Available online at www.sciencedirect.com
SCIENCE DIRECT®
Bioorganic Chemistry 32 (2004) 309–315
BIOORGANIC CHEMISTRY
www.elsevier.com/locate/bioorg

Mini-review

Unique biosynthesis of dehydroquinic acid?

Ronald W. Woodard*

Department of Medicinal Chemistry and Department of Chemistry, University of Michigan, Ann Arbor, MI 48109-1065, USA

Received 29 March 2004
Available online 2 July 2004

Abstract

A search of the genomic sequences of the thermophilic microorganisms *Aquifex aeolicus*, *Archaeoglobus fulgidus*, *Methanobacterium thermoautotrophicum*, and *Methanococcus jannaschii* for the first seven enzymes (*aroG*, *B*, *D*, *E*, *K*, *A*, and *C*) involved in the shikimic acid biosynthetic pathway reveal two key enzymes are missing. The first enzyme in the pathway, 3-deoxy-D-arabino-heptulosonic acid 7-phosphate synthase (*aroG*) and the second enzyme in the pathway, 5-dehydroquinic acid synthase (*aroB*) are "missing." The remaining five genes for the shikimate pathway in these organism are present and are similar to the corresponding *Escherichia coli* genes. The genomic sequences of the thermophiles *Pyrococcus abyssi* and *Thermotoga maritima* contain the *aroG* and *aroB* genes. Several fungi such as *Aspergillus fumigatus*, *Aspergillus nidulans*, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Pneumocystis carinii* f. sp. *carinii*, and *Neurospora crassa* contain the gene *aroM*, a pentafunctional enzyme whose overall activity is equivalent to the combined catalytic activities of proteins expressed by *aroB*, *D*, *E*, *K*, and *A* genes. Two of these fungi also lack an *aroG* gene. A discussion of potential reasons for these missing enzymes is presented.

Keywords: Biosynthesis; Thermophilic bacteria; Dehydroquinic acid; Shikimic acid; *aroG*; *aroB*; *aroM*; 3-Deoxy-D-arabino-heptulosonic acid 7-phosphate synthase; Missing genes

1. Introduction

A major difference between eukaryotes and prokaryotes is the ability of prokaryotes to synthesize amino acids, an ability seemingly lost by the majority of eukaryotes.

* Fax: 1-734-763-2022.
E-mail address: rww@umich.edu.

0045-2068/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.bioorg.2004.06.003

The biosynthesis of the aromatic amino acids phenylalanine, tyrosine, and tryptophan is a particularly interesting pathway that has been the subject of intense investigation over the past 50 years.

In all organisms studied to date, the aromatic amino acids are obtained via the shikimic acid biosynthetic pathway (see Fig. 1) [1]. While it probably would have been more appropriate to name the pathway after the branch point compound, chorismic acid, the pathway derives its name from *Illicium religiosum* (in Japanense, shikimi-no-ki) from which shikimic acid was first isolated in 1885. The first reaction in the pathway, catalyzed by 3-deoxy-D-arabino-heptulosonic acid 7-phosphate (DAH7P) synthase (*aroG*, EC 4.1.2.15), is an aldol-like condensation between

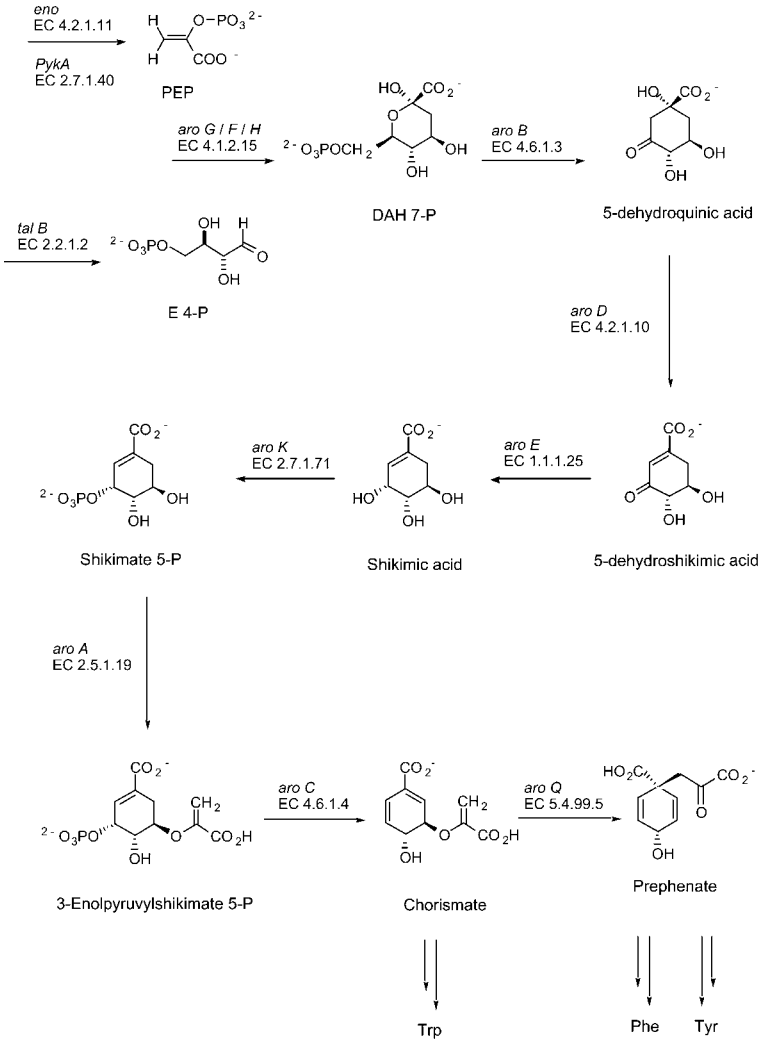

Fig. 1. The shikimate acid pathway

erythrose 4-phosphate (E4P), an intermediate of the pentose-phosphate cycle, and phosphoenolpyruvate (PEP), an intermediate of the glycolytic pathway. The second enzyme in the pathway, 5-dehydroquinic acid (DHQ) synthase (*aroB*, EC 4.6.1.3), is responsible for conversion of the phosphorylated monosaccaride DAH7P into the carbocyclic compound 5-dehydroquinic acid. In addition to the aromatic amino acids, this pathway branches from chorismic acid to furnish *p*-aminobenzoic acid (a precursor of folic acid) and *p*-hydroxybenzoic acid (a precursor of the quinones, which are members of the electron transport chain). Thus, this pathway is critical to the organism's daily maintenance.

At least four environmentally stressed microorganisms chosen for this report, seemingly missing two key steps in the shikimate biosynthetic pathway, represent a diverse set of organisms for which the complete genomic sequences have only recently become available. They are *Aquifex aeolicus* (Eubacteria, Aquificales) [2]; *Archaeoglobus fulgidus* (Archaea, Euryarchaeota, and Archaeoglobales) [3]; *Methanobacterium thermoautotrophicum* (Archaea, Euryarchaeota, and Methanobacteriales), and *Methanococcus jannaschii* (Archaea, Euryarchaeota, and Methanococcales) [4]. Two thermophiles, known to contain the *aroG* and *aroB* genes, where chosen as controls; *Pyrococcus abyssi* (Archaea, Euryarchaeota, and Thermococcales) [5], the control thermophilic archaea and *Thermotoga maritima* (Eubacteria, Thermotogales) [6], the control thermophilic eubacteria.

At present, *A. aeolicus* represents one of only a few thermophiles whose genomes have been sequenced. *A. aeolicus* is one of the most ancient of all life known and grows at 95 °C with an optimal growth temperature of 85 °C whereas most thermophiles normally grow between 50 and 70 °C [2]. Unlike the majority of thermophiles that are classified as *Archaea*, *A. aeolicus* is classified as *Eubacteria* and thus is the earliest known diverging *Eubacteria*. *A. aeolicus* grows under an atmosphere of H_2(79.5%):CO_2(19.5%):O_2(1%) and inorganic (mineral) salts. It will not grow on sugars, amino acids, or conventional media thus it is highly unlikely that the needed aromatic acids are imported from the surrounding environment [7]. Being microaerophilic, *A. aeolicus* requires lower than atmospheric levels of O_2 (7.5 ppm, ca. 30,000 times less than atmospheric) and can oxidize hydrogen ($H_2 \rightarrow 2H^+ + 2e^-$). Finally, *A. aeolicus* is a chemolithotroph, obtaining energy through oxidation of inorganic compounds, as well as an autotroph, obtaining carbon from CO_2 and nitrogen from inorganic sources.

Archaeoglobus fulgidus is the only non-*Eubacterial* sulphur-metabolizing (via sulfate reduction) organism. In addition, *A. fulgidus* is a hyperthermophile, a strict anaerobe, and grows heterotrophically on various carbon sources and lithoautotrophically on hydrogen, thiosulfate, and carbon dioxide.

Methanobacterium thermoautotrophicum is not a true hyperthermophile, but rather an anaerobic archeon that grows best at temperatures in the 65–70 °C range. Also a methanogen, *M. thermoautotrophicum* can use sulfur, sulfite, thiosulfate or ammonium sulfide as a sulfur source and glutamine or urea as a nitrogen source.

Methanococcus jannaschii, a strict anaerobe, grows at pressures of 200 atm and 83 °C. As the name suggests, the organism produces methane and is autotrophic.

As can be seen in Table 1, these four target organisms seem to be lacking both an *aroG* and an *aroB* gene. The genomic sequences of these organisms, except *A. aeolicus*, do not contain the sequence for *aroK* or *aroL*, a second type of shikimate kinase found in *Escherichia coli*, although all four have the DNA sequence for the next enzyme in the shikimate biosynthetic pathway, *aroA* (Fig. 1). The presence of the preceding gene (*aroE*) as well as the following gene (*aroA*) further supports the possibility that genes within the shikimate pathway may have evolved non-standard sequences. Even more complex is *Pyrococcus horikoshii* which seems to be missing the entire shikimate pathway; however, both *P. abyssi* and *P. furiosus* have a DNA sequence coding for an *aroA*/*Q*-like DAH7-P synthase as well as the genes for the remainder of the pathway.

From a chemical stability standpoint, thermophilic microorganisms are also rather interesting since many of the substrates and/or products synthesized have limited thermo-stability in aqueous extracellular buffered solution. This is the case for both the substrate of AroG, E4P, and the product of AroB, DHQ. At room temperature, E4P has been reported to be chemically unstable [8] and at the elevated temperature, at which these organisms grow, it is very unstable. Further, at pH 7 and 50 °C both 5-dehydroquinic acid and 5-dehydroshikimate acid (product of AroD) are quantitatively converted into 3,4-dihydrobenzoic acid. Thus, the thermophilic enzymes or enzyme-complexes needed for these transformations must have evolved to control these thermodynamic phenomena. All the thermophilic microorganisms discussed above contain the genes necessary to express the enzyme responsible for the production of E4P, transaldolase [EC 2.2.1.2], and the enzymes responsible for the synthesis of PEP, enolase [EC 4.2.1.1] and pyruvate kinase [EC 2.7.1.40]. Thus, the potential to synthesize the necessary precursors for the DAH7P synthase reaction exist. However, recent results from feeding experiments utilizing ^{13}C-labeled acetate in an acetate-requiring *Methanococcus* species, reported by Whitman et al. [9], suggest that either E4P is formed by both the non-oxidative pentose phosphate and the triose carboxylation pathways or that E4P is form ONLY by the non-oxidative pathway and is not a precursor of aromatic amino acids.

Environmentally stressed microorganisms are not the only organisms to have developed an unusual pathway for the biosynthesis of the critical aromatic amino acids [10]. The Fungi (Metazoa), *Aspergillus fumigatus*, *Aspergillus nidulans* [11], *Saccharomyces cerevisiae* [12], *Schizosaccharomyces pombe* [13], *Pneumocystis carinii f.* sp. *carinii* [14], and *Neurospora crassa* [15] contain the gene *aroM* (see Table 2). *aroM* expresses a pentafunctional polypeptide chain of $M_r = 165{,}000$ whose overall activity is equivalent to the combined enzyme activities of proteins expressed by *aroB*, *D*, *E*, *K*, *A* genes, i.e., DAH7P goes in and 5-enolpyruvylshikate 3-phosphate exits. It is interesting to note that the AroG (or F, H) activity needed to produce the initial substrate for this complex is not contained within the polypeptide, thus an *aroG* gene must exist elsewhere as a separate enzyme. However, the classical *aroG* gene in two of the Fungi, *N. crassa* and *P. carinii* is again "missing." It is tempting to speculate that the organisms discussed above, both the thermophilic eubacteria, archaea, and the fungias as well as others organisms not scanned for this report, have indeed developed different pathway(s) for the biosynthesis of DAH7P as well as 5-dehydroquinic acid.

Table 1
Sequence homologies for the first seven shikimate biosynthetic genes in several hyperthermic microorganisms as well as *B. Subtilis* and *E. coli*[a]

Microorganism	*aroG* [F, H][b] (*aroG*/*Q*)[c]	*aroB*	*aroD*	*aroE*	*aroK*	*aroA*	*aroC*
Escherichia coli	100 (26)	100 (41)	100 (33)	100 (26)	100 (33)	100 (27)	100 (37)
Bacillus subtilis Marburg 168	Absent (100)	41 (100)	53 (100)	35 (100)	33 (100)	27 (100)	37 (100)
Aquifex aeolicus	Absent (31)[d]	Absent (Absent)	33 (42)	32 (41)	36 (39)	29 (42)	38 (55)
Archaeoglobus fulgidus	Absent (Absent)	Absent (22)	26 (35)	30 (30)	Absent (Absent)	28 (30)	48 (37)
Methanococcus jannaschii	Absent (30)	Absent (Absent)	30 (42)	33 (32)	Absent (Absent)	35 (33)	46 (37)
Methanobacterium Thermoautotrophicum	Absent (31)	Absent (Absent)	29 (39)	29 (28)	29 (Absent)	33 (28)	48 (39)
Pyrococcus abyssi	Absent (54)	38 (37)	Absent (39)	35 (26)	28 (absent)	32 (27)	45 (38)
Pyrococcus furiosus	Absent (55)	38 (38)	24 (40)	36 (27)	28 (Absent)	28 (28)	43 (38)
Pyroccoccus horikoshii	Absent (Absent)	Absent (Absent)	Absent (Absent)	Absent (Absent)	Absent (Absent)	Absent (Absent)	Absent (Absent)
Aeropyrum pernix	Absent (54)	34 (34)	29 (27)	33 (28)	25 (Absent)	31 (29)	43 (38)
Thermotoga maritima[e]	Absent (53)	33 (29)	36 (32)	28 (42)	37 (30)	28 (44)	34 (41)

[a] Numbers represent amino acid sequence % identity to *E. coli* K-12 (MG1655), numbers in (parentheses) represents amino acid sequence % identity to *B. subtilis* Marburg 168.

[b] *E. coli* has three isoforms of DAH7P synthase, each being regulated by a different aromatic amino acid.

[c] The *B. subtilis* DAH7P synthase also has chorismate mutase activity, thus the designation.

[d] This is NOT DAH7P synthase but rather KDO8P synthase which is "matched" by the tBLASTn program to DAH7P synthase.

[e] In *T. maritima* the *aroB*/*K* gene is fused, expressing a protein product with the kinase portion (*aroK*) at the N-terminus. All data was obtained from TIGR.org or NCBI.NLM.NIH.gov.

Table 2
Sequence homologies for the shikimate biosynthetic genes in several fungi/metazoa

Microorganism	*aroG*[a]	*aroM* (%)[b]	*aroC*
Neurospora crassa	Absent	96	28
Aspergillus fumigatus	57	60	Absent
Aspergillus nidulans	58	58	Absent
Saccharomyces cerevisiae	60	50	45
Schizosaccharomyces pombe	60	52	42
Pneumocystis carinii	Absent	46	Absent

[a] Numbers represent amino acid % sequence identity to *E. coli* K-12 (MG1655) *aroG*.
[b] The *aroM* gene contains the same activity as the aroB, *D*, *E*, *K*, and *A* gene but in one pentafunctional polypeptide chain. All data were obtained from TIGR.org or NCBI.NLM.NIH.gov.

The increasing availability of genomic sequences from diverse biological species, the desire to assign a function to every open reading frame and the need to define metabolic pathways in these organisms has spawned intense efforts to find "missing" genes in various pathways. A recent review on the subject of "missing" enzymes has appeared and a quote from this article has been chosen to guide our experimental studies [16]. "Missing enzymes from within biochemical pathways can be explained in three possible ways: (1) they are encoded by low-similarity, novel or 'analogous' genes; (2) they are bypassed if substrates and end-products are obtained from unrelated pathways or from the surrounding medium; or (3) they are absent from genuinely incomplete pathways and have no metabolic role in the cell." We believe number one is operative in the present case and are designing and performing experiments to test that hypothesis. The first part of number two could be operative except we do not believe these organisms extract DHQ from their surroundings since at 50 °C (pH 7) DHQ (as well as dehydroshikimic acid) is quantitatively converted into 3,4-dihydroxy-benzoic acid. Finally, number three is not feasible since all organisms need aromatic amino acids and it is unlikely that these needs could be met by importing the amino acids from their environment especially for the strict chemolithoautotrophy *Aquifex aeolicus*. Work is in progress to "find" the missing *aroG* and *aroB* genes in these microorganisms.

Acknowledgment

This work was supported by GM 53069 (R.W.W.).

References

[1] B. Ganem, Tetrahedron 34 (1978) 3353–3383.
[2] G. Deckert, P.V. Warren, T. Gaasterland, W.G. Young, A.L. Lenox, D.E. Graham, R. Overbeek, M.A. Snead, M. Keller, M. Aujay, R. Huber, R.A. Feldman, J.M. Short, G.J. Olsen, R.V. Swanson, Diversa Corporation, S. D. C. U. S. A., Nature 392(6674) (1998) 353–358.
[3] H.P. Klenk, R.A. Clayton, J.F. Tomb, O. White, K.E. Nelson, K.A. Ketchum, R.J. Dodson, M. Gwinn, E.K. Hickey, J.D. Peterson, D.L. Richardson, A.R. Kerlavage, D.E. Graham, N.C. Kyrpides, R.D. Fle-

ischmann, J. Quackenbush, N.H. Lee, G.G. Sutton, S. Gill, E.F. Kirkness, B.A. Dougherty, K. McKenney, M.D. Adams, B. Loftus, J.C. Venter, et al., The Institute for Genomic Research, R. M. U. S. A., Nature 390(6658) (1997) 364–370.
[4] C.J. Bult, O. White, G.J. Olsen, L. Zhou, R.D. Fleischmann, G.G. Sutton, J.A. Blake, L.M. FitzGerald, R.A. Clayton, J.D. Gocayne, A.R. Kerlavage, B.A. Dougherty, J.F. Tomb, M.D. Adams, C.I. Reich, R. Overbeek, E.F. Kirkness, K.G. Weinstock, J.M. Merrick, A. Glodek, J.L. Scott, N.S. Geoghagen, J.C. Venter, Microbiology Department, U. o. I. C.-U. I. L. U. S. A., Science 273(5278) (1996) 1058–1073.
[5] O. Lecompte, R. Ripp, Puzos-V. Barbe, S. Duprat, R. Heilig, J. Dietrich, J.C. Thierry, O. Poch, Institut de Gâenâetique et de Biologie Molâeculaire et Cellulaire, U. P. R. I. C. U. d. S. F., Genome Res. 11(6) (2001) 981–993.
[6] K.E. Nelson, J.A. Eisen, C.M. Fraser, Institute for Genomic Research, R. M. U. S. A., Methods Enzymol. 330 (2001).
[7] P.M. Coutinho, B. Henrissat, J. Mol. Microbiol. Biotechnol. 1 (2) (1999) 307–308.
[8] P.F. Blackmore, J.F. Williams, J.K. MacLeod, FEBS Lett. 64 (1) (1976) 222–226.
[9] D.L. Tumbula, Q. Teng, M.G. Bartlett, W.B. Whitman, Department of Microbiology, U. o. G. A. U. S. A., J. Bacteriol. 179(19) (1997) 6010–6013.
[10] A.R. Hawkins, H.K. Lamb, J.D. Moore, I.G. Charles, C.F. Roberts, Department of, B., and Genetics, N. M. S. U. o. N. u. T. U. K., J. Gen. Microbiol. 139 (1993) 2891–2899.
[11] I.G. Charles, J.W. Keyte, W.J. Brammar, M. Smith, A.R. Hawkins, Nucleic Acids Res. 14 (5) (1986) 2201–2213.
[12] L.D. Graham, F.M. Gillies, J.R. Coggins, Department of Biochemistry, U. o. G.S. U. K., Biochim. Biophys. Acta 1216(3) (1993) 417–424.
[13] R. Bode, G. Kunze, Z. allgemeine Mikrobiol. 23 (5) (1983).
[14] S. Banerji, A.E. Wakefield, A.G. Allen, D.J. Maskell, S.E. Peters, J.M. Hopkin, University of Oxford Department of Paediatrics, J. R. H. H. U. K., J. Gen. Microbiol. 139 (1993) 2901–2914.
[15] J.R. Coggins, M.R. Boocock, S. Chaudhuri, J.M. Lambert, J. Lumsden, G.A. Nimmo, D.D. Smith, Methods Enzymol. 142 (1987).
[16] S.J. Cordwell, Australian Proteome Analysis Facility, L. B. F. B. M. U. S. A. s. p. o. a., Arch. Microbiol. 172(5) (1999) 269–279.