eLife

Reviewed Preprint
v1 • May 29, 2026
Not revised

For correspondence:
jotin.email@gmail.com
sankar@ccmb.res.in

Competing interests: No competing interests declared

Funding: See page 13

Reviewing editor: Martin Graña, Institut Pasteur de Montevideo, Uruguay

Bacterial ancestry of the mitochondrial ATP exporter

Jotin Gogoi[1], Rajan Sankaranarayanan[1,2]

[1]CSIR-Centre for cellular and Molecular Biology, Hyderabad, India • [2]Academy of Scientific and Innovative Research (AcSIR), Ghaziabad, India

eLife Assessment

This potentially **useful** paper presents an intriguing hypothesis about the evolutionary origins of the SLC25 family of mitochondrial carrier proteins common to all eukaryotic life, proposing that all members originated from the ADP/ATP carrier (AAC) and that AAC itself may have emerged from bacterial homologs such as CysZ and YihY. While the phylogenetic analyses and structural searches are reasonable methodologies to explore ancient evolutionary events, the evidence provided here is deemed to provide **incomplete** support for the conclusion that the mitochondrial ATP transporter is related to CysZ and Yih.

https://doi.org/10.7554/eLife.111475.1.sa4

Abstract

Mitochondria originated through endosymbiosis of an Alphaproteobacterium within an Asgard archaeal host, with ATP export to the cytosol being a key driver for the organelle integration. The mitochondrial ATP/ADP carrier (AAC), a member of the SLC25 family, performs this critical function, whose evolutionary origin was not known due to the absence of any known prokaryotic homologues and was therefore termed a eukaryotic innovation. Here, using protein tertiary structure search combined with comprehensive sequence analyses, we identify conserved bacterial inner membrane transporters, CysZ and YihY, as putative homologues of mitochondrial AAC. CysZ and YihY are structurally similar to AAC, albeit with a circular permutation of one of the six transmembrane helices. Strikingly, we could identify the conserved MCF motif—a characteristic feature of the SLC25 family—in the bacterial sulfate transporter CysZ, suggesting a common ancestry. Together, our results identify a bacterial origin for the mitochondrial ATP exporter, thus resolving a long-standing question in mitochondrial evolution and a key step required for the emergence of eukaryotic cell complexity.

Introduction

Eukaryotic cell has evolved from its Asgard archaeal ancestor around 2 billion years ago upon symbiosis of an Alphaproteobacterial endosymbiont that became mitochondria (Vosseberg et al., 2024). The energy budget afforded by mitochondria enabled the emergence of the profound eukaryotic cell complexities, which the prokaryotic cells could never evolve in around 4 billion years of evolution (Lane and Martin, 2010). The surplus production and supply of ATP to the host cell incentivised the integration of the bacterial endosymbiont leading to its transformation into mitochondria that afforded the energetic cost for evolution of eukaryotic cell complexities (Roger et al., 2017; Martin, 2025; Speijer, 2025). Thus, the evolution of the mitochondrial ATP exporting translocase, ATP/ADP carrier (AAC), is a foundational step in all the models of the emergence of mitochondria and eukaryotic cell (Gray, 2014). Notably, mitochondrial AAC is the one and only known ATP exporter that is evolutionarily distinct from other ATP/ADP translocases present in plastids and in intracellular pathogens that import ATP (Amiri et al., 2003; Lin et al., 2025) (Figure 1).

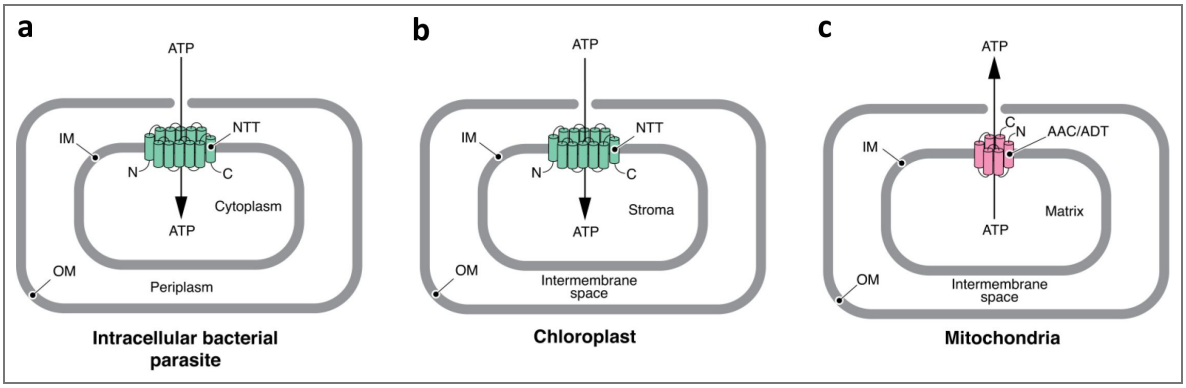


Figure 1. **ATP transporter types.**

(**a**) Nucleotide transporter (NTT) present in intracellular bacterial parasite that imports ATP from host cell cytoplasm. Parasite NTT features all alpha 12 TM containing MFS general substrate transporter fold (SCOP id: 2000248). (**b**) ATP/ADP translocator 1 (NTT1) present in plastid inner membrane responsible for importing ATP from plant cell cytoplasm. Plastid NTTs feature all alpha 12 TM containing MFS general substrate transporter fold (SCOP id: 2000248). (**c**) Mitochondrial ATP/ADP translocator (ADT/ANT/AAC) that resides in mitochondrial and mitochondria derived organelles that exports ATP to the eukaryotic cell cytoplasm. AACs feature all alpha 6 TM bundle Mitochondrial carrier fold (SCOP id: 2000266).

Mitochondrial AACs belong to the nuclear encoded six transmembrane (6TM) Solute Carrier Family 25 (SLC25), also known as mitochondrial carrier family (MCF) (Nury et al., 2006). SLC25 family is the largest mitochondrial carrier family constituted of 53 paralogues in *Homo sapiens*, 35 in *Saccharomyces cerevisiae*, 60 in *Arabidopsis thaliana* and 35 in *Andalucia godoyi* (Palmieri et al., 2011). Metabolites transported by SLC25 across the impermeable mitochondrial inner membrane support myriad of eukaryotic processes, such as: oxidative phosphorylation, amino acid metabolism, apoptosis, iron-sulphur cluster synthesis, haem synthesis, heat production, mitochondrial fusion/fission dynamics, etc. (Kunji et al., 2020; Ruprecht and Kunji, 2020; Kunji et al., 2025; Khan et al., 2025). AAC is the most extensively studied member of SLC25 family and for which the molecular mechanism of solute transport has been elucidated based on atomic structures of bovine and fungal AACs (Pebay-Peyroula et al., 2003; Ruprecht et al., 2019, 2014). SLC25 transporters are conserved in all the extant branches of eukaryotes and were present in the last eukaryotic common ancestor (LECA) (Kurland and Andersson, 2000). However, the evolutionary origin of the SLC25 mitochondrial carrier family, a crucial step in the emergence of mitochondria, remains an enigma owing to the lack of any trace of prokaryotic ancestry and therefore it is categorised as orphan or eukaryotic innovation (Karlberg et al., 2000).

Here, we harness protein tertiary structure search to screen for putative remote homolog of AAC in the recently generated AlphaFold-predicted structure database of proteome of archaeal and bacterial species. This approach leverages the principle that protein tertiary structure is greater conserved over evolutionary distance compared to protein sequence. Our protein structure-guided screen identified bacterial proteins with significant structural similarity with AACs, which can be potential remote homologs of AACs in prokaryotes. Notably, the bacterial proteins tertiary structure is related to the structure of AACs through circular permutation of one transmembrane helix at sequence level. Furthermore, our thorough sequence analysis revealed the conservation of the hallmark SLC25 sequence motif in one among the structure search hits of AACs in bacteria, CysZ — a conserved bacterial inner membrane residing sulfate transporter protein. Thus, our comprehensive analysis identifies a bacterial homologue of mitochondrial AACs on the basis of structural similarity, functional similarity and conservation of a hallmark sequence-motif. Overall, identification of a bacterial ancestry establishes endosymbiotic roots for the emergence of the mitochondrial ATP exporter – a foundational step in the evolution of mitochondria and the onset of eukaryotic cell complexity.

Results

Rooted phylogeny resolves AACs as the founder member of SLC25 carrier family

Insights into the sequential order in which the members of SLC25 carrier family emerged is crucial for understanding the evolutionary emergence of AAC during the origin of mitochondria. Inferring the root of the phylogenetic tree of SLC25 family protein sequences may resolve the order in which AACs emerged relative to the rest of SLC25 family members. For phylogenetic analysis of SLC25 family, we took advantage of the SLC25 sequences of *Andalucia godoyi* (Gray et al., 2020), a member of the eukaryotic order Jakobida under the supergroup Excavata, which is considered to be the most ancient branch of eukaryotes that is phylogenetically closest to LECA and harbours most bacterial-like mitochondria (Lang et al., 1997; Burger et al., 2013; Williamson et al., 2025). We also employed SLC25 sequences from *Saccharomyces cerevisiae* and *Paramecium tetraurelia*, which belong to the eukaryotic supergroups Opisthokonta and SAR, respectively. The above three organisms from which we employ the SLC25 sequences to infer the phylogenetic tree are representative of three distinct supergroups of eukaryotes that have diverged at the stage of LECA (Figure 2a) (Burki et al., 2020; Williamson et al., 2025). Next, we inferred phylogenetic tree of the SLC25 sequences using both

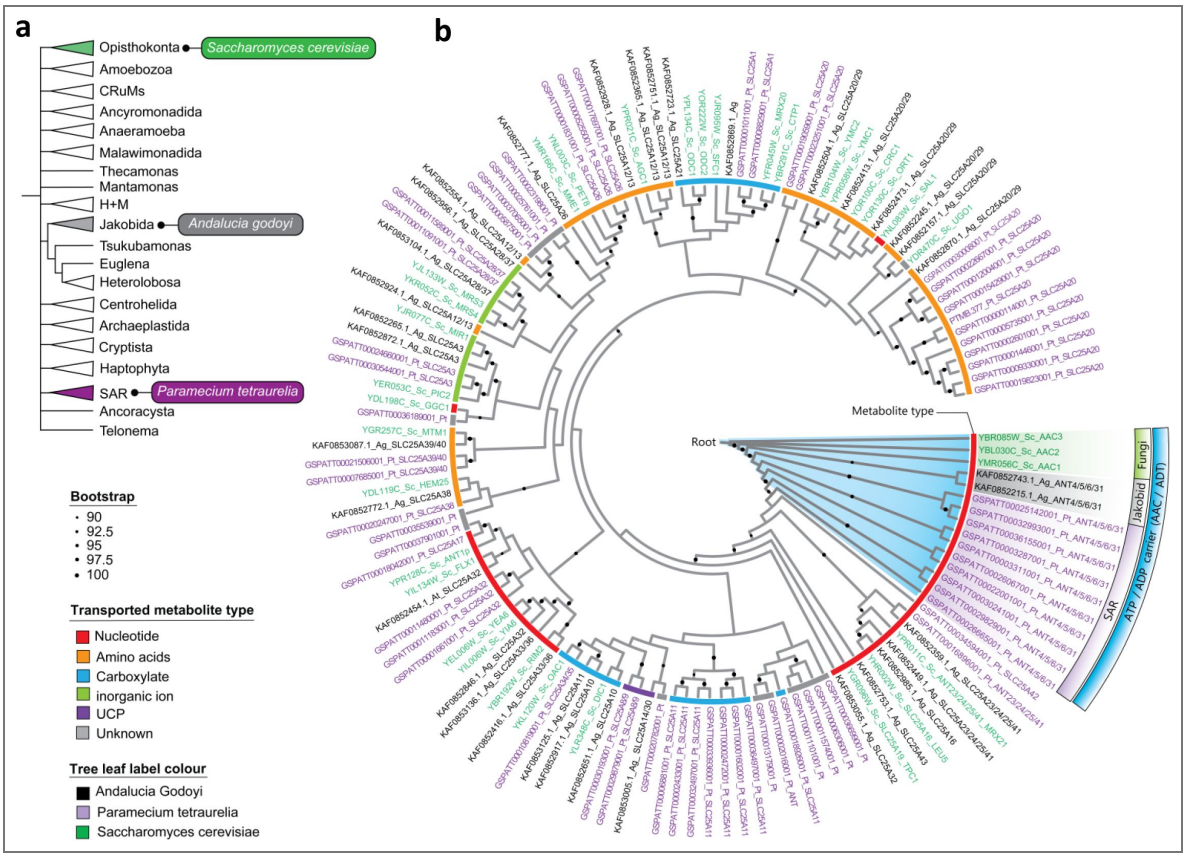


Figure 2. Mitochondrial ATP-ADP carriers (AACs) are the pioneer members of SLC25 family.

(a) Tree showing all the Eukaryotic supergroups and the phylogenetic position of *A. godoyi*, *S. cerevisiae* and *P. tetraurelia* in three evolutionarily distinct eukaryotic supergroups that have diverged at the stage of LECA. **(b)** Maximum likelihood phylogenetic tree of SLC25 carrier family members from *A. godoyi*, *S. cerevisiae* and *P. tetraurelia*. Branch support of bootstrap value ≥ 90 are indicated by black circles.

Maximum likelihood (ML) and Bayesian methods (Huelsenbeck and Ronquist, 2001; Nguyen et al., 2015). The ML phylogenetic tree of SLC25 carriers from *Andalucia godoyi*, *Paramecium tetraurelia* and *Saccharomyces cerevisiae* had all the AACs from the three organisms grouped together and also positioned at the root of the phylogenetic tree (Figure 2b). Grouping is observed also for the remaining SLC25 transporters from the three organisms that transport similar metabolites. The Bayesian phylogenetic tree of the same sequences also exhibits a similar tree topology where the AACs are positioned at the root of the tree (Supplementary Figure 1). Together, our comprehensive rooted phylogenetic analysis of SLC25 sequences from organisms belonging to three different eukaryotic supergroups that diverged at the stage of LECA resolves AACs as the evolutionarily founder member of the SLC25 carrier family.

Structure-guided screen detects putative remote homologue of AAC in archaea and bacteria

Comparative genomics studies have established that the mitochondrial AACs is part of a group of proteins that are categorised as eukaryotic innovations due to lack of any trace of prokaryotic ancestry (Karlberg et al., 2000). Protein tertiary structure is significantly more conserved over evolutionary distance than sequence (Holm and Sander, 1996; Illergård et al., 2009). Protein tertiary structure search therefore affords more confidence in detection of remote homologs well beyond the limits of sequence-based methods (Murzin et al., 1995; Orengo et al., 1997; Cheng et al., 2014). Recent advancement in *ab initio* protein structure prediction programs like AlphaFold2 has enabled the generation of high quality AlphaFold structure database (AF-DB) of the entire known protein sequence space (Baek et al., 2021; Jumper et al., 2021; Varadi et al., 2022). The predicted high quality structures in AF-DB allows for structure search-based screening of remote homologs in the *Foldome* of a target organism using programs like DALI AF-DB Search and Foldseek Search (Holm, 2022; van Kempen et al., 2024). We set out to perform structure search of AACs using X-ray crystal structure of *Saccharomyces cerevisiae* AAC (*Sc*AAC) (Pebay-Peyroula et al., 2003) and high confidence AlphaFold3 (Abramson et al., 2024) predicted structure of *Andalucia godoyi* AAC (*Ag*AAC), in the predicted structures of the whole proteome of archaeal and bacterial species in the AF-DB using DALI AF-DB Search and Foldseek Search **(**Figure 3a,b**,c)**. Both *Ag*AAC and *Sc*AAC structures were searched in predicted structures of the entire proteome of bacterial species *E. coli* (Gram-negative), *S. aureus* (Gram-positive) and archaeal species through DALI AF-DB search. AAC structures were also searched across all archaeal and bacterial proteome structures in AF-DB through Foldseek Search. The structure search hits from the DALI AF-DB and Foldseek Search, having reviewed UniProtKB entry (The UniProt Consortium, 2025), were filtered according to the degree of structural similarity to AAC based on DALI Z-score and TM-score cut-off criteria (Holm and Sander, 1993; Zhang and Skolnick, 2005), and presence of overall 6TM topology based on DeepTMHMM (Hallgren et al., 2022) and manual curation (Figure 3c). We retrieved five bacterial proteins and one archaeal protein that showed significant structural similarity to AACs, those that have DALI Z-score of more than 3 and TM-score of more than 0.4, and also features a 6TM fold topology like AACs **(**Figure 3d,e,f and Supplementary Figure 2). Next, we performed sequence search to check the presence of each archaeal and bacterial hits across archaeal and bacterial phyla respectively (Figure 4a). Only bacterial proteins YihY and CysZ were widely conserved across bacterial phyla—including Alphaproteobacteria—which asserts their likelihood of being the evolutionary ancestor of mitochondrial AACs; whereas, the rest of the archaeal and bacterial proteins were filtered out owing to their very confined distribution across selective prokaryotic phyla **(**Figure 4b,c**,d,e)**. Overall, our pipeline, comprising of protein tertiary structure search coupled with extensive curation based on phylogenetic conservation and transmembrane topology, identified bacterial proteins YihY and CysZ as potential remote bacterial homologs of AACs.

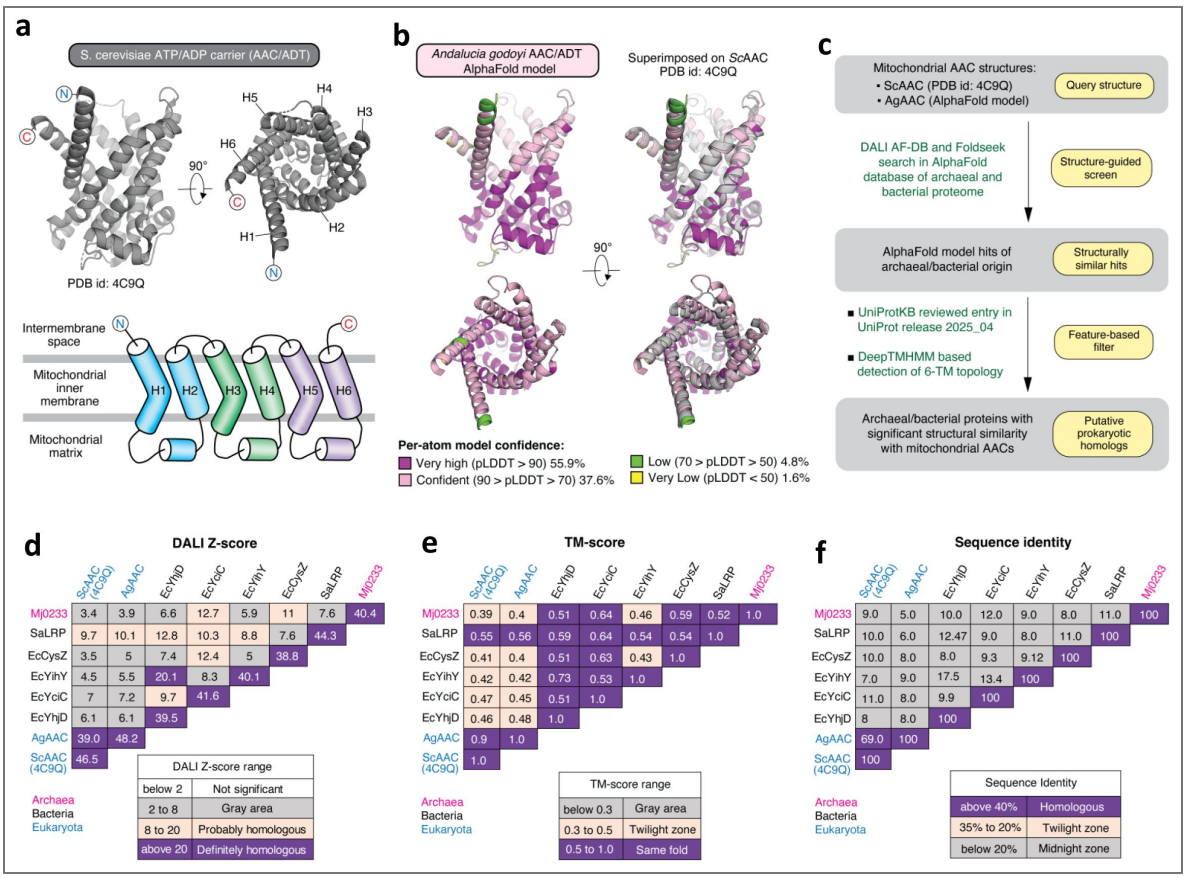


Figure 3. Structure guided screen of AACs in AlphaFold structure database of prokaryote proteome.

(a) X-ray crystal structure of *Saccharomyces cerevisiae* AAC and its secondary structure topology. **(b)** AlphaFold predicted structure model of *Andalucia godoyi* AAC. **(c)** Schematic diagram of the pipeline employed in this study leveraging protein 3D-structrure search for identification of potential remote homologue of AACs in archaea and bacteria. **(d)** DALI Z-score, **(e)** TM-score, **(f)** sequence identity of archaeal and bacterial structure search hits of AACs.

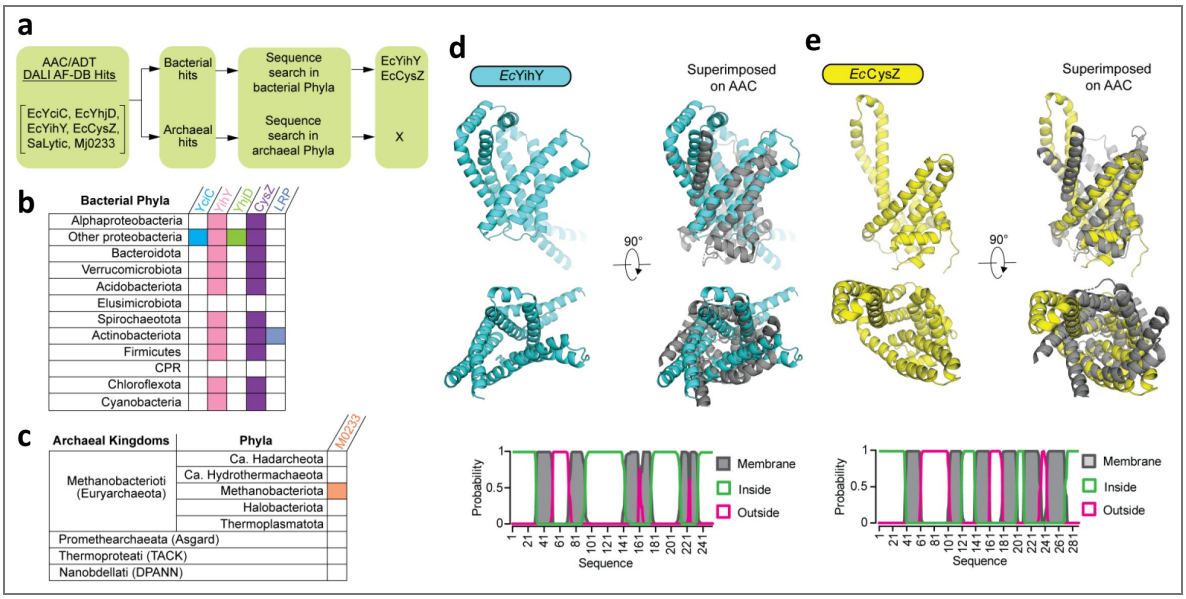


Figure 4. **Putative homologue of AAC in prokaryotes.**

(a) Schematic flowchart depicting the approach for selection of the candidate homologue of AACs in prokaryote structural hits based on presence or absence of the structural hits in the bacterial and archaeal phyla. **(b)** Gene conservation of bacterial structural search hits of AAC across bacterial phyla. **(c)** Gene conservation of archaeal hits of structural search of AAC in archaeal kingdoms and phyla. **(d)** AlphaFold predicted model of *E. coli* YihY and its predicted transmembrane topology. **(e)** AlphaFold predicted model of *E. coli* CysZ and its predicted transmembrane topology.

YihY and CysZ are structurally related to AACs through circular permutation

Structure of bacterial proteins YihY and CysZ is significantly similar to AACs, quantitated based on DALI Z-score of ~ 5 and TM-score of ~ 0.4, which both indicate probable homology (Figure 3d,e). Inspection of the structural superimposition of YihY/CysZ and AACs reveal that although there is correspondence in the spatial arrangement of six TM helices in the structures, the connectivity of the helices in the primary structure is different (Figure 5a). AACs feature a six TM-helical bundle (H1-6) with both N- and C-terminals present toward the intermembrane space (IMS). YihY and CysZ share a common structural fold featuring six TM-helical bundle with both N- and C-terminals in the cytoplasmic side. Inspection of the structural superimposition of YihY/CysZ and AAC reveal that there is circular permutation of the N- and C-terminal involving one TM helix in the protein sequence. The TM helix of YihY/CysZ structurally corresponding to the H6 of AAC has transposed to the N-terminal in the protein sequence, resulting in the shift of their N- and C-terminal at the position corresponding to the loop M1 of AAC, albeit maintaining a similar overall spatial arrangement of the six TM helices (Figure 5b). Next, we sought to explore the effects of circular permutation of sequence on overall tertiary structure of YihY/CysZ. We computationally circular permuted YihY and CysZ by transposing their 1st TM helix (h1) in the sequence to the C-terminal akin to the topology of AACs and then leveraged AlphaFold for prediction of their tertiary structure model (Figure 5c,d,e). The AlphaFold predicted structure model of circular permuted proteins, YihY^{CP} and CysZ^{CP}, superimposed on the AlphaFold model of native proteins with an RMSD of 0.38 Å over 290 Cα atoms and 0.29 Å over 253 Cα atoms respectively, suggesting that the inherent tertiary structure of this group of proteins is unaffected by alteration at sequence level by circular permutation (Figure 5g,h). Furthermore, structure of circular permuted YihY^{CP} and CysZ^{CP} is more similar to AACs as indicated by the uptick in structure similarity scores compared to their native forms i.e., DALI Z-score increased from ~ 4 to ~ 6 and TM-score increased from ~ 0.4 to ~ 0.5 (Figure 5i,j). Overall, our analysis reveals that the six TM helix bundle structure of YihY/CysZ and AACs are related by circular permutation of one TM helix, suggesting circular permutation as a potential step in the evolutionary route for the emergence of AACs from bacterial proteins YihY/CysZ.

Presence of signature MCF motif in bacterial inner membrane sulfate transporter CysZ supports homology with AACs

Homology between two proteins means that both have descended from a common ancestor. Statistically significant sequence similarity between proteins alone is considered a definitive basis for assertion of homology. In absence of significant sequence similarity for highly diverged proteins, homology can be supported based on combination of structural resemblance, conservation of sequence motifs and functional similarity (Grishin, 2001). In case of YihY/CysZ and AACs we show that they share significant structural similarity, which suggest for potential homology between them, but they are highly divergent at sequence level with sequence similarity score in the mid-night zone (Figure 3d,e,f). In order to gain clarity regarding the hypothesised homology between YihY/CysZ and AACs, we examined their functional similarity and also checked for conservation of signature sequence motif. YihY (EcoCyc id: EG11851) is a conserved bacterial inner membrane protein of unknown function. CysZ (EcoCyc id: EG10003) is a conserved bacterial inner membrane protein, which has been shown to be a high specificity sulfate transporter involved in sulfate import inside cell for cysteine biosynthesis (Assur Sanghai et al., 2018; Zhang et al., 2014; Britton et al., 1983). Thus, functionally CysZ is similar to AACs in that they both are membrane transporters of small metabolite.

Next, we sought to check for any motifs conserved between YihY/CysZ and AACs. All the members of SLC25 carrier family feature a conserved structure with six TM helices having a three-fold internal symmetry of a repeat containing two TM helices called as “mito_carr’ domain (Figure 6a) (Pebay-Peyroula et al., 2003; Nury et al., 2006). The odd-numbered helix of each repeat features a highly conserved characteristic sequence motif – PX[D/E]XX[K/R]X[K/R], known as the

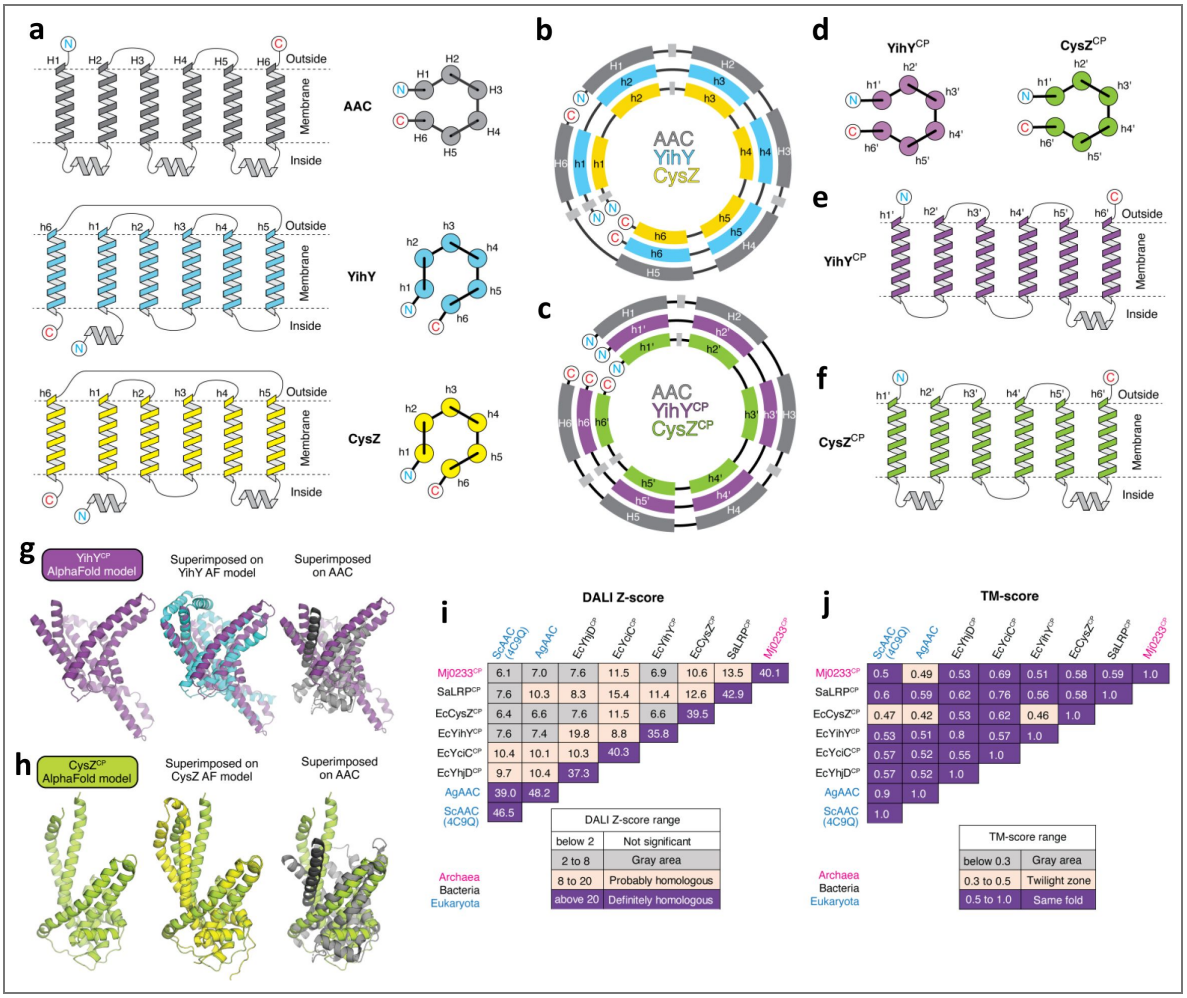


Figure 5. AAC is structurally related to prokaryotic proteins through circular permutation of a TM-helix.

(a) Secondary structure topology and six TM-helix of AAC, YihY and CysZ (lateral and axial views). **(b)** Primary structure of AAC, YihY and CysZ showing circular permutation of one TM-helix. **(c)** Hypothetical circular permuted YihY^{CP} and CysZ^{CP} featuring similar secondary structure topology as to that of AAC. (**d**) Axial view of the six TM-helix topology of YihY^{CP} and CysZ^{CP}. Secondary structure topology of (**e**) YihY^{CP} and (**f**) CysZ^{CP}. **(g)** AlphaFold predicted structure model of YihY^{CP} and its superimposition on YihY and AAC. (**h**) AlphaFold predicted structure model of CysZ^{CP} and its superimposition on CysZ and AAC. **(i)** DALI Z-score and **(j)** TM-score for structural superimposition of AAC on AlphaFold predicted structure model of circular permuted bacterial and archaeal structure search hits.

MCF-motif, which forms inter-domain salt-bridges called as the matrix salt-bridge network. The matrix salt-bridge network is an integral part of the gates that has direct role in the alternate access mechanism of solute transport by SLC25 solute carrier family (Ruprecht and Kunji, 2020). In case of YihY/CysZ, the six helical bundle can be assigned with repeat 2 and repeat 3 based on their mapping on the repeats of AACs in the structural overlap (Figure 6b,c,d,e,f,g,h,i). In order to check for conservation of sequence motif between AACs and YihY/CysZ, we separated the sequences of the repeats of AACs and YihY/CysZ and then generated a structure-based multiple sequence alignment. Structure-based multiple sequence alignment of repeats revealed the conservation of the signature motif PX[D/E]XX[K/R]X[K/R] in the repeat 3 of CysZ **(**Figure 6j,k and Supplementary Figure 3a).Thus, the bacterial sulfate transporter CysZ not only features a significantly similar structure like AACs, it also exhibits functional similarity and contains the signature MCF-motif, which altogether accentuate the support for their evolutionary relationship and thereby asserts homology between CysZ and AACs.

Discussion

Evolutionary origin of mitochondrial ATP exporter, a foundational step in all the models of the evolution of mitochondria, remains unclear (Gray, 2015). For decades, comparative genomics studies have categorised mitochondrial AACs and rest of the SLC25 family members as orphan or eukaryotic innovation owing to the lack of any trace of prokaryotic ancestry. Here, we first delineated the sequential order of evolutionary origin of the members of the SLC25 family through a comprehensive phylogenetic analysis and resolved AACs as the evolutionary founder member of SLC25 family. Then, we harnessed protein tertiary structure search of experimental as well as AlphaFold predicted structure of AACs in the AlphaFold predicted structure database of prokaryotic organisms, in tandem with thorough sequence analysis, to establish bacterial sulfate transporter CysZ as a potential evolutionary ancestor of mitochondrial AACs and the SLC25 carrier family.

AACs and all the other members of SLC25 carrier family features a three-fold pseudosymmetry of a two TM helical repeat (Kuan and Saier, 1993). SLC25 members has been hypothesised to have evolved from multiple duplication of a two TM helical repeating unit, a process dubbed as repeat *triplication* (Palmieri et al., 2011; Haferkamp and Schmitz-Esser, 2012). Due to the overall structural similarity and conservation of the signature MCF-motif, the repeat-3 of bacterial sulfate transporter CysZ makes for an ideal *arch*-repeat from which the AAC and the rest of the SLC25 carriers could have evolved through multiple duplication events.

Widely accepted syntrophy hypothesis and entangle-engulf-endogenize (E3) hypothesis for the emergence of proto-mitochondria are based on metabolic syntrophy which later transformed into energy exchange in form of ATP in mitochondria. The above metabolic syntrophy process involved sulfate as one of the metabolite exchanged between symbiont bacteria and the archaeal host (Imachi et al., 2020; López-García and Moreira, 2020; Vosseberg et al., 2024). Our finding that the AAC emerged from the conserved bacterial sulfate transporter CysZ, strikingly, suggest that a transporter involved in the initial metabolic syntrophy that spurred the formation of proto-mitochondria was eventually evolved and re-purposed for ATP export in mitochondria at the stage of LECA (Figure 7).

About half of the ~1000 mitochondrial targeted proteins lack any trace of archaeal or bacterial ancestry, which are categorised as orphan or eukaryotic innovation (Kurland and Andersson, 2000; Gray, 2015). Various evolutionary mechanisms have been implicated for the evolution of this profound eukaryotic genomic novelties (Vosseberg et al., 2024). Recently, optimization of the macromolecules for attaining compatibility among the bacterial and archaeal derived components for successful integration of various processes during the emergence of mitochondria was implicated in the emergence of the mosaic mitochondrial proteome (Gogoi et al., 2022; Kumar et al., 2022, 2023). We speculate that many of the novel eukaryotic innovations that lack any trace of prokaryotic ancestry could be a result of extreme degree of optimization of the prokaryotic proteins, in the evolutionary course of emergence of the eukaryotic proteins, that the sequence level similarity among them is eroded beyond the detection capabilities of the sequence

Figure 6. **Conservation of MCF motif in CysZ.**

(a) Secondary structure topology of AAC and the three-fold internal symmetry of two TM-helix containing repeat. **(b)** Topology of CysZ and its two TM-helix repeats. Structure of **(c)** AAC repeat 1, (**d**) AAC repeat 2 **(e)** AAC repeat 3, **(f)** CysZ repeat 2, (**g**) CysZ repeat 3. **(h)** Structural overlay of AAC repeat 2 and CysZ repeat 2. **(i)** Structural overlay of AAC repeat 3 and CysZ repeat 3. **(j)** MCF motif in AAC repeat 1, 2 and 3. **(k)** MCF motif in CysZ repeat 3.

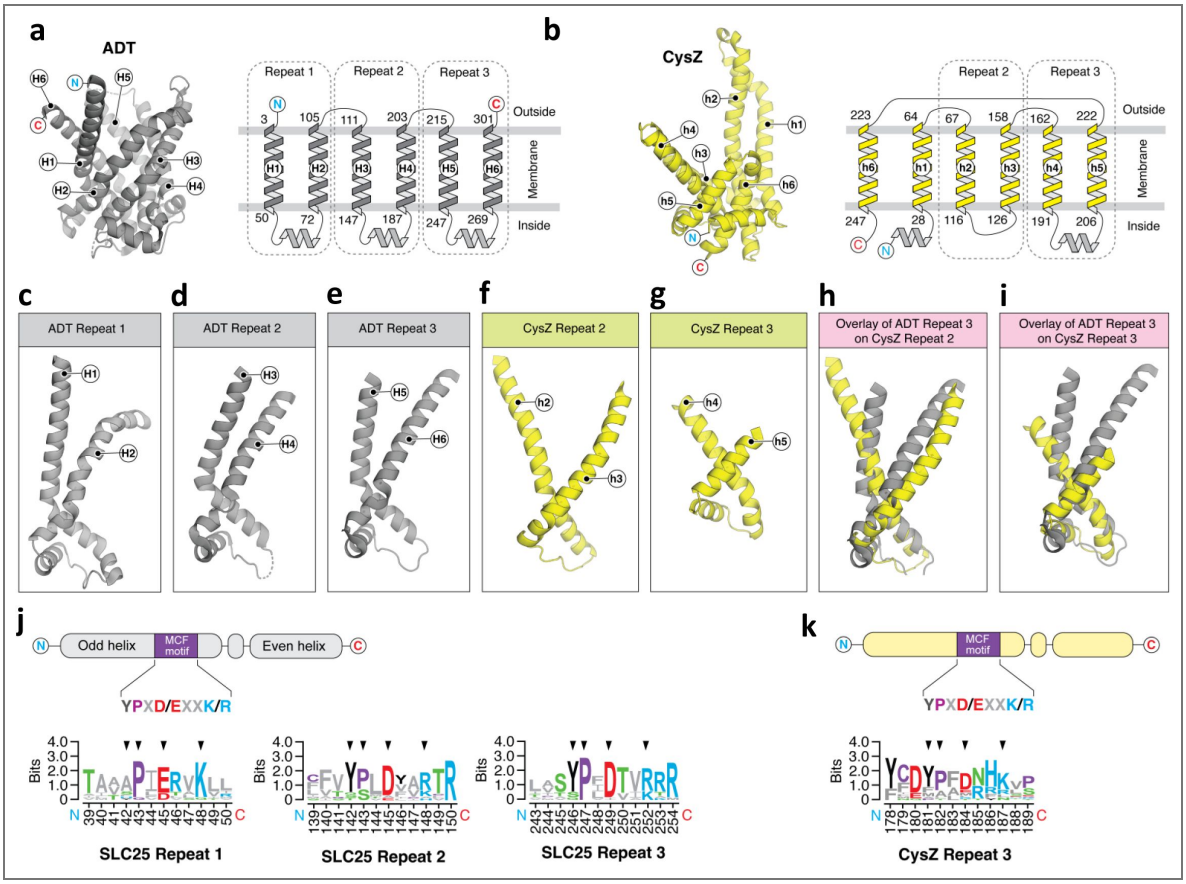
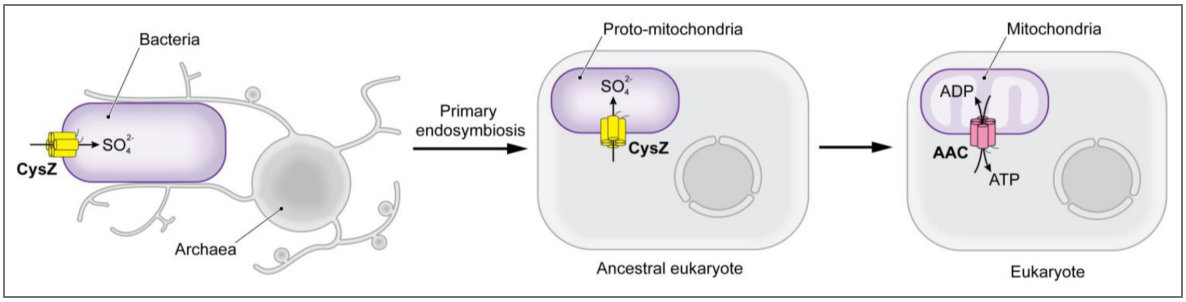


Figure 7. **Model for evolution of mitochondrial AAC from bacterial CysZ.**

Schematic model depicting proposed origin of mitochondrial ATP exporter (AAC) from the bacterial sulfate transporter CysZ during the endosymbiotic origin of mitochondria in eukaryotic cells from bacterial endosymbiont. The bacterial sulfate transporter CysZ, involved in the initial metabolic syntrophy that spurred the formation of proto-mitochondria was eventually evolved and re-purposed for ATP export in mitochondria at the stage of LECA. Evolution of AAC from bacterial sulfate transporter CysZ is a foundational step of mitochondrial emergence, and therefore the onset of eukaryotic cell complexities.



search methods. Here, we have leveraged the recently generated database of high quality AlphaFold predicted structure of the entire sequence space to screen for remote prokaryotic homolog of AACs using protein tertiary structure search programs. The retrieved structurally similar prokaryotic proteins facilitated implementation of focused sequence analysis methods to unravel the conserved MCF motif that unequivocally established the evolutionary relationship and therefore the homology between AAC and bacterial protein CysZ. We envisage that such an approach can be expanded to deorphanise many more such remote homologs that are categorised as orphan/eukaryotic innovation by unravelling the distant evolutionary relationship with their prokaryotic ancestors. Such studies may further shine a light on the evolutionary mechanisms that underpins the emergence of the eukaryotic genomic novelties and illuminate the origins of eukaryotic cellular complexities.

Methods

Sequence and annotation retrieval

All the SLC25 sequences from *Paramecium tetraurelia* and *Saccharomyces cerevisiae* were retrieved from KEGG GENES database using KEGG BLAST search (*https://www.genome.jp/tools/blast/*) of S. cerevisiae AAC3 (Uniprot id: P18238). SLC25 sequences of *Andalucia godoyi* were retrieved by performing Domain Enhanced Lookup Time Accelerated BLAST (DELTA-BLAST) of *S. cerevisiae* AAC3 (Uniprot id: P18238) sequence. The annotations details for each sequence were retrieved by searching the corresponding accession code in either Uniprot or NCBI Gene database.

Sequence alignment

Structure-based multiple sequence alignment of the SLC25 sequences was performed using *Expresso* program in T-coffee server (Di Tommaso et al., 2011) using S. cerevisiae AAC structure (PDB id: 4C9Q) as structural template. Structure-based multiple sequence alignment of repeats was created by first separating the sequences of repeats from SLC25 and CysZ and then were aligned in *Expresso* program in T-coffee server by using structure of repeat extracted from S. cerevisiae AAC structure (PDB id: 2C9Q). Alignments were curated in Jalview (Waterhouse et al., 2009).

Phylogenetic tree reconstruction

The Maximum likelihood phylogenetic tree was reconstructed from the structure-based multiple sequence alignment using online program IQ-tree (*http://iqtree.cibiv.univie.ac.at/*) (Nguyen et al., 2015). The Bayesian phylogenetic tree was reconstructed in using online program MrBayes (*https://www.phylogeny.fr/*) (Huelsenbeck and Ronquist, 2001). The phylogenetic trees were visualized in iTol (Letunic and Bork, 2019).

Atomic coordinates of the structures of AAC were downloaded from RCSB protein data bank. Structural superimposition of the structures was carried out in Coot using SSM method. Structural superimposition of bacterial and archaeal structures on AACs was carried out using CEAlign in PyMol. High quality predicted AlphaFold structure of proteins were retrieved from AlphaFold Protein Structure Database (*https://alphafold.ebi.ac.uk/*) (Varadi et al., 2022). Prediction of structures were carried out in AlphaFold Server (*https://alphafoldserver.com/*) (Abramson et al., 2024).

Structure Search

Protein structure search was carried out in DALI AF-DB search (*http://ekhidna2.biocenter.helsinki.fi/dali/*) (Holm, 2022), and Foldseek Search server (*https://search.foldseek.cojotinm/search*) (van Kempen et al., 2024).

Data availability

Atomic coordinates of the Saccharomyces cerevisiae AAC structure used in the study was obtained from RCSB PDB under the Entry id: 4C9Q. All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Material.

Acknowledgements

We thank all the members of the RS's Laboratory at CCMB for the critical discussions on the work and for careful review of the manuscript.

JG thanks University Grants Commission, India for research fellowship. RSN thanks healthcare theme projects-Fundamental and Innovative CSIR in Science of Tomorrow (FIRST; MLP-0162) and Niche Creation Project (NCP; MLP-0138) of CSIR, India; J.C. Bose Fellowship of SERB, India.

Additional information

Author contributions

JG and RSN designed the study. JG carried out the analysis. JG and RSN conceived and supervised the study. JG and RSN wrote the manuscript and reviewed it.

Funding

Funder	Grant reference number	Author
Council of Scientific and Industrial Research, India (CSIR)	MLP-0162	Rajan Sankaranarayanan
Council of Scientific and Industrial Research, India (CSIR)	MLP-0138	Rajan Sankaranarayanan
DST | Science and Engineering Research Board (SERB)	J. C. Bose Fellowship	Rajan Sankaranarayanan

Author ORCID iDs

Rajan Sankaranarayanan: https://orcid.org/0000-0003-4524-9953

Additional files

Supplementary figures.

References

Abramson J, Adler J, Dunger J, Evans R, Green T, Pritzel A, Ronneberger O, Willmore L, Ballard AJ, Bambrick J, *et al.* (2024) Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**:493-500 https://doi.org/10.1038/s41586-024-07487-w | PubMed

Amiri H, Karlberg O, Andersson SGE (2003) Deep origin of plastid/parasite ATP/ADP translocases. *Journal of Molecular Evolution* **56**:137-150 https://doi.org/10.1007/s00239-002-2387-0 | PubMed

Assur Sanghai Z, Liu Q, Clarke OB, Belcher-Dufrisne M, Wiriyasermkul P, Giese MH, Leal-Pinto E, Kloss B, Tabuso S, Love J, *et al.* (2018) Structure-based analysis of CysZ-mediated cellular uptake of sulfate. *eLife* **7**:e27829 https://doi.org/10.7554/eLife.27829 | PubMed

Baek M, DiMaio F, Anishchenko I, Dauparas J, Ovchinnikov S, Lee GR, Wang J, Cong Q, Kinch LN, Schaeffer RD, *et al.* (2021) Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **373**:871-876 https://doi.org/10.1126/science.abj8754 | PubMed

Britton P, Boronat A, Hartley DA, Jones-Mortimer MC, Kornberg HL, Parra F (1983) Phosphotransferase-mediated regulation of carbohydrate utilization in Escherichia coli K12: location of the gsr (tgs) and iex (crr) genes by specialized transduction. *Journal of General Microbiology* **129**:349-356

https://doi.org/10.1099/00221287-129-2-349 | PubMed

Burger G, Gray MW, Forget L, Lang BF (2013) Strikingly bacteria-like and gene-richmitochondrial genomes throughout jakobid protists. *Genome Biology and Evolution* **5**:418-438 https://doi.org/10.1093/gbe/evt008 | PubMed

Burki F, Roger AJ, Brown MW, Simpson AGB (2020) The New Tree of Eukaryotes. *Trends in Ecology & Evolution* **35**:43-55 https://doi.org/10.1016/j.tree.2019.08.008 | PubMed

Cheng H, Schaeffer RD, Liao Y, Kinch LN, Pei J, Shi S, Kim B-H, Grishin NV (2014) ECOD: An Evolutionary Classification of Protein Domains. *PLOS Computational Biology* **10**:e1003926 https://doi.org/10.1371/journal.pcbi.1003926 | PubMed

Di Tommaso P, Moretti S, Xenarios I, Orobitg M, Montanyola A, Chang JM, Taly JF, Notredame C (2011) T-Coffee: A web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. *Nucleic Acids Research* **39** https://doi.org/10.1093/nar/gkr245 | PubMed

Gogoi J, Bhatnagar A, Ann Kezia J, Pottabathini S, Singh R, Mazeed M, Kuncha SK, Kruparani SP, Sankaranarayanan R (2022) Switching a conflicted bacterial DTD-tRNA code is essential for the emergence of mitochondria. *Science Advances* **8**:7307 https://doi.org/10.1126/SCIADV.ABJ7307 | PubMed

Gray MW (2015) Mosaic nature of the mitochondrial proteome: Implications for the origin and evolution of mitochondria. *Proceedings of the National Academy of Sciences of the United States of America* **112**:10133-10138 https://doi.org/10.1073/pnas.1421379112 | PubMed

Gray MW (2014) The pre-endosymbiont hypothesis: A new perspective on the origin and evolution of mitochondria. *Cold Spring Harbor Perspectives in Biology* **6** https://doi.org/10.1101/cshperspect.a016097 | PubMed

Gray MW, Burger G, Derelle R, Klimeš V, Leger MM, Sarrasin M, Vlček Č, Roger AJ, Eliáš M, Lang BF (2020) The draft nuclear genome sequence and predicted mitochondrial proteome of Andalucia godoyi, a protist with the most gene-rich and bacteria-like mitochondrial genome. *BMC Biology* **18**:1-35 https://doi.org/10.1186/s12915-020-0741-6 | PubMed

Grishin NV (2001) Fold Change in Evolution of Protein Structures. *Journal of Structural Biology* **134**:167-185 https://doi.org/10.1006/jsbi.2001.4335 | PubMed

Haferkamp I, Schmitz-Esser S (2012) The Plant Mitochondrial Carrier Family: Functional and Evolutionary Aspects. *Frontiers in Plant Science* **3** https://doi.org/10.3389/fpls.2012.00002 | PubMed

Hallgren J, Tsirigos K, Pedersen MD, Almagro Armenteros JJ, Marcatili P, Nielsen H, Krogh A, Winther O (2022) DeepTMHMM predicts alpha and beta transmembrane proteins using deep neural network. *bioRxiv* https://doi.org/10.1101/2022.04.08.487609

Holm L (2022) Dali server: structural unification of protein families. *Nucleic Acids Research* **50**:W210-W215 https://doi.org/10.1093/nar/gkac387 | PubMed

Holm L, Sander C (1996) Mapping the Protein Universe. *Science* **273**:595-602 https://doi.org/10.1126/science.273.5275.595 | PubMed

Holm L, Sander C (1993) Protein structure comparison by alignment of distance matrices. *Journal of Molecular Biology* **233**:123-138 https://doi.org/10.1006/jmbi.1993.1489 | PubMed

Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics (Oxford, England)* **17**:754-755 https://doi.org/10.1093/bioinformatics/17.8.754 | PubMed

Illergård K, Ardell DH, Elofsson A (2009) Structure is three to ten times more conserved than sequence--a study of structural response in protein cores. *Proteins* **77**:499-508 https://doi.org/10.1002/prot.22458 | PubMed

Imachi H, Nobu MK, Nakahara N, Morono Y, Ogawara M, Takaki Y, Takano Y, Uematsu K, Ikuta T, Ito M, *et al.* (2020) Isolation of an archaeon at the prokaryote–eukaryote interface. *Nature* **577**:519-525 https://doi.org/10.1038/s41586-019-1916-6 | PubMed

Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, *et al.* (2021) Highly accurate protein structure prediction with AlphaFold. *Nature* **596**:583-589 https://doi.org/10.1038/s41586-021-03819-2 | PubMed

Karlberg O, Canbäck B, Kurland CG, Andersson SGE (2000) The dual origin of the yeast mitochondrial proteome. *Yeast* **17**:170-187 https://doi.org/10.1002/1097-0061(20000930)17:3<170::AID-YEA25>3.0.CO;2-V | PubMed

Khan A, Liu Y, Gad M, Kenny TC, Birsoy K (2025) Solute carriers: The gatekeepers of metabolism. *Cell* **188**:869-884 https://doi.org/10.1016/j.cell.2025.01.015 | PubMed

Kuan J, Saier MH (1993) The mitochondrial carrier family of transport proteins: structural, functional, and evolutionary relationships. *Critical Reviews in Biochemistry and Molecular Biology* **28**:209-233 https://doi.org/10.3109/10409239309086795 | PubMed

Kumar P, Babu KSD, Singh AK, Singh DK, Nalli A, Mukul SJ, Roy A, Mazeed M, Raman B, Kruparani SP, *et al.* (2023) Distinct localization of chiral proofreaders resolves organellar translation conflict in plants. *Proceedings of the National Academy of Sciences* **120**:e2219292120 https://doi.org/10.1073/pnas.2219292120 | PubMed

Kumar P, Bhatnagar A, Sankaranarayanan R (2022) Chiral proofreading during protein biosynthesis and its evolutionary implications. *FEBS letters* **596** https://doi.org/10.1002/1873-3468.14419 | PubMed

Kunji ERS, King MS, Ruprecht JJ, Thangaratnarajah C (2020) The SLC25 Carrier Family: Important Transport Proteins in Mitochondrial Physiology and Pathology. *Physiology (Bethesda, Md.)* **35**:302-327 https://doi.org/10.1152/physiol.00009.2020 | PubMed

Kunji ERS, Mavridou V, King MS, Cimadamore-Werthein C, Jaiquel Baron S, Jones SA, King AC, Springett R, Chand D, Palmer SM, *et al.* (2025) The peculiar properties of mitochondrial carriers of the SLC25 family. *The Biochemical Journal* **482**:1047-1078 https://doi.org/10.1042/BCJ20253171 | PubMed

Kurland CG, Andersson SGE (2000) Origin and Evolution of the Mitochondrial Proteome. *Microbiology and Molecular Biology Reviews* **64**:786-820 https://doi.org/10.1128/mmbr.64.4.786-820.2000 | PubMed

Lane N, Martin W (2010) The energetics of genome complexity. *Nature* **467**:929-934 https://doi.org/10.1038/nature09486 | PubMed

Lang BF, Burger G, O'Kelly CJ, Cedergren R, Golding GB, Lemieux C, Sankoff D, Turmel M, Gray MW (1997) An ancestral mitochondrial DNA resembling a eubacterial genome in miniature. *Nature* **387**:493-497 https://doi.org/10.1038/387493a0 | PubMed

Letunic I, Bork P (2019) Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Research* **47**:W256-W259 https://doi.org/10.1093/nar/gkz239 | PubMed

Lin H, Huang J, Li T, Li W, Wu Y, Yang T, Nian Y, Lin X, Wang J, Wang R, *et al.* (2025) Structure and mechanism of the plastid/parasite ATP/ADP translocator. *Nature* **641**:797-804 https://doi.org/10.1038/s41586-025-08743-3 | PubMed

López-García P, Moreira D (2020) The Syntrophy hypothesis for the origin of eukaryotes revisited. *Nature Microbiology* **5**:655-667 https://doi.org/10.1038/s41564-020-0710-4 | PubMed

Martin WF (2025) ATP requirements for growth reveal the bioenergetic impact of mitochondrial symbiosis. *Biochimica et Biophysica Acta (BBA) - Bioenergetics* **1866**:149564 https://doi.org/10.1016/j.bbabio.2025.149564 | PubMed

Murzin AG, Brenner SE, Hubbard T, Chothia C (1995) SCOP: A structural classification of proteins database for the investigation of sequences and structures. *Journal of Molecular Biology* **247**:536-540 https://doi.org/10.1016/S0022-2836(05)80134-2 | PubMed

Nguyen LT, Schmidt HA, Von Haeseler A, Minh BQ. (2015) IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution* **32**:268-274 https://doi.org/10.1093/molbev/msu300 | PubMed

Nury H, Dahout-Gonzalez C, Trézéguet V, Lauquin GJM, Brandolin G, Pebay-Peyroula E (2006) Relations Between Structure and Function of the Mitochondrial ADP/ATP Carrier. *Annual Review of Biochemistry* **75**:713-741 https://doi.org/10.1146/annurev.biochem.75.103004.142747 | PubMed

Orengo CA, Michie AD, Jones S, Jones DT, Swindells MB, Thornton JM (1997) CATH--a hierarchic classification of protein domain structures. *Structure (London, England: 1993)* **5**:1093-1108 https://doi.org/10.1016/s0969-2126(97)00260-8 | PubMed

Palmieri F, Pierri CL, De Grassi A, Nunes-Nesi A, Fernie AR. (2011) Evolution, structure and function of mitochondrial carriers: a review with new insights. *The Plant Journal* **66**:161-181 https://doi.org/10.1111/j.1365-313X.2011.04516.x | PubMed

Pebay-Peyroula E, Dahout-Gonzalez C, Kahn R, Trézéguet V, Lauquin GJ-M, Brandolin G (2003) Structure of mitochondrial ADP/ATP carrier in complex with carboxyatractyloside. *Nature* **426**:39-44 https://doi.org/10.1038/nature02056 | PubMed

Roger AJ, Muñoz-Gómez SA, Kamikawa R (2017) The Origin and Diversification of Mitochondria. *Current Biology* **27**:R1177-R1192 https://doi.org/10.1016/j.cub.2017.09.015 | PubMed

Ruprecht JJ, Hellawell AM, Harding M, Crichton PG, McCoy AJ, Kunji ERS (2014) Structures of yeast mitochondrial ADP/ATP carriers support a domain-based alternating-access transport mechanism. *Proceedings of the National Academy of Sciences* **111**:E426-E434 https://doi.org/10.1073/pnas.1320692111 | PubMed

Ruprecht JJ, King MS, Zögg T, Aleksandrova AA, Pardon E, Crichton PG, Steyaert J, Kunji ERS (2019) The Molecular Mechanism of Transport by the Mitochondrial ADP/ATP Carrier. *Cell* **176**:435-447.e15 https://doi.org/10.1016/j.cell.2018.11.025 | PubMed

Ruprecht JJ, Kunji ERS (2020) The SLC25 Mitochondrial Carrier Family: Structure and Mechanism. *Trends in Biochemical Sciences* **45**:244-258 https://doi.org/10.1016/j.tibs.2019.11.001 | PubMed

Speijer D (2025) Eukaryogenesis From FECA to LECA: Radical Steps Along the Way. *BioEssays* **47**:e70063 https://doi.org/10.1002/bies.70063 | PubMed

The UniProt Consortium (2025) UniProt: the Universal Protein Knowledgebase in 2025. *Nucleic Acids Research* **53**:D609-D617 https://doi.org/10.1093/nar/gkae1010 | PubMed

van Kempen M, Kim SS, Tumescheit C, Mirdita M, Lee J, Gilchrist CLM, Söding J, Steinegger M. (2024) Fast and accurate protein structure search with Foldseek. *Nature Biotechnology* **42**:243-246 https://doi.org/10.1038/s41587-023-01773-0 | PubMed

Varadi M, Anyango S, Deshpande M, Nair S, Natassia C, Yordanova G, Yuan D, Stroe O, Wood G, Laydon A, *et al.* (2022) AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Research* **50**:D439-D444 https://doi.org/10.1093/nar/gkab1061 | PubMed

Vosseberg J, van Hooff JJE, Köstlbacher S, Panagiotou K, Tamarit D, Ettema TJG (2024) The emerging view on the origin and early evolution of eukaryotic cells. *Nature* **633**:295-305 https://doi.org/10.1038/s41586-024-07677-6 | PubMed

Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ (2009) Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**:1189-1191 https://doi.org/10.1093/bioinformatics/btp033 | PubMed

Williamson K, Eme L, Baños H, McCarthy CGP, Susko E, Kamikawa R, Orr RJS, Muñoz-Gómez SA, Minh BQ, Simpson AGB, *et al.* (2025) A robustly rooted tree of eukaryotes reveals their excavate ancestry. *Nature* **640**:974-981 https://doi.org/10.1038/s41586-025-08709-5 | PubMed

Zhang L, Jiang W, Nan J, Almqvist J, Huang Y (2014) The *Escherichia coli* CysZ is a pH dependent sulfate transporter that can be inhibited by sulfite. *Biochimica et Biophysica Acta (BBA)-Biomembranes* **1838**:1809-1816 https://doi.org/10.1016/j.bbamem.2014.03.003 | PubMed

Zhang Y, Skolnick J (2005) TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Research* **33**:2302-2309 https://doi.org/10.1093/nar/gki524 | PubMed

Peer reviews

Reviewer #1 (Public review):

Summary:

This paper tries to address an important outstanding issue, which is the evolutionary origin of the SLC25 family of mitochondrial carrier proteins, which are common to all eukaryotic life, with few exceptions. The authors have carried out phylogenetic analyses and DALI searches of AlphaFold databases of bacterial and archaeal membrane proteins. They identify two bacterial proteins, CysZ and YhiY, and they propose that they are progenitors of SLC25 family members. Whilst the paper addresses an interesting topic, the conclusions are not supported by the data and are not presented in an unbiased manner, as they highlight only features that provide some tentative support for the hypothesis. They do not address the large number sequence and structural properties that refute the hypothesis, such as the asymmetric vs three-fold pseudo-symmetric features, hexamer vs monomer, and the complete lack of any conserved motifs with similar functions. Any resemblances between CysZ/YhiY and mitochondrial carriers thus seem to be superficial and could well be coincidental, as they represent generic properties of membrane proteins rather than specific ones, indicative of an evolutionary relationship.

Strengths:

This paper explores the evolutionary origins of the SLC25 family of mitochondrial carrier proteins, which are found across nearly all eukaryotic organisms. They were likely to be present in the last common ancestor of all eukaryotes, around two billion years ago. The question is whether they are of bacterial, archeal or eukaryotic origin. The authors propose that two bacterial proteins, CysZ and YihY, may represent ancestral forms of these carriers, based on structural comparisons of models, a sequence motif, and phylogenetic analyses. While the research addresses an important and longstanding question, the presented evidence does not convincingly support their hypothesis.

Weaknesses:

A central concern is the reliance on structural similarity searches using predicted protein models, since these models are often built using known protein structures as templates, and thus these searches may produce misleading matches. The reported similarities between CysZ, YihY, and mitochondrial carriers are weak and fall within ranges expected for unrelated membrane proteins, which commonly share general structural features, such as helical bundles. Quantitative measures of similarity are low and do not support a shared evolutionary origin. The case for YhiY is extremely poor as neither structure nor sequence features support the claim. Importantly, the opening of the YihY is towards the membrane rather than the water phase, as is the case for carriers, indicating that it has a very different structure and function. The case for CysZ is somewhat better, as it is a helical bundle with two short helices somewhat resembling the matrix helices of mitochondrial carriers, and a short sequence PXDXXK that is part of one of the known sequence motifs of mitochondrial carriers, but this is where the similarities end.

Mitochondrial carriers have a distinctive threefold pseudo-symmetrical structure and a highly complex transport mechanism involving six structural elements. This paper's hypothesis does not explain how such a high level of threefold pseudo-symmetry could have evolved from entirely asymmetric proteins. To complicate matters further, CysZ is not functional as a monomer but forms a functional hexamer, which also explains why it has two half helices rather than two transmembrane helices. Thus, the hypothesis is that CysZ, which is an asymmetric protomer of a functional hexamer, has evolved into a three-fold pseudo-

symmetric protein, which is functional as a monomer. A more convincing explanation is that the threefold pseudo-symmetrical structure arose from gene triplication and fusions, with later mutations introducing asymmetry to support diverse substrate binding. In support of this notion, mitochondrial carriers transporting large molecules, such as ATP, show more asymmetry, whereas those for small molecules remain nearly symmetrical. In general, the vast majority of transport proteins arose from gene duplications and fusions of the domains.

Although mitochondrial carriers have a similar sequence motif as found in CysZ (PXDXXK), their roles are very different. In mitochondrial carriers, this motif is located roughly in the middle of transmembrane helices H1, H3, and H5, where proline creates a pronounced kink, bringing the charged residues inward to form a salt-bridge network in the central water-filled cavity. The formation and disruption of this network is essential for the transport mechanism when switching between inward- and outward-open states. In CysZ, the motif is found at the end of a helix and in the following loop at the end of the transporter, with residues pointing outward toward the water phase. These residues are typical of membrane-water interface regions, where proline acts as a helix breaker and charged residues interact with the water phase. Thus, this motif in CysZ does not match the position or function seen in mitochondrial carriers, and its presence is likely to be coincidental, because these residues often occur in the water-membrane region. Importantly, none of the other important conserved three-fold symmetrical motifs of mitochondrial carriers is found in these bacterial proteins, such as the cytoplasmic network [YF][DE]xx[RK], cardiolipin binding sites, ER-links, and sequences of small amino acids, which are critical for its dynamic mechanism.

The phylogenetic relationship is also overstated, as there is no sequence similarity between these proteins other than that occurring because of similar biophysical properties, such as transmembrane helices. The authors suggest that a specific mitochondrial carrier represents the ancestral member of the family, but this conclusion appears to be inferred rather than rigorously demonstrated. Key aspects, such as tree rooting and taxon sampling, are not sufficiently addressed, weakening confidence in the evolutionary claims. Further, the selection of only a few bacterial and archaeal proteomes for analysis limits the study's scope. Broader searches would be necessary to support claims about conservation and ancestry. Independent sequence searches indicate that CysZ and YihY are not widely conserved in the bacterial groups most closely related to mitochondria, undermining the argument that they are plausible ancestors.

Overall, the presented similarities are superficial and can be explained by general features of membrane proteins rather than by specific adaptations to function. The hypothesis that CysZ and YihY are evolutionary precursors of mitochondrial carriers is not supported by the presented data.

https://doi.org/10.7554/eLife.111475.1.sa3

Reviewer #2 (Public review):

Summary:

Here, the authors performed a phylogenetic analysis of mitochondrial ATP/ADP carrier (AAC) proteins. They also performed a structure-based screen for remote homologs, seeking to reveal their evolutionary origins. The authors claim that AACs are found at the root of their family tree, and through a structure-based homolog search protocol, identify putative prokaryotic homologs.

The proposed evolutionary history of AACs is bold and complicated, but the phylogenetic methodology and the way in which the tree is interpreted are incomplete and unconvincing. Further, the structure-based search strategy uses very relaxed cutoffs for fold similarity, which may be fine, but it does not clearly justify this decision. This is potentially very

problematic, as I did not find the quantitative or qualitative assessments of fold similarity particularly compelling.

In summary, the authors have presented a bold and extremely interesting hypothesis for the evolution of these proteins, but there is insufficient support for their claims.

Strengths:

(1) The authors are presenting a very interesting hypothesis about the birth of these proteins, including that they may have undergone a radical rearrangement in their sequence at some point in evolution.

(2) The paper makes use of appropriate tools for structure-based homolog identification.

(3) Identification of a conserved sequence motif in these twilight zone proteins would be a rare and interesting occurrence, and could be consistent with their proposed homology.

Weaknesses:

(1) The phylogenetic analysis and its interpretations are incomplete. The authors regularly refer to the root of the tree, and its placement is given central importance. However, the methodology by which they selected the root is unexplained. This is notable, as the proposed root is curious and quite confusing. It implies that (at least) yeast and Paramecium AACs are independently paraphyletic. While certainly not impossible, this evokes quite a complicated evolutionary history. The taxonomy of this gene family, when rooted this way, does not seem to echo the phylogeny of species, suggesting an extremely complex history of duplication/loss and horizontal gene transfer, none of which the authors discuss in detail. Perhaps more clearly and specifically: I'm very surprised by the branching order at the root, where there are three independent branches of fungal proteins, followed by the excavate proteins in a monophyletic clade, followed by several independent branches of the Paramecium proteins. I very much expect incomplete lineage sorting at this evolutionary depth, but this seems extreme to the point that I question if it is accurately placed. More directly: this very much looks like an unrooted tree, presented radially.

(2) The Bayesian and ML trees seem quite incongruent, but this is not discussed. In fact, the text states that they "exhibit a similar tree topology." This is admittedly very difficult to assess without very carefully going over the tree, branch by branch, but there are nevertheless differences, the most obvious being paraphyly vs monophyly of taxon-specific AAC clades. Do the authors have any comments on this, and can they show some sort of consensus tree? How does this affect their interpretation?

(3) Presenting branch support as similarly-sized points makes it nearly impossible to actually judge the strength of support.

(4) The use of structure for remote homology detection is becoming increasingly popular, and in my opinion, is very powerful. But it is still much too early to be taken for granted. The methodology must be justified. Most importantly, the authors have not clearly described why they chose these quantitative cutoffs (I'm mostly thinking of the Dali Z-score cutoff, which here seems very low for a transmembrane protein of this size, as the Z-score is very dependent on alignment length). The authors reference categories defined by tool authors, but why a Z-score of 3, specifically? The same goes for TM scores. There are not yet any defined best practices, to my knowledge, so the authors should independently validate/justify their approach in some way and/or cite and discuss relevant literature (there have been a growing number of these screens using similar approaches in recent years).

(5) The proposed homologs have very little quantitative structural similarity to the query structure, or to each other, as shown in Figure 3 (and hence my concerns about the

methodology). Also, I did not find the structural alignments in the supplement or Figure 4 to be qualitatively compelling. They simply appear too different, and I cannot discard this qualitative assessment because the quantitative similarities are likewise very weak. It's not clear to me if this is because the folds are in fact different, or if my view of them is a presentation issue (perhaps it could be improved by visualizing more angles, or more carefully cartooning the similarities and differences).

(6) The authors point out that the alpha-helices are ordered differently in YihY and CysZ, and that their membrane orientation is flipped. Taken at face value, I would view this as evidence against homology. This could perhaps be more reasonably explained as convergent global fold similarity resulting from different underlying structures. However, the authors imply that this may be the result of the transposition of the sequences encoding these alpha helices, yet there is no convincing description or argument concerning when and how this could have occurred. I think this would be a deeply interesting phenomenon, but there is insufficient evidence and discussion to seriously consider whether or not it is homology or convergence.

(7) Following up on comment #5, the authors did perform a very interesting in silico experiment by transposing sequences to reorder the helices. They then note that structural similarity improved. This is very, very interesting, but without other evidence of homology between the transposed alpha helices, I do not think this disproves alternative hypotheses. Does any such evidence exist?

(8) The authors show in Figure 5E-F that sequence transposition flips the membrane orientation, such that YihY and CysZ have extracellular termini (which you would expect from homologs, I suppose). But it is just cartooned and not discussed. Is this computationally or experimentally supported?

(9) The putative presence of a conserved motif would be a very compelling piece of evidence consistent with homology. However, it is not clear to me in the text which proteins actually have the repeats - is it truly just CysZ? What does this mean for YihY? Further, what specifically is being proposed to be homologous? Is SLC25 repeat 2 proposed to be homologous to CysZ repeat 2 (and the same for 3 to 3)? If so, this would seem to have implications for the transposition hypothesis. The helix nomenclature (e.g., H1-6) suggests homology across the proteins (i.e, H1 is homologous to H1); however, wouldn't the presence of these conserved domains instead, for example, suggest homology between SLC H3 and CysZ H2? The authors' conclusions are not clear, and it is difficult to interpret what the implications are for assessing homology.

(10) The sequence retrieval methods are incomplete, so it is impossible to reproduce the searches or to judge their accuracy and scope. What were the E-value cutoffs and other settings used in the searches?

(11) The phylogenetic methods are incomplete. What substitution models were used, and how were they chosen? What branch support method was used? What were the stop conditions of the Bayesian analysis (e.g. did the authors monitor for convergence, and how)? How much of the Bayesian analysis was considered burn-in, if any? And echoing points 1 & 2 above, how were these phylogenies rooted?

(12) Throughout, there is a distinct lack of careful, evolutionarily informative language.

(i) In reference to the phylogeny, the authors frequently refer to "grouping," but it's not entirely clear what this means. Referring to clades and their branching order would be more informative.

(ii) The authors refer to the excavate branch as the "most ancient." Whether or not excavates most closely resemble LECA is somewhat irrelevant, because the branch itself is not the most ancient - it is equally as ancient as its sister branch, which may be all other eukaryotes.

(iii) Likewise, the authors refer to bacterial proteins as "the evolutionary ancestor of mitochondrial AACs," and state that "AAC emerged from the conserved sulfat transporter CysZ." But extant bacteria are not the ancestors of mitochondria - nor are extant proteins descended from other extant proteins. They are, perhaps more accurately, cousins.

(iv) The authors refer to AACs as "evolutionarily founder member of the SLC25 carrier family," but I'm not sure that has a clear evolutionary meaning, unless the authors mean to say that the common ancestor was more AAC-like than anything-else-like. Even if the rooting is accurate, a basal branch does not necessarily reflect the ancestral state.

https://doi.org/10.7554/eLife.111475.1.sa2

Reviewer #3 (Public review):

Summary:

The most important weakness is that the authors have avoided the direct structural comparison of experimentally determined x-ray structures of AAC and CysZ. Instead, the comparisons are made through predicted membrane topologies and predicted structural models of protein homologs, which give rise to misleading results. Direct comparison of the X-ray structures of the ADP/ATP carrier and CysZ clearly shows that these proteins have very different folds. Therefore, flaws in the methods produce results that lead to the wrong conclusions, and the authors have not achieved their aims.

Weaknesses:

(1) Figure 2. There is something very strange about how the tree is drawn, given that S. cerevisiae AAC1, AAC2 and AAC3 share about 76-83% sequence identity but appear to be very diversified in the tree. The phylogenetic trees are only based on the sequences of three species. The authors should explain in much more detail how they made the phylogenetic trees to support their statement that all mitochondrial carriers have come from an ancient AAC.

(2) There are at least three and seven X-ray structures of CysZ (with about 43% sequence identity to the E. coli homolog) and AAC, respectively, deposited in the Protein Data Bank. Therefore, there is no need for the approach using predicted structures as described in the manuscript. It is clear from direct comparison of the CysZ and AAC structures that they have very different folds, i.e. lengths of the transmembrane helices, their orientation and packing. CysZ has been suggested to form dimers or trimers of dimers (eLife 2018;7:e27829), with each protomer formed by two long transmembrane helices and four short helices that do not cross the membrane totally. Thus, CysZ has a different membrane topology and oligomeric state than AAC (monomer with six transmembrane helices). CysZ is therefore rightfully classified in a separate 3D domain fold from mitochondrial carriers in various protein family and domain databases.

(3) In the 3D structures of CysZ, the conserved QYXDYPXDNHK motif is involved in a network of hydrogen bonds and salt bridges thought to hold the helical bundle together (eLife 2018;7:e27829). This motif is similar to PX[DE]XX[KR], a part of the signature motif, typical of mitochondrial carriers, which is repeated three times in the sequences and forms a three-fold pseudo-symmetrical salt bridge network of the so-called matrix gate that opens and closes during the transport cycle. Therefore, although this single motif in CysZ is similar to those of mitochondrial carriers, it is not found in a similar structural context to those in AAC structures.

(4) It appears odd that the sulfate transporter CysZ should be more similar to nucleotide-transporting AAC than any of the other mitochondrial carriers, of which some transport

sulfate.

(5) The alphafold model of YihY is not very similar to either the crystal structures of CysZ or AAC.

(6) The authors are relying too much on the TM-score results. The values of 0.5-0.6 between AAC and CysZ or YihY probably reflect that they contain six main helices. However, as noted in point 2, they have very different folds.

https://doi.org/10.7554/eLife.111475.1.sa1

Author response:

Thank you for your decision letter with the public review and the recommendations. While we are delighted that the referees feel the work is addressing an outstanding and important issue, they have raised concerns regarding the strength of the support. We will address all the concerns in full in a revised manuscript in the due course. Please find below a couple of general points regarding the referees' concerns and a proposal as to how we plan to address them.

(1) The idea of the manuscript is to present a plausible solution for a long-standing question in the field of mitochondrial biology and evolution. The fact that the identified solution to the origin of AAC transporters is a remote structural homolog (as you will see in our later detailed response that it is better than any other sequence/structure available till date) is to be expected. If the actual similarities were any better than what we have identified (with a special case of circular permutation), they could have been identified by other simpler structural homology search methodologies.

(2) A recurrent and strong disagreement of the reviewers on the findings presented in this manuscript is rooted on the fact that the structural and sequence relatedness between AAC and CysZ detected in this work are so weak that they can be co-incidental and not an actual evolutionary link. Based on the above, we now searched carefully in all available structural databases such as SCOP, CATH, ECOD etc. whether the above fold link has been noted by others independently. We notice that in the ECOD (Evolutionary Classification of Protein Domains) database only AAC and CysZ are grouped together under a single Possible homology group (X) called 'Mitochondrial ADP/ATP carrier-like'. The ECOD database contains hierarchical classification of protein domains organized according to their evolutionary relationships and the server is maintained by Prof. Nick Grishin at The University of Texas Southwestern Medical Center.

Link to ECOD database: *http://prodata.swmed.edu/ecod/index_af2_pdb.php*

Reference: Cheng H, Schaeffer RD, Liao Y, Kinch LN, Pei J, et al. (2014) ECOD: An Evolutionary Classification of Protein Domains. PLOS Computational Biology 10(12): e1003926. *https://doi.org/10.1371/journal.pcbi.1003926*

Therefore, our study and the independent findings of the ECOD database team together offers greater confidence on the proposed remote evolutionary relationship between AAC and CysZ, and that the structural and sequence similarity we report in the manuscript are not a mere co-incidence. We will also incorporate the details of possible evolutionary relationship between AAC and CysZ identified in the ECOD database in the revised version of manuscript.

(3) One point we would like to stress is that considering all the similarities identified, it very unlikely falls into the class of 'convergent evolution'. We will make this point explicit in the revised version.

(4) Lastly, while we totally agree that the similarities are in the twilight zone, considering the importance of the problem, we feel that our work would induce researchers from the field of protein design to attempt possible interconversion of the two distantly related transporters thus providing an experimental rationale for the evolution of these transporters.

https://doi.org/10.7554/eLife.111475.1.sa0