

Coenzyme-Protein Interactions since Early Life

Reviewed Preprint

v2 • October 16, 2024

Revised by authors

Reviewed Preprint

v1 • February 26, 2024

Alma Carolina Sanchez-Rocha, Mikhail Makarov, Lukáš Pravda, Marian Novotný , Klára Hloučová 

Department of Cell Biology, Faculty of Science, Charles University, BIOCEV, Prague, 12843, Czech Republic • Protein Data Bank in Europe, European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Genome Campus, Hinxton, Cambridge CB10 1SD, UK • Institute of Organic Chemistry and Biochemistry, Czech Academy of Sciences, Prague, 16610, Czech Republic

 https://en.wikipedia.org/wiki/Open_access
 Copyright information

Abstract

Recent findings in protein evolution and peptide prebiotic plausibility have been setting the stage for reconsidering the role of peptides in the early stages of life's origin. Ancient protein families have been found to share common themes and proteins reduced in composition to prebiotically plausible amino acids have been reported capable of structure formation and key functions, such as binding to RNA. While this may suggest peptide relevance in early life, their functional repertoire when composed of a limited number of early residues (missing some of the most sophisticated functional groups of today's alphabet) has been debated.

Cofactors enrich the functional scope of about half of extant enzymes but whether they could also bind to peptides lacking the evolutionary late amino acids remains speculative. The aim of this study was to resolve the early peptide propensity to bind organic cofactors by analysis of protein-coenzyme interactions across the Protein Data Bank (PDB). We find that the prebiotically plausible amino acids are more abundant in the binding sites of the most ancient coenzymes and that such interactions rely more frequently on the involvement of the protein backbone atoms and metal ion cofactors. Moreover, we have identified a few select examples in today's enzymes where coenzyme binding is supported solely by prebiotically available amino acids. These results imply the plausibility of a coenzyme-peptide functional collaboration preceding the establishment of the Central Dogma and full protein alphabet evolution.

eLife Assessment

This study presents a **valuable** examination of the prevalence of interactions between amino acids from different periods of Earth's history and coenzymes. While the premise of this work is well founded and the analysis is **solid**, with more data, the interpretation could change. This manuscript would be of interest to evolutionary biologists and biophysicists.

<https://doi.org/10.7554/eLife.94174.2.sa3>

Introduction

Organic and inorganic cofactors occupy about half of all known protein structures, expanding across all the enzyme E.C. classes (Putignano, 2018 [DOI](#); Mukhopadhyay et al., 2019 [DOI](#)). While their role in current life is indisputable, some of the cofactors were present and apparently crucial also during life's early evolution (Chu and Zhang 2020 [DOI](#); Goldman and Kacar 2021 [DOI](#); Kirschning et al., 2021; Fried et al., 2022 [DOI](#); Kirschning, 2022 [DOI](#)). The significance of metal ions has been broadly discussed, regardless of the different origins-of-life scenarios and has somewhat overshadowed that of organic cofactors (e.g. Wächtershäuser, 1992 [DOI](#); Russell and Hall, 1997 [DOI](#); Lane and Martin, 2012 [DOI](#); Chu and Zhang, 2020 [DOI](#); Fried et al., 2022 [DOI](#)).

Diverse lines of evidence have however indicated that many of the extant organic cofactors (coenzymes) date back to the earliest life while their core chemistries have been detected in abiotic material such as recently reported by the Hayabusa2 mission (Holliday et al., 2007 [DOI](#); Fried et al., 2022 [DOI](#); Naraoka et al., 2023 [DOI](#)). At the same time, these ancient coenzymes – often of nucleotide origin– have been traced to the most ancient protein folds (such as P-loop NTPases, TIM beta/alpha-barrels, OB and Rossmann folds) that date before the Last Universal Common Ancestor (LUCA) (Goldman and Kacar, 2021 [DOI](#); Caetano-Anollés et al., 2007; Goldman et al., 2013 [DOI](#); Longo et al., 2020 [DOI](#) (a); Kessel and Ben-Tal, 2022 [DOI](#)). Within the most ancient folds, tens of peptide fragments/themes have been identified throughout seemingly unrelated structural domains, and frequently found to mediate ligand binding (Söding and Lupas, 2003 [DOI](#); Alva et al., 2015 [DOI](#); Narunsky et al., 2020 [DOI](#); Kolodny et al., 2021 [DOI](#)). Such themes may well represent the remnants of protoenzymes in a peptide-nucleotide world (Fried et al., 2022 [DOI](#)). Chu and Zhang recently proposed that cofactors could initially “select” the earliest primitive proteins from the vast sequence space by the ability to bind them (Chu and Zhang, 2020 [DOI](#)). More generally, binding of cofactors to peptides could thus determine the evolution of both protoenzyme function and folding preferences (Tokuriki and Tawfik, 2009 [DOI](#)).

Prior to the fixation of Central Dogma and ribosomal synthesis, peptides would condense from amino acids (or their alternatives) prebiotically abundant in the environment (Frenkel-Pinter et al. 2019 [DOI](#); Frenkel-Pinter et al. 2020 [DOI](#); Fried et al., 2022 [DOI](#)). Independent meta-analyses of the amino acid alphabet evolution based on different possible sources of organic material and different disciplines point towards an “early alphabet” of ~10 residues (Ala, Asp, Glu, Gly, Ile, Leu, Pro, Ser, Thr and Val) (Higgs and Pudritz, 2009 [DOI](#); Trifonov, 2000 [DOI](#); Cleaves, 2010 [DOI](#)). These could be supplemented by other prebiotically plausible non-canonical amino acids while the other half of the canonical alphabet is assumed to be the product of later biosynthesis (Wong and Bronskill, 1979 [DOI](#); Weber and Miller, 1981 [DOI](#); Burton et al., 2012 [DOI](#); Zaia et al., 2008 [DOI](#)). Typically, the early amino acids (canonical as well as non-canonical) are smaller and less complex, missing e.g. sulfur groups and aromatics. Additionally, the canonical early alphabet lacks positively charged residues. An emerging question therefore is whether coenzymes could bind to small proteins of prebiotic relevance and whether they could be bound by the prebiotically available residues. In such a scenario, cofactors would provide a palette of functional groups to the early peptide world which would nominate them relatively sophisticated structural and catalytic hubs (Milner-White and Russell, 2011 [DOI](#)). Alternatively, if coenzymes could not be bound by these simple amino acids, this would suggest that their pairing with peptide molecules would become relevant only after the evolution of the full amino acid alphabet.

Work from our group and others has recently demonstrated that in select cases, protein sequences re-engineered from the early amino acids can still bind to nucleic acid and nucleotide-based cofactors (Longo et al., 2020 [DOI](#) (b); Makarov et al., 2021 [DOI](#); Giacobelli et al., 2022 [DOI](#)). Whether this phenomenon is still seen in today's biology, its abundance and laws represent open questions. Here, we present a systematic survey of coenzyme binding throughout the PDB database. The

outcomes of our study support that the coenzyme binding characteristics by amino acids differ by their evolutionary age. Early amino acids are enriched in binding pockets of the most ancient coenzymes and the interaction relies predominantly on the protein backbone groups. Selected examples show that unlike evolutionary younger cofactors, the ancient cofactors can still be bound in proteins only by early amino acids. Our analysis therefore points to an early peptide-coenzyme significance, preceding evolution of proteosynthesis and fixation of the Central Dogma.

Results

Identification of coenzymes in PDB

We identified all the available structures from the PDB that interact with the 27 coenzyme classes as defined in [Fischer et al., 2010](#). In addition, ATP (that was not included in that study) was included here. Using these parameters, we found 25,822 protein structures and 81 nucleic acid macromolecules (Supplementary Table 1, Supplementary File 1). The protein structures were assigned to 8194 UniProt (The UniProt Consortium, 2023) codes. Those UniProt sequences were clustered by 90% identity and resulted in 7399 unique UniProt entries, corresponding to 21,317 protein structures (**Fig. 1**). In parallel, the clustering was also performed for 30% sequence identity, resulting in 3544 UniProt codes and 9645 PDB structures.

The interaction ratio method was adopted to identify the most relevant residues in coenzyme binding sites. For each protein (unique UniProt ID) we defined the cofactor binding site as a subset of amino acids that appeared to interact with the cofactor in at least 50% of the structures of that particular protein within our dataset to pinpoint the amino acids that are important for the interaction. This methodology does not consider any qualitative criteria (e.g. resolution, R-factor, Clashscore).

Our database is composed of protein structures from members of the three cellular domains – Bacteria (54.3%), Archaea (6.2%), Eukaryota (37.8%) – as well as Viruses (1.5%), metagenomes and unassigned entries (0.3%) (Supplementary Table 1).

Evolutionary classification of coenzymes

To differentiate the evolutionary age of the analyzed coenzymes, we further adapted the classification system from [Fried et al., 2022](#). This system encompasses four primary categories (or temporalities) and one additional subcategory: i) "Ancient" coenzymes, including the subcategory "Nucleotide derived"; ii) "LUCA" coenzymes; iii) "Post-LUCA" coenzymes; and iv) "Unclassified" coenzymes (**Fig. 2**).

"Ancient" coenzymes comprise those that could be prebiotically synthesized, according to available studies ([Miller and Schlesinger, 1993](#); [Keefe et al., 1995](#); [Holliday et al., 2007](#); [Kirschning, 2021](#); [Menor-Salván et al., 2022](#); [Pinna et al., 2022](#)); while the subcategory "Nucleotide derived" includes cofactors chemically derived from nucleotides ([White, 1975](#); [Monteverde et al., 2017](#)). "LUCA" coenzymes were presumably present in the last universal common ancestor (LUCA) and exhibit a universal distribution among Bacteria, Archaea, and Eukarya, although their prebiotically feasible synthesis was not established. "Post-LUCA" coenzymes likely originated only after the divergence of the three cellular domains, mirrored in their non-universal distribution. "Unclassified" coenzymes do not conform to the classification scheme. As a typical representative of the latest category, Coenzyme M has been synthesized under prebiotic conditions ([Miller and Schlesinger, 1993](#); [Kirschning, 2021](#)), nonetheless, its biosynthetic pathways in Archaea and Bacteria have been shown to arise through convergent evolution and it is mainly prevalent in methanogens ([Wu et al., 2022](#)). Factor F430 is a coenzyme only distributed in methanogens ([Thauer and Bonacker, 1994](#)), although its precursors have been synthesized prebiotically ([Seitz et al., 2021](#)). Glutathione is another example of a coenzyme

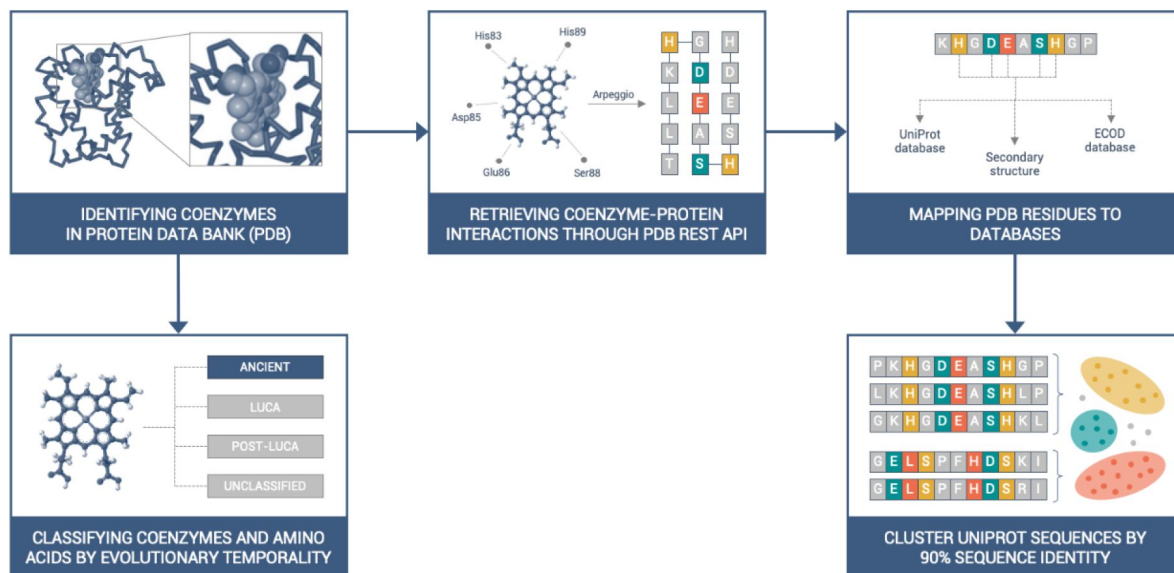


Fig 1

Workflow of this study. All available coenzymes in the PDB were identified according to the CoFactor database (Fischer et al., 2010). The PDB entries of structures bound to coenzymes were downloaded programmatically through the PDB REST API (pdbe.org/api), including the interatomic cofactor-protein interactions, calculated by Arpeggio (Jubb et al, 2017). The coenzyme binding amino acids were mapped to Uniprot databases via SIFTS (Velankar et al., 2013; Dana et al., 2019). PDB entries were grouped by UniProt code; redundancy was removed by clustering the UniProt sequences by 90% (and in parallel also 30%) sequence identity.

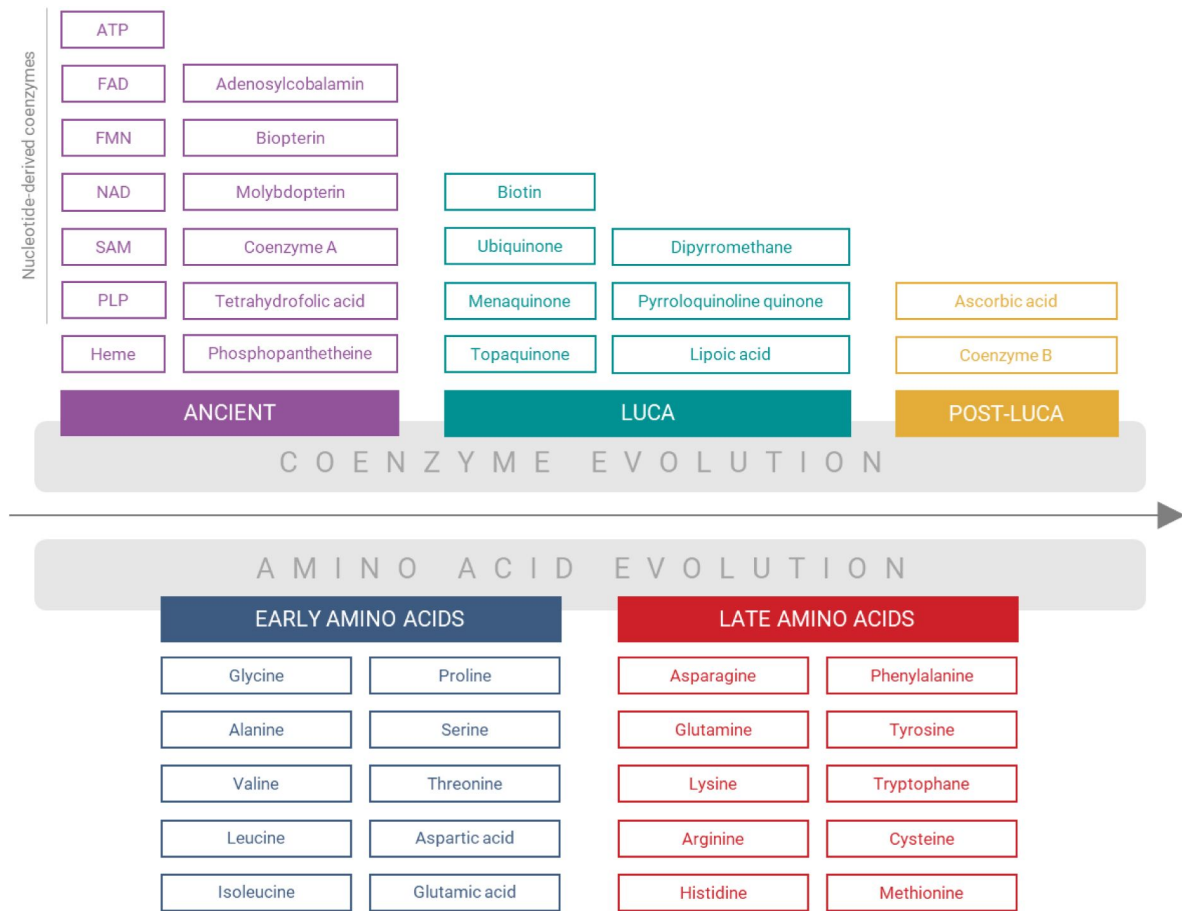


Fig 2

Classification of coenzymes and amino acids by their assumed evolutionary temporality. The “Unclassified” coenzymes Thiamine diphosphate, Coenzyme M, Factor F430 and Glutathione are not shown in the scheme.

with restricted biological distribution, being mainly in eukaryotes, Gram-negative bacteria, and one archaea phylum (Copley and Dhillon, 2002 [↗](#)) and the feasibility of its prebiotic synthesis remains unclear (Bonfio et al., 2017 [↗](#)). Thiamine diphosphate was also designated as Unclassified. Although the definitive prebiotic synthesis of Thiamine diphosphate remains unclear, preliminary investigations conducted by Aylward (2006) [↗](#) and Aylward & Bofinger (2006) [↗](#) suggest its presence in the prebiotic world. Its nucleotide nature (White, 1975) and the existence of its universal riboswitch (Barrick and Breaker, 2007 [↗](#)) provide compelling evidence of its potential status as an Ancient coenzyme.

The Ancient coenzymes represent the most abundant temporality of our PDB dataset, dominated by ATP, NAD, Heme, FAD, SAM, and CoA structures and amounting to 94% of all analyzed structures in our database grouped by UniProt codes. Within the enzyme E.C. classification, oxidoreductases and transferases represent the classes with the most abundant coenzyme content. While the LUCA, Post-LUCA and Unclassified coenzymes are typically found in specific enzyme classes, the Ancient coenzymes are distributed across all the E.C. classes (Supplementary Fig. 1).

Distribution of amino acids in the coenzyme binding sites

We hypothesized that the evolutionary significance of individual coenzyme classes would be reflected in distinct amino acid binding propensities as a smaller “early” protein alphabet apparently preceded its canonical version. The abundance of residues that compose each coenzyme binding site was analyzed and examined with respect to the order by which individual amino acids have been reported to enter the protein canonical alphabet (Higgs and Pudritz, 2009 [↗](#)) (Fig. 3 [↗](#); Supplementary Table 2). The binding site composition for both 90% and 30% identity datasets revealed that the occupancy of early amino acids is higher in Ancient coenzyme binding sites and tends to decrease in LUCA and Post-LUCA cofactor sites. Overall, for the 90% dataset the average occupancy of early vs. late amino acids in the Ancient sites is 61% vs 39% while this ratio decreases to 53% vs 47% for the LUCA and 47% vs 53% in Post-LUCA sites. These numbers follow the same trend for the 30% identity dataset and throughout the rest of the analysis, the 90% identity dataset – which includes a higher number of proteins – was evaluated for more robust statistical analysis.

To explore the contribution of individual amino acids to this effect, fractional difference (FD) for early vs. late amino acids among the Ancient, LUCA, and Post-LUCA coenzyme binding was calculated (Supplementary Table 5A). The mean FD revealed a similar trend to the amino acid composition analysis (Fig. 3 [↗](#)). The amino acids most enriched in LUCA vs. Post-LUCA are Gly, Ser, and Leu (FD of 4.4, 4.3, and 4.1 respectively), while the most depleted include Phe, Arg, and His (FD of -11, -4.2, and -3.2) (Supplementary Table 5B).

Moreover, we investigated whether the observed trend in amino acid occurrence at the binding sites was dominated by the presence of phosphate groups, which are common in many Ancient cofactors except for SAM, Tetrahydrofolic acid, Biopterin, and Heme. An additional analysis therefore excluded all phosphate-containing coenzymes indicating that while the trend is less pronounced, it remains even in the absence of phosphate groups (Supplementary Table 6).

To examine the impact that the distribution of amino acids in the coenzyme binding sites has on the binding modes, the interactions between coenzymes and individual amino acids were further inspected.

Interaction types between coenzymes and proteins

First, backbone vs side chain protein interactions of all the coenzyme classes were mapped (Fig. 4A [↗](#)). As expected, most of the interactions with coenzymes are mediated by amino acid side chains (61 %), frequently in combination with backbone (24%) (Fig. 4 [↗](#)). Nevertheless, purely

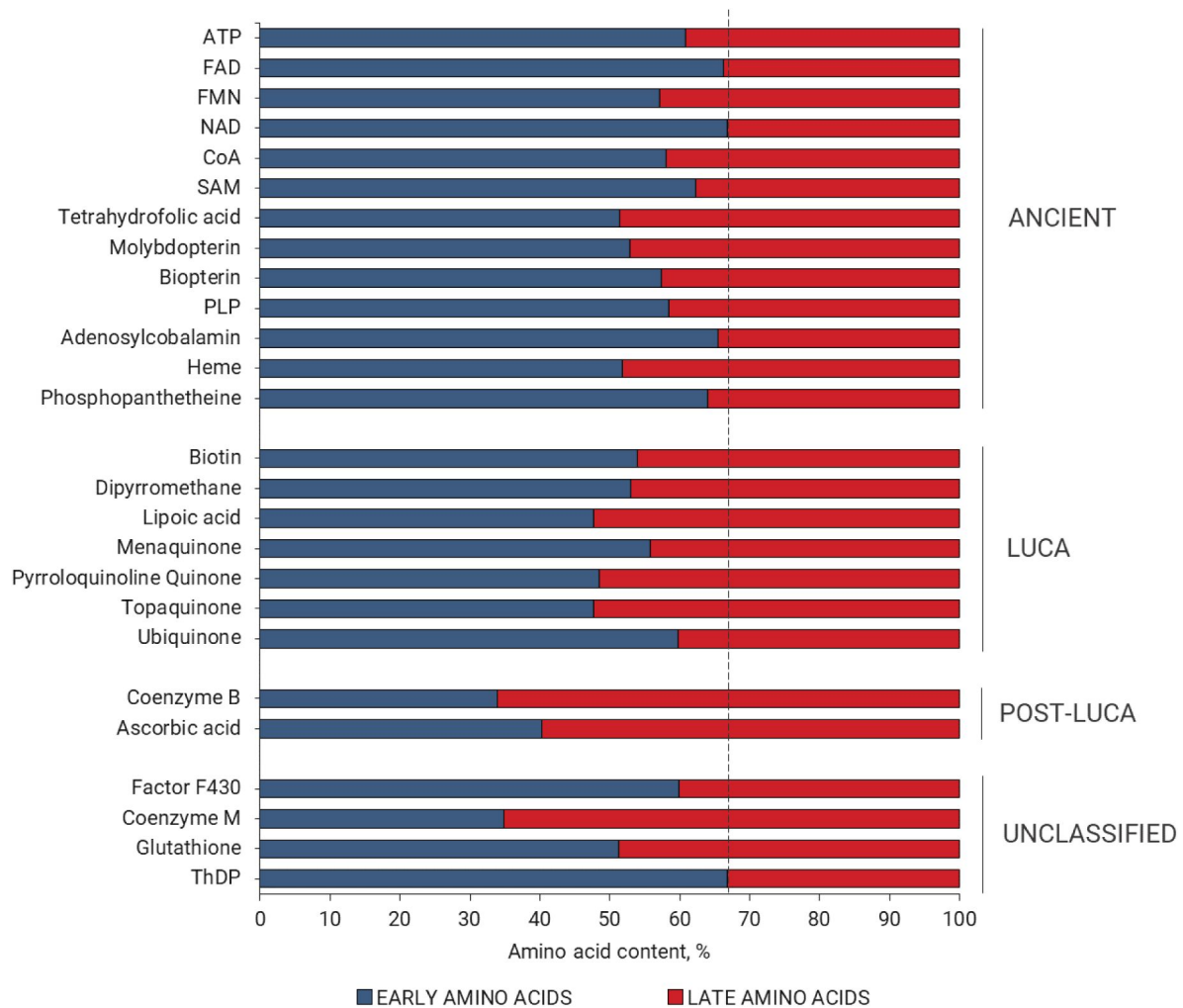


Fig 3

Early versus late amino acid composition of the coenzyme binding sites, categorized according to the evolutionary temporality of coenzymes. Early amino acids are shown in color blue and late residues in red. The dashed line corresponds to the proportion of early vs. late amino acids within the UniProt composition of the sequences derived from our database (67% early and 33% late residues). The statistical significance of the early versus late amino acid composition was assessed by a Chi-squared test ($P < 0.0001$). Detailed statistical data are listed in Supplementary Table 8.

backbone interactions prevail in Ancient coenzymes (24 %) (**Fig. 4A** [↗](#)). When backbone interactions are present throughout the different coenzyme temporalities, they are dominated by the early amino acids (**Fig. 4B** [↗](#)).

Next, we inspected the interaction types for each amino acid-coenzyme binding event employing Arpeggio (Jubb et al., 2017 [↗](#)). The analysis revealed that electrostatic interactions are dominant in all coenzyme ages (Supplementary Fig. 2). In Ancient cofactors, electrostatic interactions are more frequently mediated by early residues. This trend is more significant for the nucleotide-derived Ancient coenzymes. The second most prevalent interaction type is Van der Waals for Ancient cofactors while hydrophobic interactions are similarly frequent as Van der Waals in the LUCA and Post-LUCA classes.

Structural properties of cofactor binding sites

Structural properties of proteins have been observed to change during eons of life's evolution (Edwards et al., 2013 [↗](#); Lupas and Alva, 2017 [↗](#); Kovacs et al., 2017 [↗](#)). To map its interdependence with binding of cofactors, the secondary structure and fold classes of individual coenzyme binding sites were analyzed here.

There are detectable differences in the binding site secondary structure content among the coenzyme temporalities (**Fig. 5** [↗](#), Supplementary Fig. 3). While loops and helices dominate all the binding sites, they are less represented in the Ancient and LUCA coenzyme sites which are more rich in beta-sheet structures (**Fig. 5** [↗](#)). This distinction is found only on the level of the binding sites and not preserved on the level of the overall protein structure.

To explore the fold diversity of domains containing the coenzyme binding sites, we assigned their ECD X-groups (Cheng et al., 2014 [↗](#)) at a residue level (Supplementary Table 3). In total, 101 groups were identified. The Ancient coenzymes are associated with higher numbers of different X-groups than the LUCA and Post-LUCA cofactors (**Fig. 6** [↗](#)). Some coenzymes stand out by their large number of associated folds: the Ancient ATP (74); CoA (34); NAD (30); Heme (27) and the Unclassified cofactor Glutathione (25).

The most frequently observed X-groups include Rossmann-like, Alpha-beta plait, TIM beta/alpha-barrel, Flavodoxin-like, cradle loop barrel, HUP domain-like and beta-Grasp (**Fig. 6** [↗](#)). Among these, Rossmann-like, TIM beta/alpha-barrel and Flavodoxin-like bind to coenzymes of all ages.

Coenzyme early vs late binding sites

To further explore whether extant proteins can bind enzymes only by early or only by late residues (featuring early vs late binding sites), we looked for these specific cases and analyzed their evolutionary conservation.

We found 25 PDB entries that contain at least one chain bound to coenzymes solely by early amino acids (**Fig. 7** [↗](#); Supplementary Fig. 4). Those structures correspond to 17 different proteins, represented by unique UniProt codes. The full set of those proteins bind exclusively Ancient coenzymes: ATP, NAD and Phosphopantetheine. In comparison, 15 PDB entries, representing 12 unique proteins, bind coenzymes only by late amino acids (**Fig. 7** [↗](#); Supplementary Fig. 4). These examples include all Ancient-to-Post-LUCA coenzymes: ATP, CoA, NAD, PLP, biotin, and Ascorbic acid.

To assess the conservation of amino acids in these specific binding sites we used ConSurf (Ashkenazy et al., 2010 [↗](#); Ashkenazy et al., 2016 [↗](#)). According to the analysis, both the early and late binding sites are relatively highly conserved. Around 60% of the residues from both cases have conservation scores ≥ 7 . Furthermore, we employed the MAX AA parameter, that represents

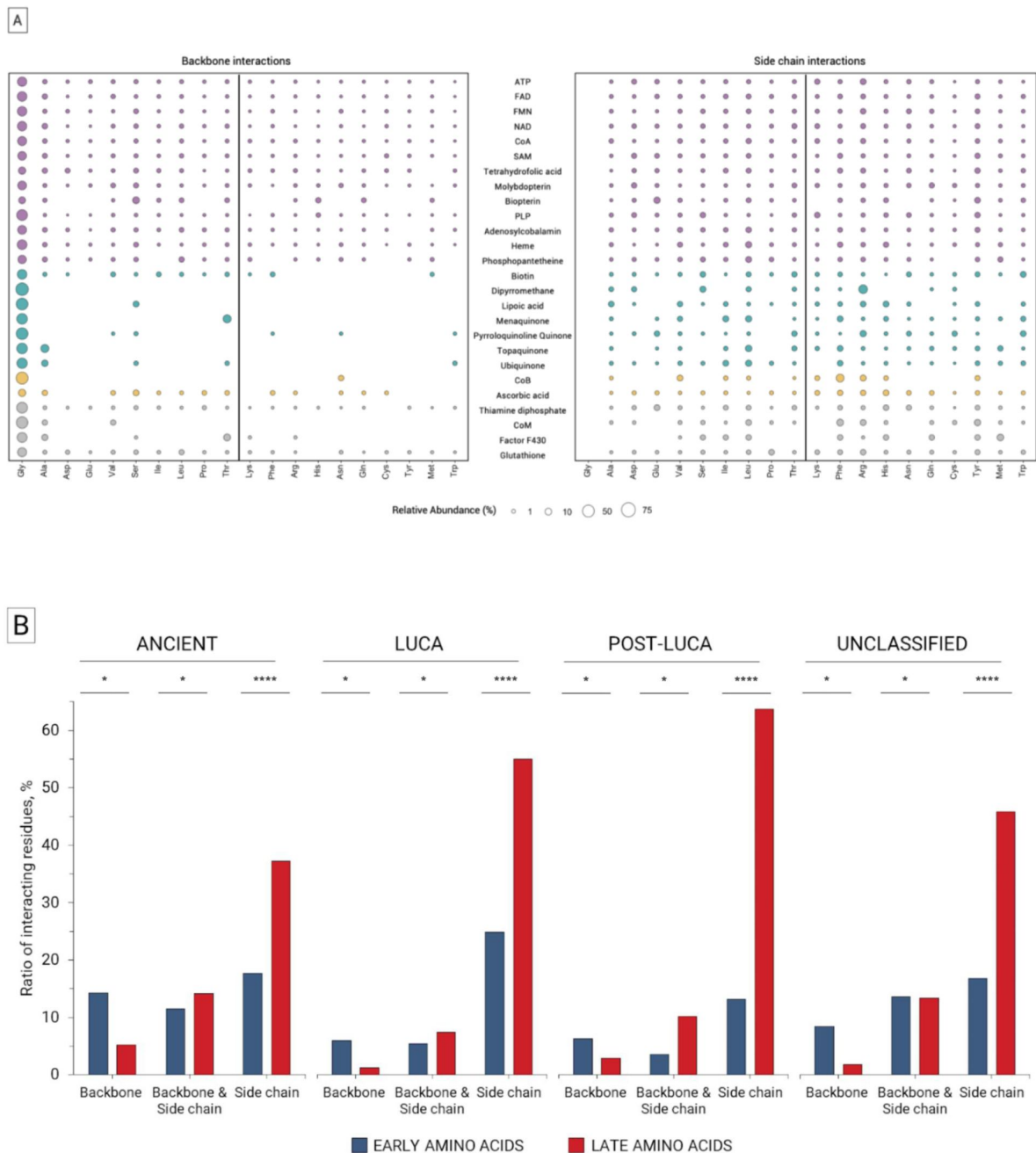


Fig 4.

Binding of coenzymes with early and late amino acids by backbone and side chain atoms. "Backbone" interactions refer to residues in the coenzyme binding sites that interact purely through amino acid backbone atoms. "Side chain" interactions involve residues that interact solely via side chain atoms. "Backbone & Side chain" residues are those that interact with the coenzyme using both their backbone and side chain atoms. (A) Abundance of amino acids in individual studied coenzymes. "Backbone & Side chain" interactions are not depicted. Unclassified cofactors are in gray, Post-LUCA in yellow, LUCA in cyan and Ancient in purple. Amino acids are ranked by the order of addition of amino acids to the genetic code (Higgs and Pudritz, 2009). (B) Proportion of early versus late residues in coenzyme categories by interaction type. In each coenzyme category, the individual proportions add up to 100%. The amino acid composition was normalized by the percentage of late residues from the UniProt sequences retrieved from our database. The statistical significance of early versus late amino acid composition for each interaction type per coenzyme temporality was determined by a Chi-squared test (*, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; ****, $P < 0.0001$). For detailed statistical analysis, refer to Supplementary Table 9.

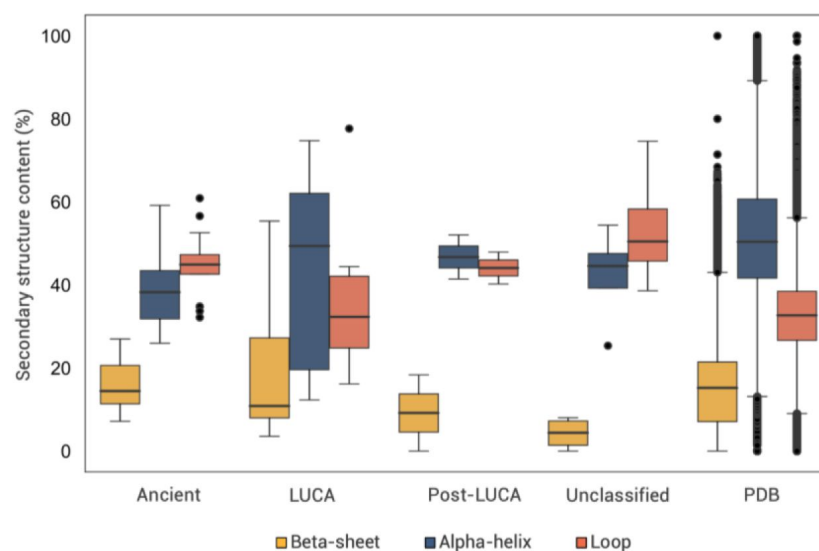


Fig 5.

Secondary structure content in coenzyme binding sites. Composition of secondary structural elements in amino acids interacting with coenzymes. The PDB category represents secondary structure content across the dataset for comparison with coenzyme binding sites. Additional statistical analyses are shown in Supplementary Table 10.

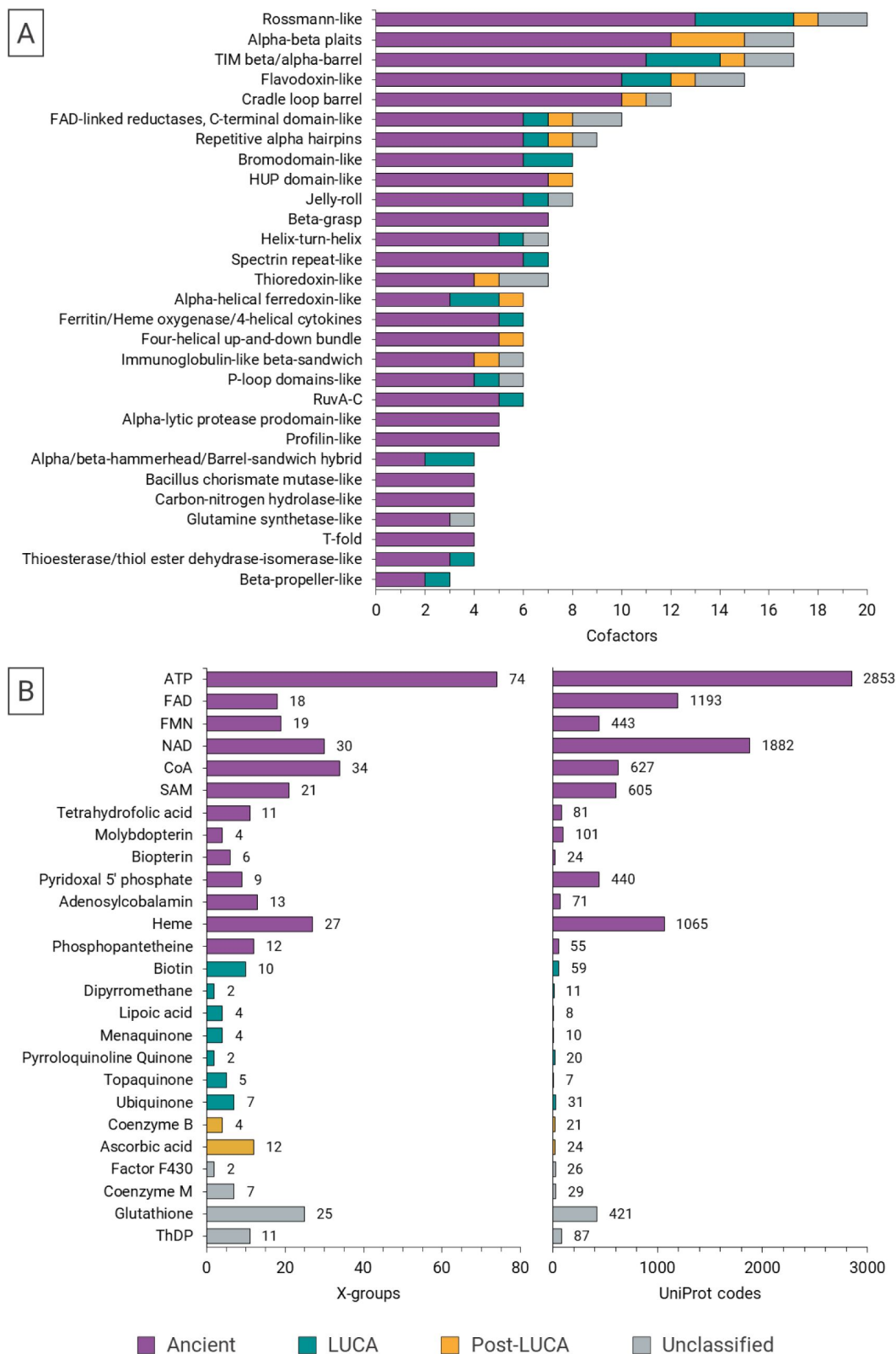


Fig 6.

Fold diversity of coenzyme binding sites. (A) Folds represented by ECOD X-groups, according to numbers of coenzyme binding sites. (B) Comparison of numbers of ECOD X-groups vs. UniProt entries per cofactor class

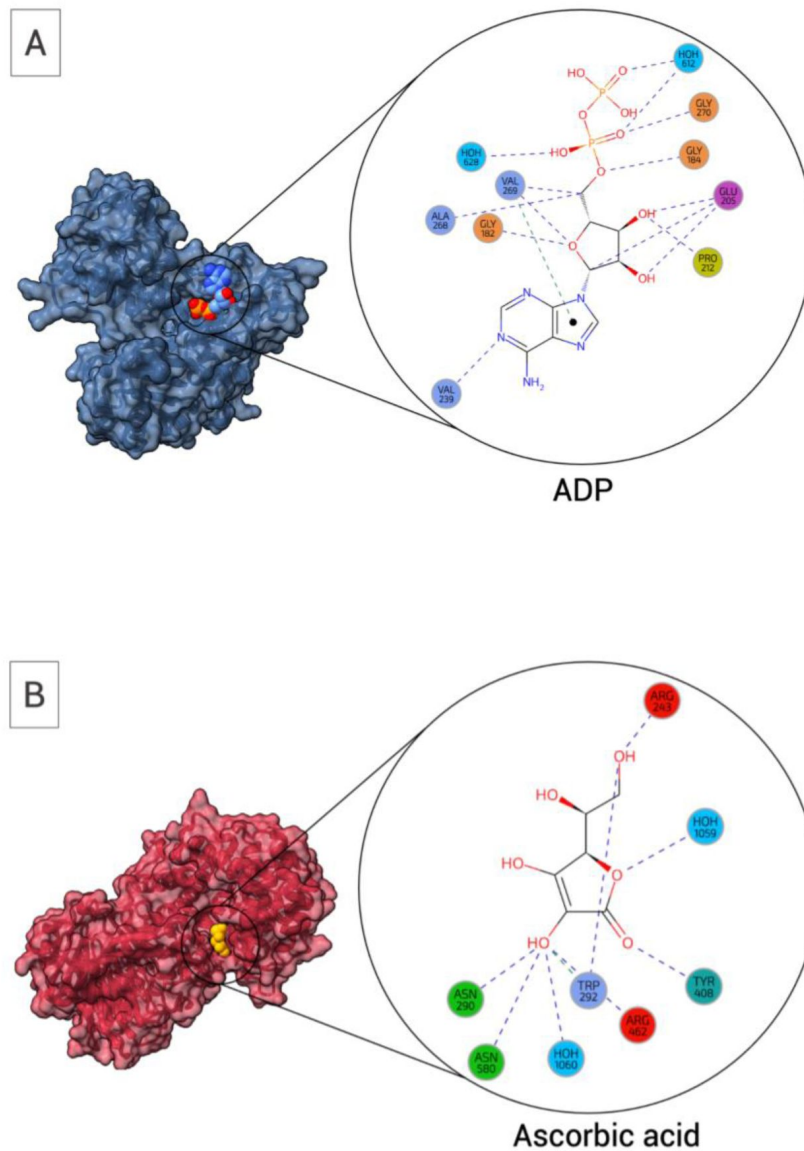


Fig 7.

Examples of coenzyme binding solely through early or late amino acids. (A) Coenzymes bound exclusively by early residues (AMP bound by ATP-phosphoribosyltransferase. PDB code 6czm (chain B) created by LIGPLOT (Laskowski and Swindells, 2011)). (B) Coenzyme, entirely bound by late residues (Ascorbic acid bound by Hyaluronate lyase. PDB code 1f9g (chain A), created by LIGPLOT).

the most abundant residue in the multiple sequence alignment of all homologs. 76 vs 72% of the residues in the early vs late binding sites are the same, which suggests their evolutionary conservation in both cases.

Coenzyme binding mediated by metal ions

Because of the significance of metal ions in both extant and early life, we also analyzed coenzyme binding via metal ions (Supplementary Table 4). Notably, this phenomenon is more frequent in Ancient cofactors, constituting approximately 24% of coenzyme binding sites that have at least one metal ion. Younger cofactors exhibit a lower requirement for metal ion binding. LUCA coenzymes exhibit a metal ion binding in approximately 13% of instances, while in Post-LUCA, 11% and in Unclassified coenzymes, about 23%. Certain coenzymes exhibit a notably high percentage of cases reliant on at least one Mg^{2+} ion (76 % in case of Thiamine diphosphate and 55% in case of ATP binding).

The subsequent most prevalent mediating ion is Ca^{2+} , found along with 65 % cases of the LUCA coenzyme Pyrroloquinoline Quinone. Following Ca^{2+} , the next most frequent metal ions mediating coenzyme binding are Mn^{2+} and Fe^{2+} .

Discussion

Enzymatic activities rely heavily on interplay with organic cofactors. Those are found at the very heart of cellular metabolism and some of them quite possibly branch deep to life's early start (White, 1976 [\[1\]](#)). The core chemistries of the most abundant and Ancient coenzymes have been repeatedly detected in material and experiments mimicking prebiotic environments (Miller and Schlesinger, 1993 [\[2\]](#); Keefe et al., 1995 [\[3\]](#); Holliday et al., 2007 [\[4\]](#); Kirschning, 2021 [\[5\]](#); Menor-Salván et al., 2022 [\[6\]](#); Pinna et al., 2022 [\[7\]](#); Fairchild et al., 2024 [\[8\]](#)). Along with metal ions and minerals, some of the extant coenzymes could probably catalyze metabolic reactions in the absence of enzymes, before their emergence (Muchowska et al. 2020 [\[9\]](#); Henriques Pereira et al. 2022 [\[10\]](#); Cvjetan et al., 2023 [\[11\]](#); Dherbassy et al. 2023 [\[12\]](#)). When and how coenzymes seeded the functional hubs of today's enzymes represents a fundamental bridge between prebiotic chemistry and biochemistry and therefore one of the central questions in the study of life's origins (Preiner et al. 2020 [\[13\]](#)).

The aim of this study was to resolve this conundrum by analyzing protein-coenzyme interactions throughout PDB with respect to amino acid and coenzyme evolutionary temporality (Fig. 2 [\[14\]](#)). While no direct inferences about distant evolutionary past can be drawn from the analysis of extant proteins, the principles guiding these interactions can imply their potential prebiotic feasibility and significance.

Ancient coenzymes are abundant in extant life and bind more frequently to early amino acids

We find that an absolute majority (94 %) of extant coenzymes that appear in PDB structures is conserved across all the life's domains and available already through prebiotic syntheses (i.e. Ancient). This coenzyme temporality bears many molecules that are derived from nucleotides (such as ATP, FAD, SAM, and NAD) and is spread throughout all the E.C. classes of protein catalysts. This is the first obvious distinction from the other (less-populated) classes of coenzymes and supports the coenzyme-peptide significance from the earliest life until today (White, 1976 [\[1\]](#)).

There are several outstanding differences in the properties of binding to proteins among the different coenzyme temporalities. First, the Ancient coenzymes bind to proteins more frequently via early amino acids. On average, this interaction presents 61 % of all Ancient coenzyme bonds to

proteins. It is only 53 and 38 % for the LUCA and Post-LUCA coenzymes, respectively.

While the Ancient nucleotide-derived and Unclassified coenzyme binding sites are dominated by residues in the loop conformation (**Fig. 5** [↗](#); Supplementary Fig. 3), there is also a substantially higher frequency of residues in beta-sheet conformations when compared to Post-LUCA coenzyme sites. Those, on the other hand, are dominated by alpha-helical conformations. Loops exhibit greater sequence variability compared to ordered structures, and their flexible nature enables them to undergo structural changes (Kessel and Ben-Tal, 2018 [↗](#), Tokuriki and Tawfik, 2009 [↗](#)). Such properties of protein active sites were previously associated with evolvability and promiscuity (Tokuriki and Tawfik, 2009 [↗](#), Corbella et al., 2023 [↗](#)). The role of loops could thus be important for the flexibility and versatility of early peptide- coenzyme binding sites. It has been noted that evolutionary benefits would be presented by sequences that could adopt closed loop conformations, providing stability and protection to early coenzymes and such hubs could truly transition the peptide-coenzyme world towards primordial enzymes (Goncareenco and Berezovsky 2011 [↗](#), Gamiz-Arco et al., 2021 [↗](#); Toledo-Patiño et al., 2022 [↗](#); Gutierrez-Rus et al., 2023 [↗](#)). Besides sequences without regular secondary structure elements, beta-sheets have been considered a more prevalent and significant motif during early stages of protein evolution than alpha helices. Beta- sheet represents the first structural motif in models of the ribosomal evolution, and it has also been observed as a mildly enriched motif in sequences formed from early amino acids (Brack and Orgel 1975 [↗](#), Lupas and Alva 2017 [↗](#), Kovacs et al., 2017 [↗](#), Tretyachenko et al., 2022 [↗](#)). Ancient coenzyme-peptide binding properties support the scenario of its significance during early stages of protein evolution.

Ancient coenzymes bind to proteins through more backbone interactions, typically assigned to early amino acids

While all the coenzymes bind preferentially to protein residue sidechains, more backbone interactions appear in the Ancient coenzyme temporality when compared to others. This supports an earlier hypothesis that functions of the earliest peptides (possibly of variable compositions and lengths) would be performed with the assistance of the main chain atoms rather than their sidechains (Milner-White and Russel 2011). A specific example of such a scenario was recently reported, where a dihydrofolate reductase activity was supported purely by protein backbone-coenzyme interactions (Lemay-St-Denis and Pelletier, 2023 [↗](#)). Finally, Longo et al., recently analyzed binding sites of different phosphate- containing ligands which were arguably of high relevance during earliest stages of life, connecting all of today's core metabolism (Longo et al., 2020 [↗](#) (b)). They observed that unlike the evolutionary younger binding motifs (which rely on sidechain binding), the most ancient lineages indeed bind to phosphate moieties predominantly via the protein backbone.

Our analysis assigns this phenomenon primarily to interactions via early amino acids that (as mentioned above) are generally enriched in the binding interface of the Ancient coenzymes. This implies that late amino acids would not be necessarily needed for the sovereignty of coenzyme-peptide interplay. To address this intriguing possibility, we next searched whether there are such examples in the PDB dataset, where coenzymes would be bound exclusively by early amino acids. We found 17 such proteins where all the coenzymes belonged to the Ancient temporality (such as ATP and NAD). Together with all of the above, this finding supports the possibility that peptide-coenzyme functional hubs could have originated before the evolution of the full canonical amino acid alphabet.

Reinforcing this, we have recently demonstrated on a select example of a ribosomal RNA-binding domain, that a negatively charged variant of the protein composed of only early amino acids is indeed capable of binding to RNA (Giacobelli et al., 2022 [↗](#)). In that case, the interaction is further supported by metal ions that were not present at the binding interface of the wild-type protein. Interestingly, the same trend was observed here throughout the PDB dataset. 24% of the Ancient

coenzymes in PDB are additionally mediated by at least one metal ion. LUCA and Post-LUCA coenzymes involved metal ions only in 13 and 11 %, respectively. It is quite probable that not all the metal ion densities are recognized or fully resolved in all the PDB structures that were used in our analysis. Nevertheless, we hypothesize that the overall trend can be attributed to the inherent negative charge of many of the Ancient coenzymes, necessitating engagement of positively charged metal ions. This would also be true for direct interaction of early peptides/proteins and metal ions, independent of organic cofactor involvement, as discussed previously by us and others (Bromberg et al., 2022 [↗](#); Frenkel-Pinter et al., 2020 [↗](#); Fried et al., 2022 [↗](#)). For example, it has been observed that coordination of prebiotically most relevant metal ions (e.g., Mg^{2+}) is more often mediated by early amino acids such as Asp and Glu, whereas metal ions of later relevance (e.g., Cu^+ and Zn^{2+}) bind more frequently via late amino acids like His and Cys (Fried et al. 2022 [↗](#)). Similarly, ancient metal binding folds have been shown to be enriched in early amino acids (Bromberg et al., 2022 [↗](#)). Along with the general adaptive properties of late amino acids in expanding the chemistry space, the late amino acids would supplement the positive charge in the residue side chains (Ilardo and Freeland, 2014 [↗](#)).

Coenzymes could have served as bridges towards protein structural and functional sovereignty from the peptide-nucleotide world

Our study further revealed that Ancient coenzymes stand out in the variety of protein structures that they bind to, as represented by the ECOD X-groups. While this may partly be caused by their general over-representation in our dataset, the significance of coenzymes has been pointed out previously throughout the most ancient protein folds, such as P-loop NTPases, TIM beta/alpha-barrels, OB and Rossmann folds (Caetano-Anollés et al., 2017 [↗](#); Goldman et al. 2013 [↗](#); Longo et al., 2020(a); Longo et al., 2020(b)). It has been implied by computer simulations that coenzymes could bind to proteins with similar propensity even before the onset of protein homochirality, despite lower structural stability and secondary structure content in heterochiral polypeptides (Skolnick et al., 2019 [↗](#)).

Phosphate-containing coenzymes truly stand out by their large number of associated folds (e.g. 74 different ECOD X-groups in case of ATP). It has been postulated that phosphate-binding loops served as the most significant precursors for contemporary enzymes (Romero Romero et al., 2018 [↗](#)). Combining ancestral sequence reconstruction and selection/protein design, short polypeptide motifs capable of poly/nucleotide binding have been recovered from the P-loop NTPase and HhH motifs, relying primarily on early amino acids (Longo et al., 2020 [↗](#) (b); Romero Romero et al., 2018 [↗](#)). Both demonstrated that these ancient motifs are highly robust to sequence variations, implying that such interaction can be encountered more easily than previously thought (Longo et al., 2020 [↗](#) (b); Keefe and Szostak, 2001 [↗](#)). Provokingly, it has also been implied that many of these motifs emerged initially as polynucleotide binders and started serving catalysis only after gaining higher structural complexity (Romero Romero et al., 2018 [↗](#)).

Our entry search of PDB coenzymes also retrieved 80 RNA structures. While these were not the primary subject of our analysis, the majority of those were found to interact with Ancient coenzymes, mainly the nucleotide derived ones, such as ATP, FMN, NAD, and SAM (Supplementary Table 7). While some of these structures are assigned as riboswitches, other coenzyme-RNA complexes belong to ribozymes, representing the potential of polynucleotides in early catalysis as discussed by many previous studies (White, 1976 [↗](#); Gilbert, 1986 [↗](#); Reyes-Prieto et al., 2012 [↗](#); Goldman and Kacar 2021 [↗](#)). If peptide- polynucleotide interactions were initially more feasible and dominant (in a putative peptide- polynucleotide world, as implied above), coenzymes could have played a key role in resolving the sovereignty of these molecules towards their tertiary structures and catalytic functions. Polypeptide- coenzyme catalysts would soon dominate in

performance (enabling more efficient catalysis) and functional repertoires, especially those that would be hard to facilitate in their absence, such as oxidoreductases and transferases (Fried et al., 2022 [↗](#); Goldman and Kacar 2021 [↗](#); Kessel and Ben-Tal 2018 [↗](#)).

Limitations of our work

The first obvious drawback of this analysis is the ambiguity that accompanies the division of coenzymes into evolutionary age temporalities. While the prebiotic availability (or lack of) is quite consensual in some cases, there are also contradictory studies and opinions in the case of some other coenzymes. Some of the coenzymes have e.g. prebiotic precursors but are not present across all the kingdoms of life. This may suggest that the coenzyme became important only Post-LUCA but it can also mean that its importance was only preserved in specific branches of life. “Unclassified” category has been included for such specific cases (presented e.g. by glutathione). Several properties of glutathione that were identified here (such as protein backbone vs. sidechain interaction, ubiquity across all E.C. enzyme classes and a high number of associated X-groups) would suggest that it is closer to Ancient coenzymes. The other “classified” coenzymes were categorized based on prevailing studies although some ambiguity remains. Despite our effort, the evolutionary ratings of coenzymes (and amino acids) are therefore not always clear cut and not all belong to the categories with the same weight (e.g. some are likely to be more ancient than others). For example, the nucleotide-derived coenzymes probably predate the others of the Ancient temporality - it has been proposed that S-adenosylmethionine emerged before the more complex heme-related porphyrin adenosylcobalamin or coenzyme B12 functionalities (Lazcano, 2013 [↗](#)).

Another possible bias of our study stems from the population differences among the three coenzyme temporalities. The Ancient coenzymes are by far the most abundant temporality in the PDB dataset. It can be argued that this is the result of their ~4 billion years of essentiality to life (Goldman and Kacar 2021 [↗](#)). Nevertheless, it may as well be contributed by the bias of structures that are deposited in the PDB and most probably does not reflect the true distribution of coenzymes in the biological protein space. Additionally, the comparison of differentially populated temporalities has challenged some aspects of the analysis presented here although care has been taken to perform all the appropriate statistical tests.

Conclusions

The findings presented here propose that early (the less complex and prebiotically plausible) amino acids are sufficient for binding to Ancient coenzymes. Consequently, coenzyme-peptide interactions might have been conceivable at a time when the amino acid alphabet was not yet evolved to its current form. In such interactions, binding modes would rely more on the protein backbone atoms and on involvement of metal ions, both of which are less frequent in interactions with evolutionarily young coenzymes.

Methodology

Identification of organic cofactors

We systematically identified all available cofactor and cofactor-like molecules in the Protein Data Bank in Europe (wwPDB Consortium, 2020) programmatically through the PDBe REST API (pdbe.org/api) (Mukhopadhyay et al., 2019 [↗](#)). All cofactor molecules were classified into 27 classes based on the CoFactor database (Fischer et al., 2010 [↗](#)). Furthermore, we included ATP and its analogs as an additional cofactor class.

The identification of all the available ligand codes from the PDB chemical component dictionary for each cofactor class was achieved by programmatic access through the “Cofactors” endpoint (<https://www.ebi.ac.uk/pdbe/api/pdb/compound/cofactors> [↗](#)) using the PDBe REST API (pdbe.org/api)

and all responses were in JSON format.

Structural database and classifications

We retrieved the PDB entries associated to each chemical component from every cofactor class using the “PDB entries containing the compound” endpoint (https://www.ebi.ac.uk/pdbe/api/pdb/compound/in_pdb/id) via the Entry-based API. The count of PDB entries for each cofactor class is provided in the supplementary information (Supplementary file 1). The information from the REST API was unavailable for two coenzyme classes, MIO and Orthoquinone, so they were excluded from the analysis.

The secondary structure assignments and (EC) numbers for all PDB structures analyzed were determined through residue-level cross references obtained from the SIFTS XML files (Velankar et al., 2013; Dana et al., 2019). Secondary structure elements include “h” for helix, “b” for strand, and “c” for coil; and they correspond to the information available in the PDB website. Only observed residues were examined.

Furthermore, we assigned all our PDB entries to the ECOD hierarchical system groups “X”, “H” and “F” (Cheng et al., 2014).

UniProt assignment and interaction ratio

The assignment of UniProt codes to our structural dataset was achieved by mapping the information with the SIFTS (Velankar et al., 2013; Dana et al., 2019) file “pdb_chain_uniprot.tsv.gz” (<https://www.ebi.ac.uk/pdbe/docs/sifts/quick.html>). Next, we mapped the UniProt residue to each of the PDB structures with the residue-level cross-reference data of SIFTS by retrieving the XML files (<https://www.ebi.ac.uk/pdbe/docs/sifts/quick.html>).

Each UniProt code represents a unique protein sequence that encompasses one or various PDB associated entries. With the aim of filtering those residues relevant to the interaction sites at the level of protein sequence, we incorporated the interaction ratio. The interaction ratio is a measure of the interaction for each ligand with all its PDB associated entries by UniProt residue. Those residue-ligand interactions that were preserved in more than 50% of the associated PDB entries were selected (we call the ratio of preserved interactions among structures of one unique protein interaction ratio).

Upon UniProt residues assignment for each residue of the PDB structures, we downloaded the calculated interaction ratio with the endpoint “UniProt- Get ligand binding residues for a UniProt accession” (https://www.ebi.ac.uk/pdbe/graph-api/uniprot/ligand_sites/:accession).

The redundancy of our database was removed by clustering the UniProt sequences using CD- HIT (Li and Godzik, 2006) with a 90% sequence identity parameter.

Analysis of coenzyme interactions

In order to analyze the amino acid-coenzyme interactions, we downloaded information of all bound molecules found in a given PDB entry using the “Get bound molecules” endpoint (https://www.ebi.ac.uk/pdbe/graph-api/pdb/bound_molecule_interactions/:pdbId/:bmid) of the Aggregated API. Then, we retrieved the ligand interactions for each bound molecule in every entry through the “PDB-Get bound ligand interactions” endpoint (https://www.ebi.ac.uk/pdbe/graph-api/pdb/bound_ligand_interactions/:pdbId/:chain/:seqId), which calculates these interactions with Arpeggio (Jubb et al., 2017). The retrieved interactions included the standard amino acid codes, water molecules and metal ions.

We classified all the interactions reported by Arpeggio into nine distinct interaction types. The classification scheme aligns with the one used by PDBe and encompasses the following categories: i) “covalent”; ii) “electrostatic”, which combines “ionic”, “hbond”, “weak_hbond”, “polar”, “weak_polar”, “xbond” and “carbonyl”; iii) “amide”, consisting in “AMIDEAMIDE” and “AMIDERING”; iv) “vdw”, denoting van der Waals interactions; v) “hydrophobic”; vi) “aromatic”, grouping “aromatic”, “FF”, “OF”, “EE”, “FT”, “OT”, “ET”, “FE”, “OE” and “EF” contacts; vii) “atom-pi”, comprised of “CARBONPI”, “CATIONPI”, “DONORPI”, “HALOGENPI”, and “METSULPHURPI”; viii) “metal” and ix) “clashes”, including “clash” and “vdw_clash” contacts. We have omitted this last category due to the limited number of interactions, most of which result from experimental errors during X-ray diffraction.

Backbone and side chains interactions were identified based on the atom identities in the coenzyme binding sites. Those atoms corresponding to the backbone of standard amino acids were identified as: “N”, “C”, “CA”, “O”. Glycine has only a hydrogen atom as its side chain; nevertheless, no side chain atom mediating any interaction was identified.

Secondary structure analysis

Statistical analysis of secondary structure content was conducted at the UniProt level. For each residue within every UniProt entry, we considered all potential secondary structure elements derived from the PDB structures associated with each UniProt code. Subsequently, we eliminated redundancy on a per-residue basis. This methodological approach enabled us to comprehensively encompass the structural diversity at each position of the protein.

Interactions mediated exclusively by early or late amino acids

To examine proteins that interacted with cofactors solely through early or late amino acids, we filtered the data to include only proteins interacting with at least two amino acids.

For the assessment of the evolutionary conservation of coenzyme-binding amino acids, we employed ConSurf (Ashkenazy et al., 2010 [DOI](#); Ashkenazy et al., 2016 [DOI](#)). Specifically, we analyzed the `msa_positional_aa_frequency` files generated for each PDB structure.

Acknowledgements

This work was supported by the Human Frontier Science Program grant HFSP-RGEC27/2023 and was carried out with the support of ELIXIR CZ Research Infrastructure (ID LM2023055, MEYS CR). A.C.S.R. and M.M. acknowledge support by the project the “Grant Schemes at CU” (reg. no. CZ.02.2.69/0.0/0.0/19_073/0016935), project no. START/SCI/148. Finally, we would like to thank Prof. Stephen Freeland and Prof. Janet Thornton for helpful discussions on this manuscript.

References

- Alva V, Söding J, Lupas AN (2015) **A vocabulary of ancient peptides at the origin of folded proteins** *eLife* **4** <https://doi.org/10.7554/eLife.09410>
- Aylward N (2006) **An ab initio computational study of thiamin synthesis from gaseous reactants of the interstellar medium** *Biophysical Chemistry* **121**:185–193 <https://doi.org/10.1016/j.bpc.2005.12.018>
- Aylward N, Bofinger N (2006) **A plausible prebiotic synthesis of pyrdoxal phosphate: Vitamin B-6 - A computational study** *Biophysical Chemistry* **123**:113–121 <https://doi.org/10.1016/j.bpc.2006.04.014>
- Ashkenazy H, Erez E, Martz E, Pupko T, Ben-Tal N (2010) **ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids** *Nucleic Acids Res* **38**:W529–33 <https://doi.org/10.1093/nar/gkq399>
- Ashkenazy H, Abadi S, Martz E, Chay O, Mayrose I, Pupko T, Ben-Tal N (2016) **ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules** *Nucleic Acids Res* **44**:W344–50 <https://doi.org/10.1093/nar/gkw408>
- Barrick JE, Breaker R (2007) **The distributions, mechanisms, and structures of metabolite-binding riboswitches** *Genome Biology* **8** <https://doi.org/10.1186/gb-2007-8-11-r239>
- Bonfio C, Valer L, Scintilla S, Shah S, Evans D, Jin L, Szostak J, Sassellov D, Sutherland J, Mansy S (2017) **UV-light-driven prebiotic synthesis of iron-sulfur clusters** *Nature Chemistry* **9**:1229–1234 <https://doi.org/10.1038/nchem.2817>
- Burton AS, Stern JC, Elsila JE, Glavin DP, Dworkin JP (2012) **Understanding prebiotic chemistry through the analysis of extraterrestrial amino acids and nucleobases in meteorites** *Chemical Society reviews* **41**:5459–72 <https://doi.org/10.1039/c2cs35109a>
- Brack A, Orgel LE (1975) **Beta structures of alternating polypeptides and their possible prebiotic significance** *Nature* **256**:383–7 <https://doi.org/10.1038/256383a0>
- Bromberg Y, Aptekmann AA, Mahlich Y, Cook L, Senn S, Miller M, Nanda V, Ferreira DU, Falkowski PG (2022) **Quantifying structural relationships of metal-binding sites suggests origins of biological electron transfer** *Science Advances* **8** <https://doi.org/10.1126/sciadv.abj3984>
- Caetano-Anollés G., Hee SK, Mitterthal JE (2007) **The origin of modern metabolic networks inferred from phylogenomic analysis of protein architecture** *Proceedings of the National Academy of Sciences of the United States of America* **104**:9358–9363 <https://doi.org/10.1073/pnas.0701214104>
- Cheng H, Schaeffer RD, Liao Y, Kinch LN, Pei J, Shi S, Kim BH, Grishin NV (2014) **ECOD: an evolutionary classification of protein domains** *PLoS Computational Biol* **10** <https://doi.org/10.1371/journal.pcbi.1003926>

Chu XY, Zhang HY. (2020) **Cofactors as molecular fossils to trace the origin and evolution of proteins** *ChemBioChem* **21**:3161–3168

Cvjetan N, Schuler L, Ishikawa T, Walde P (2023) **Optimization and Enhancement of the Peroxidase-like Activity of Hemin in Aqueous Solutions of Sodium Dodecylsulfate** *ACS Omega* **8**:42878–42899 <https://doi.org/10.1021/acsomega.3c05915>

Cleaves HJ (2010) **The origin of the biologically coded amino acids** *Journal of Theoretical Biology* **263**:490–498 <https://doi.org/10.1016/j.jtbi.2009.12.014>

Copley SD, Dhillon JK (2002) **Lateral gene transfer and parallel evolution in the history of glutathione biosynthesis genes** *Genome Biology* **3**:1–16 <https://doi.org/10.1186/gb-2002-3-5-research0025>

Corbella M, Pinto GP, Kamerlin SCL (2023) **Loop dynamics and the evolution of enzyme activity** *Nature Reviews Chemistry* **7**:536–547 <https://doi.org/10.1038/s41570-023-00495-w>

Dana JM, Gutmanas A, Tyagi N, Qi G, O'Donovan C, Martin M, Velankar S (2019) **SIFTS: updated Structure Integration with Function, Taxonomy and Sequences resource allows 40-fold increase in coverage of structure-based annotations for proteins** *Nucleic Acids Res* **47**:D482–D489 <https://doi.org/10.1093/nar/gky1114>

Dherbassy Q, Mayer R, Muchowska K, Moran J (2023) **Metal-Pyridoxal Cooperativity in Nonenzymatic Transamination** *Journal of the American Chemical Society* **145**:13357–13370 <https://doi.org/10.1021/jacs.3c03542>

Edwards H, Abeln S, Deane CM (2013) **Exploring Fold Space Preferences of New-born and Ancient Protein Superfamilies** *PLoS Computational Biology* **9** <https://doi.org/10.1371/journal.pcbi.1003325>

Fairchild J, Islam S, Singh J, Powner MW (2024) **Prebiotically plausible chemoselective pantetheine synthesis in water** *Science* **383**:911–918 <https://doi.org/10.1126/science.adk4432>

Fischer JD, Holliday GL, Thornton JM (2010) **The CoFactor database: Organic cofactors in enzyme catalysis** *Bioinformatics* **26**:2496–2497 <https://doi.org/10.1093/bioinformatics/btq442>

Frenkel-Pinter M, Haynes JW, Martin C, Petrov AS, Burcar BT, Krishnamurthy R, Hud N, Leman L, Williams LD (2019) **Selective incorporation of proteinaceous over nonproteinaceous cationic amino acids in model prebiotic oligomerization reactions** *Proceedings of the National Academy of Sciences of the United States of America* **116**:16338–16346 <https://doi.org/10.1073/pnas.1904849116>

Frenkel-Pinter M, Mousumi S, Ashkenasy G, Leman L (2020) **Prebiotic Peptides: Molecular Hubs in the Origin of Life** *Chemical Reviews* **120**:4707–4765 <https://doi.org/10.1021/acs.chemrev.9b00664>

Fried SD, Fujishima K, Makarov M, Cherepashuk I, Hlouchova K (2022) **Peptides before and during the nucleotide world: An origins story emphasizing cooperation between proteins and nucleic acids** *Journal of the Royal Society Interface* **19** <https://doi.org/10.1098/rsif.2021.0641>

Gamiz-Arco G *et al.* (2021) **Heme-binding enables allosteric modulation in an ancient TIM-barrel glycosidase** *Nature Communications* **12** <https://doi.org/10.1038/s41467-020-20630-1>

Giacobelli VG, Fujishima K, Lepšík M, Tretiyachenko V, Kadavá T, Makarov M, Hlouchová K (2022) **In Vitro Evolution Reveals Noncationic Protein-RNA Interaction Mediated by Metal Ions** *Molecular Biology and Evolution* **39**:1–11 <https://doi.org/10.1093/molbev/msac032>

Gilbert W (1986) **The RNA world superlattices point ahead** *Nature* **319**

Goldman AD, Bernhard TM, Dolzhenko E, Landweber LF (2013) **LUCApedia: A database for the study of ancient life** *Nucleic Acids Res* **41**:1079–1082 <https://doi.org/10.1093/nar/gks1217>

Goldman AD, Kacar B (2021) **Cofactors are Remnants of Life's Origin and Early Evolution** *Journal of Molecular Evolution* **89**:127–133 <https://doi.org/10.1007/s00239-020-09988-4>

Goncearenco A, Berezovsky IN (2011) **Prototypes of elementary functional loops unravel evolutionary connections between protein functions** *Bioinformatics* **27**:i497–i503 <https://doi.org/10.1093/bioinformatics/btq374>

Gutierrez-Rus LI, Gamiz-Arco G, Gavira JA, Gaucher EA, Risso VA, Sanchez-Ruiz JM (2023) **Protection of Catalytic Cofactors by Polypeptides as a Driver for the Emergence of Primordial Enzymes** *Molecular Biology and Evolution* **40**:1–8 <https://doi.org/10.1093/molbev/msad126>

Henriques DP, Leethaus J, Beyazay T, do Nascimento A, Kleineremanns K, Tüysüz H, Martin W, Preiner M. (2022) **Role of geochemical protoenzymes (geozymes) in primordial metabolism: specific abiotic hydride transfer by metals to the biological redox cofactor NAD⁺** *FEBS Journal* **289** <https://doi.org/10.1111/febs.16329>

Higgs PG, Pudritz RE (2009) **A thermodynamic basis for prebiotic amino acid synthesis and the nature of the first genetic code** *Astrobiology* **9**:483–490 <https://doi.org/10.1089/ast.2008.0280>

Holliday GL, Thornton JM, Marquet A, Smith AG, Rébeillé F, Mendel R, Schubert HL, Lawrence AD, Warren MJ (2007) **Evolution of enzymes and pathways for the biosynthesis of cofactors** *Natural Product Reports* **24**:972–987 <https://doi.org/10.1039/b703107f>

Huang F, Bugg CW, Yarus M. (2000) **RNA-Catalyzed CoA, NAD, and FAD synthesis from phosphopantetheine, NMN, and FMN** *Biochemistry* **39**:15548–15555 <https://doi.org/10.1021/bi002061f>

Ilardo MA, Freeland SJ (2014) **Testing for adaptive signatures of amino acid alphabet evolution using chemistry space** *Journal of Systems Chemistry* **5**:1–9 <https://doi.org/10.1186/1759-2208-5-1>

Ji HF, Chen L, Zhang HY (2008) **Organic cofactors participated more frequently than transition metals in redox reactions of primitive proteins** *BioEssays* **30**:766–771 <https://doi.org/10.1002/bies.20788>

Jubb HC, Higuero AP, Ochoa-Montano B, Pitt WR, Ascher DB, Blundell TL (2017) **Arpeggio: A Web Server for Calculating and Visualising Interatomic Interactions in Protein Structures** *J Mol Biol* **429**:365–371 <https://doi.org/10.1016/j.jmb.2016.12.004>

Keefe AD, Newton GL, Miller SL (1995) **A possible prebiotic synthesis of pantetheine, a precursor to coenzyme a** *Nature* **373**:683–685 <https://doi.org/10.1038/373683a0>

Keefe AD, Szostak JW (2001) **Functional proteins from a random-sequence library** *Nature* **410**:715–8 <https://doi.org/10.1038/35070613>

Kessel A, Ben-Tal N (2018) **Introduction to proteins: structure, function, and motion** *Crc Press (Taylor & Francis Group)* <https://doi.org/10.1201/9781315113876>

Kessel A, Ben-Tal N. (2022) **From Molecules to Cells: The Origin of Life on Earth**

Kirschning A (2021) **Coenzymes and Their Role in the Evolution of Life** *Angewandte Chemie - International Edition* **60**:6242–6269 <https://doi.org/10.1002/anie.201914786>

Kirschning A (2022) **On the Evolutionary History of the Twenty Encoded Amino Acids** *Chemistry - A European Journal* **28** <https://doi.org/10.1002/chem.202201419>

Kolodny R, Nepomnyachiy S, Tawfik DS, Ben-Tal N (2021) **Bridging Themes: Short Protein Segments Found in Different Architectures** *Molecular Biology and Evolution* **38**:2191–2208 <https://doi.org/10.1093/molbev/msab017>

Kovacs NA, Petrov AS, Lanier KA, Williams LD (2017) **Frozen in Time: The History of Proteins** *Molecular Biology and Evolution* **34**:1252–1260 <https://doi.org/10.1093/molbev/msx086>

Lane N, Martin WF (2012) **The origin of membrane bioenergetics** *Cell* **151**:1406–1416 <https://doi.org/10.1016/j.cell.2012.11.050>

Laurino P, Tóth-Petróczy Á, Meana-Pañeda Lin W, Truhlar DG, Tawfik DS (2016) **An Ancient Fingerprint Indicates the Common Ancestry of Rossmann-Fold Enzymes Utilizing Different Ribose- Based Cofactors** *PLoS Biology* **14**:1–23 <https://doi.org/10.1371/journal.pbio.1002396>

Laskowski RA, Swindells MB (2011) **LigPlot+: multiple ligand-protein interaction diagrams for drug discovery** *J Chem Inf Model* **51**:2778–86 <https://doi.org/10.1021/ci200227u>

Lazcano A (2013) **Planetary change and biochemical adaptation: Molecular evolution of corrinoid and heme biosyntheses** *Hematology* **17**:s7–s10 <https://doi.org/10.1179/102453312X13336169155015>

Lemay-St-Denis C, Pelletier J (2023) **From a binding module to essential catalytic activity: how nature stumbled on a good thing** *Chem. Commun* **59**:12560–12572 <https://doi.org/10.1039/D3CC04209J>

Li W, Godzik A (2006) **Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences** *Bioinformatics* **22**:1658–9 <https://doi.org/10.1093/bioinformatics/btl158>

- a. Longo LM, Jabtoiiska J, Vyas P, Kanade M, Kolodny R, Ben-Tal N, Tawfik DS. (2020) **On the emergence of p-loop ntpase and rossmann enzymes from a beta-alpha-beta ancestral fragment** *Elife* **9**:1–16 <https://doi.org/10.7554/ELIFE.64415>
- b. Longo LM, Petrovic D, Kamerlin SCL, Tawfik DS. (2020) **Short and simple sequences favored the emergence of N-helix phospho-ligand binding sites in the first enzymes** *Proceedings of the National Academy of Sciences of the United States of America* **117**:5310–5318 <https://doi.org/10.1073/pnas.1911742117>

- Lupas AN, Alva V (2017) **Ribosomal proteins as documents of the transition from unstructured (poly)peptides to folded proteins** *Journal of Structural Biology* **198**:74–81 <https://doi.org/10.1016/j.jsb.2017.04.007>
- Makarov M, Meng J, Tretyachenko V, Srb P, Březinová A, Giacobelli VG, Bednářová L, Vondrášek J, Dunker K, Hlouchová K (2021) **Enzyme catalysis prior to aromatic residues: Reverse engineering of a dephospho-CoA kinase** *Protein Science* **30**:1022–1034 <https://doi.org/10.1002/pro.4068>
- Menor-Salván C, Burcar BT, Bouza M, Fialho DM, Fernández FM, Hud NV (2022) **A Shared Prebiotic Formation of Neopterins and Guanine Nucleosides from Pyrimidine Bases** *Chemistry (Weinheim an Der Bergstrasse, Germany)* **28** <https://doi.org/10.1002/chem.202200714>
- Miller SL, Schlesinger G (1993) **Prebiotic syntheses of vitamin coenzymes: I. Cysteamine and 2- mercaptoethanesulfonic acid (coenzyme M)** *Journal of Molecular Evolution* **36**:302–7 <https://doi.org/10.1007/BF00182177>
- Milner-White EJ, Russell MJ (2011) **Functional capabilities of the earliest peptides and the emergence of life** *Genes* **2**:671–88 <https://doi.org/10.3390/genes2040671>
- Monteverde DR, Gómez-Consarnau L, Suffridge C, Sañudo-Wilhelmy SA (2017) **Life’s utilization of B vitamins on early Earth** *Geobiology* **15**:3–18 <https://doi.org/10.1111/gbi.12202>
- Muchowska KB, Varma SJ, Moran J (2020) **Nonenzymatic Metabolic Reactions and Life’s Origins** *Chemical Reviews* **120**:7708–7744 <https://doi.org/10.1021/acs.chemrev.0c00191>
- Mukhopadhyay A, Borkakoti N, Pravda L, Tyzack JD, Thornton JM, Velankar S (2019) **Finding enzyme cofactors in Protein Data Bank** *Bioinformatics* **35**:3510–3511 <https://doi.org/10.1093/bioinformatics/btz115>
- Naraoka H *et al.* (2023) **Soluble organic molecules in samples of the carbonaceous asteroid (162173) Ryugu** *Science* **379** <https://doi.org/10.1126/science.abn9033>
- Narunsky A, Kessel A, Solan R, Alva V, Kolodny R, Ben-Tal N (2020) **On the evolution of protein- adenine binding** *Proceedings of the National Academy of Sciences of the United States of America* **117**:4701–4709 <https://doi.org/10.1073/pnas.1911349117>
- Nepomnyachiy S, Ben-Tal N, Kolodny R (2017) **Complex evolutionary footprints revealed in an analysis of reused protein segments of diverse lengths** *Proceedings of the National Academy of Sciences of the United States of America* **114**:11703–11708 <https://doi.org/10.1073/pnas.1707642114>
- PDBE-KB consortium (2020) **PDBE-KB: a community-driven resource for structural and functional annotations** *Nucleic Acids Res* **48**:D344–D353 <https://doi.org/10.1093/nar/gkz853>
- Pinna S, Kunz C, Halpern A, Harrison SA, Jordan SF, Ward J, Werner F, Lane N (2022) **A prebiotic basis for ATP as the universal energy currency** *PLoS Biology* **20**:1–25 <https://doi.org/10.1371/journal.pbio.3001437>
- Putignano V, Rosato A, Banci L, Andreini C (2018) **MetalPDB in 2018: A database of metal sites in biological macromolecular structures** *Nucleic Acids Res* **46**:D459–D464 <https://doi.org/10.1093/nar/gkx989>

Preiner M *et al.* (2020) **The future of origin of life research: Bridging decades-old divisions** *Life* **10** <https://doi.org/10.3390/life10030020>

Qiu K, Ben-Tal N, Kolodny R (2022) **Similar protein segments shared between domains of different evolutionary lineages** *Protein Science* **31** <https://doi.org/10.1002/pro.4407>

Reyes-Prieto F, Hernández-Morales R, Jácome R, Becerra A, Lazcano A (2012) **Coenzymes, viruses and the RNA world** *Biochimie* **94**:1467–1473 <https://doi.org/10.1016/j.biochi.2012.01.004>

Romero Romero ML *et al.* (2018) **Simple yet functional phosphate-loop proteins** *Proceedings of the National Academy of Sciences of the United States of America* **115**:E11943–E11950 <https://doi.org/10.1073/pnas.1812400115>

Russell MJ, Hall AJ (1997) **The emergence of life from iron monosulphide bubbles at a submarine hydrothermal redox and pH front** *J Geol Soc London* **154**:377–402 <https://doi.org/10.1144/gsjgs.154.3.0377>

Seitz C, Eisenreich W, Huber C (2021) **The abiotic formation of pyrrole under volcanic, hydrothermal conditions—an initial step towards life’s first breath?** *Life* **11**:1–10 <https://doi.org/10.3390/life11090980>

Skolnick J, Zhou H, Gao M (2019) **On the possible origin of protein homochirality, structure, and biochemical function** *Proceedings of the National Academy of Sciences of the United States of America* **116**:26571–26579 <https://doi.org/10.1073/pnas.1908241116>

Söding J, Lupas AN (2003) **More than the sum of their parts: On the evolution of proteins from peptides** *BioEssays* **25**:837–846 <https://doi.org/10.1002/bies.10321>

Thauer RK, Bonacker LG (1994) **Biosynthesis of coenzyme F430, a nickel porphinoid involved in methanogenesis** *Ciba Found Symp* **180**:210–22 <https://doi.org/10.1002/9780470514535.ch12>

The UniProt Consortium (2023) **UniProt: the Universal Protein Knowledgebase in 2023** *Nucleic Acids Res* **51**:D523–D531 <https://doi.org/10.1093/nar/gkac1052>

Toledo-Patiño S, Pascarelli S, Uechi GI, Laurino P (2022) **Insertions and deletions mediated functional divergence of Rossmann fold enzymes** *Proceedings of the National Academy of Sciences of the United States of America* **119** <https://doi.org/10.1073/pnas.2207965119>

Tokuriki N, Tawfik DS (2009) **Protein dynamism and evolvability** *Science* **324**:203–7 <https://doi.org/10.1126/science.1169375>

Tretyachenko V, Vymětal J, Neuwirthová T, Vondrášek J, Fujishima K, Hlouchová K (2022) **Modern and prebiotic amino acids support distinct structural profiles in proteins** *Open Biol* **12** <https://doi.org/10.1098/rsob.220040>

Trifonov EN (2000) **Consensus temporal order of amino acids and evolution of the triplet code** *Gene* **261**:139–151 [https://doi.org/10.1016/S0378-1119\(00\)00476-5](https://doi.org/10.1016/S0378-1119(00)00476-5)

Velankar S, Dana JM, Jacobsen J, van Ginkel G, Gane PJ, Luo J, Oldfield TJ, O’Donovan C, Martin MJ, Kleywegt GJ. (2013) **SIFTS: Structure Integration with Function, Taxonomy and Sequences resource** *Nucleic Acids Res* **41**:D483–9 <https://doi.org/10.1093/nar/gks1258>

Wächtershäuser G (1992) **Groundworks for an evolutionary biochemistry: The iron-sulphur world** *Progress in Biophysics and Molecular Biology* **58**:85–201 [https://doi.org/10.1016/0079-6107\(92\)90022-X](https://doi.org/10.1016/0079-6107(92)90022-X)

Weber AL, Miller SL (1981) **Reasons for the occurrence of the twenty coded protein amino acids** *Journal of Molecular Evolution* **17**:273–84 <https://doi.org/10.1007/BF01795749>

White HB (1976) **Coenzymes as fossils of an earlier metabolic state** *Journal of Molecular Evolution* **7**:101–104 <https://doi.org/10.1007/BF01732468>

White HB (1982) **Evolution of Coenzymes and the Origin of Pyridine Nucleotides. The Pyridine Nucleotide Coenzymes** *Econometrica* **50**:1–17 <https://doi.org/10.1016/b978-0-12-244750-1.50010-5>

Wong JT, Bronskill PM (1979) **Inadequacy of prebiotic synthesis as origin of proteinous amino acids** *Journal of Molecular Evolution* **13**:115–25 <https://doi.org/10.1007/BF01732867>

Wu HH *et al.* (2022) **The pathway for coenzyme M biosynthesis in bacteria** *Proceedings of the National Academy of Sciences of the United States of America* **119** <https://doi.org/10.1073/pnas.220719011>

Zaia DA, Zaia CT, De Santana H (2008) **Which amino acids should be used in prebiotic chemistry studies?** *Orig Life Evol Biosph* **38**:469–88 <https://doi.org/10.1007/s11084-008-9150-5>

Editors

Reviewing Editor

Donald Hamelberg

Georgia State University, Atlanta, United States of America

Senior Editor

David Ron

University of Cambridge, Cambridge, United Kingdom

Reviewer #1 (Public review):

By examining the prevalence of interactions with ancient amino acids of coenzymes in ancient versus recent folds, the authors noticed an increased interaction propensity for ancient interactions. They infer from this that coenzymes might have played an important role in prebiotic proteins. By only focusing on coenzymes, the authors may have overestimated their importance. What about other small molecules that existed in the prebiotic soup? Do they also prefer such ancient amino acids? if so, this might reflect the interaction propensity of specific amino acids rather than some possible role in very ancient proteins. Or it might diminish the conjectured importance of coenzymes. The analysis, which is very straightforward, is technically correct. However, the conclusions might not be as strong as presented. This paper presents an excellent summary of contemporary thought on what might have constituted prebiotic proteins and their properties.

<https://doi.org/10.7554/eLife.94174.2.sa2>

Reviewer #2 (Public review):

This study advances the model that the first canonical amino acids to emerge in life bound the earliest cofactors and led to the first proteins. The focus is on organic/organometallic cofactors, building on previous work on metals - ie. those in the groups of Bromberg, Dupont and others as well cited in the manuscript. Studies of this type are limited both by data availability and confounding chemical effects that are exacerbated by the timescale of evolutionary inference tackled here. However, the analysis provides a solid addition to the field and complements existing metal-focused studies as well as those Longo, Russell and others (also well cited).

<https://doi.org/10.7554/eLife.94174.2.sa1>

Author response:

The following is the authors' response to the original reviews.

Reviewer #1 (Public Review):*Summary:*

By examining the prevalence of interactions with ancient amino acids of coenzymes in ancient versus recent folds, the authors noticed an increased interaction propensity for ancient interactions. They infer from this that coenzymes might have played an important role in prebiotic proteins.

Strengths:

(1) The analysis, which is very straightforward, is technically correct. However, the conclusions might not be as strong as presented.

(2) This paper presents an excellent summary of contemporary thought on what might have constituted prebiotic proteins and their properties.

(3) The paper is clearly written.

We are grateful for the kind comments of the reviewer on our manuscript. However, we would like to clarify a possible misunderstanding in the summary of our study. Specifically, analysis of "ancient versus recent folds" was not really reported in our results. Our analysis concerned "coenzyme age" rather than the "protein folds age" and was focused mainly on interaction with early vs. late amino acids in protein sequence. While structural propensities of the coenzyme binding sites were also analyzed, no distinction on the level of ancient vs. recent folds was assumed and this was only commented on in the discussion, based on previous work of others.

Weaknesses:

(1) The conclusions might not be as strong as presented. First of all, while ancient amino acids interact less frequently in late with a given coenzyme, maybe this just reflects the fact that proteins that evolved later might be using residues that have a more favorable binding free energy.

We would like to point out that there was no distinction between proteins that evolved early or late in our dataset of coenzyme-binding proteins. The aim of our analysis was purely to observe trends in the age of amino acids vs. age of coenzymes. While no direct inference can

be made from this about early life as all the proteins are from extant life (as highlighted in the discussion of our work), our goal was to look for intrinsic propensities of early vs. late amino acids in binding to the different coenzyme entities. Indeed, very early interactions would be smeared by the eons of evolutionary history (perhaps also towards more favourable binding free energy, as pointed out also by the reviewer). Nevertheless, significant trends have been recorded across the PDB dataset, pointing to different propensities and mechanistic properties of the binding events. Rather than to a specific evolutionary past, our data therefore point to a “capacity” of the early amino acids to bind certain coenzymes, and we believe that this is the major (and standing) conclusion of our work, along with the properties of such interactions. In our revised version, we will carefully go through all the conclusions and make sure that this message stands out, but we are confident that the following concluding sentences copied from the abstract and the discussion of our manuscript fully comply with our data:

“These results imply the plausibility of a coenzyme-peptide functional collaboration preceding the establishment of the Central Dogma and full protein alphabet evolution”

“While no direct inferences about distant evolutionary past can be drawn from the analysis of extant proteins, the principles guiding these interactions can imply their potential prebiotic feasibility and significance.”

“This implies that late amino acids would not be necessarily needed for the sovereignty of coenzyme-peptide interplay.”

We would also like to add that proteins that evolved later might not always have higher free energy of binding. Musil et al., 2021 (<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8294521/>) showed in their study on the example of haloalkane dehalogenase Dha A that the ancestral sequence reconstruction is a powerful tool for designing more stable, but also more active proteins. Ancestral sequence reconstruction relies on finding ancient states of protein families to suggest mutations that will lead to more stable proteins than are currently existing proteins. Their study did not explore the ligand-protein interactions specifically but showed that ancient states often show more favorable properties than modern proteins.

(2) What about other small molecules that existed in the prebiotic soup? Do they also prefer such ancient amino acids? If so, this might reflect the interaction propensity of specific amino acids rather than the inferred important role of coenzymes.

We appreciate the comment of the reviewer towards other small molecules, which we assume points mainly towards metal ions (i.e. inorganic cofactors). We completely agree with the reviewer that such interactions are of utmost importance to the origins of life. Intentionally, they were not part of our study, as these have already been studied previously by others (e.g. Bromberg et al., 2022; and reviewed in Frenkel-Pinter et al., 2020) and also us (Fried et al., 2022). For example, it is noteworthy that prebiotically relevant metal binding sites (e.g. of Mg²⁺) exhibit enrichment in early amino acids such as Asp and Glu while more recent metal (e.g. Cu and Zn) site in the late amino acids His and Cys (Fried et al., 2022). At the same time, comparable analyses of amino acid - coenzyme trends were not available.

Nevertheless, involvement of metal ions in the coenzyme binding sites was also studied here and pointed to their bigger involvement with the Ancient coenzymes. In the revised version of the manuscript, we will be happy to enlarge the discussion of the studies concerning inorganic cofactors.

The following sentence was added in the discussion of the revised manuscript:

“This would also be true for direct interaction of early peptides/proteins and metal ions, independent of organic cofactor involvement, as discussed previously by us and others

(Bromberg et al., 2022; Frenkel-Pinter et al., 2020; Fried et al., 2022). For example, it has been observed that coordination of prebiotically most relevant metal ions (e.g., Mg^{2+}) is more often mediated by early amino acids such as Asp and Glu, whereas metal ions of later relevance (e.g., Cu and Zn) bind more frequently via late amino acids like His and Cys (Fried et al. 2022). Similarly, ancient metal binding folds have been shown to be enriched in early amino acids (Bromberg et al., 2022)."

(3) Perhaps the conclusions just reflect the types of active sites that evolved first and nothing more.

We partly agree on this point with the reviewer but not on the fact why it is listed as the weakness of our study and on the "nothing more" notion. Understanding what the properties of the earliest binding sites is key to merging the gap between prebiotic chemistry and biochemistry. The potential of peptides preceding ribosomal synthesis (and the full alphabet evolution) along with prebiotically plausible coenzymes addresses exactly this gap, which is currently not understood.

Reviewer #2 (Public Review):

I enjoyed reading this paper and appreciate the careful analysis performed by the investigators examining whether 'ancient' cofactors are preferentially bound by the first-available amino acids, and whether later 'LUCA' cofactors are bound by the late-arriving amino acids. I've always found this question fascinating as there is a contradiction in inorganic metal-protein complexes (not what is focused on here). Metal coordination of Fe, Ni heavily relies on softer ligands like His and Cys - which are by most models latecomer amino acids. There are no traces of thiols or imidazoles in meteorites - although work by Dvorkin has indicated that could very well be due to acid degradation during extraction. Chris Dupont (PNAS 2005) showed that metal speciation in the early earth (such as proposed by Anbar and prior RJP Williams) matched the purported order of fold emergence.

As such, cofactor-protein interactions as a driving force for evolution has always made sense to me and I admittedly read this paper biased in its favor. But to make sure, I started to play around with the data that the authors kindly and importantly shared in the supplementary files. Here's what I found:

Point 1: The correlation between abundance of amino acids and protein age is dominated by glycine.

There is a small, but visible difference in old vs new amino acid fractional abundance between Ancient and LUCA proteins (Figure 3, Supplementary Table 3). However, the bias is not evenly distributed among the amino acids - which Figure 4A shows but is hard to digest as presented. So instead I used the spreadsheet in Supplement 3 to calculate the fractional difference $FD_{aa} = F(\text{old aa}) - F(\text{new aa})$. As expected from Figure 3, the mean FD for Ancient is greater than the mean FD for LUCA. But when you look at the same table for each amino acid $FD_{\text{cofactor}} = F(\text{ancient cofactor}) - F(\text{LUCA cofactor})$, you now see that the bias is not evenly distributed between older and newer amino acids at all. In fact, most of the difference can be explained by glycine ($FD_{\text{cofactor}} = 3.8$) and the rest by also including tryptophan ($FD_{\text{cofactor}} = -3.8$). If you remove these two amino acids from the analysis, the trend seen in Figure 3 all but disappears.

Troubling - so you might argue that Gly is the oldest of the old and Trp is the newest of the new so the argument still stands. Unfortunately, Gly is a lot of things - flexible, small, polar - so what is the real correlation, age, or chemistry? This leads to point 2.

We truly acknowledge the effort that the reviewer made in the revision of the data and for the thoughtful, deeper analysis. We agree that this deserves further discussion of our data.

As invited by the reviewer, we indeed repeated the analysis on the whole dataset. First, we would like to point out that the reviewer was most probably referring to the Supplementary Fig. 2 (and not 3, which concerns protein folds). While the difference between Ancient and LUCA coenzyme binding is indeed most pronounced for Gly and Trp, we failed to confirm that the trend disappears if those two amino acids are removed from the analysis (additional FDcofactors of 3.2 and -3.2 are observed for the early and late amino acids, resp.), as seen in Table I below. The main additional contributors to this effect are Asp (FD of 2.1) and Ser (FD of 1.8) from the early amino acids and Arg (FD of -2.6) and Cys (FD of -1.7) of the late amino acids. Hence, while we agree with the reviewer that Gly and Trp (the oldest and the youngest) contribute to this effect the most, we disagree that the trend reduces to these two amino acids.

In addition, the most recent coenzyme temporality (the Post-LUCA) was neglected in the reviewer's analysis. The difference between F (old) and F (new) is even more pronounced in Post-LUCA than in LUCA, vs. Ancient (Supplementary table 5A) and depends much less on Trp. Meanwhile, Asp, Ser, Leu, Phe, and Arg dominate the observed phenomenon (Supplementary table 5b). This further supports our lack of agreement with the reviewer's point. Nevertheless, we remain grateful for this discussion and we will happily include this additional analysis in the Supplementary Material of our revised manuscript.

The following text (and the additional data) was included in the revised manuscript version:

“To explore the contribution of individual amino acids to this effect, fractional difference (FD) for early vs. late amino acids among the Ancient, LUCA, and Post-LUCA coenzyme binding was calculated (Supplementary Table 5). The mean FD revealed a similar trend to the amino acid composition analysis (Fig. 3). The amino acids most enriched in LUCA vs. Post-LUCA are Gly, Ser, and Leu (FD of 4.4, 4.3, and 4.1 respectively), while the most depleted include Phe, Arg, and His (FD of -11, -4.2, and -3.2) (Supplementary Table 5B).”

Point 2 - The correlation is dominated by phosphate.

In the ancient cofactor list, all but 4 comprise at least one phosphate (SAM, tetrahydrofolic acid, biopterin, and heme). Except for SAM, the rest have very low Gly abundance. The overall high Gly abundance in the ancient enzymes is due to the chemical property of glycine that can occupy the right-hand side of the Ramachandran plot. This allows it to make the alternating α left- α right conformation of the P-loop forming Milner-White's anionic nest. If you remove phosphate binding folds from the analysis the trend in Figure 3 vanishes.

Likewise, Trp is an important functional residue for binding quinones and tuning its redox potential. The LUCA cofactor set is dominated by quinone and derivatives, which likely drives up the new amino acid score for this class of cofactors.

Once again, we are thankful to the reviewer for raising this point. The role of Gly in the anionic nests proposed by Milner-White and Russel, as well as the Trp role in quinone binding are important points that we would be happy to highlight more in the discussion of the revised manuscript.

Nevertheless, we disagree that the trends reduce only to the phosphate-containing coenzymes and importantly, that “the trend in Figure 3 vanishes” upon their removal. Supplementary table 6A and 6B show the data for coenzymes excluding those with phosphate moiety and the trend in Fig. 3 remains, albeit less pronounced.

The following text was included in the revised manuscript version:

“Moreover, we investigated whether the observed trend in amino acid occurrence at the binding sites was dominated by the presence of phosphate groups, which are common in many ancient cofactors except for SAM, Tetrahydrofolic acid, Biopterin, and Heme. An additional analysis therefore excluded all phosphate-containing coenzymes indicating that while the trend is less pronounced, it remains even in the absence of phosphate groups (Supplementary Table 6).”

In summary, while I still believe the premise that cofactors drove the shape of peptides and the folds that came from them - and that Rossmann folds are ancient phosphate-binding proteins, this analysis does not really bring anything new to these ideas that have already been stated by Tawfik/Longo, Milner-White/Russell, and many others.

I did this analysis ad hoc on a slice of the data the authors provided and could easily have missed something and I encourage the authors to check my work. If it holds up it should be noted that negative results can often be as informative as strong positive ones. I think the signal here is too weak to see in the noise using the current approach.

We are grateful to the reviewer for encouraging further look at our data. While we hope that the analysis on the whole dataset (listed in Tables I - IV) will change the reviewer's standpoint on our work, we would still like to comment on the questioned novelty of our results. In fact, the extraordinary works by Tawfik/Longo and Milner-White/Russell (which were cited in our manuscript multiple times) presented one of the motivations for this study. We take the opportunity to copy the part of our discussion that specifically highlights the relevance of their studies, and points out the contribution of our work with respect to theirs.

“While all the coenzymes bind preferentially to protein residue sidechains, more backbone interactions appear in the ancient coenzyme class when compared to others. This supports an earlier hypothesis that functions of the earliest peptides (possibly of variable compositions and lengths) would be performed with the assistance of the main chain atoms rather than their sidechains (Milner-White and Russell 2011). Longo et al., recently analyzed binding sites of different phosphate-containing ligands which were arguably of high relevance during earliest stages of life, connecting all of today's core metabolism (Longo et al., 2020 (b)). They observed that unlike the evolutionary younger binding motifs (which rely on sidechain binding), the most ancient lineages indeed bind to phosphate moieties predominantly via the protein backbone.

Our analysis assigns this phenomenon primarily to interactions via early amino acids that (as mentioned above) are generally enriched in the binding interface of the ancient coenzymes. This implies that late amino acids would not be necessarily needed for the sovereignty of coenzyme-peptide interplay.”

Unlike any other previous work, our study involves all the major coenzymes (not just the phosphate-containing ones) and is based on their evolutionary age, as well as age of amino acids. It is the first PDB-wide systematic evolutionary analysis of coenzyme-amino acid binding. Besides confirming some earlier theoretical assertions (such as role of backbone interactions in early peptide-coenzyme evolution) and observations (such as occurrence of the ancient phosphate-containing coenzymes in the oldest protein folds), it uncovers substantial novel knowledge. For example, (i) enrichment of early amino acids in the binding of ancient coenzymes, vs. enrichment of late amino acids in the binding of LUCA and Post-LUCA coenzymes, (ii) the trends in secondary structure content of the binding sites of coenzyme of different temporalities, (iii) increased involvement of metal ions in the ancient coenzyme binding events, and (iv) the capacity of only early amino acids to bind ancient

coenzymes. In our humble opinion, all of these points bring important contributions in the peptide-coenzyme knowledge gap which has been discussed in a number of previous studies.

Recommendations for the authors:

(1) By only focusing on coenzymes, the authors may have overestimated their importance. What about other small molecules that existed in the prebiotic soup? Do they also prefer such ancient amino acids? If so, this might reflect the interaction propensity of specific amino acids rather than some possible role in very ancient proteins. Or it might diminish the conjectured importance of coenzymes.

The following sentence was added in the discussion of the revised manuscript:

“This would also be true for direct interaction of early peptides/proteins and metal ions, independent of organic cofactor involvement, as discussed previously by us and others (Bromberg et al., 2022; Frenkel-Pinter et al., 2020; Fried et al., 2022). For example, it has been observed that coordination of prebiotically most relevant metal ions (e.g., Mg²⁺) is more often mediated by early amino acids such as Asp and Glu, whereas metal ions of later relevance (e.g., Cu and Zn) bind more frequently via late amino acids like His and Cys (Fried et al. 2022). Similarly, ancient metal binding folds have been shown to be enriched in early amino acids (Bromberg et al., 2022).”

(2) The authors should analyze whether the interactions are with similar types of amino acids in ancient versus early proteins.

While we appreciate the interesting suggestion, we would like to clarify that we did not aim to elucidate the differences between early and late protein folds - we agree that this might add an interesting perspective to our work, but we feel that it is well beyond the scope of our current study.

(3) The authors might also wish to do sequence alignments to the structures in early versus late evolving proteins to see how general this pattern of residue usage is beyond the limited set of proteins found in the PDB.

This is an interesting suggestion but similar to the previous recommendation, it is not within the scope of this study where no distinction between early and late evolving proteins has been made.

There has been a number of attempts to classify the folds as shared among Bacteria, Archea and Eukaryota or specific to one or two of these groups of organisms (<https://link.springer.com/article/10.1007/s00239-023-10136-x>, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9541633/>) - this does not however compare easily with our time scales - where ancient ligands occur well before the last common ancestor.

We also agree the set of sequences present in the PDB is biased, but perhaps it is less biased than we have thought. The recent fantastic work <https://www.biorxiv.org/content/10.1101/2024.03.18.585509v2> from Nicola Bordin and his colleagues from Orengo group attempted to classify over 200 million structures in AlphaFold database in so called Encyclopedia of Domains and they found out that nearly 80% of detected domains can be assigned to already known superfamilies in CATH (<https://www.biorxiv.org/content/10.1101/2024.03.18.585509v2>).

(4) The authors might wish to consider the results in Skolnick, H. Zhou, and M. Gao. On the possible origin of protein homochirality, structure, and biochemical function. PNAS 2019: 116(52): 26571-26579.

Based on the editorial recommendation, the following sentence was added in the discussion:

“It has been implied by computer simulations that coenzymes could bind to proteins with similar propensity even before the onset of protein homochirality, despite lower structural stability and secondary structure content in heterochiral polypeptides (Skolnick et al., 2019).”

<https://doi.org/10.7554/eLife.94174.2.sa0>