosyl transferase, group 2 family protein	ZP_02736170	34002	901	47	53.9
6	glycosyl transferase, group 2 family protein	ZP_02735160	28363	440	12	31.5
6	glycosyl transferase family 2	ZP_02732222	32359	235	13	20.1
6	glycosyl transferase family 2	ZP_02730876	39841	131	5	13.4
6 (2)	Sulphate transport system permease protein 1	ZP_02732762	39652	489	15	26.2
6 (2)	ketol-acid reductoisomerase	ZP_02737871	38760	350	10	18.6
6 (2,3)	putative peptidase	ZP_02736310	42703	734	30	35.4
6 (2)	hypothetical protein GobsU_20243	ZP_02734144	41028	478	18	26.5
6	HlyD family secretion protein, putative	ZP_02733129	37920	343	8	16.4
6 (2,3)	hypothetical protein GobsU_20643	ZP_02734224	92712	750	25	19.2
6	hypothetical protein GobsU_19334	ZP_02733963	39693	299	10	17.9
6 (2)	Aldose 1-epimerase	ZP_02734020	41900	296	7	14.5
6 (2)	DNA-directed RNA polymerase subunit alpha	ZP_02733781	37755	728	39	25.6
6	hypothetical protein GobsU_22142	ZP_02734523	40583	278	12	12.1

6	probable oxidoreductase	ZP_02734812	37679	219	8	15.4
6	FMN-dependent alpha-hydroxy acid dehydrogenase	ZP_02735804	43337	554	18	32.2
6	glyceraldehyde-3-phosphate dehydrogenase	ZP_02733850	37727	218	9	11.4
6	3-beta hydroxysteroid dehydrogenase/isomerase	ZP_02737474	35702	220	6	13.1
6	hypothetical protein GobsU_14152	ZP_02732939	44295	322	10	18.9
6 (2)	hypothetical protein GobsU_29901	ZP_02736061	51041	527	18	20.1
6 (2)	hypothetical protein GobsU_04644	ZP_02731061	47023	233	8	11.3
6	probable Zn-dependent alcohol dehydrogenase	ZP_02734440	42880	372	17	19
6	muconate cycloisomerase	ZP_02733879	43181	274	9	13.8
6	succinyl-CoA synthetase (beta subunit)	ZP_02732131	41963	418	12	17.9
6 (2,3)	oxidoreductase domain protein	ZP_02735697	47995	777	36	32.8
6	acriflavine resistance protein A	ZP_02731890	42775	1106	90	45
6	transaldolase	ZP_02736346	38880	632	20	37.8
6 (2)	serine proteinase, HtrA/DegQ/DegS family	ZP_02735252	42708	216	5	10.9
6 (2)	hypothetical protein GobsU_26261	ZP_02735338	45236	530	15	26.2
6	amine oxidase, flavin-containing	ZP_02732954	43795	538	6	32.6
6	xylose isomerase	ZP_02733161	49224	207	6	10.1
6	hypothetical protein GobsU_33144	ZP_02736706	44110	191	5	10.5
6 (2,3)	translation-associated GTPase	ZP_02730857	39663	193	6	10.4
6	D-amino acid dehydrogenase, small chain	ZP_02737095	45047	192	7	9
6	YcjX-like protein	ZP_02732078	51480	512	21	24.8
6	secretion protein HlyD	ZP_02732105	50180	403	13	16.1
6	hypothetical protein GobsU_13427	ZP_02732802	37480	155	6	10.8
6	oxidoreductase	ZP_02731486	44071	157	4	11.2
6	hypothetical protein GobsU_31434	ZP_02736366	50078	219	7	11.7
6 (2)	elongation factor Tu	ZP_02733080	47629	1210	14	55.1
6	probable protein phosphatase 1	ZP_02730410	47515	755	41	40.3
6	efflux transporter, RND family, MFP subunit	ZP_02734533	48327	391	12	15.8
6	efflux transporter, RND family, MFP subunit	ZP_02735371	43840	219	9	14
6	NADH dehydrogenase subunit D	ZP_02732662	46011	549	7	24.4
6	FAD-dependent pyridine nucleotide-disulphide oxidoreductase	ZP_02734627	47231	687	32	29.9
6	dihydrolipoamide dehydrogenase	ZP_02731257	50105	406	13	23.4
6	hypothetical protein GobsU_20203	ZP_02734136	44521	450	13	26.1
6	Enolase	ZP_02736401	46429	468	18	23.8
6 (2)	twitching mobility protein PilT	ZP_02734150	43895	507	16	30.2
6	Catalase domain protein	ZP_02733351	39161	509	39	28.1
6 (2)	Cobalamin synthesis protein/P47K	ZP_02730707	41940	317	3	17.6
6	hypothetical protein GobsU_33069	ZP_02736691	88468	1546	88	43.8
6 (2,3)	hypothetical protein GobsU_29906	ZP_02736062	76943	1018	33	31.5
6	cell division protein FtsH	ZP_02737094	76041	502	24	20.1
6	dihydrolipoamide acetyltransferase	ZP_02733334	43017	414	15	20.4
6 (2)	pyruvate dehydrogenase complex, dihydrolipoamide acetyltransferase E2 component	ZP_02735609	56948	642	21	21.2
6 (2)	twitching motility protein PilT	ZP_02733546	43722	391	16	27.8
6 (2)	6-phosphogluconate dehydrogenase	ZP_02734911	52647	392	12	17.3
6	acyl-CoA dehydrogenase domain protein	ZP_02734570	41522	273	9	12.8
6	phosphoglycerate kinase	ZP_02733849	42098	255	7	11.9
6	probable tetraacyldisaccharide 4-kinase	ZP_02736188	36745	242	7	18.2
6	hypothetical protein GobsU_07767	ZP_02731678	52109	754	29	29.4
6	hypothetical protein GobsU_29508	ZP_02735984	53183	657	30	31.2
6 (2)	hypothetical protein GobsU_23532	ZP_02734797	61448	1390	66	37.7
6	type I phosphodiesterase/nucleotide pyrophosphatase	ZP_02737031	50667	567	19	25.9
6	putative auxin-regulated protein	ZP_02734402	58942	963	35	34
6 (2)	probable auxin-responsive-like protein	ZP_02733619	63167	158	4	6.9
6	proton-dependent oligopeptide transporter family protein	ZP_02734743	64966	266	9	13.4
6	C-terminal processing peptidase S41A	ZP_02737571	25174	268	8	13.4
6	hypothetical protein GobsU_20208	ZP_02734137	52065	316	21	16
6 (2,3)	Flagellar FliF M-ring protein	ZP_02730987	54974	693	26	26
6	hypothetical protein GobsU_01882	ZP_02730517	49589	253	15	10
6 (2)	Ferredoxin	ZP_02732656	63129	1306	71	44.7
6 (2,3)	hypothetical protein GobsU_04904	ZP_02731113	61669	596	19	20.8
6	hypothetical protein GobsU_23022	ZP_02734695	57316	392	11	15.6
6 (2)	sigma-54 dependent transcriptional	ZP_02732401	53500	572	14	23.2

	regulator/response regulator					
6	FAD dependent oxidoreductase	ZP_02735201	58357	531	17	25.3
6 (2)	RNA binding S1 domain protein	ZP_02737093	118263	500	18	11.2
6	Acyl-CoA dehydrogenase	ZP_02732342	65323	1571	120	39
6 (2)	sialic acid-specific 9-O-acetylesterase	ZP_02737096	57483	497	14	15.9
6 (2)	hypothetical protein GobsU_34522	ZP_02736980	59238	443	12	16.2
6 (2)	hypothetical protein GobsU_14549	ZP_02733018	54836	443	12	20.2
6	hypothetical protein GobsU_06520	ZP_02731433	50596	338	12	13.6
6 (2,3)	hypothetical protein GobsU_38147	ZP_02737694	46334	580	36	23.1
6 (2)	WD-40 repeat	ZP_02737050	72113	870	47	22.3
6 (2)	hypothetical protein GobsU_16609	ZP_02733426	56115	758	54	27.3
6	hypothetical protein GobsU_25556	ZP_02735199	62055	691	36	20.8
6	Phytoene dehydrogenase and related protein-like protein	ZP_02736586	50255	207	5	7.4
6	hypothetical protein GobsU_28340	ZP_02735752	46838	296	10	16
6	probable PbrT protein-possibly cytochrome c	ZP_02733592	52631	312	10	12.1
6 (2)	probable DNA-directed RNA polymerase alpha chain	ZP_02736955	51701	322	7	10.7
6 (2)	Signal Transduction Histidine Kinases (STHK)	ZP_02734429	58312	495	13	16.4
6 (2)	fumarate hydratase	ZP_02735455	54808	438	17	19.8
6 (2)	succinate dehydrogenase flavoprotein subunit	ZP_02737002	70751	1198	54	32.2
6 (2)	heat shock protein GroEL	ZP_02737557	58429	348	11	17.4
6 (2)	probable chemotaxis transducer	ZP_02735177	59822	318	7	16.7
6 (2)	hypothetical protein GobsU_24941	ZP_02735076	56127	299	10	13.8
6	probable protein kinase yloP-putative serine/threonine protein kinase	ZP_02733729	56940	247	9	15.4
6 (2)	hypothetical protein GobsU_27341	ZP_02735553	61963	212	7	8.3
6	hypothetical protein GobsU_31909	ZP_02736461	59664	209	8	12
6 (2,3)	hypothetical protein GobsU_38668	ZP_02737797	58509	382	16	13.9
6 (2)	30S ribosomal protein S1	ZP_02730364	57995	218	8	6.8
6 (2,3)	30S ribosomal protein S1	ZP_02736868	69112	803	38	28.6
6 (2)	heat shock protein 70 family protein	ZP_02735562	66391	798	26	31.6
6 (2)	2-isopropylmalate synthase	ZP_02732952	57498	188	7	10.4
6 (2)	PrkA AAA domain protein	ZP_02733099	79001	1058	31	29.7
6 (2)	hypothetical protein GobsU_19384	ZP_02733973	47872	849	37	31.5
6 (2)	hypothetical protein GobsU_20458	ZP_02734187	63184	674	21	23.8
6 (2)	flagellin FliC	ZP_02737314	61923	1272	59	40.6
6	delta-1-pyrroline-5-carboxylate dehydrogenase	ZP_02735812	112374	1133	35	23.3
6	peptidase S45 penicillin amidase	ZP_02735855	85148	348	11	9.6
6	probable NADH-dependent dehydrogenase	ZP_02731432	63505	917	37	28.8
6	60 kDa chaperonin 5	ZP_02736489	60169	501	22	20.3
6 (2)	type IV fimbrial assembly protein PilB	ZP_02734151	63301	917	37	28.8
6 (2)	trigger factor	ZP_02733942	55785	443	14	16.3
6	hypothetical protein GobsU_23972	ZP_02734885	20528	245	7	30.8
6 (2)	hypothetical protein GobsU_21120	ZP_02734319	66262	192	6	8.1
6 (2)	alkaline phosphatase	ZP_02732407	66257	1360	50	44.5
6 (2)	60 kDa chaperonin	ZP_02736491	61375	250	8	10
6 (2)	hypothetical protein GobsU_30770	ZP_02736234	68086	685	21	22.7
6 (2)	glycyl-tRNA synthetase	ZP_02736496	63211	333	11	16.2
6 (2,3)	glutamine synthetase, catalytic region	ZP_02737578	79979	1721	88	43.2
6	NADH dehydrogenase (quinone)	ZP_02734324	76409	1020	41	25.5
6	probable signal peptidase I	ZP_02732748	68998	629	24	18.6
6 (2)	threonyl-tRNA synthetase	ZP_02734724	70497	461	12	14.7
6	hypothetical protein GobsU_08402	ZP_02731805	68867	369	8	13.7
6 (2)	hypothetical protein GobsU_30765	ZP_02736233	70604	580	23	16
6 (2)	GTP-binding elongation factor	ZP_02732217	67411	192	6	7.7
6 (2,3)	hypothetical protein GobsU_29154	ZP_02735914	33558	158	4	8.6
6 (3)	50S ribosomal protein L5	ZP_02734651	20975	133	3	13.1
6 (3)	Scramblase family protein	ZP_02730790	21656	388	6	39.4
6 (2,3)	hypothetical protein GobsU_20648	ZP_02734225	74282	940	29	23.7
6	hypothetical protein GobsU_38418	ZP_02737747	14344	312	22	39
6	probable ribosomal protein S6	ZP_02736394	20250	107	2	15.4
6	probable 30S ribosomal protein S17	ZP_02734654	13764	145	9	16.3
6	hypothetical protein GobsU_33074	ZP_02736692	89212	299	7	9.1
6 (2)	methylcitrate synthase	ZP_02732426	42530	164	6	9.2
6	probable ATP synthase CF1 subunit e	ZP_02731448	14248	446	37	61.2
6 (2)	hypothetical protein GobsU_30964	ZP_02736272	23254	382	9	30.4
6	hypothetical protein GobsU_11645	ZP_02732450	14806	261	12	41.3

6	hypothetical protein GobsU_36060	ZP_02737281	15530	93	3	12.2
6	hypothetical protein GobsU_21075	ZP_02734310	16557	248	10	38.3
6 (3)	hypothetical protein GobsU_14884	ZP_02733085	19370	236	7	19.7
6	hypothetical protein GobsU_35800	ZP_02737229	13495	233	12	18
6 (2)	hypothetical protein GobsU_11415	ZP_02732406	14995	220	30	27.5
6	hypothetical protein GobsU_05084	ZP_02731147	15541	209	10	24.3
6	hypothetical protein GobsU_26206	ZP_02735327	17561	161	5	20.3
6	HflC protein	ZP_02732735	37784	152	6	8.5
6	hypothetical protein GobsU_27231	ZP_02735531	12326	150	2	14.8
6	hypothetical protein GobsU_36340	ZP_02737337	14636	149	4	25.5
6	probable anti-anti-sigma regulatory factor (antagonist of anti-sigma factor)	ZP_02734268	13732	146	14	25.6
6	hypothetical protein GobsU_29040	ZP_02735892	10640	112	4	26.3
6 (2)	hypothetical protein GobsU_17456	ZP_02733593	12531	103	4	24.6
6	riboflavin synthase subunit beta	ZP_02732766	16302	171	6	20.8
6	hypothetical protein GobsU_22822	ZP_02734655	16991	101	7	17.4
6	hypothetical protein GobsU_30195	ZP_02736119	10341	99	2	20.4
6	hypothetical protein GobsU_14469	ZP_02733002	11672	91	3	13.9
6 (2)	multidrug efflux system, HlyD family subunit	ZP_02734259	47621	476	29	18
6 (2)	efflux transporter, RND family, MFP subunit	ZP_02732378	48939	302	19	11
6	hypothetical protein GobsU_21140	ZP_02734323	59330	216	6	6.4
6	hypothetical protein GobsU_18305	ZP_02733762	57398	143	8	9.3
6	Molybdopterin oxidoreductase, iron-sulfur binding subunit	ZP_02736517	125542	1687	64	32.1
6 (2)	hypothetical protein GobsU_34912	ZP_02737058	127649	1168	35	20.6
6	GAF sensor hybrid histidine kinase	ZP_02735634	125815	478	12	8
6	transporter, hydrophobe/amphiphile efflux-1 (HAE1) family protein	ZP_02732379	116833	337	10	5.5
6	hypothetical protein GobsU_12857	ZP_02732690	106371	495	11	10.3
6	cation efflux system protein CZCA	ZP_02733510	111095	299	10	6.1
6	DNA gyrase subunit A	ZP_02737492	96932	311	10	7.2
6 (2,3)	hypothetical protein GobsU_20253	ZP_02734146	26677	426	14	39.1
6 (2,3)	30S ribosomal protein S4	ZP_02736977	22489	379	15	23.5
6 (2)	ABC transporter (glutamine transport ATP-binding protein)	ZP_02735279	25792	180	6	21.7
6 (2)	hypothetical protein GobsU_26266	ZP_02735339	27891	161	8	13.3
6 (2,3)	hypothetical protein GobsU_25221	ZP_02735132	31969	150	4	12.3
6 (2)	chaperone protein HtpG	ZP_02733601	27919	138	9	11.8
6 (2)	lipoprotein releasing system ATP-binding protein lolD	ZP_02734712	24120	94	3	12.5
6	putative serine protease containing two PDZ domains	ZP_02734485	31120	116	3	7.3
Summary for Fraction 6	**184 proteins are unique for this fraction (in green) – 57.7% of total**	**34 proteins overlap with proteins from fractions 2 and 3**	**82 proteins overlap with proteins from fraction 2**	**19 proteins overlap with proteins from fraction 3**		**319 proteins in total**

* as calculated by MASCOT using WAL-1 draft genome project. NCBI data may differ from those given by WAL-1.

S2 Table. Summary of the membrane proteome analysis.

fraction	total number of proteins	proteins with predicted function	hypothetical proteins (unknown function)	unique proteins in fraction
2	271	150 (55%)	121 (45%)	119 (44%)
3 (pore fraction)	128	44 (34%)	84 (66%)	39 (30.5%)
6	319	193 (61%)	126 (39%)	184 (58%)

The proteins identified by mass-spectrometry are grouped according to the NCBI annotation as hypothetical proteins with no predicted function or with firmly predicted function (such as ribosomal proteins, ABC transporters etc).

Page 1

S4 Table. Results from structural analysis for the C-terminal region of cluster 1 (β-propeller) protein constituents*

ID	Fraction	Confidence	Coverage	PDB template
ZP_02731030	(3)	16.0%	1%	3IKM
ZP_02731113	(2,3,6)	99.9	52%	2C4D
ZP_02733245	(2,3)	99.9%	42%	2C4D
ZP_02734577	(3)	98.6%	25%	2C4D
ZP_02734776	(3)	99.9%	25%	2C4D
ZP_02734818	(3)	40.1%	5%	3FCS
ZP_02735782	(3)	1.6%	9%	3RB7
ZP_02736670	(3)	100%	76%	2C4D
ZP_02737072	(3)	99.8	39%	2C4D
ZP_02737073	(2,3)	100%	91%	2C4D
ZP_02737797	(2,3,6)	99.9%	49%	2C4D

*Models generated from full sequences except ZP_02734818 (see text). Best hits (shown) were chosen based on combined top confidence and coverage scores, and location in the alignable C-terminus (Figure supplement 13). Column 1, Genbank accessions. Columns 2 and 3, confidence and coverage scores as generated by Phyre2. Column 4, PDB template ID used by Phyre2 to generate models. Results below 95% confidence were not further analysed.

S5 Table. **Results from structural analysis of cluster 2 (pili) protein constituents***

ID	Fraction	Confidence	Coverage	PDB template
ZP_02731198	(2,3)	99.7%	21%	1OQW
ZP_02731806	(2,3,6)	99.6%	23%	1OQW
ZP_02732451	(2,3)	99.7%	24%	1OQW
ZP_02732467	(2,3,6)	99.7%	23%	1OQW
ZP_02733038	(2,3)	99.6%	25%	1OQW
ZP_02733041	(2,3,6)	99.7%	27%	1OQW
ZP_02735033	(2,3,6)	54.8%	8%	2PIL
ZP_02735132	(2,3,6)	99.7%	27%	1OQW
ZP_02735532	(2,3,6)	99.6%	20%	1OQW
ZP_02735914	(2,3,6)	99.7%	26%	1OQW
ZP_02737610	(2,3)	99.7%	26%	1OQW

*Models generated from full sequences. Best hits (shown) were chosen based on combined top confidence and coverage scores. Column 1, Genbank accessions. Columns 2 and 3, confidence and coverage scores as generated by Phyre2. Column 4, PDB template ID used by Phyre2 to generate models.

S6 Table: Results from keyword analysis of Phyre output.

ID Fraction Confidence Coverage PDB template

outer membrane lipoprotein

ID	Fraction	Confidence	Coverage	PDB template
ZP_02730892	(3)	100.00%	95.00%	2J58
ZP_02731226	(3)	77.00%	29.00%	1JCC
ZP_02733516	(3)	100.00%	75.00%	2J58
ZP_02736601	(2,3)	100.00%	70.00%	2J58

outer membrane efflux protein

ID	Fraction	Confidence	Coverage	PDB template
ZP_02730507	(3)	99.00%	67.00%	1WP1
ZP_02731891	(3)	100.00%	91.00%	1WP1
ZP_02732067	(3)	99.70%	91.00%	1WP1
ZP_02732104	(2,3)	100.00%	95.00%	1WP1
ZP_02732829	(3)	100.00%	88.00%	1WP1
ZP_02735955	(2,3)	100.00%	62.00%	1WP1
ZP_02736193	(3)	100.00%	89.00%	1WP1

transmembrane beta barrel

ID	Fraction	Confidence	Coverage	PDB template
ZP_02730574	(3)	99.90%	58.00%	3RBH
ZP_02731192	(3)	98.30%	53.00%	1I78
ZP_02732104	(2,3)	5.30%	9.00%	1QJ8
ZP_02732631	(3)	99.40%	78.00%	2O4V
ZP_02734397	(3)	96.80%	52.00%	1I78
ZP_02734840	(2,3)	97.40%	46.00%	1I78
ZP_02735776	(3)	97.90%	60.00%	2X27
ZP_02735845	(2,3)	100.00%	64.00%	3RBH
ZP_02735880	(3)	97.60%	11.00%	1T16
ZP_02735955	(2,3)	38.80%	6.00%	2MPR

porin

ID	Fraction	Confidence	Coverage	PDB template
ZP_02730574	(3)	98.60%	64.00%	3SYB
ZP_02731192	(3)	20.80%	9.00%	2O4V
ZP_02732631	(3)	99.40%	78.00%	2O4V
ZP_02734397	(3)	36.80%	61.00%	1T16
ZP_02734840	(2,3)	50.60%	58.00%	1T16
ZP_02735776	(3)	76.20%	34.00%	2WJQ
ZP_02735845	(2,3)	100.00%	64.00%	2Y0K
ZP_02735880	(3)	97.60%	11.00%	1T16
ZP_02735955	(2,3)	38.80%	6.00%	2MPR

tolc

ID	Fraction	Confidence	Coverage	PDB template
ZP_02730507	(3)	99.90%	61.00%	1TQQ

ZP_02731891	(3)	100.00%	89.00%	1TQQ
ZP_02732067	(3)	98.40%	82.00%	1TQQ
ZP_02732104	(2,3)	100.00%	99.00%	1TQQ
ZP_02732829	(3)	100.00%	80.00%	1TQQ
ZP_02735955	(2,3)	100.00%	60.00%	1TQQ
ZP_02736193	(3)	100.00%	77.00%	1TQQ

oprX

ID	Fraction	Confidence	Coverage	PDB template	Class
ZP_02730574	(3)	97.10%	61.00%	2ODJ	oprd
ZP_02731192	(3)	96.50%	48.00%	2X27	oprg
ZP_02732631	(3)	99.40%	78.00%	2O4V	oprp
ZP_02734397	(3)	91.00%	40.00%	2X27	oprg
ZP_02734840	(2,3)	58.20%	39.00%	2LHF	oprh
ZP_02735776	(3)	97.90%	60.00%	2X27	oprg
ZP_02735845	(2,3)	99.90%	64.00%	2ODJ	oprd
ZP_02735880	(3)	78.90%	5.00%	2X27	oprg
ZP_02737902	(2,3)	5.60%	10.00%	2LHF	oprh

All the structural models generated by Phyre for the 128 fraction 3 proteins were screened for hits to bacterial transmembrane proteins using the keywords showed. A protein can have hits to more than one of these transmembrane protein classes. For each protein, we list the ID, the membrane fraction containing the protein, the confidence as calculated by Phyre2, the coverage of the sequence, and the PDB template used in the model. For the oprX proteins we also list the specific class. Bold entries show proteins that were clustered together as a duo in the clustering analysis.

S Text

Bioinformatics analyses

We performed a number of bioinformatics analyses on the 128 unique proteins identified through proteomics as belonging to the pore-containing membrane fraction (fraction 3). We first searched for similarity to known proteins in the non-redundant protein database (downloaded from NCBI) using BLASTP and PHMMER.

BLASTP (2) reported 112 proteins with significant homologs (E<0.001) (S3 Table). Of these, 33 are in the membrane fraction 3 (84.6% of all fraction 3 proteins). Of the significant hits, 25 top hits are to non-identical sequences in *G. obscuriglobus*, 18 are to *Planctomyces*, and 16 to *S. acidiphila*. One sequence found no hits at all (ZP_02735547). Almost half of all hits are to hypothetical (50), unnamed (5) or probable (2) proteins. Among hits with assigned function, 6 are flagellar, 12 ribosomal, 6 are membrane-related, and 4 are efflux proteins.

PHMMER (from the HMMER package, ver. 3.0(1)), returned significant hits for 115 proteins (E<0.001) (S3 Table), of which 21 are to *G. obscuriglobus*, 20 are to *Planctomyces*, and 21 are to *S. acidiphila*. The functional distribution is similar to the BLAST results: A large fraction are hypothetical (51), unnamed (7) and probable (3), while annotated functions include flagellar (6), ribosomal (13), efflux (4) and membrane (5) hits. Notably, PHMMER and BLASTP find the exact same top hits for 64/128 proteins. 17 are non-identical *G. obscuriglobus* proteins, 11 are from *Planctomyces*, and 14 from *S. acidiphila*.

As BLASTP and PHMMER screens both identified hits to other proteins coded in the *G. obscuriglobus* genome, we examined similarity among proteins from the pore-containing fraction. We performed all against all BLAST (2) followed by a Markov clustering on the results (3) as implemented in VisBLAST (4) with default parameters E<0.001 and i-value=2.0. This yielded two large clusters (both containing 11 proteins), and a number of small triplet and doublet clusters (Fig S13). The vast majority of proteins were singletons (91 proteins; S3 Table), and the clustering was robust at higher E-value cutoffs (up to E=10) and i-values (from 1.2 to 5). One of these large clusters contained 7 proteins that were unique to fraction 3 with the remaining four proteins being divided between fractions (2,3) and (2,3,6). The second large cluster contained no proteins that were unique to fraction 3. We performed the same clustering on the full set of all 512 proteins and the same clusters were recreated. VisBLAST was used to cluster proteins based on sequence similarity. An E-value cut-off of 0.001 was used together with an i-value of 2. Interestingly, the cluster containing 8 proteins unique to fraction 3 did not change at all in this larger analysis, which indicates that it is indeed membrane-specific. The other large cluster expanded with proteins belonging to fractions (2,6), (6) and especially fraction (2). Other large clusters were found within the set of 512 proteins, however none contained fraction 3 proteins (Fig S14).

Transmembrane helix structure potential was predicted for all 128 fraction 3 sequences using TMHMM2 (5). Of these, 42 showed significant signal of one or more transmembrane helices. Seven of these are unique to fraction 3 (S3 Table). We also looked for evidence of coiled coils using Paircoil2 (6). We report 15 proteins with significant signal of a coiled coil structure (S3 Table).

We next examined fold architecture of proteins in the membrane pore fraction using Phyre2 (7). This approach uses homology modelling to infer the structure of an amino acid sequence based on resemblance to known structures. 127 of the sequences were modelled in full, but one protein

(ZP_02734818, 2558 aa) was analyzed in pieces because of its large size. The Phyre2 result for this protein is thus based on the best scoring subsequence. In the S3 Table we report the top Phyre2 structural hit for each protein. These hits are extracted automatically from the output, and we note that they may not represent the sole best hit (multiple hits often have the same highest confidence score). Based on PDB descriptions, these unfiltered results include membrane proteins (14), flagellar proteins (7) and ribosomal proteins (8), along with a number of other bacterial membrane/transport-related hits.

Comparing all structural predictions with our clustering analysis reveals a number of interesting patterns. Most significantly, cluster 1 consists of proteins modelled by Phyre2 as β-propeller-containing. This is notable given the presence of the β-propeller architecture in protein constituents of the eukaryotic nuclear pore complex (8). Of the 11 proteins in this β-propeller cluster, seven are unique to the pore-containing membrane fraction (3) (Fig S14 and S3 Table). The second large cluster (cluster 2) is dominated by pilins, and the proteins in this cluster mainly come from membrane fraction (3, 2, 6). Approximately half the structural predictions for singleton proteins showed significant structural similarity to porins and membrane proteins, ribosomal subunits, and flagellar proteins (S3 Table). The predicted triplet cluster contains flagellar proteins.

A closer investigation of the structure predictions for the 11 members of the cluster 1 showed that eight yield at least one structure prediction with a confidence >95%. In most cases, multiple predictions are made covering all parts of the sequences.

As constituency in a cluster does not establish whether all constituents share a common region of sequence similarity, we performed multiple sequence alignments across all members of both cluster 1 and cluster 2 (using MAFFT, option L-ins-i, (9). For cluster 1, all sequences displayed similarity in the C-terminal region. Fig S15 shows the alignment for the eight sequences for which we also obtained high

confidence (>95%) structures with Phyre. We evaluated the full alignment using the T-Coffee CORE program (10), which shows a moderately robust 8-way alignment with a CORE-score of 69 (where 100 is perfect alignment). It is clear that the conservation is most pronounced in the C-terminal end of the sequences from around position 850 in the alignment. Indeed, if only the C-terminal part of the alignment is analysed using T-Coffee, the score increases to 81. This corresponds well with the observation that the majority of hits retrieved when searching the non-redundant protein database using both BLAST and PHMMER are also against the C-terminal ends.

With the aim of better characterising the commonalities of cluster 1, we focused on the structural predictions associated with the common C-terminal region. Note that there is some disagreement between the top hits shown in S3 Table (which was automatically generated) and those derived from the conserved C-terminal region (S4 Table). If both coverage and confidence scores generated by Phyre 2 are considered, the structures associated with the C-terminal region (S4 Table) emerge as the best hits. Some cluster 1 proteins also yield significant predictions for their N-terminal ends, but these are not in conflict with results from their respective C-termini, indicating these may be multi-domain proteins. In all 8 cases where a significant (confidence >95%) structure model is obtained, the C-terminal predictions are for β-propeller structures that overlap with the conserved C-terminal region of the sequences (Fig S15). Furthermore, Phyre2 modeled all 8 proteins to the same PDB template (2C4D), which we interpret as independent verification that these proteins share a common structural fold. These results are not due to extensive sequence similarity, as the overall sequence identity between the queries and the PDB template ranges from 13% to 19%.

Results from our structural analysis of cluster 1 are given in S4 Table, and structural models are depicted in Fig 7C (all structures are visualized using The PyMOL Molecular Graphics System, Version 1.4.1 Schrödinger, LLC).

For the second large cluster (Fig S14), we performed the same type of analysis. All can be aligned (Fig S16), 10 of 11 sequences have significant (>95% confidence) structure predictions, and in all cases the best hit in terms of both confidence and coverage (S5 Table) was modeled against PDB file 1OQW. The predicted structures are all very similar and consist of a single α-helix.

Two proteins (ZP_02735673 and ZP_02736511) show possible α-solenoid structures with stacked α-helices (Fig S18). Both are singletons in the cluster analysis, and both are present in membrane fraction 3 (the first is in both fractions 3 and 2, and the latter is unique to fraction 3). The left-hand structure in Fig S18 (ZP_02735673) models against alpha-solenoid structures in pdb with high (>99%) confidence, with models spanning >95% of the sequence. The model shown is based on 1OYZ, a hypothetical protein from *E. coli*, which is classified in SCOP as a member of the ARM repeat superfamily. Within the top 10 hits are structures that derive from Bacteria, Archaea and Eukaryotes, including clathrin adaptor core proteins (2VGL, 1W63). The right-hand structure in Fig S18 (ZP_02736511) contains two high confidence domain models. The N-terminal region models to the same alpha-solenoid structure as seen in the left-hand structure in Fig S18. This spans 30% of the protein sequence. In the adjacent central region, Phyre2 models a response regulator (top hit: 1ZES) with high (>99%) confidence.

A possible FG repeat-containing protein (ZP_02734840) was found in the pore-containing membrane fraction and in the total nuclear membrane proteome. With 5 FGs in the first 200 residues, this conforms to a recent definition of FG-repeat nucleoporin (11) but the C-terminal half of the protein models as a transmembrane beta-barrel protein.

We also performed a more general screen for bacterial transmembrane proteins among our structural predictions. To do this, we screened results for predicted structures of the following type: “outer membrane lipoprotein”, “outer membrane efflux protein”, “transmembrane beta barrel”, “porin”, “tolc” and “oprd/g/h/p/”. The same protein might have hits in more than one of these categories. For transmembrane beta barrels, we chose the best hit to a beta barrel spanning the membrane even if the term “transmembrane” was not used to describe that particular hit (however, for all proteins in this category at least one hit is called “transmembrane”). For all other categories we chose the best hit containing the specified keyword(s). S6 Table summarises these hits with their ID, membrane fraction, confidence score and coverage (as reported by Phyre2), and the PDB template used. For the oprx group we also list the specific class.

Supplementary Text References

1. Finn RD, Clements J, Eddy SR HMMER web server: interactive sequence similarity searching. *Nucleic acids research* 2011; 39(Web Server issue):W29-37.
2. Altschul SF, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic acids research* 1997;25(17):3389-3402.
3. van Dongen S Graph clustering by flow simulation. PhD (University of Utrecht) 2000.
4. Weirather JL, Wilson ME, Donelson JE (2012) Mapping of VSG similarities in *Trypanosoma brucei*. *Molecular and biochemical parasitology* 2012;181(2):141-152.

5. Krogh A, Larsson B, von Heijne G, Sonnhammer EL Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *Journal of molecular biology* 2001;305(3):567-580.

6. McDonnell AV, Jiang T, Keating AE, Berger B Paircoil2: improved prediction of coiled coils from sequence. *Bioinformatics* 2006; 22(3):356-358.

7. Kelley LA, Sternberg MJE Protein structure prediction on the Web: a case study using the Phyre server. 2009;*Nat Protoc* 4(3):363-371.

8. Devos D*,* et al. Components of coated vesicles and nuclear pore complexes share a common molecular architecture. *PLoS Biol* 2004; 2(12):e380.

9. Katoh K, Toh H Parallelization of the MAFFT multiple sequence alignment program. *Bioinformatics* 2010; 26(15):1899-1900.

10. Notredame C, Higgins DG, Heringa J T-Coffee: A novel method for fast and accurate multiple sequence alignment. *Journal of molecular biology* 2000;302(1):205-217.

11. Degrasse JA, Devos D (2010) A functional proteomic study of the *Trypanosoma brucei* nuclear pore complex: an informatic strategy. *Methods Mol Biol* 2010;673:231-238.