Supplementary Information for

Boundary conditions for early life converge to an organo-sulfur metabolism

Joshua E. Goldford, Hyman Hartman, Robert Marsland III and Daniel Segrè

Daniel Segrè
E-mail: dsegre@bu.edu

This PDF file includes:

Supporting Information Text

Network enzymes retain features of nitrogen-free catalysts. In order for the expanded networks presented in the main text to have operated in prebiotic conditions, reactions would have been catalyzed non-enzymatically by inorganic or simple organic catalysts available in prebiotic environments. Prior work has suggested that reactions in metabolic networks that proceed spontaneously or depend on enzymes with inorganic coenzymes, such as iron-sulfur or transition metal cofactors, may have operated in prebiotic conditions (1–4). We identified reactions in KEGG that could proceed spontaneously or are dependent on one of several inorganic coenzymes (Methods), and defined this set of reactions as *plausibly pre-enzymatic*, or "PPE"-reactions (Extended Data Fig. 1a). For each proposed prebiotic scenario that lead to expansion of at-least 100 metabolites ($n = 144$, Extended Data Fig. 1a), we partitioned reactions added to the network before the inclusion of ammonia into the seed set (herein called "pre-ammonia" reactions) and reactions added to the network after ammonia was added to the seed set (or "post-ammonia" reactions). We then computed the fraction pre- and post-ammonia reactions that were classified as PPE reactions, and found that pre-ammonia reactions contained a higher proportion of PPE-reactions relative to post-ammonia reactions (one-tailed Wilcoxon sign-rank test: $P < 10^{-19}$), suggesting that the pre-ammonia reactions may have been more readily catalyzed by simple inorganic catalysts in prebiotic environments. We next hypothesized that if these enzymes evolved from a thioester-driven proto-metabolism without nitrogen, then enzymes in these networks should be depleted in enzyme-bound nitrogen-containing coenzymes. We thusly computed the fraction of pre- and post-ammonia reactions dependent on enzymes containing TPP, PLP, heme, biotin, flavin, pterin, and cobalamin (Fig. Extended Data Fig. 1c, Methods). We found that the proportion of pre-ammonia reactions associated with these coenzymes were significantly less than the proportion of post-ammonia reactions dependent on these coenzymes (Fig. Extended Data Fig. 1d, one-tailed Wilcoxon sign-rank test: $P < 10^{-24}$), which is primarily due to the large number of PLP-dependent reactions added to the network after the inclusion of ammonia.

Since only a minority of reactions in this network were categorized as PPE, simple organic or organosulfur catalysts may have been necessary in order for this network to function in prebiotic environments. Christian de Duve suggested that thioester-based polymers may have provided the necessary catalytic components of ancient metabolism in addition to inorganic catalysts (5). In modern living systems, monomers of keto acids are converted into amino acids, which are then polymerized into polypeptides either with or without the aid of the ribosome and mRNA. If prebiotic environments were severely nitrogen limited, keto acids may have been reduced to hydroxy acids, and polymerized into polyesters using thioesters as a condensing agent. Notably, in such a scenario only the polymer backbone is altered, leaving the side chains (R-groups) within today's amino acids intact. Recent work has demonstrated that polyesters may aid in the polymerization of amino acids during dry-wet cycles (6), and that the peptidyl-transferase domain on the ribosome can polymerize hydroxyacylated tRNAs to form polyesters (7, 8), suggesting that ester bond formation may have proceeded amide bond formation in living systems.

It has been proposed that enzymes retain features of early catalysts before the emergence of the genetic code and protein translation systems, and that enzyme active sites may bear resemblance to ancient catalysts. Thus, if this network represents a relic of an ancient metabolism before the biological incorporation of nitrogen, then the active sites of enzymes catalyzing reactions within the network should be depleted in amino acids with side chains containing nitrogen (Extended Data Fig. S1e). To see if the catalytic residues of the enzymes in the pre-ammonia network were depleted in amino acids with nitrogenous side chains, we first obtained a database of catalytic site residues inferred from protein structures (9). After removing entries with interactions mediated by the peptide backbone, this dataset consisted of 18,721 entries, 1,304 of which were associated with active sites of enzymes in the nitrogen-free network in a representative network. For each putative prebiotic scenario resulting in an expansion with more than 100 metabolites, we computed the fraction of active site residues that contained nitrogen in enzymes associated with both pre- and post-ammonia reactions (Extended Data Fig. 1e). We found that the proportion of nitrogenous catalytic residues associated with pre-ammonia reactions was significantly lower than the proportion of nitrogenous catalytic residues associated with post-ammonia reactions (Fig. Extended Data Fig. 1f, Wilcoxon sign-rank test: $P < 10^{-24}$).

One potential alternative explanation for these biases in amino acid composition within the active sites of extant enzymes may be the outcome of evolutionary selection: nitrogen limitation in the environment may have favored mutations that lead to less nitrogen within these enzymes. However, evidence for selection for less nitrogen usage would manifest within the entire protein sequence, rather than just the active sites. Thus, we computed the fraction of amino acids with nitrogenous side chains across the entire coding sequences, rather than specifically the active sites, for enzymes associated with pre- and post-ammonia reactions (see Methods). We found no evidence that enzymes in the pre-ammonia network had a decreased usage of amino acids with nitrogenous side chains side chains relative to enzymes added to the network after ammonia was included in the seed set (one-tailed Wilcoxon sign rank test: $P = 1$) , suggesting that the biases within the active sites are not merely a consequence evolutionary selection (see Extended Data Fig. 2).

References

1. Sousa FL, et al. (2013) Early bioenergetic evolution. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 368(1622):20130088.
2. Muchowska KB, et al. (2017) Metals promote sequences of the reverse Krebs cycle. *Nature Ecology and Evolution* 1(11):1716–1721.
3. Varma SJ, Muchowska KB, Chatelain P, Moran J (2018) Native iron reduces CO2 to intermediates and end-products of the acetyl-CoA pathway. *Nature Ecology & Evolution* 2.

4. Sousa FL, Preiner M, Martin WF (2018) Native metals, electron bifurcation, and CO2 reduction in early biochemical evolution. *Current opinion in microbiology* 43:77–83.
5. de Duve C (1991) *Blueprint for a cell: the nature and origin of life.* (Neil Patterson Publishers, Carolina Biological Supply Company, Burlington, N.C.), p. 275.
6. Forsythe JG, et al. (2015) Ester-Mediated Amide Bond Formation Driven by Wet-Dry Cycles: A Possible Path to Polypeptides on the Prebiotic Earth. *Angewandte Chemie - International Edition* 54(34):9871–9875.
7. Fahnestock S, Rich A (1971) Ribosome-Catalyzed Polyester Formation. *Science* 173(3994):340–343.
8. Ohta A, Murakami H, Suga H (2008) Polymerization of alpha-hydroxy acids by ribosomes. *Chembiochem : a European journal of chemical biology* 9(17):2773–2778.
9. Ribeiro AJM, et al. (2018) Mechanism and Catalytic Site Atlas (M-CSA): a database of enzyme reaction mechanisms and active sites. *Nucleic Acids Research* 46(D1):D618–D623.