The autotrophic origins paradigm and small-molecule organocatalysis

Eric Smith
Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501

Harold J. Morowitz
Krasnow Institute for Advanced Study, George Mason University, Fairfax, VA 22030
(Dated: October 17, 2009)

Autotrophy and heterotrophy: disentangling the issues

All origin of life paradigms suppose at some level that cells have incorporated organic molecules that were once of abiotic origin. The distinction between primordial organosynthesis by high-energy processes quite different from those of biochemistry, and geochemical processes posited to be continuous with biochemistry, has become framed as a distinction between “heterotrophic” and “autotrophic” origin scenarios. The naming draws on an analogy with heterotrophy versus autotrophy of organisms, which is convenient but potentially misleading. Core anabolism is universal at the ecosystem level, meaning that organism autotrophy or heterotrophy is an ecological distinction rather than a chemical one. To cast an analogy between the heterotrophic organism in its ecological environment, and all of emerging life in an environment of distinct organic chemistry, overemphasizes the organization of cellular hierarchy as a model for the organization of life, and underemphasizes the aspects of biochemistry not directly linked to hierarchical control.

If we suppose that modern life has preserved part of the inventory of primordial organic molecules – for whatever reasons of either their chemistry or commitments to higher-level molecular assemblies – then the major chemical distinction between autotrophic and heterotrophic paradigms concerns whether their mechanisms of synthesis were conserved or replaced, and whether the original molecular inventory was similar to the universal core today. By “mechanism” here we refer to the substrate-level architecture of the pathway and the elementary bond transitions; the extreme efficacy and selectivity of biological catalysts is a separate problem of emergence of higher-order structures, which either paradigm must address.

The autotrophy/heterotrophy divide does not involve significant disagreement about sources and forms of abiotically produced organics. All of the following sources are believed to have contributed to early-earth chemistry: low-temperature synthesis on dust and asteroidal surfaces or cometary ices delivered by impacts [1]; free radical and ionization reactions in the atmosphere [2, 3], and reducing reactions at the tectonically active lithosphere-hydrosphere interface [4–7], potentially augmented with catalytically active mineral surfaces [8]. The problems for deciding between autotrophic and heterotrophic origins therefore come from uncertainty about mechanisms of organization. Which abiotic molecules and mechanisms could ever have become incorporated into networks capable of producing high molecular complexity and of permanently colonizing the geosphere? Did such incorporation depend on information encoded in higher-order structures such as oligomers, or was it more plausibly driven by chemical kinetics without hierarchical organization or explicitly informatic molecules?

The problem of the emergence, selection, and persistence of a biosynthetic network is not easily separated from questions about higher-order organization. The rate enhancement and selectivity of oligomer RNA and polypeptide catalysts are so powerful that – if their emergence from an unsupervised chemical medium were not so hard to explain – they might seem to provide a plausible route to replace practically any abiotic synthetic mechanism or to innovate any new pathway. The organization of such a “top-down” controlled metabolism would most naturally be explained by a process of Darwinian selection. The assumption that this is the dominant organizing principle is implicit in many RNA-first scenarios for the origin of life, and has led to the explicit proposal that metabolism is a “palimpsest” of the RNA world [9]. In our other abstract (on confederacy) we have mentioned probabilistic arguments [10] that emergence and stability may be more likely for a cluster of partly autonomous metabolic modules than through a mechanism that relies heavily on top-down control.

Here we argue that a key quantitative feature of metabolism to be explained is the number of universal small core metabolites. It is about 300 [11] – a number much larger than the number of comparably complex molecules appearing with non-vanishing probability in a Gibbs equilibrium ensemble, but much smaller than the $10^7$ molecules of comparable complexity indexed by PubChem [http://pubchem.ncbi.nlm.nih.gov/]. Whatever organizational mechanism led from abiotic organosynthesis to biochemistry must have enabled significant complexity, while severely limiting the elaboration of that complexity. A restatement of the observation is that metabolism is sparse within the chemical possibility space, and that this sparseness may be seen as a result of selection. The key question then becomes what part of this selection was performed by chemical kinetics prior to hierarchy or acting independently of it, and what part was performed by Darwinian mechanisms using hierarchical integration and top-down control. Whether an organizing mechanism preserves or replaces molecules
and synthetic pathways will determine whether our understanding of extant biochemistry is relevant to reconstructing stages of origination.

We currently lack a principled, quantitative, chemically explicit theory of the requirements to stabilize a metabolic network and a self-generated control system far from thermodynamic equilibrium. However, the last ten years has seen rapid growth in knowledge and understanding of small-molecule organocatalysis [12, 13], which may offer new inputs to a theory of non-hierarchical chemical self-organization and selection. The term “organocatalysis” has been adopted to refer to catalysis by small organic molecules not making use of transition metals in the catalytic mechanism, as distinct from organo-metallic catalysis which is now an established field. It is not intended to include the study of macromolecular enzymes either, as this study is established within biochemistry.

**Organocatalysis: a bridge from self-organization to evolution?**

In autotrophic scenarios that credit a significant part of the organization of metabolism to primordially selected networks, the mechanism believed to account for the emergence and selection of reactions is autocatalysis. This may take the form of catalysis of reactions within the network by single molecules produced somewhere in the same network, or by topological motifs capable of amplifying their substrates (a phenomenon termed “network autocatalysis”). The idea is not new. It is the basis of Eigen’s concept of the hypercycle [14, 15], and has been applied in conceptual models by numerous researchers [16–22]. It is usually invoked for polymers, because their catalytic potential is widely recognized, but it is also usually represented abstractly, because the space of catalytic mechanisms is complex, and the prediction of catalytic power and selectivity from structural features is not well developed.

Recent advances in the study and generalization [13] of mechanisms of catalysis by small molecules have the potential to convert abstract models of autocatalysis into quantitatively realistic simulations that can be used to guide experiment. Early work with explicit applications to origin of life was the survey by Pizzarello and Weber of the amino acids for catalysis of aldol condensations [23, 24]. Of interest in those studies was not only catalytic efficacy, but enantioselectivity with the potential to propagate chirality through reaction networks. More recent work has shown not only that small biomolecules can catalyze reactions, but that the mechanisms of catalysis can often be related to known enzymatic mechanisms [13]. Much of the work in this field has been driven by industrial synthetic chemistry, a highly quantitative discipline which has made heavy use of metallic and organometallic catalysts, and is now seeking lower-cost, non-hazardous catalysts to control synthesis and in some cases enantioselectivity [12].

The distinct roles which the same molecule can fill within a non-hierarchical network, as catalyst or as reactant, opens the possibility for more nuanced theories of early biogenesis than a pure opposition between autotrophy and heterotrophy. Arguments for autotrophic origins often turn on the steady availability and high concentration of geochemically generated organics relative to deposits from space infall or atmospheric reactions [25]. More generally, the Oparin-Haldane conjecture [26] that exogenous molecules fed the first higher-order assemblies requires that pathways have been discovered to produce particular molecular seed species before “the soup is exhausted” [27]. While we find these arguments convincing for the reactants in a network, they are less restrictive for its catalysts. The most robust exogenous organics, such as polycyclic aromatic hydrocarbons, graphenes, etc., could contribute to organic gels when convected through hydrothermal systems, serving as differential diffusion barriers, adsorbants, dielectric contrasts to aid phase-transfer catalysis, or alignment sites for nucleobases or other planar molecules. Not being consumed by reactions, they would be under less pressure from exhaustion, and their functions could successively be replaced by endogenously generated molecules and eventually by the whole micro-environments of protocells.

The distinction between catalyst and reagent opens a quantitative question about the transition from geochemical self-organization to early evolution: are catalysts in a network more easily replaced than reagents as the network flux and molecular complexity increase? We have proposed this asymmetry as a path from protometabolism to the RNA world [28], as a continuation of our proposal for an autocatalytic loop between nucleotide biosynthesis (from amino acid precursors) and amino acid synthesis from citric-acid cycle precursors (with nucleotides acting as catalysts) [29]. This proposition has the corollary that the fitness of oligomers would be dominated by their ability to support existing pathways rather than by competitive self-replication, causing complex life to preserve rather than to over-write primordial pathways. It will be important to test this proposition both for particular molecules and as a general principle of network evolution.

**Autocatalytic small-molecule networks: data and models**

Two research thrusts are immediately suggested, to develop our knowledge of organocatalysis and to apply it to the origin of life. The first is a comprehensive survey of the small universal core metabolites [11] for catalytic efficacy, specificity, and enantioselectivity on standard organic reactions. Priority may be given to those reactions that make up the bulk of core biosynthesis and which are supported in extant life by cofactors as group-transfer agents. This survey could be coordinated by...
a small core of researchers to be performed in parallel across a wide range of teaching institutions looking for low-cost, finite scope projects. Perhaps the SETI@home sky survey [http://setiathome.ssl.berkeley.edu/] provides a model for task division and aggregation of results.

The second thrust is to develop the theory of large-network autocatalysis with realistic chemical stoichiometry. An important difference between origin-of-life questions and those that arise in industrial synthesis is that industrial organocatalysts – like their metallic precursors – are exogenously supplied and controlled, whereas positive feedbacks and non-equilibrium network growth are of primary interest to origins. Network simulation by origins researchers has so far depended heavily on numerical simulation or simplification of reaction stoichiometry, but these limitations may be surmountable. Algorithms currently exist to generate molecules in SMILES format from SMIRKS [http://www.daylight.com/] or other representations of reactions [30]. Working from a basis of such networks and a database of catalytic species or functional groups, it may be possible to estimate whether positive feedbacks are rare or common, and whether they have the effect of focusing flow of energy and materials so as to select sparse networks comparable to those in biochemistry. If the broad statistical features of large chemical networks can be characterized and empirically validated, monte carlo models and simulated dynamics may become useful to make quantitative predictions.

The common theme in our proposals is to increase the heterogeneity of model elements without sacrificing chemical realism, so that our model systems will become more plausible approximations to early earth. This theme is not limited to organocatalysis. Knowledge gained in biochemistry about mechanisms of transition-metal catalysis in organic molecules – particularly cooperative mechanisms involving atoms of two different metals – may be used to revisit catalysis by natural minerals. Surveys of mineral catalysis of biologically relevant reactions could provide both initial conditions and parallel inputs for networks organized by organocatalysts.

[25] Robert Shapiro. Small molecule interactions were central...


