# Implications of divergence of methionine adenosyltransferase in archaea

Bhanu Pratap Singh Chouhan, Madhuri Gade, Desirae Martinez, Saacnicteh Toledo-Patino and Paola Laurino (iD)

Protein Engineering and Evolution Unit, Okinawa Institute of Science and Technology Graduate University, Onna, Japan

Methionine adenosyltransferase (MAT) catalyzes the biosynthesis of S-adenosyl methionine from L-methionine and ATP. MAT enzymes are ancient, believed to share a common ancestor, and are highly conserved in all three domains of life. However, the sequences of archaeal MATs show considerable divergence compared with their bacterial and eukaryotic counterparts. Furthermore, the structural significance and functional significance of this sequence divergence are not well understood. In the present study, we employed structural analysis and ancestral sequence reconstruction to investigate archaeal MAT divergence. We observed that the dimer interface containing the active site (which is usually well conserved) diverged considerably between the bacterial/eukaryotic MATs and archaeal MAT. A detailed investigation of the available structures supports the sequence analysis outcome: The protein domains and subdomains of bacterial and eukaryotic MAT are more similar than those of archaea. Finally, we resurrected archaeal MAT ancestors. Interestingly, archaeal MAT ancestors show substrate specificity, which is lost during evolution. This observation supports the hypothesis of a common MAT ancestor for the three domains of life. In conclusion, we have demonstrated that archaeal MAT is an ideal system for studying an enzyme family that evolved differently in one domain compared with others while maintaining the same catalytic activity.

Common descent is one of the fundamental aspects of Darwinian evolution. This theory emphasizes that modern-day species share a common ancestry [1]. The same principle applies to enzymes: modern enzyme superfamilies across the three domains of life evolved from a set of enzymes that were already present in a last universal common ancestor (LUCA) that can be dated to over 3.5 billion years ago [2,3]. Many efforts have been made to infer the minimal set of LUCA enzymes [3,4]. The main hypothesis is that the evolutionary trajectory of enzymes (i.e., gene trees) is closely associated with and influenced by the evolution of their respective host organisms (i.e., species trees) [5]. However, in reality, gene evolution is much more complicated and disagreement between species trees and gene trees (noncongruence) can occur because of a wide range of factors such as gene duplication, lateral gene transfer [6,7], and hybridization [8]. Within this framework, a few studies have reported an unusual distribution of some enzymes among the three domains of life, whereby the archaeal enzyme shows a divergence from the more closely related eukaryotic

and bacterial orthologs [9–12]. One example is the methionine adenosyltransferase (MAT) enzyme.

The unusual sequence similarity of MAT enzymes among the three domains is striking; the archaeal enzyme is almost equidistant from the bacterial and eukaryotic enzymes, with ˜ 20% sequence identity, while the bacterial and eukaryotic enzymes share greater than ˜ 60% identity [13].

Methionine adenosyltransferase is a ubiquitous enzyme, present across all three domains of life, that catalyzes the biosynthesis of *S*-adenosyl methionine (SAM) from L-methionine and ATP. SAM is a primary methyl donor in nature, and it is also considered to be an ancient cofactor [14]. SAM-dependent enzymes are likely included in the LUCA [15]. Most of these ancient SAM-dependent enzymes are radical methyltransferases [15–18]. Nowadays, SAM is involved in four important functions (Scheme S1) [19]: (a) methylation of fundamental biomolecules (proteins, DNA, RNA) [20,21]; (b) polyamine synthesis [22]; (c) glutathione biosynthesis reactions [23]; and (d) radical chemical reactions [24]. Moreover, methylated molecules or macromolecules by SAM play an important role in cellular processes such as epigenetics, signal transduction, protein function, and genetics [25]. The importance of these biological processes makes the SAM-synthesizing enzyme, namely MAT, an important and essential enzyme for cell growth and survival [26].

The structure of the MAT enzyme is unique and classified by the SCOP database in α/β class [27]. Each monomer of MAT is composed of a βαββαβ secondary structure [28]. No sequence conservation is observed among the repeats. Like many other enzymes in nature, MAT is active as an oligomer (either as a homo-oligomer or as a hetero-oligomer; Fig. 1) [29]. The oligomeric state provides a clear advantage over the monomeric state in terms of fitness, stability, and functions [30,31]. In dimeric MAT, helixes are exposed to the surface and strands form a hydrophobic interface [32]. The MAT homodimer consists of monomeric subunits (α-subunits) paired together along a large flat hydrophobic interface (henceforth referred to as large interface; Fig. 1) wherein the active sites are enclosed [13]. A homotetramer exists as a dimer of homodimers; therefore, it has been suggested that the preferred functional quaternary structure is an oligomer (possibly an obligate dimer) [33].

It has been suggested that all extant MAT enzymes share a common ancestry [33] despite expressing a highly divergent form in archaea, as mentioned above. This further indicates that the adaptation of enzymes during evolution can lead to the accumulation of mutations at the surface or interface regions, especially near the active sites. This phenomenon can allow for the divergence and acceptance of different substrates or functions [34–39]. Despite this contrast, it has also been established that archaeal MAT can still perform the same catalytic reaction as their orthologous counterparts (from bacteria and eukarya); therefore, the structural and functional implications of this divergence are not well understood. These observations prompted us to take a closer look at the evolutionary trends adapted by the active site residues and the large interface region of MAT. Therefore, we probed the putative evolutionary trajectory of archaeal MAT more systematically by conducting investigations at various levels, including sequence studies and structural comparisons, assessment of physiochemical properties, and ancestral sequence reconstruction (ASR). We further elucidated this potential evolutionary trajectory by resurrecting the ancestral archaeal MAT sequences and characterizing their biochemical properties. In conclusion, we showed that the archaeal MAT can display high degrees of divergence at the sequence/ structural level and yet perform the same catalytic reactions as their orthologous counterparts.

## Results and Discussion

### Probing the MAT sequence space

In the present study, we conducted extensive searches to systematically probe the MAT sequence space across the three domains. The sequences were collected by combining the outcomes of different databases (nr db, eggnog, orthodb, etc.). We built sequence similarity networks (SSNs) upon compiling the first input dataset (˜800 sequences), which allowed us to check the MAT sequence space distribution across the three domains (Fig. 2A). During this analysis, we considered that the archaeal domain has not been sequenced as extensively as the other two domains of life [40]. Because phylogenetic trees may not be an optimum tool to highlight the overall sequence distribution pattern, we utilized two approaches, EFI and CLANS [41], to visualize the sequence distribution pattern (Fig. S1). Despite the variation in the percentage sequence identity clustering ranging from 100% to 40% ID, the overall topology remained consistent with two major clusters representing: (a) 'bacteria + eukarya' and (b) archaea. The SSN outcomes reflected the sequence identities as observed in nature, with archaea found to be nearly equidistant from eukarya and bacteria, at ˜ 20% sequence identity. The SSN method provided an overall view of the topology for the dataset of interest and underlined the uniqueness of the archaeal domain.
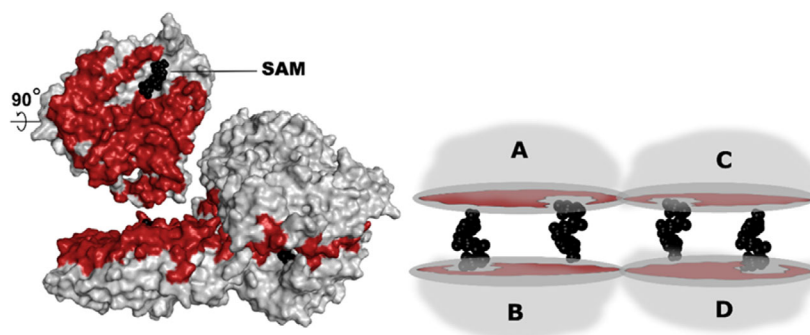
**Fig. 1.** MAT homotetramer configuration. Schematic representation of the MAT homotetramer (right) and surface/cartoon view (left)—the large interface region (between chains A and B) is highlighted in red, while SAM is highlighted in black. The large interface formed by chains A and B (also chains C and D), which have an area of ˜ 1900–2000 Å$^2$, consists of ˜ 70–100 residues. Representative structures have been utilized here, while the archaeal pfMAT structure (PDB ID: 6S83 from *Pyrococcus furiosus*) utilizes 83 residues to give rise to the large interface. These numbers are 90 residues in case of bacterial eMAT (*Escherichia coli*, PDB ID: 1RG9) and 77 in the case of eukarya hMAT2A (*Homo sapiens*, PDB ID: 4NDN).
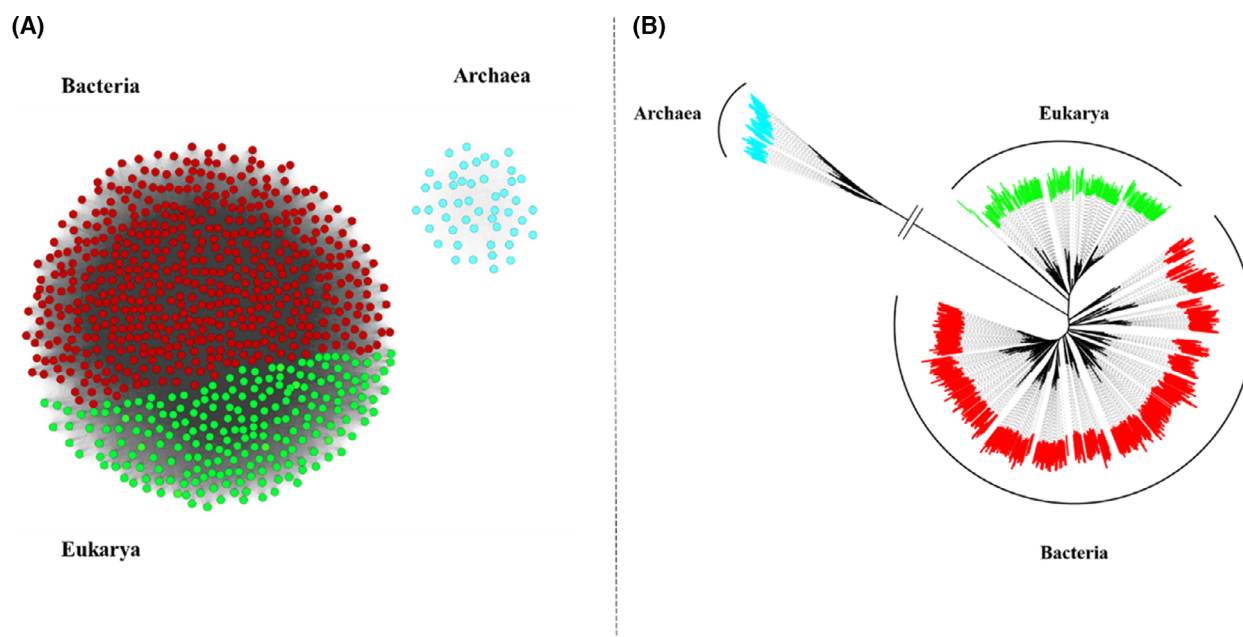


**Fig. 2.** Schematic representation of observed sequence space in MAT. (A) SSN of MATs from the three domains—the sequences from eukarya (green) and bacteria (red) form a distinct cluster, while the sequences from archaea (cyan) form a separate cluster. This SSN was created with the EFI enzyme similarity server and visualized using the Cytoscape program. Clustering pattern within the network at different levels of sequence similarity (similarity %) can be observed in the supplementary information (Fig. S1). (B) Unrooted ML tree for MATs from the three domains of life. This unrooted tree was built by applying the ML method with IQ-TREE, followed by visualization using FigTree viewer program and iTOL server. The archaeal MATs branch off, thereby highlighting their divergence, an observation that was further substantiated through SSN. The archaeal MATs are represented in cyan, eukarya in green, and bacteria in red.

## Inferring the ancestral MAT sequences

We reconstructed the phylogenetic relationships by utilizing MAT sequences from the three domains of life (Fig. 2B). In the case of archaea, two major phyla [Crenarchaea and Euryarchaea (EuryAnc)] were tested for the overall tree topology based on the MAT sequences (gene tree), as well as corresponding 16s rRNA sequences (species tree) from the SILVA database (not shown here). Herein, the two topologies clearly indicated that the tree bifurcates into two distinct clades for Crenarchaea and EuryAnc; that is, the

major phyla are well classified into their respective monophyletic groups. Subsequently, we prepared another set of trees by utilizing bacterial MAT sequences as an outgroup to extract the putative common archaeal ancestor ArchAnc (Fig. 3, violet star). We also extracted the ancestral sequences of Crenarchaea (CrenAnc) and EuryAnc MAT at the designated nodes highlighted by using a red and blue star, respectively (Fig. 3).

The sequence similarities of the ancestral sequences (including interface residues and active site residue comparisons) are detailed in the other sections.

For bacterial and eukarya MAT sequences, we prepared a separate phylogenetic tree (based only on MAT sequences). MAT sequences from these two domains constituted two distinct clusters (supported by principal component analysis clustering—data not shown), as observed in the tree topology as well (Fig. 4), which was then utilized to infer the common MAT ancestors for eukarya (EukaAnc) and bacteria (BactAnc).

## Extant MAT sequences vs ancestors: active sites and evolutionary rates

We probed the evolutionary trend(s) for amino acid sites in MAT enzymes across the three domains of life. These trends were visualized by mapping the interface sites, as well as active sites on a representative X-ray structure (Fig. 5, Fig. S2). Additionally, we probed the level of conservation for active site residues in extant and ancestral sequences.

### Active site residues

In the case of archaea, we observed that the MAT active site residues were highly conserved among the two major branches, that is, Crenarchaea and EuryAnc (residue numbers correspond to *Pyrococcus furiosus* (pfMAT) structure: Protein Data Bank (PDB): 6S83, [33]). These include the following residues involved in the interaction with SAM from chain A: H59, N61, D141, L142, and D280 and chain B: H30, D168, and Y268. Furthermore, these active site residues were completely conserved in the putative archaeal ancestor. A similar trend was observed in the cases of both eukarya and bacteria. This provides us with a key insight that despite observing differences in the active site residues, MAT enzymes from the three domains can still essentially retain the same catalytic reaction (Fig. 5).

### Evolutionary rates

To determine whether part of the protein evolved at a different rate compared with the rest of the structure,

we determined the evolutionary rate distribution of all residues in the MAT structure (Fig. S2). Interestingly, most sites located along the large interface tend to evolve slowly (highlighted in blue), while the opposite trend is observed as we move further away from the large interface. For instance, some of the residues located along the surface α-helices and loop regions tend to experience higher evolutionary rates. This pattern can be seen in MAT enzymes from all three domains of life, thereby indicating that conservation at a large interface is probably critical for enzyme function despite the domain that the enzyme belongs to.

## Large interface: charge and hydrophobicity distribution

In case of MAT enzymes, the constituent monomers pair up in an inverted configuration by exposing the α-helices toward the surface, while the β-strands interact to form a hydrophobic large interface that harbors the active site residues, thereby making this homodimer the obligate functional unit [32]. Herein, we noticed that divergence is not only limited to active residues, or the residues located close to the substrate, but most of the interdimer interface (large interface) residues also diverge across the three domains of life. For instance, in cases of bacteria (eMAT, PDB: 1RG9 [42]) and eukarya (hMAT1A, PDB: 6SW5 [43]) 38 out of the 51 aligned residues were identical. In contrast, archaea (pfMAT, PDB: 6S83) shares only 12 identical residues with bacteria (eMAT) and eukarya (hMAT1A; Figs S3 and S4). This further prompted us to perform a systematic analysis of the large interface region across the three domains of life, and thus, we assessed the charge and hydrophobicity distribution across the large interface and compared it with the ancestral sequences to gain more insights into the evolutionary patterns involved. We studied 17 experimentally solved MAT structures from PDB to probe the physiochemical properties, by creating two datasets: (a) for identification of 51 structurally aligned positions along the large interface (which includes up to five non-interface residues as well) and (b) for identification of 24 structurally aligned 'interface-only' residue positions.

### Large interface region (extant)

Based on the 51 structurally aligned positions, even at the sequence level, the large interfaces of bacterial and eukarya MATs have a striking sequence identity of ˜ 50–70% (Figs S4 and S5). However, the archaeal interface sequences are nearly equidistant from the other two domains at ˜ 15–30% sequence identity, despite having a comparable size in terms of area (˜ 1660–2970
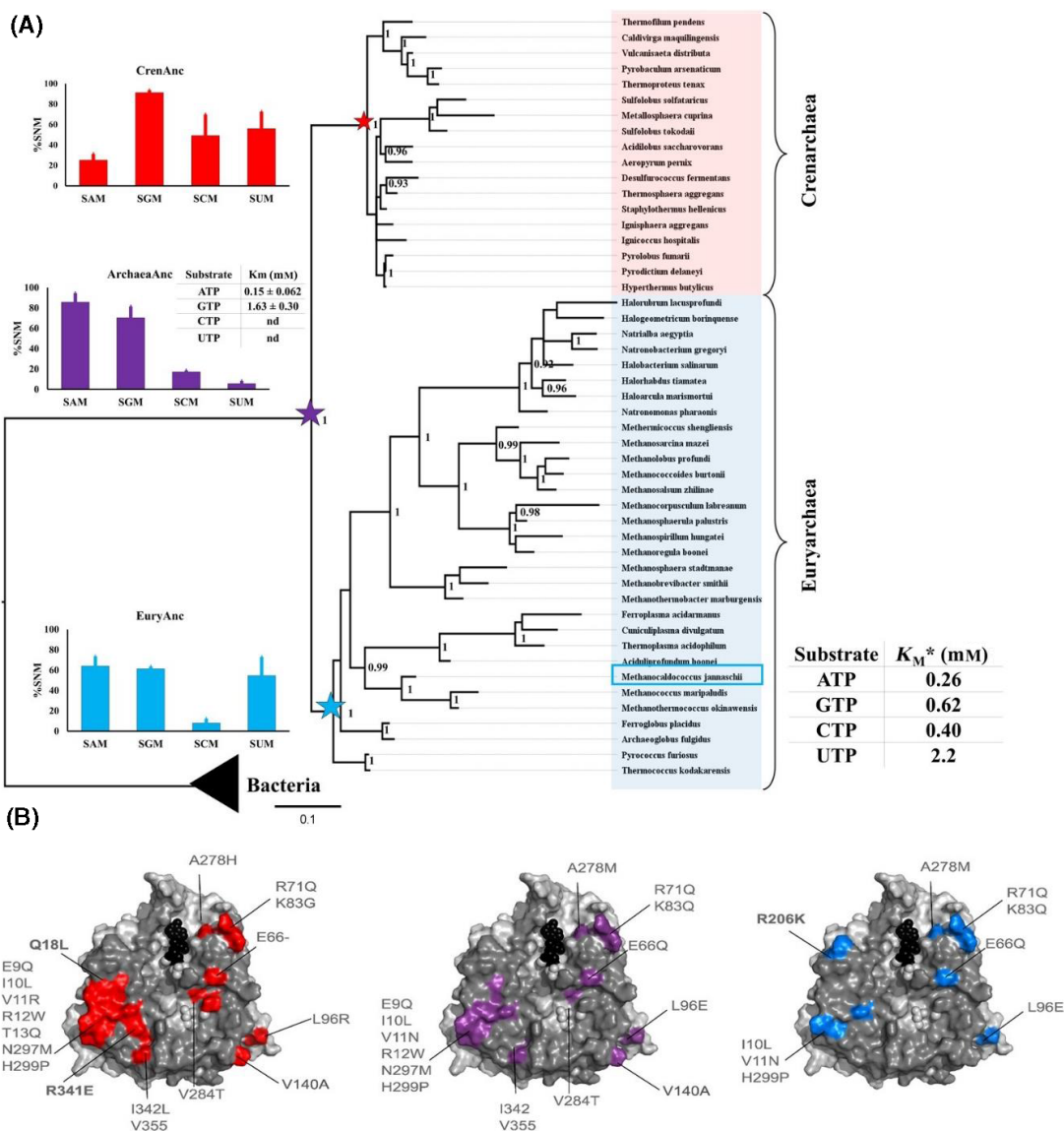
**Fig. 3.** Phylogenetic tree of archaeal MAT with bacteria domain as outgroup and activity assays of the inferred ancestors. (A) The maximum-likelihood phylogenetic tree topology of archaeal MAT with bifurcation into two major clades, that is, Crenarchaea and EuryAnc, and with bacteria as an outgroup. This topology is further supported by another phylogenetic tree based on the 16s rRNA sequences from the corresponding species. Here, *we highlight the activity of *MjMAT* as representative of domain archaea, with the data as reported previously [48]. Ancestral sequences were inferred for the ancestor of Crenarchaea MAT (CrenArc in red star), EuryAnc MAT (EuryArc in blue star), and the common ancestor for the archaeal MAT (ArcheaArc in violet star). Bayesian posterior probability values > 90% are shown. One time point assay for ancestral sequence activities was performed with corresponding enzyme (20 μM) using NTPs (5 mM) and methionine (10 mM) in HEPES (100 mM), KCl (50 mM), and $MgCl_2$(10 mM), pH 8, at 37 °C for hMAT1A and 55 °C for archaea, 1 h. For kinetics analysis, ArcheaArc MAT (0.5 μM), ATP, and GTP (0.1–2 mM) and methionine (10 mM) using the same above-mentioned buffer and conditions were used. The experiments were performed in duplicates. The error bars represent SD. Tables show the extrapolated $K_M$ for ArcheaArc. The production of SAM and SGM was analyzed by UPLC, and data were fitted to the Michaelis–Menten equation using GRAPHPAD software 7.02. (B) Mutations found by the ancestral reconstruction analysis are highlighted on the available structure of pfMAT (PDB: 6S83). The structures are represented as gray surface and show the large interface. CrenAnc MAT interface mutations are highlighted in red, EuryAnc in blue, and ArchaeaAnc in violet.
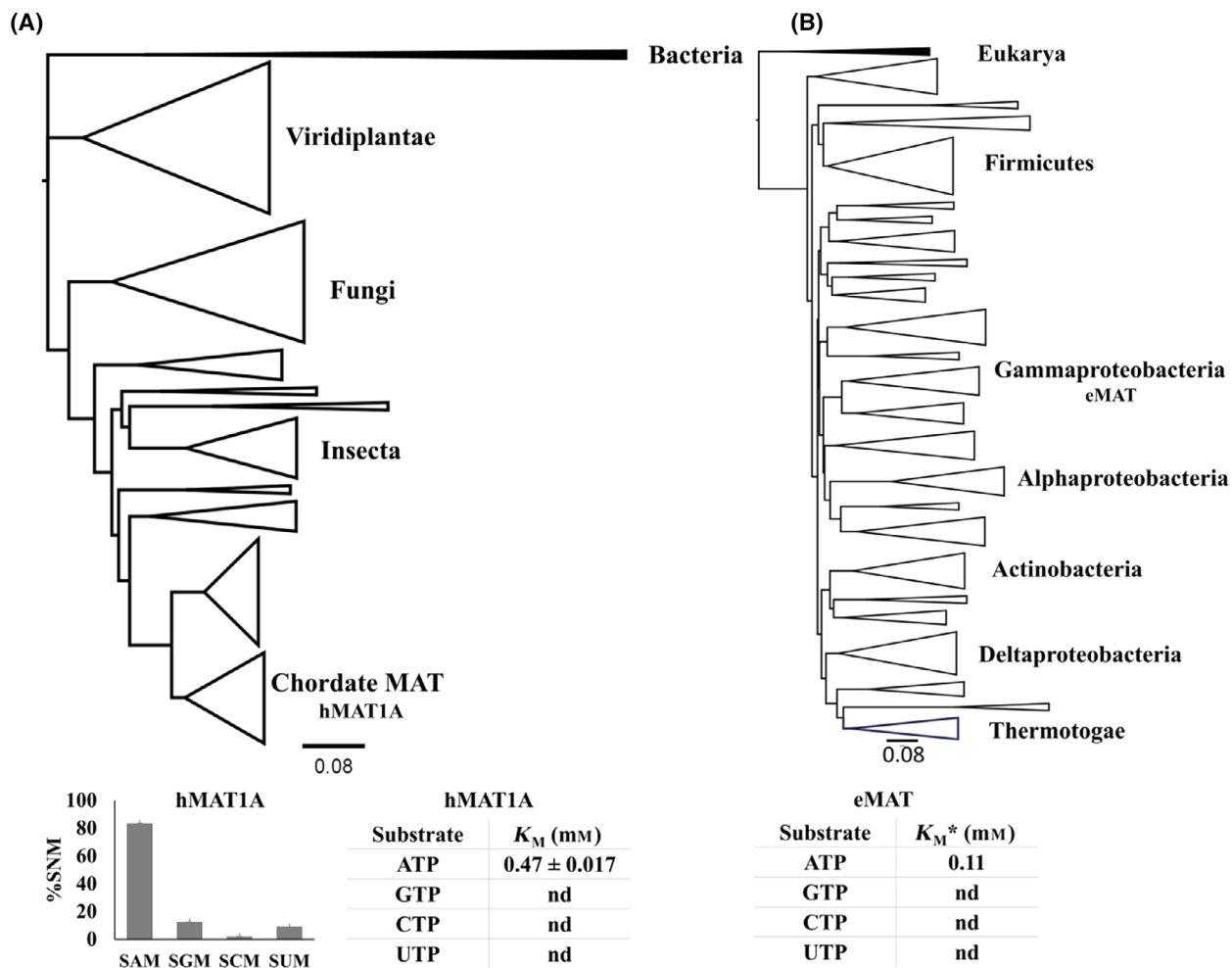
**(A)**



**(B)**



| hMAT1A | | |
| --- | --- | --- |
| Substrate | $K_M$ (mM) | |
| ATP | $0.47 \pm 0.017$ | |
| GTP | nd | |
| CTP | nd | |
| UTP | nd | |

| eMAT | | |
| --- | --- | --- |
| Substrate | $K_M$* (mM) | |
| ATP | 0.11 | |
| GTP | nd | |
| CTP | nd | |
| UTP | nd | |

**Fig. 4.** Phylogenetic trees based on the MAT sequences for eukarya and bacteria domains. (A) ML phylogenetic tree of MAT sequences from domain eukarya, with bacterial MAT sequences as an outgroup (highlighted in black). The major clades have been annotated according to their distribution in the tree topology. We reported the activity of hMAT1A as representative of domain eukarya. One time point assay for hMAT1A activities was performed using 20 μM of hMAT1A with NTPs (5 mM) and methionine (10 mM) in HEPES (100 mM), KCl (50 mM), and $MgCl_2$ (10 mM), pH 8, at 37 °C, 1 h. For kinetics analysis, hMAT1A (0.5 μM), ATP (0.1–2 mM), and methionine (10 mM) using the same above-mentioned buffer and conditions were used. The experiments were performed in duplicates. The error bars represent SD. Tables showing the extrapolated $K_M$ for hMAT1A. (B) ML phylogenetic tree of MAT sequences from bacteria domain, with eukarya MAT sequences utilized as an outgroup (highlighted in black). *We also highlight the $K_M$ of ATP from eMAT as a representative of bacteria domain, as reported previously [48]. These results suggest that both the enzymes, hMAT1A and eMAT, are not promiscuous, as they display specificity toward ATP. The trees were built using the ML method with the IQ-TREE program. Tree topologies were visualized using FigTree program.

$Å^2$). The hierarchical clustering plots (based on the pvclust package in R; Fig. 6 [44]) show that the sequence identities further dictate the physiochemical properties as well. For instance, a look at both the *P*-values provided by pvclust for comparison of the clusters, that is, approximately unbiased (AU) *P*-value and bp (bootstrap probability) value, reveals that the hydrophobicity and charge distribution are clustered together for bacterial and eukarya MAT structures, with the archaeal MAT structures clustering off separately. An additional analysis was also conducted for the second dataset, with an alignment of 24 'interface-only' residues (Fig. S6), and in this case too, we observed similar results, with the archaeal MAT structures clustering off separately.

**Ancestral vs extant (large interface)**

In this study, we compared the large interface residues extracted from the ancestral sequences against their extant counterparts. A comparison of the large interface
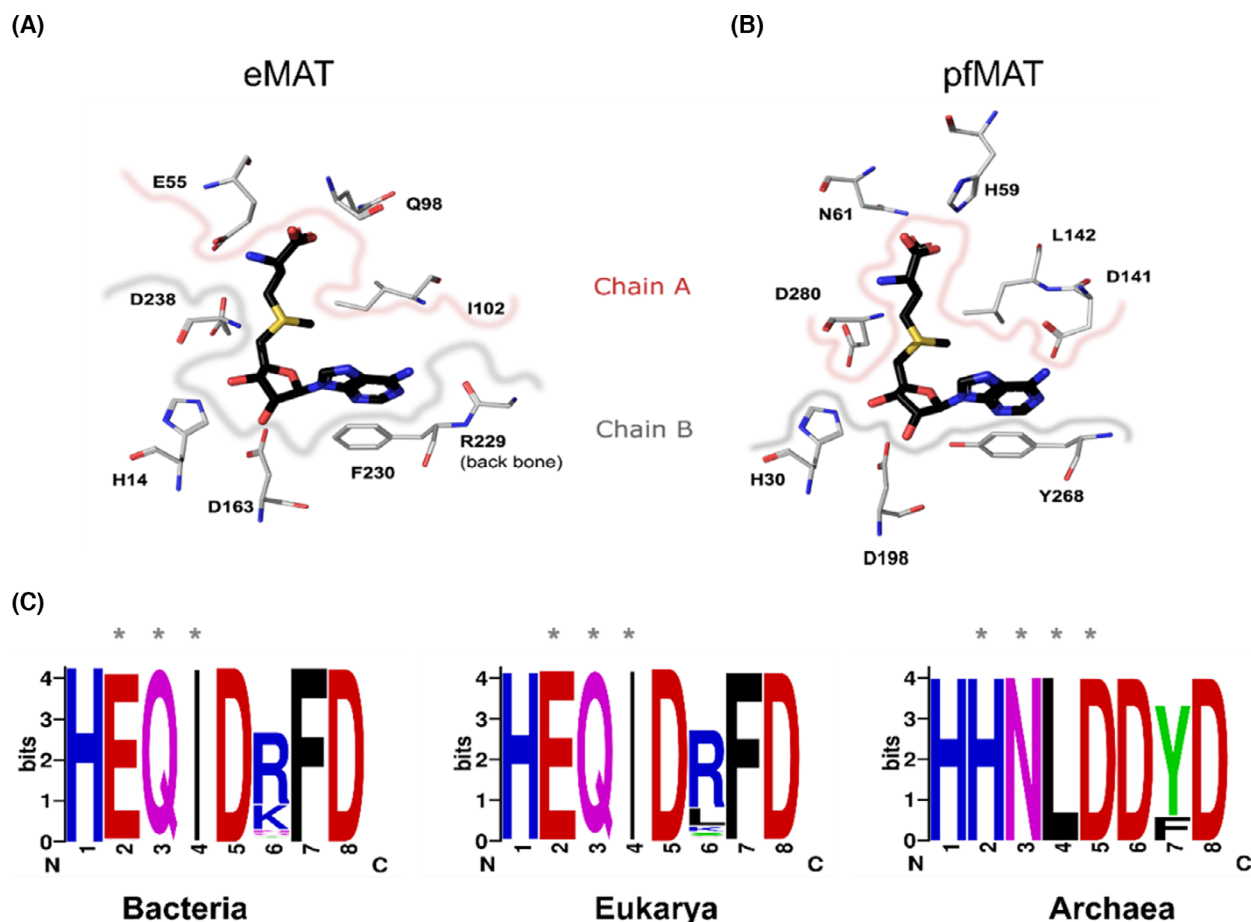
**(A)**

eMAT



**(B)**

pfMAT



**(C)**



**Fig. 5.** Comparison of the active site residues in the different domains. (A) Active site residues involved in the interaction with SAM. SAM is highlighted in black, while the active site residues from eMAT (PDB: 1RG9) have been highlighted in metal gray. The residues derived from chain A (E55, Q98, I102) have been demarcated with a pinkish border, and residues from chain B (H14, D163, R229, F230, D238) have been demarcated with a grayish border. (B) Equivalent active site residues from pfMAT [33] (PDB: 6S83) that interact with SAM have also been highlighted from chain A (H59, N61, D141, L142, D280) and chain B (H30, D198, Y268) using the same color codes. (C) The active site residues identified based on structural alignment, from all three domains of life, have been highlighted as WebLogos (https://weblogo. berkeley.edu/logo.cgi). Residues derived from the second monomeric subunit have been marked using '*'. WebLogo was generated using alignment of 526 sequences from bacteria (Appendix S1), 273 sequences from eukarya (Appendix S2), and 49 sequences from archaea (Appendix S3). Details of sequence search are in the Methods section.

residues from the bacterial MAT ancestor with the corresponding residues from eMAT highlighted a difference of 13 out of 51 residues (Fig. S4). Likewise, we also observed similar changes in eukarya: hMAT1A vs eukarya MAT ancestor, with seven residue changes (out of 51 residues); and archaea: pfMAT vs archaeal MAT ancestor, with six changes (out of 51 residues).

In conclusion, these observations provide crucial insight that sequence identities further influence the distribution of corresponding physiochemical properties. Therefore, the hydrophobicity and charge distribution, as described in the hierarchical clustering (based on pvclust), are important in understanding the evolutionary trends adapted by the large interface.

However, we observed that the active sites remained conserved to some extent across MAT enzymes.

## Structural comparison supports sequence divergence

Structural comparisons of MAT full domains from archaea, bacteria, and eukarya were performed to identify their common and unique structural features. We noticed that at a structural level, bacteria and eukarya show a higher degree of structural similarity compared with archaea (Fig. S7). In addition, an N-terminal extension in archaea segment-swaps into the C-terminal subdomain in contrast to bacteria and
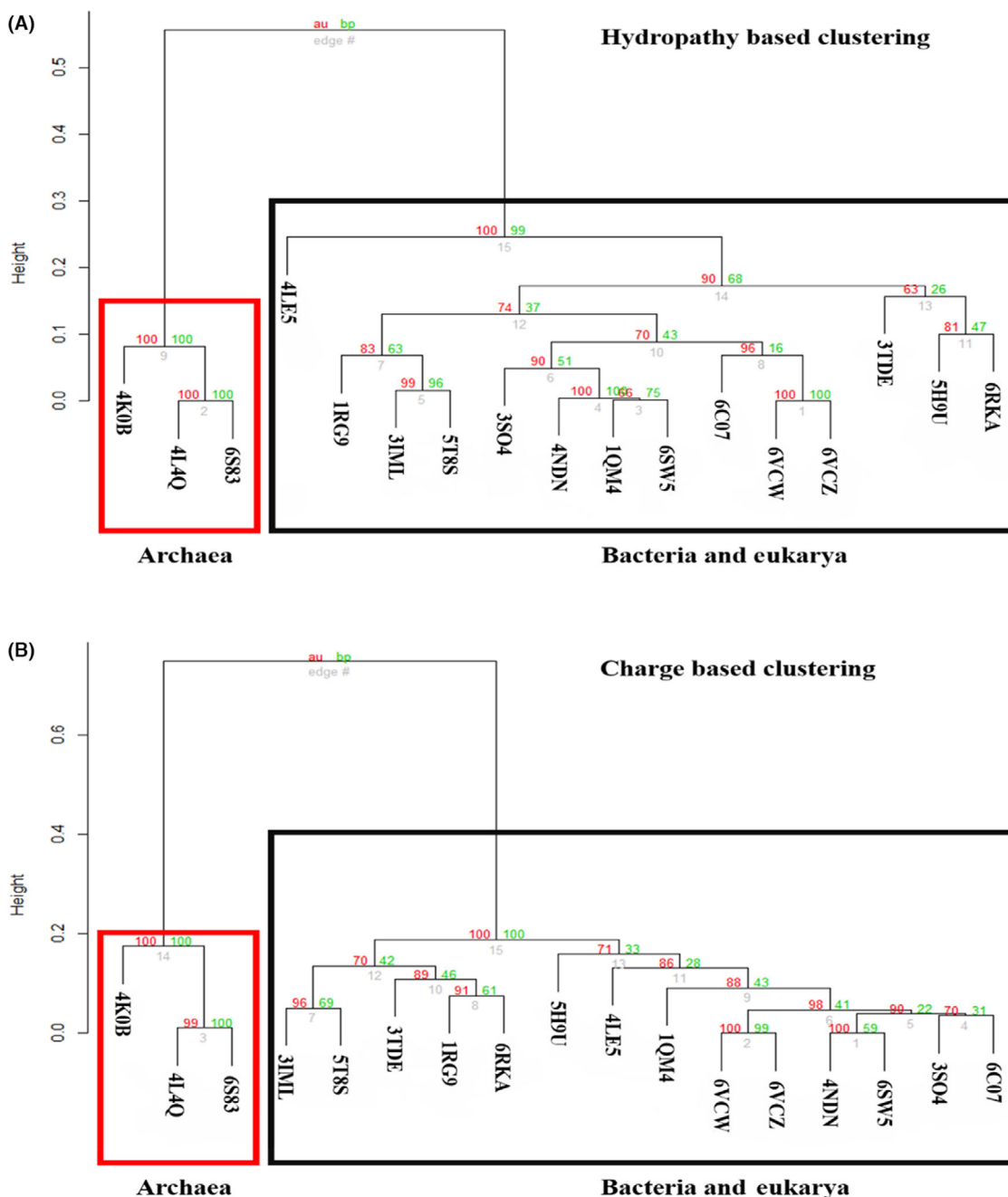
**Fig. 6.** Clustering analysis of hydropathy (A) and charge (B) of representative structures from the three domains of life. Clustering analysis of the physiochemical properties (hydropathy and charge) of 51 structurally aligned large interface residues from chain A of 17 MAT structures (PDB list in the supplementary information). Based on the 51 aligned positions, we calculated the per-site hydropathy using the Kyte–Doolittle scale from the ProtScale server. In addition, we calculated the per-site charges for the 17 MAT structures using the EMBOSS charge server. Subsequently, we conducted a clustering analysis using the pvclust package in R. This further provides a statistical score in terms of AU (AU scores depicted in red), *P*-value, and BP (bootstrap probability scores depicted in green) value, for comparison of the clusters, which reveals that the hydropathy (A) and charge distribution (B) clustered together for bacterial and eukarya MAT structures, with archaeal MAT structures clustered separately. In both cases, the archaeal cluster was provided with high support values, in terms of both the AU and BP parameters, and 100% support for a distinct archaeal cluster with respect to the two physiochemical properties.

eukarya, which lacks the N-terminal beta-strand. In return, bacterial and eukaryotic MATs display an extra C-terminal helix. These unique features may explain why archaeal MAT is not dissected into subdomains by either the SCOP or ECOD domain criteria. Topology connections according to PDBsum [45] are shown in Fig. S7.

We further extended our structural comparisons to the subdomain level. In this analysis, constituent domains (according to ECOD [46,47]) from bacterial and eukaryotic MATs were used as templates to manually dissect archaeal subdomains (Fig. S8). Constituent structural domains were then compared all-on-all within (Fig. S8), as well as across, archaea, bacteria, and eukarya (Fig. S9). Subdomain superposition showed that despite their identical topology, subdomains I to III within archaea, bacteria, and eukarya differ considerably in their secondary structure element orientation in space. In contrast, the subdomains across archaea, bacteria, and eukarya are much more similar. In this study, bacteria and eukarya displayed a higher degree of structural agreement within themselves than with archaea.

The structural comparison suggests that MAT emerged partially by duplications of subdomains within the same domain of life or those putatively shared by eukarya and bacteria, but not by archaeal enzymes. This observation supports the outcome of the analysis at the sequence level.

## Change in substrate specificity along the archaeal phylogenetic tree

It has been previously reported that the archaeal *Methanococcus jannaschii* (*MjMAT*) is promiscuous toward different nucleotides; that is, it is able to accept different nucleotide triphosphates (NTPs), while the bacterial ortholog eMAT is specific for the double ring adenine nucleotide base [13,48]. This further prompted us to study the specificity of the purine bases, ATP, and GTP within the archaeal phylogenetic tree and to determine whether this information could provide us with more clues toward the origin and evolution of MAT (Fig. 3). We expressed and tested the activity of the ancestral archaeal MAT sequences and compared them against the representative extant MAT sequences from the three domains of life (Figs 3 and 4). Data on the specificity of the representative bacterial and archaeal extant MAT have already been reported in literature [13,48]. We chose hMAT1A as a representative of eukarya. Since our purpose was to analyze the change in specificity within archaeal domain, we expressed and tested the activity of the ancestral MAT sequences for EuryAnc, CrenAnc, and the common

archaeal ancestor (ArchaeaAnc) (Figs S10–S13, Figs S14–S17). First, we performed one end-point assay with EuryAnc, CrenAnc ArchaeaAnc, and hMAT1A, where we observed CrenAnc displays similar promiscuities toward all the bases as the extant MAT enzyme (Fig. 3), while EuryAnc was losing promiscuity toward cytidine triphosphate (CTP). Interestingly, ArchaeaAnc acquires specificity for ATP and GTP ($K_M$ ATP ArchaeaAnc 0.15 mM, $K_M$ GTP ArcheaArc 1.63 mM; Fig. 3A and Figs S14 and S18). In contrast, eukarya hMAT1A shows specificity for the ATP substrate ($K_M$ ATP hMAT1A 0.47 mM (Fig. 4A and Figs S17 and S18). Unfortunately, the bacterial and eukaryotic ancestral MAT sequences could not be expressed in the soluble and insoluble fractions. Therefore, this could not be further studied. Furthermore, in the archaeal phylogenetic tree, we detected a change in specificity, starting from a more specific ancestor toward a promiscuous extant enzyme. We observed that the specificity for ATP vs GTP goes from a 10-fold difference in the ancestor ArchaeaAnc to no difference in *MjMAT*, which indicates that the ancestor was specific toward purine bases, but through the course of evolution, modern-day archaeal MAT became promiscuous toward different bases. Therefore, these data hint at the hypothesis that all the modern-day MAT enzymes from the three domains shared a common ancestor specific for the nucleobase and that archaeal MATs diverged in sequence, structure, and substrate specificity.

## Additional factors that influence the evolutionary trajectory of MAT

In this study, we have shown that various factors could have contributed to the unusual evolutionary trajectory of the MAT enzyme across the three domains of life. However, some additional aspects could have played a significant role in shaping the evolutionary trajectory of the MAT enzyme in archaea, such as codon usage bias, tRNA bias, or an alternative/distinct SAM metabolism. It has been shown that in general, the archaeal genomes display a higher GC-rich tendency in contrast to bacterial and eukaryotic genomes. This tendency has a direct influence on codon usage. Codon usage bias also has a direct association with tRNA abundance for protein translation optimization [49,50]. This includes avoiding slowly translated codons and utilizing codons with the most cognate abundant tRNAs from the genome. Additionally, in highly expressed genes, favored codons are easily recognized by abundant tRNA molecules [51–53]. This bias could be further augmented by the time and speed of gene expression, as it could also play a

critical role in the presence of more abundant and less diverse tRNAs as well [54]. Another possible factor that might have influenced the divergence of archaeal MAT is the altered SAM metabolism. Although we acknowledge the presence of a different methyl donor in archaea, tetrahydromethanopterin [54], no clear connection to the SAM metabolism was found to support the hypothesis of the alternative MAT evolutionary trajectory. Additionally, since their divergence from LUCA, the selection and adaptive forces operating on bacterial/eukaryotic and archaeal clades may have differed substantially due to a variety of factors, including different environmental and metabolic constraints, differences in the regulation of activity, changes in the oligomerization state, and intracellular turnover. It is well known that bacterial MAT is strictly dependent on GroEL [55] and this dependency might play a role as an adaptive force; however, no information is available for the archaeal MAT. Therefore, we anticipate that these additional factors could also play a key role in guiding the evolution of highly diverged archaeal MAT, as well as other enzymes.

In brief, we expected the large interface region from MAT to be well conserved across the three domains of life, primarily because of the following reasons: (a) It accommodates the active site residues; (b) facilitates the same catalytic reaction, and (c) is composed of a large flat hydrophobic region composed of β-sheets, which are known to evolve more slowly than the α-helical regions localized at the surface in the case of MAT [56,57]. However, our results clearly show that the active site interface is surprisingly divergent in the case of archaeal MAT, in contrast to bacterial/eukaryotic MAT. The present work highlights how these unique MAT features make this enzyme family an ideal system for studying one exceptional case where the canonical domain relationships are disobeyed.

## Methods

### Database

Model data are available in the PMDB database under the accession numbers PM0083365, PM0083366, and PM0083367

### Materials

ATP, GTP, CTP, uridine triphosphate (UTP), methionine, SAM, HEPES, $MgCl_2$, KCl, IPTG, Tris/HCl, $Na_2HPO_4$, $NaH_2PO_4$, potassium phosphate, NaCl, imidazole, β-mercaptoethanol, DTT, kanamycin, glycerol, NaOH, HCl, ammonium acetate, Bacto agar, Bacto tryptone, and Bacto

yeast extract all other chemicals, and HPLC grade solvents were purchased from commercial sources and used as supplied unless otherwise mentioned. Page Ruler prestained protein ladder, 10–180 kDa was purchased from Thermo Fisher Scientific (Waltham, MA, USA). BL21 (DE3) competent cells were bought from New England Biolabs (NEB, Ipswich, MA, USA). Benzonase and cOmplete His-Tag Purification Resin (NiNTA) were purchased from Sigma-Aldrich (St. Louis, MO, USA). Protein inhibitor cocktail (PIC) and lysozyme were purchased from Nacalai Tesque, Inc (Kyoto, Japan). 12% Mini-PROTEAN TGX Precast Protein Gels, 12-well, were purchased from Bio-Rad (Hercules, CA, USA). Amicon centrifugal filters were purchased from Merck. In-Fusion HD Cloning Kit was purchased from Takara Bio (Kusatsu, Japan). All experiments were performed using ultra-pure water purification system from a Milli-Q Integral MT10 type 1 (Millipore, Burlington, MA, USA).

### Sequence data collection

Methionine adenosyltransferase sequences from *M. jannaschii* (UniProt: Q58868, METE_METJA), *Escherichia coli* (UniProt: Q58605, METK_METJA), and MAT1A from *Homo sapiens* (UniProt: Q00266, METK1_HUMAN) were used to conduct NCBI BLAST [58] searches across the nonredundant database. Furthermore, the database search conditions were filtered based on the NCBI recommended e-value cutoff, that is, 1e-5, and only search results with query coverage > 90% and sequence identity > 55% were considered in constructing the initial sequence dataset. The collected sequences were then subjected to reciprocal BLAST searches to confirm orthologous relationship(s), as it is a common computational method for predicting putative orthologs. METK sequences for archaea from the phylogenetic group *asgard* (superphylum) were not considered in this study owing to the lack of taxonomic classifications associated with the sequences that were reported as hits in the NCBI BLAST searches.

### Sequence analyses

The CD-HIT program [59] was utilized to reduce sequence redundancy, by clustering at an 80% sequence identity cutoff threshold, with default settings. The Gblocks program [60] was implemented to identify highly conserved sites across sequence alignments, with assistance from the secondary structure information (from experimentally solved crystal structures) to guide the alignments. Sequence alignment was carried out using the MAFFT computer program [61], followed by a manual check to identify any errors [59]. To reduce the sequence redundancy, we used CD-HIT, by clustering at 80% sequence identity cutoff threshold with default settings. The Gblocks program [60] was used to identify highly conserved sites across sequence alignments with assistance from the secondary structure

information (from experimentally solved crystal structures) to guide the alignments. Sequence alignment was carried out using the MAFFT [61], followed by manual curation to check for any errors. Sequence alignments for WebLogo of active site residues and interface residues are provided in Appendices S1–S3.

## SSN studies

Sequence similarity networks were constructed based on the output obtained from the EFI-Enzyme Similarity Tool (EST) server [62], and subsequently, the cytoscape program (https://cytoscape.org/) was utilized to explore the SSNs via the organic layout. Here, we managed the initial dataset using CD-HIT-based clustering by selecting a sequence identity threshold of 80%. The 'organic' layout algorithm was chosen for graphical clustering.

## Phylogenetic analysis and ancestral sequence reconstruction

ModelTest computer program [63] was implemented to select the best evolutionary model for sequence alignments based on the Bayesian inference criterion. The LG model [64] with invariant sites (+I) for discrete gamma categories (+G4) was selected to construct the maximum-likelihood (ML) phylogenetic trees. Initially, we constructed ML-based trees with the IQ-TREE program.

## Phylogenetic analysis and ancestral sequence reconstruction

ModelTest [63] was used to pick out the best evolutionary model for sequence alignments based on the Bayesian inference criterion. The LG model [64] with invariant sites (+I) for discrete gamma categories (+G4) was selected to construct maximum likelihood (ML) phylogenetic trees. Initially, we constructed ML-based trees with the IQ-TREE program [65] (using 10,000 bootstrap replicates) for the datasets, to inspect the topology distribution within the dataset. The confidence level for the nodes was assessed using Felsenstein's bootstrap method [66], and the consensus tree was redrawn using FigTree (http://tree.bio.ed.ac.uk/software/figtree/). The tree file from the IQ-TREE program output was parsed to obtain the ancestral sequences at the highlighted nodes. In addition, the phylogenetic program MEGA X was consulted to check for ancestral sequences [67]. The NCBI taxonomy database was also consulted to check for major phylogenetic classifications [68].

## Molecular modeling and structural studies

Ancestral sequences (of interest) were modeled using the Swiss model at the ExPASy server (https://swissmodel.expasy.org/). 3D models were visualized using PYMOL (https://pymol.org/2), to identify, tabulate, and study the key interaction residues. The overall quality of the 3D models depends directly on the shared sequence identity between the target sequence and the sequence of the template structure; therefore, we utilized template structures with high sequence similarity. Interface residues were identified using PYMOL and the PISA server (https://www.ebi.ac.uk/pdbe/pisa/). Structural superposition studies were performed using the mtm-align server [69]. The 17 sequences for Fig. 6, Figs S4 and S5 were collected using blastp and the PDB database for sequence search. *Escherichia coli* (PDB 1RG9), human (PDB 4NDN), and pfMAT (PDB 6S83) sequences were used as input sequences for each search, and one representative sequence for each organism was manually selected. Two isoform structures were selected for human MAT, hMAT1A (6SW5), and hMAT2A (4NDN).

## Charge and hydrophobicity calculations

Per-site hydrophobicity and hydrophilicity calculations were carried out using the ProtScale program [70] from the Expasy server (https://web.expasy.org). We implemented the default 'Kyte–Doolittle' scale [71] for the calculations, with a residue window size of three, while the rest of the settings were left as default. Charge distribution per-site was calculated using the charge server from EMBOSS (http://www.bioinformatics.nl/cgi-bin/emboss/charge). A window length of size 3 was implemented, and the rest of the settings were left as default.

## Evolutionary rate calculations

The site-specific evolutionary rates were inferred using the IQ-TREE program, by implementing the estimated model parameters and applying an empirical Bayesian approach to assign site rates as mean over rate categories.

## Hierarchical clustering with pvclust

An R language package was utilized to assess the hierarchical clustering pattern, as it provides statistical support for each cluster through *P*-values. It provides two types of *P*-values: AU *P*-value and BP value. The AU *P*-value, which is computed using multiscale bootstrap resampling, is a better approximation to the unbiased *P*-value than the BP value computed using normal bootstrap resampling.

## Structural analysis

Structural superpositions were carried out employing the PDBeFold [72] server with default parameters, employing pdbs chain A of 1MXA, 2P02, and 4L4Q. Subdomain

comparison was assessed following ECOD domain delimitation and manual dissection of archaeal MAT (Fig. S8).

## Enzyme cloning

CrenArc, EuryArc, ArcheaArc MAT, and the hMAT1A codon-optimized gene for *E. coli* were ordered from Twist Bioscience, San Francisco, and cloned into pet28a vector by infusion cloning using In-Fusion HD Cloning Kit by following the kit protocol. Primers used for cloning are listed in Table 1.

## Protein expression and purification

CrenArc, EuryArc, and ArcheaArc MAT were expressed and purified following the previously published protocol [48]. Briefly, the plasmid encoding the genes for CrenArc, EuryArc, and ArcheaArc MAT in *E. coli* BL21(DE3) cells was growing at 37 °C in LB medium with 50 $\mu g \cdot mL^{-1}$ kanamycin. Once $OD_{600}$ reached 0.6, culture was cooled down to 16 °C and induction was done with 1 mM IPTG overnight at 16 °C with shaking at 230 r.p.m. Cells were harvested by centrifugation at 4648 *g* for 20 min and stored at −80 °C until use. The bacterial pellet was resuspended in lysis buffer that contains Tris/HCl (50 mM), NaCl (300 mM), imidazole (10 mM), lysozyme (0.3 $mg \cdot mL^{-1}$), benzonase (2 U), and PIC, pH 8. The cells were incubated with lysis buffer for 30 min to 1 h, followed by sonication (Digital Sonifier Models 250, 5 s, on 5 s off for 5 min) on ice. Cell debris was removed by centrifugation at 20,960 *g* for 1 h. The supernatant was loaded onto the NiNTA chelating column, which was equilibrated with lysis buffer previously. The column was washed with wash buffer 1 that contained Tris/HCl (50 mM), NaCl (300 mM), and imidazole (50 mM), pH 8. The second wash with wash buffer 2 contained Tris/HCl (50 mM), NaCl (300 mM), and imidazole (100 mM), pH 8. Finally, the column was eluted with elution buffer containing Tris/HCl (50 mM), NaCl (300 mM), imidazole (500 mM), and glycerol (10%), pH 8. The protein content was checked by the NanoDrop 2000 spectrophotometer (Thermo Scientific), and elution buffer containing proteins was dialyzed using Amicon centrifugal filters 30,000 MWCO with dialysis buffer containing Tris/HCl (50 mM), DTT (1 mM), and glycerol (10%), pH 8, and stored at −80 °C.

hMAT1A pellet was processed like archaea but used lysis buffer containing 20 mM $NaH_2PO_4$, 300 mM NaCl, 10 mM imidazole, 1 $mg \cdot mL^{-1}$ lysozyme, 1 $\mu L$ benzonase, and PIC, pH 7.8, and incubated with lysis buffer for 30 min to 1 h.

Cells were lysed by sonication (Digital Sonifier Models 250, 5 s on 5 s off for 5 min) on ice. After centrifugation at 12,000 r.p.m. for 1 h, the supernatant was loaded onto the NiNTA column. First, the column was washed with wash buffer 1 containing $NaH_2PO_4$ (50 mM), NaCl (300 mM), and imidazole (50 mM), pH 8, then washed with wash buffer 2 containing $NaH_2PO_4$ (50 mM), NaCl (300 mM), and imidazole (100 mM), pH 8. Finally, the column was eluted with elution buffer containing $NaH_2PO_4$ (50 mM), NaCl (300 mM), imidazole (500 mM), and glycerol (10%), pH 8. Fractions containing protein were concentrated with exchange buffer Tris/HCl (25 mM) and KCl (80 mM), pH 8, using Amicon centrifugal filters 30,000 MWCO. The purity of the protein was confirmed by SDS/PAGE. Protein concentration was 2.5, 6.2, 4.4, and 14 $mg \cdot mL^{-1}$ for CrenArc, EuryArc, ArchaeArc, and hMAT1A, respectively.

## MAT activity assay

Activity assay was performed as previously reported [73]. Briefly, ATP/GTP/CTP/UTP (5 mM), methionine (10 mM), HEPES (100 mM), $MgCl_2$ (10 mM), KCl (50 mM), and MAT (20 $\mu M$) were mixed in water, following which the pH of the mixture was adjusted to 8. Reactions were incubated at 37 °C for eukarya and at 55 °C for archaea in a thermomixer comfort system (Eppendorf, Hamburg Germany) for 1 h. The reaction was quenched with acetonitrile, followed by centrifugation at 12,000 r.p.m. for 5 min to precipitate the enzymes. Finally, the supernatant was filtered through a 0.22-$\mu m$ filter (Merck) and injected into the ultraperformance liquid chromatography (UPLC) for analysis. UPLC method (UPLC Acquity H class, Waters, MA USA): Diluted reaction aliquots were analyzed using UPLC with a HILIC column (SeQuant® ZIC®-cHILIC 3 $\mu m$, 100 Å 150 × 2.1 mm PEEK-coated HPLC column). An isocratic method was used with 35% solvent A (100 mM ammonium acetate, pH 5.3) and 65% solvent B (acetonitrile) for 15 min. Each injection had a volume of 3 $\mu L$, with a flow rate of 0.2 $mL \cdot min^{-1}$, and was detected at a wavelength of 260 nm.

For kinetic analysis, SGM was purified using the UPLC method mentioned above and standard curves were plotted for SGM and SAM. For kinetic assay of ArchaeaAnc (0.5 $\mu M$) and hMAT1A (0.5 $\mu M$), concentrations of the ATP and GTP were in the range of 0.1 to 2 mM and a saturating concentration of methionine 10 mM was used. Using the above-mentioned buffer, temperature was 55 °C for ArchaeaAnc and 37 °C for hMAT1A. Data were analyzed by UPLC, and the kinetic parameters were determined using

**Table 1.** Primers for infusion cloning.

| Name | Froward primer | Reverse primer |
|---|---|---|
| Pet28a_linearization | 5′−TCCGTCGACAAGCTTGCGGCCGCAC−3′ | 5′−GCGGCACCAGGCCGCTGCTGTGATG−3′ |
| MAT_amplification | 5′−GCAAGCTTGTCGACGGAGCTCGAATTC−3′ | 5′−GCAAGCTTGTCGACGGAGCTCGAATTC−3′ |

the Michaelis–Menten equation using the GRAPHPAD PRISM 7.02 (San Diego, CA, USA; https://www.graphpad.com).

## Acknowledgements

## Conflict of interest

The authors declare no conflict of interest.

## Data accessibility

The data generated/used in this study are available in the article and online supplementary materials. In addition, any other data underlying this article will be readily shared upon request from the corresponding author.

## Author contributions

BPSC and PL conceived the idea of this study. BPSC, MG, and DM performed the experiments. BPSC, MG, ST-P, and PL analyzed the data. BPSC, MG, STP, and PL wrote the manuscript and prepared the figures.

## References

1 Theobald DL. A formal test of the theory of universal common ancestry. *Nature*. 2010;**465**:219–U104.

2 Elias M, Tawfik DS. Divergence and convergence in enzyme evolution: parallel evolution of paraoxonases from quorum-quenching lactonases. *J Biol Chem*. 2012;**287**:11–20.

3 Koonin EV. Comparative genomics, minimal gene-sets and the last universal common ancestor. *Nat Rev Microbiol*. 2003;**1**:127–36.

4 Ouzounis CA, Kunin V, Darzentas N, Goldovsky L. A minimal estimate for the gene content of the last universal common ancestor–exobiology from a terrestrial perspective. *Res Microbiol*. 2006;**157**:57–68.

5 Ragan MA. Trees and networks before and after Darwin. *Biol Direct*. 2009;**4**, 43:discussion 43.

6 Dagan T, Martin W. Ancestral genome sizes specify the minimum rate of lateral gene transfer during prokaryote evolution. *Proc Natl Acad Sci USA*. 2007;**104**:870–5.

7 Beiko RG, Harlow TJ, Ragan MA. Highways of gene sharing in prokaryotes. *Proc Natl Acad Sci USA*. 2005;**102**:14332–7.

8 McBreen K, Lockhart PJ. Reconstructing reticulate evolutionary histories of plants. *Trends Plant Sci*. 2006;**11**:398–404.

9 Ammelburg M, Frickey T, Lupas AN. Classification of AAA+ proteins. *J Struct Biol*. 2006;**156**:2–11.

10 Dunin-Horkawicz S, Kopec KO, Lupas AN. Prokaryotic ancestry of eukaryotic protein networks mediating innate immunity and apoptosis. *J Mol Biol*. 2014;**426**:1568–82.

11 Ammelburg M, Hartmann MD, Djuranovic S, Alva V, Koretke KK, Martin J, *et al*. A CTP-dependent archaeal riboflavin kinase forms a bridge in the evolution of cradle-loop barrels. *Structure*. 2007;**15**:1577–90.

12 Brueckner J, Martin WF. Bacterial genes outnumber Archaeal genes in eukaryotic genomes. *Genome Biol Evol*. 2020;**12**:282–92.

13 Graham DE, Bock CL, Schalk-Hihi C, Lu ZJ, Markham GD. Identification of a highly diverged class of S-adenosylmethionine synthetases in the Archaea. *J Biol Chem*. 2000;**275**:4055–9.

14 Cantoni GL. Activation of methionine for transmethylation. *J Biol Chem*. 1951;**189**:745–54.

15 Weiss MC, Sousa FL, Mrnjavac N, Neukirchen S, Roettger M, Nelson-Sathi S, *et al*. The physiology and habitat of the last universal common ancestor. *Nat Microbiol*. 2016;**1**:16116.

16 Broderick WE, Hoffman BM, Broderick JB. Mechanism of radical initiation in the radical S-adenosyl-l-methionine superfamily. *Acc Chem Res*. 2018;**51**:2611–9.

17 Holliday GL, Akiva E, Meng EC, Brown SD, Calhoun S, Pieper U, *et al*. Atlas of the radical SAM superfamily: divergent evolution of function using a "plug and play" domain. *Methods Enzymol*. 2018;**606**:1–71.

18 Broderick JB, Duffus BR, Duschene KS, Shepard EM. Radical S-adenosylmethionine enzymes. *Chem Rev*. 2014;**114**:4229–317.

19 Ouyang Y, Wu Q, Li J, Sun S, Sun S. S-adenosylmethionine: a metabolite critical to the regulation of autophagy. *Cell Prolif*. 2020;**53**:e12891.

20 Chiang PK, Gordon RK, Tal J, Zeng GC, Doctor BP, Pardhasaradhi K, *et al*. S-Adenosylmethionine and methylation. *FASEB J*. 1996;**10**:471–80.

21 Robertson KD. DNA methylation and human disease. *Nat Rev Genet*. 2005;**6**:597–610.

22 Pegg AE. Functions of polyamines in mammals. *J Biol Chem*. 2016;**291**:14904–12.

23 Lu SC. Regulation of glutathione synthesis. *Mol Aspects Med*. 2009;**30**:42–59.

24 Horitani M, Shisler K, Broderick WE, Hutcheson RU, Duschene KS, Marts AR, *et al*. Radical SAM catalysis via an organometallic intermediate with an Fe-[5'-C]-deoxyadenosyl bond. *Science*. 2016;**352**:822–5.

25 Lu SC, Mato JM. S-adenosylmethionine in liver health, injury, and cancer. *Physiol Rev*. 2012;**92**:1515–42.

26 Hayashi T, Teruya T, Chaleckis R, Morigasaki S, Yanagida M. S-Adenosylmethionine synthetase is required for cell growth, maintenance of G0 phase, and termination of quiescence in fission yeast. *iScience*. 2018;**5**:38–51.

27 Murzin AG, Brenner SE, Hubbard T, Chothia C. SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J Mol Biol*. 1995;**247**:536–40.

28 Markham GD, Pajares MA. Structure-function relationships in methionine adenosyltransferases. *Cell Mol Life Sci*. 2009;**66**:636–48.

29 Liu S. A review on protein oligomerization process. *Int J Precis Eng Manuf*. 2015;**16**:2731–60.

30 Torshin I. Activating oligomerization as intermediate level of signal transduction: analysis of protein-protein contacts and active sites in several glycolytic enzymes. *Front Biosci*. 1999;**4**:D557–70.

31 Miller S, Lesk AM, Janin J, Chothia C. The accessible surface area and stability of oligomeric proteins. *Nature*. 1987;**328**:834–6.

32 Kleiner D, Shmulevich F, Zarivach R, Shahar A, Sharon M, Ben-Nissan G, *et al*. The interdimeric interface controls function and stability of *Ureaplasma urealiticum* methionine S-adenosyltransferase. *J Mol Biol*. 2019;**431**:4796–816.

33 Minici C, Mosca L, Ilisso CP, Cacciapuoti G, Porcelli M, Degano M. Structures of catalytic cycle intermediates of the *Pyrococcus furiosus* methionine adenosyltransferase demonstrate negative cooperativity in the archaeal orthologues. *J Struct Biol*. 2020;**210**:107462.

34 Oue S, Okamoto A, Yano T, Kagamiyama H. Redesigning the substrate specificity of an enzyme by cumulative effects of the mutations of non-active site residues. *J Biol Chem*. 1999;**274**:2344–9.

35 Arnold FH, Wintrode PL, Miyazaki K, Gershenson A. How enzymes adapt: lessons from directed evolution. *Trends Biochem Sci*. 2001;**26**:100–6.

36 Aharoni A, Gaidukov L, Khersonsky O, Mc QGS, Roodveldt C, Tawfik DS. The 'evolvability' of promiscuous protein functions. *Nat Genet*. 2005;**37**:73–6.

37 Morley KL, Kazlauskas RJ. Improving enzyme properties: when are closer mutations better? *Trends Biotechnol*. 2005;**23**:231–7.

38 Tomatis PE, Rasia RM, Segovia L, Vila AJ. Mimicking natural evolution in metallo-beta-lactamases through second-shell ligand mutations. *Proc Natl Acad Sci USA*. 2005;**102**:13761–6.

39 Liu G, Chen M, Mi N, Yang W, Li X, Wang P, *et al*. Increased oligomerization and phosphorylation of alpha-synuclein are associated with decreased activity of glucocerebrosidase and protein phosphatase 2A in aging monkey brains. *Neurobiol Aging*. 2015;**36**:2649–59.

40 Schloss PD, Girard RA, Martin T, Edwards J, Thrash JC. Status of the Archaeal and bacterial census: an update. *MBio*. 2016;**7**:e00201-16.

41 Frickey T, Lupas A. CLANS: a Java application for visualizing protein families based on pairwise similarity. *Bioinformatics*. 2004;**20**:3702–4.

42 Komoto J, Yamada T, Takata Y, Markham GD, Takusagawa F. Crystal structure of the S-adenosylmethionine synthetase ternary complex: a novel catalytic mechanism of S-adenosylmethionine synthesis from ATP and Met. *Biochemistry-Us*. 2004;**43**:1821–31.

43 Panmanee J, Antonyuk SV, Hasnain SS. Structural basis of the dominant inheritance of hypermethioninemia associated with the Arg264His mutation in the MAT1A gene. *Acta Crystallogr D*. 2020;**76**:594–607.

44 Suzuki R, Shimodaira H. Pvclust: an R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics*. 2006;**22**:1540–2.

45 Laskowski RA, Jabłońska J, Pravda L, Vařeková RS, Thornton JM. PDBsum: Structural summaries of PDB entries. *Protein Science*. 2018;**27**:129–34. https://doi.org/10.1002/pro.3289

46 Cheng H, Schaeffer RD, Liao YX, Kinch LN, Pei JM, Shi SY, *et al*. ECOD: an evolutionary classification of protein domains. *Plos Comput Biol*. 2014;**10**:e1003926.

47 Cheng H, Liao YX, Schaeffer RD, Grishin NV. Manual classification strategies in the ECOD database. *Proteins*. 2015;**83**:1238–51.

48 Lu ZJ, Markham GD. Enzymatic properties of S-adenosylmethionine synthetase from the archaeon *Methanococcus jannaschii*. *J Biol Chem*. 2002;**277**:16624–31.

49 Andersson SG, Kurland CG. Codon preferences in free-living microorganisms. *Microbiol Rev*. 1990;**54**:198–210.

50 Kotlar D, Lavner Y. The action of selection on codon bias in the human genome is related to frequency, complexity, and chronology of amino acids. *BMC Genom*. 2006;**7**:67.

51 Bibb MJ, Findlay PR, Johnson MW. The relationship between base composition and codon usage in bacterial genes and its use for the simple and reliable identification of protein-coding sequences. *Gene*. 1984;**30**:157–66.

52 McEwan CE, Gatherer D, McEwan NR. Nitrogen-fixing aerobic bacteria have higher genomic GC content than non-fixing species within the same genus. *Hereditas*. 1998;**128**:173–8.

53 Liu QP, Tan J, Xue QZ. Synonymous codon usage bias in the rice cultivar 93–11 (*Oryza sativa* L. ssp. indica). *Yi Chuan Xue Bao*. 2003;**30**:335–40.

54 Rocha EP. Codon usage bias from tRNA's point of view: redundancy, specialization, and efficient decoding for translation optimization. *Genome Res.* 2004;**14**:2279–86.

55 Kerner MJ, Naylor DJ, Ishihama Y, Maier T, Chang HC, Stines AP, *et al.* Proteome-wide analysis of chaperonin-dependent protein folding in *Escherichia coli. Cell.* 2005;**122**:209–20.

56 Roth C, Liberles DA. A systematic search for positive selection in higher plants (Embryophytes). *BMC Plant Biol.* 2006;**6**:12.

57 Schaefer C, Schlessinger A, Rost B. Protein secondary structure appears to be robust under in silico evolution while protein disorder appears not to be. *Bioinformatics.* 2010;**26**:625–31.

58 Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol.* 1990;**215**:403–10.

59 Huang Y, Niu B, Gao Y, Fu L, Li W. CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bioinformatics.* 2010;**26**:680–2.

60 Talavera G, Castresana J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst Biol.* 2007;**56**:564–77.

61 Katoh K, Misawa K, Kuma K, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* 2002;**30**:3059–66.

62 Gerlt JA, Bouvier JT, Davidson DB, Imker HJ, Sadkhin B, Slater DR, *et al.* Enzyme function initiative-enzyme similarity tool (EFI-EST): a web tool for generating protein sequence similarity networks. *Biochim Biophys Acta.* 2015;**1854**:1019–37.

63 Posada D. ModelTest server: a web-based tool for the statistical selection of models of nucleotide substitution online. *Nucleic Acids Res.* 2006;**34**:W700–3.

64 Le SQ, Gascuel O. An improved general amino acid replacement matrix. *Mol Biol Evol.* 2008;**25**:1307–20.

65 Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol.* 2015;**32**:268–74.

66 Felsenstein J. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution.* 1985;**39**:783–91.

67 Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol.* 2018;**35**:1547–9.

68 Federhen S. The NCBI taxonomy database. *Nucleic Acids Res.* 2012;**40**:D136–43.

69 Dong R, Pan S, Peng Z, Zhang Y, Yang J. mTM-align: a server for fast protein structure database search and multiple protein structure alignment. *Nucleic Acids Res.* 2018;**46**:W380–6.

70 Wilkins MR, Gasteiger E, Bairoch A, Sanchez JC, Williams KL, Appel RD, *et al.* Protein identification and analysis tools in the ExPASy server. *Methods Mol Biol.* 1999;**112**:531–52.

71 Kyte J, Doolittle RF. A simple method for displaying the hydropathic character of a protein. *J Mol Biol.* 1982;**157**:105–32.

72 Krissinel E, Henrick K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr D.* 2004;**60**:2256–68.

73 Gade M, Tan LL, Damry AM, Sandhu M, Brock JS, Delaney A, *et al.* Substrate dynamics contribute to enzymatic specificity in human and bacterial methionine adenosyltransferases. *bioRxiv.* 2021. [PREPRINT].

## Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

**Scheme S1**. *S*-adenosylmethionine (SAM) biosynthesis and its functions by secondary metabolites.

**Fig. S1**. Sequence similarity networks for MAT enzymes.

**Fig. S2**. The evolutionary trends adapted by the interface sites (located along the large interface of the homo-tetramer) as well the active sites.

**Fig. S3**. An interface view of three representative MAT structures and Weblogo of large interface residues from the three domains of life.

**Fig. S4**. Structural alignment of the residues depicting the 51 interface residues from 17 MAT crystal structures.

**Fig. S5**. Structure-based sequence alignment.

**Fig. S6**. Clustering analysis of hydropathy (A) and charge (B) of representative structures from the three domains of life.

**Fig. S7**. Structural comparison of MAT representatives from the three domains of life.

**Fig. S8**. Comparison of MAT subdomains within domains of life.

**Fig. S9**. Structural comparison of MAT subdomains across the domains of life.

**Fig. S10**. SDS/PAGE of ArchaeaAnc MAT protein purification.

**Fig. S11**. SDS/PAGE of CrenAnc MAT protein purification.

**Fig. S12**. SDA/PAGE of EuryAnc MAT protein purification.

**Fig. S13**. SDS/PAGE of hMAT1A purification.

**Fig. S14**. UPLC chromatogram of the reaction between NTP, methionine, and ArchaeaAnc.

**Fig. S15**. UPLC chromatogram of the reaction between NTP, methionine, and CrenAnc.

**Fig. S16**. UPLC chromatogram of the reaction between NTP, methionine, and EuryAnc.

**Fig. S17**. UPLC chromatogram of the reaction between NTP, methionine, and hMAT1A.

**Fig. S18**. Kinetic parameters for the SAM and SGM formation by ArchaeaAnc and hMAT1A.

**Appendix S1**. Bacteria.

**Appendix S2**. Archaea.

**Appendix S3**. Eukarya.