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

Joshua E. Goldford[1,2], Hyman Hartman[3], Robert Marsland III[2] and Daniel Segrè[1,2,4*]

[1]Bioinformatics Program, Boston University, Boston, MA 02215, USA

[2]Biological Design Center and Department of Physics, Boston University, Boston, MA 02215, USA

[3]Earth, Atmosphere and Planetary Science Department, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

[4]Department of Biomedical Engineering, Department of Biology, Boston University, Boston MA, 02215, USA

[*]Corresponding authors

Email: dsegre@bu.edu

Running Title: Boundary conditions for early metabolism

Keywords: Metabolism, evolution, network expansion, origin of life, constraint-based modeling

Abstract

It has been suggested that a deep memory of early life is hidden in the architecture of metabolic networks, whose reactions could have been catalyzed by small molecules or minerals prior to genetically encoded enzymes (1–6). A major challenge in unraveling these early steps is assessing the plausibility of a connected, thermodynamically consistent proto-metabolism under different geochemical conditions, which are still surrounded by high uncertainty. Here we combine network-based algorithms (9, 10) with physicochemical constraints on chemical reaction networks to systematically show how different combinations of parameters (temperature, pH, redox potential and availability of molecular precursors) could have affected the evolution of a proto-metabolism. Our analysis of possible trajectories indicates that a subset of boundary conditions converges to an organo-sulfur-based proto-metabolic network fueled by a thioester- and redox-driven variant of the reductive TCA cycle, capable of producing lipids and keto acids. Surprisingly, environmental sources of fixed nitrogen and low-potential electron donors seem not to be necessary for the earliest phases of biochemical evolution. We use one of these networks to build a steady-state dynamical metabolic model of a proto-cell, and find that different combinations of carbon sources and electron acceptors can support the continuous production of a minimal ancient "biomass" composed of putative early biopolymers and fatty acids.

The structure of metabolism carries a memory of its evolutionary history that may date back to before the onset of an RNA-based genetic system (1–6). Decoding this ancient evolutionary record could provide important insight into the early stages of life on our planet (2, 5–8), but constitutes a challenging problem. This challenge is due both to the difficulty of interrogating complex biochemical networks under different environmental conditions, and to the uncertainty about these conditions on prebiotic Earth. Estimates of plausible Archean environments that led to the emergence and evolution of living systems vary dramatically (21, 22), ranging from alkaline hydrothermal vents driven by chemical gradients (23) to acidic ocean seawater driven by photochemistry (3, 4). Although geochemical data support the availability of mid-potential electron donors (H_2) (24), sulfur (H_2S) and potentially fixed carbon (25) in ancient environments, several key molecules used in living systems may have been severely limiting, including a source of fixed nitrogen (26, 27) (e.g. ammonia), low-potential electron donors (28, 29) and phosphate (30–32). Rather than assuming a steady supply of these biomolecules, one can critically revisit the notion that these molecules would have been necessary for the emergence of ancient proto-metabolic systems. Indeed, we recently used network-based algorithms to ask whether early metabolism would have required a source of phosphate, and found evidence that thioesters, rather than phosphate, may have endowed ancient metabolism with key energetic and biosynthetic capacity (12). This raises the broader question of whether other molecules and physico-chemical conditions may not be as crucial as previously thought for the emergence of a proto-metabolism. Understanding these dependencies could also reveal whether specific components of metabolism are particularly robust or fragile with respect to these initial conditions.

A computational method that can help address these questions is the network expansion algorithm, which simulates the growth of a biochemical network by iteratively adding to an initial set of compounds the products of reactions enabled by available substrates, until convergence (9, 10). This algorithm, in its application to the study of ancient life (11)(12), relies on three key assumptions: *first*, that chemical reactions successful in early processes were gradually augmented with new pathways, but never replaced; *second*, that, over long time-scales, horizontal gene transfer produced abundant shuffling of biochemical reactions across different

88 organisms (19, 20), suggesting that an ecosystem-level approach to metabolism may be
89 particularly suitable for describing ancient biochemistry; and *third*, that inorganic or small
90 molecular catalysts (12, 13) could catalyze, in a weaker and less specific manner relative to
91 modern enzymes, a large number of metabolic reactions, as confirmed by an increasing body of
92 experimental evidence (14–18).
93
94 In this paper, we systematically explore a combinatorial set of molecules and parameters
95 associated with possible early Earth environments, and use an enhanced network expansion
96 algorithm to determine which proto-metabolic networks are thermodynamically reachable under
97 each of these initial conditions. We further use constraint-based flux balance modeling to
98 demonstrate the capacity of some of these networks to sustain flux, in a way that resembles
99 homeostatic growth of present-day cells. Our results suggest that a thioester-driven organic
100 network may have robustly arisen without phosphate, fixed nitrogen or low-potential electron
101 donors. This network, by supporting the biosynthesis of keto acids and fatty acids may have
102 prompted the rise of complex self-sustaining biochemical pathways, marking a key transition
103 towards the origin of life.
104
105 We first sought to systematically characterize the effect of various geochemical scenarios
106 on the possible structure of ancient metabolism. Building on prior work (11, 12), we constructed
107 a model of ancient biosphere-level metabolism based on the KEGG database (33). We first
108 modified the network (as described in Methods) to account for previously proposed primitive
109 thioester-coupling and redox reactions (12). For each possible set of environmental parameters
110 (including temperature, pH and redox potential) we computed the thermodynamic feasibility of
111 each reaction, and removed infeasible reactions (see Methods). This allowed us to implement a
112 thermodynamically-constrained network expansion algorithm (12), which iteratively adds
113 metabolites and thermodynamically-feasible reactions to a network until convergence (see
114 Methods). We performed thermodynamically-constrained network expansion (see Methods and
115 Fig. 1A) for n=672 different geochemical scenarios, systematically varying pH, temperature,
116 redox potential of primitive redox systems, and the availability of key biomolecules including
117 thiols (that subsequently form thioesters), fixed carbon (formate/acetate) and fixed nitrogen
118 (ammonia) (Methods, Fig. 1).

Of the 672 different simulated geochemical scenarios, we found that 288 (43%) expanded to networks containing over 100 metabolites (Fig. 1B). A logistic regression classifier that uses geochemical parameters as predictors (see Methods and Fig. 1C) allowed us to quantify the importance of each environmental parameter in determining whether the expanded network would reach such a large size. Surprisingly, removing the variable associated with presence/absence of ammonia did not affect predictive power of the classifier, suggesting that a source of fixed nitrogen is not an important determinant of the expansion. Consistent with the relevance of this result to ancient metabolism, we found that the enzymes that catalyze reactions in the expanded networks before the addition of ammonia were depleted in nitrogen-containing coenzymes (see Fig. S1C-D, one-tailed Wilcoxon sign rank test: $P < 10^{-24}$) and in active site amino acids with nitrogeneous side chains (see Fig. SE-F, one-tailed Wilcoxon sign rank test: $P < 10^{-24}$) relative to enzymes added after the addition of ammonia (see Supplemental Text). These results suggest that ammonia may have not been essential for the initial expansion of metabolism, and point to a thioester-coupled organo-sulfur metabolic network (Fig. 1) as a core network that deserves further attention.

Beyond the dispensability of nitrogen, the simulations described above revealed a number of relationships between plausible geochemical scenarios and the structure and size of our simulated proto-metabolic networks. First, expansion beyond 100 metabolites was feasible in the absence of a source of fixed carbon, but only when thiols were provided in the seed set, highlighting the importance for thioester-coupling for ancient carbon fixation pathways (12, 28, 29). The presence of thiols enabled the production of key biomolecules, including fatty acids and branched chain keto acids (see Fig S2). Second, we explored the effect of the primitive redox system by systematically varying the reduction potential of the electron donor in the seed set (see Methods, Fig. 2A). Unexpectedly, we found that as we increased the fixed potential of the electron donor, expansion to a large network was feasible over a broad range of reduction potentials (between -150 and 50 mV). Only upon reaching 50mV the expanded network collapsed to a much smaller solution, suggesting that the generation of low-potential electron donors from H_2 may not have been a necessary condition for the early expansion of a proto-metabolism (Fig. 3A). Thus, less stringent constraints, e.g. the presence of mid-potential redox couples and thioester-forming thiols could have enabled the emergence of an autotrophic proto-

metabolic network capable of producing key biomolecules. Notably, both functions can be potentially carried out by disulfides.

Analysis of the expanded networks without nitrogen revealed that a large number of different initial conditions converged to similar expanded organo-sulfur proto-metabolic networks, spanning variants of key pathways in central carbon metabolism (Fig. 3B). For the majority of simulations, variants of modern heterotrophic carbon assimilation pathways, including the glyoxylate cycle and TCA cycle, were highly represented in the network (Fig. 3B). Several carbon fixation pathways were highly represented in the simulated networks as well: in over half of the networks that expanded beyond 100 metabolites, we found 92 % (12/13) of the compounds (or generalized derivatives) that participate in the reductive tricarboxylic acid (rTCA) cycle, with the exception of phosphoenolpyruvate. We also found that under several geochemical conditions, all intermediates were producible for three carbon fixation pathways, including the 3-hydroxypropionate bi-cycle, the hydroxypropionate-hydroxybutylate cycle, and the dicarboxylate-hydroxybutyrate cycle (Fig. 3A). At-most, only 3 of 9 metabolites used in the Wood-Ljungdahl (WL) pathway were observed, due to the lack of nitrogen-containing pterins in the network. This does not necessarily rule out the primordial importance of the WL-pathway, as its early variants could have been radically different than today's WL-pathway, relying on native metals to facilitate reduction of CO_2 to acetate (25, 29). In addition to observing a large number of metabolites used in carbon fixation pathways, we found that a large fraction of the β-oxidation pathway was represented in our networks, which may have supported the production of fatty acids in ancient living systems by operating in the reverse direction. Lastly, we observed that the majority of intermediates involved in the production of branched-chain amino acids were also producible in the expanded networks.

A more detailed analysis of the convergent organo-sulfur proto-metabolic network reveals new possible ancestral metabolic pathways that involve previously unexplored combinations of reactions and metabolites. Fig. 3B shows a variant of the (r)TCA cycle that is a component of these expanded networks, and that may have served as the core organo-sulfur network fueling ancient living systems. Rather than using ATP-dependent reactions found in extant species (e.g. Succinyl-CoA synthetase and ATP citrate lyase), these reactions are

181 substituted with non-ATP-dependent reaction mechanisms. For instance, the production of a
182 succinyl-thioester in the extant rTCA cycle relies on Succinyl-CoA synthetase, performing the
183 following reaction: ATP + Succinate + CoA → Succinyl-CoA + ADP + P_i. However, in the
184 network presented in Fig. 3B, malyl-thioester, producible through alternative reactions, donates a
185 thiol to succinate, subsequently forming a succinyl-thioester. This (r)TCA cycle analogue is able
186 to produce eight keto acids normally serving as key intermediates and precursors to common
187 amino acids in central carbon metabolism (glyoxylate, pyruvate, oxaloacetate, 2-oxoglutarate
188 and hydroxypyruvate), as well as a few branched-chain keto acids. Additionally, long-chain
189 fatty acids like palmitate are producible in this network, driven by thioester and redox-coupling
190 rather than ATP, like in extant fatty acid biosynthesis. Thus, despite the simplicity of seed
191 compounds, several small molecular weight keto acids and fatty acids may have been producible
192 in an organo-sulfur proto-metabolism.
193
194 So far, we have focused only on the topology and global thermodynamic feasibility of
195 putative ancient metabolic networks. Inspired by recent studies on the molecular budget of
196 present-day cells, we decided to further explore whether proto-metabolic networks could support
197 steady state fluxes, and fuel primitive proto-cells with internal energy sources (e.g. thioesters),
198 redox gradients, and primitive biopolymer capable of catalysis and compartmentalization. Flux
199 balance analysis (FBA), originally developed for the study of microbial metabolism, enables the
200 prediction of systems-level properties of metabolic networks at steady-state (34). Fundamentally,
201 FBA computes possible reaction rates in a network constrained by mass and energy balance,
202 usually under the assumption that a specific composition of biomolecules is efficiently produced
203 during a homeostatic growth process. In microbial metabolism, FBA is used to simulate the
204 production of cellular biomass (e.g. protein, lipids, and nucleic acids) at fixed proportions, which
205 are derived from known composition of extant cells. We realized that the same approach could
206 help test the sustainability of a proto-metabolic biochemical system, provided that we could
207 develop a plausible hypothesis for the "biomass composition" of ancient proto-cells. As a
208 starting point, we recalled Christian de Duve's suggestion that the thioester-driven
209 polymerization of monomers producible from ancient proto-metabolism may have led to
210 "catalytic multimers," which could have served as catalysts for ancient biochemical reactions (4).
211 Under nitrogen limited conditions, keto acids producible from proto-metabolism (see Fig. 3B)

could have been reduced to α-hydroxy acids, and polymerized into polyesters using thioesters as a condensing agent (see Fig. S3). Recent work has suggested that polymers of α-hydroxy acids may have been stably produced in geochemical environments (35), and that these molecules could have served as primitive catalysts (36). These results all point to the intriguing possibility that the thioester-driven polymerization of α-hydroxy acids (producible from keto acid precursors of common amino acids) generated the first metabolically sustainable cache of ancient catalysts, leading to a collectively autocatalytic protocellular system. We employed FBA to specifically test the feasibility of such a system. Using an expanded metabolic network as a scaffold for network reconstruction (Fig. 3B), we constructed a constraint-based model of an ancient proto-cell using a biomass composition consisting of fatty acids (for proto-cellular membranes), "catalytic multimers" derived from eight keto acids (Fig. 4B), and redox and thioester-based free energy sources (Methods, Fig. 4A). We used thermodynamic metabolic flux analysis (TMFA), a variant of FBA that explicitly considers thermodynamic constraints (37) (see Methods), to determine whether homeostatic growth of the whole system was achievable. We found that growth of the proto-cell metabolic model is indeed feasible under a wide variety of assumptions regarding macromolecular compositions and input molecules (Fig. 4B). Notably, growth is achievable in simple chemoautotrophic conditions with either H_2 or H_2S, but not Fe(II), as electron donors (Fig. 4B). In this model, thiols and thioesters are not supplied as food sources, but rather are recycled during steady-state growth of the proto-cell. This reflects the possibility that thiols could have been initially supplied abiotically, followed by the rapid takeover of biotic production of mercaptopyruvate, a keto acid that could have been incorporated into primitive multimers.

While most efforts to reconstruct ancient biochemistry have traditionally relied on building qualitative models of small pathways (2, 4, 5, 7, 38), we found that quantitative modeling of larger networks can provide substantial new insight into the origin of life. By computationally mapping geochemical scenarios to plausible ancient proto-metabolic structures, we estimated which portions of extant biochemistry may have been very sensitive or very robust to initial geochemical conditions. Our approach reveals that, contrary to expectations (8, 28, 39), environmental sources of fixed nitrogen and low-potential electron donors may have not been necessary for early biochemical evolution, and a substantial degree of complexity may have

emerged prior to incorporation of nitrogen into the biosphere (3). The key catalytic role played by nitrogen in the active sites of modern enzymes may have been preceded by positively charged surfaces or metal ions (18, 25), which could have been replaced by amino/keto acids with nitrogen side chains once nitrogen became incorporated into proto-metabolism. Our simulations also cast doubts on the essential role of a low-potential electron donor in early life (8, 28, 29), consistent with the proposal that low-potential electron donors may not be necessary for acetogenesis (40), and with the possibility that energy conservation via electron bifurcation might not have been necessary in primordial metabolism. The independence of our inferred ancestral networks of low-potential electron donors and ATP, both key substrates for nitrogen fixation (41), suggests that nitrogen fixation may have evolved later throughout the history of life (42–44). A striking feature of our analysis is the convergence of multiple geochemical scenarios towards a core organo-sulfur proto-metabolic network capable of producing various keto acids and fatty acids (Fig. 3B). This feature provides a window into how thioester-driven polymerization of α-hydroxy acid monomers (derived from producible keto acids) could have added primitive macromolecular organic catalysts (4) to initial inorganic minerals or metal ion catalysts (18, 25). Further tests of this hypothesis could be pursued by measuring the capacity of these polymers to catalyze key reactions in the network, and by exploring whether these organic compounds are produced in living systems today via mechanisms similar to polyketide or non-ribosomal peptide synthesis. Finally, our constraint-based models of this core organo-sulfur proto-metabolism provide a first example of how network expansion-based predictions can be translated into dynamical models, whose capacity to estimate sustainable collective growth can drive the search for specific self-reproducing chemical networks and metabolically-driven artificial protocells.

Acknowledgements

We thank all members of the Segre` Lab for helpful discussions. We acknowledge support by the Directorates for Biological Sciences (BIO) and Geosciences (GEO) at the NSF and NASA under Agreements No. 80NSSC17K0295, 80NSSC17K0296 and 1724150 issued through the Astrobiology Program of the Science Mission Directorate.

Contributions

J.E.G., H.H. and D.S. designed the research. J.E.G. wrote code, ran simulations and performed analysis. R.M. contributed to the non-equilibrium steady-state modeling. J.G. and D.S. wrote the manuscript. All authors read and approved the final manuscript.

Competing Financial Interests

The authors declare no competing financial interests.

Corresponding Authors

Correspondence to: dsegre@bu.edu.

References

1. Eck, R. V, and Dayhoff, M. O. (1966) Evolution of the structure of ferredoxin based on living relics of primitive amino Acid sequences. *Science*. **152**, 363–366

2. Hartman, H. (1975) Speculations on the origin and evolution of metabolism. *J. Mol. Evol.* **4**, 359–370

3. Hartman, H. (1992) Conjectures and reveries. *Photosynth. Res.* **33**, 171–6

4. de Duve, C. (1991) *Blueprint for a cell: the nature and origin of life*, Neil Patterson Publishers, Carolina Biological Supply Company, Burlington, N.C.

5. Morowitz, H. J., Kostelnik, J. D., Yang, J., and Cody, G. D. (2000) The origin of intermediary metabolism. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 7704–8

6. Smith, E., and Morowitz, H. J. (2016) *The Origin and Nature of Life On Earth*, 1st Ed., Cambridge University Press, Cambridge, United Kingdom, 10.1017/CBO9781316348772

7. Smith, E., and Morowitz, H. J. (2004) Universality in intermediary metabolism. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 13168–73

8. Sousa, F. L., Thiergart, T., Landan, G., Nelson-Sathi, S., Pereira, I. a C., Allen, J. F., Lane, N., and Martin, W. F. (2013) Early bioenergetic evolution. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **368**, 20130088

9. Ebenhöh, O., Handorf, T., and Heinrich, R. (2004) Structural analysis of expanding metabolic networks. *Genome Inform.* **15**, 35–45

10. Handorf, T., Ebenhöh, O., and Heinrich, R. (2005) Expanding metabolic networks: scopes of compounds, robustness, and evolution. *J. Mol. Evol.* **61**, 498–512

11. Raymond, J., and Segrè, D. (2006) The effect of oxygen on biochemical networks and the evolution of complex life. *Science*. **311**, 1764–7

12. Goldford, J. E., Hartman, H., Smith, T. F., and Segrè, D. (2017) Remnants of an Ancient Metabolism without Phosphate. *Cell*. **168**, 1126–1134.e9

13. Goldford, J. E., and Segrè, D. (2018) Modern views of ancient metabolic networks. *Curr. Opin. Syst. Biol.* **8**, 117–124

14. Keller, M. A., Zylstra, A., Castro, C., Turchyn, A. V., Griffin, J. L., and Ralser, M. (2016) Conditional iron and pH-dependent activity of a non-enzymatic glycolysis and pentose phosphate pathway. *Sci. Adv.* **2**, e1501235

15. Keller, M. A., Turchyn, A. V, and Ralser, M. (2014) Non-enzymatic glycolysis and pentose phosphate pathway-like reactions in a plausible Archean ocean. *Mol. Syst. Biol.* **10**, 725

16. Messner, C. B., Driscoll, P. C., Piedrafita, G., De Volder, M. F. L., and Ralser, M. (2017) Nonenzymatic gluconeogenesis-like formation of fructose 1,6-bisphosphate in ice. *Proc. Natl. Acad. Sci.* **114**, 7403–7407

17. Keller, M. A., Piedrafita, G., and Ralser, M. (2015) The widespread role of non-enzymatic reactions in cellular metabolism. *Curr. Opin. Biotechnol.* **34**, 153–161

18. Muchowska, K. B., Varma, S. J., Chevallot-Beroux, E., Lethuillier-Karl, L., Li, G., and Moran, J. (2017) Metals promote sequences of the reverse Krebs cycle. *Nat. Ecol. Evol.* **1**, 1716–1721

19. Vetsigian, K., Woese, C., and Goldenfeld, N. (2006) Collective evolution and the genetic code. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 10696–701

20. Ochman, H., Lawrence, J. G., and Groisman, E. A. (2000) Lateral gene transfer and the nature of bacterial innovation. *Nature*. **405**, 299–304

21. Lazcano, A., and Miller, S. L. (1996) The origin and early evolution of life: prebiotic chemistry, the pre-RNA world, and time. *Cell*. **85**, 793–8

22. Deamer, D., and Weber, A. L. (2010) Bioenergetics and life's origins. *Cold Spring Harb. Perspect. Biol.* **2**, a004929

23. Martin, W., Baross, J., Kelley, D., and Russell, M. J. (2008) Hydrothermal vents and the origin of life. *Nat. Rev. Microbiol.* **6**, 805–14

24. Russell, M. J., Hall, A. J., and Martin, W. (2010) Serpentinization as a source of energy at the origin of life. *Geobiology*. **8**, 355–71

25. Varma, S. J., Muchowska, K. B., Chatelain, P., and Moran, J. (2018) Native iron reduces CO2 to intermediates and end-products of the acetyl-CoA pathway. *Nat. Ecol. Evol.*

10.1038/s41559-018-0542-2

26. Dörr, M., Kässbohrer, J., Grunert, R., Kreisel, G., Brand, W. A., Werner, R. A., Geilmann, H., Apfel, C., Robl, C., and Weigand, W. (2003) A possible prebiotic formation of ammonia from dinitrogen on iron sulfide surfaces. *Angew. Chem. Int. Ed. Engl.* **42**, 1540–3

27. Navarro-González, R., McKay, C. P., and Mvondo, D. N. (2001) A possible nitrogen crisis for Archaean life due to reduced nitrogen fixation by lightning. *Nature*. **412**, 61–64

28. Martin, W. F., and Thauer, R. K. (2017) Energy in Ancient Metabolism. *Cell*. **168**, 953–955

29. Sousa, F. L., Preiner, M., and Martin, W. F. (2018) Native metals, electron bifurcation, and CO2 reduction in early biochemical evolution. *Curr. Opin. Microbiol.* **43**, 77–83

30. Halmann, M. (1974) Evolution and Ecology of Phosphorus Metabolism. in *The Origin of Life and Evolutionary Biochemistry* (Dose, K., Fox, S. W., Deborin, G. A., and Pavlovskaya, T. E. eds), pp. 169–182, Springer US

31. Schwartz, A. W. (2006) Phosphorus in prebiotic chemistry. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **361**, 1743–9

32. Keefe, A. D., and Miller, S. L. (1995) Are polyphosphates or phosphate esters prebiotic reagents? *J. Mol. Evol.* **41**, 693–702

33. Kanehisa, M., and Goto, S. (2000) KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30

34. Orth, J. D., Thiele, I., and Palsson, B. O. (2010) What is flux balance analysis? *Nat Biotech*. **28**, 245–248

35. Chandru, K., Guttenberg, N., Giri, C., Hongo, Y., Butch, C., Mamajanov, I., and Cleaves, H. J. (2018) Simple prebiotic synthesis of high diversity dynamic combinatorial polyester libraries. *Commun. Chem.* **1**, 30

36. Forsythe, J. G., Yu, S. S., Mamajanov, I., Grover, M. A., Krishnamurthy, R., Fernández, F. M., and Hud, N. V. (2015) Ester-Mediated Amide Bond Formation Driven by Wet-Dry Cycles: A Possible Path to Polypeptides on the Prebiotic Earth. *Angew. Chemie - Int. Ed.* **54**, 9871–9875

37. Henry, C. S., Broadbelt, L. J., and Hatzimanikatis, V. (2007) Thermodynamics-based metabolic flux analysis. *Biophys. J.* **92**, 1792–1805

38. Wachtershauser, G. (1990) Evolution of the first metabolic cycles. *Proc. Natl. Acad. Sci.* **87**, 200–204

39. Martin, W., and Russell, M. J. (2007) On the origin of biochemistry at an alkaline hydrothermal vent. *Philos. Trans. R. Soc. London B Biol. Sci.* **362**, 1887–1926

40. Bar-Even, A. (2013) Does acetogenesis really require especially low reduction potential?

Biochim. Biophys. Acta. **1827**, 395–400

41. Duval, S., Danyal, K., Shaw, S., Lytle, A. K., Dean, D. R., Hoffman, B. M., Antony, E., and Seefeldt, L. C. (2013) Electron transfer precedes ATP hydrolysis during nitrogenase catalysis. *Proc. Natl. Acad. Sci.* **110**, 16414–16419

42. Weiss, M. C., Sousa, F. L., Mrnjavac, N., Neukirchen, S., Roettger, M., Nelson-sathi, S., and Martin, W. F. (2016) The physiology and habitat of the last universal common ancestor. *Nat. Microbiol.* **1**, 1–8

43. Gogarten, J. P., and Deamer, D. (2016) Is LUCA a thermophilic progenote? *Nat. Microbiol.* **1**, 16229

44. Weiss, M. C., Neukirchen, S., Roettger, M., Mrnjavac, N., Nelson-Sathi, S., Martin, W. F., and Sousa, F. L. (2016) Reply to "Is LUCA a thermophilic progenote?" *Nat. Microbiol.* **1**, 16230

45. Mall, A., Sobotta, J., Huber, C., Tschirner, C., Kowarschik, S., Bačnik, K., Mergelsberg, M., Boll, M., Hügler, M., Eisenreich, W., and Berg, I. A. (2018) Reversibility of citrate synthase allows autotrophic growth of a thermophilic bacterium. *Science (80-.).* **359**, 563–567

46. Nunoura, T., Chikaraishi, Y., Izaki, R., Suwa, T., Sato, T., Harada, T., Mori, K., Kato, Y., Miyazaki, M., Shimamura, S., Yanagawa, K., Shuto, A., Ohkouchi, N., Fujita, N., Takaki, Y., Atomi, H., and Takai, K. (2018) A primordial and reversible TCA cycle in a facultatively chemolithoautotrophic thermophile. *Science (80-.).* **359**, 559–563

Figures

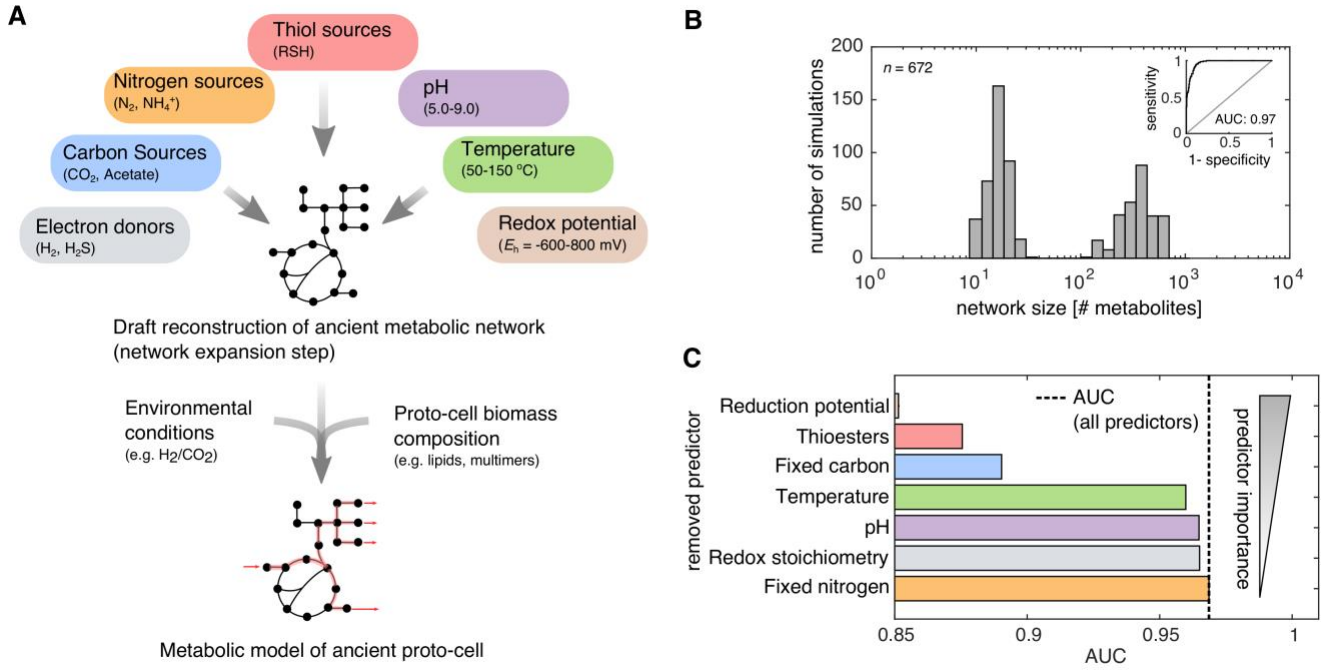


Figure 1: Nitrogen is not essential for the initial expansion of metabolism. (A) A thermodynamically-constrained network expansion algorithm was used to simulate the early expansion of metabolism under 672 scenarios, systematically varying the availability of reductants in the environment, pH, carbon sources, the presence of thiols, temperature and the availability of ammonia. (B) A histogram of network sizes (*x*-axis, number of metabolites) revealed that 43 % (288/672) of the scenarios resulted a bimodal distribution, where expansion occurred beyond 100 metabolites. (inset) A logistic regression classifier was constructed to predict whether a geochemical scenario resulted in a network that exceeded 100 metabolites, and a receiver operating curve (ROC) was plotted. The trained classifier resulted in an area under the curve (AUC) of 0.97 and leave-one out cross-validation accuracy of 0.89. (C) Models were trained without information on specific geochemical variables (*y*-axis), and the ensuing AUC was plotted as a bar-chart (*x*-axis), revealing that knowledge of the availability of fixed nitrogen offers no information on whether networks expanded.

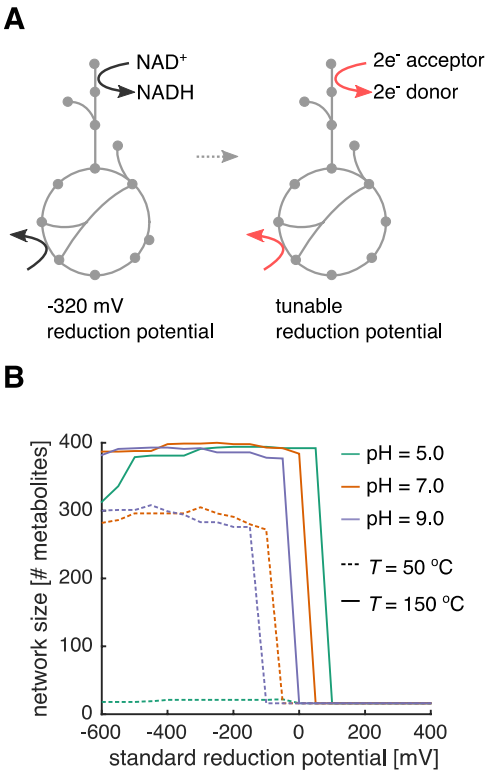
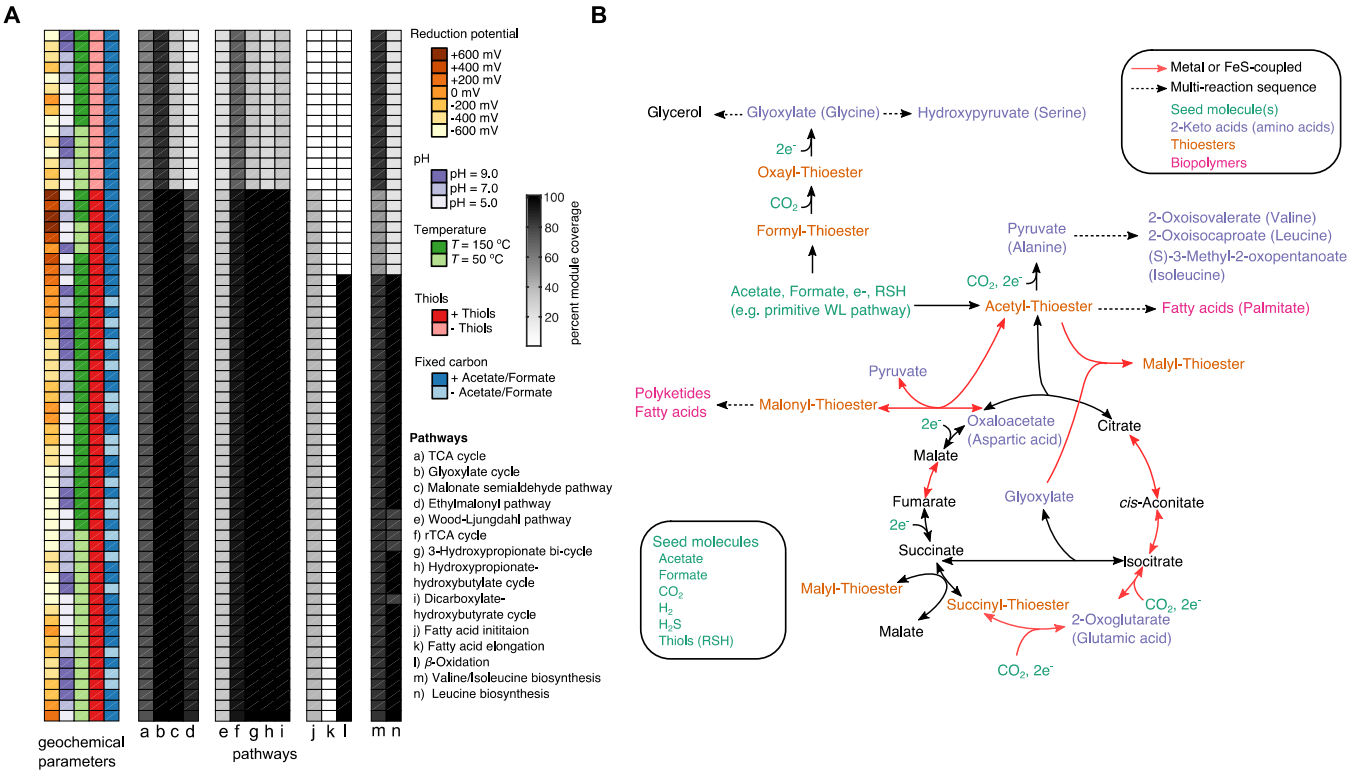


Figure 2: Reduction potential of nicotinamide and flavin substitutes influences network expansion. (A) Redox coenzymes (NAD, NADP, and FAD) were substituted with an arbitrary electron donor/acceptor at a fixed reduction potential. (B) We performed thermodynamic network expansion in acidic (pH 5), neutral (pH 7) and alkaline (pH 9) conditions at two temperatures (T =50 and 150 °C), using a two-electron redox couple at a fixed potential (x-axis) as a substitute for NAD(P)/FAD coupling in extant metabolic reactions (see Methods). We plotted the final network size across all pH and temperatures with no fixed carbon sources (e.g. only CO_2) and thiols. Notably, for these simulations, we used a base seed set of: H_2, H_2S, H_2O, HCO_3^-, H^+ and CO_2.



422

423 **Figure 3: Systematic exploration of prebiotic scenarios reveals a core organo-sulfur**
424 **network**. (A) A thermodynamically constrained network expansion algorithm was used to
425 simulate the early expansion of proto-metabolism under various scenarios, including the
426 availability of reductants in the environment, pH, temperature, and the availability of fixed
427 carbon sources and thiols. The proportion of molecules selected KEGG modules involved carbon
428 metabolism are plotted as a heatmap to the right of the parameters. (B) A representation of the
429 core network producible from a prebiotically plausible seed set without nitrogen or phosphate
430 (bottom left box). Acetyl-thioesters are first produced, potentially from a primitive Wood-
431 Ljungdahl pathway (8, 25) from acetate and thiols provided as seed molecules (green). Acetyl-
432 thioesters enable the production of all intermediates in the reductive tricarboxylic acid (rTCA)
433 cycle, with the exception of phosphoenolpyruvate. ATP-dependent reactions in the rTCA cycle
434 may have been substituted with a primitive malate sythase and transthioesterification of
435 succinate as well as the recently discovered reversible citrate synthase (45, 46). The keto acid
436 precursors for 8 common amino acids (A,D,E,G,I,L,S,V) are highlighted in purple, while routes
437 to thioester-mediated polymerization of fatty acids and polyketides are highlighted in pink.

438

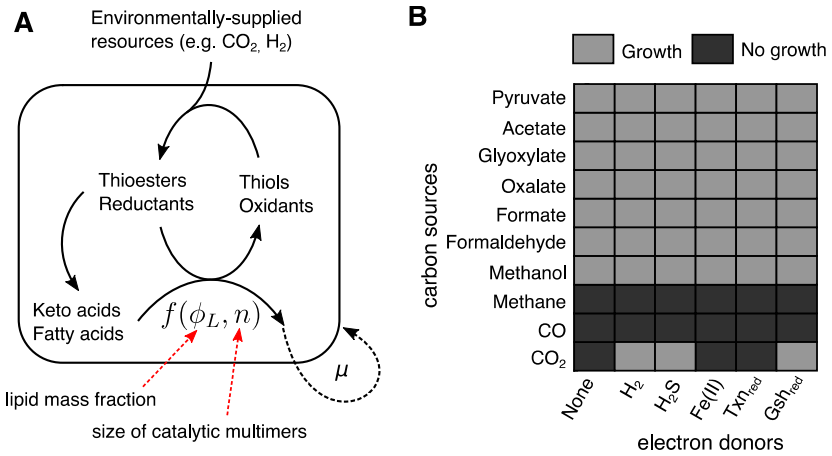


Figure 4: Constraint-based modeling of plausible ancient proto-cells. (A) We constructed a metabolic of model of plausible ancient proto-cell and used thermodynamic metabolic flux analysis (37) to simulate the maximum growth yields, steady-state fluxes, and metabolite concentrations under a variety of environmental conditions. The metabolic model was constructed using internally-generated reductants and thioesters that fueled biomass formation. The biomass composition was specified as variable fractions of fatty acids, and polymerized hydroxyacids from keto-acid precursors (see Methods). In this model, the internal redox coenzyme was assumed to be disulfide/dithiol at a standard reduction potential of -220 mV, and the production of biomass was fueled by the hydrolysis of acetyl-thioesters. We parameterized the biomass composition using a two-parameter model (see Methods), with the mass fraction of lipids in the proto-cell set to $\phi_L = 0.2$, and the the average size of a catalytic multimer, $n = 10$. (B) We simulated growth on a variety of simple carbon sources (*y*-axis) and electron donors (*x*-axis), and show environments supporting non-zero growth (light grey) or no growth (dark grey). Interestingly, H_2, H_2S and glutathione were the only reductants capable of supporting fully autotrophic growth on CO_2. Furthermore, CO and Methane could not support growth in this model, while other one-carbon sources like methanol, formate and formaldehyde could support biomass growth.