A phospholipid transporter in Asgard archaea sheds light on the origin of eukaryotic lipid transfer proteins

Nicolas-Frédéric Lipp[1§], Elida Kocharian[1], and Itay Budin[1*]

[1] Department of Biochemistry & Molecular Biophysics, University of California San Diego, La Jolla, CA 92093, USA

*Corresponding author: ibudin@ucsd.edu

[§] Current address: Laboratoire Reproduction et Développement des Plantes (RDP), Université de Lyon, ENS de Lyon, UCB Lyon 1, CNRS, INRAE, Inria, F-69342 Lyon, France

NFL ORCID: 0000-0001-7963-7415
EK ORCID: 0009-0005-3375-7418
IB ORCID: 0000-0001-9706-4294

Abstract

The evolution of eukaryotic cells necessitated the advent of machinery to transport phospholipids, the molecular building blocks of cell membranes, to support organelle proliferation. Eukaryotes share several classes of highly conserved lipid transfer proteins (LTP) that associate with donor membranes, bind individual phospholipids, and shuttle them to acceptor membranes. Because cells lacking organelles do not require extensive lipid transport networks, it is not known if this machinery pre-dated eukaryotic organelles or had to evolve alongside them. Here we describe a class of phospholipid transporters in the Asgard achaea that share key structural and functional similarities to eukaryotic LTPs in the START domain superfamily. Asgards contain three classes of START proteins, StarAsg1-3, which are conserved across most Asgard phyla. Of these, StarAsg1 family proteins contain the predicted structural features necessary for lipid transfer: large, hydrophobic binding pockets lined with amphipathic motifs for membrane docking. In contrast, StarAsg2 and StarAsg3 family proteins contain smaller binding cavities and minimal predicted membrane interactions. StarAsg1 from Lokiarchaeia interacts with anionic membranes both in vitro and in yeast cells, where it binds phospholipids. Lipid transfer assays show that StarAsg1 from Lokiarchaeia can exchange several classes of phospholipids between membranes, as eukaryotic START LTPs do. Structural phylogeny of START domains across the tree of life suggest that eukaryotic LTPs could share a common ancestry with StarAsg1 homologs, while StarAsg2 and StarAsg3 form a monophyletic group with a eukaryotic heat shock protein co-chaperone. We propose that the presence of inter-membrane lipid transporters in the ancestors of eukaryotic cells could have facilitated the development of complex intracellular organelles.

Introduction

The architecture of eukaryotic cells relies on their capacity to synthesize and transport insoluble phospholipids as the primary structural components of membranes. Organelles are bound by phospholipid membranes that are composed of distinct compositions, which tune their physicochemical properties and interactions with protein machinery. However, lipid metabolism itself is not replicated across organelles, with the endoplasmic reticulum (ER) being the primary production site for most major lipid classes. Thus, cells rely on post-metabolic transport processes to bring lipid molecules – which by their hydrophobic nature cannot diffuse through the cytosol – from the ER to all other organelles. Lipid transfer between membranes is much faster than protein trafficking[1,2], and this robust transport is essential due the high lipid fluxes needed for organelles to grow and replicate. In extant eukaryotes, these processes also act to generate compositional heterogeneities between organelles that are specialized for their function[3–7].

Intracellular lipid transport was originally thought to occur via vesicular trafficking, but in recent years it has been discovered that many distinct lipid transfer proteins instead accomplish this task. Some LTPs are channel-shaped proteins that form bridges across contact sites, where organelles sit within tens of nanometers from each other[8,9]. However, most are small, soluble proteins that transiently interact with membranes and extract individual lipid molecules and transport them between membranes[10,11]. These shuttle-like LTPs may be tethered to membranes or contact site components, yet they still rely on soluble domains with amphipathic membrane-binding sites, regulated by dynamic bound-to-soluble switching mechanisms that support second-to-minute-scale lipid transport rates[12,13]. The capacity of LTPs to shuttle lipids between opposing membranes is required for the existence of organelles that define eukaryotic cells, but our understanding of the evolution of this machinery remains limited. Lipid transfer mechanisms have been identified across the periplasmic space of gram-negative bacteria involving LolA/B and MlaC homologs[14], but no direct eukaryotic counterparts to these systems have been identified. Since all extant organelles depend on it, the acquisition of intracellular lipid transport machinery would have represented a potential bottleneck during eukaryogenesis, the evolutionary process that led to the emergence of eukaryotic cells.

Eukaryogenesis occurred via endosymbiotic event(s) between 1.5 and 2.5 billion years ago[15–18]. It had long been accepted that eukaryotes form a sister group of archaea, with mitochondria

descended from an engulfed proto-symbiont sister to alphaproteobacteria[19–21]. The discovery of Promethearchaeati[22] (hereafter called Asgards) as an archaeal superphylum that forms a monophyletic group with Eukarya changed our understanding of eukaryogenesis[23] and recent analyses suggest eukaryotes belong to Asgards *sensu stricto*[24–26]. Such a relationship is supported by the presence of Eukaryotic Signature Proteins (ESPs) in Asgards, including homologs of key components of eukaryotic membrane machinery: components of the Endosomal Sorting Complexes Required for Transport (ESCRT) system[27,28], GTPases, Roadblock/longin families of protein, and TRAnsport Protein Particle (TRAPP) complex[29–31]. Cultured Asgard, like the Lokiarchaeon *Promethearchaeum syntrophicum*[22,32], show complex membrane topologies including tubular protrusions. Emerging molecular clock data supports a late-mitochondrial model for eukaryogenesis, in which the Asgard ancestor was already quite complex[33]. Recent imaging of exponential phase Heimdallarchaeia (*Candidatus "Yibarchaeum umbracryptum"*) also shows an abundance of large, intracellular vesicles[34] that resemble eukaryotic compartments like autophagosomes that rely on LTPs for their formation and growth[35]. These findings suggest that the complex membrane biology that characterized eukaryotic cells might have roots in Asgard archaea.

The observation of internal compartments in Asgard cells suggests that they might also contain mechanisms to maintain and grow them via lipid transport, which could have been harnessed during eukaryogenesis for organelle proliferation. We thus asked if any eukaryotic LTPs classes are present in extant Asgards, which would support their presence in the ancestor of eukaryotic cells. We describe three proteins in the Steroidogenic Acute Regulatory protein-related lipid Transfer (START)-domain family in Lokiarchaeia that are widely conserved across Asgard lineages. In eukaryotes, START domains represent a widely utilized protein fold for LTPs, but also can function as regulatory proteins[36–38], transcription factors[39],, and co-chaperones[40]. We report that one Lokiarchaeia START protein is specialized for interaction with membranes and transports phospholipids in vitro, similar to other START domain LTPs. Structural and sequence alignment methods suggest a connection between this protein family and eukaryotic START LTPs and strongly support that another Asgard START family is the origin of a conserved eukaryotic co-chaperone.

104 # Results

105

106 START domain proteins in Lokiarchaeia

107

108 We first examined the distribution of 14 annotated classes of shuttle-like LTPs in bacteria,
109 archaea, viruses and eukaryotes (**Fig. 1**). Each of these LTP classes, often defined by a family
110 or a superfamily of proteins, contains a globular domain hosting one or two lipidic ligands in a
111 buried binding cavity. Nine of them are exclusive to eukaryotes with very rare occurrences in other
112 groups, suggesting that the diversity of LTPs expanded during eukaryotic evolution. In contrast,
113 proteins containing START and Sterol Carrier Protein 2 (SCP-2) domains are distributed among
114 eukaryotes and archaea, but also found in bacteria. From this observation, we identified potential
115 LTPs coding sequences in the genome of the isolated Lokiarchaeia archaeon
116 *Prometheoarchaeum syntrophicum* MK-D1. MK-D1 codes for one SCP-2 homolog and three
117 START-like domains, which we named StarLok1, StarLok2 and StarLok3. We did not observe
118 any clustering of nearby genes related to lipid metabolism or other related functions of START
119 domain proteins. However, we focused attention on these candidates because the structural
120 features and ability to bind hydrophobic ligands of START domains are well defined, in contrast
121 to other families like SCP-2[41,42].

122

123 Using computational approaches, we further characterized the predicted structural features in the
124 three Asgard START domains from the Lokiarchaeia archaeon *P. syntrophicum* MK-D1. We
125 observed a charge distribution of StarLok1 that is strongly polarized, with one side featuring a
126 large cationic patch surrounding the mouth of the protein cavity (**Fig. 2a**). This configuration
127 suggested that StarLok1 may interact with membranes, as peripheral membrane proteins are
128 often electrostatically polarized to interact with phospholipids[43–46]**.** A network of hydrophobic
129 residues (F20, L26, F23, W29, F77, V90 W104, F110, F138), two opposite tyrosines (Y54, Y81),
130 and two distant anionic residues (D21, E66), characterize the cavity of StarLok1, supporting the
131 binding of hydrophobic ligand. Notably, the cavity lacks key polar residues generally required for
132 enzymatic activity, such as ketone cyclization[47]. The Positioning of Protein in Membrane (PPM)
133 algorithm suggested that StarLok1 could dock to membranes through the insertion of five
134 hydrophobic residues, reminiscent of the binding mode of the sterol transporter STARD4[48] (**Fig.
135 2b).** In contrast, StarLok2 and StarLok3 have lower electrostatic polarization, residual cavities
136 and did not show a clear mode of membrane interaction by PPM. To test the accuracy of structural
137 predictions, we also biochemically characterized the three MK-D1 START proteins. For this, we

138 expressed and purified recombinant StarLok1, StarLok2 and StarLok3 fused to 6xHis tag from *E.*
139 *coli* (**Supplementary Fig. 1**). Folding of the three proteins was confirmed by Circular Dichroism
140 (CD) spectroscopy (**Fig. 2c**), which showed a secondary structure profile similar to those of
141 previously-profiled START domain homologs[49,50]. Among the three, StarLok1 showed reduced
142 ellipticity, StarLok2 featured an increased helical signal and reduced β-strands proportion, and
143 StarLok3 showed a reduced number of turns. These features and the overall spectra matched
144 those derived from the AlphaFold structures (**Supplementary Fig. 1h**), confirming the
145 applicability of structural predictions for these proteins. Overall, structure-sequence analyses
146 StarLok1 suggest closest homology to the plant allergen Bet v 1, while StarLok3 resembled the
147 START domain of Aha1 (**Fig. 2d and 2e**).

148

149 Distribution of START domains across Asgard archaea

150

151 To characterize the phylogeny of START domains within Asgards, we first analyzed the
152 relationships of 55 homologous sequences of this class across Asgard phyla. The resulting tree
153 indicates that StarLok1 is part of a monophyletic clade comprising 21 sequences, with moderate
154 branch support (aLRT = 0.8939) **(Fig. 3a)**. Proteins of this clade ('StarAsg1') featured predicted
155 binding cavities large enough (MK-D1 1595 $Å^3$; Heimdall AB_125, 1579 $Å^3$) to accommodate a
156 wide range of lipids, including phospholipids (~1000 $Å^3$) **(Fig. 3b)**. In contrast, the clades
157 containing StarLok2 (StarAsg2) and StarLok3 (StarAsg3) had smaller binding cavities, with
158 StarAsg3 homologs possessing residual pockets under 300 $Å^3$. StarLok2 and StarLok3 proteins,
159 but not StarAsg1, were predicted and annotated as homologs of the activators of the heat shock
160 protein 90 (Hsp90) ATPase (Aha1) by Protein Natural Language Model (ProtNLM/ EMBL-EBI)[51].
161 Thus, Asgards contain three classes of START domain proteins, represented by orthologs of
162 StarAsg1 and StarAsg2, and StarAsg3.

163

164 In a holistic analysis of Asgard proteomes, we identified 411 START domain sequences
165 distributed across 203 current species. We did not observe any START homolog in several
166 Asgard phyla, including Odin and Wukong. However, the three START domains clades
167 (StarAsg1-3) are well represented in most phyla: Kari, Hod and Heimdall phyla all display the four
168 classes of START domain, including StarAsg1 **(Fig. 3c)**. We also identified three occurrences of
169 DAPG hydrolase (PFAM 18089) as well as twenty-nine Rieske [2Fe-2S] type-protein diverging
170 from a group of START domains, which we named StarAsg4 **(Fig. 3c** and **Supplementary Fig.**
171 **2).** Overall, this analysis suggests a distinction between StarAsg1 (SH-aLRT; UFBoot = 83.3; 83),

172 StarAsg3 (73; 81) and the phylum including StarAsg4 and StarAsg2 (86.1; 71) (**Supplementary**
173 **Fig. 2**).
174
175 <u>StarLok1 binds anionic membranes</u>
176
177 Our computational analyses suggested that among the three Lokiarchaeia START domain
178 proteins, StarLok1 is uniquely suited for membrane binding. Consistent with this model, StarLok2
179 and StarLok3 showed high solubility, but StarLok1 aggregated when concentrated in the absence
180 of detergents (**Supplementary Fig. 3**). Detergent-free StarLok1 could be stabilized at acidic pH,
181 which promotes increased cationic electrostatic repulsion along the amphipathic protein face, and
182 was additionally stabilized by acetate (**Supplementary Fig. 3c-e**). StarHod1, a close ortholog of
183 StarLok1 present in Hodarchaeales **(Fig. 2a)**, showed similar predicted structural features
184 (**Supplementary Fig. 4a**) and stabilization by acetate **(Supplementary Fig. 4b)**. Asgards,
185 including Lokiarchaeia, features an acetogenic metabolism, in which carbon dioxide is fixed into
186 acetyl-CoA via the Wood-Ljungdahl carbon fixation pathway[52]. Such organisms produce acetate
187 as a byproduct of ATP from acetyl-CoA[53], which might thus influence oligomerization or other
188 interactions of StarAsg1 proteins.
189
190 We next assayed the interaction of each Lokiarchaeia START protein with lipid membranes. Using
191 a dilute solution of each protein, we measured changes in light scattering upon incubation with
192 liposomes containing the neutral lipid di-oleoyl-phosphatidylcholine (DOPC) and the anionic lipid
193 phosphatidylserine (PS). Addition of StarLok1 triggered clustering of liposomes containing 5% or
194 higher PS concentrations, which were minimized in the presence of neutral DOPC liposomes
195 (**Fig. 4a**). In contrast, StarLok2 and StarLok3 showed minimal interactions with PS-containing
196 liposomes (**Fig. 4b**). We could also detect gravity separation of StarLok1 in solutions with PS-
197 containing liposomes, but not DOPC liposomes **(Supplementary Fig. 5g**). StarLok1 also showed
198 features consistent with its hypothesized hydrophobic nature, including aggregation in the
199 absence of liposomes (**Supplementary Fig. 3**), high levels of staining with hydrophobic dyes on
200 SDS-PAGE (**Supplementary Fig. 5c**), and accumulation at air-water interfaces in density-based
201 separation experiments (**Supplementary Fig. 5g**).
202
203 We also asked if StarAsg1 proteins can interact with cell membranes. For this, we expressed
204 StarLok1, Lok2, Lok3 and Hod1 fused to a C-terminal GFP in *Saccharomyces cerevisiae*.
205 Visualization of StarLok1-GFP in yeast cells revealed a strong plasma membrane (PM)

206 association with additional enrichment at regions surrounding the bud neck of cells. More
207 specifically, StarHod1 localized exclusively in the PM in a polarized manner (**Fig. 4b,**
208 **Supplementary Fig. 6**). This distribution echoed previous observation for a PS-binding sensor,
209 the C2 domain of lactahedrin ($C2_{LACT}$) in yeast cells[54]. Expression of the mRFP-$C2_{LACT}$ sensor in
210 StarLok1-GFP expressing cells confirmed colocalization of StarLok1 with membranes enriched in
211 PS. These experiments suggest that StarAsg1 proteins have specialized for binding to anionic
212 membranes in cells. In contrast, StarLok2 and StarLok3 remained soluble in the yeast cytosol
213 (**Fig. 4c**).
214
215 <u>StarLok1 binds phospholipids and transfers them between membranes</u>
216
217 Given that StarLok1 interacts with membranes in vivo, we asked if it also extracts lipid ligands
218 when expressed in cells, an established means to identify ligands of lipid transfer proteins[55]. We
219 isolated GFP-tagged StarLok1, 2, and 3 and StarHod1 from yeast cell pellets using affinity
220 purification with beads displaying anti-GFP antibodies (**Fig. 5a**). We then extracted any bound
221 lipids from protein displaying beads and analyzed their composition and abundance by LC/MS^2
222 (lipidomics). While StarLok2 and StarLok3 showed low lipid binding, StarLok1 and StarHod1
223 extracts contained nearly equimolar bound phospholipid (110 pmol) to the capacity of bound
224 proteins on the beads (150 pmol). Lipidomics showed that StarLok1-bound ligands were enriched
225 in PS, while StarHod1 ligands were enriched in phosphatidylcholine (PC) (**Fig. 5b**). Both also
226 bound phosphatidylethanolamine (PE). In contrast, lipids extracted from StarLok2 and StarLok3
227 samples were of lower abundance, and contained predominantly phosphatidylinositol (PI) and to
228 a lesser extent PC, reflecting background whole cell phospholipids[56]. These results indicated that
229 StarAsg1 proteins extract specifically phospholipids from cell membranes.
230
231 Based on its capacity to bind phospholipids in vivo, we asked if StarLok1 could act as a bona fide
232 phospholipid transporter. The movement of NBD-labeled phospholipid analogues from donor
233 liposomes can be quantified using Förster resonance energy transfer (FRET) to Rhodamine-PE
234 that is bound to acceptor liposomes (**Fig. 5c**), as previously described for START LTPs such as
235 STARD2, STARD10, PITPα/β and CERT[57–59]. Upon addition of detergent-solubilized StarLok1 to
236 donor liposomes containing NBD-PS, NBD-PC, or NBD-PE, a robust increase in FRET was
237 observed, indicating lipid transfer (**Fig. 5d**). These three substrates correspond to the
238 phospholipids enriched in StarLok1-GFP extracts compared to background binding in StarLok2/3-
239 GFP samples (**Fig. 5b**). Phospholipid transport was not observed beyond baseline when

equivalent detergent was added, and when the protein was digested with Proteinase K (**Supplementary Fig. S7a**), the specific lipid transfer was drastically inhibited. Using equilibrated liposomes, we quantified initial transport rates ranging between 1-2 lipids·min^{-1} per StarLok1 for these phospholipids (**Fig. 5e**). We also observed that transport NBD-labeled ceramide, a lipid that lacks a phosphate headgroup, was slower than that of the phospholipids, while transport of the FRET acceptor Rhodamine-PE, which contains a very large headgroup, was not detected in control liposomes (**Supplementary Fig. S7b-c**).

Sequence and structural phylogeny of START domains across the tree of life

START domains are well represented across the tree of life, including in non-Asgard archaeal lineages (Euarchaeota, TACK group, DPANN group, Thermoplastada) (**Fig. 3c**). Given their universal nature, we sought to understand the phylogenetic relationships across START domains and test the likelihood that eukaryotic LTPs and Hsp90 co-chaperones originated from Asgards. For this, we manually curated a set of 161 representative taxa containing eukaryotes, archaea, and bacteria across the tree of life (**Fig. 6b**). This yielded a dataset containing 748 START domains containing sequences

We applied both sequence-based Maximum Likelihood (ML) approaches and structural homology analysis to generate trees. For the former, we screened several sequence alignment algorithms to optimize the overall conservation score of different sequence subsets and generated several ML trees (**Supplementary Fig. 8 to Fig. 14)** based on subsets of the dataset to address specific questions. For the latter, we employed the FoldTree[60] algorithm, which generates trees based on structural divergences measured by Foldseek[61], on the whole dataset. Because START domains often exist in tandem repeats in eukaryotic genomes or as parts of multi-domain proteins, we developed a tool (ProtDomRetriever) to isolate and align individual domains (**Fig. 6b**). While the conclusions were consistent across both approaches, FoldTree analysis revealed the evolutionary history of the START domain with deeper resolution and accurately identified the evolutionary history of known taxa (**Supplementary Fig. 15 and 16**). This is likely due to the high sequence plasticity in START domains – less than 25% of proteins passed the Chi2 composition test – that resulted in high entropy and gap rich sequence alignments that are challenging for ML analysis. However, the overall structural fold remains constant in this superfamily, making it well-suited for structural homology.

All our trees estimated an ancient origin to the START-domain that preceded eukaryogenesis. Sequence alignment-based ML phylogeny (**Fig. 6a**) and structure-based trees (**Fig. 6c**) summarized by a pruned cladogram (**Fig. 6d**) showed a distinct branching group consisting of mitochondrial Coenzyme Q-binding proteins (COQ10-like class) and Ribosome association toxin A (RatA) in bacteria, consistent with previous groupings[62]. The divergence of COQ10 from alphaproteobacterial RatA homologs was also supported by strong branch support (aLRT 0.947; UFBoot 0.94) in ML trees (**Fig. 6a**). These analyses reflect a likely bacterial origin of the nucleus-encoded but mitochondrial-localized COQ10 binding protein.

Our analyses reveal an archaeal origins of co-chaperones of Hsp90 (Aha1, ASHA1, ASHA2). Phylogenetic analyses indicate that Aha1-like proteins originate within Asgards and form a monophyletic group including nine StarLok3 asgard homologs: StarAsg3, which diverged from StarAsg2 clade in Asgards. The grouping of AHA1 and StarHod3 is strongly supported (SH-aLRT; UFBoot = 0.948; 0.9) and this clade is within the broader StarAsg3 lineage (0.938; 0.79) (**Fig. 6a**). In the full dataset, StarAsg3 and AHA1 consistently form a monophyletic group with StarAsg2-like proteins, albeit with low support (**Fig. 6a**, **Supplementary Fig. 8**). In contrast, when restricting the analysis to prokaryotic START domain only, StarAsg2 and StarAsg3 form a strongly supported monophyletic clade (0.975; 0.92; **Supplementary Fig. S13**), consistent with earlier analyses **(Fig. 1a; Supplementary Fig. 2)**. We identified several bacterial START domains with structural similarity to StarAsg2 (but not StarAsg3) **(Fig 6c)**, which also cluster weakly with StarAsg2 in ML trees (**Supplementary Fig. 14**). However, to our knowledge, no experimental evidence supports interaction between bacterial START proteins and Hsp90. Thus, AHA1-like proteins are likely to have been inherited from an ancestral StarAsg3 lineage in Asgards prior to the emergence of the first eukaryotic common ancestor.

In eukaryotes, START domains are the basis for multiple classes of LTPs that shuttle lipids intracellularly. All the LTP START domains in our dataset find their roots with protein sequences in Heimdallarchaeia (A0A1Q9NB21, A0A1Q9P3Z5, A0A1Q9NZ53 and A0A1Q9NLH4), which belong to the StarAsg1 clade. Although the available phylogenetic signal does not allow a definitive assignment of START domain LTPs to an Asgard origin, several ML reconstructions and the FoldTree analysis consistently position Heimdallarchaeia StarAsg1-like sequences near the base of eukaryotic LTP diversity. In the global dataset **(Fig. 6a**), StarHod1 (A0A1Q9NB21), together with A0A1Q9P3Z5, A0A1Q9NZ53 and A0A1Q9NLH4, forms a monophyletic group associated with eukaryotic START LTPs. In alternative reconstructions, individual StarAsg1

308 sequences cluster within the eukaryotic LTP assemblages (**Supplementary Fig. S10**). When
309 restricting the analysis to eukaryotic START LTPs and StarAsg1 homologs **(Supplementary Fig.**
310 **9)**, StarAsg1 sequences form a well-supported clade (97.9; 89) together with Bet v 1-like proteins,
311 including plant PYL1; this combined assemblage connects to the broader STARD-like LTP and
312 PITP radiation, albeit with weaker support (58.5; 53).
313
314 All StarAsg1-like sequences belong to a group characterized by an enlarged predicted binding
315 cavity, a structural feature that is enhanced in eukaryotic START LTPs **(Supplementary Fig.**
316 **17a)**. Structurally related START proteins were identified in other archaeal lineages **(Fig. 3d)**,
317 several of which grouped with StarAsg proteins in Foldseek-based structural analyses
318 **(Supplementary Fig. 17b)**; however, these non-Asgard proteins display smaller or poorly defined
319 predicted binding cavities **(Supplementary Fig. 17c-d)**, suggesting that they may not function as
320 LTPs. Based on these analyses and the established relationship of eukaryotes within Asgard
321 archaea, a parsimonious grouping is that eukaryotic START domain LTPs diversified from an
322 inherited StarAsg1 protein. Two alternative hypotheses – that START domain LTPs diversified
323 from StarAsg2/3 (**Supplementary Fig. 11**) or from bacterial-derived START domains
324 (**Supplementary Fig. 12**) – are not supported by ML trees. Horizontal gene transfer of an
325 ancestral START domain LTP to eukaryotes is also unlikely unless missing prokaryotic phyla
326 contributed to eukaryogenesis, as we did not detect taxonomic clades branching in eukaryotic
327 LTPs.
328

329 ## Discussion

330
331 Trafficking of phospholipids, given their negligible solubility, is a hallmark of eukaryotic cells and
332 is a prerequisite for the stable coexistence of multiple membrane organelles. Identification of LTPs
333 within Asgard would support a model in which eukaryotic ancestors already contained proteins
334 supporting this key eukaryotic function. To this end, we described three conserved START
335 domain proteins in Lokiarchaeia, the same fold that forms the basis for a large and diversified
336 group of eukaryotic LTPs. One, StarLok1, binds phospholipids and transports them between
337 membranes. It is a member of a large clade of Asgard START domains (StarAsg1) that contain
338 predicted structural features found in LTPs: a large, hydrophobic binding pocket lined with an
339 amphipathic and basic region for interaction with anionic lipid bilayers.
340

341 Phylogenetic and structural homology analysis suggests a broader relationship between Asgard
342 START proteins and those in eukaryotes. Structural and sequence alignments were able to
343 generate a putative tree of START domains across kingdoms suggested that STARD-like proteins
344 and similar eukaryotic LTPs emerged from an ancient START domain fold, which could have been
345 present in Asgards as StarAsg1 homologs. A second class of Asgard START domains, StarAsg3,
346 includes ancestral proteins for Activator of Heat Shock Protein 90 (Hsp90) ATPases (AHAs), an
347 essential co-chaperone for protein stability and maturation in eukaryotes. Another class of START
348 domain proteins conserved in eukaryotes, the Coenzyme-Q binding protein 10, shares a common
349 ancestry with bacterial Ribosome association toxin A proteins and was thus likely inherited from
350 the alphaproteobacterial ancestor of mitochondria, consistent with previous reports[62,63]. To
351 specialize for function as an LTP, protein families like START domains would need to adopt
352 enlarged hydrophobic binding cavities lined by a membrane-interacting motif, and that is the
353 feature we observe in StarAsg1 proteins compared to their sister groups StarAsg2/3. More
354 broadly, we observe that an increase in binding cavity size is a general feature for the
355 specialization of START domains for lipid binding across both Asgards and eukaryotes
356 (**Supplementary Fig. 17a**).
357
358 Our phylogenetic analysis also supplements previous classifications of START domain LTPs
359 within eukaryotes[42,64,65] with several additional details (**Fig. 6d**). These include orthology between
360 STARD4/5/6 and STARD1/3; both of which radiate independently in animals and their closest
361 unicellular relatives, like *Capsaspora owczarzaki*. Similarly, the two START domains containing
362 acyl-CoA thioesterases (STARD14 and STARD15) also show orthology and arose prior to the
363 divergence of animals and *C. owczarzaki*. More distantly, CERT (STARD11) is grouped as an
364 independent class. We also observed orthology between STARD8, STARD12, and STARD13, all
365 of which contain a START domain fused to a small GTPase activating protein Rho domain, which
366 are also common to choanoflagellates. We observed an earlier separation of STARD10 orthologs
367 from an ancestral EDR2-like START domain in plants, with occurrence in a broader group of
368 unicellular eukaryotes, however the position of the STARD10 group remained ambiguous. Finally,
369 our tree suggests that STARD2 (PCTP) and STARD7 form a paralogous sister group from other
370 members of the STARD group. Finally, the PITPs group, which are phosphatidylinositol and
371 phosphatidic acid transporters[66], consistently diverged from STARD proteins at the earliest stage
372 of START LTPs radiation in eukaryotes, supporting the acquisition of phosphatidylinositol as key
373 lipid in eukaryotes[67].

374 Recent advances suggest that lipid transport could be relevant in Asgard cells, even if they lack
375 complex organelles that evolved during eukaryogenesis. Heimdallarchaeia have been recently
376 observed to contain large intracytoplasmic vesicles that could require lipid transport activity to
377 support[34]. Importantly, bound ligands and transported substrates of StarLok1 include PS, a
378 phospholipid with an equivalent in archaea archaeatidylserine synthase (AS)[68]. Lokiarchaea MK-
379 D1 contains a phosphatidylserine synthases (PSS) homolog (UniProt: A0A5B9DD23) containing
380 key residues for serine binding (**Supplementary Fig. S18)** and thus is likely to natively contain
381 phospholipids with an identical polar group as PS. Although archaeal phospholipids most
382 commonly feature an opposite chirality to that of those found in eukaryotes, this aspect is unlikely
383 to be a hindrance for LTPs that bind ligands based on cavity size and hydrophobicity, and ionic
384 interactions with headgroups features. Several aspects of Asgard lipid biology are still being
385 investigated, but AS and other phospholipids represent a likely in vivo substrate for StarLok1.
386
387 Both StarLok1 and StarHod1, another StarAsg1 protein in Hodarchaeala, also bind the plasma
388 membranes in yeast cells, which are enriched in PS, as well as other anionic phospholipids. The
389 ubiquitous polarity of StarAsg1 proteins, containing basic patches near their ligand binding pocket,
390 suggests that an anionic inner membrane is widespread in Asgards, where it could be supported
391 by the activity of AS synthase and putative archaeatidylinositolphosphate synthases
392 (A0A5B9DAH1)[32]. A highly anionic cytosolic leaflet and more neutral outer leaflet is a hallmark of
393 eukaryotic cells that dictates the properties of membrane-associating proteins, like LTPs. It is
394 possible that this membrane feature also has its origins in archaeal ancestors, given the noted
395 asymmetric nature of archaeal bolalipids that contain anionic phosphoheadgroups on the
396 cytosolic leaflet and neutral glycosylations on exoplasmic leaflets[69]. In this model, eukaryotic
397 cytosolic proteins that interact with membranes, like LTPs, might have originated from the
398 archaeal ancestor, even if their backbone and phospholipid tails differ. Additionally, the
399 asymmetric distribution of anionic lipids is essential for phagocytosis and vesiculation in
400 eukaryotes[70–73], thus the presence of anionic lipid in the cytosolic leaflet of Asgard membranes
401 could be relevant for the intracellular vesicles observed in Asgards[34].
402
403 The first eukaryotic LTPs could have acted as metabolic facilitators by fortuitously transferring
404 newly synthesized lipids between membranes, enabling the proliferation of intracytoplasmic
405 compartments. Ancestral START LTPs could have fulfilled this function with their large,
406 hydrophobic binding cavity and membrane-binding capacity. Later, the emergence of more
407 sophisticated LTPs capable of coupling the transfer of multiple lipids, such as the Osh/ORP

408 proteins that counter-exchange lipid ligands for PI(4)P[74] would have enabled the generation of
409 heterotopic organelle membranes with specialized lipid compositions. Bridge-like LTPs, such as
410 those in the VPS13 class, could have emerged due to the need for rapid membrane proliferation
411 in emerging aspects of eukaryotic cell biology, including mitochondrial network extension[75], the
412 generation of autophagosomes[76] or spores[77], and were later adapted to support routine lipid
413 transfer in larger, increasingly compartmentalized cells, in which contact sites emerged as central
414 lipid transfer hubs[78,79]. In this way, the increasing complexity of eukaryotic cell biology could have
415 been enabled by biochemical advances in the ability of proteins to transport lipids. Further
416 integration of evolutionary and biochemical approaches could help resolve how adaptations in
417 cellular architecture and lipidomes co-arose during eukaryogenesis.
418

419 ## Acknowledgments

420 Yongxuan Su and the UCSD Molecular Mass Spectrometry Facility aided in analyses. Deshmukh,
421 Devaraj, Ghosh and Herzik labs provided instrumentation support. Paula Welander and Buzz
422 Baum provided helpful discussions. The work was supported by the National Institutes of Health
423 (GM142960), the Moore–Simons Project on the Origin of the Eukaryotic Cell (GBMF9734, grant
424 DOI 10.37807/GBMF9734), and the Paul G. Allen Family Foundation.
425

426 ## Data Availability

427 Alignments and trees generated are available at iTol (https://itol.embl.de/shared/BudinLab),
428 scripts used at GitHub (https://github.com/NicoFrL/ProtDomRetriever and
429 https://github.com/NicoFrL/MAFFT_ScoreNGo). Data files used to generate figures will be made
430 available on FigShare.
431

432 ## Competing interests

433 The authors have no competing interests to declare.
434

435 ## Materials and methods

436 **Nomenclature**

437 START: StAR-related transfer, it refers to START-like domain superfamily *stricto sensu (IPR:*
438 IPR023393*).* In opposition to other "Starkin" domains such as members of VAD1 Analog of StAR-
439 related lipid transfer family and PRELI/MSF1, both display a helix-grip fold analog to START
440 domains. StAR: Steroidogenic acute regulatory protein. SRPBCC: NCBI Conserved Protein

441 Domain superfamily associated to the START domain (NCBI: cl14643),
442 START/RHO_alpha_C/PITP/Bet_v1/CoxG/CalC ligand-binding domain superfamily. STARD:
443 (StAR)-related lipid transfer domain.
444
445 **Lipid transfer protein classification analysis**
446 The 14 classes of LTP we identified were previously described[80–84]. The Interpro accession
447 number of each of this class was used to determine the number of taxonomic species expressing
448 at least one of these proteins: 1) Tgt2/MlaC (Superfamily IPR042245), 2) LolA/LolB/Lppx
449 (Superfamily IPR029046), 3) START-like domain (Superfamily IPR023393), 4) SCP2 sterol-
450 binding domain (superfamily IPR036527), 5) Calycin-Lipocalin (superfamily IPR012674), 6)
451 PRELI/MSF1 (Domain IPR006797), 7) LBP/BPI/CEPT (i.e. Bactericidal permeability-increasing
452 protein, alpha/beta domain; Superfamily IPR017943), 8) nsLTP (i.e. Bifunctional inhibitor/plant
453 lipid transfer protein/seed storage helical domain; Superfamily IPR036312); 9) CRAL/Trio
454 (Superfamily IPR036865), 10) ORP/Osh (Superfamily IPR037239), 11) GM2-AP (Superfamily
455 IPR036846), 12) GLTP (Superfamily IPR036497), 13) NPC1 N-ter (i.e. Niemann-Pick C1, N-
456 terminal; Domain IPR032190), 14) VASt (i.e. VAD1 Analog of StAR-related lipid transfer; Domain
457 IPR031968).
458
459 **Expression and purification of StarLok1**
460 A codon optimized 444 bp cDNA coding for StarLok1 (UniprotKB: A0A5B9D5D6) C-terminally
461 fused to a thrombin cleavage site (LVPRGS) followed by a 6xHis tag was synthesized and cloned
462 into a pET-21a(+) vector (GeneScript) and expressed in BL21-Gold(DE3) (Agilent). StarAsg1 was
463 purified either from native soluble fraction or refolded from an insoluble fraction. For native
464 purification, a 1 L culture LB Lennox with 50 *µ*g/mL ampicillin was grown to an OD_{600} of 0.5 at
465 37°C then induced (0.2 mM IPTG) before overnight growth at 16°C. Cell pellets were resuspended
466 in cold TN buffer (50 mM Tris, 500 mM NaCl, pH 8.0, at 4°C) supplemented with protease inhibitor
467 cocktail (Roche cOmplete EDTA-free, 10 µM bestatin, 1 µg/mL pepstatin A, 10 µM
468 phosphoramidon), 10 mM imidazole, 1 mM DTT, 20 µg / mL of DNase I (Roche), 5 mM $MgCl_2$
469 and 1 mg/mL of lysozyme. Cells were then disrupted by sonication and the lysate was clarified by
470 ultracentrifugation (186,000 x *g,* 1h, 4°C). Fifty volumes of supernatant were applied to one
471 volume of Ni-NTA resin slurry (cOmplete His-Tag Purification Resin) and incubated overnight at
472 4°C. The resin suspension was applied to an Econo-Pac chromatography column (Bio-Rad)
473 successively washed with TN containing 10 mM and 20 mM imidazole. The protein was eluted
474 with TN containing 250 mM imidazole and the eluate concentrated at 2500 *x g* using Amicon Ultra

475 centrifugal unit (MWCO 3000). The protein was further purified by SEC on an Akta pure 25 FPLC
476 with a Sephacryl S200 HR column equilibrated with TN buffer at a flow rate of 1mL/min and
477 fractionation volume of 2.5 mL. Glycerol was added to the buffer (10% v/v) and protein-containing
478 fractions were pooled and concentrated up to 20 μM. Concentration was determined by both A_{280}
479 (DeNovix DS-11 spectrophotometer) and Sypro Orange (ThermoFisher) staining of SDS-PAGE
480 using BSA standards and samples were frozen in liquid N_2 and stored at -80°C.
481
482 For refolding of insoluble protein, StarLok1 expression was performed as described in the
483 previous section with minor modification. When OD_{600} reached 0.6, 1 mM IPTG was added into
484 the culture that was then incubated three hrs at 37°C before collecting the cells and preparing the
485 pellet for freezing. Pellets were resuspended in 50 mM HEPES-KOH pH 7.00 with a protease
486 inhibitor cocktail (Roche cOmplete EDTA-free, 10 µM bestatin, 1 µg/mL pepstatin A, 10 µM
487 phosphoramidon) and passed 4 times through a high-pressure homogenizer at 1000 bar
488 (EmusiFlex-c3, Avestin). 20 µg/mL of DNase I (Roche), 5 mM $MgCl_2$ were added, and the
489 inclusion bodies were then pulled down by ultracentrifugation (186,000 x *g*, 1 hr, 4°C). The
490 supernatant was discarded, and the white pellet was solubilized in HU buffer (50 mM HEPES-
491 KOH pH 7.0 with 6M urea) by mild vortexing and pipetting and one passage through high-pressure
492 homogenizer (1000 bar). After centrifugation (186,000 x *g,* 1 hr at 4°C), 10 mM imidazole was
493 added to the clear supernatant, which was applied to 50:1 to Ni-NTA resin slurry (cOmplete His-
494 Tag Purification Resin) and incubated 3 hr at 21°C. The suspension was applied to an Econo-
495 Pac chromatography column (Biorad), washed twice with 10 bed volumes of HU with 20 mM
496 imidazole, and eluted with eight bed volumes of HU with 250 mM imidazole. Refolding was
497 performed by dialysis (3.5K MWCO, Slide-A-Lyzer™ MINI Dialysis Devices) twice against 150
498 volumes of acetate buffer (164 mM acetic acid (OHAc), 36 mM sodium acetate (NaOAc), pH of
499 4.04) over 14 hr. Screening for refolding conditions was initially done by light microscopy
500 (ThermoFisher EVOS) after low-volume dialysis. Refolded samples were further purified by FPLC
501 on a HiPrep 16/60 Sephacryl S200 HR column (Cytiva) equilibrated with acetate buffer at a flow
502 rate of 1 mL/min and fractionation volume of 2.5 mL. Purification and refolding were monitored by
503 SDS-PAGE. This protocol yielded 5 to 7 mg of protein per 1 L of bacterial culture, as determined
504 by A_{280}.
505
506 For detergent solubilized protein used in lipid transfer assays, expression in *E.coli* BL21 was
507 induced with 1 mM IPTG once the OD_{600} reached 0.95 and incubated for 3 hr. Bacterial pellets
508 were collected as previously described. Bacterial pellets were resuspended in PBS with 1% N-

509 lauroyl-sarcosine (NLS) and 10 mM imidazole, 5 mM MgCl2 and 20 µg / mL of DNAse I. Bacteria
510 were lysed by sonication and the lysate was clarified by ultracentrifugation (186,000 x *g*, 1 hr,
511 4°C). The supernatant was then applied to a Ni-NTA resin slurry and incubated for 3 hr. After two
512 successive washes (PBS, 1% NLS, 10 mM imidazole), StarLok1 was eluted with 250 mM
513 imidazole. The eluate (12 mL) was adjusted with PBS buffer and no detergent, and concentrated
514 to 2 mL using an Amicon Ultra centrifugal unit (MWCO 3000), diluted again to 13 mL with PBS
515 buffer and concentrated once more until the solution started to oil-out around 500 µL. The protein
516 concentration (1.53 mM) was determined by absorbance spectrometry at 280 nm. The sample
517 was stored at 4°C, and purity and stability was determined by SDS-PAGE prior NBD-lipid transfer
518 assay.
519
520 **Expression and purification of StarLok2 and StarLok3**
521 A codon optimization 489 bp cDNA coding for StarLok2 (UniprotKB: A0A5B9DET0) and 432 bp
522 coding for StarLok3 (UniprotKB: A0A5B9DAG8), each N-terminally fused to a 6xHis tag followed
523 by a thrombin cleavage site (LVPRGS), were synthesized and cloned into a pET-28a(+) vector
524 (GeneScript) and expressed in BL21-Gold(DE3) (Agilent). Each protein was purified from 1 L
525 culture in its native form as described for StarLok1 in TN buffer supplemented with 1 mM DTT. In
526 this standard condition, the protocol yielded more than 10 mg of protein per 1 L of bacterial culture.
527 Protein concentration and purity were controlled by SDS-PAGE stained with InstantBlue (Abcam).
528
529 **Circular Dichroïsm**
530 CD was performed on a Jasco J-1500 spectrometer using a 350 µL synthetic quartz glass cell
531 with a 1 mm path length (CV1Q035AE2, ThorLabs). Measurements were taken of StarAsg1 in
532 either 20 mM Tris, 120 mM NaF, pH 7.4 (native), or in a solution prepared by diluting the stock
533 refolded protein (HOAc/NaOAc, pH 4.0) in milli-Q water. Each spectrum represents the average
534 of five continuous scans recorded from 185 to 260 nm with the following parameters: bandwidth
535 of 1 nm, step size of 0.1 nm, scan speed of 50 nm·min−1, CD scale of 200 mdeg/0.1 dOD, and
536 digital integration time (D.I.T) of 4 s. A control spectrum of buffer samples without protein was
537 subtracted from each protein spectrum. For temperature interval scans, an equilibration time of
538 20 s was set before each scan at 10°C intervals. The measured ellipticity was converted to molar
539 ellipticity, considering the cuvette path length and the mean residue concentration of proteins. For
540 secondary structure determination, the spectra were analyzed in the 185–240 nm range using the
541 SELCON3 algorithm and the SP175t dataset on the DICHROWEB server[85] and compared with

542 the algorithm available on the BeStSel web server[86]. Thermal denaturation results were displayed
543 using matplotlib.
544
545 **Protein structure analysis**
546 Structures were rendered using Pymol (v2.5.8, Schrödinger, LLC). Structure alignment
547 comparisons were conducted using the cealign algorithm and the RMSD values were reported.
548 Model-template alignments were obtained using SwissModel[87]. Cavities point clouds were
549 calculated using the CavitOmiX plugin (v. 1.0, 2022, Innophore GmbH)[88] and cavities were
550 calculated using a modified LIGSITE algorithm[89]. To compute the cavity size, a script calling
551 CavFind and LigSite modules from CavitOmiX was used to extract the number of cavities, their
552 respective volumes and ligsite score. Protein electrostatic surface potential was determined using
553 APBS (pdb2pqr)[90] with a grid spacing of 0.26. Hydrophobicity was determined using the
554 Eisenberg scale[91]. The Position of Protein on Membrane (PPM) was determined using PPM 3.0
555 web server[92], using a flat theoretical archaeal plasma membrane composed of
556 Phosphatidylglycerol, Cardiolipin, Archaeol and Menaquinone 8 at a ratio 64:7:8:21. Changing
557 the membrane composition did not significantly change the obtained results. The AlphaFold2
558 structure of StarLok1 with a thrombin cleavage linked to 6x(His)-tag was computed using
559 ColabFold v1.5.5 notebook. The secondary structure composition was determined using either
560 STRIDE[93] and DSSP algorithm[94]. The theoretical CD spectra of StarLok1 was determined using
561 either SESCA[95] and PDB2CD[96] for comparison with experimental measurement.
562
563 **Protein mass spectrometry**
564 StarLok1 sample purified in native condition was dialyzed three times (1 hr, overnight, and 1 hr)
565 against 1 L of TN buffer (10 mM Tris-HCl, 50 mM NaCl, pH 7.4) using a 0.5 mL Slide-A-Lyzer
566 cassette, to ensure complete glycerol removal. For StarLok1 purified in acetate buffer, no dialysis
567 was performed.. Protein samples were analyzed for purity and mass identification by liquid
568 chromatography with an Agilent 6230 time-of-flight mass spectrometer (TOFMS) with JetStream
569 electrospray ionization source ESI (LC-ESI-TOFMS)
570
571 **Liposome preparation**
572 Stock solutions of 1,2-dioleoyl-sn-glycero-3-phosphocholine (DOPC), 1-palmitoyl-2-{12-[(7-nitro-
573 2-1,3-benzoxadiazol-4-yl)amino]dodecanoyl}-sn-glycero-3-phosphoserine (16:0-12:0 NBD-PS),
574 1-palmitoyl-2-{12-[(7-nitro-2-1,3-benzoxadiazol-4-yl)amino]dodecanoyl}-sn-glycero-3-
575 phosphoethanolamine (16:0-12:0 NBD-PE), 1-palmitoyl-2-{12-[(7-nitro-2-1,3-benzoxadiazol-4-

576 yl)amino]dodecanoyl}-sn-glycero-3-phosphocholine (16:0-12:0 NBD-PC), N-[12-[(7-nitro-2-1,3-
577 benzoxadiazol-4-yl)amino]dodecanoyl]-D-erythro-sphingosine (C12-NBD-Ceramide) and 1,2-
578 dipalmitoyl-sn-glycero-3-phosphoethanolamine-N-(lissamine rhodamine B sulfonyl) (Rhodamine-
579 PE), 1-palmitoyl-2-oleoyl-sn-glycero-3-phospho-L-serine (POPS) in chloroform (Avanti Polar
580 Lipids) were mixed at the desired molar ratio. The solvent was evaporated under vacuum (54.9
581 mbar) for 30 min on a rotary evaporator at 37°C. Residual solvent was eliminated by further drying
582 in a vacuum chamber for at least 1 hr. The resulting lipid films were hydrated with 2 mL HK buffer
583 (50 mM HEPES-KOH, 120 mM K-acetate, pH 7.4) and vigorously vortexed with glass beads. The
584 resulting multilamellar vesicles suspension was subjected to five freeze-thaw cycles (liquid
585 nitrogen/37°C) followed by extrusion through a 0.2 µm polycarbonate filter using a mini-extruder
586 (Avanti Polar Lipids). Freshly extruded liposomes prepared on the same day of the experiment or
587 used for no more than two days were shielded from light were used for all experiments.
588
589 **Dynamic Light Scattering**
590 Measurements were performed using a DynaPro instrument with a Flex-99 correlator
591 (ProteinSolutions) and quartz cuvettes. The instrument sensitivity was set to 30% with a maximum
592 acquisition time of 10 s and a signal-to-noise threshold of 2.5. The cuvette holder temperature
593 was maintained at 25°C. The buffer refractive index was set to 1.333. For kinetics experiments,
594 20 min autocorrelation curves were acquired from 200 nm extruded liposomes (50 µM lipid
595 concentration) in degassed HK buffer (50 mM HEPES-KOH, 120 mM K-acetate, pH 7.4) with 1
596 µM of StarLok1 in acetate buffer added after 1 min. Alternatively, StarLok2 and StarLok3 were
597 tested. Data filtering and analysis were performed using DYNAMICS software (v.5.25.44) with a
598 maximum Sum Of Square (SOS) limit at 2100, under baseline limit at 0.995, over baseline limit
599 at 1.02, under amplitude limit of 0.05. For the correlation function, the first 2 coefficients were
600 ignored, and the analysis was truncated at channel 120. The autocorrelation functions were then
601 analyzed using the regularization algorithm.
602
603 **Lipid transfer assay**
604 NBD-lipid transfer was measured as previously described[59] with modifications. Fluorescence was
605 recorded using a Cary Eclipse Fluorescence Spectrophotometer (Agilent) equipped with a
606 temperature controller set to 30°C. In a 500 µL quartz SUPRASIL rectangular cuvette
607 (PerkinElmer), stirred with a magnetic bar, 200 µM donor liposomes (L_A) containing 2 mol% Rhod-
608 PE and 5 mol% NBD-labeled lipid (NBD-PC, NBD-PE, NBD-PS, or NBD-ceramide) were
609 suspended in HK buffer (50 mM Hepes, 120 mM K-acetate, pH 7.4) pre-warmed at 30°C. NBD

610 fluorescence was monitored at 538 nm (emission slit = 5 nm) with excitation at 473 nm (slit = 2.5
611 nm), photomultiplier voltage of 800 V, and acquisition interval of 0.5 s. At t = 1 min, 200 μM
612 acceptor liposomes (L_B; DOPC only) were added. After 3 min of slow spontaneous transfer,
613 StarLok1 (1.60 μL; 5 μM final concentration) was injected, resulting in a significant increase in
614 fluorescence indicative of NBD-lipid transfer from L_A to L_B. As a negative control, 1.73 μL of
615 proteinase K-digested StarLok1 was tested. Digested samples were prepared by incubating
616 StarLok1 with Proteinase K (20 mg/mL, RNA grade; Invitrogen) at a 1:12 (v/v) ratio for 1 h at
617 42°C, followed by heat inactivation for 5 min at 98°C. Complete proteolysis was confirmed by
618 SDS-PAGE analysis. The amount of NBD-lipid transferred was quantified by measuring the loss
619 of FRET between NBD and Rhod-PE, normalized as:

620 $$\text{Conc} = 5 \times ((F - F_{min}) / (F_{max} - F_{min}))$$

621 where F is the recorded fluorescence, F_{max} is the mean fluorescence over 30 s prior to protein
622 injection, and F_{min} is the mean fluorescence at equilibrium (t = 12 min, averaged over 3 min)
623 determined in an independent experiment without protein using pre-equilibrated liposome
624 populations L_A (2 mol% Rhod-PE) and L_B, both containing 2.5 mol% NBD-lipid. The coefficient, 5
625 reflects the 5 μM NBD-lipid accessible in the external leaflet of L_B. Baseline correction was
626 performed by subtracting the mean normalized signal from control experiments (n = 3) in which
627 1.64 μL of vehicle buffer (PBS containing 1% N-lauroylsarcosine) was injected instead of protein;
628 no increase in lipid transfer was observed under these conditions. Initial transfer rates (V_i ± SEM)
629 were determined from the slope of a linear fit (least-squares regression, unweighted) to data
630 collected over 3 s post-injection, treating each replicate as an individual data point. Statistical
631 analysis was performed using two-way ANOVA with Tukey's post-hoc test. All analyses were
632 conducted in GraphPad Prism (v10.2.3).
633

634 **Yeast expression**

635 *E. coli* codon optimized sequences for StarLok1, StarLok2, StarLok3 and StarHod1 were cloned
636 into the yeast integration vector pCD256 under a TEF1 promoter and fused with a C-terminal
637 yeGFP. The expression construct was integrated into the *URA3* locus of background strain
638 W303a after double digestion of pCD256 with Pme1/Pac1 and chemical transformation. Clones
639 were isolated by selective plating with G418 and grown in CSM medium with 2.0% w/v glucose.
640 The C2 domain of Bovin Lactadherin N-terminally fused to mRFP was a gifted from Sergio
641 Grinstein (Addgene plasmid #74061) was cloned into a pRS415 vector under the control of a
642 PGK1 promoter and terminator sequences. Yeast cells expressing StarLok1 were transformed
643 and selected on CSM-Leu with 250 μg/mL G418, then grown in CSM-Leu medium with 2.0% w/v

644 glucose. For imaging, stationary phase cells were added to concanavalin-1A-coated 8-well
645 coverslip dishes and imaged with a Zeiss LSM 880 confocal laser scanning microscope equipped
646 with an AxioObserver stand and a Plan-Apochromat 63×/1.4 NA Oil DIC M27 objective. Images
647 were acquired with the following settings: GFP was excited using a 488 nm argon laser at 2.3%
648 power, and fluorescence emission was collected in the 494.95-565.07 nm range using a
649 photomultiplier tube (PMT) detector set at 750 V. Brightfield (BF) images were simultaneously
650 captured using a non-descanned transmission PMT detector. Imaging parameters included a
651 pixel dwell time of 4.1 µs, 8× line averaging, and a confocal pinhole of 0.997 Airy units (50.4 µm)
652 to optimize the signal-to-noise ratio while maintaining optical sectioning. Images were digitized as
653 16-bit single optical sections with a frame size of 1024×1024 pixels and subsequently cropped to
654 512×512 pixels for further processing. The physical pixel size was 65.9 nm, yielding a field of view
655 of 33.7×33.7 µm. For super-resolution imaging, the same microscope was used with its Airyscan
656 detector. Z-stacks were collected with a step size of 0.18 µm, covering a total depth of 11.88 µm
657 (66 slices). For optimal resolution, the pixel size was set to 0.04×0.04 µm in the x-y dimension,
658 yielding voxel dimensions of 0.04×0.04×0.18 µm. For dual-channel imaging (FFP/mRFP), Z-
659 stacks were collected with 0.185 µm steps (75 slices, 13.87 µm total depth) at 0.0488×0.0488 µm
660 pixel size, yielding voxel dimensions of 0.0488×0.0488×0.185 µm. Excitation was performed
661 using 488 nm (argon laser) and 561 nm (DPSS laser) at 2% power each. Raw Airyscan images
662 were processed using the ZEISS ZEN software with the standard Airyscan processing algorithm
663 in 3D mode. In FIJI (imageJ), the "Sharpen" filter was applied once to enhance structural details.
664 No additional processing was performed on individual slices to maintain the integrity of the original
665 signal distribution. For visualization purposes, all images were cropped to 512×512 pixels to focus
666 on the regions of interest.
667
668 **Isolation of GFP-tagged Asgard archaeal proteins for lipidomic analysis**
669 StarLok1, StarLok2, StarLok3, and StarHod1 GFP-fusion proteins were expressed in S.
670 cerevisiae W303 grown in 50 mL YPD medium at 30°C for 24 h, following 1:100 inoculation from
671 an overnight culture. Cells were harvested by centrifugation (3,000 ×g, 5 min, 4°C), washed once
672 with ice-cold sterile water, and resuspended in 500 µL lysis buffer (10 mM Tris-HCl, 150 mM KCl,
673 1 mM $MgCl_2$, pH 7.2; filtered and degassed) supplemented with 1 mM DTT, Complete EDTA-free
674 protease inhibitor cocktail (1 tablet per 20 mL; Roche), and PhosSTOP phosphatase inhibitor (2
675 tablets per 20 mL; Roche). Cell lysis was performed by cryogenic grinding. The cell slurry was
676 transferred to a ceramic mortar pre-chilled and maintained at −196°C with liquid nitrogen and
677 ground to a fine powder with a pestle. The frozen powder was transferred to a 2 mL

678 microcentrifuge tube and thawed on ice. To reduce viscosity, samples were supplemented with
679 50 µL DNase I solution (20 µg/mL DNase I Grade II from bovine pancreas [Roche] in 5 mM
680 $MgCl_2$). Lysates were then clarified by centrifugation (20,000 ×g, 15 min, 4°C), yielding
681 approximately 1 mL of supernatant. For affinity purification, clarified lysate was transferred to a 2
682 mL round-bottom tube containing 50 µL GFP-Trap Magnetic Particles (M-270; Chromotek), pre-
683 washed and equilibrated in a wash buffer (10 mM Tris-HCl, 150 mM KCl, 1 mM $MgCl_2$, pH 7.2;
684 filtered and degassed) according to the manufacturer's instructions. The suspension was
685 incubated for 1 h at 4°C with end-over-end rotation. Beads were washed five times with the wash
686 buffer and stored at 4°C until lipid extraction.

687 **Analysis of lipids bound to GFP-StarAsgard proteins**

688 Total lipids bound to GFP-Star Asgard protein were extracted via an adapted BUME method[97].
689 Magnetic particles bound to GFP-StarAsgard proteins were resuspended in 500 µL of 3:1 (v/v)
690 butanol:methanol and vortexed until complete resuspension. The suspension was transferred to
691 a 12×75 mm conical glass tube, combined with 500 µL of 3:1 (v/v) heptane:ethyl acetate and 500
692 µL of 1% acetic acid, and centrifuged for 5 min at 3,100 ×g. The upper phase (450 µL) was
693 transferred to an amber glass vial with a PTFE rubber-lined cap (Wheaton) and evaporated at
694 40°C for 1 h using [evaporation method]. The dried lipid extract was resuspended in 50 µL of
695 18:1:1 (v/v/v) 2-propanol:dichloromethane:methanol and vortexed for 3 min. Samples were
696 centrifuged for 1 min at 3,100 ×g and stored at −20°C until analysis
697

698 For lipidomics analysis, deuterated internal standard (Avanti) representing each of the 15 lipid
699 classes analyzed were spiked into the resuspended samples. These allow for semi-quantification
700 of species within these classes. Other potentially present lipid species, such as ceramide
701 derivatives, glycosylated lipids, and non-sterol terpenes, were not quantified using this method.
702 Polar lipids were measured according to a previously described method[98], with slight
703 modifications. The Vanquish UHPLC (Thermo Fisher Scientific) was interfaced with a QExactive
704 mass spectrometer (Thermo Fisher Scientific). A Waters T3 1.6 µM 2.1 mm x 150 mm column
705 was used for chromatographic separation using a step gradient from 25% buffer A (10 mM
706 ammonium formate and 0.1% formic acid in water) to 100% buffer B (70/30
707 isopropanol/acetonitrile with 10 mM ammonium formate and 0.1% formic acid) over 35 min. Flow
708 rate was set at 0.3 mL/min. This LC gradient minimizes ion suppression. Lipid samples were
709 analyzed using data-dependent acquisition (DDA) with an NCE of 30 in negative mode and an
710 NCE of 25 in positive mode. Ion source parameters were: Sheath Gas 48 AU; Aux Gas 11 AU;
711 Sweet Gas 1 AU; Spray Voltage 3.5 kV for positive mode and 2.5 kV for negative mode; Capillary

712 Temp. 250°C; S-lens RF level 60 AU; Aux Gas Heater Temp. 413°C. All ions in the mass range
713 of 200-1200 m/z were monitored. MS1 resolution was set at 70,000 (FWHM at m/z 200) with an
714 automatic gain control (AGC) target of 1e6 and Maximum IT of 200 ms. MS2 resolution was set
715 to 17,500 with an AGC target of 5e4, fixed first mass of 80 m/z, and Maximum IT of 50 ms. The
716 isolation width was set at 1.2; Dynamic Exclusion was set at 3 s. Lipids were identified and
717 quantified with Lipid Data Analyzer Software.
718
719 **Interpro Sequence alignments and phylogeny of START proteins in Asgards**
720 The initial dataset consists of the 56 identified sequences corresponding to the InterPro accession
721 number IPR023393 in annotated Asgard strains including Lokiarchaeia (MK-D1, CR_4, b31, b32
722 and GC14_75), Thorarchaeia (SMTZ1_45, SMTZ1_83 and AB25), Heimdallarchaeia (including
723 LC_3 and B3_JM_08 both reclassified as Hodarchaeales, LC_2 as Kariarchaeaceae, and
724 AB_125 as Heimallarchaeiaceae[25,26]) and Helarchaeales. One sequence (A0A135VUP7) from
725 Thor was removed because it was identical to A0A135VGP2. Full length protein sequences were
726 aligned using MAFFT L-INS-i with default parameters and Blosum62 matrix[99]. The alignment
727 results was used to compute a ML phylogeny using PhyML server (PhyML 3.3.2), with Smart
728 Model Selection based on Bayesian information Criteria (SMS BIC), using a neighbor joining
729 starting tree (BioNJ)[100]. The inference was then reiterated in IQTree[101], using the same
730 substitution model, frequency parameters, and seed, and branch support was assessed using
731 SH-aLRT (1000 replicates and ultrafast bootstrap (UFBoot2, 1000 replicates). The tree was then
732 displayed and annotated using the iTOL online tool[102] and a singular sequence from
733 Helarchaeales was selected as a root.
734
735 **NCBI sequences alignment and phylogeny of START proteins in Asgards**
736 Asgard proteome sequences were retrieved from the NCBI database (GenBank Release 264),
737 comprising 711,493 Identical Protein Groups (IPGs). Protein domain analysis was performed
738 using InterProScan version 5.72-103.0[103] deployed via Docker Desktop 4.37.2 (179585) on
739 macOS. Analysis was restricted to Gene3D, Pfam, and SUPERFAMILY search programs. A
740 custom bash script divided protein sequences into smaller batches and ran analyses in parallel
741 (two concurrent processes, two CPU cores each). The pre-calculated match lookup service was
742 utilized with default configuration. Output files were organized by batch and the 508,901 newly
743 annotated sequences were concatenated for downstream analysis. After assigning taxonomic
744 ranks to each sequence using NCBI annotation, 411 START domains were identified using
745 InterPro annotations. (SSF55961; G3DSA:3.30.530.20; PF10604; PF11485; PF08982; PF06240;

746 PF19569; PF02121; PF08327; PF03364; PF01852; PF00407; IPR023393). Each sequence was
747 trimmed according to the longest non-overlapping domain coordinate given by the InterPro
748 Analysis. A subset of START domain-containing proteins in the Thorarchaeia clade contained two
749 domains; these were retained as distinct protein sequences. The trimmed sequences were
750 aligned using Muscle5[104]. ML phylogenies were inferred with IQ-Tree following model selection
751 using ModelFinder. Branch support was assessed using SH-aLRT (1000 replicates and ultrafast
752 bootstrap (UFBoot2, 1000 replicates). Five monophyletic domain classes were identified.. To
753 represent domain abundance, the average number of species in each Asgard archaeal taxon
754 was estimated using seven universal single-copy proteins (SecY, uS13, uS3, uS4, aIF2, Arginine
755 and Serine RNA ligase) among the 508,901 annotated sequences. The mean count of these
756 marker proteins per taxon was used as a proxy for species abundance. START domain
757 abundance in each taxon was normalized by the estimated species abundance.
758
759 **Large sequence dataset construction**
760 A set of 83 species of archaea including 14 Asgards, 4 Stygia, 4 Acherontia, 18 Stenoarchaea, 5
761 Methamonada, 4 Diaforarchaea, 31 TACK group members and 2 DPANN group members, was
762 collected based on recent phylogeny[105]. Similarly, 32 eukaryotes were selected including
763 Obazoa, Amoebozoa, Excavata, Cryptista, Archaeplastida and TSAR reflecting previous
764 molecular-clock analysis[106]. 46 Bacterial species were also selected a posteriori, to increase the
765 depth of the tree by including members of Acidobacteriota, Bdellovibrionota, Campylobacterota,
766 Chrysiogenota, Coprothermobacterota, Deferribacterota, FCB group, Fusobacteriota,
767 Myxococcota, Pseudomonadota (Alpha, Beta and Gammaproteobacteria), PVC group,
768 Terrabacteria group, (Actinomycetota, Deinococcota), Thermodesulfobacteriota and
769 Spirochaetota. For each species, START-containing protein sequences are collected using
770 IPR023393 accession as a request to Uniprot and then used CD-HIT (-c 0.9 -n 5) to identify
771 putative in-paralog groups[107]. For each of these groups, spurious redundant sequences were
772 manually removed depending on their annotation in genomic databases such as genbank,
773 Gramene, Flybase, EmsemblMetazoa and RefSeq. When locus annotations were lacking
774 between every in-paralog, the longest transcript was selected. From this new protein dataset, as
775 a list of accession file, ProtDomRetriever.py was executed using as an input the InterPro entries
776 IPR023393, CATH G3DSA:3.30.530.20, and SUPERFAMILY SSF55961.

Reduced sequence-based Maximum Likelihood phylogeny of START domains

A curated pool of 842 START domains derived from 774 protein sequences was assembled using ProtDomRetriever.py. From this pool, specific sequence subsets were defined to generate the phylogenetic trees presented in the main and supplementary figures, enabling comparisons across START clades or within major domain groups. The sequences were aligned using MAFFT (v7)[108]**.** To empirically determine optimal alignment parameters, we developed MAFFT ScoreNGo.py, a wrapper that systematically evaluates alternative MAFFT configurations. The following alignment strategies were tested: G-INS-i, E-INS-i, and L-INS-i, using both BLOSUM62 and BLOSUM80 scoring matrices, together with a range of gap-opening, gap-extension, and local alignment penalty parameters. In total, 337 parameter combinations were evaluated using the standard search mode. Each resulting alignment was scored using a conservation-based metric implementing Jalview's per-column conservation scheme[109], derived from the AMAS method[110]. For each alignment column, conservation was evaluated based on agreement across 10 physicochemical property classes, with columns exceeding a defined gap threshold excluded from scoring. Per-column conservation scores (0–11) were summed to obtain a global conservation score for each alignment. For each dataset, the alignment yielding the highest global conservation score was retained for downstream phylogenetic inference. This empirical optimization strategy was designed to maximize residue-property consistency across the conserved START domain core, improving alignment coherence and preserving phylogenetically informative sites. Alternative alignments generated with MUSCLE5 were also evaluated using the conservation-based scoring framework described above. In direct comparisons on selected START domain datasets, MUSCLE5 produced lower global conservation scores and more gap-dense alignments relative to MAFFT under optimized parameters, reducing coherence across conserved core residues. MAFFT was therefore used for phylogenetic inference in the tree-of-life START domain analyses. ML phylogenies were inferred using IQ-TREE with extended model selection implemented via ModelFinder. ModelFinder was used to evaluate a comprehensive set of substitution models, including mixture models (C10, C20, C40, C60), LG4M and LG4X models, alternative equilibrium frequency schemes (F, FO, FQ), nuclear substitution matrices, and multiple rate heterogeneity models (E, I, G, I+G, R, I+R; 4–20 categories). The best-fitting model was selected according to the Bayesian Information Criterion (BIC) for each dataset. Branch support was assessed using SH-aLRT (1,000 replicates) and ultrafast bootstrap (UFBoot2, 1,000 replicates). The trees were displayed and annotated with iTol[102].

Large structural phylogeny of START domains

AlphaFold structures were retrieved from Alphafold Protein Structure database. A trimming algorithm was applied to the PDB structures using the position file generated by ProtDomRetriever, to generate new .pdb files excluding any atoms that are not part of the START domain. When multiple START domains are present in a structure, each of these domains are extracted and copied in an independent new file. This set of trimmed structures was submitted to the FoldTree (build d0f0239) pipeline[60], using Google Colab Notebook on a Python3 runtime accelerated by a A100 GPU. The rooted tree was then displayed and annotated with iTol without modifying the topology. From this tree, a smaller collapsed cladogram was constructed by deleting all sequences not belonging to a known clade of functional protein. Taxonomic congruence was evaluated by comparing the START domain phylogeny with a synthetic species tree extracted from the Open Tree of Life (OTOL) database[111] using the induced_subtree API. After pruning to identical taxon sets, topological agreement among major eukaryotic clades and lineage-specific paralogous expansions was assessed using a tanglegram representation.

Structural comparison analysis

Structural comparisons were performed using Foldseek (exhaustive search mode) to compute all-versus-all TM-scores across 805 trimmed AlphaFold2 structural models. A pairwise distance matrix (1 − TM-score) was constructed and reduced to two dimensions by t-SNE following PCA pre-processing (50 components), using scikit-learn. Ligand-binding cavity volumes were derived from cavity volumes previously computed using CavitOmiX. Figures were generated using Plotly.

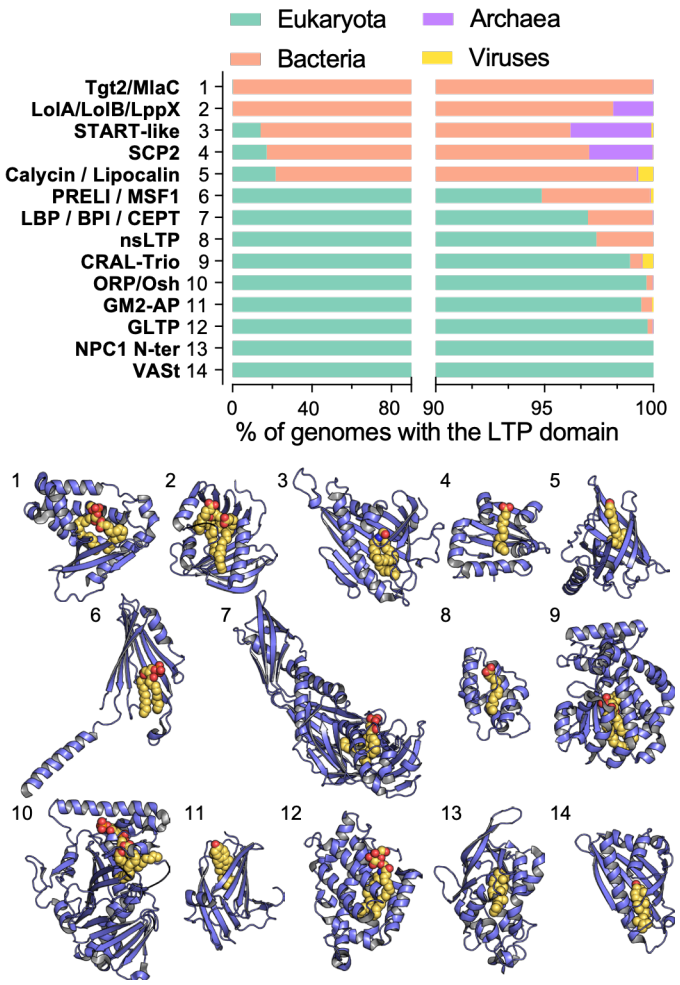


Fig. 1: Distribution and structural diversity of shuttle-like LTPs. *Top panel:* The proportion of total sequenced genomes that express shuttle-like LTP classes belonging to each domain of life. For each class, the InterPro accession number is given in the method section. Most known LTP classes are exclusive to eukaryotes; archaea contain examples of three (2-4). The data is split into two segments to highlight distributions under 10%. *Bottom panel:* Structures of example proteins from each shuttle-like LTP class show in (B) containing a bound lipid ligand (alkanes, yellow; oxygen, red; nitrogen, blue; phosphate, orange). 1. MlaC (lipid: phosphatidylethanolamine, pdb: 7VR6). 2. Lppx (α-linolenic acid and docosahexaenoic acid, 2BY0), 3. START domain of CERT (ceramide, 2E3P), 4. SCP-2 (palmitate, 4JGX), 5. RBPA (retinol, 1RBP), 6. Ups1 (phosphatidic acid, 4YTX), 7. BPI (phosphatidylcholine (PC), 1BP1), 8. nsLTP (palmitoleate, 1FK3), 9. Sec14 (PC, 3B7Z), 10. Osh4 (phosphatidylinositol 4-phosphate, 3SPW), 11. NPC2 (ergosterol, 6R4N), 12. GLTP domain of 4-phosphate adaptor protein 2 (FAPP2) (*N*-oleoyl-galactosylceramide, 5KDI), 13. N-terminal domain of NPC1 (cholesterol, 3GKI), 14. Lam2 (ergosterol, 6CAY).

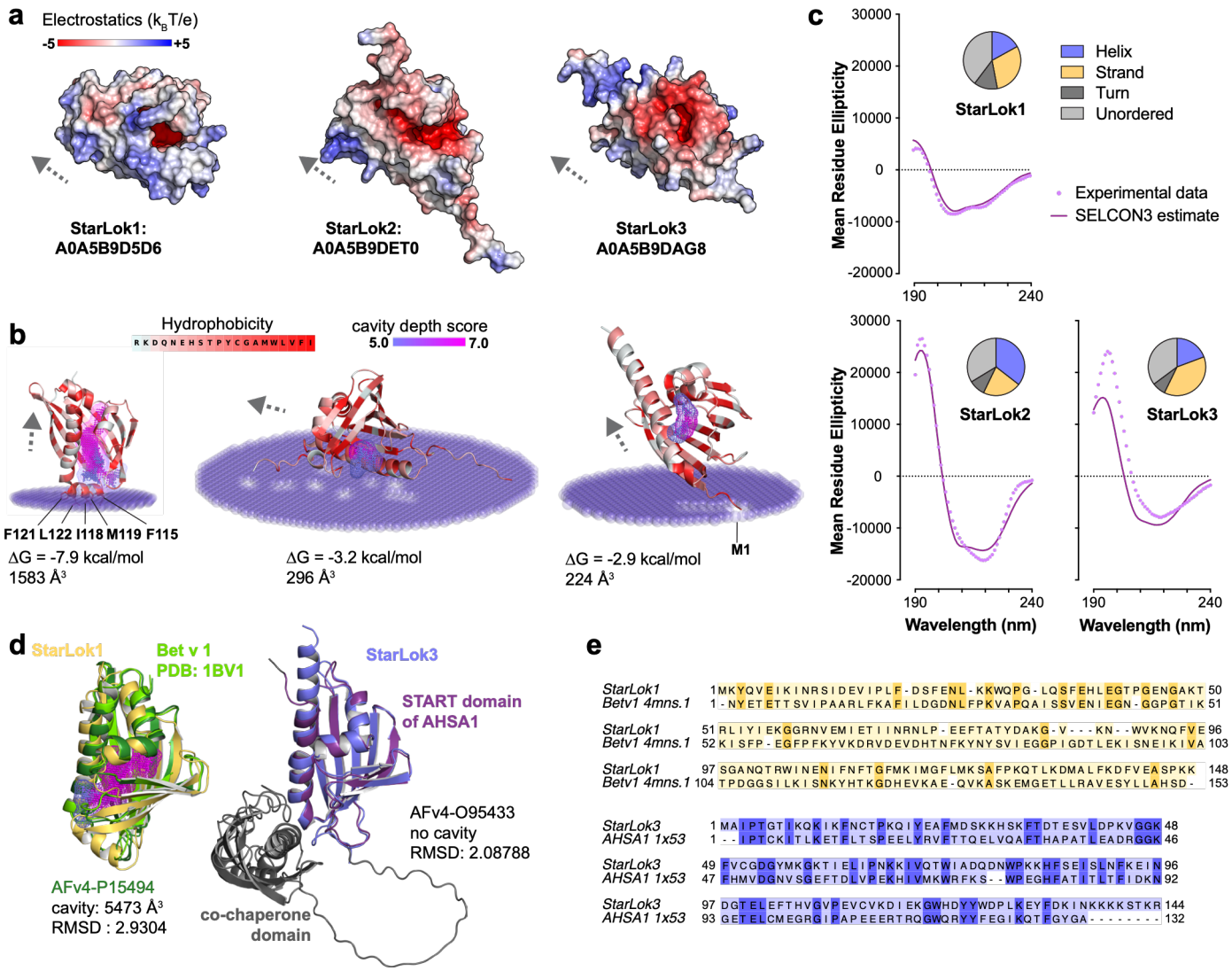


Fig. 2: START domain proteins from *Promethearcheum syntrophicum* have divergent structural features. a. Structural prediction of StarLok1, StarLok2 and StarLok3, represented as electrostatic surface potential ranging from -5 to +5 k_BT/e and shown as a color gradient from red (negative) to white (neutral) to blue (positive). The three proteins were aligned and oriented to show the cavity entrance of StarLok1. Arrows indicate the orientation of the main α-helix of each START domain, from N- to C-terminus. **b.** PPM prediction of StarLok1, 2 and 3 (from left to right) docked to a model archaeal lipid membrane. The structure is colored according to the hydrophobicity of each residue from pale cyan (low hydrophobicity) to red (high hydrophobicity). Computed cavities are visualized as a 3D point cloud, color-coded by LigSite depth scores, from periwinkle blue (lower scores) to magenta (higher scores). Protein-membrane interaction energies, computed cavity sizes are provided for each protein. Key hydrophobic residues of StarLok1 embedded in the lipid monolayer are indicated with connecting lines: phenylalanine (F),

857 leucine (L), isoleucine (I) and methionine (M). **c.** Circular dichroism spectra of StarLok1, 2 and 3.
858 Experimental mean residue ellipticity (MRE in $deg.cm^2.dmol^{-1}$) was recorded from 190 to 240 nm
859 (dotted lines) and secondary structures were assigned by SELCON3 decomposition (solid lines).
860 Pie charts summarize the estimated proportion of α-helix, β-strand, turn and unordered regions
861 for each protein. **d.** *Left panel:* Structural similarities between the AF structure of StarLok1
862 (yellow) and that of Bet v 1 (dark green). The crystal structure of Bet v 1 (pdb: 1BV1, light green)
863 was also aligned. The cavity of Bet v 1 is displayed as in panel A using LigSite depth scoreq.
864 *Right panel:* Structural similarities between the AF structure of StarLok3 (blue) and that of the
865 START domain of Aha1 (Purple). The co-chaperone binding domain of Aha1 is indicated in black.
866 RMSD of AFv4-P15494 and AFv4-O95433 aligned to StarLok1 and StarLok3, respectively, are
867 indicated. As well, the cavity of AFv4-P15494 is reported. **e.** SWISS-MODEL hit template
868 alignment of StarLok1 and Bet v 1 (yellow) and StarLok3 and Aha1 (blue) sequences.

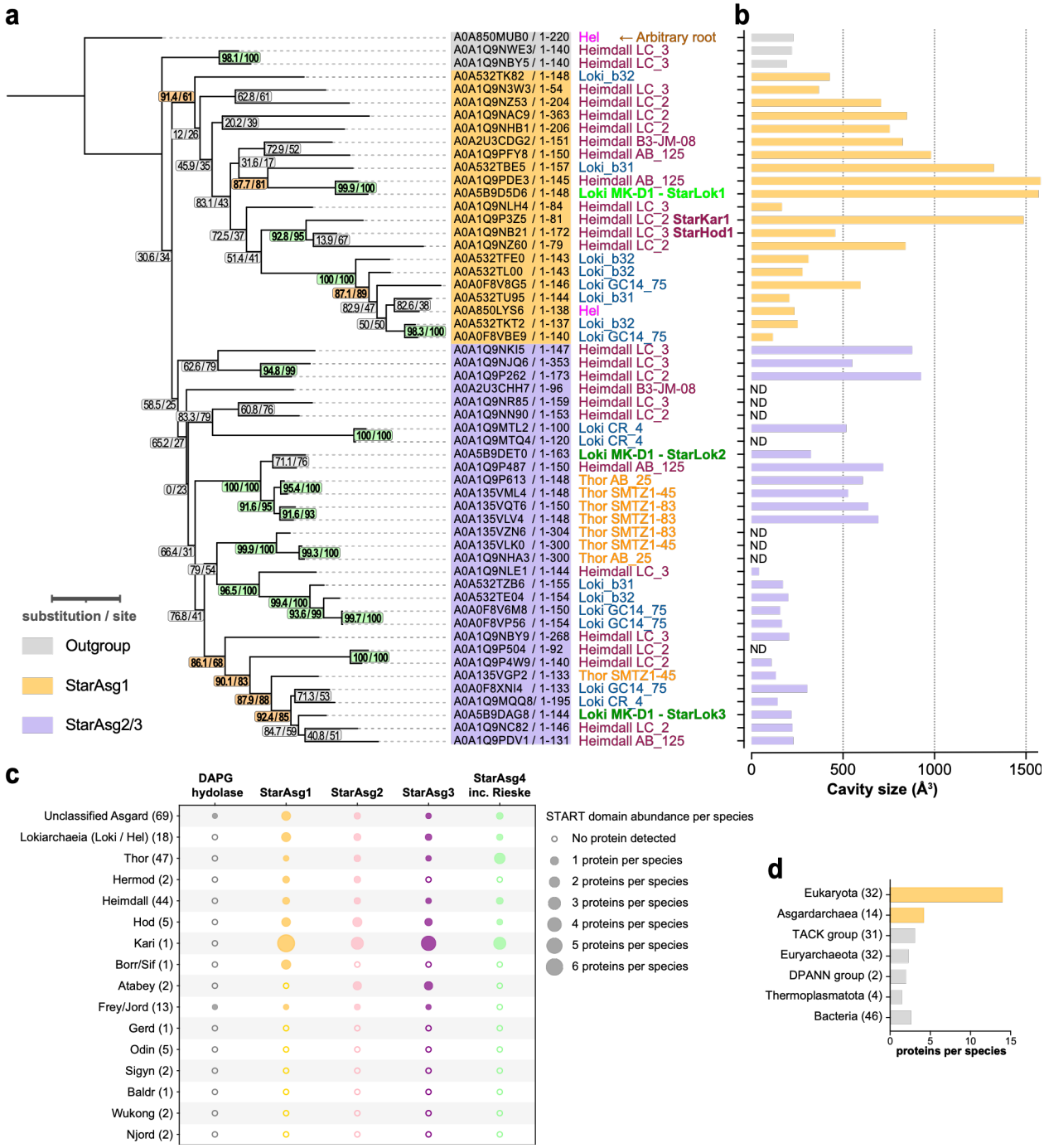


Fig. 3: Distinct clades of START domain proteins across Asgards. a. ML phylogenetic tree (PhyML Q.Pfam + R4 + F) of the START proteins identified in Asgards. Branch support values are indicated as SH-aLRT (%) and ultrafast bootstrap (%). Supports ≥ 90/90 are highlighted in green, while supports ≥ 85/60 are indicated in orange. The tree is arbitrarily rooted with a single sequence identified in the Helarchaeales (strain CR_Bin_291). Species names are color-coded according to their taxonomic groups: Helarchaeales (magenta), Heimdallarchaeia (wine),

875 Thorarchaeia (orange), Lokiarchaeia (blue), except for MK-D1 (green). Sequence accessions and
876 lengths corresponding to the monophyletic clade of StarAsg1 are highlighted in yellow, while the
877 monophyletic clade of StarAsg2 and StarAsg3 is highlighted in purple. **b.** Comparative analysis
878 of cavity sizes in StarAsg1 and its homologs in Asgards. The Uniprot accession numbers are
879 listed along the y-axis, with the corresponding cavity sizes represented by horizontal bars. Yellow
880 bars indicate proteins belonging to the StarAsg1 clade, and purple bars indicate proteins from the
881 StarAsg2 and StarAsg3 clades. ND indicates the absence of a detectable cavity. **c.** Abundance
882 of the START domain classes in different Asgard phyla. The dot size is proportional to the number
883 of identified START domains per species; an empty circle indicates the absence of START
884 domains in that phylum. The average number of determined species in the dataset is given next
885 to each phyla's name in parenthesis. **d.** Abundance of the START domains identified in a
886 representative set of archaeal, bacterial and eukaryotic species referenced in the UniProt
887 database. The number of species in each clade is indicated in parentheses. The set corresponds
888 to that used for the analyses in Fig. 6.

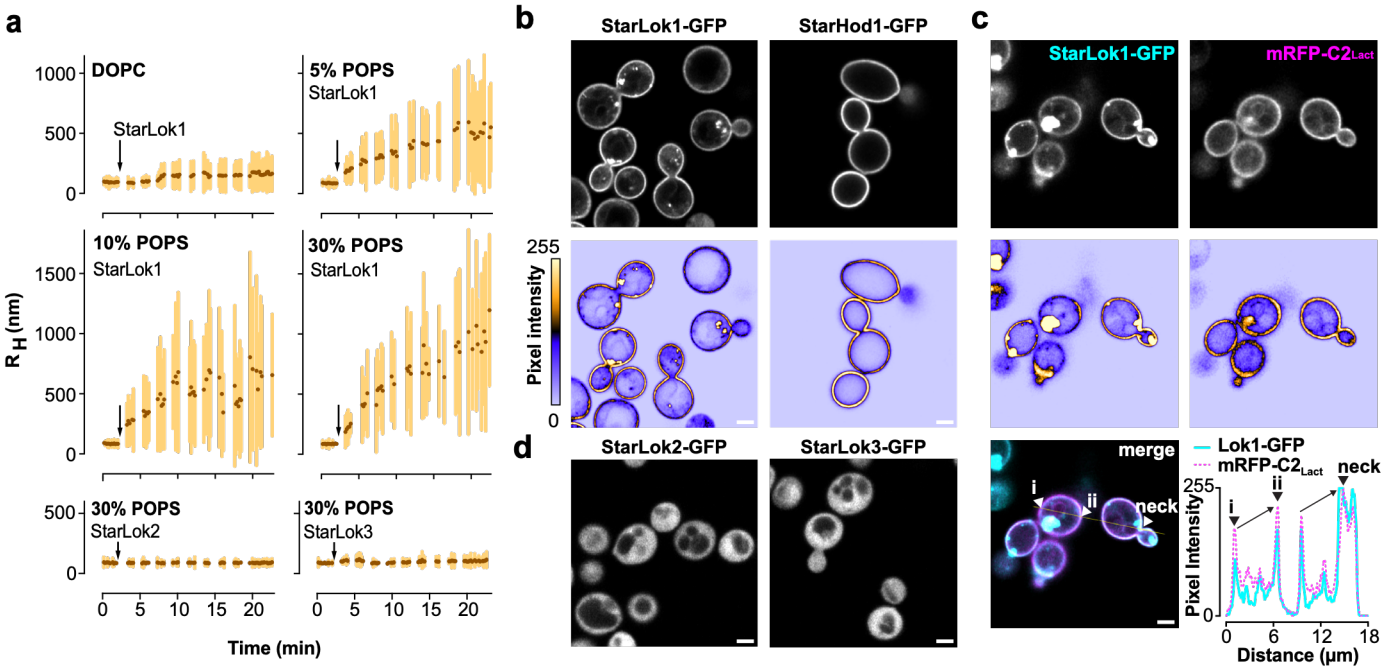


Fig. 4: StarAsg1 proteins are specialized for interaction with anionic membranes. a. DLS of StarAsg1 with liposomes of different lipid compositions. Light scattering was recorded for 2 min after the addition of 50 µM liposomes composed of pure DOPC, or DOPC doped with 5 mol% POPS, 10 mol% POPS, or 30 mol% POPS. At t = 2 min, 1 µM StarAsg1 was added to the mixture, and measurements were acquired for an additional 20 min. Increasing POPS causes protein-mediated aggregation of the liposomes. Addition of StarLok2 and StarLok3 addition to DOPC liposomes with 30 mol% POPC does not induce aggregation. **b.** StarLok1-GFP and StarHod1-GFP expressed in budding yeast binds the PM. The lower panels are intensity heat maps of the corresponding confocal image to highlight the polarized distribution of StarAsg1 in membranes surrounding the yeast bud neck and budding daughter cells. **c.** StarLok1 localization and polarization correlates with PS abundance in the OM. PS is detected with the protein sensor mRFP-C2$_{LACT}$ as previously descibed[47]. Corresponding intensity heatmaps are reported as in b. The lower left composite image of StarLok1-GFP (cyan) and mRFP-C2$_{LACT}$ (magenta) was used to determine colocalization shown on the left curve. An increase of signal from “i” to “ii” across the PM is indicated by an arrow. Scale bar, 2 µm.

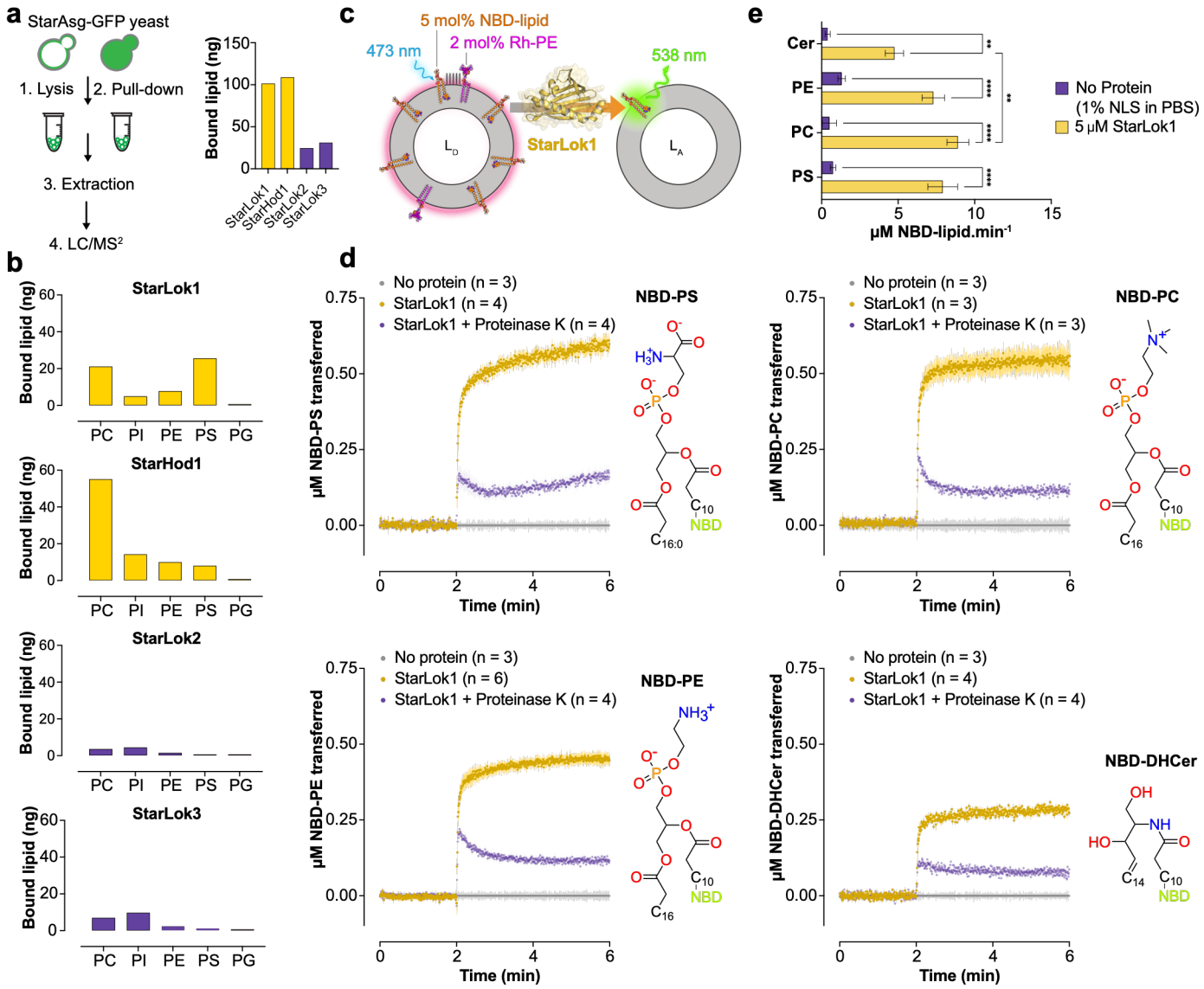


Fig. 5: Characterization of lipids bound to StarAsg proteins and lipid transfer activity. a. Schematic representation of the experimental workflow used to identify lipids bound to GFP-tagged StarAsg proteins expressed in yeast and plot showing total lipid abundance quantified by LC/MS^2. **b.** Abundance of phospholipid classes associated with StarLok1, StarHod1, and StarLok2/3. StarAsgard1 proteins show increased levels of PC, PE, and PS binding. **c**. Schematic of lipid transfer assay. donor liposomes (L_D, 200 μM total lipid) And acceptor liposomes (L_A, 200 μM total lipid) were incubated with 5μM of StarLok1 in PBS buffer (pH 7.4). **d.** NBD-lipid transfer kinetics of StarLok1 (yellow) incubated with L_D and L_A liposomes compared to a blank control containing an equivalent amount of N-lauroylsarcosine (gray), or an equivalent molar amount of proteinase K–inactivated StarLok1 with (purple). The fluorescence signal was normalized using calibration control to yield the absolute amount of NBD-lipid transferred (μM) and is represented

916 as the mean of n independent replicates ± SD. Each plot represents a different NBD-tagged
917 ligand, whose structure is shown to the right. **e.** Calculated lipid transfer rate derived from the
918 initial slopes of the transfer curves (mean ± SEM). Statistical significance was determined using
919 two-way ANOVA followed by Tukey's multiple comparisons test. ***P*< 0.01; ****P*< 0.0001.

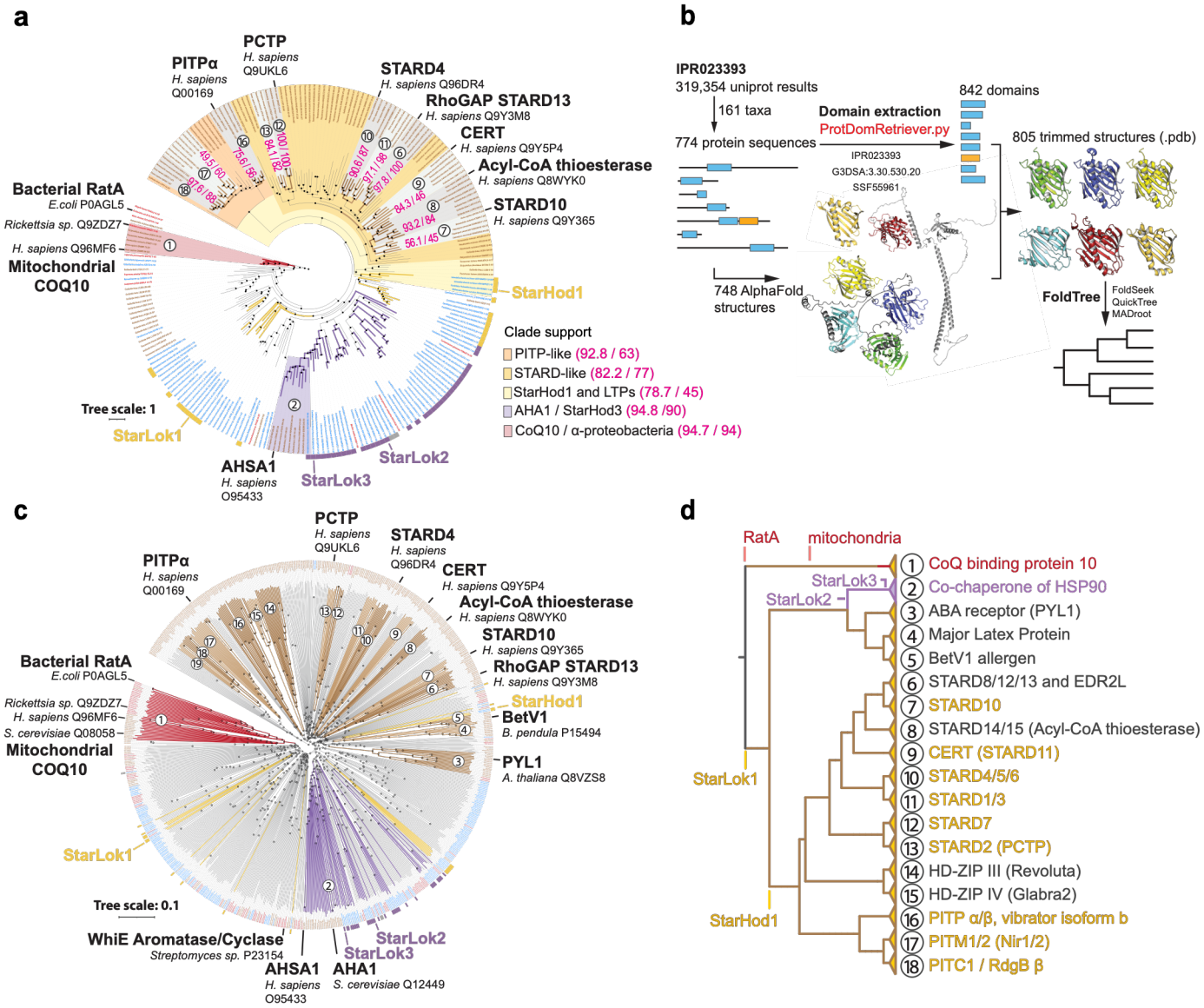
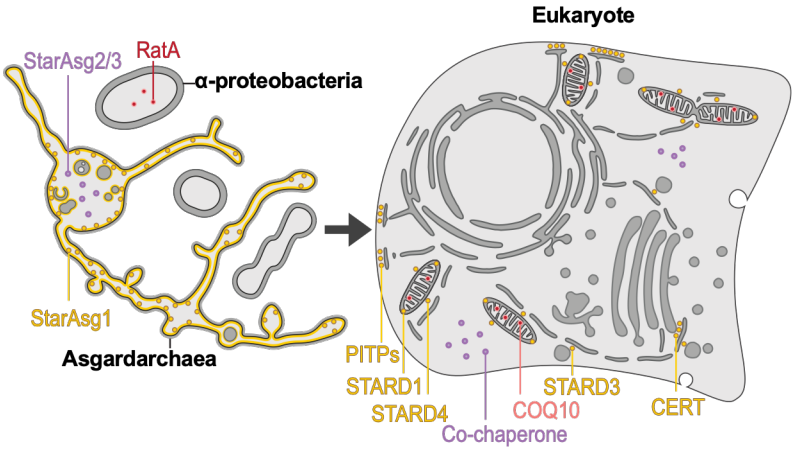


Fig. 6: Sequence and structural domain phylogeny supports orthology between StarAsg proteins and eukaryotic START domains. a. ML phylogenetic tree inferred from 227 START domain sequences, including StarAsg proteins (see **Fig. 3a**), representative eukaryotes (*Homo sapiens, Nematostella vectensis*, *Capsaspora owczarzaki*, *Dictyostelium discoideum*, *Thecamonas trahens*, *Guillardia theta*), 11 archaeal species, and 25 bacterial species. Sequences were aligned using MAFFT (L-INS-i; --localpair --maxiterate 1000) with BLOSUM80 and modified gap parameters (--op 1.53 --ep 0.5 --bl 80 --lop -1.0 --lexp -0.5). The best-fit substitution model was selected using ModelFinder according to BIC (Q.PFAM+FO+R5), and ML inference was performed with IQ-TREE. Branch support values correspond to SH-aLRT (%) and ultrafast bootstrap (%) from 1,000 replicates each and are shown in magenta for major clades only. Branch lengths represent substitutions per site. The tree is rooted using *Shigella flexneri*

931 and *Escherichia coli*. Branch-tip labels are colored according to taxonomic origin (brown,
932 eukaryotes; blue, archaea; red, bacteria). Yellow, purple, and gray stripes adjacent to branch-tip
933 labels indicate Asgard sequences belonging to the StarAsg1, StarAsg2/3, or undefined clades,
934 respectively (see **Fig. 3a**). Red branches highlight the monophyly of mitochondrial CoQ-binding
935 protein 10 homologs with bacterial RatA. Purple, yellow, and brown branches denote StarAsg2/3
936 and AHA1, StarAsg1, and eukaryotic START LTP clades, respectively. **b.** Workflow for large-
937 scale structural domain phylogeny. ProtDomRetriever was used to identify START domain
938 boundaries using InterPro superfamily accessions, CATH/gene3D annotations, and
939 SUPERFAMILY HMM library accessions. Corresponding structural domains were extracted from
940 AlphaFold models and phylogenetic inference was performed using the FoldTree pipeline. **c.**
941 Rooted structural phylogenetic tree of 805 START domains from 161 species. Visual annotations
942 (branch colors, label colors, and stripe coding) follow the scheme described in panel **a**. Each
943 monophyletic START domain class is numbered. Nodes are represented as black dots. The scale
944 bar indicates substitutions per site. The bottom right graph indicates abundance of the START
945 domains identified in representative archaeal, bacterial and eukaryotic species referenced in the
946 dataset database. The number of species in each clade is indicated in parentheses. **d.** Simplified
947 cladogram derived from panel. LTPs previously characterized *in vitro* are highlighted in yellow,
948 signifying the major eukaryotic START domain protein families. Monophyletic classes are labeled
949 using the same numbering as in panel c.



Schematic: Proposed relationships between START proteins in the prokaryotic ancestors of eukaryotic cells (left) and their localization in eukaryotes. StarAsg1 is a putative membrane-bound lipid binding protein in Asgards, which could share a common ancestry with several START domain proteins acting as LTPs. These are predominantly localized to membrane contact sites between organelles. StarAsg2 and StarAsg3, proteins of unknown functions in Asgards, share an origin with the conserved HSP90 co-chaperone in the cytosol of eukaryotes. RatA, a START domain protein in alpha-proteobacteria, shares ancestry with the mitochondrial protein COQ10.

References

1. Kaplan, M. R. & Simoni, R. D. Transport of cholesterol from the endoplasmic reticulum to the plasma membrane. *J. Cell Biol.* 101, 446–453 (1985).
2. Iglesias-Artola, J. M. *et al.* Quantitative imaging of lipid transport in mammalian cells. *Nature* 646, 474–482 (2025).
3. Bigay, J. & Antonny, B. Curvature, lipid packing, and electrostatics of membrane organelles: defining cellular territories in determining specificity. *Dev. Cell* 23, 886–895 (2012).
4. Menon, A. K. Sterol gradients in cells. *Curr. Opin. Cell Biol.* 53, 37–43 (2018).
5. Platre, M. P. *et al.* A combinatorial lipid code shapes the electrostatic landscape of plant endomembranes. *Dev. Cell* 45, 465–480.e11 (2018).
6. Banerjee, T. *et al.* Spatiotemporal dynamics of membrane surface charge regulates cell polarity and migration. *Nat. Cell Biol.* 24, 1499–1515 (2022).
7. Santinho, A., Carpentier, M., Lopes Sampaio, J., Omrane, M. & Thiam, A. R. Giant organelle vesicles to uncover intracellular membrane mechanics and plasticity. *Nat. Commun.* 15, 3767 (2024).
8. Khaddaj, R. & Kukulski, W. Piecing together the structural organisation of lipid exchange at membrane contact sites. *Curr. Opin. Cell Biol.* 83, 102212 (2023).
9. Voeltz, G. K., Sawyer, E. M., Hajnóczky, G. & Prinz, W. A. Making the connection: How membrane contact sites have changed our view of organelle biology. *Cell* 187, 257–270 (2024).
10. Wong, L. H., Gatta, A. T. & Levine, T. P. Lipid transfer proteins: the lipid commute via shuttles, bridges and tubes. *Nat. Rev. Mol. Cell Biol.* 20, 85–101 (2019).
11. Reinisch, K. M. & Prinz, W. A. Mechanisms of nonvesicular lipid transport. *J. Cell Biol.* 220, (2021).
12. Wong, L. H., Čopič, A. & Levine, T. P. Advances on the transfer of lipids by lipid transfer proteins. *Trends Biochem. Sci.* 42, 516–530 (2017).
13. Zhang, Y., Ge, J., Bian, X. & Kumar, A. Quantitative models of lipid transfer and membrane contact formation. *Contact (Thousand Oaks.)* 5, 1–21 (2022).
14. Yeow, J. & Chng, S.-S. Of zones, bridges and chaperones - phospholipid transport in bacterial outer membrane assembly and homeostasis. *Microbiology* 168, (2022).
15. Knoll, A. H., Javaux, E. J., Hewitt, D. & Cohen, P. Eukaryotic organisms in Proterozoic oceans. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 361, 1023–1038 (2006).

16. Eme, L., Spang, A., Lombard, J., Stairs, C. W. & Ettema, T. J. G. Archaea and the origin of eukaryotes. *Nat. Rev. Microbiol.* 15, 711–723 (2017).
17. Betts, H. C. *et al.* Integrated genomic and fossil evidence illuminates life's early evolution and eukaryote origin. *Nat. Ecol. Evol.* 2, 1556–1562 (2018).
18. López-García, P. & Moreira, D. The Syntrophy hypothesis for the origin of eukaryotes revisited. *Nat. Microbiol.* 5, 655–667 (2020).
19. Williams, T. A., Foster, P. G., Cox, C. J. & Embley, T. M. An archaeal origin of eukaryotes supports only two primary domains of life. *Nature* 504, 231–236 (2013).
20. Williams, T. A., Cox, C. J., Foster, P. G., Szöllősi, G. J. & Embley, T. M. Phylogenomics provides robust support for a two-domains tree of life. *Nat. Ecol. Evol.* 4, 138–147 (2020).
21. Muñoz-Gómez, S. A. *et al.* Site-and-branch-heterogeneous analyses of an expanded dataset favour mitochondria as sister to known Alphaproteobacteria. *Nat. Ecol. Evol.* 6, 253–262 (2022).
22. Imachi, H. *et al.* Promethearchaeum syntrophicum gen. nov., sp. nov., an anaerobic, obligately syntrophic archaeon, the first isolate of the lineage 'Asgard' archaea, and proposal of the new archaeal phylum Promethearchaeota phyl. nov. and kingdom Promethearchaeati regn. nov. *Int. J. Syst. Evol. Microbiol.* 74, (2024).
23. Spang, A. *et al.* Complex archaea that bridge the gap between prokaryotes and eukaryotes. *Nature* 521, 173–179 (2015).
24. Zaremba-Niedzwiedzka, K. *et al.* Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature* 541, 353–358 (2017).
25. Eme, L. *et al.* Inference and reconstruction of the heimdallarchaeial ancestry of eukaryotes. *Nature* 618, 992–999 (2023).
26. Zhang, J. *et al.* Deep origin of eukaryotes outside Heimdallarchaeia within Asgardarchaeota. *Nature* 642, 990–998 (2025).
27. Hatano, T. *et al.* Asgard archaea shed light on the evolutionary origins of the eukaryotic ubiquitin-ESCRT machinery. *Nat. Commun.* 13, 3398 (2022).
28. Souza, D. P. *et al.* Asgard archaea reveal the conserved principles of ESCRT-III membrane remodeling. *Sci. Adv.* 11, eads5255 (2025).
29. Tran, L. T., Akıl, C., Senju, Y. & Robinson, R. C. The eukaryotic-like characteristics of small GTPase, roadblock and TRAPPC3 proteins from Asgard archaea. *Commun Biol* 7, 273 (2024).
30. Vargová, R. *et al.* The Asgard archaeal origins of Arf family GTPases involved in eukaryotic organelle dynamics. *Nat. Microbiol.* 10, 495–508 (2025).

1026 31. Zhu, J. *et al.* Asgard Arf GTPases can act as membrane-associating molecular switches
1027 with the potential to function in organelle biogenesis. *Nat. Commun.* 16, 2622 (2025).
1028 32. Imachi, H. *et al.* Isolation of an archaeon at the prokaryote-eukaryote interface. *Nature*
1029 577, 519–525 (2020).
1030 33. Kay, C. J. *et al.* Dated gene duplications elucidate the evolutionary assembly of
1031 eukaryotes. *Nature* 1–12 (2025).
1032 34. MacLeod, F. I. *et al.* An Asgard archaeon with internal membrane compartments. *Cell*
1033 *Biology* (2025).
1034 35. Ghanbarpour, A., Valverde, D. P., Melia, T. J. & Reinisch, K. M. A model for a partnership
1035 of lipid transfer proteins and scramblases in membrane expansion and organelle
1036 biogenesis. *Proc. Natl. Acad. Sci. U. S. A.* 118, e2101562118 (2021).
1037 36. Clark, B. J., Wells, J., King, S. R. & Stocco, D. M. The purification, cloning, and expression
1038 of a novel luteinizing hormone-induced mitochondrial protein in MA-10 mouse Leydig
1039 tumor cells. Characterization of the steroidogenic acute regulatory protein (StAR). *J. Biol.*
1040 *Chem.* 269, 28314–28322 (1994).
1041 37. Klingler, J. P., Batelli, G. & Zhu, J.-K. ABA receptors: the START of a new paradigm in
1042 phytohormone signalling. *J. Exp. Bot.* 61, 3199–3210 (2010).
1043 38. Frey, Y., Lungu, C. & Olayioye, M. A. Regulation and functions of the DLC family of
1044 RhoGAP proteins: Implications for development and cancer. *Cell. Signal.* 125, 111505
1045 (2025).
1046 39. Husbands, A. Y. *et al.* The START domain potentiates HD-ZIPIII transcriptional activity.
1047 *Plant Cell* 35, 2332–2348 (2023).
1048 40. Mondol, T. *et al.* Aha1 regulates Hsp90's conformation and function in a stoichiometry-
1049 dependent way. *Biophys. J.* 122, 3458–3468 (2023).
1050 41. Iyer, L. M., Koonin, E. V. & Aravind, L. Adaptations of the helix-grip fold for ligand binding
1051 and catalysis in the START domain superfamily. *Proteins* 43, 134–144 (2001).
1052 42. Radauer, C., Lackner, P. & Breiteneder, H. The Bet v 1 fold: an ancient, versatile scaffold
1053 for binding of large, hydrophobic ligands. *BMC Evol. Biol.* 8, 286 (2008).
1054 43. Ferguson, K. M., Lemmon, M. A., Sigler, P. B. & Schlessinger, J. Scratching the surface
1055 with the PH domain. *Nat. Struct. Biol.* 2, 715–718 (1995).
1056 44. Murray, D., Arbuzova, A., Honig, B. & McLaughlint, S. The role of electrostatic and
1057 nonpolar interactions in the association of peripheral proteins with membranes. in *Current*
1058 *Topics in Membranes* 277–307 (Elsevier, 2002).
1059 45. Iaea, D. B., Dikiy, I., Kiburu, I., Eliezer, D. & Maxfield, F. R. STARD4 membrane

1060 interactions and sterol binding. *Biochemistry* 54, 4623–4636 (2015).
1061 46. Lipp, N.-F. *et al.* An electrostatic switching mechanism to control the lipid transfer activity
1062 of Osh6p. *Nat. Commun.* 10, 3926 (2019).
1063 47. Caldara-Festin, G. *et al.* Structural and functional analysis of two di-domain
1064 aromatase/cyclases from type II polyketide synthases. *Proc. Natl. Acad. Sci. U. S. A.* 112,
1065 E6844–51 (2015).
1066 48. Talandashti, R. *et al.* Membrane specificity of the human cholesterol transfer protein
1067 STARD4. *J. Mol. Biol.* 436, 168572 (2024).
1068 49. Husslik, F. *et al.* Folded or not? Tracking Bet v 1 conformation in recombinant allergen
1069 preparations. *PLoS One* 10, e0132956 (2015).
1070 50. McBride, J. K., Cheng, H., Maleki, S. J. & Hurlburt, B. K. Purification and characterization
1071 of pathogenesis related class 10 panallergens. *Foods* 8, 609 (2019).
1072 51. Gane, A. *et al.* ProtNLM: Model-based natural language protein annotation.
1073 https://storage.googleapis.com/brain-genomics-
1074 public/research/proteins/protnlm/uniprot_2022_04/protnlm_preprint_draft.pdf.
1075 52. Orsi, W. D. *et al.* Metabolic activity analyses demonstrate that Lokiarchaeon exhibits
1076 homoacetogenesis in sulfidic marine sediments. *Nat. Microbiol.* 5, 248–255 (2020).
1077 53. Ragsdale, S. W. & Pierce, E. Acetogenesis and the Wood-Ljungdahl pathway of CO(2)
1078 fixation. *Biochim. Biophys. Acta* 1784, 1873–1898 (2008).
1079 54. Fairn, G. D., Hermansson, M., Somerharju, P. & Grinstein, S. Phosphatidylserine is
1080 polarized and required for proper Cdc42 localization and for development of cell polarity.
1081 *Nat. Cell Biol.* 13, 1424–1430 (2011).
1082 55. Chung, J. *et al.* INTRACELLULAR TRANSPORT. PI4P/phosphatidylserine
1083 countertransport at ORP5- and ORP8-mediated ER-plasma membrane contacts. *Science*
1084 349, 428–432 (2015).
1085 56. Kim, H. & Budin, I. Intracellular sphingolipid sorting drives membrane phase separation in
1086 the yeast vacuole. *J. Biol. Chem.* 300, 105496 (2024).
1087 57. Olayioye, M. A. *et al.* StarD10, a START domain protein overexpressed in breast cancer,
1088 functions as a phospholipid transfer protein. *J. Biol. Chem.* 280, 27436–27442 (2005).
1089 58. Panagabko, C., Baptist, M. & Atkinson, J. In vitro lipid transfer assays of
1090 phosphatidylinositol transfer proteins provide insight into the in vivo mechanism of ligand
1091 transfer. *Biochim. Biophys. Acta Biomembr.* 1861, 619–630 (2019).
1092 59. Tsai, Y.-T. *et al.* 1-deoxysphingolipids dysregulate membrane properties and cargo
1093 trafficking in the early secretory pathway. *Cell Chem. Biol.* 33, 45–58.e8 (2026).

1094 60. Moi, D. *et al.* Structural phylogenetics unravels the evolutionary diversification of
1095 communication systems in gram-positive bacteria and their viruses. *Nat. Struct. Mol. Biol.*
1096 32, 2492–2502 (2025).
1097 61. van Kempen, M. *et al.* Fast and accurate protein structure search with Foldseek. *Nat.*
1098 *Biotechnol.* 42, 243–246 (2024).
1099 62. Tsui, H. S. *et al.* Human COQ10A and COQ10B are distinct lipid-binding START domain
1100 proteins required for coenzyme Q function. *J. Lipid Res.* 60, 1293–1310 (2019).
1101 63. Fino, C. *et al.* PasT of Escherichia coli sustains antibiotic tolerance and aerobic respiration
1102 as a bacterial homolog of mitochondrial Coq10. *Microbiologyopen* 9, e1064 (2020).
1103 64. Alpy, F. & Tomasetto, C. Give lipids a START: the StAR-related lipid transfer (START)
1104 domain in mammals. *J. Cell Sci.* 118, 2791–2801 (2005).
1105 65. Wyckoff, G. J., Solidar, A. & Yoden, M. D. Phosphatidylinositol transfer proteins: sequence
1106 motifs in structural and evolutionary analyses. *J. Biomed. Sci. Eng.* 3, 65–77 (2010).
1107 66. Shadan, S. *et al.* Dynamics of lipid transfer by phosphatidylinositol transfer proteins in
1108 cells. *Traffic* 9, 1743–1756 (2008).
1109 67. Esposti, M. D. Eukaryotes inherited inositol lipids from bacteria: implications for the
1110 models of eukaryogenesis. *FEBS Lett.* 597, 2484–2496 (2023).
1111 68. Centola, M., van Pee, K., Betz, H. & Yildiz, Ö. Crystal structures of phosphatidyl serine
1112 synthase PSS reveal the catalytic mechanism of CDP-DAG alcohol O-phosphatidyl
1113 transferases. *Nat. Commun.* 12, 6982 (2021).
1114 69. Morii, H. & Koga, Y. Asymmetrical topology of diether- and tetraether-type polar lipids in
1115 membranes of Methanobacterium thermoautotrophicum cells. *J. Biol. Chem.* 269, 10492–
1116 10497 (1994).
1117 70. Yeung, T. *et al.* Contribution of phosphatidylserine to membrane surface charge and
1118 protein targeting during phagosome maturation. *J. Gen. Physiol.* 134, i1–i1 (2009).
1119 71. Uchida, Y. *et al.* Intracellular phosphatidylserine is essential for retrograde membrane
1120 traffic through endosomes. *Proc. Natl. Acad. Sci. U. S. A.* 108, 15846–15851 (2011).
1121 72. Xu, P., Baldridge, R. D., Chi, R. J., Burd, C. G. & Graham, T. R. Phosphatidylserine
1122 flipping enhances membrane curvature and negative charge required for vesicular
1123 transport. *J. Cell Biol.* 202, 875–886 (2013).
1124 73. Hirama, T. *et al.* Membrane curvature induced by proximity of anionic phospholipids can
1125 initiate endocytosis. *Nat. Commun.* 8, 1393 (2017).
1126 74. Antonny, B., Bigay, J. & Mesmin, B. The oxysterol-binding protein cycle: Burning off
1127 PI(4)P to transport cholesterol. *Annu. Rev. Biochem.* 87, 809–837 (2018).

1128 75. Guillén-Samander, A. *et al.* VPS13D bridges the ER to mitochondria and peroxisomes via
1129 Miro. *J. Cell Biol.* 220, (2021).
1130 76. Valverde, D. P. *et al.* ATG2 transports lipids to promote autophagosome biogenesis. *J.*
1131 *Cell Biol.* 218, 1787–1798 (2019).
1132 77. Park, J.-S. & Neiman, A. M. VPS13 regulates membrane morphogenesis during
1133 sporulation in Saccharomyces cerevisiae. *J. Cell Sci.* 125, 3004–3011 (2012).
1134 78. Kumar, N. *et al.* VPS13A and VPS13C are lipid transport proteins differentially localized at
1135 ER contact sites. *J. Cell Biol.* 217, 3625–3639 (2018).
1136 79. Park, J.-S. *et al.* Yeast Vps13 promotes mitochondrial function and is localized at
1137 membrane contact sites. *Mol. Biol. Cell* 27, 2435–2449 (2016).
1138 80. Horenkamp, F. A., Valverde, D. P., Nunnari, J. & Reinisch, K. M. Molecular basis for sterol
1139 transport by StART-like lipid transfer domains. *EMBO J.* 37, (2018).
1140 81. Khafif, M., Cottret, L., Balagué, C. & Raffaele, S. Identification and phylogenetic analyses
1141 of VASt, an uncharacterized protein domain associated with lipid-binding domains in
1142 Eukaryotes. *BMC Bioinformatics* 15, 222 (2014).
1143 82. Chiapparino, A., Maeda, K., Turei, D., Saez-Rodriguez, J. & Gavin, A.-C. The orchestra of
1144 lipid-transfer proteins at the crossroads between metabolism and signaling. *Prog. Lipid*
1145 *Res.* 61, 30–39 (2016).
1146 83. Missaoui, K. *et al.* Plant non-specific lipid transfer proteins: An overview. *Plant Physiol.*
1147 *Biochem.* 171, 115–127 (2022).
1148 84. Miliara, X. *et al.* An intermolecular hydrogen bonded network in the PRELID-TRIAP
1149 protein family plays a role in lipid sensing. *Biochim. Biophys. Acta: Proteins Proteomics*
1150 1871, 140867 (2023).
1151 85. Miles, A. J., Ramalli, S. G. & Wallace, B. A. DichroWeb, a website for calculating protein
1152 secondary structure from circular dichroism spectroscopic data. *Protein Sci.* 31, 37–46
1153 (2022).
1154 86. Micsonai, A. *et al.* BeStSel: webserver for secondary structure and fold prediction for
1155 protein CD spectroscopy. *Nucleic Acids Res.* 50, W90–W98 (2022).
1156 87. Waterhouse, A. *et al.* SWISS-MODEL: homology modelling of protein structures and
1157 complexes. *Nucleic Acids Res.* 46, W296–W303 (2018).
1158 88. Gruber, K., Steinkellner, G. & Gruber, C. Determining novel enzymatic functionalities
1159 using three-dimensional point clouds representing physico chemical properties of protein
1160 cavities. *US Patent* (2020).
1161 89. Hendlich, M., Rippmann, F. & Barnickel, G. LIGSITE: automatic and efficient detection of

1162 potential small molecule-binding sites in proteins. *J. Mol. Graph. Model.* 15, 359–63, 389
1163 (1997).
1164 90. Dolinsky, T. J., Nielsen, J. E., McCammon, J. A. & Baker, N. A. PDB2PQR: an automated
1165 pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res.*
1166 32, W665–7 (2004).
1167 91. Eisenberg, D., Schwarz, E., Komaromy, M. & Wall, R. Analysis of membrane and surface
1168 protein sequences with the hydrophobic moment plot. *J. Mol. Biol.* 179, 125–142 (1984).
1169 92. Lomize, A. L., Todd, S. C. & Pogozheva, I. D. Spatial arrangement of proteins in planar
1170 and curved membranes by PPM 3.0. *Protein Sci.* 31, 209–220 (2022).
1171 93. Frishman, D. & Argos, P. Knowledge-based protein secondary structure assignment.
1172 *Proteins* 23, 566–579 (1995).
1173 94. Kabsch, W. & Sander, C. Dictionary of protein secondary structure: pattern recognition of
1174 hydrogen-bonded and geometrical features. *Biopolymers* 22, 2577–2637 (1983).
1175 95. Nagy, G., Hoffmann, S. V., Jones, N. C. & Grubmüller, H. Reference data set for circular
1176 dichroism spectroscopy comprised of validated intrinsically disordered protein models.
1177 *Appl. Spectrosc.* 78, 897–911 (2024).
1178 96. Mavridis, L. & Janes, R. W. PDB2CD: a web-based application for the generation of
1179 circular dichroism spectra from protein atomic coordinates. *Bioinformatics* 33, 56–63
1180 (2017).
1181 97. Löfgren, L., Forsberg, G.-B. & Ståhlman, M. The BUME method: a new rapid and simple
1182 chloroform-free method for total lipid extraction of animal tissue. *Sci. Rep.* 6, (2016).
1183 98. Quehenberger, O. *et al.* Lipidomics reveals a remarkable diversity of lipids in human
1184 plasma. *J. Lipid Res.* 51, 3299–3305 (2010).
1185 99. Katoh, K., Rozewicki, J. & Yamada, K. D. MAFFT online service: multiple sequence
1186 alignment, interactive sequence choice and visualization. *Brief. Bioinform.* 20, 1160–1166
1187 (2017).
1188 100. Guindon, S. *et al.* New algorithms and methods to estimate maximum-likelihood
1189 phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* 59, 307–321 (2010).
1190 101. Minh, B. Q. *et al.* IQ-TREE 2: New models and efficient methods for phylogenetic
1191 inference in the genomic era. *Mol. Biol. Evol.* 37, 1530–1534 (2020).
1192 102. Letunic, I. & Bork, P. Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic
1193 tree display and annotation tool. *Nucleic Acids Res.* 52, W78–W82 (2024).
1194 103. Jones, P. *et al.* InterProScan 5: genome-scale protein function classification.
1195 *Bioinformatics* 30, 1236–1240 (2014).

1196 104. Edgar, R. C. Muscle5: High-accuracy alignment ensembles enable unbiased assessments
1197 of sequence homology and phylogeny. *Nat. Commun.* 13, 6968 (2022).
1198 105. Aouad, M. *et al.* A divide-and-conquer phylogenomic approach based on character
1199 supermatrices resolves early steps in the evolution of the Archaea. *BMC Ecol. Evol.* 22, 1
1200 (2022).
1201 106. Strassert, J. F. H., Irisarri, I., Williams, T. A. & Burki, F. A molecular timescale for
1202 eukaryote evolution with implications for the origin of red algal-derived plastids. *Nat.*
1203 *Commun.* 12, 1879 (2021).
1204 107. Li, W. & Godzik, A. Cd-hit: a fast program for clustering and comparing large sets of
1205 protein or nucleotide sequences. *Bioinformatics* 22, 1658–1659 (2006).
1206 108. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7:
1207 improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780 (2013).
1208 109. Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. & Barton, G. J. Jalview
1209 Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics*
1210 25, 1189–1191 (2009).
1211 110. Livingstone, C. D. & Barton, G. J. Protein sequence alignments: a strategy for the
1212 hierarchical analysis of residue conservation. *Comput. Appl. Biosci.* 9, 745–756 (1993).
1213 111. Rees, J. A. & Cranston, K. Automated assembly of a reference taxonomy for phylogenetic
1214 data synthesis. *Biodivers. Data J.* 5, e12581 (2017).

1215