ANNUAL REVIEWS

Annual Review of Biochemistry

Metabolite–Enzyme Coevolution: From Single Enzymes to Metabolic Pathways and Networks

Lianet Noda-Garcia,[1] Wolfram Liebermeister,[2,3] and Dan S. Tawfik[1]

[1]Department of Biomolecular Sciences, Weizmann Institute of Science, Rehovot 76100, Israel; email: Lianet.Noda@weizmann.ac.il, Dan.Tawfik@weizmann.ac.il

[2]INRA, Unité MaIAGE, 78352 Jouy en Josas, France; email: wolfram.liebermeister@inra.com

[3]Institute of Biochemistry, Charité Universitätsmedizin, Berlin, 10117 Berlin, Germany

Annu. Rev. Biochem. 2018. 87:187–216

The *Annual Review of Biochemistry* is online at biochem.annualreviews.org

https://doi.org/10.1146/annurev-biochem-062917-012023

Keywords

coevolution, enzyme evolution, evolutionary biochemistry, metabolism, pathway evolution, promiscuity

Abstract

How individual enzymes evolved is relatively well understood. However, individual enzymes rarely confer a physiological advantage on their own. Judging by its current state, the emergence of metabolism seemingly demanded the simultaneous emergence of many enzymes. Indeed, how multicomponent interlocked systems, like metabolic pathways, evolved is largely an open question. This complexity can be unlocked if we assume that survival of the fittest applies not only to genes and enzymes but also to the metabolites they produce. This review develops our current knowledge of enzyme evolution into a wider hypothesis of pathway and network evolution. We describe the current models for pathway evolution and offer an integrative metabolite–enzyme coevolution hypothesis. Our hypothesis addresses the origins of new metabolites and of new enzymes and the order of their recruitment. We aim to not only survey established knowledge but also present open questions and potential ways of addressing them.

Contents

1. INTRODUCTION

As beautifully illustrated by Hochachka & Somero's book (1, 2), evolutionary biochemistry addresses the origins of biochemical adaptations, that is, how various biochemical traits, from single enzymes to entire metabolic networks, evolved to meet specific physiological and environmental challenges. Evolutionary biochemistry is an integrated approach that combines different specialized disciplines, including biochemistry, biophysics, molecular biology, molecular evolution, population genetics, and comparative genomics (3–6). This integration also yields a deeper understanding of biochemical processes and the biomolecules that mediate them. The term evolutionary biochemistry was originally used in the context of origins of life, emphasizing that the chemistry of life evolved alongside its biological elements (7, 8). Along this vein, we argue that biochemical adaptations should not be seen as enzymes evolving to fit a given preexisting set of metabolites and reactions. Rather, the enzymes and the metabolites coevolved, incrementally, to yield the biochemistry we see today.

The evolution of single proteins, including enzymes, is the most well studied aspect of evolutionary biochemistry. However, proteins rarely confer a physiological advantage on their own. Individual enzymes are typically part of a pathway or even a larger network that operates in concert to yield advantageous phenotypes. Biosynthetic pathways, for example, comprise a series of interlocked enzymes—the loss of any individual enzyme abolishes pathway productivity. Thus, judging by their current state, the emergence of metabolic pathways demanded a simultaneous recruitment of multiple enzymes. How such complex, interlocked systems emerged is largely unknown. Although a gradual emergence of pathways of specialized metabolism could be envisaged, also given that these pathways evolved at advanced stages with many different enzymes already present, the emergence of the pathways that now comprise core and central metabolism, or even intermediate metabolism, is particularly enigmatic (7).

Foremostly, this review describes the currently known models for pathway evolution (Section 4) and integrates them into a unifying metabolite–enzyme coevolution hypothesis (Section 5). To lay the background for this hypothesis, we summarize the current knowledge with respect to the evolutionary divergence of individual enzymes (Section 2) while focusing on aspects relevant to the evolution of the higher levels of biochemical complexity, namely, metabolic pathways and networks. Although, as described in Section 3, contemporary metabolism is best described as a network of thousands of interconnected metabolites and enzymes. These networks, however, evolved from a much simpler beginning: individual founder enzymes and reaction modules or pathways. We thus outline the four currently known models for pathway evolution and the evidence in support of each model, starting from Horowitz's retrograde model and ending with the patchwork model (Section 4). We then propose a metabolite–enzyme coevolution model that integrates the current models into a unifying hypothesis (Section 5). The last part, Section 6, describes more far-fetched implications of evolutionary biochemistry and some standing questions and potential ways of addressing them.

Core metabolism: the main carbon and energy flow (e.g., tricarboxylic acid, pentose-phosphate, glycolysis)

Central metabolism: biosynthesis and catabolism of major amino acids, fatty acids, and nucleotides

Intermediate metabolism: biosynthesis and degradation of common cellular components, such as cofactors and coenzymes

Founder enzyme: a primordial enzyme that emerged individually, that is, prior to the appearance of other enzymes that presently support the same pathway(s)

Reaction module: a sequence of two or more chemical transformations performed on a variety of chemically similar substrates

2. THE DIVERGENCE OF INDIVIDUAL ENZYMES

This topic has many facets. We apologize for providing a brief extract addressing the biochemical aspect only: how enzymes with new reaction or/and substrate specificities diverge from existing ones. At this stage, we focus on the principles; specific examples follow, where the models of pathway evolution are described (Section 4).

2.1. Generalist–Specialist Transitions

New enzymes originate from promiscuous activities of existing enzymes (4, 9–11). Enzymes, even the most specific ones, exhibit coincidental, latent enzymatic activities that have no physiological relevance. Promiscuity involves the same chemistry performed on alternative substrates, but also alternative reactions, as elaborated in Section 2.3. Once a promiscuous activity provides some physiological advantage, it comes under selection and may further evolve. Typically, at the early stages of divergence, a bifunctional enzyme arises that catalyzes both the original and the new reaction (12–14). Bifunctional or multifunctional enzymes are dubbed generalists, and many enzymes stay generalist (15). Generalists tend to exhibit a wider range of promiscuous activities, including novel activities that were not present in the original enzyme prior to divergence (16, 17). These novel activities open the door to yet other new enzymes (16, 17). At later stages, via gene duplication, specialization can occur whereby the ancestral function splits between the original gene and its duplicated copy (subfunctionalization) (14). Duplication and divergence to specialists is a dominant trend, also apparent in an increasing number of paralog genes along evolutionary time. However, generalists can prevail for long periods, and specialist enzymes also diverge to

become generalists (18). Many of today's enzymes are generalists (19–21), and as discussed below, the ancestral, founder enzymes are likely to have been generalists as well (Section 4.3).

Shared chemistry: divergence of new enzymes by coopting a key chemical step and the key active-site residues that catalyze this step

Shared substrate: divergence of new enzymes by virtue of the existing and new enzyme sharing a common substrate motif

2.2. Enzyme Families, Superfamilies, and Classes

The enzyme repertoire expanded via specialist-to-generalist transitions (promiscuous functions coming under selection) and via the reverse process, that is, generalist-to-specialist transitions (duplications and subfunctionalization). This trend is also visible in the hierarchical classification of the current enzyme repertoire into families, superfamilies, and classes (22). Families encompass enzymes that catalyze the same reaction using identical or very similar substrates. Sequences of enzyme family members show relatively high identity across the entire protein length (typically $\geq$30%), and structures are nearly identical. Superfamilies combine several, or sometimes many, families that share the same overall structure (fold) and a key chemical step that corresponds to a common active-site architecture. The latter is manifested in conserved sequence motifs that are observed within superfamilies despite a low overall sequence similarity. Given these common features, superfamilies represent a gradual expansion from a common generalist ancestor that diverged to give a range of different families with different functions (23).

Enzyme classes comprise the top of this evolutionary hierarchy. A class typically encompasses several superfamilies that share a fold and a functional motif that typically involves few residues (e.g., a catalytic dyad, or cofactor binding residues). Common ancestry is evident in several enzyme classes, such as the Rossmann-fold enzyme class that includes >300 different families (24). However, divergence of enzyme classes from a single common ancestor is harder to prove because folds and functional motifs can also arise independently through convergent evolution. For example, many enzyme superfamilies share the triosephosphate isomerase (TIM)-barrel fold. However, while some of these superfamilies likely share common ancestry, others seem to have independently converged onto this fold (25, 26).

2.3. Shared Chemistry and the Divergence of Substrate Binding

Shared active-site architecture (i.e., a set of key residues arranged in a given geometry), and hence shared catalytic chemistry (hereafter, shared chemistry), is the hallmark of enzyme superfamilies (23, 27–29). Accordingly, promiscuous acceptance of substrates other than the original one (substrate ambiguity) is far more frequent than catalysis of a different reaction type (catalytic promiscuity) (9, 30). However, a given catalytic chemistry can mediate many different types of reactions (e.g., 31). For example, the enolase superfamily shares the abstraction of a proton from a carbon next to a carboxylate group. Yet the enzymatic functions are diverse, including isomerases, racemases, dehydratases, and lactone-forming enzymes (28).

Although shared chemistry is considered the hallmark of enzyme superfamilies, substrate binding is a prerequisite for the emergence of a new reaction. Thus, a substrate motif that is frequently recognized (e.g., a phosphate moiety) greatly increases the likelihood that a substrate will promiscuously bind preexisting enzymes (30). Shared substrate binding motifs often mediate the use of cofactors or cosubstrates. For example, the Rossmann-fold enzyme class utilizes various cosubstrates, such as NAD(P), FAD, *S*-adenosylmethionine (SAM), and NTPs, and accordingly includes enzymes that catalyze completely different chemical reactions, such as oxidation–reduction, methylation, and phosphoryl transfer. These cosubstrates are all ribonucleosides. The shared Rossmann motif is, accordingly, a ribose binding motif (an Asp/Glu that binds the ribose 2′ and 3′ hydroxyls via a bidentate interaction) (24). Promiscuous binding of cofactors or cosubstrates may therefore enable the divergence of an enzyme with a new chemistry (24, 32, 33). That shared substrate

binding can drive enzyme evolution is also manifested in the abundance of self-inhibition, i.e., of metabolites often inhibiting the enzymes that produce them (34).

Overall, divergence of substrate specificity while maintaining the reaction's chemistry is the common scenario in enzyme evolution (e.g., 23, 28, 35). As elaborated in Section 4, this mode of divergence prevails in the evolution of some pathways (Section 4.3). However, in others, one pathway enzyme may have diverged into another one by virtue of a shared substrate binding motif (Sections 4.1 and 4.2).

3. METABOLISM: FROM NETWORKS TO MODULES

Metabolism is often viewed as a collection of individual enzymes assembled in linear pathways or cycles, but metabolism is actually a broad network of thousands of enzymes and metabolites that are extensively interconnected. This network architecture complicates the understanding of how metabolism evolved, since pathways cannot be considered in isolation. Still, certain characteristics of metabolic networks today indicate their gradual evolutionary expansion from a much simpler beginning.

3.1. The Connectivity of Metabolic Networks Is Highly Skewed

Metabolic networks consist of metabolites and the biochemical reactions that connect them. To study their topology, we can depict metabolites as nodes and link them via the reactions they participate in (36, 37). The connectivity of each metabolite describes the number of metabolic reactions in which it takes part, and hub metabolites are those utilized by many enzymes. The average connectivity of metabolic networks is high (38, 39). For example, in a reconstructed metabolic model of *Escherichia coli*, the average connectivity is 5.8. However, the median connectivity is 2, indicating a highly skewed distribution. The three most connected metabolites (ATP, ADP, and phosphate) participate in >1,000 reactions. On the lower end, >500 different metabolites each participate in only a few reactions (**Figure 1**). Alternatively, enzymes can be depicted as nodes and linked if they share a reactant. Accordingly, the most connected enzyme in *E. coli*'s metabolic network is ATP synthase (37).

Overall, like many other biological traits (36), the connectivity distributions for reactions, enzymes, and metabolites seem to obey a power law, as illustrated in **Figure 1**. The connectivity values span over three orders of magnitude only, and hence they may fit alternative models (41). Nonetheless, in the evolutionary context, the distribution of connectivity supports a rich-get-richer mode of evolutionary expansion of metabolic networks.

3.2. Increased Complexity and Connectivity with Evolutionary Time

The power law distribution of metabolic networks has been associated with design principles such as robustness to physiological and genetic perturbations (36). However, this distribution, and many other traits of metabolic networks, may result from the evolutionary mechanisms that led to their emergence and expansion, mechanisms by which preferential attachment is the most accessible route of divergence. New enzymes diverge from existing ones via promiscuous acceptance of alternative substrates and catalysis of alternative reactions (Section 2.1). The most commonly used metabolites are recognized by many preexisting enzymes, and therefore they are also likely to be promiscuously accepted and generate new products of evolutionary potential. These promiscuous enzymes will then diverge to give rise to new enzymes that utilize the same, commonly used metabolite—thus, the rich get richer. Further, divergence of new enzymes typically occurs

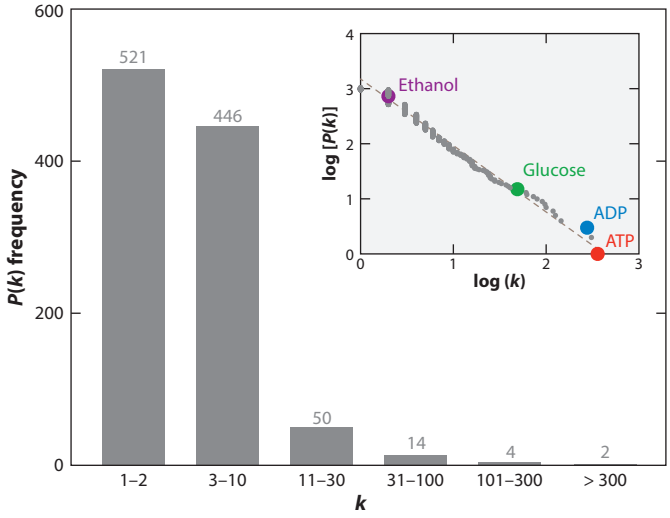


Figure 1

The distribution of metabolite connectivities. Data were derived from a stoichiometric metabolic model of *Escherichia coli* (6,007 different reactions and 1,037 different metabolites) (40). Metabolites were categorized by the connectivity k, the number of enzymatic reactions in which they participate. The value $k = 1$ designates end product metabolites, which are not utilized as substrates by any of the enzymes included in the model; the maximal value, $k = 359$, is for ATP. The number of different metabolites with a given k value [$P(k)$] is presented on the y axis. (*Inset*) A linear fit of the log values of k and $P(k)$ (power-law; although other models may fit equally well). Cosubstrates (FAD, NAD, ATP, CoA, etc.) were included; cofactors that are not part of the reaction per se, for example, pyridoxal phosphate or metal ions, were excluded, as were protons and water.

via generalist intermediates. Multifunctional enzymes are therefore common (15, 42), further increasing connectivity as metabolic networks expand.

The rich-get-richer mode of evolutionary expansion is also apparent in the fact that strictly linear metabolic pathways, that is, pathways whose intermediate metabolites do not participate in any other pathway, are a rare exception. Approximately 93% of *E. coli's* metabolic reactions involve metabolites that participate in ≥10 reactions (**Figure 1**). Accordingly, trends of recent evolutionary extensions to older enzymes have been observed (43). Thus, although we lack systematic phylogeny-based analyses, it seems that over evolutionary time the vast majority of new pathways evolved via recruitment of the most highly utilized metabolites on the one hand, combined with newly emerging metabolites on the other.

3.3. Underground Metabolism

The above discussion of network architectures and their evolution relates to the known, official metabolic networks. However, underneath the known enzymatic reactions that occur with measurable fluxes, there exists a so-called underground network of latent side reactions (via enzyme promiscuity, as discussed in Section 2.1, and nonenzymatic reactions, as discussed in Section 5.2). Within underground metabolism, a set of preexisting enzymes may produce the same end product via alternative ways (44–46), and may also produce a myriad of yet unidentified compounds that may not have a physiological function. As discussed in Section 5, this molecular messiness can be a basis for new pathways to emerge (47). As metabolic networks expand, the potential for cross-wiring, cross-reactions, and mutual inhibition also increases. Thus, further expansion of metabolic networks demands tighter enzyme regulation and subcompartmentalization (34).

3.4. Evolutionary Hierarchy of Metabolism: Founder Enzymes and Reaction Modules

The gradual expansion of metabolism can be schematically divided into three levels. The highest, third level, is the metabolic network itself. Pathways and reaction modules comprise the second hierarchical level. Pathways are somewhat artificially defined by the choice of starting and ending metabolites. Thus, while it is operatively useful to refer to pathway evolution, metabolic pathways as defined now (lysine biosynthesis, glycolysis, etc.) were not necessarily the earliest building blocks of metabolism. The second level therefore includes reaction modules, that is, relatively short sequences of chemical transformations that can be performed on a variety of chemically similar substrates. Accordingly, the same reaction module can be found in different parts of different contemporary pathways (48–50). As elaborated in the next section, we surmise that in evolution, reaction modules comprise the earliest multienzyme building blocks. The first level comprises founder enzymes, ancestral enzymes that emerged in a much simpler context whereby their presence per se (i.e., in the absence of enzymes catalyzing the connected reactions) was hugely beneficial. The most connected enzymes in contemporary metabolism, and their substrates/products, are likely to have been founders.

3.5. Reaction Modules: Ancient Evolutionary Building Blocks

While the number of different enzymatic reactions is huge, the number of reaction types is much smaller. For example, all sugar kinases catalyze the transfer of a phosphate group, typically from ATP, to a sugar's hydroxyl group. Further, in hundreds of different enzymatic reactions, a phosphate group is transferred from a donor (an activated phosphate group) to a hydroxyl acceptor, which can be a sugar hydroxyl, a carboxylate, another phosphate group, or water. Additionally, ATP, other NTPs, and other activated phosphoesters (e.g., acetyl phosphate) can serve as donors. All these reactions not only share the same type of chemistry, but there exists a single class of enzymes, P-loop NTPases, that can catalyze this entire range of phosphoryl transfer reactions (51). Thus, we may assume a founder enzyme that catalyzed the transfer of a phosphate group from a range of phosphorylated nucleoside donors to a variety of different acceptors.

In agreement with the notion of there being few types of founder reactions, there are also relatively few types of reaction modules. Nature seems to reuse certain sequences of reactions on a variety of different starting substrates to yield different products. Perhaps the simplest and most common reaction module is the carboxyl reduction module (50) (**Figure 2**).

A systematic classification of reaction modules is beyond the scope of this review. Rather, we focus on an insightful example, the so-called C1 elongation module (for brevity, C1 module) (**Figure 3**). As elaborated below, there is ample evidence indicating that the enzymes that mediate

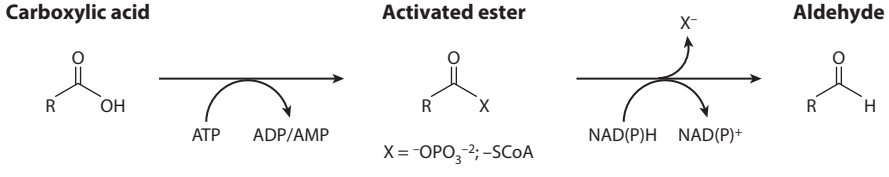


Figure 2

Certain sequences of reactions appear in different metabolic pathways, for example, the carboxyl reduction module (50). A carboxylic acid is activated via the formation of an ester (typically a phosphoro- or CoA ester) that is subsequently reduced to aldehyde. This module appears in many different pathways, including glycolysis/glycogenesis (50).

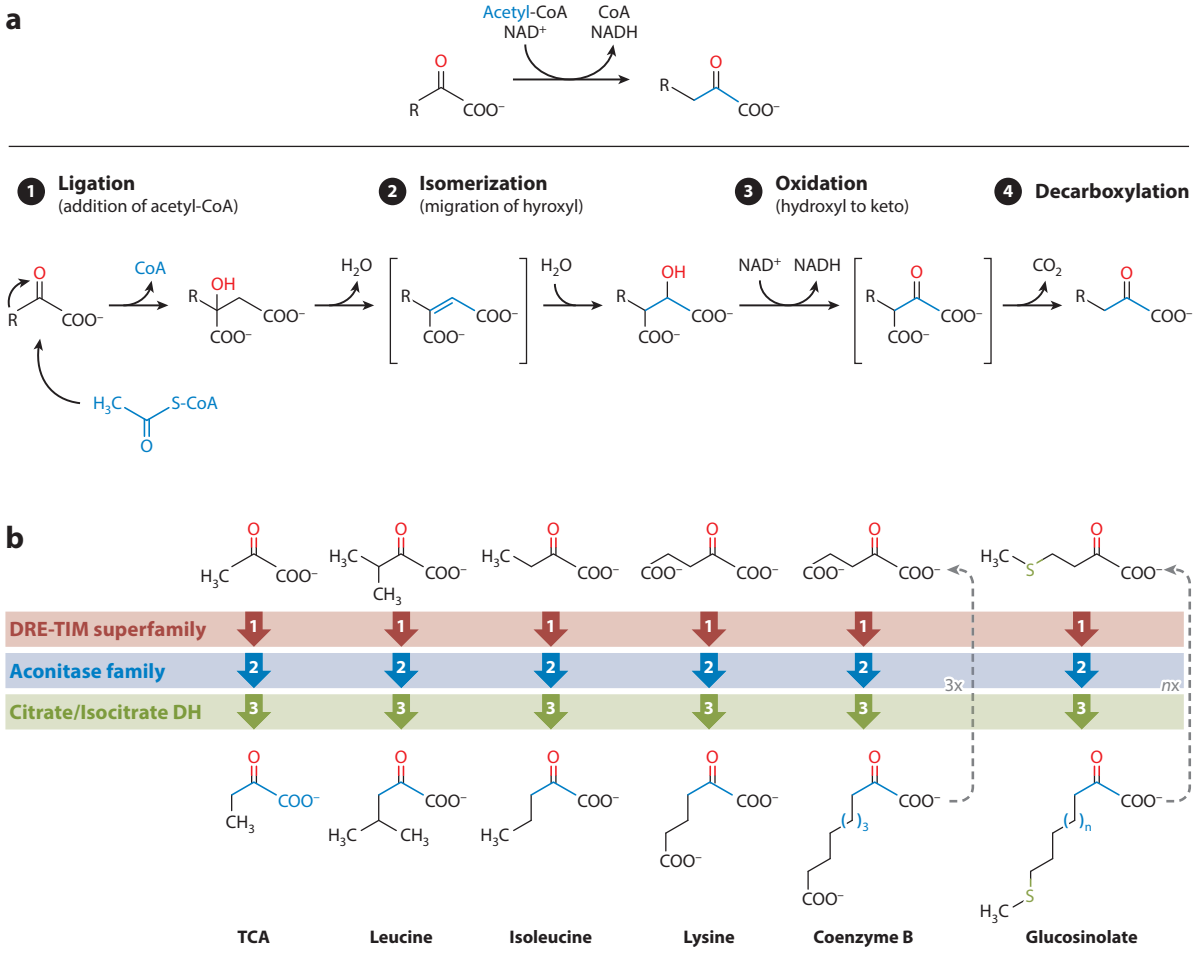


Figure 3

(*a*) The C1 reaction module begins with an α-ketoacid and ends with a one-carbon-longer α-ketoacid. (❶) Acetyl-coenzyme A (CoA) donates two new carbons (*blue*), but the α-ketoacid substrate loses one carbon via decarboxylation, thus resulting in a net gain of one methylene group. (❷) The brackets indicate intermediates that may not exist in certain species; for example, in the tricarboxylic acid (TCA) cycle and lysine biosynthesis, a single enzyme catalyzes both dehydration and rehydration. (❸) Oxidation and (❹) decarboxylation are generally catalyzed by one enzyme, but the latter may also occur nonenzymatically and is thus shown separately. (*b*) The C1 module is part of at least six pathways (*left* to *right*): one core pathway (the TCA cycle), three central pathways (leucine, isoleucine, and lysine biosynthesis), and two specialized pathways (biosynthesis of coenzyme B, a redox cofactor in methanogens, and of glucosinolate, a natural product of pungent plants) (48, 49, 52). In the latter two pathways, the C1 module is used recursively (*dashed line*). The first enzyme catalyzes the acetyl-CoA ligation and belongs to the DRE-TIM superfamily (53, 54). The second step, isomerization, is catalyzed by members of the aconitase superfamily (sometimes by two consecutive enzymes) (55). The third step, oxidative decarboxylation, is catalyzed by members of the citrate/isocitrate dehydrogenase (DH) superfamily (56).

the C1 module in what are now unrelated pathways share common ancestry (Section 4.3) (48, 52). Overall, it appears that reaction modules comprise an elementary evolutionary building block of metabolism. However, in principle, the constraints underlying the evolutionary emergence of reaction modules and of pathways are the same, as both comprise interlocked building blocks. Accordingly, in the text below, the term pathway is often used (for clarity or citation accuracy) even though, in effect, we mean pathway or reaction module.

3.6. What Makes a Reaction Module?

Modules comprise a cascade of two or more reactions that are energetically and chemically feasible as a whole (57). Thermodynamic driving force is crucial, as it directs the reaction module to the desired product and provides favorable kinetics at reasonable enzyme doses (58). For example, activated ester formation drives the reduction of carboxylic acids to aldehydes (**Figure 2**), which would in itself be highly unfavorable (50). Decarboxylation drives the oxidation step in the C1 module (**Figure 3*a***, Step 3) because oxidation of α-hydroxyl to α-keto (alongside the accompanying reduction of NAD+ to NADH) is unfavorable. In fact, decarboxylation results in the entire C1 module being thermodynamically favored, and hence capable of fast conversion rates (58). Moreover, many metabolites are unstable, or may leak out of the cell. Coupling the formation of these intermediates to favorable reactions that utilize them is also crucial. A fast pathway flux minimizes the impact of potential side reactions, as nearly every starting or intermediate metabolite can react in multiple ways. The architecture of metabolic pathways, and even of entire networks, can be rationalized in the light of physicochemical constraints (57). For example, that thermodynamic driving force is crucial is manifested in ATP being the most connected metabolite (**Figure 1**). ATP hydrolysis is also linked to both the carboxyl reduction (**Figure 2**) and the C1 modules (**Figure 3**; CoA ester synthesis is driven by ATP hydrolysis). Thus, flux considerations also dictate pathway evolution, and our model, described in Section 5.3, addresses them.

Abiotic reaction: a reaction that occurs in the absence of enzymes yet yields a key natural metabolite

Nonenzymatic reaction: a reaction that is enzyme-catalyzed in contemporary organisms yet can be performed without enzymes

4. PATHWAY EVOLUTION: THE CURRENT MODELS

That the interlocked nature of metabolic pathways presents a challenge with respect to their evolutionary emergence seems to have been first raised by Norman Horowitz (1915–2005). Horowitz defined metabolic pathways as a sequence of consecutive reactions, each catalyzed by an individual enzyme that in turn corresponds to an individual gene (59). Ever since, the question of how pathways evolved has been repetitively addressed. The proposed models are presented below in historical order (as far as we could track them down).

4.1. Retrograde Evolution (Horowitz)

In 1945, a year after delineating his groundbreaking pathway hypothesis with A.M. Srb (60), Horowitz published his now more widely known paper describing the retrograde pathway evolution model (61). From his findings on the arginine biosynthesis pathway, Horowitz (61, pp. 153–54) deduced that "intermediate substances cannot, in general, be assumed to have physiological significance, and the ability to produce them does not of itself confer a selective advantage." However, simultaneous recruitment of all pathway enzymes is extremely unlikely (a macroevolutionary event in Horowitz's terms). Horowitz (61, p. 155) suggested that a stepwise recruitment could occur in "a special kind of chemical environment; namely, one in which end products and potential intermediates are available." In such an environment, the end product of the yet-to-become pathway (A in Horowitz's annotation) is initially imported from the environment to support growth. Growth depletes A, rapidly making it a limiting resource. The emergence of an enzyme that catalyzes the formation of A from any available precursor will give an immediate advantage. The subsequent depletion of A's precursors (B + C) will prompt the recruitment of the next pathway enzyme, and so on. Consequently, the sequence of enzymes in the fully evolved pathway we observe today is therefore a mirror image of their historical order of recruitment (the last pathway enzyme was recruited first, etc.).

Retrograde recruitment depends on abiotic reactions and nonenzymatic reactions, as discussed in Section 5.2, and also on the fact that the earliest enzymes (i.e., the final steps in

present-day pathways) could provide a starting point for the later-evolving steps (i.e., the enzyme now catalyzing the earlier pathway step). Obviously, in 1946, this suggestion could not be related to sequence homology. Twenty years later, however, Horowitz (62) suggested that evidence for the retrograde model could be found in pathways whose enzymes exhibit sequence homology. Analyses of metabolic networks indicated enzymes that catalyze consecutive metabolic steps and appear to share a common evolutionary origin; however, this trend is rather weak, as it concerns at most 5% of all possible enzyme pairs (63). Indeed, enzymes that belong to the same pathway typically belong to different superfamilies or classes and thus do not share an evolutionary origin (Section 2.2). However, extant organisms do not necessarily reflect the state of the primordial metabolic pathways as hypothesized by Horowitz.

Specialized metabolism: biosynthesis or degradation of metabolites in specific cells or tissues, and in relatively limited quantities (e.g., natural products)

Few cases of consecutive pathway enzymes that share common ancestry are known, which thus supports Horowitz's model. One classical example is TrpF, TrpC, and TrpA in L-tryptophan biosynthesis (39). The three reactions are fundamentally different: TrpF is an isomerase; TrpA catalyzes the indole's ring formation; and TrpC releases it (**Figure 4*a***). Nonetheless, common ancestry can be inferred. The three enzymes share the TIM-barrel fold (**Figure 4*b***) as well as a key functional motif: The ribose 5′-phosphate moieties of their substrate/product are bound in the very same mode (**Figure 4*c–f***). Low yet statistically significant sequence homology can also be inferred. Tryptophan biosynthesis therefore represents a case of three consecutive enzymes that diverged from a common ancestor (64). Notably, substrate binding is the common denominator, while the catalytic mechanism has diverged. Further, the substrates/products of these three enzymes seem to take part in only this pathway (i.e., have connectivity $k = 2$; Section 2.2), at least in central metabolism (see also core metabolism, intermediate metabolism, and specialized metabolism). They thus fit Horowitz's key assumption, namely, that pathway intermediates have no use in themselves.

Few other such examples are known. Although these cases were described before as supporting Horowitz's model of retrograde evolution (37, 65), we surmise that, in fact, they support an alternative model of forward stepwise recruitment, as described in the next section.

4.2. Forward Pathway Evolution (Granick)

As far as we know, the alternative model of forward stepwise recruitment was first proposed by Sam Granick (1909–1977), a biochemist who studied the biosynthesis of heme and chlorophyll. In 1965, Granick addressed the evolution of these two key molecules (66). He noted that the biochemical functionalities of chlorophyll and heme are already present in the metal-bound protoporphyrin precursors. Thus, although present-day pathways include multiple steps that convert these precursors into functional end products, these precursors were probably the functional products at earlier stages of evolution. Granick thus proposed a model of forward stepwise recruitment: "[E]ach step had to be an end product of the biosynthetic chain and serve a useful function. Later, an additional step would be elaborated to produce a new end product that could carry out the same function more efficiently, but the previous step would still be essential, for it would serve as an intermediate" (66, p. 74).

What Granick proposed is profoundly important, namely, that the evolution of metabolism involves not only the emergence of new enzymes but, most importantly, the emergence of new metabolites. In heme and chlorophyll, this is evident: Starting from a common ancestral precursor, these metabolites diverged to fit two different biochemical functions (oxygen binding versus light absorbance and transfer) and cellular environments (protein complexes versus membranes). Their specialization is likely to have occurred incrementally, whereby the end product of each intermediate modification step was functional at some point in evolutionary history.

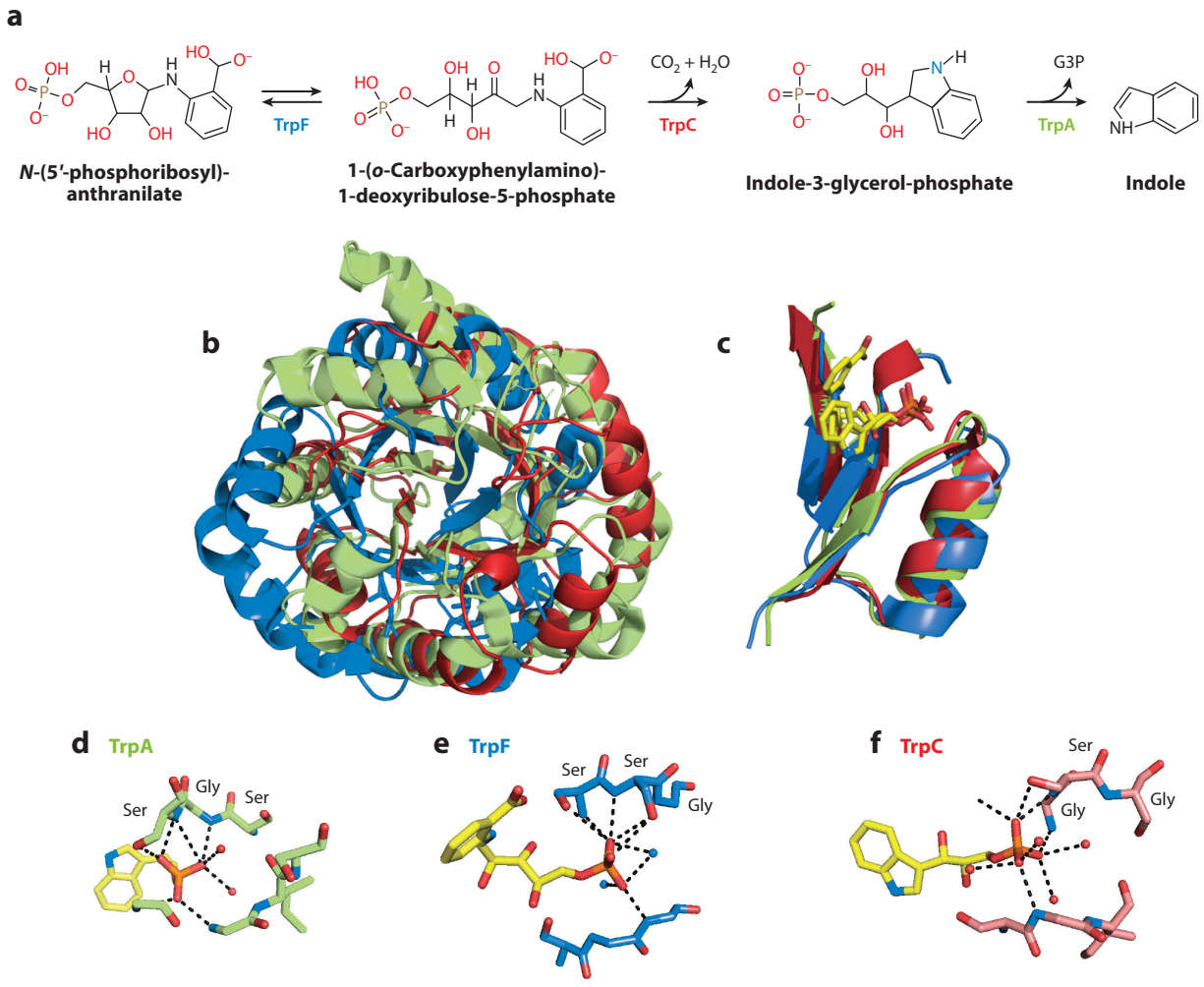


Figure 4

Pathway enzymes may diverge from one another via shared substrate binding. TrpF is shown in blue, TrpC in red, and TrpA in green throughout this figure. (*a*) In the L-tryptophan biosynthesis pathway, the three enzymes leading from 5′-phosphoribosyl anthranilate to indole share a common evolutionary origin (64). (*b*) The three enzymes share the same fold [triosephosphate isomerase (TIM) barrel] and low, yet statistically significant, sequence similarity. (*c*) Primarily, a common origin is indicated by the identical binding site for the 5′-phosphate moiety present in the substrates of these three enzymes. The phosphate binding motif resides at the tip of β-strands 7 and 8. (*d–f*) Detailed schematics of the phosphate binding motifs. Protein Data Bank codes: TrpA, 1V7Y; TrpF, 1LBM; TrpC, 1A53.

Granick noted two other key points: (*a*) Certain consecutive biosynthetic steps are chemically similar, acting on incrementally modified substrates (recursive chemistry), and (*b*) new genes arise by duplication. Indeed, two events of gene duplication were detected in the evolution of the chlorophyll biosynthesis pathway (67). Further, in both heme and chlorophyll biosynthesis, a single enzyme performs two, or three, recursive modification steps (two consecutive methylations in heme, or three consecutive reductions of the geranylgeranyl side chain in chlorophyll). A generalist ancestor that catalyzes recursive modifications, and that could later diverge to give specialized enzymes, each catalyzing one reaction, is thus a feasible evolutionary intermediate.

Recursive chemistry: consecutive steps of a pathway that involve the same or a very similar reaction performed iteratively

Although the forward evolution hypothesis is relatively unknown, there seems to be ample evidence supporting it. In particular, cases of common ancestry of consecutive pathway enzymes are traditionally cited in support of Horowitz's retrograde model (63, 68). However, they in fact indicate forward rather than retrograde evolution. Most supportive is the case of four consecutive ATP-driven condensation steps in the biosynthesis of the peptidoglycan building block (37, 65) (**Figure 5**). Most importantly, mature peptidoglycan is formed by conjugation of the glycan building blocks (with release of the UDP groups), while the peptide moiety, which is the end product of this four-step pathway, remains largely unchanged. Thus, the evolutionary intermediates of this pathway could comprise functional glycan chains with increasingly elaborate peptide moieties. Overall, recursive chemistry is common in specialized metabolism (69) but is relatively rare in core and central metabolism. MetB/C in methionine biosynthesis is one exception, an example of two consecutive enzymes with nearly identical active sites and similar catalytic chemistry (70).

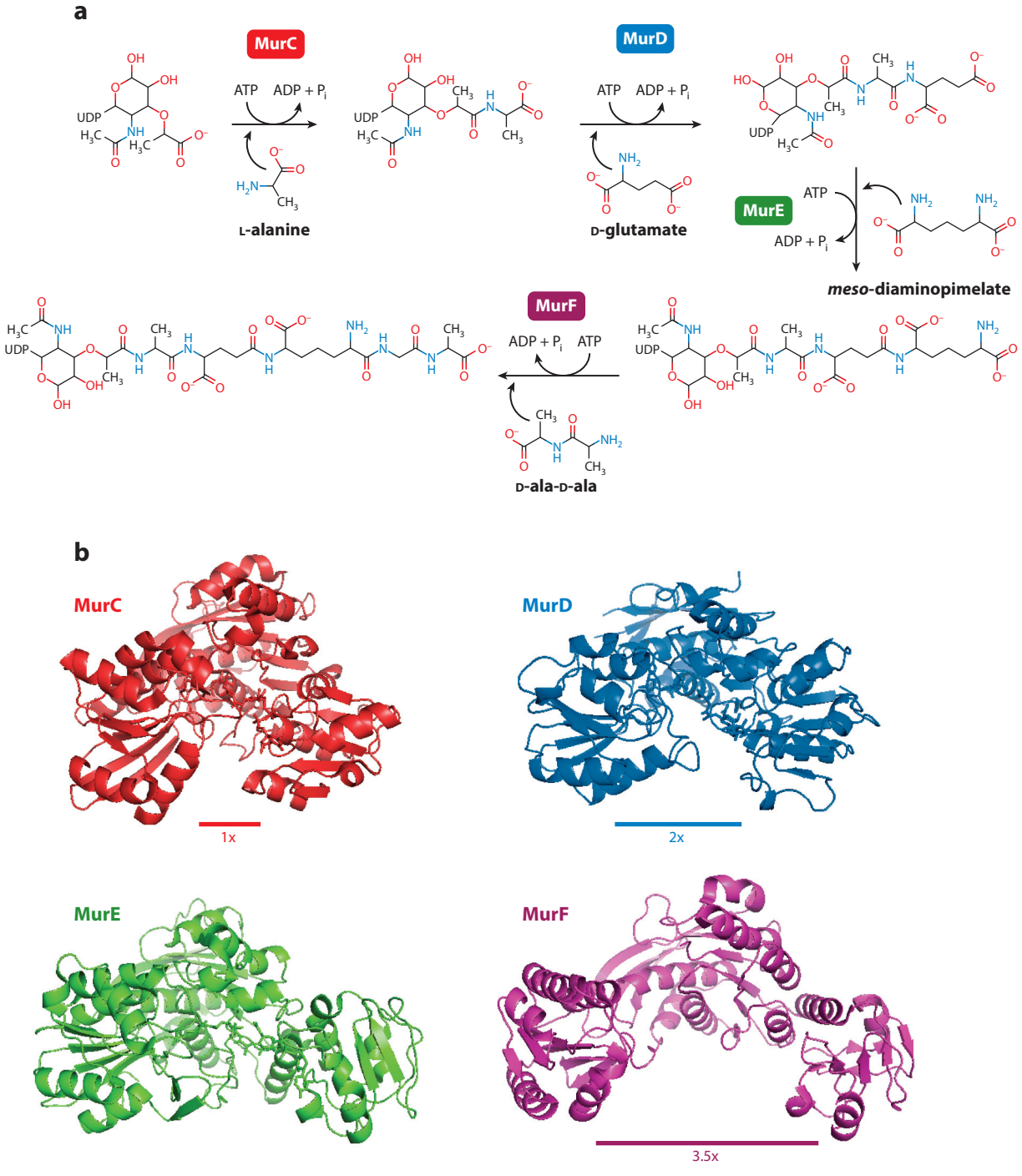
In general, Granick's model differs from Horowitz's in assuming physiologically functional intermediates generated by enzymatic action rather than physiologically unusable intermediates supplied by the environment. Granick's model does not make the assumption that nonenzymatic reactions preceded enzymatic ones. Nonetheless, coevolution of metabolites and enzymes is embedded in both models. Thus, although forward evolution seems feasible mostly in the context of intermediate and specialized metabolism, it may also apply to central and possibly to core metabolism, as elaborated in Section 5.

4.3. Pathway Duplication–Divergence: Generalists-to-Specialists (Ycas and Jensen)

Both Ycas (52) and Jensen (48) were inspired by the observation that the C1 reaction module comprises a series of analogous reactions that occur in multiple pathways on similar substrates (**Figure 3**). They suggested independently that these pathways diverged from ancestral enzymes that exhibited broad specificity and thus yielded multiple end products (48, 52). However, an identical set of consecutive reactions does not necessarily imply divergence from a common origin. As discussed above, reaction modules are the outcome of physicochemical constraints (Section 3.6) that in turn could drive convergence, that is, independent emergence of the very same reaction module due to chemical necessity. Thus, phylogenetic evidence is required to support divergence, as exemplified below for the C1 module.

More than 40 years later, there is ample evidence indicating that the C1 module that inspired both Ycas and Jensen is indeed an example of the emergence of multiple pathways via divergence of a single, ancestral reaction module. Further, multifunctional enzymes were proven to be a key evolutionary factor (Section 2). However, the phylogenetic picture with respect to C1 is complex, especially with regard to the divergence of the core pathways, an event that had presumably taken place >3.7 bya (billion years ago) and was followed by multiple incidents of gene gain, loss, and lateral transfer (**Figure 6**). The emergence of new alternative pathways that replaced the presumed ancestral pathways is also observed. For example, few of today's organisms synthesize lysine via the C1 module; the vast majority use an alternative pathway (72). Indeed, if Ycas and Jensen were to submit their papers today, they would have probably been rejected due to lack of systematic evidence with respect to the core pathways they had originally indicated. However, more recent evolutionary events, as in coenzyme B biosynthesis in Archaea (73, 74), provide clear-cut evidence for duplication and divergence of the C1 module (**Figure 7**). Moreover, bifunctional C1 enzymes, supporting in parallel two different C1 reaction modules, also exist in contemporary organisms (42).

Other recently evolved pathways in specialized metabolism lend support to the Ycas–Jensen model. For example, the tomato cholesterol biosynthesis pathway (12 steps in total) emerged via



(Caption appears on following page)

Figure 5 (*Figure appears on preceding page*)

In forward evolution, pathway enzymes catalyzing recursive steps may diverge from one another. (*a*) Peptidoglycan biosynthesis includes four consecutive ATP-mediated condensation steps, each adding one additional amino acid to the growing glyco-peptide building block. (*b*) The corresponding enzymes share the same fold and catalytic mechanism, and they also exhibit sequence homology, thus indicating common ancestry (Protein Data Bank codes: MurC, 1P3D; MurD, 2UAG; MurE, 1E8C; MurF, 1GG4) (71). The bars denote the gradual opening of the active-site pocket to accommodate the increasingly growing substrates. Abbreviation: UDP, uridine diphosphate.

gene duplication and divergence of six of the phytosterol biosynthetic enzymes. The remaining six enzymes are generalists that take part in both the original context (phytosterol biosynthesis) and the new one (78). Tomato cholesterol biosynthesis is therefore a vivid example of evolution in action, indicating emergence via broad-specificity generalist ancestors followed by duplication and divergence of specialists. Note, however, that while the first two models address stepwise enzyme recruitment, the Ycas–Jensen model describes the simultaneous recruitment of multiple enzymes, as does the patchwork model described below.

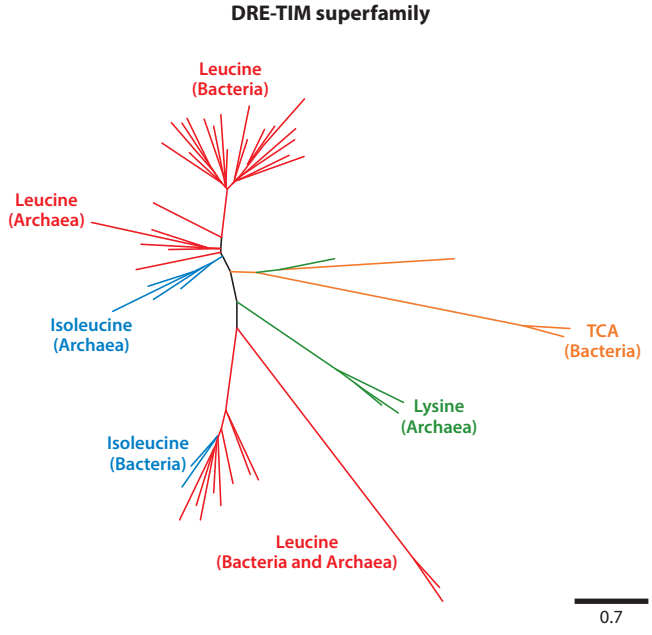
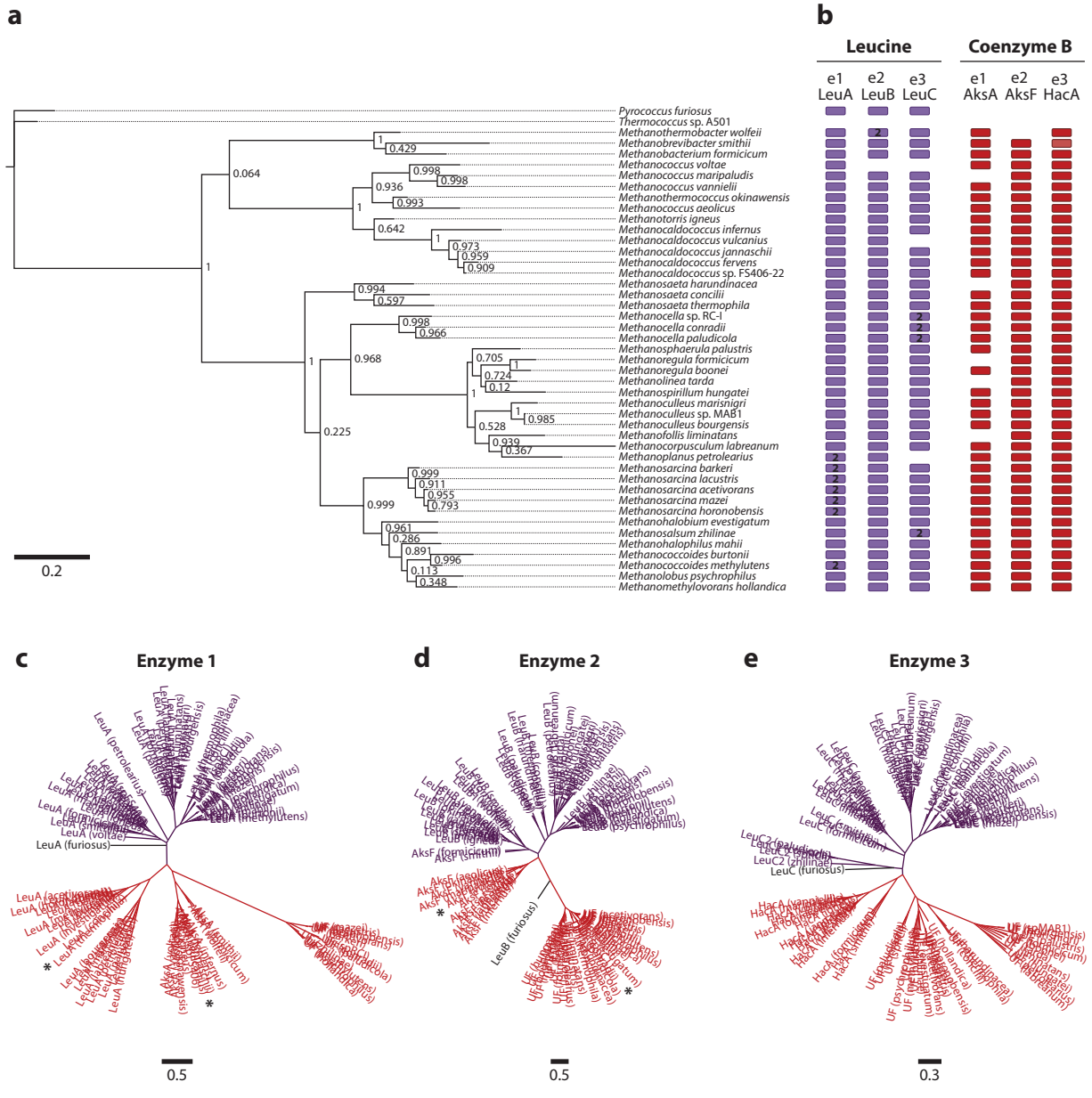


Figure 6

Multiple pathways emerged from an ancestral C1 module via duplication and divergence. The schematic tree indicates the common origins of the first enzyme of the C1 module that catalyzes the acetyl-coenzyme A ligation (**Figure 3**). Today, only leucine biosynthesis (*red*) and the tricarboxylic acid (TCA) cycle (*orange*) are ubiquitous pathways. Two different enzyme families can be found that catalyze the first step of the TCA cycle, the Si- and Re-citrate synthases. However, Si-citrate synthase, which is the most abundant contemporary enzyme, does not share a common origin with the other C1 enzymes. Re-citrate synthase shares a common origin with other first enzymes in the C1 module yet can only be found in a few anaerobic bacteria (49). Consequently, the orange clade (TCA cycle) is scarcely populated compared to the red one (leucine biosynthesis). Similarly, lysine (*green*) and isoleucine (*blue*) biosynthetic pathways containing the C1 module are found in few organisms (75, 76). It also seems that isoleucine biosynthesis, or at least its first C1 enzyme (*blue*), emerged at least twice in independent lineages. The phylogenetic mapping of the second and third enzyme families (**Figure 3**) is even more complex (55, 56).

4.4. Patchwork, or Patch Assembly (Lazcano & Miller)

The notion of multifunctional generalist enzymes gave rise to the patchwork, or the patch assembly, model. According to this model, enzymes can be recruited from different pathways. The patchwork model has earlier roots, yet it was formalized primarily by Lazcano & Miller (79). There is no doubt that numerous extant enzymes are multifunctional, catalyzing more than one transformation (typically the same reaction with related substrates) and thus taking part in multiple pathways (15). Latent, promiscuous activities give birth to new enzymes, as discussed in Section 2.



(Caption appears on following page)

Figure 7 (*Figure appears on preceding page*)

Divergence by duplication of a new C1 module. A phylogenetic reconstruction shows the evolutionary history of pathways. The C1 module involved in coenzyme B biosynthesis in some methanogenic Archaea was analyzed (73, 74). (*a*) All complete methanogenic archaeal genomes were retrieved from the Joint Genome Institute database. The organismal tree (tree of life) was constructed for a fixed set of 45 organisms using RNA polymerase B as reference. Two nonmethanogenic Archaea (*Pyrococcus furiosus* and *Thermococcus* sp. A501) were used as out-groups. (*b*) All putative C1 enzymes were identified by searching homologs of LeuA, LeuC, and LeuB, the first, second, and third enzymes of the C1 module in leucine biosynthesis. Nearly all organisms contain a duplicated set of these three genes, but the out-group does not. (*c–e*) An independent phylogenetic reconstruction for these three different enzyme families shows the same topology. A large clade whose topology largely mirrors the organismal tree shown in panel *a* is observed; this clade likely corresponds to the essential leucine biosynthesis genes (*purple*). The second set of paralogs is distributed in two or three clades (*red*). Experimentally validated enzymes involved in coenzyme B biosynthesis are marked with an asterisk (74). Phylogenies were reconstructed using PhyML software (77). The degree of sequence divergence is indicated by the scale bars at the bottom of the figure panels.

Thus, once a relatively large repertoire of enzymes is available to an evolving organism (in its own genome or via lateral gene transfer), enzymes can be readily recruited from various pathways to assemble a new pathway. Further, rather than applying the very same reaction sequence on different substrates, as exemplified by the C1 module (**Figure 3**), mix and match of different reactions can give rise to completely new reaction modules and pathways. This is beautifully exemplified in pathways that evolved to catabolize contaminating man-made chemicals (80–82), and also in the potential to engineer synthetic pathways (83).

The current picture of metabolism is mostly compatible with the patchwork model: The vast majority of pathways comprise enzymes of independent evolutionary origins (65). This patchy state also indicates convergence of some pathways and reaction modules due to chemical necessity. However, a present-day patchy state of a pathway does not necessarily indicate that it originally emerged via patchwork. Rather, pathways may have initially evolved via the above-described mechanisms (Sections 4.1–4.3), and multiple events of gene gain, loss, and lateral transfer gave rise to the patchy patterns seen today (**Figure 6**).

5. HOW DO PATHWAYS EVOLVE? AN INTEGRATED METABOLITE–ENZYME COEVOLUTION MODEL

We propose an integrated metabolite–enzyme coevolution model that combines different elements from the above-described models (Section 4) as well as some new elements, particularly with respect to the order of enzyme recruitment. Our model addresses the origins of metabolites (Sections 5.1 and 5.2) and the origins of enzymes and their order of recruitment (Section 5.3).

5.1. Enzyme Promiscuity Yields New Metabolites and Reactions

Metabolites are chemical compounds with a physiological role. Metabolites may originate via different routes: (*a*) through changes in the environment that result in the appearance of new extracellular compounds that may be recruited by organisms, (*b*) by enzyme promiscuity, or (*c*) via nonenzymatic reactions (reactions that are not mediated by enzymes) that may generate new intracellular compounds from preexisting metabolites. Enzymes promiscuously accept substrates that differ from their physiological substrates, thus generating side products (9). Alternatively, a side reaction may promiscuously occur on the native substrate, also yielding side products (84). Both phenomena are widely observed, as is evidenced by the common presence of damage metabolites and of enzymes that evolved to eliminate side products whose accumulation has deleterious effects (85). However, many side products do not accumulate to a degree that affects the cell. Thus, mutations that have no or little effect on an enzyme's native function can induce dramatic changes in its promiscuous activities, yielding some new side metabolites (9).

Thus, alongside the known metabolites, many unknown metabolites likely exist, as well as compounds that are produced in cells, primarily via enzyme promiscuity, yet have no physiological role. To name but one example, certain S-adenosylmethionine (SAM, or AdoMet) synthases readily accept NTPs and dNTPs other than their native substrate, ATP. Thus, alternative metabolites such as S-guanosylmethionine or S-deoxyadenosylmethionine likely exist (86). The level of chemical messiness [metabolic dark matter (87)] is hard to estimate but likely to be significant. Only a small percentage of the molecular masses seen in untargeted high-throughput metabolomics has so far been identified (87, 88). Overall, enzyme promiscuity leads to the generation of metabolic side products and also provides an evolutionary starting point for a specialized enzyme that generates this product.

5.2. The Role of Nonenzymatic Reactions

Aside from enzyme promiscuity, the emergence of new pathways may involve abiotic reactions or nonenzymatic reactions. These two terms have slightly different meanings. Starting with the Urey–Miller experiment, the majority of abiotic reactions described so far yield key natural metabolites such as amino acids via reactions, which often fundamentally differ from the enzyme-catalyzed reactions that produce the same metabolites in living organisms (89). In contrast, nonenzymatic reaction is a term used here to describe reactions that are enzyme catalyzed yet can occur in the presence of inorganic catalysts (water, bases, acids), including metal ions (transition metal ions such as Fe^{2+}, Mn^{2+}, or Zn^{2+} can be highly potent catalysts), or simple organic catalysts such as amino acids or short peptides. Abiotic reactions are central to understanding the origins of life, a topic that is outside the scope of this review. Nonenzymatic reactions, on the other hand, are directly relevant to the emergence of metabolic pathways, including recently evolved ones.

Contemporary life is utterly dependent on enzymes. Nonetheless, some reactions occur within living cells without enzyme catalysis (90). Certain metabolites exhibit intrinsically high reactivity, and their nonenzymatic reactions can have important regulatory roles (91, 92). Accordingly, nonenzymatic reactions are also an integral part of pathways in core metabolism and central metabolism. For example, the decarboxylation in the C1 module (**Figure 3**, Step 4) occurs nonenzymatically in many cases [i.e., the keto-acid intermediate produced in Step 3 is released and forms the end product outside the dehydrogenase's active site (42)]. In fact, in quite a few enzyme-catalyzed reactions, only one step is catalyzed in the enzyme's active site [e.g., dehydration by certain serine deaminases gives 2-aminoprop-3-enolate that breaks down outside the active site to give ammonia and pyruvate (93)].

Although nonenzymatic reactions exist, they are rare in core and central metabolism. However, specialized metabolism, the "panda's thumb" of evolutionary biochemistry, indicates a crucial role for nonenzymatic reactions. Specialized pathways evolved relatively recently and sporadically, and they typically involve low fluxes. As such, the tinkering nature of evolution is vividly visible. The enzymes show broad specificity (21) and are relatively slow [k_{cat} values of secondary metabolic enzymes are, on average, ~30-fold lower than the enzymes of core metabolism (94)]. Nonenzymatic reactions are commonly observed in specialized metabolism (95–97). Betalain (plant pigments) biosynthesis, for example, involves several reactions that occur with no enzymes, some at very high rates (95). These specialized pathways evolved when an ample repertoire of preexisting enzymes was available (land plants emerged only ~0.5 bya). Nonetheless, nonenzymatic reactions played a key role in their emergence (96, 97).

That nonenzymatic reactions have an evolutionary role is also supported by a laboratory experiment in which an alternative pyridoxal-5′-phosphate (PLP) biosynthesis pathway was unraveled (98). Here also, the tinkering power of evolution is beautifully revealed. The newly emerging

pathway drew an intermediate of serine biosynthesis to make the missing PLP precursor via four steps (**Figure 8**). Three steps are enzyme catalyzed, by a generalist enzyme, by the promiscuous activity of another enzyme, and by a protein of unknown function. However, one step, the conversion of hydroxypyruvate to glycolaldehyde, occurs nonenzymatically. A nonenzymatic step was recruited despite the fact that *E. coli* has a repertoire of ~1,500 preexisting enzymes that offer multiple opportunities for promiscuous enzyme-catalyzed reactions. The rate of conversion is slow at ~0.015 h^{-1} (half-life of ~2 days) but nonetheless adequate for production of PLP at a sufficient rate (~5 μM/h).

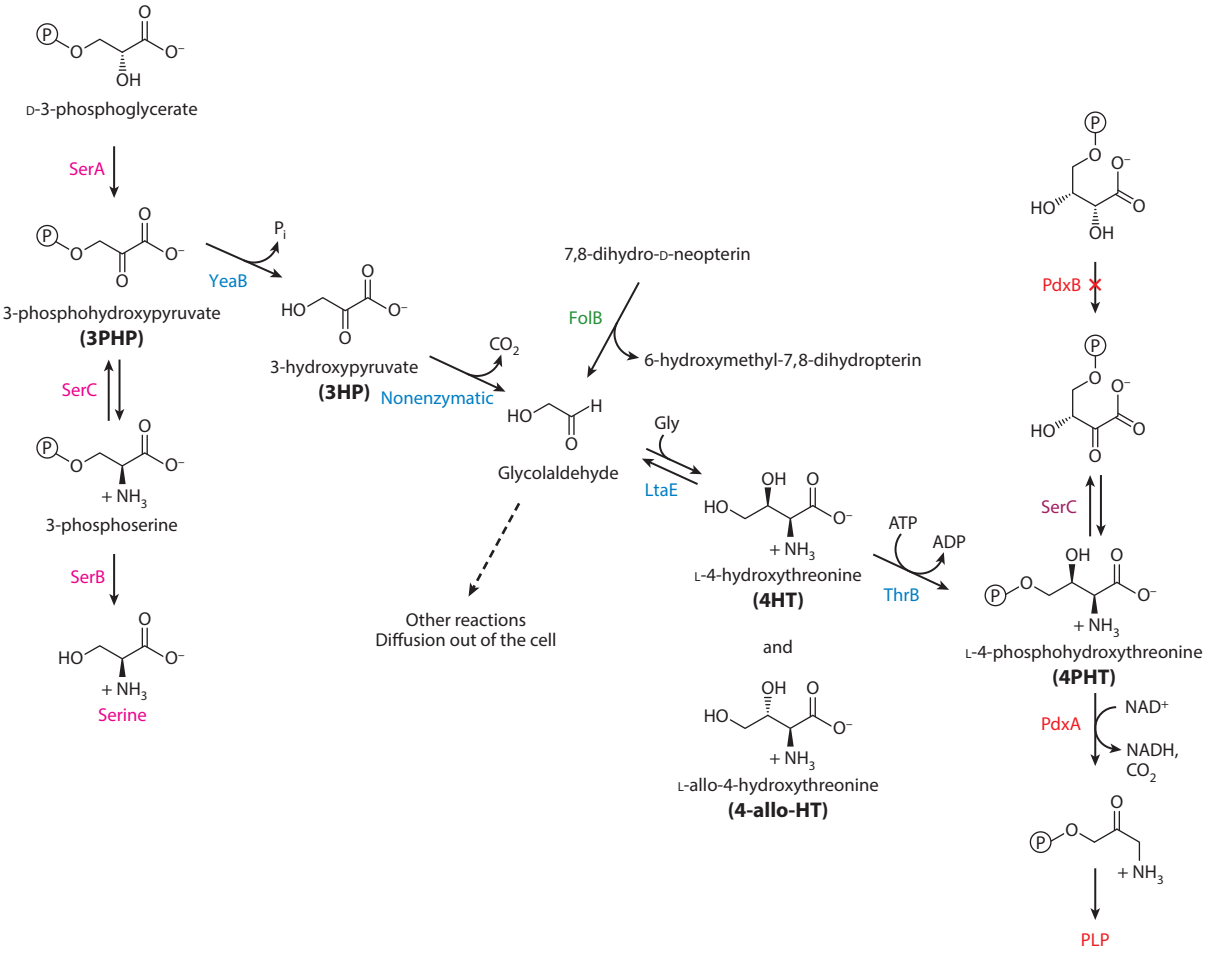


Figure 8

The emergence of a de novo pathway via recruitment of a nonenzymatic reaction. The figure shows one out of three latent or underground pathways that restored the synthesis of pyridoxal-5′-phosphate (PLP) in a strain of *Escherichia coli* that lacked 4-phosphoerythronate dehydrogenase (ΔPdxB). The pathways were unraveled by identifying *E. coli* genes, YeaB and ThrB in the pathway shown here, whose overexpression allowed growth. The pathway shown here (*blue*) converts 3-phosphohydroxypyruvate, an intermediate in serine biosynthesis, to L-4-phosphohydroxythreonine, an intermediate in the PLP synthesis pathway. The second step is a nonenzymatic conversion of 3-hydroxypyruvate to glycolaldehyde, which is a known alternative precursor of PLP (99). Figure based on data from Reference 98.

It is also becoming increasingly apparent that a surprisingly high number of enzyme-catalyzed reactions can occur at measurable rates without an enzyme, either by general acid–base catalysis (protons or hydroxyls) or via catalysis by simple inorganic (e.g., metal ions) or organic catalysts (e.g., amino acids). For example, most of the reaction steps of glycolysis (and hence of gluconeogenesis) and of the pentose phosphate pathway have been shown to occur under anaerobic conditions in the presence of ferrous iron (Fe^{2+}), some with a surprisingly high yield and rate (up to ~10% conversion per hour) (100, 101). The gluconeogenic aldol condensation to yield fructose 1,6-bisphosphate is accelerated by individual amino acids in ice (102). Likewise, nonenzymatic tricarboxylic acid (TCA) cycle reactions (103) and nonenzymatic synthesis of the cosubstrate SAM (104) have also been described. In fact, two nonenzymatic routes of SAM synthesis have been discovered. One reaction involves methionine and ATP, as in the biosynthetic pathway, while the alternative reaction occurs at low pH and uses adenosine instead of ATP. Thus, a compound might arise via an abiotic reaction, be recruited as a metabolite, and later be biosynthesized via an alternative reaction (104).

Altogether, the above examples lend support to the notion that pathways emerge from a myriad of preexisting, latent, coincidental reaction sequences from an underground network. These sequences may include promiscuous enzymatic reactions, native enzymatic reactions that occur out of their original context (underground metabolism), and nonenzymatic and abiotic reactions. Emergence via latent reaction sequences alleviates the need for simultaneous recruitment of multiple enzymes. Rather, a gradual optimization of the pathway's flux may take place via a stepwise recruitment of one enzyme at a time.

5.3. The Order of Enzyme Recruitment: Rate-Determining Steps Come First

Stepwise recruitment is the key element in only two of the existing models for pathway evolution (Sections 4.1 and 4.2). We surmise, however, that it is a critical element regardless of the model of emergence. Enzyme recruitment is an inherently incremental process. It involves duplication or horizontal gene transfer, and these events typically occur one gene at a time. Even if whole operon or genome duplications occur, a new pathway that emerges from a set of preexisting enzymes [via duplication-divergence (Ycas–Jensen) or via the patchwork model; Section 4.3 or 4.4, respectively] still demands the divergence of gene expression, and the latter occurs via mutations, one gene at a time. Thus, a unifying model of stepwise recruitment is needed.

We surmise that the recruitment of enzymes crucially depends on their potential contribution to the overall flux of the evolving pathway. The degree to which an increase in an enzyme's activity changes a pathway's flux is quantified by the so-called flux control coefficient. Depending on their kinetic constants and metabolite concentrations, enzymes share the control over the pathway's flux. However, sometimes, one enzyme stands out, and thus catalyzes a rate-limiting reaction step. This would certainly be the case when a latent, underground pathway that proceeds with very low flux comes to be under selection.

Hence, by our unifying model, stepwise recruitment may occur either by a forward order or by a flux-determining order. Forward order applies to those cases where each of the pathway intermediates served at some stage as a selectable end product (Granick's model; Section 4.2). In all other cases, when only the end product provides a selectable advantage, the order of recruitment is dictated by the rates and equilibrium constants of the various reaction steps that provide the intermediates for the emerging pathway. Flux-determining order prevails in all scenarios, whether we assume a latent, preexisting sequence of nonenzymatic reactions (Horowitz's model basically) or a recruitment of enzyme-catalyzed reactions, either of a pathway or module en bloc (the Ycas–Jensen model; Section 4.3) or of enzymes from different pathways (patchwork model; Section 4.4).

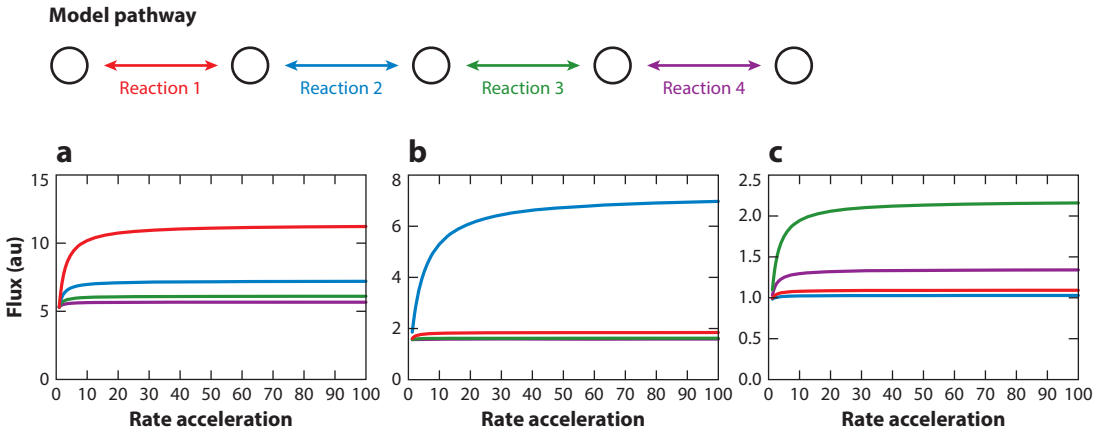


Figure 9

The flux-determining order of enzyme recruitment. In an evolutionary scenario in which enzyme levels are fixed, the order of enzyme recruitment (i.e., of evolutionary improvement of enzymatic rate constants) can be expected to follow the potential of individual reaction steps to provide a higher end-product flux. To illustrate the principle, we simulated emerging pathways of four consecutive, reversible reactions that occur at an arbitrary forward rate and reverse rate of 1 ($K_{eq} = 1$). Standard reversible mass-action kinetics is assumed for all steps, and the pathway is assumed to be at steady state (constant in-flow of the starting metabolite, S_0, and removal of end product, S_f, and assuming $S_0 = 10 \times S_f$ to maintain a positive flux) (105, 106). The figure shows the steady-state rate of end product formation (flux, in arbitrary concentration units per time) as a function of the enhancement in the rate of the recruited enzyme, per each of the four steps. We simulated three different scenarios: (*a*) All four steps initially occur with the same rate constant. As can be seen, increasing the rate constant of the first reaction has the highest effect on end product formation, while the last step has almost no effect. (*b*) If the second step initially has a lower rate constant, it is rate determining and hence is recruited first (simulated as occurring at a tenfold slower rate constant compared to all other steps, in both forward and reverse). (*c*) The second step is assumed to occur with a forward rate constant of 1 but with a reverse rate of 10 ($K_{eq} = 0.1$ instead of 1). Consequently, an enzyme catalyzing the third step is likely to be recruited first.

Broadly speaking, an enzyme that catalyzes a rate-limiting step is likely to emerge first. To illustrate this notion, we simulated the gradual stepwise emergence of a pathway from a preexisting set of reactions that lead to a selectable end product [largely using the approaches described by Heinrich & Hoffman (105) and Heinrich & Klipp (106)]. Initially, these reactions occur nonenzymatically or are promiscuously catalyzed by enzymes belonging to other pathways, and hence occur at slow rates. By assessing flux control, we then predict which enzyme will be recruited first. This model has numerous potential configurations, but we consider only a few simple scenarios to illustrate the principle.

If all steps are reversible and initially occur at the same basal rate constant, recruitment of the first enzyme provides the highest advantage, that is, the largest increase in end product formation (**Figure 9*a***). Instead, if one step is substantially slower than all others (a rate-determining step), recruitment of an enzyme for this step results in the highest increase in pathway flux (**Figure 9*b***). Alternatively, if one step is closer to its equilibrium state compared to all other states, the enzyme catalyzing the following step is most likely to be recruited because it will draw the unfavorable reaction in the forward direction (**Figure 9*c***). Notably, at the initial recruitment stage, a modest increase in activity, caused by a single mutation in a preexisting enzyme or by the recruitment of an enzyme with a weak promiscuous activity, results in a parallel flux increase (i.e., the emerging enzyme exhibits a flux control coefficient close to 1).

As the recruitment of enzymes continues, each recruited enzyme changes the control coefficients, thus creating the selection pressure for the recruitment of another enzyme. In effect,

recruitment occurs via fixation of mutations that increase either enzyme rates or/and levels under conditions when the new pathway operates, or/and increase the enzyme's catalytic efficiency for the new reaction. Our hypothesis that the rate-limiting enzyme evolves first does not necessarily contradict Horowitz's retrograde evolution model. For example, if the last intermediate is present at high concentrations but is slowly converted into the end product, the last enzyme will have the highest control on product formation and is likely to emerge first.

The above examples are oversimplified in terms of ignoring branching reactions. Given a more realistic scenario, that is, substrates or/and intermediates that can react in more than one way, the most critical contribution, as shown by flux control coefficients, is likely to be from an enzyme that funnels the reaction sequence toward the desired end product. In a complex chemical environment, the formation rate of an end product would also be greatly improved by blocking alternative reactions. Thus, which is the rate-limiting step depends on the metabolic network at the time of emergence.

In contemporary metabolism, the enzymes with the highest contribution to end product formation can be detected by their current flux control coefficient (107). However, the control coefficients of contemporary pathway enzymes do not testify on their contributions at the early stages of emergence. The demands for catalytic efficiency change between the initial recruitment stage and the later stages when additional pathway enzymes have been recruited to support higher flux; accordingly, the ancestors of some metabolic enzymes may have exhibited higher rates than their contemporary descendants (108).

Overall, our knowledge regarding the historical order of recruitments is limited. Cases where enzymes belonging to the same pathway share common ancestry are rare (Section 4.1) and in most of them, forward evolution is a feasible scenario (Section 4.2). In the laboratory-evolved PLP biosynthesis pathway discussed above (**Figure 8**), pathway emergence was initiated by recruitment of either the last enzyme (ThrB) or the first one (YeaB) (98). In another case, a retrograde strategy for pathway engineering was implemented; introduction of the enzyme catalyzing the last step first resulted in the identification of a shortcut that could not be foreseen (109).

In summary, stepwise recruitment is key to pathway emergence. The order of recruitment is primarily dictated by whether and which intermediates can provide an advantage at the time of emergence, and by the relative contribution of individual steps to the rate of formation of the selectable end product. Overall, selection acts on the performance of entire systems. Metabolic performance is therefore not a property of individual enzymes but of a system. To link these two levels (single enzyme parameters, and overall metabolic performance), it is insufficient to link enzymatic steps like Lego bricks. Rather, the specifics of each reaction sequence dictate how changes in the rates of individual enzymes translate to changes in the rate of formation of its utilizable end product.

5.4. Metabolite–Enzyme Coevolution: Implications and Potential Validations

The four models presented in Section 4 were described as alternatives, but in fact these models are complementary rather than contradictory. We thus propose a unifying model that combines various elements of these models with some new elements. It is, foremost, a model of stepwise coevolution of enzymes and metabolites. The key to the evolution of metabolism is that the composition of metabolites changes alongside the appearance of new enzymes. New compounds appear initially as nonphysiological by-products (Section 5.1). If utilized in a way that provides a selective advantage, that is, as metabolites, they represent an incentive for recruitment of enzymes that promote their synthesis. Upon selection toward a new function (16, 17), enzymes promiscuously give rise to yet other new compounds. Accordingly, metabolic networks expand gradually,

not only in terms of enzymes but also of metabolites (Section 3). Survival of the fittest therefore applies not only to genes (i.e., enzymes) but also to metabolites and the reactions that form and transform them. Selection against interfering metabolites is also likely to act, thus preventing expansion in certain directions.

Pathway evolution is initiated from a sequence of latent, underground reactions. However, as a pathway evolves and is lifted from the underground, some reactions may be replaced by others that afford higher end product yield, also by minimizing toxic and inhibitory intermediates (reaction loss and gain, equivalent to gene loss and gain) (110). The enzymes are recruited incrementally, one at a time. The order of recruitment—forward order (Section 4.2), flux-determining order (Section 5.2), or reverse order (Section 4.1)—may vary. Regardless of their recruitment order, when pathway enzymes diverge from one another, multifunctional enzyme intermediates that simultaneously support two different pathway steps are critical. Thus, generalist enzymes are also a key element of the combined model (Sections 2.1 and 4.3). Given multifunctional enzymes, founder reaction modules such as the C1 module are likely founders of multiple modern pathways via divergence en bloc (Section 4.3). Finally, another critical component of our model is that alongside shared chemistry, shared substrate binding (Section 2.3) is a key factor in scenarios where enzyme pathways diverge from one another. Shared binding may involve common substrate moieties such as phosphate (e.g., **Figure 4**) and also cofactor or cosubstrate cross-binding (ATP, NAD^+, etc.). Alternatively, shared chemistry applies in recursive pathway reactions (Section 4.2).

Metabolite–enzyme coevolution is most obvious when a molecule that is functional in itself is optimized by a series of incremental derivatizations, as exemplified by heme, chlorophyll, or peptidoglycan (**Figure 5**). This scenario is relevant to numerous natural products, such as antibiotics or plant scents and colors. These metabolites evolved alongside the biosynthetic pathways that produced them in a forward fashion (Section 4.2). The current intermediates originally served as end products and were later replaced by fitter derivatives (111, 112). Accordingly, contemporary pathway intermediates of some antibiotics, for example, exhibit function that is lower than the end products yet measurable (113, 114).

Whether metabolite–enzyme coevolution applies to core metabolism and central metabolism, and specifically to the key biosynthetic pathways (e.g., glycogenesis, or amino acids or nucleotide biosynthesis) is not easy to establish. Nonetheless, there are some obvious examples. Thymidine (dT) is a DNA nucleoside synthesized from uridine, an RNA metabolite that likely emerged first. Another clear example is NADP, a metabolite generated via phosphorylation of NAD [NAD kinases are present in all three kingdoms of life (115)]. Conversely, NADP(H)-utilizing dehydrogenases evolved from NAD(H)-utilizing dehydrogenases (116), and switches in NAD-NADP selectivity can be initiated by a few mutations (117). Thus, a new metabolite (NADP) may have appeared coincidentally via promiscuous kinase action and been promiscuously utilized by a preexisting NAD-dependent dehydrogenase. Eventually, two parallel redox systems, with independent metabolites and enzymes, had diverged: NAD is primarily used in oxidations and NADPH in reductions, thus enabling these opposing processes to occur in parallel.

In other cases, contemporary damage metabolites could be vestiges of earlier, functional metabolites, also because promiscuous activities in extant enzymes often reflect ancestral functions (9). For example, enzymes of the C1 module in leucine biosynthesis promiscuously produce 2-ketovalerate, which via the promiscuous action of aminotransferase yields norvaline, a nonproteogenic amino acid (118, 119). However, norvaline may have been part of early-stage proteins. Similarly, recursive C1 cycles produce 2-oxoadipate, which in turn yields ornithine, currently an intermediate in the biosynthesis of polyamines and arginine. However, ornithine may have originally emerged as a proteogenic amino acid that preceded lysine or/and arginine. Norvaline,

ornithine, and many other metabolites, therefore, might represent bridging metabolites, that is, end products of ancestral pathways that have later diverged to produce a different end product.

6. STANDING QUESTIONS AND CHALLENGES

The evolution of metabolism presents a breadth of questions. In this last section, we mention a few more aspects, discuss standing questions and challenges, and point out implications for areas such as synthetic biology.

6.1. How Metabolism Evolved: A Phylogenetic Approach

We now have a fairly clear picture of how individual enzymes evolve (Section 2). However, we know much less about the next two levels in the evolution of metabolism, namely, metabolic pathways and networks. In 2002, when reviewing the basic models and theories, as we do here, Rison & Thornton (68, p. 374) anticipated that "as metabolic structure and sequence space are further explored, it should become easier to trace the finer details of pathway development and understand how complexity has evolved." However, since 2002, there has been relatively little progress, especially with respect to how core metabolic pathways emerged. The pioneering analyses of *E. coli*'s metabolic network uniquely combined pathway and network information with gene sequence information (63, 65, 68, 120). However, it appears that beyond the observation of homology in individual *E. coli* enzyme pairs, or for that matter in any other individual species, phylogenies of pathways have not been systematically explored. Further, explorations of certain homologous enzyme pairs indicated that, in fact, homology stems from gene duplications that occurred well after their pathways emerged (121). As demonstrated in **Figures 6** and **7**, detecting phylogenetically consistent trends for specific enzymes (122) and for their combinations in pathways can be informative. That certain enzymes and reaction modules do not show phylogenetic consistency does not necessarily indicate that they do not share common ancient origins. However, those that do follow a consistent phylogeny provide further insight and may even enable ancestral sequence reconstruction (123).

The pathways that now comprise core and central metabolism remain the holy grail of evolutionary research (Section 6.2). Nonetheless, recently emerged pathways, typically in specialized metabolism of plants (78) and other organisms (e.g., **Figure 7**), offer intriguing case studies because the evolutionary record is still preserved in present-day genomes. Similarly, recently emerged catabolic, xenobiotic-degrading pathways in microorganisms (80) also teach us about pathway evolution. However, given that these pathways are sporadic and niche-specific, tracing their origins is like finding a needle in a haystack.

6.2. The Early Metabolic Pathways

Over evolutionary time, metabolism evolved toward immensely high complexity and specialization. However, the composition of the much simpler, early metabolic pathways is largely unknown. Summarized in this section are a few aspects related to the exploration of the early metabolic pathways.

6.2.1. Let there be LUCA. In principle, the composition of the earliest living forms could be extrapolated from contemporary life forms. We can infer, for example, that a gene, cofactor, or metabolite present in all current species was present in the so-called LCA, or LUCA (last universal common ancestor). Accordingly, the ribosome, ATP, and elongation factor (EF-Tu) are

likely LUCA entities. However, since LUCA, fundamental changes in the chemical composition of this planet have occurred, leading to massive extinction of some species and takeover by others (124). LUCA was likely an obligate anaerobe, and the appearance of oxygen had a dramatic effect on the evolution of metabolism (125). Indeed, considering LUCA's geochemical context yielded a new, unorthodox picture of LUCA's metabolism (126). Overall, a key remaining challenge in the area of evolutionary biochemistry is the description of LUCA's metabolism and its metabolites, reactions, and enzymes.

6.2.2. Reconstruction of founder enzymes. The most ancient enzyme classes seem to have been reliably identified. These include classes that are highly abundant in modern proteomes such as the P-loop nucleotide triphosphate hydrolases and NAD-utilizing Rossmann enzymes (24, 127, 128). These classes also make use of the most abundant cosubstrates and cofactors and thus are likely candidates for the founder enzymes. The actual reconstruction of such founder enzymes is, however, nontrivial. Although ancestral sequence reconstruction is widely applied (123), most reconstructions relate to relatively late evolutionary stages. The few inferred reconstructions from the Archaean age yielded proteins that are essentially identical in structure and function to their modern counterparts (123). A change in cofactor specificity of isocitrate dehydrogenase, from NAD(H) to NADP(H), is to our knowledge the most ancient divergence event explored so far (116). Reconstructions of founder enzymes have not yet been described, primarily because traditional ancestral sequence reconstruction is inapplicable when sequence identity is low (e.g., **Figures 3–6**). However, the approach taken toward the reconstruction of prototypes of primordial amino acid tRNA synthetases (urzymes) seems feasible; computational design has allowed us to surgically remove structural extensions that occurred across more than 3.7 billion years of evolution and thus to construct a minimal, core enzyme (129, 130).

6.2.3. The space of nonenzymatic reactions. The focus of abiotic chemistry research has been chemical reactions that produce the most basic building blocks of life, such as amino acids and nucleosides. However, a better understanding of how contemporary metabolism evolved demands the identification of reactions that are enzyme-catalyzed in contemporary metabolism yet can proceed in the presence of much simpler catalysts (89). Such catalysts, in combination with reconstructed founder enzymes, could give rise to rudimentary metabolic networks. Nonenzymatic reactions also seem relevant to pathway emergence at the later stages of evolution, and hence their investigation has wider implications.

6.3. What We Do Understand, We Can Create

Metabolic engineering is a growing field that also provides new insight into how natural pathways evolved. Mix and match of existing enzymes, including promiscuous ones, is the most flexible strategy (83). Retrograde recruitment can also be a powerful engineering strategy if the desired end product presents a selectable phenotype. Given one heterologously expressed enzyme (the one catalyzing the last pathway reaction), the shortest pathway that the endogenous enzyme repertoire offers can be identified (109). Design and implementation of synthetic pathways also provide deeper understanding of how natural pathways evolved by identifying the simplest reaction sequences that yield a desired product (83, 131, 132) and the thermodynamic and kinetic demands that a physiologically viable reaction sequence must fulfill (57, 58). By employing both enzyme promiscuity and underground reactions, new pathways that replace natural ones can be unraveled in the laboratory (45, 98, 133) (**Figure 8**). These studies are also instrumental in guiding the engineering of new, synthetic pathways.

Finally, metabolite–enzyme coevolution could be applied to develop new compounds, be they materials or pharmaceuticals. Mix and match of various precursor compounds and modifying enzymes of broad specificity can yield large repertoires of derivatives. These repertoires can be screened to isolate biologically active derivatives alongside the enzymes that produce them. Iterations of this process and directed enzyme evolution can then yield a new, optimized compound alongside the pathway that synthesizes it.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

This review has been inspired by numerous discussions with colleagues. We are, foremost, grateful to Ron Milo and his current and former group members, including Dan Davidi, Avi Flamholz, and Arren Bar-Even for insightful inputs. We also thank Adi Goldenzweig, Jagoda Jablonska, Uwe Sauer, and Markus Ralser for comments that were critical in shaping this review. We thank Dan Davidi for guiding the making of **Figure 1**.

LITERATURE CITED

1. Hochachka PW, Somero GN. 1984. *Biochemical Adaptation*. Princeton, N.J.: Princeton Univ. Press
2. Hochachka PW, Somero GN. 2002. *Biochemical Adaptation: Mechanism and Process in Physiological Evolution*. New York: Oxford Univ. Press
3. Harms MJ, Thornton JW. 2013. Evolutionary biochemistry: revealing the historical and physical causes of protein properties. *Nat. Rev. Genet.* 14:559–71
4. Copley SD. 2015. An evolutionary biochemist's perspective on promiscuity. *Trends Biochem. Sci.* 40:72–78
5. Cornish-Bowden A, Pereto J, Cardenas ML. 2014. Biochemistry and evolutionary biology: two disciplines that need each other? *J. Biosci.* 39:13–27
6. Vianello A, Passamonti S. 2016. Biochemistry and physiology within the framework of the extended synthesis of evolutionary biology. *Biol. Direct* 11:7
7. Kocherezhkin VG. 1967. [Evolutionary biochemistry and the origin of life]. *Izv. Akad. Nauk SSSR Biol.* 2:310–13 (in Russian)
8. Wachtershauser G. 1992. Groundworks for an evolutionary biochemistry: the iron-sulphur world. *Prog. Biophys. Mol. Biol.* 58:85–201
9. Khersonsky O, Tawfik DS. 2010. Enzyme promiscuity: a mechanistic and evolutionary perspective. *Annu. Rev. Biochem.* 79:471–505
10. Pandya C, Farelli JD, Dunaway-Mariano D, Allen KN. 2014. Enzyme promiscuity: engine of evolutionary innovation. *J. Biol. Chem.* 289:30229–36
11. O'Brien PJ, Herschlag D. 1999. Catalytic promiscuity and the evolution of new enzymatic activities. *Chem. Biol.* 6:R91–105
12. Hughes AL. 1994. The evolution of functionally novel proteins after gene duplication. *Proc. Biol. Sci.* 256:119–24
13. Piatigorsky J. 2003. Gene sharing, lens crystallins and speculations on an eye/ear evolutionary relationship. *Integr. Comp. Biol.* 43:492–99
14. Soskine M, Tawfik DS. 2010. Mutational effects and the evolution of new protein functions. *Nat. Rev. Genet.* 11:572–82
15. Nam H, Lewis NE, Lerman JA, Lee DH, Chang RL, et al. 2012. Network context and selection in the evolution to enzyme specificity. *Science* 337:1101–4

16. Matsumura I, Ellington AD. 2001. In vitro evolution of beta-glucuronidase into a beta-galactosidase proceeds through non-specific intermediates. *J. Mol. Biol.* 305:331–39
17. Rockah-Shmuel L, Tawfik DS. 2012. Evolutionary transitions to new DNA methyltransferases through target site expansion and shrinkage. *Nucleic Acids Res.* 40:11627–37
18. Noda-Garcia L, Camacho-Zarco AR, Medina-Ruiz S, Gaytan P, Carrillo-Tripp M, et al. 2013. Evolution of substrate specificity in a recipient's enzyme following horizontal gene transfer. *Mol. Biol. Evol.* 30:2024–34
19. Ferla MP, Brewster JL, Hall KR, Evans GB, Patrick WM. 2017. Primordial-like enzymes from bacteria with reduced genomes. *Mol. Microbiol.* 105:508–24
20. Jia B, Cheong GW, Zhang S. 2013. Multifunctional enzymes in archaea: promiscuity and moonlight. *Extremophiles* 17:193–203
21. Weng JK, Noel JP. 2012. The remarkable pliability and promiscuity of specialized metabolism. *Cold Spring Harb. Symp. Quant. Biol.* 77:309–20
22. Dawson NL, Lewis TE, Das S, Lees JG, Lee D, et al. 2017. CATH: an expanded resource to predict protein function through structure and sequence. *Nucleic Acids Res.* 45:D289–95
23. Baier F, Copp JN, Tokuriki N. 2016. Evolution of enzyme superfamilies: comprehensive exploration of sequence-function relationships. *Biochemistry* 55:6375–88
24. Laurino P, Toth-Petroczy A, Meana-Paneda R, Lin W, Truhlar DG, Tawfik DS. 2016. An ancient fingerprint indicates the common ancestry of Rossmann-fold enzymes utilizing different ribose-based cofactors. *PLOS Biol.* 14:e1002396
25. Nagano N, Orengo CA, Thornton JM. 2002. One fold with many functions: the evolutionary relationships between TIM barrel families based on their sequences, structures and functions. *J. Mol. Biol.* 321:741–65
26. Copley RR, Bork P. 2000. Homology among $(\beta\alpha)_8$ barrels: implications for the evolution of metabolic pathways. *J. Mol. Biol.* 303:627–41
27. Baier F, Tokuriki N. 2014. Connectivity between catalytic landscapes of the metallo-β-lactamase superfamily. *J. Mol. Biol.* 426:2442–56
28. Gerlt JA, Babbitt PC, Jacobson MP, Almo SC. 2012. Divergent evolution in enolase superfamily: strategies for assigning functions. *J. Biol. Chem.* 287:29–34
29. Knutson ST, Westwood BM, Leuthaeuser JB, Turner BE, Nguyendac D, et al. 2017. An approach to functionally relevant clustering of the protein universe: active site profile-based clustering of protein structures and sequences. *Protein Sci.* 26:677–99
30. Khersonsky O, Malitsky S, Rogachev I, Tawfik DS. 2011. Role of chemistry versus substrate binding in recruiting promiscuous enzyme functions. *Biochemistry* 50:2683–90
31. Soo VW, Yosaatmadja Y, Squire CJ, Patrick WM. 2016. Mechanistic and evolutionary insights from the reciprocal promiscuity of two pyridoxal phosphate-dependent enzymes. *J. Biol. Chem.* 291:19873–87
32. Rossmann MG, Moras D, Olsen KW. 1974. Chemical and biological evolution of nucleotide-binding protein. *Nature* 250:194–99
33. Goldman AD, Beatty JT, Landweber LF. 2016. The TIM barrel architecture facilitated the early evolution of protein-mediated metabolism. *J. Mol. Evol.* 82:17–26
34. Alam MT, Olin-Sandoval V, Stincone A, Keller MA, Zelezniak A, et al. 2017. The self-inhibitory nature of metabolic networks and its alleviation through compartmentalization. *Nat. Commun.* 8:16018
35. Newton MS, Guo X, Soderholm A, Nasvall J, Lundstrom P, et al. 2017. Structural and functional innovations in the real-time evolution of new $(\beta\alpha)_8$ barrel enzymes. *PNAS* 114:4727–32
36. Jeong H, Tombor B, Albert R, Oltvai ZN, Barabasi AL. 2000. The large-scale organization of metabolic networks. *Nature* 407:651–54
37. Light S, Kraulis P. 2004. Network analysis of metabolic enzyme evolution in *Escherichia coli*. *BMC Bioinform.* 5:15
38. Pfeiffer T, Soyer OS, Bonhoeffer S. 2005. The evolution of connectivity in metabolic networks. *PLOS Biol.* 3:e228
39. Guimera R, Nunes Amaral LA. 2005. Functional cartography of complex metabolic networks. *Nature* 433:895–900

40. Orth JD, Conrad TM, Na J, Lerman JA, Nam H, et al. 2011. A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism—2011. *Mol. Syst. Biol.* 7:535
41. Lima-Mendez G, van Helden J. 2009. The powerful law of the power law and other myths in network biology. *Mol. Biosyst.* 5:1482–93
42. Vorobieva AA, Khan MS, Soumillion P. 2014. *Escherichia coli* D-malate dehydrogenase, a generalist enzyme active in the leucine biosynthesis pathway. *J. Biol. Chem.* 289:29086–96
43. Light S, Kraulis P, Elofsson A. 2005. Preferential attachment in the evolution of metabolic networks. *BMC Genom.* 6:159
44. D'Ari R, Casadesus J. 1998. Underground metabolism. *BioEssays* 20:181–86
45. Notebaart RA, Szappanos B, Kintses B, Pal F, Gyorkei A, et al. 2014. Network-level architecture and the evolutionary potential of underground metabolism. *PNAS* 111:11762–67
46. Kurakin A. 2007. Self-organization versus Watchmaker: ambiguity of molecular recognition and design charts of cellular circuitry. *J. Mol. Recognit.* 20:205–14
47. Tawfik DS. 2010. Messy biology and the origins of evolutionary innovations. *Nat. Chem. Biol.* 6:692–96
48. Jensen RA. 1976. Enzyme recruitment in evolution of new function. *Annu. Rev. Microbiol.* 30:409–25
49. Marcheschi RJ, Li H, Zhang K, Noey EL, Kim S, et al. 2012. A synthetic recursive "+1" pathway for carbon chain elongation. *ACS Chem. Biol.* 7:689–97
50. Bar-Even A, Flamholz A, Noor E, Milo R. 2012. Thermodynamic constraints shape the structure of carbon fixation pathways. *Biochim. Biophys. Acta* 1817:1646–59
51. Leipe DD, Koonin EV, Aravind L. 2003. Evolution and classification of P-loop kinases and related proteins. *J. Mol. Biol.* 333:781–815
52. Ycas M. 1974. On earlier states of the biochemical system. *J. Theor. Biol.* 44:145–60
53. Li F, Hagemeier CH, Seedorf H, Gottschalk G, Thauer RK. 2007. Re-citrate synthase from *Clostridium kluyveri* is phylogenetically related to homocitrate synthase and isopropylmalate synthase rather than to Si-citrate synthase. *J. Bacteriol.* 189:4299–304
54. Casey AK, Hicks MA, Johnson JL, Babbitt PC, Frantom PA. 2014. Mechanistic and bioinformatic investigation of a conserved active site helix in α-isopropylmalate synthase from *Mycobacterium tuberculosis*, a member of the DRE-TIM metallolyase superfamily. *Biochemistry* 53:2915–25
55. Gruer MJ, Artymiuk PJ, Guest JR. 1997. The aconitase family: three structural variations on a common theme. *Trends Biochem. Sci.* 22:3–6
56. Aktas DF, Cook PF. 2009. A lysine-tyrosine pair carries out acid-base chemistry in the metal ion-dependent pyridine dinucleotide-linked β-hydroxyacid oxidative decarboxylases. *Biochemistry* 48:3565–77
57. Bar-Even A, Flamholz A, Noor E, Milo R. 2012. Rethinking glycolysis: on the biochemical logic of metabolic pathways. *Nat. Chem. Biol.* 8:509–17
58. Noor E, Bar-Even A, Flamholz A, Reznik E, Liebermeister W, Milo R. 2014. Pathway thermodynamics highlights kinetic obstacles in central metabolism. *PLOS Comput. Biol.* 10:e1003483
59. Metzenberg RL. 2005. Norman Harold Horowitz, 1915–2005. *Genetics* 171:1445–48
60. Srb AM, Horowitz NH. 1944. The ornithine cycle in *Neurospora* and its genetic control. *J. Biol. Chem.* 154:129–39
61. Horowitz NH. 1945. On the evolution of biochemical syntheses. *PNAS* 31:153–57
62. Horowitz NH. 1965. The evolution of biochemical synthesis—retrospect and prospect. In *Evolving Genes and Proteins*, ed. V Bryson, HJ Vogel, pp. 15–23. New York: Academic
63. Rison SC, Teichmann SA, Thornton JM. 2002. Homology, pathway distance and chromosomal localization of the small molecule metabolism enzymes in *Escherichia coli*. *J. Mol. Biol.* 318:911–32
64. Wilmanns M, Hyde CC, Davies DR, Kirschner K, Jansonius JN. 1991. Structural conservation in parallel β/α-barrel enzymes that catalyze three sequential reactions in the pathway of tryptophan biosynthesis. *Biochemistry* 30:9161–69
65. Diaz-Mejia JJ, Perez-Rueda E, Segovia L. 2007. A network perspective on the evolution of metabolism by gene duplication. *Genome Biol.* 8:R26
66. Granick S. 1965. The evolution of heme and chlorophyll. In *Evolving Genes and Proteins*, ed. V Bryson, HJ Vogel, pp. 67–88: New York: Academic

67. Dailey HA, Dailey TA, Gerdes S, Jahn D, Jahn M, et al. 2017. Prokaryotic heme biosynthesis: multiple pathways to a common essential product. *Microbiol. Mol. Biol. Rev.* 81:e00048-16
68. Rison SC, Thornton JM. 2002. Pathway evolution, structurally speaking. *Curr. Opin. Struct. Biol.* 12:374–82
69. Austin MB, O'Maille PE, Noel JP. 2008. Evolving biosynthetic tangos negotiate mechanistic landscapes. *Nat. Chem. Biol.* 4:217–22
70. Belfaiza J, Parsot C, Martel A, de la Tour CB, Margarita D, et al. 1986. Evolution in biosynthetic pathways: two enzymes catalyzing consecutive steps in methionine biosynthesis originate from a common ancestor and possess a similar regulatory region. *PNAS* 83:867–71
71. Kouidmi I, Levesque RC, Paradis-Bleau C. 2014. The biology of Mur ligases as an antibacterial target. *Mol. Microbiol.* 94:242–53
72. Velasco AM, Leguina JI, Lazcano A. 2002. Molecular evolution of the lysine biosynthetic pathways. *J. Mol. Evol.* 55:445–59
73. Howell DM, Harich K, Xu H, White RH. 1998. α-Keto acid chain elongation reactions involved in the biosynthesis of coenzyme B (7-mercaptoheptanoyl threonine phosphate) in methanogenic Archaea. *Biochemistry* 37:10108–17
74. Howell DM, Graupner M, Xu H, White RH. 2000. Identification of enzymes homologous to isocitrate dehydrogenase that are involved in coenzyme B and leucine biosynthesis in methanoarchaea. *J. Bacteriol.* 182:5013–16
75. Fondi M, Brilli M, Emiliani G, Paffetti D, Fani R. 2007. The primordial metabolism: an ancestral interconnection between leucine, arginine, and lysine biosynthesis. *BMC Evol. Biol.* 7(Suppl. 2):S3
76. Xu H, Zhang Y, Guo X, Ren S, Staempfli AA, et al. 2004. Isoleucine biosynthesis in *Leptospira interrogans* serotype lai strain 56601 proceeds via a threonine-independent pathway. *J. Bacteriol.* 186:5400–9
77. Lefort V, Longueville JE, Gascuel O. 2017. SMS: smart model selection in PhyML. *Mol. Biol. Evol.* 34:2422–24
78. Sonawane PD, Pollier J, Panda S, Szymanski J, Massalha H, et al. 2016. Plant cholesterol biosynthetic pathway overlaps with phytosterol metabolism. *Nat. Plants* 3:16205
79. Lazcano A, Miller SL. 1999. On the origin of metabolic pathways. *J. Mol. Evol.* 49:424–31
80. Copley SD. 2009. Evolution of efficient pathways for degradation of anthropogenic chemicals. *Nat. Chem. Biol.* 5:559–66
81. Trivedi VD, Jangir PK, Sharma R, Phale PS. 2016. Insights into functional and evolutionary analysis of carbaryl metabolic pathway from *Pseudomonas* sp. strain C5pp. *Sci. Rep.* 6:38430
82. Shapir N, Mongodin EF, Sadowsky MJ, Daugherty SC, Nelson KE, Wackett LP. 2007. Evolution of catabolic pathways: genomic insights into microbial *s*-triazine metabolism. *J. Bacteriol.* 189:674–82
83. Erb TJ, Jones PR, Bar-Even A. 2017. Synthetic metabolism: metabolic engineering meets enzyme design. *Curr. Opin. Chem. Biol.* 37:56–62
84. Patrick WM, Quandt EM, Swartzlander DB, Matsumura I. 2007. Multicopy suppression underpins metabolic evolvability. *Mol. Biol. Evol.* 24:2716–22
85. Linster CL, van Schaftingen E, Hanson AD. 2013. Metabolite damage and its repair or pre-emption. *Nat. Chem. Biol.* 9:72–80
86. Lu ZJ, Markham GD. 2002. Enzymatic properties of *S*-adenosylmethionine synthetase from the archaeon *Methanococcus jannaschii*. *J. Biol. Chem.* 277:16624–31
87. da Silva RR, Dorrestein PC, Quinn RA. 2015. Illuminating the dark matter in metabolomics. *PNAS* 112:12549–50
88. Zampieri M, Sekar K, Zamboni N, Sauer U. 2017. Frontiers of high-throughput metabolomics. *Curr. Opin. Chem. Biol.* 36:15–23
89. Sutherland JD. 2017. Opinion: Studies on the origin of life—the end of the beginning. *Nat. Rev. Chem.* 1:0012
90. Keller MA, Piedrafita G, Ralser M. 2015. The widespread role of non-enzymatic reactions in cellular metabolism. *Curr. Opin. Biotechnol.* 34:153–61
91. Piedrafita G, Keller MA, Ralser M. 2015. The impact of non-enzymatic reactions and enzyme promiscuity on cellular metabolism during (oxidative) stress conditions. *Biomolecules* 5:2101–22

92. Moellering RE, Cravatt BF. 2013. Functional lysine modification by an intrinsically reactive primary glycolytic metabolite. *Science* 341:549–53
93. Xu XL, Grant GA. 2016. Mutagenic and chemical analyses provide new insight into enzyme activation and mechanism of the type 2 iron-sulfur L-serine dehydratase from *Legionella pneumophila*. *Arch. Biochem. Biophys.* 596:108–17
94. Bar-Even A, Noor E, Savir Y, Liebermeister W, Davidi D, et al. 2011. The moderately efficient enzyme: evolutionary and physicochemical trends shaping enzyme parameters. *Biochemistry* 50:4402–10
95. Bar-Even A, Tawfik DS. 2013. Engineering specialized metabolic pathways—is there a room for enzyme improvements? *Curr. Opin. Biotechnol.* 24:310–19
96. Gravel E, Poupon E. 2010. Biosynthesis and biomimetic synthesis of alkaloids isolated from plants of the *Nitraria* and *Myrioneuron* genera: an unusual lysine-based metabolism. *Nat. Prod. Rep.* 27:32–56
97. Poupon E, Gravel E. 2015. Manipulating simple reactive chemical units: fishing for alkaloids from complex mixtures. *Chemistry* 21:10604–15
98. Kim J, Kershner JP, Novikov Y, Shoemaker RK, Copley SD. 2010. Three serendipitous pathways in *E. coli* can bypass a block in pyridoxal-5′-phosphate synthesis. *Mol. Syst. Biol.* 6:436
99. Tani Y, Dempsey WB. 1973. Glycolaldehyde is a precursor of pyridoxal phosphate in *Escherichia coli* B. *J. Bacteriol.* 116:341–45
100. Keller MA, Turchyn AV, Ralser M. 2014. Non-enzymatic glycolysis and pentose phosphate pathway-like reactions in a plausible Archean ocean. *Mol. Syst. Biol.* 10:725
101. Keller MA, Zylstra A, Castro C, Turchyn AV, Griffin JL, Ralser M. 2016. Conditional iron and pH-dependent activity of a non-enzymatic glycolysis and pentose phosphate pathway. *Sci. Adv.* 2:e1501235
102. Messner CB, Driscoll PC, Piedrafita G, De Volder MFL, Ralser M. 2017. Nonenzymatic gluconeogenesis-like formation of fructose 1,6-bisphosphate in ice. *PNAS* 114:7403–7
103. Keller MA, Kampjut D, Harrison SA, Ralser M. 2017. Sulfate radicals enable a non-enzymatic Krebs cycle precursor. *Nat. Ecol. Evol.* 1:0083
104. Laurino P, Tawfik DS. 2017. Spontaneous emergence of *S*-adenosylmethionine and the evolution of methylation. *Angew. Chem. Int. Ed.* 56:343–45
105. Heinrich R, Hoffmann E. 1991. Kinetic parameters of enzymatic reactions in states of maximal activity; an evolutionary approach. *J. Theor. Biol.* 151:249–83
106. Heinrich R, Klipp E. 1996. Control analysis of unbranched enzymatic chains in states of maximal activity. *J. Theor. Biol.* 182:243–52
107. Thomas S, Fell DA. 1998. The role of multiple enzyme activation in metabolic flux control. *Adv. Enzyme Regul.* 38:65–85
108. Newton MS, Arcus VL, Patrick WM. 2015. Rapid bursts and slow declines: on the possible evolutionary trajectories of enzymes. *J. R. Soc. Interface* 12:20150036
109. Birmingham WR, Starbird CA, Panosian TD, Nannemann DP, Iverson TM, Bachmann BO. 2014. Bioretrosynthetic construction of a didanosine biosynthetic pathway. *Nat. Chem. Biol.* 10:392–99
110. Braakman R, Smith E. 2012. The emergence and early evolution of biological carbon-fixation. *PLOS Comput. Biol.* 8:e1002455
111. Firn RD, Jones CG. 2000. The evolution of secondary metabolism—a unifying model. *Mol. Microbiol.* 37:989–94
112. Benderoth M, Textor S, Windsor AJ, Mitchell-Olds T, Gershenzon J, Kroymann J. 2006. Positive selection driving diversification in plant secondary metabolism. *PNAS* 103:9118–23
113. Park JW, Park SR, Nepal KK, Han AR, Ban YH, et al. 2011. Discovery of parallel pathways of kanamycin biosynthesis allows antibiotic manipulation. *Nat. Chem. Biol.* 7:843–52
114. Hamed RB, Gomez-Castellanos JR, Henry L, Ducho C, McDonough MA, Schofield CJ. 2013. The enzymes of beta-lactam biosynthesis. *Nat. Prod. Rep.* 30:21–107
115. Kawai S, Murata K. 2008. Structure and function of NAD kinase and NADP phosphatase: key enzymes that regulate the intracellular balance of NAD(H) and NADP(H). *Biosci. Biotechnol. Biochem.* 72:919–30
116. Zhu G, Golding GB, Dean AM. 2005. The selective cause of an ancient adaptation. *Science* 307:1279–82
117. Cahn JK, Werlang CA, Baumschlager A, Brinkmann-Chen S, Mayo SL, Arnold FH. 2017. A general tool for engineering the NAD/NADP cofactor preference of oxidoreductases. *ACS Synth. Biol.* 6:326–33

118. Cvetesic N, Palencia A, Halasz I, Cusack S, Gruic-Sovulj I. 2014. The physiological target for LeuRS translational quality control is norvaline. *EMBO J.* 33:1639–53
119. Cvetesic N, Semanjski M, Soufi B, Krug K, Gruic-Sovulj I, Macek B. 2016. Proteome-wide measurement of non-canonical bacterial mistranslation by quantitative mass spectrometry of protein modifications. *Sci. Rep.* 6:28631
120. Buchan DW, Shepherd AJ, Lee D, Pearl FM, Rison SC, et al. 2002. Gene3D: structural assignment for whole genes and genomes using the CATH domain structure database. *Genome Res.* 12:503–14
121. Gophna U, Bapteste E, Doolittle WF, Biran D, Ron EZ. 2005. Evolutionary plasticity of methionine biosynthesis. *Gene* 355:48–57
122. Martin WF, Weiss MC, Neukirchen S, Nelson-Sathi S, Sousa FL. 2016. Physiology, phylogeny, and LUCA. *Microb. Cell* 3:582–87
123. Gumulya Y, Gillam EM. 2017. Exploring the past and the future of protein evolution with ancestral sequence reconstruction: the 'retro' approach to protein engineering. *Biochem. J.* 474:1–19
124. Fischer WW, Hemp J, Valentine JS. 2016. How did life survive Earth's great oxygenation? *Curr. Opin. Chem. Biol.* 31:166–78
125. Raymond J, Segre D. 2006. The effect of oxygen on biochemical networks and the evolution of complex life. *Science* 311:1764–67
126. Weiss MC, Sousa FL, Mrnjavac N, Neukirchen S, Roettger M, et al. 2016. The physiology and habitat of the last universal common ancestor. *Nat. Microbiol.* 1:16116
127. Alva V, Soding J, Lupas AN. 2015. A vocabulary of ancient peptides at the origin of folded proteins. *eLife* 4:e09410
128. Goncearenco A, Berezovsky IN. 2015. Protein function from its emergence to diversity in contemporary proteins. *Phys. Biol.* 12:045002
129. Li L, Francklyn C, Carter CW Jr. 2013. Aminoacylating urzymes challenge the RNA world hypothesis. *J. Biol. Chem.* 288:26856–63
130. Sapienza PJ, Li L, Williams T, Lee AL, Carter CW Jr. 2016. An ancestral tryptophanyl-tRNA synthetase precursor achieves high catalytic rate enhancement without ordered ground-state tertiary structures. *ACS Chem. Biol.* 11:1661–68
131. Delepine B, Libis V, Carbonell P, Faulon JL. 2016. SensiPath: computer-aided design of sensing-enabling metabolic pathways. *Nucleic Acids Res.* 44:W226–31
132. Bar-Even A. 2016. Formate assimilation: the metabolic architecture of natural and synthetic pathways. *Biochemistry* 55:3851–63
133. Szappanos B, Fritzemeier J, Csorgo B, Lazar V, Lu X, et al. 2016. Adaptive evolution of complex innovations through stepwise metabolic niche expansion. *Nat. Commun.* 7:11607

Annual Review of Biochemistry

Volume 87, 2018

Contents

Errata

An online log of corrections to *Annual Review of Biochemistry* articles may be found at http://www.annualreviews.org/errata/biochem