Unitary origin of life versus origination through modules

We begin with the observation that all experimental and theoretical approaches to origins of life, which attempt to study the emergence of order in isolated components of the living state, are de facto commitments to describing life in terms of modules. Laboratory conditions or hypothesized early worlds provide contexts sufficiently less complex than modern cells to be tractable to us, in which the requirements for order within different subsystems can be studied independently. If we assume that lessons learned about subsystem organization remain informative about order in living systems, we have assumed that in real life the subsystems interact either sufficiently weakly or through sufficiently simple interfaces that the internal subsystem dynamics are not qualitatively unlike those we can reproduce.

All of these assumptions can be doubted, since we do not observe the heterogeneous, refined, particular chemical and processes of life separately in any non-living system, or in any less complex contexts than fully integrated autotrophic bacterial or archaeal cells. Yet as a practical matter the study of modules is unavoidable. Moreover, the assumption that life has a modular architecture and that it emerged in a sequence of stages has strong likelihood (in a Bayesian sense) based on our understanding of random processes, and it is the only one to which scientific method can be applied. The likelihood argument has been cleanly presented by Simon [1]: the only route for stochastic processes to produce structures of considerable complexity from unstable elementary steps (such as those deviating from thermodynamic equilibrium), without invoking events of extreme improbability, is through formation of modules which are stable intermediate points of construction.

The two central questions in choosing experiments or theories are then: 1) which subsystem boundaries will be most informative about the major transitions that led to life as we now see it, and 2) does a large enough quantitative difference exist between intra- and inter-system interactions to permit useful study of modules in simplified contexts? The likelihood (in the sense of Simon’s argument) for a putative modular description depends not only tractability of study, but on the probability that the proposed modules could have been stable holding points for a stochastic process in conditions plausible for the early earth.

From this perspective the difference between genes-first and metabolism-first approaches to origins arises from whether individuals and population dynamics, or physiological and ecological universals, are preferred as grounds for decomposition. The genes-first approach draws on a heavy (if often unstated) influence of the Central Dogma in thinking about molecular biology, and of population genetics on thinking about evolutionary dynamics. These two thought systems mesh coherently, because both take the Darwinian individual (capable of preserving heritable variation, replicating, and perishing) as the natural unit, with information input through selective processes at the population level, and phenotypes produced more or less mechanistically from genotypes. Limitations of this point of view are a somewhat impoverished treatment of the constraints of development and ecology which may structure the space of possible phenotypes a priori, and a presumption of a quite complex entity capable of functioning as a Darwinian individual. To some degree the latter complexity can be displaced from the individual to its context, as when the replication-selection model is applied to RNA oligomers in an Eigen hypercycle, competing for activated monomers in a primordial soup.

The contrary feature of metabolism first – as we will represent the idea here – is a diminished emphasis on both the individual and any distinctively Darwinian dynamic, and much stronger emphasis on coherence in biosynthetic pathways across organisms, spanning the organism/eco-system distinction, and connecting biochemistry to geochemical context. As for genes-first paradigms, we will argue that feedbacks between modules essential to maintaining modern cellular life could have been weak or absent in earlier life. However, rather than draw a boundary between a controller (RNA) and a controlled substrate (the metabolome) and remove the feedback by supplying the metabolome exogenously (a primordial soup), we will argue that the feedbacks between universal biosynthetic modules arise when these modules form micro-environments for each other energetically and chemically, and we will propose that it was
the support from such environments which was originally provided exogenously and perhaps independently by geo-chemical environments.

A further test for the appropriateness of a modular decomposition of the living state is that modules proposed to be primordial should retain some recognizable autonomy from one another in extant life. The autonomy may be manifest as partial independence from control through hierarchical systems (such as the genome or descent within lineages), or as independent response to environmental pressures that is regular across systems. Partial autonomy is to be expected as more than a vestige of a past state. A corollary to the Simon likelihood argument is that modules are effective routes to complexity precisely because they do not depend on integration within higher-order systems for their stability. Instead, they contribute autonomous stability to the systems comprising them, reducing the complexity of the higher-order assembly problem (while at the same time limiting the forms it can take).

Both genes-first and metabolism-first decompositions meet this criterion, although in different ways. Support for a division between the genome and cell physiology can be drawn from the partial independence of viral or transposon dynamics from host-cell lineages. For our metabolically rooted modularity, the support is physiological and evolutionary plasticity between pathway networks and cellular energy systems: for example, the way diverse and heterogeneous redox couples are converted to a universal currency of phosphate esters through the machinery of oxidative phosphorylation, or the way sugars have been substituted for amino-sugars as structural elements in nitrogen-limited plants versus nitrogen-sufficient bacteria. An important question about extant life which would contribute to assessing the relative likelihoods of gene-first versus metabolism-first decompositions – which we do not yet have tools to answer – is how much of the stability of biosynthetic pathways owes to feedback through population-genetic mechanisms depending on the hierarchy of genomic control, and how much reflects environmental constraints that preclude other solutions, perhaps expressed through a form of absolute normalizing selection or lack of evolutionary divergence.

We believe that explicitly regarding life as a confederacy, and integrating the study of its modules with attempts to reproduce stages of origin, reframes existing problems and opens new ones in useful ways. It emphasizes diverse contributions to the stability of the living state, which range from chemical kinetics to trophic ecology. It suggests that the emergence of individuality – of different kinds, at several events – is only one aspect of the emergence of life, but it is the particular aspect that mediates the emergence of Darwinian dynamics as distinct from other forms of geochemical self-organization. Finally, it suggests that we must explain both the existence and the limits to hierarchies of control systems, and that retracing the difficulties of origins may be a useful way to do so.

Suggestions of modular organization in modern life

Descriptions of life as hierarchal or modular may be made in many ways. We will emphasize modules which follow common divisions in the network topology, chemistry, and energetics of core biosynthesis, and which are then recapitulated at higher levels of cellular organization and physiology, and evolutionary diversification. In all cases we propose modules at the deepest level which appear to be universals of all known life, suggesting either that they antedate all forms of adaptive variation and were “frozen” into the structure of life by the dependence of other systems upon them, or that they are unique solutions to certain problems of function within a self-maintaining system. In the latter case they would define a grey area where strongly convergent evolution becomes indistinguishable from “physical constraints” against evolutionary variation.

The network topology of core biosynthesis suggests a natural decomposition into reaction groups with high internal inter-dependence, separated by “gateway” molecules or reactions. Following the “robust, yet fragile” topology observed for a minimal biosynthetic chart derived from *Aquifex aeolicus* [2, 3], we suggest the following as functionally integrated modules. Fig. 1 shows the major reactions in four carbon fixation pathways and a subset of the universal biosynthetic chains. 1) Citric acid or TCA cycle arcs or loop for arriving at core carbon skeletons. 2) The gluconeogenic pathway connected to the TCA reactions through pyruvate or phosphoenol-pyruvate, with 3) the reductive pentose-phosphate (Calvin-Benson) network as an elaboration of aldol and retro-aldol condensations about the gluconeogenic (or its reverse, the glycolytic) pathway. (Note that whether or not the Calvin-Benson pathway is used for carbon fixation, several of its reactions are key steps in the synthesis of electron-transfer cofactors and aromatic amino acids.) 4) The fatty-acid synthesis pathway from malonate and 5) the isoprenoid synthesis pathway from acetate represent two recursive chain-elongation pathways linked to the TCA reactions through acetate. The biosynthesis of amino acids is more complex but still regularly structured, clustering into simple syntheses from citric-acid intermediates with small number of reactions, and more complex syntheses drawing on more remote modules of the metabolic chart such as the module through chorismate for the aromatics [2, 4].

Redundant chemistry, chemical energetics, and serving cofactors often align with the network modules. Reduction reactions within the TCA cycle take CO₂ to acetate, an exergonic transformation. Further exergonic reductions then begin from acetate in fatty acid and isoprene synthesis, linking the a-keto-acid and lipid networks energetically [5]. Thioester to phosphate substrate-level phosphorylation [6] at two key steps (carboxylation of acetyl-CoA and of succinyl-CoA) is a key feature of redox-to-phosphate energy conversion in the TCA reactions. Phosphorylation is a distinctive feature of
the gluconeogenic pathway and the aldol condensations that branch from it, often removing OH groups from the network of available aldol condensations, as when glyceraldehyde-3-phosphate and dihydroxyacetone phosphate condense to fructose-1,6-bis-phosphate. A recurrently used small set of phosphoryl transfers, amino transfers, and reductive aminations, used in the early stages of amino acid biosynthesis, is remarkably stereotypically arranged in the same patterns as base assignments in the genetic code [4].

Cofactors, which mediate the recurrently used organic reactions as catalysts or group-transfer agents, are correspondingly used unequally across modules. Coenzyme-A functions in thioester formation in some TCA reactions and fatty acid synthesis. ATP as phosphoryl donor is active in dehydrations. NAD is the primary source of reductants, with flavins and deazaflavins secondary. Ammonia enters the organic nitrogen cycle almost exclusively through the formation of glutamate and glutamine, which act as amine transfer agents in most secondary aminations. S-adenosyl-methionine, folates, and pterins act as carriers of C\(_1\) groups in various states of reduction.

We attach significance to the chemical and bienergetic divisions among modules because they mimic divisions among energy sources and energy carriers in geochemical processes, particularly hydrothermal processes acting at the interface of the hydrosphere with the tectonically active lithosphere. Geothermal mantle convection brings reduced metals into contact with seawater, producing copious reductant, and volcanic activity produces dehydrated phosphates at least in surface environments. Strong pH gradients are produced in most hydrothermal systems, with vent effluents ranging from highly acidic for magma-hosted systems to highly basic for peridotite-hosted systems [7]. Thus biochemistry has preserved the distinctive chemical structure-forming capacities of the three major geological energy sources.

The module boundaries we have suggested may be seen again (more weakly) in signatures of evolutionary conservation. Elaborations of carbon fixation (six forms are now known; four are shown in Fig. 1) have conserved the core modules. TCA arcs run either as a loop autocatalytically in reductive TCA organisms, or in parallel along oxidative and reductive branches in acetogens and methanogens, suggesting that a commitment to these precursors was formed before divergence of these two deepest-rooted carbon fixation strategies. The 3-hydroxypropionate pathway parallels the reductive TCA arc from a precursor in malonate that is the gateway to fatty acid synthesis, and then uses the glyoxylate bypass to reach other synthetic precursors [8]. As noted, the Calvin-Benson network reverses the direction of the gluconeogenic pathway to feed reduced carbon into TCA reactions via pyruvate. Enzyme conservation across clades shows similar boundary distinctions. Attempts to use phylogenetic weighting to arrive at principled likely gene inventories in the LUCA (Ref. [9], Fig. 8) show strikingly little variability for enzymes governing core biosynthesis from acetate, and much greater variability for chemically homologous reactions from succinate, recapitulating the apparently very old divergence between TCA and acetyl-CoA carbon fixation, where the loop is either maintained or broken at this point.

Finally the energetic, geometric, and topological organization of the cell recapitulates in many respects the modularity of core biosynthesis. Cellular energy is carried on three systems: redox couples, proton-motive force, and phosphate esters. Substrate-level phosphorylation has been proposed [6, 10, 11] as the earliest direct coupling between redox and phosphate energy carriers, but it is a mechanism dependent on compatible bond structure and energies. More flexible coupling is now mediated by protons, which form a “classical” energy currency decoupling the quantum transitions of oxidation/reduction and phosphoryl transfer. This system is entirely dependent on the capacitance and proton insolation/semiconductor geometry of cell membranes, and on the topology of cellular compartments (e.g. the periplasmic space) to maintain pH and voltage differences.

A proposal for modular origins

In the modular organization that we have suggested governs extant life, the biosynthetic modules appear as both the simplest and the most exclusively self-referential, in that they require from higher levels only the provision of catalysts and the orchestration and buffering of energy systems. (Even for energy systems, some degree of independence is attained through substrate-level phosphorylation.) Universality at all higher levels depends on that in the biosynthetic core, directly for the cofactors and mediated by more complex constructions for cell form and physiology. We therefore interpret the modularity recapitulated across the hierarchy as one rooted in and constrained by biosynthesis. Because the biosynthetic modularity in turn mimics modularity in geochemical energy sources, we interpret it as the preservation of a transition stage between geochemical organization and the first biochemical organization.

We suggest that core biosynthetic pathways first formed as geosynthetic pathways dependent separately on externally provided chemical energy sources. Because the synthesized organics are not long-lived, these groups of reactions must have sometimes occurred in the same places, but their energy sources need not have originated locally, if these could be preserved and transported in mineral substrates (G. Cody, pers. comm.). The emergence of biochemistry as a distinct form of organization came when these modules formed “micro-environments” for each other, exchanging the dependence on exogenous geochemical sources for dependence on their collocation and coupling. The most explicit such proposal is for the coupling of redox and phosphate energy systems in the
earliest protocells, which would have freed biosynthesis from anhydrous phosphate—a short-lived reagent once exposed to seawater—and permitted it to depend only on the more diverse and more ubiquitous environmentally available redox couples, as cells do today [8].

A biochemical modularity originally coupled to geological processes suggests a path to an RNA world in which RNA catalysts could originally be selected as replacements for prior mineral catalysts [12]. Some form of template-directed replication would be necessary for the preservation of sequence information, but the primary determinant of fitness could have been increased capacity of pre-existing biosynthetic subsystems, rather than Darwinian competition among RNAs for self-replication within an externally supplied resource pool. The shift in emphasis we propose places greater separation between mechanisms for molecular replication and determinants of fitness (an idea also recently pursued by Nowak and Ohtsuki [13]), and suggests that the most plausible path to an incipient RNA world would have been one in which RNA catalysts preserved the structure of core biochemistry rather than over-wrote it. In this sequence, an RNA replication system need never have been autonomous, but could have co-evolved with polypeptide and biosynthetic machinery from the start. While it is analytically more complex to consider, it is probably more realistic to expect that biochemical organization formed through a bootstrapping sequence of replacement of catalysts—and perhaps less frequently of pathways or core reagents—with the major classes of small-molecule constituents such as cofactors and aptamers taking on their key functional roles during this transition toward an RNA world [12].

Similarly, cellular organization within this interpretation would have been driven by the energetic and catalytic advantages of compartmentalization, and only later taken on aspects of Darwinian individuality. The view of a late-emerging cellular individuality, and one partly independent of the individuality of RNA sequence lineages, is compatible with the notion of a progenote advocated by Woese [14–16], and with the modern understanding of the relation of viral to free-living cell lineages. It also suggests that biofilms, whether self-created or partly exogenous (discussed in our other abstract) may have been important from the very earliest transition from geochemical to cellular organization (J. Baross, pers. comm., see also [17–21]), an idea compatible with proposals for mineral-to-cell transitions [11]. However, the most plausible route to the emergence of cells remains very unclear, because phase-separation could have been an early mechanism leading to macromolecular synthesis [22, 23], and phylogenetic analyses suggest that all major membrane systems were present in the organism or community that constituted the LUCA and subsequently differentiated into the major domains [24].

Tests for residues of primordial modularity

The search for biosynthetic modules to be reproduced in laboratory experiments can be guided in part by reconstructions of plausible earth geochemistry, and in part by testing extant cells for partial subsystem independence. The latter tests may be performed physiologically (experimentally or with flux-balance analyses) or through phylogenetic inference of the history of organism and ecosystem compositions.

Physiological tests for independence of subsystems within the biosynthetic network may look for the minimum level of regulation through gene expression or allosteric enzyme response preserving organism viability, or alternatively for autonomous response of biosynthetic sub-networks to changes in energy carriers and substrate concentrations. Minimally regulated states may be sought, together with reduced but still viable biosynthetic networks, by global transposon mutagenesis [25] as part of ongoing efforts to produce minimal organisms. For enzymes in which allosteric response to substrate levels is part of a regulatory system, a more complex targeted mutation might be introduced to maintain catalytic function at the active site but to reduce response to signals.

At the same time as regulation mechanisms can be removed to see whether cells can remain viable without them, cell phenotype can be tested for response to changes in environmental condition that are independent of active regulation. Growth-rate and flux-model responses to changes in medium composition may be compared. In both studies, numerical knockout in models may be used cooperatively with experimental minimization to see how the degree of independence in subsystem response varies with whole-network complexity. We find autotrophs preferable model systems for such studies, because they are metabolically complete.

Metabolic reconstruction from the metagenomes of ecosystems, which might be termed meta-metabolomics, offers a comparative mode of analysis complementary to physiological studies. We may ask to what extent autotrophic organisms are stereotypical as self-contained ecosystems, and whether there are constrained elements within biosynthesis that have been completely preserved over the course of evolutionary history. To understand the degree of plasticity of ecosystem-level metabolism, and so interpret observations of total preservation, we may reconstruct the routes by which pathways have been gained or lost by organisms [26, 27], and understand the constraints on complementary specialization to form ecosystems. Combining metabolic universality with species-level plasticity and ecological constraint should clarify the roles that emerging individuality actually played in the origin of life.


FIG. 1: The universal core of carbon biosynthesis (heavy lines) and major variations. Solid circles represent chemical reagents, and dashed lines represent chemical reactions, in a diagrammatic similar to that proposed by Sinanoglu [28, 29]. Shaded large ellipses indicate recursively defined pathways: malonate decarboxylation to extend fatty acid chains by hydrocarbon \((\text{CH}_2)_n\) groups, and isoprene \((\text{C}_5\text{H}_8)\) condensation. Citric-acid cycle, Wood-Ljungdahl, Calvin-Benson, and 3-hydroxypropionate pathways shown, with phosphate and sulfur chemistry omitted and only CHO backbones indicated. Abbreviations: ACE, acetate; PYR, pyruvate; OXA, oxaloacetate; MAL, malate; FUM, fumarate; SUC, succinate; AKG, \(\alpha\)-ketoglutarate or 2-oxoglutarate; OXS, oxalosuccinate; ISC, isocitrate; cAC, cis-aconitinate; CIT, citrate; MLN, malonate; AcACE, acetoacetate; HMG, hydroxymethyl glutarate; MEV, mevalonate; IPT, isopentenyl; DMA, di-methylallyl; GLT, glycerate; GLA, glyceraldehyde; DHA, di-hydroxyacetone; FRC, fructose; ERY, erythrose; SED, sedoheptulose; XYL, xylulose; RBL, ribulose; RIB, ribose; MSA, malonate semialdehyde; 3HP, 3-hydroxypropionate; ACR, acrylate; PRP, propionate; MEM, methyl-malonate; GLX, glyoxy- late.