Proc. Natl. Acad. Sci. USA
Vol. 79, pp. 3794–3797, June 1982
Genetics

Overlapping evolutionary affinities revealed by comparison of amino acid compositions

(mutation/recombination/oligonucleotide substitution/gene conversion/enzymes)

W. K. YEH, C. SHIH, AND L. N. ORNSTON

Department of Biology, Box 6666, Yale University, New Haven, Connecticut 06511

Communicated by Clement L. Markert, February 1, 1982

ABSTRACT **Comparison of the amino acid compositions of purified proteins indicates the presence of overlapping evolutionary affinities among enzymes of the β-ketoadipate pathway. Isofunctional enzymes from different bacterial genera share a common evolutionary origin. Moreover, enzymes that mediate isofunctional or chemically analogous reactions within an organism appear to be evolutionarily homologous. Most remarkably, closely similar amino acid compositions are found in enzymes that mediate the following consecutive metabolic steps: lactonization, decarboxylation, hydrolysis, and transfer of a thioester bond.**

We have reported similar amino acid sequences in the primary structures of enzymes that mediate catalytically diverse reactions in the β-ketoadipate pathway (refs. 1–5; Fig. 1). The inference drawn from these findings was that a novel form of mutation, the substitution of oligonucleotide sequences from one gene into another, took place as different genes for the pathway coevolved (6, 7). Recently, similar mechanisms (7, 8) involving the substitution of "minigene" sequences (9) into different regions of immunoglobulin genes have been proposed to account in part for antibody diversity. It is important to learn the extent to which oligonucleotide substitution mutations took place in metabolic evolution, and a comprehensive approach to this question would be comparison of the primary structures of all of the enzymes of the β-ketoadipate pathway. Some years will pass before this evidence accumulates, but a related form of evidence, the amino acid compositions of the proteins (10), is more easily obtained and allows extensive comparisons to be made. Techniques of comparison tend to exaggerate differences in amino acid composition: thus, the methods may fail to reveal existing homology, but it is quite unusual for them to suggest a homology that does not exist (11).

In this report we use a method of compositional comparison to explore relationships among enzymes of the β-ketoadipate pathway. The measure of relatedness, $S\Delta Q$, is the sum of the square of the difference in mole fraction of the 16 common amino acids that can be readily and quantitatively determined in a protein hydrolysate. Marchalonis and Weltman (12) conducted more than 5,000 pairwise comparisons and found that a $S\Delta Q$ of <50 invariably reflected a structural homology that could be documented at the level of amino acid sequence. The amino acid compositions of some enzymes of the β-ketoadipate pathway have been published (1–5), and we used described procedures (2) to determine the amino acid composition of other enzymes of the pathway. Pairwise comparison of these compositions with those of 25 unrelated proteins* revealed an $S\Delta Q$ of <50 in only 2 of 624 comparisons; the $S\Delta Q$ exceeded 100 in 93% of the comparisons. As described herein, comparisons among enzymes of the β-ketoadipate pathway often yielded low $S\Delta Q$s, suggesting far-reaching structural relationships.

Isofunctional enzymes from different bacterial genera

The most comprehensive compositional comparisons can be made with enzymes from two bacterial groups, the genus *Acinetobacter* and fluorescent *Pseudomonas* species. The two groups differ widely in their nutritional properties and in the G+C content of their DNA; the organisms also exercise different transcriptional controls in their regulation of the β-ketoadipate pathway (15). Nevertheless, intergeneric comparisons indicate that identical steps in the pathway of the two groups are mediated by evolutionarily homologous enzymes. As summarized in Table 1, five of the eight possible comparisons reveal evolutionary homology as indicated by serological crossreaction or by amino acid sequence. In two instances [carboxymuconolactone decarboxylase (4-carboxymuconolactone carboxy-lyase, EC 4.1.1.44) and β-ketoadipate–succinate CoA-transferase (succinyl-CoA:3-oxoadipate CoA-transferase, EC 2.8.3.6); see Table 1], the conclusion of homology is strengthened by observation of an $S\Delta Q$ of <50. In addition, an $S\Delta Q$ of <50 is observed in comparison of the respective carboxymuconate- and muconate-lactonizing enzymes (Table 1). Thus, the catechol oxygenases from the two bacterial genera are the only isofunctional enzymes of the pathway for which the inference of evolutionary homology cannot be drawn.

Analogous or isofunctional enzymes that coevolved in a single cell line

The two branches of the β-ketoadipate pathway are chemically analogous, and isofunctional enzymes mediate terminal reactions in the pathway of *Acinetobacter*. These observations raised the possibility that parallel steps in the pathway are mediated by evolutionarily homologous enzymes (20), and this interpretation is favored by comparison of both the amino acid compositions and the NH_2-terminal amino acid sequences of the decarboxylase–isomerase pair and *Acinetobacter* enol-lactone hydrolases I and II (Table 2). Compositional comparison of the *Pseudomonas* lactonizing enzymes yields an $S\Delta Q$ of 37 (Table 2), and this low number indicates that they share regions of similar amino acid sequence. Because the amino acid compositions of the respective carboxymuconate- and muconate-lactonizing enzymes of *Acinetobacter* and *Pseudomonas* yield $S\Delta Q$ of <50

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U. S. C. §1734 solely to indicate this fact.

* *Pseudomonas* rubredoxin, *Pseudomonas* putidaredoxin, azurin, deoxyribonuclease, elastase, papain, basic trypsin inhibitor, prolactin, secretin, histone H4, κ-casein, cytochrome *c*, ribonuclease, hemoglobin α chain, hemoglobin β chain, myoglobin, alcohol dehydrogenase, animal lysozyme, bacteriophage lysozyme, trypsin inhibitor, proinsulin, tobacco mosaic virus coat protein, flagellin serum, albumin, avidin, and acyl carrier protein. Amino acid compositions were reported by Dayhoff (13, 14).

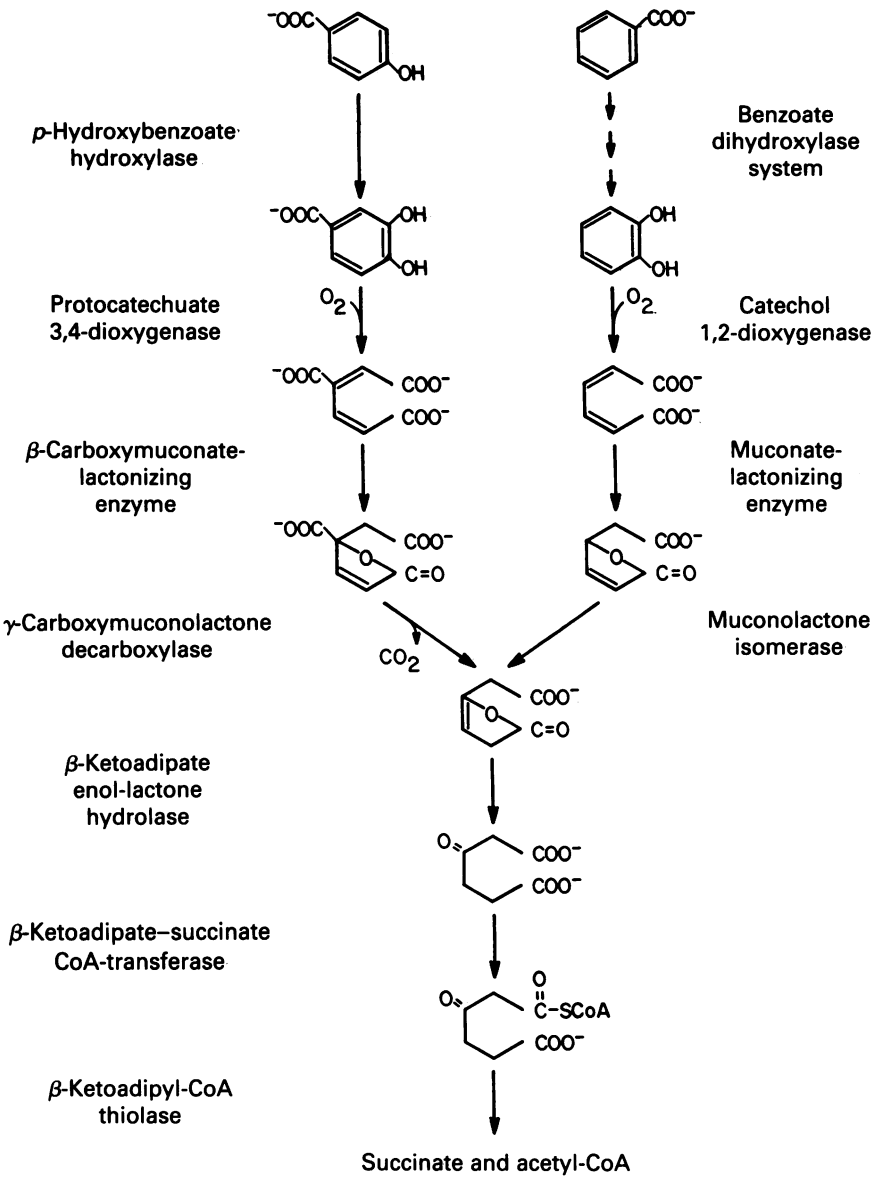

FIG. 1. The β-ketoadipate pathway. The enol-lactone hydrolase and transferase reactions are mediated by isofunctional enzymes in *Acinetobacter*: set I is induced with the protocatechuate pathway and set II is induced with the catechol pathway. A single set of enzymes is found in *Pseudomonas*.

(Table 1), it appears that all four lactonizing enzymes derived sequences from a common ancestor.

Coevolving enzymes that mediate different steps in the β-ketoadipate pathway

The compositional data allow 26 pairwise comparisons of enzymes that mediate two different reactions in one of the branches of the β-ketoadipate pathway in a single cell line (*Acinetobacter* or *Pseudomonas*). Eight of the $S\Delta Q$s emerging from the 26 comparisons are <50. The most striking results are those that emerge from comparison of enzymes of the protocatechuate pathway in *Acinetobacter*. As shown in Table 3, all of the pairwise comparisons (excluding those involving the oxygenase) yield an $S\Delta Q$ of less than 50. Thus, the compositional data predict that similar amino acid sequences will be found in a lactonizing enzyme, a decarboxylase, a hydrolase, and a transferase. When taken in conjunction with the data presented in Tables 1 and 2, the evidence suggests that segments of similar primary structure are widely distributed among enzymes of the β-ketoadipate pathway.

Discussion of genetic origins of homologous sequences

Most studies of molecular evolution have focused upon isofunctional genes or gene products in divergent organisms. In many cases comparisons have allowed the construction of evolutionary trees tracing divergent macromolecules to a common ancestor (21). Thus, it is not surprising that isofunctional enzymes from *Acinetobacter* and *Pseudomonas* (Table 1) share similar sequences, apparently conserved as the proteins diverged from an ancestral protein with the same activity. Enzymes that subject analogous substrates to similar chemical transformations have been shown to be members of enzyme families (22, 23)—homologous proteins that acquired different substrate specificity as they diverged. Thus, it elicits no surprise that muconate- and carboxymuconate-lactonizing enzymes possess similar amino acid compositions (Table 2): this probably is an indication of homologous primary structures conserved from an ancestral amino acid sequence.

A remarkable finding is the presence of closely similar amino acid compositions in enzymes with different catalytic activities

Table 1. Intergeneric comparison of isofunctional enzymes

Enzyme*	SΔQ	Independent evidence for sequence homology	Ref.
Protocatechuate oxygenase	121	Immunological crossreaction	16
Carboxymuconate-lactonizing enzyme	38	Not available	
Carboxymuconolactone decarboxylase	32	NH_2-terminal amino acid sequence	4
β-Ketoadipate enol-lactone hydrolase	55	NH_2-terminal amino acid sequence	3
β-Ketoadipate–succinate CoA-transferase	14	Immunological crossreaction	17
Catechol oxygenase	56	Not available	
Muconate-lactonizing enzyme	38	Not available	
Muconolactone isomerase	58	NH_2-terminal amino acid sequence	1

* Enzymes were obtained from *Acinetobacter calcoaceticus* and from closely related *Pseudomonas* species in the fluorescent group: protocatechuate oxygenase from *Pseudomonas aeruginosa* (18); catechol oxygenase from *Pseudomonas arvilla* (19); and the rest of the enzymes from *Pseudomonas putida*.

(Table 3). On the basis of this evidence, it is tempting to conclude that enzymes of the protocatechuate pathway are members of a different kind of enzyme family, a group of enzymes that possess sequence similarity due to their shared evolutionary history as participants in a single set of metabolic reactions. A mechanism to account for the structural similarity, envisioned by Horowitz (24), is that a series of genetic duplications gave rise to the ancestors of enzymes that subject similar substrates to different chemical transformations in a pathway. As described elsewhere (6, 25), this interpretation poses several difficulties in the context of the β-ketoadipate pathway. Here we wish to call attention to the fact that at least several enzymes of the β-ketoadipate pathway are likely to have been recruited (26, 27), selected on the basis of catalytic activity, from other metabolic sequences: β-ketoadipate–succinate CoA-transferase closely resembles acetoacetate–acetate CoA-transferase (17), and muconolactone isomerase, homologous with carboxymuconolactone decarboxylase (Table 2), shares regions of sequence similarity with ketosteroid isomerase from *Pseudomonas testosteroni* (28). Thus, we observe enzymes that may be traced to different

Table 3. Compositional relationships among enzymes of the protocatechuate pathway in *Acinetobacter*

	SΔQ when compared with			
Enzyme	Oxygenase	Lactonizing enzyme	Decarboxylase	Hydrolase
Lactonizing enzyme	77			
Decarboxylase	64	24		
Hydrolase	54	18	37	
Transferase	78	19	35	32

trees—recombinant proteins that bear the marks of different ancestors. In principle, intergeneic recombination could involve either the insertion or the substitution of DNA sequences. In the former case, the structural gene is lengthened by accretion of the inserted DNA. In the latter case, the length of the structural gene is not altered because the length of the acquired sequence matches the length of the deleted sequence. The available evidence indicates that the latter mechanism, oligonucleotide substitution, was responsible for intergeneic sequence exchange among the structural genes we have examined (5, 6); evidence presented here indicates that such exchange was widespread.

Mutation by oligonucleotide substitution appears to occur in diverse biological systems. This form of genetic exchange appears to have contributed to the evolution of *Bacillus* spore proteins (29). Gene conversion, a form of oligonucleotide substitution that depends upon homology in recombining sequences, has been documented in several eukaryotic systems and has been suggested as a mechanism contributing to hypervariability in antibody genes (7, 8). It might be wise to refer to the latter process as "oligonucleotide substitution" rather than "gene conversion" because the degree to which it depends upon sequence homology has not been established.

We thank G. Davis for skilled technical assistance, S. Mazumdar for samples of purified *Acinetobacter* muconate-lactonizing enzyme, and W. Davidson for useful discussion of methods of compositional comparison. Our work was supported by National Science Foundation Grant PCM 7724884, National Institutes of Health Grant GM25487, and the Celanese Corporation.

Table 2. Analogous or isofunctional enzymes from a single cell line

Enzyme	SΔQ	Independent evidence for sequence homology	Ref.
	*Pseudomonas**		
Protocatechuate oxygenase/ catechol oxygenase	91	Not available	
Carboxymuconate-lactonizing enzyme/ muconate-lactonizing enzyme	37	Not available	
Carboxymuconolactone decarboxylase/ muconolactone isomerase	96	NH_2-terminal amino acid sequence	4
	Acinetobacter		
Carboxymuconate-lactonizing enzyme/ muconate-lactonizing enzyme	83	Not available	
Carboxymuconolactone decarboxylase/ muconolactone isomerase	96	NH_2-terminal amino acid sequence	4
β-Ketoadipate enol-lactone hydrolase I/ β-ketoadipate enol-lactone hydrolase II	54	NH_2-terminal amino acid sequence	3

* *Pseudomonas* protocatechuate oxygenase was from *P. aeruginosa* (18), and *Pseudomonas* catechol oxygenase was from *P. arvilla* (19). The other *Pseudomonas* enzymes were from *P. putida*.

1. Yeh, W. K., Davis, G., Fletcher, P. & Ornston, L. N. (1978) *J. Biol. Chem.* **253,** 4920–4923.
2. McCorkle, G. M., Yeh, W. K., Fletcher, P. & Ornston, L. N. (1980) *J. Biol. Chem.* **255,** 6335–6341.
3. Yeh, W. K., Fletcher, P. & Ornston, L. N. (1980) *J. Biol. Chem.* **255,** 6342–6346.
4. Yeh, W. K., Fletcher, P. & Ornston, L. N. (1980) *J. Biol. Chem.* **255,** 6347–6354.
5. Ornston, L. N. & Yeh, W. K. (1979) *Proc. Natl. Acad. Sci. USA* **76,** 3996–4000.
6. Yeh, W. K. & Ornston, L. N. (1980) *Proc. Natl. Acad. Sci. USA* **77,** 5365–5369.
7. Egel, R. (1981) *Nature (London)* **290,** 191–192.
8. Baltimore, D. (1981) *Cell* **24,** 592–594.
9. Kabat, E. A., Wu, T. T. & Bilofsky, H. (1978) *Proc. Natl. Acad. Sci. USA* **75,** 2429–2433.
10. Metzger, H., Shapiro, M. B., Mosimann, J. E. & Vinton, J. E. (1968) *Nature (London)* **219,** 1166–1168.
11. Cornish-Bowden, A. (1981) *Trends Biochem. Sci. Pers. Ed.* **6,** 217–219.
12. Marchalonis, J. J. & Weltman, J. K. (1971) *Comp. Biochem. Physiol. B* **38,** 609–625.
13. Dayhoff, M. O., ed. (1972) *Atlas of Protein Sequence and Structure* (National Biomedical Research Foundation, Silver Spring, MD), Vol. 5.
14. Dayhoff, M. O., ed. (1976) *Atlas of Protein Sequence and Structure* (National Biomedical Research Foundation, Silver Spring, MD) Vol. 5, Suppl. 2.
15. Ornston, L. N. & Parke, D. (1977) *Curr. Top. Cell. Regul.* **12,** 209–262.
16. Durham, D. R., Stirling, L. A., Ornston, L. N. & Perry, J. J. (1980) *Biochemistry* **19,** 149–155.
17. Yeh, W. K. & Ornston, L. N. (1981) *J. Biol. Chem.* **256,** 1565–1569.
18. Yoshida, R., Hori, K., Fujiwara, M., Saeki, Y., Kagamiyama, H. & Nozaki, M. (1976) *Biochemistry* **15,** 4048–4053.
19. Nakai, C., Kagamiyama, H., Saeki, Y. & Nozaki, M. (1979) *Arch. Biochem. Biophys.* **195,** 12–22.
20. Ornston, L. N. & Stanier, R. Y. (1966) *J. Biol. Chem.* **241,** 3776–3786.
21. Wilson, A. C., Carson, S. S. & White, T. J. (1977) *Annu. Rev. Biochem.* **46,** 573–639.
22. Mortlock, R. P. & Wood, W. A. (1964) *J. Bacteriol.* **88,** 838–844.
23. Hartley, B. S. (1974) *Symp. Soc. Gen. Microbiol.* **24,** 151–182.
24. Horowitz, N. H. (1945) *Proc. Natl. Acad. Sci. USA* **31,** 153–157.
25. Ornston, L. N. & Yeh, W. K. (1981) *Microbiology (Washington, DC),* **1981,** 140–143.
26. Hegeman, G. D. & Rosenberg, S. L. (1970) *Annu. Rev. Microbiol.* **24,** 429–462.
27. Jensen, R. A. (1976) *Annu. Rev. Microbiol.* **30,** 409–425.
28. Ornston, L. N. & Yeh, W. K. (1982) in *Biodegradation and Detoxification of Environmental Pollutants,* ed. Chakrabarty, A. M. (CRC, Boca Raton, FL), in press.
29. Yuan, K., Johnson, W. C., Tipper, D. J. & Setlow, P. (1981) *J. Bacteriol.* **146,** 965–971.