Proc. Natl. Acad. Sci. USA
Vol. 77, No. 9, pp. 5365–5369, September 1980
Genetics

Origins of metabolic diversity: Substitution of homologous sequences into genes for enzymes with different catalytic activities

(primary structure/sequence homology/mutation/oligonucleotide/evolution)

W. K. YEH AND L. N. ORNSTON

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

Communicated by Norman H. Giles, May 23, 1980

ABSTRACT **Similar amino acid sequences were found in portions of bacterial enzymes that mediate different biochemical transformations. Reactions catalyzed by the enzymes include oxygenation, decarboxylation, isomerization, and hydrolysis. The proteins share a common evolutionary history because they participate in an overall catabolic process known as the β-ketoadipate pathway. One interpretation of the sequence similarities might be that duplication of a single gene gave rise to ancestral genes for the enzymes with different catalytic activities. According to this view, homologous sequences from the ancestral gene were conserved as the proteins diverged to assume different functions. This hypothesis is vitiated by comparison of the NH_2-terminal amino acid sequences of sets of enzymes that mediate identical or analogous metabolic reactions within an organism. Gene duplications giving rise to the enzymes within each set must have followed duplication of a putative ancestral gene for all the sets. Yet the amino acid sequences of the proteins within each set have diverged widely, and against this background of divergence the conservation of sequences from an ancestor common to all the enzymes is unlikely. Rather, it appears that most regions of sequence similarity shared by enzymes from different sets were acquired subsequent to their divergence from any common ancestor. In some cases it appears that relatively short regions of sequence homology were achieved by mutations causing the transfer of sequence information from one set of structural genes to structural genes in another set. Alignment of homologous amino acid sequences within any single set requires the introduction of few gaps. Because gaps are required to align sequences that have been altered by the insertion of genetic material, the evidence indicates that copies of oligonucleotides were exchanged by genetic substitution among different structural genes as they coevolved.**

Over the last several years, we have examined structural relationships among bacterial enzymes that mediate a group of convergent catabolic transformations collectively known as the β-ketoadipate pathway (1, 2). As shown in Fig. 1, the enzymes mediate different catalytic processes: five of the six major type reactions classified by the Enzyme Commission are represented. In recent reports (3–6) we described evolutionary relationships revealed by comparison of the NH_2-terminal amino acid sequences of γ-carboxymuconolactone decarboxylase (EC 4.1.1.44), muconolactone isomerase (EC 5.3.3.4), and β-ketoadipate enol-lactone hydrolase (EC 3.1.1.24). The NH_2-terminal amino acid sequence comparisons revealed a significant fraction of the proteins' primary structures because each of the enzymes is an oligomer composed of identical subunits containing fewer than 120 amino acid residues. Protocatechuate oxygenase (EC 1.13.11.3) is formed by association of two larger, nonidentical, subunits (7). The complete primary structures of the protocatechuate oxygenase α (8) and β (9) subunits from

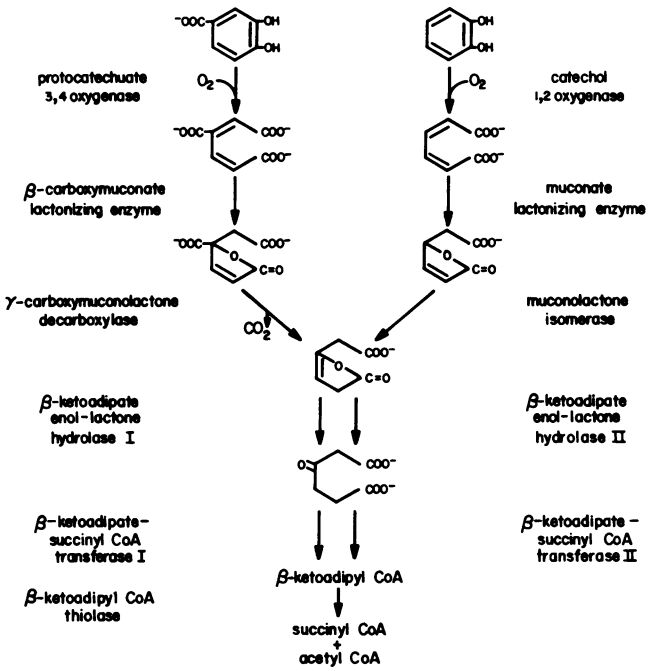

FIG. 1. The β-ketoadipate pathway. Diverse reaction types are represented in the consecutive reactions of the pathway. The two convergent catabolic branches are chemically analogous. For example, γ-carboxymuconolactone decarboxylase and muconolactone isomerase, subjects of this report, mediate similar intralactonic rearrangements. Members of *Acinetobacter* elaborate isofunctional β-ketoadipate enol-lactone hydrolases: enol-lactone hydrolase I is induced with enzymes of the protocatechuate branch, and enol-lactone hydrolase II is induced with enzymes of the catechol branch. Representatives of the fluorescent *Pseudomonas* group form a single β-ketoadipate enol-lactone hydrolase.

Pseudomonas aeruginosa have been determined. In this report we summarize information suggesting common evolutionary origins for amino acid sequences found within the decarboxylase, isomerase, hydrolase, and oxygenase.

Homology of Isofunctional Enzymes from Different Bacterial Genera. Enzymes of the β-ketoadipate pathway are inducible, and comparison of their induction patterns has revealed diverse modes of biosynthetic control. For example, γ-carboxymuconolactone decarboxylase (Fig. 1) is induced by protocatechuate in *Acinetobacter calcoaceticus* (10) and by β-ketoadipate in fluorescent *Pseudomonas* species (11). Despite the differences in their transcriptional regulation, the structural genes for isofunctional enzymes from *Acinetobacter* and *Pseudomonas* are clearly homologous: the first 36 positions of the aligned decarboxylase sequences have 18 identical residues (6), and the first 50 residues of the aligned muconolactone isomerase sequences have 28 identical residues (3).

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.

Acinetobacter Hydrolase I	1 cys	2 glu	3 ile	4 HIS	5 ser	6 ile	7 met	8 ile	9 thr	10 asn	11 arg	12 GLN	13 gly	14 lys	
Acinetobacter Hydrolase II	1 pro	2 val	3 phe	4 HIS	5 phe	6 lys	7 asp	8 thr	9 leu	10 thr	11 ala	12 GLN	13 asp	14 val	
Acinetobacter Hydrolase I	15 thr	16 LEU	17 ser	18 val	19 glx	20 ile	21 asx	22 tyr	23 pro	24 glx	25 ASX	26 pro ?	27 PRO	28 ALA ?	29 ile
Acinetobacter Hydrolase II	ala	LEU	asp	tyr	ala	thr	phe	gly	gln	ala	ASP	arg	PRO	ALA	leu

FIG. 2. The NH_2-terminal amino acid sequences of *Acinetobacter* β-ketoadipate enol-lactone hydrolases I and II. Capital letters designate identical residues occupying the same position in both sequences. Questions marks indicate uncertainty about the assignment of positions 26 and 28 in the hydrolase I sequence.

Divergence of Isofunctional Enzymes Within a Single Cell Line. Representatives of *Acinetobacter* elaborate isofunctional enol-lactone hydrolases under different transcriptional control: enol-lactone hydrolase I is induced by protocatechuate, whereas enol-lactone hydrolase II is induced by *cis,cis*-muconate (Fig. 1). Insofar as is known, the isofunctional enzymes are not regulated at the level of their activity, and therefore selection of protein sites for the binding of different chemical ligands cannot have contributed to their evolutionary divergence. Nevertheless, the NH_2-terminal amino acid sequences of the isofunctional enol-lactone hydrolases diverged widely as they coevolved in *Acinetobacter:* the structures possess a sequence identity of about 20%, sharing no more than 6 residues in the first 29 positions (Fig. 2). A single β-ketoadipate enol-lactone hydrolase is formed in *Pseudomonas putida,* and 14 of the first 24 residues in the NH_2-terminal amino acid sequence of this protein (4) are identical with corresponding residues in one or both sequences of the enol-lactone hydrolases from *Acinetobacter* (5). Therefore all three enol-lactone hydrolases have diverged from a common origin.

Divergence of Enzymes That Mediate Chemically Analogous Reactions. γ-Carboxymuconolactone decarboxylase and muconolactone isomerase (Fig. 1) mediate similar intralactonic rearrangements, and sequence evidence suggests that the

Acinetobacter Isomerase	26 val	27 glu	28 lys	29 ala	30 tyr	31 ser	32 gln	33 glu	34 leu	-Δ-	35 gln	36 arg	37 gln	38 gly	39 lys
Acinetobacter Isomerase	9 val	10 his	11 leu	12 pro	13 val	14 ser	15 met	16 pro	17 thr	18 asp	19 gln	20 ala	21 -Δ-	22 asn	gln
Acinetobacter Hydrolase I	1 cys	2 glu	3 ile	4 HIS	5 ser	6 ile	7 met	8 ile	9 thr	10 asn	-Δ-	11 arg	12 GLN	13 gly	14 lys
Acinetobacter Hydrolase II	1 pro	2 val	3 phe	4 HIS	5 phe	6 lys	7 asp	8 thr	9 leu	10 thr	-Δ-	11 ala	12 GLN	13 asp	14 val
Pseudomonas Isomerase	1 met	2 leu	3 phe	4 his	5 val	6 lys	7 met	8 thr	9 val	10 lys	11 leu	12 pro	13 val	14 asp	15 met

Acinetobacter Isomerase	40 trp	41 arg	42 his	43 ile	44 trp	45 arg	46 ile	47 thr	48 gly	49 gln	50 tyr				
Acinetobacter Isomerase	23 ile	24 lys	25 ser	26 val	27 glu	28 lys	29 ala	30 tyr	31 ser	32 gln	33 glu	34 leu	35 gln	36 arg	37 gln
Acinetobacter Hydrolase I	15 thr	16 LEU	17 ser	18 val	19 glx	20 ile	21 asx	22 tyr	23 pro	24 glx	25 ASX	26 pro ?	27 PRO	28 ALA ?	29 ile
Acinetobacter Hydrolase II	15 ala	16 LEU	17 asp	18 tyr	19 ala	20 thr	21 phe	22 gly	23 gln	24 ala	25 ASP	26 arg	27 PRO	28 ALA	29 leu
Pseudomonas Isomerase	16 asp	17 pro	18 ala	19 lys	20 ala	21 thr	22 gln	23 leu	24 lys	25 ala	26 asp	27 glu	28 lys	29 glu	30 leu

FIG. 3. Homologous amino acid sequences shared by muconolactone isomerases with *Acinetobacter* enol-lactone hydrolases. As in Fig. 2, identical residues occupying the same position in both hydrolase sequences are represented by capital letters; boxes enclose identical residues occupying corresponding positions in the hydrolase and isomerase sequences. The hydrolases resemble the isomerases more than they resemble each other.

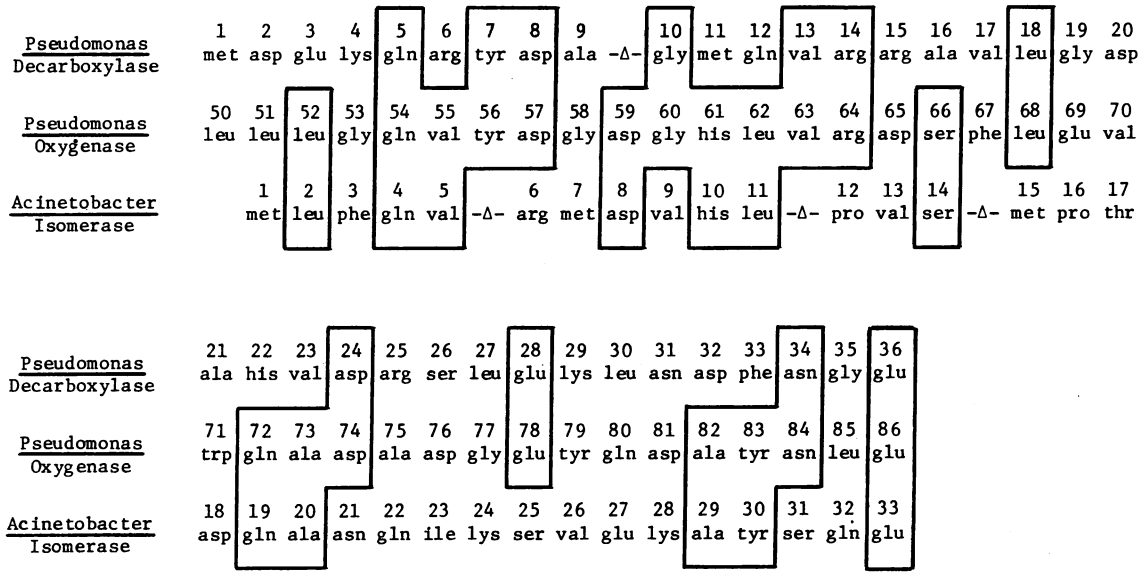

FIG. 4. Part of a computer-facilitated alignment that demonstrates segments of homologous amino acid sequences in *Pseudomonas* γ-carboxymuconolactone decarboxylase, *Acinetobacter* muconolactone isomerase, and the α subunit of *Pseudomonas* protocatechuate oxygenase. Boxes enclose identical residues occupying corresponding positions in the oxygenase sequence and the sequence of either the decarboxylase or the isomerase. The computer program, designed by T. F. Smith (12), aligns sequences to achieve optimal similarity; the program takes into account the minimal mutational distance between nonidentical residues and exacts a penalty for gaps. For the region between residues 51 and 100 of the oxygenase sequence the computer calculated homology matrices of 1082 and 944 for the respective decarboxylase–oxygenase and isomerase–oxygenase comparisons. In a comparative study of 20 unrelated proteins the computer calculated homology matrices between 400 and 600 for optimal alignments.

structural genes for the enzymes have diverged widely from a common origin (6). The genes evolved in somewhat different directions in *Acinetobacter* and *Pseudomonas* and, as a result, comparison of the NH_2-terminal amino acid sequences of the analogous enzymes from both genera reveals overall sequence homology (6). Four pairwise comparisons of the decarboxylase and isomerase sequences can be made: after the introduction of several gaps, the average percentage of identical residues in corresponding positions is 20% (6), the same percentage of sequence identity observed in comparison of the isofunctional enol-lactone hydrolases from *Acinetobacter* (Fig. 2).

Homologous Amino Acid Sequences in Enzymes with Different Catalytic Activities: Isomerases and Hydrolases. Muconolactone isomerases and β-ketoadipate enol-lactone

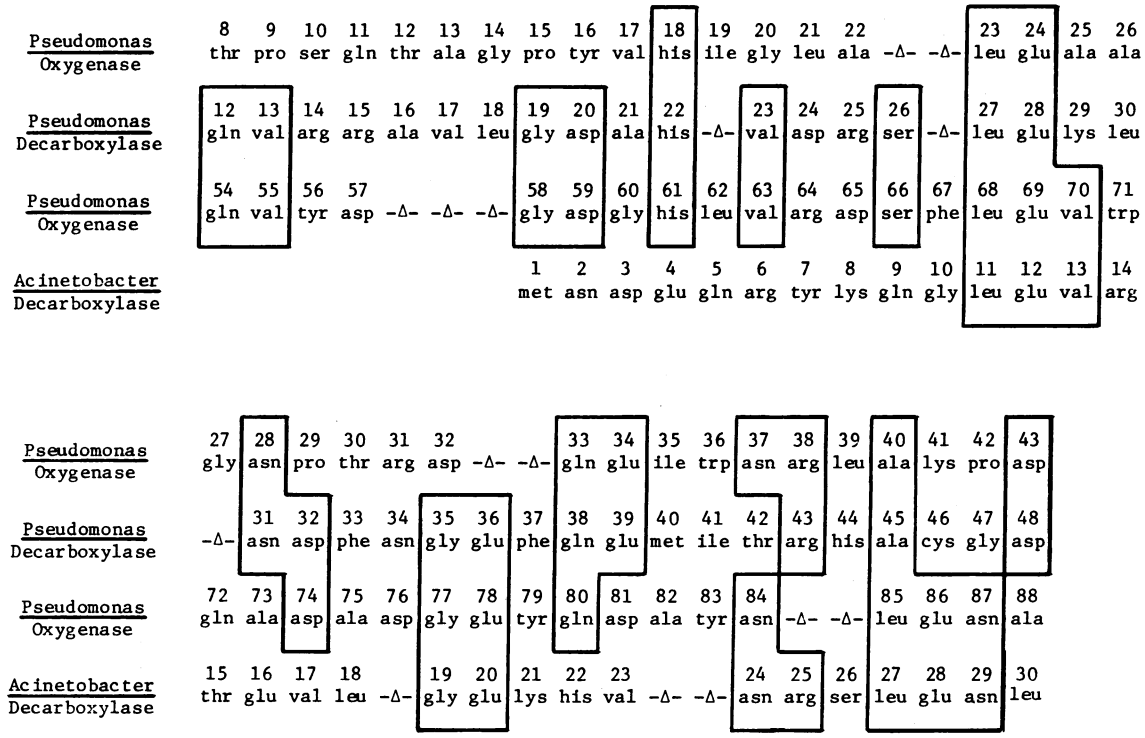

FIG. 5. Sequence homologies evident as the decarboxylase and oxygenase amino acid sequences are aligned in different phases from those depicted in Fig. 4.

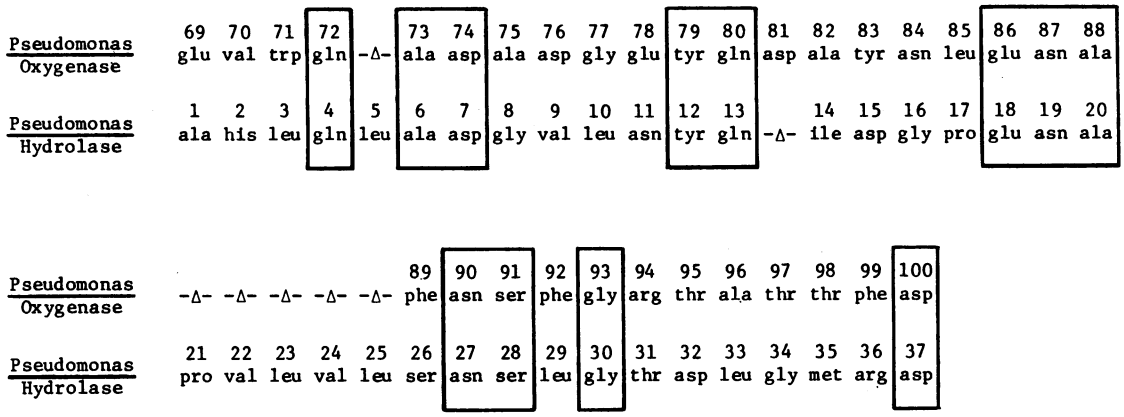

FIG. 6. Amino acid sequences common to *Pseudomonas* enol-lactone hydrolase and the α subunit of *Pseudomonas* protocatechuate oxygenase.

hydrolases mediate different biochemical reactions (Fig. 1). Therefore it is remarkable that portions of *Acinetobacter* enol-lactone hydrolase I and II sequences resemble muconolactone isomerase sequences more closely than they resemble each other (Fig. 3). We would like to call attention to three aspects of the isomerase–hydrolase comparisons.

First, optimal alignment of the isomerase sequences with respect to the hydrolase sequences requires shifts away from the phase that allows optimal alignment of the isomerase sequences with respect to each other. Pairwise comparison of *Acinetobacter* and *Pseudomonas* muconolactone isomerase sequences reveals a sequence identity of 56% when the NH_2-terminal residues are aligned (3), yet the isomerase sequences must be moved from this orientation in order to achieve optimal sequence identity in the isomerase–hydrolase comparisons shown in Fig. 3. Thus it appears that the isomerase and hydrolase sequences are not homologous in their entirety, but different regions within the respective genes are homologous.

Second, only a single gap has been introduced into four of the five sequences in order to achieve the isomerase–hydrolase alignments shown in Fig. 3. Gaps indicate the accretion or loss of genetic material throughout the evolution of the compared sequences. The infrequency of gaps in the isomerase–hydrolase comparisons indicates that the similar sequences were not achieved by the insertion of DNA segments into preexisting structural genes.

Third, the enol-lactone hydrolase sequences do not share most of the sequences that they have in common with the isomerases. This fact suggests that the shared isomerase-hydrolase sequences were absent from the ancestral hydrolase gene and leads to the conclusion that the sequences were substituted by mutation into the hydrolase genes as they diverged from each other. The substantial divergence of the hydrolase sequences eliminates the interpretation that the shared isomerase-hydrolase sequences are products of the convergent evolution of some property shared by the isomerases and hydrolases.

Homologous Amino Acid Sequences in Enzymes with Different Catalytic Activities: Decarboxylase, Isomerase, and Oxygenase. The suggestion that homologous DNA sequences may have been exchanged among coevolving genes is fortified by comparison of the NH_2-terminal amino acid sequences of γ-carboxymuconolactone decarboxylase and muconolactone isomerase with amino acid sequences found within the structure of the protocatechuate oxygenase subunit. Part of a computer-facilitated alignment (12), depicted in Fig. 4, demonstrates sequences found within the decarboxylase and isomerase that resemble sequences within the oxygenase. As in the isomerase–hydrolase comparisons, different sets of homologous sequences become evident as the alignment of primary structures is shifted (Fig. 5). The decarboxylases and isomerases have diverged widely (6), so the sequences these enzymes share with the oxygenase are not likely to have been conserved from a putative ancestor of all three enzymes.

Homologous Amino Acid Sequences in Enzymes with Different Catalytic Activities: Oxygenase, Hydrolase, and Decarboxylase. Sequences common to the primary structures of *Pseudomonas* protocatechuate oxygenase and *Pseudomonas* β-ketoadipate enol-lactone hydrolase are shown in Fig. 6. The enzymes exercise different catalytic activities on chemically remote substrates, and the biochemical transformations are separated by two metabolic steps (Fig. 1). Thus the enzymes

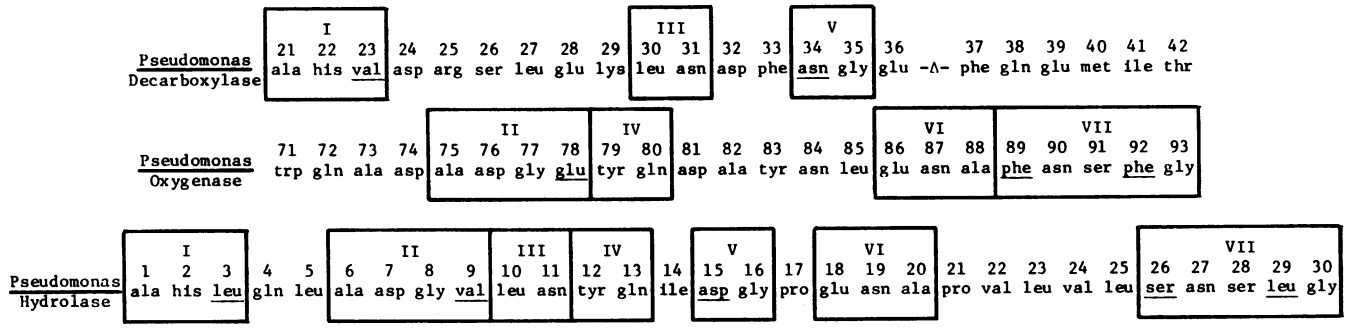

FIG. 7. Similar sequence fragments found in the primary structure of *Pseudomonas* enol-lactone hydrolase and either *Pseudomonas* γ-carboxymuconolactone decarboxylase or the α subunit of protocatechuate oxygenase. The decarboxylase and oxygenase sequences were aligned as depicted in Fig. 4. Boxes enclose sequence fragments that appear to be shared by either of these proteins and by the hydrolase. Underlines in boxes with the same number indicate residues whose codons differ by only a single base.

meet none of the criteria suggested by Horowitz (13, 14) in his proposal that ancestral gene duplications might give rise to structural genes for enzymes that mediate consecutive metabolic reactions. On the other hand, there is no obstacle to the interpretation that substitution of the same sequences into nonhomologous backgrounds could have given rise to the observed pattern of sequence similarity.

Additional evidence, based in part on the computer-facilitated alignment (Fig. 4), illustrates how the exchange of information among juxtaposed nonhomologous genetic regions could either create novel sequences or repair errors in established genes. Part of the decarboxylase–oxygenase alignment is reproduced in Fig. 7, where sequence segments derived from both the decarboxylase and the oxygenase are traced to the hydrolase. This exercise should not be stretched to form the conclusion that the hydrolase gene was formed by intermingling of oxygenase and decarboxylase sequences. The available data suggest a complex pattern of information flow among different genes but do not allow conclusions about past or present directions of information transfer.

Mutation and Selection as Sources of Order and Diversity. There is a tendency in evolutionary thought to regard mutations as the sole cause of genetic diversity; order is the presumed product of selection for a desired trait. Interplay of these forces is illustrated in the commonly accepted view of how new genes evolve. Gene copies may be created and transposed by mutation; variability within genes may be developed by point mutations or by mutations leading to small deletions or insertions. A product of a mutant gene might possess a selectively advantageous trait, an altered binding capacity or a novel catalytic site, and organisms containing such genes would be enriched. Additional mutations might create genetic variants selected because they contained an altered gene product that offered the organism an even greater selective benefit.

To leave the description as outlined above overlooks interactions among coevolving genes, and this consideration shifts our perspective on the relative contribution of mutation and selection to order and diversity. Potential interactions include the loss of homologous genes mediated by recombination enzymes and the interference of slightly divergent gene products in their accurate self-assembly (15). These problems make it likely that duplicated genes diverge rapidly as they become established within a genome. Our data suggest that rapid divergence may be achieved by mutations that substitute DNA sequences with copies of oligonucleotides, either from within the mutating gene or from another gene. Intragenic substitution mutations would cause repeated amino acid sequences within a protein (16), and intergenic substitutions would place segments of homologous amino acid sequences in enzymes derived from different genetic backgrounds.

Substitution of oligonucleotides would allow fragments of existing information to be placed in new contexts and therefore would be a process more orderly than mononucleotide replacement. Additional genetic order may be achieved if oligonucleotide substitution serves as a basis for repair (16). On the other hand, selection *away* from lengthy DNA homologies or *away* from amino acid residues associated in self-assembly of protein subunits probably is far less discriminating than selection *for* a specific binding or catalytic site; it is likely that any of a diverse set of modifications would accommodate a need for rapid divergence. Therefore, as homologous genes become established within a genome, natural selection favors the development of genetic diversity.

We thank Dr. T. F. Smith of Northern Michigan University for providing computer programs that analyze the significance of sequence comparisons and C. Shih for his expert assistance in application of the programs. Our sequence data were gathered in collaboration with Dr. P. Fletcher of the Yale Cell Biology Section. This work was supported by grants from the National Science Foundation (PCM 24884) and from the National Institutes of Health (GM 25487).

1. Stanier, R. Y. & Ornston, L. N. (1973) in *Advances in Microbial Physiology*, eds. Rose, A. H. & Tempest, D. W. (Academic, London), Vol. 9, pp. 89–151.
2. Ornston, L. N. & Parke, D. (1977) in *Current Topics in Cellular Regulation*, eds. Horecker, B. L. & Stadtman, E. R. (Academic, New York), Vol. 12, pp. 209–262.
3. Yeh, W. K., Davis, G., Fletcher, P. & Ornston, L. N. (1978) *J. Biol. Chem.* **253**, 4920–4923.
4. McCorkle, G. M., Yeh, W. K., Fletcher, P. & Ornston, L. N. (1980) *J. Biol. Chem.* **255**, 6335–6341.
5. Yeh, W. K., Fletcher, P. & Ornston, L. N. (1980) *J. Biol. Chem.* **255**, 6342–6346.
6. Yeh, W. K., Fletcher, P. & Ornston, L. N. (1980) *J. Biol. Chem.* **255**, 6347–6354.
7. Durham, D. R., Stirling, L. A., Ornston, L. N. & Perry, J. J. (1980) *Biochemistry* **19**, 149–155.
8. Kohlmiller, N. A. & Howard, J. B. (1979) *J. Biol. Chem.* **254**, 7302–7308.
9. Iwaki, M., Kagamiyama, H. & Nozaki, M. (1979) *J. Biochem. (Tokyo)* **86**, 1159–1162.
10. Canovas, J. L. & Stanier, R. Y. (1967) *Eur. J. Biochem.* **1**, 289–300.
11. Ornston, L. N. (1966) *J. Biol. Chem.* **241**, 3800–3810.
12. Smith, T. F., Fitch, W. M. & Waterman, M. F. (1980) *J. Mol. Evol.*, in press.
13. Horowitz, N. H. (1945) *Proc. Natl. Acad. Sci. USA* **31**, 153–157.
14. Horowitz, N. H. (1965) in *Evolving Genes and Proteins*, eds. Bryson, V. & Vogel, H. L. (Academic, New York), pp. 15–23.
15. Ornston, L. N. & Parke, D. (1976) *Biochem. Soc. Trans.* **4**, 468–473.
16. Ornston, L. N. & Yeh, W. K. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 3996–4000.