PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Evolution of Glycolytic Enzymes [and Discussion]

M. G. Rossman, A. D. B. Malcolm and W. F. Bodmer

Phil. Trans. R. Soc. Lond. B 1981 **293**, 191-203
doi: 10.1098/rstb.1981.0072

Phil. Trans. R. Soc. Lond. B **293**, 191–203 (1981)
Printed in Great Britain

Evolution of glycolytic enzymes

By M. G. Rossman

Department of Biological Sciences, Purdue University, West Lafayette, Indiana 47907, *U.S.A.*

The requirements for glycolysis are examined in relation to other essential metabolic processes in the most primitive organisms. The construction of more complex enzymes from primitive domain building blocks is assessed with respect to glycolytic enzymes. Special attention is given to the evolution of the NAD binding domain in dehydrogenases and the related, frequently observed nucleotide binding domain. An attempt is made to differentiate between convergence and divergence of frequently observed domains. Consideration is given to the structure–function relation of these domains and the development of quaternary structure in later stages of evolution. Some attention is also given to the evolution of structural adaptation to extreme environments as a means of differentiating between essential functions and specific modifications.

Introduction

Three modes for the evolution of enzymes have been suggested from time to time. The first is the independent development of classes, each catalysing different types of reaction (e.g. hydrogen transfer, phosphate transfer, isomerization), 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 (1969) 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 (1974), 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 that 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 (Woese 1967, p. 179). 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 (Yčas 1974; Woese 1965).

The second possibility has been supported by Horowitz (1945). 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 as well as the potential intermediates of a given pathway. Furthermore, Horowitz (1965) has suggested that all of the enzymes involved in a given pathway evolved from one another. Waley (1969), although he

concedes that this is a plausible scheme for the evolution of pathways, takes exception to 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 (Yčas 1974).

The third possibility has been considered by Waley (1969). Watts (1965) 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 (1969) has suggested that

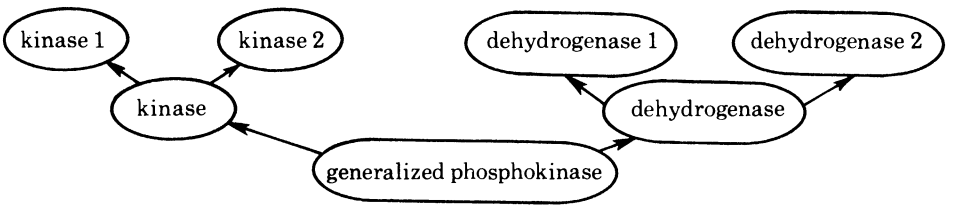

FIGURE 1. General scheme for the evolution of enzymes according to Waley (1969).

the corresponding enzymes may be related. Furthermore, he suggests that the general phosphokinase, postulated by Haldane (1965) 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 (1974) 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 non-haem iron and haem proteins as well as flavin and NAD-dependent enzymes might have a common origin.

All of the above suggestions are consistent with the view that the present-day highly specific enzymes have probably evolved from ancestral forms of much broader specificity by gene duplication and subsequent mutation. Fortunately the structure determination of polypeptide folds can, to some considerable extent, support and provide detail to the above generalizations. However, before the subject of structure can be pursued further, it is necessary to consider aspects of polypeptide organization.

DOMAINS AS EVOLUTIONARY BUILDING BLOCKS

The structural hierarchy implied by the terms primary, secondary, tertiary and quaternary structure is well known. Similarly, the term 'domain' is much used to describe structural units within a polypeptide, but has seldom been clearly defined. Indeed, an accurate definition is not entirely possible. However, some or all of the following properties can be ascribed to domains.

1. Homologous amino acid sequences can be found either within the polypeptide or in a different molecule.
2. Similar structures (of at least two consecutive secondary structural elements) are found in either the same polypeptide or in different molecules.
3. Domains within a polypeptide are spatially separated from each other, and can therefore often be enzymically cleaved from each other if suitable residues are available in the 'hinge' region.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

4. A special function can be associated with specific domain structures, such as nucleotide or polysaccharide binding.

5. The active centre of molecules lies at the interface between domains.

Not all these properties need be applicable simultaneously. For instance, chymotrypsin has two domains of similar structure, but no individual function can be attributed to any one of the domains, thus violating condition 4. There have been attempts to define domains by using only the third criterion above (Wetlaufer 1973; Wetlaufer *et al.* 1976; Crippen 1978). Others

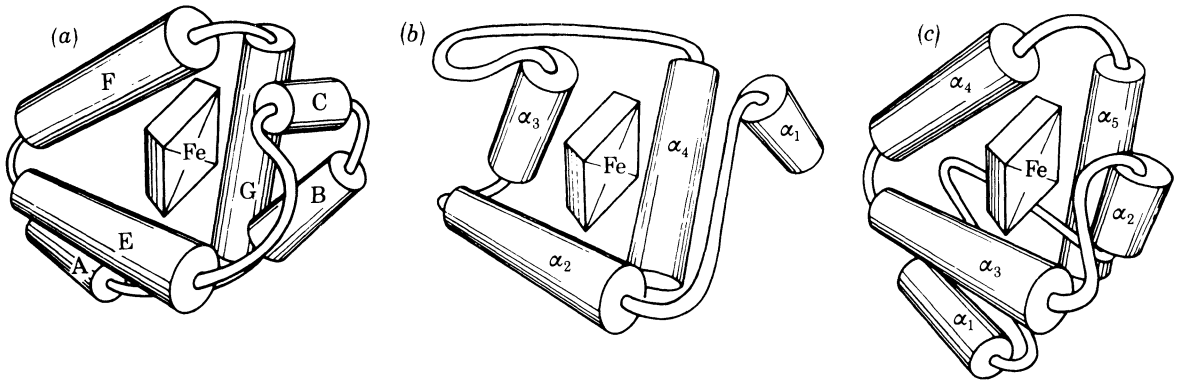

FIGURE 2. Diagrammatic representation of the similar polypeptide chain topology in (*a*) haemoglobin β chain, (*b*) cytochrome c_{551} and (*c*) cytochrome b_5. The common structure contains most of the haem-binding environment and corresponds to the central exon of the globins. From Argos & Rossmann (1979).

(Levitt & Chothia 1976) have assigned domains on a rather intuitive basis. Although no unique answer will be obtained by applying the above qualitative criteria, they nevertheless will attain the intuitively obvious.

Blake (1979) has suggested that domain integrity is preserved by their coding within exons at least in eukaryotes. This is supported, for instance, by the common helical folding pattern observed in globin, cytochrome c_{551} and cytochrome b_5 (figure 2), which has a haem-binding function (Craik *et al.* 1980) and corresponds to the middle exon of the β globin chain (Argos & Rossmann 1979). Since domains might thus be the result of independent genetic development, they may also represent the basis for protein folding. Furthermore, gene fusion of a variety of exons, corresponding to domains with simple functions, will then give rise to a sophisticated enzyme (Rossmann & Liljas 1974).

DOMAINS IN GLYCOLYTIC ENZYMES

Attempts at taxonomic classifications of proteins are still primitive (cf. Levitt & Chothia 1976; Sternberg & Thornton 1977; Richardson 1977). Nevertheless they all depend on an intuitive recognition of domains. The variety of domains expressed by glycolytic proteins is extremely limited (figure 3). Indeed, all known structures of glycolytic enzymes are of the 'β-α-β' taxonomic class described by Levitt & Chothia. Furthermore, they all contain domains reminiscent of the flat, six-stranded, parallel, β-pleated sheet, NAD-binding domain in dehydrogenases (figure 4) or the eight-stranded 'TIM barrel'.

An initial inspection of figure 3 might provoke the notion that glycolysis has evolved from a

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

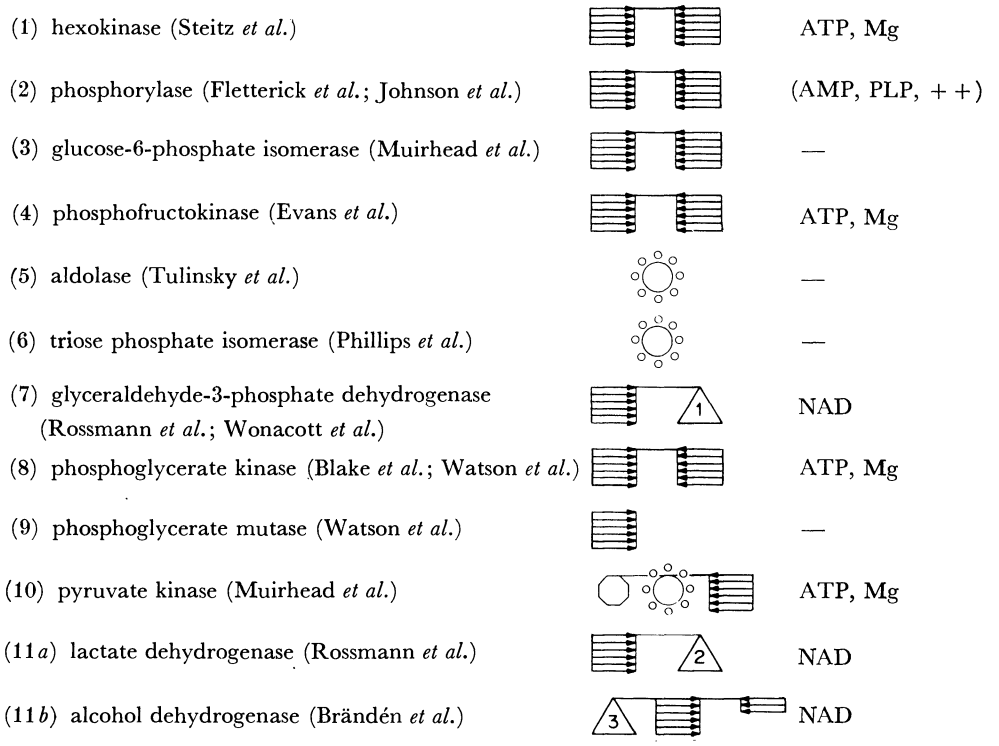

FIGURE 3. Gross simplification of variety of domains in known glycolytic enzyme structures. Shown also are the principal investigators of each enzyme. Primary references to their works have not been included.

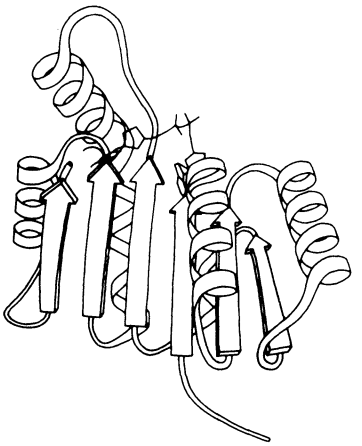

FIGURE 4. Schematic drawing of the NAD-binding domain in LDH.

general glycolytic enzyme based on the very limited domains. However, this concept breaks down when it is recognized that numerous other enzymes, not involved in glycolysis, also have similar domains. For instance rhodanese (Ploegman *et al.* 1978) also contains two domains arranged such that the carboxy ends of the β-sheets tend to face each other. Perhaps even more significantly, malate dehydrogenase, an enzyme of the citric acid cycle, is closely similar to lactate dehydrogenase (LDH).

Nevertheless, some crude generalizations might be made.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

1. There is a tendency for those enzymes that require nucleotides as cofactors to possess at least one flat, parallel, β-pleated sheet domain. The nucleotide invariably binds to the carboxy end of this sheet.

2. There is a tendency for those enzymes that require no nucleotide cofactors to have 'eight-stranded, parallel barrels'. The phosphorylated substrate invariably binds to the carboxy end of this sheet structure.

3. There is a tendency for kinases to contain two flat, parallel, β-pleated sheet 'domains' with their carboxy ends facing each other. The substrate tends to bind between the domains. Steitz's results (Anderson *et al.* 1979; Pickover *et al.* 1979) on the functional requirement of lobe movements in kinases are clearly related to these structural properties.

4. Dehydrogenases invariably contain one flat, parallel, β-pleated sheet domain and one catalytic domain. The latter's taxonomic classification invariably belongs to the 'α + β' Levitt & Chothia class.

These observations suggest that enzyme evolution proceeded by independent development of groups of enzyme classes, interrelated by the need to bind a common cofactor or substrate. It therefore makes little sense to consider the evolution of the glycolytic pathway on its own. Rather, its development must be closely related to other essential functions in early biological developments. As domains evolved in their functions, so enzymes in different metabolic pathways were derived by the chance fusion of suitable genes of selected domains.

Possible origin of the basic domains found in glycolytic enzymes

Richardson (1976) and Sternberg & Thornton (1976) have both shown that the right-handed crossover structure (figure 5) is a frequently recurring theme in β-α-β folds. They both argue in favour of this representing a stable super-secondary fold (Rao & Rossmann 1973). The importance of β-α-β structures for nucleotide binding was pointed out by Rossmann *et al.* (1974) and Ohlsson *et al.* (1974), but it was Hol *et al.* (1978) who suggested that this particular structure may have properties useful for the binding of phosphates.

The recent publication by Shoham & Steitz (1980) makes this point very strongly. They describe the binding of ATP to hexokinase in the catalytic domain. Their observations agree well with the prediction of Rossmann & Argos (1977) based on a comparison of the hexokinase structure with the NAD-binding domain of LDH, particularly in as far as the P_α and P_β positions are concerned (table 1).

Two consecutive one-crossover structures are by far the most common repeat within parallel sheets (Richardson 1977). This has sometimes been referred to as the mononucleotide binding fold (Rao & Rossmann 1973) as such folds bind AMP and NMN in the NAD-binding domain of LDH (figure 6). Two mononucleotide folds together can form either a flat, six-stranded, parallel, β-pleated sheet domain or an eight-stranded, parallel β-barrel (figure 7). Thus the two principal types of domain found in glycolytic enzymes may represent different combinations of similar entities.

Convergence and divergence

The foregoing discussion has tacitly assumed a process of divergent evolution from one or a few ancestral enzymes used for glycolysis and other essential metabolic processes in the most

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

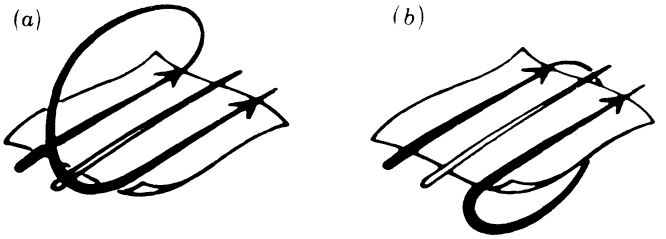

FIGURE 5. (*a*) A right-handed ±2x crossover connection; (*b*) a left-handed ±2x crossover connection. Direction is not indicated for the skipped strand, since it may be either parallel or anti-parallel to the others. From Richardson (1976).

TABLE 1. DIFFERENCES IN THE OBSERVED AND PREDICTED VALUES IN THE BINDING OF ATP TO THE CATALYTIC DOMAIN OF HEXOKINASE

	observed values†			predicted values‡			difference
	x	*y*	*z*	*x*	*y*	*z*	Å
adenine	45.8	−12.2	−4.4	56.6	−7.4	−8.8	12.6
ribose	47.6	−16.0	−3.3	55.6	−12.3	−7.5	9.8
P_α	52.2	−14.6	−4.4	51.7	−14.6	−5.9	1.6
P_β	54.2	−16.5	−3.1	51.5	−15.9	−3.9	2.9

† From Shoham & Steitz (1980).
‡ From Rossmann & Argos (1977).

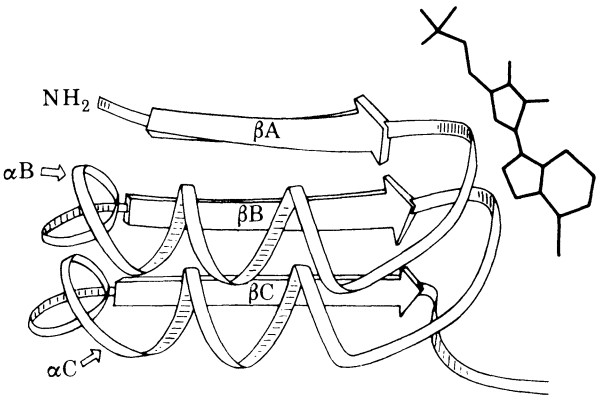

FIGURE 6. A possible primordial 'mononucleotide binding fold'. From Rossmann *et al.* (1977).

primitive prokaryotes. The possibility of convergence to functionally useful and stable folds should also be considered. Obviously the formation of α-helices occurs because they are an especially stable fold. The occurrence of α-helices in numerous protein structures does not imply divergence from some ancestral α-helical protein. The presence of eight successive helices arranged in space as in globin, however, is a sufficiently complex and rare structure that its occurrence in more than one structure immediately suggests 'development from a common genetic precursor' (Perutz *et al.* 1960). How, then, is it possible to differentiate between convergence and divergence?

Unless there is a clear historical record, all such differentiations must depend on probabilities. Their calculations are based on counting similar and dissimilar characters. The greater the

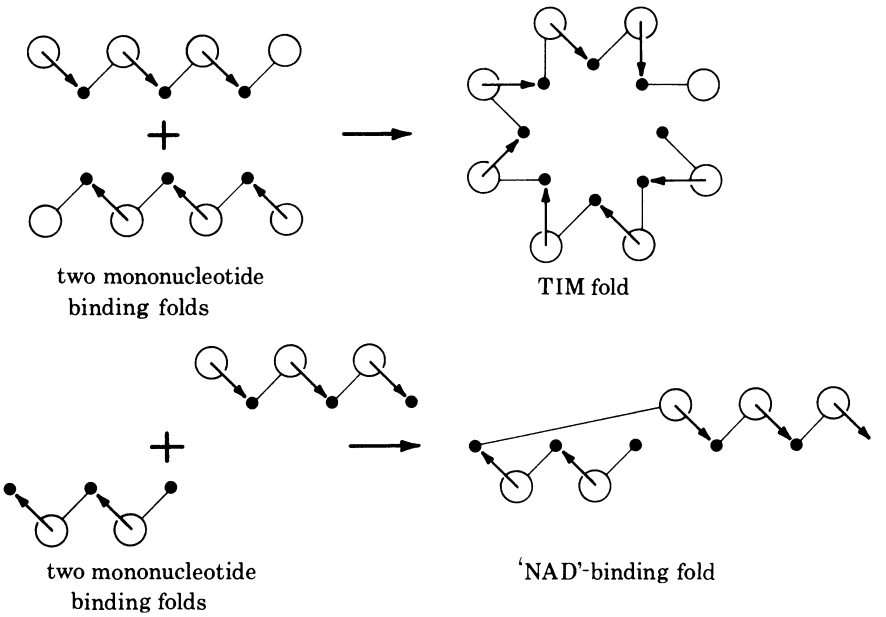

FIGURE 7. The NAD-binding domain and the triose phosphate isomerase subunit each contain approximately two 'mononucleotide binding folds'.

TABLE 2. CRITERIA FOR ASSESSING STRUCTURAL EQUIVALENCES

(The first three examples represent well documented standards with respect to which other comparisons can be calibrated.)

	number of equivalences	percentage of equivalences molecule 1	molecule 2	r.m.s. distance Å	m.b.c.c.
Hb-α–Hb-β	139	83	91	1.6	0.70
LDH–GAPDH NAD domains	83	58	56	—	1.24
T4–HEW lysozyme	64	39	50	4.1	1.53
HK1–HK2	78	35		—	—
PGK1–PKG2	62	34		—	—
LDH–PGK1	50	30		—	—
LDH–PGK2	63	31		—	—
LDH–HK1	61	40		—	—
LDH–HK2	95	55		—	—

proportion of similar characters within the total organism or molecule, the greater the likelihood of a divergent evolutionary relation. Such considerations were used by Schulz & Schirmer (1974) and Sternberg & Thornton (1976) to compute the probability of convergence of various nucleotide binding structures. A more comprehensive approach has been taken by Rossmann & Argos (1977), who compare the spatial arrangement of every amino acid between pairs of proteins. They derive a variety of criteria to estimate molecular similarity dependent in part on the number of topologically equivalent residues, the number of non-alignable residues and, hence, the frequency and size of insertions and deletions. A few of these estimates are shown in table 2. Standards of similarity are given in terms of the comparison between the α and β haemoglobin chains, lysozyme from phage and from hen egg white, and the NAD domains of LDH and glyceraldehyde-3-phosphate dehydrogenase (GAPDH). These can be contrasted

with the comparison between the two 'flat, β-pleated sheet' domains in hexokinase (HK1, HK2), between such domains in phosphoglycerate kinase (PGK1, PGK2) and between each of these domains and the NAD domain in LDH. It is apparent that the similarity of structure between flat, parallel, β-pleated sheet domains in at least some of the glycolytic enzymes is of the same order as that between the two different lysozymes. Similar analyses between triose phosphate isomerase and pyruvate kinase have been performed by Levine *et al.* (1978). A detailed analysis of all the similarities and dissimilarities suggested by figure 3 has not yet been undertaken.

TABLE 3. MINIMUM BASE CHANGE PER CODON (m.b.c.c.) FOR STRUCTURALLY EQUIVALENT RESIDUES

(Those residues that are most accurately superimposed are also those residues with greatest amino acid similarity.)

C_α–C_α separation ...	0–1 Å	1–2 Å	2–3 Å	3 Å or more
LDH–GAPDH				
m.b.c.c.	1.26	1.20	1.46	1.31
number of residues	23	31	13	16
haemoglobin β–cytochrome b_5				
m.b.c.c.	1.21	—	1.25	1.75
number of residues	14	—	20	12

Similarity of fold among glycolytic enzymes is far greater than homology of amino acid sequence. This can be interpreted as a loss of memory of the original sequence, while retaining approximate fold, during divergent evolution, or a lack of similar sequence requirements for convergent evolution. Whenever some semblance of amino acid similarity can be recognized. then these can act as characters to determine similarity. It is instructive that in some cases (table 3) the minimum base change per codon (m.b.c.c.) is well below random (random is approximately 1.4 m.b.c.c.) for those residues that have the most accurate structural equivalence. Thus, structure can be used for alignment to identify polypeptide segments that might have sequence homology.

THE NAD-BINDING DOMAINS IN DEHYDROGENASES

The NAD-binding domains in dehydrogenases are structures sufficiently similar in topology, in function and in amino acid sequence that a divergent evolutionary history is moderately probable. These structures therefore represent an opportunity to study divergent relationships for enzymes of similar function but extending beyond the requirements of glycolysis alone. Figure 8 shows, for example, the results of amino acid sequence alignments for the secondary structural elements βA and αB. Shown also are the alignments of other β-α-β structures. It is immediately clear that the glycine in the first position of αB is universally conserved. It must be concluded that this is a necessity of folding and may, therefore, be the result of convergent evolution. However, the conservation of other amino acids among the NAD-binding domains is significantly below random and below that of other such structures (table 3). Four residues are explicitly conserved in the known NAD-binding domains, as a requirement either of folding or of function (table 4). The function of the Asp (e.g. LDH residue 53) that binds the O-2′

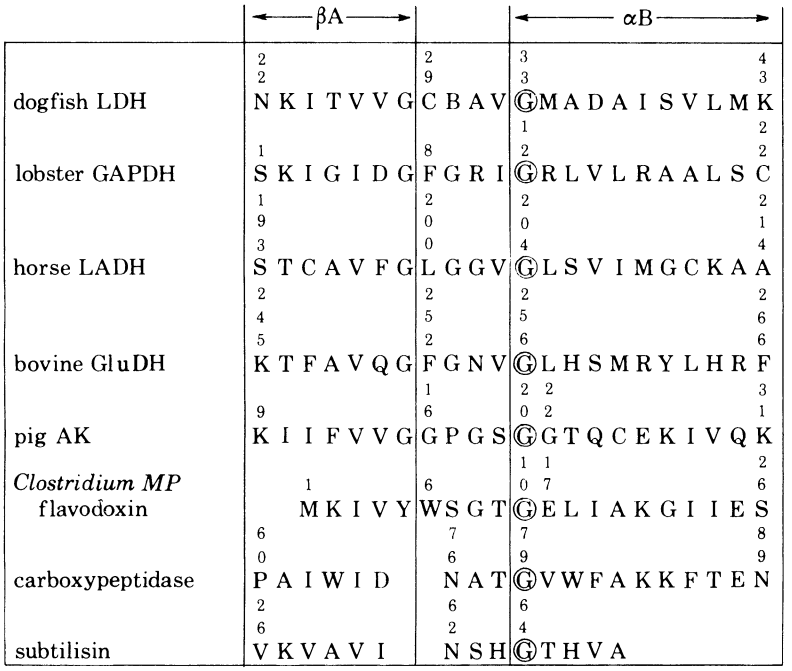

FIGURE 8. Amino acid sequence alignments for the secondary structural elements βA and αB in several β-α-β structures (for nomenclature of βA and αB see figure 4).

TABLE 4. INVARIANT RESIDUES IN THE NAD-BINDING DOMAINS OF LDH, GAPDH AND LIVER ALCOHOL DEHYDROGENASE (LADH)

amino acid	Gly	Gly	Asp	Gly
position	βA	αB	βB	βD
sequence no.				
In LDH	28	33	53	99
In GAPDH	7	12	32	97
In LADH	199	204	223	270
binds to	A-Rib	A-PO_4	A-Rib 2′-OH	N-Rib

of the adenine ribose in NAD is of particular interest. A residue in a structurally (but not topologically) equivalent position of dihydrofolate reductase is an arginine where the O-2′ is replaced by a phosphate, thus providing specificity for NADP (Matthews *et al.* 1978).

RECENT EVOLUTION

Comparison of amino acids rather than fold gives guidance to recent evolutionary events, say during the last 10^9 years, for the rather slow acceptance of mutational events in glycolytic enzymes. Three-dimensional structure is a useful measure only for events longer ago, where amino acid sequences show no significant correlation but where polypeptide topology can still be recognizably similar. On the other hand, there is little structural alteration during relatively short time periods, whereas amino acid differences can then provide a sensitive measure of evolutionary events.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

The formation of oligomers must depend on the surface shape and charges of its component monomers. Hence, quaternary structure relates to specific tertiary folds. The aggregates seen today must necessarily succeed the evolution of the component parts. An analysis of the amino acid conservation for tetrameric GAPDH is shown in table 5, where it will be seen that the highest degree of conservation occurs for the residues in the active centre. Residues in the subunit contact surfaces are less conserved than the active centre residues but more than the mean of the complete molecule. Similar results are found for tetrameric LDH (Eventoff *et al.* 1977).

TABLE 5. ANALYSIS OF THE CONSERVATION OF AMINO ACID RESIDUES IN TETRAMERIC GAPDH

domain		*	*P*	*Q*	*R*	D	A
first	number of conserved residues	17	0	3	5	9	71
	total number in contact	19	0	3	5	13	148
	percentage conserved	90	—	100	100	69	48
second	number of conserved residues	11	27	2	10	2	124
	total number in contact	12	33	5	12	5	186
	percentage conserved	92	82	40	83	40	67
both	number of conserved residues	28	27	5	15	11	195
	total number in contact	31	33	8	17	18	334
	percentage conserved	90	82	63	88	61	58

Symbols: *, involved in the active centre; *P*, in a *P*-axis generated contact between subunits; *Q*, in a *Q*-axis generated contact between subunits; *R*, in a *R*-axis generated contact between subunits; D, in a domain–domain contact within a subunit; A, amino acids involved.

The most rapid sequence changes occur as a result of selective pressure to changing environmental pressure on the organism. Extreme environments such as high salt concentration or unusually cold or hot temperatures may thus provide some information on the nature of such changes. Once again, work on some glycolytic enzymes, in particular LDH and GAPDH, has given us the most information yet available on such topics. An analysis of a number of such sequences (Argos *et al.* 1979) for thermally stable enzymes shows the preference of certain amino acids over others (figure 9). These can be related to various physical properties of the protein (table 6) such as stabilizing of helix structures, increasing internal hydrophobicity and improving internal packing organization of residues.

CONCLUSIONS

A study of the enzymes in the glycolytic pathway has provided information not only on the mechanistic constraints put on each enzyme by its structure, but also on the evolution of such enzymes. It is clear that there was neither a primordial general glycolytic enzyme, nor a series of primordial specific kinases, mutases, etc. Rather, evolution is dependent on the use of domains with simple functions from which are built the enzymes in many metabolic processes. Whether such domains are the result of convergence toward stable structures of simple function or divergence from some ancient basic domains must be further investigated as more data become available. The subsequent evolution of formed enzymes is controlled by requirements such as oligomeric aggregation and environmental conditions.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

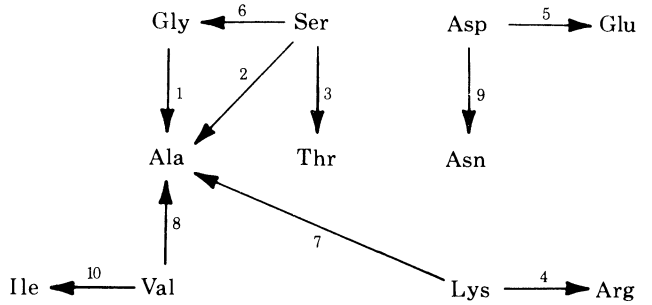

FIGURE 9. Direction of observed preferred exchanges. Arrows point from the mesophilic to the thermophilic protein. Numbers indicate the ranking of the significance for the given exchange. From Argos *et al.* (1979).

TABLE 6. BENEFIT OF CERTAIN AMINO ACID EXCHANGES TO VARIOUS PHYSICAL PROPERTIES IN THE STABILIZATION OF PROTEINS

(From Argos *et al.* (1979).)

rank	exchange cold → hot	helical region	sheet region	internal hydrophobicity	external polarity	internal packing
1	G → A	+ +		+ +		+ +
2	S → A	+ +		+		+
3	S → T		+			−
4	K → R	−	+	+ +	+ +	+
5	D → E	+				−
6	S → G	−			+	−
7	K → A		+	+ +	− −	+ +
8	V → A	+ +	−	− −	+	− −
9	D → N	−		+	− −	−
10	V → I		−	+		+

I am grateful to my assistant, Sharon Wilder, for her diligent help in the preparation of this manuscript, to the organizers of this Discussion Meeting for their invitation to present this paper, and to the National Institutes of Health (grant no. GM 10704) and the National Science Foundation (grant no. PCM78-16584) for financial support.

REFERENCES (Rossmann)

Anderson, C. M., Zucker, F. H. & Steitz, T. A. 1979 Space-filling models of kinase clefts and conformation changes. *Science, N.Y.* **204**, 375–380.

Argos, P. & Rossmann, M. G. 1979 Structural comparisons of heme binding proteins. *Biochemistry, Wash.* **18**, 4951–4960.

Argos, P., Rossmann, M. G., Grau, U. M., Zuber, H., Frank, G. & Tratschin, J. D. 1979 Thermal stability and protein structure. *Biochemistry, Wash.* **18**, 5698–5703.

Baltscheffsky, H. 1974 A new hypothesis for the evolution of biological electron transport. *Origins Life* **5**, 387–395.

Blake, C. C. F. 1979 Exons encode protein functional units. *Nature, Lond.* **277**, 598.

Craik, C. S., Buchman, S. R. & Beychok, S. 1980 Characterization of globin domains: heme binding to the central exon product. *Proc. natn. Acad. Sci. U.S.A.* **77**, 1384–1388.

Crippen, G. M. 1978 The tree structural organization of proteins. *J. molec. Biol.* **126**, 315–332.

Eventoff, W., Rossmann, M. G., Taylor, S. S., Torff, H. J., Meyer, H., Keil, W. & Kiltz, H. H. 1977 Structural adaptations of lactate dehydrogenase isozymes. *Proc. natn. Acad. Sci. U.S.A.* **74**, 2677–2681.

Haldane, J. B. S. 1965 Data needed for a blueprint of the first organism. In *Origins of pre-biological systems and of their molecular matrices* (ed. S. W. Fox), pp. 11–23. New York: Academic Press.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Hol, W. G. J., van Duijnen, P. T. & Berendsen, H. J. C. 1978 The α-helix dipole and the properties of proteins. *Nature, Lond.* **273**, 443–446.
Horowitz, N. H. 1945 The evolution of biochemical syntheses. *Proc. natn. Acad. Sci. U.S.A.* **31**, 153–156.
Horowitz, N. H. 1965 The evolution of biochemical syntheses – retrospect and prospect. In *Evolving genes and proteins* (ed. V. Bryson & H. J. Vogel), pp. 15–26. New York: Academic Press.
Levine, M., Muirhead, H., Stammers, D. K. & Stuart, D. I. 1978 Structure of pyruvate kinase and similarities with other enzymes: possible implications for protein taxonomy and evolution. *Nature, Lond.* **271**, 626–630.
Levitt, M. & Chothia, C. 1976 Structural patterns in globular proteins. *Nature, Lond.* **261**, 552–558.
Matthews, D. A., Alden, R. A., Bolin, J. T., Filman, D. J., Freer, S. T., Hamlin, R., Hol, W. G. J., Kisliuk, R. L., Pastore, E. J., Plante, L. T., Xuong, N. & Kraut, J. 1978 Dihydrofolate reductase from *Lactobacillus casei*. X-ray structure of the enzyme·methotrexate·NADPH complex. *J. biol. Chem.* **253**, 6946–6954.
Ohlsson, I., Nordström, B. & Brändén, C. I. 1974 Structural and functional similarities within the coenzyme binding domains of dehydrogenases. *J. molec. Biol.* **89**, 339–354.
Perutz, M. F., Rossmann, M. G., Cullis, A. F., Muirhead, H., Will, G. & North, A. C. T. 1960 Structure of haemoglobin. A three-dimensional Fourier synthesis at 5.5-Å. resolution, obtained by X-ray analysis. *Nature, Lond.* **185**, 416–422.
Pickover, C. A., McKay, D. B., Engelman, D. M. & Steitz, T. A. 1979 Substrate binding closes the cleft between the domains of yeast phosphoglycerate kinase. *J. biol. Chem.* **254**, 11323–11329.
Ploegman, J. H., Drent, G., Kalk, K. H. & Hol, W. G. J. 1978 Structure of bovine liver rhodanese. I. Structure determination at 2.5 Å resolution and a comparison of the conformation and sequence of its two domains. *J. molec. Biol.* **123**, 557–594.
Rao, S. T. & Rossmann, M. G. 1973 Comparison of super-secondary structures in proteins. *J. molec. Biol.* **76**, 241–256.
Richardson, J. S. 1976 Handedness of crossover connections in β sheets. *Proc. natn. Acad. Sci. U.S.A.* **73**, 2619–2623.
Richardson, J. S. 1977 β-Sheet topology and the relatedness of proteins. *Nature, Lond.* **268**, 495–500.
Rossmann, M. G. & Argos, P. 1977 The taxonomy of protein structure. *J. molec. Biol.* **109**, 99–129.
Rossmann, M. G., Garavito, R. M. & Eventoff, W. 1977 Conformational adaptations among dehydrogenases. In *Pyridine nucleotide-dependent dehydrogenases* (ed. H. Sund), pp. 3–30. Berlin: Walter de Gruyter.
Rossmann, M. G. & Liljas, A. 1974 Recognition of structural domains in globular proteins. *J. molec. Biol.* **85**, 177–181.
Rossmann, M. G., Moras, D. & Olsen, K. W. 1974 Chemical and biological evolution of a nucleotide-binding protein. *Nature, Lond.* **250**, 194–199.
Schulz, G. E. & Schirmer, R. H. 1974 Topological comparison of adenyl kinase with other proteins. *Nature, Lond.* **250**, 142–144.
Shoham, M. & Steitz, T. A. 1980 Crystallographic studies and model building of ATP at the active site of hexokinase. *J. molec. Biol.* **140**, 1–14.
Sternberg, M. J. E. & Thornton, J. M. 1976 On the conformation of proteins: the handedness of the β-strand–α-helix–β-strand unit. *J. molec. Biol.* **105**, 367–382.
Sternberg, M. J. E. & Thornton, J. M. 1977 On the conformation of proteins: the handedness of the connection between parallel β-strands. *J. molec. Biol.* **110**, 269–283.
Waley, S. G. 1969 Some aspects of the evolution of metabolic pathways. *Comp. Biochem. Physiol.* **30**, 1–13.
Watts, D. C. 1965 Evolutionary implications of enzyme structure and function. In *Studies in comparative biochemistry* (ed. K. A. Munday), pp. 162–165. London: Pergamon Press.
Wetlaufer, D. B. 1973 Nucleation, rapid folding, and globular intrachain regions in proteins. *Proc. natn. Acad. Sci. U.S.A.* **70**, 697–701.
Wetlaufer, D. B., Rose, G. D. & Taaffe, L. 1976 Orientation of structural segments in globular proteins. *Biochemistry, Wash.* **15**, 5154–5157.
Woese, C. R. 1965 On the evolution of the genetic code. *Proc. natn. Acad. Sci. U.S.A.* **54**, 1546–1550.
Woese, C. R. 1967 *The genetic code: the molecular basis for genetic expression*. New York: Harper & Row.
Yčas, M. 1974 On earlier states of the biochemical system. *J. theor. Biol.* **44**, 145–160.

Discussion

A. D. B. Malcolm (*Department of Biochemistry, St Mary's Hospital Medical School, London, U.K.*). I do not think that studying the DNA from a thermophilic bacterium will tell us very much. The change of a few AT base pairs into GC will not have any significant effect on the stability of the DNA. I think it is much more likely that an effect might be observed on the stability of the mRNA. In this connection it is interesting to recall some calculations carried out by Andrew Ball (*J. theor. Biol.* **36**, 313 (1972)). He showed that codon usage in the phages MS2 and R17

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

was consistent with the theory that maximizing secondary structure in the RNA was an important selective pressure. Such ideas have been applied to the RNAs for globin and cytochrome *c* (*Science* **175**, 1264 (1972)).

M. G. Rossmann. Argos *et al.* (1979) have attempted to demonstrate that the observed amino acid changes for a given enzyme taken from mesophiles and thermophiles can be directly related to physical properties that affect the stability of a protein molecule. Indeed such a relation must be the case, whether or not the nature of the changes can be understood with respect to the molecular properties! I would, however, agree that the evolutionary selection of changes that bring about protein stability might simultaneously bring about greater stability of the corresponding mRNA. It would be interesting to determine whether those changes which have been determined as being important in improving thermal stability would also increase the base-pairing possibilities of the mRNA. For this it will be necessary to await the DNA sequences and their comparisons between mesophiles and thermophiles. Clearly such simultaneous requirements for mRNA and protein stability would greatly reduce the number of possible acceptable mutations.

A. D. B. Malcolm. Alpha and beta globin have been used by Professor Rossmann as examples of closely related proteins and it is generally assumed that they have evolved from a common ancestor. The genes for these two proteins occur on different chromosomes; I should like to ask Dr Bodmer how this fits with his ideas about the evolution of gene clusters.

W. F. Bodmer, F.R.S. It has for some time been assumed that the haemoglobin α and β clusters are derived from a common precursor DNA sequence, which presumably was duplicated. Following this there was either duplication of the whole chromosome or translocation or transposition of a part of the duplicated cluster to another chromosome, followed by divergence to give the two sets of gene products. Recently, Jeffreys *et al.* (*Cell* **21**, 555–564 (1980)) have shown that in *Xenopus laevis*, the major adult α and β globin genes are closely linked in tandem, suggesting in fact that the divergence of these products may have arisen within a cluster and their separation into two linkage groups followed later. It is, in general, a feature of gene clusters that within them products tend to be mutually exclusively expressed, in the sense that a cell at any given time is likely to be expressing only one member of the gene set within the cluster. Between the two haemoglobin clusters, for example, there is by contrast always a need for one chain from each cluster. The arrangement by which the two clusters are unlinked may well have evolved because of some selective advantage for mutually related control of gene expression.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES