THE EVOLUTION OF DEHYDROGENASES AND KINASES

Authors: **William Eventoff**
Michael G. Rossmann
Department of Biological Sciences
Purdue University
West Lafayette, Indiana

Referee: Carl-Ivar Brändén
Department of Chemistry
Agricultural College
Uppsala, Sweden

I. STRUCTURE AND FUNCTION

A. Introduction*

Three modes for the evolution of enzymes have been suggested from time to time. The first is the independent development of classes, each catalyzing different types of reactions (e.g., hydrogen transfer, phosphate transfer, isomerization, etc.), followed by the gradual specialization of enzymes within a class. The second is the independent development of biochemical pathways with all of the members of a given pathway related to one another. The third is the independent development of a group of enzyme classes interrelated by the need to bind a common or similar cofactor.

In support of the first of these hypotheses, Waley[1] has suggested that the preservation of the conformation of the polypeptide chain is of overriding importance in the evolution of enzymes. He derives this premise by observing that (1) the evolutionary process consists of the stepwise replacement of amino acids in the polypeptide chain; (2) the conformation of the polypeptide chain is determined largely or entirely by the sequence; and (3) there are strict requirements for the folding of the linear sequence into a compact stable conformation. Yčas,[2] in a similar vein, has suggested that the mechanism for the evolution of modern metabolic pathways involved the specialization of a smaller set of primordial enzymes which had a much broader specificity. These ideas are supported by the suggestion that the translation process itself must have evolved from something more primitive and less accurate.[3] This would have necessarily led to the production of families of closely related proteins, which taken as a whole would have had a broad specificity, although the individual members may have been highly specific.[2,4]

The second possibility has been supported by Horowitz.[5] He postulated that the biosynthetic pathways evolved in a backward manner, one step at a time, by assuming that the first organism existed in an environment rich in the end product

**Abbreviations used* – LDH: lactate dehydrogenase (EC 1.1.1.27); GAPDH: glyceraldehyde-3-phosphate dehydrogenase (EC 1.2.1.12); LADH: liver alcohol dehydrogenase (EC 1.1.1.1); YADH: yeast alcohol dehydrogenase (EC 1.1.1.1); s-MDH: soluble malate dehydrogenase (EC 1.1.1.37); GluDH: glutamate dehydrogenase (EC 1.4.1.3); AK: adenylate kinase (EC 2.7.4.3); PGK: phosphoglycerate kinase (EC 2.7.2.3); HK: hexokinase (EC 2.7.1.1); Flavo: flavodoxin; Subt: subtilisin (EC 3.4.4.16).

as well as the potential intermediates of a given pathway. Furthermore, Horowitz[6] has suggested that all of the enzymes involved in a given pathway evolved from one another. Waley,[1] although he concedes that this is a plausible scheme for the evolution of pathways, takes exception with the idea that the enzymes in a given pathway have a common origin. Horowitz's primary assumption that all of the metabolic intermediates were in the environment has also been criticized.[2]

The third possibility has been considered by Waley.[1] Watts[7] has pointed out that one sixth of all known enzymes require ATP. In addition, several other coenzymes are structurally related to ATP. Based on these common features, Waley[1] has suggested that the corresponding enzymes may be related. Furthermore, he suggests that the general phosphokinase postulated by Haldane[8] as the only enzyme required by the original organism for the production of polypeptides and polynucleotides was the ancestor of these classes of enzymes (Figure 1). This suggestion is consistent with the idea of primitive enzymes having broad specificity. Baltscheffsky[9] has come to the same concept by considering the redox potential required to derive essential metabolic processes under varying Earth conditions. Thus, he suggests that nonheme iron and heme proteins as well as flavin and NAD-dependent enzymes might have a common origin. However, it is not clear whether the development of different classes from a common precursor could meet Waley's[1] criteria of the preservation of conformation during the evolutionary process.

Thus, while all of the above suggestions are consistent with the view that the present highly specific enzymes probably evolved from ancestral forms of much broader specificity by gene duplication and subsequent mutation, the relationship between the classes of functionally diverse enzymes is not clear.

Recently the solution of the three-dimensional structure of a number of dehydrogenases and kinases has revealed a structural similarity in portions of the polypeptide chain. These developments, described in Part I and quantized in Part II of this review, have helped to shed light on the functional and evolutionary relationship between the two classes of enzymes.

B. Dehydrogenases

The three-dimensional structures of dogfish LDH, porcine s-MDH, lobster GAPDH, and horse LADH have been determined (Table 1). The conformation of the bound coenzyme has also been found in all four cases. The structure of one subunit of these enzymes consists of two or more domains of specialized function.[35] Evolutionary relationships among the dehydrogenases have been reviewed by Rossman et al.[36,37] and Ohlsson et al.[38]

1. Tertiary Structural Comparisons

The similarity in the folding of the s-MDH and LDH subunits was the first structural correlation to emerge from these investigations.[16,39] This may be seen by comparing both the nucleotide binding (Figures 2A and 2B) and catalytic regions (Figures 3A and 3B) of both enzymes. The only significant difference in conformation between the two subunits is the first 20 residues at the amino end of LDH, which are important in the stabilization of quaternary structure (see Section I.B.2). Rao and Rossmann[39] have quantitatively compared these two structures. Their method involves determining a rotation matrix and translation vector which minimizes the sum of the squared differences between structurally equivalent atoms. Once the structures have been so

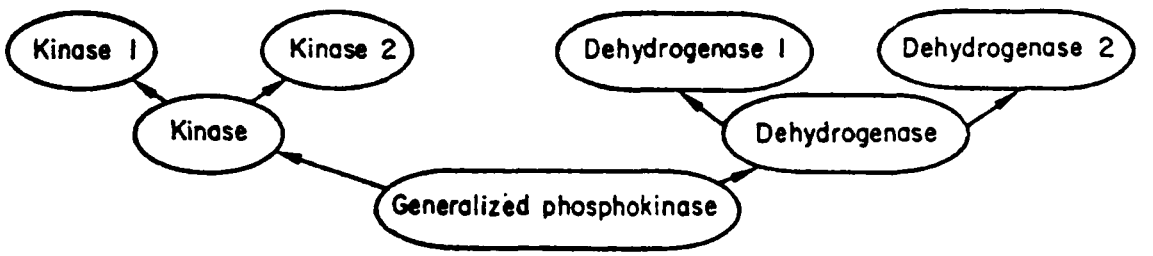


FIGURE 1. General scheme for the evolution of enzymes according to Waley. (From Waley, S. G., *Comp. Biochem. Physiol.*, 30, 1, 1969. With permission.)

TABLE 1

Crystallographic and Sequence Data Available for the Dehydrogenases and Kinases

Enzyme	Types of structures studied	Number of subunits	References: Three-dimensional structure	References: Primary structure
Dehydrogenases				
LDH	Apo, holo, LDH:NAD-pyruvate, and other ternary complexes	4	10–12	13, 14
s-MDH	Holo	2	15, 16	
LADH	ADP-ribose, apo, and a substrate inhibitor	2	17–20	21
GAPDH	Holo	4	22, 23	24
GluDH		6		25–27
Kinases				
PGK	Apo, holo (ADP-Mn, ADP-Mg, ATP-Mg, ATP)	1	28, 29	
AK	Apo	1	30	31
HK	Apo, holo (AMP) complexes with glucose	2	32–34	

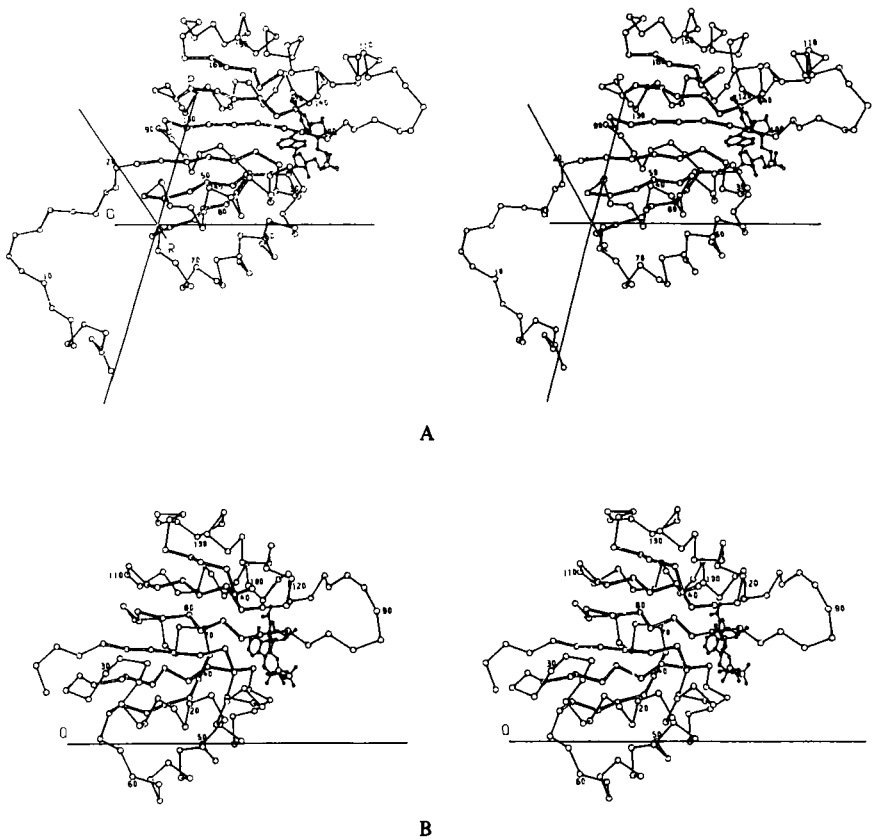


FIGURE 2. Stereo views of nucleotide binding domains in (A) LDH, (B) s-MDH, (C) GAPDH, (D) LADH. (From Rossmann, M. G., Liljas, A., Brändén, C.-I., and Banaszak, L. J., in *The Enzymes*, 3rd ed., Vol. 11, Boyer, P. D., Ed., Academic Press, New York, 1975, in press. With permission.)

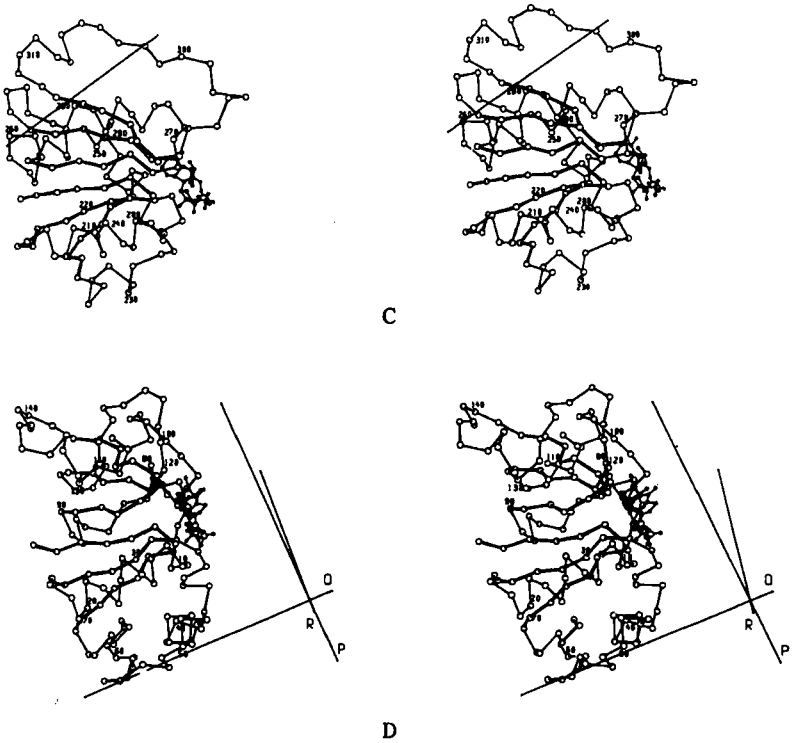


FIGURE 2 (continued)

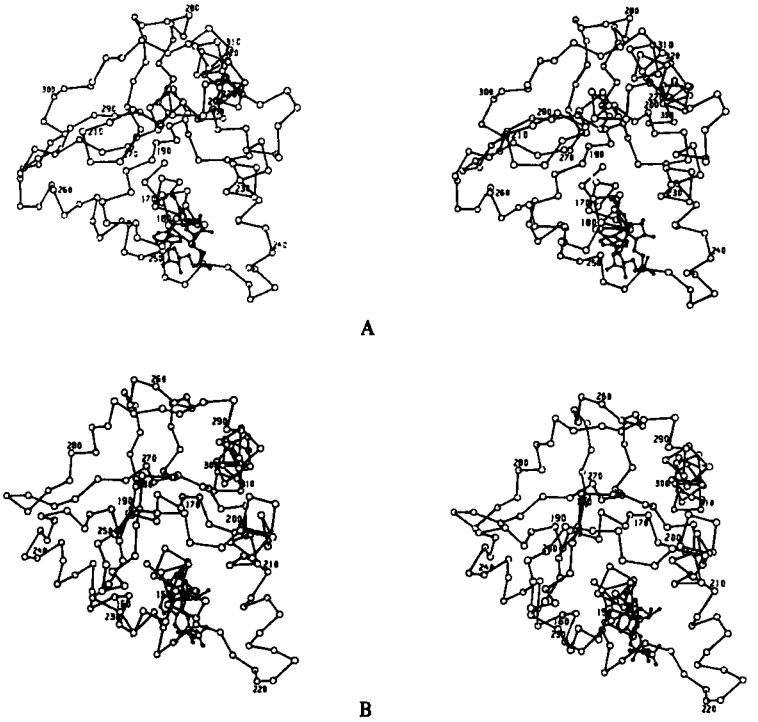


FIGURE 3. Stereo views of catalytic domains in (A) LDH, (B) s-MDH, (C) GAPDH, (D) LADH. (From Rossmann, M. G., Liljas, A., Brändén, C.-I., and Banaszak, L. J., in *The Enzymes*, 3rd ed., Vol. 11, Boyer, P.D., Ed., Academic Press, New York, 1975, in press. With permission.)

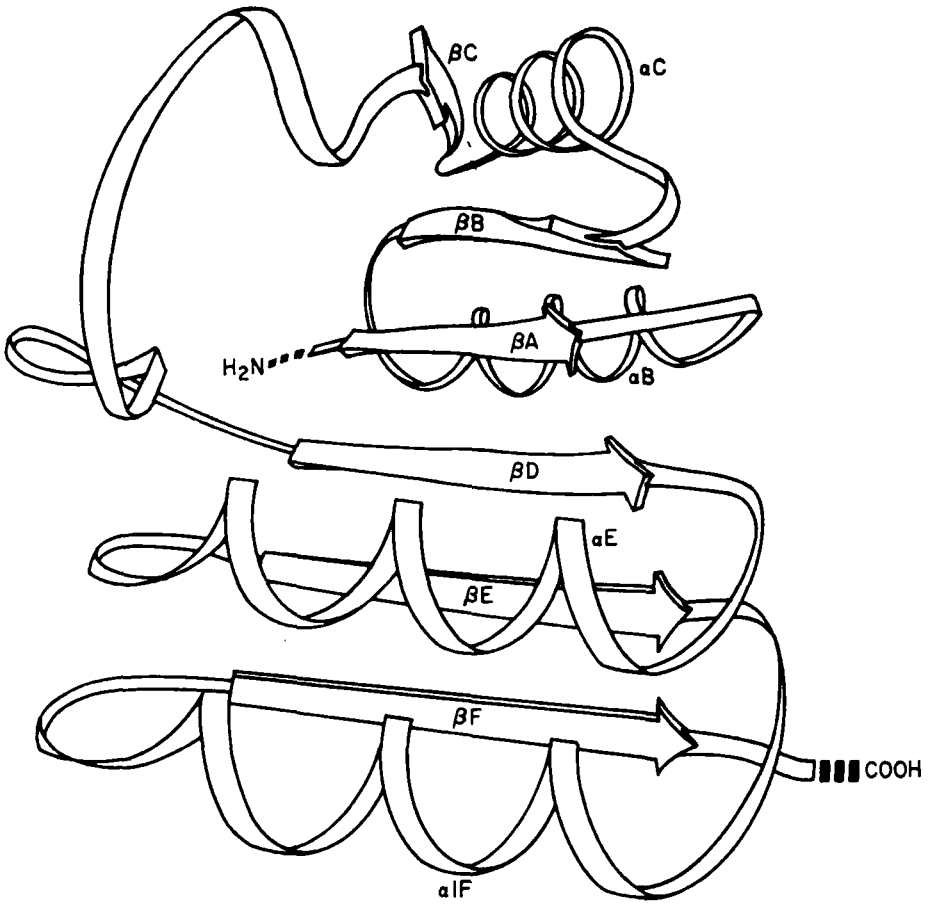
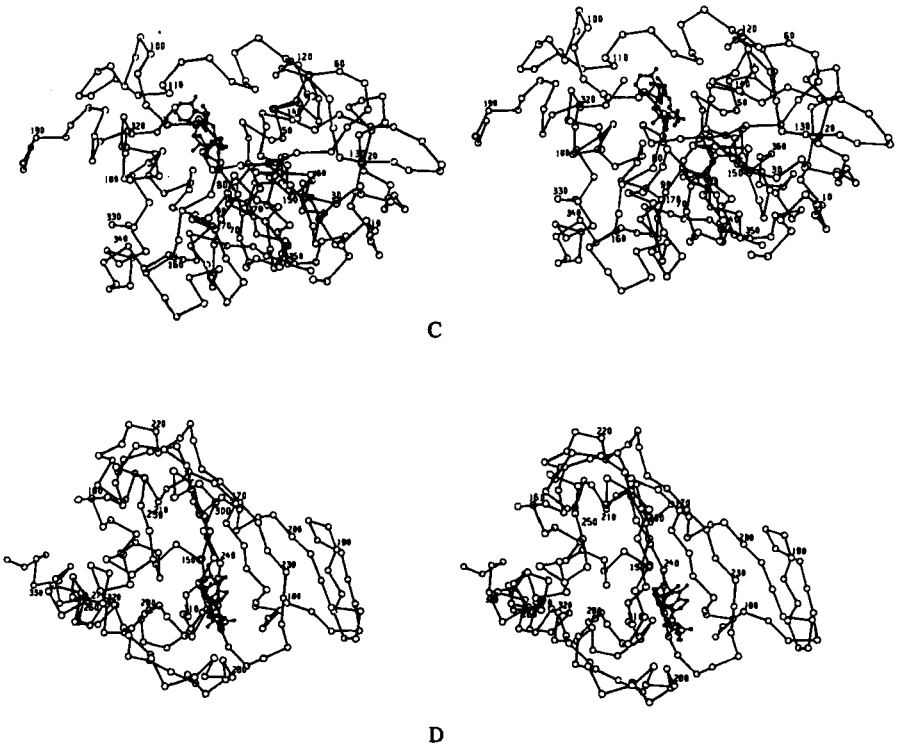


FIGURE 4. Schematic diagram of nucleotide binding domain. (From Ohlsson, I., Nordström, B., and Brändén, C.-I., *J. Mol. Biol.*, 89, 339, 1974. With permission.)

comprising each domain, 147 in GAPDH, 143 in LDH, and 126 in LADH, which can be correlated with the presence of extra structural elements in LDH and GAPDH. In LDH, residues 100 to 120 form a flexible loop which is absent in LADH and GAPDH, while in GAPDH, there is an additional antiparallel sheet excursion between αC and βC which is not present in the other dehydrogenases.

Ohlsson et al.[38] have used the method of Rao and Rossmann[39] to obtain a quantitative measure of the similarities between these coenzyme binding domains. The equivalent atoms in the two structures which were used for alignment were determined by using a method suggested by Rossmann et al.[37] This consists of using the hydrogen bonding scheme within the parallel pleated sheet region. When these patterns are aligned with due regard for the polarity of the peptide chains, the orientation of each C_β atom in the amino acid side chains must also be aligned, thus providing a sensitive method of determining equivalent amino acids. The procedure was shown to lead to a superpositioning of homologous amino acids in LDH, GAPDH, and LADH. The orientation parameters defined by these equivalent atoms were then used by Ohlsson et al.[38] to determine the similarity of structure in the coenzyme binding domains and the bound nucleotide cofactors. The excellent agreement between equivalent atoms in the central parallel pleated sheet demonstrates that the left-handed twist (approximately 100° between the extreme strands βC and βF) is essentially constant for all known dehydrogenases.[36,38] The central region of the dinucleotide binding domain was found to be better conserved than the outside regions. While it has already been mentioned that the conformations of the bound dinucleotides were exceedingly similar in the various structures (see Reference 15 for s-MDH; 18 for LADH; 23 for GAPDH; 11 and 12 for LDH), the spatial alignment of these cofactors, based on the alignment of the protein fold, shows that the



C

D

FIGURE 3 (continued)

aligned, the rms distance between equivalent atoms (σ) represents a measure of their similarity. In the case of s-MDH they showed that for the 288 equivalent atoms in the crystallographically independent subunits, σ equaled 1.8 Å. A comparison of one subunit of LDH with the apo and holo s-MDH subunits gave σ values of 2.7 and 2.9, respectively, for the approximately 250 equivalent atoms used in each case. Although a close relationship between these two enzymes was not expected based on their physical properties,[40] their structural homology suggests that they may have evolved from a common precursor of much broader substrate specificity. This is in agreement with the suggestion by Biellmann,[41] who postulates that due to the exclusive A-side specificity, all α-hydroxy acid dehydrogenases evolved from a common precursor and that modifications in the region involved in substrate binding are responsible for the present variation in specificity.

A structural comparison of LDH and s-MDH with GAPDH[23] and LADH[38] shows that whereas the catalytic domains have different structures (Figure 3), the overall conformations of the domains responsible for the binding of nucleotide cofactors display a striking similarity in all of the subunits (Figure 2). This common structure consists of six strands of parallel pleated sheet (βA, βB, βC, βD, βE, βF) and four helices (αB, αC, αE, α1F), labeled according to the nomenclature first suggested by Adams et al.[10] and later extended by Hill et al.[16] A schematic diagram is shown in Figure 4.

Starting from the amino end, the first structural element is βA. The helices αB and αC form the connections between βA and βB, and βB and βC, respectively, and are on the same side of the sheet. This structure forms the AMP binding domain. From βC, the polypeptide chain passes back to the amino end of the sheet and lays down βD next to βA. The sequence βD, αE, βE, α1F, βF forms the NMN binding domain. As shown by Figure 4, there is an approximate twofold axis parallel to βA and βD relating the two mononucleotide binding domains.[39]

Apart from the overall similarity of these structures, there are a number of differences. Perhaps the most apparent is that they are formed by different sections of the polypeptide chain, residues 1 to 148 in GAPDH, 22 to 164 in LDH, and 193 to 318 in LADH. In addition, they differ slightly in the overall number of residues

TABLE 2

Distances in A Between Equivalent Coenzyme Atoms in Dehydrogenases When Proteins are Superimposed

	LDH-GAPDH	LDH-LADH	GAPDH-LADH
Mean distance	1.8	3.3	2.8
Maximum distance	3.0	5.6	3.9
Minimum distance	0.9	1.4	1.6

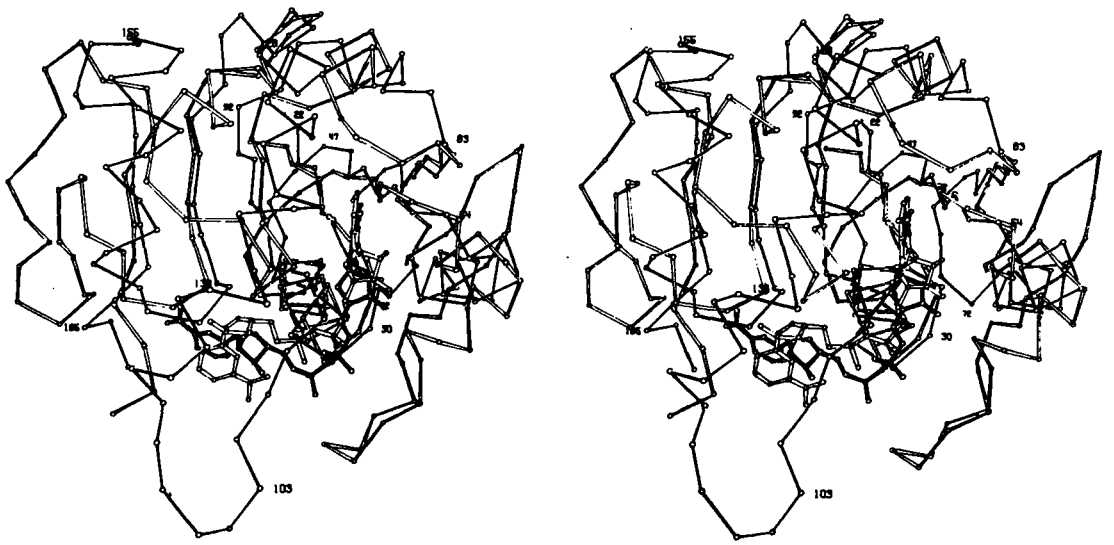


FIGURE 5. Stereo view comparing the NAD binding domains of LDH and GAPDH including bound nucleotide. The GAPDH structure is shown with open bonds, whereas LDH is represented with black bonds. (From Rossmann, M. G., Moras, D., and Olsen, K. W., *Nature*, 250, 194, 1974. With permission.)

positioning of the cofactor within the dinucleotide binding domain is also essentially constant. The mean deviations for the ADP-ribose part of the cofactor for the LADH-LDH, LADH-GAPDH, and LDH-GAPDH comparisons are 2.8, 3.3, and 1.8 Å, respectively (Table 2).[36,38] These comparisons are particularly impressive when shown visually, as has been done by Rossmann et al.,[36,37] Buehner et al.,[23] and Ohlsson et al.[38] A typical example is the comparison of LDH and GAPDH shown in Figure 5.

2. Quaternary Structure

The P, Q, and R axes nomenclature, originally devised for describing the molecular symmetry axes of LDH and s-MDH[42] and later extended to include the GAPDH structure,[22] will be used in discussing the quaternary structure of the dehydrogenases.

The Q axis of LDH is conserved in s-MDH,[16,39] although in the latter only a dimeric structure is formed, probably due to the lack of the N-terminal arm (residues 1 to 21 in LDH). The apparent conservation of the Q axis in GAPDH[22] is probably fortuitous in view of the very few intersubunit contacts generated by this axis in GAPDH as compared with LDH or s-MDH. Furthermore, although the P and R axes are similarly oriented in LDH and GAPDH with respect to the nucleotide binding structure, the interactions across these axes are completely different (Figure 6). In GAPDH the coenzyme binding sites are close together across the R-axis interface, whereas in LDH they are separated. This permits cooperativity between subunits in the latter enzyme.[22,23]

In LADH the subunits are associated in a completely different manner.[17,19] The molecular twofold axis runs approximately perpendicular to the plane defined by βE and βF, causing the two βF strands in adjacent subunits to form an antiparallel pleated sheet.

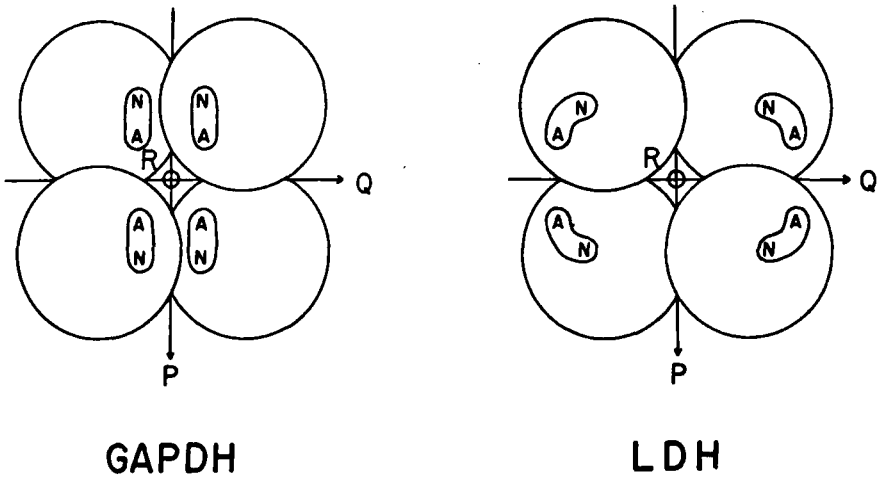


FIGURE 6. Diagrammatic representation of the P, Q, and R axes in GAPDH and LDH. (From Buehner, M., Ford, G. C., Moras, D., Olsen, K. W., and Rossmann, M. G., *Proc. Natl. Acad. Sci. U.S.A.*, 70, 3052, 1973. With permission.)

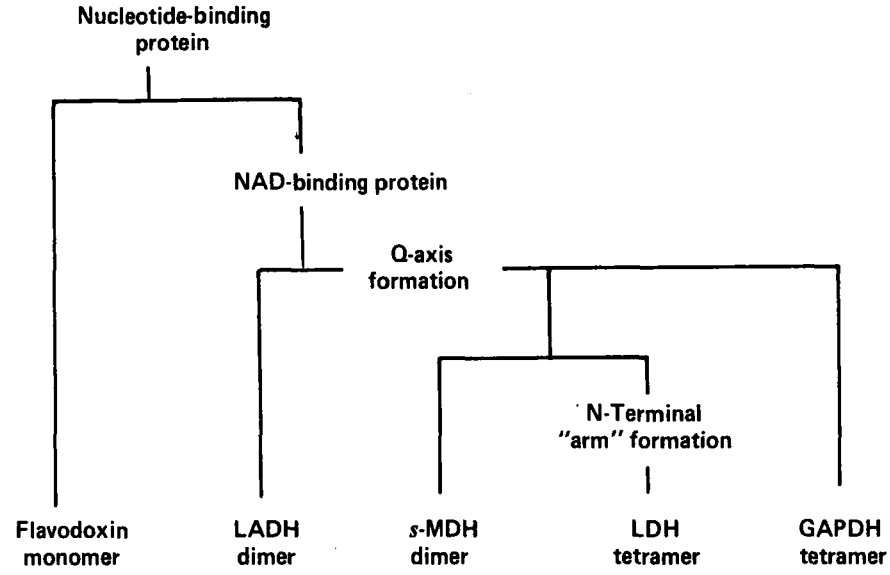


FIGURE 7. Evolutionary scheme for the evolution of dehydrogenases and their relation to flavodoxin as suggested by Buehner et al. (From Buehner, M., Ford, G. C., Moras, D., Olsen, K. W., and Rossmann, M. G., *Proc. Natl. Acad. Sci. U.S.A.*, 70, 3052, 1973. With permission.)

The conservation of the coenzyme binding domains, as well as the similarities and differences in the catalytic domains and quaternary structures of these dehydrogenases, led Buehner et al.[22] to suggest a scheme for their evolution (Figure 7).

3. Sequence Comparisons

In the absence of structural information, a number of sequence comparisons between the dehydrogenases have been made. However, some of these alignments are questionable in light of the structural information now available.

Harris[43] and Fondy et al.[44] used the essential cysteine residue to align the structures of LDH, GAPDH, LADH, and YADH. These results were cited by Taylor[13] and extended by Dayhoff.[45] In view of the known structure, this is only valid for LDH and GAPDH since the essential cysteine, 46, in LADH is in the middle of the catalytic domain. Furthermore, as the essential cysteine is at the

junction of the catalytic and nucleotide binding domains (at the end of βF) in LDH and GAPDH, comparison of these two sequences beyond this residue may not be valid.

Jörnvall's alignment of LADH with GAPDH[46] relates the catalytic domain of LADH and the nucleotide binding domain of GAPDH and thus probably does not reflect a common ancestral relationship between them. Similarly, the homology between GAPDH and GluDH suggested by Engel,[47] an extension of a comparison made by Smith,[48] is unreasonable in view of the proposed repeat of part of the GAPDH catalytic domain in two sections of the GluDH polypeptide chain. This comparison has also been strongly criticized on purely statistical grounds by Williams and Wilkins.[49] Engel[50] has also related regions of GAPDH and LADH, but these do not correspond to the structural equivalence between the two enzymes. However, he does suggest a gene duplication in GAPDH which roughly corresponds to the duplication of the mononucleotide binding region. Bennett[51] makes equally unfortunate comparisons among LDH, GAPDH, LADH, GluDH, and dihydrofolate reductase. These examples point out the difficulty involved in determining homologies between related proteins purely on the basis of amino acid sequences when the divergence is great.

In contrast, Rossmann et al.,[36,37] Ohlsson et al.,[38] and Buehner et al.[22] have all used the structural equivalences of the nucleotide binding domain in dehydrogenases, kinases (*vide infra*), and some other nucleotide binding proteins to obtain sequence homologies (Table 3). Their results show the conservation of functionally important amino acids required in the binding of coenzyme and of structurally important residues required in the folding of the domain. Glycine 28 and glycine 99 (all numbering refers to the LDH sequence) are conserved in all the dehydrogenase structures of known three-dimensional structure. Any larger residue would cause steric hindrance to the binding of the adenine and nicotinamide ribose moieties, respectively. The conservation of glycine 33 is probably due to its position in αB, immediately opposite βB, since any larger residue may affect the relative positioning of this central helix opposite the β-pleated sheet. The conservation of aspartate 53 is related to its functional role of binding the O'_2 atom of the adenine ribose.

Although no three-dimensional structure is known for GluDH, Rossmann et al.[37] compared the sequences of GAPDH and GluDH using the method of Jukes and Cantor.[52] They identified a region in GluDH (residues 245 to 280) which corresponded to the sequence βA, αB, βB. Within this region all of the functionally important amino acids as well as the character of the amino acids in the hydrophobic pocket are conserved, except for aspartate 53, which is changed to glutamic acid in GluDH. Wootton[53] has used a different method for identifying the nucleotide binding regions in the homologous *Neurospora* and beef GluDH sequences. He used the procedure of Chou and Fasman[54,55] to predict the location of helices and sheets in GluDH and then applied a variety of criteria, including conservation of functional residues and the sequence of alternating helices and sheets, to identify two nucleotide binding domains in GluDH (domain-1 and domain-3). Part of domain-3 is that which had previously been identified by Rossmann et al.[37] Domain-3 of *Neurospora* GluDH was also found to have a good degree of homology with the NAD binding domain of GAPDH. He therefore suggests an evolutionary tree which differs from that of Rossmann et al.[37] in that the closer relationship between GAPDH and GluDH has also been considered (Figure 8). One of the difficulties of this prediction is the deletion of the structurally important central strand βA in domain-1 of *Neurospora* GluDH.

4. Summary

The studies of both primary and three-dimensional structures of the dehydrogenases suggest that an NAD binding domain has evolved from a common ancestor which duplicated and then fused with a variety of other genes. It is interesting to note that the structural Zn in LADH is tetrahedrally coordinated by four cysteines in a manner reminiscent of both rubredoxin and ferredoxin.[19] It is thus conceivable that LADH includes structural domains the functions of which relate to two different energy transfer processes.

C. Kinases

The structure and sequence information on kinases is not as complete as for the dehydrogenases (Table 1).

The structure of PGK from both horse[28] and yeast[29] consists of two distinct lobes (related by an approximate twofold axis).[28] Each lobe is structurally reminiscent of the NAD binding domain in the dehydrogenases. One of these lobes

TABLE 3

Sequence Comparisons of Nucleotide Binding Domains

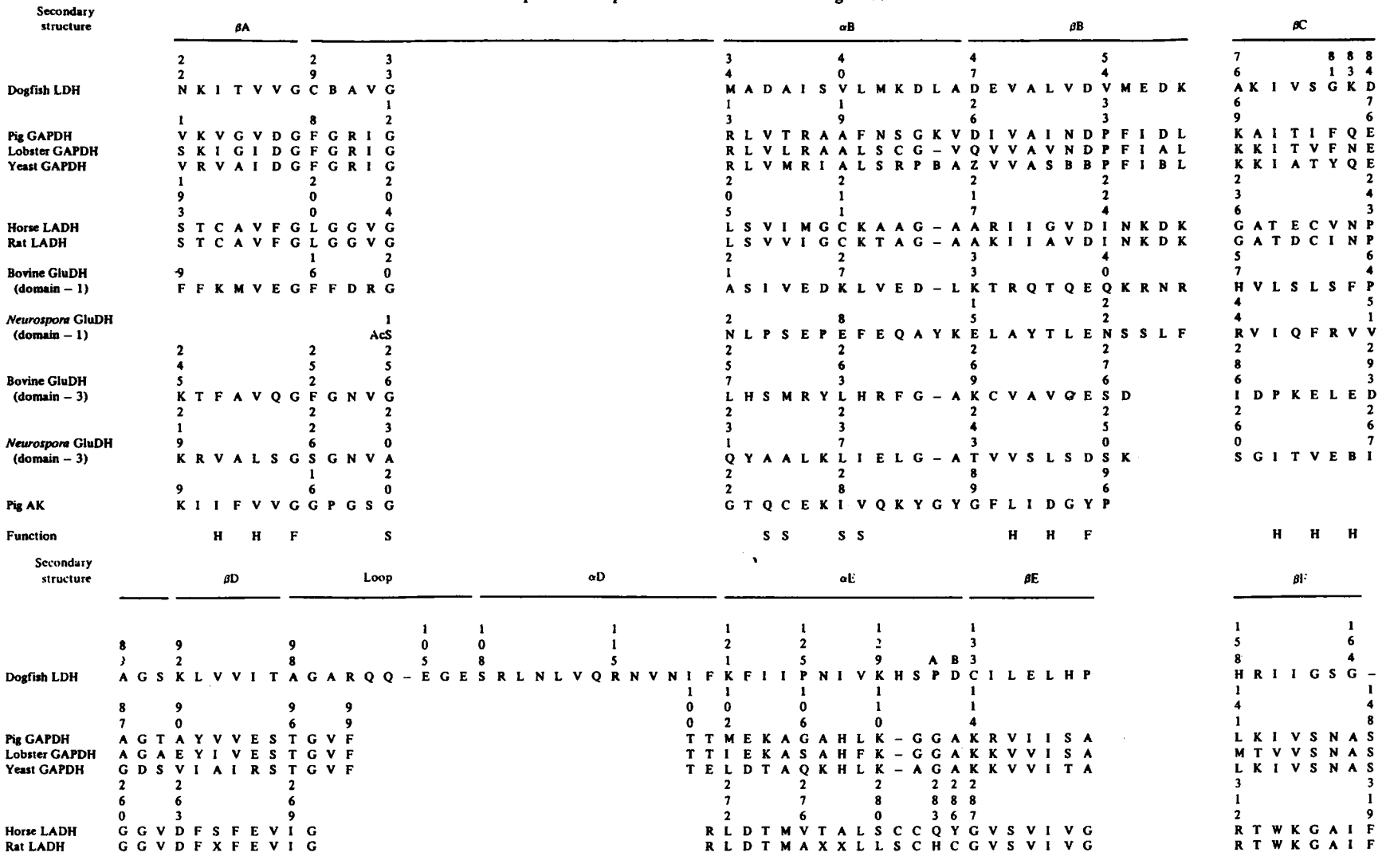


TABLE 3 (continued)

Sequence Comparisons of Nucleotide Binding Domains

Secondary structure		βD	Loop	αD	αE	βE	βF
	6	7	7		8 8 8 9	9	1 2
Bovine GluDH	9	2	8		3 5 9 3	7	5 2
(domain – 1)	D G S	W E V I E G	Y R A Q		S H Q R T P C K G G I – R Y S	T D V S V D E	A V V D V P F G G
	5	5	6		6 7 7 7	8	0 1
Neurospora GluDH	6	9	5		9 1 5 9	4	3 0
(domain – 1)	D G N	V Q V N R G	Y R V Q		F N S A L G P Y K G G L R L H	P S V N L S I	A L T G L S M G G
	2	3	3		3	3	3 3
	9	0	0		1	1	4 5
Bovine GluDH	8	1	7		1	6	4 1
(domain – 3)	H G T	I L G F P K	A K		I Y E G S I L E	V D C D I L I	I A E G A N G P T
	2	2	2		2 2 2	3	3 3
	7	8	8		8 9 9	0	3 3
Neurospora GluDH	7	0	6		9 3 7	2	2 9
(domain – 3)	R Q S	L T S F Q H	A G		H L K W I E G A R P W L H V	G K V D I A L	V A E G S N M G C
	1	1	1 1		1 1 1	1	
	1	1	1 2		5 5 6	6	
	0	3	9 4		5 9 3	8	
Pig AK	G Q P	T L L L Y V	D A G P E T		K A T E P V I A F Y E K R	G I V R K V N	
Function		H H	F	F	S S S S S	H H	H H H

Notes:

1. The LDH, GAPDH, LADH and AK alignments are taken from Rossmann et al.[37] and the GluDH alignments from Wootton.[53]
2. The abbreviations used in denoting the function of the various residues are defined as follows:
   - H represents sheet residues which are in hydrophobic pockets facing helices;
   - S denotes helical residues in this pocket which face the sheet;
   - F defines those residues which have a functional role in the binding of coenzyme.

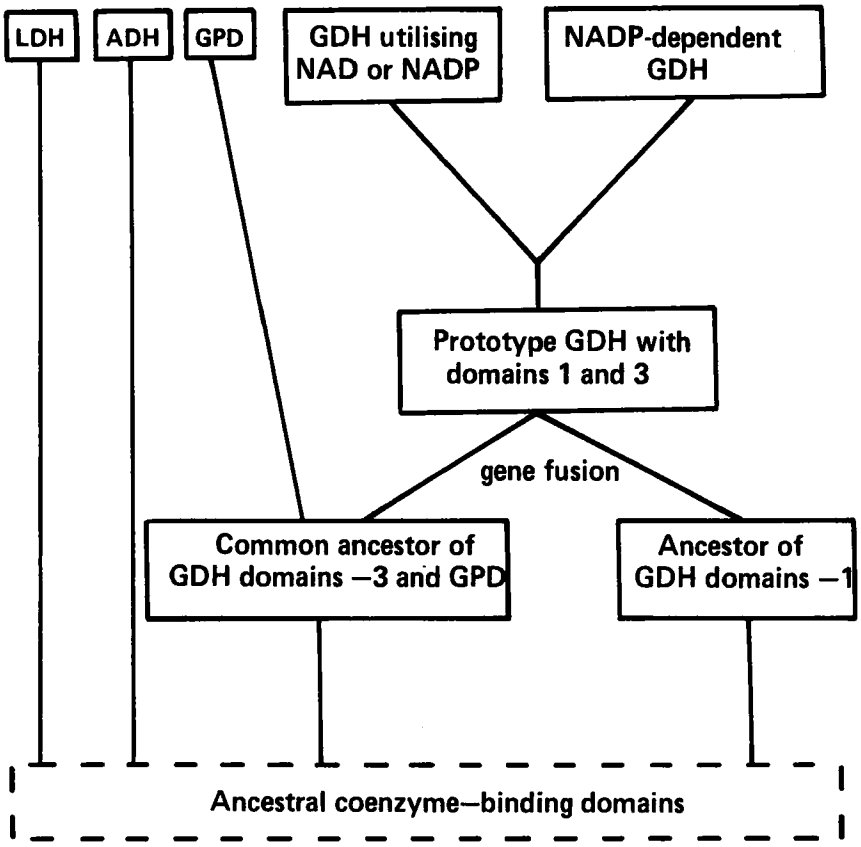


FIGURE 8. Evolutionary scheme for the evolution of dehydrogenases as suggested by Wootton. (From Wootton, J. C., *Nature*, 252, 542, 1974. With permission.)

has been associated with the binding of Mg-ADP, Mn-ADP, Mg-ATP, and ATP.[28] These cofactors bind in a manner similar to the AMP portion of NAD in the dehydrogenases (Figure 9). The metal is liganded to the α and β phosphates. The substrate binding site has been tentatively identified as being near the γ phosphate position in the other lobe. Its position would then correspond roughly to the phosphates of the NAD if this were an NAD binding domain of the dehydrogenases. It is thus of some significance that each subunit of phosphoglycerate mutase possibly has the structure of an NAD binding domain in which the substrate binds to the phosphate sites.[56] Thus, PGK could represent a coenzyme binding domain preceded by a domain specializing in binding phosphoglycerate, where the latter is a modest adaptation of the former.

The coenzyme binding domain of PGK, as also in AK and LDH, has an extended loop between βD and αE. Both PGK and AK have one extra small helix immediately following βD. There is evidence both for PGK[28] and AK[57] that this loop, as in the case of LDH, undergoes conformational changes on binding ligands to the active center. Furthermore, this region in AK has three arginines and is otherwise acidic, which is reminiscent of the corresponding loop peptide in LDH.

Porcine AK has strong structural similarities to PGK and the dehydrogenases,[58] although the sequences of βB and βC are interchanged and βF is missing. A cleft in the molecule suggests the same binding site for the ATP coenzyme and AMP substrate. Schulz et al.[30] have suggested ATP and AMP binding sites by considering catalytic residues and hydrophobic pockets. These positions are consistent with conformational changes (detected by NMR measurements) induced by binding these nucleotides.[57] It would seem probable that ATP binds in a manner comparable to the AMP position of NAD, while AMP corresponds to the NMN position. The larger cleft of AK is necessary to accommodate the required four phosphates rather than the two phosphates of NAD.

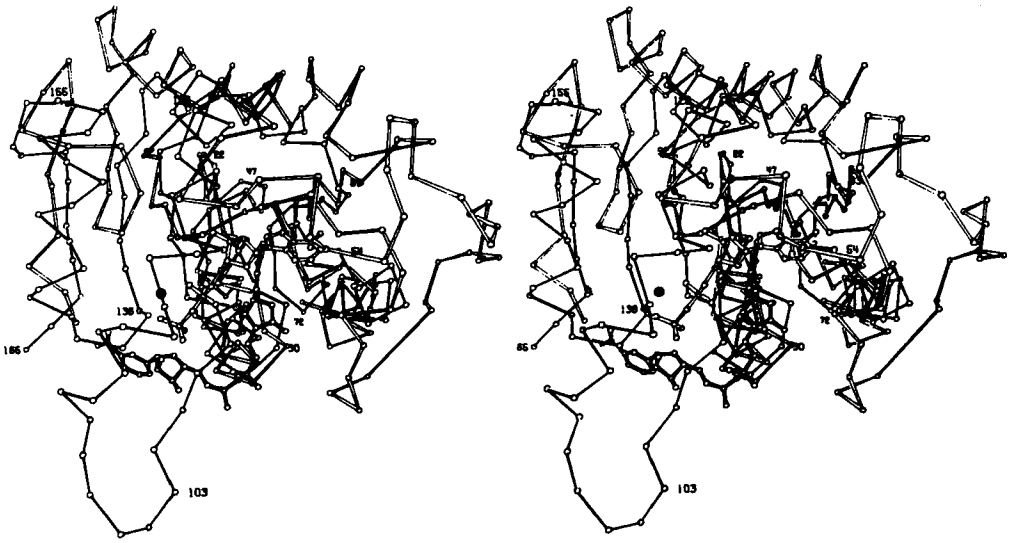


FIGURE 9. Comparison of the ADP binding domain of PGK with the NAD binding domain of LDH. The LDH and PGK structures are shown with solid and open bonds, respectively. (From Rossmann, M. G., Moras, D., and Olsen, K. W., *Nature*, 250, 194, 1974. With permission.)

Rossmann et al.[36] have extended the sequence comparison of the dehydrogenases to include AK. They show that glycine 28 and 33 are conserved. The calculated minimum base changes per codon for structurally equivalent amino acids indicate, however, that AK is more distantly related to the dehydrogenases than the dehydrogenases are to each other. This is in agreement with the decreased structural similarity of AK to the NAD binding domain in dehydrogenases.

Three different subunit associations have been reported for HK.[32-34] Although some substrate and cofactor studies have been made in all cases, a preliminary 2.7-Å resolution structure has been completed for only one of them. However, no definitive results are available at this time. A region which binds AMP and diiodofluorescein[54] is reminiscent of the AMP binding domain in dehydrogenases and other kinases. Thompson et al.[60] have developed an empirical method for recognizing nucleotide binding domains by using blue dextran attached to sepharose. They show that HK does not bind to this column, in contrast to the other enzymes discussed here. This could suggest that HK differs from other kinases and dehydrogenases.

Much discussion has arisen[56,61] as to whether the β-pleated sheet structures in mutases, kinases, and dehydrogenases represent simply an energetically favored conformation or an evolutionary remnant.[26,37,38,58] It suffices to say here that the similarity of structures supports the idea of a common origin for many of the enzymes in the glycolytic pathway. The solution of the structures of enolase and pyruvate kinase will thus be awaited with much interest. Further analysis of the convergent versus divergent evolutionary concepts will be treated in Part II of this review.

II. MOLECULAR TAXONOMY

A. Background

The Linnaean nomenclature of biological species depends on the intuitive recognition of both common and unique characters. The methods of numerical taxonomy depend on the recognition of as many characters as possible.[62] Their presence or absence can then be correlated between different species, permitting the determination of similarity matrices and the drawing of dentograms or Wagner trees. Taxonomists have been careful to avoid suggesting that such trees are necessarily related to divergent evolutionary processes, as similarity might be the consequence of convergence of many characters to provide necessary functions. Schulz et al.[58] made a reasonable attempt to determine the probability of divergence versus convergence in the comparison of certain protein molecules. They treated the position of each strand of a polypeptide chain in

the nucleotide binding protein as an independent character and then computed the probability that the observed character combination might occur by chance. Clearly the greater the number of similar characters between species, the less probable it is that these have all converged to be the same.

During the last two decades there has become available a very large amount of amino acid sequence information on a wide variety of molecules.[63] This has been used extensively to compile phylogenetic trees by comparing sequences of molecules with the same function in differing organisms. Such methods were pioneered particularly by Fitch and Margoliash,[64] who showed the close correlation between phylogenetic trees based on fossil data and on amino acid sequences. Jukes and Cantor,[52] among others, suggested tests for measuring the significance of any given comparisons based on the number of amino acids in the sequence (number of characters). Many, including Fitch and Farris[65] and Dayhoff and Eck,[66] have considered methods of constructing phylogenetic trees from such data. The importance of amino acid sequences is that they are directly related to genetic events. However, it is also apparent that amino acid sequences vary much faster than either the function or the three-dimensional structure of molecules. Thus, to investigate evolutionary processes over the longest possible period of time it is necessary to consider changes in molecular architecture or molecular function. An attempt will now be made to develop methods for measuring the similarity (or dissimilarity) of folding of protein molecules and to use such measurements in the construction of evolutionary trees. The procedure is a taxonomic study of molecular shape.

B. Comparison of Molecular Structure: Method

If two molecules are to be compared, then one molecule must be rotated and translated on top of the other in order to count the structurally equivalent residues. A method for determining structural equivalence will be given below. Most simply it is desired to produce a six-dimensional search of three translational and three rotational parameters and then to plot a map of the number of equivalent residues for each combination of parameters. If there is a single large peak, significantly higher than any other peak, the two structures can be said to have a unique similarity. In practice, however, such a search would be impossibly slow, and instead a reasonable start can be suggested from visual inspection of the molecular backbones. This can then be refined by minimizing the sum of the squares of the distances between equivalenced α carbon atoms.[39,67]

The following programmed sequence has been used here to quantify the similarity among various nucleotide binding structures:

1. Designate at least four pairs of structurally equivalent residues by visual inspection or any other means.
2. Determine a reasonable starting rotation matrix and translation vector.[39,67]
3. Adjust the three Eularian angles and three translational components in order to minimize the sum of the squared distances between C_α atoms of equivalenced residues. Three or four iterative cycles of nonlinear least squares are usually sufficient. The probability of equivalence can be used as weighting among the observational equations.
4. Redetermine structurally equivalenced amino acids between the two molecules with respect to the revised rotational and translational parameters. If there have been any changes go back to step 3, otherwise stop.

In practice, a reasonable alignment was found within the first two cycles around steps 3 and 4. Subsequent refinement causes only the acceptance or rejection of a variety of marginal equivalences. However, there may often be 20 or more iterations before a combination of equivalences has been found which causes no further alterations. The above procedure differs from that used by Ohlsson et al.[38] in that it does not necessarily retain all the originally equivalenced residues or, indeed, even the original alignment.

The determination of structural equivalence between two amino acids itself proceeds in two steps: assignment of probabilities and the imposition of topological conditions of similar folding.

As the first step, the probability, P_{ij}, of equivalence between residue i in the first molecule and residue j in the second molecule was considered as the product of the probabilities P_1, P_2, P_3 relating three independent conditions. Thus,

$$P_{ij} = P_1 P_2 P_3$$

and

$$P_1 \propto \exp - \frac{d_{ij}^2}{2E_1^2}$$

$$P_2 \propto \exp - \frac{S_M^2}{2E_2^2}$$

$$P_3 \propto \exp - \frac{S_R^2}{2E_3^2}$$

Here d_{ij} is the distance between the ith C_α atom on molecule 1 and the jth C_α atom on molecule 2. Thus, P_1 measures the probability of the chance superposition of two residues.

S_M is the rms scatter of the distances $d_{i-1,j-1}$, d_{ij}, $d_{i+1,j+1}$ from their mean. Thus, P_2 determines the probability of similar orientation of the polypeptide chain around a given residue pair.

S_R is the rms scatter of the distances from their mean between all the corresponding atoms of the ith and jth residues in molecules 1 and 2, respectively. P_3 is thus another, possibly not entirely independent, determination of similar orientation of the two paired residues. In many comparisons, $P_3 = 1$ has been found quite adequate.

Values of E_1, E_2, and E_3 are determined as the corresponding rms values among those residues equivalenced in the previous cycle. These values are continuously updated, although they vary little from cycle to cycle. Usual values are $E_1 \cong 2.5$ Å, $E_2 \cong 1.0$ Å, and $E_3 \cong 0.7$ Å when comparing dehydrogenases.

These probabilities thus relate a residue, i, in the first molecule, $R_{1,i}$, and any residue, j, in the second, $R_{2,j}$. They fill a symmetric matrix, from which a sequence must be selected which satisfies the topological requirement of a "similar fold" and the genetic requirement of gene linearity. Thus, if

$$R_{1,i} \equiv R_{2,j}$$

then we must have "the progression rule"

$$R_{1,i+n} \equiv R_{2,j+m}$$

where $n \geqslant 1$ and $m \geqslant 1$. If there are no insertions or deletions, $n = m$.

In the second step of assigning structural equivalences, the probabilities, P_{ij}, are sorted into descending order for each value of i. If the molecules are exceedingly similar, then the maximum probabilities for each value of i should also satisfy the progression rule. If this is not the case, then there will be, nevertheless, local "runs" where the progression rule is satisfied. These runs (a run being two or more consecutive residues where the maximum probabilities obey the progression rule) are identified and associated with their total probability.

Adjacent runs can then be compared in pairs. Both runs are accepted if they obey the progression rule, but if that is not the case, that run with the smaller total probability, T, is rejected. Emphasis is thus given to the longer and more similar folds of the protein. This process of rejection is continued until every run, and hence every residue, obeys the progression rule. Finally the runs are extended at both ends by applying the progression rule. Extension is stopped when either the probabilities fall below a given value (say 0.05) or the extension meets another run. When two extensions meet, the extension with the larger individual probabilities is accepted. The purpose of run extension is to include those parts of the protein fold where there may be amino acid homology in the absence of precise structural equivalence, particularly at bends.

C. Comparison of Molecular Structures: Results

In Table 4 are given the results of comparisons among the NAD binding domains of dogfish LDH, horse LADH, lobster GAPDH and the nucleotide binding proteins porcine AK, flavodoxin, and subtilisin BP'. (See Rossmann et al.[36,37] for discussion of other nucleotide binding proteins.) Initial alignments were determined by inspection of the sheet residues. Coordinates were kindly supplied by Drs. Carl Brändén, Georg Schulz, Martha Ludwig, and Joseph Kraut for LADH, AK, flavodoxin, and subtilisin, respectively. Table 4 shows equivalent sequence numbers. It should be noted that residue 82 and 104 of dogfish LDH are nonexistent[12] and residue 24 of lobster GAPDH is a deletion. Although comparison can be made only in pairs, the alignments were consistent except in a few places. In such cases some very small adjustment was made in order to obtain a single alignment set. The results are, however, not entirely consistent with those given by Rossmann et al.[36,37] or Ohlsson et al.[38] Differences arise at corners between secondary structural elements.

TABLE 4

Structural Equivalence Among Various Nucleotide Binding Proteins

	Comparisons with LDH										Comparisons with LADH				Comparison with AK		
LDH	GAPDH		LADH		AK		Subtilisin		Flavodoxin		GAPDH		Flavodoxin		Subtilisin		Secondary structure
22	1	3.7	193	1.1	9	1.3	26	1.2							26	1.6	
23	2	2.1	194	0.9	10	1.0	27	0.8							27	1.1	
24	3	0.8	195	0.7	11	1.0	28	1.0	1	2.1	3	0.4	1	2.0	28	0.7	
25	4	1.0	196	0.4	12	0.9	29	0.4	2	1.9	4	1.2	2	1.1	29	1.8	βA
26	5	0.6	197	0.8	13	1.1	30	0.7	3	0.8	5	1.0	3	0.7	30	3.1	
27	6	0.6	198	1.1	14	1.1	31	0.7	4	1.3	6	0.9	4	1.1	31	2.4	
28	7	1.1	199	0.2	15		32	0.8			7	1.6			32	3.2	
29	8	0.8	200	1.1			33	2.7			8	1.1					
30	9	0.9	201	1.3			62	4.5			9	1.4					
31	10	0.7	202	1.9			63	4.2	14	2.0	10	1.5	14	2.3			
32	11	1.5	203	1.4			64	4.1	15	0.8	11	2.1	15	0.6			
33	12	1.2	204	0.7			65	3.2	16	0.7	12	1.0	16	0.7	65	4.5	
34	13	0.7	205	0.7	22	1.9	66	1.9	17	1.0	13	0.3	17	0.7	66	4.1	
35	14	1.1	206	1.5	23	2.6	67	1.8	18	0.6	14	1.0	18	1.4	67	1.5	
36	15	1.0	207	1.6	24	1.6	68	2.0	19	0.4	15	1.0	19	1.3	68	2.0	
37	16	0.4	208	1.9	25	1.6	69	0.7	20	0.5	16	1.7	20	1.1	69	1.6	αB
38	17	1.0	209	1.8	26	1.7	70	1.5	21	0.5	17	1.9	21	1.9	70	2.2	
39	18	0.6	210	2.9	27	2.6	71	1.9	22	0.3	18	3.2	22	3.0	71	2.9	
40	19	0.9	211	2.4	28	2.1	72	1.6	23	0.3	19	2.6	23	2.2	72	2.3	
41	20	1.1	212	0.6	29	2.3	73	1.6	24	1.3	20	1.5	24	0.9	73	2.0	
42	21	1.1	213	3.9	30	2.9			25	2.2	21	3.0	25	2.8	83	2.7	
43	22	1.9	214	4.9	31	2.7	84	3.7	26	1.4	22	3.9	26	3.9	84	2.9	
44	23	3.8	215	2.2	32	3.6	85	3.7	27	1.8	23	1.6	27	3.4	85	3.2	
45	25	1.7	216	3.4	33	5.2	86	4.1	28	3.4	25	2.4			86	1.9	
46	26	2.1	217	3.9	34	5.1	87	4.6			26	2.3	28	2.6	87	4.0	
47	27	3.3	218	3.9	89	3.1					27	3.9	29	3.4	88	1.6	
48					90	2.1	89	1.3	29	4.8					89	1.3	
49	28	0.3	219	0.5	91	1.8	90	0.7	30	3.3			30	1.6	90	1.4	
50	29	1.3	220	1.3	92	1.4	90	0.8	31	4.3	29	0.7	31	1.9	91	1.3	βB
51	30	0.6	221	1.2	93	1.4	92	1.3	32	2.9	30	0.7	32	2.3	92	2.6	
52	31	1.4	222	1.1	94		93	1.3	33	2.6	31	2.3	33	1.0	93	2.7	

TABLE 4 (continued)

Structural Equivalence Among Various Nucleotide Binding Proteins

	Comparisons with LDH								Comparisons with LADH		Comparison with AK	
LDH	GAPDH		LADH		AK		Subtilisin		Flavodoxin		GAPDH	Flavodoxin	Subtilisin	Secondary structure
53	32	0.9	223	1.5			96	5.1	34	1.5	32 2.5			
54	33	1.5	224	1.7			97	5.6	35	4.4	33 1.7			
55	34	1.5	225	1.6					36	5.5	34 2.3			
56	36	0.9	226	0.9							38 3.5			
57	37	2.4	227	1.5							39 4.1			
58			228	1.6							40 4.1			
59														
60	38	3.9												
61	39	2.4												
62	40	2.5												
63	41	5.2	229	2.7							41 3.7			
64			230	3.1							42 3.9			
65			231	4.3							43 4.6			αC
66			232	3.5										
67			233	1.6										
68			234	3.8										
69			235	5.9										
70			236	4.7										
71														
72														
73														
74														
75			237								68 7.6			
76	69	2.9	238								69 4.9			
77	70	1.5												
78	71	1.3												
79	72	0.8	239	1.4							72 2.0			
80	73	0.9	240	0.5							73 1.7			βC
81	74	1.5	241	1.0							74 2.2			
82	(non-existent in LDH)													
83	75	1.7	242	2.5	96	4.5			38	2.3	75 1.7			

TABLE 4 (continued)

Structural Equivalence Among Various Nucleotide Binding Proteins

Comparisons with LDH											Comparisons with LADH				Comparison with AK		Secondary structure
LDH	GAPDH		LADH		AK		Subtilisin		Flavodoxin		GAPDH		Flavodoxin		Subtilisin		
84					97	3.1			39	2.1							
85			251	4.9	98	5.8			40	2.0			40	1.8			
86			252	5.0	100	4.2			41	2.2			41	1.8			
87			253	4.0	101	3.8			42	2.9			42	2.4			
88			254	2.9	102	5.5			43	2.4			43	2.3			
89			255	2.2					44	3.6			44	3.9			
90			256		111		118	1.5	45	0.7			45	6.5	118	1.5	
91	89	2.7	262	0.4	112	1.4	119	1.6	46	0.9			46	1.0	119	1.3	
92	90	2.3	263	1.1	113	1.6	120	1.9	47	1.2	90	2.0	47	1.6	120	0.9	
93	91	1.7	264	0.3	114	1.5	121	1.7	48	1.1	91	1.2	48	1.7	121	1.0	
94	92	0.8	265	0.4	115	1.4	122	0.9	49	0.7	92	0.7	49	1.2	122	1.2	βD
95	93	1.3	266	0.7	116	1.1	123	0.7	50	1.4	93	1.1	50	0.9	123	2.7	
96	94	0.6	267	0.9	117	1.0	124	0.8	51	1.0	94	1.5	51	1.3	124	2.9	
97	95	3.5	268	1.3	118	1.5	125	2.9	52	0.4	95	3.9	52	1.5	125	2.5	
98	96	2.1	269	1.9	119	1.7			53	2.4	96	3.9	53	2.0	126	1.8	
99	97	5.6			120	2.1			54	4.3	97	3.8			127	2.9	
100	98	5.5			121						98	3.9			128	2.5	
101																	
102																	loop
103																	
104	(nonexistent in LDH)																
105																	
106																	
107																	
108																	
109																	
110																	
111									56	6.3							
112									57	5.5							
113									58	4.2							
114									59	4.3							αD

TABLE 4 (continued)

Structural Equivalence Among Various Nucleotide Binding Proteins

Comparisons with LDH										Comparisons with LADH				Comparison with AK		
LDH	GAPDH		LADH		AK		Subtilisin		Flavodoxin		GAPDH		Flavodoxin		Subtilisin		Secondary structure
115									60	4.9							
116																	
117									61	3.1							
118					155	8.7			62	4.3							
119					156	8.9			65	3.8							
120			271	2.8	157	7.5			66	3.2							
121			272	3.0	158	5.3			67	1.7			67	2.2	132	7.1	
122			273	2.3	159	3.1			68	1.3			68	3.0	133	8.1	
123			274	2.2	160	4.2			69	2.2			69	4.0	138	4.0	
124	105		275	4.5	161	3.3			70	1.6	105	3.3	70	3.4	139	3.0	
125	106		276	5.1	162	4.3			71	3.3	106	3.9	71	5.1	140	2.6	
126	107		277	2.8					72	3.7	107	4.3	72	5.7			
127	108	1.8	278	2.7					73	3.5	108	2.4	73	5.9			αE
128	109	1.6	279	4.3					74	4.0	109	3.8	74	5.8			
129	110	0.8	280	4.8	163	4.0			75	4.3	110	3.5	75	6.7	141	1.9	
130	111	0.9	281	3.1	164	4.4			76	6.6	111	3.3			142	2.5	
131	112	1.7	282	3.0	165	5.9			77	5.0	112	3.4	77	5.4	143	2.0	
132A	113	4.8	283	2.6	166	6.1			78	5.4	113	4.9	78	7.0	144	2.0	
132B	114	3.2	286	5.0	167	5.0	146	1.5	79	5.4			79	5.6	145	2.0	
133			287	4.3	168		147	1.0					80	4.2	146	1.7	
134	115	0.7	288	3.2	169	1.7	148	0.8	81	1.5	115	2.0	81	1.5	147	2.3	
135	116	0.8	289	0.9	170	1.7	149	1.2	82	1.6	116	0.4	82	1.2	148	1.7	
136	117	1.0	290	1.2	171	1.6	150	1.3	83	1.4	117	1.5	83	2.4			βE
137	118	1.5	291	0.8	172	1.3	151	1.1	84	1.1	118	2.1	84	2.1			
138	119	0.8	292	1.6	173	0.9	152		85	0.5	119	1.1	85	1.6	152	1.9	
139			293	2.6	174		153								153	2.4	
140			294	2.6	175		154								154	3.8	
141					176		155								155	2.8	
142	121	1.8	296						96	3.5							
143	122	2.5	297				165	2.9	97	1.9							
144	123	1.7	298				166	3.0	98	2.4							

TABLE 4 (continued)

Structural Equivalence Among Various Nucleotide Binding Proteins

Comparisons with LDH									Comparisons with LADH		Comparison with AK	
LDH	GAPDH		LADH		AK	Subtilisin		Flavodoxin	GAPDH	Flavodoxin	Subtilisin	Secondary structure
145	124	6.0	299	8.1		167	1.6	99	1.8				
146	125	4.1	300	8.9		168	3.7	100	1.2				
147			301	6.7		169	5.0	101	1.7				
148			302	5.4		170	5.0	102	2.9		αlF		
149			302	5.4		171	4.7	103	3.9				
150			303	5.9		172	4.4	104	3.5				
151			304	5.6				105	2.6				
152			305	7.2									
153													
154													
155								106	3.6				
156								107	3.1				
157	140	5.8						108	5.6				
158	141	2.3	312	2.4		109	4.1	141	3.9				
159	142	2.8	313	2.8		110	5.2	142	2.4				
160	143	2.1	314	2.3				143	2.0				
161	144	1.9	315	2.3		175	0.3	144	1.3		βF		
162	145	0.6	316	2.1		176	1.3	145	1.4				
163	146	1.3				177	1.5						
164	147	1.4				178	2.3						
165	148	5.2				179	4.3						

Note: Left-hand column for each protein is the amino acid sequence number and right-hand column is the distance, in A, between the indicated C_α atoms. Sequence numbers without associated distances show alignment with another column, although these residues were not found to be structurally equivalent with LDH.

TABLE 5

Conformational Parameters of Bound Coenzyme

Parameter	s-MDH (holo)	s-MDH (apo)	LDH (ternary)	LADH (ADPR binary)	GAPDH (holo, average)
$P_a - N9_a$ (A)	6.9	6.9	6.9	6.5	7.1
$P_a - C4'_a$ (A)	4.0	3.9	3.9	3.6	4.3
$P_n - N1_a$ (A)	6.3	6.5	6.7	–	5.5
$P_n - C4'_n$ (A)	3.9	4.0	4.0	3.4	3.6
$C6_a - C2_n$ (A)	14.3	14.2	14.1	–	15.8
$N1_a - N4_n$ (A)	16.7	16.2	16.6	–	17.6
$\alpha°$	−82.8	−81.3	−58.3	–	−85.6

Notes:

a. $C6_a - C6_n$ to compensate for the B-side specificity in GAPDH.
b. α is the dihedral angle between the planes defined by the atoms $N9_a$, P_a, P_n, and by N1, P_a, P_n.

The reader should compare Tables 3 and 4 to observe that differences occur primarily at bends and, to a lesser extent, in helices. For example, the effect on alignment of the deletion in GAPDH of residue 24 (Table 4), relative to other GAPDH species and to LDH, indicates that the purely structural comparisons are probably less reliable at corners. However, systematic differences in the alignment of helices show that structural equivalences might be more accurate here. These observations are underlined by observing the increased distances between equivalenced C_α atoms (Table 4) where the two procedures differ.

In evaluating the possibilities of divergent or convergent evolution in terms of the number of structurally equivalent amino acids, a further character set can often be introduced, namely, the equivalence of a bound coenzyme or prosthetic group. Each parameter describing the conformation of the bound molecule might be considered a character. Hence, an NAD molecule could be considered to represent another ten characters.[68] When the structural relationship between the three dehydrogenases LDH, GAPDH, and LADH was determined as described above, their coenzymes were found to be also closely related. In Table 2 are shown mean, maximum, and minimum distances for corresponding atoms, and in Table 5 are values of parameters which give an overall description of the coenzyme conformation.[15,36] Table 6 gives the distances between equivalent phosphate sites (or the carboxy site of a subtilisin inhibitor) as representing the most conserved position for the general class of nucleotide binding domains.

TABLE 6

Distance in A Between Equivalent Phosphate Sites (or Carboxy Site of Subtilisin Inhibitor)

	LDH	LADH	GAPDH	Flavo	Subt
LDH		3.3	1.8	4.7	5.4
LADH			2.8	1.9	
GAPDH				5.5	
Flavodoxin					
Subtilisin					

D. Establishing Evolutionary Trees for Nucleotide Binding Proteins

Fitch and Margoliash[64] were among the first to develop phylogenetic trees from amino acid sequence differences. Similarly, it should be possible to establish evolutionary trees from taxonomic differences among related proteins.

In Matrix 1 are given the number of structurally equivalent residues between any pair of the six nucleotide binding proteins shown in Table 4. The larger numbers show greater similarity. In constructing trees it is more useful to have smaller numbers to show greatest similarity. Thus, in Matrix 2 all the numbers of Matrix 1 have been subtracted from 143, the number of amino acids in the nucleotide binding domain of LDH. This gives some "standardization" with respect to the LDH structure. A tree can now be constructed

Matrix 1

Similarity matrix: Number of structurally equivalent residues between pairs of nucleotide binding proteins.

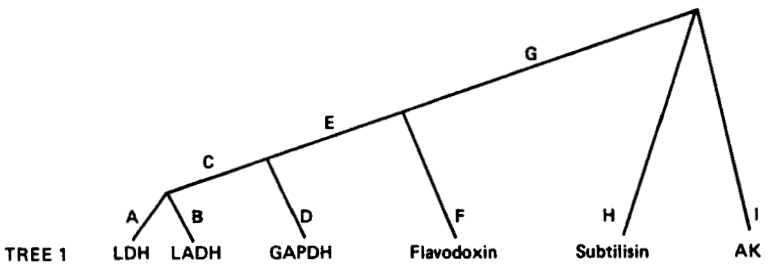
	LDH	LADH	GAPDH	Flavo	Subt	AK
LDH		99	93	85	73	67
LADH			79	71	64	61
GAPDH				65	62	55
Flavodoxin					59	61
Subtilisin						56
AK						

Matrix 2

Dissimilarity matrix: Number of amino acids in the LDH nucleotide binding protein (143) minus the number of equivalent residues between any pair of proteins.

	LDH	LADH	GAPDH	Flavo	Subt	AK
LDH		44	50	58	70	76
LADH			64	72	79	82
GAPDH				78	81	88
Flavodoxin					84	82
Subtilisin						82
AK						

which is most parsimonious with the data in Matrix 2. It contains the implied assumption that evolutionary processes proceed by the least number of steps. Visual inspection suggests a tree such as



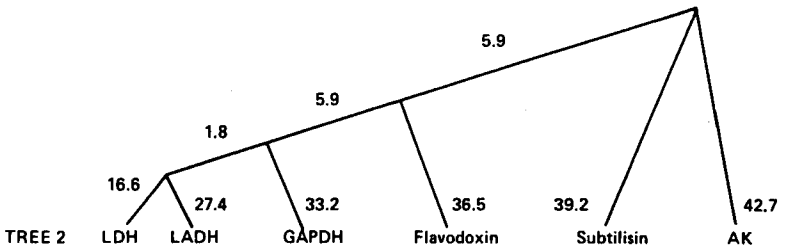
where A, B, C, D, E, F, G, H, I are the lengths of the arms. It is now possible to set up 15 observational equations of the form

A + B = 44 (relating LDH and LADH)

or

B + C + D = 64 (relating LADH and GAPDH) etc.

A linear least squares solution can then be found for the nine parameters A, B, C, . . . I. When this procedure is applied to the data in Matrix 2, it produces Tree 2.



Matrix 3

Dissimilarity matrix: Based on distances given in Evolutionary Tree 2.

	LDH	LADH	GAPDH	Flavo	Subt	AK
LDH		44.0	51.6	60.5	69.2	72.5
LADH			62.4	71.6	79.8	83.3
GAPDH				75.6	83.8	87.3
Flavodoxin					81.2	84.7
Subtilisin						81.9
AK						

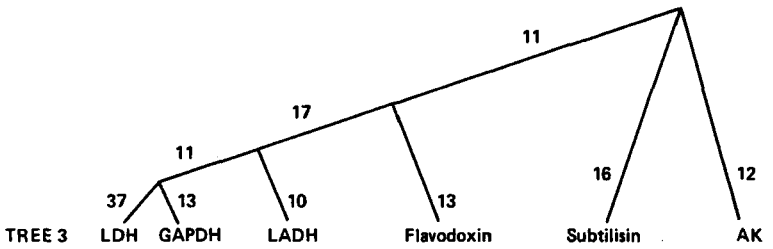
Matrix 4

Differences between observed Matrix 2 and computed Matrix 3.

	LDH	LADH	GAPDH	Flavo	Subt	AK
LDH		0.0	−1.6	−2.5	0.8	3.5
LADH			1.6	0.4	−0.8	−1.3
GAPDH				2.4	−2.8	0.6
Flavodoxin					2.8	−2.7
Subtilisin						0.1
AK						

Tree 2 can then be translated into the "calculated" dissimilarity Matrix 3. When the observed Matrix 2 is subtracted from Matrix 3, the differences Δ shown in Matrix 4 are obtained. The agreement between the tree and the matrix can then be expressed as $\Sigma|\Delta|/\Sigma$(observed elements) and is found to be 2.2%. Any other tree gives larger values, which shows the above tree to be the most parsimonious.

An alternative approach to constructing a tree is the ancestral structure technique (corresponding to the ancestral sequence technique). This has been used in the analysis of amino acid sequences by Dayhoff et al.[69] and by Boulter et al.[70] Farris[71] has also used a similar approach in studying classical taxonomic data. We have used the latter's method with a program devised by F. J. Rohlf of SUNY at Stony Brook. Here the objective is to find a set of "hypothetical taxonomic units" for each node in the network which minimizes the difference of this set with respect to the three surrounding sets. In terms of this jargon, the amino acids of the known structures are the "operational taxonomic units," whereas the historical sequences at the nodes are the "hypothetical taxonomic units." By this method Tree 3 was determined. The lengths along each arm now represent the number of amino acids that are *different* between any two nucleotide binding proteins. The weakness of this approach is seen in Matrix 5 (based on Table 3). The following matrix counts only those residues that are structurally dissimilar between any two proteins after comparison with LDH. Insertions or



Matrix 5

The number of dissimilar amino acids between pairs of nucleotide binding proteins shown in Table 3.

	LDH	LADH	GAPDH	Flavo	Subt	AK
LDH		44	50	58	70	76
LADH			34	40	44	44
GAPDH				50	40	50
Flavodoxin					40	36
Subtilisin						28
AK						

deletions which do not relate to LDH are essentially ignored. For example, the numbers in Matrix 5 do not refer to the first eight AK residues which cannot be compared with LDH. In contrast, the numbers in Matrix 1 and 2 relate to the structurally similar fragments. The method does produce the possible ancestral Tree 3, which, however, is only one of many probable trees. When the structure and sequence of a larger number of nucleotide binding proteins become available, then ancestral structures can also be used to determine ancestral sequences. Such sequences relate, presumably, to precellular proteins in which the amino acid composition may have been different. Those amino acids present might perhaps correlate to an earlier genetic code[72] and to amino acids synthesized in models of a primitive Earth atmosphere, in some meteorites, or on lunar soil.[73,74]

Both methods produce reasonable results (Trees 2 and 3). The dehydrogenases are the most closely associated, radiating from almost a single node in Tree 2. Furthermore, in Tree 2 (which probably represents the better tree) LDH and LADH, both of which are specific to the A-side of the nicotinamide ring, are more closely related than the B-side specific GAPDH. It has been suggested that A- and B-side specificity is related to evolutionary development.[41] On the other hand, Tree 3 links more closely LDH and GAPDH. This might reflect the special properties of LADH, such as its need for both catalytic and structural zinc atoms,[19] as well as the position of the NAD binding domain towards the end rather than at the beginning of the polypeptide chain. It would also seem meaningful that the proteins involved in energy transfer pathways (LDH, GAPDH, LADH, and flavodoxin) are clustered more closely together.

The trees determined here are similar to those given by Buehner et al.[22] (see Figure 7) and Wootton[53] (see Figure 8). However, by quantization it is possible to make a more precise analysis not possible by visual inspection of structure or by functional analysis of amino acid sequences.

E. Evidence for Evolutionary Divergence in the Genetic Code

It would seem reasonable to assume that those residues which bear greater structural equivalence are also the more likely to be evolutionarily related. Unfortunately, however, there is almost no trace of a common ancestor in the amino acid sequence of distantly related proteins. If, however, the genetic code is considered, there are three chances that one of the bases has remained unchanged as opposed to only one chance that the

TABLE 7

Minimum Base Change per Codon (MBC/Codon) for Structurally Equivalent Residues Between LDH and GAPDH

a. In terms of separation of C_α atoms:

Distance	0–1 A	1–2 A	2–3 A	3 or more A
MBC/Codon	1.26	1.20	1.46	1.31
Number	23	31	13	16

b. In terms of probability of equivalence (E_1 = 2.5 A, E_2 = 1.0 A):

Probability	1.0–0.7	0.7–0.4	0.4–0.0
MBC/Codon	1.10	1.69	1.36
Number	42	13	28

amino acid is unaltered. Thus, the minimum base change per codon has often been used to assess distant relationships among proteins.[52,65,75]

In Tables 7 and 8 a comparison has been made between minimum base changes per codon as a function of probable structural equivalence. Estimates of equivalence are made both in terms of the separation of the equivalent C_α atoms and in terms of the probability calculations. The greater correspondence between the genetic code for those residues with better structural equivalence is particularly good in terms of the probability calculations. Thus, structural and genetic characters have good correlation, strengthening the case for evolutionary divergence. While Table 7 compares LDH and GAPDH, Table 8 shows a similar situation for the evolution of a heme binding pocket as seen in the globins and cytochrome b_5.[67] Apparently, therefore, this is a general property of structurally related proteins.

Tables 7 and 8 and Figure 10 show other characteristics when comparing protein structures. These are (1) that there are progressively fewer equivalenced atoms as the distance between C_α atoms increases and (2) that the number of equivalenced atoms decreases as the distance between C_α atoms approaches zero. The first of these is a consequence of the definition of structural equivalence, while the second is a consequence of error in measurement of the C_α atoms. If there is error it is highly improbable that two atoms are placed directly on top of each other in space. The maximum between these limiting conditions relates to the accuracy of the coordinates used in the comparison and the degree of equivalence between the structures compared. Tables 7 and 8 also show that there remains a residual in the genetic code corresponding to an ancient nucleotide binding protein. Hence, it may

TABLE 8

Minimum Base Change per Codon (MBC/Codon) for Structurally Equivalent Residues Between Cytochrome b_5 and Horse Hemoglobin Chain

a. In terms of separation of C_α atoms:

Distance	0–2 A	2–4 A	4 or more A
MBC/Codon	1.21	1.25	1.75
Number	14	20	12

b. In terms of probability of equivalence (E_1 = 4.0 A, E_2 = 0.7 A):

Probability	1.0–0.7	0.7–0.4	0.4–0.0
MBC/Codon	1.10	1.26	1.65
Number	10	13	17

be possible to search present-day genes for other residual information. One such example is given in Table 9, where the start and stop codons are related to the initial and final amino acid codes of various nucleotide binding proteins. Clearly one base had to be changed during gene fusion in order to eliminate the start and stop. The significant observation is that apparently only one base has changed among a number of sequences consistent with the mean minimum base change of not much more than one per codon. Should this be a correct analysis, it would explain why there remains a reactive cysteine at the end of the nucleotide binding domain of LDH which is not important in the catalytic mechanism.

As a speculative conclusion to this article, we would like to suggest that the possibility exists that the catalytic domains of dehydrogenases represent a frameshift mutation of the nucleotide binding domain and could be tested in view of the apparent conserved residual in the base sequences. This hypothesis would explain the repetition of the "dinucleotide binding fold" in PGK (i.e., no

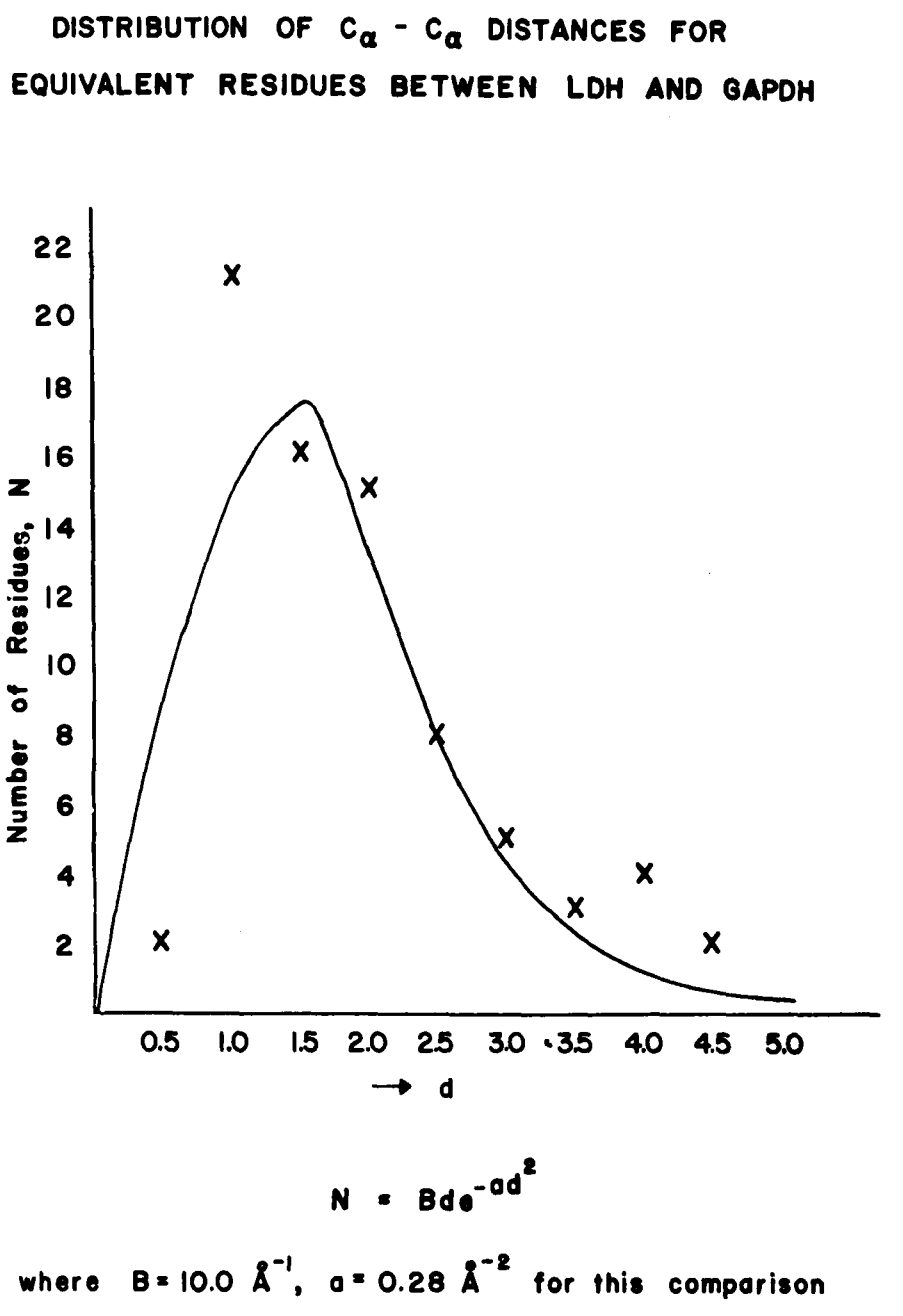


FIGURE 10. Frequency (N) of distances (d) between equivalenced residues of LDH and GAPDH. Curve shows the best fit of the observed points with the theoretically reasonable distribution curve $N = Bde^{-ad^2}$, where B and a are constants. Values found were $B = 10.0\ A^{-1}$ and $a = 0.28\ A^{-2}$.

TABLE 9

Evidence for Gene Fusion

Starting Codes for Nucleotide Binding Protein Gene

Protein	Amino acid	Possible codons	Start codon
LDH dogfish	Lys (23)	AAG	
GAPDH pig	Lys (2)	AAG	
lobster	Lys (2)	AAG	
yeast	Arg (2)	AAG	
LADH horse	Thr (194)	ACG	AUG
GluDH–domain 1			
bovine	Phe (10)	UUC or UUU	
Neurospora	–	–	
GluDH–domain 3			
bovine	Thr (246)	ACG	
Neurospora	Arg (220)	AAG	
AK horse	Ile (10)	AUU, AUC or AUA	
Subtilisin	Val (121)	GUG	

Stop Codes for Nucleotide Binding Protein Gene

Protein	Amino acid	Possible codons	End codon
LDH dogfish	Cys (165)	UGU or UGC	
GAPDH pig	Cys (149)		
lobster	Cys (149)	UGU or UGC	
yeast	Cys (149)		UGA
LADH horse	Gly (320)	GGA	
Glu DH–domain 1			
bovine	Gly (124)	GGA	
Neurospora	Gly (111)		
GluDH–domain 3			
bovine	Thr (351)	ACA	
Neurospora	Cys (340)	UGU or UGC	

frameshift has occurred in copying the primordial gene), the similarity of size of the catalytic and nucleotide binding domains in GAPDH and LDH, and the two sequentially similar structures in the catalytic domain of LDH which would correspond to the two sequential mononucleotide binding domains.

III. CONCLUSION

Structural comparisons among calcium binding proteins,[76,77] iron sulfur proteins,[74,78] heme binding proteins,[67] as well as the similarity between superoxide dismutase[79] and immunoglobulins have led to evolutionary relationships similar to those described here. It is probable that in the future the comparison of protein molecules will be studied with ever-increasing vigor. The reward will be an understanding of the evolution of protein families and, by implication, the processes leading to the simplest prokaryotes. In addition, these studies will provide an appreciation of the important functional groups in conserved domains and the nature of the folding mechanism in polypeptides.

REFERENCES

1. Waley, S. G., Some aspects of the evolution of metabolic pathways, *Comp. Biochem. Physiol.*, 30, 1, 1969.
2. Yčas, M., On earlier states of the biochemical system, *J. Theor. Biol.*, 44, 145, 1974.
3. Woese, C. R., *The Genetic Code: The Molecular Basis for Genetic Expression*, Harper and Row, New York, 1967, 179.
4. Woese, C. R., On the evolution of the genetic code, *Proc. Natl. Acad. Sci. U.S.A.*, 54, 1546, 1965.
5. Horowitz, N. H., The evolution of biochemical syntheses, *Proc. Natl. Acad. Sci. U.S.A.*, 31, 153, 1945.
6. Horowitz, N. H., The evolution of biochemical syntheses – retrospect and prospect, in *Evolving Genes and Proteins*, Bryson, V. and Vogel, H. J., Eds., Academic Press, New York, 1965, 15.
7. Watts, D. C., Evolutionary implications of enzyme structure and function, in *Studies in Comparative Biochemistry*, Munday, K. A., Ed., Pergamon Press, London, 1965, 162.
8. Haldane, J. B. S., Data needed for a blueprint of the first organism, in *Origins of Pre-biological Systems and of Their Molecular Matrices*, Fox, S. W., Ed., Academic Press, New York, 1965, 11.
9. Baltscheffsky, H., A new hypothesis for the evolution of biological electron transport, *Origins of Life*, 5, 387, 1974.
10. Adams, M. J., Ford, G. C., Koekoek, R., Lentz, P. J., Jr., McPherson, A., Jr., Rossmann, M. G., Smiley, I. E., Schevitz, R. W., and Wonacott, A. J., Structure of lactate dehydrogenase at 2.8 A resolution, *Nature*, 227, 1098, 1970.
11. Rossmann, M. G., Adams, M. J., Buehner, M., Ford, G. C., Hackert, M. L., Lentz, P. J., Jr., McPherson, A., Jr., Schevitz, R. W., and Smiley, I. E., Structural constraints on possible mechanisms of lactate dehydrogenase as shown by high resolution studies of the apo-enzyme and a variety of enzyme complexes, *Cold Spring Harbor Symp. Quant. Biol.*, 36, 179, 1971.
12. Holbrook, J. J., Liljas, A., Steindel, S. J., and Rossmann, M. G., Lactate dehydrogenase, in *The Enzymes*, 3rd ed., Vol. 11, Boyer, P. D., Ed., Academic Press, New York, 1975, in press.
13. Taylor, S. S., Oxley, S. S., Allison, W. S., and Kaplan, N. O., Aminoacid sequence of dogfish M_4 lactate dehydrogenase, *Proc. Natl. Acad. Sci. U.S.A.*, 70, 1790, 1973.
14. Taylor, S. S., Allison, W. S., and Kaplan, N. O., The amino acid sequence of the tryptic peptides isolated from dogfish M_4 lactate dehydrogenase, *J. Biol. Chem.*, 1975, submitted for publication.
15. Webb, L. E., Hill, E. J., and Banaszak, L. J., Conformation of nicotinamide adenine dinucleotide bound to cytoplasmic malate dehydrogenase, *Biochemistry*, 12, 5101, 1973.
16. Hill, E. J., Tsernoglou, D., Webb, L. E., and Banaszak, L. J., Polypeptide conformation of cytoplasmic malate dehydrogenase from an electron density map at 3.0 A resolution, *J. Mol. Biol.*, 72, 577, 1972.
17. Brändén, C.-I., Eklund, H., Nordström, B., Boiwe, T., Söderlund, G., Zeppezauer, E., Ohlsson, I., and Åkeson, A., Structure of liver alcohol dehydrogenase at 2.9 A resolution, *Proc. Natl. Acad. Sci. U.S.A.*, 70, 2439, 1973.
18. Abdallah, M. A., Biellmann, J.-F., Nordström, B., and Brändén, C.-I., The conformation of adenosine diphosphoribose and 8-bromo adenosine diphosphoribose when bound to liver alcohol dehydrogenase, *Eur. J. Biochem.*, 50, 475, 1975.
19. Eklund, H., Nordström, B., Zeppezauer, E., Söderlund, G., Ohlsson, I., Boiwe, T., and Brändén, C.-I., The structure of horse liver alcohol dehydrogenase, *FEBS Lett.*, 44, 200, 1974.
20. Brändén, C.-I., Jörnvall, H., Eklund, H., and Furugren, B., Alcohol dehydrogenase, in *The Enzymes*, 3rd ed., Vol. 11, Boyer, P. D., Ed., Academic Press, New York, 1975, in press.
21. Jörnvall, H., Horse liver alcohol dehydrogenase: on the primary structures of the isoenzymes, *Eur. J. Biochem.*, 16, 41, 1970.
22. Buehner, M., Ford, G. C., Moras, D., Olsen, K. W., and Rossmann, M. G., D-Glyceraldehyde-3-phosphate dehydrogenase: three-dimensional structure and evolutionary significance, *Proc. Natl. Acad. Sci. U.S.A.*, 70, 3052, 1973.
23. Buehner, M., Ford, G. C., Moras, D., Olsen, K. W., and Rossmann, M. G., Three-dimensional structure of D-glyceraldehyde-3-phosphate dehydrogenase, *J. Mol. Biol.*, 90, 25, 1974.
24. Jones, G. M. T. and Harris, J. I., Glyceraldehyde-3-phosphate dehydrogenase: amino acid sequence of enzyme from bakers yeast, *FEBS Lett.*, 22, 185, 1972.
25. Moon, K., Piszkiewcz, D., and Smith, E. L., Glutamate dehydrogenase: amino acid sequence of the bovine enzyme and comparison with that from chicken liver, *Proc. Natl. Acad. Sci. U.S.A.*, 69, 1380, 1972.
26. Wootton, J. C., Chambers, G. K., Holder, A. A., Baron, A. J., Taylor, J. G., Fincham, J. R. S., Blumenthal, K. M., Moon, K., and Smith, E. L., Amino acid sequence of NADP-specific glutamate dehydrogenase of *Neurospora crassa*, *Proc. Natl. Acad. Sci. U.S.A.*, 71, 4361, 1974.
27. Moon, K. and Smith, E. L., Sequence of bovine liver glutamate dehydrogenase. VIII. Peptides produced by specific chemical cleavages; the complete sequence of the protein, *J. Biol. Chem.*, 248, 3082, 1973.
28. Blake, C. C. F. and Evans, P. R., Structure of horse muscle phosphoglycerate kinase: some results on the chain conformation, substrate binding and evolution of the molecule from a 3 A Fourier map, *J. Mol. Biol.*, 84, 585, 1974; private communication, 1975.
29. Bryant, T. N., Watson, H. C., and Wendell, P. L., Structure of yeast phosphoglycerate kinase, *Nature*, 247, 14, 1974.
30. Schulz, G. E., Elzinga, M., Marx, F., and Schirmer, R. H., Three-dimensional structure of adenyl kinase, *Nature*, 250, 120, 1974.

31. **Heil, A., Müller, G., Noda, L., Pinder, T., Schirmer, H., Schirmer, I., and von Zabern, I.,** The amino acid sequence of porcine adenylate kinase from skeletal muscle, *Eur. J. Biochem.*, 43, 131, 1974.
32. **Steitz, T. A., Fletterick, R. J., and Hwang, K. J.,** Structure of yeast hexokinase. II. A 6 Å resolution electron density map showing molecular shape and heterologous interaction of subunits, *J. Mol. Biol.*, 78, 551, 1973.
33. **Anderson, W. F., Fletterick, R. J., and Steitz, T. A.,** Structure of yeast hexokinase. III. Low resolution structure of a second crystal form showing a different quaternary structure, heterologous interaction of subunits and subunit binding, *J. Mol. Biol.*, 86, 261, 1974.
34. **Fletterick, R. J., Bates, D. J., and Steitz, T. A.,** The structure of a yeast hexokinase monomer and its complexes with substrates at 2.7 Å resolution, *Proc. Natl. Acad. Sci. U.S.A.*, 72, 38, 1975.
35. **Rossmann, M. G. and Liljas, A.,** Recognition of structural domains in globular proteins, *J. Mol. Biol.*, 85, 177, 1974.
36. **Rossmann, M. G., Liljas, A., Brändén, C.-I., and Banaszak, L. J.,** Evolutionary and structural relationships among dehydrogenases, in *The Enzymes*, 3rd ed., Vol. 11, Boyer, P. D., Ed., Academic Press, New York, 1975, in press.
37. **Rossmann, M. G., Moras, D., and Olsen, K. W.,** Chemical and biological evolution of a nucleotide binding protein, *Nature*, 250, 194, 1974.
38. **Ohlsson, I., Nordström, B., and Brändén, C.-I.,** Structural and functional similarities within the coenzyme binding domains of dehydrogenases, *J. Mol. Biol.*, 89, 339, 1974.
39. **Rao, S. T. and Rossmann, M. G.,** Comparison of super-secondary structures in proteins, *J. Mol. Biol.*, 76, 241, 1973.
40. **Fondy, T. P. and Holohan, P. C.,** Structural similarities within groups of pyridine nucleotide-linked dehydrogenases, *J. Theor. Biol.*, 31, 229, 1971.
41. **Biellmann, J.-F. and Rosenheimer, N.,** Dogfish lactate dehydrogenase: the stereochemistry of hydrogen transfer, *FEBS Lett.*, 34, 143, 1973.
42. **Rossmann, M. G., Adams, M. J., Buehner, M., Ford, G. C., Hackert, M. L., Liljas, A., Rao, S. T., Banaszak, L. J., Hill, E. J., Tsernoglou, D., and Webb, L. E.,** Molecular symmetry axes and subunit interfaces in certain dehydrogenases, *J. Mol. Biol.*, 76, 533, 1973.
43. **Harris, J. I.,** Structure and catalytic activity of alcohol dehydrogenases, *Nature*, 203, 30, 1964.
44. **Fondy, T. P., Everse, J., Driscoll, G. A., Castillo, F., Stolzenbach, F. E., and Kaplan, N. O.,** The comparative enzymology of lactic dehydrogenase. IV. Function of sulfhydryl groups in lactic dehydrogenases and the sequence around the essential group, *J. Biol. Chem.*, 240, 4219, 1965.
45. **Dayhoff, M. O., Barker, W. C., and Hardman, J. K.,** Evolutionary conservatism: enzyme structure and function, *Atlas Protein Sequence Struct.*, 5, 53, 1972.
46. **Jörnvall, H.,** Horse liver alcohol dehydrogenase: the primary structure of the protein chain of the ethanol-active isoenzyme, *Eur. J. Biochem.*, 16, 25, 1970.
47. **Engel, P. C.,** Evolution of enzyme regulator sites: evidence for partial gene duplication from amino acid sequence of bovine glutamate dehydrogenase, *Nature*, 241, 118, 1973.
48. **Smith, E. L., Landon, M., Piszkiewcz, D., Brattin, W. J., Langley, T. J., and Melamed, M. D.,** Bovine liver glutamate dehydrogenase: tentative amino acid sequence; identification of a reactive lysine; nitration of a specific tyrosine and loss of allosteric inhibition of guanosine triphosphate, *Proc. Natl. Acad. Sci. U.S.A.*, 67, 724, 1970.
49. **Williams, J. and Wilkins, A. G.,** On the evidence for partial gene duplication from amino acid sequence of bovine glutamate dehydrogenase, *Nature*, 247, 556, 1974.
50. **Engel, P. C.,** Sequence homologies among pyridine nucleotide-linked dehydrogenases: possible partial gene duplications in glyceraldehyde-3-phosphate dehydrogenase, *FEBS Lett.*, 33, 151, 1973.
51. **Bennett, C. D.,** Similarity in the sequence of *Escherichia coli* dihydrofolate reductase with other pyridine nucleotide-requiring enzymes, *Nature*, 248, 67, 1974.
52. **Jukes, T. H. and Cantor, C. R.,** Evolution of protein molecules, in *Mammalian Protein Metabolism*, Munro, H. N., Ed., Academic Press, New York, 1969, 21.
53. **Wootton, J. C.,** The coenzyme-binding domains of glutamate dehydrogenases, *Nature*, 252, 542, 1974.
54. **Chou, P. Y. and Fasman, G. D.,** Conformational parameters for amino acids in helical, β-sheet and random coil regions calculated from proteins, *Biochemistry*, 13, 211, 1974.
55. **Chou, P. Y. and Fasman, G. D.,** Prediction of protein conformation, *Biochemistry*, 13, 222, 1974.
56. **Campbell, J. W., Watson, H. C., and Hodgson, G. I.,** Structure of yeast phosphoglycerate mutase, *Nature*, 250, 301, 1974.
57. **Cohen, M.,** Magnetic resonance studies of specificity in binding and catalysis of phosphotransferases, *CIBA Found. Symp.*, 1975, in press.
58. **Schulz, G. E. and Schirmer, R. H.,** Topological comparison of adenyl kinase with other proteins, *Nature*, 250, 142, 1974.
59. **Wasserman, P. M. and Lentz, P. J., Jr.,** The interaction of tetra-iodofluorescein with dogfish muscle lactate dehydrogenase: a chemical and X-ray crystallographic study, *J. Mol. Biol.*, 60, 509, 1971.
60. **Thompson, S. T., Cass, K. H., and Stellwagen, E.,** Blue dextran-sepharose: an affinity column for the dinucleotide fold in proteins, *Proc. Natl. Acad. Sci. U.S.A.*, 72, 669, 1975.
61. **Steitz, T. A.,** private communication, 1975.
62. **Sokal, R. A. and Sneath, P. H. A.,** *Principles of Numerical Taxonomy*, W. H. Freeman and Company, San Francisco, 1963.

63. **Dayhoff, M. O.,** *Atlas Protein Sequence Struct.*, 5, 1972.
64. **Fitch, W. M. and Margoliash, E.,** Construction of phylogenetic trees, *Science*, 155, 279, 1967.
65. **Fitch, W. M. and Farris, J. S.,** Evolutionary trees with minimum nucleotide replacements from amino acid sequences, *J. Mol. Evol.*, 3, 263, 1974.
66. **Dayhoff, M. O. and Eck, R. V.,** Tracing biochemical evolution, *Atlas Protein Sequence Struct.*, 5, 1, 1972.
67. **Rossmann, M. G. and Argos, P.,** A comparison of the heme binding pocket in globins and cytochrome b_5, *J. Biol. Chem.*, 1975, in press.
68. **Chandrasekhar, K., McPherson, A., Jr., Adams, M. J., and Rossmann, M. G.,** Conformation of coenzyme fragments when bound to lactate dehydrogenase, *J. Mol. Biol.*, 76, 503, 1973.
69. **Dayhoff, M. O., Park, C. M., and McLaughlin, P. J.,** Building a phylogenetic tree: cytochrome c, *Atlas Protein Sequence Struct.*, 5, 7, 1972.
70. **Boulter, D., Ramshaw, J. A. M., Thompson, E. W., Richardson, M., and Brown, R. H.,** A phylogeny of higher plants based on the amino acid sequences of cytochrome c and its biological implications, *Proc. R. Soc. Lond.*, B181, 441, 1972.
71. **Farris, J. S.,** Methods for computing Wagner trees, *Syst. Zool.*, 19, 83, 1970.
72. **Jukes, T. H.,** Possibilities for the evolution of the genetic code from a preceding form, *Nature*, 246, 22, 1973.
73. **Rohlfing, D. L.,** Evolution of models for evolution, in *The Origin of Life and Evolutionary Biochemistry*, Dose, K., Fox, S. W., Deborin, G. A., and Pavlovskaya, T. E., Eds., Plenum Press, New York, 1974, 397.
74. **Hall, D. O., Cammack, R., and Rao, K. K.,** The iron-sulphur proteins: evolution of a ubiquitous protein from model systems to higher organisms, in *Origin of Life*, Dose, K., Fox, S. W., Deborin, G. A., and Pavlovskaya, T. E., Eds., Plenum Press, New York, 1973, 363.
75. **Fitch, W. M.,** Toward defining the course of evolution: minimum change for a specific tree topology, *Syst. Zool.*, 20, 406, 1971.
76. **Tufty, R. M. and Kretsinger, R. H.,** Troponin and parvalbumin calcium binding regions predicted in myosin light chain and T4 lysozyme, *Science*, 187, 167, 1975.
77. **Weeds, A. G. and McLachlan, A. D.,** Structural homology of myosin alkali light chains, troponin C and carp calcium binding protein, *Nature*, 252, 646, 1974.
78. **Carter, C. W., Jr., Kraut, J., Freer, S. T., Alden, R. A., Sieker, L. C., Adman, E., and Jensen, L. H.,** A comparison of Fe_4S_4 clusters in high potential iron protein and in ferredoxin, *Proc. Natl. Acad. Sci. U.S.A.*, 69, 3526, 1972.
79. **Richardson, D. C.,** private communication, 1975.