The compositional and evolutionary logic of metabolism

TOPICAL REVIEW

The compositional and evolutionary logic of metabolism

Rogier Braakman and Eric Smith

Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, USA

E-mail: rogier@santafe.edu and desmith@santafe.edu

Abstract
Metabolism is built on a foundation of organic chemistry, and employs structures and
interactions at many scales. Despite these sources of complexity, metabolism also displays
striking and robust regularities in the forms of modularity and hierarchy, which may be
described compactly in terms of relatively few principles of composition. These regularities
render metabolic architecture comprehensible as a system, and also suggests the order in
which layers of that system came into existence. In addition metabolism also serves as a
foundational layer in other hierarchies, up to at least the levels of cellular integration including
bioenergetics and molecular replication, and trophic ecology. The recapitulation of patterns
first seen in metabolism, in these higher levels, motivates us to interpret metabolism as a
source of causation or constraint on many forms of organization in the biosphere. Many of the
forms of modularity and hierarchy exhibited by metabolism are readily interpreted as stages in
the emergence of catalytic control by living systems over organic chemistry, sometimes
recapitulating or incorporating geochemical mechanisms.
We identify as modules, either subsets of chemicals and reactions, or subsets of functions, that
are re-used in many contexts with a conserved internal structure. At the small molecule
substrate level, module boundaries are often associated with the most complex reaction
mechanisms, catalyzed by highly conserved enzymes. Cofactors form a biosynthetically and
functionally distinctive control layer over the small-molecule substrate. The most complex
members among the cofactors are often associated with the reactions at module boundaries in
the substrate networks, while simpler cofactors participate in widely generalized reactions.
The highly tuned chemical structures of cofactors (sometimes exploiting distinctive properties
of the elements of the periodic table) thereby act as 'keys' that incorporate classes of organic
reactions within biochemistry.
Module boundaries provide the interfaces where change is concentrated, when we catalogue
extant diversity of metabolic phenotypes. The same modules that organize the compositional
diversity of metabolism are argued, with many explicit examples, to have governed long-term
evolution. Early evolution of core metabolism, and especially of carbon-fixation, appears to
have required very few innovations, and to have used few rules of composition of conserved
modules, to produce adaptations to simple chemical or energetic differences of environment
without diverse solutions and without historical contingency. We demonstrate these features of
metabolism at each of several levels of hierarchy, beginning with the small-molecule metabolic
substrate and network architecture, continuing with cofactors and key conserved reactions, and
culminating in the aggregation of multiple diverse physical and biochemical processes in cells.

1. Introduction

The chemistry of life is distinguished both by its high degree of order and by its essential dependence on a number of far-from-equilibrium reactions [1]. While in some cases reactions may be treated as isolated subsystems with equilibrium approximations [2, 3], such isolations are themselves cumulative deviations far from equilibrium, reflecting the system-level properties of life as a whole. The dynamical order of life's chemistry is maintained by the non-equilibrium transfer of electrons through the biosphere. Free energy from potential differences between electron donors and acceptors can be derived from a variety of biogeochemical cycles [4], but within cells electron transfer is mediated by a small number of universal electron carriers which drive a limited array of organic reactions [5]. Together these reactions make up metabolism, which governs the chemical dynamics both within organisms and across ecosystems. The universal and apparently conserved metabolic network transcends all known species diversification and evolutionary change [6, 7], and distinguishes the biosphere within the major classes of planetary processes [8]. We identify metabolism with the quite specific substrate architecture and hierarchical control flow of this network, which provide the most essential characterization of the chemical nature of the living state.

Understanding the structure of metabolism is central to understanding how physics and chemistry constrain life and evolution. The polymerization of monomers into selected functional macromolecules, and the even more complex integration and replication of complete cells, form a well-recognized hierarchy of coordination and information-carrying processes. However, in the sequence of biosynthesis these processes come late, and they involve a much smaller and simpler set of chemical reactions than core metabolism, the network in which all basic monomer components of biomass are created from environmental inputs.

Because the core is the origin of all biomass, its flux is perforce higher than that in any secondary process; only membrane electron transport (reviewed in [4]) has higher energy flux. For example, [9] notes that, over a broad sample of enzymes collected from the literature, those for secondary metabolic reactions have rates $\sim 1/30$ the typical rates of enzymes for core reactions.

The combined effects of a higher diversity of constraints from chemistry and physics and a higher density of mass flux within core metabolism relative to other processes in living systems have major impacts on the large-scale structure of evolution, as we will show. Metabolism is the sub-space of organic chemistry over which life has gained catalytic control, and because in the construction and optimization of biological phenotypes all matter flows through this sub-space, its internal structure imposes a very strong filter on evolution.

In this review we identify a number of organizing principles behind the major universal structures and functions of metabolism. They provide a simple characterization of metabolic architecture, particularly in relation to microbial metabolism, ecology, and phylogeny, and the major (biogeochemical) transitions in evolution. We often find the same patterns of organization recapitulated at multiple scales of time, size, or complexity, and can trace these to specific underlying chemistry, network topology, or robustness mechanisms. Acting as constraints and sources of adaptive variation, they have governed the evolution of metabolism since the earliest cells, and some of them may have governed its emergence. They allow us to make plausible reconstructions of the history of metabolic innovations and also to explain certain strong evolutionary convergences and the long-term persistence of the core components of metabolic architecture.

Many structural motifs in both the substrate and control levels of metabolism may be interpreted as functional modules. By isolating effects of perturbation and error, modularity can both facilitate emergence, and support robust function, of hierarchical complex systems [10, 11]. It may also affect the large-scale structure of evolution by favoring variation in the regulation and linkage between modules, while conserving and thereby minimizing disruption of their internal architecture and stability [12, 13]. This can enhance evolvability through two separate effects. An increased *phenotypic* (i.e. structural or functional, as opposed to *genotypic* or sequence) robustness of individual modules gives access to larger genetic neutral spaces and thus a greater number of novel phenotypes at the boundaries of these spaces [14]. At the same time, concentrating change at module interfaces, and allowing combinatorial variation at the module level, can decrease the amount of genetic variation needed to generate heritable changes in aggregate phenotypes [15, 16]. It has been argued that asymmetries in evolutionary constraints can be amplified through direct selection for evolvability, and that this is a central source of modularity and hierarchy within biological systems [15–18].

These functional consequences of modularity lead us to expect that metabolism will be modular as a reflection of the requirements of emergence and internal stability. Certainly we observe this empirically; many topological analyses of metabolic networks find a modular and hierarchical structure [19–21]. Because of the higher mass flux and more diverse chemistry in core pathways, we also expect that modularity in their subnetwork will have the greatest influence on evolutionary dynamics. In section 3 we will review a range of evidence supporting this expectation, which suggests that innovations in core CO_2 fixation were a large part of the cause for major divergences in the deep tree of life [22].

To understand the origin and evolutionary consequences of modularity in metabolism, however, we will need system-level representations that go beyond topology, to include sometimes quite particular distinctions of function. Details of substrate chemistry, enzyme grouping and conservation, and phylogenies of metabolic modules, in particular, are rich sources of functional information and context. While we will find it significant that some module boundaries are recapitulated at many levels, differences between levels will also help to distinguish modularity originating in reaction mechanisms and network topology of the small-molecule metabolic substrate, from possibly independent higher-level forms of modularity in the regulation of flux rates or

phenotypic expression by the macromolecular components of cells. As an example of the second kind of regulatory control, it has been argued that the modular constraints observed in amino acid biosynthetic pathways are due to evolutionary optimization of the overall kinetics and dynamic responses of these pathways [23–25]. These forms of modularity arise from mechanisms such as allosteric inhibition of enzymes and the tuning of enzyme specific activities, which are brought into existence by the underlying network topology and molecular inventory of metabolites. We will return in section 6.1.2 to ways in which regulation of networks may have been essential to the stability of their underlying architecture. Recognizing the distinct character of architectural motifs and control mechanisms at different levels will enable us to reconstruct steps in metabolic evolution and identify their environmental drivers.

1.1. Hierarchy in metabolism, and the role of individuals and ecosystems

While most metabolic conversions are performed within cells, there is still a significant number that take place at the cell population level (for example those involving siderophores and secreted enzymes). In general it is important to appreciate that a complete accounting of biochemical fluxes not only will span many levels of biological organization, but also may incorporate multiple distinct internal modes of organization. In addition to the standard ecological distinction between autotrophy and heterotrophy, scientists working in the area of bioremediation, for example, have coined the term *epi-metabolome* to refer to those compounds that due to their slow degradation are freely diffusible across microbial communities [26]. Thus, the causes and roles of evolutionary changes, even though they arise within cellular lineages, may be only partly explained by organization at the cellular or species level. Other levels that must also be considered include the meta-metabolome of trophic ecosystems [27–30], and the links to geochemistry [31–38]. The great biogeochemical cycles—of carbon, nitrogen, phosphorus, or many metals—combine physiological, ecological and even geochemical links such as mantle convection or continental weathering [39, 40].

An important additional empirical observation is that the deepest universal features of metabolism are reliably seen not at the individual, but at the ecosystem level [7, 41]. The single-organism metabolisms among members of complex ecosystems may vary extremely widely [42], because different organisms perform different segments of biosynthetic or degradative pathways, using trophic links (predation, parasitism, symbiosis, syntrophy, saprophyty) to obtain what they do not make. The aggregate, or net, pathways to which these individual metabolisms contribute, once assembled through their trophic links, mostly remain within standard networks as reflected in databases such as KEGG [43] or UniProt [44]. Appreciating that the redundancy in metabolism would permit the assembly of a comprehensible metabolic chart despite the bewildering variety of species was the major contribution of Nicholson [45]. A dynamical

interpretation of the universality of metabolism that may be more important for understanding evolution is that ecosystems have dynamics inherent to their own level of aggregation that is not captured in their descriptions merely as assembled communities of species. Such dynamics are expressed as limits on, or long-range evolutionary convergences of, innovations within metabolism.

The corollary, that individuality is a derived characteristic of living systems within a larger framework of metabolic regularity, fits well with the modern understanding that individuality takes many forms which must be explained within their contexts [46]. Alternatively, in more conventional genetic descriptions of evolution [47, 48], metabolic completeness, trophic as well as physiological flux balance, and network-level response to fluctuations are explicit features contributing to an organism's fitness within a co-evolving or constructed environment [49].

We can to a considerable extent disentangle the inherent chemical hierarchy of metabolism from the evolutionary hierarchy of species by studying variations in the anabolic (biosynthetic) versus catabolic (degradative) pathways within organisms, along with the relations of autotrophy (self-feeding) versus heterotrophy (feeding from others) in the ecological roles of species. We can argue for the existence of a universal anabolic, autotrophic network [50, 51] that comprises the chemistry essential to life. We can then separate the structural requirements and evolutionary history of the universal network from secondary complexities, which we will argue originate in the diversification of species and the concurrent processes of assembly of ecological communities.

Within the universal (and apparently essential) network we may identify further layers, with distinct functions and plausibly distinct origins. A functioning metabolism is both a network of fluxes through substrate molecules, and a set of hierarchical relations in which some of the more complex structures control the kinetics of flows within the network. Within the substrate network, distinguishable subnetworks include the core network to synthesize CHO backbones, networks radiating from the core that incorporate N, S, P or metals, higher-order networks that assemble complex organics from 'building blocks', and still others that synthesize all forms of polymers from small organic monomers. Within the control hierarchy, the layers of cofactors, oligomer catalysts, and integrated cellular energetic and biosynthetic subsystems are qualitatively distinct.

The foundation of autotrophy—and more generally the anchor that embeds the biosphere within geochemistry—is *carbon-fixation*, the transformation of CO_2 into small organic molecules (see figure 1). A recent study [22] combining evidence from phylogeny and metabolic network reconstruction—an approach we refer to as 'phylometabolic' reconstruction—showed that all carbon fixation phenotypes may be related by an evolutionary tree with very high (nearly perfect) parsimony, and a novel but sensible phenotype at the root. The branches representing innovations in carbon fixation were found to trace the standard deep divergences of bacteria and archaea. More striking, this work showed that likely
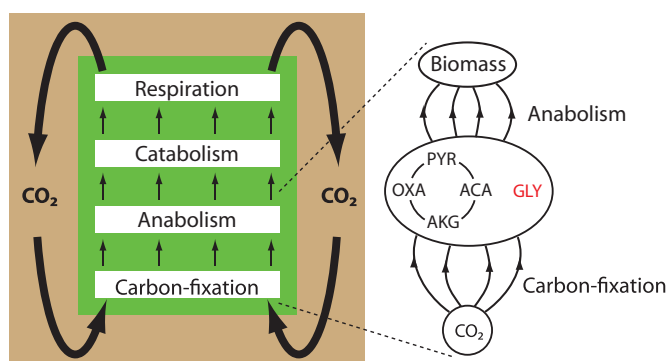
Figure 1. The metabolic structure of the biosphere. The biosphere as a whole (green) can be described as implementing a global biological carbon cycle based on CO_2, with carbon-fixation as the metabolic foundation that anchors it within geochemistry (brown). The small organic molecules produced during fixation of CO_2 are subsequently transformed and built up into the full diversity of known biomolecules through the process of anabolism, before ultimately being broken back down through catabolism and re-released as CO_2 through respiration. The striking modularity of metabolism is expressed in the fact that the interface between carbon-fixation and anabolism consists of a very small number of small organic molecules (shown schematically at right-center). The key observation that, in addition to intermediates in the citric acid cycle—from which nearly all anabolic pathways emanate [7]—glycine (red) should be included in this set, allows a complete reconstruction of the evolutionary history of carbon-fixation [22]. These points are explained in detail in section 3. Abbreviations: acetyl-CoA (ACA); pyruvate (PYR); oxaloacetate (OXA); α-ketoglutarate (AKG); glycine (GLY).

environmental drivers could be identified for most divergences, suggesting that deep evolution reflects first incursions into novel geochemical environments [22]. The tight coupling of the reconstructed phylogeny to geochemical variety suggests that constraints from chemistry and energetics drove early evolution in predictable ways, leaving little need to invoke historical contingency. We will discuss these points, placing them in context, in detail in section 3.

1.2. Catalytic control and origins of modularity in metabolism

While carbon-fixation draws on all levels of biological organization (requiring integration and control of many cellular components), evolution in the network of its small-molecule substrate has consisted only of changes in the use of a small number of clearly defined reaction sequences. The disruption, disconnection or reversal of these modules accounts for the full diversity of modern carbon-fixation. As we will show below, the module structure is further defined by a distinction between two types of chemistry. Within modules, the reactions are mainly (de-)hydration or (de-)hydrogenation reactions, catalyzed by enzymes from common and highly-diversified families. The interfaces between modules are created (and distinguished) by key carboxylation reactions, catalyzed by highly conserved enzymes, often involving special metal centers and/or complex organic cofactors. The congruence of phylogenetic branching with topological and

chemical module boundaries suggests that a very small number of catalytic innovations were the key bottlenecks to evolutionary diversification, against a background of facile and readily re-used organic chemistry.

Topological modularity in the small-molecule substrate network is often associated with functional divisions in the more complex molecules that control metabolism, particularly the cofactors, showing that their metabolic role is also an evolutionary role. As carriers of electrons or essential functional groups, cofactors regulate kinetic bottlenecks in metabolic networks. Here again we wish to distinguish the chemical complexity of metabolic reaction mechanisms, and the role of cofactors in those reactions, from higher level control of kinetics through regulation of concentrations, which may optimize pathway robustness against perturbations [52]. The structurally most elaborate cofactors tend to facilitate the chemically most complex reaction mechanisms. As a result, the distinction between presence and absence of these cofactors is effectively the absolute presence or absence of reaction mechanisms (a 'topological' network property), as contrasted with a finite rate adjustment. Thus, the appearance and diversification of families of biosynthetically related cofactors introduced functions which served as 'keys' to domains in organic chemistry, incorporating these within biochemistry. As a result we may often map biosynthetic pathway diversification of cofactors onto particular lineage divergences in the tree of life. We will show examples of this in section 4. Cofactor biosynthetic networks are themselves often modular, with multiple biosynthetic pathways in a family using closely related enzymes that enable structures characteristic of the cofactor class.

The quite sharply defined roles of many modules enable us to understand strong evolutionary convergences that have occurred within fundamental biochemistry, and in some cases we can relate the functioning of an entire class of substrate or control molecules to specific chemical properties of elements or small chemical groups. Several important module boundaries are aligned at the same points in their substrate networks and their control layers. This suggests to us that lower-level substrate-reaction networks introduced constraints on the accessible or robust forms of catalysis and aggregation that it was later possible to build up over them. From repeated motifs within the substructure of modules, and from patterns of re-use or convergence, we may identify chemical constraints on major transitions in metabolic evolution, and we may separate the early functions of promiscuous catalysts as enablers of chemistry, from later restrictions of reactants as catalysts were made more specific. The remarkable fact that such low-level chemical distinctions (in elements, reactions or small-molecule networks) should have created constraints on innovation well into the Darwinian era of modern cells suggests these as relevant constraints also in the pre-cellular era.

1.3. Manuscript outline

Our main message is twofold: (1) that the structure of biosynthetic networks and their observed variation,

even though the networks are elaborate, has a compact representation in terms of a small collection of rules for composition, and (2) that the same rules we abstract from composition have a natural interpretation as constraints on evolutionary dynamics, which as a generating process has produced the observed variants. We intend the expression 'logic of metabolism' to refer to the collection of architectural motifs and functions that have apparently been necessary for persistence of the biosphere, that have led to modularity in the physics and chemistry of life, and that have determined its major evolutionary contingencies and convergences.

After a short description of the important global features of metabolism in section 2, we will construct these at ascending levels in the hierarchy, beginning in section 3 with the networks of core carbon fixation and the lowest levels of intermediary metabolism. We will then, in section 4, consider cofactors as the intermediate level of structure and the first level of explicit control in biochemistry, illustrating how key cofactor classes govern the fixation and transfer of elementary carbon units, and introduce control over reductants and redox state. Both in the metabolic substrate and in the cofactor domain, it will be possible to suggest a specific historical order for many major innovations. For the substrate network this will capture conditional dependences in the innovation of carbon fixation strategies. For cofactors it will allow us to approximately place the emergence of specific cofactor functionalities within the expansion of metabolic networks from inorganic inputs.

In section 5 we consider the processes by which innovation occurs, specifically interplay of the introduction of general reaction mechanisms versus selectivity over substrates. The modular substructure and evolutionary sequence of many of our reconstructed innovations favors an early role for non-specific catalysts, with substrate selectivity appearing later. In section 6 we then list candidates for the major organizing constraints on integration of metabolism within cells. These include the role of compartments in linking energy systems, as well as the coupling of physiological and genetic individuality, which permit species differentiation, and complementary specialization within ecological assemblies. Finally, in section 7 we discuss how the various observations made throughout the paper may be used to provide context in assessing scenarios for the emergence of life. Because we draw from several areas of research which do not have fully-shared vocabulary, a glossary with some frequently used terms is provided in appendix A.

2. An overview of the architecture of metabolism

2.1. Anabolism and catabolism in individuals and ecosystems

Metabolic networks within organisms are commonly characterized as having three classes of pathways: (1) catabolic pathways that break down organic food to provide chemical 'building blocks' or energy; (2) core pathways through which nearly all small metabolites pass during primary synthesis or ultimate breakdown, and (3) anabolic pathways that build up all complex chemicals from components originating in

the core. The motif of three-stage pathways—catabolic, core, anabolic—between typical pairs of metabolites has been abstracted into a paradigm of 'bowtie' architecture for metabolism [53–55]. This qualitative characterization (which may be complicated by salvage pathways and other cross-linkages) is supported by a strong statistical observation that most minimal pathways connecting pairs of metabolites consist of a catabolic and an anabolic segment connected through the core [55]. Thus, relatively speaking, the catabolic and anabolic pathways are less densely crosslinked than pathways within the core, from which they radiate.

The reason for this lack of cross-linking can be understood from the explanation of path lengths in terms of number theory and string chemistries in [55]. Lengths of typical optimal paths between pairs of metabolites in *E. coli* are logarithmic in carbon count, because they decompose molecules into small prime 'factors' in the core which are then modified by single carbons to other prime factors and re-assembled. Thus, optimal conversions within the bowtie consist of finding common molecular 'divisors' of input and output metabolites, which in actual metabolic chemistry are familiar 2-, 3-, and 5-carbon groups. We will argue that, when other chemical and phylogenetic evidence is taken into account, the fact that short paths exist from most metabolites to a small set of building blocks is more likely a reflection of the prior role of the core (where building blocks are created) in defining the possibilities for later anabolism and thus the metabolites reached by the bowtie.

Catabolic pathways in a cell may be fed through physiological or trophic links to other cells or organisms, or they may break down food produced previously by the same cell and then stored. Figure 2 illustrates schematically the relation of the three classes. Both catabolic and anabolic pathways may be large and somewhat diversified; the core itself constitutes no more than a few hundred small metabolites [50, 51], most of which have functions that are universal throughout the biosphere.

Whole-organism metabolisms are conventionally divided into two classes—autotrophic and heterotrophic—according to the ways they combine anabolic and catabolic pathways [5]. Autotrophs synthesize all required metabolites from inorganic precursors, and can function without catabolism, using only the core and anabolic pathways radiating from it. Establishing the metabolic self-sufficiency of a putatively autotrophic organism can prove challenging, however [56]. Heterotrophs, in contrast, are organisms that must obtain organic inputs from their environments because they lack essential biosynthetic pathways. Autotrophy and heterotrophy are best understood as modes of metabolism, between which some individual species may switch depending on circumstances, and which may even be mixed at the level of sub-networks within a given organism. Many organisms are obligate autotrophs or heterotrophs, but others are facultative autotrophs that can switch between fully-autotrophic and heterotrophic metabolic states, while still others are mixotrophs that concurrently use both CO_2 and organic carbon inputs to synthesize different parts of their biomass [5]. The important distinction for what follows is that autotrophs and heterotrophs play fundamentally different ecological roles.
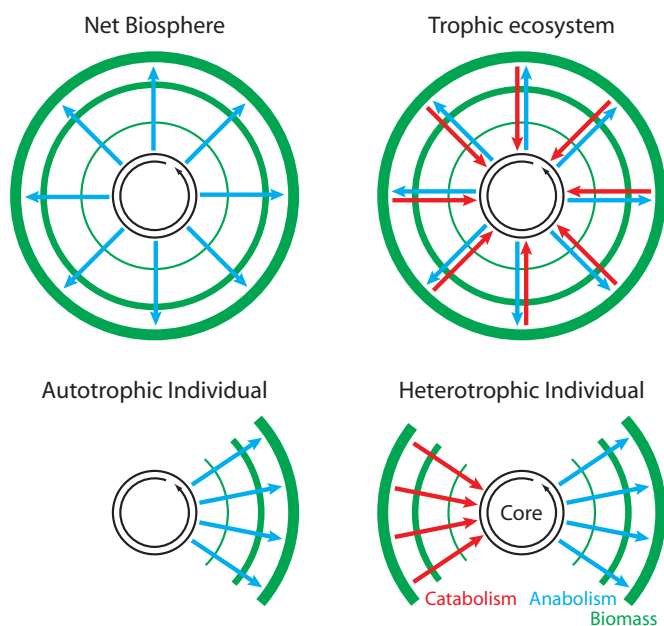
Figure 2. Global structure of metabolism. Anabolic pathways (blue) build biomass and catabolic pathways (red) break it down. Catabolic pathways may be direct or approximate reverses of anabolism; see main text for further discussion on this point. Carbon enters through the core (black), which is the starting point of anabolism and also the endpoint of respiration. Because carbon enters the biosphere as CO_2 (see figure 1), the biosphere today is autotrophic as a whole and anabolism is functionally prior to catabolism. As a result, both single organisms and assemblies of organisms can function using metabolic modes consisting only of anabolic pathways that fan outward from the core (blue and green). By partitioning pathway directions between anabolic and catabolic (joined at the core), organisms can use metabolic modes consisting of the familiar 'bowtie' architecture of derived metabolism (red with blue). Their assembly into trophic ecosystems (blue and red radial graph) then both builds and degrades organic compounds actively, cycling carbon between environmental CO_2 and biomass (green). In these graphs, concentric (green) shells reflect sequential steps in biosynthesis leading to a hierarchy of increasing molecular complexity.

Autotrophic metabolism forms the lowest trophic level in the biosphere, fixing CO_2 into organic matter, while heterotrophic metabolism forms all subsequent levels, determining the structure of flows of organic compounds in trophic webs [57], and actively cycling carbon from biomass back to environmental CO_2. While all biological free energy passes at some stage through redox couples, autotrophs capture a part of this energy by transferring electrons from high energy reductants to CO_2 [7]. Heterotrophs may exploit incomplete use of this free energy through internal redox reactions (fermentation), or they may re-oxidize organic matter back to CO_2 (respiration).

The role of catabolism in most organisms is closely tied to their ecological role as heterotrophs. Heterotrophy provides enormous opportunity for metabolic diversification [42], in the evolution of catabolic pathways and the partitioning of essential anabolic reactions among the constituent species within ecosystems. However, the study of metabolism

restricted to particular heterotrophic organisms[1] can obscure much of its universality: heterotrophs may differ widely, but the aggregate anabolic networks that sustain them at the level of ecosystems are largely invariant. Autotrophs show that much of this diversity is not essential to life, allowing us to conceptually separate the requirements for biosynthesis from complexities that originate in processes of individual specialization and ecological assembly [58].

The 'bowtie' motif [53, 54]—a paradigm derived from the study of heterotrophs[2]—can be misleading, as it combines universal metabolic pathway dependences with widely variable physiological or ecological specializations. The core and anabolism are essential (and we argue more ancestral), and the reduction in cross-linking with distance from the core may be seen to reflect an entirely outgoing radial 'fan' of anabolism. Biomass is organized in a sequence of concentric shells spanned by the radial pathways, which count the number and complexity of biosynthetic steps [59, 51]. Organisms, in particular autotrophs, exist which can function without catabolism, but only the most derived parasites lack anabolism. For example, members of the genus *Mycoplasma* can function with remarkably small genomes, having given up nearly all genes associated with the *de novo* synthesis of amino acids, cofactors, nucleotides and lipids [60–62], because they live as intracellular parasites in hosts that synthesize these.

Most catabolic pathways are also, in varying degrees, reversals of widespread anabolic pathways. In some cases the reversal is exact, often for short pathways, as in the case of glycine metabolism that we discuss in detail in the next section. In other cases, such as gluconeogenesis and glycolysis [63] or fatty acid metabolism, catabolic pathways resemble their anabolic counterparts closely but differ in a few intermediates, cofactors, or enzymes, usually for thermodynamic reasons [64]. Finally, in some cases catabolism reflects genuine innovations, as in the metabolism of the branched-chain amino acids [65] or of nucleotides [66], or some salvage pathways. We find it significant that even in cases where reversals are only approximate due to variation in some of the substrates or catalysts, the overall *sequences* of reactions at the substrate-level are often nearly completely preserved. In such cases, substitutions, which may appear to be large differences from the perspective of enzyme homology, clearly are often local alterations in energy flow usually involving interchanged reaction orders. An example of this is the variable order of thioesterification to form succinyl-coenzyme-A (succinyl-CoA) in reductive, fermentative, and

---

[1] Almost all model organisms have been heterotrophs, because these are accessible and are usually connected to humans as symbionts, pathogens or cultivars. *E. coli* (in which operons were discovered) is a phenotypically and trophically very plastic organism as this is required for its complex life cycle. No known multicellular organisms can reduce the triple-bond of N_2, making them reliant either directly on microbial nitrogen fixers for reduced nitrogen (NH_4^+) or on mineralized forms such as NO_3^- derived ultimately from microbial nitrogen metabolism (or outputs of human technological processes). The only known autotrophic organisms are bacteria and archaea, and none of these is developed nearly to the level that standard heterotrophic model systems (such as *E. coli*) are.

[2] The paradigm of the metabolic bowtie is also in part a borrowing from a conventional paradigm in engineering [53], motivated by applications to human physiology and medicine [310].

oxidative tricarboxylic acid (TCA) cycles [67–69]. In the last two cases, organisms may use succinate as an intermediate in the formation of succinyl-CoA from α-ketoglutarate, rather than directly performing an oxidative decarboxylation, which would constitute strict reversal of the reductive TCA reaction. Succinate acts as an electron sink, pulling the reaction forward, but an additional ATP hydrolysis is then needed to form succinyl-CoA from it.

The preserved reaction sequences may be 'channels' within organic chemistry with optimal path length or connectivity [70–72] that were easier to bring or maintain under catalytic control, or for later innovations they may reflect lock-in by requirements of secondary metabolism. Finally, many reversals from anabolism to catabolism can be explained as consequences of ecological change, with finer distinctions arising as adaptations to specific ecological or geochemical environments.

The conceptual difference and asymmetry between autotrophy and heterotrophy becomes clearer when we examine the metabolic structure of ecosystems at increasing scales of aggregation. Entire ecosystems, to the extent that they are approximately closed, function chemically as autotrophs. The biosphere as a whole (see figure 1) is not only approximately, but fully autotrophic, as today it does not depend significantly on extraterrestrially, atmospherically or geologically produced organics. This observation still admits two possibilities for the emergence of aggregate metabolism: Either the biosphere has been autotrophic since its inception, or it was originally heterotrophic and later switched to using CO_2 as its sole carbon source. We have recently shown that assuming autotrophy at least as far back as the era of a common metabolic ancestor leads to a highly parsimonious reconstruction for the evolution of carbon-fixation pathways [22]. The congruence of our tree of carbon-fixation phenotypes with standard phylogenies [73], which places modern autotrophs as the conservative descendents of deep branches [32], together with numerous arguments drawing evidence from biochemistry and geochemistry that thermophilic autotrophs are the most plausible models for deep-ancestral bacteria and archaea [33, 74–77], permits quite specific and consistent biochemical proposals for an autotrophic deep-ancestral stage of life. To our knowledge, there is no equivalent body of evidence leading to specific and consistent predictions of heterotrophic forms at the earliest evolutionary times.

For all of these reasons we will interpret the core and anabolic pathways as the base layer and skeleton of the most fundamental constraints on metabolism, and will consider the problem of emergence and early evolution of fully autotrophic systems. Reconstructing the emergence of autotrophic metabolism provides important context to the emergence of life, to which we return in section 7. We restrict the discussion here to the structure and evolution of metabolism, and to conclusions that can be drawn from biochemistry, phylogenetics and geochemical and ecological context. These conclusions do not depend on speculations about what chemical stages may have preceded the emergence of anabolism.

As long as we do not conflate the chemical condition of autotrophy (complete anabolism) with assumptions about individuality (whether complete anabolisms are contained within the regulatory control of individual organisms) [58], and as long as we recognize the ecosystem as potentially the correct level of aggregation to define autotrophy, we need not assume that the first fully functioning autotrophic metabolism consisted of individual cells. Our interpretation extends equally to populations of organisms that were physiologically as well as genetically incomplete and functioned cooperatively [78–81]. Once organism-level and species-level organization has been put aside as a separate question, the chemical distinction between heterotrophy and autotrophy is one between metabolic partial-systems with unknown and highly variable boundary conditions, versus whole-systems required to subsist on CO_2 and reductant. If we wish to understand the structure of the biosphere and to interpret the sequence of innovations in core carbon fixation, the added constraint of autotrophy provides a framework to do this.

2.2. Network topology, self-amplification, and levels of structure

Understanding either the emergence of metabolism, or the robust persistence of the biosphere, requires understanding life's capacity for exponential growth. Exponential growth results from proportional self-amplification of metabolic and other networks that have an 'autocatalytic' topology [82–87] (see figure 3). Network autocatalysis is a term used to describe a topological (stoichiometric) property of the substrate network of chemical reactions. In a *catalytic* network, one or more of the network intermediates is needed as a substrate to enable the pathway to connect to its inputs or to convert them to outputs, but the catalytic species is regenerated by the stage at which the pathway completes. Network-catalytic pathways must therefore incorporate feedback and comprise one or more loops with regard to the internally produced molecules. An *autocatalytic* network is a catalytic network augmented by further reactions that convert outputs to additional copies of the network catalyst, rendering the pathway self-amplifying.

Molecular autocatalysis—the property that intermediates in a pathway serve as conventional molecular catalysts for other reactions in the pathway—may be understood as a restricted form of network autocatalysis in which the reaction to which some species is an essential input is the same reaction that regenerates that species. Some chemists prefer to use the term 'network autoamplification' for the general case, restricting 'autocatalysis' to apply only when species are traditionally-defined molecular catalysts. We will use 'autocatalysis' for the general case, to reflect the property of stoichiometry that a pathway regenerates essential inputs. For us the distinction between autocatalysis at the single molecule versus more general network level mainly effects the kinetics and regulation of pathways.

Network autocatalysis is necessary to maintain dynamical ordered states, by re-concentrating inputs into a finite number of intermediates, against the disordering effects of thermodynamic decay and continual external perturbation.
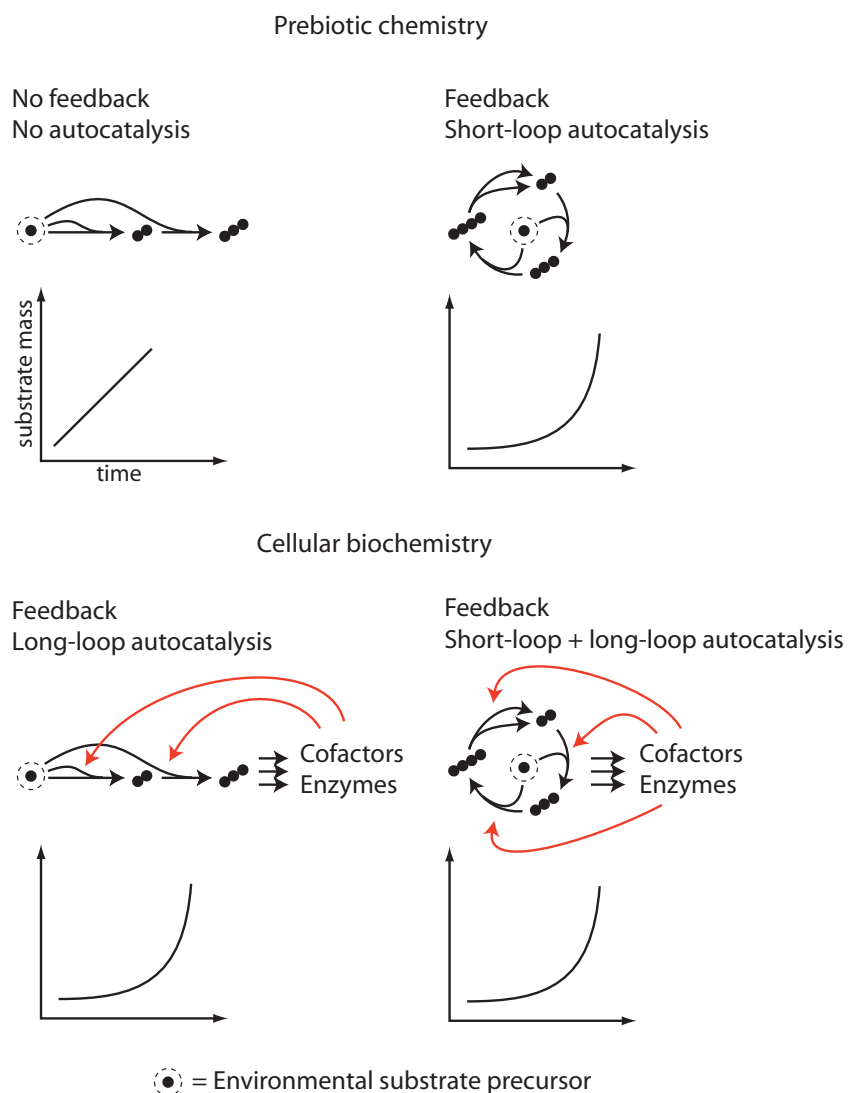
Figure 3. Upper limit growth rate curves for chemical reaction networks of different classes. To highlight the role of network topologies, the chemistry is simplified with only C–C bond forming and cleaving reactions shown. Each of the growth curves qualitatively represents the upper limits for mass accumulation within the participating substrates of the pathways. When fully integrated within modern cellular biochemistry, both linear and cyclic pathways are network autocatalytic and capable of exponential growth. This form of network autocatalysis, however, derives from feedback provided by cofactors and enzymes, both of which have elaborate synthesis pathways, and is thus classified as 'long-loop' autocatalysis. In an abiotic world in which reaction-level catalysis is limited to external sources, only cyclic pathways with feedback topologies at the substrate level—correspondingly classified as 'short-loop' autocatalysis—are capable of exponential growth, while linear pathways are limited to linear growth. We contend that an early presence of short-loop autocatalysis is important because it provides a mechanism to concentrate mass flux within abiotic chemical networks, preventing excessive dilution with increasing size and complexity of organic molecules, and in turn giving easier and more robust access to subsequent long-loop feedback closures.

Therefore all observed persistent material flows in the biosphere can only be products of autocatalytic networks, though they may require hard-to-recognize feedbacks ranging from the level of cell metabolism to trophic ecology for full regeneration. This *ex post* observation does not, however, explain why self-amplification was possible in abiotic chemistry to give rise to a biosphere. In addition to topologies enabling feedback, the latter would have required that intermediates in the network be produced at rates higher than those at which they were removed.

The significant observations about autocatalysis in the extant biosphere, which may also contain information about its emergence, concern the complexity, number and particular form of levels in which autocatalytic feedback can be found. Where the hierarchical modules of metabolic structure or function follow the boundaries required for feedback closure of different autocatalytic sub-networks, it may be possible to order the appearance of those sub-networks in time. It may also be possible to infer the geochemical supports they required for stability and self-amplification, before those supports were attained through integration into cellular biochemistry.

We wish, in these characterizations, to recognize what we might call 'conditional' as well as strict autocatalysis. In extant

organisms, where (essentially) all reactions are catalyzed by macromolecules, and most cofactors (reductants, nucleoside-triphosphates, coenzymes) are recharged by cellular processes, strict autocatalysis of any network is only satisfied in the context of the full complement of integrated cellular processes. If, however, inputs provided by cofactors, macromolecules and energy systems in modern cells could have been provided externally in earlier stages of life, for instance by minerals or geochemical processes, then identifying networks in extant biochemistry that, although simple, *would be* autocatalytic if given these supports, may give information about intermediate stages of emergence (see figure 3). The strong modularity of extant metabolism and its congruence with such conditionally autocatalytic topologies suggests that a separation into layers corresponding to stages of emergence may be sensible.

We will argue that the two most important and functionally distinct early layers are the small-molecule substrate of core metabolism itself, and the organic cofactors synthesized downstream from this core, which feed back through network and molecular catalysis to form a control layer over the core network (as well as, later, secondary networks). The picture of extant metabolism as the outcome of layers of emergence has been advanced in many forms [88, 59, 6], and the central importance of feedback through catalysis has also been emphasized [89]. Here we consider the very specific relations between reductive core pathways and cofactors, as evidence that intermediary metabolism is the result of kinetic stabilization and selection of the core that arose previously. In addition to reconstructing historical stages, the mechanisms leading to autocatalysis in different sub-systems may suggest important geochemical contexts or sources of robustness still exploited in modern metabolism.

2.3. Network-autocatalysis in carbon-fixation pathways

At the chemically simplest level of description—that of the small-molecule metabolic substrates and their reaction-network topologies—carbon fixation pathways form two classes. Five of the six known pathways are autocatalytic loops, while one is a linear reaction sequence. (All uses of autocatalysis in this section refer to *conditional* autocatalysis as explained above.) The loop pathways condense CO_2 or bicarbonate onto their substrate molecules, lengthening them. Each condensation is accompanied or followed by a reduction, making the average oxidation state of carbon in the pathway substrate lower than that of the input CO_2, and resulting in a negative net free energy of formation in a reducing environment [7]. (Reducing power may originate in the geochemical environment, but in modern cells electrons are transferred endergonically to more powerful reductants such as NADH, NADPH, $FADH_2$, or reduced ferredoxin.) Each fixation loop contains one reaction where the maximal-length substrate is cleaved to produce two intermediates earlier in the same pathway, resulting in self-amplification of the pathway flux. As long as pathway intermediates are replenished faster than they are drained by parasitic or anabolic side reactions, the loop current remains above the autocatalytic threshold. However, the threshold is fragile, as pathway kinetics provide

no inherent barrier against flux falling below threshold and subsequently collapsing. The autocatalytic threshold and dynamics of growth, saturation, or collapse are considered in section 3.4.

At the level of network topology, the linear *Wood–Ljungdahl* (WL) fixation pathway [90–92] is strikingly unlike the five loop pathways. Instead of covalently binding CO_2 onto pathway substrates, which then serve as platforms for reduction, the WL reactions directly reduce one-carbon (C_1) groups, and then distribute the partly- or fully-reduced intermediates to other anabolic pathways where they are incorporated into metabolites. The linear sequence of reductions has no feedback, and the C_1 groups at intermediate oxidation states do not increase in complexity. Instead, these reductions (leading to intermediate C_1 states that would be unstable in solution) are carried out on evolutionarily refined folate cofactors [93]. The topology of the WL pathway becomes self-amplifying only if the larger and more complex biosynthetic network for these cofactors is considered together with that of the C_1 substrate, requiring that a longer feedback loop be maintained than the mere substrate loop in the other fixation pathways. In the network context of the WL fixation mechanism, the folate cofactors have an intermediate role between network catalysts and molecular catalysts, as they are passive carriers, but form stable molecular intermediates rather than mere complexes as are formed by enzymes with their substrates. We will characterize this distinction between the loop-fixation pathways and WL as a distinction between *short-loop* and *long-loop* autocatalysis (see figure 3).

The network catalysts that could be said to 'select' the short-loop pathways are the reaction intermediates themselves. The key metabolites that have the corresponding selection role for WL are the folate cofactors produced in a secondary biosynthetic network. Short-loop and long-loop pathways are therefore distinguished both by the number of reactions that must be maintained and regulated, and by the fact that WL spans substrates and the biosynthesis of cofactors, which we will argue in section 4 are naturally interpreted as qualitatively distinct layers within biochemistry.

The appearance of different features suggesting simplicity or primordial robustness, in different fixation pathways, together with aspects of their phylogenetic distribution, have led to diverse proposals about the order of their emergence [94, 95]. WL is the only carbon-fixation pathway found in both bacteria and archaea, and its reactions have been shown to have abiotic mineral analogues [74, 75, 94], suggesting a prebiotic origin. Yet WL is not self-amplifying and so lacks the capacity for chemical 'competitive exclusion' (equivalent to the capacity for exponential growth). The cofactors that make it self-amplifying are complex, and the simple pathway structure of C_1 reduction does not suggest what would have supported their formation.

In contrast, autocatalysis within the small-molecule substrate networks of the loop pathways suggests the inherent capacity for self-amplification, exponential growth and chemical competitive exclusion. This is an appealing explanation [7] for the role, particularly of the intermediates in the reductive citric acid cycle [96, 6] (discussed in section 3)

as precursors of biomass. Arcs within this pathway have also been reproduced experimentally in mineral environments [97], though a self-amplifying system has not yet been demonstrated. However, self-amplification requires complete loops, and even the most compelling candidate for a primordial form (reductive citric acid cycling) is found only in a subset of bacterial clades.

We argue in the next section that a joint fixation pathway incorporating both WL and reductive citric acid cycling resolves many of these ambiguities in a way that no modern fixation pathway can. Proposals have previously been made in [59] and [75] for WL fixation followed by the use of citric-acid cycle pathways. However, in [59] the TCA cycle is suggested to run oxidatively, while the primordial networks proposed in [75] are forms of acetogenesis. Neither therefore emphasizes self-amplification and short-loop autocatalysis as essential early requirements. As a phylogenetic root, a fully connected network combining WL and the rTCA cycle defines a template from which the fixation pathways in all modern clades could have diverged. As a candidate for a primordial metabolic network, in turn, it provides both chemical selection of biomass precursors by short-loop autocatalysis, and a form of protection against the fragility of the autocatalytic threshold. We will first describe the biochemistry and phylogenetics of carbon-fixation pathways in the current biosphere, and then show how their patterns of modularity and chemical redundancy provide a framework for historical reconstruction.

3. Core carbon metabolism

Currently six carbon fixation pathways are known [76, 77]. While they are distinct as complete pathways, they have significant overlaps at the level of individual reactions, and even greater redundancy in local-group chemistry. They are also, as shown in figure 4, tightly integrated with the main pathways of core carbon metabolism, including lipid synthesis, gluconeogenesis, and pentose-phosphate synthesis.

An extensive analysis of their chemistry under physiologically relevant conditions has shown that individual fixation pathways contain two groups of thermodynamic bottlenecks: carboxylation reactions, and carboxyl reduction reactions [98]. In isolation these reactions generally require ATP hydrolysis to proceed, and the way pathways deal with (or avoid) these costs has been concluded to form an important constraint on their internal structure [98]. We will further show how the elaborate and complex catalytic mechanisms associated with these reactions form essential *evolutionary* constraints on metabolism.

We will first describe the biochemical and phylogenetic details of the individual pathways, and then diagram their patterns of redundancy, first at the level of modular reaction sequences, and then in local-group chemistry. Finally we will use this decomposition together with evidence from gene distributions to propose their historical relation and identify constraints that could have spanned the Darwinian and pre-cellular eras.

3.1. Carbon fixation pathways

3.1.1. Overview of pathway chemistries, phylogeny and environmental context. *Wood–Ljungdahl.* The WL pathway [90–92, 94] consists of a sequence of five reactions that directly reduce one CO_2 to a methyl group, a parallel reaction reducing CO_2 to CO, and a final reaction combining the methyl and CO groups with each other and with a molecule of CoA to form the thioester acetyl-CoA. The reactions are shown below in figure 5, and discussed in detail in section 4. The five steps reducing CO_2 to $-CH_3$ make up the core pathway of folate (vitamin B_9) chemistry and its archaeal analogue, which we consider at length in section 3.2. The reduction to CO, and the synthesis of acetyl-CoA, are performed by the bi-functional CO-dehydrogenase/acetyl-CoA synthase (CODH/ACS), a highly conserved enzyme complex with Ni–[Fe_4S_5] and Ni–Ni–[Fe_4S_4] centers [99–102]. Methyl-transfer from pterins to the ACS active site is performed by a corrinoid iron–sulfur protein (CFeSP) in which the cobalt-tetrapyrrole cofactor cobalamin (vitamin B_{12}) is part of the active site [103, 104].

Phylogenetically, WL is a widely distributed pathway, found in a variety of both bacteria and archaea, including acetogens, methanogens, sulfate reducers, and possibly anaerobic ammonium oxidizers [76]. The full pathway is found only in strict anaerobes, because the CODH/ACS is one of the most oxygen-sensitive enzymes known [105, 106]. However, as we have argued in [22], the folate-mediated reactions form a partly-independent sub-module. This module combines with the equally-distinctive CODH/ACS enzyme to form the complete WL pathway, but can serve independently as partial carbon-fixation pathways even in the absence of the final step to acetyl-CoA (see figure 5). In this role it is found almost universally among deep bacterial clades. In addition to its being highly oxygen sensitive, recent results [309] suggest that the CODH/ACS is also sensitive to sulfides and perhaps other oxidants, a point to which we will return in section 3.4.2 when discussing evolutionary divergences between the complete and incomplete forms of WL.

All carbon fixation pathways in extant organisms employ some essential and apparently unique enzymes and most also rely in essential ways on certain cofactors. For example, the 3-hydroxypropionate pathway relies on biotin for reactions shared with (or homologous to) those in fatty acid synthesis. The reductive citric-acid cycle relies on reduced ferredoxin [107], a simple iron–sulfur enzyme, and on thiamin in its reductive carbonyl-insertion reaction [108], and also on biotin for its β-carboxylation steps [109, 110] (all of these examples will also be discussed along with the pathways in which they are used below). Among the uses of cofactors in carbon-fixation pathways, however, the function provided by pterin cofactors in WL is distinct and arguably the most complex. (*Pterin* is a name referring to the class of cofactors including folates and the methanopterins, which are both derived from a neopterin precursor.) Whereas most cofactors act as transfer agents cycling between two or three states, pterins undergo elaborate multi-step cycles, mediating capture of formate, reduction of carbon bound to one or two nitrogen atoms and transfer of formyl, methylene, or methyl groups. This
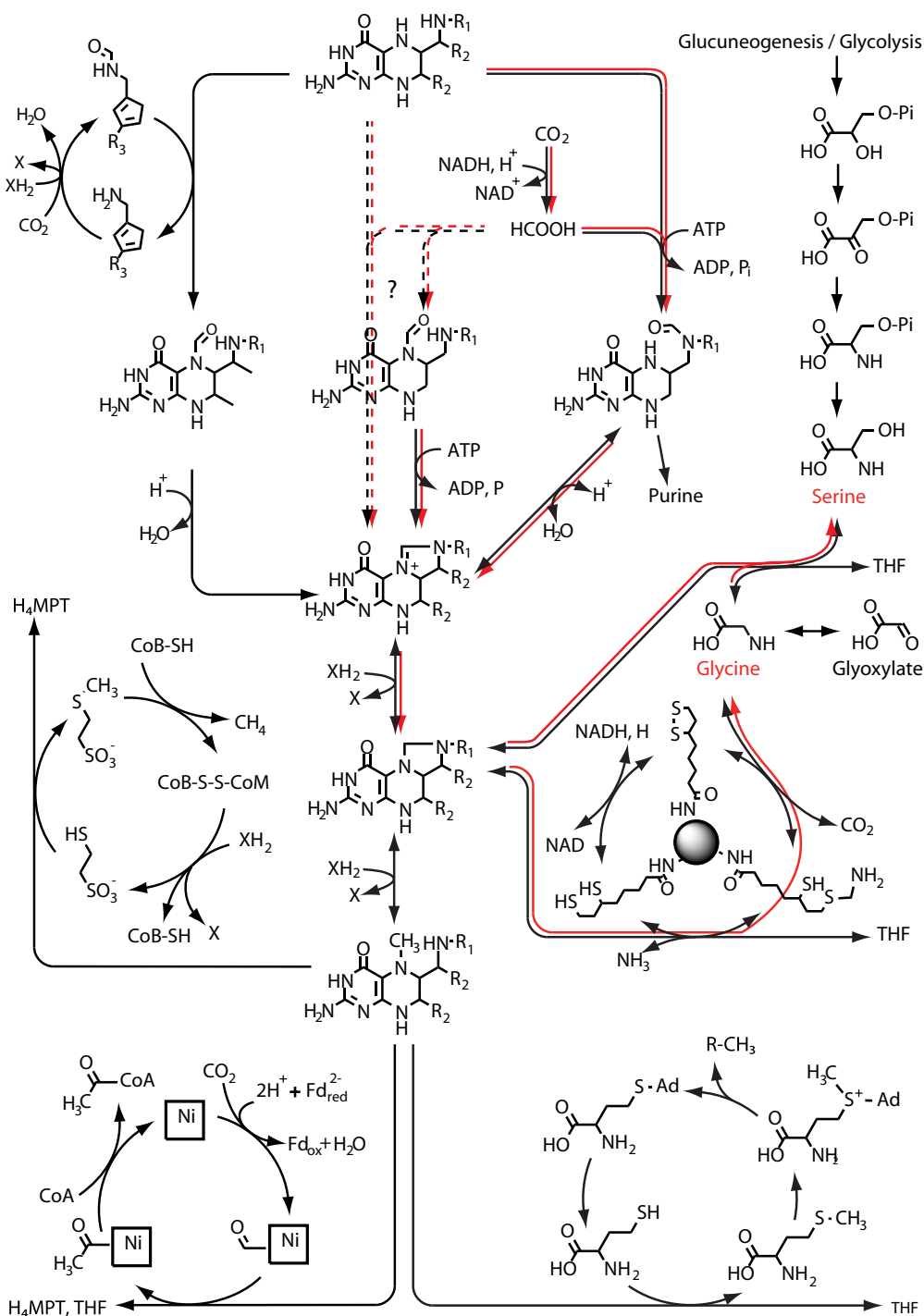
Figure 4. The projection of the complete network for core carbon anabolism onto its CHO components. Phosphorylated intermediates and thioesters with coenzyme-A are not shown explicitly. The bipartite graph notation used to show reaction stoichiometry is explained in appendix B. Arcs of the reductive citric acid cycle and gluconeogenesis are bold, showing that these pathways pass through the universal biosynthetic precursors. The Wood–Ljungdahl (labeled WL) pathway, without its cofactors and reductants shown, is represented by the last reaction of the acetyl-CoA synthase, which is the inverse of a disproportionation. Abbreviations: acetate (ACE); pyruvate (PYR); oxaloacetate (OXA); malate (MAL); fumarate (FUM); succinate (SUC); α-ketoglutarate (AKG); oxalosuccinate (OXS); isocitrate (ISC); cis-aconitate (CAC); citrate (CIT); malonate (MLN); malonate semialdehyde (MSA); 3-hydroxypropionate (3HP); acrolyate (ACR); propionate (PRP); methylmalonate (MEM); succinate semialdehyde (SSA); 4-hydroxybutyrate (4HB); crotonate (CRT); 3-hydroxybutyrate (3HB); acetoacetate (AcACE); butyrate (BUT); hydroxymethyl-glutarate (HMG); glyoxylate (GLX); methyl-malate (MML); mesaconate (MSC); citramalate (CTM); glycerate (GLT); glyceraldehyde (GLA); dihydroxyacetone (DHA); fructose (FRC); erythrose (ERY); sedoheptulose (SED); xylulose (XYL); ribulose (RBL); ribose (RIB).

diversity of roles has led to the folate pathway being termed the 'central superhighway' of C_1 chemistry [93]. The distinctive use of cofactors within WL continues with the dependence of the acetyl-CoA synthesis on cobalamin, a highly reduced tetrapyrrole capable of two-electron transfer [103]. In this sense the simple network topology of direct C_1 reduction seems to require a more elaborate dependence on cofactors than is seen in other pathways.

Reductive citric-acid cycle. The reductive citric-acid (reductive TCA or rTCA) cycle [111, 96] is the reverse of the oxidative Krebs cycle. It is a sequence of eleven intermediates and eleven reactions, highlighted in figure 4, which reduce two

molecules of CO_2, and combine these through a substrate-level phosphorylation with CoA, to form one molecule of acetyl-CoA. In the cycle, one molecule of oxaloacetate grows by condensation with two CO_2 and is reduced and activated with CoA. The result, citryl-CoA, undergoes a retro-aldol cleavage to regenerate oxaloacetate and acetyl-CoA. Here we separate the formation of citryl-CoA from its subsequent retro-aldol cleavage, as this is argued to be the original reaction sequence, and the one displaying the closest homology in the substrate-level phosphorylation with that of succinyl-CoA [112, 113]. A second arc of reactions, sometimes termed *anaplerotic* [5], then condenses two further CO_2 with acetyl-CoA to

Figure 5. The reactions in the WL pathway of direct C_1-reduction. The main sequence on pterins is shown, with five outputs for formyl, methylene, or methyl groups. The semi-independent submodule often used to directly synthesize glycine and serine from CO_2, even when acetyl-CoA synthesis is absent, is highlighted in red. Alternative pathways to glycine and serine, from 3-phosphoglycerate in gluconeogenesis/glycolysis and glyoxylate, are shown in the upper right quadrant. Finally, the dashed arrows represent a suggested alternative form of formate uptake based on binding at N^5 rather than N^{10} of folate before cyclization to methenyl-THF [22].

produce a second molecule of oxaloacetate, completing the network-autocatalytic topology and making the cycle self-amplifying. The distinctive reaction in the rTCA pathway is a carbonyl insertion at a thioester (acetyl-CoA or succinyl-CoA), performed by a family of conserved ferredoxin-dependent oxidoreductases which are triple-Fe_4S_4-cluster proteins [108]. The cycle is found in many anaerobic and microaerophilic

bacterial lineages, including Aquificales, Chlorobi and δ- and ϵ-proteobacteria.

Enzymes from reductive TCA reactions are very widely distributed among bacteria, where in addition to complete cycles they support fermentative pathways that break cycling and use intermediates such as succinate as terminal electron acceptors [68]. The distribution of full rTCA cycling correlates

with clades whose origins are placed in a pre-oxygenic earth, while fermentative TCA arcs and the oxidative Krebs cycle are found in clades that by phylogenetic position (such as γ-, β- and α-proteobacteria) and biochemical properties (membrane quinones and intracellular redox couples involving them [114]) arose during or after the rise of oxygen. The co-presence of enzymes descended from both the reductive and oxidative cycles [67, 69, 115] in members of clades that at a coarser level are known to straddle the rise of oxygen, such as Actinobacteria and proteobacteria [114], may provide detailed mechanistic evidence about the reversal of core metabolism from ancestral reductive modes to later derived oxidative modes.

Functionally, the transition to an oxidizing earth is complex in two ways. First, oxygenation is unlikely to have occurred homogeneously within the oceans [116–118]. For bacteria not harbored in sheltered environments, this likely created complex needs to responsive phenotypes. Second, the co-presence of strong oxidants with geochemical reductants could in some cases provide an energy source during oxygenation, but as we see on the current Earth as well as in the history of banded iron, oxidants eventually scavenge reducing equivalents from most environments, likely leaving fermentative pathways as a fallback energy source for many organisms. It is plausible that both of these geochemical responses served as pre-adaptations enabling the complex host-associated lifecycles of bacteria and archaea descended within clades that straddled this transition.

Earlier work [71] that also recognized the central position of TCA reactions has attempted to argue for an ancestral *oxidative* TCA cycle. However, the optimal-path part of this argument, and similar path-length optimality arguments in [55, 70, 72], apply broadly to intermediary metabolism as a consequence of its limited reaction types or cross-linking as explained in section 2.1, without implying directionality. Furthermore, the functional criterion of acetate oxidation to produce reducing equivalents to drive ATP production through oxidative phosphorylation relies on the assumption that the organic interconversion of acetate to CO_2 was the main redox couple of early metabolism. Prior to the rise of oxygen, however, we find many inorganic electron acceptors such as elemental sulfur in redox couples with inorganic electron donors such as H_2 and not with organics of stoichiometry CH_2O [119–121]. For these reasons, combined with the strict absence of oxidative TCA as a plausible genetically reconstructed form in deep-branching clades, we find arguments for an ancestral rTCA cycle [7, 41, 122], replaced possibly via fermentative intermediates, by a later oxidative Krebs cycle, more convincing.

Dicarboxylate/4-hydroxybutyrate cycle. The dicarboxyate/4-hydroxybutyrate (DC/4HB) cycle [94, 123], illustrated in figure 6 is, like rTCA, a single-loop network-autocatalytic cycle, but has a simpler form of autocatalysis in which acetyl-CoA rather than oxaloacetate is the network catalyst. Only two CO_2 molecules are attached in the course of the cycle to form acetoacetyl-CoA, which is then thioesterified at the second acetyl moiety and cleaved to directly regenerate two molecules of acetyl-CoA. An extra copy of the network

catalyst is thus directly regenerated (with suitable CoA activation) without the need for anaplerotic reactions. The cycle has so far been found only in anaerobic crenarchaeota, but within this group it is believed to be widely distributed phylogenetically [94, 123]. The first five reactions in the cycle (from acetyl-CoA to succinyl-CoA) are identical to those of rTCA. The second arc of the cycle begins with reactions found also in 4-hydroxybutyrate and γ-aminobutyrate fermenters in the Clostridia (a subgroup of Firmicutes within the bacteria), and terminates in the reverse of reactions in the isoprene biosynthesis pathway. The DC/4HB pathway thus uses the same ferredoxin-dependent carbonyl-insertion reaction used in rTCA (though only at acetyl-CoA), along with distinctive reactions associated with 4-hydroxybutyrate fermentation. In particular, the dehydration/isomerization sequence from 4-hydroxybutyryl-CoA to crotonyl-CoA is performed by a flavin-dependent protein containing an [Fe4-S4] cluster, and involves a ketyl-radical intermediate [124, 125].

3-hydroxypropionate bicycle. The 3-hydroxypropionate (3HP) bicycle [106], highlighted in figures 6 and 7, has the most complex network topology of the fixation pathways, using two linked cycles to regenerate its network catalysts and to fix carbon. The network catalysts in both loops are acetyl-CoA and the outlet for fixed carbon is pyruvate. The reactions in the cycle begin with the biotin-dependent carboxylation of acetyl-CoA to form malonyl-CoA, from the fatty-acid synthesis pathway, followed by a distinctive thioesterification [126] and a second, homologous carboxylation of propionyl-CoA (to methylmalyl-CoA) followed by isomerization to form succinyl-CoA. The first cycle then proceeds as the *oxidative* TCA arc, followed by retro-aldol reactions also found in the glyoxylate shunt (described below). A second cycle is initiated by an aldol condensation of propionyl-CoA with glyoxylate from the first cycle to yield β-methylmalyl-CoA, which follows a sequence of reduction and isomerization through an enoyl intermediate (mesaconate) similar to the second cycle of rTCA or the 4HB pathway. This complex pathway was discovered in the Chloroflexi and is believed to represent an adaptation to alkaline environments in which the CO_2/HCO_3^- (bicarbonate) equilibrium strongly favors bicarbonate. All carbon fixations proceed through activated biotin, thus avoiding the carbonyl insertion of the rTCA and DC/4HB pathways. While topologically complex, the bicycle makes extensive use of relatively simple aldol chemistry, which we will argue in section 3.2.4 made its evolutionary innovation less improbable than the topology alone might suggest.

3-hydroxypropionate/4-hydroxybutyrate cycle. The 3-hydroxypropionate/4-hydroxybutyrate (3HP/4HB) cycle [127], shown in figure 6, is a single-loop pathway in which the first arc is the 3HP pathway, and the second arc is the 4HB pathway. Like DC/4HB, 3HP/4HB uses acetyl-CoA as network catalyst and fixes two CO_2 to form acetoacetyl-CoA. The pathway is found in the Sulfolobales (Crenarchaeota), where it combines the crenarchaeal 4HB pattern of autotrophic carbon fixation with the bicarbonate adaptation of the 3HP pathway. Like the 3HP bicycle, the 3HP/4HB pathway is thought to be an adaptation to alkalinity, but because the 4HB
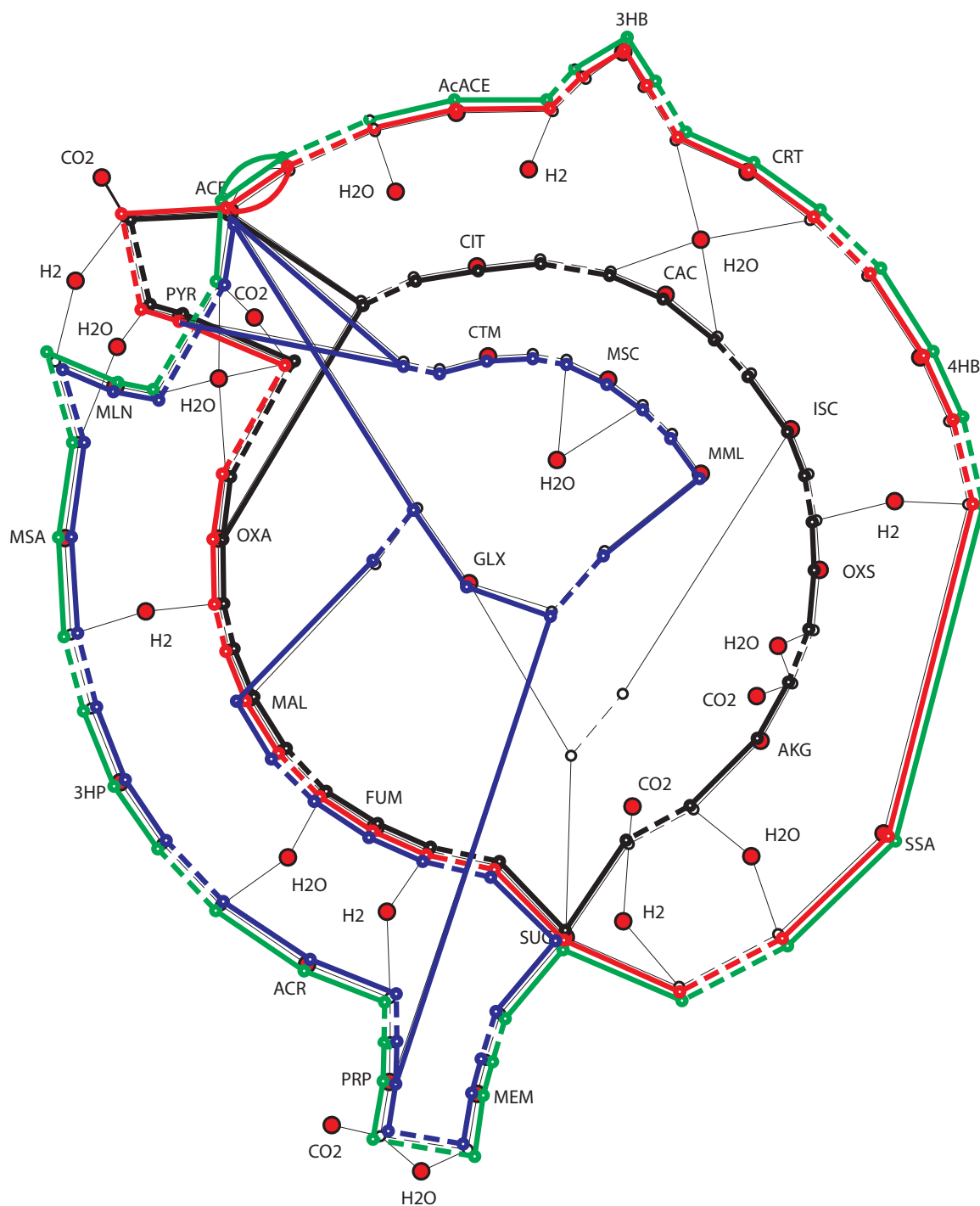
Figure 6. The four loop carbon-fixation pathways that pass through some or all of the universal biosynthetic precursors, from the graph of figure 4. rTCA is black, DC/4HB is red, 3HP-bicycle is blue, and 3HP/4HB is green. The one aldol reaction from the glyoxylate shunt that is not part of the 3HP-bicycle is shown in fine lines. The module-boundary nature of acetate (ACE) and succinate (SUC) is shown by the intersection of multiple paths in these compounds. Radially aligned reactions are homologous in local-group chemistry; deviations from strict homology in different pathways appear as excursions from concentric circles.

arc does not fix additional carbon, this adaptation resulted in a simpler pathway structure than the bicycle.

Calvin–Benson–Bassham cycle. The Calvin–Benson–Bassham (CBB) cycle [128, 129] is responsible for most known carbon fixation in the biosphere. In the same way

as WL adds only the distinctive CODH/ACS reaction to an otherwise very-widely-distributed folate pathway [22], the CBB cycle adds a single reaction to the otherwise-universal network of aldol reactions among sugar-phosphates that make up the gluconeogenic pathway to fructose 1,6-bisphosphate
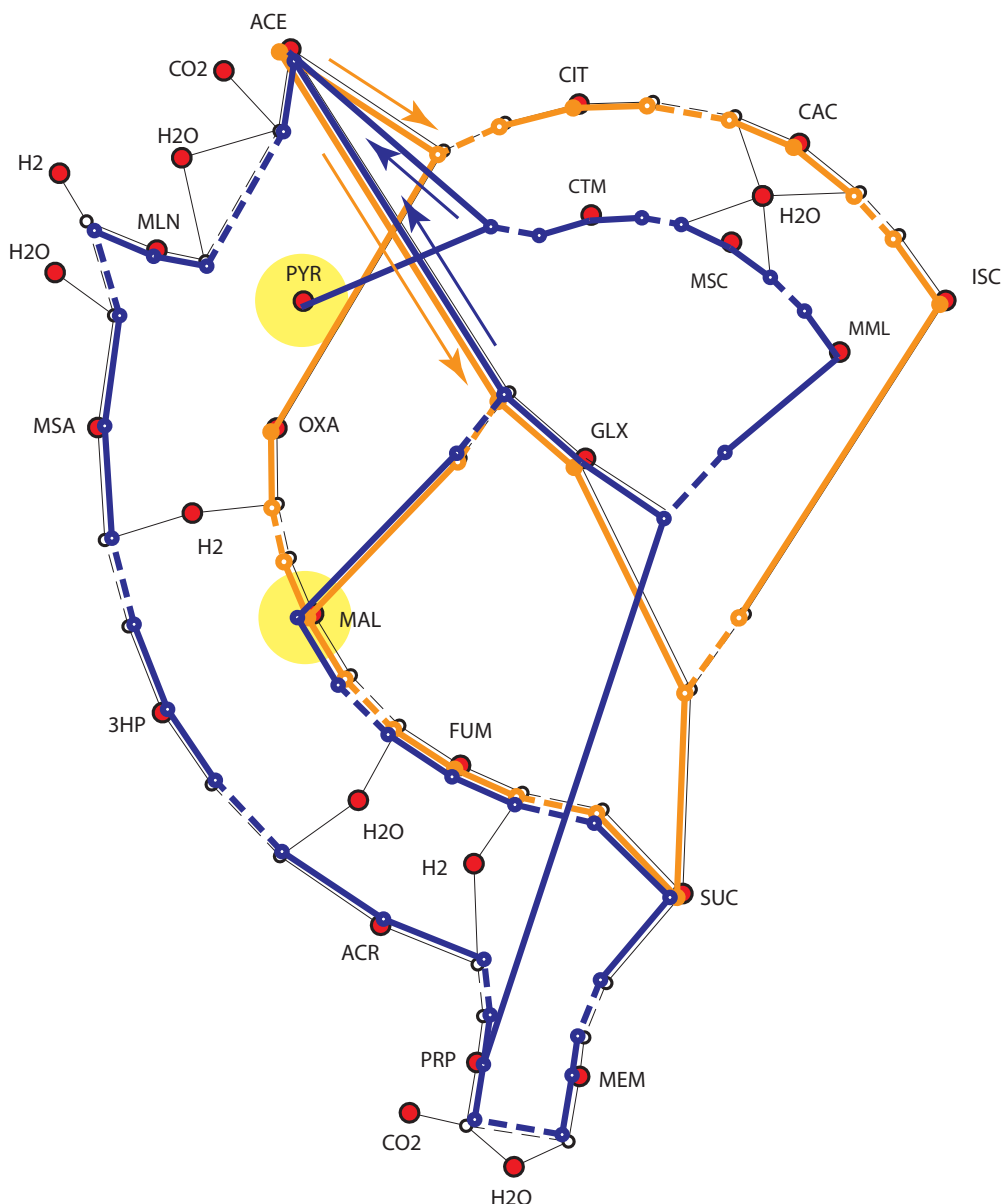
Figure 7. The 3-hydroxypropionate bicycle (blue) and the glyoxylate shunt (orange) compared. Directions of flow are indicated by arrows on the links to acetate (ACE). The common core that enables flux recycling in both pathways is the aldol reaction between glyoxylate (GLX), acetate and MAL. The four other aldol reactions (labeled by their cleavage direction) are from ISC, methyl-malate (MML), CIT and citramalate (CTM). Malate is a recycled network catalyst in both pathways. Carbon is fixed in the 3HP-bicycle as pyruvate (PYR), so the cycle only becomes autocatalytic if pyruvate can be converted to MAL through anaplerotic (rTCA) reactions.

and the reductive pentose phosphate pathway to ribose and ribulose 1,5-bisphosphate[3]. The distinctive CBB reaction that extends reductive pentose-phosphate synthesis to a carbon fixation cycle is a carboxylation performed by the ribulose 1,5-bisphosphate carboxylase/oxygenase (RubisCO), together with cleavage of the original ribulose moiety to produce two molecules of 3-phosphoglycerate. The Calvin cycle resembles

the 4HB pathways in regenerating two copies of the network catalyst directly, not requiring separate anaplerotic reactions for autocatalysis. In addition to carboxylation, RubisCO can react with oxygen in a process known as *photorespiration* [130–132] to produce 2-phosphoglycolate (2PG), a precursor to glyoxylate that is independent of rTCA-cycle reactions. The CBB cycle is widely distributed among cyanobacteria, in chloroplasts in plants and in some secondary endosymbionts.

The glyoxylate shunt. Although it is not an autotrophic carbon-fixation pathway, the glyoxylate shunt (or glyoxylate bypass) is of interest because it shares intermediates and reactions with many of the above fixation pathways, and because it resembles a fixation pathway in certain topological features. The pathway is shown in figure 7. All aldol reactions

---

[3] The universality of this network requires some qualification. We show a canonical version of the network in figure 4, and some variant on this network is present in every organism that synthesizes ribose. However, the $(CH_2O)^n$ stoichiometry of sugars, together with the wide diversity of possible aldol reactions among sugar-phosphates, make sugar re-arrangement a problem in the number theory of the small integers, with solutions that may depend sensitively on allowed inputs and outputs. Other pathways within the collection of attested pentose-phosphate networks are shown in [87].

that can be performed starting from rTCA intermediates appear in this pathway, either as cleavages or as condensations. In addition to condensation of acetate and oxaloacetate to form citrate (CIT), these include cleavage of isocitrate (ISC) to form glyoxylate and succinate, and condensation of glyoxylate and acetate to form malate (MAL). The shunt is a weakly oxidative pathway (generating one H_2-equivalent from oxidizing succinate to fumarate), and is otherwise a network of internal redox reactions. It is therefore a very widely-used facultative pathway under conditions where carbon for biosynthesis, more than reductant, is limiting.

Two of the arcs of the shunt overlap with arcs in the oxidative Krebs cycle, but the entire pathway is a bicycle much like the 3HP-bicycle, sharing many of the same intermediates, but running in the opposite direction. Oxidative pathways such as the Krebs cycle are ordinarily catabolic, and hence not self-maintaining. The glyoxylate shunt may be regarded as a network-autocatalytic pathway for intake of acetate, using MAL as the network catalyst and regenerating a second molecule of MAL from two acetate molecules. This may be part of the reason that the shunt is up-regulated in the Deinococcus–Thermus family of bacteria in response to radiation exposure [133], providing additional robustness from network topology under conditions when metabolic control is compromised.

3.1.2. Thermodynamic constraints on pathway structure. The central energetic costs of carbon-fixation pathways are associated with carboxylation reactions in which CO_2 molecules are added to the growing substrate, and the subsequent reactions in which the carboxyl group is reduced to a carbonyl [98]. In isolation these reactions require ATP hydrolysis, but these costs can be avoided in several ways. In some cases a thioester intermediate is used to effectively couple together a carboxyl reduction and a subsequent carboxylation, allowing the two reactions to be driven by a single ATP hydrolysis. An unfavorable (endergonic) reaction can also be coupled to a highly favorable (exergonic) reaction, allowing the reactions to proceed without ATP hydrolysis.

Individual pathways employ such couplings to varying degrees, resulting in a range of ATP costs associated with carbon fixation. At the low end, WL eliminates nearly all use of ATP through its unique pathway chemistry, requiring only a single ATP in the synthesis of pyruvate from CO_2. This ATP is associated with the attachment and activation of formate on folates. Reducing CO_2 to free formate prior to attachment, and further reducing the activated formate on folates prior to incorporation into growing substrates saves one ATP associated with carboxylation. Additional ATP costs are saved by coupling the endergonic reduction of CO_2 to CO to the subsequent exergonic synthesis of acetyl-CoA. Finally, the activated thioester bond of acetyl-CoA allows the subsequent carboxylation to pyruvate to also proceed without additional ATP. In methanogens even the ATP cost of attaching formate to folates has been eliminated by modifying the structure of tetrahydrofolate (THF) to that of H_4MPT [22], enabling a membrane-bound iron–sulfur system to serve as energy source. Similarly, rTCA has high energetic efficiency

as a result of extensive reaction coupling, requiring only two ATP to synthesize pyruvate from CO_2 [94, 98]. Two ATP are saved by coupling carboxyl reductions to subsequent carboxylations using thioester intermediates, and an additional ATP is saved by coupling the carboxylation of α-ketoglutarate to the subsequent carbonyl reduction leading to ISC.

At the high end of energetic cost of carbon-fixation are pathways that couple unfavorable reactions less effectively, or not at all, or even hydrolyze ATP for reactions other than carboxylation or carboxyl reduction. Both the DC/4HB pathway and the 3HP bicycle decouple one or more of the thioester-mediated carboxyl reduction + carboxylation sequences of the kind used in rTCA, and neither couples endergonic carboxylations to exergonic reductions. As a result, DC/4HB requires five ATP and the 3HP bicycle seven ATP to synthesize pyruvate from CO_2 [94, 98]. The 3HP/4HB pathway has the highest cost of any fixation pathway, with nine ATP required to synthesize pyruvate from CO_2. This is partly because it also decouples thioester-mediated carboxyl reduction + carboxylation sequences, and partly because pyruvate is synthesized by diverting and ultimately *decarboxylating* succinyl-CoA [77, 94, 98]. Finally, CBB is also at the high end in terms of cost, requiring seven ATP to synthesize pyruvate from CO_2. Although this pathway avoids the cost of carboxylation reactions by coupling them to exergonic cleavage reactions, CBB is the only fixation pathway that invests ATP hydrolysis in chemistry other than carboxylations or carboxyl reductions, thereby increasing its relative cost [98].

3.1.3. Centrality and universality of the reactions in the citric-acid cycle, and the pillars of anabolism. The apparent diversity of six known fixation pathways is unified by the role of the citric-acid cycle reactions, and secondarily by that of gluconeogenesis and the pentose-phosphate pathways. Figure 4 showed the C, H, O stoichiometry for a network of reactions that includes all six known pathways. Here, by *stoichiometry* we refer to the mole-ratios of reactants and products for each reaction, with molecules represented by their CHO constituents, and attached phosphate or thioester groups omitted. Where phosphorylation or thioesterification mediates a net dehydration, we have represented the dehydration directly in the figure. The network contains only 35 organic intermediates, because many intermediates and reactions appear in multiple pathways. Hydroxymethyl-glutarate and butyrate are also shown, to indicate points of departure to isoprene and fatty acid synthesis, respectively.

In figure 4 the TCA cycle and the gluconeogenic pathway are highlighted. Beyond being mere points of departure for alternative fixation pathways and for diversifications in intermediary metabolism, they are invariants under diversification because they determine carbon flow among the universal precursors of biosynthesis.

Almost all anabolic pathways in extant organisms originate in one of five intermediates in the TCA cycle— acetate (as acetyl-CoA), pyruvate, oxaloacetate, succinate (or succinyl-CoA) or α-ketoglutarate—which have been dubbed the 'pillars of anabolism' [51]. Succinyl-CoA can serve as
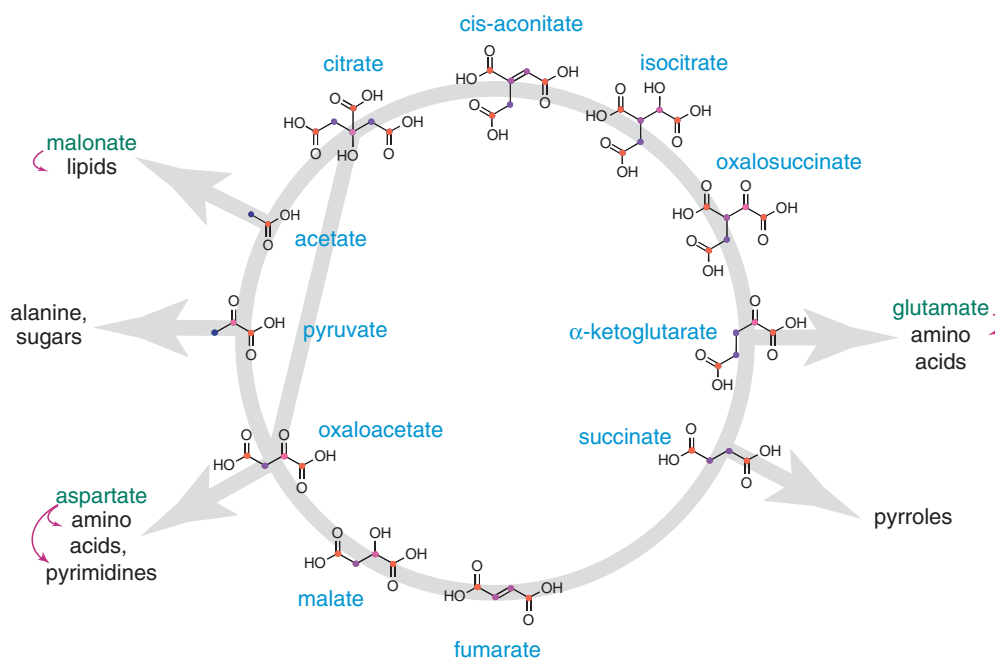
Figure 8. The pillars of anabolism, showing lipids, sugars, amino acids, pyrimidines and purines, and tetrapyrroles from either succinate or AKG. Molecules with homologous local chemistry are at opposite positions on the circle. Oxidation states of internal carbon atoms are indicated by color (red = oxidized, blue = reduced).

the precursor to pyrroles (metal-coordinating groups in many cofactors)—mainly in α-proteobacteria and mitochondria—but these are more commonly made from α-ketoglutarate via glutamate in what is known as the C5 pathway [134]. A phylogenetic analysis of these pathways confirms that the C5 pathway is the most plausible ancestral route to pyrrole synthesis [135]. (An even stronger claim has been made, from a similar combination of biochemical and phylogenetic inputs, that it is a likely relic of the RNA world [136].) Thus as few as four TCA intermediates provide the organic inputs to all anabolic pathways. Figure 8 shows the major molecule classes associated with each intermediate. The only exceptions to this universality, which form a biosynthetic sequence, are glycine, serine, and a few compounds synthesized from them; this sequence can be initiated directly from CO_2 outside the pillars (see figure 5). This observation becomes key in reconstructing the evolutionary history of carbon-fixation (see section 3.3). The gluconeogenic pathway then forms a similarly unique bridge between the TCA intermediate pyruvate (in the activated form phosphoenolpyruvate) and the network of sugar-phosphate reactions known as the *pentose-phosphate pathway*.

Carbon-fixation pathways must reach all four (or five) of the universal anabolic starting compounds. They may do this either by producing them as pathway intermediates, or by means of secondary reactions converting pathway intermediates into the essential precursors. The degree to which a pathway passes through all essential biosynthetic precursors may suggest its antiquity. In metabolism-first theories of the origin of life [6], the limited set of compounds selected and made available in high concentration by proto-metabolism determined the opportunities for further biosynthesis, thus establishing themselves as the precursors of anabolism.

Among the five network-autocatalytic fixation pathways, the CBB pathway is unique in not passing through *any* universal anabolic precursors. When used as a fixation pathway, CBB reactions must thus be connected to the rest of anabolism through several reactions in the glycolytic pathways connecting 3-phosphoglycerate (3PG) to pyruvate. Pyruvate is then connected to the remaining precursors through partial TCA sequences. The glycolytic pathway is the primary connection of CBB to the anabolic precursors, but 2-phosphoglycolate (2PG) produced during photorespiration may also be converted to glyoxylate and subsequently to glycine and serine (see figure 5). Glycine synthesis from photorespiration is not itself a carbon-fixing process, but rather a salvage pathway to compensate for poor discrimination of the enzyme RuBisCO. 2PG is produced from ribulose-1,5-bisphosphate when O_2 replaces CO_2 in the RuBisCo uptake reaction. This toxin inhibits RuBisCO, and would require excretion if it could not be recycled, leading to a net loss of carbon from the pentose-phosphate network. However, most RuBisCO uptake events do fix CO_2, and the carbon circulating in the pentose-phosphate pathway in CBB organisms is the product of these successful fixation cycles. Thus, photorespiration with glycine salvage amounts to a variant elaboration of the fixation pathway to include null cycles, and the connection of this more complex process to the precursor set.

Among the remaining loop-fixation pathways, only rTCA passes through all five anabolic pillars. Through its partial overlap with rTCA, DC/4HB passes through four, excluding α-ketoglutarate. The 3HP-bicycle further bypasses oxaloacetate, while the 3HP/4HB loop and WL include only acetyl-CoA. All of the latter pathways require anaplerotic reactions in the form of incomplete (either oxidative or reductive) TCA arcs; when

these combine (in various ways) with WL carbon fixation, they are known collectively as the *reductive acetyl-CoA pathways*.

The most parsimonious explanation for the universality of the TCA arcs as anaplerotic reactions is lock-in by downstream anabolic pathways, to which metabolism was committed by the time carbon-fixation strategies diverged. This can also be understood as a direct extension of the metabolism-first assumption that anabolic pathways themselves formed around proto-metabolic selection of the rTCA intermediates[4]. (A similar but later form of commitment has been argued to convert basal gene regulatory networks in metazoan development into *kernels*, which admit no variation and act as constraints on subsequent evolutionary dynamics [137, 138].) If lock-in provides the correct interpretation of TCA universality, then much of the burden of accounting for the inventory of small metabolites is shifted away from Darwinian selection for function in a post-RNA world, and onto constraints of biosynthetic simplicity and network context. We show below that phylogenetic reconstruction is consistent with a selective role for rTCA cycling in the root metabolism of cellular life, though only as part of a larger network than the modern rTCA cycle.

3.2. Modularity in the internal structure and mutual relationships of the known fixation pathways

Figure 4 shows that the number of molecules and reactions required to include all carbon fixation pathways is much smaller than might have been expected from their nominal diversity, because many reactions are used in multiple pathways, and all pathways remain close to the universal precursors. We have already noted in the previous section that this re-use goes beyond the requirements of autocatalysis, to the anaplerotic role of rTCA arcs adapting variant fixation pathways to an invariant set of biosynthetic precursors.

The aggregate network also shows many kinds of structure: clusters, concentric rings and ladders reflecting parallel sequences of the same inputs and outputs in different pathways. We will show in this section that these result from re-use of local-group chemistry in transformations of distinct molecules.

At the end of the section we will describe a third form of re-use not represented in the aggregate graph. The folate-mediated direct C_1 reduction sequence of WL, responsible for production of the methyl group used in the synthesis of acetyl-CoA in figure 4, is also found as a free-standing fixation pathway across the bacterial tree, often as one component in a disconnected autotrophic network using one of the loop fixation pathways as its other component.

Because of such extensive redundancy, little innovation is required to explain the extant diversity of carbon fixation. All known carbon fixation strategies can be described as assemblies of a small number of strongly-defined *modules*, which govern not only the function of pathways, but also their evolution.

---

[4] Harold Morowitz summarizes this assumption with the phrase *metabolism recapitulates biogenesis* [6].

3.2.1. Modularity in carbon fixation loops from re-use of pathway segments.
Figure 6 shows the sub-network from figure 4 containing the four loop-autotrophic carbon fixation pathways that pass through some or all universal precursors, together with reactions in the glyoxylate shunt. The four loop pathways are shown in four colors, with the organic pathway-intermediates (but not environmental precursors or reductants) highlighted.

The figure shows that these pathways re-use intermediates by combining entire *pathway segments*. The combinatorial assembly of these segments is possible because they all pass through acetate (as acetyl-CoA), succinate (usually as succinyl-CoA), and all except the second loop of the 3HP bicycle pass through both. Thus the conserved reactions among the autocatalytic loop carbon-fixation pathways are shared within strictly preserved sequences, which have key molecules as the boundaries at which segments may be combined.

3.2.2. Homologous local-group chemistry across pathway segments.
In addition to the re-use of complete reactions in pathway segments, variant carbon-fixation pathways have extensively re-used transformations at the level of local functional groups. The network of figure 6 is arranged in concentric rings, in which the arcs of the rTCA cycle align with the 3HP or 4HB pathways, or with the mesaconate arc of the 3HP bicycle. The 'ladder' structure of inputs and outputs of reductant (H_2) or water between these rings shows the similar stoichiometric progression in these alternative pathways. Figure 9 decomposes the aggregate network into two pairs of short-molecule and long-molecule arcs, and the mesaconate arc, and shows the pathway intermediates in each arc. The figure makes clear that, both *within* the arcs of the loop pathways, and *between* alternate pathways, the type, sequence and position of reactions are highly conserved. In particular, the reduction sequence from α-ketones or semialdehydes, to alcohols, to isomerization through enoyl intermediates, is applied to the same bonds on the same carbon atoms from input acetyl moieties in rTCA, 3HP and 4HB pathways, and to analogous functional groups in the bicycle. Finally, in the cleavage of both citryl-CoA and citramalyl-CoA, the bond that has been isomerized through the enoyl intermediate is the one cleaved to regenerate the network catalyst.

Even the distinctive step to crotonyl-CoA in the 4HB pathway creates an aconate-type intermediate, and the enzyme responsible has high homology to the acrolyl-CoA synthetase [139, 140], whose output (acrolyl-CoA) follows the standard pattern. Only the position of the double bond breaks the strict pattern in crotonyl-CoA, and the abstraction of the un-activated proton required to produce this bond requires the unique ketyl-radical intermediate [141]. From crotonyl-CoA, the sequence to 3-hydroxybutyrate is then followed by a surprising *oxidation* and re-hydration, resulting in a five-step, redox-neutral, sequence. The net effect of this sequence is to shift the carbonyl group (of succinate semialdehyde, SSA) by one carbon (in acetoacetate, AcACE). Because the 4HB pathway takes in no new CO_2 molecules, this isomerization enables regeneration of the network catalysts in the same way
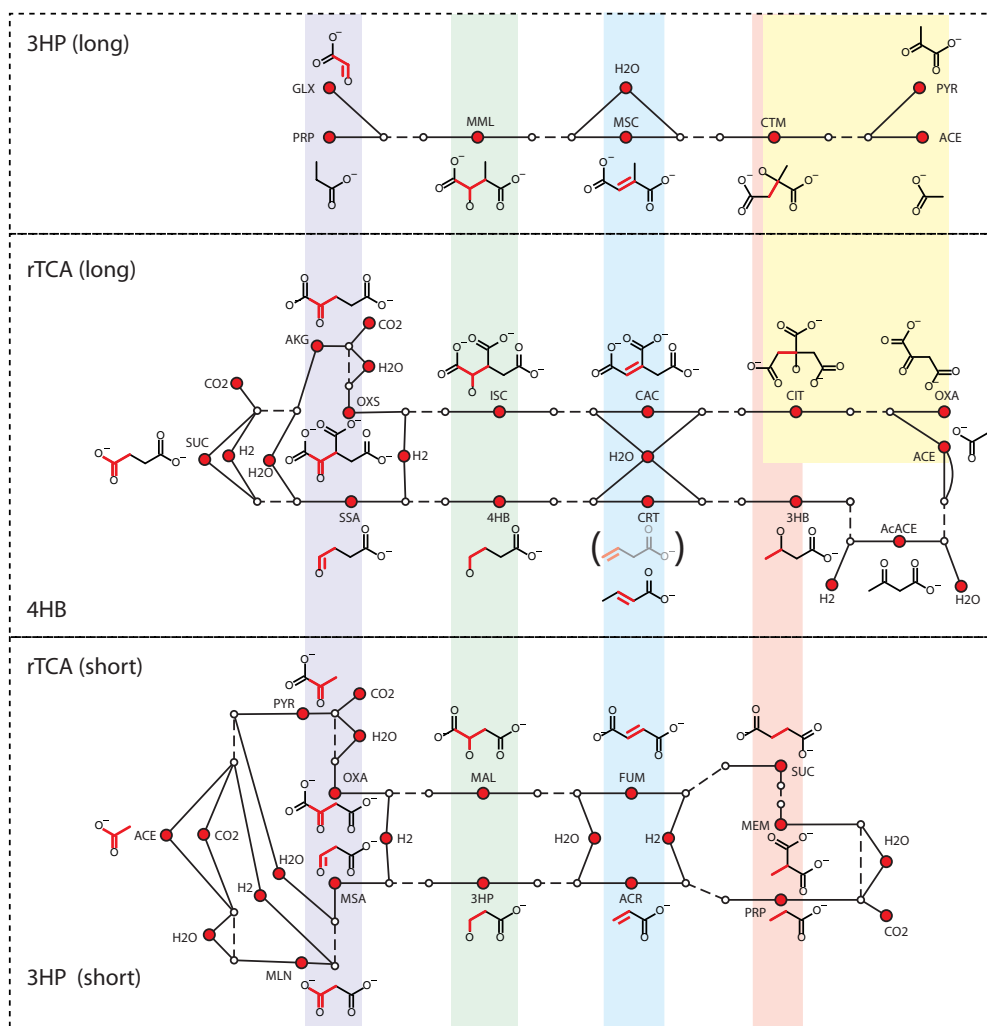
Figure 9. Comparison of redundant reactions in the loop carbon fixation pathways. Pathways are divided into 'long-molecule' (upper ranks) and 'short-molecule' (lower ranks) segments; long-molecule segments occupy roughly the upper-right half-plane in figure 6, and abbreviations are as in figure 4. Molecule forms are shown next to the corresponding tags. Bonds drawn in red are the active acetyl or semialdehyde moieties in the respective segments. Vertical colored bars align homologous carbon states. The yellow block shows retro-aldol cleavages of CIT or citramalate. Two molecules are shown beneath the tag CRT (crotonate): the grayed-out molecule in parentheses would be the homologue to the other aconitase-type reactions; actual crotonate (full saturation) displaces the double bond by one carbon, requiring the abstraction of the α-proton in 4-hydroxybutyrate via the ketyl-radical mechanism that is distinctive of this pathway.

the reduction/aldol-cleavage sequence enables regeneration for rTCA or the 3HP bicycle.

Duplication of reaction sequences in diverse fixation pathways appears to have resulted from retention of gene sets as organism clades diverged. Duplication of local-group chemistry in diverse reactions similarly appears to have resulted (at least in most cases) from retention of reaction mechanisms as enzyme families diverged. All enoyl intermediates are produced by a widely diversified family of aconitases [142], while biotin-dependent carboxylations are performed by homologous enzymes acting on pyruvate and α-ketoglutarate [110], and substrate-level phosphorylation and thioesterification are similarly performed by homologous enzymes on CIT and succinate in rTCA [112, 113]. Similar to the synthesis of citryl-CoA we separate here the carboxylation of α-ketoglutarate from the subsequent reduction of oxalosuccinate to ISC—performed by a single enzyme in most organisms—because it is argued to be the

ancestral form [143, 144]. However, the thioesterification of propionate in the 3HP pathway is performed by distinct enzymes in bacteria and archaea, an observation that has been interpreted to suggest convergent evolution [106, 127]. The widespread use of a few reaction types by a small number of enzyme classes/homologues may reflect their early establishment by promiscuous catalysts [95, 145], followed by evolution toward increasing specificity as intermediary metabolic networks expanded and metabolites capable of participating in carbon fixation diversified.

A functional identification of modules that seeks to minimize influence from historical effects (such as lock-in) has been carried out by Noor *et al* [72], and identifies similar module boundaries. Using as data the first three numbers of the EC classification of enzymes—which distinguish reaction types but coarse-grain over both substrate specificity and enzyme homology—they show that many pathways in core metabolism are the shortest routes possible between inputs

and products. This work builds on earlier studies showing that under very simple rules, the pentose phosphate pathway uses the minimal number of steps to connect inputs to outputs [70, 146]. In each of those studies, the authors suggested that Darwinian selective pressure may have led to a such minimal pathways as optimal network connections between *given* pairs of metabolites, with the implication that the selection of metabolites was based on some aspects of phenotypic function aside from their network positions.

If we do not presume that phenotypic selection preceded metabolism, however, the problem of pathway optimization ceases to be one with fixed endpoints, and causation may even run from pathways to the metabolite inventory. In this view minimal pathways may have been selected because their kinetics and regulation were easier to control. Starting points of downstream intermediary metabolism could then have been selected from the intermediates made available by fixation pathways. We argue in favor of a selection hierarchy of this form in section 3.4.4: shorter fixation pathways capable of attaining autocatalytic feedback offer fewer opportunities for dilution by parasitic side-reactions [147], and (reaction chemistry otherwise being equal) require less regulatory control. They may thus be the only sustainable forms.

Where the pathways analyzed by Noor *et al* overlap with those we have shown, many of their minimal sequences overlap with the modules in figure 6, as well as with others in gluconeogenesis which we do not consider here. Thus, from the perspective of an emerging metabolism, it may be that historical retention of a small number of reaction types reflects facility of the substrate-level chemistry, and that this has placed time-independent constraints on evolution.

The functional-group homology shown in figure 9 allows us to separate stereotypical sequences of widely diversified reactions from key reactions that distinguish pathways. The stereotypical sequences lie downstream of reactions such as the ferredoxin-dependent carbonyl insertion (rTCA), or biotin-dependent carboxylation (3HP), which are associated with highly conserved enzymes or cofactors. The downstream reactions are also more 'elementary', in the sense that they are common and widely diversified in biochemistry, compared to the pathway-distinguishing reactions.

3.2.3. Association of the initiating reactions with transition-metal sulfide mineral stoichiometries and other distinctive metal–ligand complexes. The observation that alternative fixation pathways are not distinguished by their internal reaction sequences, but primarily by their initiating reactions, suggests that these reactions were the crucial bottlenecks in evolution, and perhaps reflect the limiting diversity of chemical mechanisms for carbon bond formation. Mechanisms of organosynthesis in aqueous solution are especially limited by the instability of radical intermediates, which may be stabilized by association with metal centers. The distinctive use of metals in the (often highly conserved) enzymes and cofactors for these initiating reactions may thus suggest a direct link between prebiotic mineral and metal–ligand chemistry [148], and constraints inferable from the long-term structure of later cellular evolution.

Several enzyme iron–sulfur centers have been recognized [149, 150] to use strained versions of the unit cells found in Fe–S minerals, particularly Mackinawite and Greigite. These are particular instances within a wider use of transition-metal-sulfide chemistry in core-metabolic enzymes.

Pyruvate:ferredoxin oxidoreductase (PFOR), which catalyzes the reversible carboxylation of acetyl-CoA to pyruvate, contains three $[Fe_4S_4]$ clusters and a thiamin pyrophosphate (TPP) cofactor. The $[Fe_4S_4]$ clusters and TPP combine to form an electron transfer pathway into the active site, and the TPP also mediates carboxyl transfer in the active site [108].

The bifunctional carbon monoxide dehydrogenase/acetyl-CoA synthase (CODH/ACS) enzyme that catalyzes the final acetyl-CoA synthesis reaction in the WL pathway employs even more elaborate transition-metal chemistry. Like PFOR, this enzyme uses $[Fe_4S_4]$ clusters for electron transfer, but its active sites contain additional, more unusual metal centers. The CODH active site contains an asymmetric $Ni-[Fe_4S_5]$ cluster on which CO_2 is reduced to CO [99], while the ACS active site contains a $Ni-Ni-[Fe_4S_4]$ cluster on which CO (from CODH) and a methyl group from folates are joined to form acetyl-CoA [100–102]. It was originally thought that a variant form of the ACS active site contains a $Cu-Ni-[Fe_4S_4]$ cluster [151, 152], but it was subsequently shown that the Cu-containing cluster represents an inactivated form of ACS [101]. Similarly, it has also been shown that the open form of the Ni-Ni ACS may switch to a closed, inactivated, form by exchanging one of the nickel atoms for a zinc atom [100].

Finally, methyl-group transfer to the ACS active site mediated by the corrinoid iron–sulfur protein (CFeSP) containing the cofactor cobalamin also involves elaborate metal chemistry [103, 104]. In accepting the methyl-group from folates, the cobalt in cobalamin becomes oxidized from the Co^+ to the Co^{3+} state. Donation of the methyl group to the ACS active site restores the Co^+ state, while in turn oxidizing the active nickel in the ACS active site from the Ni^0 to the Ni^{2+} state. Release of acetyl-CoA then reduces the active nickel back to the Ni^0 state, allowing the cycle to start over [104].

Perhaps not surprisingly, all these examples of metal-cluster enzymes concern catalysis not just of the formation of C–C bonds, but of the incorporation of the small gas-phase molecule CO_2. In general, enzymes involved in the processing of small gas-phase molecules (including H_2 and N_2) are among the most unique enzymes in biology—all but one of the known Nickel-containing enzymes belong to this group [153]—always containing highly complex metal centers in their active sites [154–159]. This indicates both the difficulty of controlling the catalysis of these reactions, and the importance of understanding their functions in the context of the emergence of metabolism [150].

3.2.4. Complex network closures: diversity and opportunity created by aldol reactions. The network closures that retain carbon flux and enable autocatalysis in rTCA, DC/4HB, and 3HP/4HB pathways are all topologically rather simple, and are quite similar due to the homology among most of the pathway intermediates. Their module boundaries also are all defined by

acetate and succinate, and at least in the case of acetate, were probably facilitated by its multiple pre-existing roles as the redox-drain of the rTCA cycle [41] and the starting point for both isoprenoid and fatty-acid lipid biosynthesis.

In contrast, the topology of the 3HP-bicycle appears complex, and perhaps an improbable solution to the problem of recycling all carbon flux through core pathways. This form of complexity arises from the requirement to complete an autocatalytic network topology while avoiding reactions based on CO_2 in favor of those based on HCO_3^-. It is thus different from the topological complexity within the bowtie, where dense cross-linking in the core arises at the intersection of many minimal pathways. If we are to argue that the emergence or evolution of network closures such as that in the bicycle is facilitated by a form of modularity, it must exist at the level of reaction mechanisms that render the evolutionary innovation of such topologties plausible. For the 3HP bicycle and the related glyoxylate shunt—and to a lesser degree also for rTCA—the mechanism of interest is the aldol reaction.

The aldol reaction is an internal oxidation-reduction reaction, which means that it exploits residual free energy from organosynthesis, and also that it can take place independently of external electron donors or acceptors. Many aldol reactions are also kinetically *facile*, occurring at appreciable rates without the aid of catalysts. We therefore expect that among compounds capable of participating in them, aldol reactions would have been common in the prebiotic world, providing opportunities for pathway generation. Since their diversity is difficult to suppress except by special mechanisms [160], we expect that potential aldol reactions among metabolites would either have become regulated (perhaps through phosphoryl occupation of hydroxyl groups) or else incorporated into actively-used biochemical pathways.

Aldol reactions are important generators of diversity in organic chemistry, notorious for the very complex network known as the *formose reaction* [161–164], initiated from formaldehyde and glycolaldehyde. Many aldol reactions are possible for sugars, and the reductive pentose phosphate pathway is indeed a network of selected aldol condensations and cleavages among sugar-phosphates [87].

Fewer aldol reactions are possible among intermediates of the rTCA cycle and their homologues such as methyl-malate (MML) or citramalate in other carbon-fixation pathways, but all possibilities are indeed used either in intermediary metabolism or in carbon fixation. The complex topology of the 3HP bicycle therefore also suggests that a diverse inventory of pathway segments were available at the time of its emergence. Figure 7 shows the overlap between the bicycle and the closely-related glyoxylate shunt, which is thus a possible precursor to the bicycle. In both pathways, the network topologies that regenerate all carbon flux or achieve autocatalysis are created by aldol reactions. The retention of carbon within the shunt appears to be a reason for its widespread distribution and frequent use [133, 165, 166], even when energetically more efficient pathways such as the Krebs cycle exist as alternatives within organisms. The re-use of reaction mechanisms on different substrates is a distinct form of simplicity and redundancy that we consider more generally in section 5.

3.2.5. Re-use of the direct C1 reduction pathway and hybrid fixation strategies. A unique form of re-use is found for the sequence of reactions that directly reduce one-carbon (C_1) groups on pterin cofactors. We have argued elsewhere [22] that even when a complete, autotrophic WL pathway is not present due to the loss of the oxygen-sensitive CODH/ACS enzyme, the direct C_1-reduction sequence on pterins is often still present and being used as a partial fixation pathway. The reaction sequence supplies the diverse methyl-group chemistry mediated by S-adenosyl-methionine (SAM), and the direct synthesis of glycine and serine from methylene groups, reductant and ammonia. Serine then serves as a precursor to cysteine and tryptophan. The pathway may exist in either a complete (8-reaction) or a previously unrecognized but potentially widespread (7-reaction) form that involves uptake on N^5 rather than N^{10} of THF [22] (see figure 5.)

The widely distributed and diversified form of direct C_1 reduction functions much as auxiliary catabolic pathways function in mixotrophs [5], operating in parallel to an independent 'primary' fixation pathway, with the primary and the direct-C_1 pathway supplying carbon to different subsets of core metabolites. In many cases where the CODH/ACS is lost, this loss disconnects the primary and direct-C_1 pathway segments, creating the novel feature of a *disjoint* carbon fixation pathway. The existence of parallel fixation pathways in autotrophs had previously been recognized only in one (relatively late-branching) γ-proteobacterium, the uncultured endosymbiont of the deep-sea tube worm *Riftia pachyptila*, which was found to be able to use both the rTCA and CBB cycles [167]. In that case, however, the two pathways are not disjointed, but rather connected through intermediates in the glycolytic/gluconeogenic pathways. In addition, the capacity for using either cycle is thought to reflect an ability to adapt to variation in the availability of environmental energy sources, with an apparent up-regulation of the more efficient rTCA cycle under energy-poor conditions [167]. Our phylogenetic reconstruction [22], however, indicates that parallel disjoint pathways were the majority phenotype in the deep tree of life, in which a reductive C1 sequence to glycine and serine is preserved in combination with rTCA in Aquificales and Nitrospirae, with CBB in Cyanobacteria, with the 3HP bicycle in Chloroflexi (all bacteria), and with DC/4HB in Desulfurococcales and Acetolobales, and the 3HP/4HB cycle in Sulfolobales (all archaea). In contrast, the full WL pathway is found only in a subset of lineages of bacteria (especially acetogenic Firmicutes) and archaea (methanogenic Euryarcheota).

Apparently as a result of the flexibility enabled by parallel carbon inputs to core metabolites, the direct C_1 reduction sequence is more universally distributed than any of the other loop-networks (whether paired with C_1 reduction or used as exclusive fixation pathways), or than the complete WL pathway. The status of the pterin-mediated sequence as a module appears more fundamental than its integration into the full WL pathway, and comparable to the arcs identified within rTCA, which may function as parts of fixation pathways or alternatively as anaplerotic extensions to other pathways. The two types of pathways also serve similar functional roles

in our phylogenetic reconstruction of a root carbon-fixation phenotype, as the key components enabling and selecting the core anabolic precursors.

The reductive synthesis of glycine furnishes a potent reminder of the importance of taking evolutionary context into account when interpreting results from studies of metabolism. The goal of understanding human physiology and disease states has historically been a major driver in the study of biochemistry and metabolism. Although microbial biochemistry is currently better understood (because it is simpler) than human biochemistry [168], model systems and interpretations have continued to emphasize heterotrophy. An example of the interpretive bias that can result is the common reference to the reductive citric acid cycle as the 'reverse' citric acid cycle, despite its likely being the original form as we (and many others) have argued. Similarly, the 'glycine cleavage system' (GCS) was originally studied in rat and chicken livers [169], before being recognized as phylogenetically widespread. The distribution of this system is now known to be nearly universal across the tree of life (with methanogens being the main systematic exception, for reasons explained elsewhere [22]), suggesting that it was present already in the LUCA. The lipoyl-protein based system has long been known to be fully reversible [169–171], and has nearly neutral thermodynamics at physiological conditions [98]. Thus, the LUCA could have used this system either to synthesize or to cleave glycine. From this perspective the former possibility (synthesis) seems a more likely interpretation, even without additional data. From a perspective less strongly focused on heterotrophs, the choice between these alternatives might have become clear much sooner.

3.3. A coarse-graining of carbon-fixation pathways

We can combine all the previous observations on modularity in carbon fixation—the sharing of arcs between loop pathways, the re-use of TCA and reductive C1 sequence to complete the set of anabolic pillars—to perform a 'coarse-graining' of carbon fixation. Combining the decomposition of figure 6 with the gluconeogenic and WL pathways in figure 4, we may list the seven modules from which all known autotrophic carbon-fixation pathways are assembled: (1) direct one-carbon reduction on folates or related compounds, with or without the CODH/ACS terminal reaction of WL; (2) the short-molecule rTCA arc from acetyl-CoA to succinyl-CoA; (3) the long-molecule rTCA arc from succinyl-CoA to citryl-CoA; (4) the gluconeogenic/reductive pentose-phosphate pathway, with or without the RubisCO reaction of CBB; (5) the 3HP arc from acetyl-CoA to succinyl-CoA; (6) the long-molecule 4HB pathway from succinyl-CoA to acetoacetyl-CoA; (7) the glyoxylate-shunt/mesaconate pathway to citramalate, which is the long-molecule loop in the 3HP bicycle. Figure 10 shows the summary of these modules at the pathway level, as well as their different combinations to form complete autotrophic carbon-fixation pathways.

The importance of including glycine in the set of anabolic pillars immediately becomes clear in this coarse-grained view. The general similarity among different carbon-fixation

pathways increases significantly, while finer distinction between forms becomes possible. In particular, both of the pathways that have been most commonly discussed in the context of ancestral carbon fixation and the origin of life, WL and rTCA [7, 33, 77, 122], separate into deep- and late-branching forms. The increased similarity of the deep-branching forms of these pathways suggests an underlying template that combines both WL and rTCA in a fully connected network. WL and rTCA differ from this linked network by single reactions associated either with energy (ATP) economy or oxygen (or perhaps sulfide or other oxidant) sensitivity. Combining information on the synthesis, structural variation, ecology and phylogenetics of the pterin molecules upon which direct C_1 reduction is performed similarly suggests a distinction between the acetogenic (bacterial) and methanogenic (archaeal) forms of WL associated with energy economy [22]. A 'proto-tree' of carbon-fixation emerges from the pooling of these different observations, which in turn makes it possible to reconstruct a complete phylometabolic tree of carbon fixation, as discussed in detail in section 3.4 below.

3.3.1. How the inventory of elementary modules has constrained innovation and evolution. The essential invariance across the biosphere of the seven sub-networks listed above allows us to represent all carbon-fixation phenotypes in terms of the presence or absence, connectivity and direction of these basic modules. In this representation, metabolic innovation at the modular level retains the character of individual discrete events, even if the pathway segments involved incorporate multiple genes. In cases where multiple genes must be acquired to constitute a module, this innovation may take place at higher levels of metabolism, after which their incorporation as fixation pathways appears as a single innovation. For example, most of the reaction sequences used in the autotrophic 4HP pathway appear in diversified forms in fermentative secondary metabolism from hydroxybutyrate or aminobutyrate, which is both outside the rTCA/folate core and ecologically heterotrophic. It is plausible (and we think likely [22]) that these pathways were recruited for autotrophy from an organism similar to *Clostridium kluyveri*. (See also discussions in section 3.4.2).

Because the module boundaries are defined by particular (often universal) molecular species (e.g., acetyl-CoA, succinyl-CoA, and ribulose-1,5-bisphosphate) it often remains true that innovation can be traced to the change in single genes. This is true for the loss of the CODH/ACS from acetyl-CoA phenotypes, the innovation of RubisCO in CBB bacteria, or the loss of substrate-level phosphorylation to acetyl-CoA or succinyl-CoA in acetogens. A case with only slightly greater complexity is the apparently repeated, convergent evolution of an oxidative pathway to form serine from 3-phosphoglycerate (3PG), which involves three common and widely diversified reactions: a dehydrogenation, a reductive transamination and a dephosphorylation. The evolution of this bridge pathway creates a secondary connection between the previously disjoint carbon-fixation pathways described in section 3.2.5. As a full evolutionary reconstruction (described next) shows, such a bridge may permit subsequent loss of direct-C_1 reduction as
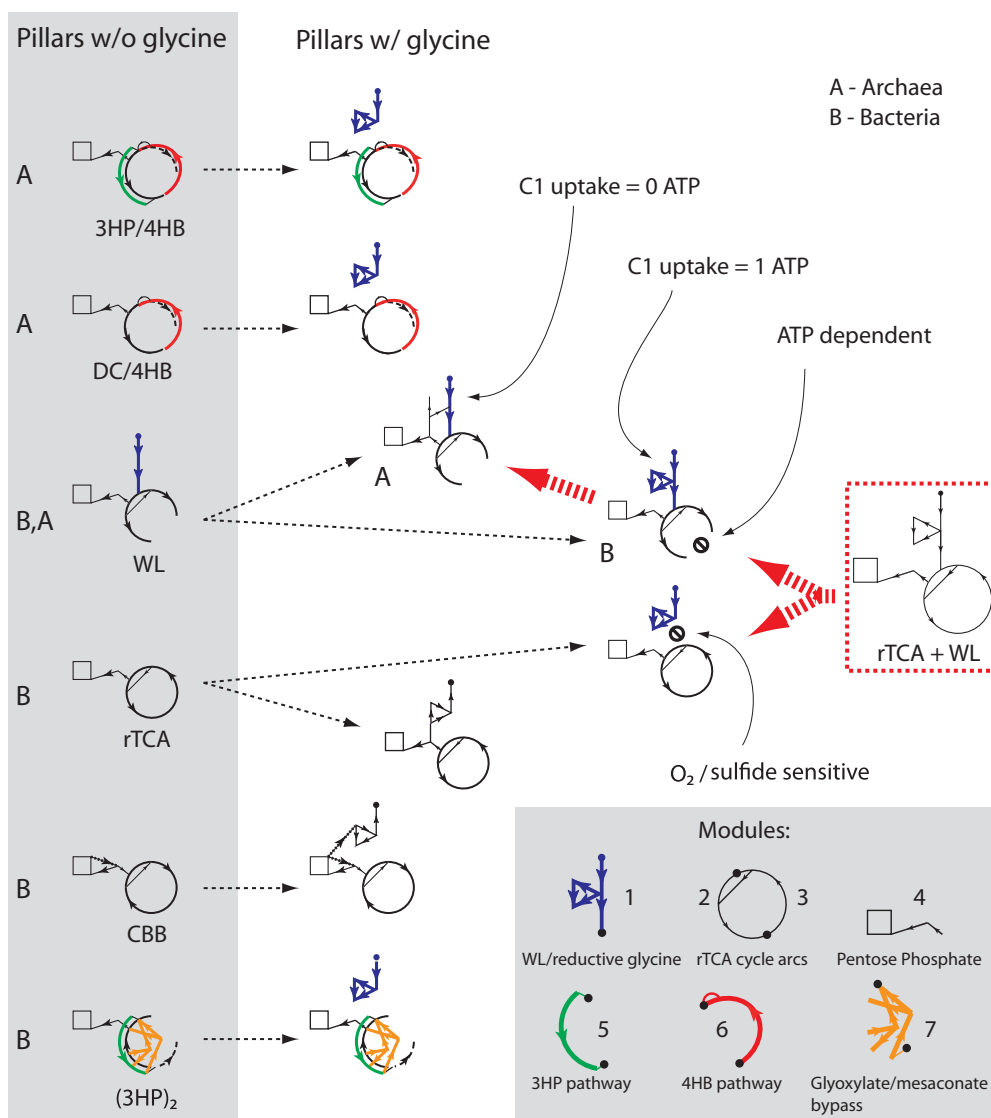
Figure 10. Coarse-grained summary of carbon-fixation pathways. The left panel shows the six pathways as they are known from extensive laboratory characterization. Including glycine along with the anabolic pillars as the molecules that must be reached in carbon fixation then adds resolution, allowing finer distinctions among forms and generally increasing their similarity. As a result, underlying evolutionary templates and patterns begin to emerge. The panel on the bottom right shows the modules from which all carbon-fixation pathways are constructed, as outlined in the main text.

a fixation route, as in the proteobacteria, or it may release a constraint, permitting change in pterin cofactor chemistry as in methanogens.

At the module level, we may represent changes in carbon fixation pathways between closely-related phenotypes in terms of single connections, disconnections or overall changes of direction within the subsets of the seven modules which are present. The change of direction within modules is usually complete, even if it is partial or intermediate at the level of whole pathways. An example is the switch from autotrophic rTCA to fermentative TCA using a reductive small-molecule arc and an oxidative large-molecule arc [68]. Such fermentative pathways may alternate with fully oxidative TCA (Krebs) cycling, and they often occur in organisms that carry homologues to genes for both oxidative and reductive pathway directions [67, 69, 115].

An important exception to this pattern is the partial reversal of the formyl-to-methylene sequence on folates, between its carbon-fixation role and its role in the catabolic cleavage of glycine. We refer in [22] to the module formed by combining the GCS with the methylene-serine transferase as the *glycine cycle*. The combination of the complex free energy landscape provided by the folates [93] with the reversibility and nearly neutral thermodynamics of the glycine cycle [98, 169] permits a high degree of flexibility within this module. Carbon can enter either directly through CO_2, through serine (from 3PG), or through glycine (from glyoxylate), and from any of these sources it may be redirected to all of C_1 chemistry. The topology of the main reaction sequence is preserved in all of the above cases of reversal, though new enzymes or cofactors may be recruited to reverse some reactions.

A representative example of complete module reversal (and in this case, complete cycle reversal) enabled by reductant and cofactor substitution is given by the relation between reductive and oxidative TCA cycles. The electron donor in rTCA, reduced ferredoxin, is replaced by lipoic acid as an electron acceptor in the Krebs cycle, in the TPP-dependent oxidoreductase reaction. The enzymes catalyzing the retro-aldol cleavage of rTCA, which have undergone considerable re-arrangement even within the reductive world [112, 113], are further replaced by the oxidative citryl-CoA synthetase. Finally, the change from fumarate reduction to succinate oxidation may require a substitution of membrane quinones [114]. Yet the underlying carbon skeletons over the whole pathway are completely retained, and apart from some details of reaction ordering for thioesters, and possibly the use of phospho-enol intermediates, the energetic side groups are also the same.

3.4. Reconstructed evolutionary history

3.4.1. Phylogeny suggests little historical contingency of deep evolution within the modular constraints. The small number of modules that contribute to carbon fixation, and the even smaller number of 'gateway' molecules that serve as interfaces between most of them, permit free recombination into many phenotypes satisfying the constraints of autotrophy. An important consequence of free recombination is that a constraint of overall autotrophy only enforces network completeness—the existence of *some* connection between gateway molecules. Because there exist multiple modules that can be used to satisfy these constraints, autotrophy alone therefore does not lock in dependences within networks over distances longer than the modules themselves. Homology across intra-modular reaction sequences—especially if it is due to catalytic promiscuity—further weakens any lock-in effect created by selection for metabolic completeness. Through these mechanisms modularity promotes innovation-sharing [172] and rapid and reliable adaptation [18] to environmental conditions, but reduces standing variation among individuals sharing a common environment.

However, despite the potential for free recombination in principle, distinct carbon-fixation pathway modules have very different couplings to the chemical environment, as we reviewed in section 3. The genome distributions reported in [22] show that they also have very uneven phylogenetic distribution. For example, TCA arcs and intermediates, as well as direct C_1-reduction, are nearly universally distributed, while the 3HP arcs are restricted to specific bacterial or archaeal clades living in alkaline environments. Finally, we note that not all module combinations consistent with autotrophy have been observed in extant organisms.

By combining these observations it is possible to arrange autotrophic phenotypes on a graph according to their degree of similarity, and to assign environmental factors as correlates of phenotypic changes over most links. The graph projects onto a tree with very high parsimony and therefore requires invoking almost no horizontal gene transfer or convergent evolution from distinct lineages. Instead, all divergences may

be interpreted as independent simple innovations driven by environmental factors. Finally, the directionality of these links (divergences) and the overlap of the tree with bacterial and archaeal phylogeny motivates a natural choice of root. The lack of reticulation in a tree of innovations in autotrophy—at first surprising when compared to highly-reticulated gene phylogenies [73] covering the same period—becomes sensible as a record of invasion and adaptation to new chemical environments by organisms capable of maintaining little long-standing variation.

3.4.2. A parsimony tree for autotrophic metabolism, and causation on links. The tree of autotrophic carbon-fixation phenotypes from [22] is shown in figure 11. All nodes in the tree satisfy the constraint that all five universal anabolic precursors plus glycine can be synthesized directly from CO_2. We have defined parsimony by requiring single changes over links at the level of pathway modules, as explained above, rather than at the level of single genes, in cases where the two criteria differ. (This definition separates the evolution of genetic backgrounds, such as 4-hydroxybutyrate fermentation, from the events at which organisms came to rely on complete pathways for autotrophy.)

A complete-parsimony tree for the known phenotypes is not possible, so we chose a tree in which the only violations are duplicate innovation of serine synthesis from 3-phosphglycerate (3-PG), and duplication or transfer of the short-molecule 3HP pathway. The synthesis of serine from 3-PG involves reactions—the dehydrogenation of an alcohol to a carbonyl, the transamination of this carbonyl to an amino group and a dephosphorylation—that are common throughout metabolism and are performed by highly diversified enzyme families. We have therefore regarded multiple occurrences of this event as not attaching a large probability penalty to parsimony violation. We make a similar judgment for the 3HP pathway. This pathway contains two key biotin-carboxylase enzymes, one of which (acetyl-CoA carboxylase) is also part of fatty acid synthesis, which is suggested to have been present already in the LUCA [173]. Sequence analysis of propionyl-CoA carboxylase has in turn been used to suggest convergent evolution as an explanation for the multiple occurrence of this enzyme across bacterial and archaeal domains [127]. The remaining reactions in this pathway are again common metabolic reactions performed by highly diversified enzyme families. An alternative hypothesis is transfer: this complete pathway occurs in environments shared by the bacteria and archaea that harbor it, and this environment also contains a stressor (alkalinity) that may induce gene transfer [174]. Thus, both gene transfer and convergence are plausible explanations for why this phenotype should be paraphyletic.

Any tree in which either of these phenotypes was made monophyletic would require more extensive parsimony violations than the tree we chose, involving innovations for which convergence or transfer are also less plausible. Such trees would require major sub-branches to contain both bacterial and archaeal members, and within these, repeated divergences of major domain-specific differences. Common descent would fail to account for the exclusivity of rTCA-based
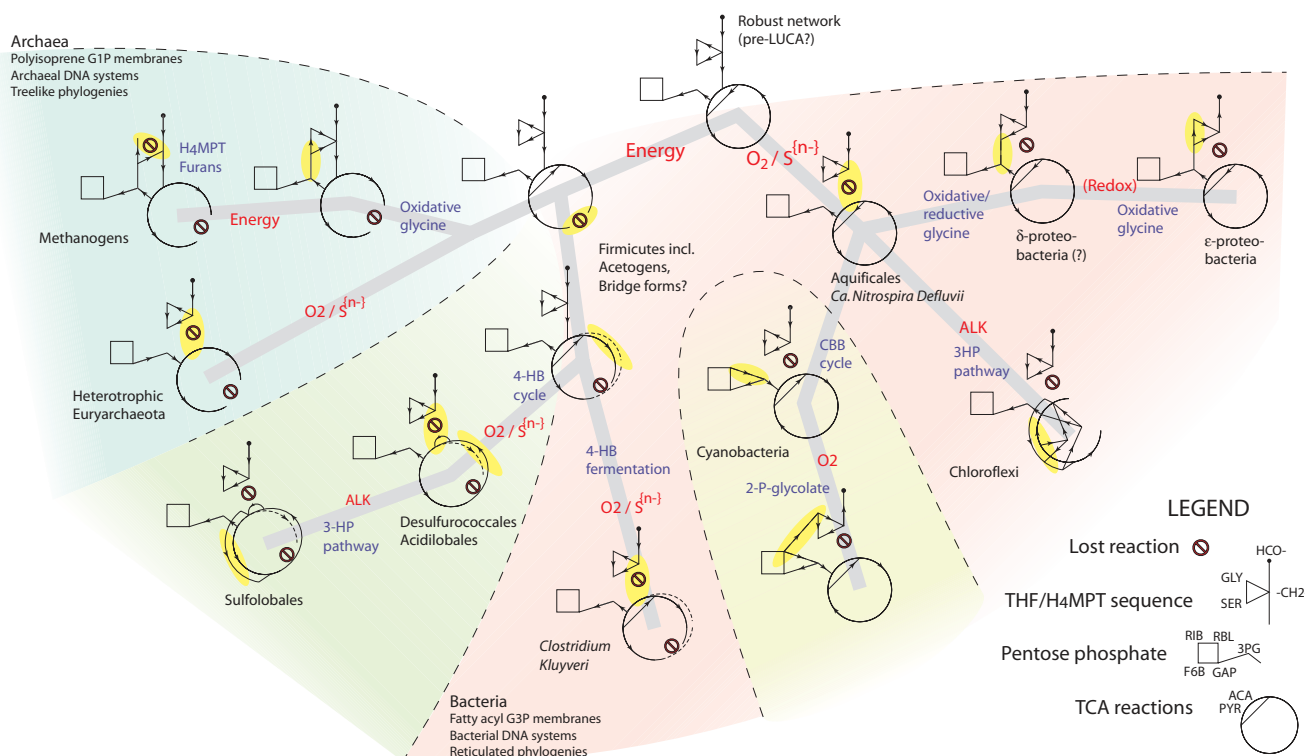
Figure 11. A parsimony-based reconstruction of the innovations linking the major carbon-fixation phenotypes, from [22]. Nodes in the tree are autotrophic phenotypes, following the coarse-grained notation introduced in figure 10, and summarized in the legend. Grey links are transitions in the maximum-parsimony phylometabolic reconstruction, and yellow-highlighted regions in the diagrams are innovations following each link. Organism names or clades in which these phenotypes are found are given in black; fixation pathways innovated along each link are shown in blue, and imputed evolutionary causes are shown in red. $S^{\{n-\}}$ refers to sulfides of different oxidation states. Dashed lines separate regions in which the clades by phylometabolic parsimony follow standard phylogenetic divisions. Abbreviations: formyl (HCO−); methylene (−CH$_2$−); acetyl-CoA (ACA); pyruvate (PYR); serine (SER); 3-phosphoglycerate (3PG); glyceraldehyde-3-phosphate (GAP); fructose-1,6-bisphosphate (F6B); ribose-phosphate (RIB); ribulose-phosphate (RBL); akalinity (ALK). Arrows indicate reaction directions; dashed line connecting 3PG to SER indicate intermittent or bi-directional reactions.

phenotypes within bacteria, and among archaea either (non-THF) pterin-based one-carbon chemistry, or isoprene-related hydroxybutarate reactions would be required to have arisen several times. It is corroborating evidence for this parsimony argument that the tree we propose preserves the monophyly of bacteria and archaea, and is consistent with the most robust signals in purely statistical gene phylogenies, including the greatest congruence of firmicutes with the archaea, among the bacterial branches [175, 176].

The nodes in the tree of figure 11 are all phenotypes of extant organisms, with one important exception, which is the node between the Aquificale branch and the Firmicute/Archaea branch. Aquificales and all phenotypes descending from them lack the CODH/ACS enzyme, while firmicutes and archaea lack one or more ATP-dependent acyl-CoA (citryl-CoA or succinyl-COA) synthases. Therefore, if we seek a connected tree of life, two changes—the gain of one enzyme and loss of the other—are required to connect these branches. Since any organism lacking both enzymes could not fix carbon autotrophically, we have chosen the order of gain and loss so that the intermediate node has *both* the CODH/ACS and the acyl-CoA synthases. It therefore has both a complete WL pathway and an autocatalytic rTCA loop, connected through their shared intermediate acetyl-CoA. Losses (but not re-acquisitions) of either of these enzymes occur at multiple

points on the tree, and both have likely explanations in either environmental chemistry or energetics. For this reason and several others given below, although a parsimony tree is (*a priori*) unrooted, we will regard the joint WL/rTCA phenotype as not only a bridging node but the root of the tree of autotrophs.

There is one further, unproblematic exception to the assignment of extant phenotypes to nodes in the tree, which is the insertion of an acetogenic phenotype with a facultative oxidative pathway to serine at the root of the Euryarchaeota. Since methanogens use this pathway, and since an acetogenic pathway lacking oxidative serine synthesis is the most plausible ancestral form for all archaea as well as for Firmicutes within the bacteria, we infer that such an intermediate state did or does exist. This fixation pathway is consistent with forms observed in extant organisms, and these proposals would be supported if such a phenotype were to be discovered or to result from reclassification of genes in an existing organism.

In the evolution of carbon fixation from a joint WL/rTCA root, the primary division is between the loss of the CODH/ACS, resulting in rTCA loop-fixation phenotypes, and the loss of the acetyl-CoA or succinyl-CoA synthetases, resulting in acetogenic phenotypes. Very low levels of oxygen permanently inactivate the CODH/ACS, so its loss is probable

even under microaerobic conditions. Although the dominant mineral buffers for oxygen in the Archaean remain a topic of significant uncertainty [117, 177–179], it appears unlikely that molecular oxygen was the toxin responsible for loss of the CODH/ACS much before the 'great oxidation event' (GOE)[5]. Therefore the sensitivity of the CODH/ACS to sulfides or perhaps other oxidants [309] remains a possibly important factor in the early divergences of carbon fixation.

Alternatively, among strict WL-anaerobes, the loss of citryl-CoA or succinyl-CoA synthetase saves one ATP per carbon fixed, and all acetogenic phenotypes break rTCA cycling only through the loss of one or the other of these enzymes. We therefore interpret the loss of rTCA cycling as a result of selection for energy efficiency. The failure to regain either of these enzymes by acetogens which subsequently also lost the CODH/ACS is perhaps surprising given the inferred homology of the ancestral citryl-CoA and succinyl-CoA synthetases [112, 113], but explains the absence of rTCA cycling in either Firmicutes or any Archaea.

The remaining autotrophic phenotypes are derived from either rTCA cycling or acetogenesis in natural stages due to plausible environmental factors. Oxidative serine synthesis (from 3PG) is associated with the rise of the Proteobacteria, whose differentiation in many features tracks the rise of oxygen and the transition to oxidizing rather than reducing environments. RubisCO and subsequently photorespiration arise within the Cyanobacteria. The innovation of the 3HP bicycle from the malonate pathway arises within the Chloroflexi. In the Firmicutes (bacteria), 4-hydroxybutyrate (or closely related 4-aminobutyrate) fermentation is more or less developed. Closure of the fermentative arcs to form a ring, again driven by elimination of the CODH/ACS [22] leads to the DC/4HB pathway in Crenarchaeota, which is then specialized in the Sulfolobales to the alkaline 3HP/4HB pathway. The Euryarchaeota are distinguished by the absence of an alternative loop-fixation pathway to rTCA, so that all members are either methanogens or heterotrophs.

Similarly, the innovation of the 3HP pathways, using biotin, emerges as a specialization to invade extreme but relatively rare environments. A particularly interesting case is the modification of folates in archaea, leading from THF in ancestral nodes to tetrahydromethanopterin in the methanogens, which enables initial fixation of formate (formed by hydrogenation of CO_2) in an ATP-free system [22, 93]. The root position of rTCA explains the preservation of rTCA arcs both in reductive acetyl-CoA pathways, and in anaplerotic appendages to other fixation pathways, and the root position of direct C_1 reduction explains its near-universal distribution.

3.4.3. Parsimony violation and the role of ecological interactions. A tree is by construction a summary statistic for the relations among the phenotypes which are its leaves

or internal nodes. It is not inherently a map of species descent, and takes on that interpretation only when common ancestry is shown to explain the conditional independence of branches given their (topological) parent nodes. This caution is especially important for the interpretation of figure 11, which shows high parsimony in the deepest branches where horizontal gene transfer is generally believed to have been most intense [78, 79]. We have argued that this behavior is consistent in a tree of successive optimal adaptations to varied environments, by organisms that could maintain little persistent variation. Violations of parsimony that are improbable by evolutionary convergence contain information about contact among historically separated lineages. Under this interpretation the separation is primarily environmental, with the subsequent contact identifying ecological co-habitation. As explained above, the possible transfer of genes for the 3HP pathway is especially plausible, as the organisms involved may have shared the same extreme (alkaline) environments and been under common selection pressure, which when severe is known to accelerate rates of gene transfer [174, 187].

Our methods in [22] include flux-balance analysis of core networks, where the boundaries of analysis are defined to be carbon input solely from CO_2 and the output of the universal precursors we have listed as the interface between carbon fixation and anabolism, as shown in figure 1. We do not model cellular-level mechanisms of either regulation or heredity, nor full downstream intermediary metabolism. Our system of metabolic flux constraints therefore does not distinguish between individual species and ecosystems. It does not, of course, *exclude* the possibility of representing individual organisms. The general agreement with robust phylogenetic signatures from many different genomic phylogenies [73, 175, 188] may thus still suggest a dominant role for vertical descent among autotrophic *organisms* (and not merely consortia) in the early evolution of carbon fixation.

3.4.4. A non-modern but plausible form of redundancy in the root node. The joint WL/rTCA network was introduced into figure 11 to produce a connected tree containing only autotrophic nodes. It also gives the most parsimonious interpretation of the nearly universal distributions of both reductive C_1 chemistry and of citric acid cycle components, and receives further circumstantial support from the identifiability of both plausible and specific environmental driving forces for most subsequent branches. The constraints which jointly required the insertion of a linked WL/rTCA network at the root have led us to propose a kind of redundancy not found in extant fixation pathways. Either WL or rTCA alone is self-maintaining (in a modern organism) so a network that incorporates both is *redundantly* autocatalytic. This is an important and speculative departure from known phenotypes, but it can be argued to have conferred a selective advantage under the more primitive conditions of early cells, because the pathway topology itself possesses a form of inherent *robustness*. The redundant network topology of the root phenotype would have allowed it to better cope with both internal and external perturbation in an era when
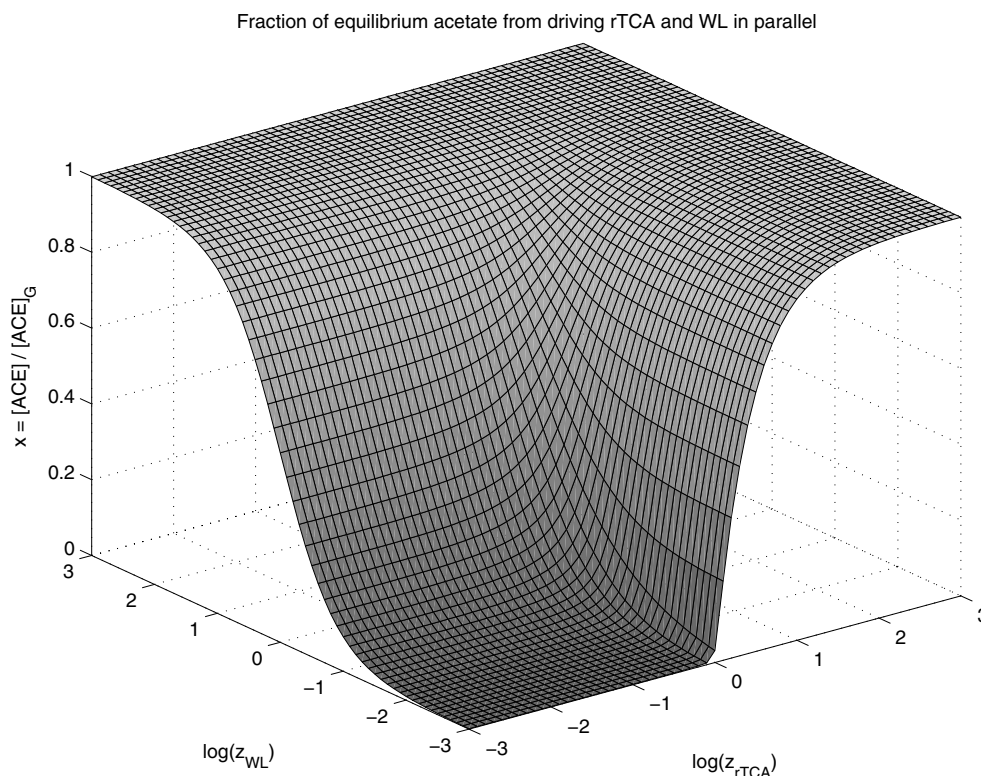
---

[5] The GOE is usually dated at 2.5 GYA, which may be relevant dates to compare to genetically estimated loss events in later branches of the Archaea or possibly in the Clostridia, but they are unlikely as dates for the first branching in the tree of figure 11. Arguments have been made for low levels of oxygen preceding the GOE by as much as 50–100 million years [180], as well as a transient rise in oxygen as far back as 2.9 GYA [181, 182], but both are actively being debated [183–186].

Fraction of equilibrium acetate from driving rTCA and WL in parallel

Figure 12. Network dynamics of connected WL and rTCA pathways. Graph of solutions to equation (B.44) from appendix B is shown versus base-10 logarithms of z_{rTCA} and z_{WL}. The quantity x on the z-axis is the fraction of the acetate concentration [ACE] relative to the value it would take in an equilibrium ensemble with carbon dioxide, reductant and water. The value $x = 1$ corresponds to an asymptotically zero impedance of the chemical network, compared to the rate of environmental drain. The parameter z_{rTCA} is a monotone function of the non-equilibrium driving chemical potential to synthesize acetate and z_{WL} measures the conductance of the 'feeder' WL pathway. At $z_{WL} \to 0$, the WL pathway contributes nothing, and the rTCA network has a sharp catalytic threshold at $z_{rTCA} = 1$. For nonzero z_{WL}, the transition is smoothed, so some excess population of rTCA intermediates occurs at any driving chemical potential.

regulation and kinetic control were probably less sophisticated and refined than they are today. In that respect it is a more plausible phenotype for a universal ancestor than any modern network.

The enhanced robustness of the joint network follows from the interaction of short-loop and long-loop autocatalysis. The threshold for autocatalysis in the rTCA loop, fragile against parasitic side reactions or unconstrained anabolism, is supported and given a recovery mode when fed by an independent supply of acetyl-CoA from WL. In turn, the production of a sufficient concentration of folates to support direct C_1 reduction, fragile if the long biosynthetic pathway is unreliable, is augmented by additional carbon fixed in rTCA. These arguments are topological, and do not make specific reference to whether the catalysts for the underlying reactions are enzymes. They may provide context for (perhaps multi-stage) models of transition from primordial mineral catalysis [74, 189] to the eventual support of carbon fixation by biomolecules.

Figure 12 shows a numerical solution for the flux through a minimal version of the joint WL/rTCA network, with lumped-parameter representations of parasitic side reactions and the net free energy of formation of acetate. (The exact rate equations used, and their interpretation, are developed in appendix B.) In the absence of a WL 'feeder' pathway, rTCA has a sharp threshold for the maintenance of flux through the network as a

function of the free energy of formation of its output acetate. The existence of such a sharp threshold depending on the rate of parasitism, below which the cycle supports *no* transport, has been one of the major sources of criticism of network-autocatalytic pathways as models for proto-metabolism [190]. When WL is added as a feeder, however, the threshold disappears, and some nonzero flux passes through the pathway at any negative free energy of formation of the outputs. The existence of a pathway that supports some organosynthetic flux at any positive driving chemical potential—dependent on external catalysts but not contingent on the pathway's own internal state—has been one of the major reasons WL has been favored as a protometabolic pathway by molecular biologists and chemists [33, 74, 77].

The reason (beyond evidence from reconstructions) that we regard a linear pathway modeled on the modern acetyl-CoA pathways as an incomplete answer to the needs of incipient metabolism is that it offers an avenue for *production* of organics, but does not by itself offer a chemical mechanism for the kind of *selection* and *concentration* of fluxes that is equally central to the sparse network of extant core metabolism [191]. Chemical self-amplification, if it can be demonstrated experimentally, is the most plausible mechanism by which the biosphere can concentrate all energy flows and material cycles through a small, stable set of organic compounds. It supplies the molecules that are within the loop—and secondarily

those that are made from loop intermediates—above the concentrations they would have in a Gibbs equilibrium distribution, as a result of flow through the network. The fact that self-amplification is permitted to act in the model of figure 12, even below the chemical-potential difference where the rTCA loop alone is self-sustaining, provides a mechanism by which the loop intermediates could have been supplied in excess in the earliest stages of the emergence of metabolism. We return in section 6 to a related form of robustness and selection, which applies as anabolic pathways begin to form from loop intermediates.

A surprising observation suggested by our reconstructed history is how conservative the biosphere has been in its intermediate stages of innovation, as a consequence of geochemical niche diversity on earth. Except for the root node, we have not needed to invoke extinct ancestral forms to explain extant diversity, an argument that even Darwin [192] expected to be required frequently for cases where modern, optimized forms outcompeted their more primitive ancestors and erased direct evidence about the past. The one case where we do invoke an essential extinct ancestral form is the root node, and its character suggests reasons why it should have become extinct that are more chemically basic and biochemically consequential than the secondary physiological or ecological distinctions that modern evolutionists use to explain extinct ancestral forms. The topological robustness of the root node comes at the combined costs of sub-optimal energy efficiency and oxidant sensitivity. The fitness advantage to shedding either of these costs would have increased significantly as organisms obtained more sophisticated macromolecular components and correspondingly greater control over their internal chemistry, lessening (and ultimately removing) the selective advantage of a redundant carbon-fixation strategy even in the absence of external biogeochemical perturbations.

Without the ability to culture and analyze a population of LUCA organisms, the amount we can conclude from mathematical analyses of general network properties is of course less, but it is still within the range commonly used to assess proposals for early metabolism. For example: proposals for autocatalysis in geochemical networks with crude catalysts are routinely criticized on the basis of their *shared topological feature* of feedback and its associated threshold fragility [190]. These criticisms emphasize that parasitic side-reactions are a likely problem, although the corollary that autocatalysis is generally ruled out requires the strong claim that side-reactions are a problem in *all* plausible environments, which extends well beyond current experimental knowledge. In the opposite direction (and in this case based on particular and well-understood side-reactions), it is argued for the formose network [160] that without some severe pruning mechanism, the reactions are *too productive*, creating mixtures too complicated to be relevant to biochemistry. General arguments of both kinds contribute to a negative hypothesis behind the hope [193] that catalytic RNA will be a sufficient solution to both problems of productivity and selection. But the larger points in both arguments are important, and should be applied at many other places in hypotheses about the emergence and early elaboration

of metabolism. They emphasize that both robustness *and* selectivity are needed features of any mechanism responsible for the earliest cellular organosynthesis. At the same level of generality as they are raised as criticisms, these criteria create meaningful distinctions between topologies for early carbon-fixation pathways, to which the lumped-parameter model of our root node gives a quantative form.

3.5. The rise of oxygen, and changes in the evolutionary dynamics of core metabolism

The limits of the phylometabolic tree we show in figure 11 fall on a horizon that coincides with the rise of oxygen. More precisely: we do not show branches that phylogenetically trace lineage divisions later than this horizon, because no known divisions in carbon fixation distinguish such later branches. Many of the late branches contain only heterotrophs, and to the extent that post-oxygen lineage divisions follow divisions in metabolism, they are divisions in forms of heterotrophy. The rise of oxygen seems to have put an end to innovation in carbon fixation, and led to a florescence of innovation in carbon sharing. By 'sharing' we refer to general exchanges in which organic compounds are re-used without *de novo* synthesis; we do not intend only symbiotic associations. At the level of aggregate-ecosystem net primary production, the exchange of organics with incomplete catabolism may, however, reduce the free energy cost of the *de novo* synthesis of biomass that supports a given level of phenotypic diversity or specialization, allowing ecologies of complementary specialists to partially displace ecologies of generalist autotrophs.

On the same horizon, the high parsimony of the tree we have shown ends, and it becomes necessary to explain complex metabolisms as a consequence of transfer of metabolic modules among clades in which they had evolved separately. We no longer expect that it would be possible to explain—and to some extent to predict—these innovations given only constraints of chemistry and invasion of new geochemical environments. Instead, they rely chemically on ecologically determined carbon flows, and genetically on opportunities for transfer of genes or pathway segments. Therefore any explanation will require some explicit model of ecological dynamics, and may require invoking some accidents of historical contingency. This contrast of phylometabolic reconstructions, between later and earlier periods, illustrates our association of parsimony violation with the role of ecosystems and explicit contributions of multilevel dynamics to evolution.

It is perhaps counterintuitive, but we believe consistent, that the phylometabolic tree is more tree-like in the earlier era of more extensive single-gene lateral transfers, and becomes *less* tree-like and more reticulated, in the era of complex ecosystems enabled by oxygenic metabolisms, which may have come as much as 1.5 billion years later. For reticulation to appear in a tree of reconstructed metabolisms, it is necessary that variants which evolved independently—as we have argued, under distinct selection pressures—be maintained in new environments where they can be brought into both

contact and interdependence. The maintenance of standing variation is facilitated both by the evolution of more advanced mechanisms to integrate genomes and limit horizontal transfer, and by the greater power density of oxygenic metabolisms.

The *serine cycle* used by some methylotrophic proteobacteria, shown in figure 13, provides an example of the structure and complex inheritance of a post-oxygen, heterotrophic pathway. Methylotrophs possess both an H_4MPT system transferred from methanogenic archaea [194, 195], and a conserved THF system ancestral to the proteobacteria (and we argue, to the universal common ancestor). In methylotrophs, H_4MPT is primarily used for the oxidation of formaldehyde to formate, while THF can be used in both the oxidative direction as part of the demethylation of various reduced one-carbon compounds and in the reduction of formate. C_1 compounds are then assimilated either as CO_2 in the CBB cycle, as methylene groups and CO_2 in the serine cycle or as formaldehyde in the ribulose monophosphate (RuMP) cycle, in which formaldehyde is attached to ribulose-5-phosphate to produce fructose-6-phosphate [196, 197].

The full substrate network of the most complex assimilatory pathway of methylotrophy is a bicycle in which the serine cycle is coupled to the *glyoxylate regeneration cycle*. This full network employs segments of all four loop-autotrophic pathways, as well as reactions in glycolysis, and part of the 'glycine cycle'. Carbon enters the pathway at several points. Methylene groups enter through the glycine cycle, combining with glycine to form serine. Serine is then deaminated and reduced to pyruvate, which is combined with a CO_2 in a carboxylation to enter the core of TCA reactions. TCA arcs are performed reductively from pyruvate to MAL, and oxidatively from succinate to MAL, following the pattern of the 3HP pathway plus anaplerotic reactions from its output pyruvate. The short-molecule arc of 3HP is run as in the autotrophic carbon-fixation pathway starting from propionate, but part of the long-molecule arc of 3HP is reversed in the glyoxylate regeneration cycle. The 4HB pathway arc, transferred from archaea, is also reversed to feed this glyoxylate cycle, and is followed by a final additional carboxylation unique to this pathway [198, 199].

The serine/glyoxylate cycle of methylotrophy is a remarkable 'Frankenstein's monster' of metabolism, stitched together from parts of all pre-existing pathways, but requiring almost nothing new in its own local chemistry. Notably, the modules in this bacterial pathway which have been inherited from archaea are all reversed from the archaeal direction.

3.6. Summary: catalytic control as a central source of modularity in metabolism

Focusing on the metabolic foundation of the biosphere— carbon fixation and its interface with anabolism—we have seen many examples of ways in which catalytic control is a central organizing principle in metabolism. The most complex and conserved reaction mechanisms in carbon fixation often have unique and very elaborate metal centers and cofactors associated with them, reflecting the difficulty (or at least

unique requirements) of the catalytic problem being solved. Not surprisingly, these reactions form the boundaries at which the various modules making up carbon fixation are connected. Carbon fixation is the precursor to all biosynthesis, and in the context of a fan/bowtie network where it is part of the core, it is therefore also a strong constraint. Finally, under comparative analysis we find the maximum-parsimony assignment of innovation events in the compositional structure of the network of possibilities to coincide closely with robust signals from genomic phylogenies. We interpret the convergence of these diverse observations to mean that innovations in carbon fixation were at least a large factor in the major early evolutionary divergences of bacteria and archaea. The preservation of this evolutionary signal over very long periods and the very small diversity of pathway innovations suggests that they have also been some of the strongest long-term constraints on evolution. The fixation-module boundaries act as 'turnstiles' along which the flow of carbon into the biosphere is redirected upon biogeochemical perturbations, and they are preconditions for higher-level diversification.

The catalytic control of classes of organic reactions also leads to a secondary source of modularity, the locking in of various core pathways by the elaboration of downstream intermediary metabolism. The most striking example of lock-in is the origination of all anabolic pathways in only a very small number of molecules, mostly within the TCA cycle, even when different carbon-fixation strategies are used. The suggested interpretation is that much of intermediary metabolism had elaborated prior to the divergences in carbon fixation. A related, but slightly different form of lock-in is found in the construction of methylotrophic pathways, which circumvents innovations in the catalytic control of difficult chemistry by re-using a wide range of parts from pre-existing carbon-fixation pathways.

4. Cofactors, and the emergence and centralization of metabolic control

Cofactors form a unique and essential class of components within biochemistry, both as individual molecules and as a distinctive level in the control over metabolism. In synthesis and structure they tend to be among the most complex of the metabolites, and unlike amino acids, nucleotides, sugars and lipids, they are not primary structural elements of the macromolecular components of cells. Instead, cofactors provide a limited but essential inventory of functions, which are used widely and in a variety of macromolecular contexts. As a result they often have the highest connectivity (forming topological 'hubs') within metabolic networks, and are required in conjunction with key inputs or enzymes [200–202] to complete the most elaborate metabolisms.

Cofactor chemistry is in its own right an essential component of the logic underlying metabolic architecture and evolution. We argued in section 2.2 that part of the structure of the small-molecule substrate network is explained by reaction mechanisms and autocatalysis in short-loop pathways, which may once have been supported by external mineral catalysts. At least since the first cells, however, all such pathways

Figure 13. The serine cycle/glyoxylate-regeneration cycle of methylotrophy. Left panel shows the stoichiometric pathway overlaid on the autotrophic loop pathways from figure 6. Right panel gives a projection of the serine cycle and glyoxylate regeneration cycle showing pathway directions; overlaps with the predecessor autotrophic pathways are labeled. Abbreviations: hydroxy pyruvate (HPY); ethyl-malonate (EMA); methyl-succinate (MSU), others as in figure 4.

have been realized only with the essential participation of intermediates from the hierarchically and functionally distinct cofactor class, which add a second layer of network-catalytic feedback. The more structurally complex cofactors tend to be associated with more catalytically complex functions within carbon fixation. Their long synthetic pathways result in long feedback loops, creating new needs for pathway stabilization and control. Because cofactors often mediate kinetic bottlenecks in metabolism, their inventory of functions may constrain the evolutionary possibilities for new pathways, so innovations in cofactor synthesis can have dramatic consequences for the large-scale structure of evolution.

As we note below, cofactors are among the less well-understood components of metabolism. Our ability to decompose cofactor functions and reconstruct the likely history of their elaboration is therefore less comprehensive than the analysis we have given of the small-molecule substrate. However, many functions that divide the cofactors into groups, which seem also to have been responsible for cases of convergent evolution and have perhaps stabilized major functional categories, relate directly to properties of particular chemical elements. Others are molecular properties shared as a consequence of derivation from a common precursor. In this section we select aspects of cofactor chemistry that seem to us most essential to overall metabolic architecture and evolution, with the goal of framing as much as of answering questions. As our understanding of cofactor chemistry improves through laboratory studies, so will our ability to integrate the observations in this section into a more complete theory of metabolic architecture and evolution.

4.1. Introduction to cofactors as a group, and why they define an essential layer in the control of metabolism

4.1.1. Cofactors as a class in extant biochemistry. The biosynthesis of cofactors involves some of the most elaborate and least understood organic chemistry used by organisms. The pathways leading to several major cofactors have only very recently been elucidated or remain to be fully described, and their study continues to lead to the discovery of novel reaction mechanisms and enzymes that are unique to cofactor synthesis [203–205]. While cofactor biosynthetic pathways often branch from core metabolic pathways, their novel reactions may produce special bonds and molecular structures not found elsewhere in metabolism. These novel bonds and structures are generally central in their catalytic functions.

Structurally, many cofactors form a class in transition between the core metabolites and the oligomers. They contain some of the largest directly-assembled organic monomers (pterins, flavins, thiamin, tetrapyrroles), but many also show the beginnings of polymerization of standard amino acids, lipids or ribonucleotides. These may be joined by the same phosphate ester bonds that link RNA oligomers or aminoacyl-tRNA, or they may use distinctive bonds (e.g. 5′–5′ esters) found only in the cofactor class [206].

The polymerization exhibited within cofactors is distinguished from that of oligomers by its heterogeneity. Srinivasan and Morowitz [51] have termed cofactors 'chimeromers', because they often include monomeric components from several molecule classes. Examples are CoA, which includes several peptide units and an ATP; folates, which join a pterin moiety to para-aminobenzoic acid (PABA); quinones, which join a PABA derivative to an isoprene lipid

tail; and a variety of cofactors assembled on phosphoribosyl-pyrophosphate (PRPP) to which RNA 'handles' are esterified.

We may understand the border between small and large molecules, where most cofactors are found, as more fundamentally a border between the use of heterogeneous organic chemistry to encode biological information in covalent structures, and the transition to homogeneous phosphate chemistry, with information carried in sequences or higher-order non-covalent structures. The chemistry of the metabolic substrate is mostly the chemistry of organic reactions. Phosphates and thioesters may appear in intermediates, but their role generally is to provide energy for leaving groups, enabling formation of the main structural bonds among C, N, O and H. One of the striking characteristic scales in metabolism is that its organic reactions, the near-universal mode of construction for molecules of 20 to 30 carbons or less, cease to be used in the synthesis of larger molecules. Even siderophores, among the most complex of widely-used organic compounds, are often elaborations of functional centers that are small core metabolites, such as CIT [207, 208]. Large oligomeric macromolecules are almost entirely synthesized using the dehydration potential of phosphates [209] to link monomers drawn from the inventory [50] of small core metabolites. Many cofactors have structure of both kinds, and they are the smallest molecules that as a class commonly use phosphate esters as permanent structural elements [210].

Finally, cofactors are distinguished by structure–function relations determined mostly at the single-molecule scale. The monomers that are incorporated into macromolecules are often distinguished by general properties, and only take on more specific functional roles that depend strongly on location and context [211, 212]. In contrast, the functions of cofactors are specific, often finely tuned by evolution [93], and deployable in a wide range of macromolecular contexts. Usually they are carriers or transfer agents of functional groups or reductants in intermediary metabolism [213]. Nearly half of enzymes require cofactors as coenzymes [210, 213]. If we extend this grouping to include chelated metals [214, 215] and clusters, ranging from common iron–sulfur centers to the elaborate metal centers of gas-handling enzymes [104, 150], more than half of enzymes require coenzymes or metals in the active site.

The universal reactions of intermediary metabolism depend on only about 30 cofactors [213] (though this number depends on the specific definition used). Major functional roles include (1) transition-metal-mediated redox reactions (heme, cobalamin, the nickel tetrapyrrole F_{430}, chlorophylls[6]), (2) transport of one-carbon groups that range in redox state from oxidized (biotin for carboxyl groups, methanofurans for formyl groups) to reduced (lipoic acid for methylene groups, SAM, coenzyme-M and cobalamin for methyl groups), with some cofactors spanning this range and mediating interconversion of oxidation states

(the folate family interconverting formyl to methyl groups), (3) transport of amino groups (pyridoxal phosphate, glutamate, glutamine), (4) reductants (nicotinamide cofactors, flavins, deazaflavins, lipoic acid and coenzyme-B), (5) membrane electron transport and temporary storage (quinones), (6) transport of more complex units such as acyl and amino-acyl groups (pantetheine in CoA and in the acyl-carrier protein (ACP), lipoic acid, thiamine pyrophosphate), (7) transport of dehydration potential from phosphate esters (nucleoside di- and tri-phosphates), and (8) sources of thioester bonds for substrate-level phosphorylation and other reactions (pantetheine in CoA).

4.1.2. Roles as controllers, and consequences for the emergence and early evolution of life. Cofactors fill roles in network or molecular catalysis below the level of enzymes, but they share with all catalysts the property that they are not consumed by participating in reactions, and therefore are key loci of control over metabolism. Cofactors as transfer agents are essential to completing many network-catalytic loops. In association with enzymes, they can create channels and active sites, and thus they facilitate molecular catalysis. An example of the creation of channels by cofactors is given by the function of cobalamin as a C_1 transfer agent to the nickel reaction center in the acetyl-CoA synthase from a corrinoid iron–sulfur protein [216–218]. An example of cofactor incorporation in active sites is the role of TPP as the reaction center in the pyruvate-ferredoxin oxidoreductase (PFOR), which lies at the end of a long electron-transport channel formed by Fe–S clusters [108]. Through the limits in their own functions or in the functional groups they transport through networks, they may impose constraints on chemical diversity or create bottlenecks to evolutionary innovation. The previous sections have shown that many module boundaries in carbon fixation and core metabolism are defined by idiosyncratic reactions, and we have noted that many of these idiosyncrasies are associated with specific cofactor functions.

Cofactors, as topological hubs, and participants in reactions at high-flux boundaries in core and intermediary metabolism, are focal points of natural selection. The adaptations available to key atoms and bonds include altering charge or pKa, changing energy level spacing through non-local electron transport, or altering orbital geometry through ring strains. Divergences in low-level cofactor chemistry may alter the distribution of functional groups and thereby change the global topology of metabolic networks, and some of these changes map onto deep lineage divergences in the tree of life. A well-understood example is the repartitioning of C_1 flux from methanopterins versus folates [22, 93]. The same adaptation that enables formylation of methanopterins within an exclusively thioester system, where the homologous folate reaction requires ATP, reduces the potential for methylene-group transfer, and necessitates the oxidative formation of serine from 3PG in methanogens, which is not required of acetogens.

Most research on the origin of life has focused either on the metabolic substrate [6, 219] or catalysis by RNA [193], but we believe the priority of cofactors deserves (and
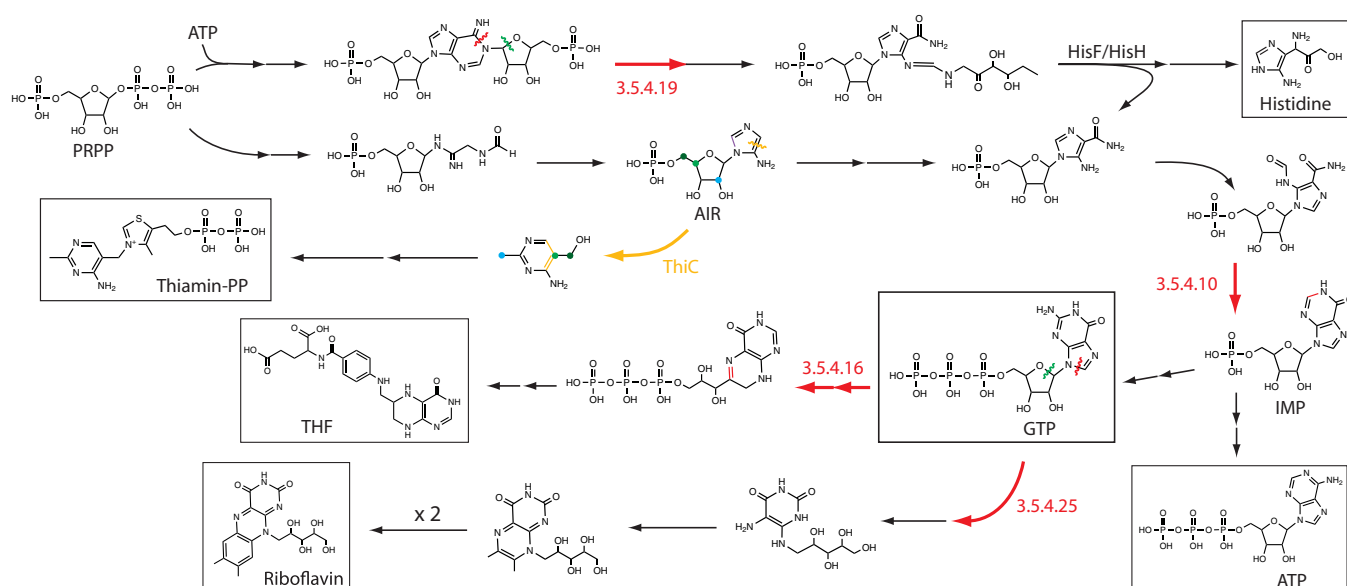
---

[6] It is natural in many respects to include ferredoxins (and related flavodoxins) in this list. Although not cofactors by the criteria of size and biosynthetic complexity, these small, widely diversified, ancient and general-purpose Fe_2S_2, Fe_3S_4 and Fe_4S_4-binding polypeptides are unique low-potential (high-energy) electron donors. Reduced ferredoxins are often generated in reactions involving radical intermediates in iron–sulfur enzymes, described below in connection with electron bifurcation.

Figure 14. Key molecular re-arrangements in the network leading from AIR to purines and the purine-derived cofactors. The 3.5.4 class of cyclohydrolases (red) convert FAICAR to IMP (precursor to purines), and subsequently convert GTP to folates and flavins by opening the imidazole ring. Acting on the six-member ring of ATP and on a second attached PRPP, the enzyme 3.5.4.19 initiates the pathway to histidinol. The thiamin pathway, which uses the unclassified enzyme ThiC to hydrolyze imidazole and ribosyl moieties, is the most complex, involving multiple group rearrangements (indicated by colored atoms). This complexity, together with the subsequent attachment of a thiazole group, lead us to place thiamin latest in evolutionary origin among these cofactors.

is beginning to receive) greater consideration [89, 220]. In the expansion of metabolic substrates from inorganic inputs, the pathways to produce even such complex cofactors as folates *et alia* are comparable in position and complexity to those for purine RNA, while some for functional groups such as nicotinamide [89] or chorismate are considerably simpler. Therefore, even though it is not known what catalytic support or memory mechanisms enabled the initial elaboration of metabolism, any solutions to this problem should also support the early emergence of at least the major redox and C- and N-transfer cofactors. Conversely, the pervasive dependence of biosynthetic reactions on cofactor intermediates makes the expansion of protometabolic networks most plausible if it was supported by contemporaneous emergence and elaboration of cofactor groups. In this interpretation cofactors occupy an intermediate position in chemistry and complexity, between the small-metabolite and oligomer levels [89]. They were the transitional phase when the reaction mechanisms of core metabolism came under selection and control of organic as opposed to mineral-based chemistry, and they provided the structured foundation from which the oligomer world grew.

We argue next that a few properties of the elements have governed both functional diversification and evolutionary optimization of many cofactors, especially those associated with core carbon fixation. We focus on heterocycles with conjugated double bonds incorporating nitrogen, and on the groups of functions that exploit special properties of bonds to sulfur atoms. The recruitment of elements or special small-molecule contexts constitutes an additional distinct form of modularity within metabolism. Like the substrate network, cofactor groups often share or re-use synthetic reaction sequences. However, unlike the small-molecule network, cofactors can also be grouped by criteria of catalytic

similarity that are independent of pathway recapitulation. For example, alkyl-thiol cofactors, which comprise diverse groups of molecules, all make essential use of distinctive properties of the sulfur bond to carbon, which appear nowhere else in biochemistry. As an example involving elements in specific contexts, a large group of cofactors employing C-N heterocycles all arise from a single sub-network whose reactions are catalyzed by related enzymes, and the transport and catalytic functions performed by the heterocycles are distinctive of this cofactor group.

4.2. The cofactors derived from purine RNA

Most of the cofactors that use heterocycles for their primary functions have biosynthetic reactions closely related to those for purine RNA. These reactions are performed by a diverse class of cyclohydrolase enzymes, which are responsible for the key ring-formation and ring-rearrangement steps. The cyclohydrolases can split and reform the ribosyl ring in PRPP, jointly with the 5- and 6-membered rings of guanine and adenine. Five biosynthetically related cofactor groups are formed in this way. Three of these—the folates, flavins and deazaflavins—are formed from GTP, while one—thiamin—is formed from a direct precursor to GTP, as shown in figure 14.

Folates. The folates are structurally most similar to GTP, but have undergone the widest range of secondary specializations, particularly in the Archaea. They are primarily responsible for binding C_1 groups during reduction from formyl to methylene or methyl oxidation states, and their secondary diversifications are apparently results of selection to tune the free-energy landscape of these oxidation states.

Flavins and deazaflavins. The flavins are tricyclic compounds formed by condensation of two pterin groups,

while deazaflavins are synthesized through a modified version of this pathway, in which one pterin group is replaced by a benzene ring derived from chorismate. Flavins are general-purpose reductants, while deazaflavins are specifically associated with methanogenesis.

Thiamin. Thiamin combines a C–N heterocycle common to the GTP-derived cofactors with a thiazole group (so incorporating sulfur), and shares functions with both the purine cofactor group and the alkyl-thiol group reviewed in the next subsection.

Histidine. The last 'cofactor' in this group is the amino acid histidine, synthesized from ATP rather than GTP but using similar reactions. Histidine is a general acid–base catalyst with unique pKa, which in many ways functions as a 'cofactor in amino acid form' [51].

We will first describe in detail the remarkable role of the folate group in the evolutionary diversification of the WL pathway, and then return to general patterns found among the purine-derived cofactors, and their placement within the elaboration of metabolism and RNA chemistry.

4.2.1. Folates and the central superhighway of C_1 metabolism. Members of the folate family carry C_1 groups bound to either the N^5 nitrogen of a heterocycle derived from GTP, an exocyclic N^{10} nitrogen derived from a PABA, or both. The two most common folates are THF, ubiquitous in bacteria and common in many archaeal groups, and tetrahydromethanopterin (H_4MPT), essential for methanogens and found in a small number of late-branching bacterial clades. Other members of this family are exclusive to the archaeal domain and are structural intermediates between THF and H_4MPT. Two kinds of structural variation are found among folates, as shown in figure 15. First, only THF retains the carbonyl group of PABA, which shifts electron density away from N^{10} via the benzene ring and lowers its pKa relative to N^5 of the heterocycle. All other members of the family lack this carbonyl. Second, all folates besides THF incorporate one or two methyl groups that impede rotation between the pteridine and aryl-amine planes, changing the relative entropies of formation among different binding states for the attached C_1 [22, 93, 221].

Folates mediate a diverse array of C_1 chemistry, various parts of which are essential in the biosynthesis of all organisms [93]. The collection of reactions, summarized in figure 5, has been termed the 'central superhighway' of one-carbon metabolism. Functional groups supplied by folate chemistry, connected by interconversion of C_1-oxidation states along the superhighway, include (1) formyl groups for synthesis of purines, formyl-tRNA, and formylation of methionine (fMet) during translation, (2) methylene groups to form thymidilate, which are also used in many deep-branching organisms to synthesize glycine and serine, forming the ancestral pathway to these amino acids [22], and (3) methyl groups which may be transferred to SAM as a general methyl donor in anabolism, to the acetyl-CoA synthase to form acetyl-CoA in the WL pathway, or to coenzyme-M where the conversion to methane is the last step in the energy system of methanogenesis.

The variations among folates, shown in figure 15, leave the charge, pKa and resulting C–N bond energy at N^5 roughly unaffected, while the N^{10} charge, pKa and C–N bond energy change significantly across the family. This charge effect, together with entropic effects due to steric hindrance from methyl groups, can sharply vary the functional roles that different folates play in anabolism.

The biggest difference lies between THF and H_4MPT. In THF, the N^{10} pKa is as much as 6.0 natural-log units lower than that of N^5 [222]. The resulting higher-energy C-N bond cannot be formed without hydrolysis of one ATP, either to bind formate to N^{10} of THF, or to cyclize N^5-formyl-THF to form N^5,N^{10}-methenyl-THF (see figure 5). This latter reaction is the mirror image of the cyclization of N^{10}-formyl-THF, and as we will argue below, a plausibly conserved evolutionary intermediate in the attachment of formate onto folates. After further reduction, the resulting methylene is readily transferred to lipoic acid to form glycine and serine, in what we have termed the 'glycine cycle' [22] (the lipoyl-protein based cycle on the right in figure 5).

In contrast, in H_4MPT the difference in pKa between N^{10} and N^5 is only 2.4 natural-log units. The lower C–N^{10} bond energy permits spontaneous cyclization of N^5-formyl-H_4MPT, following (also ATP-independent) transfer of formate from a formyl-methanofuran cofactor. Through this sequence, methanogens fix formate in an ATP-independent system using only redox chemistry. The initial free energy to attach formate to methanofuran is provided by the terminal methane released in methanogenesis (the Co-M/Co-B cycle in figure 5). The resulting downstream methylene group, however, has too little energy as a leaving group to transfer to an alkyl-thiol cofactor, so methanogens sacrifice the ability to form glycine and serine by direct reduction of formate.

The reconstructed ancestral use of the 7–9 reactions in figure 5 is to reduce formate to acetyl-CoA or methane. However, the reversibility of many reactions in the sequence, possibly requiring substitution of reductant/oxidant cofactors, allows folates to accept and donate C_1 groups in a variety of oxidation states, from and into many pathways including salvage pathways. Methylotrophic proteobacteria which have obtained H_4MPT through horizontal gene transfer [195, 196] may run the full reaction sequence in reverse. They may use either H_4MPT to oxidize formaldehyde or THF to oxidize various methylated C_1 compounds, in both cases leading to formate, or other intermediary oxidation states (from THF) as inputs to anabolic pathways. In many late-branching bacteria, some archaea and eukaryotes, the THF based pathway may run in part oxidatively and in part reductively, through connections to either gluconeogenesis/glycolysis or glyoxylate metabolism. In these organisms serine (derived through oxidation, amination and dephosporylation from 3-phosphoglycerate) or glycine (derived through amination of glyoxylate) become the sources of transferable methyl groups in anabolism. This versatility has preserved the folate pathway as an essential module of biosynthesis in all domains of life, and at the same time has made it a pivot of evolutionary variation.

4.2.2. Refinement of folate-C_1 chemistry maps onto lineage divergence of methanogens. The structural and functional
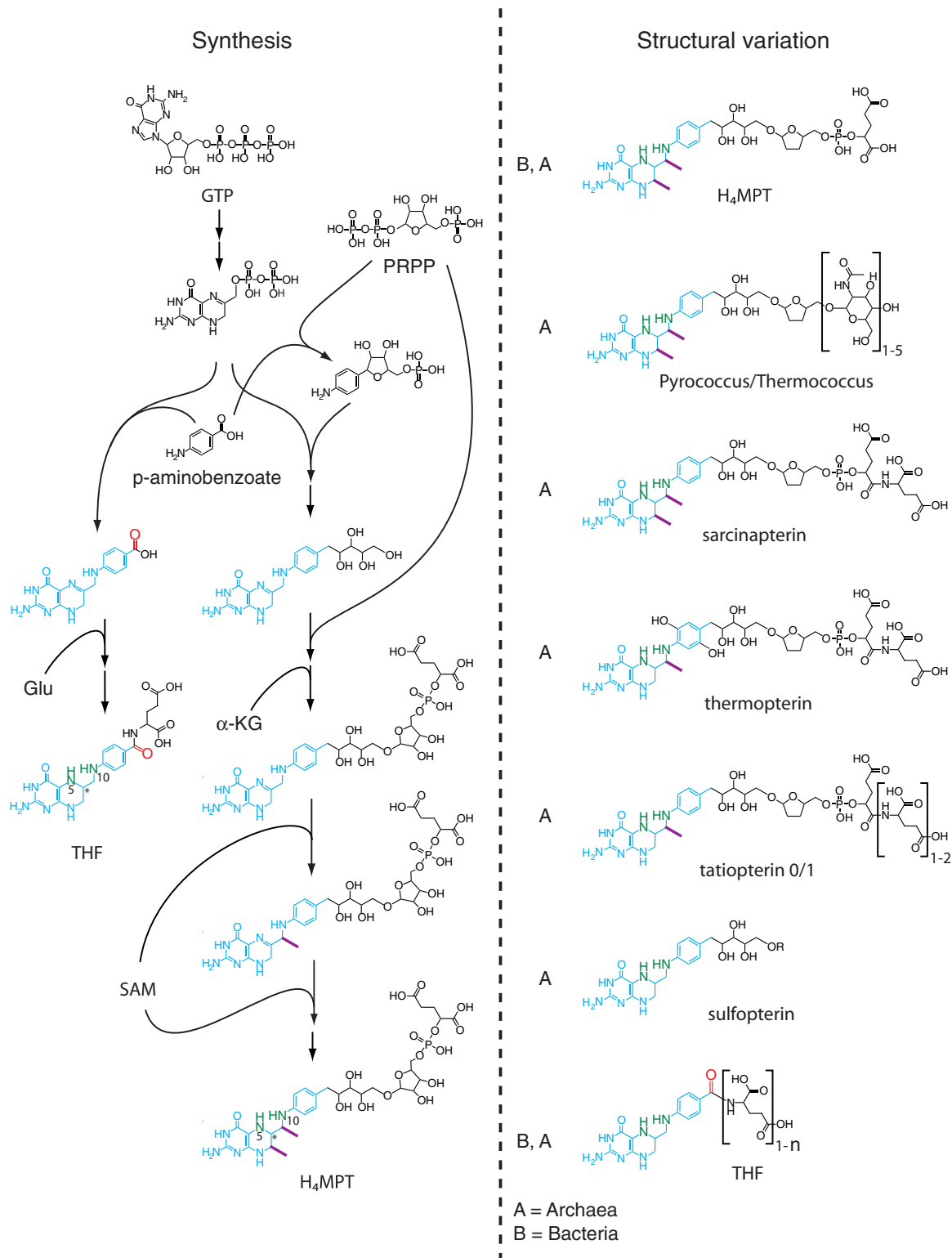
Figure 15. Structural variants among cofactors in the folate family, shown with the biosynthetic pathways that produce these variations, from [22]. Pteridine and benzene groups are shown in blue, active nitrogens are shown in green, electron-withdrawing carbonyl groups are shown in red and methyl groups that regulate steric hindrance are shown in purple.

variation within the folate family illustrates the way that selection, acting on cofactors, can create large-scale re-arrangements in metabolism, enabling adaptations that are reflected in lineage divergences. The free-energy cascade described in the last section, linking ATP hydrolysis, the charge and pKa of the N^{10} nitrogen, and the leaving-group activity of the resulting bound carbon for transfer to alkyl-thiol cofactors or other anabolic pathways, is a fundamental long-range constraint of folate-C_1 chemistry. A comparative analysis of gene profiles in

pathways for glycine and serine synthesis, explained in [22], shows that while the constraint cannot be overcome, its impact on the form of metabolism can vary widely depending on the structure of the mediating folate cofactor.

The annotated role for ATP hydrolysis in WL autotrophs is to attach formate to N^{10} of THF, initiating the reduction sequence. However, many deep-branching bacteria and archaea show no gene for this reaction, while multiple lines of evidence indicate that THF nonetheless functions as a carbon-fixation cofactor in these organisms [22]. In almost all cases where an ATP-dependent N^{10}-formyl-THF synthase is absent, an ATP-dependent N^5-formyl-THF cycloligase [223, 224] is found. This is another case where a broad evolutionary context allows an alternate interpretation. N^5-formyl-THF cycloligase was originally discovered in mammalian systems, where its function has been highly uncertain and hypothesized to be the salvage mechanism as part of a futile cycle [223, 224], before being found to be widespread across the tree of life [22]. If we deduce by reconstruction, however, that ancestral folate chemistry operated in the fully reductive direction, and that in H_4MPT systems formate is attached at the N^5 position, while in THF systems formate is attached at the N^{10} position, the widespread distribution of the cycloligase takes on a different possible meaning. It is plausible that the N^5-formyl-THF cycloligase allows a formate incorporation pathway that is an evolutionary intermediate between the commonly recognized pathway using THF and its evolutionary derivative using H_4MPT (see figure 5). The ATP-dependent cycloligase produces N^5,N^{10}-methenyl-THF from N^5-formyl-THF, which may potentially form spontaneously due to the higher N^5-pKa [224]. ATP hydrolysis is thus specifically linked to the N^{10}-carbon bond, which is the primary donor for carbon groups from folates. Methanogens, in contrast, escape the dependence on ATP hydrolysis by decarboxylating PABA before it is linked to pteridine to form methanopterin (see figure 15), but they sacrifice methyl-group donation from H_4MPT to most anabolic pathways, making methanogenesis viable only in clades that evolved the oxidative pathway to serine from 3-phosphoglycerate.

We noted in section 3.4 that the elimination of one ATP-dependent acyl-CoA synthase in acetogens reduces the free energy cost of carbon fixation relative to rTCA cycling. The decoupling of the formate-fixation step on methanopterins from ATP hydrolysis is a further significant innovation, lowering the ATP cost for uptake of CO_2. This divergence of H_4MPT from THF, and a related divergence of deazaflavins from flavins (see figure 16), follow phylogenetically (and we believe, were responsible for) the divergence of the methanogens from other euryarcheota [22].

We regard this example as representative of the way that innovations in cofactor chemistry more generally mediated large-scale rearrangements in metabolism, and corresponding evolutionary (and ecological) divergences of clades. Another similar example comes from the quinones, a diverse family of cofactors mediating membrane electron transport [225]. [114] found that the synthetic divergence of mena- and ubiquinone follows the pattern of phylogenetic diversification within

proteobacteria. δ- and ϵ-proteobacteria use menaquinone, γ-proteobacteria use both mena- and ubiquinone, and α- and β-proteobacteria use only ubiquinone. Because mena- and ubiquinone have different midpoint potentials, it was suggested that their distribution reflects changes in environmental redox state as the proteobacteria diversified during the rise of oxygen [114, 226].

4.2.3. Relation of the organic superhighway to minerals. An interpretive frame for many of these observations is the proposal that metabolism is an outgrowth of geochemistry [41, 74, 149], which came under the control of living organisms [58] (see section 7 for dedicated discussion). If we wish to judge this proposal, then it is informative to look for parallels and differences between biochemical and plausible geochemical reaction sequences. The distinctive features of biochemical C_1 reduction are the attachment of formate to tuned heterocyclic or aryl-amine nitrogen atoms for reduction, and the transfer of reduced C_1 groups to sulfhydryl groups (of SAM, lipoic acid or CoM). In the mineral-origin hypothesis for direct reduction, the C_1 were adsorbed at metals and either reduced through crystal oxidation [227] or by reductant in solution. The transfer of reduced C_1 groups to alkyl-thiol cofactors may show continuity with reduction on metal-sulfide minerals. However, the mediation of reduction by nitrogens appears to be a distinctively biochemical innovation.

4.2.4. Cyclohydrolases as the central enzymes in the family, and the resulting structural homologies among cofactors. The common reaction mechanism unifying the purine-derived cofactors is an initial hydrolysis of both purine and ribose rings performed by cyclohydrolases assigned EC numbers 3.5.4 (see figure 14). All cyclohydrolases within this EC family are used for biosynthesis or conversions within this class of molecules. They are responsible for the synthesis of inosine-monophosphate (IMP, precursor to AMP and GMP) from 5-formamidoimidazole-4-carboxamide ribonucleotide (FAICAR), for the first committed steps in the syntheses of both folates and flavins from GTP, and for the initial ring-opening step in the synthesis of histidine from ATP and PRPP. Figure 14 shows the key steps in the network synthesizing both purines and the pterins, folates, flavins, thiamin and histidine.

The common function of the 3.5.4 cyclohydrolases is hydrolysis of rings on adjacent nucleobase and ribose groups, or the formation of cycles by ligation of ring fragments. In all cases, the ribosyl moieties come from PRPP. In the synthesis of pterins from GTP and of histidinol from ATP, both a nucleobase cycle and a ribose are cleaved. In pterin synthesis, the imidazole of guanine and the purine ribose are cleaved. In histidine synthesis, the six-membered ring of adenine is cleaved (at a different bond than the one synthesized from FAICAR), and the ribose comes from a secondary PRPP.

By far the most complex synthesis in this family is that of thiamin from aminoimidazole ribonucleotide (AIR). This sequence begins with an elaborate molecular rearrangement, performed in a single step by the enzyme ThiC [205]. (Eukaryotes use an entirely different pathway, in which the
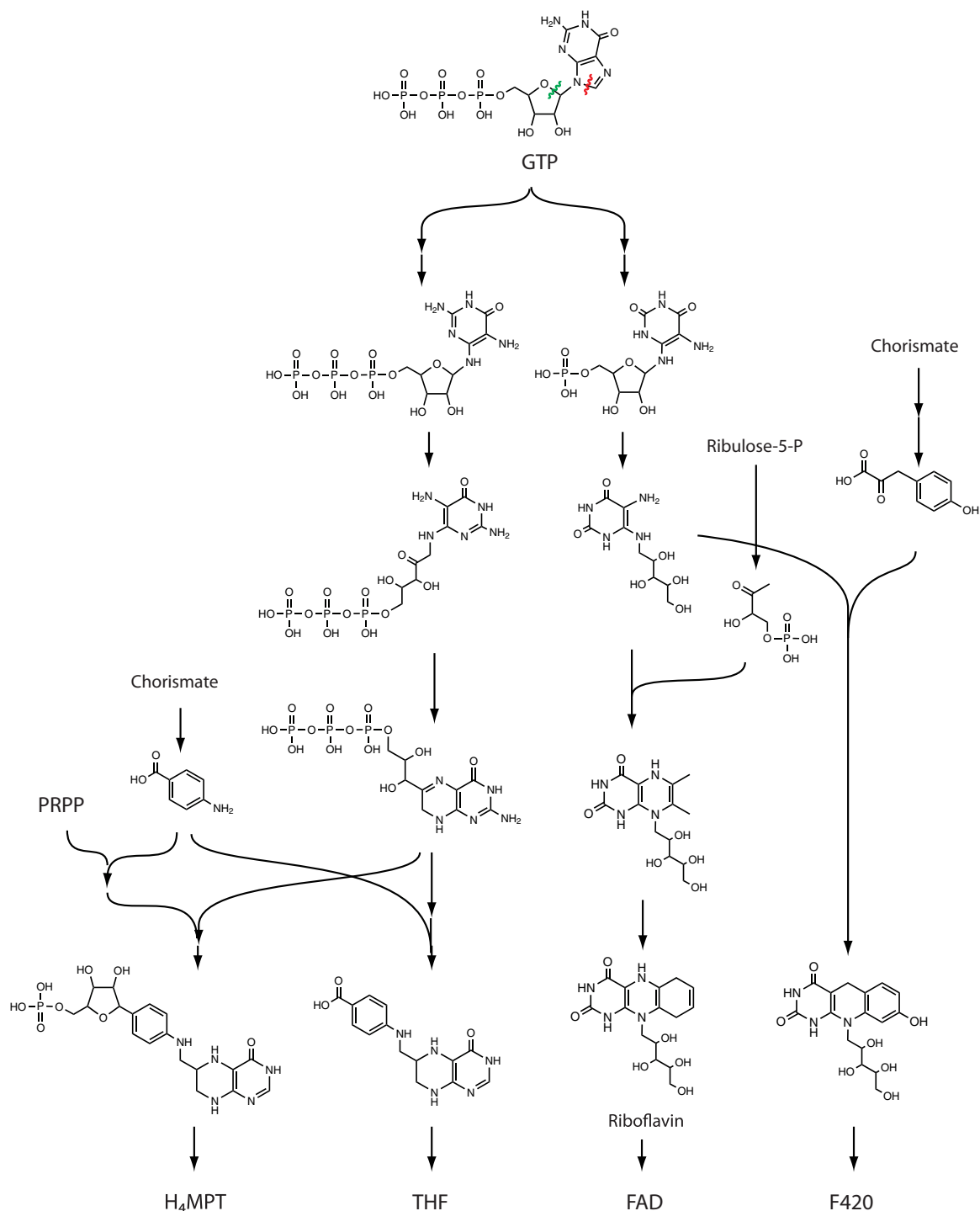
Figure 16. The substrate modifications leading from GTP to the four major cofactors H$_4$MPT, THF, riboflavin (in FAD) and the archaeal homologue deazaflavin F$_{420}$. The branches indicating substrate diversification may also reflect evolutionary lineage.

pyrimidine is synthesized from histidine and pyridoxal-5-phosphate [228].) While the ThiC enzyme is unclassified, and its reaction mechanism incompletely understood, it shares apparent characteristics with members of the 3.5.4 cyclohydrolases. As in the first committed steps in the synthesis of folates and flavins from GTP, both a ribose ring and a 5-member heterocycle are cleaved and subsequently (as in folate synthesis) recombined into a 6-member heterocycle. The complexity of this enzymatic mechanism makes a pre-enzymatic homologue to ThiC difficult to imagine, and

suggests that thiamin is both of later origin, and more highly derived, than other cofactors in this family. This derived status is supported by the fact that the resulting functional role of thiamin is not performed on the pyrimidine ring itself, but rather on the thiazole ring to which it is attached, and which is likewise created in an elaborate synthetic sequence [205].

Figure 16 shows the detailed substrate re-arrangement in the sub-network leading from GTP to methanopterins, folates, riboflavin and the archaeal deazaflavin F$_{420}$. In the pterin branch, both rings of neopterin are synthesized directly

from GTP, and an aryl-amine originating in PABA provides the second essential nitrogen atom. PABA is either used directly (in folates) or decarboxylated with attachment of a PRPP (in methanopterins) to vary the pKa of the amine. In contrast, the flavin branch is characterized by the integration of either ribulose (in riboflavin) or chorismate (in F_{420}) to form the internal rings. Two 6,7-dimethyl-8-(D-ribityl)lumazine are condensed to form riboflavin, whereas a single GTP with chorismate forms F_{420}.

The cyclohydrolase reactions can be considered the key innovation enabling the biosynthesis of this whole family of cofactors and, importantly, of purine RNA itself. The heterocycles that are formed or cleaved by these reactions provide the central structural components of the active parts of the final cofactor molecules. In this sense, except for TPP, the distinctions among purine-derived cofactors can be considered secondary modifications on a background structured by PRPP and C–N heterocycles. If we consider sub-networks of metabolism as producing key structural or functional components, in this case for the synthesis of cofactors, then this family draws on only two such developed sub-networks. The first of these is purine synthesis and the other is synthesis of chorismate, the precursor to PABA and the unique source of single benzene rings in biochemistry [229]. Flexibility in the ways that chorismate is modified to control electron density, and the way the benzene ring is combined with other heterocycles, contributes to the combinatorial elaboration within the family.

4.2.5. Placing the members of the class within the network expansion of metabolism. The following observations suggest to us the possibility that most of the purine-derived cofactors (perhaps excepting thiamin) were available contemporaneously with monomer purine RNA.

For some reactions, the abstraction of enzyme mechanisms is advanced enough to identify small-molecule organocatalysts that could have provided similar functions [230, 231]. The current understanding of cyclohydrolase proteins, however, does not suggest other simpler mechanisms by which similar reactions might first have been catalyzed, leaving us almost wholly uncertain about how RNA was first formed. Unless non-enzymatic mechanisms are discovered which are both plausible and *selective*, our previous arguments about the permissiveness of crude catalysts lead us to expect that, at whatever stage catalysts capable of interconverting AIR, AICAR, FAICAR and IMP first became available, pteridines would have been formed contemporaneously and possibly played a role in the elaboration of the metabolic network. (See section 5 for further discussion on promiscuous versus selective enzymes.) If the chorismate pathway (which begins in the sugar-phosphate network) had also arisen by that stage, the same arguments suggest that folates and flavins may also have been available. In this supposition we are treating the first three EC numbers as an appropriate guide to reaction mechanism without restriction of the molecular substrate. Whether the first RNA were produced in this way, or through structurally very dissimilar stages, is a currently active question [232].

As in our discussion of the root node in section 3.4.4, we consider it important to apply *ubiquitously* the premise that enabling network throughput and pruning network diversity were concurrent ongoing requirements in the co-evolution of substrate reactions and their catalysts. Most often [160, 190, 233], the inability to prune networks is recognized as a problem for the early formation of order. In the case of the purine-derived cofactors, it may offer both clues to help explain the structure of the biosynthetic network, and a way to break down the problem of early metabolic evolution into simpler steps with intermediate criteria for selection.

The patterns that characterize current metabolism as a recursive network expansion [200, 201] about inorganic inputs are most easily understood as a reflection of the organic-chemical possibilities opened by cofactors. Pterins, as donors of activated formyl groups, support (among other reactions) the synthesis of purines, forming a short autocatalytic loop. Similarly, flavins would have augmented redox reactions. Finally, it has long been recognized that acid/base catalysis is uniquely served by histidine, which has a pKa ≈ 6.5 on the ε-nitrogen, a property not found among any biological ribonucleotides (though possible for some substituted adenine derivatives) [234].

Within the class of GTP-derived cofactors, a sub-structure may perhaps be suggested: the dimer condensation that forms riboflavin is a hierarchical use of building blocks formed from GTP. Although simple and consisting of a single key reaction, this could reflect a later stage of refinement. It is recognized [235] that flavins are somewhat specialized reductants, both biosynthetically and functionally more specific than the much simpler nicotinamide cofactors, which plausibly preceded them [89].

4.2.6. Purine-derived cofactors selected before RNA itself, as opposed to having descended from an RNA world defined through base pairing? The overlap between RNA and cofactor biosynthesis, and the incorporation of AMP in several cofactors (where is serves primarily as a 'handle' for docking), has been noticed and given the interpretation that cofactors are a degenerated relic of an *oligomer* RNA world [210]. While *monomer* RNA is of comparable complexity to small-molecule cofactors, oligomer RNA is significantly more complex. The only significant logical motivation to place oligomer RNA prior to small-molecule cofactors, is therefore the premise that RNA base pairing and replication is the *least*-complex plausible mechanism supporting (specifically, Darwinian) selection and persistence of catalysts that are hypothesized to have been required for the elaboration of biosynthesis.

This is still a complex premise, however, as it requires not only organosynthesis of oligomer RNA, but also chiral selection and mechanisms to enable base pairing and (presumably template-directed) ligation [236]. A particular problem for RNA replication is the steric restriction to 3′-5′ phosphate esters, over the kinetically favored 2′-5′ linkage. In comparison, small-molecule catalysis by either RNA [237] or related cofactors may be considered in any context that

supports their synthesis[7]. If chemical mechanisms are found which support structured organosynthesis and selection—a requirement for any metabolism-first theory of the origin of life—the default premise may favor simplicity: that heterocycles were first selected as cofactors, and that purine RNA, only one among many species maintained by the same generalized reactions, was subsequently selected for chirality, base-pairing and ligation.

4.3. The alkyl-thiol cofactors

The major chemicals in this class include the sulfonated alkane-thiols coenzyme-B (CoB) and coenzyme-M (CoM), cysteine and homocysteine including the activated forms S-adenosyl-homocysteine (which under methylation becomes SAM), lipoic acid and pantetheine or pantothenic acid, including pantetheine-phosphate. The common structure of the alkyl-thiol cofactors is an alkane chain terminated by one or more sulfhydryl (SH) groups. In all cases except lipoic acid, a single SH is bound to the terminal carbon; in lipoic acid two SH groups are bound at sub-adjacent carbons. Differences among the alkyl-thiol cofactors arise from their biosynthetic context, the length of their alkane chains, and perhaps foremost the functional groups that terminate the other ends of the chains. These may be as simple as sulfones (in CoB) or as complex as peptide bonds (in CoA).

Cofactors in this class serve three primary functions, as reductants (cysteine, CoB, pantetheine and one sulfur on lipoic acid), carriers of methyl groups (CoM, SAM), and carriers of larger functional groups such as acyl groups (lipoic acid in lipoyl protein, phosphopantetheine in ACP). A highly specialized role in which H is a leaving group is the formation of thioesters at carboxyl groups (pantethenic acid in CoA, lipoic acid in lipoyl protein) This function is essential to substrate-level phosphorylation [241], and appears repeatedly in the deepest and putatively oldest reactions in core metabolism. A final function closely related to reduction is the formation and cleavage of S–S linkages by cysteine in response to redox state, which is a major controller of both committed and plastic tertiary structure in proteins. The sulfur atoms on cysteine often form coordinate bonds to metals in metallo-enzymes, a function that we may associate with protein ligands, in contrast to the more common nitrogen atoms that coordinate metals in pyrrole cofactors.

The properties of the alkyl-thiol cofactors derive largely from the properties of sulfur, which is a 'soft' period-3 element

---

[7] The relative importance of synthesis and selection depends on whether opening access to a space of reactions, or concentrating flux within a few channels in that space, is the primary limit on the emergence of order at each phase in the elaboration of metabolism. Following our earlier arguments about the need for autocatalysis, selection will be essential in some stages, and this remains an important problem for metabolism-first premises [190]. Chemical selection criteria derived from differential growth rate pose no problem in the domain of small-molecule organocatalysis, but the identification of plausible mechanisms to preserve selected differences remains an important area of work. Most mechanisms that do not derive from RNA base pairing involve separation by spatial geometry or material phases, including porous-medium processes akin to invasion percolation [189], or more general proposals for compositional inheritance [238–240], abstracted from models of coacervate chemistry.

[242] that forms relatively unstable (usually termed 'high-energy') bonds with the hard period-2 element carbon. For the alkyl-thiol cofactors in which sulfur plays direct chemical roles, three main bonds dictate their chemistry: S–C, S–S, and S–H. Sulfur can also exist in a wide range of oxidation states, and for this reason often plays an important role in energy metabolism [243], particularly for chemotrophs, and due to its versatility has been suggested to precede oxygen in photosynthesis [244]. The electronic versatility of sulfur and the high-energy C–S bonds combine with the large atomic radius of sulfur to give access to additional geometrical, electronic and ring-straining possibilities not available to CHON chemistry.

Although not alkyl-thiol compounds as categorized above, two additional cofactors that make important indirect use of sulfur are thiamin and biotin. In neither case is sulfur the element to which transferred C_1 groups are bound. For reactions involving TPP the C_1-unit is bound to the carbon between sulfur and the positively charged nitrogen, while in biotin C_1-units are bound to the carboxamide nitrogen in the (non-aromatic) heterocycle opposite the sulfur-containing ring. It seems likely, however, that the sulfur indirectly contributes to the properties of the binding carbon or nitrogen, through some combination of electrostatic, resonance or possibly ring-straining interactions. The importance of the sulfur to the focal carbon or nitrogen atom is suggested by the complexity of the chemistry and enzymes involved in its incorporation into these two cofactors [205, 245].

4.3.1. Biochemical roles and phylogenetic distribution. Transfer of methyl or methylene groups.
The S atoms of CoM, lipoic acid, and S-adenosyl-homocysteine accept methyl or methylene groups from the nitrogen atoms of pterins. Considering that transition-metal sulfide minerals are the favored substrates for prebiotic direct-C_1 reduction [147, 149, 246], a question of particular interest is how, in mineral scenarios for the emergence of carbon fixation, the distinctive relation between tuned nitrogen atoms in pterins as carbon carriers, and alkyl-thiol compounds as carbon acceptors, would have formed.

Reductants and co-reductants. CoB and CoM act together as methyl carrier and reductant to form methane in methanogenesis. In this complex transfer [150], the fully-reduced (Ni^+) state of the nickel tetrapyrrole F_{430} forms a dative bond to $-CH_3$ displacing the CoM carrier, effectively re-oxidizing F_{430} to Ni^{3+}. Reduced F_{430} is regenerated through two sequential single-electron transfers. The first, from CoM–SH, generates a Ni^{2+} state that releases methane, while forming a radical $CoB^{\cdot}$–S–S–CoM intermediate with CoB. The radical then donates the second electron, restoring Ni^+. The strongly oxidizing heterodisulfide CoB–S–S–CoM is subsequently reduced with two NADH, regenerating CoM–SH and CoB–SH.

A similar role as methylene carrier and reductant is performed by the two SH groups in lipoic acid. CoM is specific to methanogenic archaea [247], while lipoic acid and S-adenosyl-homocysteine are found in all three domains [22, 248]. Lipoic acid is formed from octanoyl-CoA, emerging

from the biotin-dependent malonate pathway to fatty acid synthesis, and along with fatty acid synthesis [109], may have been present in the universal common ancestor. The previously noted universal distribution of the glycine cycle supports this hypothesis.

Role in the reversal of citric-acid cycling. Lipoic acid becomes the electron acceptor in the oxidative decarboxylation of α-ketoglutarate and pyruvate in the oxidative Krebs cycle, replacing the role taken by reduced ferredoxin in the rTCA cycle. Thus the prior availability of lipoic acid was an enabling precondition for reversal of the cycle in response to the rise of oxygen.

Carriers of acyl groups. Transport of acyl groups in the ACP proceeds through thioesterification with pantetheine phosphate, similar to the thioesterification in fixation pathways. In fatty acid biosynthesis acyl groups are further processed while attached to the pantetheine phosphate prosthetic group.

Electron bifurcation. The heterodisulfide bond of CoB–S–S–CoM has a high midpoint potential ($E'_0 = -140$ mV), relative to the H^+/H_2 couple ($E'_0 = -414$ mV), and its reduction is the source of free energy for the *endergonic* production of reduced ferredoxin (Fd^{2-}, E'_0 *in situ* unknown but between -520 mV and -414 mV) [249], which in turn powers the initial uptake of CO_2 on H_4MPT in methanogens. The remarkable direct coupling of exergonic and endergonic redox reactions through splitting of binding pairs into pairs of radicals, which are then directed to paired high-potential/low-potential acceptors, is known as *electron bifurcation* [140]. Variant forms of bifurcation are coming to be recognized as a widely-used strategy of metal-center enzymes, either consuming oxidants as energy sources to generate uniquely biotic low-potential reductants such as Fd^{2-} [249, 250–252], or to 'titrate' redox potential to minimize dissipation and achieve reversibility of redox reactions involving reductants at diverse potentials, e.g. by combining low-potential (Fd^{2-}, $E'_0 = -420$ mV) and high-potential (NADH, $E'_0 = -300$ mV) reductants to produce intermediate-potential reductants (NADPH, $E'_0 = -360$ mV) [253]. Together with substrate-level phosphorylation (SLP), electron bifurcation may be the principal chemical mechanism (contrasted with membrane-mediated oxidative phosphorylation) for interconverting biological energy currencies, and along with SLP [241], a mechanism of central importance in the origin of metabolism [254]. Small metabolites including such heterodisulfides of cofactors, which can form radical intermediates exchanging single electrons with Fe–S clusters (typically via flavins) are essential sources and repositories of free energy in pathways using bifurcation. Both electron bifurcation and the stepwise reduction of F$_{430}$ (above) illustrate the central role of metals as mediators of single-electron transfer processes in metabolism.

4.3.2. Participation in carbon fixation pathway modules. The similarity between the glycine cycle and methanogenesis in figure 5 emphasizes the convergent roles of alkyl-thiol cofactors. In the glycine cycle, methylene groups are accepted by the terminal sulfur on lipoic acid, and the subadjacent

SH serves as reductant when glycine is produced, leaving a disulfide bond in lipoic acid. The disulfide bond is subsequently reduced with NADH. In methanogenesis, a methyl group from H_4MPT is transferred to CoM, with the subsequent transfer to F$_{430}$, and the release from F$_{430}$ as methane in the methyl-CoM reductase, coupled to formation of CoB–S–S–CoM. The heterodisulfide is again reduced with NADH, but employs a pair of electron bifurcations to retain the excess free energy in the production of Fd^{2-} rather than dissipating it as heat [249]. Methanogenesis is thus associated with seven distinctive cofactors beyond even the set known to have diversified functions within the archaea [5], again suggesting the derived and highly optimized nature of this Euryarchaeal phenotype. The striking similarity of these two methyl-transfer systems, mediated by independently evolved and structurally quite different cofactors, suggests evolutionary convergence driven specifically by properties of alkyl thiols.

A curious pattern, which we note but do not attempt to interpret, is the association of non-sulfur, nitrogen-heterocycle cofactors with WL carbon fixation, contrasted with the use of sulfur-containing heterocycles in carboxylation reactions of the rTCA cycle. The non-sulfur cofactors THF and H_4MPT are used in the reactions of the WL pathway, while the biosynthetically-related but sulfur-containing cofactor thiamin mediates the carbonyl insertion (at a thioester) in rTCA [108, 255]. Biotin—which has been generally associated with malonate synthesis in the fatty-acid pathway (and derivatives such as propionate carboxylation to methyl-malonate in 3HP [109])—mediates the subsequent β-carboxylation of pyruvate and of α-ketoglutarate [110, 256, 257]. Thus the two cofactors we have identified as using sulfur indirectly to tune properties of carbon or nitrogen C_1-bonding atoms mediate the two chemically quite different sequential carboxylations in rTCA.

4.4. Carboxylation reactions in cofactor synthesis

Carboxylation reactions can be classified as falling into two general categories: those used in core carbon 'uptake', and those used exclusively in the synthesis of specific cofactors. In addition to carboxylation reactions in carbon-fixation pathways, the former category includes the carboxylation of crotonyl-CoA in the glyoxylate regeneration cycle. This cycle is a mixotrophic rather than an autotrophic pathway, but this reaction does form a distinct entry point for CO_2 into the biosphere. The carboxylation of acetyl-CoA to malonyl-CoA further serves a dual purpose, in being both the starting point for fatty acid synthesis, as well as a key step in the 3HP pathway used in several carbon-fixation pathways. All these carboxylation reactions thus have in common that they are used at least in some organism as the central source for cellular carbon. All other carboxylation reactions that are not used as part of core carbon uptake, are used in the synthesis of the biotin cofactor, and the purine and pyrimidine nucleotides (see figure 17).

If we consider the sequences in which these carboxylation reactions are used to synthesize biotin, purine and pyrimidine, they also form a distinct class of chemistry. In all three cases
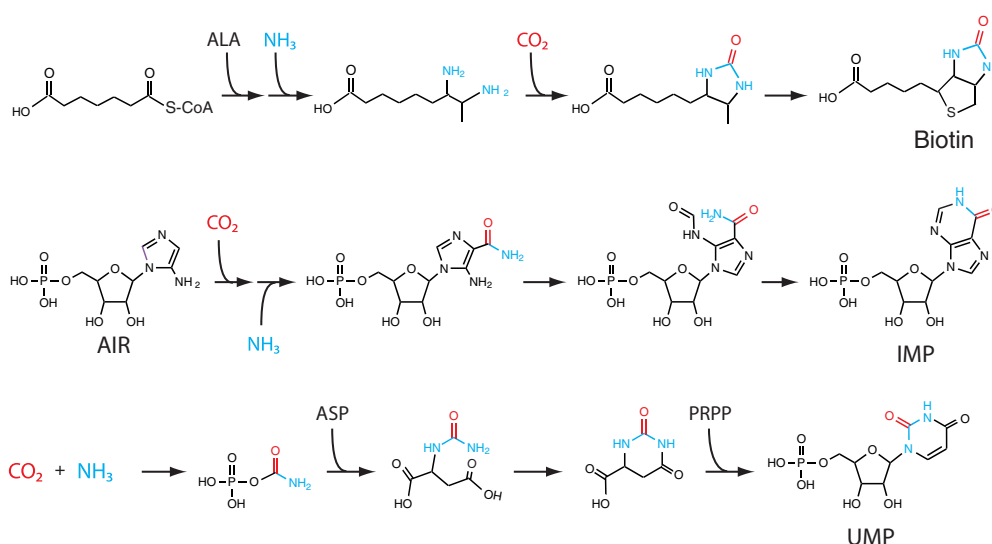
Figure 17. Carboxylation reactions in the synthesis of cofactors. The sequences show the immediate amination of the carboxyl group to a carboxamide group, which is then preserved into the final heterocyclic structure. As the only carboxylations not used in core carbon uptake, these reaction sequences form a distinct class of chemistry. Amination reactions are shown as net additions of ammonia, which may be derived from other sources (such as glutamine, aspartate or SAM). Abbreviations: alanine (ALA); aspartate (ASP); phosphoribosyl pyrophosphate (PRPP).

the resulting carboxyl group is immediately aminated, either as part of the carboxylation reaction, or in the following reaction, and the carboxamide group is subsequently maintained into the final heterocyclic structure. In addition we previously saw that IMP becomes the source for the folate and flavin family (through GTP). Carboxylation reactions are thus either a general source for cellular carbon in core metabolism, or a specific source of carboxamide groups in the synthesis of cofactors that are part of the catalytic control of core metabolism.

4.5. The chorismate pathway in both amino acid and cofactor synthesis

Chorismate is the sole source of single benzene rings in biochemistry [229]. The non-local π-bond resonance is used in a variety of charge-transfer and electron transfer and storage functions, in functional groups and cofactors derived from chorismate. We have noted the charge-transfer function of PABA in tuning N^{10} of folates, and its impact on C_1 chemistry. The para-oriented carbonyl groups of quinones may be converted to partially- or fully-resonant orbitals in the benzene ring, enabling fully oxidized (quinone), half-reduced (semiquinone), or fully reduced (hydroquinone) states [235]. Finally, the aromatic ring in tryptophan (a second amino acid which behaves in many ways like a cofactor) has at least one function in the active sites of enzymes as a mediator of non-local electron transfers [258].

5. Innovation: promiscuous catalysis, serendipitous pathways

The previous sections argued for the existence of low-level chemical and cofactor/catalyst constraints on metabolic innovations, and presented evolutionary divergences that either respected these as constraints, or were enabled by

the diversification of cofactor and catalytic functions. In this section we consider the dynamics by which innovation occurs, and its main organizing principles. Innovation in modern metabolism occurs principally by duplication and divergence of enzyme function [145, 259, 260]. Often it relies on similarity of functions among paralogous enzymes, but in some cases may exploit more distant or accidental overlap of functions.

Innovation always requires some degree of enzymatic promiscuity [145], which may be the ability to catalyze more than one reaction (catalytic promiscuity) or to admit more than one substrate (substrate ambiguity). Pathway innovation also requires *serendipity* [261], which refers to the coincidence of new enzymatic function with some avenue for pathway completion that generates an advantageous phenotype from the new reaction. Although most modern enzymes are highly specific, broad substrate-specificity is no longer considered rare, and is even explained as an expected outcome in cases where costs of refinement are higher than can be supported by natural selection, and in other cases by positive selection for phenotypic plasticity [260, 262]. However, when enzymes are specific—whether due to structure or due to evolved regulation—they are of necessity diversified in order to cover the broad range of metabolic reactions used in the modern biosphere. Serendipitous pathways assembled from a diversified inventory of specific enzymes will in most cases be strongly historically contingent as they depend on either overlap of narrow affinity domains or on 'accidental' enzyme features not under selection from pre-existing functions. Such pathways therefore seem unpredictable from first principles; whether they are rare will depend on the degree to which the diversity of enzyme substrate-affinities compensates for their specificity.

A key question for early metabolic evolution is whether the trade-off between specificity and diversity was different in the deep past than in the present, either in degree or in

structure, in ways that affected either the discovery of pathway completions or the likelihood that new metabolites could be retained within existing networks. These structural aspects of promiscuity and serendipity determine the regulatory problem faced by evolution in balancing the elaboration of metabolism with its preservation and selection for function.

5.1. Creating reaction mechanisms and restricting substrates, while evolving genes

Metabolism is characterized at all levels by a tension between creating reaction mechanisms that introduce new chemical possibilities, and then pruning those possibilities by selectively restricting reaction substrates. Whether this tension creates a difficult or an easy problem for natural selection to solve depends at any time on whether the accessible changes in catalytic function, starting from integrated pathways, readily produce new integrated pathways whose metabolites can be recycled in autocatalytic loops. We argue that the conservation of pathway mechanisms, particularly when these are defined by generic functional groups such as carboxyls, ketones and enols, with promiscuity coming from substrate ambiguity with respect to molecular properties away from the reacting functional group, favors the kind of orderly pathway duplication that we observe in the extant diversity of core metabolism. Therefore we expect that serendipitous pathway formation was both facile in those instances in the early phases of metabolic evolution where innovations in radical-based mechanisms for carbon incorporation occurred, and structured according to the same local-group chemistry around which the substrate network is organized.

Modern enzymes both create reaction mechanisms and restrict substrates, but the parts of their sequence and structure that are under selection for these two categories of function may be quite different, so the two functions can evolve to a considerable degree independently. Active-site mechanisms in enzymes for organic reactions will often depend sensitively on a small number of highly conserved catalytic residues in a relatively fixed geometry, while substrate selection can depend on a wide range of properties of enzyme shape or conformation dynamics [260], on local functional-group properties of the substrate that have been termed 'chemophores' [263], as well as (in some cases) on detailed relations between the substrate and active-site geometry or residues. An extreme example of the potential for separability between reaction mechanism and substrate selection is found in the polymerases. A stereotypical reaction mechanism of attack on activating phosphoryl groups requires little more than correct positioning of the substrates. In the case of DNA polymerases, at least six known categories (A, B, C, D, X and Y) with apparently independent sequence origin have converged on a geometry likened to a 'right hand' which provides the required orientation [264, 265].

At the same time as evolving enzymes needed to provide solutions to the biosynthetic problem of enabling and regulating metabolic network expansion, they were themselves dependent on the evolving capabilities of genomic and translation systems for maintaining complexity and diversity. Jensen [259] originally argued that high enzymatic specificity was no more plausible in primitive cells than highly

diversified functionality[8], and that enzymatic promiscuity was both evolutionarily necessary and consistent with what was known at that time about substrate ambiguity and catalytic promiscuity. Modern reviews [145, 260, 263] of the mechanisms underlying functional diversity, promiscuity and serendipity confirm that substrate ambiguity is the primary source of promiscuity that has led to the diversification of enzyme families. It is striking that, even in cases where substrate affinity has been the conserved property while alternate reaction mechanisms or even alternate active sites have been exploited, it is often local functional groups on one or more substrates that appear to determine much of this affinity [260].

5.2. Evidence in our module substructure that early innovation was governed principally by local chemistry

The substructure of modules, and the sequence of innovations, we have sketched in section 3 appears to be dominated by substrate ambiguity in enzymes or enzyme families with conserved reaction mechanisms. The key reactions in carbon fixation are of two types: crucial reactions typically involve metal centers or cofactors that could have antedated enzymes, and it is primarily reaction sites, not molecular selectivity, that distinguishes pathways at the stage of these reactions. Recall that the enzymes that have been argued to be the ancestral forms of both the acetyl- and succinyl-CoA ligases and the pyruvate and α-ketoglutarate biotin-carboxylases show very close sequence homology [110, 112], suggesting shared ancestral enzymes for both. The shared internal sequence of reductions and isomerizations common to modules (figure 9) are very broadly duplicated, and the molecular specificity in their enzymes today is not correlated with significant reaction-sequence changes in the internal structure of pathways. These pathways could plausibly function much as they do today with less-specific hydrogenases and aconitases.

A quantitative reconstruction of early evolutionary *dynamics* will require merging probability models for networks and metabolic phenotypes with those for sequences and structure of enzyme families. The goal is a consistent model of the temporal sequence of ancestral states of catalyst families, and of the substrate networks on which they acted.

6. Integration of cellular systems

The features of metabolism that display a 'logic' of composition, which is then reflected in their evolutionary history, are those with few and robust responses to environmental conditions that can be inferred from present diversity. These are the subsystems whose evolution has

---

8 This argument was largely a rebuttal of an earlier proposal by Horowitz [266] for 'retrograde evolution' of enzyme functions. The 1940s witnessed the rise of an overly-narrow interpretation of 'one gene, one enzyme, one substrate, one reaction' (a rigid codification of what would become Crick's *Central Dogma* [267]), which in the context of complex pathway evolution appeared to be incompatible with natural selection for function of intermediate states. The Horowitz solution was to depend on an all-inclusive 'primordial soup' [268], in which pathways could grow backward from their final products, propagating selection stepwise downward in the pathway until a pre-existing metabolite or inorganic input was found as a pathway origin.

been simplified and decoupled by modularity. Their relative immunity from historical contingency, resulting in more 'thermodynamic' modes of evolution, results from rapid, high-probability convergence in populations that can share innovations [172].

The larger roles for standing variation and historical contingency that are so often emphasized [269] in accounts of evolutionary dynamics are made possible by longer-range correlations that link modules, creating mutual dependences and restricting viable changes [137, 138]. The most important source of such linkage in extant life is the unification of metabolic substrates and control processes within cells [270]. Cellular death or reproduction couples fitness contributions from many metabolic-phenotype traits, together with genome replication systems. This enables the accumulation of diversity as genomes capture and exploit gains from metabolic control, complementary specialization [271] and the emergence of ecological assemblies of specialists as significant mediators of contingent aspects of evolutionary innovation (as we illustrated with the example of methylotrophy).

We consider in this section several important ways in which aggregation of metabolic processes within cells follows its own orderly hierarchy and progression. We note that even a single cell does not impose only one type of aggregation, but at least three types, and that these are the bases for different selection pressures and could have arisen at different times. Within cellular subsystems, the coupling of chemical processes is often mediated by coupling of their energy systems, which has probably developed in stages that we may be able to identify. Finally, even where molecular replication is coupled to cellular physiology, in the genetic code, strong and perhaps surprising signatures of metabolic modularity are recapitulated.

6.1. Cells provide at least three functionally distinct forms of compartmentation

Under even the coarsest functional abstraction, the cell provides not one form of compartmentation, but at least three [272, 273]. The geometry and topology of closed spheres or shells, and the capacitance and proton impermeability of lipid bilayers, permit the buildup of pH and voltage differences, and thus the coupling of redox and phosphate energy systems through intermediate proton-motive (or in many cases, sodium-motive) force [274]. The concentration of catalysts with substrates enhances reactions that are second-order in organic species, and the equally important homeostatic control of the cytoplasm regulates metabolic reaction rates and precludes parasitic reactions. Finally, the cell couples genetic variations to internal biochemical and physiological variations much more exclusively than they are coupled to shared resources such as biofilms or siderophores, leading to the different evolutionary dynamics of development from niche construction [49].[9] The perspective that this is an active coupling, which defines one of the forms of *individuality*

---

[9] For an argument that somatic development and niche construction are variants on a common process, distinguished by the genome's level of control and exploitation of the constructed resources, see [275].

rather than providing a complete characterization of the nature of the living state, is supported by the complex ecosystems including viral RNA and DNA that are partly autonomous of the physiology of particular cells [276, 277]. Each of these different forms of coupling affects the function and evolution of the modules we have discussed.

6.1.1. Coupling of redox and phosphate energy systems may have been the first form of compartmentation selected.
Biochemical subsystems driven, respectively, by redox potential or phosphoanhydride-bond dehydration potential, cannot usually be directly coupled to one another due to lack of 'transducer' reactions that draw on both energy systems. In addition to the ultimate physical constraint of limits to free energy, biochemistry also operates under additional proximate constraints from the chemical and quantum-mechanical substrates in which that free energy it is carried. The notable exception to the general lack of direct coupling between energy systems is the exchange of phosphate and sulfur groups in substrate-level phosphorylation [235] from thioesters (which may proceed in either direction depending on conditions). Although it provides a less flexible mode of coupling than membrane-mediated oxidative phosphorylation, this crucial reaction type, which occurs in some of the deepest reactions in biochemistry (those employing CoA, including all those in the six carbon fixation pathways), has been proposed as the earliest coupling of redox and phosphate [241], and the original source of phosphoanhydride potential [75] enabling pathways that require both reduction and dehydration reactions.

Phosphate concentration limits growth of many biological systems today, and phosphate concentrations appear to be even lower in vent fluids [278] than on average in the ocean, making it difficult to account for the emergence of many metabolic steps in hydrothermal vent scenarios for the origin of life. Serpentinization and other rock–water interactions that produce copious reductants—and are believed to have been broadly similar at least from the early archean to the present [74, 279]—also scavenge phosphates into mineral form. Unless new mechanisms are discovered that could have produced an increased amount of phosphate for early vents, it thus appears doubtful that phosphates were abundant in the environments otherwise most favorable to geochemical organosynthesis. What little phosphate is found in water is primarily orthophosphate, because the phosphoanhydride bond is unstable to hydrolysis. Therefore the retention of orthophosphate, and the continuous regeneration of pyrophosphate and polyphosphates [280–284], may have been essential to the spread of early life beyond relatively rare geochemical environments.

The membrane-bound ATP-synthetase, which couples phosphorylation to a variety of redox reactions [5] through proton or sodium pumping, is therefore essential in nearly all biosynthetic pathways, and must have been among the first functions of the integrated cell. Without a steady source of phosphate esters, none of the three oligomer families could exist. The ATP synthetase itself is homologous in all organisms, providing one strong argument (among many [109, 173]) for a membrane-bound last common ancestor.

Proton-mediated phosphorylation (best known through oxidative phosphorylation in the respiratory chain [235]) requires a topologically enclosed space to function as a proton capacitor [274]. However, as shown by gram-negative bacteria [5] and their descendants mitochondria and plastids, which acidify the periplasmic space or thylakoid lumen, the proton capacitor need not be (and generally is not) the same compartment as the cytoplasm containing enzymatic reactions. Because the coupling of energy systems is a different function from regulating reaction rates catalytically, the phosphorylation system should not generally have been subject to the same set of evolutionary pressures and constraints as other cellular compartments, and need not have arisen at the same time. We note that, because it may have lower osmotic pressure than the cytoplasm, the acidified space required for proton-driven phosphorylation may not have required a cell wall, greatly simplifying the number of concurrent innovations required for compartmentalization, compared to those for the cytoplasm. Therefore we conjecture that proton-mediated phosphorylation could have been the first function leading to selection for lipid-bilayer compartmentalization, allowing other cellular functions to accrete at later times.

6.1.2. Regulation of biosynthetic rates may have been prerequisite for the optimization of loop-autocatalytic cycles.

The second function of cellular compartments, and the one most emphasized in vesicle theories of the origin of life [6, 285, 286], is the enhancement of second-order reactions by collocation of catalysts and their substrates. Here we note another role that we have not seen mentioned, which is more closely related to the functions of the cell that *inhibit* reactions. Organisms employing autocatalytic-loop carbon fixation pathways must reliably limit their anabolic rates to avoid drawing off excess network catalysts into anabolism, resulting in passage below the autocatalytic threshold for self-maintenance, and collapse of carbon fixation and metabolism. Regulating anabolism to maintain viability and growth may have been an early function of cells.

We noted in section 3.4.4 the fragility of autocatalytic-loop pathways to parasitic side-reactions, and the way the addition of a linear pathway such as WL stabilizes loop autocatalysis in the root node of figure 11. For proto-metabolism, spontaneous abiotic side-reactions may be hazardous, if catalysts in the main fixation pathway do not sufficiently accelerate their reaction rates, creating a separation of timescales relative to the uncatalyzed background. Within the first cells, the same hazard is posed by secondary anabolism, as its reaction rates become enhanced by catalysts similar to those in the core. This fact was clearly noted already in [147]. It may thus be that the optimizations in either branch of the carbon-fixation tree were not possible until rates of anabolism were sufficiently well-regulated to protect supplies of loop intermediates or essential cofactors. Therefore, while the root node is plausible as a pre-cellular [147] or an early cellular (but non-optimized) form, either branch from it may have required the greater control afforded by quite refined cellular regulation of reaction rates. It is here that we envision a crucial role for feedback regulation at the genomic level

[24, 25] as a support for the *architectural stability* of the underlying substrate network, prior even to its service in homeostasis in complex environments or in phenotypic plasticity.

6.2. Coupling of metabolism to molecular replication, and signatures of chemical regularity in the genetic code

Among the subsystems coupled by modern cells, perhaps none is more elaborate than the combined apparatus of amino acid and nucleotide biosynthesis and protein coding. The most remarkable chemical aspect of the protein-coding system is that it is an *informational* system: a sophisticated machinery of transcription, tRNA formation and aminoacylation, and ribosomal translation separates the chemical properties of DNA and RNA from those of proteins, permitting almost free selection of sequences in both alphabets in response to requirements of heredity and protein function[10]. The interface at which this separation occurs is the genetic code. From the informational suppression of chemical details that defines the coding system, the code itself might have been expected to be a random map, but empirically the code is known to contain many very strong regularities related to amino acid biosynthesis and chemical properties, and perhaps to the evolutionary history of the aminoacyl-tRNA synthetases.

Many explanations have been advanced for redundancy in the genetic code, as a source of robustness of protein properties against single-point mutations [172, 287–289], but in all of these the source of selection originates in the elaborate and highly evolved function of coding itself. In many cases the redundancy of amino acids at adjacent coding positions reflects chemical or structural similarities, consistent with this robustness-criterion for selection, but in nearly all cases redundancy of bases in the code correlates even more strongly with shared elements of *biosynthetic pathways* for the amino acids. The co-evolutionary hypothesis of Wong [290] accounts for the correlation of the first base-position with amino-acid backbones as a consequence of duplication and divergence of amino acid biosynthetic enzymes together with aminoacyl-tRNA synthetases (aaRS). The stereochemical hypothesis of Woese [291] addresses a correlation of the second coding position with a measure of hydrophobicity called the *polar requirement*. The remarkable fact that both correlations are highly significant relative to random assignments, but that they are *segregated* between first and second codon bases, is not specifically addressed in either of these accounts. Copley *et al* [237] address the same regularities as both the Wong and Woese hypotheses, but link them to much more striking redundancies in biosynthetic pathways, which they propose are consequences of small-molecule organo-catalytic roles of dimer RNA in the earliest biosynthesis of amino acids.

We note here a further chemical regularity in the genetic code, which falls outside the scope of the previous explanations, and possibly relates to the biosynthetic pathways of the purine cofactors as discussed in section 4.2. This regularity concerns triplet codons with purines at the second position, and takes one of two forms. Several amino acids

---

[10] The observation that enzymes acting on DNA have evolved to actively mitigate chemical differences in the bases, to enable a more nearly neutral combinatorial alphabet, is due to Schuster [311].

that use GTP-derived cofactors in their biosynthetic pathways are associated with triplet codons containing guanine at the second position, while another amino acid (histidine) that in its synthesis is directly derived from ATP is associated with triplet codons containing adenine at the second position. This association is much more comprehensive for G-second codons than for A-second codons, and it does not suggest the same kinds of mechanistic relations in the two cases. However, it further compresses the description of patterns in the code that were not addressed in [237], in terms of similar chemical and biosynthetic associations.

The correlation between the glycine cycle for amino acid biosynthesis from C_1 groups on folate cofactors, and codons XGX, where X is any base and G is guanine is strong. (In what follows we abbreviate wobble-base positions y for pyrimidines U and C, or u for purines A or G.) This group includes glycine (GGX), serine (AGy), cysteine (UGy), and tryptophan (UGu)[11]. We do not propose a specific mechanism for such an association here, but our earlier argument that folates would have been contemporaneous with GTP suggests that biosynthesis through the glycine cycle was the important source of these amino acids at the time they became incorporated into the code. Some of these amino acids satisfy multiple regularities, as in the correlation of glycine with GXX $\leftrightarrow$ reductive transamination, or cysteine with UXX $\leftrightarrow$ pyruvate backbone, proposed in [237].

The position (CAy) of histidine, synthesized from ATP, is the only case we recognize of a related correlation in XAX codons. For this position, the availability of ATP seems to have been associated with the synthesis of histidine directly through the cyclohydrolase function (rather than through secondary cofactor functions), at the time this amino acid became incorporated into the code.

Much more than correlation is required to impute causation, and all existing theories of cause for regularities in the genetic code are either highly circumstantial or require additional experimental support. Therefore we limit the aspects of these observations that we regard as significant to the following three points:

The existence of a compression. The idealized adaptive function of coding is to give maximum evolutionary plasticity to aspects of phenotype derived from protein sequence, uncoupled from constraints of underlying biosynthesis. The near-wholesale transition from organic chemistry to polymer chemistry around the C_{20} scale suggests that this separation has been effectively maintained by evolution. Strong regularities which make the description of the genetic code *compressible* relative to a random code reflect *failures* of this separation which have transmitted selection pressure across levels, during either the emergence or maintenance of the code. These include base-substitution errors, whether from mutation or in the transcription and translation processes, but also apparently chemical relations between nucleobases and amino acids.

The segregation of the roles of different base positions and in some cases different bases in terms of their biochemical correlates. The genetic code is like a 'rule book' for steps in the biosynthesis of many amino acids, but the chemical

---

[11] Both purines are used in the mitochondrial code and only UGG is used in the nuclear code.

correlations which are its rules are of many kinds. The correlations in the code may be understood as rules because the biosynthetic pathways may be placed on a decision tree, with branches labeling alternative reactions at several stages of synthesis, and branching directions indicated by the position-dependent codon bases [237]. Beyond the mere existence of those rules, and their collective role as indices of regularity threading the code, we must explain why rules of different kinds are so neatly *segregated* over different base positions and sometimes over different bases (as in the XGX and XAX codons).

A compression that references process rather than property. the role of biosynthetic pathways as correlates of regularities makes this compression of the genetic code a reference to the *process* and metabolic network context within which amino acids are produced, and not merely to their *properties*. (Many of the chemical properties recognized as criteria of selection, whether size or hydrophobicity, are shared at least in part because they result from shared substrates or biosynthetic steps.) We think of the function of coding as separating biosynthetic process from phenotype: transcription and translation are 'Markovian' in the sense that the only information from the biosynthetic process which survives to affect the translated protein is what is inherent in the structure of the amino acid. In technical terms, one says that the phenotype is *conditionally independent* of the biosynthetic pathway, *given* the amino acid. Thus selection on post-translation phenotypes should only be responsive to the finished amino acids. The existence of regularities in the genetic code which show additional correlation with intermediate steps in the biosynthetic process therefore requires either causes other than selection on the post-coding phenotype (including its robustness), or a history-dependence in the formation of the code that reflects earlier selection on intermediate pathway states. If they reflect causal links to metabolic chemistry, these 'failures' of the separation between biosynthetic constraint and selection of polymers for phenotype may have broken down the emergence of molecular replication into a sequence of simpler, more constrained, and therefore more attainable steps.

7. The extrapolation of metabolic logic to questions of emergence

Comparative analysis and its formal extension to diachronic reconstruction simultaneously estimate two properties of systems: a model for a *generating process* including constraints or laws that have operated over the system's history, and a collection of *idiosyncratic or historically contingent events* that make each history distinct and are not assumed to be reproducible or predictable. In our review up to this point the reconstructed period has been the genomic era, and snapshots of ancestral states in this era do not directly carry information about pre-LUCA or pre-cellular forms except through constraints that we can argue were common both pre- and post-LUCA. The surprising feature of the compositional and evolutionary logic of metabolism that we have sketched is how much of it apparently reflects constraints

from low-level organic or organometallic chemistry that are not distinctively biological, or homologies and energetic contacts with geochemistry that were arguably broadly continuous through the emergence of life [32, 116, 279]. In several places prebiotic scenarios have been mentioned as interpretive frames for our observations, so here we summarize which features of the logic of metabolism we think are strong constraints on theories of the emergence of life.

7.1. Autotrophic versus heterotrophic origins

In section 2.1 we summarized reasons to regard carbon fixation from CO_2 and anabolism as both the ancestral pathways of cellular metabolism and the set of prior constraints around which catabolic pathways, and the diverse array of heterotrophic metabolisms they enable, subsequently evolved. The evidence in favor of this view is highly multi-factorial, including historical reconstruction, inferred geochemical context, pathway chemistry and degree of universality. The most striking property of this evidence is that it leads to quite specific inferences about ancestral autotrophic phenotypes, whereas we do not know of comparable proposals about ancestral heterotrophy that are similarly specific and that unify a similar diversity of observations.

It is then natural to ask: was the early role of autotrophy and anabolism in the cellular era a continuation of geochemical processes of similar character, or was it the outcome of a reversal of earlier pre-LUCA or pre-cellular metabolisms fed by organics from pathways unrelated to those in extant biochemistry? The distinction is not quite the same as that between autotrophic and heterotrophic organisms (though these terms are often borrowed) [58]. Rather, it is a distinction between a hypothesis of continuity with geochemistry which was gradually brought under autonomous control of bio-organic chemistry, and a hypothesis of discontinuity requiring that early organisms have evolved the mechanisms and networks of biochemistry *de novo*.

The arguments for geochemical continuity [6, 7, 33, 77, 122, 292] are founded first on detailed accounts of the capacity of a range of geochemical energy systems to support extant life [31, 34]. A subset of the entries in table 1 of [31], involving Fe^{2+} reduction or autotrophic methanogenesis, can be applied directly to early-earth environments. (Note, however, that many entries in their table of environments involve sulfates, nitrates, ferric iron, or small amount of molecular oxygen (the Knallgas reaction) as terminal electron acceptors. The breadth of organic conversions detailed in the paper is meant to provide a basis for habitability analysis today, so plausible pathways in the Hadean must be understood as having been limited by the available terminal electron acceptors.) Where the continuity hypothesis supposes that extant life has 'enfolded' prior geochemical mechanisms, it cites detailed similarities between transition-metal/sulfide mineral unit cells and metallo-enzyme active sites [74, 149, 279], which may reflect [254] mineral precursors to the widespread use of radical mechanisms in reactions catalyzed by metal-center enzymes [150], as we have mentioned previously. The richness of hydrothermal vent environments in particular, in geometry, surface catalysis

[122, 227, 292], thermal and pH gradients, and the overall similarity of the aqueous redox environment of hydrothermal fluids to biochemistry [33, 94, 293, 294], provides specific locations where catalysis and also other functions such as containment or selective diffusion would have been provided. Finally, the geochemical hypothesis has been circumstantially supported by experimental evidence that minerals can catalyze reactions in the citric-acid cycle [97], and an extensive range of reductions [246, 295], including synthesis of acetyl-thioesters [147], which for a variety of reasons we have noted in this review are among the most-central compounds of core metabolism.

The specificity of the links which the continuity hypothesis is in a sense *required* to propose derives from the very restrictive boundary condition of CO_2 as sole carbon source, the same constraint that permitted specific claims in our reconstruction of cellular autotrophy. Our approach of gathering formal evidence about the structure and strength of constraints, and of testing these for consistency within both organism physiology and ecology, is very similar in spirit to the approach of [136] to a 'breakout organism' from the RNA world. However, we aim at fewer and chemically lower-level facts that plausibly reach further back to pre-cellular or pre-RNA times. (The details reconstructed in [136] are also very compatible with our reconstructions of early carbon fixation, and we regard our proposal of an even earlier role for some cofactors than for RNA base pairing to be very much in the general spirit of their arguments.)

In contrast, heterotrophic-origins stories are largely objections to problems with geochemical organosynthesis *and selection* requiring 'something else' in its place. They may be quite unrestrictive about what the original organic inputs were, as in the original proposals of Oparin [296] and Haldane [268]. Their most restrictive quantitative constraints (such as pathway minimization) may not directly determine pathway direction [71], and may presume an optimization problem different from the one performed if the molecular inventory was not pre-fixed. They may also show only limited overlap with extant biochemistry [297], without suggestions for how missing components were filled in or abiotic components were pruned.

7.2. The joint WL/rTCA network as a pre-cellular form

The importance of balancing considerations of accessibility and robustness with selectivity in incipient and early-cellular biochemistry were mentioned in section 3.4.4 and section 4.2. As a solution to the problem of reconstructing history, the root WL/rTCA network was put forward as a quantitative example in which the consequences of topology for both robustness and selectivity could be analyzed. That treatment was essentially backward-looking, asking how well our proposed root node meets multiple criteria required by inference from the present, such as pathway distribution, plausible causes for innovations, and selection of the extant biosynthetic precursors. The converse question in the pre-cellular era is dynamical and echos Leibniz's question 'Why is there something instead of nothing?'.

In its earliest forms, such a joint network would be presumed to have mineral or perhaps soluble metal–ligand catalysts for both direct C_1 reduction and rTCA cycling, perhaps already showing the distinctions between the functions of the nitrogenous cofactors for C_1 reduction and the functions of sulfur-containing cofactors in rTCA. We may ask, would a hybrid pathway out-compete alternatives *chemically* as a kinetic channel for carbon reduction by H_2 (or perhaps directly by reduced iron)? To this we argue that a feeder augmented by a loop outcompetes an unaided feeder on average by virtue of autocatalytic self-amplification [298]. A loop with a feeder outcompetes a bare loop in the context of loss or fluctuations because of greater robustness and recovery (self-re-ignition). The important observation is that chemical selection already shows features common to Darwinian selection: fitness can come both from average behavior and from stability under perturbations, and different components of a pathway may provide different elements of fitness.

7.3. A synthetic description: geochemistry, the metabolic substrate and catalysis

While many mechanisms and components—particularly catalysts [89]—must have been replaced in a sequence leading from prebiotic geochemistry to the earliest cellular biochemistry, the three elements we have emphasized of a logic of metabolism should have remained continuous across the transition. We favor scenarios in which chemical networks at the aggregate scale of the biosphere originated in an abundant supply of CO_2 and H_2, containing driving redox potential resulting in an accumulation of stress, which became coupled to a robust concentration mechanism within organic chemistry, forming networks that in turn became increasingly stable with the emergence of intermediary metabolism and the appearance of complex cofactors and additional long-loop feedback mechanisms.

8. Conclusions

We have argued that the fundamental problem of electron transfer in aqueous solution leads to a qualitative division between catalytically 'hard' and 'easy' chemistry, and that this division in one form or another has led to much of the architecture and long-term evolution of metabolism and the biosphere. Hard chemistry involves electron transfers whose intermediate states would be unstable or energetically inaccessible in water if not mediated by transition-metal centers in metal–ligand complexes and/or elaborate and structurally complex organic cofactors. Easy chemistry involves hydrogenations and hydrations, intramolecular redox reactions and a wide array of acid–base chemistry. Easy chemistry is promiscuously re-used and provides the internal reactions within modules of core metabolism. Hard chemistry defines the module boundaries and the key constraints on evolutionary innovation. These simple ideas underlie a modular decomposition of carbon fixation that accounts for all known diversity, largely in terms of unique adaptations to chemically simple variations in the abiotic environment. On the foundation of core metabolism laid by carbon fixation, the remainder of biosynthesis is arranged as a fan of increasingly independent anabolic pathways. The unifying role of the core permits diverse anabolic pathways to independently reverse and become catabolic, and the combinatorics of possible reversals in communities of organisms determines the space of evolutionary possibilities for heterotrophic ecology.

We have emphasized the role of feedback in biochemistry, which takes different forms at several levels. Network autocatalysis, if we take as a separate question the origin of external catalytic and cofactor functions, is found as a property internal to the small-molecule substrate networks for many core pathways. A qualitatively different form of feedback is achieved through cofactors, which may act either as molecular or as network catalysts. As network catalysts they differ from small metabolites because their internal structure is not changed except at one or two bonds, over the reactions they enable. The cofactors act as 'keys' that incorporate domains of organic chemistry within biochemistry, and this has made them both extraordinarily productive and severely limiting. No extant core pathways function without cofactors, and cofactor diversification appears to have been as fundamental as enzyme diversification in some deep evolutionary branches. We have therefore argued for a closely linked co-evolution of cofactor functions with the expansion of the universal metabolic network from inorganic inputs, and attempted to place key cofactor groups within the dependency hierarchy of biosynthetic pathways, particularly in relation to the first ability to synthesize RNA.

The most important message we hope to convey is the remarkable imprint left by very low-level chemical constraints, even up to very high levels of biological organization. Only seven carbon fixation modules, mostly determined by distinctive, metal-dependent carboxylation reactions, cover all known phylogenetic diversity and provide the building blocks for both autotrophic and heterotrophic metabolic innovation. A similar, small collection of organic or organometallic cofactor families have been the gateways that determine metabolic network structure from the earliest cells to the present. The number of these cofactors that we consider distinct may be somewhat further reduced if we recognize biosynthetic relatedness that leads to functional relatedness (as in the purine-derived or chorismate-derived cofactors), or cases of evolutionary convergence dominated by properties of elements (as for lipoic acid and the CoB-CoM system).

We believe that these regularities should be understood as laws of biological organization. In a proper, geochemically-embedded theory of the emergence of metabolism, such regularities should be predictable from the properties of the underlying organic chemistry. As our understanding of relevant organic chemistry continues to expand, particular forms, such as distinctive metal chemistry or convergent uses of nitrogen and sulfur, should become predictable from their distinct catalytic properties. Properties of distributions, as in the use of network modules or the diversity of cofactors, should in turn be predictable from asymmetries in catalytic constraints that are likely to arise within a large and diverse possibility space of organic chemistry. Moreover, this lawfulness should

have been expected: the factors that reduce (or encrypt) the role of laws in biology, and lead to unpredictable historical contingencies, arise from long-range correlation. Correlation of multiple variables leads to large spaces of possibility and entangles the histories of different traits, making the space difficult to sample uniformly. But correlation in biology is in large part a *constructed* property; it has not been equally strong in all eras and its persistence depends on timescales. Long-term evolution permits recombination even in modern integrated cells and genomes. Early life, in contrast, with its less-integrated cells and genomes, and its more loosely-coupled traits, had constructed less long-range correlation. These are the domains where the simpler but invariant constraints of underlying chemistry and physics should show through.

Appendix A. Glossary of some terms used in the text

Term	Usage
Carbon fixation	Any process by which organisms convert CO_2 (or another inorganic one-carbon source such as bicarbonate or formic acid) into molecules possessing C–C bonds. All biosynthesis rests ultimately on carbon fixation, because the biosphere does not rely on organic carbon from abiotic sources.
Anabolism	Biochemical processes that build up molecule size. We will be concerned particularly with buildup by reactions of organic chemistry (as opposed to phosphate-driven polymerization, which is chemically simple and homogeneous). Anabolism is a net consumer of reductants, as biomass is more reduced than its input CO_2.
Catabolism	Biochemical processes that break down organic molecules taken as inputs. The breakdown may provide energy or biosynthetic precursors to other anabolic reactions.
Autotrophy	A self-sufficient mode of metabolism in which all biomolecules can be synthesized using CO_2 as sole carbon source. In strict usage the term denotes self-sufficiency of the metabolic network of an *organism*. For purposes of understanding the constraints implied by flux-balance analysis, we will extend the scope to include appropriate consortia of organisms.
Heterotrophy	A mode of metabolism in which the focal network must draw carbon from some organic source, because it lacks necessary reactions to synthesize some essential metabolites starting from CO_2.
Compositional logic	Principles of assembly of metabolic networks in organisms or consortia which capture regularities in the structure of the resulting networks, or of their dynamics which are responsible for phenotype or ecological role.
Evolutionary logic	Principles of selection or constraints which compactly express regularities in evolutionary branching and relate these to aspects of phenotype which may have constrained innovations or determined fitness.
Autocatalysis	A property of reaction networks, that intermediates or outputs of the reaction system act to catalyze earlier reactions in the system (their own biosynthetic pathways or others), leading to self-amplification of the reaction fluxes. Autocatalysis may be provided by individual molecules such as enzymes, or may result from completing network cycles that connect inputs to outputs while regenerating network intermediates. For formalizations see [85, 86].
Catalytic control	The bringing-into-existence, or the regulation, of particular input–output characteristics of a reaction system through introduction, or control of the concentrations, specificities or activities, of catalysts for its reactions. Control may be through mutation, concentration, physical location, multiple-unit interactions or allosteric regulation. Cofactors as well as enzymes may be control elements over catalysis. Less directly, assembly of catalysts for several reactions to form a network may lead to new input-output characteristics through the formation of network-catalytic pathways.
Topological modularity	Used interchangeably with 'network modularity' in this review. A property of the connectivity in a network that permits its decomposition into a collection of *clusters* or *communities*, with greater link density among members within a community than between members in distinct communities. Many measures such as *network modularity* [21] or Girvan–Newman community detection based on betweenness centrality [299] may be used. Examples: a link whose removal separates the pentose-phosphate network from the core network containing universal biosynthetic precursors is the synthesis of 3-phosphoglycerate from phosphoenolpyruvate in gluconeogenesis; a node whose removal decomposes many loop carbon-fixation pathways is succinyl-CoA.
Robustness	Preservation of some property of structure or function under incident perturbations. These may be external, such as concentration fluctuations, or internal, such as removal of a reaction or reduction in its flux due to fluctuation in concentration of a catalyst. A property that autotrophic systems require to be robust is the ability to *produce* all members in a key set of metabolites with CO_2 as the only carbon source.
Maximum parsimony	A criterion for constructing trees of relatedness that minimizes repeated instances of the same innovation over links. Strict parsimony is well-defined, but ranking among solutions that do require some repeated instances is not defined by the parsimony criterion alone. In practice, ordering of solutions by parsimony is often accompanied by judgments of the probability penalty that a richer method such as maximum-likelihood or Bayesian reconstruction would attach to repeated innovations.
Hypergraph	(Used in the appendix) A generalization from the concept of a graph. The edges in hypergraphs (called *hyperedges*) possess sets of nodes as their boundary, rather than pairs of nodes as for simple graphs. Directed hypergraphs are necessary to capture the stoichiometric relations of chemical reaction networks. See [300].

Appendix B. Bipartite graph representations for chemical reaction networks

The stoichiometry of a chemical reaction may be represented by a *directed hypergraph* [300]. A hypergraph differs from a simple graph in that, where each edge of a simple graph has two points as its boundary, in a hypergraph, a hyper-edge may have a set of points as its boundary. In a directed hypergraph, the input and output sets in the boundary are distinguished. For the application to chemistry presented here, each hyper-edge corresponds to a reaction, and its input and output boundary sets correspond to moles of the reactant and product molecules.

It is possible to display the hypergraphs representing chemical reactions as *doubly-bipartite* simple graphs, meaning that both nodes and edges exist in two types, and that well-formed graphs permit only certain kinds of connections of nodes to edges. The bipartite graph representation of a reaction has an intuitive similarity to the conventional chemical-reaction notation (shown in figure B.1), but it makes more explicit reference to the chemical mass-action law as well as to the reaction stoichiometry. For appropriately constructed graphs, graph-rewrite rules correspond one-to-one with evaluation steps of mass-action kinetics, permitting simplification of complex reaction networks to isolate key features, while retaining correspondence of the visual and mathematical representations.

We use graph representations of reaction networks in the text where we need to show relations among multiple pathways that may connect the same inputs and outputs (such as acetyl-CoA and succinyl-CoA), and may draw from the same input and output species (such as CO_2, reductant, and water). Parallel input and output sequences appear as 'ladder' topology in these graphs, and for the particular pathways of biological carbon fixation, this is due to the recurrence of identical functional-group reaction sequences in multiple pathways, as discussed in section 3.2.

In this appendix we define the graph representation used in the text, introduce graph-reduction procedures and prove that they satisfy the mathematical property of associativity, and provide solutions for the particular simplification of interacting rTCA and WL pathways in a diluting environment.

All examples in this appendix use the same simplified projection onto the CHO sector that is used in diagrams in the main text. Actual reaction free energies will be driven by coupled energies of hydrolysis of ATP or oxidation of thiols to thioesters. The graph-reduction methods described in the next section may be used to include such effects into lumped-parameter representations of multi-reagent reaction sequences that regenerate energetic intermediates such as ATP or CoA in a network where these are made explicit.

B.1. Definition of graphic elements

B.1.1. Basic elements and well-formed graphs.
The elements in a bipartite-graph representation of a chemical reaction or reaction network are defined as follows:
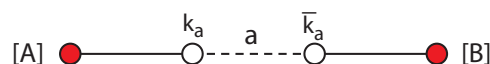
Figure B1. Basic reaction graph. [A] and [B] are concentrations associated with the two colored nodes. Forward and backward rate constants k_a and $\bar{k}_a$ are associated with the two unfilled circles. The associated reaction state current is J_a.

- Filled dots represent concentrations of chemical species. Each such dot is given a label indicating the species, such as

$$\overset{\text{ACE}}{\bullet} \leftrightarrow [\text{ACE}], \tag{1}$$

  used to refer to acetate in the text.
- Dashed lines represent transition states of reactions. Each is given a label indicating the reaction, as in $\text{-}\text{-}\underset{\text{b}}{\text{-}}\text{-}$.
- Hollow circles indicate inputs or outputs between molecular species and transition states, as in

$$\overset{\text{Ace}}{\underset{\text{CO}_2}{\text{H}_2}} \!\!\!\!\diagdown\!\!\!\!\bigcirc\text{-}\underset{\text{b}}{\text{-}} . \tag{2}$$
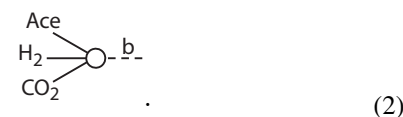
  Each circle is associated with the complex of reactants or products of the associated reaction, indicated as labeled line stubs.
- Hollow circles are tied to molecular concentrations with solid lines $\underline{\text{ACE}}$; one line per mole of reactant or product participating in the reaction. (That is, if m moles of a species A enter a reaction b, then m lines connect the dot corresponding to [A] to the hollow circle leading into reaction b. This choice uses graph elements to carry information about stoichiometry, as an alternative to labeling input- or output-lines to indicate numbers of moles.)
- Full reactions are defined when two hollow circles are connected by the appropriate transition state, as in

$$\overset{\text{ACE}}{\underset{\text{CO2}}{\text{H2}}} \!\!\!\!\diagdown\!\!\!\!\bigcirc\text{-}\underset{\text{b}}{\text{-}}\bigcirc\!\!\!\!\diagup\!\!\!\!\overset{\text{PYR}}{\underset{\text{H2O}}{}} , \tag{3}$$
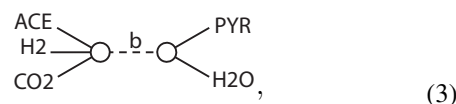
  describing the reductive carboxylation of acetate to form pyruvate.
- The bipartite graph for a fully specified reaction takes the form

$$(B.1)$$
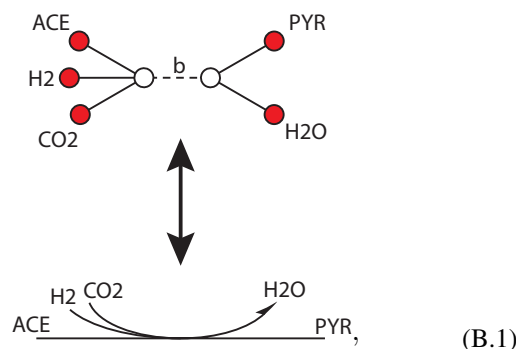
  where labeled stubs are connected to filled circles by mole-lines. The bipartite-graph corresponds to the standard chemical notation for the same reaction as shown.

B.1.2. Assignment of graph elements to terms in the mass-action rate equation. The mass-action kinetics for a graph such as the reductive carboxylation of acetate is given in terms of two half-reaction currents, which we may denote with the reaction label and an arbitrary sign as

$$j_b^+ = k_b \,[\text{ACE}]\,[\text{CO}_2]\,[\text{H}_2]$$
$$j_b^- = \bar{k}_b \,[\text{PYR}]\,[\text{H}_2\text{O}]. \tag{B.2}$$

k_b and $\bar{k}_b$ denote the forward and reverse half-reaction rate constants. The total reaction current $J_b \equiv j_b^+ - j_b^-$ is related to the contribution of this reaction to the changes in concentration as

$$[\dot{\text{ACE}}] = [\dot{\text{CO}}_2] = [\dot{\text{H}}_2] = -J_b$$
$$[\dot{\text{PYR}}] = [\dot{\text{H}_2\text{O}}] = J_b, \tag{B.3}$$

where the overdot denotes the time derivative. Reaction currents on graphs do not have inherent directions, reflecting the microscopic reversibility of reactions. All sources of irreversibility are to be made explicit in the chemical potentials that constitute the boundary conditions for reactions.

Each term in the mass-action rate equation may be identified with a specific graphical element in the bipartite representation. The half-reaction rate constants k_b, $\bar{k}_b$ are associated with the hollow circles, and the current J_b (which is the time-derivative of the coordinate giving the 'extent of the reaction') is associated with the transition-state dashed line. Concentrations, as noted, are associated with filled dots, and stoichiometric coefficients are associated with the multiplicities of solid lines.

B.2. Graph reduction for reaction networks in steady state

Networks of chemical reactions in steady state satisfy the constraints that the input and output currents to each chemical species (including any external sources or sinks) sum to zero. These constraints are the basis of stoichiometric flux-balance analysis [301–304], but they can also be used to eliminate internal nodes as explicit variables, leading to lumped-parameter expressions for entire sub-networks as 'effective' vertices or reactions. With appropriate absorption of externally buffered reagents into rate constants, this network reduction can be done exactly, without loss of information. An example of such a reduction is the Michaelis–Menton representation of multiple substrate binding at enzymes. Systematic methods for network reduction were one motivation behind Sinanǒglu's graphic methods [305, 306]. More sophisticated stochastic approaches have recently been used to include fluctuation properties in effective vertices, generalizing the Michaelis relation beyond mean field [307].

The map we have given of mass-action rate parameters to graphic elements allows us to represent steady-state network reduction in terms of graph reduction. In this approach, rewrite rules for the removal of graph elements are mapped to composition rules for half-reaction rate constants and stoichiometric coefficients. These composition rules can be proved to be associative, leading to an algebra for graph reduction. Here we sketch the rewrite rules relevant to reduction of the citric-acid cycle graph. In the next subsection we will reduce the graph, to the form used in the text.
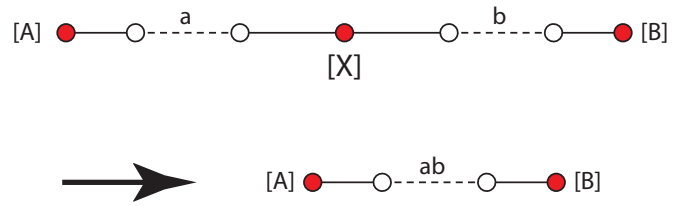
Figure B2. Removal of an internal species X from a diagram with elementary reactions. Rate constant pairs $(k_a, \bar{k}_a)$, $(k_b, \bar{k}_b)$ are used to define new rate constants $(k_{ab}, \bar{k}_{ab})$ for the effective transition state ab.

B.2.1. The base composition rule for removal of a single internal species. The simplest reduction is removal of an intermediate chemical species that is the sole output to one reaction, and the sole input to another, in a linear chain. Examples in the TCA cycle include MAL and ISC, produced by reductions and consumed by dehydrations. They also include CIT itself, produced by the hydration of aconitate and consumed by retro-aldol cleavage.

For a single linear reaction as shown in figure B1, the mass-action law is

$$[A]k_a - [B]\bar{k}_a = J_a, \tag{B.4}$$

and concentrations change as

$$[\dot{A}] = -J_a$$
$$[\dot{B}] = J_a. \tag{B.5}$$

The equilibrium constant for the reaction A → B is

$$K_{A\to B} = \frac{k_a}{\bar{k}_a}. \tag{B.6}$$

For two such reactions in a chain, as shown in figure B2, the mass-action laws are

$$[A]k_a - [X]\bar{k}_a = J_a$$
$$[X]k_b - [B]\bar{k}_b = J_b, \tag{B.7}$$

and the conservation laws become

$$[\dot{A}] = -J_a$$
$$[\dot{X}] = J_a - J_b$$
$$[\dot{B}] = J_b. \tag{B.8}$$

Under the steady-state condition $[\dot{X}] = 0$, we wish to replace the equations (B.7, B.8) with a rate equation

$$[A]k_{ab} - [B]\bar{k}_{ab} = J_{ab} \tag{B.9}$$

and a conservation law expressed in terms of $J_a = J_{ab} = J_b$. The rate constants in equation (B.9) are to be specified through a composition rule

$$(k_a, \bar{k}_a) \circ (k_b, \bar{k}_b) = (k_{ab}, \bar{k}_{ab}) \tag{B.10}$$

derived from the graph rewrite. Removing [X] from the mass-action equations using $[\dot{X}] = 0$, we derive that the rate constants satisfying equation (B.9) are given by

$$k_{ab} = \frac{k_a k_b}{\bar{k}_a + k_b}$$
$$\bar{k}_{ab} = \frac{\bar{k}_a \bar{k}_b}{\bar{k}_a + k_b}. \tag{B.11}$$
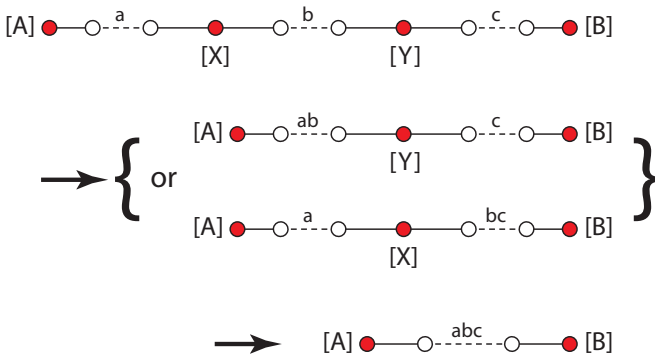
Figure B3. Composition of three reactions a, b, c can proceed by elimination of either X or Y first.

The associated equilibrium constant correctly satisfies the relation

$$\frac{k_{ab}}{\bar{k}_{ab}} = \frac{k_a}{\bar{k}_a}\frac{k_b}{\bar{k}_b}. \tag{B.12}$$

B.2.2. Associativity of the elementary composition rule. The composition rule (B.12) is associative, meaning that internal nodes may be removed from chains of reactions in any order, as shown in figure B3. All composition rules derived in the remainder of this appendix will be variants on the elementary rule (with additional buffered concentration variables added), so we demonstrate associativity for the base case as the foundation for other cases.

From equation (B.12) for $(k_a, \bar{k}_a) \circ (k_b, \bar{k}_b)$, followed by the equivalent expressions for $(k_{ab}, \bar{k}_{ab}) \circ (k_c, \bar{k}_c)$, $(k_a, \bar{k}_a) \circ (k_{bc}, \bar{k}_{bc})$, and $(k_b, \bar{k}_b) \circ (k_c, \bar{k}_c)$, we derive the sequence of reductions

$$\begin{aligned}
k_{abc} &= \frac{k_{ab}k_c}{\bar{k}_{ab} + k_c} \\
&= \frac{k_a k_b k_c}{\bar{k}_a \bar{k}_b + (\bar{k}_a + k_b)k_c} \\
&= \frac{k_a k_b k_c}{\bar{k}_a(\bar{k}_b + k_c) + k_b k_c} \\
&= \frac{k_a k_{bc}}{\bar{k}_a + k_{bc}}, \tag{B.13}
\end{aligned}$$

and a similar equation follows for $\bar{k}_{abc}$. Thus we have

$$[(k_a, \bar{k}_a) \circ (k_b, \bar{k}_b)] \circ (k_c, \bar{k}_c) = (k_a, \bar{k}_a) \circ [(k_b, \bar{k}_b) \circ (k_c, \bar{k}_c)]. \tag{B.14}$$

B.2.3. Removal of internal nodes that require other inputs or outputs. Next we consider the elimination of an internal node [X] that is produced or consumed together with other products or reactants. Conservation $[\dot{X}] = 0$ implies relations among the currents of these other species as well. All remaining graph reductions that we will perform for the TCA cycle are of this kind. In some cases both the secondary product and reactant are the solvent (water), as in the aconitase reactions (repeated in TCA, 3HB, 4HB and bicycle pathways). In other cases they
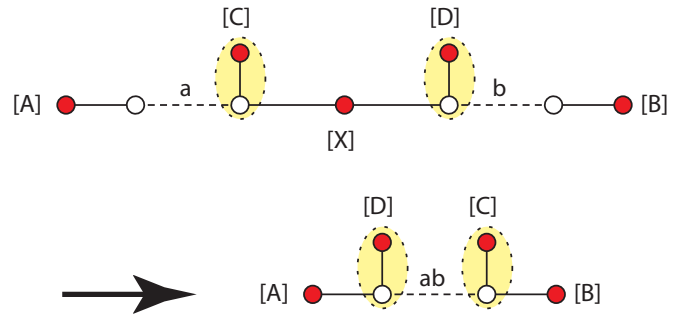
Figure B4. Representation of a composite graph with internal connections other than those to X as an effective elementary graph. Highlights denote the absorption of other species into modifications of effective rate constants coupled to X at a and b. These are then used to define the elementary-form rate constants $\hat{k}_{ab}$ and $\hat{\bar{k}}_{ab}$ in the reduced graph.

are reductants or inputs such as CO_2 that we consider buffered in the environment.

The pair of mass action equations we wish to reduce are[12]

$$\begin{aligned}
[A]k_a - [X][C]\bar{k}_a &= J_a \\
[X][D]k_b - [B]\bar{k}_b &= J_b, \tag{B.15}
\end{aligned}$$

and the desired reduced form is

$$[A][D]k_{ab} - [C][B]\bar{k}_{ab} = J_{ab}. \tag{B.16}$$

We first reduce equation (B.15) to the base case of the previous section, by absorbing the concentrations not to be removed into a pair of effective rate constants

$$\begin{aligned}
[A]k_a - [X]([C]\bar{k}_a) &= J_a \\
[X]([D]k_b) - [B]\bar{k}_b &= J_b. \tag{B.17}
\end{aligned}$$

From these we derive a composition equation

$$[A]\hat{k}_{ab} - [B]\hat{\bar{k}}_{ab} = J_{ab}, \tag{B.18}$$

corresponding to the graph representation in figure B4. We may then define $\hat{k}_{ab}$ and $\hat{\bar{k}}_{ab}$ by the elementary composition rule (B.10)

$$(k_a, [C]\bar{k}_a) \circ ([D]k_b, \bar{k}_b) = (\hat{k}_{ab}, \hat{\bar{k}}_{ab}), \tag{B.19}$$

giving the transformation[13]

$$\begin{aligned}
\hat{k}_{ab} &= \frac{k_a[D]k_b}{[C]\bar{k}_a + [D]k_b} \\
\hat{\bar{k}}_{ab} &= \frac{[C]\bar{k}_a\bar{k}_b}{[C]\bar{k}_a + [D]k_b}. \tag{B.20}
\end{aligned}$$

[12] In this and the following examples, we consider single additional species [C] and [D]. These may readily be generalized to a variety of cases in which the additional reagents are $\prod_{k=1}^{p}[C_k]$ and $\prod_{l=1}^{q}[D_l]$.

[13] Note that if [C] and [D] are the same species these cancel in the numerator and denominator of equation (B.20), and the same applies to common factors in products $\prod_{k=1}^{p}[C_k]$ and $\prod_{l=1}^{q}[D_l]$. Therefore these factors may simply be removed before the graph reduction if desired, because they encoded redundant constraints with the conservation law already implied by $[\dot{X}] = 0$. The irrelevance of redundant species in the graph reduction for removal of [X] is radically different from the graphically similar-looking role of a network catalyst which is both an input and an output of the same *reaction*. Network catalysts are essential to the determination of reaction rates.
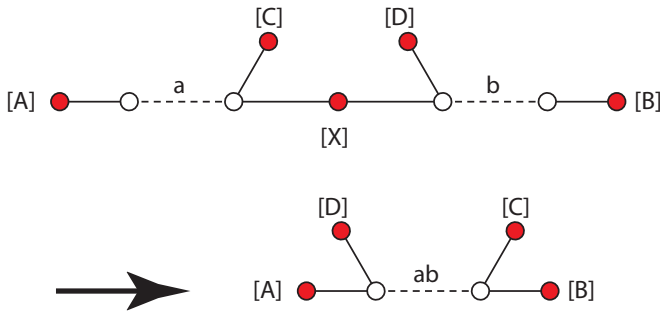
Figure B5. The composite graph corresponding to the reduction from equation (B.15) to equation (B.16).
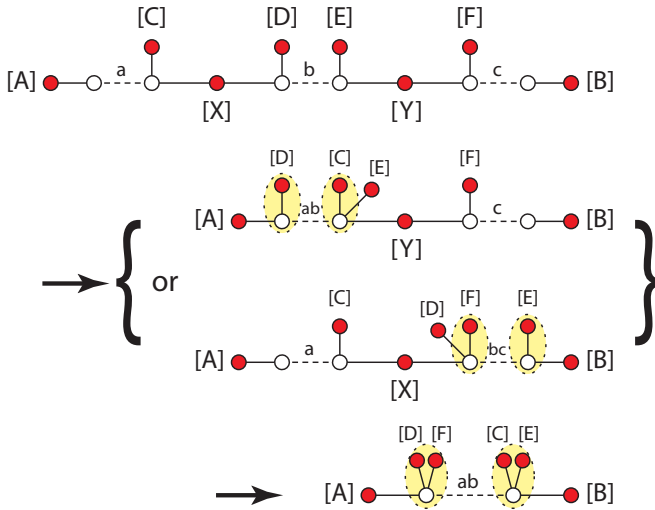

Figure B6. A two-step reduction with other internal connections, which may be performed by removing either X or Y first.

Now removing the factors of [C] and [D] used to define the hatted rate constants,

$$
\hat{k}_{ab} = [D]k_{ab}
$$
$$
\bar{\hat{k}}_{ab} = [C]\bar{k}_{ab}, \qquad (B.21)
$$

we obtain a direct expression for the composition rule in equation (B.18), of

$$
k_{ab} = \frac{k_a k_b}{[C]\bar{k}_a + [D]k_b}
$$
$$
\bar{k}_{ab} = \frac{\bar{k}_a \bar{k}_b}{[C]\bar{k}_a + [D]k_b}, \qquad (B.22)
$$

which is the interpretation of the graph reduction shown in figure B5.

B.2.4. Associativity for composite graphs. Associativity for composite graphs follows from the associativity of the elementary composition rule (B.14), via the grouping (B.19). To show how this works, we demonstrate associativity for the minimal case shown in figure B6. The important features are that the graph 're-wiring' follows from composition of the rule demonstrated in figure B5, and the composition rule for rate constants permits consistent removal of the necessary factors of reagent concentrations.

The application of the elementary reduction to remove X, corresponding to the second line in figure B6, yields equations (B.19) and (B.20). An equivalent removal of Y first (the third line of figure B6) gives

$$
(k_b, [E]\bar{k}_b) \circ ([F]k_c, \bar{k}_c) = (\hat{k}_{bc}, \bar{\hat{k}}_{bc}), \qquad (B.23)
$$

with rule

$$
\hat{k}_{bc} = \frac{k_b[F]k_c}{[E]\bar{k}_b + [F]k_c}
$$
$$
\bar{\hat{k}}_{bc} = \frac{[E]\bar{k}_b \bar{k}_c}{[E]\bar{k}_b + [F]k_c}. \qquad (B.24)
$$

The two equivalent rules for removing whichever internal node was not removed in the first reduction are

$$
(\hat{k}_{ab}, [E]\bar{\hat{k}}_{ab}) \circ ([F]k_c, \bar{k}_c) = (\hat{k}_{abc}, \bar{\hat{k}}_{abc}),
$$
$$
(k_a, [C]\bar{k}_a) \circ ([D]\hat{k}_{bc}, \bar{\hat{k}}_{bc}) = (\hat{k}_{abc}, \bar{\hat{k}}_{abc}). \qquad (B.25)
$$

Composing these rules for intermediate rate constants, we may check that

$$
\hat{k}_{abc} = \frac{\hat{k}_{ab}[F]k_c}{[E]\bar{\hat{k}}_{ab} + [F]k_c}
$$
$$
= \frac{(k_a[D]k_b)[F]k_c}{[C]\bar{k}_a[E]\bar{k}_b + ([C]\bar{k}_a + [D]k_b)[F]k_c}
$$
$$
= \frac{k_a[D](k_b[F]k_c)}{[C]\bar{k}_a([E]\bar{k}_b + [F]k_c) + [D]k_b[F]k_c}
$$
$$
= \frac{k_a[D]\hat{k}_{bc}}{[C]\bar{k}_a + [D]\hat{k}_{bc}}, \qquad (B.26)
$$

and a similar equation follows for $\bar{\hat{k}}_{abc}$. Converting the hatted forms to the normal reaction form produces the rate equation

$$
[A][D][F]k_{abc} - [C][E][B]\bar{k}_{abc} = J_{abc}. \qquad (B.27)
$$

We may directly obtain the rate constants $k_{abc}, \bar{k}_{abc}$ with the composition rule

$$
(k_{abc}, \bar{k}_{abc}) = (k_{ab}, \bar{k}_{ab}) \circ (k_c, \bar{k}_c)
$$
$$
= (k_a, \bar{k}_a) \circ (k_{bc}, \bar{k}_{bc}), \qquad (B.28)
$$

using the appropriate version of the graph-dependent evaluation rule (B.22) in each step. The resulting composition (B.28) is automatically associative, because it satisfies the conversion

$$
\hat{k}_{abc} = [D][F]k_{abc}
$$
$$
\bar{\hat{k}}_{abc} = [C][E]\bar{k}_{abc} \qquad (B.29)
$$

with equation (B.26), which is associative. As a final check, the equilibrium constants in the normal reaction form satisfy the necessary chain rule

$$
\frac{k_{abc}}{\bar{k}_{abc}} = \frac{k_a}{\bar{k}_a} \frac{k_b}{\bar{k}_b} \frac{k_c}{\bar{k}_c}. \qquad (B.30)
$$

Intermediate (hatted) rate constants have been used here to show how associativity is inherited from the base case. The examples below work directly with the actual (un-hatted) rate constants, which keep the network in its literal form at each reduction.

B.3. Application to the citric-acid cycle reactions

Using this graph representation and the associated graph reductions, we may express the qualitative kinetics associated with network autocatalysis in the rTCA cycle. We use a minimal model network in which only the cycle intermediates are represented explicitly, and only the CHO stoichiometry is retained. As noted above, phosphorylated intermediates and thioesters, including the energetically important substrate-level phosphorylation of CIT and succinate, are not represented. External sources or sinks are used to buffer only four compounds in the network, which are CO_2, H_2, H_2O and a pool of reduced carbon which we take to be acetate (ACE or CH_3COOH) because it has the lowest free energy of formation of cycle intermediates under reducing conditions (following [308]) and is the natural drain compound [7].

The purpose of network reduction in such a model is to produce a graph in which each element corresponds to a specific control parameter for the interaction of conservation laws with non-equilibrium boundary conditions. CO_2, H_2 and H_2O provide sources of carbon and reductant, and an output for reduced oxygen atoms. Because they comprise different ratios of three elements, any set of concentrations is consistent with a Gibbs equilibrium, and the chemical potentials corresponding to the elements are preserved by the conservation laws of arbitrary reactions. A fourth boundary condition for acetate cannot be linearly independent in equilibrium, and drives the steady-state reaction flux.

Such a model is limited in many ways. The replacement of explicit (and unknown) parasitic side reactions, from all cycle intermediates, by a single loss rate for acetate may fail to capture concentration-dependent losses, in a way that cannot simply be absorbed into lumped rate constants. Moreover, the rate constants themselves depend on catalysts, and reasonable values for these in a prebiotic or early-cellular context are unknown. Therefore all critical properties of the model are expressed relative to these rate constants. The reduction remains meaningful, however, because the lumped-parameter rate constants are controlled by the three buffered environmental compounds CO_2, H_2, and H_2O, leaving the network flux to be controlled by the disequilibrium concentration of acetate.

B.3.1. The graph reduction sequence.
The bipartite graph for the minimal rTCA network in CHO compounds is shown in figure B7. All networks in the text are generated by equivalent methods. Highlighted nodes are those that can be removed by the base reduction in section appendix B.2.1. Reactions are labeled with lowercase Roman letters, and relative to the elementary half-reaction rate constants, the lumped-parameter rate constants are given by

$$k_{de} = \frac{k_d k_e}{\bar{k}_d + k_e} \quad \bar{k}_{de} = \frac{\bar{k}_d \bar{k}_e}{\bar{k}_d + k_e}$$

$$k_{ij} = \frac{k_i k_j}{\bar{k}_i + k_j} \quad \bar{k}_{ij} = \frac{\bar{k}_i \bar{k}_j}{\bar{k}_i + k_j}$$

$$k_{ka} = \frac{k_k k_a}{\bar{k}_k + k_a} \quad \bar{k}_{ka} = \frac{\bar{k}_k \bar{k}_a}{\bar{k}_k + k_a}, \quad (B.31)$$
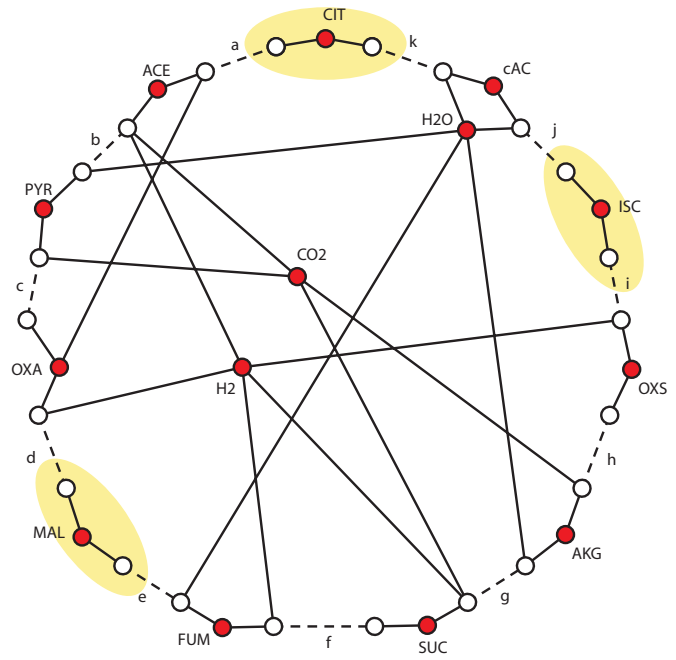

Figure B7. The projection of the TCA cycle onto CHO compounds. Phosphates and thioesters are omitted, and the stoichiometry of all acids refers to the protonated forms, so that H_2 stands for general two-electron reductants. Omission of explicit representations of substrate-level phosphorylation to form citryl-CoA and succinyl-CoA causes water elimination to accompany carboxylation of acetate and succinate in this graph, where in the actual cycle it would occur outside the graph, in the formation of pyrophosphates. Highlighted species are sole outputs and sole inputs of their associated reactions, and can be removed with the elementary composition rule (B.11) of section appendix B.2.1. Legend: acetate (ACE), pyruvate (PYR), oxaloacetate (OXA), malate (MAL), fumarate (FUM), succinate (SUC), α-ketoglutarate (AKG), oxalosuccinate (OXS), cis-aconitate (cAC), isocitrate (ISC), citrate (CIT).

with equivalent expressions for the $\bar{k}$s. These define the elementary reactions in the reduced graph of figure B8. Here and below, we give formulae only for the forward half-reaction rate constants k. Formulae for the backward half-reaction rate constants $\bar{k}$ have corresponding forms as shown in the preceding sections.

One further reduction that follows the elementary rule in figure B8 is removal of cis-aconitate (cAC), which involves a common factor of the solvent $[H_2O]$. The resulting lumped-parameter rate constants are given by

$$k_{ijka} = \frac{k_{ij} k_{ka}}{\bar{k}_{ij} + k_{ka}} \quad \bar{k}_{ijka} = \frac{\bar{k}_{ij} \bar{k}_{ka}}{\bar{k}_{ij} + k_{ka}}. \quad (B.32)$$

These lead to the graph of figure B9.

All further graph reductions require the composition rules of appendix B.2.3, and result in changes of the input or output stoichiometries of the unreduced nodes. All highlighted compounds in figure B9 may be removed, and the resulting lumped-parameter rate constants are given by

$$k_{bc} = \frac{k_b k_c}{[H_2O]\bar{k}_b + [CO_2]k_c}$$
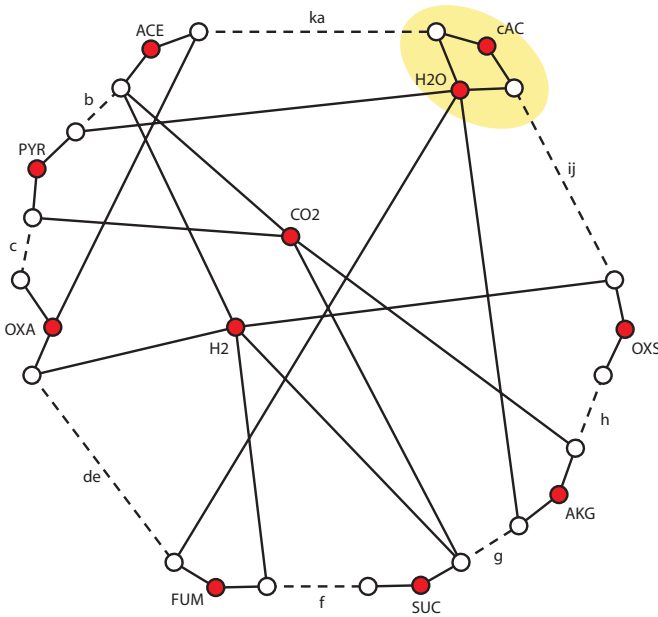
$$k_{def} = \frac{k_{de} k_f}{[H_2O]\bar{k}_{de} + [H_2]k_f}$$

Figure B8. Graph of figure B7 with its highlighted species removed. Cis-aconitate (cAC highlighted) has common factors of [H_2O], and is the next internal node to be removed, by the rewrite rules of appendix B.2.3, but with the simplifying feature that common factors cancel, so they resemble the base case.

$$k_{defg} = \frac{k_{def}k_g}{[H_2O]\bar{k}_{def} + [H_2][CO_2]k_g}$$

$$k_{defgh} = \frac{k_{defg}k_h}{[H_2O]^2\bar{k}_{defg} + [CO_2]k_h}$$

$$k_{defghijka} = \frac{k_{defgh}k_{ijka}}{[H_2O]^2\bar{k}_{defgh} + [H_2]k_{ijka}}. \qquad (B.33)$$

These define the maximal reduction of the original rTCA graph, to the graph shown in figure B10.

The lumped-parameter rate equations for figure B10, parametrized by lumped-parameter rate constants, are

$$J_{bc} = [ACE][H_2][CO_2]^2 k_{bc}$$
$$\quad - [OXA][H_2O]\bar{k}_{bc}$$
$$J_{defghijka} = [OXA][H_2]^4[CO_2]^2 k_{defghijka}$$
$$\quad - [OXA][ACE][H_2O]^2\bar{k}_{defghijka}. \quad (B.34)$$

In steady state $J_{bc} = 0$ and [OXA] may be replaced with the equilibrium function

$$[OXA] = \frac{k_{bc}}{\bar{k}_{bc}}\frac{[H_2][CO_2]^2}{[H_2O]}[ACE]. \qquad (B.35)$$

B.3.2. Network reaction fluxes and their control parameters.
For the remainder of the appendix we replace the subscript $_{defghijka}$ with designation $_{rTCA}$ in currents J, half-reaction rate constants k, $\bar{k}$, and equilibrium constants K. Dimensionally, the rate constants require the concentration of OXA in the mass-action law, and so presume that the anaplerotic segment $_{bc}$ has been handled.

Plugging equation (B.35) into the second rate equation of equation (B.34), and supposing [OXA] is in equilibrium with [ACE] at a (non-equilibrium) steady state for the network
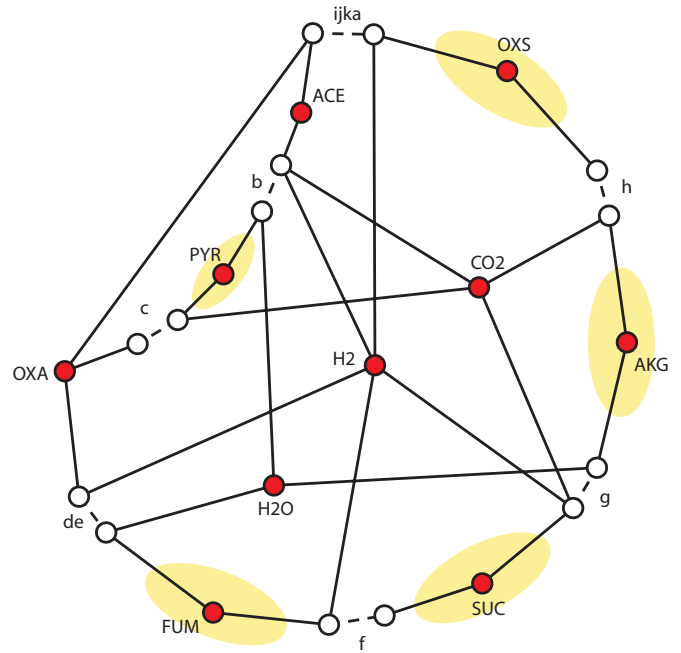
Figure B9. Graph of figure B8 with cAC and its parallel links to water removed. For all remaining species except acetate (ACE), neither sources nor sinks are assumed, and these may be removed with non-trivial instances of the composition rule of appendix B.2.3. Each of these removals changes the degree of the remaining reactions, and thus changes the topology of the graph.
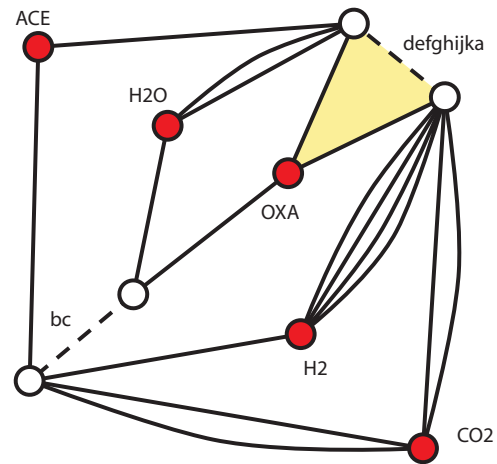

Figure B10. Graph of figure B9 with all internal nodes from linear chains removed. [H_2O], [H_2], [CO_2], and [ACE] are the four molecular concentrations to which boundary sources are coupled. [OXA] is retained as the last representation of the network catalysis of the loop, indicated by highlighting of the reaction in which OXA is input and output with equal stoichiometry. In steady state, OXA is in equilibrium with ACE, because it is not coupled to external currents.

as a whole, we obtain the only independent mass-action rate equation for the reduced network. This is the current producing acetate:

$$J_{rTCA} = \bar{k}_{rTCA}\frac{k_{bc}}{\bar{k}_{bc}}[H_2][CO_2]^2[H_2O][ACE]$$
$$\times \left(\frac{k_{rTCA}}{\bar{k}_{rTCA}}\frac{[H_2]^4[CO_2]^2}{[H_2O]^2} - [ACE]\right). \qquad (B.36)$$

The first term in parenthesis in equation (B.36) is the concentration at which acetate would be in equilibrium with the inorganic inputs, which we denote

$$[ACE]_G \equiv \frac{k_{rTCA}}{\bar{k}_{rTCA}} \frac{[H_2]^4[CO_2]^2}{[H_2O]^2}. \qquad (B.37)$$

Therefore the network response is proportional to the offset of [ACE] from its equilibrium value, with a rate constant that depends on the particular contributions of chemical potential from $[CO_2]$ and reductant. Although the lumped-parameter rate constant in this relation appears complex, the consistency conditions with single-reaction equilibrium constants ensure that $k_{rTCA}/\bar{k}_{rTCA}$ is independent of synthetic pathway and equal to the exponential of the Gibbs free energy of formation.

B.4. Interaction of Wood–Ljungdahl with rTCA

We may envision an early WL 'feeder' pathway to acetyl-CoA as a reaction with the same stoichiometry as rTCA for the creation of acetate, but fixed half-reaction rate constants that do not depend on the internal concentrations in the network. This may be a pre-pterin mineral pathway [147], in which rate constants are determined by the abiotic environment, or an early pathway using pterin-like cofactors, if the concentrations of these are somehow buffered from the instantaneous flows through the reductive pathway. Labeling this 'linear' effective reaction WL, the rate equation becomes

$$J_{WL} = \bar{k}_{WL}[H_2O]^2 \left(\frac{k_{WL}}{\bar{k}_{WL}} \frac{[H_2]^4[CO_2]^2}{[H_2O]^2} - [ACE] \right). \qquad (B.38)$$

Note that $k_{WL}/\bar{k}_{WL} = k_{rTCA}/\bar{k}_{rTCA}$ because both are expressions for the equilibrium constant which depends only on the free energy of reaction.

To understand the performance of a joint network in the presence of losses, as the simplest case introduce a reaction labeled Env standing for dilution of acetate to an environment at zero concentration. The dilution current becomes

$$J_{Env} = k_{Env}[ACE]. \qquad (B.39)$$

At a non-equilibrium steady state the total losses must equal the total supply currents, so that

$$J_{Env} = J_{rTCA} + J_{WL}. \qquad (B.40)$$

The un-reduced equation for steady-state currents can be written

$$J_{rTCA} + J_{WL} = [H_2O]^2$$
$$\times \left\{ \sqrt{k_{rTCA}\bar{k}_{rTCA}} \frac{K_{bc}}{K_{rTCA}} \frac{[CO_2]}{[H_2]}[ACE]_G^{1/2}[ACE] + \bar{k}_{WL} \right\}$$
$$\times ([ACE]_G - [ACE])$$
$$= J_D = k_D[ACE] \qquad (B.41)$$

The graph corresponding to this model for rate laws is shown in figure B11.

The variable that characterizes the 'impedance' of a chemical reaction network, and displays thresholds for
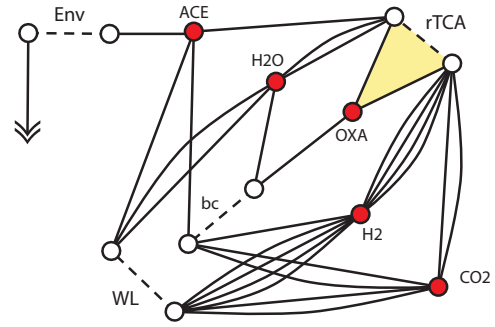

Figure B11. Hypergraph model for parallel reactions through the rTCA and WL pathways, coupled to a linear drain reaction representing dilution of acetate by the environment.

autocatalysis when these exist, is the ratio of the output acetate concentration to the value that would exist in a Gibbs equilibrium with the inputs:

$$x \equiv \frac{[ACE]}{[ACE]_G}. \qquad (B.42)$$

For a network with no reaction barriers (either in rate constants or due to limitations of network catalysts, the output $x \to 1$.

The two control parameters that govern the relative contributions of the rTCA loop and the direct WL feeder are

$$z_{rTCA} = \frac{\sqrt{k_{rTCA}\bar{k}_{rTCA}}}{k_{Env}} \frac{K_{bc}}{K_{rTCA}} \frac{[CO_2][H_2O]^2}{[H_2]}[ACE]_G^{3/2}$$
$$z_{WL} = \frac{\bar{k}_{WL}[H_2O]^2}{k_{Env}}. \qquad (B.43)$$

Each control parameter is a ratio of lumped half-reaction rates that feed [ACE] to the environment dilution constant k_{Env} through which it drains.

In terms of z_{WL} and z_{rTCA}, the normalized concentration x—which is proportional by k_{Env} to the total current through the system—satisfies

$$x = \frac{1}{2} \left(1 - \frac{1 + z_{WL}}{z_{rTCA}} \right) + \sqrt{\frac{z_{WL}}{z_{rTCA}} + \frac{1}{4} \left(1 - \frac{1 + z_{WL}}{z_{rTCA}} \right)^2}. \qquad (B.44)$$

The solution to equation (B.44) is shown versus base-10 logarithms of z_{rTCA} and z_{WL} in figure 12 in the main text. The critical (unsupported) response of the rTCA loop occurs at $z_{WL} \to 0$ and $z_{rTCA} = 1$. It is identified with the discontinuity in the derivative $\partial x / \partial z_{rTCA}$ at $z_{rTCA} = 1$ and the exactly zero value of x for $z_{rTCA} < 1$. As z_{WL} increases from zero, the transition becomes smooth, and a nonzero concentration x is maintained against dilution at *all* values of z_{rTCA}.

References

[1] Schrödinger E 1992 *What is Life? The Physical Aspect of the Living Cell* (New York: Cambridge University Press)
[2] Smith E 2008 Thermodynamics of natural selection: I. Energy and entropy flows through non-equilibrium ensembles *J. Theor. Biol.* **252** 185–97

[3] Smith E 2008 Thermodynamics of natural selection: II. Chemical Carnot cycles *J. Theor. Biol.* **252** 198–212

[4] Falkowski P G, Fenchel T and Delong E F 2008 The microbial engines that drive earth's biogeochemical cycles *Science* **320** 1034–9

[5] Lengeler J W, Drews G and Schlegel H G 1999 *Biology of the Prokaryotes* (New York: Blackwell)

[6] Morowitz H J 1992 *Beginnings of Cellular Life* (New Haven, CT: Yale University Press)

[7] Smith E and Morowitz H J 2004 Universality in intermediary metabolism *Proc. Natl Acad. Sci. USA* **101** 13168–73

[8] Rankama K and Sahama T G 1950 *Geochemistry* (Chicago, IL: University of Chicago Press)

[9] Bar-Even A, Noor E, Savir Y, Liebermeister W, Davidi D, Tawfik D S and Milo R 2011 The moderately efficient enzyme: evolutionary and physicochemical trends shaping enzyme parameters *Biochemistry* **50** 4402–10

[10] Simon H A 1962 The architecture of complexity *Proc. Am. Phil. Soc.* **106** 467–82

[11] Simon H A 1973 The organization of complex systems *Hierarchy Theory: The Challenge of Complex Systems* ed H H Pattee (New York: George Braziller) pp 3–27

[12] Ancel L W and Fontana W 2000 Plasticity, evolvability and modularity in RNA *J. Exp. Zool.* B **288** 242–83

[13] Fontana W 2002 Modeling 'evo-devo' with RNA *Bioessays* **24** 1164–77

[14] Wagner A 2008 Robustness and evolvability: a paradox resolved *Proc. R. Soc. B: Biol. Sci.* **275** 91–100

[15] Wagner G P and Altenberg L 1996 Perspective: complex adaptations and the evolution of evolvability *Evolution* **50** 967–76

[16] Kirschner M and Gerhart J 1998 Evolvability *Proc. Natl Acad. Sci.* **95** 8420–7

[17] Gerhart J and Kirschner M 1997 *Cells, Embryos and Evolution* (New York: Wiley)

[18] Gerhart J and Kirschner M 2007 The theory of facilitated variation *Proc. Natl Acad. Sci. USA* **104** 8582–9

[19] Ravasz E, Somera A L, Mongru D A, Oltvai Z N and Barabasi A-L 2002 Hierarchical organization of modularity in metabolic networks *Science* **297** 1551–5

[20] Guimera R and Amaral L A N 2005 Functional cartography of complex metabolic networks *Nature* **433** 895–900

[21] Newman M E J 2006 Modularity and community structure in networks *Proc. Natl Acad. Sci.* **103** 8577–82

[22] Braakman R and Smith E 2012 The emergence and early evolution of biological carbon fixation *PLoS Comput. Biol.* **8** e1002455

[23] Monod J, Changeux J-P and Jacob F 1963 Allosteric proteins and cellular control systems *J. Mol. Biol.* **6** 306–29

[24] Savageau M A 1974 Optimal design of feedback control by inhibition *J. Mol. Evol.* **4** 139–56

[25] Savageau M A 1975 Optimal design of feedback control by inhibition *J. Mol. Evol.* **5** 199–222

[26] de Lorenzo V 2008 Systems biology approaches to bioremediation *Curr. Opin. Biotechnol.* **19** 579–89

[27] Elser J J, Sterner R W, Gorokhova E, Fagan W F, Markow T A, Cotner J B, Harrison J F, Hobbie S E, Odell G M and Weider L W 2000 Biological stoichiometry from genes to ecosystems *Ecol. Lett.* **3** 540–50

[28] Venter J C *et al* 2004 Environmental genome shotgun sequencing of the Sargasso Sea *Science* **304** 66–74

[29] Borenstein E, Kupiec M, Feldman M W and Ruppin E 2008 Large-scale reconstruction and phylogenetic analysis of metabolic environments *Proc. Natl Acad. Sci. USA* **105** 14482–7

[30] Klitgord N and Segrè D 2011 Ecosystems biology of microbial metabolism *Curr. Opin. Biotechnol.* **22** 541–6

[31] Amend J P and Shock E L 2001 Energetics of overall metabolic reactions of thermophilic and hyperthermophilic archaea and bacteria *FEMS Microbiol. Rev.* **25** 175–243

[32] Reysenbach A-L and Shock E 2002 Merging genomes with geochemistry in hydrothermal ecosystems *Science* **296** 1077–82

[33] Martin W, Baross J, Kelley D and Russell M J 2008 Hydrothermal vents and the origin of life *Nature Rev. Microbiol.* **6** 805–14

[34] Shock E L 2009 Minerals as energy sources for microorganisms *Econ. Geol.* **104** 1235–48

[35] Erwin D H, Laflamme M, Tweedt S M, Sperling E A, Pisani D and Peterson K J 2011 The Cambrian conundrum: early divergence and later ecological success in the early history of animals *Science* **334** 1091–7

[36] Erwin D H and Tweedt S 2012 Ecological drivers of the Ediacaran-Cambrian diversification of metazoa *Evol. Ecol.* **26** 417–33

[37] Erwin D H and Valentine J W 2013 *The Cambrian Explosion: The Construction of Animal Biodiversity* (Greenwood Village, CO: Roberts)

[38] Erwin D H 2012 Macroevolution: dynamics of diversity *Curr. Biol.* **21** R1000–1

[39] Redfield A C 1934 On the proportions of organic derivatives in sea water and their relation to the composition of plankton *James Johnstone Memorial Volume* (Liverpool: University of Liverpool Press) pp 176–192

[40] Redfield A C 1958 The biological control of chemical factors in the environment *Am. Sci.* **46** 205–21

[41] Morowitz H J and Smith E 2007 Energy flow and the organization of life *Complexity* **13** 51–9 SFI preprint # 06-08-029

[42] Rodrigues J F M and Wagner A 2009 Evolutionary plasticity and innovations in complex metabolic reaction networks *PLoS Comput. Biol.* **5** e1000613

[43] Kanehisa M, Goto S, Hattori M, Aoki-Kinoshita K F, Itoh M, Kawashima S, Katayama T, Araki M and Hirakawa M 2006 From genomics to chemical genomics: new developments in KEGG *Nucleic Acids Res.* **34** D354–7

[44] The UniProt Consortium 2011 Ongoing and future developments at the universal protein resource *Nucleic Acids Res.* **39** D214–9 (Suppl. 1)

[45] Dagley S and Nicholson D E 1970 *An Introduction to Metabolic Pathways* (Oxford: Blackwell)

[46] Buss L W 2007 *The Evolution of Individuality* (Princeton, NJ: Princeton University Press)

[47] Fisher R A 2000 *The Genetical Theory of Natural Selection* (London: Oxford University Press)

[48] Ewens W J 2004 *Mathematical Population Genetics* 2nd edn (Heidelberg: Springer)

[49] Odling-Smee F J, Laland K N and Feldman M W 2003 *Niche Construction: The Neglected Process in Evolution* (Princeton, NJ: Princeton University Press)

[50] Srinivasan V and Morowitz H J 2009 The canonical network of autotrophic intermediary metabolism: minimal metabolome of a reductive chemoautotroph *Biol. Bull.* **216** 126–30 PMID: 19366923

[51] Srinivasan V and Morowitz H J 2009 Analysis of the intermediary metabolism of a reductive chemoautotroph *Biol. Bull.* **217** 222–32 PMID: 20040747

[52] Coelho P M B M, Salvador A and Savageau M A 2009 Quantifying global tolerance of biochemical systems: design implications for moiety-transfer cycles *PLoS Comput. Biol.* **5** e1000319

[53] Csete M and Doyle J 2004 Bow ties, metabolism and disease *Trends Biotechnol.* **22** 446–50

[54] Zhao J, Tao L, Hong Yu, Luo J H, Cao Z W and Li Yi Xue 2007 Bow-tie topological features of metabolic networks and the functional significance *Chin. Sci. Bull.* **52** 1036–45

[55] Riehl W J, Krapivsky P L, Redner S and Segrè D 2010 Signatures of arithmetic simplicity in metabolic network architecture *PLoS Comput. Biol.* **6** e1000725

[56] Srinivasan V, Morowitz H J and Huber H 2012 What is an autotroph? *Arch. Microbiol.* **194** 135–40

[57] Sterner R W and Elser J J 2002 *Ecological Stoichiometry: The Biology of Elements from Molecules to the Biosphere* (Princeton, NJ: Princeton University Press)

[58] Smith E and Morowitz H J 2010 The autotrophic origins paradigm and small-molecule organocatalysis *Orig. Life Evol. Biosph.* **40** 397–402

[59] Hartman H 1975 Speculations on the origin and evolution of metabolism *J. Mol. Evol.* **4** 359–70

[60] Fraser C M *et al* 1995 The minimal gene complement of mycoplasma genitalium *Science* **270** 397–404

[61] Himmelreich R, Hilbert H, Plagens H, Pirkl E, Bi-Chen L and Herrmann R 1996 Complete sequence analysis of the genome of the bacterium Mycoplasma pneumoniae *Nucleic Acids Res.* **24** 4420–49

[62] Razin S, Yogev D and Naot Y 1998 Molecular biology and pathogenicity of mycoplasmas *Microbiol. Mol. Biol. Rev.* **62** 1094–156 PMID: 9841667

[63] Say R F and Fuchs G 2010 Fructose 1,6-bisphosphate aldolase/phosphatase may be an ancestral gluconeogenic enzyme *Nature* **464** 1077–81

[64] Metzler D E 2003 *Biochemistry: The Chemical Reactions of Living Cells* 2nd edn (New York: Academic)

[65] Massey L K, Sokatch J R and Conrad R S 1976 Branched-chain amino acid catabolism in bacteria *Bacteriol. Rev.* **40** 42–54 PMCID: PMC413937

[66] Vogels G D and Van der Drift C 1976 Degradation of purines and pyrimidines by microorganisms *Bacteriol. Rev.* **40** 403–468 PMCID: PMC413962

[67] Baughn A D, Garforth S J, Vilchèze C and Jacobs W R Jr 2009 An anaerobic-type α-ketoglutarate ferredoxin oxidoreductase completes the oxidative tricarboxylic acid cycle of *Mycobacterium tuberculosis PLoS Pathogens* **5** e1000662

[68] Watanabe S, Zimmermann M, Goodwin M B, Sauer U, Barry 3rd C E and Boshoff H I 2011 Fumarate reductase activity maintains an energized membrane in anaerobic *Mycobacterium tuberculosis PLoS Pathogens* **7** e1002287

[69] Zhang S and Bryant D A 2011 The tricarboxylic acid cycle in cyanobacteria *Science* **334** 1551–3

[70] Meléndez-Hevia E, Waddell T G and Montero F 1994 Optimization of metabolism: the evolution of metabolic pathways toward simplicity through the game of the pentose phosphate cycle *J. Theor. Biol.* **166** 201–20

[71] Meléndez-Hevia E, Waddell T and Cascante M 1996 The puzzle of the Krebs citric acid cycle: assembling the pieces of chemically feasible reactions, and opportunism in the design of metabolic pathways during evolution *J. Mol. Evol.* **43** 293–303

[72] Noor E, Eden E, Milo R and Alon U 2010 Central carbon metabolism as a minimal biochemical walk between precursors for biomass and energy *Mol. Cell* **39** 809–20

[73] Puigbo P, Wolf Y and Koonin E 2009 Search for a 'Tree of Life' in the thicket of the phylogenetic forest *J Biol.* **8**

[74] Russell M J and Martin W 2004 The rocky roots of the acetyl-CoA pathway *Trends Biochem. Sci.* **29** 358–63

[75] Martin W and Russell M J 2007 On the origin of biochemistry at an alkaline hydrothermal vent *Phil. Trans. R. Soc.* B **362** 1887–926

[76] Hügler M and Seivert S M 2011 Beyond the Calvin cycle: autotrophic carbon fixation in the ocean *Ann. Rev. Mar. Sci.* **3** 261–89

[77] Fuchs G 2011 Alternative pathways of carbon dioxide fixation: insights into the early evolution of life? *Ann. Rev. Microbiol.* **65** 631–58

[78] Woese C R 1998 The universal ancestor *Proc. Natl Acad. Sci. USA* **95** 6854–9

[79] Woese C R 2000 Interpreting the universal phylogenetic tree *Proc. Natl Acad. Sci. USA* **97** 8392–6

[80] Woese C R 2002 On the evolution of cells *Proc. Natl Acad. Sci. USA* **99** 8742–7

[81] Goldenfeld N and Woese C 2011 Life is physics: evolution as a collective phenomenon far from equilibrium *Ann. Rev. Condens. Matter Phys.* **2** 375–99

[82] Eigen M and Schuster P 1977 The hypercycle: part a. The emergence of the hypercycle *Naturwissenschaften* **64** 541–65

[83] Eigen M and Schuster P 1978 The hypercycle: part c. The realistic hypercycle *Naturwissenschaften* **65** 341–69

[84] Kauffman S 1993 *The Origins of Order: Self-Organization and Selection in Evolution* (London: Oxford University Press)

[85] Zachar I and Szathmary E 2010 A new replicator: a theoretical framework for analysing replication *BMC Biol.* **8** 21

[86] Hordijk W, Kauffman S A and Steel M 2012 Required levels of catalysis for emergence of autocatalytic sets in models of chemical reaction systems *Int. J. Mol. Sci.* **12** 3085–101

[87] Andersen J L, Flamm C, Merkle D and Stadler P F 2012 Maximizing output and recognizing autocatalysis in chemical reaction networks is NP-complete *J. Syst. Chem.* **3** 1

[88] Szent-Gyorgyi A 1972 *The Living State: With Observations on Cancer* (New York: Academic)

[89] Copley S D, Smith E and Morowitz H J 2007 The origin of the RNA world: co-evolution of genes and metabolism *Bioorg. Chem.* **35** 430–43

[90] Utter M F and Wood H G 1951 Mechanisms of fixation of carbon dioxide by heterotrophs and autotrophs *Adv. Enzymol. Relat. Areas Mol. Biol.* **12** 41–152

[91] Ljungdahl L, Irion E and Wood H G 1965 Total synthesis of acetate from CO_2: I. Co-methylcobyric acid and CO-(methyl)-5-methoxybenzimidazolylcobamide as intermediates with *Clostridium thermoaceticum Biochemistry* **4** 2771–80

[92] Ljungdahl L and Wood H G 1965 Incorporation of C^{14} from carbon dioxide into sugar phosphates, carboxylic acids and amino acids by *Clostridium thermoaceticum J. Bacteriol.* **89** 1055–64

[93] Maden B E H 2000 Tetrahydrofolate and tetrahydromethanopterin compared: functionally distinct carriers in C_1 metabolism *Biochem. J.* **350** 609–29

[94] Berg I A, Kockelkorn D, Ramos-Vera W H, Say R F, Zarzycki J, Hügler M, Alber B E and Fuchs G 2010 Autotrophic carbon fixation in archaea *Nature Rev. Microbiol.* **8** 447–60

[95] Peretó J 2012 Out of fuzzy chemistry: from prebiotic chemistry to metabolic networks *Chem. Soc. Rev.* **41** 5394–403

[96] Buchanan B B and Arnold D I 1990 A reverse Krebs cycle in photosynthesis: consensus at last *Photosynth. Res.* **24** 47–53

[97] Cody G D, Boctor N Z, Filley T R, Hazen R M, Scott J H, Sharma A and Yoder H S Jr 2000 Primordial carbonylated iron–sulfur compounds and the synthesis of pyruvate *Science* **289** 1337–40

[98] Bar-Even A, Flamholz A, Noor E and Milo R 2012 Thermodynamic constraints shape the structure of carbon fixation pathways *Biochim. Biophys. Acta (BBA) - Bioenerg.* **1817** 1646–59

[99] Dobbek H, Svetlitchnyi V, Gremer L, Huber R and Meyer O 2001 Crystal structure of a carbon monoxide dehydrogenase reveals a [Ni-4Fe-5S] cluster *Science* **293** 1281–5

[100] Darnault C, Volbeda A, Kim E J, Legrand P, Vernede X, Lindahl P A and Fontecilla-Camps J C 2003 Ni-Zn-[Fe4-S4] and Ni-ni-[Fe4-S4] clusters in closed and open α subunits of acetyl-coA synthase/carbon monoxide dehydrogenase *Nature Struct. Mol. Biol.* **10** 271–9

[101] Seravalli J, Xiao Y, Weiwei G, Cramer S P, Antholine W E, Krymov V, Gerfen G J and Ragsdale S W 2004 Evidence that NiNi acetyl-coA synthase is active and that the CuNi enzyme is not *Biochemistry* **43** 3944–55

[102] Svetlitchnyi V, Dobbek H, Meyer-Klaucke W, Meins T, Thiele B, Römer P, Huber R and Meyer O 2004 A functional Ni-ni-[4Fe-4S] cluster in the monomeric acetyl-coA synthase from Carboxydothermus hydrogenoformans *Proc. Natl Acad. Sci. USA* **101** 446–51

[103] Banerjee R and Ragsdale S W 2003 The many faces of vitamin B12: catalysis by cobalamin-dependent enzymes *Annu. Rev. Biochem.* **72** 209–47

[104] Bender G, Pierce E, Hill J A, Darty J E and Ragsdale S W 2011 Metal centers in the anaerobic microbial metabolism of CO and CO$_2$ *Metallomics* **3** 797–815

[105] Ragsdale S W, Clark J E, Ljungdahl L G, Lundie L L and Drake H L 1983 Properties of purified carbon monoxide dehydrogenase from *Clostridium thermoaceticum*, a nickel, iron–sulfur protein *J. Biol. Chem.* **258** 2364–9
http://www.jbc.org/content/258/4/2364.abstract

[106] Zarzycki J, Brecht V, Müller M and Fuchs G 2009 Identifying the missing steps of the autotrophic 3-hydroxypropionate CO$_2$ fixation cycle in *Chloroflexus aurantiacus* *Proc. Natl Acad. Sci. USA* **106** 21317–22

[107] Kim J D, Rodriguez-Granillo A, Case D A, Nanda V and Falkowski P G 2012 Energetic selection of topology in ferredoxins *PLoS Comput. Biol.* **8** e1002463

[108] Chabrière E, Charon M H, Volbeda A, Pieulle L, Hatchikian E C and Fontecilla-Camps J C 1999 Crystal structures of the key anaerobic enzyme pyruvate:ferredoxin oxidoreductase, free and in complex with pyruvate *Nature Struct. Biol.* **6** 182–90

[109] Lombard J and Moreira D 2011 Early evolution of the biotin-dependent carboxylase family *BMC Evol. Biol.* **11** (232) 1–22

[110] Aoshima M, Ishii M and Igarashi Y 2004 A novel biotin protein required for reductive carboxylation of 2-oxoglutarate by isocitrate dehydrogenase in *Hydrogenobacter thermophilus* TK-6 *Mol. Microbiol.* **51** 791–8

[111] Evans M C W, Buchanan B B and Arnon D I 1966 A new ferredoxin dependent carbon reduction cycle in photosynthetic bacterium *Proc. Natl Acad. Sci. USA* **55** 928–34

[112] Aoshima M, Ishii M and Igarashi Y 2004 A novel enzyme, citryl-coA synthetase, catalysing the first step of the citrate cleavage reaction in *Hydrogenobacter thermophilus* TK-6 *Mol. Microbiol.* **52** 751–61

[113] Aoshima M, Ishii M and Igarashi Y 2004 A novel enzyme, citryl-coA lyase, catalysing the second step of the citrate cleavage reaction in *Hydrogenobacter thermophilus* TK-6 *Mol. Microbiol.* **52** 763–70

[114] Schoepp-Cothenet B, Lieutaud C, Baymann F, Verméglio A, Friedrich T, Kramer D M and Nitschke W 2005 Menaquinone as a pool quinone in a purple bacterium *Proc. Natl Acad. Sci. USA* **106** 8549–54

[115] Tian J, Bryk R, Itoh M, Suematsu M and Nathan C 2005 Variant tricarboxylic acid cycle in *Mycobacterium tuberculosis*: identification of α-ketoglutarate decarboxylase *Proc. Natl Acad. Sci. USA* **102** 10670–5

[116] Fenchel T 2002 *Origin and Early Evolution of Life* (New York: Oxford University Press)

[117] Kasting J F 2006 Ups and downs of ancient oxygen *Nature* **443** 643–4

[118] Shields-Zhou G and Och L 2011 The case for a neoproterozoic oxygenation event: chemical evidence and biological consequences *GSA Today* **21** 4–11

[119] Guiral M, Tron P, Aubert C, Gloter A, Iobbi-Nivol C and Giuici-Orticoni M-T 2005 A membrane-bound multienzyme, hydrogen-oxidizing and sulfur-reducing complex from the hyperthermophilic bacterium *Aquifex aeolicus* *J. Biol. Chem.* **280** 42004–15

[120] Guiral M, Aubert T and Giudici-Orticoni MT 2005 Hydrogen metabolism in the hyperthermophilic bacterium *Aquifex aeolicus* *Int. Hydrogenases Conf. (Reading, UK, 24–29 Aug. 2004); Biochem. Soc. Trans.* **33** 22–4

[121] Guiral M, Prunetti L, Lignon S, Lebrun R, Moinier D and Giuici-Orticoni M-T 2009 New insights into the respiratory chains of the chemolithoautotrophic and hyperthermophilic bacterium *Aquifex aeolicus* *J. Proteome Res.* **8** 1717–30

[122] Wächtershäuser G 1990 Evolution of the first metabolic cycles *Proc. Natl Acad. Sci. USA* **87** 200–4

[123] Huber H, Gallenberger M, Jahn U, Eylert E, Berg I A, Kockelkorn D, Eisenreich W and Fuchs G 2008 A dicarboxylate/4-hydroxybutyrate autotrophic carbon assimilation cycle in the hyperthermophilic archaeum *Ignicoccus hospitalis* *Proc. Natl Acad. Sci. USA* **105** 7851–6

[124] Müh U, Cinkaya I, Albracht S P J and Buckel W 1996 4-Hydroxybutyryl-CoA dehydratase from *Clostridium aminobutyricum*: characterization of FAD and iron–sulfur clusters involved in an overall non-redox reaction *Biochemistry* **35** 11710–8

[125] Martins B M, Dobbek H, Cinkaya I, Buckel W and Messerschmidt A 2004 Crystal structure of 4-hydroxybutyryl-coA dehydratase: radical catalysis involving a [4Fe-4S] cluster and flavin *Proc. Natl Acad. Sci. USA* **101** 15645–9

[126] Alber B E, Kung J W and Fuchs G 2008 3-Hydroxypropionyl-coenzyme A synthetase from *Metallosphaera sedula* an enzyme involved in autotrophic CO$_2$ fixation *J. Bacteriol.* **190** 1383–9

[127] Teufel R, Jung J W, Kockelkorn D, Alber B E and Fuchs G 2009 3-hdyroxypropionyl-coenzyme A dehydratase and acroloyl-coenzyme A reductase, enzymes of the autotrophic 3-hydroxypropionate/4-hydroxybutyrate cycle in the *Sulfolobales* *J. Bacteriol.* **191** 4572–81

[128] Bassham J A, Benson A A, Kay L D, Harris A Z, Wilson A T and Calvin M 1954 The path of carbon in photosynthesis: XXI. The cyclic regeneration of carbon dioxide acceptor *J. Am. Chem. Soc.* **76** 1760–70

[129] Tabita F A 2004 Research on carbon dioxide fixation in photosynthetic microorganisms (1971–present) *Photosynth. Res.* **80** 315–32

[130] Eisenhut M, Kahlon S, Hasse D, Ewald R, Lieman-Hurwitz J, Ogawa T, Ruth W, Bauwe H, Kaplan A and Hagemann M 2006 The plant-like C2 glycolate cycle and the bacterial-like glycerate pathway cooperate in phosphoglycolate metabolism in cyanobacteria *Plant Physiol.* **142** 333–42

[131] Eisenhut M, Ruth W, Haimovich M, Bauwe H, Kaplan A and Hagemann M 2008 The photorespiratory glycolate metabolism is essential for cyanobacteria and might have been conveyed endosymbionically to plants *Proc. Natl Acad. Sci. USA* **105** 17199–204

[132] Foyer C H, Bloom A J, Queval G and Noctor G 2009 Photorespiratory metabolism: genes, mutants, energetics, and redox signaling *Annu. Rev. Plant Biol.* **60** 455–84

[133] Liu Y *et al* 2003 Transcriptome dynamics of deinococcus radiodurans recovering from ionizing radiation *Proc. Natl Acad. Sci.* **100** 4191–6

[134] von Wettstein D, Gough S and Kannangara C G 1995 Chlorophyll biosynthesis *Plant Cell* **7** 1039–57

[135] Braakman R and Smith E 2012 Whole-genome reconstruction and evolutionary context of the deep-branching autotrophic metabolism of *Aquifex aeolicus* in preparation

[136] Benner S A, Ellington A D and Tauer A 1989 Modern metabolism as a palimpsest of the RNA world *Proc. Natl Acad. Sci. USA* **18** 7054–8

[137] Davidson E H and Erwin D H 2006 Gene regulatory networks and the evolution of animal body plans *Science* **311** 796–800

[138] Erwin D H and Davidson E H 2009 The evolution of hierarchical gene regulatory networks *Nature Rev. Genet.* **10** 141–8

[139] Hetzel M, Brock M, Selmer T, Pierik A J, Golding B T and Buckel W 2003 Acryloyl-CoA reductase from *Clostridium propionicum*: an enzyme complex of propionyl-coA dehydrogenase and electron-transferring flavoprotein *Eur. J. Biochem.* **270** 902–10

[140] Herrmann G, Jayamani E, Mai G and Buckel W 2008 Energy conservation via electron-transferring flavoprotein in anaerobic bacteria *J. Bacteriol.* **190** 784–91

[141] Buckel W 1996 Unusual dehydrations in anaerobic bacteria: considering ketyls (radical anions) as reactive intermediates in enzymatic reactions *FEBS Lett.* **389** 20–4

[142] Gruer M J, Artymiuk P J and Guest J R 1997 The aconitase family: three structural variations on a common theme *Trends Biochem. Sci.* **22** 3–6

[143] Aoshima M and Igarashi Y 2006 A novel oxalosuccinate-forming enzyme involved in the reductive carboxylation of 2-oxoglutarate in *Hydrogenobacter thermophilus* TK-6 *Mol. Microbiol.* **62** 748–59

[144] Aoshima M and Igarashi Y 2008 Nondecarboxylating and decarboxylating isocitrate dehydrogenases: oxalosuccinate reductase as an ancestral form of isocitrate dehydrogenase *J. Bacteriol.* **190** 2050–5

[145] Copley S D 2003 Enzymes with extra talents: moonlighting functions and catalytic promiscuity *Curr. Opin. Chem. Biol.* **7** 265–72

[146] Meléndez-Hevia E and Isidoro A 1985 The game of the pentose phosphate cycle *J. Theor. Biol.* **117** 251–63

[147] Huber C and Günter W 2000 Activated acetic acid by carbon fixation on (Fe,Ni)S under primordial conditions *Science* **276** 245–7

[148] Morowitz H J, Srinivasan V and Smith E 2010 Ligand field theory and the origin of life as an emergent feature of the periodic table of elements *Biol. Bull.* **219** 1–6 PMID: 20813983

[149] Russell M J and Hall A J 2006 The onset and early evolution of life *Geol. Soc. Am. Mem.* **198** 1–32

[150] Fontecilla-Camps J C, Amara P, Cavazza C, Nicolet Y and Volbeda A 2009 Structure-function relationships of anaerobic gas-processing metalloenzymes *Nature* **460** 814–22

[151] Doukov T I, Iverson T M, Seravalli J, Ragsdale S W and Drennan C L 2002 A Ni-Fe-Cu center in a bifunctional carbon monoxide dehydrogenase/acetyl-coA synthase *Science* **298** 567–72

[152] Seravalli J, Weiwei G, Tam A, Strauss E, Begley T P, Cramer S P and Ragsdale S W 2003 Functional copper at the acetyl-coA synthase active site *Proc. Natl Acad. Sci.* **100** 3689–94

[153] Ragsdale S W 2009 Nickel-based enzyme systems *J. Biol. Chem.* **284** 18571–5

[154] Volbeda A, Charon M-H, Piras C, Hatchikian E C, Frey M and Fontecilla-Camps J C 1995 Crystal structure of the nickel-iron hydrogenase from *Desulfovibrio gigas* *Nature* **373** 580–7

[155] Peters J W, Lanzilotta W N, Lemon B J and Seefeldt L C 1998 X-ray crystal structure of the Fe-only hydrogenase (CpI) from *Clostridium pasteurianum* to 1.8 Angstrom resolution *Science* **282** 1853–8

[156] Georgiadis M M, Komiya H, Chakrabarti P, Woo D, Kornuc J J and Rees D C 1992 Crystallographic structure of the nitrogenase iron protein from *Azotobacter vinelandii* *Science* **257** 1653–9

[157] Kim J and Rees D C 1992 Structural models for the metal centers in the nitrogenase molybdenum–iron protein *Science* **257** 1677–82

[158] Lancaster K M, Roemelt M, Ettenhuber P, Yilin H, Ribbe M W, Neese F, Bergmann U and DeBeer S 2011 X-ray emission spectroscopy evidences a central carbon in the nitrogenase iron–molybdenum cofactor *Science* **334** 974–7

[159] Spatzal T, Aksoyoglu M, Zhang L, Andrade S L A, Schleicher E, Weber S, Rees D C and Einsle O 2011 Evidence for interstitial carbon in nitrogenase FeMo cofactor *Science* **334** 940

[160] Ricardo A, Carrigan M A, Olcott A N and Benner S A 2004 Borate minerals stabilize ribose *Science* **303** 196

[161] Weber A L 2000 Sugars as the optimal biosynthetic carbon substrate of aqueous life throughout the universe *Orig. Life Evol. Biosph.* **30** 33–43

[162] Weber A L 2001 Sugar model of the origin of life: Catalysis by amines and amino acid products *Orig. Life Evol. Biosph.* **31** 71–86

[163] Weber A L 2002 Chemical constraints governing the origin of metabolism: The thermodynamic landscape of carbon group transformations under mild aqueous conditions *Orig. Life Evol. Biosph.* **32** 333–57

[164] Weber A L 2004 Kinetics of organic transformations under mild aqueous conditions: implications for the origin of life and its metabolism *Orig. Life Evol. Biosph.* **34** 473–95

[165] Fischer E and Sauer U 2003 A novel metabolic cycle catalyzes glucose oxidation and anaplerosis in hungry *Escherichia coli* *J. Biol. Chem.* **278** 46446–51

[166] Beste D J V, Bonde B, Hawkins N, Ward J L, Beale M H, Noack S, Nöh K, Kruger N J, Ratcliffe R G and McFadden J 2011 13C metabolic flux analysis identifies an unusual route for pyruvate dissimilation in mycobacteria which requires isocitrate lyase and carbon dioxide fixation *PLoS Pathogens* **7** e1002091

[167] Markert S *et al* 2007 Physiological proteomics of the uncultured endosymbiont of *Riftia pachyptila* *Science* **315** 247–50

[168] Duarte N C, Becker S A, Jamshidi N, Thiele I, Mo M L, Vo T D, Srivas R and Palsson B O 2007 Global reconstruction of the human metabolic network based on genomic and bibliomic data *Proc. Natl Acad. Sci. USA* **104** 1777–82

[169] Kikuchi G 1973 The glycine cleavage system: composition, reaction mechanism, and physiological significance *Mol. Cell. Biochem.* **1** 169–87

[170] Barker H A and Beck J V 1941 The fermentative decomposition of purines by *Clostridium acidi-urici* and *Clostridium cylindrosporum* *J. Biol. Chem.* **141** 3–27 http://www.jbc.org/content/141/1/3

[171] Waber L J and Wood H G 1979 Mechanism of acetate synthesis from CO₂ by *Clostridium acidi-urici J. Bacteriol.* **140** 468–78 PMCID: PMC216671

[172] Vetsigian K, Woese C and Goldenfeld N 2006 Collective evolution and the genetic code *Proc. Natl Acad. Sci. USA* **103** 10696–701

[173] Peretó J, López-García P and Moreira D 2004 Ancestral lipid biosynthesis and early membrane evolution *Trends Biochem. Sci.* **29** 469–77

[174] Brazelton W J and Baross J A 2009 Abundant transposases encoded by the metagenome of a hydrothermal chimney biofilm *ISME J.* **3** 1420–4

[175] Ciccarelli F D, Doerks T, von Mering C, Creevey C J, Snel B and Bork P 2006 Toward automatic reconstruction of a highly resolved tree of life *Science* **311** 1283–7

[176] Skophammer R G, Servin J A, Herbold C W and Lake J A 2007 Evidence for a gram-positive, eubacterial root of the tree of life *Mol. Biol. Evol.* **24** 1761–8

[177] Kasting J F, Eggler D H and Raeburn S P 1993 Mantle redox evolution and the oxidation state of the archean atmosphere *J. Geol.* **101** 245–57

[178] Hazen R M, Papineau D, Bleeker W, Downs R T, Ferry J M, McCoy T J, Sverjensky D A and Yang H 2008 Mineral evolution *Am. Mineral.* **93** 1693–720

[179] Trail D, Watson E B and Tallby N D 2011 The oxidation state of hadean magmas and implications for early earth's atmosphere *Nature* **480** 79–82

[180] Anbar A D *et al* 2007 A whiff of oxygen before the great oxidation event? *Science* **317** 1903–6

[181] Ono S, Beukes N J, Rumble D and Fogel M L 2006 Early evolution of atmospheric oxygen from multiple-sulfur and carbon isotope records of the 2.9 Ga Mozaan Group of the Pongola Supergroup, Southern Africa *South Afr. J. Geol.* **109** 97–108

[182] Ohmoto H, Watanabe Y, Ikemi H, Poulson S R and Taylor B E 2006 Sulphur isotope evidence for an oxic archaean atmosphere *Nature* **442** 908–11

[183] Farquhar J, Peters M, Johnston D T, Strauss H, Masterson A, Wiechert U and Kaufman A J 2007 Isotopic evidence for Mesoarchaean anoxia and changing atmospheric sulphur chemistry *Nature* **449** 706–9

[184] Buick R 2008 When did oxygenic photosynthesis evolve? *Phil. Trans. R. Soc. B: Biol. Sci.* **363** 2731–43

[185] Kump L R 2008 The rise of atmospheric oxygen *Nature* **451** 277–8

[186] Sessions A L, Doughty D M, Welander P V, Summons R E and Newman D K 2009 The continuing puzzle of the great oxidation event *Curr. Biol.* **19** R567–74

[187] Brazelton W J 2010 Ecology of archaeal and bacterial biofilm communities at the Lost City hydrothermal field *PhD Thesis* University of Washington

[188] Pace N R 1997 A molecular view of microbial diversity and the biosphere *Science* **276** 734–40

[189] Martin W and Russell M J 2003 On the origin of cells: an hypothesis for the evolutionary transitions from abiotic geochemistry to chemoautotrophic prokaryotes, and from prokaryotes to nucleated cells *Phil. Trans. R. Soc. Lond.* B **358** 27–85

[190] Orgel L E 2008 The implausibility of metabolic cycles on the early earth *PLoS Biol.* **06** 0005–13

[191] Copley S D, Smith E and Morowitz H J 2010 The emergence of sparse metabolic networks *Abiogenesis and the Origins of Life* ed M Russell (Cambridge, MA: Cosmo. Science Publishers) pp 175–91

[192] Darwin C 1859 *On the Origin of Species* (London: John Murray)

[193] Gesteland R F, Cech T R and Atkins J F (ed) 2006 *The RNA World* 3rd edn (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press)

[194] Chistoserdova L, Vorholt J A, Thauer R K and Lidstrom M E 1998 C₁ transfer enzymes and coenzymes linking methylotrophic bacteria and methanogenic archaea *Science* **281** 99–102

[195] Vorholt J A, Chistoserdova L, Stolyar S M, Thauer R K and Lidstrom M E 1999 Distribution of tetrahydromethanopterin-dependent enzymes in methylotrophic bacteria and phylogeny of methenyl tetrahydromethanopterin cyclohydrolases *J. Bacteriol.* **181** 5750–7 http://jb.asm.org/content/181/18/5750.short

[196] Chistoserdova L, Kalyuzhnaya M G and Lidstrom M E 2009 The expanding world of methylotrophic metabolism *Annu. Rev. Microbiol.* **63** 477–99

[197] Chistoserdova L 2011 Modularity of methylotrophy, revisited *Environ. Microbiol.* **13** 2603–22

[198] Erb T J, Berg I A, Brecht V, Müller M, Fuchs G and Alber B E 2007 Synthesis of C5-dicarboxylic acids from C2-units involving crotonyl-coA carboxylase/reductase: the ethylmalonyl-CoA pathway *Proc. Natl Acad. Sci.* **104** 10631–6

[199] Erb T J, Brecht V, Fuchs G, Müller M and Alber B E 2009 Carboxylation mechanism and stereochemistry of crotonyl-CoA carboxylase/reductase, a carboxylating enoyl-thioester reductase *Proc. Natl Acad. Sci.* **106** 8871–6

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

[201] Raymond J and Segré D 2006 The effect of oxygen on biochemical networks and the evolution of complex life *Science* **311** 1764–7

[202] Schütte M, Skupin A, Segré D and Ebenhöh O 2010 Modeling the complex dynamics of enzyme-pathway coevolution *Chaos* **20** 045115

[203] Graham D E and White R H 2002 Elucidation of methanogenic coenzyme biosyntheses: from spectroscopy to genomics *Nat. Prod. Rep.* **19** 133–47

[204] Begley T P, Chatterjee A, Hanes J W, Hazra A and Ealick S E 2008 Cofactor biosynthesis still yielding fascinating new biological chemistry *Curr. Opin. Chem. Biol.* **12** 118–25

[205] Jurgenson C T, Begley T P and Ealick S E 2009 The structural and biochemical foundations of thiamin biosynthesis *Annu. Rev. Biochem.* **78** 569–603

[206] Huang F., Bugg C W and Yarus M 2000 RNA-catalyzed CoA, NAD, and FAD synthesis from phosphopantetheine, NMN, and FMN *Biochemistry* **39** 15548–55

[207] Crosa J H and Walsh C T 2002 Genetics and assembly line enzymology of siderophore biosynthesis in bacteria *Microbiol. Mol. Biol. Rev.* **66** 223–49

[208] Butler A 2005 Marine siderophores and microbial iron mobilization *BioMetals* **18** 369–74

[209] Westheimer F H 1987 Why nature chose phosphates *Science* **235** 1173–8

[210] White H B 1976 Coenzymes as fossils of an earlier metabolic state *J. Mol. Evol.* **7** 101–4

[211] Petsko G A and Ringe D 2003 *Protein Structure and Function* (London: New Science Press)

[212] Gutteridge A and Thornton J M 2005 Understanding nature's catalytic toolkit *Trends Biochem. Sci.* **30** 622–9

[213] Fischer J D, Holliday G L, Rahman S A and Thornton J M 2010 The structures and physicochemical properties of organic cofactors in biocatalysis *J. Mol. Biol.* **403** 803–24

[214] Andreini C, Bertini I, Cavallaro G, Holliday G L and Thornton J M 2008 Metal ions in biological catalysis: from enzyme databases to general principles *J. Biol. Inorg. Chem.* **13** 1205–18

[215] Andreini C, Bertini I, Cavallaro G, Holliday G L and Thornton J M 2009 Metal-MACiE: a database of

metals involved in biological catalysis *Bioinform. Appl.* **25** 2088–9

[216] Ragsdale S W and Wood H G 1991 Enzymology of the acetyl-coa pathway of CO$_2$ fixation *Crit. Rev. Biochem. Mol. Biol.* **26** 261–300

[217] Ragsdale S W and Kumar M 1996 Nickel-containing carbon monoxide dehydrogenase/acetyl-CoA synthase *Chem. Rev.* **96** 2515–40

[218] Ragsdale S W 2008 Enzymology of the Wood–Ljungdahl pathway of acetogenesis *Ann. NY Acad. Sci.* **1125** 129–36

[219] Fry I 2000 *The Emergence of Life on Earth: A Historical and Scientific Overview* (New Brunswick, NJ: Rutgers University Press)

[220] Yarus M 2011 Getting past the RNA world: the initial Darwinian ancestor *Cold Spring Harbor Perspectives in Biology* ed J Atkins, R Gesteland and T Cech (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press) pp 1–8

[221] MacKenzie R E 1984 Biogenesis and interconversion of substituted tetrahydrofolates *Folates and Pterins: Chemistry and Biochemistry of Folates* vol 1 ed R L Blakely and S J Benkovic (New York: Wiley) pp 255–306

[222] Kallen R G and Jencks W P 1966 The dissociation constants of tetrahydrofolic acid *J. Biol. Chem.* **241** 5845–50 PMID: 5654362

[223] Stover P and Schirch V 1993 The metabolic role of leucovorin *Trends Biochem. Sci.* **18** 102–6

[224] Huang T and Schirch V 1995 Mechanism for the coupling of atp hydrolysis to the conversion of 5-formyltetrahydrofolate to 5,10-methenyltetrahydrofolate *J. Biol. Chem.* **270** 22296–300

[225] Collins M D and Jones D 1981 Distribution of isoprenoid quinone structural types in bacteria and their taxonomic implication *Microbiological Rev.* **45** 316–54 PMCID: PMC281511

[226] Nitscke W, Kramer D M, Riedel A and Liebl U 1995 From naptho- to benzoquinones-(r)evolutionary reorganizions of electron transfer chains *Photosynthesis: From Light to the Biosphere* vol 1 ed P Mathis (Dordrecht: Kluwer) pp 945–50

[227] Wächtershäuser G 1992 Groundworks for an evolutionary biochemistry: the iron–sulphur world *Prog. Biophys. Molec. Biol.* **58** 85–201

[228] Tazuya K, Azumi C, Yamada K and Kumaoka H 1995 Pyrimidine moiety of thiamin is biosynthesized from pyridoxine and histidine in *Saccharomyces cerevisiae*, *Biochem. Mol. Biol. Int.* **36** 883–8

[229] Bentley R and Haslam E 1990 The shikimate pathway a metabolic tree with many branches *Crit. Rev. Biochem. Mol. Biol.* **25** 307–84

[230] Barbas C F III 2008 Organocatalysis lost: modern chemistry, ancient chemistry, and an unseen biosynthetic apparatus *Angew. Chemie Int. Ed.* **47** 42–47

[231] MacMillan D W C 2008 The advent and development of organocatalysis *Nature* **455** 304–8

[232] Powner M W, Béatrice G and Sutherland J D 2009 Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions *Nature* **459** 239–42

[233] Benner S A, Carrigan M A, Ricardo A and Frye F 2006 Setting the stage: the history, chemistry, and geobiology behind RNA *The RNA World* ed R F Gesteland, T R Cech and J F Atkins (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press) pp 1–22

[234] Décout J-L and Maurel M-C 1993 N6-substituted adenine derivatives and RNA primitive catalysts *Orig. Life Evol. Biosph.* **23** 299–306

[235] Stryer L 1981 *Biochemistry* 2nd edn (San Francisco, CA: Freeman)

[236] Lincoln T A and Joyce G F 2009 Self-sustained replication of an RNA enzyme *Science* **323** 1229–32

[237] Copley S D, Smith E and Morowitz H J 2005 A mechanism for the association of amino acids with their codons and the origin of the genetic code *Proc. Natl Acad. Sci. USA* **102** 4442–7

[238] Segré D and Lancet D 1999 A statistical chemistry approach to the origin of life *Chemtracts –Biochem. Mol. Biol.* **12** 382–97

[239] Segré D, Ben-Ali D and Lancet D 2000 Compositional genomes: Prebiotic information transfer in mutually catalytic noncovalent assemblies *Proc. Natl Acad. Sci. USA* **97** 4112–7

[240] Segré D, Shenhav B, Kafri R and Lancet D 2001 The molecular roots of compositional inheritance *J. Theor. Biol.* **213** 481–91

[241] de Duve C 1991 *Blueprint for a Cell* (Burlington, NC: Neil Patterson)

[242] Gray H B 1994 *Chemical Bonds: An Introduction to Atomic and Molecular Structure* (Sausalito, CA: University Science Press)

[243] Wald G 1962 Life in the second and third periods: or why phosphorus and sulfur for high-energy bonds *Horizons in Biochemistry* ed M Kasha and B Pullman (New York: Academic) pp 127–42

[244] Hohmann-Marriott M F and Blankenship R E 2011 Evolution of photosynthesis *Annu. Rev. Plant. Biol.* **62** 515–48

[245] Berkovitch F, Nicolet Y, Wan J T, Jarrett J T and Drennan C L 2004 Crystal structure of biotin synthase, an s-adenosylmethionine-dependent radical enzyme *Science* **303** 76–9

[246] Cody G D, Boctor N Z, Hazen R M, Brandeis J A, Morowitz H J and Jr Yoder H S 2001 Geochemical roots of autotrophic carbon fixation: hydrothermal experiments in the system citric acid, H$_2$O-(±FeS)(±NiS) *Geochim. Cosmochim. Acta.* **65** 3557–76

[247] Balche W E and Wolfe R S 1979 Specificity and biological distribution of coenzyme M (2-mercaptoethanesulfonic acid) *J. Bacteriol.* **137** 256–63 PMID: 104960

[248] Danson M J 1993 Central metabolism of the archaea *The Biochemistry of Archaea* ed M Kates, D J Kushner and A T Matheson (Amsterdam: Elsevier) pp 1–24

[249] Kaster A-K, Moll J, Parey K and Thauer R K 2012 Coupling of ferredoxin and heterodisulfide reduction via electron bifurcation in hydrogenotrophic methanogenic archaea *Proc. Natl Acad. Sci. USA* **108** 2981–6

[250] Garhardt A, Çinkaya I, Linder D, Hulsman G and Buckel W 2000 Fermentation of 4-aminobutyrate by *Clostridium aminobytyricum*: cloning of two genes involved in the formation and dehydration of 4-hydroxybutyryl-coA *Arch. Microbiol.* **174** 189–99

[251] Schutz M, Schoepp-Cothenet B, Lojou E, Woodstra M, Lexa D, Tron P, Dolla A, Durand M-C, Stetter K O and Baymann F 2003 The naphthoquinol oxidizing cytochrome BC1 complex of the hyperthermophilic knallgasbacterium *aquifex aeolicus*: properties and phylogenetic relationships *Biochemistry* **42** 10800–8

[252] Fuli L, Hinderberger J, Seedorf H, Zhang J, Buckel W and Thauer R K 2008 Coupled ferredoxin and crotonyl coenzyme A (CoA) reduction with NADH catalyzed by the butyryl-CoA dehydrogenase/ETF complex from *Clostridium kluyveri* *J. Bacteriol.* **190** 843–50

[253] Wang S, Huang H, Moll J and Thauer R K 2010 NADP+ reduction with reduced ferredoxin and NADP+ reduction with NADH are coupled via an electron-bifurcating enzyme complex in *Clostridium kluyveri* *J. Bacteriol.* **192** 5115–23

[254] Martin W F 2012 Hydrogen, metals, bifurcating electrons, and proton gradients: the early evolution of biological energy conservation *FEBS Lett.* **586** 485–93

[255] Kesen M, Hutchins A, Sung S-J S and Adams M W W 1997 Pyruvate ferredoxin oxidoreductase from the hyperthermophilic archaeon, *Pyrococcus furiosus* functions as a CoA-dependent pyruvate decarboxylase *Proc. Natl Acad. Sci. USA* **94** 9608–13

[256] Cronan J E Jr and Waldrop G L 2002 Multi-subunit acetyl-CoA carboxylases *Prog. Lipid Res.* **41** 407–35

[257] Zhang H, Yang Z, Shen Y and Tong L 2003 Crystal structure of the carboxyltransferase domain of acetyl-coenzyme a carboxylase *Science* **299** 2064–7

[258] Kuki A and Wolynes P G 2000 Electron tunneling paths in proteins *Science* **236** 1647–52

[259] Jensen R A 1976 Enzyme recruitment in evolution of new function *Annu. Rev. Microbiol.* **30** 409–25

[260] Khersonsky O and Tawfik D S 2010 Enzyme promiscuity: a mechanistic and evolutionary perspective *Annu. Rev. Biochem.* **79** 471–505

[261] Kim J, Kershner J P, Novikov Y, Shoemaker R K and Copley S D 2010 Three serendipitous pathways in *E. coli* can bypass a block in pyridoxal-5′-phosphate synthesis *Mol. Sys. Biol.* **6** 436

[262] Nam H, Lewis N E, Lerman J A, Lee D-H, Chang R L, Kim D and Palsson B O 2012 Network context and selection in the evolution to enzyme specificity *Science* **337** 1101–4

[263] Khersonsky O, Malitsky S, Rogachev I and Tawfik D S 2011 Role of chemistry versus substrate binding in recruiting promiscuous enzyme functions *Biochem.* **50** 2683–90

[264] Braithwaite D K and Ito J 1993 Compilation, alignment, and phylogenetic relationships of DNA polymerases *Nucl. Acids Res.* **21** 787–802

[265] Bailey S, Wing R A and Steitz T A 2006 The structure of *T. aquaticus* DNA polymerase III is distinct from eukaryotic replicative DNA polymerases *Cell* **126** 893–904

[266] Horowitz N H 1945 On the evolution of biochemical synthesis *Proc. Natl Acad. Sci. USA* **31** 153–7

[267] Crick F 1970 Central dogma of molecular biology *Nature* **227** 561–3

[268] Haldane J B S 1967 *The Origin of Life* ed J D Bernal (London: Weidenfeld and Nicolson) pp 242–9

[269] Gould S J 2002 *The Structure of Evolutionary Theory* (Cambridge, MA: Harvard University Press)

[270] Wilson E B 1925 *The Cell in Development and Inheritance* 3rd edn (New York: Macmillan)

[271] Poissot T, Bever J D, Nemri A , Thrall P H and Hochberg M E 2011 A conceptual framework for the evolution of ecological specialization *Ecol. Lett.* **14** 841–51

[272] Nelson D L and Cox M M 2004 *Lehninger Principles of Biochemistry* 4th edn (San Francisco, CA: Freeman)

[273] Alberts B 2002 *Molecular Biology of the Cell* 4th edn (New York: Garland Science)

[274] Morowitz H J 1987 *Foundations of Bioenergetics* (New York: Academic)

[275] Bershad A K, Fuentes M A and Krakauer D C 2008 Developmental autonomy and somatic niche construction promotes robust cell fate decisions *J. Theor. Biol.* **254** 408–16

[276] Claverie J-M 2006 Viruses take center stage in cellular evolution *Genome Biol.* **7** 1–5

[277] Forterre P 2010 Defining life: the virus viewpoint *Orig. Life Evol. Biosph.* **40** 151–60

[278] Kelley D S, Baross J A and Delaney J R 2002 Volcanoes, fluids, and life at mid-ocean ridge spreading centers *Annu. Rev. Earth Planet. Sci.* **30** 385–491

[279] Russell M J and Hall A J 1997 The emergence of life from iron monosulphide bubbles at a submarine hydrothermal redox and pH front *J. Geol. Soc. Lond.* **154** 377–402

[280] Baltscheffsky H, von Stedingk L-V, Heldt H-W and Klingenberg M 1966 Inorganic pyrophosphate: formation in bacterial photophosphorylation *Science* **153** 1120–2

[281] Yamagata Y, Watanabe H, Saitoh M and Namba T 1991 Volcanic production of polyphosphates and its relevance to prebiotic evolution *Nature* **352** 516–9

[282] Kornberg A, Rao N N and Ault-Riché D 1999 Inorganic polyphosphate: a molecule of many functions *Annu. Rev. Biochem.* **68** 89–125

[283] Baltscheffsky M, Schultz A and Baltscheffsky H 1999 H$^+$-PPases: a tightly membrane-bound family *FEBS Lett.* **457** 527–33

[284] Brown M R W and Kornberg A 2004 Inorganic polyphosphate in the origin and survival of species *Proc. Natl Acad. Sci. USA* **101** 16085–7

[285] Hanczyk M M, Fujikawa S M and Szostak J W 2003 Experimental models of primitive cellular compartments: encapsulation, growth, and division *Science* **302** 618–22

[286] Luisi P L 2006 *The Emergence of Life: from Chemical Origins to Synthetic Biology* (London: Cambridge University Press)

[287] Freeland S J and Hurst L D 1998 The genetic code is one in a million *J. Mol. Evol.* **47** 238–48

[288] Knight R D, Freeland S J and Landweber L F 1999 Selection, history and chemistry: the three faces of the genetic code *Trends Biochem. Sci.* **24** 241–7

[289] Yi L and Freeland S 2006 On the evolution of the standard amino-acid alphabet *Gen. Biol.* **7** 102

[290] Wong T-F 1975 A co-evolution theory of the genetic code *Proc. Natl Acad. Sci. USA* **72** 1909–12

[291] Woese C R, Dugre D H, Saxinger W C and Dugre S A 1966 The molecular basis for the genetic code *Proc. Natl Acad. Sci. USA* **55** 966–74

[292] Wächtershäuser G 1988 Before enzymes and templates: a theory of surface metabolism *Microbiol. Rev.* **52** 452–84 PMCID: PMC373159

[293] Shapiro R 2006 Small molecule interactions were central to the origin of life *Q. Rev. Biol.* **81** 105–25

[294] Shapiro R 2007 A simpler origin for life *Sci. Am.* **12** 24–31

[295] Heinen W and Lauwers A M 1996 Organic sulfur compounds resulting from the interaction of iron sulfide, hydrogen sulfide and carbon dioxide in an anaerobic aqueous environment *Orig. Life Evol. Biosph.* **26** 131–50 10.1007/BF01809852

[296] Oparin A I 1967 *The Origin of Life* ed J D Bernal (London: Weidenfeld and Nicolson) pp 199–234

[297] Miller S L 1953 Production of amino acids under possible primitive earth conditions *Science* **117** 528–9

[298] Hordijk W and Steel M 2012 Autocatalytic sets extended: dynamics, inhibition, and a generalization *J. Syst. Chem.* **3** 5

[299] Girvan M and Newman M E J 2002 Community structure in social and biological networks *Proc. Natl Acad. Sci. USA* **99** 7821–6

[300] Berge C 1973 *Graphs and Hypergraphs* (Amsterdam: North-Holland) (revised edn)

[301] Price N D, Famili I, Beard D A and Palsson B O 2002 Extreme pathways and Kirchhoff's second law *Biophys. J.* **83** 2879–82

[302] Famili I and Palsson B O 2003 Systemic metabolic reactions are obtained by singular value decomposition of genome-scale stoichiometric matrices *J. Theor. Biol.* **224** 87–96

[303] Palsson B O 2006 *Systems Biology* (Cambridge: Cambridge University Press)

[304] Feist A M, Herrgard M J, Thiele I, Reed J L and Palsson B O 2009 Reconstruction of biochemical networks in microorganisms *Nature Rev. Microbiol.* **7** 129–43

[305] Sinanŏglu O 1975 Theory of chemical reaction networks. All possible mechanisms or synthetic pathways with given number of reaction steps or species *J. Am. Chem. Soc.* **97** 2309–20

[306] Sinanŏglu O 1984 On the algebraic construction of chemistry from quantum mechanics. A fundamental valency vector field defined on the Euclidean 3-space and its relation to the Hilbert space *Theor. Chim. Acta* **65** 243–8

[307] Sinitsyn N A, Hengartner N and Nemenman I 2009 Adiabatic coarse-graining and simulations of stochastic biochemical networks *Proc. Natl Acad. Sci. USA* **106** 10546–51

[308] Miller S L and Smith-Magowan D 1990 The thermodynamics of the Krebs cycle and related compounds *J. Phys. Chem. Ref. Data.* **19** 1049–73

[309] Ragsdale S 2012 private communication

[310] Doyle J 2010 private communication

[311] Schuster P 2001 private communication