

8.02 Evolution and the Enzyme

Shelley D. Copley, University of Colorado at Boulder, Boulder, CO, USA

© 2010 Elsevier Ltd. All rights reserved.

8.02.1	Introduction	9
8.02.2	The Earliest Enzymes: Getting the Basic Chemistry in Place	10
8.02.2.1	Enzymes in the LUCA	10
8.02.2.2	Properties of Early Generalist Enzymes	11
8.02.3	Exploring Fold Space	15
8.02.3.1	Protein Domains in the LUCA and in Extant Life	15
8.02.3.2	Mechanisms for Accessing New Folds	16
8.02.3.3	Novel Associations of Domains and Subunits Provide New Opportunities for Catalysis, Substrate Binding, and Regulation	18
8.02.4	Elaborating on the Basics	20
8.02.4.1	Structural Elaboration via Indels	20
8.02.4.2	The Power of Point Mutations	22
8.02.4.3	Multienzyme Complexes	25
8.02.4.4	Regulation	28
8.02.4.5	Beyond Structure – The Role of Motion	29
8.02.5	Enzyme Evolution in the Context of the Cell	35
8.02.5.1	The Influence of the Environment	35
8.02.5.2	Gene Duplication and Constraints on Enzyme Evolution Imposed by the Need to Maintain an Original Activity	37
8.02.5.3	Topological Effects in Metabolic Networks	40
8.02.6	Enzyme Evolution in the Modern Age	40
References		44

8.02.1 Introduction

Enzymes catalyze almost every metabolic reaction in extant cells. A few unusually facile reactions, such as cyclization of L-glutamate γ -semialdehyde to form pyrroline-5-carboxylate in the proline biosynthesis pathway and decarboxylation of 2-amino-3-oxo-4-phosphonooxybutyrate in the pyridoxal phosphate (PLP) synthesis pathway, do not require acceleration to satisfy the demands of the cell. For all other reactions, catalysis is required because the rates of nonenzymatic reactions are very slow. Modern enzymes are marvelous catalysts. They accelerate reactions by up to 20 orders of magnitude,¹ prevent side reactions of reactive intermediates, and catalyze stereoselective and stereospecific reactions. Further, they are often exquisitely regulated by small molecule ligands.

Enzymes have played a key role in the adaptation of microbes to an extraordinary range of environmental niches and in the competition between species that has led to ever more efficient and well-regulated metabolic systems. The evolutionary processes that shaped the prodigious catalysts of extant organisms are of great interest to evolutionary biologists interested in understanding the origin and diversification of life, and to protein engineers interested in evolving novel catalysts for industrial uses.

Catalysts played an important role in the emergence of life on Earth nearly 4 billion years ago. Catalysis by mineral surfaces and small molecules enabled the emergence of a proto-metabolic network that, in turn, enabled the emergence of the RNA world. The first macromolecular catalysts may have been ribozymes, an idea first proposed by Carl Woese² that gained credence with the discovery of catalytic RNAs by Cech³ and Altman.⁴ Subsequently, ribozymes generated by *in vitro* evolution methods have been shown to catalyze a wide range of reactions involved in metabolism, including amino acid activation;⁵ formation of coenzyme A (CoA), nicotinamide adenine dinucleotide (NAD), and flavin adenine dinucleotide (FAD)

from 4'-phosphopantetheine, nicotinamide mononucleotide (NMN), and flavin mononucleotide (FMN), respectively;⁶ peptide bond synthesis;⁷ and aldol condensation.⁸ It is possible that ribozymes, in conjunction with catalytic auxiliaries such as metal ions, organic cofactors, and peptides, could have catalyzed all of the reactions required to both maintain metabolism and replicate genetic information.

This chapter will focus on catalysts in a later era, after the advent of genetically encoded proteins, and specifically on enzymes involved in metabolism in microbes. The earliest protein enzymes were probably generalists with broad substrate specificity⁹ and consequently rather poor catalytic efficiencies. We can speculate about the properties of the earliest enzymes based upon reconstructions of the proteome of the last universal common ancestor (LUCA) and studies of extant generalist enzymes; the second section of this chapter addresses these issues. The third section describes mechanisms by which the structural diversity of enzymes seen in modern life arose. The fourth section discusses mechanisms by which enzymes with higher catalytic efficiency, stricter substrate specificity, and more sophisticated regulation emerged from inefficient progenitors. The fifth section examines how enzyme evolution is intimately tied to both the physiological conditions within cells and the environmental conditions, which are often changeable, in which microbes exist. The final section describes examples of enzyme evolution in the modern era caused by the introduction of anthropogenic compounds that exert new selective pressures on microbes.

8.02.2 The Earliest Enzymes: Getting the Basic Chemistry in Place

8.02.2.1 Enzymes in the LUCA

Proteins emerged before the LUCA, since rRNA and tRNA genes, as well as the genetic code itself, are conserved in all known forms of life. The emergence of the LUCA marks the point at which vertical transmission of genetic information became possible. Before the LUCA, life may have consisted of communities of proto-organisms that shared metabolites and genetic information.¹⁰ It is difficult to know what the genetic and metabolic capabilities of these proto-organisms were, as we can obtain only a rather fuzzy picture of the LUCA itself. The LUCA clearly had DNA, ribosomes, proteins, and a well-developed metabolic network. Variability in gene structure and strategies for DNA replication in the three kingdoms of life suggests that replication and transcription processes were refined after the LUCA. Further, the structure of the cell wall had not yet been firmly established in the LUCA, as different strategies were ultimately adopted in different kingdoms of life. Bacteria and eukaryotes build their membranes from fatty acids linked to *sn*-glycerol-3-phosphate through an acyl linkage, whereas Archaea build their membranes primarily from isoprenoid lipids linked to *sn*-glycerol-1-phosphate through ether linkages.

Estimating the metabolic capabilities of the LUCA is a formidable problem because gene loss and gene gain, both by horizontal gene transfer and by emergence of new genes via gene duplication and divergence, have been rampant on an evolutionary time scale. **Figure 1** illustrates the difficulty. It is often possible to generate multiple scenarios that account for the presence or absence of genes in various lineages. In some scenarios, an ancestral gene may have been lost in some lineages, while in others, a gene that is not in fact ancestral may have spread by horizontal gene transfer between kingdoms.

Various authors using different data sets, algorithms, and assumptions about the frequencies of gene loss and gene gain have come to differing conclusions about the proteome of the LUCA. The number of universal

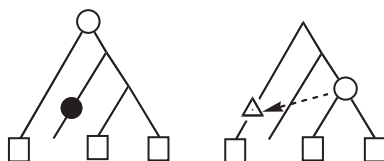


Figure 1 Patchy patterns of gene presence in extant organisms can obscure the evolutionary history of a gene and make it difficult to predict whether the gene was present in the LUCA and was subsequently lost in some lineages (left) or originated later and was transferred to another lineage (right). (○), first appearance of a gene; (●), gene loss; (△), gene acquired by horizontal gene transfer; (□), gene acquired from an ancestor.

protein domains found in all forms of life is estimated to be 190 by Abeln and Deane,¹¹ 219 by Lee *et al.*,¹² and 140 by Ranea *et al.*¹³ Since these universal domains are primarily involved in transcription and translation, it has been suggested that the LUCA had only rudimentary metabolism, and relied upon amino acids and nucleotides supplied by a primordial soup.¹³ This suggestion is at odds with the striking conservation of core metabolic pathways for the synthesis of amino acids, sugars, nucleotides, lipids, and cofactors in all domains of life. The core of metabolism must have been laid down by the LUCA, since it is inconceivable that nearly identical pathways would have emerged independently after the divergence of the three kingdoms of life.

Recently, more sophisticated analyses that deal explicitly with the impact of the frequencies of gene loss and gene gain on the predicted proteome of the LUCA have been developed. The results obtained depend on the topology of the tree and the value chosen for the relative frequency of gene gain and loss, as well as technical details of the algorithms. Thus, even the most sophisticated approaches produce results with some uncertainty. Nevertheless, interesting conclusions emerge. Mirkin *et al.* estimated that the LUCA contained 572 genes.¹⁴ In a later study with a larger dataset, Ouzounis *et al.* estimated that the LUCA contained 669 genes.¹⁵ Genes encoding enzymes for the synthesis of amino acids, nucleotides, sugars, fatty acids, and cofactors are predicted to have been present in the LUCA in both cases. It is quite remarkable that a few hundred enzymes capable of catalyzing reactions in core biosynthetic pathways emerged within the relatively short period between the end of the Late Heavy Bombardment about 3.9 billion years ago¹⁶ and the appearance of life possibly as early as 3.8 billion years ago.¹⁷

8.02.2.2 Properties of Early Generalist Enzymes

The first enzymes were probably generalists that catalyzed similar reactions using a variety of substrates.⁹ A relatively small number of generalist enzymes might have sufficed to support the metabolic network of the first proto-organisms. For example, extant pathways for the synthesis of guanosine monophosphate (GMP) starting from 5-phosphoribosyl-1-pyrophosphate (PRPP) involve 12 steps. The reaction shown in **Figure 2** (activation of a carboxylate via phosphoryl transfer from adenosine triphosphate (ATP), followed by attack of a nucleophilic amine) occurs 5 times. Similar reactions occur in other metabolic pathways, as well. Thus, a single nonspecific catalyst might have accelerated several physiologically significant reactions.

The hypothesis that the earliest enzymes were generalists is supported by the evolutionary history of PLP-dependent enzymes. PLP must have been available in the LUCA, since it is used by enzymes in all kingdoms of life. PLP facilitates a number of transformations of amino acids. Each of these reactions begins with a common step, attack of the amino group of a substrate upon an 'internal' aldimine formed between PLP and a lysine on the protein, resulting in formation of an 'external' aldimine. The cofactor serves as an electron sink for delocalization of electrons during cleavage of one of three bonds in the substrate. The orientation of the substrate determines the subsequent steps.¹⁸ The bond to be cleaved is oriented by interactions with the active site to be orthogonal to the ring of the cofactor, allowing maximal overlap of the highest occupied molecular orbital (HOMO) of the labile bond and the lowest unoccupied molecular orbital (LUMO) of the cofactor. Based on the orientation of the substrate, the result can be transamination, decarboxylation, racemization, β - γ elimination, or β -replacement. Primordial PLP enzymes might have provided little more than a lysine to attach to PLP and a rudimentary binding site for substrates, which would have allowed acceleration of multiple

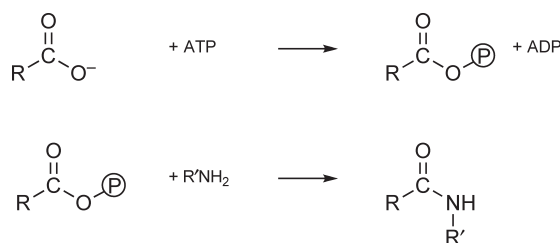


Figure 2 Activation of a carboxylate by formation of an acyl phosphate, followed by attack of an amine to form an amide, occurs 5 times in the 12 steps needed to synthesize GMP starting from PRPP.

reactions using a range of substrates. This hypothesis is consistent with the phylogenetic tree of the α -family of PLP enzymes,¹⁹ which suggests that a nonspecific primordial enzyme diverged into reaction-specific subfamilies, and that substrate specificity emerged later within these subfamilies. Notably, many of these enzymes are found in all domains of life, suggesting that substantial divergence of PLP enzymes had occurred even before the LUCA.

Unfortunately, we cannot access specific sequence information about enzymes from the LUCA to allow resurrection of very ancient enzymes and evaluation of their properties. Phylogenetic reconstruction of ancestral protein sequences does not allow us to look back billions of years; beyond a certain point, there is too much ambiguity due to sequence divergence to allow a trustworthy estimation of the ancestral sequence.

A catalytic antibody (33F12) that catalyzes aldol additions and condensations provides a striking demonstration of the potential for catalysis of multiple reactions by a generalist enzyme.²⁰ A lysine in the active site of the antibody can form an enamine with several different aldehydes and ketones. This enamine can then react with several different aldehyde and ketone acceptors. Because the active site admits so many structurally different substrates, the antibody catalyzes over 100 different reactions with accelerations of 10^5 – 10^7 over the rate of the uncatalyzed reaction (see **Figure 3**). The structural basis for the broad specificity of the antibody was revealed by X-ray crystallography.²⁰ The catalytic lysine resides at the bottom of a deep hydrophobic pocket (see **Figure 4**) that accommodates structurally diverse substrates. Early generalist proteins may have used this strategy, providing binding interactions primarily for functional groups near the site at which chemistry needs to happen, while the rest of the substrate was allowed to protrude into the solvent or into a capacious cavity.

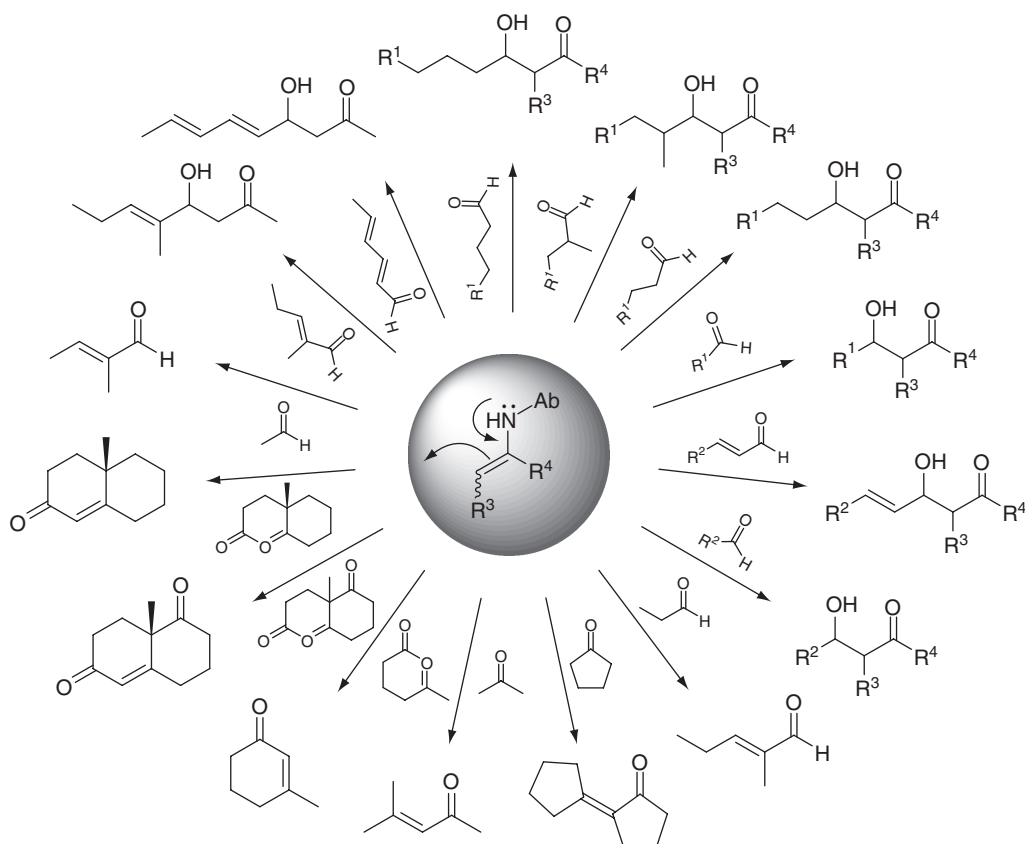


Figure 3 Broad substrate specificity in an antibody aldolase allows catalysis of over 100 different aldol additions and condensations. R¹, 4-acetamidobenzyl; R², 4-nitrobenzyl; R³ and R⁴, various alkyl groups. Reproduced with permission from C. F. Barbas, III; A. Heine; G. Zhong; T. Hoffman; S. Gramatikova; R. Björnstedt; B. List; J. Anderson; E. A. Stura; I. A. Wilson; R. A. Lerner, *Science* **1997**, 278, 2085–2092.