1 **Bacterial Major Vault Protein homologs shed new light on origins of the enigmatic**
2 **organelle**

3 Tymofii Sokolskyi¹

4 ¹ Trinity College of Arts and Sciences, Duke university, Durham, NC, USA, 27708

5

6 Corresponding author: tymofii.sokolskyi@duke.edu

7

8 **Keywords:** vaults, major vault protein, Cyanobacteria, lateral gene transfer, eukaryogenesis

9

10

Abstract

11 Vaults are large cone-shaped and highly conservative ribonucleoprotein complexes present in the
 12 cells of most major eukaryote clades. However, despite their wide distribution, their functions
 13 and evolutionary dynamics still remain enigmatic. Several minor functions in modulating
 14 signaling cascades and multidrug resistance phenotypes were previously discovered for
 15 eukaryotic vaults, yet nothing is known about bacterial homologs of the major vault protein
 16 (MVP), a protein that comprises the entirety of vault external surface. Using gene and protein
 17 BLAST searches in NCBI and UniProt databases we identified a number of bacterial species in
 18 prokaryotic taxa Myxococcales, Cytophagales and Oscillatoriales with >50% identity to
 19 eukaryotic MVP sequences. Interestingly, all of these species are characterized by one common
 20 feature – gliding type of motility. Secondary structures of the identified proteins were predicted
 21 using RAPTORX web service and aligned via jFATCAT-flexible algorithm in the RCSB PDB
 22 Java Structure Alignment tool to elucidate structural identity. Coiled coil domain at the MVP C-
 23 terminus of all studied bacterial species resembled TolA protein of *Escherichia coli* by both
 24 structure and sequence. We also showed that MVP sequences from chemotrophic bacteria
 25 Myxococcales and Cytophagales contain a domain homologous to eukaryotic band-7 domain,
 26 unlike cyanobacterial and eukaryotic major vault proteins. As expected, maximum-likelihood
 27 phylogenetic trees for MVP sequences separate studied taxa into two clades – first clade contains
 28 Oscillatoriales (Cyanobacteria) and Eukaryotes and the second one contains chemotrophic
 29 bacteria. In addition, binding prediction via RAPTORX showed great multiplicity GMP and
 30 CMP nucleoside monophosphate binding pockets in Myxococcales and Cytophagales MVP,
 31 unlike eukaryotic and cyanobacterial proteins which had much lower affinity to these substrates.

32 Due to high similarity of eukaryotic and cyanobacterial MVP sequences and a pattern of its
 33 phylogenetic distribution, we can speculate that the most likely scenario for vault appearance in
 34 eukaryotes is horizontal gene transfer from cyanobacteria. Presence of GMP and CMP binding
 35 pockets in MVP could also point to a function in depleting cytosolic nucleotide concentration
 36 which would be beneficial, for instance, during a viral infection. Further research is necessary to
 37 uncover potential functions of this enigmatic protein in bacteria and to determine its evolutionary
 38 patterns. In addition, a correlation between MVP presence and gliding motility in bacteria could
 39 also lead to elucidating selective pressures on the early evolution of this protein. Unfortunately,
 40 this topic has been largely neglected in recent literature and it can lead us to a much better
 41 understanding of not only current physiological processes but also eukaryogenesis, and even
 42 broader – origins of cellular life.

43 Introduction

44 Vaults are large 13 MDa ribonucleoprotein complexes present in cells of many Eukaryota
 45 species (Kedersha et al., 1991). They consist of 3 types of proteins – major vault protein (MVP),
 46 vault poly-ADP ribose polymerase (vPARP), telomerase-associated protein (TEP1) and vault
 47 RNA. TEP1 is also shared with the telomerase complex, part of a so-called TROVE module that
 48 is shared between some ribonucleoproteins and mediates RNA binding (Bateman & Kickhoefer,
 49 2003). The outer surface of a vault particle consists of MVP monomers forming two connected
 50 dome-shaped structures (Mikyas et al., 2004). It has been shown that vault particles are also
 51 capable of opening, possibly to transport other particles inside them (Querol-Audi et al., 2009).
 52 Vaults seem to be very conservative structures present among various Metazoa, Fungi and
 53 Protozoa taxa (Kong et al., 1999), meaning they could perform or could have performed a
 54 globally significant function. Multiple specific functions of this structure known in Metazoa are
 55 implicated in signaling pathway regulation, multidrug resistance, immunity, etc. (Berger et al.,
 56 2009). Currently vaults are even tested as drug and probe delivery vectors (Benner et al., 2017).
 57 These functions include:

- 58 1. Nuclear import of phosphoinositide phosphatase PTEN.
- 59 2. Scaffold in EGFR/MAPK cascade.
- 60 3. Apoptotic suppression via interactions with COP1 protein.
- 61 4. Nuclear import of activated estrogen receptors.
- 62 5. Unidentified role in axonal transport, possibly RNA transport (Li et al., 1999).
- 63 6. Possible regulation of poly-ADP ribosylation to facilitate DNA reparation.
- 64 7. Unidentified role in sea urchin ontogenesis (Stewart et al., 2005).
- 65 8. Multidrug resistance in mammal cells (Izquierdo et al., 1996).
- 66 9. Possible role in Epstein-Barr virus immunity.

67 A hypothesis was proposed for vaults to be an efficient way of amino acid and nucleotide storage
 68 (Shaik, 2013). This could explain their weird phylogenetic distribution – mostly in taxa that lost
 69 essential amino acid biosynthesis pathways and their upregulation during pathogen invasion –
 70 they could act as nutrient sequesters from the cytoplasm (Shaik, 2013). However, if this
 71 explanation is viable, then many questions regarding vault composition arise – what is the
 72 purpose of TEP1 and vPARP presence in the complex, why vaults contain relatively few RNAs,
 73 why are they shaped like they are and specifically why are they hollow inside.

74 Vault's outer surface completely consists of major vault protein monomers. Recently its
 75 homologs were identified in several bacterial taxa, namely in some representatives of
 76 Cyanobacteria, Deltaproteobacteria and Bacteroidetes (Shaik, 2013). However, until now they

77 have not been investigated neither *in silico* nor *in vivo*. Analysis of structure and evolution of
78 bacterial MVP-like proteins may shed new light on vault function and evolution.

79 **Materials and methods**

80 BLAST searches for the Homo sapiens MVP in Uniprot KB Bacteria database uncovered various
81 proteins, several of which, representing different bacterial clades were selected for further
82 investigation: *Moorea producens* uncharacterized protein F4Y3B4_9CYAN, *Saprospira grandis*
83 uncharacterized protein H6L4P8_SAPGL, *Microscilla marina* uncharacterized protein
84 A1ZGE7_9BACT, and *Enhygromyxa salina* uncharacterized protein A0A0C2D5V5_9DELT.
85 Subsequent searches for these proteins found similarities with TolA proteins for various
86 Proteobacteria and Bacteroidetes, and band 7 (SPFH) domain-containing proteins.

87 Then chosen MVP-like sequences from the first BLAST iteration were aligned with various
88 TolA protein sequences found in the UniProt database for the respective taxa using local
89 pairwise alignment tool EMBOSS Water (default settings, Smith-Waterman algorithm; Rice et
90 al., 2000). Most common region of similarity was identified – sequence of approximately 200-
91 250 amino acids. Maximum-likelihood phylogenetic tree with 1000 bootstrap replications were
92 constructed in MEGA 7.0, including some sequences of eukaryotic MVP (Tamura et al., 2007).
93 Secondary structure and binding prediction for isolated regions of some of the studied proteins
94 was performed with the help of RAPTORX web service (Källberg et al., 2012). Acquired
95 structure data in the PDB format was then subjected for pairwise structure alignment via
96 jFATCAT-flexible algorithm in the RCSB PDB Java Structure Alignment tool to confirm
97 structural identity (Prlić et al., 2010; Ye & Godzik, 2003).

98 **Results**

99 **Comparison to TolA.** BLAST searches for these sequences uncovered various TolA proteins;
100 highest identity of 36% had *Gilliamella* TolA A0A1B9K2Y3_9GAMM. Series of alignments
101 between the chosen sequences and TolA and TolA-like proteins of different bacterial species
102 (*Enhygromyxa*, *Myxococcus*, *Sandaracinus*, *Escherichia*, *Haemophilus*, *Methylophaga*,
103 *Mangrovibacter*, *Gilliamella*, *Frischella*, *Galibacterium*) showed that the highest similarity is
104 located in the coiled coil region of the selected proteins and domain II of TolA. Identity variation
105 between 22% and 37% and alignment of TolA sequences not only with the coiled coil region, but
106 also with adjacent fragments of MVP shoulder and N-terminal domain suggests that it is most
107 likely a result of homologous relationship between TolA domain II and MVP.

Results of the MOTIF search service on Genome Net (Bioinformatics Center at Kyoto University, <http://www.genome.jp/tools/motif/>), Pfam database (Finn et al., 2013) support this hypothesis showing TolA domains with E-values up to $1e-07$ when analyzing bacterial MVP-like proteins. Schematic structure comparisons of *Saprospira* MVP-like protein and *Escherichia* TolA are shown in **Fig.1, A**.

Fig.1 Structural features of MVP-like proteins. **A.** Schematic representation of *Saprospira grandis* uncharacterized protein (upper block) and *Escherichia coli* TolA protein (lower block) sequences. Aligned regions are marked with red line (portion of domain II from *E. coli* TolA and coiled coil with a small portion of MVP shoulder from *S. grandis* protein). Identity 27,6%, similarity 42,0%. **B.** Location of band 7 domain in *Saprospira* protein H6L4P8_SAPGL (result of MOTIF search). **C.** Graphic representation of a predicted GMP-binding site of *Saprospira* MVP-like protein.

Structure prediction. Some of the most similar sequence fragments, corresponding to coiled coils of MVP-like proteins and domain II of TolA for some of the studied proteins were subjected to structure prediction via RAPTORX web service. Pairwise structural alignment was then conducted to support the idea of homology between major vault protein and TolA. Chosen sequences include: *Haemophilus* TolA TOLA_HAEIN (70-275); *Homo* Major Vault Protein MVP_HUMAN (729-805); *Moorea* MVP-like protein F4Y3B4_9CYAN (750-820); *Saprospira* uncharacterized protein H6L4P8_SAPGL (701-780). Acquired results, summarized in **table 1**, confirm the conclusion of evolutionary relationship between TolA and MVP-like proteins because of relatively low P-values and RMSD.

	<i>Moorea</i>	<i>Homo</i>	<i>Haemophilus</i>	<i>Saprospira</i>
<i>Moorea</i>				

<i>Homo</i>	1,79e-04; 3,17			
<i>Haemophilus</i>	2,03e-05; 2,96	9,93e-04; 1,56		
<i>Saprospira</i>	1,94e-08; 3,05	2,19e-06; 1,17	5,34e-06; 1,76	

130

131 **Table 1.** Results of jFATCAT-flexible structure alignment of structures, explained in text. First
132 number represents P-value, second – RMSD value.

133 **Comparison with band 7 domains.** Other than TolA, some BLAST searches for MVP
134 shoulder-like domains in KB Bacteria database recovered more than 30% identity with band 7
135 proteins (*Halothermothrix* protein B8D0R4_HALOH showed the highest result of 33%). In
136 addition, results of the MOTIF search service also identified SPFH / Band 7 domain in 509-598
137 region of *Saprospira* sequence H6L4P8_SAPGL with E-value=0.00039 (**Fig. 1, B**).
138 Interestingly, same results were observed with other studied bacterial proteins, except
139 cyanobacterial MVP-like proteins and eukaryotic MVP. This similarity was first reported by
140 Daly et al., 2013.

141 Secondary structures for corresponding fragments of these two proteins along with *Pongo* MVP
142 MVP_PONAB shoulder domain (519-647) and *Homo* stomatin STOM_HUMAN band 7 domain
143 (56-227) were predicted and aligned using jFATCAT-flexible algorithm. Resulting RMSD and
144 P-value are shown in **table 2**. P-values and RMSD are low enough to conclude likely homology.
145 *Saprospira* MVP shoulder-like domain exhibits significantly higher structural similarity to band
146 7 domain of stomatin and *Halothermothrix* protein than to MVP shoulder of *Pongo* MVP.

	<i>Halothermothrix</i>	<i>Homo</i>	<i>Pongo</i>	<i>Saprospira</i>
<i>Halothermothrix</i>				
<i>Homo</i>	1,18e-12; 1,55			
<i>Pongo</i>	1,57e-01; 3,01	1,55e-01; 3,03		
<i>Saprospira</i>	1,18e-14; 1,36	6,38e-10; 1,82	2,68e-01; 3,26	

147

148 **Table 2.** Results of jFATCAT-flexible alignment of structures, explained in text. First number
149 represents P-value, second – RMSD value.

150 **Binding prediction.** Binding prediction via RAPTOR-X server was conducted for *Saprospira*,
151 *Enhygromyxa*, *Moorea* and *Homo* MVP-like proteins. *Saprospira* protein showed high-

152 multiplicity binding with GMP and CMP. Multiplicity represents the frequency with which the
 153 selected pocket was found in a set of ligand-binding protein structures – generally, if the value is
 154 above 40 the pocket is highly likely true. Multiplicity for the first GMP-binding site is 102, for
 155 the second – 35 (shown on **Fig. 1, C**), for CMP-binding site – 68. For *Enhyngromyxa* it is 102
 156 and 28 for GMP and 72 for CMP. For *Moorea* protein it is only 46 and 12 for GMP and 28 for
 157 CMP; *Homo* MVP shows 37 and 14 for 2 possible GMP-binding sites, 38 for UMP and none for
 158 CMP. All of the possible nucleotide binding sites are located in 500-700 sequence, that
 159 corresponds to the part that aligns with MVP shoulder/band 7 domain. TolA proteins were also
 160 subjected for similar analysis. Neither *Haemophilus* nor *Gilliamella* TolA show any nucleotide
 161 binding *in silico*.

162 **Discussion**

163 **Origin of vaults.** We found 2 likely homologies for bacterial MVP-like proteins: TolA and band
 164 7 domain proteins. A maximum-likelihood phylogenetic tree was constructed for complete
 165 protein sequences of various bacterial MVP-like proteins (**Fig. 2**). Two macroclades can be
 166 defined: first one contains Bacteroidetes and Deltaproteobacteria, the second one – eukaryotic
 167 MVP and cyanobacterial MVP-like proteins. The main structural difference between them is
 168 presence of band 7 domain in the sequences of all first clade taxa, and its absence in all second
 169 clade taxa. *Hyalangium* protein A0A085WQF1_9DELT is the only found non-cyanobacterial
 170 MVP-like protein lacking band 7 domain, no similar deltaproteobacterial sequences were found.
 171 Band 7 domains are generally known for association with lipid membranes (Tavernarakis et al.,
 172 1999).

173

175 **Fig. 2** Maximum-likelihood phylogenetic tree derived for various bacterial MVP-like proteins.
 176 *Haemophilus TolA* sequence was used as an outgroup.

177 One of these bacteria, specifically *Saprospira grandis*, displays an unusual characteristic – rod-
 178 shaped protein structures in the cytoplasm, called raphidosomes (Saw et al., 2012). Their
 179 functions are currently unknown. One of the raphidosome proteins, SGRA_0791, was shown to
 180 contain Band 7-like domain (Saw et al., 2012). However, alignment of its 25-232 sequence that
 181 corresponds to Band 7 domain, according to Uniprot, with 500-600 sequence of *Saprospira* Band
 182 7-like domain displayed only 20% identity, that is very low for such a short sequence, especially
 183 compared to >50% identities of MVP-like proteins between each other. At least for now there is
 184 no evidence of connection between raphidosomes and MVP-like proteins.

185 Major vault protein homologs were not found in archaea, that are closely related to eukaryotes.
 186 Therefore, MVP-like proteins were likely to appear in the bacterial lineage after
 187 Bacteria/Archaea divergence and then experience secondary loss in some lineages. Alternatively,
 188 they MVP-like proteins were present in the Last Universal Common Ancestor and then lost in
 189 Archaea and retained in Bacteria. However, if that is the case, then the loss should have occurred
 190 after the divergence of Eukarya and Archaea and then simultaneously happen in all
 191 contemporary archaeal clades which seems improbable. Therefore, we suggest that the most
 192 likely scenario of MVP appearance in Eukarya is lateral transfer from a bacterial lineage, most
 193 likely Cyanobacteria (Oscillatoriales), because they possess MVP-like protein with more than

194 50% identity to MVP. This event probably has happened before the divergence of major protist
195 groups, animals, plants and fungi (Daly et al., 2013). This hypothesis is congruent with the
196 phylogeny shown on **Fig. 2** that groups together cyanobacterial and eukaryotic MVP-like
197 proteins.

198 Collected information suggests a probable evolutionary scenario for major vault protein,
199 schematically depicted in **Fig. 3**. Pre-MVP appeared in the common ancestor of most bacterial
200 phyla and contained band 7-like (SPFH) and TolAII-like domains (possibility of independent
201 appearance of both domain types in many distantly related phyla is very low). Consequently, a
202 duplication and a series of deletions occurred resulting in TolA protein, that lost band 7 domain
203 in favor of a short transmembrane helix and acquired a variable N-terminal TolAIII domain.
204 Some lineages, including Bacteroidetes, Cyanobacteria (at least Oscillatoriales) and
205 Deltaproteobacteria retained the original pre-MVP. Ability to bind NMP appeared prior to
206 Cyanobacteria/Bacteroidetes/Proteobacteria divergence, because it can be found in both *Moorea*
207 and *Saprospira* proteins.

208 Cyanobacterial pre-MVP evolution made it move away from cell membrane transforming its
209 band 7 domain into MVP shoulder. Cyanobacterial proteins could represent the first evolutionary
210 steps towards the vault complex formation. Then, TolAII-like domain evolved to vault coiled
211 coils that facilitate MVP aggregation into vaults, as demonstrated in modern eukaryotes (van
212 Zon et al., 2002). Similar structure of *Hyalangium* protein could have been acquired
213 convergently or due to a separate lateral gene transfer event. Then sequence of the resulting
214 protein was laterally transmitted from cyanobacteria to early eukaryotes (most likely due to
215 phagotrophic lifestyle of the latter ones (Shaik, 2013)) or due to an endosymbiotic event and was
216 subsequently lost in various lineages possibly including plants and fungi (Daly et al., 2013).

Figure 3. Most likely scenario of major vault protein evolution. Probably, NMP-binding appeared in Cyanobacteria and Deltaproteobacteria/Bacteroidetes MVP common ancestor, and then started waning in Cyanobacteria.

Future directions. All bacteria with MVP-like proteins identified here despite belonging to highly different taxa share a common characteristic: gliding motility (Engene et al., 2012; Iizuka et al., 2003 [1, 2]; Reichenbach, 2015). Specifically, these taxa include Deltaproteobacteria (only order Myxococcales), Bacteroidetes (specifically Cytophagales) and Cyanobacteria (only Oscillatoriales). This type of locomotion usually is prevalent among bacteria living in low-water content environments, including biofilms, mats, soils or hypersaline habitats, that is also true for the species studied here (Spormann, 1999). Interestingly, TolA protein that is partially homologous to bacterial MVP-like proteins is involved in flagellar locomotion (Cascales et al., 2001). Therefore MVP-like proteins could have played a role in the transition to gliding motility in selected taxa, mechanism of which is poorly studied (McBride, 2001).

Observation of MVP-like proteins NMP binding also supports the idea of vaults functioning as sequesters and regulators, as was proposed by Shaik, 2013. Bacterial MVP-like proteins could perform the same function vaults do, but without necessary aggregation into vaults. One of the interesting possible early functions of MVP-like proteins can be pathogen defense. By sequestering nucleotides and amino acids from the cytoplasm they can at least decrease the rate of bacteriophage reproduction. Nucleotides are required for bacteriophage replication, while they are not immediately necessary for bacterium during the infection process because no DNA replication is occurring (**Fig. 4**). It is possible that MVP demonstrates a lot weaker nucleotide binding than MVP-like proteins because of eukaryotic compartmentalization – deoxynucleotide

240 concentration is a lot smaller in eukaryotic cytosol than in bacterial, meaning there is less sense
241 in sequestering it during bacterial/viral infection. However, eukaryotes face wider systematic
242 range of pathogens, meaning that a new way of spatial nutrient isolation should have been
243 developed – that is aggregation into large hollow structures (vaults).

244 Membrane-association via band 7 domains also could help them maintain this function. It has
245 been shown that band 7 domain proteins can form oligomeric ring structures in mitochondria
246 (Tatsuta et al., 2005) and cyanobacteria (Boehm et al., 2009) that are membrane-associated. This
247 leads to an idea that MVP-like proteins could have also been associated with toxin export,
248 similar to modern vaults involved in multidrug resistant phenotypes (Izquierdo et al., 1996).
249 Alternatively, they could have also performed direct detoxification of substances such as toxic
250 anions, as was demonstrated for modern MVP and tellurite anions (Suprenant et al., 2007).
251 However, these hypotheses are purely speculative and require experimental verification.

253 **Figure 4.** Possible functions of bacterial MVP-like proteins in cell protection. 1 – nucleotide
254 sequestering from the cytosol to prevent viral replication; 2 – amino acid sequestering from the
255 cytosol to affect viral protein synthesis or to conserve energy; 3 – export of toxic substances due
256 to oligomerization through band 7 domains and membrane binding; 4 – direct detoxification.

257 MVP is one of the very few protein groups that are conserved among both Bacteria and Eukarya
258 (Kedersha et al., 1990). It has been suggested that they were already present in last eukaryotic
259 common ancestor, however their discovery in bacteria significantly expands their importance
260 (Daly et al., 2013). In addition, the potential of cyanobacteria-eukaryote lateral gene transfer is
261 intriguing and deserves further investigation. Unfortunately, the vault system has been largely

neglected in the scientific community showing lack of publications even despite advances in molecular techniques over the past 35 years since the discovery of vaults (Kedersha & Rome, 1986). However, due to its conservativity and wide distribution, elucidating functions of vaults and MVP-like proteins in both prokaryotes and eukaryotes and their evolutionary patterns is necessary to understand eukaryogenesis and the broader picture of origins of cellular life. More *in silico*, *in vivo* and *in vitro* studies need to be done on this topic so that we can understand why these systems were retained by so many different taxa and how did they influence our evolution.

References

- Asfa Alli Shaik. The unique phylogenetic distribution of vault particles reveals its functional roles. A thesis submitted for the degree of doctor of Philosophy, Department of biological sciences, National university of Singapore 2013
- Bateman, A., & Kickhoefer, V. (2003). The TROVE module: a common element in Telomerase, Ro and Vault ribonucleoproteins. *BMC bioinformatics*, 4(1), 49.
- Benner, N. L., Zang, X., Buehler, D. C., Kickhoefer, V. A., Rome, M. E., Rome, L. H., & Wender, P. A. (2017). Vault nanoparticles: chemical modifications for imaging and enhanced delivery. *ACS nano*, 11(1), 872-881.
- Berger, W., Steiner, E., Grusch, M., Elbling, L., & Micksche, M. (2009). Vaults and the major vault protein: novel roles in signal pathway regulation and immunity. *Cellular and molecular life sciences*, 66(1), 43.
- Boehm, M., Nield, J., Zhang, P., Aro, E. M., Komenda, J., & Nixon, P. J. (2009). Structural and mutational analysis of band 7 proteins in the cyanobacterium *Synechocystis* sp. strain PCC 6803. *Journal of bacteriology*, 191(20), 6425-6435.
- Cascales, E., Lloubès, R., & Sturgis, J. N. (2001). The TolQ–TolR proteins energize TolA and share homologies with the flagellar motor proteins MotA–MotB. *Molecular microbiology*, 42(3), 795-807.
- Corda, D., & Di Girolamo, M. (2003). Functional aspects of protein mono-ADP-ribosylation. *The EMBO journal*, 22(9), 1953-1958.
- Daly, T. K., Sutherland-Smith, A. J., & Penny, D. (2012). Beyond BLASTing: tertiary and quaternary structure analysis helps identify major vault proteins. *Genome biology and evolution*, 5(1), 217-232.

292 Daly, T. K., Sutherland-Smith, A. J., & Penny, D. (2013). In silico resurrection of the Major
293 Vault Protein suggests it is ancestral in modern eukaryotes. *Genome biology and*
294 *evolution*, 5(8), 1567-1583.

295 Engene, N., Rottacker, E. C., Kaštovský, J., Byrum, T., Choi, H., Ellisman, M. H., ... &
296 Gerwick, W. H. (2012). *Moorea producens* gen. nov., sp. nov. and *Moorea bouillonii*
297 comb. nov., tropical marine cyanobacteria rich in bioactive secondary
298 metabolites. *International journal of systematic and evolutionary microbiology*, 62(Pt
299 5), 1171.

300 Finn, R. D., Bateman, A., Clements, J., Coghill, P., Eberhardt, R. Y., Eddy, S. R., ... &
301 Sonnhammer, E. L. (2013). Pfam: the protein families database. *Nucleic acids*
302 *research*, 42(D1), D222-D230.

303 Iizuka, T., Jojima, Y., Fudou, R., Hiraishi, A., Ahn, J. W., & Yamanaka, S. (2003).
304 *Plesiocystis pacifica* gen. nov., sp. nov., a marine myxobacterium that contains
305 dihydrogenated menaquinone, isolated from the Pacific coasts of Japan. *International*
306 *journal of systematic and evolutionary microbiology*, 53(1), 189-195.

307 Iizuka, T., Jojima, Y., Fudou, R., Tokura, M., Hiraishi, A., & Yamanaka, S. (2003).
308 *Enhygromyxa salina* gen. nov., sp. nov., a slightly halophilic myxobacterium isolated
309 from the coastal areas of Japan. *Systematic and applied microbiology*, 26(2), 189-196.

310 Izquierdo, M. A., Scheffer, G. L., Flens, M. J., Giaccone, G., Broxterman, H. J., Meijer, C. J.,
311 ... & Scheper, R. J. (1996). Broad distribution of the multidrug resistance-related vault
312 lung resistance protein in normal human tissues and tumors. *The American journal of*
313 *pathology*, 148(3), 877.

314 Källberg, M., Wang, H., Wang, S., Peng, J., Wang, Z., Lu, H., & Xu, J. (2012). Template-
315 based protein structure modeling using the RaptorX web server. *Nature*
316 *protocols*, 7(8), 1511.

317 Kedersha, N. L., & Rome, L. H. (1986). Isolation and characterization of a novel
318 ribonucleoprotein particle: large structures contain a single species of small
319 RNA. *The Journal of cell biology*, 103(3), 699-709.

320 Kedersha, N. L., Heuser, J. E., Chugani, D. C., & Rome, L. H. (1991). Vaults. III. Vault
321 ribonucleoprotein particles open into flower-like structures with octagonal
322 symmetry. *The Journal of cell biology*, 112(2), 225-235.

323 Kedersha, N. L., Miquel, M. C., Bittner, D., & Rome, L. H. (1990). Vaults. II.
 324 Ribonucleoprotein structures are highly conserved among higher and lower
 325 eukaryotes. *The Journal of cell biology*, 110(4), 895-901.

326 Kong, L. B., Siva, A. C., Rome, L. H., & Stewart, P. L. (1999). Structure of the vault, a
 327 ubiquitous cellular component. *Structure*, 7(4), 371-379.

328 Lewin, R. A. (1997). *Saprospira grandis*: A Flexibacterium That Can Catch Bacterial Prey
 329 by "Ixotrophy". *Microbial ecology*, 34(3), 232-236.

330 Li, J. Y., Volkandt, W., Dahlstrom, A., Herrmann, C., Blasi, J., Das, B., & Zimmermann, H.
 331 (1999). Axonal transport of ribonucleoprotein particles (vaults). *Neuroscience*, 91(3),
 332 1055-1065.

333 McBride, M. J. (2001). Bacterial gliding motility: multiple mechanisms for cell movement
 334 over surfaces. *Annual Reviews in Microbiology*, 55(1), 49-75.

335 Mikiyas, Y., Makabi, M., Raval-Fernandes, S., Harrington, L., Kickhoefer, V. A., Rome, L.
 336 H., & Stewart, P. L. (2004). Cryoelectron microscopy imaging of recombinant and
 337 tissue derived vaults: localization of the MVP N termini and VPARP. *Journal of*
 338 *molecular biology*, 344(1), 91-105.

339 Prlić, A., Bliven, S., Rose, P. W., Bluhm, W. F., Bizon, C., Godzik, A., & Bourne, P. E.
 340 (2010). Pre-calculated protein structure alignments at the RCSB PDB
 341 website. *Bioinformatics*, 26(23), 2983-2985.

342 Querol-Audí, J., Casañas, A., Usón, I., Luque, D., Castón, J. R., Fita, I., & Verdager, N.
 343 (2009). The mechanism of vault opening from the high resolution structure of the
 344 N-terminal repeats of MVP. *The EMBO journal*, 28(21), 3450-3457.

345 Reichenbach, H. 2015. Hyalangiium gen. nov. *Bergey's Manual of Systematics of Archaea*
 346 *and Bacteria*. 1–4.

347 Rice, P., Longden, I., & Bleasby, A. (2000). EMBOSS: the European molecular biology
 348 open software suite.

349 Saw, J. H., Yuryev, A., Kanbe, M., Hou, S., Young, A. G., Aizawa, S. I., & Alam, M.
 350 (2012). Complete genome sequencing and analysis of *Saprospira grandis* str. Lewin, a
 351 predatory marine bacterium. *Standards in Genomic Sciences*, 6(1), 84.

352 Spormann, A. M. (1999). Gliding Motility in Bacteria: Insights from Studies of *Myxococcus*
 353 *xanthus*. *Microbiol. Mol. Biol. Rev.*, 63(3), 621-641.

354 Stewart, P. L., Makabi, M., Lang, J., Dickey-Sims, C., Robertson, A. J., Coffman, J. A., &
355 Suprenant, K. A. (2005). Sea urchin vault structure, composition, and differential
356 localization during development. *BMC developmental biology*, 5(1), 3.

357 Suprenant, K. A., Bloom, N., Fang, J., & Lushington, G. (2007). The major vault protein is
358 related to the toxic anion resistance protein (TelA) family. *Journal of Experimental*
359 *Biology*, 210(6), 946-955.

360 Tamura, K., Dudley, J., Nei, M., & Kumar, S. (2007). MEGA4: molecular evolutionary
361 genetics analysis (MEGA) software version 4.0. *Molecular biology and*
362 *evolution*, 24(8), 1596-1599.

363 Tatsuta, T., Model, K., & Langer, T. (2005). Formation of membrane-bound ring complexes
364 by prohibitins in mitochondria. *Molecular biology of the cell*, 16(1), 248-259.

365 Tavernarakis, N., Driscoll, M., & Kyrpides, N. C. (1999). The SPFH domain: implicated in
366 regulating targeted protein turnover in stomatins and other membrane-associated
367 proteins. *Trends in biochemical sciences*, 24(11), 425-427.

368 van Zon, A., Mossink, M. H., Schoester, M., Scheffer, G. L., Scheper, R. J., Sonneveld, P., &
369 Wiemer, E. A. (2002). Structural domains of vault proteins: a role for the coiled coil
370 domain in vault assembly. *Biochemical and biophysical research*
371 *communications*, 291(3), 535-541.

372 Ye, Y., & Godzik, A. (2003). Flexible structure alignment by chaining aligned fragment pairs
373 allowing twists. *Bioinformatics*, 19(suppl_2), ii246-ii255.