Modern views of ancient metabolic networks
Joshua E. Goldford\(^1\) and Daniel Segrè\(^{1,2,3,4}\)

**Abstract**

Metabolism is a molecular, cellular, ecological and planetary phenomenon, whose fundamental principles are likely at the heart of what makes living matter different from inanimate one. Systems biology approaches developed for the quantitative analysis of metabolism at multiple scales can help understand metabolism’s ancient history. In this review, we highlight work that uses network-level approaches to shed light on key innovations in ancient life, including the emergence of protometabolic networks, collective autocatalysis and bioenergetics coupling. Recent experiments and computational analyses have revealed new aspects of this ancient history, paving the way for the use of large datasets to further improve our understanding of life’s principles and abiogenesis.

**Addresses**

\(^1\) Bioinformatics Program and Biological Design Center, Boston University, Boston, MA 02215, United States
\(^2\) Department of Biology, Boston University, Boston MA 02215, United States
\(^3\) Department of Biomedical Engineering, Boston University, Boston MA 02215, United States
\(^4\) Department of Physics, Boston University, Boston MA 02215, United States

Corresponding author: Segrè, Daniel, Bioinformatics Program, Boston University, Boston, MA 02215, United States. (dsegre@bu.edu)

__Current Opinion in Systems Biology__ 2018, 8:117–124

_This review comes from a themed issue on Regulatory and metabolic networks (2018)_

Edited by Uwe Sauer and Bas Teusink

For a complete overview see the [Issue](https://doi.org/10.1016/j.coisb.2018.01.004) and the [Editorial](https://doi.org/10.1016/j.coisb.2018.01.004)

Available online 2 February 2018

2452-3100© 2018 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

**Keywords**

Metabolism, Evolution, Network expansion, Origin of life.

**Introduction**

The metabolic network of a cell transforms free energy and environmentally available molecules into more cells, moving electrons step by step along gradients in a complex energetic landscape \(^1,2\). The ability of a cell to efficiently and simultaneously manage hundreds of metabolic processes so as to accurately balance the production of its internal components constitutes a very complex resource allocation problem. In fact, it is only through recent systems biology research that we have begun to quantitatively assess this resource allocation problem at the whole-cell level \(^3\). A common perspective in the analysis of cellular self-reproduction is the notion that the genome, with its crucial information-storage role, is the central molecule of the cell, and that everything else can be collectively regarded as the machinery whose role is to produce a copy of the DNA. It is therefore not surprising that, as we struggle with the fascinating question of how life started on a lifeless planet, it is tempting to look for how a single information-containing molecule could arise spontaneously from prebiotic compounds. However, in spite of the appeal of thinking of DNA (or its historically older predecessor, RNA) as the central molecule that is being replicated in the cell, no molecule in the cell really self-replicates: the cell is a network of chemical transformations capable of collective autocatalytic self-reproduction. Collective autocatalysis is the capacity for an ensemble of chemicals to enhance or catalyze the synthesis or import of its own components, enabling a positive feedback mechanism that can lead to their sustained amplification. Combining this systems-level view of a cell with the argument of what is usually called the “metabolism first” view of the origin of life, one could propose that the ability of a chemical network to produce more of itself (or to grow autocatalytically) is and has always been a key hallmark of life \(^5\). An interesting modern version of this same principle is embedded in one of the most popular systems biology approaches for the study of whole cell metabolism: this approach, based on reaction network stoichiometry and efficient constraint-based optimization algorithms, is commonly known as flux balance analysis (FBA) \(^6\). FBA solves mathematically the resource allocation problem that every living cell needs to solve in order to transform available nutrients into the macromolecular building blocks that are necessary for maintenance and reproduction. When an FBA calculation estimates the maximal growth capacity of a cell, it essentially computes the set of reaction network fluxes that enable optimally efficient autocatalytic self-reproduction. While in cellular life this process is finely regulated and controlled, ancient life must have gone through many different stages of similar, but much less organized collectively autocatalytic processes. Thus, one of the key problems of the origin of life is the question of how an initially random path in the space of possible chemical transformations driven far from thermodynamic equilibrium could have ended up being dynamically “trapped” in a collectively autocatalytic state.

The focus on cellular self-reproduction as the fundamental level at which life and its origin should be
understood is however too narrow. An exciting recent development in systems biology of metabolism is the rise of methods to extend FBA models from the genome scale to the ecosystem level [10–12]. In addition to solving the resource allocation problem of metabolism for individual organisms in a given environment, these approaches take into account the fact that metabolites can be exchanged across species, giving rise to metabolically-driven ecological networks [11]. These advances suggest that metabolism may be best understood as an ecosystem-level phenomenon (Figure 1b), where the collective biochemical capabilities of multiple co-existing organisms may reflect — better than any individual metabolic network — an optimal capacity of life to utilize resources present in a given environment [9,13]. The ecosystem-level nature of metabolism is another feature of present-day life whose roots likely date back to the early stages of life on our planet. For example, the chemical networks that gradually gave rise to reproducing protocells may have wandered for quite some time in a broader chemical space, effectively generating molecular ecosystems before the rise of spatially and chemically well-defined cellular structures.

At an even larger scale, metabolism can be viewed as operating not just at the level of individual cells or ecosystems, but as a planetary phenomenon, in which cellular processes collectively affect (and are affected by) the flow of molecules at geological scales (Figure 1c). The strong coupling between the metabolic processes of ecosystems and planetary-scale geochemistry [14,15] suggest that biosphere-level metabolism should be viewed as one of the natural scales for the study of life’s history. A paramount challenge in the study of the origin of metabolic networks is to gain insight on the structure of metabolic networks before LUCA and before the rise of genetic coding (Figure 2). At the core of this challenge is the question of whether and how metabolic reactions, which depend on genome-encoded enzymes in modern cells,
could have been carried out without such enzymes, resulting in a classical origin of life “chicken-and-egg” problem. One possible way out of this conundrum is the possibility that some of these metabolic reactions were initially catalyzed by less sophisticated and less specific catalysts, such as small organic molecules, metal ions, minerals, short RNA polymers, prebiotic amino acids or peptides. These small molecules could have persisted throughout evolution, gradually becoming incorporated into protein enzymes as catalytic cores or cofactors [25,26]. Adding on to a large body of evidence on individual metabolic reactions being catalyzed by small molecules [27,28], recent experimental work has demonstrated that several key pathways found in modern day metabolic networks can be catalyzed non-enzymatically [29–34]. For instance, Ralser and colleagues [29–32,35] have shown the feasibility of non-enzymatic networks that resemble modern day biochemical pathways, including the TCA cycle, glycolysis and gluconeogenesis. In addition, Moran and colleagues have demonstrated that metals can selectively catalyze and drive portions of non-enzymatic reductive TCA cycle (rTCA) [33]. These experimental results support the hypothesis that the catalytic cores of some modern enzymes may represent evolved variants of simple geochemically available prebiotic catalysts like transition metals, iron-sulfur clusters or organic cofactors [36]. Despite these important advances, the known instances of non-enzymatic catalysis are still the tip of the iceberg relative to the large...
number of possible catalyst-reaction pairs. Future high-throughput experiments could greatly expand the scope of possible prebiotic chemical networks and test the limits of non-enzymatic catalysis, shedding important light on the complexity of prebiotic chemistry obtainable before the availability of proteinaceous enzymes.

From non-enzymatic catalysis to collective autocatalysis

The above examples illustrate the fact that chemical reactions, and whole pathways, typically viewed in biological context as feasible only in the presence of protein enzymes, could take place under much more primitive conditions, through the catalytic action of small molecules, minerals or even non-covalent supramolecular assemblies [37]. As mentioned above, however, a major leap in the history of life must have involved the rise of a collectively autocatalytic chemical system. The feasibility of pre-enzymatic chemistry suggests a possible path for the rise of such collective autocatalysis: if the molecules produced by these reactions are themselves good catalysts, or if these reactions contribute to solubilize inorganic catalysts, then it is reasonable to hypothesize that early metabolic network growth, such as the removal of likely toxic intermediates. Although further experimental and theoretical work has uncovered generic constraints of autocatalytic networks [38], and offered plausible biophysical mechanisms leading to sustained autocatalysis of biopolymer ensembles [39,40], Experimentally, Whitesides and colleagues have constructed an autocatalytic chemical network based on simple, biologically relevant organic compounds [41]. In particular, by using a continuous flow of nutrients into and out of their reaction vessel, they showed that simple mixtures of thiols and thioesters could display a wide range of dynamical properties, such as bistability, oscillations and autocatalysis. Notably, this work demonstrated the possibility that a thioester-based chemistry could fuel autocatalytic networks [41]. Future approaches could extend the use of network expansion models by incorporating additional constraints on metabolic network growth, such as the removal of likely toxic intermediates. Although further experimental and theoretical work is required to fully address the scope, implications, and fundamental limitations of an early phosphate-free biosphere, the use of the network expansion algorithm to explore plausible routes of abiogenesis represents an interesting research direction.

Navigating possible paths from primordial to present-day networks

Studies in evolutionary biology suggest that biological systems evolve by partially building on prior innovations. If this principle extends back to the origin of metabolic networks, then it is reasonable to hypothesize that early proto-metabolic networks were based on previously accessible chemistry. Such logic leads to the conjecture that the structure of metabolic networks encodes the evolutionary history of metabolism, and that the chemistry of core metabolism is similar to the initial abiotic chemical networks that lead to life’s emergence [43,44]. This conjecture is supported, as discussed above, by experimental work demonstrating that a significant portion of core metabolism is accessible without the use of proteinaceous enzymes. While these concepts have been heavily utilized in origin of life research, recent efforts have transformed this conceptual paradigm into an algorithmic and quantitative framework using metabolic network modeling [19,45]. A modeling approach recently used to explore the plausible evolutionary history of very early stages of biochemistry is the network expansion algorithm (Figure 2b), which iteratively simulates the growth of new metabolites and reactions starting from an initial seed set [46–48]. We used the network expansion algorithm to construct a model for ancient prebiotic metabolism, specifically addressing the question of whether any portion of current biochemistry could have possibly emerged in the absence of phosphate (and thus prior to transcription/translation) [45]. Models of prebiotic networks were constructed starting from minimal sets of compounds thought to have been readily available on early Earth. Notably, even if these initial compounds did not include any phosphate-containing molecule, a surprisingly large expanded network could ensue, covering several pathways that are part of central metabolism today, and of previously proposed models of biogenesis [44]. This finding is consistent with the possibility that thioesters, sulfur-based energy rich chemical moieties, could have predated phosphates as energy carriers in the cell, providing the required thermodynamic driving force. Interestingly, recent work has experimentally demonstrated the possibility that a thioester-based chemistry could fuel autocatalytic networks [41]. Future approaches could extend the use of network expansion models by incorporating additional constraints on metabolic network growth, such as the removal of likely toxic intermediates. Although further experimental and theoretical work is required to fully address the scope, implications, and fundamental limitations of an early phosphate-free biosphere, the use of the network expansion algorithm to explore plausible routes of abiogenesis represents an interesting research direction.

Although the majority of chemical reactions important in early living systems may still be encoded in modern day living systems, there is also a possibility that key reactions and compounds initially critical for living systems were lost throughout the course of evolution. Even more broadly, it is plausible that a much bigger space of chemically possible reactions could have given rise to an organized metabolism [49,50]. As shown in recent elegant experimental work, molecules important for life as we know it may in principle be producible through
Beyond these realistic chemical spaces, biochemical organization has been studied extensively using simplified toy models based on artificial chemical rules, such as the so-called “string chemistries” \[54\]. Similar approaches were the foundation of some of the early work on collectively autocatalytic networks \[8,55\]. More recently, a very simple string chemistry, simulated and analyzed using systems biology approaches (including FBA \[3\]) yielded a family of optimally efficient pathways, some of which resemble functionally and topologically the rTCA cycle network \[56\]. An artificial chemistry which incorporated catalytic polymers with a toy folding process was recently shown to be helpful towards explaining the emergence of polymer-based structures within a compositional inheritance world \[39\]. In addition to serving as a basis to explore possible scenarios for the emergence of metabolism, abstract chemistry models can be very helpful in the exploration of statistical physics-based models of non-equilibrium chemical systems \[57\].

**Overcoming energy barriers, then and now**

Whether realistic or abstract, ancient or modern, any metabolism can operate only if kept far from thermodynamic equilibrium by an external free energy source. Thus, to achieve a working theory for the origin of metabolism, one should identify not only sources of materials, but also sources of free energy consistent with geochemical data. Effectively, even if early life may have extensively used abiogenic organic material heterotrophically, this question largely hinges on our understanding of what free energy source could have fueled the production of electron donors capable of reducing abundant gases like CO$_2$ and N$_2$ into the reduced forms readily used by biological systems. Two potential sources include chemical energy from hydrothermal vents, and photochemical energy from solar (especially UV) radiation \[58\]. The former scenario is consistent with recent phylogenetic studies \[22\], where chemical energy in the form of molecular hydrogen is used to fix carbon dioxide using a variant of the Wood-Ljungdahl pathway in LUCA. However, it is unclear whether this scenario would be compatible with thioesters as a key component for free energy transduction, given that these molecules have been recently shown to be highly unstable in simulated hydrothermal systems \[59\]. Interestingly, UV light can support the synthesis of organic molecules \[60,61\] and iron-sulfur clusters \[62\] as well as drive the reductive steps in the rTCA cycle \[63\]. Future work exploring the potential roles of various energy sources to fuel non-enzymatic prebiotic networks will be important in determining plausible models for ancient metabolism.

Even if a source of free energy is available, a major open question in the evolution of bioenergetics is the rise of coupling between driving forces and driven reactions. Through this coupling, currently enabled by large proteins, reactions that dissipate free energy (e.g. thioester or phosphodiester bond breaking) drive reactions that require a free energy input. Such couplings have recently been proposed to universally operate as a “Brownian ratchet”, in which enzyme complexes rely on the step-wise, gated mechanism of highly coordinated multi-domain enzymes \[2\]. Martin and colleagues proposed that electron bifurcation, the most recently discovered energy conserving process \[64,65\], may have been the first mechanism through which ancient metabolic networks coupled free energy sources to drive endergonic reactions. Electron bifurcation is a mechanism that enables coupling between available, mid-potential electron donors (e.g. H$_2$) and acceptors (e.g. CO$_2$) to generate low-potential electron donors. This mechanism is for example capable of producing reduced ferrodoxin, an energy source common in diverse biochemical pathways like photosynthesis and methanogenesis \[66\]. In general, identifying the scope of non-enzymatic analogs for such free energy coupling processes remains an open challenge, and efforts to this end will undoubtedly shed light on the earliest phases of bioenergetic evolution.

**Towards data-driven origin of life research**

As the above examples clearly illustrate, origin of life research is a multidisciplinary endeavor, requiring consideration of multiple, increasingly large datasets (chemical, geological, biological, physical) for both experimental and computational analyses \[67\]. Currently available databases that may be useful for the study of ancient life range from collections of genetic and phenotypic diversity of microbial species and communities \[68\], to “knowledge-base” resources available for exploring metabolites, reactions and biochemical pathways \[69\]. As origin of life research may require data from broader categories of molecules and reactions beyond present-day biochemistry \[51\], databases of known organic and inorganic chemicals \[70\] and reactions \[50\] will constitute important components of future attempts to reconstruct the first biochemical processes. Other categories of data relevant to the ancient history of metabolism are available in more specialized databases \[71,72\]. Future efforts could assemble other databases useful for the computational analysis of prebiotic chemistry, including a database of documented prebiotic chemistry experiments. Furthermore, and most importantly, a standardization of
experimental and computational results would enable comparisons across different efforts, allowing researchers to build more systematically on previous work. Integrating data from various sources, ranging from prebiotic chemistry experiments to inferred early Earth geochemical data, could allow for the construction of large-scale models of ancient metabolic states at unprecedented levels of resolution.

Future work aimed at understanding early life will increasingly benefit from ongoing synthetic biology efforts towards the implementation of minimal living systems, and from quantitative approaches developed for systems biology of metabolism [73]. It would be highly beneficial for origin of life research to embrace theory and modeling as essential tools for transforming data and hypotheses into testable, nontrivial predictions, i.e. predictions whose outcome may not be known a priori, and whose validation or falsification may be clearly achievable, even if technologies may be years away from feasibility. Conversely, the study of early metabolism has a chance to provide new tools and ideas for how to move systems biology approaches beyond the current paradigms. For example, the exploration of putative early metabolic pathways not known in present-day organisms bears some similarities with the huge and challenging efforts of annotating metabolic enzyme functions in newly sequenced genomes and metagenomes [74]. Furthermore, biosphere-level analyses of ancient metabolism [45,48] could inspire new approaches for studying the collective biochemistry of microbial ecosystems.

Acknowledgements

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.

References

Papers of particular interest, published within the period of review, have been highlighted as:
* of special interest
** of outstanding interest


2. Branscomb E, Biancalani T, Goldenfeld N, Russell M: Escape-moment mechanisms and the conversion of disequilibria; the engines of creation. Phys Rep 2017, 677:1–60. Branscomb et al. provide critical insight on the attributes of all life’s molecular machines that harvest disequilibria, demonstrating that “brownian ratchet” mechanisms are widespread and critical for the fidelity of reactions that coupling endergonic reactions with dissipative exergonic reactions.


Moore et al. integrate geochemical data and enzymatic information across the tree of life to demonstrate the strong coupling between the availability of metal ions in the Archean ocean with the metal usage in modern enzymes, highlighting metabolic pathway innovations that reshaped the major biogeochemical electrical circuits on geological time scales.


The authors use a phylogenetic approach to estimate the metabolic repertoire of LUCA, suggesting that LUCA contained at least 355 protein families. This study suggests that LUCA was complex and highly evolved, including genes involved in carbon fixation (WL-pathway) and nitrogen fixation, and indicates that LUCA lived in thermophilic environments with sources of iron, sulfur and hydrogen gas, consistent with proposed ancient hydrothermal vent systems.

This paper demonstrates that a chemical network of small molecules and random peptides: an unexplored reservoir of potentially functional primitive organocatalysts. The case of seryl-histidine. Life 2017, 7:19.


This study used high-throughput metabolomics to screen for potential catalysts that can interconvert compounds in the TCA cycle, finding that iron sulfides and sulfate radicals could serve as catalysts for reactions that span the TCA cycle, glyoxylate shunt and the succinate semialdehyde pathway.


This paper demonstrates that glyoxylate and hydrogen peroxide can fuel reaction cycles capable of producing organic molecules (including malate and oxaloacetate) typically observed in the TCA cycle, suggesting that protometabolism may have utilized pathways different than the current ones.


This paper demonstrates that a chemical network of small molecules and thioesters is capable of autocatalytic and oscillatory dynamics, suggesting that prebiotic chemical networks may have been capable of both “memory” and dynamics typically associated with complex enzyme-based biological networks.

In this work, Mansy and colleagues demonstrate that iron-sulfur clusters can be formed from mixtures of Fe$^{3+}$ and thiolates (e.g. glutathione) upon irradiation using UV light. The presence of sulfide leads to both [2Fe–2S] and [4Fe–4S] complexes, which are components of enzymes proposed to be part of ancient metabolism.


This paper demonstrates that the synthesis of clay minerals, which are plausible prebiotic catalysts, can be catalyzed by the TCA cycle intermediate succinate.


