æ

19. ENZYMES AND METABOLIC PATHWAYS

25 September 2021

Virtually every aspect of cell biology relies on intermolecular interactions, some of which extend into long, multi-step pathways. In many pairwise interactions, the interacting partners represent a coevolutionary loop with each participant constituting part of the selective environment of the other. This is true, for example, of crosstalk between transcription factors and their binding sites, and between proteins conferring specificity in vesicle trafficking in eukaryotes. Metabolic pathways are different in that there is typically no direct physical contact between the major players, with each simply imposing a chemical modification on a nonevolvable substrate, which is then passed on to another enzyme. Most cellular processes involving resource acquisition and biosynthesis involve multiple steps of this sort.

Catabolic pathways break down complex organic molecules into simpler ones for a variety of reasons: the release of energy for use in other cellular processes; reduction to simpler precursors for use in specific biosynthetic pathways; and detoxification of harmful substances. Sometimes an intermediate metabolite in a pathway has no function other than to serve as a link between two compounds that are useful. Frequently, a link serves multiple purposes, e.g., generating a useful precursor molecule for a biosynthetic pathway, and in the process regenerating ATP and/or NADH. In contrast, anabolic pathways, which require energy input, are responsible for the synthesis and elaboration of the constituents of proteins, nucleic acids, lipids, and other molecules from which cells are built.

Metabolic pathways motivate several fundamental evolutionary questions. First, what limits the level of perfection of their enzymatic constituents? Given that biomass and energy production are central to all of biology, and given that the costs of protein machineries are substantial, one would expect metabolic enzymes to have high efficiencies and specificities so as to minimize the infrastructure costs. Second, what are the sources of the enzymes that enable the emergence of novel metabolic features? One expects a descent-with-modification scenario, with functional alterations requiring moderate numbers of steps, but how does a specialized enzyme integrated into metabolism become free to take on a new activity? Third, what determines the structure of metabolic pathways, and how recalcitrant are the participants in such pathways to replacement? Again, one might expect such features to be among the most conserved aspects of cell biology.

Comparative analyses of metabolic machinery provide surprising answers to all of these questions. Despite the impressive performance of enzymes, their catalytic efficiencies are generally far below the biophysical limits, whereas inherent inaccuracies provide latent potential for the emergence of new catalytic functions. Despite

the centrality of metabolic pathways to all of life, both their structure and the identities of the underlying participants exhibit substantial phylogenetic variation. Thus, as with many other cell biological features, numerous aspects of the metabolic machinery appear to be influenced by the intervention of nonadaptive processes.

Enzymes

Essentially all metabolic pathways are driven by enzymatic proteins whose primary roles are to facilitate the transformation of substrates into products. Thus, any attempt to understand the mechanisms of metabolic-pathway evolution must begin with a basic appreciation of the fundamental features of enzymes. First, enzymes generally do not create entirely novel biochemical reactions, but simply magnify transformation rates beyond what happens spontaneously. The level of enhancement of catalytic rates can be enormous, typically in the range of 10^7 to $10^{19}\times$ the corresponding rates in the absence of enzyme (Wolfenden and Snider 2001). Second, enzymes are recyclable – after they disengage with a substrate/product molecule, they are free to enter into a new reaction. Third, most enzymes carry out very simple reactions, usually involving no more than two substrates. Such simplicity likely results from the diminishingly small probabilities of joint encounters of multiple sets of molecules. Third, there are usually no more than one or two substrate modifications (in the form of chemical-bond changes) per enzymatic step. A natural consequence of such simplicity is the necessity of multistep pathways for the complete breakdown or synthesis of complex molecules.

Basic enzymology. Although most enzymes catalyze simple chemical reactions, these are almost always targeted to highly specialized substrates. Biochemists generally classify enzymes into six categories according to the reaction types in which they engage, although layers of subdivision also enter the overall classification scheme. Oxidoreductases (e.g., dehydrogenases and oxidases) transfer electrons from one substrate to another, whereas transferases (e.g., transaminases and kinases) transfer larger functional groups (e.g., amine or phosphate groups) from one substrate to another. Hydrolases (e.g., lipases, phosphatases, and peptidases) cleave chemical bonds by addition of water, leading to the breakdown of one substrate into two product molecules, whereas lyases (e.g., decarboxylases) break bonds in a nonhydrolytic fashion. Isomerases make molecular changes within a single molecule, whereas ligases (e.g., synthetases) join two molecules.

Although the structural features of enzymes dictate their chemical potential and specificity, the actual reaction rates are highly dependent on the substrate concentration. The basic issues are outlined in Foundations 19.1, where the famous two-parameter formulation of Michaelis and Menten (1913) to describe this relationship is derived. Under this model, an enzymatic reaction rate is viewed as the net outcome of two processes: 1) the rate of productive encounters between enzyme and substrate, which is mainly a function of physical factors and molecular densities; and 2) the rate of the ensuing chemical transformation, which is a function of enzyme structure and the nature of the interaction. These features define two key quantitative parameters for every enzyme-substrate interaction: 1) the maximum

reaction rate (per unit enzyme concentration) at saturating substrate concentration, $k_{\rm cat}$, which is a measure of catalytic efficiency once engaged with a substrate molecule (often also referred to as the enzyme turnover number); and 2) the substrate concentration at which the reaction rate equals half the maximum rate, $K_{\rm S}$, which is a function of the physical rates of association and dissociation between enzyme and substrate (Figure 19.1). The hyperbolic Michaelis-Menten relationship between the rate of product formation and substrate composition arises in many other contexts in biology, including nutrient-uptake and growth responses to nutrient concentration (Chapters 9 and 18).

To understand how the kinetic parameters of enzyme performance are guided by natural selection, it is useful to know the typical substrate concentrations within the intracellular environment. At sufficiently high concentrations, such that the time intervals between substrate encounter rates are negligible, an enzymatic reaction rate is defined by $k_{\rm cat}$ alone, whereas at sufficiently low concentrations the reaction rate is a function of the composite parameter, $k_{\rm cat}/K_{\rm S}$, which is the slope of the response to low substrate concentrations (generally called the kinetic efficiency; Foundations 19.1). If substrate concentrations are typically $\gg K_{\rm S}$, selection for higher reaction rates would largely be focused on the catalytic efficiency. However, at low substrate concentrations, selection should also favor a reduction in $K_{\rm S}$.

Given the underlying determinants of K_S, Equation 19.1.1c, it is clear that selection can enhance the flux rate of a reaction in three ways: 1) increasing the rate of association of enzyme and substrate ($k_{\rm a}$); 2) decreasing the backwards rate of disengagement ($k_{\rm d}$); and 3) increasing the forward rate of catalysis ($k_{\rm cat}$). Crowley (1974) refers to such changes as influences on affinity, efficiency, and velocity. Unless there are physical tradeoffs in the mechanisms allowing improvement in these three parameters, selection is likely to act in a directional fashion on each of them. However, as will be noted below, there are biophysical limits to what can be accomplished. Moreover, as can be seen from Equations 19.1.1c and 19.1.2, changes in $k_{\rm cat}$ influence both the maximum reaction rate and the half-saturation constant, so these two features are not entirely independent (see also Foundations 18.1).

Only a few attempts have been made to estimate the intracellular concentrations of various metabolites. In *E. coli,* one of the few species for which kinetic parameters are available for a large number of enzymes, for cells in exponential-growth phase on several different carbon substrates, most metabolite concentrations are in excess of the $K_{\rm S}$ values of their associated enzymes (Bennett et al. 2009). Indeed, $\sim 60\%$ of metabolites (including ATP and NAD^+) have concentrations more than tenfold higher than their associated $K_{\rm S}$, implying saturation-level flux rates. Given the costs of protein biosynthesis, at least for this set of enzymes, this suggests that selection puts a premium on catalytic-core features that maximize the flux per unit enzyme (high $k_{\rm cat}$). Nonetheless, as will be shown below, $k_{\rm cat}$ is often far below what seems possible biophysically.

Degree of molecular perfection. A perfect enzyme would be one with such a high catalytic efficiency that the reaction rate is limited only by the rate of encounters with substrate molecules. In this case, the reaction would be effectively instantaneous following a productive colocalization of enzyme and substrate molecules. Thus, understanding the determinants of enzyme-substrate encounter rates is an

essential starting point to evaluating the limits of what is possible vs. what natural selection has actually accomplished.

The rate at which randomly and homogeneously distributed particles engage with each other in an aqueous medium can be viewed as the product of three functions: 1) the rate of encounter by diffusion-like processes (k_e); 2) a potential multiplier to account for further attractive (or repulsive) charges between enzyme and substrate ($m_e > 1$ or < 1, respectively); and 3) the fraction of encounters that are in the correct orientation for proper engagement (p_e),

$$k_{\text{enc}} = k_{\text{e}} m_{\text{e}} p_{\text{e}}. \tag{19.1}$$

Assuming that encounters between enzyme (E) and substrate (S) occur via a three-dimensional diffusion process, and that both types of molecules can be approximated as spheres, with the latter being much smaller than the former, the diffusion limit is

$$k_{\text{e}} \simeq (1.7 \times 10^9) \left(\frac{r_{\text{E}}}{r_{\text{S}}} \right), \tag{19.2}$$

in units of $\text{M}^{-1}\text{sec}^{-1}$, where r_{E}, r_{S} are the effective radii of enzyme and substrate particles, and M denotes the number of moles per liter (Foundations 7.2). The assumption that $r_{\text{E}} \gg r_{\text{S}}$ is entirely reasonable for metabolic enzyme functions. For example, the twenty amino acids have individual masses ranging from 75 to 204 g/mol, and most enzymes contain on the order of a few hundred amino acids, whereas the molecular weight (g/mol) of glucose is 180, of NADH is 663, of ATP is 507, and of nucleosides A, G, C, and T is 135, 151, 111, and 126 g/mol, respectively. Thus, noting that a radius is roughly proportional to the one-third power of mass, depending on the enzyme and substrate, for metabolic reactions $r_{\text{E}}/r_{\text{S}}$ will typically be in the neighborhood of 10 to 100, implying a k_{e} on the order of 10^{10} to 10^{11} $\text{M}^{-1}\text{sec}^{-1}$.

Although numerous factors can cause the overall encounter rate, k_{enc}, to deviate from the expectation based on diffusion alone, much less is known about the two other components in Equation 19.1. Early attempts to estimate the fraction of productive encounters, p_{e}, generally assumed that overlap between a small circular patch on the surface of the enzyme (and usually the substrate) is critical for proper adhesion (Berg and von Hippel 1985; Janin 1997; Camacho et al. 2000). However, simply multiplying the overall encounter rate by the fraction of properly oriented molecules yields rates that are too low (Berg 1985; Schlosshauer and Baker 2004), probably because some initially unfavorable encounters undergo rotational diffusion, with the two molecules sliding around each other rather than completely disengaging. If at least one contact point is correct, rotational movement may quickly locate a second favorable contact, thereby facilitating the remaining fine tuning (Northrup and Erickson 1992). All of these things taken together, orientation limitations can still reduce k_{enc} by as much as 100 to 1000 fold (Schlosshauer and Baker 2004), i.e., p_{e} is commonly < 0.01. Less clear is the degree to which electrostatic-charge effects enhance an enzyme's likelihood of productive encounters (Zhou 1993). For a very well-studied enzyme, barnase, and its inhibitor, barstar, Janin (1997) estimated m_{e} to be $\simeq 10^5$, but because $p_{\text{e}} \simeq 10^{-5}$ for this particular interaction, electrostatic effects and hydrophobic steering essentially compensate for each other, rendering $k_{\text{enc}} \simeq k_{\text{e}}$.

One final point to consider is that the internal milieu of a cell is hardly the open-water environment assumed in most diffusion theory (Chapter 7). Rather, 20 to 40% of the cytoplasmic volume of a typical cell is occupied by proteins and other macromolecules (Zimmerman and Trach 1991; Luby-Phelps 2000; Ellis 2001). As a consequence, the distance between proteins is on the order of the width of the proteins themselves. On the one hand, molecular crowding reduces the volume that must be searched to locate a small solute, but on the other hand, transient molecular confinement can inhibit free diffusion. Although the net effects on small metabolites are minor, the diffusion coefficients for proteins are reduced by 10 to 50× in *E. coli* (Elowitz et al. 1999; Konopka et al. 2006; Nenninger et al. 2010), and toward the lower end of this range in eukaryotic cells (Luby-Phelps 2000; Dix and Verkman 2008).

The preceding considerations allow a rough calculation of the maximum achievable rate of an enzymatic reaction, conditioned on the internal state of today's cells. As noted above, the ideal situation would be one in which the overall reaction rate is dictated only by the rate of encounter, which is expected to be on the order of 10^{10} to 10^{11} $M^{-1}sec^{-1}$ in a purely aqueous environment, assuming spherical particles with all contacts being productive regardless of orientation. Roughly speaking, after accounting for inappropriate orientations and the effects of molecular crowding in a cellular environment, the best that may be physically achievable inside cells is an association rate of order 10^8 to 10^9 $M^{-1}sec^{-1}$.

How does this biophysical limit to the catalytic rate compare with the evolved capacities of enzymes? Returning to the Michaelis-Menten formula, Equation 19.1.4, we see that at low substrate concentrations (the situation under diffusion limitation), the reaction rate is closely approximated by $k_{cat}[E_T][S]/K_S$, where the quantities in brackets denote the molar concentrations (M, in moles per liter) of enzyme and substrate. Because $K_S \simeq k_{cat}/k_e$ (Foundations 19.1), this reduces to $k_e[E_T][S]$, showing that the preceding arguments yield an upper limit to the kinetic efficiency, k_{cat}/K_S, an observable composite parameter.

Using this logic and early observations on triosephosphate isomerase, Albery and Knowles (1976a,b) argued that many enzymes have been pushed to this level of perfection by natural selection, and hence have reached the end of their "evolutionary development." However, summarizing kinetic data for thousands of enzymes in dozens of species, Bar-Even et al. (2011) and Davidi et al. (2018) reached a substantially different conclusion. The average enzyme has a $k_{cat}/K_S \simeq 10^5$ $M^{-1}sec^{-1}$, several orders of magnitude below the apparent diffusion limit (Figure 19.2). Although most rate estimates are derived from *in vitro* aqueous environments, they are highly correlated with *in vivo* estimates (Davidi et al. 2016), with the average value of the latter for specific enzymes being somewhat lower than the *in vitro* estimates but generally by no more than 10×.

The simplest interpretation of these observations is that, on average, only one in 10^3 to 10^4 collisions between an enzyme and a substrate molecule typically results in a productive interaction (Bar-Even et al. 2015). In *E. coli,* the average catalytic rate of surveyed enzymes is $k_{cat} \simeq 10$ sec^{-1}, with an approximate range of 10^{-2} to 10^3 (Davidi et al. 2016, 2018; Chen and Nielsen 2021), and a survey based on a smaller number of enzymes in the plant *Arabidopsis* yields k_{cat} estimates of a similar order of magnitude. Based on a number of biochemical considerations (e.g., electrostatic

and hydrogen-bond interactions, bond cleavage, exclusion of water molecules), the approximate upper bound of k_{cat} for metabolic reactions is $\sim 10^6$ sec^{-1} (Hammes 2002; Benkovic and Hammes-Schiffer 2003). Remarkably, there is no correlation between an enzyme's catalytic rate and the degree to which the reaction goes forward spontaneously, i.e., the activation barrier, nor is there a correlation with the *in vivo* flux rate of the reaction or with the expression level of the enzyme (Davidi et al. 2018).

Taken together, these observations suggest that enzyme efficiencies are generally far below what should be physically possible, one of the most dramatic examples being the RuBisCo protein discussed in the preceding chapter. The most compelling explanation for such striking disparities is a form of the drift-barrier hypothesis (Hartl et al. 1985; Heckmann et al. 2018), which postulates a sort of diminishing-returns epistasis, such that as catalytic rates reach higher and higher levels, additional improvements enhance organismal fitness to a lesser and lesser extent. This requires that the mapping of fitness on catalytic rate eventually reaches a plateau, much like the Michaelis-Menten relationship itself. One reason why such a relationship is likely to exist will be discussed below in the context of the sensitivity of the fluxes through enzymatic pathways – such networks dilute the selection intensity on the individual components, although not entirely so, and not equally for all pathway constituents.

Consistent with this theory, Bar-Even et al. (2011) found that enzymes involved in secondary metabolism are on average $\sim 30\times$ less efficient than those involved in central metabolism, as expected if selection operates less efficiently on downstream enzymes, assuming these have variants with diminished fitness effects owing to their lower degree of entrenchment in cellular biochemistry. In addition, prokaryotic enzymes have slightly better kinetics than those from eukaryotes (Bar-Even et al. 2011), as expected for species with higher effective population sizes, although the differences are not large, possibly because of uncertainties with data drawn from a wide array of enzymes, species, and experimental conditions.

An additional issue with respect to the drift-barrier hypothesis for the low catalytic rates of enzymes is that the flux rate of a reaction is equal to the product of the catalytic rate per enzyme molecule and the enzyme concentration. Proteins are energetically costly (Chapter 17), so any increase in the catalytic rate per enzyme molecule would presumably reduce the number of molecules required for reactions, thereby freeing up resources for other cellular functions. This further emphasizes the surprisingly low catalytic efficiency of enzymes. An alternative argument for low catalytic rates is that enzymes face biophysical compromises, such that any improvement in efficiency comes at the cost of enzyme stability, solubility, or accuracy. However, although such tradeoffs surely exist, they are not universal (Nguyen et al. 2017; Klesmith et al. 2017). Thus, the relatively low efficiencies of enzymes remains one of the big unsolved mysteries of evolutionary biochemistry.

Enzyme promiscuity. In addition to being far from perfect in terms of catalytic rates, most enzymes also bind to inappropriate substrates at fairly high frequencies, the case of RuBisCo discussed in the previous chapter again providing a dramatic example. Although quantitative analyses have been largely confined to bacteria, error rates at normal cellular concentrations of off-target substrates are commonly

in the range of 0.1% to well over 1.0% (e.g., Wilks et al. 1988; Yano et al. 1998; Rothman and Kirsch 2003; Rakus et al. 2008; Ge et al. 2014; Fernandes et al. 2015). A review of ~ 250 enzymes in *E. coli* suggest an average error rate of 0.0045 (Notebaart et al. 2014), with 37% of metabolic enzymes engaging with multiple substrates at biologically significant rates (Nam et al. 2012)

Additional large-scale studies in *E. coli* illustrate the dramatic reservoir of latent activities of proteins. For example, $\sim 20\%$ of single-gene knockouts in this species can be rescued by overexpression of at least one non-cognate gene (Patrick et al. 2007). Starting with a full library of *E. coli* constructs, each overexpressing a single endogenous protein residing on a plasmid, Soo et al. (2011) found measurable resistance to 237 novel (and likely never-before encountered) toxins and antibiotics in $\sim 0.4\%$ of cases. In addition, using repeated cycles of mutagenic PCR, cloning, transformation, and selective challenge, Aharoni et al. (2005) generated mutants for phosphotriesterase capable of hydrolyzing substrates that had been rarely, if ever, encountered before. In this latter study, depending on the substrate selected for, the ancestral gene function sometimes declined but in other cases increased, the end result being proteins with substantially remodeled catalytic sites with their own novel profiles of promiscuous functions.

The nature of enzyme promiscuity ranges from substrate ambiguity (catalysis of the same reaction with alternative substrates) to catalytic ambiguity (catalysis of different reactions with the same substrate) (Copley 2003; Khersonsky et al. 2006; Tokuriki and Tawfik 2009; Babtie et al. 2010; Jia et al. 2013). In some cases promiscuity can be condition dependent, with the specific reactions and substrates differing depending on the local chemistry. Even completely randomly designed proteins have a high probability of having inherent functions, as they often harbor surface pockets capable of a variety of catalytic activities (Skolnick and Gao 2013).

Although the cost of enzyme promiscuity for cell performance and fitness remains unclear, its ubiquity implies that when confronted with new biochemical challenges, adaptation need not await the mutational origin of new functions from scratch. Indeed, despite the relatively low kinetic rates of promiscuous interactions, they are still generally orders of magnitude greater than rates of noncatalyzed reactions, thereby providing a well-endowed (if far from perfect) starting point for adaptive exploitation. Not surprisingly, it has been repeatedly found that given sufficient selection pressure and a large enough population size, cells are capable of evolving novel metabolic features in remarkably short periods of time (often in < 1000 cell divisions in laboratory evolution experiments).

Often, these sorts of experiments start with complete deletions of a key enzyme, and then reveal the emergence of compensatory structural and/or regulatory changes at seemingly unrelated loci (Kim et al. 2010; Blank et al. 2014). Evolutionary advancement typically initiates with the fortuitous appearance of just a few mutations enhancing a baseline level of promiscuity followed by a relatively rapid remodeling of the catalytic core (Matsumura and Ellington 2001; Copley 2014, 2020; Kim et al. 2019; Morgenthaler et al. 2019). Other long-term evolution studies with *E. coli* have established novel pathways for the uptake and utilization of nonnative carbon resources, through broadened substrate utilization mechanisms involving specific enzymes (Boronat et al. 1983; Lee and Palsson 2010) and/or by the emergence of novel transporters for resource uptake (Blount et al. 2012; Quandt et al. 2014).

With a > 3-billion year legacy of evolution distributed over countless proteins in an enormous number of phylogenetically independent lineages, there may be few enzymatic reactions whose evolution is absolutely dependent on *de novo* mutations. Nowhere is the evidence for this more striking than in the field of industrial enzymology, where substantial progress has been made in developing enzymes with novel specificities, activities, and stability, again typically starting with pre-existing proteins with promiscuous functions and progressively modifying these via strategies involving random mutagenesis and screening and/or directed evolution (Johannes and Zhao 2006; Hult and Berglund 2007; Brustad and Arnold 2011; Bornscheuer et al. 2012; Abatemarco et al. 2013).

Pathway Flux Control

The construction/deconstruction of metabolites often involves a series of steps carried out by different enzymes, the glycolytic pathway being one of many notable examples (Figure 19.3). This raises questions about the degree to which the overall flux rate of the final product is dictated by the properties of the individual pathway constituents. Some aspects of these issues can be evaluated by treating the flux of metabolites through a pathway as a steady-state process with the metabolite entering the pathway having a constant concentration (as would be the case with a stable rate of nutrient intake) and no feedback inhibition from the final product. Under such a framework, the internal metabolites in the pathway will also be maintained at equilibrium concentrations, so that the rate of entry into the pathway must equal the rates of flux through each subsequent step, including the final output from the system.

As outlined in Foundations 19.2, the total flux rate through a linear pathway is a function of the enzyme kinetics operating at each step. The transition rates between all pairs of adjacent metabolites are in turn defined by both the concentrations and the biochemical features of each enzyme in the pathway. The general expression for the equilibrium flux rate is a fairly complicated function of the features at individual steps, showing that there are innumerable ways to regulate overall pathway flux. However, one fairly simple question can be asked (and potentially answered) – How dependent is the overall flux rate on the features of enzymes at successive positions in the pathway? Mathematical analysis (Foundations 19.3) leads to the prediction that the sensitivity of the overall flux rate declines dramatically (exponentially) with increasing downstream position of enzymes in the pathway. This prediction appears to also hold for segments of branched pathways, with enzymes appearing earlier in branches being under stronger selective constraint (Rausher 2012). If this hypothesis is correct, enzymes higher up in a pathway should exhibit signs of stronger selection for the level of molecular refinement.

Pathway position and the strength of selection. Although there is need for more comparative work, the existing data are at least superficially consistent with this expectation. For example, consistent with the flux-control model, for the linear pathways for amino-acid biosynthesis in *E. coli,* the response rates and the maximum response levels to amino-acid depletion decline with the position of enzymes in the

pathway (Zaslaver et al. 2004). In the anthocyanin production pathway in plants, upstream enzymes evolve more slowly than those downstream (Rausher et al. 1999; Lu and Rausher 2003), and this pattern has been observed repeatedly with other biosynthetic pathways in plants and other eukaryotes (Vitkup et al. 2006; Ramsay et al. 2009; Wright and Rausher 2010). Arguably, these differences reflect a reduction in the efficiency of purifying selection operating to remove mildly deleterious mutations from downstream genes. Genes whose products are associated with pathways with higher flux rates also exhibit low rates of amino-acid sequence evolution (Colombo et al. 2014).

The significant influence of pathway position is also illustrated by numerous studies in yeast and mammals showing that genes whose products have more interacting partners evolve more slowly than those with fewer interactors, i.e., are more constrained from accepting amino-acid changes (Fraser et al. 2002; Vitkup et al. 2006; Yang et al. 2009; Montanucci et al. 2011). The only known exception to the predicted pattern arises in a signal-transduction pathway in *Drosophila* (Alvarez-Ponce et al. 2009). Such patterns are consistent with the general flux-control model, given that highly connected genes must on average enter earlier into pathways than more downstream genes with fewer interacting partners. It should be noted, however, that enzymes entering into a pathway upstream of a large number of branchpoints will also tend to have more pleiotropic effects, which could magnify the overall strength of purifying selection for reasons other than pathway location and flux control (Ramsay et al. 2009).

Finally, this prediction of the evolution of greater control of overall pathway flux by upstream enzymes is not absolute. All other things being equal, changes in upstream enzymes are likely to be more effective. But all other things may not always be equal. For example, some enzymes may have architectures that make them more or less vulnerable to change for reasons unassociated with pathway position, and enzymes that are utilized in multiple contexts may be constrained with respect to overall cellular concentrations.

Speed vs. efficiency. Beyond the issue of the relatively low catalytic rates of many enzymes, most catabolic pathways are much less efficient at extracting energy from substrate molecules than seems possible. For example, glycolysis, a significant mechanism for ATP production in most cells, makes just two ADP-to-ATP conversions per glucose molecule consumed, whereas the energy contained within glucose is sufficient for up to four such conversions. As a consequence, as glycolysis converts single glucose molecules to two lactic acid molecules, only about 50% of the energy released is restored in the form of ATPs produced, the remainder being lost as heat to the environment (as will be noticed in any well-nourished compost pile).

Given that energy can be hard to come by, it would seem that metabolism would wring as much energetic advantage out of feed molecules as possible, so why is so much change left on the counter? One possibility, outlined in Foundations 19.3, relates to the intrinsic tradeoff between the rate of a chemical reaction and the efficiency of product formation.

As noted above, assuming constant concentrations of substrate and product molecules, a system of metabolic reactions will reach a steady-state flux rate, with the rate of resource consumption being balanced by the rate of product output. Just

as the rate of diffusion of molecules across a boundary layer is proportional to the concentration gradient, the rate of flux through a chemical reaction is proportional to the energy gradient across the pathway. Thus, with increasing efficiency of energy extraction, there is reduced flux through the overall system, the result being that the total rate of energy harvesting (the product of the flux rate of the carbon source and the efficiency of conversion to ATP per molecule consumed) is maximized as some intermediate level of efficiency. In the case of glycolysis, the flux rate is maximized when ~ 2 ATPs are produced per glucose consumed (Foundations 19.3). Whereas, in principle, four ATPs could be produced from the energy of one glucose molecule, this would reduce the total rate of ATP production by 90%. Thus, energy dissipation in terms of heat loss seems to be an intrinsic requirement if natural selection is to maximize the rate of flux as opposed to the efficiency of resource utilization.

Note that this kind of pathway outcome represents a kind of digital trait in cell biology. One molecule of glucose might yield 1, 2, 3, or 4 ATPs, but not 1.5 or 2.5, even if such an output might be mathematically optimal. Glycolytic conversion of glucose to lactic acid yields ~ 205 kJoules/mol glucose consumed, whereas each conversion of ADP to ATP stores 50 kJoules/mol. Fortuitously perhaps, the optimal (but not possible) rate of ATP production of 2.05 per glucose molecule (Foundations 19.3) is very close to the realized value of 2, and $\sim 25\%$ better (in terms of the total ATP production rate) than would be the case if 1 or 3 ATPs were produced. In principle, however, depending on the overall energetics of a pathway, the optimum strategy of energy harvesting could be further from the digital possibilities (e.g., 2.50 in the previous example, rather than 2.05). Such deviations, which are likely larger for some substrate types, represent a real limitation of biological systems.

Although concepts of evolutionary tradeoffs or constraints pervade the evolutionary ecology literature, for the most part such invocations are at best based on intuition (which need not be correct) or at worst entirely *ad hoc.* Justified or not, such arguments necessarily follows whenever one assumes that natural selection has unlimited power, save for biophysical constraints. Here, however, we are confronted with real compromises derived from first principles. These kinds of speed vs. efficiency tradeoffs exist in a number of other contexts in cell biology. For example, microbes that exhibit high yields per carbon molecule consumed also have lower levels of heat dissipation (von Stockar and Liu 1999). Wagoner and Dill (2019) have considered this kind of trade-off for various molecular machines, finding that cases such as ATP synthase, optimization has placed a premium on efficiency, perhaps not surprisingly given that the main mission of this machine is energy (ATP) production. (In contrast, glycolysis produces intermediate metabolites that serve as precursors for various biosynthetic pathways).

Pathway Expansion and Contraction

One of the most striking aspects of metabolism, at least to a nonbiochemist, is the large number of incremental pathway steps often used to accomplish what superficially seem to be relatively simple molecular alterations. Many enzymes produce intermediate metabolites whose sole role is to be passed on to another type of enzyme for further processing. Consider, for example, glycolysis, which requires

ten separate enzymes to convert one six-carbon/six-oxygen glucose molecule to two three-carbon/three-oxygen pyruvates (Figure 19.3).

In the case of glycolysis, a few of the intermediates do serve as precursors for other pathways for building-block production, but pyruvate is a key entry point for cellular energy production and the ten steps to get there might seem excessive. However, Bar-Even et al. (2012) argue that intrinsic constraints of biochemistry (thermodynamic limitations, limited availability of enzymatic mechanisms, and physicochemical properties of pathway intermediates) make any alternative routes from glucose to pyruvate implausible. In other words, their view, which was extended to all of central metabolism by Noor et al. (2010), is that given the evolved structure of biochemistry and the intermediate metabolites that must be relied upon, there is no shorter pathway between glucose and pyruvate.

Of course, this begs the question as to why biochemistry evolved to have such a structure. As will be noted below, there are cases in which individual proteins are capable of carrying out multiple reactions. Moreover, as it is highly unlikely that the extended catabolic and anabolic pathways known in today's organisms were present at the moment of life's origin, deeper questions concern how complex metabolic pathways become assembled evolutionarily. As an entrée to this particular problem, we will start with a consideration of the theoretical issues, and then make a transition to empirical data bearing on the theory.

Stochastic meandering of pathway architecture. Consider the situation outlined in Figure 19.4, where a focal resource, S, is potentially obtainable either from the environment or via input from some other cellular activity that produces a precursor metabolite P, which in the presence of an appropriate enzyme can be converted into S. If P is present in sufficient quantity, the shuttling of a fraction of an appropriately converted product (S) to a downstream pathway might be an advantageous strategy. The addition of an enzymatic mechanism for producing and feeding P into a pre-existing pathway would be viewed as a forward or backward extension depending on one's perspective. Either way, if the initial ability to obtain S directly from the environment is lost, the pathway will have increased in length.

Finally, as discussed in other contexts in previous chapters, provided there are nonzero probabilities of forward and reverse evolutionary transitions between all adjacent pathway states, for a certain set of conditions there is a steady-state probability that a population will reside in any one of the alternative states if it were to be sampled at enough intervals over evolutionary time (Foundations 19.4). Borenstein et al. (2008) provide evidence for such evolution, showing that the turnover rate of usage of external precursor compounds (upstream metabolites that cannot be synthesized and are exogenously acquired from the environment) is substantial among microbes, with phylogenetic shifts likely driven by environmental availabilities of alternative substrates.

To determine the population-genetic and ecological conditions that are most conducive to the evolution of the three alternative pathways in Figure 19.4, two basic issues need to be considered: 1) the relative selective advantages of the alternative pathways; and 2) the relative rates of origin and loss of pathway links via mutational mechanisms. As outlined in Chapter 5, the transition rate between two adjacent states is equal to the product of the numbers of relevant mutations arising

per generation in the population and their probability of fixation. The long-term probabilities of occupancy of the alternative states are then functions of the relative biases of mutation and selection. If the bias in mutation pressure in one particular direction is sufficiently strong relative to any opposing selection pressures, the state of the population will be driven primarily by mutation, with the most frequently occupied state not necessarily being that which yields the highest fitness.

Three factors can contribute to the differential fitness of the pathways illustrated in Figure 19.4. First, the direct effects of resource S on fitness must be considered. If R_S is the amount of resource available for direct uptake, and R_P is the additional amount that can be added by the novel enzymatic step, then $(R_S + R_P)$ is the total amount of resource available to a mutant possessing both input mechanisms. Assuming that fitness, $W(R)$, is an increasing function of resource availability, approaching an asymptotic maximum value at high R (Figure 19.5), then network 2 has the highest direct payoff, whereas that of pathway 3 can be greater or smaller than that of pathway 1 depending on the relative availability and convertibility of precursor P. Because natural selection operates on relative fitnesses, it is sufficient to denote the direct selective advantages of pathways 1, 2, and 3 as deviations from the maximum (state 2) value: $-r_1$, 0, and $-r_3$. The magnitudes of the two negative coefficients will depend on the resource availability via the nonredundant pathways, asymptotically approaching zero as resources approach saturating levels, i.e., as R_S, $R_P \to \infty$, and $W(R)$ plateaus.

Second, each pathway will incur a baseline cost in terms of the required energetic investments for the production and maintenance of the enzymatic machinery and/or acquisition of any additional cofactors needed for uptake/production of S. These costs, which can be denoted $-c_1$, $-(c_1 + c_3)$, and $-c_3$, will necessarily be greater for pathway 2.

Third, each pathway can be silenced by degenerative mutations. Let u_1 be the rate of mutational loss of the enzymatic connection with the environmental source of S, and u_3 be the rate of mutational loss of the connection to P. Assuming the final product to be essential, for pathways 1 and 3, respectively, a fraction of u_1 and u_3 newborn individuals are eliminated by selection, as they imply the loss of access to S. On the other hand, single deactivating mutations incurred by the redundant state (2) do not lead to the inviability of individuals, but simple transformation to either state 1 or 3.

Taking all three factors into consideration, the selective disadvantages of the three pathways are $s_1 = r_1 + c_1 + u_1$, $s_2 = c_1 + c_3$, and $s_3 = r_3 + c_3 + u_3$. The key point is that the relative fitnesses of alternative mechanisms of resource acquisition are functions of: 1) the amounts of resource directly (S) and indirectly (P) available; 2) the bioenergetic costs of building and maintaining the two alternative mechanisms; and 3) the mutation rates to defective phenotypes. The redundant pathway (2) is advantageous with respect to resource acquisition and mutational vulnerability, but more costly with respect to maintenance. Not included in this expression is one other possibility – the precursor molecule P might have a pre-existing function in the cell, in which case allocation of a fraction of P towards the production of S could impose a cost via the primary function of P, which would detract from the fitnesses of states 2 and 3.

There are two additional points to consider. First, as noted in Chapters 4 and

5, a key issue is the ratio of the magnitude of the strength of selection to the power of drift, $N_e s_i$, where N_e is the effective population size (Foundations 19.4). This term influences the probability distribution of alternative states exponentially, as $e^{N_e s_i}$, which is the ratio of fixation probabilities for mutations entering and leaving class i. When $N_e s_i \ll 1/$, $e^{N_e s_i} \simeq 1.0$, showing that selection is ineffective unless the effective population size is sufficiently large.

Second, the steady-state probabilities of alternative pathways also depend on the ratios of rates of gain and loss of the end states by mutation. The essential result is that relative to the intermediate (redundant) state 2, the probability of each of the end states is simply defined by the net pressure of mutation and selection producing versus exiting these states (Foundations 19.4). In the extreme case of effective neutrality (all $N_e s_i < 1$), the probabilities of alternative states depend only on the degree of mutation bias.

Based on these considerations, several conclusions can be drawn (Figure 19.6). First, because the rate of mutational production of a new mechanism of resource acquisition is likely to be substantially smaller than the mutational rate of inactivation, there is a tendency for mutation to drive the redundant pathway to the two extreme states. This means that unless the selective advantage of the redundant state is much larger than the power of drift, and also large enough to offset mutational pressure, populations will most likely reside in one of the end states. Second, the nonredundant pathway that is most likely to emerge (uptake-dependent vs. precursor-dependent) depends on the relative magnitudes of the ratios of loss:gain rates of the two types.

Finally, it is important to note that all of the preceding arguments assume a nonrecombining genome, such that offspring have genotypes identical to those of their parents in the absence of mutation. Suppose, however, that recombination can occur between the genes underlying the two alternative pathways. Of the three types of encounters between alternative genotypes, 1-3, 1-2, and 2-3, only the 1-3 combination results in novel progeny genotypes. This is because a redundantly encoded genome recombined with a single-feature genome still results in progeny with a 1:1 ratio of parental features, whereas a 1-3 recombination event results in one genome of type 2 and one completely lacking in function. This potential for recombinational breakdown between single-feature genotypes has the potential to promote the maintenance of genotypes with redundant acquisition pathways. However, such an effect can be partially or entirely offset if the redundant pathway has sufficiently low fitness from a metabolic perspective, or if the population is small enough that the essential ingredient for its operation (the simultaneous presence of state 1 and 3 genotypes) is absent (Lynch 2007).

Although this model provides a glimpse into the understanding of how simple single-step changes might lead to among-species differentiation in metabolic pathways, many metabolic pathways go far beyond the simple systems just explored. Consider, for example, the mysteries posed by the amino-acid biosynthetic pathways (Figure 19.7). First, there is the matter of the order of emergence of the pathways leading to each of the twenty amino acids. Only three of these (glutamate, aspartate, and alanine) are a single step removed from a metabolic precursor, with several others being derived from these primary products via one or more additional steps. For example, asparagine is one step removed from aspartate, whereas

lysine, methionine, and threonine are several steps removed, and isoleucine is still more steps removed from threonine. Is the distance of each amino acid from central metabolic pathways an indicator of its age, i.e., are the more derived amino acids evolutionarily younger?

Second, many of the intermediate precursor products leading to the production of downstream amino acids simply serve as pathway stepping stones with no other function. If this is the case, how could the steps leading to such products be established evolutionarily prior to completion of the pathway, if only the end product is of utility? As a potential solution to this dilemma, Horowitz (1945) proposed a retrograde model for the evolution of biosynthetic pathways, postulating that upstream steps to a pathway are added as the environmental availability of downstream metabolites become limited. Under this hypothesis, the final step in a pathway is the first to have been acquired, with the preceding step being established second, and so on. Such a model assumes that molecular intermediates were at one time freely available in the environment (as in the progression in Figure 19.4), but is challenged by the fact that intermediate metabolites often exhibit high levels of chemical instability.

The origin of novel enzymes. Although the theory outlined above clarifies the population-genetic requirements for shortening or lengthening metabolic pathways, a deeper appreciation of the actual biological underpinnings of such transitions is necessary to achieve a mechanistic understanding of how pathways expand. In the preceding example, a transition from state 1 to 2 requires the emergence of an enzymatic function for converting precursor molecule P to the final resource S, whereas a transition from state 3 to 2 requires the appearance of a mechanism for direct acquisition of S. As it is quite unlikely that new molecular mechanisms arise *de novo* in highly refined states, more gradual mechanisms must be sought.

As already noted, all enzymes make errors (i.e., bind to inappropriate substrates) at some low frequency. A fortuitous consequence of such promiscuity is the predisposition of many enzymes to respond evolutionarily in the face of sufficient selection pressure to utilize alternative substrates. On the other hand, because the refinement of a previously deleterious side effect may come at the expense of the pre-existing primary function, additional genomic changes will typically be essential to allow the emergence of an efficient novel function. Gene duplication provides a powerful route for evolutionarily remodeling an enzyme without relinquishing key ancestral functions, by allowing each copy to become more specialized to individual tasks (Copley 2012).

As discussed in Chapter 6, for the simplest case in which a gene is completely duplicated, it will go through an initial phase of functional redundancy, and unless maintained by selection for increased dosage, will generally suffer one of three fates: loss via a nonfunctionalizing mutation; preservation by subfunctionalization (with the two sister copies becoming specialized to independently mutable subfunctions); or preservation by neofunctionalization (evolution of an entirely novel function by one copy, and retention of the original key functions by the other copy).

As noted above, at least in the case of metabolic-function evolution, there may be very few cases of pure neofunctionalization, given the promiscuity of individual enzymes. This would mean that almost all cases of long-term preservation of du-

plicate metabolic genes are likely driven by either dosage demands or selection for refinements and losses of pre-existing functions (i.e, enzyme diversification). The latter mechanism was, in fact, suggested by Kacser and Beeby (1984; see also Pfeiffer et al. 2005) as a potential explanation for how the earliest enzymes (presumably with much lower substrate fidelity than today's enzymes) might have diversified in function. When a gene is amplified in copy number, this provides additional mutational targets (free of compromising constraints) for the refinement of at least one copy to alternative, more specialized subfunctions. Quite similar arguments for the origin of new functions were subsequently promoted by Piatigorsky and Wistow (1991), Hughes (1994), and Bergthorsson et al. (2007).

There are some subtle distinctions between alternative verbal models that have suggested in this area. For example, Hughes' (1994) adaptive-conflict hypothesis envisions situations in which an ancestral gene carries out two or more essential but compromised functions that then become specialized after duplication, whereas Bergthorsson et al. (2007) focuses more on situations in which a specialized ancestral gene acquires a novel mutation inducing weak functional promiscuity, which then undergoes adaptive refinement following a duplication event that increases its exposure to natural selection by amplifying its expression level. Both hypotheses contain an element of neofunctionalization, and both involve a sort of subfunctionalization, the primary difference being the degree of essentiality of the original duplicate functions.

Voordeckers et al. (2012) obtained strong evidence for just such an innovation-amplification-divergence scenario in yeast where an ancestral glucosidase protein, active primarily with maltose, had weak promiscuous attraction to isomaltose. Upon duplication, one copy lost maltose activity while becoming optimized for isomaltose in a process that involved just three key amino-acid changes. Likewise, Blount et al. (2012) found that the evolution of the ability to utilize citrate in an experimental population of *E. coli* was initiated by a regulatory change that increased the expression of a citrate transporter, followed by amplification to a multi-copy array of the modified locus. In a comparative study that involved the resurrection of ancestral states, Rauwerdink et al. (2016) found that plant cyanohydrin lyases, which produce insecticidal cyanides, evolved from an initially weak promiscuous side effect of plant esterases, only becoming refined after duplication. Significantly, all of these studies illustrate dramatic shifts in metabolic functions being driven by just a few mutational changes.

Pathway Participant Remodeling

Although most of the pathways associated with central metabolism are typically presented in biochemistry textbooks as invariant, formal analyses have generally been carried out in no more than a handful of taxa. Yet, existing comparative analyses amply demonstrate that the structures and constituents of metabolic pathways sometimes vary dramatically among species. As an entrée into this area, we now consider how a cellular process even as fundamental as amino-acid biosynthesis can be subject to substantial modifications among phylogenetic lineages.

Two tyrosine-biosynthesis pathways are known, both starting with prephenate

as a precursor, but having different intermediates (Jensen and Pierson 1975; Song et al. 2005). Both pathways involve transamination and dehydrogenation steps, but in reverse order (Figure 19.8, top). Among bacteria, many different lineages use one or the other pathway, whereas species that are capable of both may employ a single broad-specificity enzyme. The cofactors deployed in the key reactions, e.g., NAD^+ vs. $NADP^+$, are also subject to phylogenetic changes.

Cysteine is produced by at least three different pathways in various organisms, two of which involve completely nonoverlapping reactions. In animals, the sulfhydryl group of cysteine is derived from methionine, whereas in plants and prokaryotes that have been studied, inorganic sulfide provides the source (Figure 19.8, middle). In the budding yeast *S. cerevisiae*, there is a modification of the animal pathway, such that the methionine → homocysteine step is replaced by two others. However, there is a good deal of variation among fungi, with the fission yeast *S. pombe* having the pathway initiating with serine, and many species having both types of pathways (Hébert et al. 2011).

Lysine biosynthesis provides still another example of pathway diversification. Two major modes of production exist for this amino acid: the diaminopimelate (DAP) pathway and the α-aminoadipate (AAA) pathway, the first starting from aspartate and the second from α-ketoglutarate, and with no overlap in the intermediate products. Both pathways are quite long, typically with eight to nine steps, although short-length variants that bypass some internal steps are known (Nishida et al. 1999; Hudson et al. 2006; Curien et al. 2008; Torruella et al. 2009). The DAP pathway (and variants of it) is present in bacteria, plants, and some algae, and the AAA pathway has been found in fungi, a few bacteria, and some protists. A few taxa may have both pathways, and lateral gene transfer has been suggested (Nishida and Nishiyama 2012).

Perhaps the most spectacular example of diversification of amino-acid biosynthetic pathways involves methionine production (Figure 19.8, bottom). In all cases, the starting point is homoserine, but three alternative first-step metabolites can be produced (acetylated, phosphorylated, or succinylated homoserine). From these points, the addition of cysteine can lead to cystathione (a sulfur source) in a single step, which is then converted to homocysteine in the penultimate step to methionine. However, single-step pathways from both acetyl-homoserine and succinyl-homoserine to homocysteine also exist, utilizing H_2S and bypassing cystathione production. Finally, there are two alternative final-step mechanisms for converting homocysteine to methionine. Taken together, these observations suggest the possibility of as many as ten partially overlapping pathways. At least eight of these have been observed (Gophna et al. 2005), and some microbial species are capable of multiple production pathways. As some archaea appear to be completely lacking proteins for homoserine activation, additional pathways likely await discovery. Moreover, evolutionary shifts between alternative methionine biosynthetic pathways do not appear to be difficult. For example, a single amino-acid change is sufficient to alter the first step of homoserine activation. As a consequence, these first-step enzymes are not monophyletic with respect to function (Zubieta et al. 2008).

These four examples are by no means an exclusive list of variations in amino-acid biosynthesis pathways. For example, although glutamate is one of the simpler amino acids, being just one step (the addition of an NH_3) removed from α-ketoglutarate,

the source of NH_3 in the reaction can vary, and the enzyme involved (glutamate synthase) utilizes different cofactors in different phylogenetic contexts (Dincturk and Knaff 2000; Suzuki and Knaff 2005). Just within the bacterial genus *Pseudomonas,* two nonoverlapping pathways are known for phenylalanine production, with some species having one of the pathways and others having both (Byng et al. 1983). The first step in the biosynthesis of arginine is the acetylation of glutamine, but this is done by apparently unrelated enzymes in various phylogenetic lineages (Xu et al. 2007). The branchpoint intermediate in the path leading to methionine vs. threonine production is O-phosphohomoserine in plants, whereas it is one step higher in *E. coli* and *S. cerevisiae*, at homoserine (Curien et al. 2008). There are at least two largely nonoverlapping pathways for serine production (Melcher et al. 1995; Ho et al. 1998; Shimizu et al. 2008), three for isoleucine (Hochuli et al. 1999), and two for glycine (Sakuraba et al. 2005; Ramazzina et al. 2010). Finally, the tryptophan operon, which is often used as a textbook example of gene regulation, shows extensive variation in modes of regulation among bacterial species, with some having fusions of the various genes in the pathway (Xie et al. 2003; Merino et al. 2008).

This broad set of examples, mostly drawn from prokaryotes but by no means exclusive to amino-acid biosynthesis (e.g., Hoshino and Gaucher 2018), highlights the flexibility of cells to evolutionarily modify the underlying machinery associated with key metabolic processes, presumably while keeping the overall performance of the pathway nearly constant. Several additional points are clear. First, even the pathways that one might expect to be highly conserved are subject to complete remodeling to the point of sharing no reaction steps. Second, even when species share the same pathway structure, the participating enzymes can be nonorthologous. Third, redundant pathways can often coexist in the same organism. Finally, alternative pathways often differ in efficiency (Du et al. 2018), e.g., in utilization of ATP, and it remains unclear as to why energetically suboptimal variants exist.

These general observations, combined with the common existence of enzyme promiscuity and gene duplication, lead to a plausible set of scenarios for metabolic pathway evolution involving intermediate stages of functional redundancy, fully consistent with the model outlined above (Figures 19.4-6). Many variants of such a model can be envisioned, all of which fit well within the general theme of the patchwork model of pathway evolution (Ycas 1974; Jensen 1976; Copley 2000), under which enzymes with substrate ambiguity can sometimes be stitched together to establish novel pathways and/or to confer redundancy with preexisting molecular mechanisms. We now consider the underlying issues in somewhat greater detail, again motivating each discussion with a series of examples, primarily to emphasize just how common pathway remodeling is.

Nonorthologous gene replacement. As alluded to in the previous section, in addition to the rewiring of pathway structure, many examples are known in which metabolic pathway structures remain stable, while the underlying participants turn over. As a case in point, consider glycolysis, the major path to sugar breakdown in the cells of eukaryotes and many prokaryotes (Figure 19.3). Involving a series of ten reactions, this pathway converts six-carbon glucose molecules into three-carbon pyruvate molecules, which can then be further processed (e.g., by entering the citric

acid cycle, or by being reduced to lactic acid).

Despite the centrality of glycolysis, extraordinary variation exists in the nature of all of the major players, with comparative analyses suggesting a very complex evolutionary history both within and among the major domains of life (Canback et al. 2002; Oslancová and Janecek 2004; Siebers and Schönheit 2005). For example, bacterial phosphofructose kinase (step 2) and phosphoglycerate kinase (step 7) appear to be nonorthologous to the homologous enzymes used in eukaryotes (Galperin et al. 1998), and most of the pathway enzymes in archaea appear to be nonhomologous to those in either eukaryotes or bacteria (Verhees et al. 2001; Kawai et al. 2005; Siebers and Schönheit 2005). Two types of fructose bisphosphate aldolase (step 4) are known, each deploying a different molecular mechanism, one group in animals and land plants, and the other in fungi, some algae, and bacteria. Various protist lineages harbor a diversity of sequences from both classes, with the complexity of the phylogenetic distribution strongly suggesting substantial ancient horizontal transfer (Siebers et al. 2001; Sánchez et al. 2002; Rogers and Keeling 2004; Allen et al. 2012). Clear examples of inter-kingdom horizontal transfers also exist for glyceraldehyde-3-phosphate dehydrogenase (step 6) (Qian and Keeling 2001; Takishita and Inagaki 2009). Finally, phosphoglycerate mutase (PGM, step 8) is found in two forms, apparently unrelated to each other and also differing in cofactor dependency (Liapounova et al. 2006). Several members of the archaea and bacteria harbor both types of PGM, consistent with the idea of the ancient origin of the two types (van der Oost et al. 2002; Johnsen and Schönheit 2007). This list could be expanded, but by now the central point should be clear – as with many of the amino-acid biosynthetic pathways, across the Tree of Life probably all of the enzymes involved in glycolysis have evolved more than once from apparently unrelated ancestral enzymes and then been substituted for their nonorthologous analogs.

Multiple mechanisms exist for such nonorthologous gene replacement. As noted above, phylogenetic analyses clearly implicate horizontal gene transfer in a number of instances, although most of the well-established examples are restricted to quite ancient periods in the Tree of Life. Although inter-kingdom transfers must be confronted with a number of selective challenges, these may be eased for enzymes universally conserved to have the same function. Indeed, occasions might arise in which the enzyme from a potential donor species has acquired slightly beneficial kinetic, assembly, and/or stability properties relative to that of the recipient, e.g., because of a population-genetic environment more conducive to efficient selection.

One situation that can encourage nonorthologous gene replacement (including that via horizontal transfer) arises during periods in which a resource that must normally be synthesized becomes freely available in the environment (e.g., Figure 19.4). Such a condition might relax selection on the maintenance of the internal catabolic pathway, thereby leading to the effectively neutral loss of one or more components via degenerative mutations. The resultant condition, known as auxotrophy, which is an inability to synthesize an essential organic compound. Consistent with such a scenario, it is common for some isolates within bacterial species to be missing parts of particular pathways and hence deficient with respect to certain kinds of metabolism (Monk et al. 2014). This phenomenon that has been repeatedly seen in endosymbiotic bacteria that inhabit the well-provisioned cells of insects (McCutcheon and Moran 2007), an extreme example being chimeric biosynthetic pathways in which

different enzyme components reside in different genomes of the mutualists (Bublitz et al. 2019). The enzymes most likely to be lost involve steps (usually downstream) with minimal interactions with other pathways (Hittinger et al. 2004; Pál et al. 2006). Reintroduction to an environment lacking the resource would then impose strong selection for pathway reconstruction, e.g., replacement of a missing step with a duplication from a nonorthologous gene with a suitably promiscuous function.

As noted above, further opportunities for pathway diversification exist when a broad-specificity enzyme is duplicated, particularly if different subfunctions are independently mutable, as this provides a route by which a gene from one pathway might move to another. By this means, through progressive gains and losses of subfunctions, genes might be free to wander across pathways, provided this is accomplished in ways that do not erode the operation of individual pathways. For example, *Mycobacterium tuberculosis* deploys a bifunctional enzyme in both histidine and tryptophan biosynthesis (Due et al. 2011), and experimental work demonstrates that simple amino-acid substitutions can establish tryptophan-pathway activity from histidine-pathway genes (Leopoldseder et al. 2004; Näsvall et al. 2012). In a number of species that have experienced genome reductions, one or the other of the amino-acid biosynthetic paths has been lost, and the bifunctional enzyme has reverted to monofunctionality (Juárez-Vázquez et al. 2017).

Notably, indirect evidence suggests that entire pathways may sometimes have arisen by duplication of most or all constituent proteins. For example, Velasco et al. (2002) suggest that the DAP pathway for lysine production is related to the arginine biosynthesis pathway, in that many of the constituent enzymes appear to be related. A link between the AAA pathway for lysine and the leucine biosynthesis pathway has also been suggested, leading to the idea that pathways for the production of all three amino acids – leucine, arginine, and lysine – share a common ancestry, possibly derived from a state in which all three were synthesized by enzymes with lower specificity (Nishida et al. 1999; Miyazaki et al. 2001; Fondi et al. 2007). Remarkable similarities between key steps in the TCA cycle and the paths for lysine, isoleucine, and leucine biosynthesis have also been suggested (Jensen 1976). Once a redundant set of pathways arises, reciprocal loss by subfunctionalization can occur in descendant sister taxa, leaving little or no record of past redundancy. For example, Jensen (1985) highlights hydrogenases in bacteria whose ancestral state involved joint NAD^+ and $NADP^+$ cofactor specificity but became fixed for one or the other alternatives in descendant sister taxa.

Returning to the situation with the glycolytic pathway, the potential to shift pathway affiliations seems particularly large, given the widespread moonlighting of such enzymes in functions completely unrelated to glycolysis (Kim and Dang 2005). Just a few examples are noted here. G3PDH (step 6) is known to bind DNA and telomeres, plays a role in cell-cycle regulation, and has many other functions associated with nuclear export, membrane fusion, phosphorylation, and DNA repair (Seidler 2013). Many examples are known in which FBP (step 4) and triosephosphate isomerase (TPI, step 5) are involved in membrane fusion in pathogenic bacteria (Gozalbo et al. 1998; Modun and Williams 1999; Alvarez et al. 2003; Tunio et al. 2010). One of the two hexokinase (HEX, step 1) genes in *S. cerevisiae* also acts as a repressor of glucose metabolism; and HEX has been implicated in anti-apoptotic activity in mammals. Phosphopyruvate hydratase (PPH, also known as enolase, step

9) has also been found to have numerous secondary functions – it is a major lens protein in some vertebrates (Wistow et al. 1988), plays a role in thermal tolerance in yeast (Iida and Yahara 1985), and can be involved in cell-cell communication (Miles et al. 1991; Castaldo et al. 2009) and transcriptional regulation (Chang et al. 2003; Kim and Dang 2005).

These kinds of observations extend to each of the remaining enzymes involved in glycolysis, as well as to many other enzymes involved in central metabolism. Often ignored, the types of gene loss and substitution implied by nonorthologous gene replacement have significant implications for the use of biological observations on the genes in model organisms to infer gene functions in more distantly related species (Gabaldón and Koonin 2013).

Internal pathway expansion via multifunctional enzymes. A fundamental feature of biochemical pathways is that consecutive enzymes must interact with one of the same ligands – the product of an upstream enzymatic reaction becomes the substrate for the next. Some enzymes exploit this linkage by engaging in two adjacent interactions in a pathway. A good example of such a bifunctional enzyme is fructose-1,6,bisphosphate aldolase/phosphatase, which in most archaea and many bacteria is used in gluconeogenesis (the reverse of the glycolytic pathway) to first convert dihydroxyacetone phosphate (DHAP) to FBP and then to dephosphorylate the latter to fructose-6-phosphate (Du et al. 2011). Unlike many bifunctional enzymes, both reactions are carried out by the same catalytic domain, leading to an efficient system in which the intermediate metabolite (FBP) is never released.

Situations like this open the opportunity for pathway growth when an ancestral protein capable of carrying out two steps in a pathway becomes duplicated into two daughter copies, each of which loses an alternative function, locking the two into permanent preservation by subfunctionalization. A possible example of such a situation exists in *E. coli*, where two consecutive enzymes in the methionine biosynthetic pathway appear to have been derived from a common gene by duplication that then underwent subsequent specialization (Belfaiza et al. 1986; Parsot et al. 1987). Here, what is done in two steps in *E. coli* is accomplished by one related enzyme in *S. cerevisiae*. It has also been argued that a single enzyme converting O-acetyl-serine to cysteine is related to the two taking O-acetyl-homoserine to homocysteine in the methionine pathway (Figure 19.8; Parsot et al. 1987). Consecutive genes in the threonine biosynthetic pathway also appear to have arisen by gene duplication in bacteria (Parsot 1986; Parsot et al. 1987), and similar structures for several of the sequential proteins in tryptophan biosynthesis suggest a common ancestry (Wilmanns et al. 1991; List et al. 2011). Arguments have also been made that a number of the enzymes in the purine biosynthesis pathway have arisen by duplication and divergence (Kanai and Toh 1999; Zhang et al. 2008).

A few enzymes are known to be capable of catalyzing more than two reactions (Roy 1999), e.g., dehydroquinate synthase (Carpenter et al. 1998), and carbamoyl phosphate synthetase (Raushel et al. 1998), either in the same site or with active sites connected by molecular tunnels. In principle, such enzymes provide rare opportunities for an entire pathway to be carried out by a single protein, as well as for repeated processes of duplication and subfunctionalization to lead to pathway extension. Of course, there seems to be no reason to rule out the reverse – pathway

retraction, as a monofunctional enzyme becomes bifunctional by taking on the task of an adjacent step.

Summary

- Virtually all metabolic reactions are carried out by enzymes. Although these proteins enhance reaction rates by many orders of magnitude above background rates, their catalytic rates are generally orders of magnitude below the biophysical limits, suggesting that the degree of perfection is stalled at a drift barrier.

- Generally highly specialized for particular functions, most enzymes are also quite promiscuous, commonly engaging with inappropriate substrate molecules up to 1% of the time. Although this must impose a burden on cells, in the long-run it also provides a launching pad for the evolution of novel enzyme functions through the refinement of latent capacities.

- Dozens of experimental-evolution studies have demonstrated that, combined with gene duplication, substrate-utilization promiscuity facilitates the origins of new metabolic features, commonly via just a few key amino-acid substitutions.

- Because individual enzymes generally carry out simple chemical reactions, typically involving only single chemical bonds, this necessitates the use of complex, stepwise metabolic pathways in the transformation of organic material.

- Theory suggests that the sensitivity of flux rates of metabolic pathways declines exponentially from the entry-level to final steps in the pathway, and this appears to be reflected in a relaxation of selection on the amino-acid sequences of downstream enzymes.

- Catabolic pathways are commonly highly inefficient energetically in the sense that a substantial amount of energy derived from the breakdown of substrate molecules is simply released as heat rather than converted to bioenergetic currency (ATP). Flux-rate theory suggests that such inefficiency is an intrinsic requirement for maximizing the speed of product formation. However, the digital nature of molecular transactions is an intrinsic constraint that necessitates energy loss.

- Despite the near universality of many resource-utilization and biosynthetic pathways, both the structure and underlying enzyme participants of such pathways can differ dramatically among phylogenetic lineages. This suggests the existence of multiple degrees of freedom for pathway remodeling without significant neg-

ative fitness consequences. Gene duplication, combined with subdivision and divergent refinement of ancestral gene functions, plays a central role in such reconfigurations.

Foundations 19.1. Michaelis-Menten enzyme kinetics. The simplest approach to enzyme kinetics starts with the assumption that molecules of enzyme (E) and substrate (S) interact in two steps to yield a product (P). The first step involves the production of a noncovalent intermediate complex (ES) and is assumed to be reversible, governed by association-dissociation kinetics (Foundations 13.1). The second step involves transformation of the substrate into the final product, after which the released enzyme is free to engage in another reaction. The overall reaction can be described as

$$\mathrm{E} + \mathrm{S} \rightleftharpoons \mathrm{ES} \rightarrow \mathrm{E} + \mathrm{P}$$

The intent here is to describe the relationship between the net conversion rate of S to P and the concentration of S.The rate of product formation is a function of the three rates associated with the arrows in the preceding formulation. Let k_a be the rate of association of E and S to form ES, k_d be the rate of dissociation of ES into E and S, and k_cat be the catalytic rate of production of P from ES (also known as the turnover rate). Under the assumption of a steady-state process (e.g., with the concentrations of all components remaining constant, owing to the constant replenishment of substrate and recycling of enzyme), the rates of gain and loss of the intermediate state must be equal, so that

$$k_\mathrm{a}[\mathrm{E}][\mathrm{S}] = (k_\mathrm{d} + k_\mathrm{cat})[\mathrm{ES}], \tag{19.1.1a}$$

where brackets denote concentrations. Note that the rate of production of ES is a function of the concentrations of the two constituents and the rate of association per unit concentration. Rearranging, we find that

$$[\mathrm{ES}] = \frac{[\mathrm{E}][\mathrm{S}]}{K_\mathrm{S}}, \tag{19.1.1b}$$

where the composite parameter is defined as

$$K_\mathrm{S} = (k_\mathrm{d} + k_\mathrm{cat})/k_\mathrm{a}. \tag{19.1.1c}$$

Denoting the rate of production of P as $V = k_\mathrm{cat}[\mathrm{ES}]$, substitution of Equation 19.1.1b for [ES] implies

$$V = \frac{k_\mathrm{cat}[\mathrm{E}][\mathrm{S}]}{K_\mathrm{S}} \tag{19.1.2}$$

There is one final complication in that [E] is the concentration of free enzyme, whereas the more easily measured total amount of enzyme in the system, $[\mathrm{E_T}]$, is the sum of the amounts freely circulating and tied up in the enzyme-substrate complex. Again using Equation 19.1.1b,

$$[\mathrm{E_T}] = [\mathrm{E}] + [\mathrm{ES}] = [\mathrm{E}]\left(1 + \frac{[\mathrm{S}]}{K_\mathrm{S}}\right). \tag{19.1.3}$$

Solving this equation for [E], and substituting into Equation 19.1.2 yields an equation of the same form presented by Michaelis and Menten (1913),

$$V = \frac{k_\mathrm{cat}[\mathrm{E_T}][\mathrm{S}]}{K_\mathrm{S} + [\mathrm{S}]}. \tag{19.1.4}$$

Notably, the derivation utilized above was first presented by Briggs and Haldane (1925), which although yielding the same mathematical structure, involves some subtle

distinctions from the approach used by Michaelis and Menten (1913). For interesting historical commentary on this and other matters, see Gunawardena (2012), as well as Johnson and Goody (2011), who also generated an English translation of Michaelis and Menten (1913). Formulae with the same form appear in many other areas of cell biology (Wong et al. 2018).

There are several notable features of this expression. First, for a constant level of total enzyme, Equation 19.1.4 predicts a hyperbolic relationship between the substrate concentration, [S], and the rate of product formation (Figure 19.1). At high substrate concentrations, V asymptotically approaches a maximum level $V_{\max} = k_{\text{cat}}[\text{E}_\text{T}]$, which shows that k_{cat} is the maximum rate of product formation per unit of total enzyme concentration.

Second, K_S is equivalent to the substrate concentration at which the reaction rate is half the maximum, as can be seen by substituting K_S for [S] in Equation 19.1.4. Thus, K_S is usually referred to as the half-saturation constant, although as can be seen from Equation 19.1.1c, K_S is not mathematically independent of k_{cat}.

Third, if the catalytic rate k_{cat} is much smaller than the dissociation rate k_d, then the former can be ignored in Equation 19.1.1c, yielding $K_\text{S} \simeq k_\text{d}/k_\text{a}$. This ratio of dissociation to association rates is usually referred to as the dissociation constant and often denoted by K_D.

Finally, we note that as $[\text{S}] \to 0$, $V \to k_{\text{cat}}[\text{E}_\text{T}][\text{S}]/K_\text{S}$. The ratio $\phi_\text{E} = k_{\text{cat}}/K_\text{S}$ is often referred to as the kinetic efficiency, as it defines the innate capacity of an enzyme when the substrate is at nonsaturating levels.

Two additional points are worth noting about the Michaelis-Menten equation. First, inspired by their own kinetic observations and those made earlier by others, Michaelis and Menten employed a hypothetical intermediate enzyme-substrate complex in their formulation as a means for achieving the hyperbolic relationship between enzyme reaction rates and substrate concentration, several decades passed before any such complexes were actually observed. Thus, this early exercise in biochemical modeling is a beautiful example of a theoretical construct predicting a previously unseen phenomenon. Second, the enzyme pathway deployed assumes that product formation is irreversible, unlike the ES complex. Although this is a reasonable approximation if P remains rare (as would be the case if P were consumed in some other cellular process), as P becomes increasingly common, negative feedback will cause a reduction in the net forward reaction rate, necessitating further modification of the Michaelis-Menten formula (Gunawardena 2014).

Foundations 19.2. Evolutionary sensitivity of pathway steps. Metabolic pathways often consist of a linear series of steps, each transforming a metabolite to a new product that then serves as a substrate for a subsequent enzyme. Such a chain can be represented as follows:

$$\text{S}_0 \rightleftharpoons \text{S}_1 \rightleftharpoons \text{S}_2 \cdots\cdots\cdots \rightleftharpoons \text{S}_{n-1} \rightleftharpoons \text{S}_n,$$

with S_0 denoting the initial substrate, and the S_i denoting the intermediate metabolites en route to the final product S_n. The right/left arrows between metabolites denote the forward and reverse reactions, which occur at rates k_{+i} and k_{-i} between substrates $i-1$ and i. In the following, we will assume that the initial substrate is kept at a constant intracellular concentration $[\text{S}_0]$, and that the final metabolite is utilized (and hence removed from the cytosol) at rate γ. Assuming that the rates on the transition arrows in the reaction scheme remain constant, such a system will yield a set of equilibrium concentrations for each of the metabolites.

Consider a simple two-enzyme system, for which the time course of concentration changes can be written as

$$\frac{d[\mathrm{S}_1]}{dt} = k_{+1}[\mathrm{S}_0] - (k_{-1} + k_{+2})[\mathrm{S}_1] + k_{-2}[\mathrm{S}_2] \tag{19.2.1a}$$

$$\frac{d[\mathrm{S}_2]}{dt} = k_{+2}[\mathrm{S}_1] - (k_{-2} + \gamma)[\mathrm{S}_2] \tag{19.2.1b}$$

The three terms in the first equation respectively denote the conversion of the initial substrate to S_1, the loss of S_1 to the two alternative substrates (S_0 and S_2), and the gain from the reverse reaction involving S_2. The steady-state concentrations are found by setting the derivatives equal to zero and solving the pair of equations,

$$[\widetilde{\mathrm{S}}_1] = \frac{K_1[\mathrm{S}_0]\{1 + (\gamma/k_{-2})\}}{1 + (\gamma K_2/k_{-1}) + (\gamma/k_{-2})} \tag{19.2.2a}$$

$$[\widetilde{\mathrm{S}}_2] = \frac{K_1 K_2[\mathrm{S}_0]}{1 + (\gamma K_2/k_{-1}) + (\gamma/k_{-2})} \tag{19.2.2b}$$

The terms involving upper-case Ks are the ratios of forward and reverse reaction rates, often referred to as the equilibrium constants,

$$K_i = k_{+i}/k_{-i}, \tag{19.2.3}$$

as they would define the equilibrium concentrations of two adjacent metabolites in the absence of any other steps in the system,

$$k_{+i}[\widetilde{\mathrm{S}}_{i-1}] = k_{-i}[\widetilde{\mathrm{S}}_i]. \tag{19.2.4}$$

The rate of flux through the system (i.e., the amount of final product drawn off for use in cellular functions) is equal to the product of the concentration of the final metabolite and its utilization rate, $\gamma[\widetilde{\mathrm{S}}_2]$,

$$F = \frac{\gamma K_1 K_2[\mathrm{S}_0]}{1 + (\gamma K_2/k_{-1}) + (\gamma/k_{-2})}. \tag{19.2.5}$$

The two other fluxes in the system, from S_0 to S_1 and from S_1 to S_2, occur at net rates $(k_{+1}[\mathrm{S}_0] - k_{-1}[\widetilde{\mathrm{S}}_1])$ and $(k_{+2}[\widetilde{\mathrm{S}}_1] - k_{-2}[\widetilde{\mathrm{S}}_2])$, respectively, when the system is in equilibrium. After substituting the equilibrium concentrations from Equations 19.2.2a,b, these flux rates are found to be identical to the equilibrium exit rate from the pathway, Equation 19.2.5, as must be the case for a steady-state system.

Although the algebra gets more tedious with longer pathways, Heinrich and Rapoport (1974) obtained the general solution for a pathway of length n,

$$[\widetilde{\mathrm{S}}_i] = [\mathrm{S}_0] \prod_{j=1}^{i} K_j \cdot \frac{1 + \gamma \sum_{l=i+1}^{n} (1/k_{-l}) \prod_{m=l+1}^{n} K_m}{1 + \gamma \sum_{l=1}^{n} (1/k_{-l}) \prod_{m=l+1}^{n} K_m}, \tag{19.2.6}$$

with the product terms being set equal to 1.0 when $m > n$. With this expression, the total flux rate can be written as the product of γ and the concentration of S_n,

$$F = \frac{\gamma[\mathrm{S}_0] \prod_{j=1}^{n} K_j}{1 + \gamma \sum_{l=1}^{n} (1/k_{-l}) \prod_{m=l+1}^{n} K_m}, \tag{19.2.7}$$

which reduces to Equation 19.2.5 when $n = 2$. Similar results were obtained independently by Kacser and Burns (1973).

We are now in a position to address some key questions regarding the degree to which the overall flux through a pathway is influenced by the individual component reactions. To simplify the analysis, we will assume that all of the enzymatic reactions have equal reversion rates k_- and equilibrium constants K. Because the summation in the denominator of Equation 19.2.7 is expected to be much greater than 1 (owing to the forward nature of the pathway), the 1 can be dropped, and substitution of the constant parameters leads to

$$F \simeq \frac{k_- K^n [\mathrm{S}_0]}{\sum_{i=0}^{n-1} K^i}. \tag{19.2.8a}$$

Still further simplification is possible because the equilibrium constant is likely to be $\gg 1$, in which case the final term in the series in the denominator dominates, leading to

$$F \simeq k_- K [\mathrm{S}_0]. \tag{19.2.8b}$$

This shows that provided K is larger than 10 or so, the steady-state flux rate is nearly independent of the number of steps in the pathway.

Finally, we consider the sensitivity of the flux rate to changes in the enzymes responsible for the individual steps. Suppose the equilibrium constant for a single step is multiplied by the factor x. Equation 19.2.8a defines F_* as the pre-perturbation flux rate, and F_j (the post-perturbation rate) is obtained from Equation 19.2.7 with all equilibrium constants equal to K, except that for the jth step, which is set to xK. The sensitivity coefficient for step i is then defined to be the fractional change in the total flux rate relative to its initial value, $(F_j - F_*)/F_*$, scaled by the fractional change in the equilibrium constant relative to its initial value $(x - 1)$,

$$C_j = \frac{F_j - F_*}{F_*(x-1)}, \tag{19.2.9a}$$

Substitution of the expressions for F_* and F_j into Equation 19.2.9a, followed by some algebra, yields

$$C_j = \frac{\sum_{i=0}^{n-j} K^i}{\sum_{i=0}^{n-1} K^i}, \tag{19.2.9b}$$

and again assuming $K \gg 1$, so that only the final terms in each summation predominate, this further simplifies to

$$C_j \simeq \frac{1}{K^{j-1}}. \tag{19.2.9c}$$

This shows that the relative sensitivity of the flux rate to a change in features influencing K at a particular step (e.g., enzyme concentration, substrate affinity, rate of reactivity) declines by a factor of $1/K$ with each increasing step, starting with $C_1 = 1$ (complete transmission of the effect) for a change at the first step.

Foundations 19.3. Optimization of the glycolytic flux rate. Glycolysis, a key energy-generating process found in most cells, involves a linear pathway of ten steps

starting with a glucose molecule and culminating in the production of two pyruvates (which can be further reduced to lactic acid) and two ATPs (Figure 19.3). At typical physiological concentrations of the participating molecules, the glucose to lactic-acid reaction releases $\sim$ 205 kJoules/mol, whereas each of the ADP to ATP conversions stores 50 kJoules/mol, meaning that 105 kJoules/mol is simply lost as heat.

Why has such an apparently wasteful process been universally retained, rather than the production of four ATPs (equivalent to a total of 200 kJoules/mol), or even three, being squeezed out of the reaction? One explanation is that instead of maximizing the efficiency of resource utilization, natural selection commonly puts a premium on the total rate of ATP production, which is a function of both efficiency and speed. Letting $\Delta G_{\rm gly} = -205$ kJoules/mol be the energy released by a glycolytic reaction, $\Delta G_{\rm ATP} = 50$ kJoules/mol be the amount retained per net ATP produced (the negative and positive signs denote energy release and gain, respectively), and n be the number of ATPs produced per glucose molecule, the efficiency of the energy recapture can be written as

$$\eta = \frac{n \cdot \Delta G_{\rm ATP}}{|\Delta G_{\rm gly}|}. \tag{19.3.1}$$

Thus, in terms of efficiency of resource utilization, just 49% of the energy endowment of glucose is retained by the organism. (Notably, the anaerobic process of glycolysis is also much less efficient than aerobic respiration via the Kreb's (TCA) cycle, which generates $\sim$ 30 ATPs per glucose molecule (Chapter 17), although here we focus on glycolysis alone).

The speed of a reaction (the flux rate) depends on the concentrations of reactants and products, as well as on the net energy differential between the two. In this case, this differential is equal to $\Delta G_{\rm gly} + (n \cdot |\Delta G_{\rm ATP})$ (recalling that $\Delta G_{\rm gly}$ is negative). The flux rate through the system is

$$F_{\rm glu} = L \cdot [\,|\Delta G_{\rm gly} + (n \cdot \Delta G_{\rm ATP})|\,], \tag{19.3.2}$$

where L is a constant that scales the flux rate to the net energy differential. Note that Equation 19.3.2 has the same form as the diffusion equation discussed in Chapter 7, where the diffusion rate was equal to the product of a diffusion coefficient and a local concentration gradient. Here, L is analogous to the diffusion coefficient, and the term in brackets is analogous to the concentration gradient. The actual value of L can depend on a number of cellular features, but is irrelevant to the following analysis.

The rate of ATP production is the product of the efficiency of conversion, Equation 19.3.1, and the glucose flux rate, Equation 19.3.2,

$$F_{\rm ATP} = L \cdot \frac{n \cdot \Delta G_{\rm ATP}}{\Delta G_{\rm gly}} \cdot [\Delta G_{\rm gly} + (n \cdot \Delta G_{\rm ATP})]. \tag{19.3.3}$$

Here, the absolute signs have been removed, as the negatives cancel out. Note, however, that the first (efficiency) term increases with n, whereas the second (flux-rate) term becomes closer to zero with increasing n (because $\Delta G_{\rm gly}$ is negative). This implies that there must be some intermediate value of n that maximizes the rate of ATP production, $F_{\rm ATP}$. This behavior reflects the intrinsic tradeoff between speed and efficiency.

By taking the first derivative of Equation 19.3.3 with respect to n, setting this equal to zero (the peak of a nonlinear function), and rearranging, we obtain

$$n^* = \frac{|\Delta G_{\rm gly}|}{2 \cdot \Delta G_{\rm ATP}} \tag{19.3.4}$$

as the ratio of ATP to glucose molecules that should maximize the total rate of ATP production. Substituting from above for the energy terms yields $n^* = 2.05$, remarkably close to the actual glycolytic ratio of 2.

Using Equation 19.3.3, we can further inquire as to the consequences of extracting different numbers of ATPs in this reaction. For $n = 1$, 2, 3, and 4, $F_{\text{ATP}} = 37.8L$, $51.2L$, $40.3L$, and $4.9L$, respectively. The number of ATPs produced per glucose molecule, n, can only take on integer values, but given that the adopted value of 2 is very close to the optimum, it follows that the alternatives of $n = 1$, 3, and 4 are expected to yield 26, 21, and 90% reductions in flux rates.

Further details on these issues can be found in Heinrich et al. (1997, 1999), with particularly lucid explanations given in Waddell et al. (1997) and Aledo and del Valle (2002).

Foundations 19.4. Extension / contraction of a metabolic pathway. Here we consider a situation in which there are two alternative pathways by which an organism can acquire an essential molecule S. Under state 1, S is only obtained by direct environmental uptake, whereas under state 3, S can only be obtained indirectly by conversion of an upstream precursor P. Both mechanisms of uptake are possible under state 2 (Figure 19.4). Under this model, evolutionary transitions occur between the alternative states by gains/losses of particular molecular mechanisms, with the probability of long-term population occupancy of each site being dictated by its selective advantage and mutation accessibility.

Following the scheme

$$\text{Pathway 1} \rightleftharpoons \text{Pathway 2} \rightleftharpoons \text{Pathway 3},$$

the equilibrium probabilities of the reliance on these alternative pathways in a phylogenetic lineage depend on the evolutionary transition rates between adjacent states. Given constant and nonzero transition rates over a long evolutionary time period, a lineage would be expected to slowly wander from state to state. With sufficient time, regardless of the starting state, the time-averaged state probabilities would then reach an equilibrium defined by their relative selective advantages and mutational interconversion rates (Foundations 5.2 and 5.3).

Each of the four transition rates (associated with the arrows in the preceding scheme) is equal to the product of the rate of origin of the appropriate mutation and the subsequent probability of fixation. For example, transitions from state 1 to 2 depend on: the rate of origin of mutations that allow conversion of P to S, Nv_1, where N is the number of individuals in the population (assumed to be haploid), v_1 is the per-individual mutational rate of origin of the upstream pathway, and the probability of fixation is denoted by $\phi(s_2 - s_1)$, where $(s_2 - s_1)$ is the selective advantage of pathway 2 over that of pathway 1. Extending this approach to all four coefficients leads to

$$1 \rightarrow 2: \qquad \theta_{12} = Nv_1\phi(s_2 - s_1) \tag{19.4.1a}$$
$$2 \rightarrow 1: \qquad \theta_{21} = Nu_1\phi(s_1 - s_2) \tag{19.4.1b}$$
$$2 \rightarrow 3: \qquad \theta_{23} = Nu_3\phi(s_3 - s_2) \tag{19.4.1c}$$
$$3 \rightarrow 2: \qquad \theta_{32} = Nv_3\phi(s_2 - s_3) \tag{19.4.1d}$$

where u_1 is the mutational rate of loss of the precursor pathway, and v_3 and u_3 denote the mutational rates of gain and loss of the environmental uptake mechanism. (Here we assume that state transitions are rare enough that the simultaneous presence of all three states within a population is negligible, i.e., the sequential model in the parlance of Chapter 5).

Two mathematical simplifications lead to a relatively straight-forward solution. First, the equilibrium frequencies of each of the alternative states in a linear array can

be determined by multiplying the coefficients on all of the arrows pointing toward the state from above and below. This means that

$$\widetilde{P}_1 \propto \theta_{21}\theta_{32} \tag{19.4.2a}$$

$$\widetilde{P}_2 \propto \theta_{12}\theta_{32} \tag{19.4.2b}$$

$$\widetilde{P}_3 \propto \theta_{12}\theta_{23} \tag{19.4.2c}$$

where $\propto$ means "proportional to." Second, the fixation probabilities are defined by the standard formula for newly arising mutations (Equation 4.1b), and have the property that the ratio of these probabilities for advantageous and deleterious mutations with the same absolute effects is

$$\phi(s)/\phi(-s) = e^{2N_e s} \tag{19.4.3}$$

for a haploid population with effective size N_e. Dividing Equations 19.4.2a-c by $\theta_{12}\theta_{32}$, and applying Equation 19.4.3 leads to

$$\widetilde{P}_1 = C \cdot \beta_1 e^{2N_e(s_1 - s_2)} \tag{19.4.4a}$$

$$\widetilde{P}_2 = C \cdot 1 \tag{19.4.4b}$$

$$\widetilde{P}_3 = C \cdot \beta_3 e^{2N_e(s_3 - s_2)} \tag{19.4.4c}$$

where $\beta_i = u_i/v_i$, and C is a normalization constant equal to the reciprocal of the sum of the three terms to the right of C in Equations 19.4.4a,c (usage of C insures that $\widetilde{P}_1 + \widetilde{P}_2 + \widetilde{P}_3 = 1$). As β_i defines the ratio of the mutational rates of loss to gain of the two alternative mechanisms of substrate acquisition, these expressions show that the probabilities of alternative states are simple functions of the net directional flux into the end states by mutational and selection processes.

Literature Cited

Abatemarco, J., A. Hill, and H. S. Alper. 2013. Expanding the metabolic engineering toolbox with directed evolution. Biotechnol. J. 8: 1397-1410.

Aharoni, A., L. Gaidukov, O. Khersonsky O, S. M. Gould, C. Roodveldt, and D. S. Tawfik. 2005. The 'evolvability' of promiscuous protein functions. Nat. Genet. 37: 73-76.

Albery, W. J., and J. R. Knowles. 1976a. Evolution of enzyme function and the development of catalytic efficiency. Biochemistry 15: 5631-5640.

Albery, W. J., and J. R. Knowles. 1976b. Free-energy profile of the reaction catalyzed by triosephosphate isomerase. Biochemistry 15: 5627-5631.

Aledo, J. C., and A. E. del Valle. 2002. Glycolysis in Wonderland: the importance of energy dissipation in metabolic pathways. J. Chem. Educ. 79: 1336-1339.

Allen, A. E., A. Moustafa, A. Montsant, A. Eckert, P. G. Kroth, and C. Bowler. 2012. Evolution and functional diversification of fructose bisphosphate aldolase genes in photosynthetic marine diatoms. Mol. Biol. Evol. 29: 367-379.

Alvarez, R. A., M. W. Blaylock, and J. B. Baseman. 2003. Surface localized glyceraldehyde-3-phosphate dehydrogenase of *Mycoplasma genitalium* binds mucin. Mol. Microbiol. 48: 1417-1425.

Alvarez-Ponce, D., M. Aguadé, and J. Rozas. 2009. Network-level molecular evolutionary analysis of the insulin/TOR signal transduction pathway across 12 *Drosophila* genomes. Genome Res. 19: 234-242.

Babtie, A., N. Tokuriki, and F. Hollfelder. 2010. What makes an enzyme promiscuous? Curr. Opin. Chem. Biol. 14: 200-207.

Bar-Even, A., A. Flamholz, E. Noor, and R. Milo. 2012. Rethinking glycolysis: on the biochemical logic of metabolic pathways. Nat. Chem. Biol. 8: 509-517.

Bar-Even, A., R. Milo, E. Noor, and D. S. Tawfik. 2015. The moderately efficient enzyme: futile encounters and enzyme floppiness. Biochemistry 54: 4969-4977.

Bar-Even, A., E. Noor, Y. Savir, W. Liebermeister, D. Davidi, D. S. Tawfik, and R. Milo. 2011. The moderately efficient enzyme: evolutionary and physicochemical trends shaping enzyme parameters. Biochemistry 50: 4402-4410.

Belfaiza, J., C. Parsot, A. Martel, C. B. de la Tour, D. Margarita, G. N. Cohen, and I. Saint-Girons. 1986. Evolution in biosynthetic pathways: two enzymes catalyzing consecutive steps in methionine biosynthesis originate from a common ancestor and possess a similar regulatory region. Proc. Natl. Acad. Sci. USA 83: 867-871.

Benkovic, S. J., and S. Hammes-Schiffer. 2003. A perspective on enzyme catalysis. Science 301: 1196-1202.

Bennett, B. D., E. H. Kimball, M. Gao, R. Osterhout, S. J. Van Dien, and J. D. Rabinowitz. 2009. Absolute metabolite concentrations and implied enzyme active site occupancy in *Escherichia coli.* Nat. Chem. Biol. 5: 593-599.

Berg, O. G. 1985. Orientation constraints in diffusion-limited macromolecular association: the role

of surface diffusion as a rate-enhancing mechanism. Biophys J. 47: 1-14.

Berg, O. G., and P. H. von Hippel. 1985. Diffusion-controlled macromolecular interactions. Annu. Rev. Biophys. Biophys. Chem. 14: 131-160.

Bergthorsson, U., D. I. Andersson, and J. R. Roth. 2007. Ohno's dilemma: evolution of new genes under continuous selection. Proc. Natl. Acad. Sci. USA 104: 17004-17009.

Blank, D., L. Wolf, M. Ackermann, and O. K. Silander. 2014. The predictability of molecular evolution during functional innovation. Proc. Natl. Acad. Sci. USA 111: 3044-3049.

Blount, Z. D., J. E. Barrick, C. J. Davidson, and R. E. Lenski. 2012. Genomic analysis of a key innovation in an experimental *Escherichia coli* population. Nature 489: 513-518.

Borenstein, E., M. Kupiec, M. W. Feldman, and E. Ruppin. 2008. Large-scale reconstruction and phylogenetic analysis of metabolic environments. Proc. Natl. Acad. Sci. USA 105: 14482-14487.

Bornscheuer, U. T., G. W. Huisman, R. J. Kazlauskas, S. Lutz, J. C. Moore, and K. Robins. 2012. Engineering the third wave of biocatalysis. Nature 485: 185-194.

Boronat, A., E. Caballero, and J. Aguilar. 1983. Experimental evolution of a metabolic pathway for ethylene glycol utilization by *Escherichia coli.* J. Bacteriol. 153: 134-139.

Briggs, G. E., and J. B. Haldane. 1925. A note on the kinetics of enzyme action. Biochem. J. 19: 338-339.

Bublitz, D. C., G. L. Chadwick, J. S. Magyar, K. M. Sandoz, D. M. Brooks, S. Mesnage, M. S. Ladinsky, A. I. Garber, P. J. Bjorkman, V. J. Orphan, et al. 2019. Peptidoglycan production by an insect-bacterial mosaic. Cell 179: 703-712.

Brustad, E. M., and F. H. Arnold. 2011. Optimizing non-natural protein function with directed evolution. Curr. Opin. Chem. Biol. 15: 201-210.

Byng, G. S., R. J. Whitaker, and R. A. Jensen. 1983. Evolution of L-phenylalanine biosynthesis in rRNA homology group I of *Pseudomonas.* Arch. Microbiol. 136: 163-168.

Camacho, C. J., S. R. Kimura, C. DeLisi, and S. Vajda. 2000. Kinetics of desolvation-mediated protein-protein binding. Biophys. J. 78: 1094-1105.

Canback, B., S. G. Andersson, and C. G. Kurland. 2002. The global phylogeny of glycolytic enzymes. Proc. Natl. Acad. Sci. USA 99: 6097-6102.

Carpenter, E. P., A. R. Hawkins, J. W. Frost, and K. A. Brown. 1998. Structure of dehydroquinate synthase reveals an active site capable of multistep catalysis. Nature 394: 299-302.

Castaldo, C., V. Vastano, R. A. Siciliano, M. Candela, M. Vici, L. Muscariello, R. Marasco, and M. Sacco. 2009. Surface displaced α-enolase of *Lactobacillus plantarum* is a fibronectin binding protein. Microb. Cell Fact. 8: 14.

Chang, Y. S., W. Wu, G. Walsh, W. K. Hong, and L. Mao. 2003. Enolase-α is frequently down-regulated in non-small cell lung cancer and predicts aggressive biological behavior. Clin. Cancer Res. 9: 3641-3644.

Chen, Y., and J. Nielsen. 2021 *In vitro* turnover numbers do not reflect *in vivo* activities of yeast enzymes. Proc. Natl. Acad. Sci. USA 118: e2108391118.

Colombo, M., H. Laayouni, B. M. Invergo, J. Bertranpetit, and L. Montanucci. 2014. Metabolic flux is a determinant of the evolutionary rates of enzyme-encoding genes. Evolution 68: 605-613.

Copley, S. D. 2000. Evolution of a metabolic pathway for degradation of a toxic xenobiotic: the patchwork approach. Trends Biochem. Sci. 25: 261-265.

Copley, S. D. 2003. Enzymes with extra talents: moonlighting functions and catalytic promiscuity. Curr. Opin. Chem. Biol. 7: 265-272.

Copley, S. D. 2012. Toward a systems biology perspective on enzyme evolution. J. Biol. Chem. 287: 3-10.

Copley, S. D. 2014. An evolutionary perspective on protein moonlighting. Biochem. Soc. Trans. 42: 1684-1691.

Copley, S. D. 2020. The physical basis and practical consequences of biological promiscuity. Phys. Biol. 17: 051001.

Crowley, P. H. 1975. Natural selection and the Michaelis constant. J. Theor. Biol. 50: 461-475.

Curien, G., V. Biou, C. Mas-Droux, M. Robert-Genthon, J. L. Ferrer, and R. Dumas. 2008. Amino acid biosynthesis: new architectures in allosteric enzymes. Plant Physiol. Biochem. 46: 325-339.

Davidi, D., L. M. Longo, J. Jabłońska, R. Milo, and D. S. Tawfik. 2018. A bird's-eye view of enzyme evolution: chemical, physicochemical, and physiological considerations. Chem. Rev. 118: 8786-8797.

Davidi, D., E. Noor, W. Liebermeister, A. Bar-Even, A. Flamholz, K. Tummler, U. Barenholz, M. Goldenfeld, T. Shlomi, and R. Milo. 2016. Global characterization of *in vivo* enzyme catalytic rates and their correspondence to *in vitro* k_{cat} measurements. Proc. Natl. Acad. Sci. USA 113: 3401-3406.

Dincturk, H. B., and D. B. Knaff. 2000. The evolution of glutamate synthase. Mol. Biol. Rep. 27: 141-148.

Dix, J. A., and A. S. Verkman. 2008. Crowding effects on diffusion in solutions and cells. Annu. Rev. Biophys. 37: 247-263.

Du, B., D. C. Zielinski, J. M. Monk, and B. O. Palsson. 2018. Thermodynamic favorability and pathway yield as evolutionary tradeoffs in biosynthetic pathway choice. Proc. Natl. Acad. Sci. USA 115: 11339-11344.

Du, J., R. F. Say, W. Lü, G. Fuchs, and O. Einsle. 2011. Active-site remodelling in the bifunctional fructose-1,6-bisphosphate aldolase/phosphatase. Nature 478: 534-537.

Due, A. V., J. Kuper, A. Geerlof, J. P. von Kries, and M. Wilmanns. 2011. Bisubstrate specificity in histidine/tryptophan biosynthesis isomerase from *Mycobacterium tuberculosis* by active site metamorphosis. Proc. Natl. Acad. Sci. USA 108: 3554-3559.

Ellis, R. J. 2001. Macromolecular crowding: obvious but underappreciated. Trends Biochem. Sci. 26: 597-604.

Elowitz, M. B., M. G. Surette, P. E. Wolf, J. B. Stock, and S. Leibler. 1999. Protein mobility in the cytoplasm of *Escherichia coli*. J. Bacteriol. 181: 197-203.

Fernandes, P., H. Aldeborgh, L. Carlucci, L. Walsh, J. Wasserman, E. Zhou, S. T. Lefurgy, and E. C. Mundorff. 2015. Alteration of substrate specificity of alanine dehydrogenase. Protein Eng. Des. Sel. 28: 29-35.

Fondi, M., M. Brilli, G. Emiliani, D. Paffetti, and R. Fani. 2007. The primordial metabolism: an

ancestral interconnection between leucine, arginine, and lysine biosynthesis. BMC Evol. Biol. 7 Suppl. 2: S3.

Fraser, H. B., A. E. Hirsh, L. M. Steinmetz, C. Scharfe, and M. W. Feldman. 2002. Evolutionary rate in the protein interaction network. Science 296: 750-752.

Gabaldón, T., and E. V. Koonin. 2013. Functional and evolutionary implications of gene orthology. Nat. Rev. Genet. 14: 360-366.

Galperin, M. Y., D. R. Walker, and E. V. Koonin. 1998. Analogous enzymes: independent inventions in enzyme evolution. Genome Res. 8: 779-790.

Ge, Y. D., P. Song, Z. Y. Cao, P. Wang, and G. P. Zhu. 2014. Alteration of coenzyme specificity of malate dehydrogenase from *Streptomyces coelicolor* A3(2) by site-directed mutagenesis. Genet. Mol. Res. 13: 5758-5766.

Gophna, U., E. Bapteste, W. F. Doolittle, D. Biran, and E. Z. Ron. 2005. Evolutionary plasticity of methionine biosynthesis. Gene 355: 48-57.

Gozalbo, D., I. Gil-Navarro, I. Azorín, J. Renau-Piqueras, J. P. Martínez, and M. L. Gil. 1998. The cell wall-associated glyceraldehyde-3-phosphate dehydrogenase of *Candida albicans* is also a fibronectin and laminin binding protein. Infect. Immun. 66: 2052-2059.

Gunawardena, J. 2012. Some lessons about models from Michaelis and Menten. Mol. Biol. Cell 23: 517-519.

Gunawardena, J. 2014. Time-scale separation – Michaelis and Menten's old idea, still bearing fruit. FEBS J. 281: 473-488.

Hammes, G. G. 2002. Multiple conformational changes in enzyme catalysis. Biochemistry 41: 8221-8228.

Hartl, D. L., D. E. Dykhuizen, and A. M. Dean. 1985. Limits of adaptation: the evolution of selective neutrality. Genetics 111: 655-674.

Hébert, A., S. Casaregola, and J. M. Beckerich. 2011. Biodiversity in sulfur metabolism in hemiascomycetous yeasts. FEMS Yeast Res. 11: 366-378.

Heckmann, D., D. C. Zielinski, and B. O. Palsson. 2018. Modeling genome-wide enzyme evolution predicts strong epistasis underlying catalytic turnover rates. Nat. Commun. 9: 5270.

Heinrich, R., E. Meléndez-Hevia, F. Montero, J. C. Nuño, A. Stephani, and T. G. Waddell. 1999. The structural design of glycolysis: an evolutionary approach. Biochem. Soc. Trans. 27: 294-298.

Heinrich, R., F. Montero, E. Klipp, T. G. Waddell, and E. Meléndez-Hevia. 1997. Theoretical approaches to the evolutionary optimization of glycolysis: thermodynamic and kinetic constraints. Eur. J. Biochem. 243: 191-201.

Heinrich, R., and T. A. Rapoport. 1974. A linear steady-state treatment of enzymatic chains: general properties, control and effector strength. Eur. J. Biochem. 42: 89-95.

Hittinger, C. T., A. Rokas, and S. B. Carroll. 2004. Parallel inactivation of multiple GAL pathway genes and ecological diversification in yeasts. Proc. Natl. Acad. Sci. USA 101: 14144-14149.

Ho, C. L., M. Noji, M. Saito, M. Yamazaki, and K. Saito. 1998. Molecular characterization of plastidic phosphoserine aminotransferase in serine biosynthesis from *Arabidopsis*. Plant J. 16:

443-452.

Hochuli, M., H. Patzelt, D. Oesterhel, K. Wüthrich, and T. Szyperski. 1999. Amino acid biosynthesis in the halophilic archaeon *Haloarcula hispanica.* J. Bacteriol. 181: 3226-3237.

Horowitz, N. H. 1945. On the evolution of biochemical syntheses. Proc. Natl. Acad. Sci. USA 31: 153-157.

Hoshino, Y., and E. A. Gaucher. 2018. On the origin of isoprenoid biosynthesis. Mol. Biol. Evol. 35: 2185-2197.

Hudson, A. O., B. K. Singh, T. Leustek, and C. Gilvarg. 2006. An LL-diaminopimelate aminotransferase defines a novel variant of the lysine biosynthesis pathway plants. Plant Physiol. 140: 292-301.

Hughes, A. L. 1994. The evolution of functionally novel proteins after gene duplication. Proc. Roy. Soc. Lond. B 256: 119-124.

Hult, K., and P. Berglund. 2007. Enzyme promiscuity: mechanism and applications. Trends Biotechnol. 25: 231-238.

Iida, H., and I. Yahara. 1985. Yeast heat-shock protein of Mr 48,000 is an isoprotein of enolase. Nature 315: 688-690.

Janin, J. 1997. The kinetics of protein-protein recognition. Proteins 28: 153-161.

Jensen, R. A. 1976. Enzyme recruitment in evolution of new function. Annu. Rev. Microbiol. 30: 409-425.

Jensen, R. A. 1985. Biochemical pathways in prokaryotes can be traced backward through evolutionary time. Mol. Biol. Evol. 2: 92-108.

Jensen, R. A., and D. L. Pierson. 1975. Evolutionary implications of different types of microbial enzymology for L-tyrosine biosynthesis. Nature 254: 667-671.

Jia, B., G. W. Cheong, and S. Zhang. 2013. Multifunctional enzymes in archaea: promiscuity and moonlight. Extremophiles 17: 193-203.

Johannes, T. W., and H. Zhao. 2006. Directed evolution of enzymes and biosynthetic pathways. Curr. Opin. Microbiol. 9: 261-267.

Johnsen, U., and P. Schönheit. 2007. Characterization of cofactor-dependent and cofactor-independent phosphoglycerate mutases from Archaea. Extremophiles 11: 647-657.

Johnson, K. A., and R. S. Goody. 2011. The original Michaelis constant: translation of the 1913 Michaelis-Menten paper. Biochemistry 50: 8264-8269.

Juárez-Vázquez, A. L., J. N. Edirisinghe, E. A. Verduzco-Castro, K. Michalska, C. Wu, L. Noda-García, G. Babnigg, M. Endres, S. Medina-Ruíz, J. Santoyo-Flores, et al. 2017. Evolution of substrate specificity in a retained enzyme driven by gene loss. eLife 6: e22679.

Kacser, H., and R. Beeby. 1984. Evolution of catalytic proteins or on the origin of enzyme species by means of natural selection. J. Mol. Evol. 20: 38-51.

Kacser, H., and J. A. Burns. 1973. The control of flux. Symp. Soc. Exp. Biol. 27: 65-104.

Kanai, S., and H. Toh. 1999. Identification of new members of the GS ADP-forming family from the *de novo* purine biosynthesis pathway. J. Mol. Evol. 48: 482-492.

Kawai, S., T. Mukai, S. Mori, B. Mikami, and K. Murata. 2005. Hypothesis: structures, evolution, and ancestor of glucose kinases in the hexokinase family. J. Biosci. Bioeng. 99: 320-330.

Khersonsky, O., C. Roodveldt, and D. S. Tawfik. 2006. Enzyme promiscuity: evolutionary and mechanistic aspects. Curr. Opin. Chem. Biol. 10: 498-508.

Kim, J., J. J. Flood, M. R. Kristofich, C. Gidfar, A. B. Morgenthaler, T. Fuhrer, U. Sauer, D. Snyder, V. S. Cooper, C. C. Ebmeier, et al. 2019. Hidden resources in the *Escherichia coli* genome restore PLP synthesis and robust growth after deletion of the essential gene pdxB. Proc. Natl. Acad. Sci. USA 116: 24164-24173.

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

Kim, J. W., and C. V. Dang. 2005. Multifaceted roles of glycolytic enzymes. Trends Biochem. Sci. 30: 142-150.

Klesmith, J. R., J. P. Bacik, E. E. Wrenbeck, R. Michalczyk, and T. A. Whitehead. 2017. Trade-offs between enzyme fitness and solubility illuminated by deep mutational scanning. Proc. Natl. Acad. Sci. USA 114: 2265-2270.

Konopka, M. C., I. A. Shkel, S. Cayley, M. T. Record, and J. C. Weisshaar. 2006. Crowding and confinement effects on protein diffusion *in vivo*. J Bacteriol. 188: 6115-6123.

Küken, A., K. Gennermann, and Z. Nikoloski. 2020. Characterization of maximal enzyme catalytic rates in central metabolism of *Arabidopsis thaliana*. Plant J. 103: 2168-2177.

Lee, D. H., and B. Ø. Palsson. 2010. Adaptive evolution of *Escherichia coli* K-12 MG1655 during growth on a Nonnative carbon source, L-1,2-propanediol. Appl. Environ. Microbiol. 76: 4158-4168.

Leopoldseder, S., J. Claren, C. Jürgens, and R. Sterner. 2004. Interconverting the catalytic activities of $(\beta\alpha)(8)$-barrel enzymes from different metabolic pathways: sequence requirements and molecular analysis. J. Mol. Biol. 337: 871-879.

Liapounova, N. A., V. Hampl, P. M. Gordon, C. W. Sensen, L. Gedamu, and J. B. Dacks. 2006. Reconstructing the mosaic glycolytic pathway of the anaerobic eukaryote Monocercomonoides. Eukaryot. Cell 5: 2138-2146.

List, F., R. Sterner, and M. Wilmanns. 2011. Related $(\beta\alpha)8$-barrel proteins in histidine and tryptophan biosynthesis: a paradigm to study enzyme evolution. Chembiochem.12: 1487-1494.

Lu, Y., and M. D. Rausher. 2003. Evolutionary rate variation in anthocyanin pathway genes. Mol. Biol. Evol. 20: 1844-1853.

Luby-Phelps, K. 2000. Cytoarchitecture and physical properties of cytoplasm: volume, viscosity, diffusion, intracellular surface area. Int. Rev. Cytol. 192: 189-221.

Lynch, M. 2007. The evolution of genetic networks by nonadaptive processes. Nature Rev. Genetics 8: 803-813.

Matsumura, I., and A. D. Ellington. 2001. *In vitro* evolution of beta-glucuronidase into a beta-galactosidase proceeds through non-specific intermediates. J. Mol. Biol. 305: 331-339.

McCutcheon, J. P., and N. A. Moran. 2007. Parallel genomic evolution and metabolic interdependence in an ancient symbiosis. Proc. Natl. Acad. Sci. USA 104: 19392-19397.

Melcher, K., M. Rose, M. Künzler, G. H. Braus, and K. D. Entian. 1995. Molecular analysis of the yeast SER1 gene encoding 3-phosphoserine aminotransferase: regulation by general control and serine repression. Curr. Genet. 27: 501-508.

Merino, E., R. A. Jensen, and C. Yanofsky. 2008. Evolution of bacterial trp operons and their regulation. Curr. Opin. Microbiol. 11: 78-86.

Michaelis, L., and M. L. Menten. 1913. Die Kinetik der Invertinwirkung. Biochem. Z. 49: 333-369.

Miles, L. A., C. M. Dahlberg, J. Plescia, J. Felez, K. Kato, and E. F. Plow. 1991. Role of cell-surface lysines in plasminogen binding to cells: identification of alpha-enolase as a candidate plasminogen receptor. Biochemistry 30: 1682-1691.

Miyazaki, J., N. Kobashi, M. Nishiyama, and H. Yamane. 2001. Functional and evolutionary relationship between arginine biosynthesis and prokaryotic lysine biosynthesis through alpha-aminoadipate. J. Bacteriol. 183: 5067-5073.

Modun, B., and P. Williams. 1999. The staphylococcal transferrin-binding protein is a cell wall glyceraldehyde-3-phosphate dehydrogenase. Infect. Immun. 67: 1086-1092.

Monk, J., J. Nogales, and B. O. Palsson. 2014. Optimizing genome-scale network reconstructions. Nat. Biotechnol. 32: 447-452.

Montanucci, L., H. Laayouni, G. M. Dall'Olio, and J. Bertranpetit. 2011. Molecular evolution and network-level analysis of the N-glycosylation metabolic pathway across primates. Mol. Biol. Evol. 28: 813-823.

Morgenthaler, A. B., W. R. Kinney, C. C. Ebmeier, C. M. Walsh, D. J. Snyder, V. S. Cooper, W. M. Old, and S. D. Copley. 2019. Mutations that improve efficiency of a weak-link enzyme are rare compared to adaptive mutations elsewhere in the genome. eLife 8: e53535.

Nam, H., N. E. Lewis, J. A. Lerman, D. H. Lee, R. L. Chang, D. Kim, and B. O. Palsson. 2012. Network context and selection in the evolution to enzyme specificity. Science 337: 1101-1104.

Näsvall, J., L. Sun, J. R. Roth, and D. I. Andersson. 2012. Real-time evolution of new genes by innovation, amplification, and divergence. Science 338: 384-387.

Nenninger, A., G. Mastroianni, and C. W. Mullineaux. 2010. Size dependence of protein diffusion in the cytoplasm of *Escherichia coli*. J. Bacteriol. 192: 4535-4540.

Nguyen, V., C. Wilson, M. Hoemberger, J. B. Stiller, R. V. Agafonov, S. Kutter, J. English, D. L. Theobald, and D. Kern. 2017. Evolutionary drivers of thermoadaptation in enzyme catalysis. Science 355: 289-294.

Nishida, H., and M. Nishiyama. 2012. Evolution of lysine biosynthesis in the phylum deinococcus-thermus. Int. J. Evol. Biol. 2012: 745931.

Nishida, H., M. Nishiyama, N. Kobashi, T. Kosuge, T. Hoshino, and H. Yamane. 1999. A prokaryotic gene cluster involved in synthesis of lysine through the amino adipate pathway: a key to the evolution of amino acid biosynthesis. Genome Res. 9: 1175-1183.

Noor, E., E. Eden, R. Milo, and U. Alon. 2010. Central carbon metabolism as a minimal biochemical walk between precursors for biomass and energy. Mol. Cell 39: 809-820.

Northrup, S. H., and H. P. Erickson. 1992. Kinetics of protein-protein association explained by Brownian dynamics computer simulation. Proc. Natl. Acad. Sci. USA 89: 3338-3342.

Notebaart, R. A., B. Szappanos, B. Kintses, F. Pál, Á. Györkei, B. Bogos, V. Lázár, R. Spohn, B. Csörgo, A. Wagner, E. Ruppin, C. Pál, and B. Papp. 2014. Network-level architecture and the evolutionary potential of underground metabolism. Proc. Natl. Acad. Sci. USA 111: 11762-11767.

Oslancová, A., and S. Janecek. 2004. Evolutionary relatedness between glycolytic enzymes most frequently occurring in genomes. Folia Microbiol (Praha) 49: 247-258.

Pál, C., B. Papp, M. J. Lercher, P. Csermely, S. G. Oliver, and L. D. Hurst. 2006. Chance and necessity in the evolution of minimal metabolic networks. Nature 440: 667-670.

Parsot, C. 1986 Evolution of biosynthetic pathways: a common ancestor for threonine synthase, threonine dehydratase and D-serine dehydratase. EMBO J. 5: 3013-3019.

Parsot, C., I. Saint-Girons, and G. N. Cohen. 1987. Enzyme specialization during the evolution of amino acid biosynthetic pathways. Microbiol. Sci. 4: 260-262.

Patrick, W. M., E. M. Quandt, D. B. Swartzlander, and I. Matsumura. 2007. Multicopy suppression underpins metabolic evolvability. Mol. Biol. Evol. 24: 2716-2722.

Pfeiffer, T., O. S. Soyer, and S. Bonhoeffer. 2005. The evolution of connectivity in metabolic networks. PLoS Biol. 3: e228.

Piatigorsky, J., and G. Wistow. 1991. The recruitment of crystallins: new functions precede gene duplication. Science 252: 1078-1079.

Qian, Q., and P. J. Keeling. 2001. Diplonemid glyceraldehyde-3-phosphate dehydrogenase (GAPDH) and prokaryote-to-eukaryote lateral gene transfer. Protist 152: 193-201.

Quandt, E. M., D. E. Deatherage, A. D. Ellington, G. Georgiou, and J. E. Barrick. 2014. Recursive genomewide recombination and sequencing reveals a key refinement step in the evolution of a metabolic innovation in *Escherichia coli.* Proc. Natl. Acad. Sci. USA 111: 2217-2222.

Rakus, J. F., A. A. Fedorov, E. V. Fedorov, M. E. Glasner, B. K. Hubbard, J. D. Delli, P. C. Babbitt, S. C. Almo, and J. A. Gerlt. 2008. Evolution of enzymatic activities in the enolase superfamily: L-rhamnonate dehydratase. Biochemistry 47: 9944-9954.

Ramazzina, I., R. Costa, L. Cendron, R. Berni, A. Peracchi, G. Zanotti, and R. Percudani. 2010. An aminotransferase branch point connects purine catabolism to amino acid recycling. Nat. Chem. Biol. 6: 801-806.

Ramsay, H., L. H. Rieseberg, and K. Ritland. 2009. The correlation of evolutionary rate with pathway position in plant terpenoid biosynthesis. Mol. Biol. Evol. 26: 1045-1053.

Raushel, F. M., J. B. Thoden, G. D. Reinhart, and H. M. Holden. 1998. Carbamoyl phosphate synthetase: a crooked path from substrates to products. Curr. Opin. Chem. Biol. 2: 624-632.

Rausher, M. D. 2012. The evolution of genes in branched metabolic pathways. Evolution 67: 34-48.

Rausher, M. D., R. E. Miller, and P. Tiffin. 1999. Patterns of evolutionary rate variation among genes of the anthocyanin biosynthetic pathway. Mol. Biol. Evol. 16: 266-274.

Rauwerdink, A., M. Lunzer, T. Devamani, B. Jones, J. Mooney, Z. J. Zhang, J. H. Xu, R. J. Kazlauskas, and A. M. Dean. 2016. Evolution of a catalytic mechanism. Mol. Biol. Evol. 33: 971-979.

Rogers, M., and P. J. Keeling. 2004. Lateral transfer and recompartmentalization of cycle enzymes

of plants and algae. J. Mol. Evol. 58: 367-375.

Rothman, S. C., and J. F. Kirsch. 2003. How does an enzyme evolved *in vitro* compare to naturally occurring homologs possessing the targeted function? Tyrosine aminotransferase from aspartate aminotransferase. J. Mol. Biol. 327: 593-608.

Roy, S. 1999. Multifunctional enzymes and evolution of biosynthetic pathways: retro-evolution by jumps. Proteins 37: 303-309.

Sakuraba, H., S. Fujiwara, and T. Noguchi. 1996. Metabolism of glyoxylate, the end product of purin degradation, in liver peroxisomes of fresh water fish. Biochem. Biophys. Res. Commun. 229: 603-606.

Sánchez, L., D. Horner, D. Moore, K. Henze, T. Embley, and M. Müller. 2002. Fructose-1,6-bisphosphate aldolases in amitochondriate protists constitute a single protein subfamily with eubacterial relationships. Gene 295: 51-59.

Schlosshauer, M., and D. Baker. 2004. Realistic protein-protein association rates from a simple diffusional model neglecting long-range interactions, free energy barriers, and landscape ruggedness. Protein Sci. 13: 1660-1669.

Seidler, N. W. 2013. Functional diversity. Adv. Exp. Med. Biol. 985: 103-147.

Shimizu, Y., H. Sakuraba, K. Doi, and T. Ohshima. 2008. Molecular and functional characterization of D-3-phosphoglycerate dehydrogenase in the serine biosynthetic pathway of the hyperthermophilic archaeon *Sulfolobus tokodaii.* Arch. Biochem. Biophys. 470: 120-128.

Siebers, B., H. Brinkmann, C. Dörr, B. Tjaden, H. Lilie, J. van der Oost, and C. H. Verhees. 2001. Archaeal fructose-1,6-bisphosphate aldolases constitute a new family of archaeal type class I aldolase. J. Biol. Chem. 276: 28710-28718.

Siebers, B., and P. Schönheit. 2005. Unusual pathways and enzymes of central carbohydrate metabolism in Archaea. Curr. Opin. Microbiol. 8: 695-705.

Skolnick, J., and M. Gao. 2013. Interplay of physics and evolution in the likely origin of protein biochemical function. Proc. Natl. Acad. Sci. USA 110: 9344-9349.

Song, J., C. A. Bonner, M. Wolinsky, and R. A. Jensen. 2005. The TyrA family of aromatic-pathway dehydrogenases in phylogenetic context. BMC Biol. 3: 13.

Soo, V. W., P. Hanson-Manful, and W. M. Patrick. 2011. Artificial gene amplification reveals an abundance of promiscuous resistance determinants in *Escherichia coli.* Proc. Natl. Acad. Sci. USA 108: 1484-1489.

Suzuki,A., and D. B. Knaff. 2005. Glutamate synthase: structural, mechanistic and regulatory properties, and role in the amino acid metabolism. Photosynth. Res. 83: 191-217.

Takishita, K., and Y. Inagaki. 2009. Eukaryotic origin of glyceraldehyde-3-phosphate dehydrogenase genes in *Clostridium thermocellum* and *Clostridium cellulolyticum* genomes and putative fates of the exogenous gene in the subsequent genome evolution. Gene 441: 22-27.

Tokuriki, N., and D. S. Tawfik. 2009. Protein dynamism and evolvability. Science 324: 203-207.

Torruella, G., H. Suga, M. Riutort, J. Peretó, and I. Ruiz-Trillo. 2009. The evolutionary history of lysine biosynthesis pathways within eukaryotes. J. Mol. Evol. 69: 240-248.

Tunio, S. A., N. J. Oldfield, A. Berry, D. A. Ala'Aldeen, K. G. Wooldridge, and D. P. Turner. 2010.

The moonlighting protein fructose-1, 6-bisphosphate aldolase of *Neisseria meningitidis:* surface localization and role in host cell adhesion. Mol. Microbiol. 76: 605-615.

van der Oost, J., M. A. Huynen, and C. H. Verhees. 2002. Molecular characterization of phosphoglycerate mutase in archaea. FEMS Microbiol. Lett. 212: 111-120.

Velasco, A. M., J. I. Leguina, and A. Lazcano. 2002. Molecular evolution of the lysine biosynthetic pathways. J. Mol. Evol. 55: 445-459.

Verhees, C. H., J. E. Tuininga, S. W. Kengen, A. J. Stams, J. van der Oost, and V. M. de Vos. 2001. ADP-dependent phosphofructokinases in mesophilic and thermophilic methanogenic archaea. J. Bacteriol. 183: 7145-7153.

Vitkup, D., P. Kharchenko, and A. Wagner. 2006. Influence of metabolic network structure and function on enzyme evolution. Genome Biol. 7: R39.

von Stockar, U., and J. Liu. 1999. Does microbial life always feed on negative entropy? Thermodynamic analysis of microbial growth. Biochim. Biophys. Acta 1412: 191-211.

Voordeckers, K., C. A. Brown, K. Vanneste, E. van der Zande, A. Voet, S. Maere, and K. J. Verstrepen. 2012. Reconstruction of ancestral metabolic enzymes reveals molecular mechanisms underlying evolutionary innovation through gene duplication. PLoS Biol. 10: e1001446.

Waddell, T. G., P. Repovic, E. Meléndez-Hevia, R. Heinrich, and F. Montero. 1997. Optimization of glycolysis: a new look at the efficiency of energy coupling. Biochem. Education 25: 204-205.

Wagoner, J. A., and K. A. Dill. 2019. Opposing pressures of speed and efficiency guide the evolution of molecular machines. Mol. Biol. Evol. 36: 2813-2822.

Wilks, H. M., K. W. Hart, R. Feeney, C. R. Dunn, H. Muirhead, W. N. Chia, D. A. Barstow, T. Atkinson, A. R. Clarke, and J. J. Holbrook. 1988. A specific, highly active malate dehydrogenase by redesign of a lactate dehydrogenase framework. Science 242: 1541-1544.

Wilmanns, M., C. C. Hyde, D. R. Davies, K. Kirschner, and J. N. Jansonius. 1991. Structural conservation in parallel beta/alpha-barrel enzymes that catalyze three sequential reactions in the pathway of tryptophan biosynthesis. Biochemistry 30: 9161-9169.

Wistow, G. J., and J. Piatigorsky. 1988. Lens crystallins: the evolution and expression of proteins for a highly specialized tissue. Annu. Rev. Biochem. 57: 479-504.

Wolfenden, R., and M. J. Snider. 2001. The depth of chemical time and the power of enzymes as catalysts. Acc. Chem. Res. 34: 938-945.

Wong, F., A. Dutta, D. Chowdhury, and J. Gunawardena. 2018. Structural conditions on complex networks for the Michaelis-Menten input-output response. Proc. Natl. Acad. Sci. USA 115: 9738-9743.

Wright, K. M., and M. D. Rausher. 2010. The evolution of control and distribution of adaptive mutations in a metabolic pathway. Genetics 184: 483-502.

Xie, G., N. O. Keyhani, C. A. Bonner, and R. A. Jensen. 2003. Ancient origin of the tryptophan operon and the dynamics of evolutionary change. Microbiol. Mol. Biol. Rev. 67: 303-342.

Xu, Y., B. Labedan, and N. Glansdorff. 2007. Surprising arginine biosynthesis: a reappraisal of the enzymology and evolution of the pathway in microorganisms. Microbiol. Mol. Biol. Rev. 71: 36-47.

Yang, Y. H., F. M. Zhang, and S. Ge. 2009. Evolutionary rate patterns of the gibberellin pathway genes. BMC Evol. Biol. 9: 206.

Yano, T., S. Oue, and H. Kagamiyama. 1998. Directed evolution of an aspartate aminotransferase with new substrate specificities. Proc. Natl. Acad. Sci. USA 95: 5511-5515.

Ycas, M. 1974. On earlier states of the biochemical system. J. Theor. Biol. 44: 145-160.

Zaslaver, A., A. E. Mayo, R. Rosenberg, P. Bashkin, H. Sberro, M. Tsalyuk, M. G. Surette, and U. Alon. 2004. Just-in-time transcription program in metabolic pathways. Nat. Genet. 36: 486-491.

Zhang, Y., M. Morar, and S. E. Ealick. 2008. Structural biology of the purine biosynthetic pathway. Cell. Mol. Life Sci. 65: 3699-3724.

Zhou, H. X. 1993. Brownian dynamics study of the influences of electrostatic interaction and diffusion on protein-protein association kinetics. Biophys J. 64: 1711-1726.

Zimmerman, S. B., and S. O. Trach. 1991. Estimation of macromolecule concentrations and excluded volume effects for the cytoplasm of *Escherichia coli*. J. Mol. Biol. 222: 599-620.

Zubieta, C., K. A. Arkus, R. E. Cahoon, and J. M. Jez. 2008. A single amino acid change is responsible for evolution of acyltransferase specificity in bacterial methionine biosynthesis. J. Biol. Chem. 283: 7561-7567.

Figure 19.1. Michaelis-Menten enzyme kinetics as a function of substrate concentration, [S], as defined in Equation 19.1.4. The maximum rate $k_{\mathrm{cat}}[\mathrm{E}]$ is arbitrarily set to equal 1.0, with k_{cat} being the maximum catalytic rate, [E] the enzyme concentration, and K_{S} the half-saturation constant (here also set equal to 1.0). The plot appears hyperbolic on an arithmetic scale and sigmoid on a logarithmic scale.

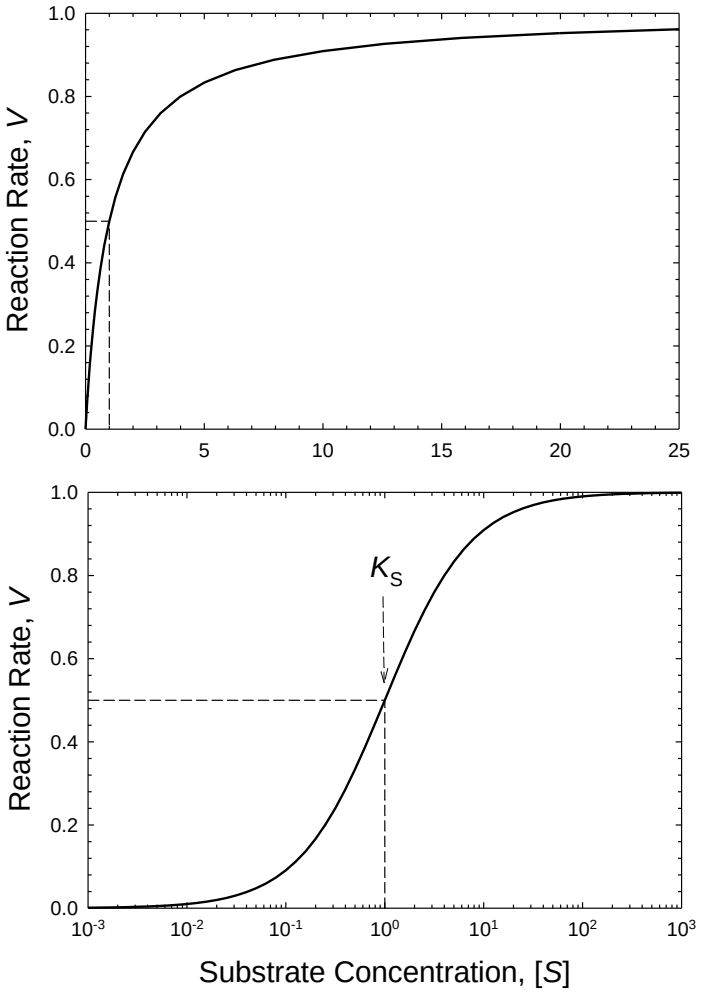


Figure 19.2. Distributions of kinetic features of enzymes surveyed across the Tree of Life. Sample sizes are 3500, 12576, and 4451 for the three parameters, almost all determined *in vitro*. Median values are given in the insets. From Davidi et al. (2018).

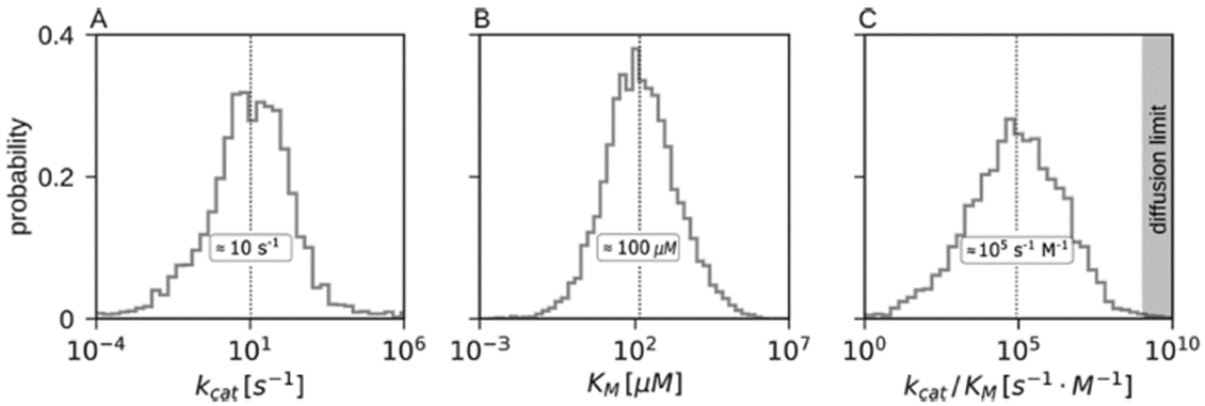


Figure 19.3. The canonical pathway of glycolysis, with ten different enzymes operating in a stepwise fashion to produce two pyruvates from a glucose molecule (only one pyruvate is shown). As noted by the side arrows, a number of the intermediate metabolites are used as entry points to various biosynthetic pathways. Two ATPs are consumed early in the process, but four are produced in downstream steps (two for each pyruvate), giving a net yield of two ATPs per glucose molecule consumed.

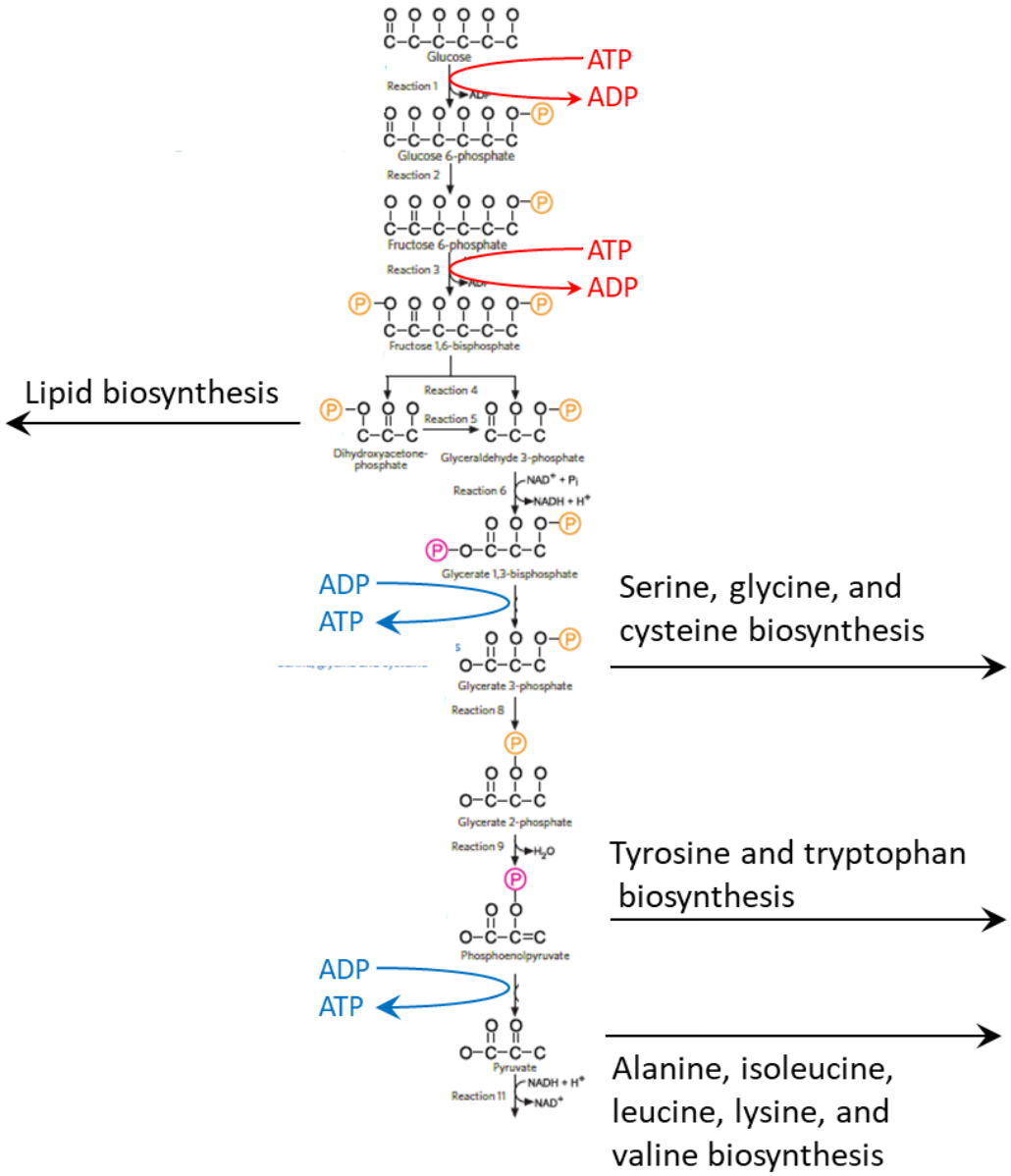


Figure 19.4. Alternative pathway topologies for acquisition of a key metabolite S. **1)** At one extreme, the sole source of S is environmental (blue). **2)** From this perspective, upstream pathway growth occurs if a mechanism (green arrow) arises that converts another substrate (P) to the focal metabolite (S). Initially, this results in two (at least partially redundant) pathways to metabolite S, and the output of the pre-existing pathway descending from P (which may or may not be of relevance to cellular fitness) will be reduced accordingly. **3)** If the environmental-uptake mechanism for S is then lost, acquisition of S will have become obligatorily dependent on the precursor pathway, with the linear pathway now having an additional step. In principle, this evolutionary trajectory can be reversed if a population in the third state acquires a capacity for direct environmental uptake of S.

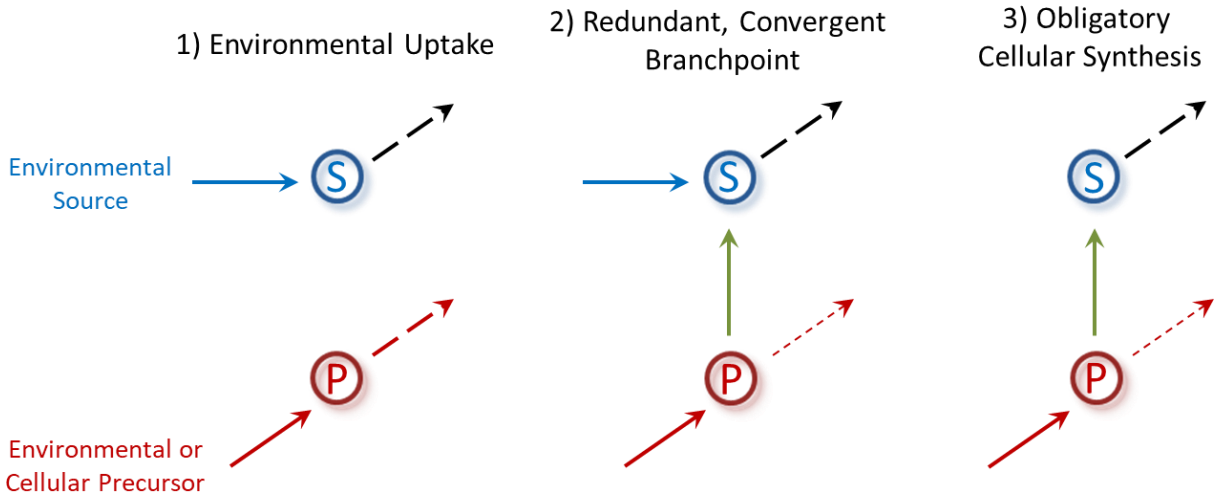


Figure 19.5. A hyperbolic fitness function, with R_S and R_P, respectively, denoting amounts of resources available directly from the environment and via a precursor molecule, and r_1 and r_3 denoting reductions in fitness relative to the case in which both sources of substrate molecule are available.

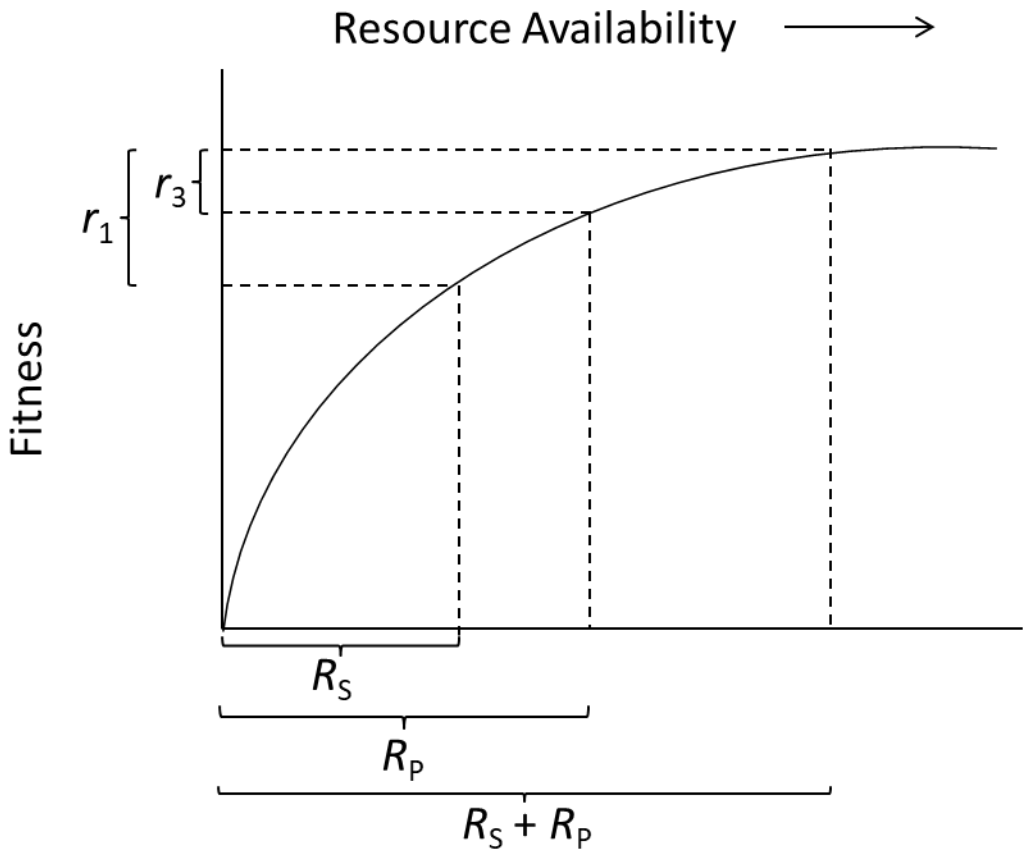
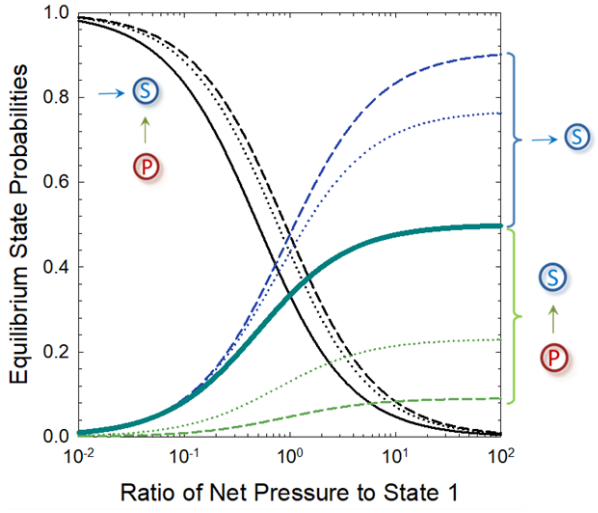


Figure 19.6. Left) Long-term evolutionary probabilities of the three alternative pathway states in Figure 19.4, as a function of the relative net forces of mutation and selection from and to the intermediate (redundant) state 2. The x axis is the ratio of the net pressure towards state 1 relative to 2; Solid lines denote the situation in which the two extreme states (1 and 2) experience identical mutation/selection pressures (in this case, the incidences of states 1 and 3 are identical, and given as the thick cyan line). Dotted and dashed lines respectively denote situations in which the bias toward the extended-pathway state (3) is 0.3 and 0.1× that toward the external dependent state (1); with ratios of 3.33 and 10.0, the positions of the state-1 (blue) and state-3 (green) curves would simply be reversed. All curves follow directly from Equations 19.4.4a-c. **Right)** The phase diagram denoting which of the three alternative pathway types will have the highest likelihood of occurrence is entirely a function of two composite parameters, the ratios of the net rates of loss/gain of the two end states.



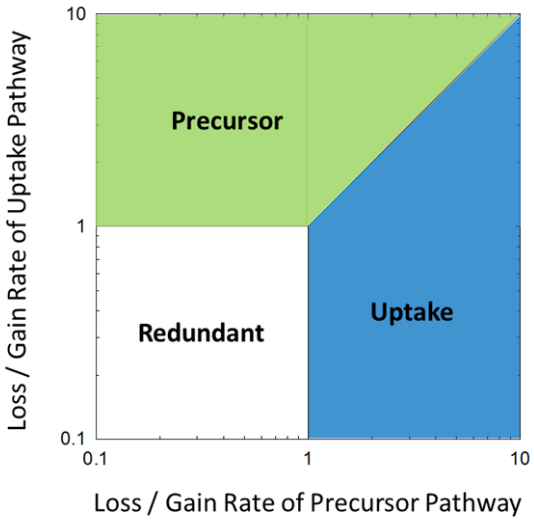


Figure 19.7. The primary pathways of amino-acid biosynthesis. Not all pathways are present in all organisms, and in a number of cases variant pathways leading to the same amino acid exist in different phylogenetic lineages (see main text).

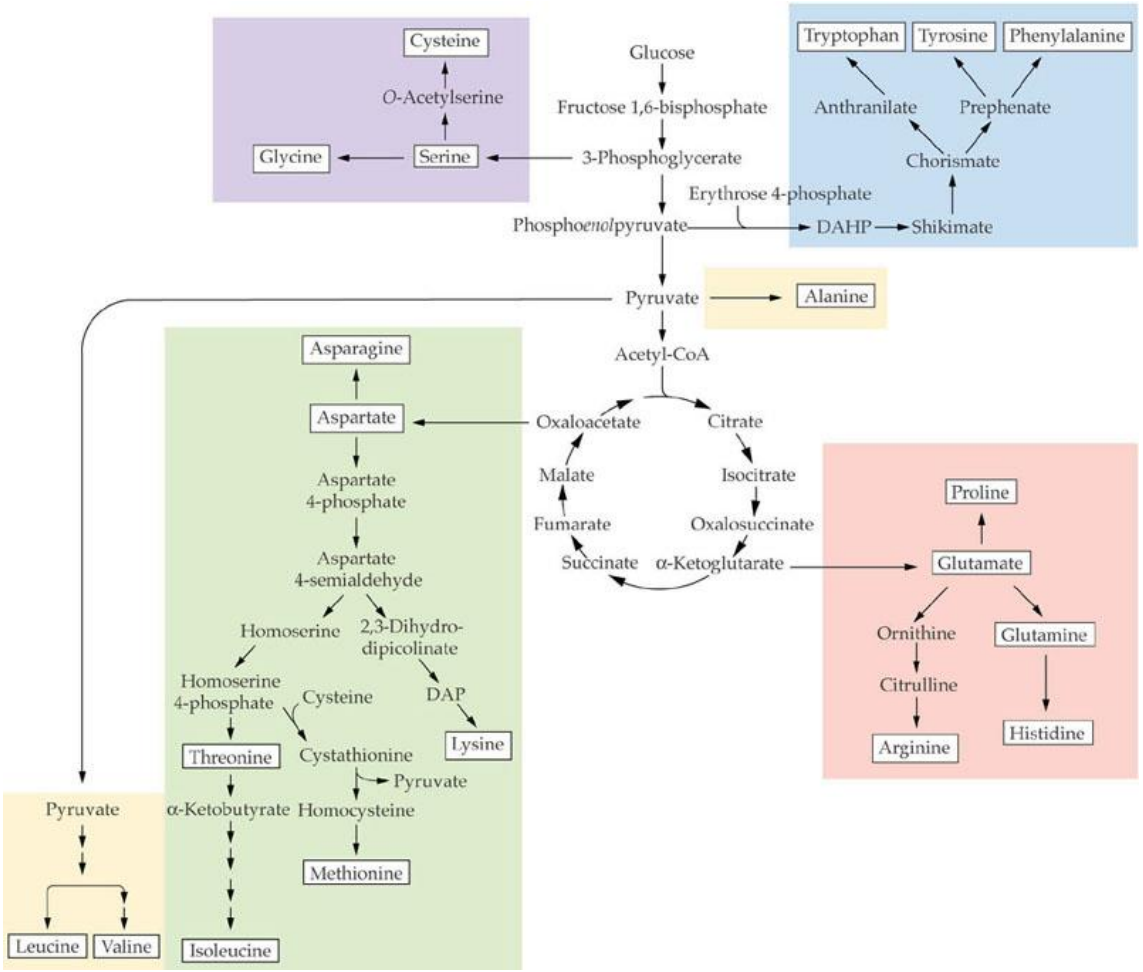


Figure 19.8. Alternative biosynthetic pathways known to exist in different phylogenetic lineages for three amino acids.