Enzymes and Metabolic Pathways

- Enzymes.
  - Basic enzymology
  - Degree of molecular perfection.
  - Enzyme promiscuity.

- Pathway flux control.
  - Pathway position and the strength of selection.
  - Speed vs. efficiency.

- Pathway expansion and contraction.
  - Stochastic meandering of pathway architecture.
  - The origin of novel enzymes.

- Pathway participant remodeling.
  - Nonorthologous gene replacement.

Principles of Metabolic Reactions

- Enzymes do not generally create entirely novel biochemical reactions, but simply enhance catalytic rates, often 10^7 to 10^{19} x background spontaneous rates.

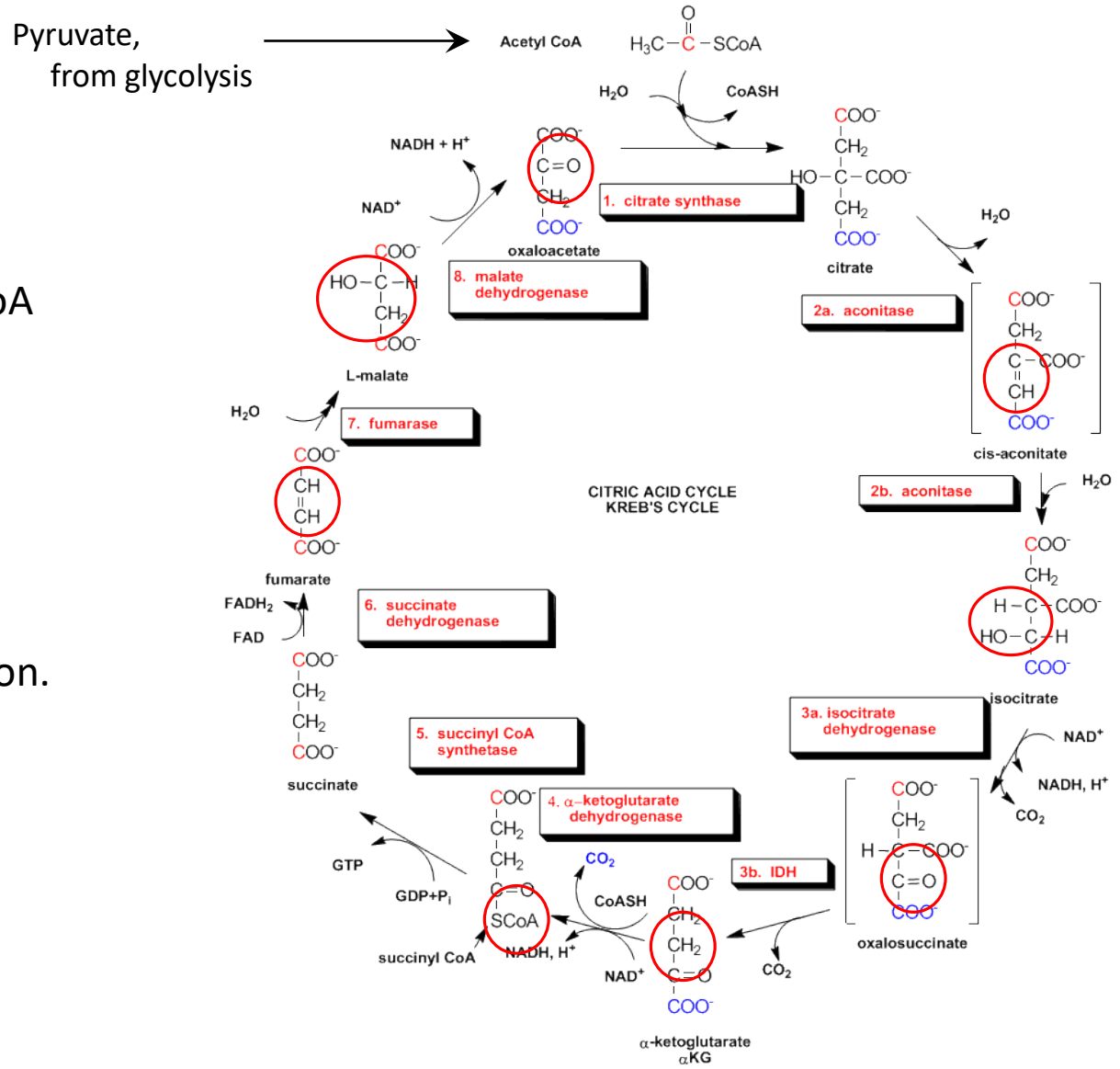
- Enzymes are recyclable – after disengaging with a substrate/product molecule, free to enter into a new reaction.

- Most enzymes carry out very simple reactions, usually engaging with no more than two substrates, and implementing just one or two simple molecular changes per reaction.
  - One hypothetical reason – the diminishingly small probability of joint encounters of multiple molecules.
  - A necessary consequence – the need for multistep pathways for the complete breakdown of organic resources or the synthesis of complex features.

Metabolic Pathways / Cycles Generally Impose Small Molecular Changes at Each Step

The Kreb's (TCA, citric acid) Cycle: joins 2-carbon acetyl CoA from sugar breakdown to 4-carbon oxaloacetate, which it ultimately restores.

- Precursors for amino-acid synthesis are built, CO_2 is released, and energy is stored in the form of NADH for use in the electron-transport pathway for ATP production.



Six Primary Categories of Enzymes

- **Oxidoreductases** (e.g., dehydrogenases and oxidases) transfer hydrogen or oxygen atoms or electrons from one substrate to another.
- **Transferases** (e.g., transaminases and kinases) transfer larger functional groups (such as methyl, amine, or phosphate groups) from one substrate to another.
- **Hydrolases** (e.g., lipases, phosphatases, and peptidases) cleave a chemical bond by addition of water, leading to the breakdown of one substrate into two product molecules.
- **Lyases** (e.g., decarboxylases) break bonds in a nonhydrolytic fashion.
- **Isomerases** lead to molecular changes within a single molecule.
- **Ligases** (e.g., synthetases) join two molecules together with the use of energy derived from ATP.

Key Evolutionary Questions

- How far are levels of enzyme efficiency from the limits of molecular perfection?
- What are the sources of enzymes that enable the emergence of novel metabolic features?
- How do metabolic pathways grow over evolutionary time – from the top down, or from the bottom up?
- To what extent do metabolic pathways in different phylogenetic groups share the same structures and participant contents?

Selection on Enzyme Kinetic Features

$$\mathrm{E} + \mathrm{S} \underset{k_d}{\overset{k_a}{\rightleftharpoons}} \mathrm{ES} \xrightarrow{k_{cat}} \mathrm{E} + \mathrm{P}$$

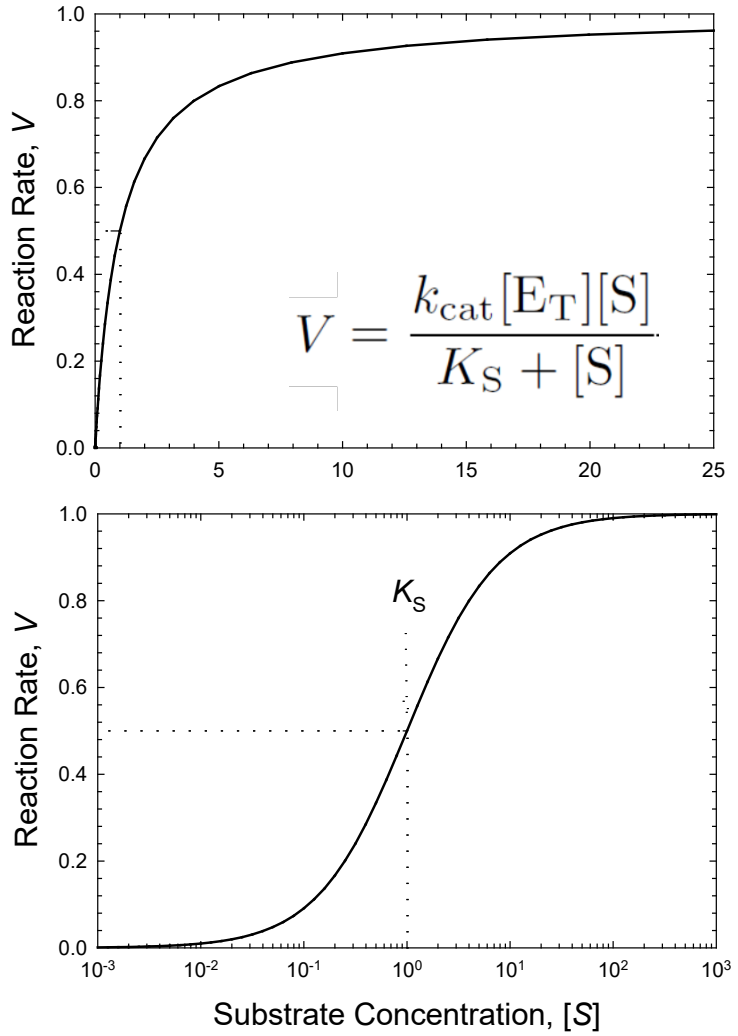
Selection can increase the flux rate of a reaction by:

1) Increasing the forward turnover rate, k_{cat}

2) Increasing the association rate k_a

3) Decreasing the backwards rate of dissociation k_d

- Low substrate concentrations, selection operates on K_{cat}/K_s.
- High substrate concentrations, selection operates on k_{cat}.



How do internal metabolite concentrations compare with K_S?

- 60% of *E. coli* metabolites have concentrations >10x their K_S.

- High substrate concentrations ensure that most enzyme molecules are occupied, reducing the energetic investment in protein production?

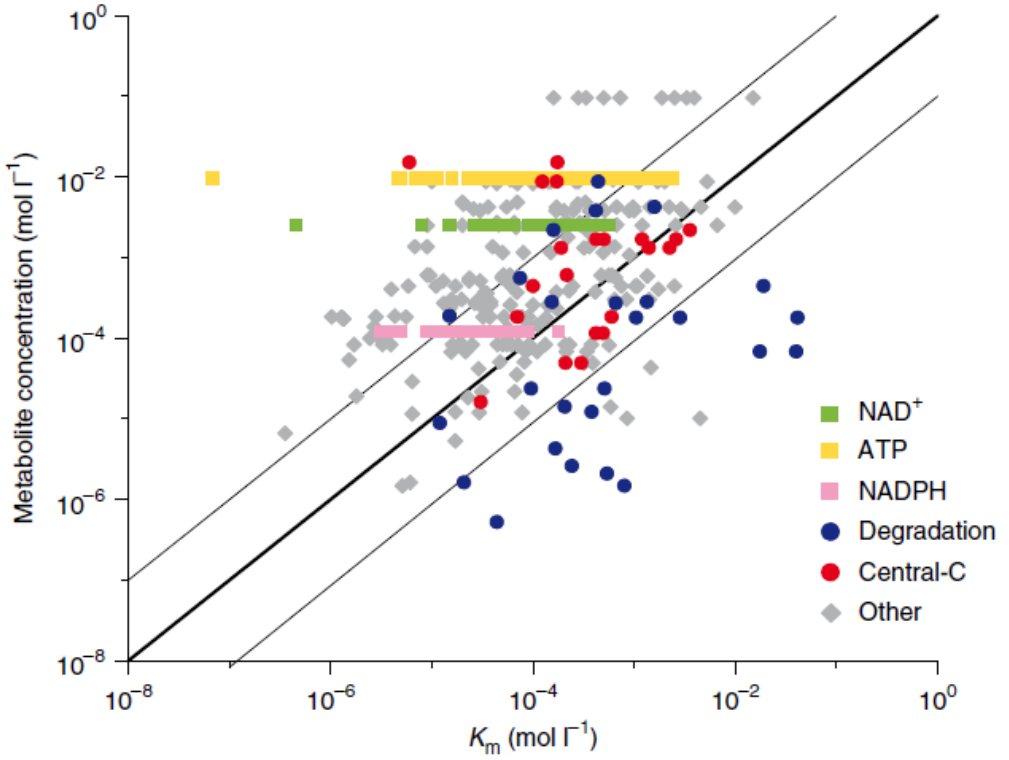


Figure 2 Implied enzyme active site saturation. The relationship of metabolite concentrations and the K_m values of their consuming enzymes in glucose-grown *E. coli*. NAD$^+$ is shown as green squares, ATP as yellow squares, NADPH as pink squares, degradation reactions as blue circles, and reactions in central carbon metabolism (glycolysis, the pentose-phosphate pathway and the tricarboxylic acid cycle) as orange circles. All other data are shown as gray diamonds. The dark line is the line of unity (where concentration = K_m), and the light lines denote a tenfold deviation from the line of unity.

Bennett et al. 2009, Nature Chemical Biology)

Estimating the Biophysical Limits on Enzyme Performance

Rate of enzymatic reaction =

$$\left[\frac{1}{\text{rate of productive enzyme-substrate encounters}} + \frac{1}{\text{rate of chemical transformation}}\right]^{-1} \cdot E \cdot S$$

← function of physical factors and molecular densities

← function of enzyme structure (≈ ∞ for a perfect enzyme, limited only by diffusion)

Rate of productive encounters: $k_{enc} = k_e m_e p_e$

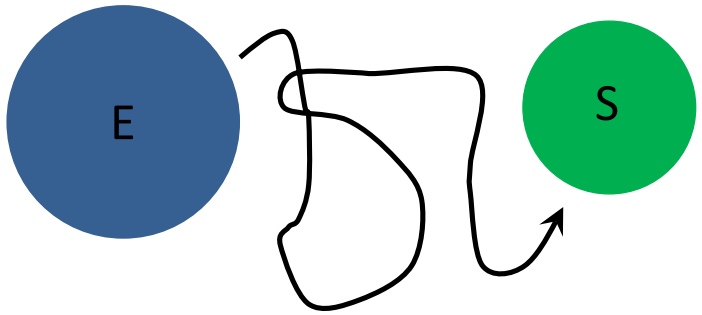
k_e = rate of encounter by diffusion processes

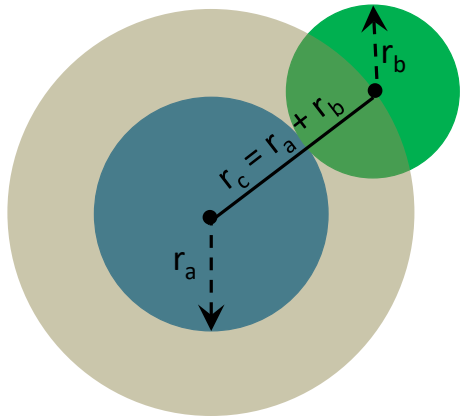
m_e = magnifier resulting from repulsive and/or attractive interactions

p_e = fraction of encounters in the correct orientation for proper engagement

An enzyme has achieved molecular perfection when $k_{enc} \approx k_e$.

The encounter rate by diffusion: $k_e \approx 4\pi(D_a + D_b)(r_a + r_b)$





Sum of diffusion coefficients: $$D = D_a + D_b = \frac{RT(r_a + r_b)}{6\pi\eta(r_a r_b)}$$

R is the Boltzmann constant (1.381×10^{-16} cm$^2 \cdot$ g $\cdot$ sec^{-2}, which relates energy at the particle level to temperature T in degrees Kelvin)

η is the viscosity $\simeq 10^{-2}$ g $\cdot$ cm$^{-1} \cdot$ sec^{-1}

To standardize to molar concentrations of solute, we further multiply by 6.02×10^{20} molecules/M/cm^3 (one thousandth of **Avogadro's number**),

$$k_e = \left(\frac{2RT}{3\eta}\right)\left(\frac{(r_a + r_b)^2}{r_a r_b}\right) = (1.66 \times 10^9)\left(\frac{(r_a + r_b)^2}{r_a r_b}\right)$$

Because enzymes are much larger than their substrates, $r_E >> r_S$:

$$k_e \simeq (1.66 \times 10^9)\left(\frac{r_E}{r_S}\right) \qquad \text{with units } \mathrm{M}^{-1}\cdot\ \mathrm{sec}^{-1}$$

How large are enzymes and their substrates?

amino acids = 75 to 204 g/mol
glucose = 180 g/mol
NADH = 663 g/mol
nucleotides = 111 to 523 g/mol

⇨

Assuming ~300 AAs/ protein, and mass ~r^3,

$r_E/r_S \approx 3.5$ to 10.0

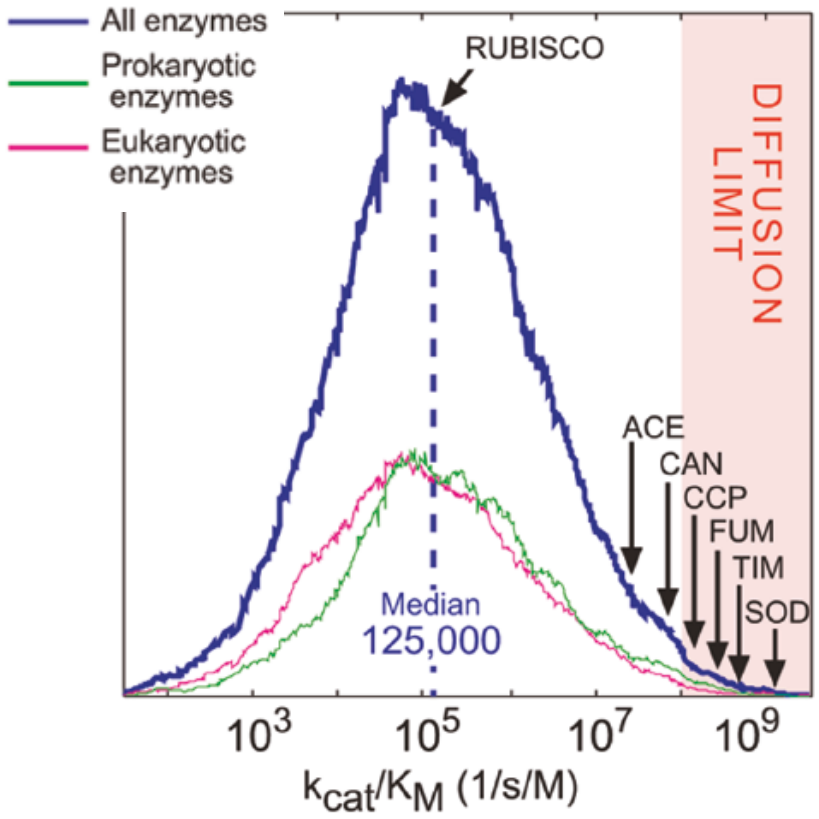
Substituting into the above, we obtain the diffusion limit to a molecular reaction:

$$k_e \approx 10^{10}\ \mathrm{M}^{-1} \cdot \mathrm{sec}^{-1}$$

Note that at low substrate concentrations, the reaction rate is k_{cat}/k_S.

How does this compare with k_e?

Enzyme Efficiencies Tend to be Orders of Magnitude Below the Diffusion Limit



- With an average enzyme efficiency of ~10^5 sec^{-1} M^{-1}, relative to the diffusion limit of ~10^{10} sec^{-1} M^{-1}, only ~1 in 100,000 collisions between an average enzyme and substrate results in a productive interaction.

- Prokaryotic enzymes appear to have slightly better kinetics than those from eukaryotes.

Distribution of enzyme kinetic parameters from the literature (Brenda database; Bar-Even et al. 2011)

Figure 1. Distributions of kinetic parameters: (A) k_{cat} values ($N = 1942$), (B) k_{cat}/K_M values ($N = 1882$), and (C) K_M values ($N = 5194$). Only values referring to natural substrates were included in the distributions (Supporting Information). Green and magenta lines correspond to the distributions of the kinetic values of prokaryotic and eukaryotic enzymes, respectively. The location of several well-studied enzymes is highlighted: ACE, acetylcholine esterase; CAN, carbonic anhydrase; CCP, cytochrome *c* peroxidase; FUM, fumarase; Rubisco, ribulose-1,5-bisphosphate carboxylase oxygenase; SOD, superoxide dismutase; TIM, triosephosphate isomerase.

Enzymes Involved in Secondary Metabolism are on Average Slower Than Those in Central Metabolism

Central CE = carbohydrate, energy metabolism

Central AFN = amino acid, nucleotide, fatty acid metabolism

Intermediate = biosynthesis and degradation of common cellular components, cofactors, and coenzymes

Secondary = metabolites produced in specific cells

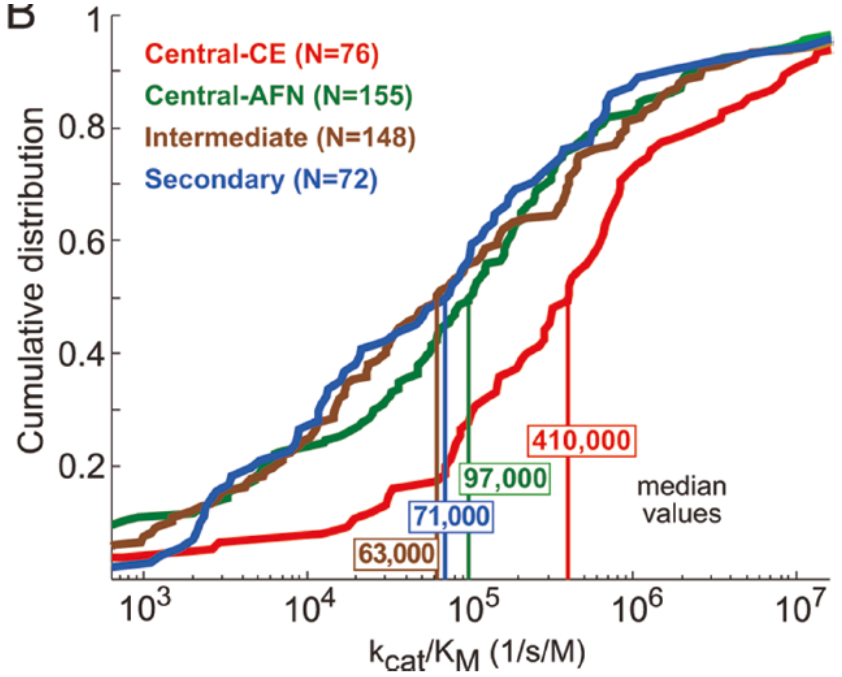


Figure 2. Enzymes operating within different metabolic groups have significantly different k_{cat} and k_{cat}/K_M values. (A) Distribution of k_{cat} values for enzyme–substrate pairs belonging to different metabolic contexts. All distributions are significantly different with a p value of <0.0005 (rank-sum test), except for intermediate versus secondary metabolisms ($p < 0.05$). (B) Distribution of k_{cat}/K_M values for enzyme–substrate pairs belonging to different metabolic contexts. Central-CE (carbohydrate and energy) metabolism has significantly higher k_{cat}/K_M values than all other metabolic groups [$p < 0.0005$ (rank-sum test)]. Abbreviations: CE, carbon and energy; AFN, amino acids, fatty acids, and nucleotides. Numbers in parentheses represent the numbers of enzyme–substrate pairs included in each set.

A Drift-Barrier Hypothesis for Enzyme Efficiency (Hartl et al. 1985)

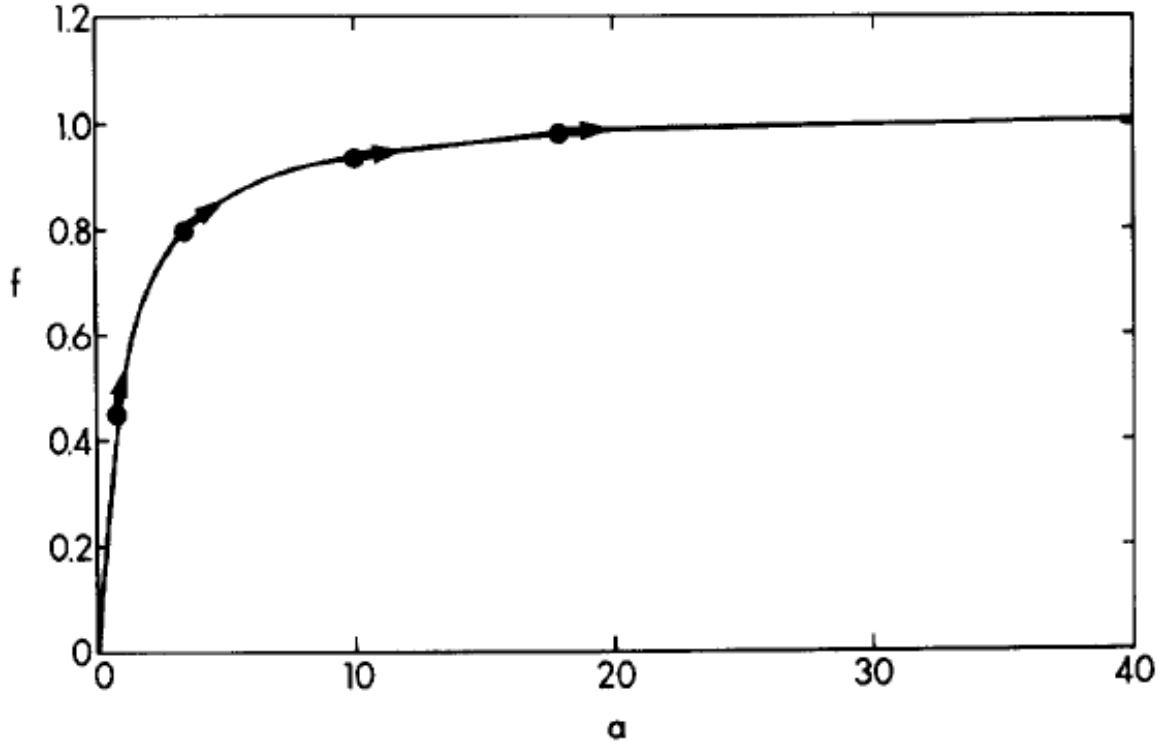


FIGURE 1.—Standardized Michaelis-Menten saturation equation $f(a) = a/(1 + a)$ scaled to equal 1 at $a = 30$.

Diminishing returns once the flux rate is out on the plateau.

- Proposed as an example of how natural selection can lead to an evolutionary scenario of selective neutrality.

Enzyme Promiscuity

- In 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).
  - For ~250 enzymes in *E. coli*, the average error rate is 0.0045 (Notebaart et al.\ 2014).

- The ubiquity of enzyme promiscuity implies that when confronted with new biochemical challenges, adaptation need not await the mutational origin of new functions from scratch.
  - Despite the relatively low kinetic rates of promiscuous interactions, they are still generally orders of magnitude greater than rates of noncatalyzed reactions, providing a well-endowed starting point for adaptive exploitation.

Promiscuous Enzyme Functions Are Common

- ~20% of single-gene knockouts in *E. coli* can be rescued by overexpression of at least one non-cognate gene (Patrick et al., 2007, Mol. Biol. Evol.)

Table 1
Multicopy Suppressors Rescue 21 Conditional Auxotrophs, via 8 Proposed Mechanisms

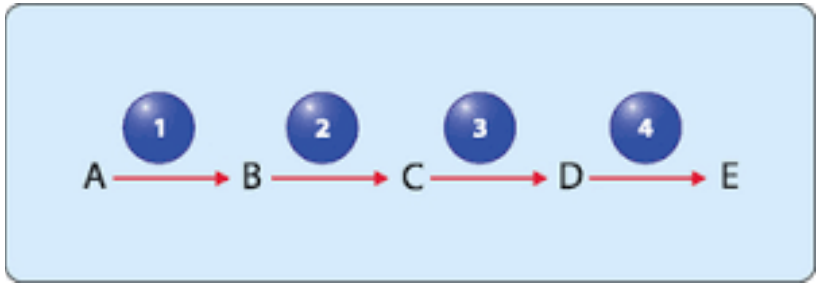
Strain	Protein Deficiency	MS	Rescuing Gene Product
Isozyme overexpression			
Δ*glyA*	Glycine hydroxymethyltransferase	*ltaE*	L-*allo*-threonine aldolase
Δ*ilvA*	Threonine dehydratase (biosynthetic)	*tdcB*	Threonine dehydratase (catabolic)
Δ*ilvE*	Branched-chain amino acid transferase	*avtA*	Valine-pyruvate AT
Δ*metC*	Cystathionine β-lyase/cysteine desulfhydrase	*malY*	Cystathionine β-lyase/maltose regulon repressor
Substrate ambiguity			
Δ*glnA*	Glutamine synthetase	*asnB*	Asparagine synthetase
Δ*pdxB*	Erythronate-4-phosphate dehydrogenase	*tdh*	Threonine dehydrogenase
Δ*serB*	Phosphoserine phosphatase	*gph*	Phosphoglycolate phosphatase
Δ*serB*	Phosphoserine phosphatase	*hisB*	IGP dehydratase/histidinol phosphate phosphatase
Δ*serB*	Phosphoserine phosphatase	*ytjC*	Predicted bisphosphatase; uncharacterized
Ambiguity in metabolite transport			
Δ*fes*	Enterochelin esterase	*setB*	Predicted lactose/IPTG efflux transporter
Δ*ilvA*	Threonine dehydratase (biosynthetic)	*emrD*	Multidrug efflux transporter
Δ*ppc*	Phosphoenolpyruvate carboxylase	*ecfM*	Predicted extracytoplasmic factor; uncharacterized
Δ*ppc*	Phosphoenolpyruvate carboxylase	*yccT*	Predicted outer membrane protein; uncharacterized
Δ*ptsI*	Carbohydrate PTS system, enzyme I	*fucP*	Fucose transporter
Δ*ptsI*	Carbohydrate PTS system, enzyme I	*xylE*	Xylose transporter
Catalytic promiscuity			
Δ*fes*	Enterochelin esterase	*thiL*	Thiamine monophosphate kinase
Δ*metC*	Cystathionine β-lyase/cysteine desulfhydrase	*alr*	Alanine racemase
Δ*pabA*	Aminodeoxychorismate synthase, small subunit	*menF*	Isochorismate synthase
Δ*pabB*	Aminodeoxychorismate synthase, large subunit	*menF*	Isochorismate synthase
Δ*pdxB*	Erythronate-4-phosphate dehydrogenase	*purF*	Glutamine PRPP amidotransferase
Δ*purK*	Carboxyaminoimidazole ribonucleotide synthetase	*anmK*	Anhydro-*N*-acetylmuramic acid kinase

Evolution of Pathway Flux Control

- Pathway position and the strength of selection.

- Speed vs. efficiency.

Pathway Position of an Enzyme and Influence on the Total Flux Rate



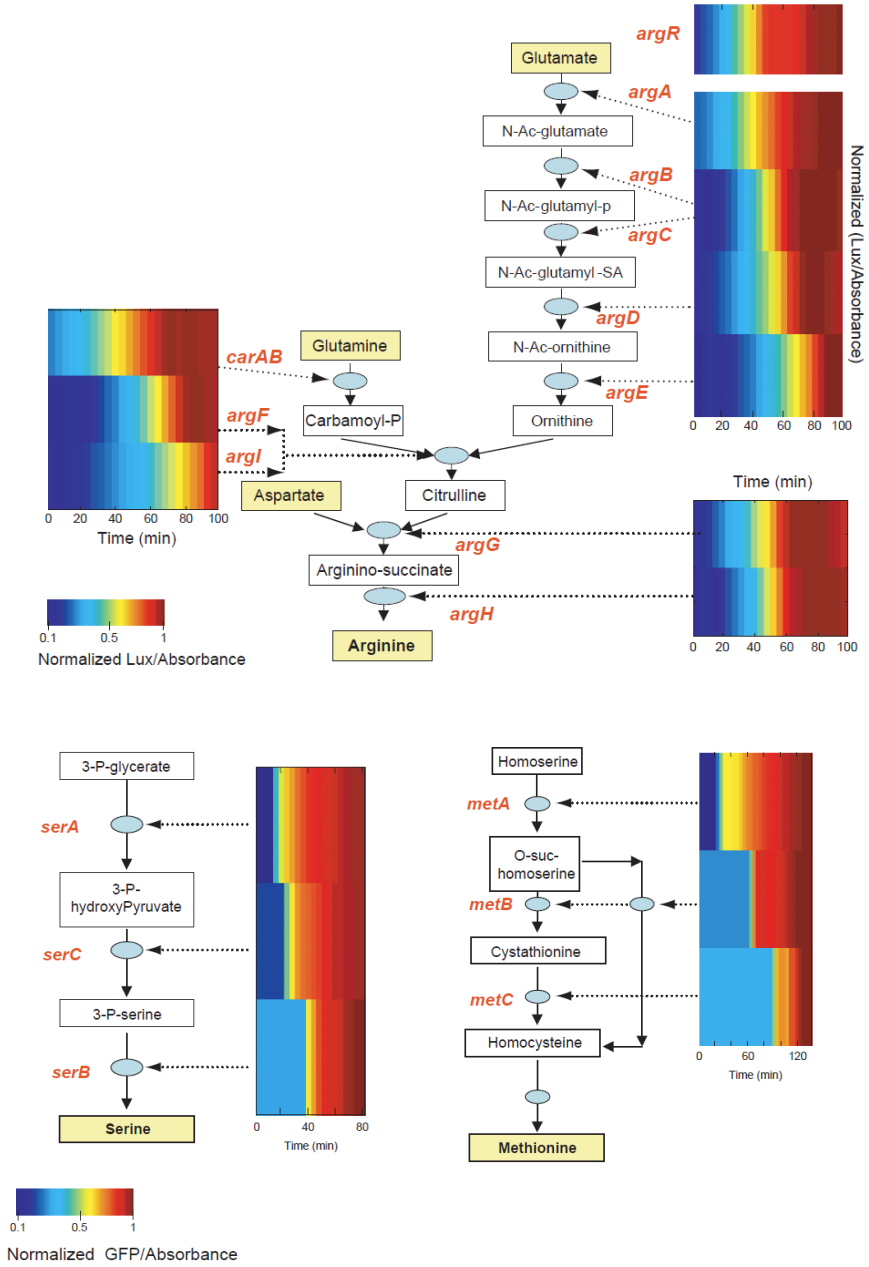
- The total flux rate through a linear pathway is a function of the enzyme kinetics operating at each step, including both the forward and reverse reaction rates between all pairs of adjacent metabolites.

- Mathematical analysis leads to the prediction that the sensitivity of the overall flux rate declines dramatically (exponentially) with increasing downsteam position of enzymes in the pathway.

- If this hypothesis is correct, enzymes higher up in a pathway should exhibit signs of stronger selection for the level of molecular refinement.

Enzymes Higher in Unbranched Pathways Evolve Shorter Response Times and Higher Maximum Promoter Activities

Figure 3 Promoter activity profiles of the nine arginine biosynthesis operons. Lux reporter strains were grown in defined medium containing all amino acids and diluted into the same medium lacking arginine. Shown is the ratio of Lux activity to absorbance at each time point divided by the highest ratio for each strain. Blue and red indicate low and high expression, respectively. The pathway diagram illustrates the metabolic steps of the arginine biosynthesis pathway and the dashed arrows indicate the stages at which each gene product participates. Similar results were obtained using GFP reporter strains.



Zaslaver et al. 2004, Nature Genetics

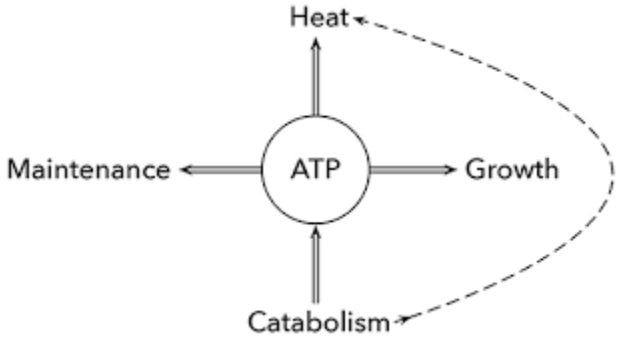
Are enzymes at higher levels in a pathway under stronger selective constraint?

- In the anthocyanin pathway of plants, upstream enzymes evolve more slowly than those downstream, putatively because the former are above branchpoints in networks (Rausher et al. 1999; Lu and Rausher 2003).

- Genes with more interacting partners in yeast and mammalian cells evolve more slowly than those with fewer interactors, regardless of their actual positions in pathways (Fraser et al. 2002; Vitkup et al. 2006; Yang et al. 2009; Montanucci et al. 2011).

Why is Metabolism So Inefficient?

- 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.



  - Glucose to lactic acid reactions release ~205 kJoules / mol.
  - Each ADP-to-ATP conversion stores ~50 kJoules / mol.
  - Efficiency of energy recovery ≈ 50% (from 2 x 50 / 205), the remaining energy being released as heat.

Why is so Much Change Left on the Table? Speed vs. Efficiency

- Up to N = 4 ATPs can be squeezed out of a glucose molecule undergoing glycolysis.
- Evolution of the pathway has led to N = 2 ATPs.

- Flux Rate for Energy Storage in ATP is proportional to: Efficiency x Overall Speed of Conversion to ATP

Efficiency = (N x energy stored in ATP) / energy released by glycolysis = N x 50 /205.

Energy gradient = Energy released by glycolysis - Energy stored in ATP = 205 – (N x 50).

Product is maximized when N = 2.05.

Digital Nature of Metabolism

N	Relative Rate of Energy Gain
1	0.74
2	1.00
3	0.79
4	0.10

Pathway Expansion and Contraction

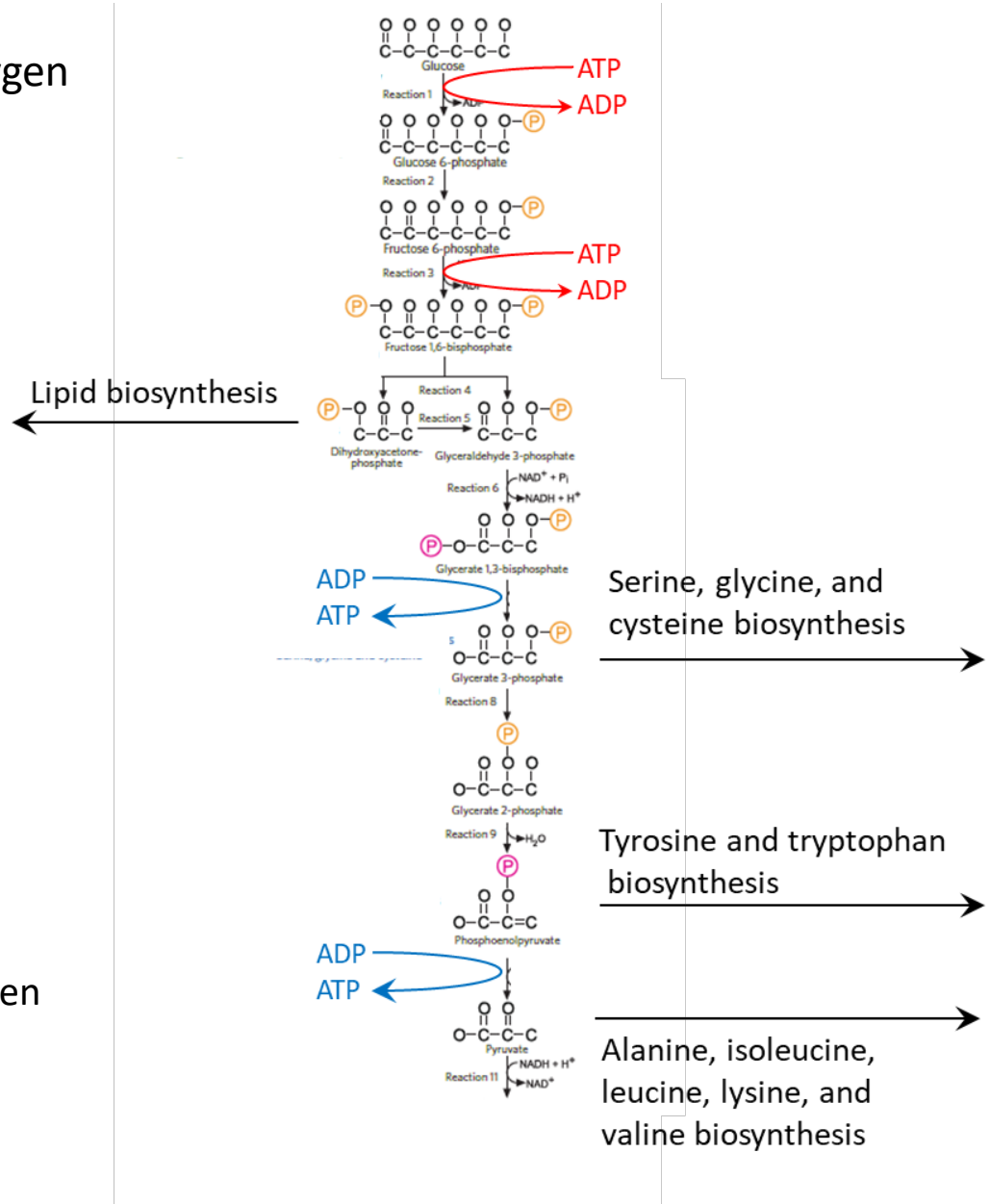
- Evolution of pathway architecture.
- The origin of novel enzymes.

Large Numbers of Incremental Steps Often Used to Accomplish Superficially Simple Molecular Alterations

- Glycolysis uses 10 separate enzymes to convert one 6-carbon/6-oxygen glucose molecule to two 3-carbon/3-oxygen pyruvates.

- Many enzymes produce intermediate metabolites whose sole role is to be passed on to another type of enzyme for further processing.

- 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 (Bar-Even et al. 2012).
  - Given the evolved structure of biochemistry and the intermediate metabolites that must be relied upon, there is no shorter pathway between glucose and pyruvate.



The Paradox of Pathway Precursors

- How can the steps leading to a precursor product (with no other function) become established in a pathway?

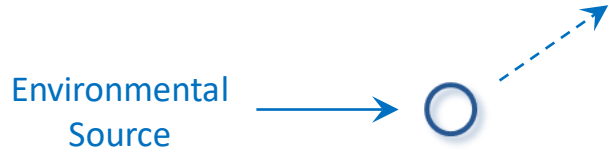
- Horowitz's (1965) retrograde evolution hypothesis – upstream steps in biosynthetic pathways are added as the environmental availability of downstream metabolites become limited – **a bottom-up view**.
  - The bottom step in a pathway was the first to be acquired, with the preceding step having been acquired second, and so on.

- Jensen's (1976) substrate ambiguity hypothesis – a multifunctional enzyme might evolve by duplication into the next enzyme by maintaining the ligand-binding residues but altering the catalytic residues – **a top-down view.**
  - A key feature of biochemical pathways is that consecutive enzymes must interact with the same ligand – the product of one enzymatic reaction is the substrate for the next.

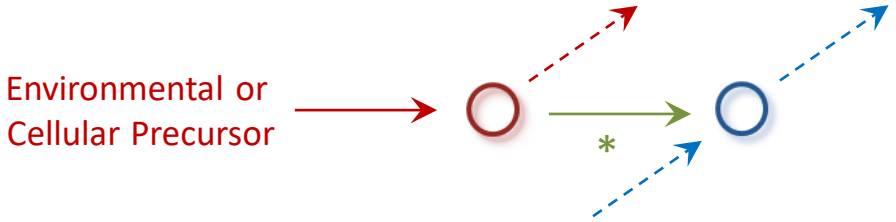
  Another potential example of **subfunctionalization**.

Pathway Growth Topologies: Why and how does the structure of biochemical pathways evolve?

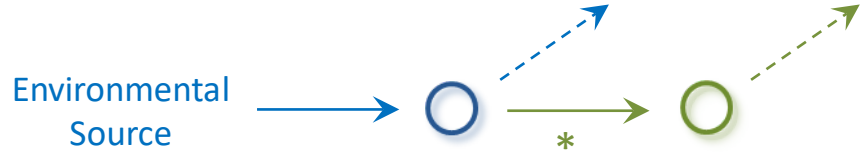
Initial state:



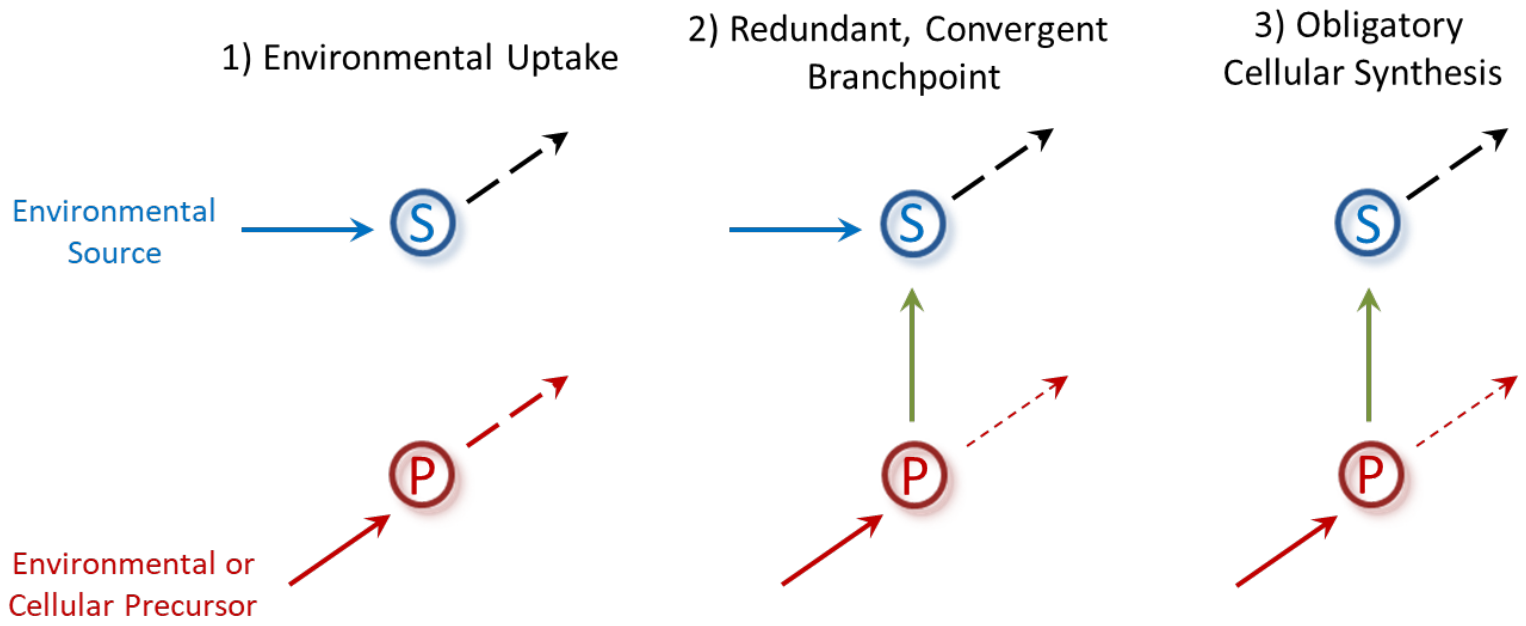
Incorporation of an upstream metabolite:



Extension to produce a novel downstream metabolite:

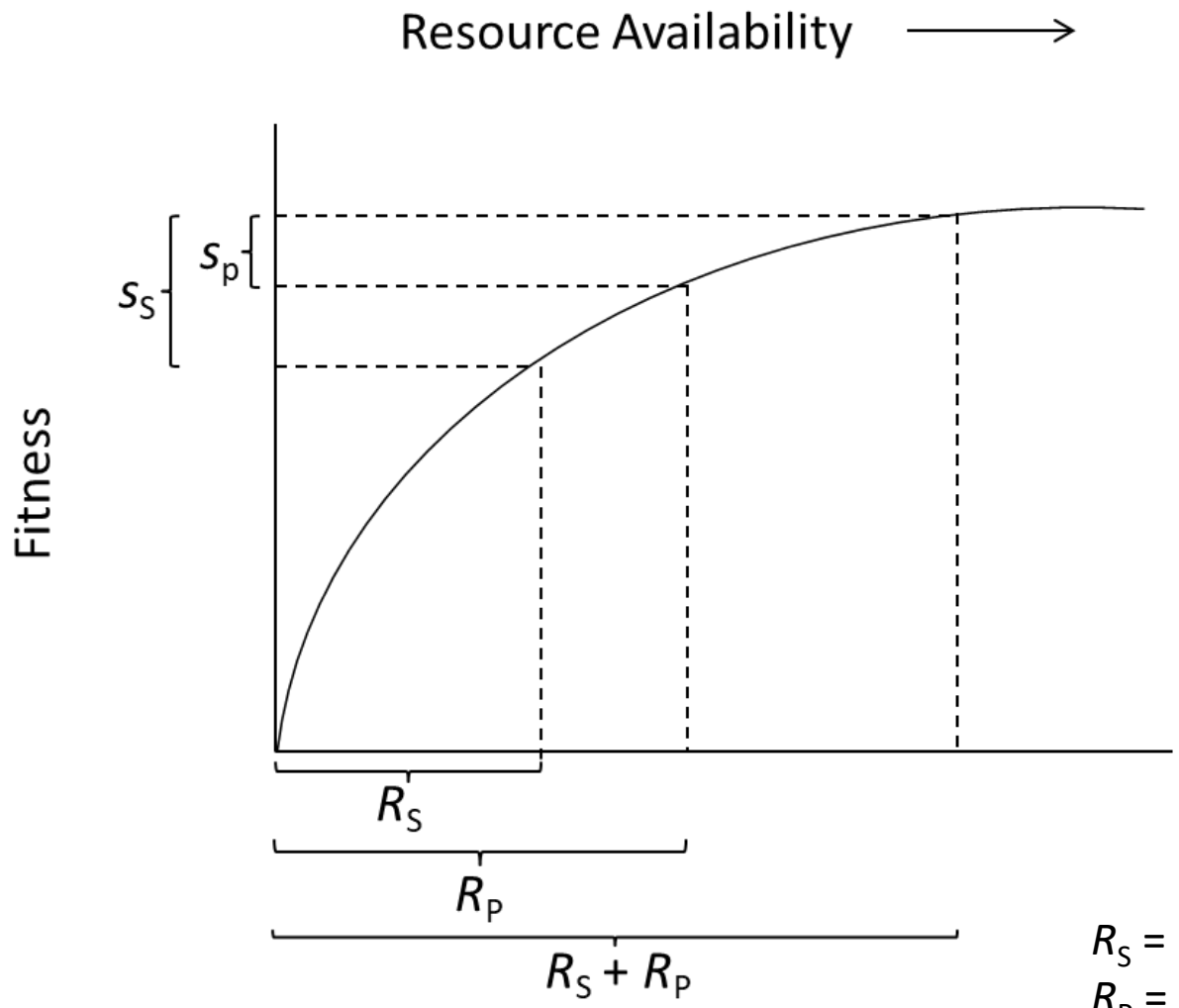


Alternative Pathway Topologies for Acquisition of a Key Metabolite S



- Population-genetic and ecological conditions conducive to the evolution of the three alternative pathways:
  - the relative selective advantages of the alternative pathways;
  - the relative rates of origin and loss of pathway links via mutational mechanisms.

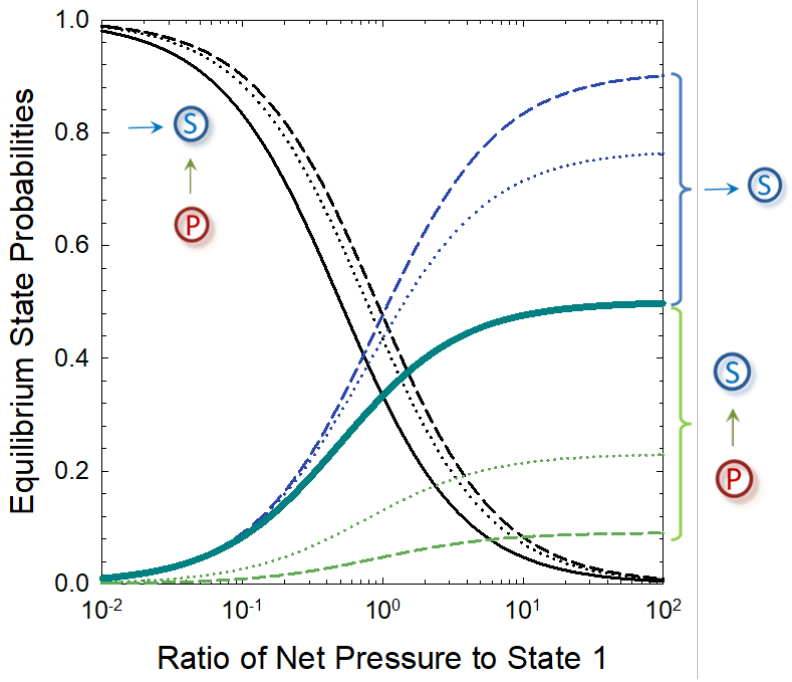
Advantages of Alternative Modes of Resource Acquisition Depend on the Location on the Fitness Function

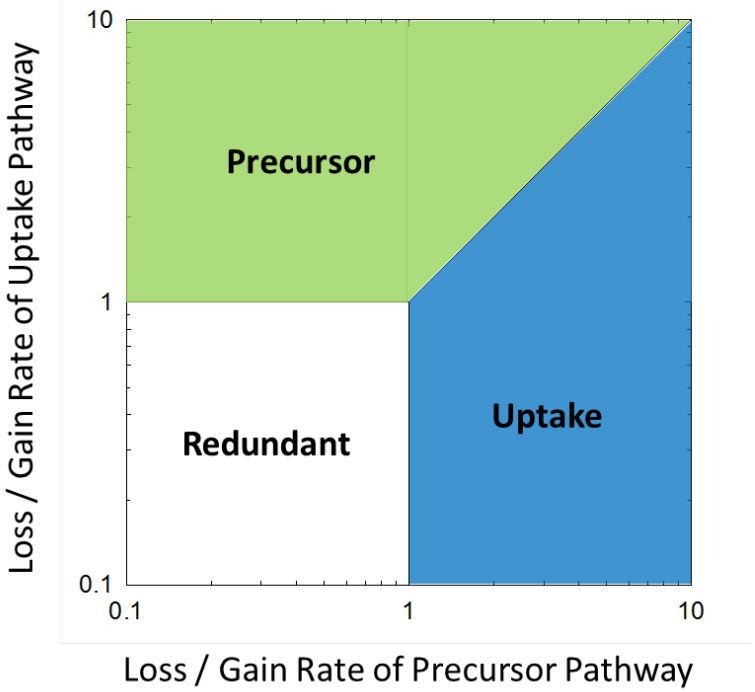


- The redundant pathway has the greatest access to resources.
- Alternative pathways will have different construction costs.
- Alternative pathways will have different mutation vulnerabilities.

R_S = resource from substrate
R_P = resource from precursor
$R_S + R_P$ = redundant pathway

Long-term Steady-State Probabilities of Alternative Pathway States





- Because degenerative mutations are more common than gain-of-function mutations, populations will generally reside in nonredundant states unless there is a strong advantage of redundancy.

The Evolution of a New Function May Often Proceed with Little Effect on the Native Function

Starting enzyme, PON1 = phosphotriesterase, catalyzes the hydrolysis of lactones and a range of other substrates with no known physiological relevance.

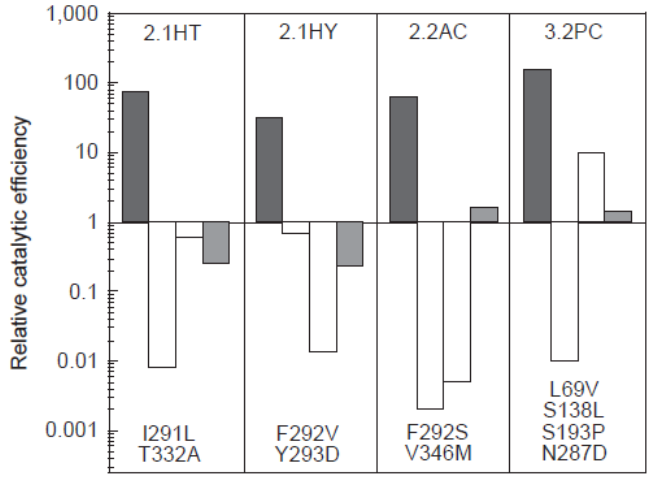


Figure 1 Changes in activities of the newly evolved PON1 variants. The ratios of k_{cat}/K_M values relative to wild-type PON1 for four evolved variants and the mutations observed in them are plotted. Black bars, the promiscuous substrate for which each variant was evolved (2.1HT was evolved for γ-butyryl thiolactone, 2.1HY for 2-naphtyl octanoate, 2.2AC for *O*-acetoxy-7-hydroxycoumarin and 3.2PC for the organophosphate 7-*O*-diethylphosphoryl-3-cyano-7-hydroxycoumarin); white bars, two other promiscuous substrates (phenyl acetate, left; paraoxon, right); gray bars, dihydrocoumarin representing the native lactonase activity of PON1. The

- Selection for a new enzyme function using mutagenic PCR followed by cloning and selective challenge in *E. coli*.
- Depending on the substrate selected for, the native function (grey bars) may decline somewhat or even increase, while the affinity for other prior promiscuous functions (white bars) often declines.
- The end result is a remodeled protein with altered sets of functions.

Aharoni et al. (2005, Nature Genetics)

The Origin of Novel Protein Functions via Promiscuous Interactions

- Using the ASKA (A Complete Set of *E. coli* Archive) library, which consists of every *E. coli* ORF on a plasmid, the authors challenged *E. coli* cells to 237 novel toxins and antibiotics.
- Overexpression of these genes conferred resistance in 86 cases, and 61 unique genes were found to be involved.
- Conclusion: increased expression improves the promiscuous secondary activity while maintaining the primary function.

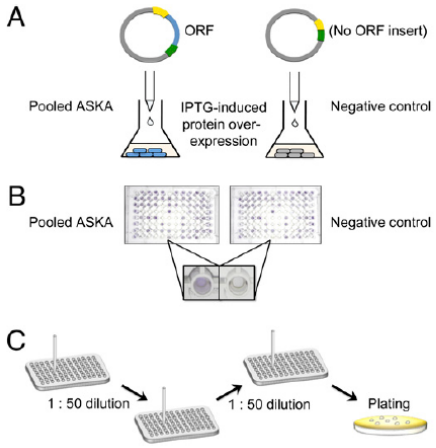


Fig. 1. The library-on-library screen and subsequent isolation of ASKA-encoded resistance genes. (*A*) *E. coli* cells harboring the pooled plasmids of the ASKA ORF collection (blue) and the negative control clone (gray) were grown to midlog phase in parallel, and protein overexpression was induced by the addition of IPTG (50 μM). (*B*) Both cultures were used to inoculate every well of two sets of PM plates. A well was scored positive when the ASKA pool out-grew the negative control, as shown by more rapid tetrazolium color development (*Inset*). (*C*) For each positive PM well, the fittest ASKA clone (or clones) was isolated by two rounds of serial transfer before an aliquot of the enriched culture was plated on nonselective medium to allow identification of the ASKA ORF (or ORFs).

Table 1. Summary of hits obtained from the library-on-library screen

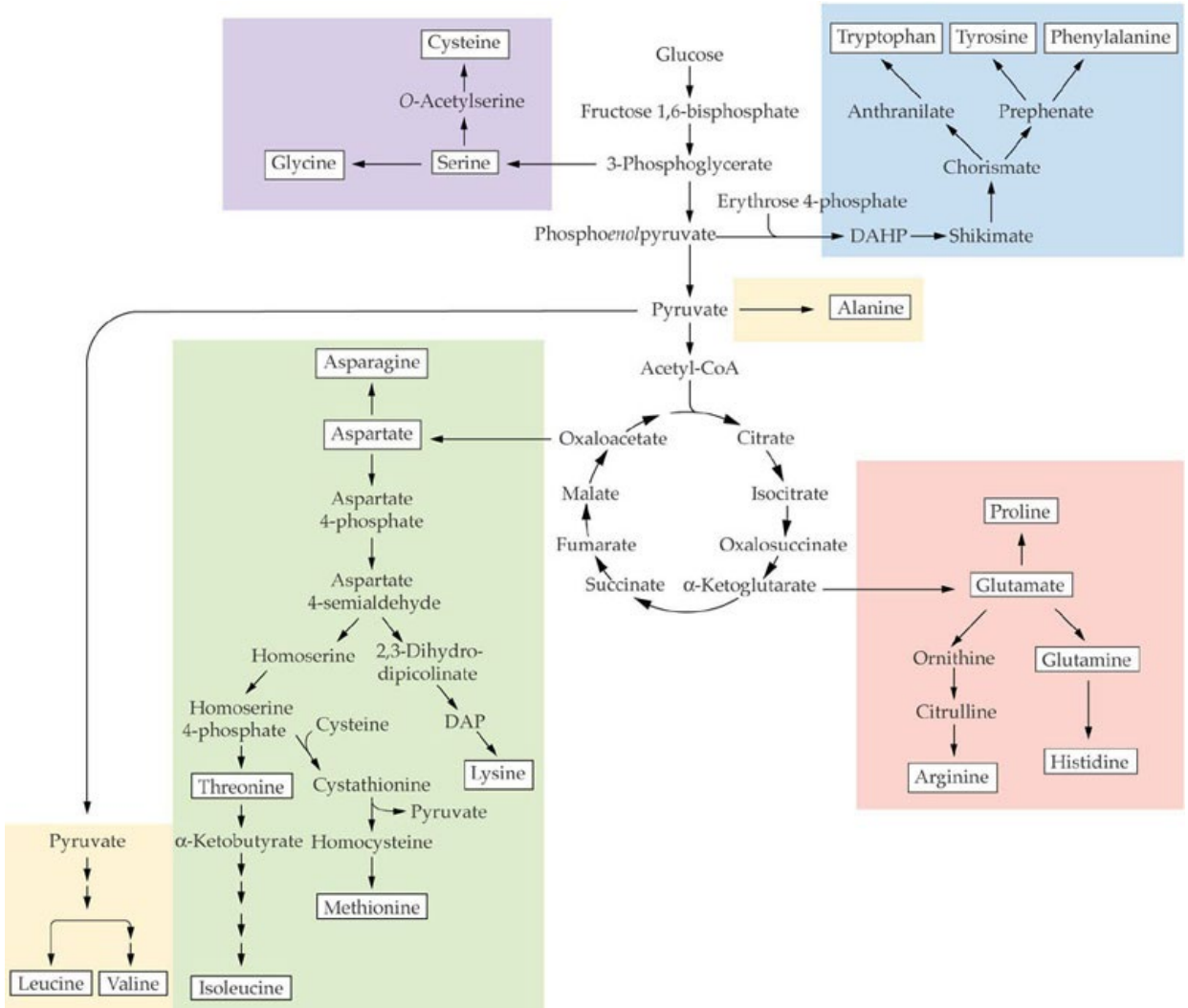
Class of toxic compound	Number of positives*	Number of ASKA hits†
Antibacterials		
β-Lactams	11 (of 22)	14
Aminoglycosides	4 (of 14)	4
Anti-folates	7 (of 9)	10
Quinolones	4 (of 9)	6
Tetracyclines	3 (of 8)	4
MLS antibiotics	6 (of 7)	10
Glycopeptides	1 (of 3)	1
Nitrofurans	3 (of 3)	4
Rifamycins	1 (of 2)	2
Steroid antibacterial	1 (of 1)	1
Other toxins	45 (of 159)	59
Total	86 (of 237)	115

*The number of toxic compounds for which screening and serial enrichment yielded at least one ASKA-encoded resistance gene. The total number of toxins that were screened in each class is listed in parentheses.
†Total number of ASKA ORFs that were isolated from PM wells containing toxins of each compound class.

- Under the IAD (Innovation-Amplification-Divergence) model (Bergthorsson et al. 2007), such promiscuous functions are expected to be refined further as multiple gene copies provide more targets of mutations conferring further improvements. This might ultimately result in a gene entirely specialized to a novel function.

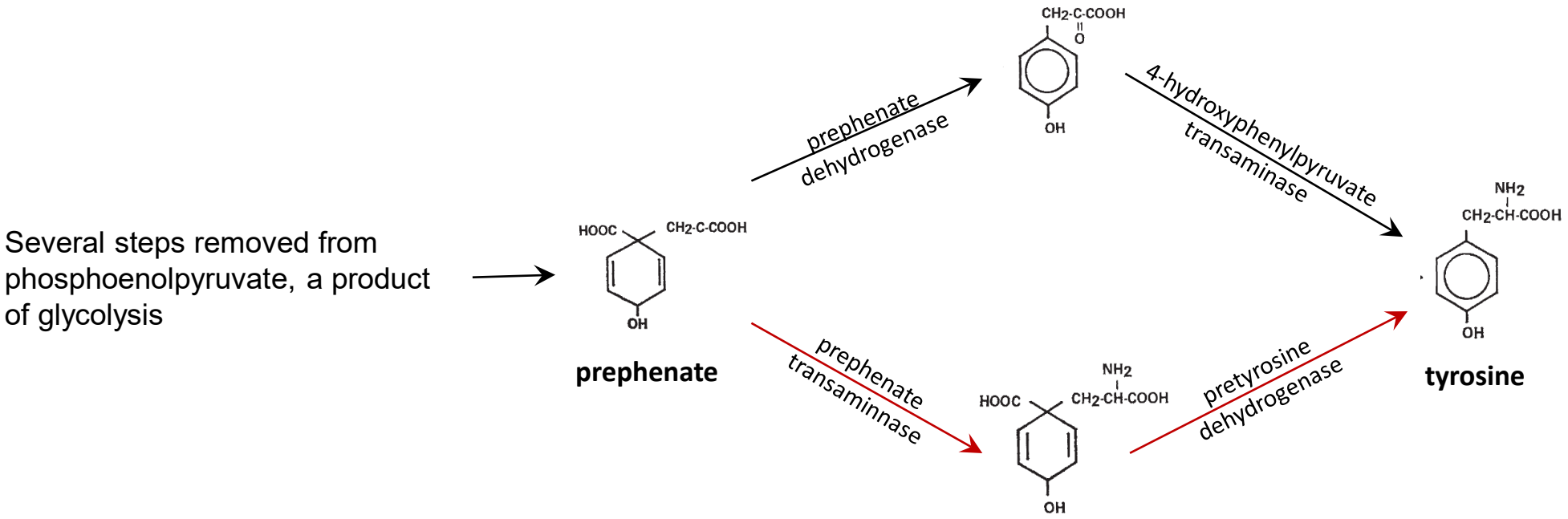
Soo et al. (2011, PNAS)

Pathway Participant Remodeling: variation in the primary pathways of amino-acid biosynthesis.



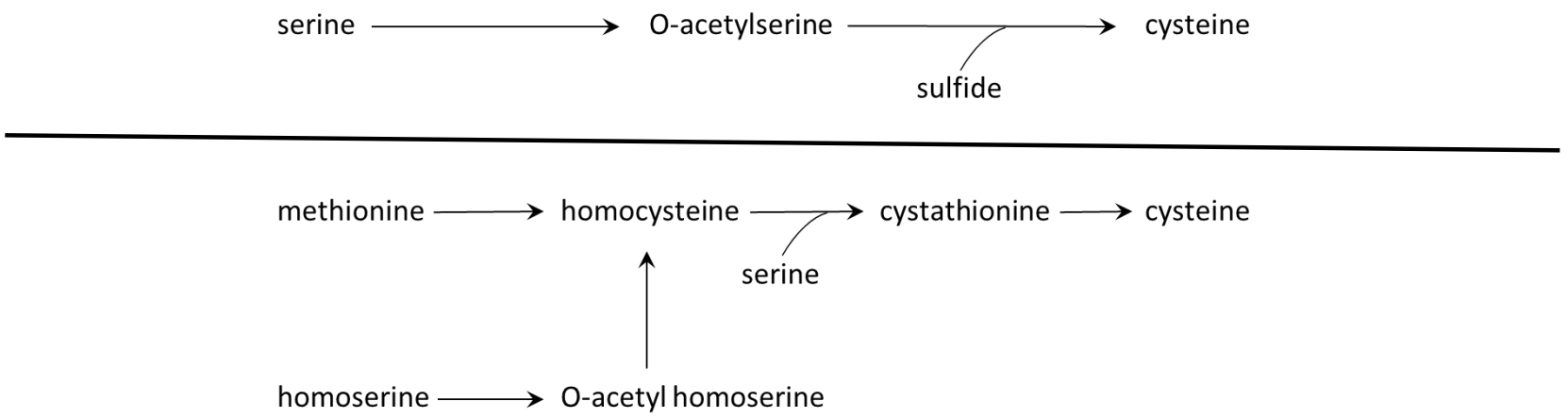
Tyrosine Biosynthesis

- Two pathways are known in bacteria, with transamination and dehydrogenation steps in reverse order, using apparently unrelated enzymes; a few use both.
- Alternative cofactors, NAD+ and NADP+, are also used.



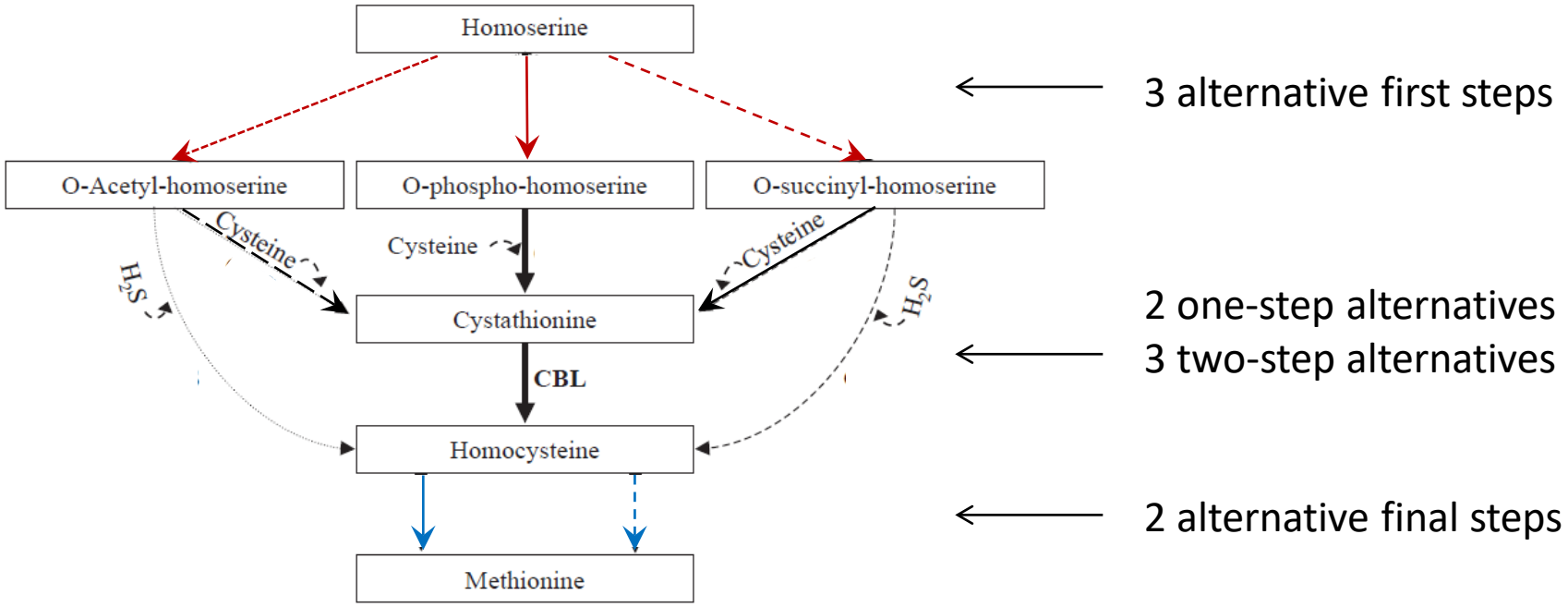
Cysteine is Biosynthesized by at Least Three Non-Overlapping Pathways

- Animals derive sulfur from methionine; plants and prokaryotes from inorganic sulfide.
- In budding yeast *S. cerevisiae*, an animal-like pathway is used, but the methionine → homocysteine step is replaced by two others.
- The fission yeast *S. pombe* uses the pathway initiating with serine.
- Many fungal species have both pathways (Hebert et al. 2011).



Methionine Biosynthesis

- Three alternative first-step metabolites can be produced (acetylated, phosphorylated, or succinylated homoserine).
- Including the alternative downstream steps, ten possible pathways exist, eight of which have been observed in different species (Gophna et al. 2005).
- Some microbes encode for multiple pathways.



Nonorthologous Gene Replacement / Constant Pathway Structure

- Enzymes for nearly every step in glycolysis are known to have unrelated variants in one or more lineages,
  e.g., bacteria vs. archaea vs. animals vs. various fungal and protist lineages.

- Possible pathways:

  - Periods in which an intermediate metabolite is environmentally abundant might lead to loss of a key enzyme (auxotrophy), with subsequent substrate scarcity imposing selection for pathway reconstruction, e.g., replacement of the missing step with a duplication from a nonorthologous gene with a suitably promiscuous function.