NEW PROSPECTS FOR DEDUCING THE EVOLUTIONARY HISTORY OF METABOLIC PATHWAYS IN PROKARYOTES: AROMATIC BIOSYNTHESIS AS A CASE-IN-POINT[1]

SUHAIL AHMAD and ROY A. JENSEN

Department of Microbiology and Cell Science, 1059 McCarty Hall, Institute of Food and Agricultural Sciences, University of Florida, Gainesville, FL 32611, U.S.A.

(Received 19 November, 1986)

Abstract. Metabolic pathways of prokaryotes are more biochemically diverse than is generally recognized. Distinctive biochemical features are shared by phylogenetic clusters. The hierarchical levels of character-state clustering depends upon evolutionary events which fortuitously became fixed in the genome of a common ancestor. Prokaryotes can now be ordered on a phylogenetic tree. This allows the evolutionary steps that underlie the construction and regulation of appropriately complex biochemical pathways to be traced in an evolutionary progression of prokaryote types that house these pathways. Essentially the approach is to deduce ancestral character states at ever deeper phylogenetic levels, utilizing logical principles of maximum parsimony. The current perspective on the evolution of the biochemical pathway for biosynthesis of aromatic amino acids is developed as a case-in-point model for analyses that should be feasible with many major metabolic systems. Phenylalanine biosynthesis probably arose prior to the addition of branches leading to tyrosine and tryptophan. An evolutionary scenario is developed that begins with non-enzymatic reactions which may have operated in primitive systems, followed by the evolution of an enzymatic system that pre-dated the divergence of major lineages of modern eubacteria (Gram-positive bacteria, Gram-negative purple bacteria, and cyanobacteria).

1. Introduction

The biochemical pathways utilized in nature for the biosynthesis of aromatic amino acids and other important aromatic metabolites display ideal features for evolutionary studies. This complicated, multi-branched pathway presents a diversity of alternative character states that cluster at a given hierarchical level in a phylogenetic grouping (Byng *et al.*, 1982). Unlike genes that encode resistance to environmental agents (such as antibiotics or heavy metals) or that encode unessential catabolic functions, the essential genes of aromatic biosynthesis can be assumed to reflect the 'vertical' evolution of the chromosomal genome (in contrast to possible lateral transfer of non-chromosomal genomes).

In this review we build upon previous speculations about the non-enzymatic origin of key mid-pathway steps (Stenmark *et al.*, 1974; Jensen and Stenmark, 1975; Jensen and Pierson, 1975). Sufficient information has accumulated about the division of Gram-negative bacteria that includes *Escherichia coli* (Superfamily B) to allow reasonable initial deductions about the evolutionary history of aromatic biosynthesis back to a common Superfamily-B ancestor (Jensen, 1985). The level of analysis is now sufficiently advanced that specific genetic events of gene duplication, inversion and fusion can be postulated. Although such scenarios are presently tentative and in-

[1] Florida Agricultural Experiment Station, Journal Series No. 8251.

Origins of Life and Evolution of the Biosphere **18** (1988) 41–57.

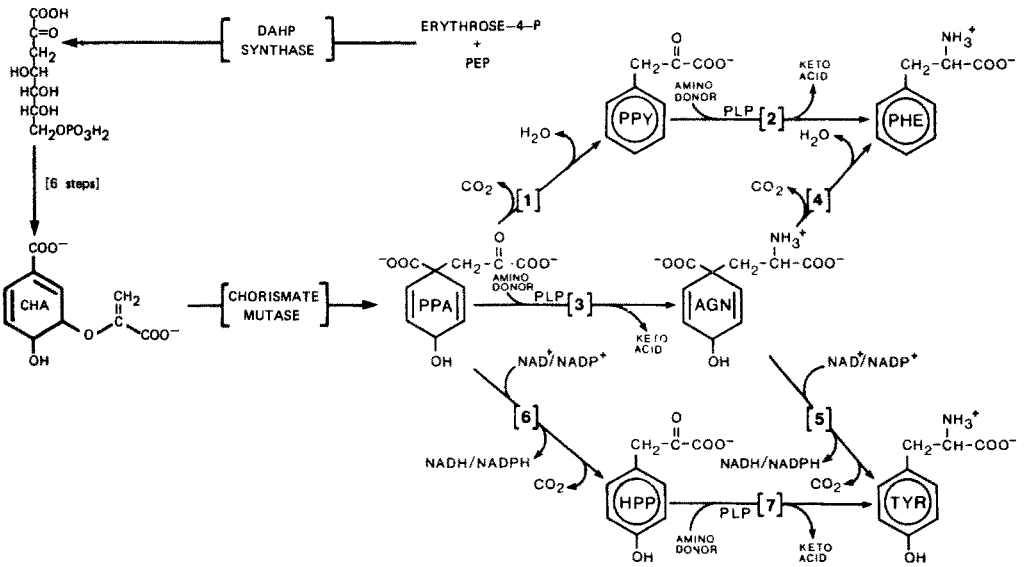

Fig. 1. Biochemical pathway for biosynthesis of aromatic amino acids. Erythrose-4-phosphate and phosphoenolpyruvate arising through central carbohydrate metabolism are condensed to form 3-deoxy-*D-arabino*-heptulosonate 7-phosphate (DAHP) via the action of DAHP synthase. Chorismate (CHA) is the last common precursor of all three aromatic amino acids and of numerous additional endproducts. The post-prephenate pathways to phenylalanine and tyrosine, where considerable biochemical diversity has been found in nature, are shown as a composite scheme. Virtually every conceivable combination of the two dual-pathway arrangements shown have been delineated in a variety of microorganisms and plants. Other abbreviations: PPA, prephenate; AGN, *L*-arogenate; PPY, phenylpyruvate; HPP, 4-hydroxyphenylpyruvate; PHE, *L*-phenylalanine; TYR, *L*-tyrosine; PLP, pyridoxal-5'-phosphate. Enzymes are numbered as follows: [1] prephenate dehydratase; [2] phenylpyruvate aminotransferase; [3] prephanate aminotransferase; [4] arogenate dehydratase; [5] arogenate dehydrogenase; [6] prephenate dehydrogenase; [7] 4-hydroxyphenylpyruvate aminotransferase. If a single dehydratase catalyzes both reactions 1 and 4, it is denoted cyclohexadienyl dehydratase; if a single dehydrogenase catalyzes both reactions 5 and 6, it is denoted cyclohexadienyl dehydrogenase. The bifunctional P-protein, distributed in nature as diagrammed in Figures 3 and 4, catalyzes the sequential reactions of chorismate mutase and prephenate dehydratase. The bifunctional T-protein (see Figure 3) catalyzes the sequential reactions of chorismate mutase and prephenate dehydrogenase.

complete, they generate testable predictions and serve to illustrate a general basis for successively more refined conclusions.

2. The Aromatic Pathway

The biochemical pathway that mediates the biosynthesis of aromatic amino acids in microorganisms and plants is a divergent, multi-branched sequence which begins with the condensation of erythrose-4-phosphate and phosphoenolpyruvate (PEP) to form the 7-carbon sugar, 3-deoxy-*D-arabino*-heptulosonate 7-phosphate (DAHP). The 7-step common pathway produces chorismate, the last common precursor of all amino acid and vitamin endproducts formed. Figure 1 highlights the post-prephenate portion of the pathway that has proven to be diverse with respect to enzymic steps

employed, cofactor specificity of dehydrogenase enzymes, allosteric regulation, and existence of multifunctional enzymes (Byng *et al.*, 1982). The pathway is metabolically expensive. Each molecule of *L*-tyrosine or *L*-phenylalanine demands the expenditure of 2 PEP's and 1 ATP, while each molecule of *L*-tryptophan (not shown in Figure 1) requires 2 PEP's, 1 ATP and 1 PRPP (plus 1 ATP for synthesis of 1 glutamine).

The aromatic pathway is centrally important in nature beyond its role as the source of 3 of the 20 protein amino acids. It is the source of iron-binding siderophores, vitamin-like compounds (folate, vitamin K, ubiquinones), microbial pigments, and antibiotics. In higher plants a quantitatively masssive and diverse network of 'secondary' metabolism depends upon aromatic biosynthesis to supply connecting routes to phenylpropanoid, alkaloid and hormone products. As with branched-chain amino acids (isoleucine, valine and leucine), phenylalanine and tyrosine are formed by parallel sequences of compounds that are simple analogues of one another. This leads to interesting evolutionary considerations with respect to what the origin of these two amino-acid pathways have in common and how specialization of enzymes and their regulation have made phenylalanine and tyrosine so distinct in contemporary systems.

3. Scenario for Evolution of the First Aromatic Pathway

It is generally accepted that aromatic amino acids were among the latest of the 20 amino acids to become fixed in the codon catalog, the order of entry being phenylalanine, tyrosine and then tryptophan (Lacey *et al.*, 1985). The thesis that biosynthetic availability and codon allocations in the genetic code co-evolved (Wong, 1984) seems likely. *L*-Phenylalanine concentrations in the prebiotic environment from pyrolysis of hydrocarbons could have been substantial (Weber and Miller, 1981). The earliest aromatic pathway probably existed as an unbranched conversion sequence to phenylalanine. The formation of *L*-phenylalanine from chorismate (Figure 2) would indeed appear to be the simplest evolutionary progression of biochemical steps.

However, the initial dilemma is how chorismate, a most novel metabolite, ever became available as a starting point for yet-to-emerge pathway branches to the aromatic amino acids. One possibility illustrated at the top of Figure 2 is that the first significance of chorismate was as precursor of one or more of the vitamin-like products produced by contemporary organisms. Compounds such as folate, vitamin K, or ubiquinones may have been crucial entities for the operation of the earliest and most central biochemical pathways. If it is argued that cofactors such as ubiquinone and vitamin K would not have been needed in the anaerobic period when amino acid synthesis presumably began, other chorismate derivatives may have been important. For example, iron-binding siderophores arising from chorismate may have been important at an evolutionary stage that pre-dated aromatic amino acid biosynthesis.

This still leaves unanswered the question of how the common aromatic pathway evolved. Jensen (1976) has addressed the question of how multi-step enzyme se-

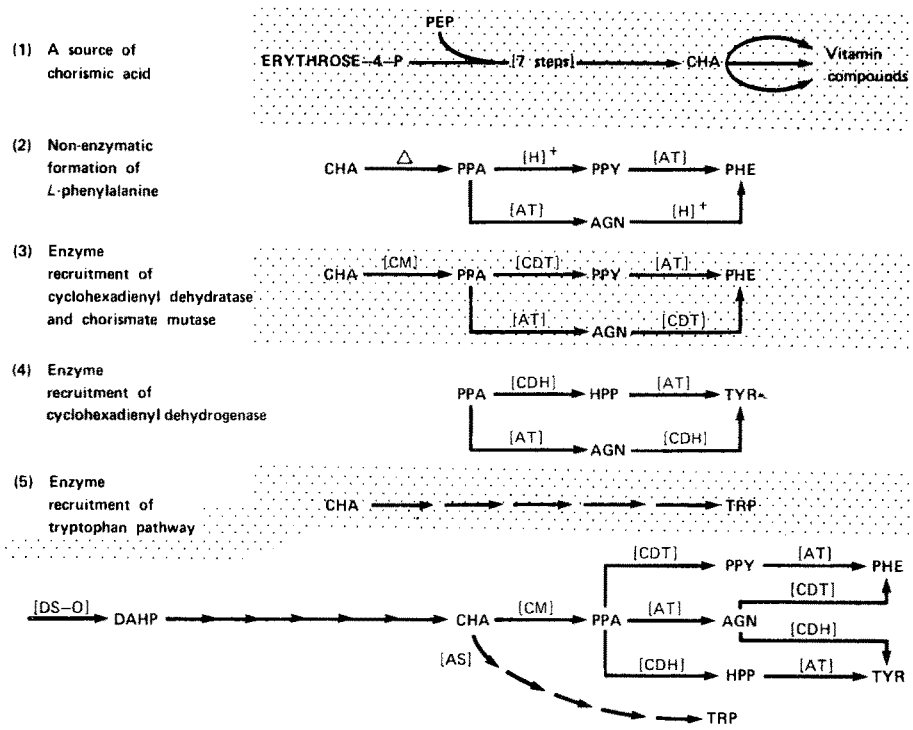

Fig. 2. Scenario for evolution of biosynthetic capability to synthesize aromatic amino acids. Phenylalanine biosynthesis is postulated to have occurred first (steps 1–3), followed by tyrosine biosynthesis (step 4), and then tryptophan biosynthesis (step 5). Abbreviations: [AT] aminotransferase; [CM], chorismate mutase; [CDT], cyclohexadienyl dehydratase; [CDH], cyclohexadienyl dehydrogenase; [DS-0], allosterically insensititive species of DAHP synthase. See legend of Figure 1 for other abbreviations.

quences evolved in view of the expectation that during the intermediate stages of one-by-one acquisition of genes encoding the full complement of enzymes, their preservation would not be ensured by selection pressure. It was proposed that pre-existing enzymes of ancient systems exhibited less specificity for substrates than modern enzymes, essentially catalyzing whole families of related reactions. Thus, the kinase and dehydrogenase reactions of the pre-chorismate pathway, for example, may have been minor and essentially erroneous reactions carried out by kinase and dehydrogenase enzymes that were primarily dedicated to catalyzing reactions in other pathways. Any minor metabolite formed via an *en bloc* series of fortuitous enzyme reactions that proved to be of selective value could then be amplified by gene duplications that might enhance limiting steps. Importantly, the new duplicate gene would have the luxury to specialize in the direction of specificity for the previously irrelevant substrate.

Horowitz (1945) proposed that as the putative supply of pathway intermediates in the primordial soup became depleted, biochemical pathways evolved backward in reverse order of the biosynthetic sequences. This idea anticipated that each gene corresponding to an enzyme step would be homologous to the gene catalyzing the succeeding step (since it arose by tandem gene duplication followed by appropriate modification of the duplicate gene). This hypothesis was seemingly supported by the initial findings of operon organizations for pathways such as histidine and tryptophan in *Escherichia coli*. However, tryptophan-pathway enzymes have not, in fact, proven to be homologous (Yanofsky, 1984). This should not be surprising since the separate reactions of the tryptophan pathway, like many pathways, employ dissimilar reaction mechanisms. Whole-pathway operon organization has also proven to be the exception rather than the rule for microbial pathways. The Horowitz hypothesis might perhaps apply to occasional simple pathways where all pathway intermediates and products were likely to have been present in the primordial soup and where sequential reaction steps are sufficiently similar, but it is frankly untenable to think that generally unstable precursors such as those that comprise the histidine, tryptophan and pre-chorismate pathways would have been present in the primordial soup.

Given the existence of biochemical capability to form chorismate, occasional molecules of *L*-phenylalanine could then be formed without any new enzyme machinery as proposed by Jensen and Pierson (1975). Chorismate undergoes facile rearrangement to prephenate at physiological temperatures (70 °C is optimal, but conversion is substantial at 25 °C) as shown by Gibson (1964). Even mildly acidic pH converts prephenate quantitatively to phenylpyruvate (Zamir *et al.*, 1983). Enzyme-mediated transamination would still have been needed for phenylalanine formation. Since even contemporary aminotransferases are notorious for their multiplicity and overlapping substrate specificities (Jensen and Calhoun, 1981), it seems likely that one or more ancient aminotransferases would have possessed sufficient substrate ambiguity to transaminate both prephenate and phenylpyruvate as shown in Figure 2. Glutamate or aspartate would already have been available as amino-donor reactants since these were surely among the earliest amino acids utilized in biological systems. *L*-Arogenate exhibits a degree of acid lability approaching that of prephenate (Zamir *et al.*, 1983). Hence, there were two non-enzymatic routes that any given molecule of prephenate could follow during conversion to *L*-phenylalanine.

The third level of evolutionary advancement would be the recruitment of enzymes catalyzing the phenylalanine-pathway reactions. These include chorismate mutase, prephenate dehydratase, and arogenate dehydratase. The two analogous dehydratase reactions were probably catalyzed by a common 'cyclohexadienyl dehydratase'. Although substrate-specific prephenate dehydratases and arogenate dehydratases exist today, substrate-ambiguous cyclohexadienyl dehydratases are widespread as well (Jensen and Fischer, 1986). Improved transamination of prephenate and/or phenylpyruvate may have evolved following modification of substrate specificity in a gene duplicate encoding the broad-specificity aminotransferase of the hypothetical ancestor.

The origin of tyrosine biosynthesis must have required the evolutionary recruitment of a cyclohexadienyl dehydrogenase, catalyzing the conversion of prephenate to 4-hydroxyphenylpyruvate and/or conversion of *L*-arogenate to *L*-tyrosine. The transamination steps already in place following evolution of the phenylalanine pathway should have been fully adequate for tyrosine-pathway reactions since aminotransferases capable of transaminating phenylpyruvate inevitably transaminate 4-hydroxyphenylpyruvate (its close analogue) as well. The gene encoding cyclohexadienyl dehydrogenase could perhaps have originated from a duplicate of the gene specifying the cyclohexadienyl dehydratase since the reactions have much in common. A pyridine nucleotide-binding domain would, of course, have had to be acquired. However, it seems more likely to us that cyclohexadienyl dehydrogenase evolved from an ancestral gene for glutamate dehydrogenase or for 2-ketoglutarate dehydrogenase. Prephenate and *L*-arogenate can be regarded as analogues of 2-ketoglutarate and *L*-glutamate, respectively. Like the cyclohexadienyl dehydrogenase reactions, the 2-ketoglutarate dehydrogenase reaction is one of oxidative decarboxylation.

Presumably the 5-step conversion of chorismate to *L*-tryptophan was the last segment of the aromatic pathway to evolve. If *p*-aminobenzoate (PABA) synthase (the first enzyme of folate synthesis) already existed, it probably produced occasional molecules of anthranilate (*o*-aminobenzoate), the first specific metabolite of tryptophan biosynthesis. The prediction of homology between genes specifying anthranilate synthase and PABA synthase (Jensen, 1976) has recently been demonstrated (Goncharoff and Nichols, 1984). The earliest PABA synthase and anthranilate synthase enzymes probably functioned as aminases (ammonia-utilizing) rather than as amidotransferases (glutamine-utilizing) since glutamine was probably not a primitive amino acid of biological systems (Weber and Miller, 1981). The glutamine-binding subunits of these enzymes (Kane *et al.*, 1972) were probably a late evolutionary addition. Each molecule of *L*-trytophan requires the input of one serine molecule in the tryptophan synthase reaction. Because serine was a product of one of the earliest amino acid pathways to evolve, it would have been available. Some enzymes of tryptophan biosynthesis could in fact have been recruited from the serine biosynthetic pathway since some of the reactions of serine biosynthesis are analogous to reactions of tryptophan biosynthesis (Jensen, 1976).

The outcome of the above-specified steps – an intact but as yet unregulated pathway of aromatic amino acid biosynthesis – is shown at the bottom of Figure 2. An allosterically insensitive DAHP synthase (DS-0) catalyzed the formation of DAHP which was ultimately converted to chorismate. The divergent flow options of chorismate molecules were influenced by competition between chorismate mutase and anthranilate synthase, while the fate of prephenate molecules was dictated by competition between the dehydratase and the dehydrogenase enzymes. Within the context of the coevolution theory for codon distribution (Wong, 1984), note the possibility that *L*-arogenate might originally have had a codon allocation that was later assumed by phenylalanine or tyrosine.

4. Evolution of Pathway Regulation

If the pre-chorismate pathway made the transition from being the supply route to minor vitamin-like endproducts to quantitatively major amino-acid endproducts as supposed in Figure 2, then amplification of rate-limiting steps by mechanisms of gene duplication would have been maintained by selective pressure. Newly recruited enzymes such as those catalyzing chorismate mutase, cyclohexadienyl dehydratase or cyclohexadienyl dehydrogenase reactions might also have developed as isozyme families due to the selective advantage of cells that amplified these activities through gene duplication. The common use of differentially regulated isozymes at branchpoint locations such as DAHP synthase and chorismate mutase may reflect a further evolutionary exploitation of already-present isozyme genes. These genes needed only to undergo changes resulting in the acquisition of allosteric sites to become a family of differentially regulated isozymes.

Under conditions of limited pathway flux, the co-evolving properties of chorismate mutase and anthranilate synthase must have been coordinated to promote balanced partitioning of chorismate toward tryptophan on the one hand, and toward phenylalanine and tyrosine on the other hand. Likewise, the kinetic properties of cyclohexadienyl dehydrogenase and cyclohexadienyl dehydratase must have co-evolved to promote the balanced distribution of prephenate or arogenate molecules into the phenylalanine and tyrosine branches. As the quantitative potential of the pathway to produce amino acid endproducts became fully developed, allosteric control devices must have become important to prevent wasteful bleeding of high-energy metabolites into the aromatic pathway. Feedback inhibition of anthranilate synthase, cyclohexadienyl dehydratase and cyclohexadienyl dehydrogenase by tryptophan, phenylalanine and tyrosine, respectively, would accomplish appropriate channeling to endproducts according to need. Chorismate mutase, at its mid-branch location, was another appropriate target of allosteric control by endproducts. The most important focal point of allosteric control is at the initial pathway step, DAHP synthase. In modern prokaryotes allosteric regulation is not always highly developed at terminal-branch or mid-pathway positions (e.g., in cyanobacteria), but is inevitably well developed at the level of DAHP synthase. As discussed by Jensen and Hall (1982), early-pathway regulation alone appears to be adequate in organisms (e.g., cyanobacteria) which obtain virtually all of their aromatic amino acids from endogenous resources. In contrast, organisms which exploit the exogenous environment for aromatic amino acids whenever possible (e.g., *E. coli*) seem to have evolved allosteric control specificities at every branchpoint position. It is difficult to say whether the earliest cells possessing an intact enzymatic pathway existed in a milieu that offered exogenous aromatic amino acids or not.

A greater number of allosteric patterns of control have been described for DAHP synthase than for any other enzyme (Byng *et al.*, 1982). Inevitably where initial studies suggested lack of allosteric regulation, a new pattern of control mediated by unexpected allosteric agents has been found. Thus, the primary allosteric effector for

TABLE I

Evolutionary recruitment of allosteric sites for modern DAHP synthase enzymes starting with an unregulated species of DAHP synthase (DS-0) originating as diagrammed at the bottom of Figure 2

Ancestral catalytic entity	Enzyme gene recruited for new allosteric sites	Gene fusion product[a]	Evolved allosteric sites[b]	Exemplifying organism
DAHP synthose-0	Chorismate mutase	[DS-ppa]	CHA, PPA	*Bacillus subtilis*
DAHP synthase-0	Anthranilate synthase	[DS-trp]	CHA, TRP	*Pseudomonas aeruginosa*
DAHP synthase-0	Chorismate mutase	[DS-aro]	ARO	*Brevibacterium flavum*
DAHP synthase-0	Prephenate aminotransferase	[DS-Mn]	PPA, AGN	Tobacco chloroplasts

[a] DS-ppa indicates that prephenate (ppa) is the primary allosteric effector in *Bacillus subtilis* (Jensen and Nester, 1965), and indeed throughout *Bacillus, Staphylococcus* and *Streptococcus* (Jensen and Rebello, 1970), now known to comprise a phylogenetic cluster. DS-trp within Superfamily-B prokaryotes is exemplified by *Pseudomonas aeruginosa* (Whitaker *et al.*, 1982) where *L*-trypophan is probably the physiologically significant allosteric effector. However, some Superfamily-B organisms (*Xanthomonas, Lysobactor*) have eliminated other DAHP synthase isozymes and maximized the allosteric sensitivity of the remaining DS-trp isozyme to chorismate. In enteric bacteria, on the other hand, allosteric sensitivity of DS-trp to *L*-tryptophan has been maximized in an isozyme system where DS-trp is a minor fractional entity (Ahmad *et al.*, 1987). In *Brevibacterium flavum* a bifunctional DAHP synthase/chorismate mutase protein [DS-aro] is activated by *L*-tryptophan and inhibited synergistically by *L*-phenylalanine and *L*-tyrosine (Shiio and Sigimoto, 1979). In mung bean (Rubin and Jensen, 1985) and in tobacco and spinach chloroplasts (Jensen, 1986) a manganese-stimulated isozyme of DAHP synthase [DS-Mn] is inhibited by prephenate and by *L*-arogenate, the latter being the primary allosteric agent.

[b] *Abbreviation*; CHA, chorismate; PPA, prephenate; TRP, *L*-tryptophan; ARO, combination of all three aromatic amino acids; AGN, *L*-arogenate.

DAHP synthase in *Xanthomonas campestris* (Whitaker *et al.*, 1985), *Bacillus subtilis* (Jensen and Nester, 1965), and plant chloroplasts (Jensen, 1986) proved to be chorismate, prephenate and *L*-arogenate, respectively. These latter examples all illustrate variations of a pathway control pattern termed sequential feedback inhibition, and at least in these cases the circuits of feedback inhibition present in the terminal branchlets must have preceded the evolution of the final early-pathway feedback loop.

If speculations summarized in Table I are correct, early systems were impressively creative in recruiting allosteric sites for DAHP synthase by fusing the gene with other aromatic-pathway genes that specified appropriate binding sites. In *B. subtilis* DAHP synthase and chorismate mutase activities are carried on a single bifunctional protein. The catalytic sites for chorismate and prephenate within the chorismate mutase domain are in fact also the allosteric sites for DAHP synthase. In other cases presented in Table I new allosteric sites are hypothesized to have arisen from the indicated enzymes on the strength of the precedent in *B. subtilis* and the unexpected combination of allosteric binding sites established. Thus, in Superfamily-B bacteria the

feedback-regulated isozyme DAHP synthase-trp (DS-trp) generally exhibits allosteric specificity for both chorismate and *L*-tryptophan, highly suggestive of origin by fusion of the gene specifying the unregulated DS-0 isozyme (Ahmad *et al.*, 1986) with a duplicate of the anthranilate synthase gene (aminase-encoding cistron). Note the implication that the evolution of feedback inhibition of anthranilate synthase by *L*-tryptophan preceded the evolution of feedback inhibition of DS-trp. Phylogenetic subclusters of Superfamily-B prokaryotes possess DS-trp isozymes that range from one extreme where chorismate is the primary effector (*Lysobacter enzymogenes*) to another extreme where *L*-tryptophan is the only effector (*E. coli*). This evolutionary progression of isozyme types in its phylogenetic context has been discussed by Ahmad *et al.* (1986) and by Ahmad and Jensen (1987). In *Brevibacterium flavum* DAHP synthase and chorismate mutase co-exist as a bifunctional enzyme (Shiio and Sugimoto, 1979), as in *Bacillus* and a number of other Gram-positive genera. In *B. flavum.* however, allosteric agents for DAHP synthase are not chorismate and prephenate, but all three aromatic amino acids. This would appear to be an alternative mechanism that coordinates the activity of DAHP synthase and chorismate mutase in response to overall endproduct levels. It seems likely that the binding sites for the aromatic amino acids are conferred by a small, dissociable subunit needed (Shiio and Sugimoto, 1979) for chorismate mutase activity (although the authors conclude otherwise). Finally, DAHP synthase present in higher plant chloroplasts is feedback inhibited by prephenate and *L*-arogenate. This is suggestive of evolution of the allosteric sites from a gene duplicate of prephenate aminotransferase, an extraordinarily specific aminotransferase present in higher plants (Bonner and Jensen, 1985). Although chloroplast enzymes presumably originated from ancestral endosymbiotic prokaryotes, no cyanobacterial or other prokaryote DAHP synthases subject to allosteric control by *L*-arogenate have yet been detected.

5. Biochemical-Pathway Evolution with Phylogenetic Groupings

Jensen (1985) proposed that within a set of organisms having established positions on a phylogenetic tree, evolution of biochemical pathways can be deduced provided that the pathway under analysis exhibits a distribution of alternative character states. In prokaryotes this assertion had not been anticipated on two counts. Firstly, the "unity of biochemical pathways in nature" concept has been over-emphasized. The biochemical pathway for synthesis of aromatic amino acids, for example, is complex and exhibits a rich diversity of character states – the number of which is still increasing (Byng *et al.*, 1982). Yet it was widely believed that the patterns of regulation (until the late-1960's) and the pathway construction (until 1974, Stenmark *et al.*, 1974) were identical with that originally described in *E. coli.* There are increasing indications that aromatic biosynthesis is not an isolated case among biochemical pathways of prokaryotes with respect to the existence of diverse character-state features. For example, substantial diversity in arginine biosynthesis and metabolism has been reviewed by Cunin *et al.* (1986).

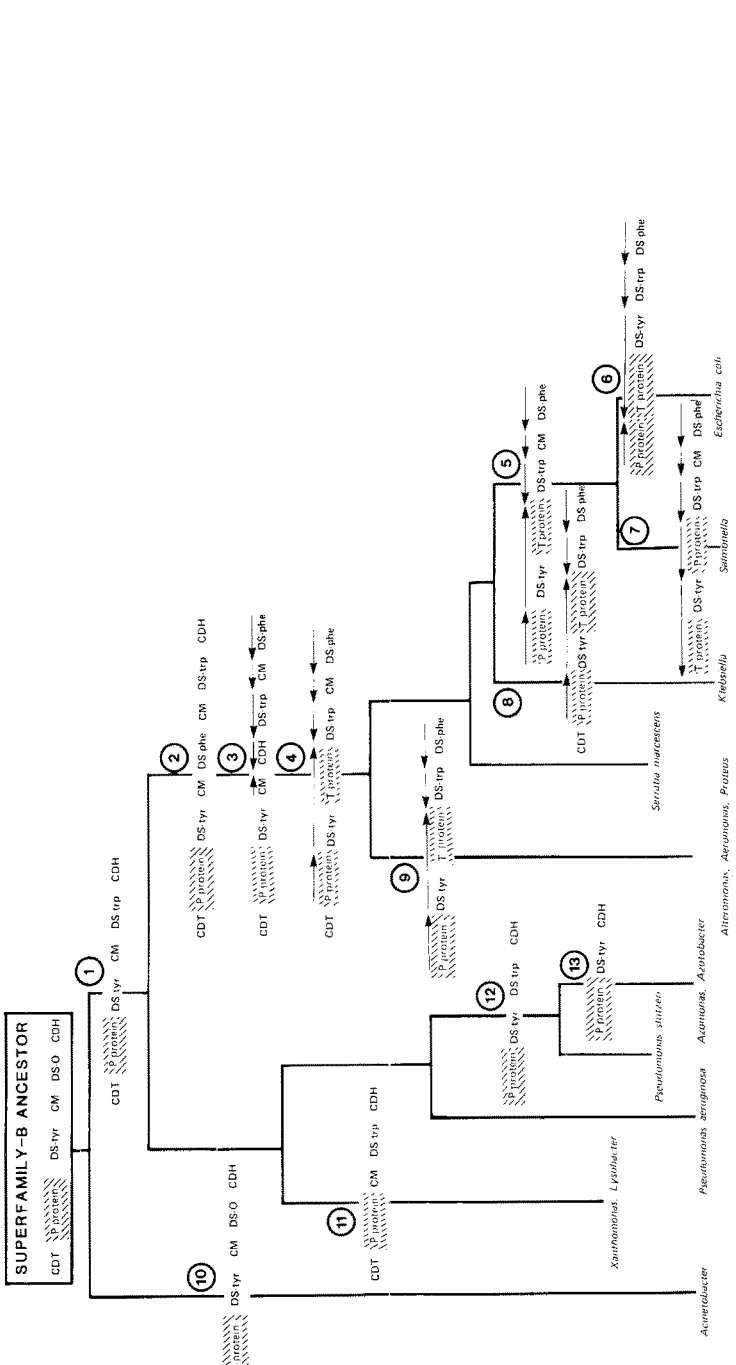

Fig. 3. Evolution of the pathway of aromatic amino acid biosynthesis in Superfamily-B prokaryotes. Genes specifying the gene products indicated are shown, beginning at the top with a hypothetical common ancestor. Any comtemporary organism named at the bottom is the product of the number of changes encountered as the dendrogram line is followed to the top of the diagram. Thus, *Acinetobacter* is the product of only one change (step 10), as is *Pseudomonas aeruginosa* (step 1). *Serratia marcescens* is the product of steps 1–4, while *E. coli* is the product of steps 1–6. The horizontal arrows indicating direction of transcription are hypothetical except for genes encoding the P-protein, the T-protein and DS-tyr in *E. coli* and *Salmonella* (lower right). *Character-state alterations*: 1, DS-0 acquires regulatory site for L-tryptophan (and chorismate), becoming DS-trp; 2, unequal crossing over between homologous DS-tyr and DS-trp genes generates a hybrid gene (DS-phe) and a duplication of the gene for CM; 3, the DNA segment encoding [DS-phe CM DS-trp CDH] is inverted; 4, fusion of genes encoding CM and CDH occurs to form *tyrA*, a gene specifying the bifunctional T-protein and transcribed from the original CM-gene promoter; 5, the gene for CDT is lost; 6, the gene encoding CM is lost and the DNA segment encoding [DS-tyr T-protein] is inverted; 7, the DNA segment encoding [P-protein DS-tyr T-protein] is inverted; 8, the gene encoding CM is lost; 9, the genes encoding CDT and CM are lost; 10, the gene specifying CDT is lost; 11, DS-tyr is lost and DS-trp acquires greater sensitivity to chorismate (Ahmed *et al.*, 1986); 12, genes encoding both CDT and CM are lost; 13, DS-trp is lost. *Character-state abbreviations*: CDT, cyclohexadienyl dehydratase; CM, chorismate mutase; DS-0, unregulated DAHP synthase; CDH, cyclohexadienyl dehydrogenase; PDT, prephenate dehydratase; DS-phe, phenylalanine-inhibited DAHP synthase; ADH, arogenate dehydrogenase; DS-ppa, prephenate-inhibited DAHP synthase; PDH, prephenate dehydrogenase; DS-tyr, tyrosine-inhibited DAHP synthase; P-protein, bifunctional prephenate dehydratase/chorismate mutase.

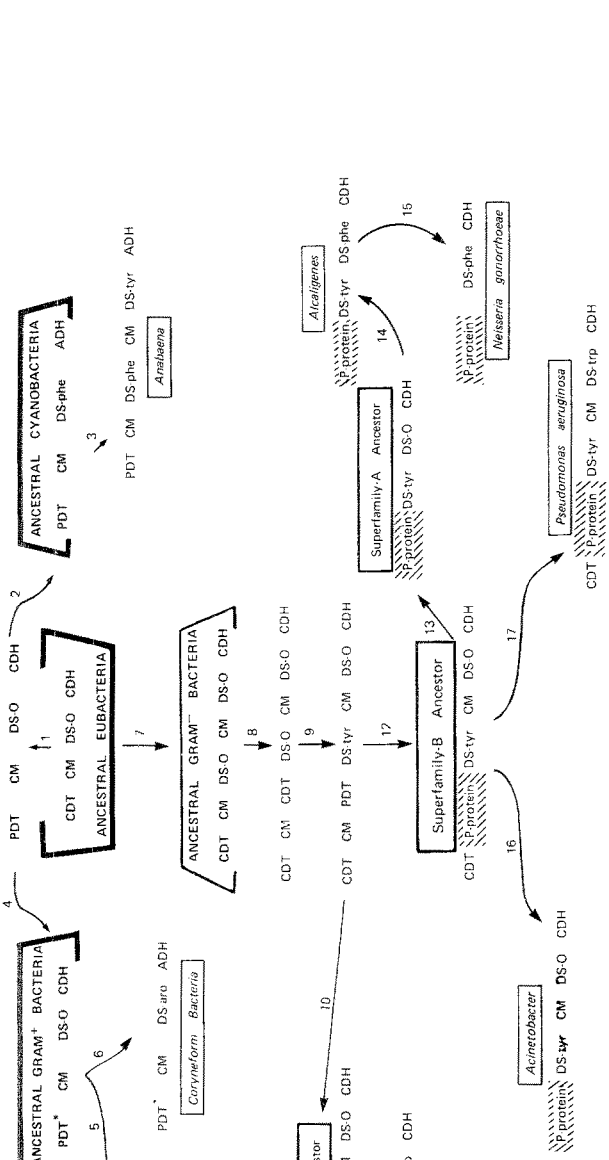

Fig. 4. Evolution of aromatic amino acid biosynthesis in eubacteria. The scenario shown includes three major eubacterial lineages: cyanobacteria (right), Gram-positive bacteria (left), and purple bacteria (denoted Gram-negative bacteria, middle). The major lineages excluded (green sulfur bacteria and spirochaetes) have received too little attention thus far to formulate interpretations. *Character-state alterations*: 1, CDT becomes specialized for prephenate; 2, DS-0 acquires an allosteric site for phenylalanine and CDH becomes specialized for *L*-arogenate; 3, CM and DS-phe undergo gene duplication and one DS-phe gene duplicate is modified at its allosteric site to become the gene for DS-tyr; 4, PDT acquires property of activation by hydrophobic amino acids (PDT*); 5, fusion of chorismate mutase and DS-0 genes produce the bifunctional DS-ppa/CM protein and CDH becomes specific for prephenate; 6, the DS-0 gene fuses with the gene for chorismate mutase producing a bifunctional enzyme that associates with a protein binding all three aromatic amino acids (this complex is denoted DS-aro), and CDH becomes specific for *L*-arogenate; 7, genes on the [CM DS-0] DNA segment are duplicated; 8, a transposon moves the CDT gene between genes for CM and DS-0 – leaving the original CDT gene in place; 9, CDT becomes specific for prephenate and DS-tyr arises from a copy of DS-0; 10, the remaining CDT gene becomes specialized for arogenate; 11, genes for DS-tyr and CM are lost while DS-0 acquires an allosteric site for *L*-tryptophan (note that unequal crossing over could in a single step have deleted the original genes for DS-tyr, CM and DS-0 – leaving behind a hybrid gene for DAHP synthase); 12, CM and PDT genes fuse to form *pheA*, the gene specifying the bifunctional P-protein; 13, CDT and CM genes are lost; 14, DS-0 evolves an allosteric site for phenylalanine; 15, the DS-tyr gene is lost; 16, the CDT gene is lost; 17, DS-0 evolves an allosteric site for *L*-tryptophan (and chorismate).

Secondly, prior to the yeoman efforts of Woese and his many collaborators, there prevailed a general sense that a prokaryote phylogeny was impossible to determine. With the advent of nucleic-acid sequencing techniques, it has now become possible to assign phylogenetic relationships based upon comparisons of genotypic data. Considering the enormity of the sequence space in genomes, homologies in genetic sequences of two organisms can be interpreted in terms of common ancestry at some hierarchical level. The most extensive base of information has been obtained by analysis of 16S rRNA (Fox *et al.*, 1980; Woese, 1987).

6. Pathway Evolution in Superfamily-B Prokaryotes

Although comparative biochemical data have been collected within a broad range of microbial groupings, the Superfamily-B division of Gram-negative prokaryotes is thus far the only phylogenetic assemblage that has been studied in a systematic and comprehensive fashion. Although not all of the major lineages have been characterized, the results illustrate how ancestral character states can be deduced from evaluation of contemporary character states present in representatives of major Superfamily lineages (Jensen, 1985).

Figure 3 shows an organization of selected Superfamily-B organisms, arranged according to their phylogenetic-based dendrogram positions. The cyclohexadienyl dehydrogenase of *Acinetobacter calcoaceticus* is specific for $NADP^+$, while specificity for NAD^+ is characteristic of all other Superfamily-B bacteria. A case has been made that the ancestral Superfamily-B organism possessed a cyclohexadienyl dehydrogenase accepting either NAD^+ of $NADP^+$ (cofactor ambiguity). The bifunctional P-protein (chorismate mutase/prephenate dehydratase) has been a stable character-state feature, present in every Superfamily-B organism examined. Cyclohexadienyl dehydratase (CDT), on the other hand, has been lost frequently in different lineages (steps 5, 9, 10, and 12 of Figure 3). CDT is assumed to have been lost as a character state in step 10 of Figure 3, rather than acquired as a new character state in step 1, because of its presence in Superfamily-C prokaryotes (Figure 4). CDT and chorismate mutase-F (CM) of *Pseudomonas aeruginosa* have been characterized as components of an unregulated, second pathway to phenylalanine that 'overflows' when high intracellular levels of chorismate are available (Fiske *et al.*, 1983). In some cases, CDT and CM may have undergone joint elimination (steps 9 and 12 of Figure 3 and in step 13 of Figure 4). Although not shown as such in our diagrams, this may possibly suggest the contiguous placement of cistrons encoding CDT and CM in an ancestral organism – an arrangement that may still persist in modern organisms such as *Pseudomonas aeruginosa* or *Serratia marcescens*.

Evolution of the DAHP synthase isozyme family within Superfamily B has been dynamic (Ahmed *et al.*, 1986). The putative ancestral DS-0 isozyme, still present in contemporary *Acinetobacter*, eventually acquired allosteric sites for both chorismate and *L*-tryptophan (see Table I) and is denoted DS-trp. In general like *Xanthomonas*

(Whitaker *et al.*, 1985) and *Lysobacter* (Ahmad *et al.*, 1986) DS-trp became the sole DAHP synthase species and evolved greater sensitivity to chorismate (more aptly termed DS-cha), thereby completing the full transition from an endproduct pattern of isozyme control to one of sequential feedback inhibition. The DS-phe isozyme evolved only in the enteric lineage (Ahmed *et al.*, 1987). In step 2 we propose that the DS-phe gene was generated by a mechanism of unequal crossing over between homologous genes encoding DS-tyr and DS-trp. This mechanism offers an attractive way to explain the surprising result that DS-phe exhibits more homology with DS-trp than with DS-tyr (Schultz *et al.*, 1981; Hudson and Davidson, 1984). One would expect that the allosteric site for phenylalanine would most easily have arisen from modification of the allosteric site on DS-tyr. Since unequal crossing over produces a third gene that is a hybrid of the paired homogolous genes, DS-phe could be a duplication of most of DS-trp but still have acquired the sequence encoding the allosteric site of DS-tyr. It is even possible that the unique conformation of the hybrid gene product resulted in better binding of phenylalanine than of tyrosine from the start. The mechanism of unequal crossing over shown also results in duplication of chorismate mutase, yielding a total of three isozymes (the mutase component of the bifunctional P-protein, and two monofunctional mutases – one destined to become a portion of the bifunctional T-protein and the other destined to become the species denoted CM-F. Such examples of a third chorismate mutase within the enteric lineage (e.g., in *Serratia marcescens* and *Erwinia herbicola*) were a surprise. We had anticipated (Jensen, 1985) a simpler scenario: namely that the bifunctional T-protein arose from fusion of the ancestral genes encoding CDH and CM-F, these latter genes being present unfused in the contemporary *P. aeruginosa* (and indeed much of Superfamily B).

The gene orders shown in Figure 3 for *pheA* (P-protein), *tyrA* (T-protein) and *aroF* (DS-tyr) have been established in *E. coli* (Bachmann, 1983) and *Salmonella typhimurium* (Sanderson and Roth, 1983). The direction of transcription indicated by horizontal arrows has been demonstrated for the phe operon (*pheA*) and for the tyr operon (*aroF tyrA*) of both *E. coli* and *S. typhimurium*. All other directions of transcription and gene orders are hypothetical. The purpose of depicting gene orders and directions of transcription in Figure 3 is to illustrate the future feasibility of deducing evolutionary events at this ultimate level of genetic organization. Since *Escherichia* and *Salmonella* are very closely related, we undertook the exercise of formulating a reasonable common ancester (the product of step 5) that would fit most simply at its position in the dendrogram. Such a beginning model generates testable predictions, which if incorrect, can undergo successive modifications. Thus, the model shown predicts an organization of the phe and tyr operons that are similar in *Klebsiella, Serratia, Proteus, Aeromonas* and *Alteromonas* – all differing from the known genetic organization utilized by either *Escherichia* or *Salmonella*. Ultimately, the determination of gene order and direction of transcription will become another dimension of the evolutionary analysis.

The inversion shown in step 3 moves the CDH cistron adjacent to the leftward CM

cistron. The fusion event pictured in step 4 results in a new T-protein gene (*tyrA*) that employs the original chorismate mutase promoter. The chorismate mutase components of both the P-protein and the T-protein are located at the amino terminus (Hudson and Davidson, 1985), and the evolutionary progressions made in Figure 3 to arrive at the known genetic arrangements of modern *E. coli* and *Salmonella* take this constraint into account.

7. The Ancestral Eubacterial Pathway

The stage of pathway evolution achieved at the bottom of Figure 2 – an unregulated, enzyme-based sequence of reactions, some of the enzymes being only broadly specific – is the starting point shown in Figure 4 for aromatic pathway evolution in the eubacteria. Almost no information is available about the archaebacterial kingdom at this time. Only the character states of DAHP synthase (early-pathway), chorismate mutase (mid-pathway), and cyclohexadienyl dehydrogenase/cyclohexadienyl dehydratase (late-pathway) are included.

Cyanobacteria are a major, exceedingly diverse prokaryote grouping, and Hall *et al.* (1982) published data indicating distinctive character states that typify a number of sublineages. However, evolution of aromatic amino acid biosynthesis does not seem to have been nearly as dynamic in cyanobacteria as in Superfamily-B prokaryotes, for example. The general cyanobacterial pattern is one of utilization of prephenate for phenylalanine biosynthesis, but of *L*-arogenate for tyrosine biosynthesis. Arogenate dehydrogenase is usually specific for $NADP^+$. A single DAHP synthase is usually feedback inhibited by *L*-phenylalanine which is also a potent inhibitor of prephenate dehydratase. *L*-Tyrosine does not inhibit arogenate dehydrogenase but does activate prephenate dehydratase. The overall regulatory pattern indicates that tyrosine must be made preferentially. However, when tyrosine is present in excess, activation of prephenate dehydratase switches the flow preference to phenylalanine, which in turn shuts the early-pathway down through inhibition of DAHP synthase. Variation of this theme occurs, e.g., a single DS-tyr species is not uncommon. In genera like *Anabaena* isozymes of both DAHP synthase (Hall and Jensen, 1981a) and chorismate mutase (Hall and Jensen, 1981b) have arisen. No multifunctional proteins have been described to date in cyanobacteria. Allosteric regulation at the early-pathway step appears to have been emphasized heavily in cyanobacterial genera (Jensen and Hall, 1982).

Within the Gram-positive bacteria the phenylalanine-pathway dehydratase has become specific for prephenate in *Bacillus* species, in coryneform bacteria (glutamic-acid bacteria), and in mycoplasmas (Berry *et al.*, 1987). It is interesting that prephenate dehydratase in all three of these groups is activated by hydrophobic amino acids such as *L*-leucine (Pierson and Jensen, 1974; Fazel and Jensen, 1980; Berry *et al.*, 1987). The possible distribution of this character state among the entire Gram-positive assemblage would be of interest. In *Bacillus* (and the phylogenetically near *Staphylococcus* and *Streptococcus* genera) DAHP synthase and chorismate mutase co-exist as a bifunctional protein, and prephenate is a potent feedback inhibitor (Table I). In coryneform bacteria DAHP synthase and chorismate mutase also co-exist

as a bifunctional protein. However, tyrosine and phenylalanine feedback inhibit activity while tryptophan activates activity (probably through binding to a small protein which associates with the bifunctional enzyme (Table I). Again, it would be of interest to follow the distribution of the bifunctional protein throughout the entire Gram-positive grouping. In coryneform bacteria the tyrosine-pathway dehydrogenase has become specific for *L*-arogenate (and $NADP^+$), in contrast to *Bacillus* where the dehydrogenase has become specific for prephenate and NAD^+.

The Gram-negative (purple) bacteria have been studied fairly extensively. The distribution of character states within Superfamily-B organisms has produced a reasonably firm indication of what the character states in the Superfamily-B common ancestor must have been (Figure 3). Steps 16 and 17 of Figure 4 correspond to steps 10 and 1 of Figure 3. When similarly comprehensive analyses of Superfamilies A and C are completed, a more certain deduction about the common ancestor of all three Superfamilies can be made. The presence of the P-protein in both Superfamilies A and B, but not in Superfamily C, indicates that Superfamily C diverged from a common ancestor of the other two Superfamilies prior to their evolutionary divergence (Ahmad and Jensen, 1986). It is most pleasing that this order of branching has been recently determined by Oyaizu and Woese (1985). Allosteric specificity of DAHP synthase for tyrosine seems to have arisen before that for either tryptophan or phenylalanine. Note that by our scheme in Figure 4 DS-phe in Superfamily A arose from DS-0 present after step 13, whereas in Superfamily B (Figure 3) DS-0 became DS-trp and DS-phe was postulated to have arisen as a hybrid of the genes for DS-trp and DS-tyr. Phenylalanine-pathway dehydratase existed at an early stage as both prephenate dehydratase and cyclohexadienyl dehydratase. Presumably prephenate dehydratase arose by gene duplication and substrate specialization of an ancestral CDT species. In contrast, a single dehydrogenase capable of utilizing either prephenate or *L*-arogenate may have persisted without generation of isozyme forms. Even the contemporary T-proteins (Figure 3) which must use prephenate as the dehydrogenase sunstrate *in vivo* (because of the channeling mechanism) are able to utilize *L*-arogenate to varying degreess *in vitro* (Ahmad, unpublished data).

8. Tentative Generalizations of Pathway Evolution in Eubacteria

The total information available to date is skeletal, but it is sufficient to allow mention of a few emerging generalizations that apply to the three divisions of eubacteria depicted in Figure 4: cyanobacteria, Gram-positive eubacteria and Gram-negative eubacteria. In Gram-negative bacteria the cyclohexadienyl dehydrogenase (CDH) enzyme of tyrosine synthesis has remained stable throughout the passage of evolutionary time, whereas changes in the enzymatic basis for phenylalanine biosynthesis have been dynamic as outlined above. The opposite appears to be true of Gram-positive bacteria. All of the latter so far studied possess a specific prephenate dehydratase (that is activated by hydrophobic amino acids). In contrast to this uniformity, the tyrosine-pathway dehydrogenase is a cyclohexadienyl

dehydrogenase (CDH) in mycoplasma (Berry *et al.*, 1987), a prephenate dehydrogenase in *Bacillus* (Champney and Jensen, 1970; Ahmad, unpublished data), and an arogenate dehydrogenase in coryneform bacteria (Fazel *et al.*, 1980). In cyanobacteria a stable evolutionary arrangement exists whereby a specific prephenate dehydratase (PDT) mediates phenylalanine biosynthesis and a specific arogenate dehydrogenase mediates tyrosine biosynthesis.

9. Perspective

The tentative and fragmentary schemes for evolution of aromatic biosynthesis shown in Figures 3 and 4 will themselves evolve with the availability of more data focussed upon the skeleton of interpretation just begun. The basic approach is to determine the character states of a manageable phylogenetic cluster of prokaryotes – the more comprehensive, the more accurate should be the deduced ancestral state. If the analysis is extended to nearby phylogenetic clusters, those common ancestral states can in like manner be deduced – and then the common ancestor of these ancestral states, etc.

These evolutionary scenarios generate testable predictions, e.g., will anthranilate synthase (the predicted progenitor of the allosteric portion of DS-trp in Superfamily-B prokaryotes) prove to be homologous with one end of DS-trp? Will the *pheA* and *tyrA aroF* operons exhibit the orientation predicted in Figure 3 for *Serratia marcescens*?

It should be fascinating to compare the map locations of aromatic-pathway genes along the DNA genome in key organisms such as *Acinetobacter*, *Pseudomonas aeruginosa*, and *Xanthomonas* for comparison with detailed maps already available in such as *E. coli* and *S. typhimurium*. The genetic maps of the closely related *Escherichia* and *Salmonella* genera exhibit a multitude of gene rearrangements (Riley, 1985). Genes thought to have arisen by tandem gene duplication by virtue of their homology have usually been dispersed. Such rearrangement may stabilize newly arisen genes from loss by reversal of the original duplication event. Perhaps the dynamics of gene movement will have been too great to follow them progressively among the membership of phylogenetic clusters. However, Riley has also pointed out that the overall genetic maps of *E. coli* and *S. typhimurium* are remarkably congruent considering their divergence millions of years ago. The organization of the genic units of the prokaryotic genome is clearly maintained with considerable stability over very long spans of evolutionary time. Reason for optimism along these lines comes from detailed molecular-genetic information available about differences in gene-enzyme relationships and genetic organization of cistrons participating in tryptophan and PABA biosynthesis in scattered Superfamily-B bacteria (reviewed in Jensen, 1985).

Acknowledgements

These studies have been supported by grant PCM 8316275 from the National Science Foundation. We appreciate the helpful comments of Dr. David H. Calhoun.

References

Ahmad, S. and Jensen, R. A.: 1986, *Trends Biochem. Sci.* **11**, 108.
Ahmad, S. and Jensen, R. A.: 1987, *Arch. Microbiol.* **147**, 8.
Ahmad, S., Rightmire, B., and Jensen, R. A.: 1986, *J. Bacteriol.* **165**, 146.
Ahmad, S., Johnson, J. L., and Jensen, R. A.: 1987, *J. Mol. Evol.* **25**.
Bachmann, B. J.: 1983, *Microbiol. Rev.* **47**, 180.
Berry, A., Ahmad, S., Liss, A., and Jensen, R. A.: 1987, *J. Gen. Microbiol.* **133**.
Bonnner, C. A. and Jensen, R. A.: 1985, *Arch. Biochem. Biophys.* **238**, 237.
Byng, G. S., Kane, J. F., and Jensen, R. A.: 1982, *Crit. Rev. Microbiol.* **9**, 277.
Champney, W. S. and Jensen, R. A.: 1970, *J. Biol. Chem.* **245**, 3763.
Cunin, R., Glansdorff, N., Pierard, A., and Stalon, V.: 1986, *Microbiol. Rev.* **50**, 314.
Fazel, A. M. and Jensen, R. A.: 1980, *Arch. Biochem. Biophys.* **200**, 165.
Fazel, A. M., Bowen, J. R., and Jensen, R. A.: 1980, *Proc. Natl. Acad. Sci. USA* **77**, 1270.
Fiske, M. J., Whitaker, R. J., and Jensen, R. A.: 1983, *J. Bacteriol.* **154**, 623.
Fox, G. E., Stackebrandt, E., Hespell, R. B., Gibson, J., Maniloff, J., Dyer, T. A., Wolfe, R. S., Balch, W. E., Tanner, R. S., Magrum, L. J., Zablen, L. B., Blakemore, R., Gupta, R., Bonen, L., Lewis, B. J., Stahl, D. A., Luehrsen, K. R., Chen, K. N., and Woese, C. R.: 1980, *Science* **209**, 457.
Gibson, F.: 1964, *Biochem. J.* **90**, 256.
Goncharoff, P. and Nichols, B. P.: 1984, *J. Bacteriol.* **159**, 57.
Hall, G. C. and Jensen, R. A.: 1981a, *J. Bacteriol* **148**, 361.
Hall, G. C. and Jensen, R. A.: 1981b, *Curr. Microbiol.* **6**, 189.
Hall, G. C., Flick, M. B., Gherna, R. L., and Jensen, R. A.: 1982, *J. Bacteriol.* **149**, 65.
Horowitz, N. H.: 1945, *Proc. Natl. Acad. Sci. USA* **31**, 153.
Hudson, G. S. and Davidson, B. E.: 1984, *J. Mol. Biol.* **180**, 1023.
Jensen, R. A.: 1976, *Ann. Rev. Microbiol.* **30**, 409.
Jensen, R. A.: 1985, *Mol. Biol. Evol.* **2**, 92.
Jensen, R. A.: 1986, *Rec. Adv. Phytochem.* **20**, 57.
Jensen, R. A. and Calhoun, D. H.: 1981, *Crit. Rev. Microbiol.* **8**, 229.
Jensen, R. A., and Fischer, R.: 1986, in S. P. Colowick and N. O. Kaplan (eds.), *Methods in Enzymology – Metabolism of Aromatic Amino Acids and Amines*, Academic Press, Inc. (Orlando, FL). **142**, 472.
Jensen, R. A. and Hall, G. C.: 1982, *Trends Biochem. Sci.* **7**, 177.
Jensen, R. A. and Nester, E. W.: 1965, *J. Mol. Biol.* **12**, 468.
Jensen, R. A. and Pierson, D. L.: 1975, *Nature* **254**, 667.
Jensen, R. A. and Rebello, J. L.: 1970, in C. J. Corum (ed.), *Indust. Microbiol.*, American Institute of Biological Sciences, Washington, D. C., Vol. 11, p. 105.
Jensen, R. A. and Stenmark, S. L.: 1975, *J. Mol. Evol.* **4**, 249.
Jung, E. and Jensen, R. A.: 1986, *Proc. Natl. Acad. Sci. USA* **83**, 7231.
Kane, J. F., Holmes, W. F., and Jensen, R. A.: 1972, *J. Biol. Chem.* **247**, 1587.
Lacey, J. C., Hall, L. M., and Mullins, D. W.: 1985, *Orig. Life* **16**, 69.
Oyaizu, H. and Woese, C. R.: 1985, *System. Appl. Microbiol.* **6**, 257.
Pierson, D. L. and Jensen, R. A.: 1974, *J. Mol. Biol.* **90**, 563.
Riley, M.: 1985, in M. K. Hecht, B. Wallace, and G. T. Prance (eds.), *Evolutionary Biology*, Plenum Pub. Corp. Vol. 19, pp. 1–36.
Rubin, J. L. and Jensen, R. A.: 1985, *Plant Physiol.* **79**, 711.
Sanderson, K. E. and Roth, J. R.: 1983, *Microbiol. Rev.* **47**, 410.
Shiio, I. and Sugimoto, S.: 1979, *J. Biochem. (Tokyo)* **86**, 17.
Shultz, J., Hermodson, M. A., and Herrmann, K. M.: 1981, *FEBS Lett.* **131**, 108.
Stenmark, S. L., Pierson, D. L., Glover, G. I., and Jensen, R. A.: 1974, *Nature* **247**, 290.
Weber, A. L. and Miller, S. L.: 1981, *J. Mol. Evol.* **17**, 273.
Whitaker, R. J., Berry, A., Byng, G. S., Fiske, M. J., and Jensen, R. A.: 1985, *J. Mol. Evol.* **21**, 139.
Woese, C. R.: 1987, *Microbiol. Rev.* **51**, 221.
Whitaker, R. J., Fiske, M., and Jensen, R. A.: 1982, *J. Biol. Chem.* **257**, 12789.
Wong, J. T.-F.: 1984, in J. Ricard and Cornish-Bowden (eds.), *Dynamics of Biochemical Systems*, pp. 247–258.
Zamir, L. O., Tiberio, R., and Jensen, R. A.: 1983, *Tet. Lett.* **28**, 2815.
Yanofsky, C.: 1984, *Mol. Biol. Evol.* **1**, 143.